Journal of Machine Learning Research 22 (2021) 1-89 Submitted 11/18; Revised 8/20; Published 2/21

Mixing Time of Metropolis-Hastings for Bayesian Community Detection

Bumeng Zhuo BUMENGZHUO@UCHICAGO.EDU
Chao Gao CHAOGAO@GALTON.UCHICAGO.EDU
Department of Statistics
University of Chicago
Chicago, IL 60637, USA

Editor: XuanLong Nguyen

Abstract

We study the computational complexity of a Metropolis-Hastings algorithm for Bayesian community detection. We first establish a posterior strong consistency result for a natural prior distribution on stochastic block models under the optimal signal-to-noise ratio condition in the literature. We then give a set of conditions that guarantee rapidly mixing of a simple Metropolis-Hastings algorithm. The mixing time analysis is based on a careful study of posterior ratios and a canonical path argument to control the spectral gap of the Markov chain.

Keywords: Stochastic block model, Markov chain Monte Carlo, Bayesian computation, Posterior strong consistency, Algorithmic complexity

1. Introduction

Markov Chain Monte Carlo (MCMC) is a popular sampling technique, in which the equilibrium distribution of the Markov chain matches the target distribution. Most attentions to date have been focused on Bayesian applications in order to sample from posterior distributions. Despite its popularity in Bayesian statistics and many other areas, its theoretical properties are not well understood, not to mention the limited theory for computational efficiency of MCMC algorithms, where the pivotal interest lies in the analysis of mixing time. The mixing time of a Markov chain is the number of iterations required to get close enough to the target distribution, in the sense that the total variation distance is bounded above by some small constant ε. We call the Markov chain rapidly mixing (resp. slowly mixing) if the mixing time grows at most polynomially (resp. exponentially) with respect to the sample size of the problem. One central research interest is to determine whether a designed Markov chain is rapidly mixing or slowly mixing. A series of studies have made efforts to design efficient Markov chains Thawornwattana et al. (2018); Wu et al. (2004); Møller et al. (2006); Hutter et al. (2014); Tu and Zhu (2002); Wu et al. (2004) without providing theoretical guarantees. Over the past fifteen years, a surge of research has led to breakthroughs in the understandings of geometric ergodicity of Markov chains, as well as exploration of the intimate relationship between the spectral gap of a Markov chain and its mixing time Roberts et al. (1997); Meyn et al. (1994); Diaconis and Stroock (1991); Sinclair (1992). Based on graphical theory, several elegant techniques were further developed to lower bound spectral gap and characterize the mixing property Bubley and Dyer (1997); Guruswami (2016); Diaconis and Stroock (1991); Sinclair (1992); Meyn and Tweedie (2012); Randall (2006); Levin and Peres (2017).

Traditional literature on the theory of MCMC is mostly about the study of Markov chains and its convergence to the stationary distributions, but the implications of such theory on the statistical and algorithmic properties in specific high-dimensional problems are not clear. It is still a very challenging problem to understand the computational complexity of MCMC in Bayesian high-dimensional estimation. There exist several obstacles associated with the mixing time analysis of MCMC in Bayesian high-dimensional models. First of all, an MCMC algorithm often needs to explore a state space of exponential complexity in a high-dimensional estimation setting. Whether MCMC algorithms can still rapidly converge to the stationary distribution with complexity depending on both the sample size and the dimension polynomially remains an unsettled theoretical problem. Secondly, even if the algorithm converges, whether the stationary distribution can recover the underlying knowledge of data in a high-dimensional setting is also not obvious.

To the best of our knowledge, the only paper that addresses both the computational and statistical challenges of MCMC algorithms is Yang et al. (2016). The paper Yang et al. (2016) studies the mixing property of MCMC in the setting of sparse linear regression, and chooses a Metropolis-Hastings algorithm as the representative of a broader family. Two important results are obtained in the paper. First, it proves that the Metropolis-Hastings converges to its stationary distribution in polynomial time. Second, the stationary distribution selects the correct active set of variables with high probability. Both conclusions are non-trivial, because the Metropolis-Hastings algorithm for variable selection is combinatorial in nature. In order to bound the mixing time, the paper Yang et al. (2016) adopted a canonical path technique developed in Sinclair (1992); Diaconis and Stroock (1991). The method of canonical path heavily relies on the graph structure, and the construction of canonical paths is very hard especially for a high-dimensional problem. The paper Yang et al. (2016) took advantage of the sparse linear regression structure, and designed a set of canonical paths between all pairs of points such that no edge is "overloaded" (congested), which yielded sharp quantitative bounds for mixing time of the Metropolis-Hastings algorithm.

The work of Yang et al. (2016) naturally leads to the following question. Is the polynomial-time mixing property of the Metropolis-Hastings algorithm due to the spectral structure of sparse linear regression, or is the good computational property of the Metropolis-Hastings algorithm a more general phenomenon for other Bayesian high-dimensional problems as well?

In this paper, we show that the Metropolis-Hastings algorithm also has good theoretical property for community detection in Stochastic Block Model (SBM). We prove that the Metropolis-Hastings algorithm converges to its stationary distribution in polynomial time. Moreover, we show that the stationary distribution achieves strong consistency of community detection with high probability. These results, together with the work Yang et al. (2016), show that the Metropolis-Hastings algorithm has good theoretical properties in two very different high-dimensional problems, variable selection and clustering, which suggests that this can be a more general phenomenon. The results we have obtained in this paper are benchmarked by the state-of-the-art theoretical work in the literature of community detection. To be specific, both the statistical and algorithmic properties of the Metropolis-Hastings algorithm in this paper are established under the optimal signal-to-noise ratio threshold of community detection (Abbe, 2016; Abbe et al., 2014; Zhang et al., 2016).

To close this section, we discuss several new challenges of analyzing the Metropolis-Hastings algorithm for community detection. First, different from the variable selection problem in a regression setting (Yang et al., 2016), the non-identifiability of clustering labels of our problem leads to additional technical difficulties in analyzing the Markov chain. We overcome this difficulty by relating the Markov chain in the space of clustering labels to a Markov chain in the space of clustering structure according to Figure 1.

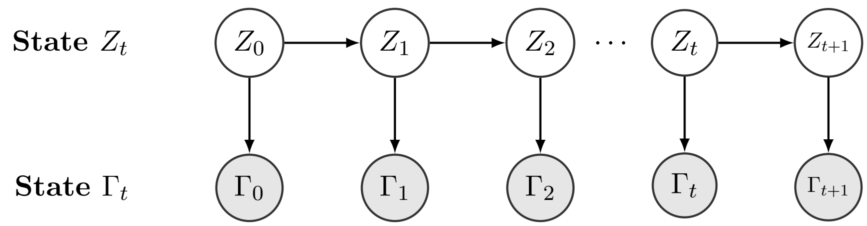


Figure 1: Updating process of $\Gamma(Z_t)$.

Here, $\{Z_t\}_{t\geq 0}$ is the sequence of labels obtained from the algorithm, and $\{\Gamma_t\}_{t\geq 0}$ is the Markov chain of clustering structures we are interested in. This additional treatment is important, and sets ourselves apart from the previous work. Then, on top of canonical path technique, we construct a super martingale to prove that the number of mistakes of labels is upper bounded within any polynomial running time, which forms a good region of clustering space for the construction of canonical paths. Last but not least, in order to prove that the statistical and algorithmic properties of Metropolis-Hastings hold simultaneously under the optimal signal-to-noise ratio condition, we introduce a scaled posterior distribution with a suitable temperature parameter to carefully upper bound any posterior ratios along the canonical path. The analysis of this paper is based on a frequentist point of view that the data are generated from an underlying true model.

Organization. The rest of the paper is organized as follows. Section 2 formally sets up the community detection problem and presents the posterior strong consistency property. Section 3 introduces a Metropolis-Hastings algorithm and provides an explicit mixing time bound, followed by some numerical results demonstrating its competitive performance on simulated data sets in Section 4. Section 5 is devoted to the proofs of the technical results of the paper.

Notations. We close this section by introducing some notations. For an integer d, we use $[d]$ to denote $\{1, 2, \ldots, d\}$. For a set S, we write $\mathbb{I}\{S\}$ as its indicator function and $|S|$ as its cardinality. For a vector $v \in \mathbb{R}^d$, its norms are defined by $\|v\|_1 = \sum_{i=1}^d |v_i|$, $\|v\|^2 = \sum_{i=1}^d v_i^2$, and $\|v\|_\infty = \max_{1\leq i\leq d} |v_i|$. The Hamming error of two binary vectors $v_1, v_2 \in \{0,1\}^d$ is defined by $H(v_1, v_2) = \sum_{i=1}^d \mathbb{I}\{v_1(i) \neq v_2(i)\}$. For a matrix $A \in \mathbb{R}^{K\times K}$, its norms are defined by $\|A\|_\infty = \max_{i,j\in[K]} |A_{ij}|$, and $\|A\|_1 = \sum_{i,j\in[K]} |A_{ij}|$. The notation $\mathbb{P}$ and $\mathbb{E}$ are generic probability and expectation operators whose distribution is determined from the context. We use $o(1)$ to denote any positive sequence tending to 0. Throughout the paper, unless otherwise noticed, we use C, c and their variants to denote absolute constants, and the values may vary from line to line. For any two distributions P and Q, the total variation distance is defined by $\|P-Q\|_{\mathsf{TV}} = \frac{1}{2}\int |dP - dQ|$, and the KL divergence is defined by $D\left(P\|Q\right) = \int dP \log \frac{dP}{dQ}$. For simplicity, we write $D\left(p\|q\right)$ to denote $D\left(\text{Bernoulli}(p)\|\text{Bernoulli}(q)\right)$ for $p, q \in [0,1]$. For any two numbers a and b, we use $a \wedge b$ and $a \vee b$ to denote $\min\{a,b\}$ and $\max\{a,b\}$ respectively.

2. Bayesian Community Detection

Networks have arisen in various areas of applications and have attracted a surge of research interests in fields such as physics, computer science, social sciences, biology, and statistics Goldenberg et al. (2010); Newman (2010); Wasserman and Faust (1994); Fortunato (2010); Chen and Yuan (2006). In the realm of network analysis, community detection has emerged as a fundamental task that provides insights of the underlying structure. Great advances have been made on community detection recently with a remarkable diversity of models and algorithms developed in different areas Girvan and Newman (2002); Newman and Leicht (2007); Handcock et al. (2007). Among various statistical models, the stochastic block model (SBM), first proposed in Holland et al. (1983), is one of the most prominent generative model that depicts the network topologies and incorporates the community structure. It is arguably the simplest model of a graph with communities and has been widely applied in social, biological and communication networks. Much effort has been devoted to SBM-based methods and their asymptotic properties have also been studied recently Bickel and Chen (2009); Celisse et al. (2012); Bickel et al. (2013). A Bayesian point of view of community detection was first suggested in Snijders and Nowicki (1997) with only two communities, motivated by the computational advantage of Gibbs sampling. The approach was further extended in Nowicki and Snijders (2001); Hofman and Wiggins (2008) to incorporate adjusted priors on community proportions as well as edge probabilities and allow for the cases of more than two communities. There has been little theoretical analysis of Bayesian community detection until very recently, when the consistency results of posterior distribution were obtained by van der Pas et al. (2017); Geng et al. (2019). However, they required the expected degree of a node to be at least of order $\log^2 n$ to ensure the strong consistency, which is a suboptimal condition for strong consistency Abbe (2016); Abbe et al. (2014); Zhang et al. (2016).

In this section, we give a precise formulation of the community detection problem and introduce a Bayesian approach. Then, we present the posterior strong consistency result.

2.1 Problem formulation

Consider an unweighted and undirected network with n nodes and K communities. The adjacency matrix is denoted by $A \in \{0,1\}^{n\times n}$, $A = A^T$, and $A_{ii} = 0$, for all $i \in [n]$. The edges are independently generated as Bernoulli variable with $\mathbb{E}A_{ij} = P_{ij}$, for all $i < j$. Here, P_{ij} denotes the connectivity probability for nodes i and j, and depends on the communities that the two nodes are assigned to. In this paper, we focus on a homogeneous SBM and assume $P_{ij} = p$ if two nodes are from the same community and $P_{ij} = q$ otherwise. We call p (resp. q) as the within-community (resp. between-community) connectivity probability and assume $p > q$ to satisfy the "assortative" property. Extensions to heterogenous SBMs are straightforward, but will not be considered in the paper for the sake of the presentation.

Let $Z \in [K]^n$ denote a label assignment vector, where Z_i is the community label for the ith node. Let $B \in [0,1]^{K \times K}$ be a symmetric connectivity probability matrix and thus $P_{ij} = B_{Z_i Z_j}$ with $B_{aa} = p$ for all $a \in [K]$, and $B_{ab} = q$ for all $a \neq b$. According to the description of the model, the likelihood formula can be written as

$$p(A|Z,B) = \prod_{i<j} B_{Z_i Z_j}^{A_{ij}} \left(1 - B_{Z_i Z_j}\right)^{1-A_{ij}}. \tag{1}$$

We use Z^* to denote the underlying true label assignment vector, and further assume that

$$\frac{n}{\beta K} \leq \sum_{i=1}^{n} \mathbb{I}\left\{Z_i^* = k\right\} \leq \frac{\beta n}{K}, \text{ for all } k \in [K], \tag{2}$$

where $\beta \geq 1$ is an absolute constant. It indicates that the all community sizes are of the same order. When $\beta = 1 + o(1)$, all communities have almost the same sizes. Furthermore, we assume K is a known constant, $p, q \to 0$ and $p \asymp q$ throughout the paper. To conclude, this paper focuses on a sparse homogeneous SBM with a finite number of communities.

Note that community detection is a clustering problem, and thus any label assignment gives an equivalent result after a label permutation. To be specific, let

$$\Gamma(Z) = \{\sigma \circ Z : \sigma \in \mathcal{P}_K\}, \tag{3}$$

where $\mathcal{P}_K$ stands for the set of all permutations on $[K]$, and then any $Z' \in \Gamma(Z)$ leads to an equivalent clustering structure. Hence, with the identifiability issue, our ultimate goal is to reconstruct the community structure, or equivalently, to recover the community label assignment Z^* up to a label permutation.

2.2 A Bayesian model for community detection

In addition to the likelihood formula of the adjacency matrix A given in (1), we put a uniform prior on Z over a set S_α, where S_α is the set of all feasible label assignments depending on a hyperparameter α. The connectivity probabilities B_{ab} for $1 \leq a \leq b \leq K$ receive independent Beta priors. More precisely, the Bayesian model is given by

stochastic block model:	$p(A\|Z,B) = \prod_{i<j} B_{Z_i Z_j}^{A_{ij}} \left(1 - B_{Z_i Z_j}\right)^{1-A_{ij}},$
label assignment prior:	$\pi(Z) \propto \mathbb{I}\left\{Z \in S_\alpha\right\},$
connectivity probability prior:	$B_{ab} \overset{\text{iid}}{\sim} \text{Beta}(\kappa_1, \kappa_2), \quad 1 \leq a \leq b \leq K,$

where $\kappa_1, \kappa_2 > 0$ measure the prior information of the connectivity probabilities and have negligible effects on the results when the sample size is large enough. This is the same set-up in van der Pas et al. (2017), except that we introduce a uniform prior over set S_α.

The key set S_α is defined by

$$S_\alpha = \left\{ Z \in [n]^K : \sum_{i=1}^n \mathbb{I}\{Z_i = k\} \in \left[\frac{n}{\alpha K}, \frac{\alpha n}{K}\right], \text{ for all } k \in [K] \right\}, \tag{4}$$

where the hyperparameter α controls the size of the feasible set S_α, which rules out those models whose group sizes differ too much. Here, α can be any sufficiently large constant, and we only require $\alpha > \beta$ so that $Z^* \in S_\alpha$. As will be clarified in Section 4, this additional constraint seems to be necessary for the rapidly mixing according to our practical experiments.

The induced posterior distribution can be expressed as

$$\begin{aligned} \Pi(Z|A) &\propto \int_{[0,1]^{K(K+1)/2}} \prod_{a \le b} B_{ab}^{O_{ab}(Z)} (1 - B_{ab})^{n_{ab}(Z) - O_{ab}(Z)} d\Pi(B) \\ &\propto \prod_{a \le b} \text{Beta}(O_{ab}(Z) + \kappa_1, n_{ab}(Z) - O_{ab}(Z) + \kappa_2), \qquad \text{for } Z \in S_\alpha, \end{aligned}$$

and it follows that for $Z \in S_\alpha$,

$$\log \Pi(Z|A) = \sum_{a \le b} \log \text{Beta}(O_{ab}(Z) + \kappa_1, n_{ab}(Z) - O_{ab}(Z) + \kappa_2) + Const, \tag{5}$$

where $n_{ab}(Z) = n_a(Z) n_b(Z)$, $n_{aa}(Z) = n_a(Z)(n_a(Z) - 1)/2$ for all $a \ne b \in [K]$. We use $n_a(Z)$ to denote the size of community a, i.e., $n_a(Z) = |\{i : Z_i = a\}|$. We use $O_{ab}(Z)$ to denote the number of connected edges between communities a and b, which takes the formula $O_{ab}(Z) = \sum_{i,j} A_{ij} \mathbb{I}\{Z_i = a, Z_j = b\}$ and $O_{aa}(Z) = \sum_{i<j} A_{ij} \mathbb{I}\{Z_i = Z_j = a\}$ for all $a \ne b \in [K]$. Note that the posterior distribution is permutation symmetric, i.e.,

$$\Pi(Z|A) = \Pi(Z'|A), \quad \text{for all } Z' \in \Gamma(Z). \tag{6}$$

2.3 Posterior strong consistency

Before stating the theoretical properties of the proposed Bayesian model, we introduce some useful quantities. The first quantity I plays a crucial part in the minimax theory Zhang et al. (2016),

$$I = -2 \log(\sqrt{pq} + \sqrt{(1-p)(1-q)}),$$

which is the Rényi divergence of order $1/2$ between Bernoulli(p) and Bernoulli(q). It can be shown that when $p, q \to 0$,

$$I = (1 + o(1))(\sqrt{p} - \sqrt{q})^2.$$

Then, we introduce an effective sample size to simplify the presentation of the results. As mentioned in Zhang et al. (2016), the minimax misclassification error rate is determined by that of classifying two communities of the smallest sizes. When $K = 2$, the hardest case is when one has two communities of the same size $n/2$. When $K > 2$, the hardest case is when one has two communities of sizes $n/K\beta$. Thus, we define

$$\bar{n} = \begin{cases} \dfrac{n}{2}, & \text{for } K = 2, \\ \dfrac{n}{K\beta}, & \text{for } K > 2, \end{cases} \tag{7}$$

as the effective sample size of the problem. The following result characterizes the statistical performance of the posterior distribution $\Pi(Z|A)$ under mild conditions.

Theorem 1 (Posterior strong consistency) *Recall that* $\Gamma(Z) = \{\sigma \circ Z : \sigma \in \mathcal{P}_K\}$, *where* $\mathcal{P}_K$ *stands for the set of all permutations on* $[K]$. *Suppose that*

$$\liminf_{n\to\infty} \frac{\bar{n}I}{\log n} > 1, \tag{8}$$

and the feasible set S_α *satisfies that* $\alpha - \beta$ *is a positive constant. Then, we have that*

$$\mathbb{E}[\Pi(Z \in \Gamma(Z^*)|A)] \geq 1 - n\exp(-(1-\eta_n)\bar{n}I) = 1 - o(1)$$

for a large n *and some positive sequence* η_n *tending to 0 as* $n \to \infty$, *and the expectation is with respect to the data-generating process.*

We defer the proof of the theorem to Section 5. It is worth noting that the condition required in Theorem 1 is identical to the fundamental limits required for exact label recovery Abbe (2016); Abbe et al. (2014); Zhang et al. (2016). In the special case of two communities of equal sizes, we require $nI > 2\log n$ to guarantee the strong consistency result. Hence, Theorem 1 implies that under our Bayesian framework, posterior strong consistency holds under the optimal condition.

We can also compare the statistical performance of our model with other Bayesian community detection approaches. The first Bayesian SBM was suggested by Snijders and Nowicki (1997), who considered two communities and proposed a uniform prior for both community proportions and the connectivity probabilities. It was further extended for more communities with Dirichlet priors on community proportions and Beta priors on the connectivity probabilities. However, the field of Bayesian SBM grows in a slow pace due to lack of theoretical analysis in terms of statistical consistency. Recently, van der Pas and van der Vaart van der Pas et al. (2017) proved that the strong consistency of MAP holds under a condition that the average expected degree satisfies $\lambda_n \gg \log^2 n$. With our notation, they require that $n(p-q)^2/p \gg \log^2 n$, much stronger than the optimal threshold order $n(p-q)^2/p \gtrsim \log n$. In contrast, our model introduces a feasible set S_α and proposes a uniform prior for label assignment Z on set S_α. It results in the strong consistency of posterior distribution under the condition that $n(p-q)^2/p \gtrsim \log n$, much weaker than the condition required in van der Pas et al. (2017).

3. Rapidly mixing of a Metropolis-Hastings algorithm

In this section, we propose a modified Metropolis-Hastings random walk, and analyze its statistical performance as well as the computational complexity. Due to the identifiability of the problem, the rapidly mixing property is analyzed in clustering space that will be defined in the sequel.

3.1 A Metropolis-Hastings algorithm

A general Metropolis-Hastings algorithm is an iterative procedure consisting of two steps:

Step 1 For the current state X_t, generate an $X' \sim q(x|X_t)$, where $q(x|X_t)$ is the proposal distribution defined on the same state space.

Step 2 Move to the new state X' with acceptance probability $\rho(X_t, X')$, and stay in the original state X_t with probability $1 - \rho(X_t, X')$, where the acceptance probability is given by

$$\rho(X_t, X') = \min\left\{1, \frac{p(X')q(X_t|X')}{p(X_t)q(X'|X_t)}\right\},$$

where $p(\cdot)$ is the target distribution.

In this paper, we are sampling the community label assignment $Z \in [K]^n$. In particular, we take the *single flip update* as the proposal distribution, which is to choose an index $j \in [n]$ uniformly at random, and then randomly choose $c \in [K] \setminus \{Z_t(j)\}$ to assign a new label. The whole algorithm is presented in Algorithm 1.

Algorithm 1: A Metropolis-Hastings algorithm for Bayesian community detection

---

Input: Adjacency matrix $A \in \{0,1\}^{n\times n}$,
number of communities K,
initial community assignment Z_0,
inverse temperature parameter ξ,
maximum number of iterations T.

Output: Community label assignment Z_T.

for each $t \in \{0, 1, 2, \ldots, T\}$ **do**

Choose an index $j \in [n]$ uniformly at random;

Randomly assign a new label for index j from the set $[K] \setminus \{Z_t(j)\}$ to get a new assignment Z';

$Z_{t+1} = Z'$ with probability

$$\rho(Z_t, Z') = \min\left\{1, \frac{\Pi^\xi(Z'|A)}{\Pi^\xi(Z_t|A)}\right\},$$

otherwise set $Z_{t+1} = Z_t$.

The Markov chain induced by Algorithm 1 is characterized by the transition matrix, which takes the form as

$$P(Z, Z') = \begin{cases} \dfrac{1}{n(K-1)} \min\left\{1, \dfrac{\Pi^\xi(Z'|A)}{\Pi^\xi(Z|A)}\right\}, & \text{if } H(Z, Z') = 1, \\ 1 - \displaystyle\sum_{\tilde{Z}:\tilde{Z}\neq Z} P(Z, \tilde{Z}), & \text{if } Z' = Z, \\ 0, & \text{if } H(Z, Z') > 1, \end{cases} \tag{9}$$

where $H(Z, Z')$ is the Hamming error between the two label assignments Z, Z'. The inverse temperature parameter ξ satisfies that $\xi \geq 1$. The algorithm is sampling from the scaled distribution $\widetilde{\Pi}(Z|A)$, where $\widetilde{\Pi}(Z|A) \propto \Pi^\xi(Z|A)$ for any $Z \in [K]^n$. As $\xi \to \infty$, the probability mass of $\widetilde{\Pi}(\cdot|A)$ concentrates on the global maximum of $\Pi(\cdot|A)$, in which case the algorithm is deterministic and reduces to a label switching algorithm, as discussed in Bickel and Chen (2009). When $\xi = 1$, asymptotically the algorithm is sampling from the true posterior distribution. The possible choices of ξ will be discussed in the sequel.

Corollary 2 *Under the conditions of Theorem 2.1, if $\xi \geq 1$, then we have*

$$\mathbb{E}\left[\sum_{Z \in \Gamma(Z^*)} \widetilde{\Pi}(Z|A)\right] \geq 1 - n\exp(-(1-o(1))nI),$$

where the expectation is with respect to the data-generating process.

Corollary 2 shows that posterior strong consistency also holds for $\widetilde{\Pi}(Z|A)$, and it holds by the same proof of Theorem 1, considering that $\xi \geq 1$ and the distribution concentrates more mass on large values. The parameter T in Algorithm 1 is the total number of iterations required. As long as the Markov chain mixes before T, according to Theorem 1, Z_T recovers the true community label assignment up to a label permutation with high probability, i.e., $Z_T \in \Gamma(Z^*)$ where $\Gamma(Z^*)$ is as defined in (3). Even though Theorem 1 is only stated for $\xi = 1$, it is easy to see that its conclusion also holds for $\widetilde{\Pi}(\cdot|A)$ for a general $\xi \geq 1$.

Due to the identifiability issue, the theoretical analysis of mixing time will be performed in the clustering space $\{\Gamma(Z) : Z \in S_\alpha\}$, where $\Gamma(Z)$ is defined in (3). We denote the state in the clustering space at time t as $\Gamma_t = \Gamma(Z_t)$, where Z_t is generated from Algorithm 1. The graphical model of the sequence $\{\Gamma_t\}_{t\geq 0}$ is illustrated in Figure 1.

Proposition 3 *The sequence $\{\Gamma_t\}_{t\geq 0}$ induced by Algorithm 1 is a Markov chain.*

Proof The proof relies on the permutation symmetry of the posterior distribution given by (6). We first introduce a distance between two clustering structures Γ and Γ', defined by

$$\breve{H}(\Gamma, \Gamma') = \min_{Z \in \Gamma, Z' \in \Gamma'} H(Z, Z'). \tag{10}$$

When $\breve{H}(\Gamma_{t+1}, \Gamma_t) \leq 1$, we have

$$\mathbb{P}\{\Gamma_{t+1} \mid \Gamma_s,\ s \leq t\} = \sum_{Z \in \Gamma_t} \mathbb{P}\{\Gamma_{t+1} \mid Z_t = Z\} \cdot \mathbb{P}\{Z_t = Z \mid \Gamma_s,\ s \leq t\}. \tag{11}$$

The equality holds since given Z_t, Γ_{t+1} and $\{\Gamma_s : s \leq t\}$ are independent. We proceed to calculate $\mathbb{P}\{\Gamma_{t+1} \mid Z_t = Z\}$. In the case of $\breve{H}(\Gamma_{t+1}, \Gamma_t) \leq 1$, it is obvious that for any $Z \in \Gamma_t$, there exists a unique $Z' \in \Gamma_{t+1}$ such that $H(Z, Z') \leq 1$. Thus, we have that

$$\mathbb{P}\{\Gamma_{t+1} \mid Z_t = Z\} = \sum_{\widetilde{Z} \in \Gamma_{t+1}} P(Z, \widetilde{Z}) = P(Z, Z'). \tag{12}$$

By (9), the transition probability $P(Z, Z')$ only depends on the ratio of $\Pi(Z'|A)$ and $\Pi(Z|A)$, and

$$\frac{\Pi(Z'|A)}{\Pi(Z|A)} = \frac{\Pi(\Gamma(Z')|A)}{\Pi(\Gamma(Z)|A)} = \frac{\Pi(\Gamma_{t+1}|A)}{\Pi(\Gamma_t|A)}, \tag{13}$$

which only depends on Γ_t and Γ_{t+1}. It follows that

$$\mathbb{P}\left\{\Gamma_{t+1} \mid Z_t\right\} = \mathbb{P}\left\{\Gamma_{t+1} \mid \Gamma_t\right\}.$$

Hence, plug the above identity into (11), and we have

$$\begin{aligned}
\mathbb{P}\left\{\Gamma_{t+1} \mid \Gamma_s,\ s \le t\right\} &= \sum_{Z \in \Gamma_t} \mathbb{P}\left\{\Gamma_{t+1} \mid \Gamma_t\right\} \cdot \mathbb{P}\left\{Z_t = Z \mid \Gamma_s,\ s \le t\right\} \\
&= \mathbb{P}\left\{\Gamma_{t+1} \mid \Gamma_t\right\} \cdot \sum_{Z \in \Gamma_t} \mathbb{P}\left\{Z_t = Z \mid \Gamma_s,\ s \le t\right\} \\
&= \mathbb{P}\left\{\Gamma_{t+1} \mid \Gamma_t\right\}.
\end{aligned}$$

When $\breve{H}(\Gamma_{t+1}, \Gamma_t) > 1$, it is obvious that

$$\mathbb{P}\left\{\Gamma_{t+1} \mid \Gamma_s,\ s \le t\right\} = 0 = \mathbb{P}\left\{\Gamma_{t+1} \mid \Gamma_t\right\}.$$

Therefore, $\{\Gamma_t\}_{t \ge 0}$ is a Markov chain by combining the conclusions of the two cases. ■

According to the above proposition and its proof, we can define the transition matrix $\breve{P}$ from state Γ to Γ' as

$$\breve{P}(\Gamma, \Gamma') = \begin{cases} \frac{1}{n(K-1)} \min\left\{1, \left[\frac{\Pi(\Gamma'|A)}{\Pi(\Gamma|A)}\right]^{\xi}\right\}, & \text{if } \breve{H}(\Gamma, \Gamma') = 1, \\ 1 - \sum_{\tilde{\Gamma}: \tilde{\Gamma} \neq \Gamma} \breve{P}(\Gamma, \tilde{\Gamma}), & \text{if } \Gamma' = \Gamma, \\ 0, & \text{if } \breve{H}(\Gamma, \Gamma') > 1. \end{cases} \tag{14}$$

We perform the analysis of mixing time for the Markov chain $\{\Gamma_t\}_{t \ge 0}$. Write $\breve{S}_\alpha = \{\Gamma(Z) : Z \in S_\alpha\}$ for simplicity, and we define the target distribution in the clustering space as $\breve{\Pi}(\Gamma|A) = \sum_{Z \in \Gamma} \widetilde{\Pi}(Z|A)$ for any $\Gamma \in \breve{S}_\alpha$. We directly have the following corollary by Corollary 2.

Corollary 4 *Under the conditions of Theorem 1, we have*

$$\mathbb{E}\left[\breve{\Pi}(\Gamma = \Gamma(Z^*)|A)\right] \ge 1 - n\exp(-(1 - o(1))\bar{n}I),$$

where the expectation is with respect to the data-generating process.

We show in the next section that $\{\Gamma_t\}_{t \ge 0}$ is rapidly mixing to the target distribution $\breve{\Pi}(\cdot|A)$.

3.2 Main results

Before stating the main theorem, we first review the definition of ε-mixing time, as well as the loss function that we need for the community detection problem.

ε-mixing time. Let $\Gamma_0 = \Gamma(Z_0)$ be the initial state of the chain. The total variation distance to the stationary distribution after t iterations is

$$\Delta_{Z_0}(t) = \left\| \breve{P}^t(\Gamma_0, \cdot) - \breve{\Pi}(\cdot|A) \right\|_{\mathsf{TV}} = \frac{1}{2} \sum_{\Gamma \in \{\Gamma(Z): Z \in S_\alpha\}} \left| \breve{P}^t(\Gamma_0, \Gamma) - \breve{\Pi}(\Gamma|A) \right|,$$

where $\breve{P}^t(\Gamma_0, \cdot)$ and $\breve{\Pi}(\cdot|A)$ are both distributions defined in the clustering space. The ε-mixing time for Algorithm 1 starting at Z_0 is defined by

$$\tau_\varepsilon(Z_0) = \min\left\{t \in \mathbb{N} : \Delta_{Z_0}(t') \leq \varepsilon \text{ for all } t' \geq t\right\}. \tag{15}$$

It is the minimum number of iterations required to ensure the total variation distance to the stationary distribution is less than some tolerance threshold ε.

Loss function. We introduce the misclassification proportion as a loss function, which is defined by

$$\ell(Z, Z^*) = \frac{1}{n} \breve{H}(\Gamma(Z), \Gamma(Z^*)), \tag{16}$$

where $\breve{H}(\cdot, \cdot)$ in defined in (10).

To this end, let us show that the proposed modified Metropolis-Hastings algorithm in Section 3.1 gives a rapidly mixing Markov chain $\{\Gamma_t\}_{t\geq 0}$ under the following conditions.

Condition A *There exist some positive sequences $\eta = \eta(n)$ and $\gamma_0 = \gamma_0(n)$ such that*

$$\inf_{B, Z^*} \mathbb{P}\left\{\ell(Z_0, Z^*) \leq \gamma_0\right\} \geq 1 - \eta.$$

We proceed to state the conditions for γ_0.

Condition B *Suppose the sequence γ_0 in Condition A satisfies one of the following cases:*

- *Case 1: there are only two communities, i.e., $K = 2$, and*

$$(1 - K\gamma_0)^4 nI \to \infty, \quad (1 - K\gamma_0)(1 - K\beta\gamma_0)n \to \infty, \tag{17}$$

  where $\beta \geq 1$ is defined in (2).

- *Case 2: there are more than 2 communities, i.e., $K \geq 3$, and*

$$\gamma_0 = o(1). \tag{18}$$

Condition A and Condition B require that the misclassification number of the initial label assignment is less than the minimum community size $n/K\beta$ with high probability. Consider the special situation where $K = 2$ and $\beta = 1 + o(1)$, i.e., the underlying two community share the same size asymptotically, Condition B is satisfied when the initial misclassification proportion is $1/2 - \varepsilon$ for some sequence $\varepsilon \to 0$. When $K \geq 3$, we require a stronger condition that the initial label assignment need to be weakly consistent, i.e., the initial misclassification error goes to 0 as $n \to \infty$. The initial condition can be easily satisfied by algorithms such as spectral clustering McSherry (2001); Rohe et al. (2011); Coja-Oghlan (2010); Fishkind et al. (2013).

Condition C *Suppose* $\limsup_{n\to\infty} \log n/\bar{n}I = 1-\varepsilon_0$. *With the hyperparameter* ξ *defined in Algorithm 1, one of the following cases holds:*

- *Case 1: there are only two communities, i.e.,* $K = 2$, *and*

$$\xi > (1-\varepsilon_0)\left\{\frac{1}{2\varepsilon_0} \vee \frac{\alpha^2}{(1-K\gamma_0)^4}\right\}, \tag{19}$$

where α *is defined in* (4), *and* γ_0 *is defined in Condition A.*

- *Case 2: there are more than 2 communities, i.e.,* $K \geq 3$, *and*

$$\xi > \frac{1-\varepsilon_0}{2\varepsilon_0}. \tag{20}$$

Note that the condition for the inverse temperature hyperparameter ξ also depends on the signal condition (ε_0) and initialization condition (γ_0). The condition of ξ is needed for technical proof to ensure the strong rapidly mixing property in the worst scenario. Note that with stronger initialization condition for the case of $K \geq 3$, the condition of ξ is slightly weaker than the case of $K = 2$. Simulation studies of how different values of ξ influence the performance of the algorithm are deferred to Section 4.

Here are some intuitive understandings of Condition C. Theorem 1 shows that posterior strong consistency holds under the condition $\liminf_{n\to\infty} \bar{n}I/\log n > 1$. Suppose for two label assignments Z_1, Z_2 that $\Pi(Z_1|A) > \Pi(Z_2|A)$, with hyperparameter $\xi \geq 1$, the posterior ratio $\Pi^\xi(Z_1|A)/\Pi^\xi(Z_2|A)$ gets enlarged, and the Markov chain is more certain to move towards the maximum point. However, the value of ξ is also constrained by the initialization Z_0. With a larger ξ, $\Pi^\xi(Z_0|A)$ is smaller and it takes longer for the Markov chain $\{\Gamma_t\}_{t\geq 0}$ to get mixed. The special case is that when the initialization Z_0 is weakly consistent, or equivalently, $\ell(Z_0, Z^*) \to 0$ as $n \to \infty$, then the value of ξ only depends on ε_0. It gives the following alternative condition that can replace Condition B and Condition C.

Condition D *Denote* $\limsup_{n\to\infty} \log n/\bar{n}I = 1-\varepsilon_0$. *The positive sequence* γ_0 *defined in Condition A and the hyperparameter* ξ *satisfy that*

$$\gamma_0 = o(1), \quad \xi \geq \frac{1-\varepsilon_0}{2\varepsilon_0}. \tag{21}$$

Theorem 5 (Rapidly mixing) *The initial label assignment is denoted by* Z_0. *Suppose Conditions (A, B, C) or Conditions (A, D) are satisfied. Then, the* ε*-mixing time of the modified Metropolis-Hastings algorithm is upper bounded by*

$$\tau_\varepsilon(Z_0) \leq 4Kn^2 \max\left\{\gamma_0, n^{-\tau}\right\} \cdot \left(\xi \log\left(\Pi(Z_0|A)^{-1}\right) + \log(\varepsilon^{-1})\right) \tag{22}$$

with probability at least $1 - C_1 n^{-C_2} - \eta$ *for some constant* $C_1, C_2 > 0$, *where* τ *is a sufficiently small constant, and* η *is defined in Condition A.*

Remark 6 *It is classical to perform theoretical analysis on a lazy version of Markov chain, which has probability 1/2 of staying unchanged, and the other probability 1/2 of updating the state. Theorem 5 is proved for the lazy Markov chain induced by Algorithm 1, i.e.,* *the corresponding transition matrix is $(\check{P}+I)/2$. The same tricks are widely used in Yang et al. (2016); DABBS (2009); Berestycki (2016); Montenegro et al. (2006). It is worth noting that this is only for the proof, and in practice, we still use the original transition matrix in Algorithm 1.*

Theorem 5 implies the mixing time depends on the initialization Z_0 and the choice of ξ. In order to show that the mixing time is at most a polynomial of n, we still need the following lemma to lower bound the initial posterior value $\Pi(Z_0|A)$.

Lemma 7 *Under the conditions of Theorem 1, we have*

$$\log \Pi(Z_0|A) \geq -C_3 n^2 I \cdot \ell(Z_0, Z^*), \tag{23}$$

with probability at least $1 - C_4 n^{-C_5}$ for some positive constants C_3, C_4, C_5.

Theorem 5 and Lemma 7 jointly imply that $\tau_\varepsilon(Z_0) \lesssim n^2(n^2 I + \log(\varepsilon^{-1}))$ with high probability, which demonstrates that the Markov chain of Metropolis-Hastings algorithm is rapidly mixing. To the best of our knowledge, (22) is the first explicit upper bound on the mixing time of the Markov chain for Bayesian community detection.

Note that the target distribution of Algorithm 1 is $\widetilde{\Pi}(\cdot|A) \propto \Pi^\xi(\cdot|A)$. Since $\xi \geq 1$ and the posterior strong consistency property still holds for $\widetilde{\Pi}(\cdot|A)$, Theorem 5 shows that Algorithm 1 will find the maximum a posteriori in polynomial time with high probability.

Corollary 8 *Under the condition of Theorem 5, for any iteration number T such that $T \geq C_6 n^2(n^2 I + \log(\varepsilon^{-1}))$ for some constant C_6, the output Z_T of the Algorithm 1 satisfies that $Z_T \in \Gamma(Z^*)$ with high probability, or equivalently, $\ell(Z_T, Z^*) = 0$.*

The following corollary focuses on the case of $\xi = 1$, and gives explicit conditions for the Markov chain to converge to the posterior distribution $\Pi(\cdot|A)$.

Corollary 9 *When $nI/\log n \to \infty$, suppose Condition B holds, and we can take $\xi = 1$ in Algorithm 1, which reduces to the standard Metropolis-Hastings algorithm sampling from $\Pi(\cdot|A)$. We have that the ε-mixing time of the Markov chain is upper bounded by $O(n^2(n^2 I + \log(\varepsilon^{-1})))$ with high probability.*

The conditions of the above results can be weakened in the case where the connectivity probability matrix B is known. When B is known, there is no need to put a prior on B. Thus, the posterior distribution can be simplified as

$$\begin{aligned} \log \Pi(Z|A) = &\log \frac{p(1-q)}{q(1-p)} \sum_{i<j} A_{ij} \mathbb{I}\{Z_i = Z_j\} - \\ &\log \frac{1-q}{1-p} \sum_{i<j} \mathbb{I}\{Z_i = Z_j\} + Const, \qquad \text{for } Z \in S_\alpha. \end{aligned} \tag{24}$$

The posterior formula is essentially the same as likelihood, while we restrict Z inside the feasible set S_α. It can be shown that the posterior strong consistency property still holds in this case.

Theorem 10 (posterior strong consistency) *Suppose that* $\limsup_{n\to\infty} \bar{n}I/\log n > 1$, *and the feasible set* S_α *satisfies that* $\alpha - \beta$ *is a positive constant, then it follows that*

$$\mathbb{E}\left[\Pi(Z \in \Gamma(Z^*)|A)\right] \geq 1 - o(1),$$

with high probability, and the expectation is with respect to the data-generating process.

Condition E *Suppose* $\limsup_{n\to\infty} \log n/\bar{n}I = 1 - \varepsilon_0$. *Assume the positive sequence* γ_0 *defined in Condition A and the hyperparameter* ξ *satisfy one of the following conditions:*

- *Case 1:*

$$\gamma_0 = o(1), \quad \xi \geq \frac{1-\varepsilon_0}{2\varepsilon_0}. \tag{25}$$

- *Case 2:*

$$(1 - K\alpha\gamma_0)^2 nI \to \infty, \quad \xi > \begin{cases} (1-\varepsilon_0)\left(\dfrac{1}{2\varepsilon_0} \vee \dfrac{\alpha}{4(1-K\alpha\gamma_0)}\right), & \text{for } K = 2, \\ (1-\varepsilon_0)\left(\dfrac{1}{2\varepsilon_0} \vee \dfrac{\alpha}{4\beta(1-K\alpha\gamma_0)}\right), & \text{for } K \geq 3. \end{cases} \tag{26}$$

Condition E yields the rapidly mixing property when the connectivity matrix B is known.

Theorem 11 (Rapidly mixing) *Suppose we start the algorithm at* Z_0, *and Conditions (A, E) hold. Then, the* ε*-mixing time of the Metropolis-Hastings algorithm induced by Equation* (24) *is upper bounded by*

$$\tau_\varepsilon(Z_0) \leq 4Kn^2 \max\left\{\gamma_0, n^{-\tau}\right\} \cdot \left(\xi \log \Pi(Z_0|A)^{-1} + \log(\varepsilon^{-1})\right),$$

with probability at least $1 - C_7 n^{-C_8} - \eta$ *for some constants* C_7, C_8, *where* τ *is a sufficiently small constant, and* η *is as defined in Condition A.*

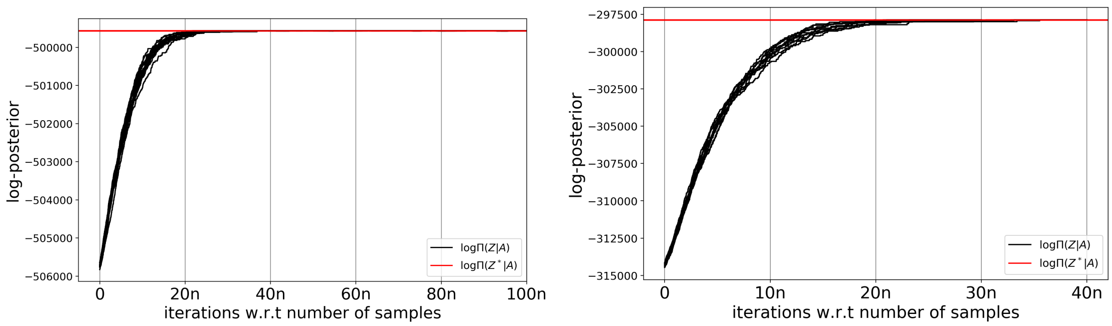
Compare the result with Theorem 5, and we can see that Theorem 11 obtains the same upper bound with slightly weaker conditions for the initialization Z_0 and hyperparameter ξ.

4. Numerical Results

In this section, we study the numerical performance of the Metropolis-Hastings algorithm [1], and the inverse temperature parameter ξ is set to be 1 unless otherwise specified. The initial label assignment vector is chosen such that: half samples are labeled correctly, and the other half samples are labeled randomly. The same mechanism is also mentioned in paper Bickel and Chen (2009).

1. The code is available on `https://github.com/zhuobumeng/MH_bayes_SBM`.

Balanced networks. In this setting, we generate networks with 2500 nodes, and 5 communities, each of which consists of 500 nodes. Figure 2 shows the trajectories of the Markov chains (each denoted by a black line). By posterior strong consistency, the true label assignment receives the highest posterior probability (denoted by the red line), and the Markov chains converge rapidly to the stationarity (within $40n$ iterations), demonstrating the rapidly mixing property.



(a) (b)

Figure 2: Log-posterior probability versus the number of iterations. Each black curve corresponds to a trajectory of the chain (20 chains in total), and the red horizontal line represents the log-posterior probability at the true label assignment. (a) A network with $p = 0.48$ and $q = 0.32$. (b) A network with $p = 0.3$ and $q = 0.1$.

Heterogeneous networks. In this setting, we generate networks with 2000 nodes and 4 communities of sizes 200, 400, 600, and 800, respectively. The connectivity matrix is set as

$$B = \begin{pmatrix} 0.50 & 0.29 & 0.35 & 0.25 \\ 0.29 & 0.45 & 0.25 & 0.30 \\ 0.35 & 0.25 & 0.50 & 0.35 \\ 0.25 & 0.30 & 0.35 & 0.45 \end{pmatrix}.$$

The algorithm still performs well. As shown in Figure 3, the posterior strong consistency still hold, and the Markov chains rapidly converge to the stationarity.

Study of inverse temperature parameter ξ. We conduct experiment with two different settings. The one is sparse homogeneous balanced network with 2500 nodes, and the other one is heterogeneous unbalanced network. We aim to study how performance changes with different values of ξ in Algorithm 1. The result is shown below.

From Figure 4, the algorithm performs well when $\xi \geq 0.5$. In both cases, when $\xi = 0.1$, Markov chains fail to mix to the target distribution within $120n$ iterations. It seems that Condition C and Condition D are mainly for technical proof of rapid mixing property, and in practice, we observe that Markov chains mix rapidly by simply choosing $\xi = 1$.

Necessity of the initialization condition. We show that the initialization condition required by our main theorems is necessary by numerical experiments. Consider the network

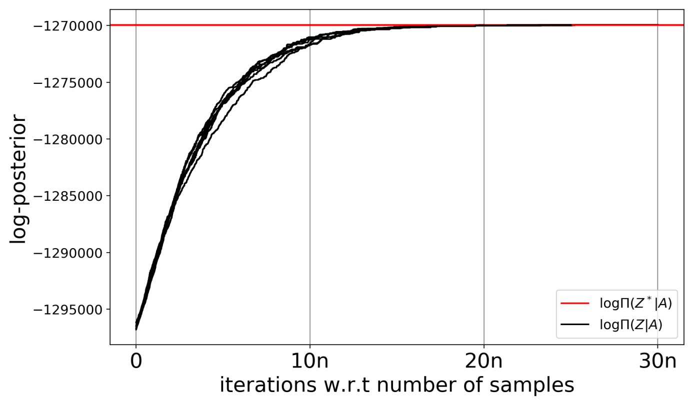
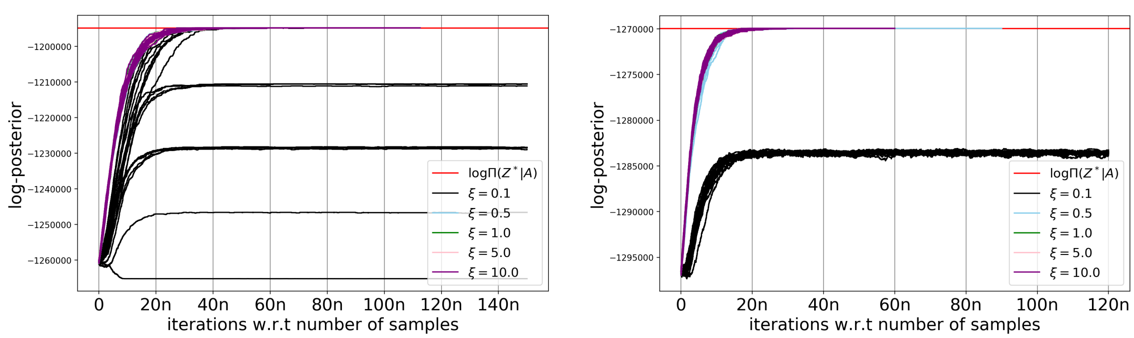


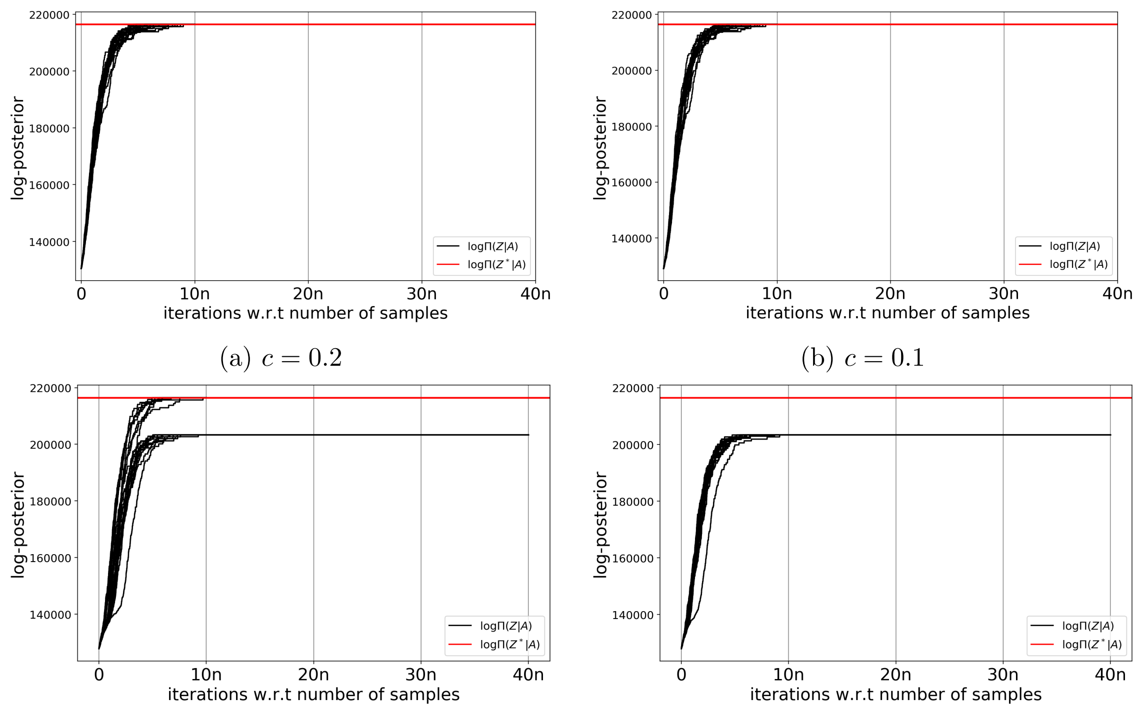
Figure 3: Log-posterior probability versus the number of iterations. Each black curve corresponds to a trajectory of the chain (20 chains in total), and the red horizontal line represents the log-posterior probability at the true label assignment.



(a) (b)

Figure 4: Log-posterior probability versus the number of iterations. Each curve is a trajectory of Markov chain (20 chains for each color), and the red horizontal line represents the log-posterior probability at the true label assignment. Different colors represent chains with different value of ξ. (a) Homogeneous network with $p = 0.3$ and $q = 0.1$. (b) Heterogeneous network designed the same as the second experiment in simulation studies.

with two communities of size 270 and 460, and the connectivity probabilities are set to be $p = 10^{-1}$, $q = 10^{-8}$. The initial label assignment Z_0 satisfies $\ell(Z_0, Z^*) = (1-c)/2\alpha$, and then Condition E is equivalent to $c > 0$ and $c^2 nI \to \infty$. In simulations, we run experiments for $c = 0.2, 0.1, -0.1, -0.2$, and the results are shown as below.



(a) $c = 0.2$ (b) $c = 0.1$

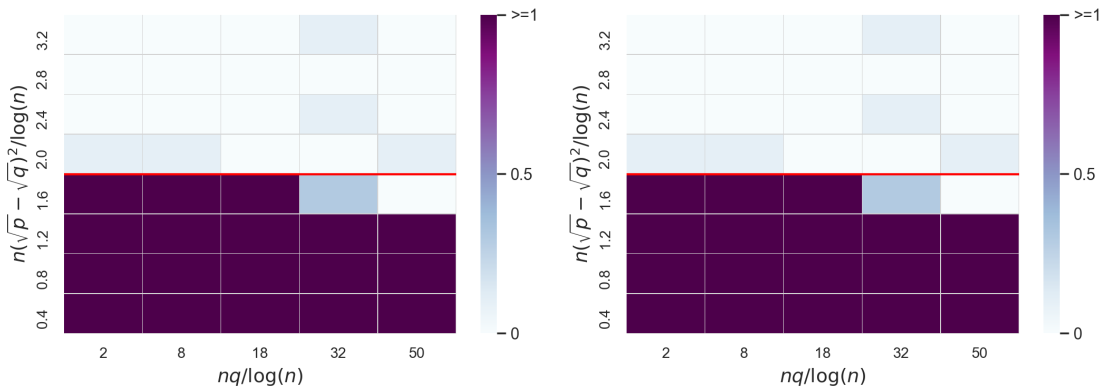
(c) $c = -0.1$ (d) $c = -0.2$

Figure 5: Log-posterior probability versus the number of iterations. The initial label assignment Z_0 is constructed so that the labels of the community of size 270 are all correct, and there are $n(1-c)/2\alpha$ labels in the community of size 460 are incorrect. Each black curve corresponds to a trajectory of the chain (20 chains in total), and the red horizontal line represents the log-posterior probability at the true label assignment.

Figure 5 shows that when $c < 0$, it is very likely for the algorithm to get stuck at some local maximum, and does not converge to the stationary distribution.

Fundamental limit of the signal condition. We check that the fundamental limit of the signal condition can be achieved by the Metropolis-Hastings algorithm. We generate homogeneous networks with 1000 nodes and 2000 nodes, and each has two communities of equal sizes. Figure 6 is the heatmap of the number of misclassified samples, where every rectangular block represents one setting with different values of p and q. In each setting, we run 20 experiments with independent initializations and adjacency matrices, and the value of each block is the average number of misclassified samples in the 20 experiments. Figure

6 shows that when $nI > 2\log n$, we are able to exactly recover the underlying true label assignment, and the result of simulation coincides with the posterior strong consistency property in Section 2.3.



(a) Network with 1000 nodes

(b) Network with 2000 nodes

Figure 6: The heatmap of the number of misclassified samples. The red line in each plot represents the fundamental limit with $K = 2$.

5. Proofs

The posterior strong consistency property, Theorem 1 and Theorem 10, is proved in Section 5.1. The main result of the paper, Theorem 5, is proved in Section 5.2.

5.1 Proof of posterior strong consistency

We first state the proof in the case where the connectivity probability matrix B is known (Theorem 11). Then, by similar techniques, we have the result of Theorem 1. To distinguish the two cases, we denote the posterior distribution as $\Pi_0(\cdot|A)$ with a known connectivity probability matrix. In this section, we use $d(Z, Z^*) = n\ell(Z, Z^*) = m(Z)$ to denote the number of mistakes for the label assignment Z. For simplicity, we also write m for $m(Z)$ with a slight abuse of notation.

5.1.1 Proof of Theorem 10

We first state a lemma in order to prove the theorem.

Lemma 12 (Lemma 5.4 in Zhang et al. (2016)) *For any constants $\alpha > \beta \geq 1$, let $Z \in S_\alpha$ be an arbitrary assignment satisfying that $d(Z, Z^*) = m$ with $0 < m < n$. Then, for the $\Pi_0(Z|A) = \Pi(Z|A)$ defined in (24), we have*

$$\mathbb{P}\left\{\Pi_0(Z|A) > \Pi_0(Z^*|A)\right\} \leq \begin{cases} \exp\left(-(\bar{n}m - m^2)I\right), & m \leq \dfrac{n}{2K}, \\ \exp\left(-d_{\alpha,\beta}\dfrac{nmI}{K}\right), & m > \dfrac{n}{2K}, \end{cases}$$

where $d_{\alpha,\beta}$ is some positive constant that only depends on α, β.

Proof [Proof of Theorem 10] Recall that for any $Z' \in \Gamma(Z)$, we have $\Pi_0(Z'|A) = \Pi_0(Z|A)$. For each $Z \in S_\alpha \cap \Gamma^c(Z^*)$, let $G_Z = \{\Pi_0(Z|A) > \Pi_0(Z^*|A)\}$, and define $G = \cup_{Z \in S_\alpha \cap \Gamma^c(Z^*)} G_Z$. Let P_Z denote the likelihood function for the assignment Z. With the uniform prior on S_α, we have

$$\begin{aligned}
\mathbb{E}\Pi_0(Z \notin \Gamma(Z^*)|A) =& P_{Z^*}\Pi_0(Z \notin \Gamma(Z^*)|A)\mathbb{I}\{G^c\} + P_{Z^*}\Pi_0(Z \notin \Gamma(Z^*)|A)\mathbb{I}\{G\} \\
\leq& P_{Z^*} \sum_{Z \notin \Gamma(Z^*)} \frac{P_Z}{P_{Z^*} + \sum_Z P_Z}\mathbb{I}\{G_Z^c\} + \sum_{Z \notin \Gamma(Z^*)} P_{Z^*}(G_Z) \\
\leq& P_{Z^*} \sum_{Z \notin \Gamma(Z^*)} \frac{P_Z}{P_{Z^*}}\mathbb{I}\{G_Z^c\} + \sum_{Z \notin \Gamma(Z^*)} P_{Z^*}(G_Z) \\
=& \sum_{Z \notin \Gamma(Z^*)} P_Z(G_Z^c) + P_{Z^*}(G_Z) \\
=& 2 \sum_{Z \notin \Gamma(Z^*)} \mathbb{P}\{\Pi_0(Z|A) > \Pi_0(Z^*|A)\}.
\end{aligned}$$

In the first inequality, the former part holds due to the uniform prior over Z and

$$\Pi_0(Z \notin \Gamma(Z^*)|A) = \sum_{Z \notin \Gamma(Z^*)} \frac{P_Z}{P_Z^* + \sum_Z P_Z},$$

where P_Z is the likelihood function given label assignment Z, and we also have $\mathbb{I}\{G^c\} \leq \mathbb{I}\{G_Z^c\}$. The latter part holds due to $\Pi_0(Z \notin \Gamma(Z^*)|A) \leq 1$ and $\mathbb{I}\{G\} \leq \sum_{Z \notin \Gamma(Z^*)} \mathbb{I}\{G_Z\}$ by uniform bound.

The second inequality holds by removing $\sum_Z P_Z$ in the denominator. The final equality holds by symmetry. We also have

$$|\{\Gamma : \exists Z \in \Gamma, s.t.\ d(Z, Z^*) = m\}| \leq \binom{n}{m}(K-1)^m \leq \min\left\{\left(\frac{enK}{m}\right)^m, K^n\right\}.$$

Note that $\{Z : Z \notin \Gamma(Z^*)\}$ is equivalent to set $\{Z : m(Z) \geq 1\}$. With the condition that $\bar{n}I > \log n$, it follows by Lemma 12 that

$$\begin{aligned}
\mathbb{E}\Pi_0(Z \notin \Gamma(Z^*)|A) &\leq 2 \sum_{1 \leq m \leq n/2K} \binom{n}{m} K^m \exp\left(-(\bar{n}m - m^2)I\right) + 2 \sum_{m > n/2K} K^n \exp\left(-d_{\alpha,\beta} mnI/K\right) \\
&\leq 2 \sum_{1 \leq m \leq n/2K} \binom{n}{m} K^m \exp\left(-(\bar{n}m - m^2)I\right) + 2nK^n \exp\left(-Cn^2 I\right)
\end{aligned} \tag{27}$$

for some constant C. We proceed to upper bound the first term in (27). It follows that

$$\sum_{1 \leq m \leq n/2K} \binom{n}{m} K^m \exp\left(-(\bar{n}m - m^2)I\right) \leq \sum_{1 \leq m \leq n/2K} (enK)^m \exp\left(-(\bar{n}m - m^2)I\right) = \sum_m P_m$$

where $P_m = (enK)^m \exp\left(-(\bar{n}m - m^2)I\right)$. The ratio of P_m and P_1 is calculated as

$$\frac{P_m}{P_1} = (enK)^{m-1}\exp(-\bar{n}I(m-1) + (m^2-1)I) = (enK\exp(-\bar{n}I + (m+1)I))^{m-1}.$$

Define $m' = \varepsilon' n$ for some positive sequence $\varepsilon' = \varepsilon'_n$ with $\varepsilon' \to 0$ and $\varepsilon' nI \to \infty$. Then, $\sum_{1\le m\le n/2K} P_m$ can be split into summation of $\sum_{1\le m<m'} P_m$ and $\sum_{m'\le m\le n/2K} P_m$, where

$$\begin{aligned}\sum_{m=1}^{m'-1} P_m = P_1 \sum_{m=1}^{m'-1}\frac{P_m}{P_1} &\le P_1\sum_{m=1}^{m'-1}\left(enK\exp\left(-\bar{n}I + m'I\right)\right)^{m-1} \\ &\le enK\exp(-\bar{n}I + I)\cdot(1 + 2enK\exp(-\bar{n}I + \varepsilon' nI)),\end{aligned}$$

and there exists some constant C such that

$$\sum_{m'<m\le n/2K} P_m \le nK^n\exp(-\varepsilon'(C-\varepsilon')n^2 I) \le \exp(-n).$$

Hence, by combining all parts and based on the condition that $\bar{n}I > \log n$, we have $\Pi_0(Z \notin \Gamma(Z^*)|A) \le Cn\exp(-\bar{n}I)$ for some constant C and for a large n.

■

5.1.2 Proof of Theorem 1

Lemma 13 *Let $Z \in S_\alpha$ be an arbitrary assignment with $d(Z, Z^*) = m > 0$. If $p, q \to 0$ and $p \asymp q$, there exists some positive sequence $\gamma = \gamma_n$ with $\gamma \to 0$ and $\gamma^2 nI \to \infty$, such that for the $\Pi(Z|A)$ defined in (5), we have*

$$\mathbb{P}\left\{\max_{Z\in S_\alpha : m>\gamma n}\log\frac{\Pi(Z|A)}{\Pi(Z^*|A)} \ge -C_1\gamma n^2 I\right\} \le 4\exp(-n),$$

and

$$\mathbb{P}\left\{\max_{Z\in S_\alpha : m\le\gamma n}\log\frac{\Pi(Z|A)}{\Pi(Z^*|A)} - \log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)} - C_2\gamma mnI > 0\right\} \le n\exp\left(-(1-o(1))\bar{n}I\right),$$

for some constants C_1, C_2. Here, $\Pi_0(Z|A)$ is the posterior probability with known connectivity probabilities.

The proof of Lemma 13 is deferred to Section 5.5. We now state another lemma that is based on Proposition 5.1 in Zhang et al. (2016).

Lemma 14 (Proposition 5.1 in Zhang et al. (2016)) *For any $Z \in S_\alpha$ where $d(Z, Z^*) = m < n/2K$,*

$$\mathbb{E}\sqrt{\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}} \le \exp\left(-\bar{n}mI + m^2 I\right).$$

Proof [Proof of Theorem 1]

With Lemma 13 and Lemma 14, we divide S_α into a large mistake region and a small mistake region according to whether $m > \gamma n$, where γ is a positive sequence defined in Lemma 13.

Large mistake region. For $m > \gamma n$, by Lemma 13, with probability at least $1 - 4\exp(-n)$,

$$\sum_{Z \in S_\alpha : m > \gamma n} \frac{\Pi(Z|A)}{\Pi(Z^*|A)} \le nK^n \exp(-C_1 \gamma n^2 I) \le \exp(-n).$$

Denote $\mathcal{E} = \{\sum_{Z \in S_\alpha : m > \gamma n} \frac{\Pi(Z|A)}{\Pi(Z^*|A)} \le \exp(-n)\}$, and it follows directly that

$$\mathbb{E}\left[\sum_{Z \in S_\alpha : m > \gamma n} \Pi(Z|A)\right] \le \mathbb{E}\left[\Pi(Z : m > \gamma n|A)\mathbb{I}\left\{\mathcal{E}\right\}\right] + \mathbb{P}\left\{\mathcal{E}^c\right\} \le 5\exp(-n).$$

Small mistake region. For $m \le \gamma n$, let $G_Z = \{\Pi_0(Z|A) > \Pi_0(Z^*|A)\}$. Let θ denote all unknown parameters and θ_0 denote the underlying true parameters respectively. Define $\mathcal{F} = \{\max_{Z \in S_\alpha : m \le \gamma n} \log \frac{\Pi(Z|A)}{\Pi(Z^*|A)} - \log \frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)} - C_2 \gamma m n I > 0\}$ as in Lemma 13. Then, we have

$$\begin{aligned}
&\mathbb{E}\Pi(Z : 1 \le m \le \gamma n|A) \\
\le& P_{Z^*,\theta_0}\Pi(Z : 1 \le m \le \gamma n|A)\mathbb{I}\left\{\mathcal{F}^c\right\} + \mathbb{P}\left\{\mathcal{F}\right\} \\
\le& \sum_{Z : 1 \le m \le \gamma n} P_{Z^*,\theta_0} \frac{\Pi(Z|A)}{\Pi(Z^*|A)} \mathbb{I}\left\{G_Z^c, \mathcal{F}^c\right\} + \sum_{Z : 1 \le m \le \gamma n} P_{Z^*,\theta_0}\mathbb{I}\left\{G_Z\right\} + \mathbb{P}\left\{\mathcal{F}\right\} \\
\le& \sum_{Z : 1 \le m \le \gamma n} P_{Z^*,\theta_0} \frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)} \mathbb{I}\left\{G_Z^c\right\} \exp\left(C_2 \gamma m n I\right) + \sum_{Z : 1 \le m \le \gamma n} P_{Z^*,\theta_0}\mathbb{I}\left\{G_Z\right\} + \mathbb{P}\left\{\mathcal{F}\right\} \\
\le& \sum_{Z : 1 \le m \le \gamma n} \exp\left(C_2 \gamma m n I\right) P_{Z,\theta_0}\mathbb{I}\left\{G_Z^c\right\} + \sum_{Z : 1 \le m \le \gamma n} P_{Z^*,\theta_0}\mathbb{I}\left\{G_Z\right\} + \mathbb{P}\left\{\mathcal{F}\right\} \\
\le& n\exp\left(-(1 - o(1))\bar{n}I\right).
\end{aligned}$$

Recall that $\Pi_0(\cdot|A)$ denotes the posterior distribution with knowledge of the connectivity probabilities. The second inequality is due to $\Pi(Z^*|A) \le 1$. The third inequality is due to the definition of the event $\mathcal{F}$. The last two inequalities hold by Lemma 12 and symmetry.

Combine the two regions, and then

$$\mathbb{E}\Pi(Z \notin \Gamma(Z^*)|A) \le n\exp(-(1 - o(1))\bar{n}I).$$

The proof is complete. ■

5.2 Proofs of Theorem 5 and Theorem 11

We first list the main challenges of the proof in the context of community detection, as well as the novel proof techniques on top of previous developments based on the canonical path ensemble. Then, the detailed proofs of mixing time are presented from Section 5.2.2 to Section 5.2.5.

5.2.1 Main challenges and proof techniques

In this section, we highlight the main challenges in our work, and for each challenge, a new proof technique is introduced on top of canonical paths technique.

First of all, due to identification issue, the permutation of labels gives the same clustering structure, and one challenge is to do mixing time analysis in the clustering space. Since the Metropolis-Hastings algorithm only outputs the label assignment vectors, we perform all the theoretical analyses of mixing time for the states of clustering structures $\{\Gamma(Z) : Z \in S_\alpha\}$ according to the graphical model in Figure 1. Recall that in Figure 1, $\Gamma_t = \Gamma(Z_t)$ is the clustering state at time t, where Z_t is generated from the Metropolis-Hastings algorithm. Hence, both results of posterior strong consistency and rapidly mixing are referring to the clustering space.

Secondly, to analyze the Metropolis-Hastings algorithm in Bayesian community detection, another big challenge is to define a canonical path $\mathcal{T}$ with small path congestion parameter $\rho(\mathcal{T})$, where the mixing time of Markov chain scales as $\rho(\mathcal{T})$. In this work, we did not design a set of canonical paths in the whole clustering space, since the path between two clustering states with large mistakes are overloaded ("congested"). Hence, to overcome this challenge, we first construct a super martingale to prove that within any polynomial running time, the number of mistakes of label assignments sampled from the Metropolis-Hastings algorithm is upper bounded, which forms a good region of clustering space with small mistakes. Then, we construct canonical paths between every two clustering states Γ and Γ' within such good region. In this way, it rules out extreme clustering cases with large number of mistakes, and yields polynomial bound for $\rho(\mathcal{T})$.

Last but not least, it is not good enough for us to establish rapid mixing of the Metropolis-Hastings algorithm, but we need to show rapid mixing under the optimal signal to noise ratio condition of community detection. In other words, the statistical and algorithmic properties of the Metropolis-Hastings algorithm meed to be established simultaneously. We use two techniques to overcome this challenge. First, we use Chernoff bound to carefully study the posterior ratio along the canonical path. Then, we introduce scaled posterior distribution with temperature parameter ξ, and upper bound the path congestion parameter $\rho(\mathcal{T})$ by the order of n.

5.2.2 Backgrounds on mixing time

Consider a reversible, irreducible, and aperiodic Markov chain on a discrete space Ω that is completely specified by a transition matrix $P \in [0,1]^{|\Omega|\times|\Omega|}$ with stationary distribution Π. Let $\omega \in \Omega$ be the initial state of the chain, and then the total variation distance to the stationary distribution after t iterations is

$$\Delta_\omega(t) = \|P^t(\omega, \cdot) - \Pi\|_{\mathsf{TV}},$$

where $P^t(\omega, \cdot)$ is the distribution of the chain after t iterations. The ε-mixing time starting at ω is given by

$$\tau_\varepsilon(\omega) = \min\left\{t \in \mathbb{N} : \Delta_\omega(t') \le \varepsilon \text{ for all } t' \ge t\right\}.$$

With this notation, we say a Markov chain is rapidly mixing if $\tau_\varepsilon(\omega)$ is $O\left(\text{poly}(\log(|\Omega|/\varepsilon))\right)$ in the case where $|\Omega|$ scales exponentially to the problem size n. This means we only need to update the Markov chain for poly(n) steps in order to obtain *good* samples from the stationary distribution. The explicit bound for the mixing time through the spectral gap is

$$\tau_\varepsilon(\omega) \le \frac{-\log \Pi(\omega) + \log(1/\varepsilon)}{Gap(P)}, \tag{28}$$

where $Gap(P)$ represents the spectral gap of the transition matrix P, defined by $Gap(P) = 1 - \max\{|\lambda_2(P)|, |\lambda_{\min}(P)|\}$, where $\lambda_2(P)$, $\lambda_{\min}(P)$ are the second largest and the smallest eigenvalues of the transition matrix P. See the paper Woodard et al. (2013) for this bound.

5.2.3 Preparation

Suppose $P(\cdot,\cdot)$ in (9) is the transition matrix introduced in Algorithm 1 defined in the label assignment space S_α, and $\breve{P}(\cdot,\cdot)$ in (14) is the transition matrix of $\{\Gamma_t\}_{t\ge 0}$ defined in the clustering space $\breve{S}_\alpha = \{\Gamma(Z) : Z \in S_\alpha\}$. The stationary distribution for P and $\breve{P}$ are denoted as $\widetilde{\Pi}$ and $\breve{\Pi}$ respectively. We require a good initializer, and use the following lemma to guarantee that all possible states visited by the algorithm remain in a good region with high probability.

Lemma 15 *Suppose we start at a fixed initializer Z_0 with $\ell(Z_0, Z^*) \le \gamma_0$ where γ_0 satisfies Condition B, D, or E. Then, the number of misclassified nodes in any polynomial running time can be upper bounded by*

$$m = n \cdot \ell(Z, Z^*) \le n \max\left\{\gamma_0, n^{-\tau}\right\} + \log^2 n, \tag{29}$$

with probability at least $1 - \exp(-\log^2 n)$, where $\tau > 0$ is a sufficiently small constant.

The proof of Lemma 15 is deferred to Section 5.3. Note that Lemma 15 is stated conditioning on a fixed initial label assignment Z_0 with $\ell(Z_0, Z^*) \le \gamma_0$. This is slightly different from the original initialization conditions where we use Z_0 dependent on data. A simple union bound will lead to the final conclusion. Lemma 15 quantifies the maximum possible number of classification mistakes when starting at a good initializer. Here, $(\log n)^2$ is chosen for simplicity and can be replaced by any sequence $\nu_n \gg \log n$.

Let $\mathcal{G}(\gamma_0)$ denote a good region with respect to the initial misclassification proportion, defined by

$$\mathcal{G}(\gamma_0) = \{Z \in S_\alpha : m \le n \max\{\gamma_0, n^{-\tau}\} + \log^2 n\}, \tag{30}$$

where τ is a sufficiently small constant. Accordingly, we can define a good region in the clustering space as

$$\breve{\mathcal{G}}(\gamma_0) = \{\Gamma(Z) : Z \in \mathcal{G}(\gamma_0)\},$$

and Lemma 15 ensures that for any T that is a polynomial of n, $\{\Gamma_t\}_{0\leq t\leq T}$ stays inside $\breve{\mathcal{G}}(\gamma_0)$ with high probability. Sometimes we write $\mathcal{G}(\gamma_0)$ and $\breve{\mathcal{G}}(\gamma_0)$ as $\mathcal{G}$ and $\breve{\mathcal{G}}$ for simplicity. Then, we modify the distributions and transition matrices according to the regions $\mathcal{G}$ and $\breve{\mathcal{G}}$. Denote the modified distributions as $\widetilde{\Pi}_g(Z|A) \propto \Pi^{\xi}(Z|A)\mathbb{I}\{Z \in \mathcal{G}\}$ for all $Z \in S_\alpha$, and $\breve{\Pi}_g(\Gamma|A) \propto \breve{\Pi}(\Gamma|A)\mathbb{I}\left\{\Gamma \in \breve{\mathcal{G}}\right\}$ for all $\Gamma \in \breve{S}_\alpha$. Define in the label assignment space the new transition matrix $P_g(\cdot,\cdot)$ corresponding to $\widetilde{\Pi}_g(\cdot|A)$, by replacing $\Pi^{\xi}(\cdot|A)$ with $\widetilde{\Pi}_g(\cdot|A)$ in (9). Define in the clustering space the new transition matrix $\breve{P}_g(\cdot,\cdot)$ corresponding to $\breve{\Pi}_g(\cdot|A)$, by replacing $\Pi(\cdot|A)$ with $\Pi(\cdot|A)\mathbb{I}\left\{\cdot \in \breve{\mathcal{G}}\right\}$ in (14).

With these notations, we proceed to bound the total variation error between the distribution of Γ_T and $\breve{\Pi}(\cdot|A)$ after T steps for some T that is a polynomial of n.

Lemma 16 (TV difference)

$$\mathbb{E}\|\breve{\Pi}_g(\cdot|A) - \breve{\Pi}(\cdot|A)\|_{\mathsf{TV}} \leq n\exp(-(1-o(1))\bar{n}I).$$

Proof We have

$$\mathbb{E}\|\breve{\Pi}_g(\cdot|A) - \breve{\Pi}(\cdot|A)\|_{\mathsf{TV}} \leq 2\mathbb{E}\breve{\Pi}(\Gamma \notin \breve{\mathcal{G}}|A) \leq 2\mathbb{E}\breve{\Pi}(\Gamma \neq \Gamma(Z^*)|A) \leq n\exp(-(1-o(1))\bar{n}I),$$

where the first inequality is due to Lemma 52. The second inequality is due to the definition of $\breve{\mathcal{G}}$. The last inequality directly follows by Theorem 1 or Theorem 11, and the condition that $\xi \geq 1$. ■

Thus, by triangle inequality, we can decompose the total variation bound at time T as

$$\begin{aligned}
&\|\breve{P}^T(\Gamma_0,\cdot) - \breve{\Pi}(\cdot|A)\|_{\mathsf{TV}} \\
\leq &\|\breve{P}^T(\Gamma_0,\cdot) - \breve{P}_g^T(\Gamma_0,\cdot)\|_{\mathsf{TV}} + \|\breve{P}_g^T(\Gamma_0,\cdot) - \breve{\Pi}_g(\cdot|A)\|_{\mathsf{TV}} + \|\breve{\Pi}_g - \breve{\Pi}(\cdot|A)\|_{\mathsf{TV}},
\end{aligned} \tag{31}$$

where T is the number of iterations. Lemma 15 implies that the first term is 0 with high probability for $T \leq \text{poly}(n)$, since the algorithm stays in the region $\mathcal{G}(\gamma_0)$. The third term can be upper bounded by Lemma 16. Therefore, the remaining proof is to adopt the canonical path approach to bound the second term in (31).

For the purpose of the proof, we replace the transition matrix $\breve{P}_g$ by its lazy version, which has a probability of 1/2 at staying at its current state, and another probability of 1/2 at updating the state. The same technique can be also found in Yang et al. (2016); DABBS (2009); Berestycki (2016); Montenegro et al. (2006). It is worth noting that this technique is only for the proof.

5.2.4 Canonical path

Given an ergodic Markov chain $\mathcal{C}$ induced by the lazy transition matrix $\breve{P}_g$ in the discrete state space $\breve{\mathcal{G}}$, we define a weighted directed graph $G(\mathcal{C}) = (V, E)$, where the vertex set $V = \breve{\mathcal{G}}$ and an edge between an ordered pair (Γ, Γ') is included in E with weight $Q(\Gamma, \Gamma') = \breve{\Pi}_g(\Gamma)\breve{P}_g(\Gamma, \Gamma')$ whenever $\breve{P}_g(\Gamma, \Gamma') > 0$. A canonical path ensemble $\mathcal{T}$ is a collection of simple paths $\{\mathcal{T}_{x,y}\}$ in the graph $G(\mathcal{C})$, one between each ordered pair (x, y) of distinct vertices. As shown in Sinclair (1992), for any choice of canonical path $\mathcal{T}$, the spectral gap of the transition matrix $\breve{P}_g$ can be lower bounded by

$$Gap(\breve{P}_g) \geq \frac{1}{\rho(\mathcal{T})\ell(\mathcal{T})},$$

where $\ell(\mathcal{T})$ is the length of the longest path in $\mathcal{T}$, and $\rho(\mathcal{T})$ is the *path congestion* parameter defined by $\rho(\mathcal{T}) = \max_{(\Gamma,\Gamma')\in E(\mathcal{C})} \frac{1}{Q(\Gamma,\Gamma')} \sum_{\mathcal{T}_{x,y}\ni(\Gamma,\Gamma')} \breve{\Pi}_g(x)\breve{\Pi}_g(y)$.

In order to apply the approach, we need to construct an appropriate canonical path ensemble $\mathcal{T}$ in the discrete state space $\breve{\mathcal{G}}$. First, we construct a unique canonical path from any clustering Γ to the underlying true clustering Γ^*, where $\Gamma^* = \Gamma(Z^*)$. Suppose for any label assignment Z, we define a function $g : S_\alpha \to S_\alpha$ such that

$$g(Z) = \begin{cases} \underset{Z'\in\mathcal{B}(Z)\cap S_\alpha}{\operatorname{argmax}} \ \Pi(Z'|A), & \text{if } Z \notin \Gamma(Z^*), \\ Z, & \text{if } Z \in \Gamma(Z^*), \end{cases} \tag{32}$$

where

$$\mathcal{B}(Z) = \{Z' : d(Z', Z^*) = d(Z, Z^*) - 1, H(Z, Z') = 1\}.$$

We use $\mathcal{B}(Z) \cap S_\alpha$ to denote the set of available states that have fewer mistakes than the current state Z. By Lemma 42, $\mathcal{B}(Z) \cap S_\alpha$ is always non-empty for $Z \notin \Gamma(Z^*)$. Here, $g(Z)$ is the *optimal* state in $\mathcal{B}(Z)$ in the sense that $g(Z)$ maximizes the posterior distribution. Then, for any current state $\Gamma \in \breve{S}_\alpha$, we define the next state $\breve{g}(\Gamma)$ to be

$$\breve{g}(\Gamma) = \{g(Z) : Z \in \Gamma\}$$

Since for any $Z \in \Gamma$, $g(Z)$ gives the equivalent result, and thus $\breve{g}(\Gamma) \in \breve{S}_\alpha$ is well defined. Hence, the canonical path from any current state $\Gamma \neq \Gamma(Z^*)$ is a greedy path, and the number of mistakes keeps decreasing along the canonical path.

Second, we construct a unique canonical path between any two states Γ and $\widetilde{\Gamma}$, defined by $\mathcal{T}_{\Gamma,\widetilde{\Gamma}} = \mathcal{T}_{\Gamma,\Gamma^*} - \mathcal{T}_{\widetilde{\Gamma},\Gamma^*}$. The operations on simple paths are the same as defined in Abbe (2016). It is worth noting that the construction of the canonical path is data dependent, i.e., for different adjacency matrix A, the construction of the canonical path might be different.

Let $\Lambda(\Gamma) = \left\{\widetilde{\Gamma} \in \breve{\mathcal{G}} : \Gamma \in \mathcal{T}_{\widetilde{\Gamma},\Gamma^*}\right\}$ denote the set of all precedents states before Γ along the canonical path. Let $\mathcal{E} = \{(\Gamma, \Gamma') \in E(\mathcal{C}) : \Gamma \in \Lambda(\Gamma')\}$ denote the ordered adjacent pairs along the canonical path. It follows that

$$\begin{aligned} \rho(\mathcal{T}) &= \max_{(\Gamma,\Gamma')\in E(\mathcal{C})} \frac{1}{Q(\Gamma,\Gamma')} \sum_{\mathcal{T}_{x,y}\ni(\Gamma,\Gamma')} \breve{\Pi}_g(x)\breve{\Pi}_g(y) \\ &\leq \max_{(\Gamma,\Gamma')\in\mathcal{E}} \frac{1}{Q(\Gamma,\Gamma')} \sum_{x\in\Lambda(\Gamma),y\in\breve{\mathcal{G}}} \breve{\Pi}_g(x)\breve{\Pi}_g(y) \\ &= \max_{(\Gamma,\Gamma')\in\mathcal{E}} \frac{\breve{\Pi}_g(\Lambda(\Gamma))}{Q(\Gamma,\Gamma')} = \max_{\Gamma\in\breve{\mathcal{G}}} \frac{\breve{\Pi}_g(\Lambda(\Gamma))}{Q(\Gamma,\breve{g}(\Gamma))}, \end{aligned}$$

where we simply take maximum only over all states in the discrete space $\breve{\mathcal{G}}$. By the definition of Algorithm 1 and the lazy transition matrix, $Q(\Gamma, \Gamma')$ can be expressed as

$$Q(\Gamma, \breve{g}(\Gamma)) = \breve{\Pi}_g(\Gamma)\breve{P}_g(\Gamma, \breve{g}(\Gamma)) = \frac{1}{2(K-1)n} \min\left\{\breve{\Pi}_g(\Gamma), \breve{\Pi}_g(\breve{g}(\Gamma))\right\}.$$

It leads to the bound for the *congestion parameter* as

$$\begin{aligned}
\rho(\mathcal{T}) &\leq 2(K-1)n \max_{\Gamma\in\breve{\mathcal{G}}} \frac{\breve{\Pi}_g(\Lambda(\Gamma))}{\min\left\{\breve{\Pi}_g(\Gamma), \breve{\Pi}_g(\breve{g}(\Gamma))\right\}} \\
&= 2(K-1)n \max_{\Gamma\in\breve{\mathcal{G}}} \frac{\breve{\Pi}(\Lambda(\Gamma)|A)}{\min\left\{\breve{\Pi}(\Gamma|A), \breve{\Pi}(\breve{g}(\Gamma)|A)\right\}} \\
&= 2(K-1)n \max_{\Gamma\in\breve{\mathcal{G}}} \left\{\frac{\breve{\Pi}(\Lambda(\Gamma)|A)}{\breve{\Pi}(\Gamma|A)} \cdot \max\left\{1, \frac{\breve{\Pi}(\Gamma|A)}{\breve{\Pi}(\breve{g}(\Gamma)|A)}\right\}\right\},
\end{aligned}$$

where $\breve{\Pi}(\Gamma|A) = \sum_{Z\in\Gamma} \widetilde{\Pi}(Z|A)$ for all $\Gamma \in \breve{S}_\alpha$.

Lemma 17 *Recall that $g(Z)$ is the next optimal state of Z defined in* (32)*, and $\mathcal{G}(\gamma_0)$ is defined in* (30)*. Suppose Z_0 is given with $\ell(Z_0, Z^*) \leq \gamma_0$ where γ_0 satisfies Condition B, D, or E. Suppose ξ satisfies Condition C, D or E. Then, we have*

$$\max_{Z\in\mathcal{G}(\gamma_0)} \frac{\widetilde{\Pi}(Z|A)}{\widetilde{\Pi}(g(Z)|A)} \leq \exp(-C\bar{n}I)$$

for some constant $C > 1 - \varepsilon_0$ with probability at least $1 - C_1 n^{-C_2}$.

The proof of Lemma 17 is deferred to Section 5.6 and Section 5.7. By Lemma 17 and by permutation symmetry, we have

$$\max_{\Gamma\in\breve{\mathcal{G}}} \frac{\breve{\Pi}(\Gamma|A)}{\breve{\Pi}(\breve{g}(\Gamma)|A)} = \max_{Z\in\mathcal{G}} \frac{\widetilde{\Pi}(Z|A)}{\widetilde{\Pi}(g(Z)|A)} \leq \exp(-C\bar{n}I) \leq \exp(-C\bar{n}I),$$

for some constant $C > 1 - \varepsilon_0$ with high probability. Denote $\widetilde{m} = n\max\{\gamma_0, n^{-\tau}\} + \log^2 n$ for simplicity, and it follows that

$$\begin{aligned}
\max_{\Gamma\in\breve{\mathcal{G}}} \frac{\breve{\Pi}(\Lambda(\Gamma)|A)}{\breve{\Pi}(\Gamma|A)} \leq& 1 + \max_{m\leq\widetilde{m}} \sum_{l=1}^{\widetilde{m}-m} \binom{n-m}{l} (K-1)^l \exp(-Cl\bar{n}I) \\
\leq& 1 + C'n\exp(-C\bar{n}I),
\end{aligned}$$

for some constants C' and $C > 1 - \varepsilon_0$. Then, we have

$$\begin{aligned}
\rho(\mathcal{T}) &\leq 2(K-1)n \max_{\Gamma\in\breve{\mathcal{G}}} \left\{\frac{\breve{\Pi}(\Lambda(\Gamma)|A)}{\breve{\Pi}(\Gamma|A)} \cdot \max\left\{1, \frac{\breve{\Pi}(\Gamma|A)}{\breve{\Pi}(g(\Gamma)|A)}\right\}\right\} \\
&\leq 2(K-1)n(1 + C'n\exp(-C\bar{n}I)).
\end{aligned}$$

Furthermore, since the canonical path is defined within $\breve{\mathcal{G}}$, we can upper bound the length of the longest path by

$$\ell(\mathcal{T}) \leq 2n \max\{\gamma_0, n^{-\tau}\} + 2\log^2 n.$$

Recall that $\Gamma_0 = \Gamma(Z_0)$. By Lemma 15 and Lemma 16, together with (28) and the strong consistency property of $\breve{\Pi}_g(\cdot|A)$, we have that for any constant $\varepsilon \in (0,1)$,

$$\|\breve{P}^T(\Gamma_0, \cdot) - \breve{\Pi}(\cdot|A)\|_{\mathsf{TV}} \leq \varepsilon, \tag{33}$$

holds for any

$$T \geq 4(K-1)n^2 \max\{\gamma_0, n^{-\tau}\} \left(-\xi \log \Pi(Z_0|A) + \log \varepsilon^{-1}\right)(1 + o(1)), \tag{34}$$

for large n with probability at least $1 - C_3 n^{-C_4}$ for some constants C_3, C_4, where $\Pi^\xi(Z_0|A) \leq \breve{\Pi}_g(\Gamma_0|A)$ always holds. Finally, if $\mathbb{P}\{\ell(Z_0, Z^*) \leq \gamma_0\} \geq 1 - \eta$, then the conclusions of Theorem 5 and Theorem 11 can be obtained by a simple union bound argument.

5.2.5 Coupling

We require T to be at most a polynomial of n so that Lemma 15 holds. Thus, the previous total variation bound (33) holds only for $T \leq \text{poly}(n)$. In order to bound the mixing time defined in (15), we further use coupling approach to show the total variation bound holds for any $t \geq T$.

We call a probability measure w over $\Omega \times \Omega$ is a coupling of (u, v) if its two marginals are u and v respectively. Before the proof, we first state the following lemma to relate the total variation to the coupling.

Lemma 18 (Proposition 4.7 in Levin and Peres (2017)) *For any coupling w of (u, v), if the random variables (X, Y) is distributed according to w, then we have*

$$\|u - v\|_{\mathsf{TV}} \leq \mathbb{P}\{X \neq Y\}.$$

Back to our problem, in order to upper bound $\|\breve{P}^t(\Gamma_0, \cdot) - \breve{\Pi}(\cdot|A)\|_{\mathsf{TV}}$ for any $t \geq T$, we first create a coupling of these two distributions as follows. Consider two copies of the Markov chain X_t and Y_t both with the transition matrix $\breve{P}$:

- Let $X_0 = \Gamma_0$, and $Y_0 \sim \breve{\Pi}(\cdot|A)$.
- If $X_t \neq Y_t$, then sample X_{t+1} and Y_{t+1} independently according to $\breve{P}(X_t, \cdot)$ and $\breve{P}(Y_t, \cdot)$ respectively.
- If $X_t = Y_t$, then sample $X_{t+1} \sim \breve{P}(X_t, \cdot)$ and set $Y_{t+1} = X_{t+1}$.

Thus, it is obviously that for any $t \geq 1$, $Y_t \sim \breve{\Pi}(\cdot|A)$, and $X_t \sim \breve{P}^t(\Gamma_0, \cdot)$. Set $T = 4Kn^2 \max\{\gamma_0, n^{-\tau}\} \left(-\xi \log \breve{\Pi}(Z_0|A) + \log \varepsilon^{-1}\right)(1 + o(1))$ defined in (34). By Lemma 18 and (34), we have for any $t \geq T$,

$$\begin{aligned}\|\breve{P}^t(\Gamma_0,\cdot) - \breve{\Pi}(\cdot|A)\|_{\mathsf{TV}} &\leq \mathbb{P}\{X_t \neq Y_t\} \leq \mathbb{P}\{X_T \neq Y_T\} \\ &= 1 - \mathbb{P}\{X_T = Y_T\} \\ &\leq 1 - \mathbb{P}\{X_T = Y_T = \Gamma^*\} \\ &\leq 2 - \mathbb{P}\{X_T = \Gamma^*\} - \mathbb{P}\{Y_T = \Gamma^*\}.\end{aligned}$$

By (33), we have

$$\|\breve{P}^T(\Gamma_0,\cdot) - \breve{\Pi}(\cdot|A)\|_{\mathsf{TV}} = \max_S \left|\breve{P}^T(\Gamma_0,S) - \breve{\Pi}(S|A)\right| \leq \varepsilon$$

with high probability. Together with the strong consistency result, it yields

$$\begin{aligned}\|\breve{P}^t(\Gamma_0,\cdot) - \breve{\Pi}(\cdot|A)\|_{\mathsf{TV}} &\leq 2 - \mathbb{P}\{X_T = \Gamma^*\} - \mathbb{P}\{Y_T = \Gamma^*\} \\ &\leq 1 - \left(\breve{\Pi}(\Gamma^*|A) - \varepsilon\right) - \breve{\Pi}(\Gamma^*|A) \\ &\leq \varepsilon(1 + o(1)),\end{aligned}$$

with probability at least $1 - Cn^{-C'}$ for some constants C, C'. Here, the high probability statement is with respect to the data generation process, i.e., adjacency matrix A.

To combine, we reach the result that for any constant $\varepsilon \in (0,1)$,

$$\tau_\varepsilon(Z_0) \leq 4Kn^2 \max\left\{\gamma_0, n^{-\tau}\right\} \cdot \left(\xi \log \Pi(Z_0|A)^{-1} + \log(\varepsilon^{-1})\right),$$

with high probability where $\tau_\varepsilon(Z_0)$ is defined in (15).

5.3 Proof of Lemma 15

For any state $Z \in S_\alpha$, we define

$$\begin{aligned}\mathcal{N}(Z) &= \left\{Z' : H(Z', Z) = 1\right\}, \\ \mathcal{A}(Z) &= \left\{Z' \in \mathcal{N}(Z) : d(Z', Z^*) = d(Z, Z^*) + 1\right\}, \\ \mathcal{B}(Z) &= \left\{Z' \in \mathcal{N}(Z) : d(Z', Z^*) = d(Z, Z^*) - 1\right\},\end{aligned} \tag{35}$$

where $\mathcal{N}(Z)$ denotes the neighborhood states of Z with only one sample classified differently, and $\mathcal{A}(Z)$ (resp. $\mathcal{B}(Z)$) denotes the set of states with more mistakes (resp. fewer mistakes) in the neighbor. We further define

$$\begin{aligned}p_m(Z) = P(Z, \mathcal{A}(Z)) &= \frac{1}{2(K-1)n} \sum_{Z' \in \mathcal{A}(Z) \cap S_\alpha} \min\left\{1, \frac{\widetilde{\Pi}_g(Z'|A)}{\widetilde{\Pi}_g(Z|A)}\right\}, \\ q_m(Z) = P(Z, \mathcal{B}(Z)) &= \frac{1}{2(K-1)n} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \min\left\{1, \frac{\widetilde{\Pi}_g(Z'|A)}{\widetilde{\Pi}_g(Z|A)}\right\},\end{aligned} \tag{36}$$

where $p_m(Z), q_m(Z)$ are the probabilities of Z jumping to states with the number of mistakes equal to $m+1$, $m-1$ respectively. We have the following lemma to bound the ratio of $p_m(Z)$ and $q_m(Z)$ for any $Z \in \mathcal{G}$. Recall that $\mathcal{G} = \mathcal{G}(\gamma_0)$ is defined in (30).

Lemma 19 *Suppose γ_0 satisfies Condition B, D, or E. Let τ be any sufficiently small constant τ, and denote $\mathcal{G}^* = \{Z : n^{-\tau} \leq \ell(Z, Z^*) \leq \max\{\gamma_0, n^{-\tau}\} + (\log n)^2/n\}$. Then, we have*

$$\mathbb{P}\left\{\max_{Z\in\mathcal{G}^*}\frac{p_m(Z)}{q_m(Z)} \geq \eta\right\} \leq \exp\left(-n^{1-\tau}\right)$$

for some small constant $4\tau < \eta < 1$. The probability is with respect to the data-generating process, i.e., the adjacency matrix A.

The proof of Lemma 19 is deferred to Section 5.8. We take the τ in Lemma 19 to be the same as τ defined in $\mathcal{G}$. In order to show that the Markov chain will stay in $\mathcal{G}$ with high probability, we transform the original problem into an one dimensional random walk problem. Lemma 19 shows that the probability ratio of the one dimensional random walk on the region $\mathcal{G}^*$ can be bounded with high probability. All the following analysis is conditioning on the adjacency matrix A such that the event

$$\mathcal{E}(A) = \left\{\max_{Z\in\mathcal{G}^*}\frac{p_m(Z)}{q_m(Z)} \leq \eta\right\}$$

happens. We construct the following three types of Markov chains in order to prove Lemma 15.

Type I MarKov chain. Consider a particle starting at the initial position u on the x-axis where $0 < u < b$ at time $t = 0$, and it moves one unit to the left, to the right, or stay at the current position at time $t = 1, 2, \ldots$ with probability q_t, p_t, or $1 - q_t - p_t$, where $\eta > p_t/q_t$ for all time t. It stops once it reaches the left or the right boundary, and we are interested in the probability of its stopping at the boundary b or the boundary 0.

Suppose the position of the particle at time t is X_t, and $X_{t+1} = X_t + \xi_t$, where ξ_t follows the distribution

$$\mathbb{P}\{\xi_t = 1\} = p_t, \quad \mathbb{P}\{\xi_t = -1\} = q_t, \quad \mathbb{P}\{\xi_t = 0\} = 1 - p_t - q_t.$$

We define $Y_t = \exp((b - X_t)\log\eta)$. It is easy to verify that the stochastic process Y_t is a super-martingale, due to the fact that

$$\begin{aligned}\mathbb{E}\left(Y_{t+1}|Y_t\right) &= \exp((b - X_t)\log\eta)\cdot\mathbb{E}\left(\exp(-\xi_t\log\eta)\right)\\ &= \exp((b - X_t)\log\eta)\cdot(1 - p_t - q_t + p_t/\eta + q_t\cdot\eta)\\ &\leq Y_t.\end{aligned}$$

Let $\tau = \min\{t \geq 1 : X_t = b \text{ or } X_t = 0\}$, and τ is *stopping time* of this random walk. It is evident that $|Y_{t\wedge\tau}| \leq 1$ since $X_\tau < b$ for all time t. By Doob's optional stopping time theorem, it follows that $\mathbb{E}\left(Y_\tau\right) \leq \mathbb{E}\left(Y_0\right)$, i.e.,

$$\mathbb{P}\{X_\tau = b\} \leq \mathbb{P}\{X_\tau = 0\}\cdot\eta^b + \mathbb{P}\{X_\tau = b\} \leq \eta^{b-u}, \tag{37}$$

where the first inequality holds since $\mathbb{P}\{X_\tau = 0\} \geq 0$, and the second inequality is due to Doob's optional stopping time theorem. By (37), the probability of the particle reaching boundary b first is upper bounded by η^{b-u}, where u is the starting position. Let $P_u^b = \mathbb{P}\{X_0 = u, X_\tau = b\}$ for $u \in (0, b)$ denote the probability of starting at u and stopping at b. Then, we have $P_1^b \le \eta^{b-1}$.

Now suppose a particle starts at 0, i.e., $X_0 = 0$. It moves to the right or stay at the current position with some fixed probability p_0 or $1 - p_0$. Let $P_0^0 = \mathbb{P}\{X_0 = 0,\ X_\tau = 0\}$, and $P_0^b = \mathbb{P}\{X_0 = 0,\ X_\tau = b\}$, where τ is the stopping time as defined before. Then, we have that

$$P_0^0 = p_0 \cdot P_1^0 + (1 - p_0), \quad P_0^b = p_0 \cdot P_1^b, \quad P_0^0 + P_0^b = 1. \tag{38}$$

Type II Markov chain. We now define another Markov chain that is similar to the previous one. Consider a particle starting at position 0 at time 0, and follows the same updating rule as the previous chain but different stopping rule. We use W_t to denote the position of the particle at time t. The particle will only stop when it reaches the boundary b. When it is at the position 0, it still moves to the right or stay at 0 with fixed probability p_0 or $1 - p_0$ (the same probability as defined in Type I Markov chain). Thus, this newly defined Markov chain is a *reflected random walk*.

It is worth noting that the Type II Markov chain can always be decomposed into several Type I Markov chains. We use τ_W to denote the stopping time of Type II Markov chain, defined by $\tau_W = \min\{t \ge 1 : W_t = b\}$.

Type III Markov chain. Now return to our original problem and construct Type III Markov chain. Let $m_0 = n\ell(Z_0, Z^*)$, and we use $H_t = n\ell(Z_t, Z^*) - m_0$ to denote the position of the particle, where Z_t is the label assignment after t steps. The state space is all integers between $-m_0$ and b, where we take $b = \max\{0, n^{1-\tau} - m_0\} + \log^2 n$. When $H_t \in (0, b)$, the particle moves to the left, to the right, or stay at the current position with probability q_t, p_t, or $1 - q_t - p_t$, which is the same as the Type II chains. The particle will only stop when it reaches the boundary b. The stopping time of Type III Markov chain is defined by $\tau_H = \min\{t \ge 1 : H_t = b\}$.

Proof [Proof of Lemma 15] Recall that $b = \max\{0, n^{1-\tau} - m_0\} + \log^2 n$. In order to prove Lemma 15, it is equivalent to show that, for any T that is a polynomial of n, the event $\{H_t < b,\ t \le T\}$ happens with high probability, i.e., $\{\tau_H > T\}$ happens with high probability. By the definition of Type II and III Markov chains we have that

$$\mathbb{P}\{\tau_H \le T\} \le \mathbb{P}\{\tau_W \le T\}.$$

The above inequality holds since $H_0 = W_0$, $H_t \ge 0$ for all time t, and the updating rule of H_t and W_t are exactly the same when $W_t, H_t \in (0, b)$.

We now connect the Type II Markov chain with multiple Type I chains. The event $\{\tau_W \le T\}$ means that the particle starts at 0 and reaches the boundary b within T steps, and it can be written as

$$\{\tau_W \le T\} = \bigcup_{k=1}^{T} \left\{ \left[\bigcap_{i=1}^{k-1} \left\{ X_0^{(i)} = 0,\ X_{\tau_i}^{(i)} = 0 \right\} \right] \bigcap \left\{ X_0^{(k)} = 0,\ X_{\tau_i}^{(k)} = b \right\} \bigcap \left\{ \sum_{i=1}^{k} \tau_i \le T \right\} \right\}, \tag{39}$$

where we use $X^{(i)}$ to denote the ith Type I Markov chain, and τ_i is the stopping time of $X^{(i)}$. Note $X^{(i)}$ is independent with $X^{(j)}$ for $i \neq j$. The right hand side of (39) can be interpreted as that the particle reaches the boundary 0 for $k-1$ times with $k \leq T$ before reaching the boundary b, and the total number of steps is less than T. Therefore, it follows directly that

$$\begin{aligned}
\mathbb{P}\left\{\tau_W \leq T\right\} &= \mathbb{P}\left\{\bigcup_{k=1}^{T}\left\{\left[\bigcap_{i=1}^{k-1}\left\{X_0^{(i)}=0,\ X_{\tau_i}^{(i)}=0\right\}\right]\bigcap\left\{X_0^{(k)}=0,\ X_{\tau_i}^{(k)}=b\right\}\bigcap\left\{\sum_{i=1}^{k}\tau_i \leq T\right\}\right\}\right\} \\
&\leq \sum_{k=1}^{T}\mathbb{P}\left\{\left[\bigcap_{i=1}^{k-1}\left\{X_0^{(i)}=0,\ X_{\tau_i}^{(i)}=0\right\}\right]\bigcap\left\{X_0^{(k)}=0,\ X_{\tau_i}^{(k)}=b\right\}\bigcap\left\{\sum_{i=1}^{k}\tau_i \leq T\right\}\right\} \\
&\leq \sum_{k=1}^{T}\mathbb{P}\left\{\left[\bigcap_{i=1}^{k-1}\left\{X_0^{(i)}=0,\ X_{\tau_i}^{(i)}=0\right\}\right]\bigcap\left\{X_0^{(k)}=0,\ X_{\tau_i}^{(k)}=b\right\}\right\} \\
&= \sum_{k=1}^{T}(P_0^0)^{k-1}P_0^b.
\end{aligned}$$

The first inequality holds by a union bound. The third inequality is by the independence. By (38), we have that

$$\begin{aligned}
\sum_{k=1}^{T}P_0^b(P_0^0)^{k-1} &= 1-(P_0^0)^T \\
&\leq -T\log P_0^0 \leq T\cdot\frac{1-P_0^0}{P_0^0} = T\cdot\frac{p_0\cdot P_1^b}{p_0\cdot P_1^0+1-p_0} \\
&\leq T\cdot\frac{P_1^b}{P_1^0} \leq \exp\left(\log\frac{\eta^{b-1}}{1-\eta^{b-1}}+\log T\right).
\end{aligned}$$

Since T is a polynomial of n, and $b \gg \log n$, then it follows that

$$\mathbb{P}\left\{\tau_H \leq T\right\} \leq \mathbb{P}\left\{\tau_W \leq T\right\} \leq \exp\left(-(1-o(1))b\log\frac{1}{\eta}\right).$$

Thus, based on the result of Lemma 19, for any given initial label assignment Z_0 with $\ell(Z_0, Z^*) \leq \gamma_0$ where γ_0 satisfies Condition B, D, or E, we have that in any polynomial running time, the number of mistakes is upper bounded by

$$m \leq m_0 + b = \max\{m_0, n^{1-\tau}\} + \log^2 n \leq n\max\{\gamma_0, n^{-\tau}\} + \log^2 n,$$

with probability at least $1-\exp(-\log^2 n)$. ■

5.4 Some preparations before the proofs of Lemma 13 and Lemma 17

In this section, we will define some events and introduce some quantities to simplify the main proof.

5.4.1 Basic events

For any $Z \in S_\alpha$, recall that $O_{ab}(Z) = \sum_{i,j} A_{ij}\mathbb{I}\{Z_i = a, Z_j = b\}$, and define $X_{ab}(Z) = O_{ab}(Z) - \mathbb{E}[O_{ab}(Z)]$ for any $a, b \in [K]$. Let $X(Z)$ denote a $K \times K$ matrix with its (a,b)th element equal to $X_{ab}(Z)$ for any $a, b \in [K]$. For any positive sequences $\bar{\varepsilon} = \bar{\varepsilon}_n$, $\gamma = \gamma_n$, and $\theta = \theta_n$ satisfying that $\bar{\varepsilon}, \gamma, \theta \to 0, \bar{\varepsilon}^2 nI \to \infty$, and $\theta^2 \gamma nI \to \infty$, consider the following events:

$$\begin{aligned}
\mathcal{E}_1(\bar{\varepsilon}) &= \left\{ \max_{Z \in S_\alpha} \|X(Z)\|_\infty \le \bar{\varepsilon} n^2 (p-q) \right\}, \\
\mathcal{E}_2 &= \left\{ \max_{Z \in S_\alpha} \|X(Z) - X(Z^*)\|_\infty - Cmn(p-q)/K \le 0 \right\}, \\
\mathcal{E}_3(\bar{\varepsilon}) &= \left\{ \max_{Z \in S_\alpha} \|X(Z) - X(Z^*)\|_\infty \le \bar{\varepsilon} n^2 (p-q) \right\}, \\
\mathcal{E}_4(\gamma, \theta) &= \left\{ \max_{Z \in S_\alpha : m > \gamma n} \|X(Z) - X(Z^*)\|_\infty \le \theta mn(p-q) \right\}, \\
\mathcal{E}_5 &= \left\{ \max_{Z \in S_\alpha} \frac{1}{|\mathcal{B}(Z) \cap S_\alpha|} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \sum_{a \le a'} \left| X_{aa'}(Z) - X_{aa'}(Z') \right| \le Cn(p-q) \right\}, \\
\mathcal{E}_6(\gamma, \theta) &= \left\{ \max_{Z \in S_\alpha : m > \gamma n} \frac{1}{|\mathcal{B}(Z) \cap S_\alpha|} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \sum_{a \le a'} \left| X_{aa'}(Z) - X_{aa'}(Z') \right| \le \theta n(p-q) \right\},
\end{aligned} \tag{40}$$

where $\mathcal{B}(Z)$ is defined in (35). We may also use $\mathcal{E}_1$ to denote $\mathcal{E}_1(\bar{\varepsilon})$ for simplicity, and such simplification also applies to any other event. Denote

$$\mathcal{E} = \mathcal{E}(\bar{\varepsilon}, \gamma, \theta) = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4 \cap \mathcal{E}_5 \cap \mathcal{E}_6. \tag{41}$$

By Lemmas 27, 29, 30, 31, 32, and 33, it follows that for any $\bar{\varepsilon}, \gamma, \theta$ satisfying the conditions,

$$\mathbb{P}\{\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)\} \ge 1 - n\exp(-(1 - o(1))\bar{n}I).$$

5.4.2 Likelihood modularity

The posterior distribution is hard to deal with directly. Hence, we first analyze the performance of the likelihood modularity function, and then bound the difference between the likelihood modularity function and the posterior distribution to simplify the proof.

Likelihood modularity is first introduced in Bickel and Chen (2009), which takes the form as

$$Q_{LM}(Z, A) = \sum_{a \le b} n_{ab}(Z) \tau\left(\frac{O_{ab}(Z)}{n_{ab}(Z)}\right), \tag{42}$$

where $\tau(x) = x\log(x) + (1-x)\log(1-x)$. This criterion replaces the connectivity probabilities by maximum likelihood estimates. Instead of comparing the direct difference between the Bayesian expression and the likelihood modularity as in van der Pas et al. (2017), we have the following lemma to bound the relative difference.

Lemma 20 *Under the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$ defined in (41), we have*

$$\max_{Z \in S_\alpha} \left| \log \frac{\Pi(Z|A)}{\Pi(Z^*|A)} - (Q_{LM}(Z, A) - Q_{LM}(Z^*, A)) \right| \leq C_{LM}$$

for some constant C_{LM} only depending on K, α, β.

The above lemma is rephrased and proved in Lemma 34.

5.4.3 Discrepancy matrix

For any label assignment $Z \in S_\alpha$, let R_Z be a *discrepancy matrix*, which takes the form of

$$R_Z(a, b) = \sum_{i=1}^{n} \mathbb{I}\left\{Z_i = a, Z_j^* = b\right\}, \quad a, b \in [K], \tag{43}$$

where $R_Z(a, b)$ is the number of samples misclassified to group a but actually from group b based on the true label assignment. Note that the true label assignment is only unique up to a label permutation, and thus we always permute the rows of R_Z to minimize the off diagonal sum. Later we write $R_Z(k, l)$ as R_{kl} for simplicity.

Using the discrepancy matrix R_Z, we have

$$\begin{aligned}
\mathbb{E}\left[O_{ab}(Z)\right] &= \mathbb{E}\left[\sum_{i,j} A_{ij} \mathbb{I}\left\{Z_i = a, Z_j = b\right\}\right] = (RBR^T)_{ab}, \quad \text{for } a \neq b \in [K], \\
\mathbb{E}\left[O_{aa}(Z)\right] &= \mathbb{E}\left[\sum_{i<j} A_{ij} \mathbb{I}\left\{Z_i = Z_j = a\right\}\right] = \frac{1}{2}\left((RBR^T)_{aa} - \sum_k B_{kk} R_{ak}\right), \quad \text{for } a \in [K].
\end{aligned} \tag{44}$$

5.5 Proof of Lemma 13

Before proving the lemma, we need to present some notations that will be frequently used:

$$\begin{aligned}
O_s(Z) &= \sum_{a \in [K]} O_{aa}(Z) = \sum_{i<j} A_{ij} \mathbb{I}\left\{Z_i = Z_j\right\}, \\
n_s(Z) &= \sum_{a \in [K]} n_{aa}(Z) = \sum_{i<j} \mathbb{I}\left\{Z_i = Z_j\right\}, \\
\boldsymbol{\Delta} \widetilde{O}_{ab} &= O_{ab}(Z) - O_{ab}(Z^*), \quad \boldsymbol{\Delta} \widetilde{O}_s = \sum_{a \in [K]} \boldsymbol{\Delta} \widetilde{O}_{aa}, \\
\boldsymbol{\Delta} \widetilde{n}_{ab} &= n_{ab}(Z) - n_{ab}(Z^*), \quad \boldsymbol{\Delta} \widetilde{n}_s = \sum_{a \in [K]} \boldsymbol{\Delta} \widetilde{n}_{aa},
\end{aligned}$$

and we may write $n_{ab}(Z^*)$, $O_{ab}(Z^*)$ as n_{ab}, O_{ab} for simplicity.

Proof [Proof of Lemma 13] For any positive sequences $\gamma = \gamma_n$ and $\theta = \theta_n$ such that $\gamma \to 0$, $\gamma^2 nI \to \infty$, and $\theta^2 \gamma nI \to \infty$, we can construct the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$ defined in (41) by setting $\bar{\varepsilon} = \gamma$, and perform analysis on a large mistake region and a small mistake region separately.

Small mistake region. For $m \le \gamma n$, by some calculations, it follows that

$$\begin{aligned}
&Q_{LM}(Z,A) - Q_{LM}(Z^*,A) \\
=&\log \Pi_0(Z|A) - \log \Pi_0(Z^*|A) - \sum_{a\le b} n_{ab} \cdot D\left(\frac{O_{ab}}{n_{ab}} \middle\| \frac{O_{ab}(Z)}{n_{ab}(Z)}\right) + \\
&\underbrace{\sum_{a\le b} \boldsymbol{\Delta}\widetilde{O}_{ab}\left(\log\frac{O_{ab}(Z)}{n_{ab}(Z)} - \log B_{ab}\right) + (\boldsymbol{\Delta}\widetilde{n}_{ab} - \boldsymbol{\Delta}\widetilde{O}_{ab})\left(\log\frac{n_{ab}(Z)-O_{ab}(Z)}{n_{ab}(Z)} - \log(1-B_{ab})\right)}_{(Error)},
\end{aligned} \tag{45}$$

where by (24)

$$\log \Pi_0(Z|A) - \log \Pi_0(Z^*|A) = 2t^*\left(\boldsymbol{\Delta}\widetilde{O}_s - \lambda^* \boldsymbol{\Delta}\widetilde{n}_s\right), \tag{46}$$

and

$$\lambda^* = \log\frac{1-q}{1-p} \Big/ \log\frac{p(1-q)}{q(1-p)}, \qquad t^* = \frac{1}{2}\log\frac{p(1-q)}{q(1-p)}. \tag{47}$$

Under the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$, by Lemma 37, we have $(Error) \le C\gamma mnI$ for some constant C. Hence, for any fixed $Z \in S_\alpha$, by Lemma 20, we have that

$$\begin{aligned}
&\log\frac{\Pi(Z|A)}{\Pi(Z^*|A)} - \log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)} \\
\le& Q_{LM}(Z,A) - Q_{LM}(Z^*,A) + C_{LM} - \log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)} \\
\le& C\gamma mnI + C_{LM}.
\end{aligned}$$

Thus, there exists some constant C_2 such that under the event $\mathcal{E}$,

$$\max_{Z\in S_\alpha : m<\gamma n} \log\frac{\Pi(Z|A)}{\Pi(Z^*|A)} - \log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)} - C_2\gamma mnI \le 0, \tag{48}$$

which proves the second statement in Lemma 13.

Large mistake region. For $m > \gamma n$, we have that

$$\begin{aligned}
Q_{LM}(Z,A) - Q_{LM}(Z^*,A) &= \sum_{a\le b} n_{ab}(Z)\tau\left(\frac{O_{ab}(Z)}{n_{ab}(Z)}\right) - n_{ab}(Z^*)\tau\left(\frac{O_{ab}(Z^*)}{n_{ab}(Z^*)}\right) \\
&= (G(Z) + \Delta(Z)) - (G(Z^*) + \Delta(Z^*)),
\end{aligned} \tag{49}$$

where we write

$$\begin{aligned}
G(\cdot) &= \sum_{a\le b} n_{ab}(\cdot)\tau\left(\frac{\mathbb{E}\left[O_{ab}(\cdot)\right]}{n_{ab}(\cdot)}\right), \\
\Delta(\cdot) &= \sum_{a\le b} n_{ab}(\cdot)\left(\tau\left(\frac{O_{ab}(\cdot)}{n_{ab}(\cdot)}\right) - \tau\left(\frac{\mathbb{E}\left[O_{ab}(\cdot)\right]}{n_{ab}(\cdot)}\right)\right).
\end{aligned}$$

Let $\widetilde{B}_{ab} = \mathbb{E}\left[O_{ab}(Z)\right]/n_{ab}(Z)$ for any $a, b \in [K]$. By (44), it follows that

$$
\begin{aligned}
2G(Z) =& 2\sum_{a\le b} n_{ab}(Z)\tau\left(\widetilde{B}_{ab}\right)\\
=& 2\sum_{a\le b} \mathbb{E}\left[O_{ab}(Z)\right]\log \widetilde{B}_{ab} + (n_{ab}(Z) - \mathbb{E}\left[O_{ab}(Z)\right])\log(1-\widetilde{B}_{ab})\\
=& \sum_{a,b,k,l} R_{ak}R_{bl}(B_{kl}\log \widetilde{B}_{ab} + (1-B_{kl})\log(1-\widetilde{B}_{ab})) - \sum_{a} n_a(Z)(p\log \widetilde{B}_{aa} + (1-p)\log(1-\widetilde{B}_{aa})).
\end{aligned}
$$

Then, we have that

$$
\begin{aligned}
&2G(Z) - 2G(Z^*)\\
=&\sum_{a,b,k,l} R_{ak}R_{bl}(B_{kl}\log \widetilde{B}_{ab} + (1-B_{kl})\log(1-\widetilde{B}_{ab})) - \sum_{a,b,k,l} R_{ak}R_{bl}(B_{kl}\log B_{kl} + (1-B_{kl})\log(1-B_{kl}))\\
&-\sum_{a} n_a(Z)\left((p\log \widetilde{B}_{aa} + (1-p)\log(1-\widetilde{B}_{aa})) - (p\log p + (1-p)\log(1-p))\right)\\
=&-\sum_{a,b,k,l} R_{ak}R_{bl}D\left(B_{kl}\middle\|\widetilde{B}_{ab}\right) + \sum_{a} n_a(Z)D\left(p\middle\|\widetilde{B}_{aa}\right).
\end{aligned}
\tag{50}
$$

By Lemma 38 and Lemma 40 that bound the above two terms separately, we have

$$2G(Z) - 2G(Z^*) \le -CmnI,$$

for some constant C. Under the events $\mathcal{E}_1(\bar{\varepsilon}), \mathcal{E}_3(\bar{\varepsilon}), \mathcal{E}_4(\gamma, \theta)$ in (40), by Lemma 34 and Lemma 41, we have that

$$\max_{Z\in S_\alpha}\left|\log\frac{\Pi(Z|A)}{\Pi(Z^*|A)} - (Q_{LM}(Z,A) - Q_{LM}(Z^*,A))\right| \le C_{LM},$$

and

$$\max_{Z\in S_\alpha: m>\gamma n} |\Delta(Z) - \Delta(Z^*)| \le \varepsilon mnI,$$

for some $\varepsilon \to 0$. Hence, it follows that there exists some constant C_1 such that

$$\mathbb{P}\left\{\max_{Z\in S_\alpha: m>\gamma n} \log\frac{\Pi(Z|A)}{\Pi(Z^*|A)} > -C_1 mnI\right\} \le 4\exp(-n).$$

Combining the result of two regions directly gives Lemma 13.

■

5.6 Proof of Lemma 17 with known connectivity probabilities

In order to distinguish from the case where probabilities are unknown, we use $\Pi_0(\cdot|A)$ to denote the posterior distribution in this case, and define $\widetilde{\Pi}_0(\cdot|A)$ as the scaled distribution proportional to $\Pi_0^\xi(\cdot|A)$. It suffices to prove the following lemma.

Lemma 21 *Recall that $g(Z)$ is the next state of Z. Suppose γ_0 satisfies Condition B, D, or E. Then, there exists some positive sequence $\gamma \to 0$ such that, with probability at least $1 - C_1 n^{-C_2}$,*

$$\frac{\Pi_0(Z|A)}{\Pi_0(g(Z)|A)} \leq \begin{cases} \exp\left(-\varepsilon \bar{n} I\right), & \text{if } m \leq \gamma n, \\ \exp\left(-4\left(1/K\alpha - \gamma_0\right) nI(1-o(1))\right), & \text{if } m > \gamma n, \end{cases}$$

holds uniformly for all $Z \in \mathcal{G}(\gamma_0)$ defined in (30). Here, ε is any constant satisfying $\varepsilon < 2\varepsilon_0$ with ε_0 defined in Condition E, and C_1, C_2 are two constants depending on ε. Furthermore, if ξ satisfies Condition E, then by choosing $\varepsilon \in ((1-\varepsilon_0)/\xi, 2\varepsilon_0)$, we have

$$\max_{Z \in \mathcal{G}(\gamma_0)} \frac{\widetilde{\Pi}_0(Z|A)}{\widetilde{\Pi}_0(g(Z)|A)} \leq \exp(-C\bar{n}I)$$

for some constant $C > 1 - \varepsilon_0$ with probability at least $1 - C_3 n^{-C_4}$.

Proof [Proof of Lemma 21] Recall the definitions of $\mathcal{A}(Z)$, $\mathcal{B}(Z)$, and $\mathcal{N}(Z)$ in (35). We introduce some notations first to simplify the proof. For any $a, b \in [K]$ and any two label assignments Z, Z', we write

$$\begin{aligned} &\boldsymbol{\Delta} O_{ab} = O_{ab}(Z) - O_{ab}(Z'), && \boldsymbol{\Delta} O_s = \sum_a \boldsymbol{\Delta} O_{aa}, \\ &\boldsymbol{\Delta} n_{ab} = n_{ab}(Z) - n_{ab}(Z'), && \boldsymbol{\Delta} n_s = \sum_a \boldsymbol{\Delta} n_{aa}, \\ &\Delta_n(Z, Z') = \boldsymbol{\Delta} O_s - \lambda^* \boldsymbol{\Delta} n_s, && \lambda^* = \log \frac{1-q}{1-p} \Big/ \log \frac{p(1-q)}{q(1-p)}. \end{aligned}$$

Suppose the current state is Z, and we randomly choose one misclassified sample from group a and move to its true group b. Denote the new state as Z'. It follows that $R_{Z'}(a,b) = R_Z(a,b) - 1$, and $R_{Z'}(b,b) = R_Z(b,b) + 1$. Write $R_Z(a,b)$ as R_{ab} for simplicity. Furthermore, let $\{x_l\}_{l \geq 1}, \{\widetilde{x}_l\}_{l \geq 1}$ be i.i.d. copies of Bernoulli(q) and $\{y_l\}_{l \geq 1}, \{\widetilde{y}_l\}_{l \geq 1}$ be i.i.d. copies of Bernoulli(p). We have that

$$\Delta_n(Z, Z') = \sum_{l=1}^{R_{aa} + \sum_{k \neq a,b} R_{ak}} (x_l - \lambda^*) - \sum_{l=1}^{R_{bb}} (y_l - \lambda^*) + \sum_{l=1}^{R_{ab}-1} (\widetilde{y}_l - \lambda^*) - \sum_{l=1}^{\sum_{k \neq b} R_{bk}} (\widetilde{x}_l - \lambda^*). \quad (51)$$

By (24), it directly follows that

$$\log \frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)} = 2t^* \Delta_n(Z, Z'), \quad t^* = \frac{1}{2} \log \frac{p(1-q)}{q(1-p)}. \quad (52)$$

For some positive sequence $\gamma = \gamma_n \to 0$ to be specified later, we can again divide S_α into two regions.

Small mistake region. For $m \le \gamma n$, we have

$$\mathbb{E}\left[\sqrt{\frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)}}\right] = \mathbb{E}\left[e^{t^*\Delta_n(Z,Z')}\right]$$

$$=\mathbb{E}\exp\left[t^*\left(\sum_{l=1}^{R_{aa}}(x_l-\lambda^*) - \sum_{l=1}^{R_{bb}}(y_l-\lambda^*)\right)\right] \tag{53}$$

$$\cdot\mathbb{E}\exp\left[t^*\left(\sum_{l=1}^{R_{ab}-1}(\widetilde{y}_l-\lambda^*) - \sum_{l=1}^{\sum_{k\neq b}R_{bk}}(\widetilde{x}_l-\lambda^*) + \sum_{l=1}^{\sum_{k\neq a,b}R_{ak}}(x_l-\lambda^*)\right)\right]. \tag{54}$$

Based on Proposition 5.1 in Zhang et al. (2016), for any positive integers n_1, n_2, we have

$$\mathbb{E}\text{exp}\left[t^*\left(\sum_{i=1}^{n_1}(x_i-\lambda^*) - \sum_{i=1}^{n_2}(y_i-\lambda^*)\right)\right] \le \exp\left(-\frac{(n_1+n_2)I}{2}\right),$$

and it leads to

$$(53) \le \exp\left(-\frac{(R_{aa}+R_{bb})I}{2}\right) \le \exp\left(-\frac{(n_a+n_b-m)I}{2}\right) \le \exp\left(-(1-c\gamma)\bar{n}I\right),$$

for some constant c. By Lemma 54, we have

$$(54) \le (\exp(C_0 I))^m = \exp(C_0 m I) \le \exp(C_0\gamma n I),$$

for some constant C_0. It follows that

$$\mathbb{E}\left[\sqrt{\frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)}}\right] \le \exp\left(-(1-C_1\gamma)\bar{n}I)\right) \tag{55}$$

for some constant C_1 depending on C_0. Let ε be any small constant satisfying $\varepsilon < 2\varepsilon_0$, and $w = \varepsilon\bar{n}I/2t^*$. By Lemma 42, since $\gamma < c_{\alpha,\beta}$, we have $\mathcal{B}(Z) \subset S_\alpha$. Then,

$$\begin{aligned}
\mathbb{P}\left\{\min_{Z'\in\mathcal{B}(Z)\cap S_\alpha}\Delta_n(Z,Z') \ge -w\right\} &\le \mathbb{P}\left\{\sum_{Z'\in\mathcal{B}(Z)}\Delta_n(Z,Z') \ge -mw\right\} \\
&= \mathbb{P}\left\{\exp\left(t^*\sum_{Z'\in\mathcal{B}(Z)}\Delta_n(Z,Z')\right) \ge \exp\left(-t^*mw\right)\right\} \\
&\le \mathbb{E}\left[\exp\left(t^*\sum_{Z'\in\mathcal{B}(Z)}\Delta_n(Z,Z')\right)\cdot\exp\left(t^*mw\right)\right] \\
&= \mathbb{E}\left[\exp\left(t^*\sum_{Z'\in\mathcal{B}(Z)}\Delta_n(Z,Z')\right)\cdot\exp\left(\varepsilon m\bar{n}I/2\right)\right],
\end{aligned} \tag{56}$$

The first inequality holds because the minimum is smaller than the average. The second inequality is due to Markov's inequality. We now proceed to bound (56) by $\exp(-(1-o(1))m\bar{n}I)$.

We first define set $\mathcal{C}(Z)=\{i: Z_i=Z_i^*\}$, which is the set of samples that are correctly classified. Thus, we have $|\mathcal{C}(Z)|=\sum_{a\in[K]}R_{aa}=n-m$. Suppose $Z'\in\mathcal{B}(Z)$ corrects kth sample from a misclassified group a, where $Z_k=a$, to its true group b, where $Z'_k=b$. Then, we must have $k\in[n]\setminus\mathcal{C}(Z)$, and by Lemma 43, we can rewrite

$$\begin{aligned}\Delta_n(Z,Z') &= \sum_{i\in[n]}(A_{ik}\mathbb{I}\left\{Z'_i=Z_k\right\}-\lambda^*)-\sum_{i\in[n]}(A_{ik}\mathbb{I}\left\{Z_i=Z'_k\right\}-\lambda^*)\\ &= \underbrace{\sum_{i\in\mathcal{C}(Z)}(A_{ik}\mathbb{I}\left\{Z'_i=Z_k\right\}-\lambda^*)-\sum_{i\in\mathcal{C}(Z)}(A_{ik}\mathbb{I}\left\{Z_i=Z'_k\right\}-\lambda^*)}_{(A_k)}\\ &+\underbrace{\sum_{i\notin\mathcal{C}(Z)}(A_{ik}\mathbb{I}\left\{Z'_i=Z_k\right\}-\lambda^*)-\sum_{i\notin\mathcal{C}(Z)}(A_{ik}\mathbb{I}\left\{Z_i=Z'_k\right\}-\lambda^*)}_{(B_k)}.\end{aligned}$$

Here, A_k, B_k correspond to summations in (53) and (54) respectively. We further have

$$\sum_{Z'\in\mathcal{B}(Z)}\Delta_n(Z,Z')=\sum_{k\in[n]\setminus\mathcal{C}(Z)}(A_k+B_k).$$

It is obvious that $A_k\perp A_j$ for $k,j\in[n]\setminus\mathcal{C}(Z)$, and $\sum_{k\in[n]\setminus\mathcal{C}(Z)}A_k$ can be written as the independent sum of the random variable A_{ij} for some $i\in[n]\setminus\mathcal{C}(Z)$ and $j\in C(Z)$. As for $\sum_{k\in[n]\setminus\mathcal{C}(Z)}B_k$, it is the summation of A_{ij} for some $i,j\in[n]\setminus\mathcal{C}(Z)$. For each random variable A_{ij}, the coefficient is at most 2 (since it can only be added twice or canceled out), and the total number of random variables is at most $\binom{m}{2}$. Hence, by the argument from (53) to (55), we can bound (56) by

$$(56)\leq\exp\left(-\left(1-C\gamma-\varepsilon/2\right)m\bar{n}I\right),$$

for some constant C. By the definition of $g(Z)$ defined in (32), we have

$$\begin{aligned}\mathbb{P}\left\{\max_{Z\in\mathcal{G}(\gamma_0):m\leq\gamma n}\frac{\Pi_0(Z|A)}{\Pi_0(g(Z)|A)}\geq\exp(-\varepsilon\bar{n}I)\right\} &= \mathbb{P}\left\{\max_{Z\in\mathcal{G}(\gamma_0):m\leq\gamma n}\min_{Z'\in\mathcal{B}(Z)}\Delta_n(Z,Z')\geq -w\right\}\\ &\leq\sum_{m=1}^{\gamma n}\binom{n}{m}(K-1)^m\exp\left(-\left(1-C\gamma-\varepsilon/2\right)m\bar{n}I\right)\\ &\leq\sum_{m=1}^{\gamma n}\left(enK\exp\left(-\left(1-C\gamma-\varepsilon/2\right)\bar{n}I\right)\right)^m\\ &= n\exp\left(-\left(1-\varepsilon/2\right)\bar{n}I(1-o(1))\right),\end{aligned}\tag{57}$$

where we require $\varepsilon<2\varepsilon_0$ in order for the last equation going to 0 as n tends to infinity.

Large mistake region. For $m > \gamma n$, recall that $\mathcal{G}(\gamma_0) \subset S_\alpha$, and S_α is defined in (4). If Z' corrects one sample from group a to group b, by (51), we have $\boldsymbol{\Delta} n_s = n_a' - n_b' - 1$. By 51, we have $\Delta_n(Z, Z') = \boldsymbol{\Delta} O_s - \lambda^* \boldsymbol{\Delta} n_s$. Let $\lambda = (p+q)/2$, $m_b = \sum_{k \neq b} R_{kb}$, $n_a' = n_a(Z)$, and $n_b' = n_b(Z)$ for simplicity. Thus, it follows that

$$\begin{aligned}
\mathbb{E}\left[\Delta_n(Z, Z')\right] &= \mathbb{E}\left[\boldsymbol{\Delta} O_s\right] - \lambda^* \boldsymbol{\Delta} n_s = \mathbb{E}\left[\boldsymbol{\Delta} O_s\right] - \lambda \boldsymbol{\Delta} n_s + (\lambda - \lambda^*) \boldsymbol{\Delta} n_s \\
&= -\frac{p-q}{2}(n_a' + n_b' - 2m_b - 2R_{ab}) + \frac{p+q-2\lambda^*}{2}(n_a' - n_b) - (p - \lambda^*) \\
&\leq -\frac{p-q}{2}(n_a' + n_b' - 2m) + \frac{p+q-2\lambda^*}{2}(n_a' - n_b') \\
&= -\frac{p-q}{2}(n_a' + n_b' - 2m - C_\lambda(n_a' - n_b')) \\
&= -\frac{p-q}{2}(n_b' + C_\lambda n_b' + (1 - C_\lambda) n_a' - 2m) \\
&\leq -\left(\frac{n}{K\alpha} - m\right)(p-q),
\end{aligned}$$

where $C_\lambda = 2(\lambda - \lambda^*)/(p-q)$. It is easy to verify that $C_\lambda \in (0,1)$, and C_λ tends to 1 (resp. tends to 0) when $(p-q)/p$ tends to 1 (resp. tends to 0). If γ_0 satisfies that $(1 - K\alpha\gamma_0)^2 nI \to \infty$, it follows that

$$\max_{Z \in \mathcal{G}(\gamma_0): m > \gamma n} \max_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \left(\mathbb{E}\left[\boldsymbol{\Delta} O_s\right] - \lambda^* \boldsymbol{\Delta} n_s\right) \leq -\left(\frac{1}{K\alpha} - \gamma_0\right) n(p-q)(1 - o(1)). \tag{58}$$

Denote $\delta_0 = 1/K\alpha - \gamma_0$ for simplicity. Then, it follows that (58) $\leq -\delta_0 n(p-q)(1-o(1))$. Since $\delta_0^2 nI \to \infty$, there exist some positive sequences γ and θ such that $\gamma, \theta \to 0$, $\theta^2 \gamma nI \to \infty$, and $\theta \ll \delta_0$. To be specific, we may take $\gamma = 1/(\delta_0 \sqrt{nI})$, $\theta = \delta/(\delta\sqrt{nI})^{1/4}$. Hence, we can construct the event $\mathcal{E}_6(\gamma, \theta)$ as defined in (40). Note that $X(Z) - X(Z^*) = \boldsymbol{\Delta} O(Z) - \mathbb{E}\left[\boldsymbol{\Delta} O(Z)\right]$. Under the event $\mathcal{E}_6$, we have

$$\begin{aligned}
&\max_{Z \in S_\alpha: m > \gamma n} \min_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \left|\boldsymbol{\Delta} O_s - \mathbb{E}\left[\boldsymbol{\Delta} O_s\right]\right| \\
\leq &\max_{Z \in S_\alpha: m > \gamma n} \frac{1}{|\mathcal{B}(Z) \cap S_\alpha|} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \left|\boldsymbol{\Delta} O_s - \mathbb{E}\left[\boldsymbol{\Delta} O_s\right]\right| \\
\leq &\max_{Z \in S_\alpha: m > \gamma n} \frac{1}{|\mathcal{B}(Z) \cap S_\alpha|} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \sum_{a \in [K]} \left|\boldsymbol{\Delta} O_{aa} - \mathbb{E}\left[\boldsymbol{\Delta} O_{aa}\right]\right| \\
\leq &\max_{Z \in S_\alpha: m > \gamma n} \frac{1}{|\mathcal{B}(Z) \cap S_\alpha|} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \sum_{a \leq a'} \left|X_{aa'}(Z) - X_{aa'}(Z')\right| \\
\leq &\theta n(p-q).
\end{aligned}$$

Then, it follows that

$$
\begin{aligned}
&\max_{Z\in\mathcal{G}(\gamma_0):m>\gamma n}\ \min_{Z'\in\mathcal{B}(Z)\cap S_\alpha} \Delta_n(Z,Z') \\
=&\max_{Z\in\mathcal{G}(\gamma_0):m>\gamma n}\ \min_{Z'\in\mathcal{B}(Z)\cap S_\alpha} (\boldsymbol{\Delta} O_s-\lambda^*\boldsymbol{\Delta} n_s) \\
\le&\max_{Z\in\mathcal{G}(\gamma_0):m>\gamma n}\ \min_{Z'\in\mathcal{B}(Z)\cap S_\alpha} (\boldsymbol{\Delta} O_s-\mathbb{E}\left[\boldsymbol{\Delta} O_s\right]) + \max_{Z\in S_\alpha:m>\gamma n}\ \max_{Z'\in\mathcal{B}(Z)\cap S_\alpha} (\mathbb{E}\left[\boldsymbol{\Delta} O_s\right]-\lambda^*\boldsymbol{\Delta} n_s) \\
\le&\theta n(p-q)-\delta_0 n(p-q)(1-o(1)) \\
=&-\delta_0(1-o(1))n(p-q).
\end{aligned} \tag{59}
$$

By the definition of $g(Z)$ and by (52), we have that

$$
\begin{aligned}
\max_{Z\in\mathcal{G}(\gamma_0):m>\gamma n} \log\frac{\Pi_0(Z|A)}{\Pi_0(g(Z)|A)} &= \log\frac{p(1-q)}{q(1-p)}\left[\max_{Z\in\mathcal{G}(\gamma_0):m>\gamma n}\ \min_{Z'\in\mathcal{B}(Z)\cap S_\alpha} \Delta_n(Z,Z')\right] \\
&\le -\log\left(\frac{p}{q}\right)\delta_0 n(p-q)(1-o(1)).
\end{aligned}
$$

Furthermore, since

$$
I=(\sqrt{p}-\sqrt{q})^2(1+o(1)), \qquad \frac{(p-q)\log\left(\frac{p}{q}\right)}{\left(\sqrt{p}-\sqrt{q}\right)^2}\ge 4,
$$

we have

$$
\log\left(\frac{p}{q}\right)(p-q)\ge 4(\sqrt{p}-\sqrt{q})^2 = 4I(1-o(1)).
$$

Then, it follows directly

$$
\max_{Z\in\mathcal{G}(\gamma_0):m>\gamma n}\frac{\Pi_0(Z|A)}{\Pi_0(g(Z)|A)}\le \exp\left(-4\delta_0 nI(1-o(1))\right).
$$

By Lemma 33, we have that $\mathcal{E}_6$ happens with probability at least $1-\exp(-n)$.

Combining results of two regions directly gives the result of Lemma 21. ■

5.7 Proof of Lemma 17 with unknown connectivity probabilities

It suffices to prove the following lemma when the connectivity probabilities are unknown.

Lemma 22 *Suppose γ_0 satisfies Condition B, D, or E. Then, there exists some positive sequence $\gamma\to 0$ such that, with probability at least $1-C_1 n^{-C_2}$,*

$$
\frac{\Pi(Z|A)}{\Pi(g(Z)|A)}\le\begin{cases}\exp\left(-\varepsilon\bar{n}I(1-o(1))\right), & \text{if } m\le\gamma n,\\ \exp\left(-\dfrac{(1-K\gamma_0)^4 nI}{2\alpha^2}(1-o(1))\right), & \text{if } m>\gamma n,\end{cases}
$$

holds uniformly for all $Z \in \mathcal{G}(\gamma_0)$ *defined in* (30)*. Here,* ε *is any constant satisfying* $\varepsilon < \varepsilon_0$*, and* C_3, C_4 *are constants depending on* ε*. Furthermore, if* ξ *satisfies Condition C or D correspondingly, then by choosing* $\varepsilon \in ((1-\varepsilon_0)/\xi, 2\varepsilon_0)$*, we have*

$$\max_{Z \in \mathcal{G}(\gamma_0)} \frac{\widetilde{\Pi}(Z|A)}{\widetilde{\Pi}(g(Z)|A)} \leq \exp(-C\bar{n}I),$$

for some constant $C > 1 - \varepsilon_0$ *with probability at least* $1 - C_3 n^{-C_4}$*.*

Note that when the connectivity probabilities are unknown, the initial conditions are different for the case of two communities and the case of more than two communities. In order to prove Lemma 22, we again divide S_α into a small mistake region and a large mistake region, according to whether $m > \gamma n$, where $\gamma \to 0$ is a positive sequence to be specified later. It is worth noting that we always start from the likelihood modularity, and then bound the exact posterior distribution.

Proof [Proof of Lemma 22] Under the conditions of Lemma 22, let $\varepsilon_{\gamma_0} = 1 - K\gamma_0$ for simplicity, and we have $\varepsilon_{\gamma_0}^4 nI \to \infty$, $\varepsilon_{\gamma_0}(1-K\beta\gamma_0)n \to \infty$. Then, for any positive sequences $\bar{\varepsilon}, \gamma, \theta \to 0$ satisfying that $\bar{\varepsilon}^2 nI \to \infty$, $\theta^2\gamma nI \to \infty$, $\varepsilon_{\gamma_0}^2 \gg \bar{\varepsilon}$, and $\varepsilon_{\gamma_0}^3 \gg \theta\bar{\varepsilon}$. To be specific, we can set $\bar{\varepsilon} = \varepsilon_{\gamma_0}^2/(\varepsilon_{\gamma_0}^4 nI)^{1/4}$, $\gamma = \varepsilon_{\gamma_0}^2$, $\theta = 1/\sqrt{\gamma nI}$.

All the following analyses are based on the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$.

Small mistake region. We write $\mathcal{M}_s = \{Z \in \mathcal{G}(\gamma_0) : m \leq \gamma n\}$. By Lemma 42, since $\gamma < c_{\alpha,\beta}$, we have that for any $Z \in \mathcal{M}_s$, $\mathcal{B}(Z) \subset S_\alpha$. By Lemma 35, under the event $\mathcal{E}_1(\bar{\varepsilon})$, we have

$$\begin{aligned}
&\max_{Z \in M_s} \min_{Z' \in \mathcal{B}(Z)} \log \frac{\Pi(Z|A)}{\Pi(Z'|A)} \\
&\leq \max_{Z \in M_s} \frac{1}{m} \sum_{Z' \in \mathcal{B}(Z)} \log \frac{\Pi(Z|A)}{\Pi(Z'|A)} \\
&\leq \max_{Z \in M_s} \frac{1}{m} \sum_{Z' \in \mathcal{B}(Z)} \big(Q_{LM}(Z,A) - Q_{LM}(Z',A)\big) + \varepsilon_{LM}.
\end{aligned} \tag{60}$$

Thus, we proceed to upper bound $Q_{LM}(Z,A) - Q_{LM}(Z',A)$. By some calculations, we have

$$\begin{aligned}
&Q_{LM}(Z,A) - Q_{LM}(Z',A) \\
=&\log \frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)} - \sum_{a \leq a'} n_{aa'} \cdot D\left(\frac{O_{aa'}}{n_{aa'}} \middle\| \frac{O_{aa'}(Z)}{n_{aa'}(Z)}\right) + \\
&\underbrace{\sum_{a \leq a'} \boldsymbol{\Delta} O_{aa'} \left(\log \frac{O_{aa'}(Z)}{n_{aa'}(Z)} - \log B_{aa'}\right) + (\boldsymbol{\Delta} n_{aa'} - \boldsymbol{\Delta} O_{aa'}) \left(\log \frac{n_{aa'}(Z) - O_{aa'}(Z)}{n_{aa'}(Z)} - \log(1 - B_{aa'})\right)}_{Err(Z,Z')},
\end{aligned} \tag{61}$$

where $\log[\Pi_0(Z|A)/\Pi_0(Z'|A)]$ is calculated in (52). Now suppose we correct one sample from a misclassified group b to its true group b'. Then, by Lemma 43, we have

$$\begin{aligned}
&\boldsymbol{\Delta} O_{b'b'} + \boldsymbol{\Delta} O_{bb} + \boldsymbol{\Delta} O_{bb'} = 0, \quad \boldsymbol{\Delta} O_s = \boldsymbol{\Delta} O_{bb} + \boldsymbol{\Delta} O_{b'b'}, \\
&\boldsymbol{\Delta} O_{ab} + \boldsymbol{\Delta} O_{ab'} = 0, \quad \boldsymbol{\Delta} O_{aa'} = 0, \quad \text{for any } a, a' \in [K] \setminus \{b, b'\}.
\end{aligned}$$

Denote $\widetilde{B}_{aa'} = \mathbb{E}[O_{aa'}(Z)]/n_{aa'}(Z)$ and $\widehat{B}_{aa'} = O_{aa'}(Z)/n_{aa'}(Z)$ for any $a, a' \in [K]$. By Lemma 36, we have

$$\frac{\left|\widetilde{B}_{aa'} - B_{aa'}\right|}{p-q} = \begin{cases} \dfrac{\sum_{k\neq l} R_{ak}R_{al}}{n'_a(n'_a-1)}, & \text{if } a = a', \\ \dfrac{\sum_k R_{ak}R_{a'k}}{n'_a n'_{a'}}, & \text{if } a \neq a', \end{cases} \tag{62}$$

and $\|\widetilde{B} - B\|_\infty \leq 2K\alpha m(p-q)/n$. Under the event $\mathcal{E}_1(\bar{\varepsilon})$ defined in (40), by Lemma 28, we have

$$\|\widehat{B} - B\|_\infty \leq \|\widehat{B} - \widetilde{B}\|_\infty + \|\widetilde{B} - B\|_\infty \leq \left(C\bar{\varepsilon} + \frac{2K\alpha m}{n}\right)(p-q) \lesssim (\gamma + \bar{\varepsilon})(p-q). \tag{63}$$

We then bound $Err(Z, Z')$ in (61) under the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$. Since $p \asymp q$, by some calculations, we have

$$\sum_{Z'\in\mathcal{B}(Z)} Err(Z, Z') = \sum_{a\leq a'} \log\frac{\widehat{B}_{aa'}(1-B_{aa'})}{B_{aa'}(1-\widehat{B}_{aa'})} \sum_{Z'\in\mathcal{B}(Z)} (\boldsymbol{\Delta} O_{aa'} - \lambda^*_{aa'}\boldsymbol{\Delta} n_{aa'}) \tag{64}$$

$$\leq \frac{C}{p}\|\widehat{B} - B\|_\infty \underbrace{\sum_{a\leq a'} \left|\sum_{Z'\in\mathcal{B}(Z)} (\boldsymbol{\Delta} O_{aa'} - \lambda^*_{aa'}\boldsymbol{\Delta} n_{aa'})\right|}_{(A)} \tag{65}$$

for some constant C, where

$$\lambda^*_{aa'} = \log\frac{1-B_{aa'}}{1-\widehat{B}_{aa'}} \Big/ \log\frac{\widehat{B}_{aa'}(1-B_{aa'})}{B_{aa'}(1-\widehat{B}_{aa'})} \in \left[B_{aa'} \wedge \widehat{B}_{aa'}, B_{aa'} \vee \widehat{B}_{aa'}\right].$$

Under the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$, for any $Z \in \mathcal{M}_s$, we bound the above term (A) by the following three terms separately,

$$\sum_{a\leq a'} \left|\sum_{Z'\in\mathcal{B}(Z)} (\boldsymbol{\Delta} O_{aa'} - \mathbb{E}[\boldsymbol{\Delta} O_{aa'}])\right| \lesssim mn(p-q), \tag{66}$$

$$\sum_{a\leq a'} \left|\sum_{Z'\in\mathcal{B}(Z)} (\mathbb{E}[\boldsymbol{\Delta} O_{aa'}] - B_{aa'}\boldsymbol{\Delta} n_{aa'})\right| \leq \sum_{Z'\in\mathcal{B}(Z)} \sum_{a\leq a'} |\mathbb{E}[\boldsymbol{\Delta} O_{aa'}] - B_{aa'}\boldsymbol{\Delta} n_{aa'}| \leq 2Kmn(p-q), \tag{67}$$

$$\sum_{a\leq a'} \left|\sum_{Z'\in\mathcal{B}(Z)} (B_{aa'} - \lambda^*_{aa'})\boldsymbol{\Delta} n_{aa'}\right| \leq \sum_{Z'\in\mathcal{B}(Z)} \sum_{a\leq a'} |\boldsymbol{\Delta} n_{aa'}| \cdot \|\widehat{B} - B\|_\infty \lesssim (\gamma + \bar{\varepsilon})mn(p-q). \tag{68}$$

The first inequality directly follows by Lemma 32. The second inequality is due to that for each fixed Z', there are at most $2K$ pairs of groups contributing to the summations of the absolute values, and for each summation, there are at most n random variables associated. The third inequality follows by (63) and Lemma 43. Hence, under the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$, we have

$$\sum_{Z' \in \mathcal{B}(Z)} Err(Z, Z') \leq C(\gamma + \bar{\varepsilon}) m \bar{n} I,$$

for some constant C where $\gamma, \bar{\varepsilon} \to 0$ as defined in the beginning of the proof. Hence, it follows that

$$\begin{aligned}
&\mathbb{P}\left\{\max_{Z \in \mathcal{M}_s} \min_{Z' \in \mathcal{B}(Z)} \log \frac{\Pi(Z|A)}{\Pi(Z'|A)} > -\varepsilon \bar{n} I\right\} \\
\leq &\mathbb{P}\left\{\max_{Z \in \mathcal{M}_s} \frac{1}{m} \sum_{Z' \in \mathcal{B}(Z)} \log \frac{\Pi(Z|A)}{\Pi(Z'|A)} > -\varepsilon \bar{n} I\right\} \\
\leq &\mathbb{P}\left\{\max_{Z \in \mathcal{M}_s} \frac{1}{m} \sum_{Z' \in \mathcal{B}(Z)} Q_{LM}(Z, A) - Q_{LM}(Z', A) > -\varepsilon \bar{n} I - \varepsilon_{LM}, \mathcal{E}\right\} + \mathbb{P}\left\{\mathcal{E}^c\right\} \\
\leq &\mathbb{P}\left\{\max_{Z \in \mathcal{M}_s} \frac{1}{m} \sum_{Z' \in \mathcal{B}(Z)} \left(\log \frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)} + Err(Z, Z')\right) > -\varepsilon \bar{n} I - \varepsilon_{LM}, \mathcal{E}\right\} + \mathbb{P}\left\{\mathcal{E}^c\right\} \\
\leq &\mathbb{P}\left\{\max_{Z \in \mathcal{M}_s} \frac{1}{m} \sum_{Z' \in \mathcal{B}(Z)} \log \frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)} > -\varepsilon \bar{n} I - C(\gamma + \bar{\varepsilon}) \bar{n} I - \varepsilon_{LM}\right\} + \mathbb{P}\left\{\mathcal{E}^c\right\} \\
= &\mathbb{P}\left\{\max_{Z \in \mathcal{M}_s} t^* \sum_{Z' \in \mathcal{B}(Z)} \Delta_n(Z, Z') > -m \varepsilon \bar{n} I/2 - Cm(\gamma + \bar{\varepsilon}) \bar{n} I/2 - m \varepsilon_{LM}/2\right\} + \mathbb{P}\left\{\mathcal{E}^c\right\}.
\end{aligned} \tag{69}$$

where $\varepsilon_{LM} \to 0$ is defined in (87), and ε is any small constant satisfying $\varepsilon < 2\varepsilon_0$. A simple union bound and following the argument from (56) to (57) lead to that

$$\mathbb{P}\left\{\max_{Z \in \mathcal{M}_s} \min_{Z' \in \mathcal{B}(Z)} \log \frac{\Pi(Z|A)}{\Pi(Z'|A)} > -\varepsilon \bar{n} I\right\} \leq C' n \exp\left(-(1 - \varepsilon/2) \bar{n} I (1 - o(1))\right),$$

for some constant C'.

Remark 23 *Before performing the analysis for the large mistake region, it is worth noting that for the small mistake region, the proof works for any sequence $\gamma \to 0$. Thus, in the case of more than two communities ($K \geq 3$), if $\gamma_0 \to 0$, by Lemma 15, $\mathcal{G}(\gamma_0) \subset \mathcal{M}_s$ for some sequence $\gamma \to 0$, and then the proof is complete. Therefore, we only need to analyze the large mistake region for $K = 2$.*

Large mistake region. We write $\mathcal{M}_l = \{Z \in \mathcal{G}(\gamma_0) : m > \gamma n\}$. By the same argument in (60), we start with $Q_{LM}(Z, A) - Q_{LM}(Z', A)$. For $K = 2$ and $Z \in \mathcal{M}_l$, by some calculations, we have

$$
\begin{aligned}
&Q_{LM}(Z,A)-Q_{LM}(Z',A)\\
&=\sum_{a\le a'} n_{aa'}(Z)\tau\left(\frac{O_{aa'}(Z)}{n_{aa'}(Z)}\right)-\sum_{a\le a'} n_{aa'}(Z')\tau\left(\frac{O_{aa'}(Z')}{n_{aa'}(Z')}\right)\\
&=\sum_{a\le a'} \boldsymbol{\Delta} O_{aa'}\log\frac{O_{aa'}(Z)}{n_{aa'}(Z)}+(\boldsymbol{\Delta} n_{aa'}-\boldsymbol{\Delta} O_{aa'})\log\left(1-\frac{O_{aa'}(Z)}{n_{aa'}(Z)}\right)-\sum_{a\le a'} n_{aa'}(Z')D\left(\frac{O_{aa'}(Z')}{n_{aa'}(Z')}\middle\|\frac{O_{aa'}(Z)}{n_{aa'}(Z)}\right)\\
&\le\underbrace{\sum_{a\le a'} \boldsymbol{\Delta} O_{aa'}\log\widetilde{B}_{aa'}+(\boldsymbol{\Delta} n_{aa'}-\boldsymbol{\Delta} O_{aa'})\log\left(1-\widetilde{B}_{aa'}\right)+}_{P(Z,Z')+Err_1(Z,Z')}\\
&\quad\underbrace{\sum_{a\le a'} \boldsymbol{\Delta} O_{aa'}\left(\log\frac{O_{aa'}(Z)}{n_{aa'}(Z)}-\log\widetilde{B}_{aa'}\right)+(\boldsymbol{\Delta} n_{aa'}-\boldsymbol{\Delta} O_{aa'})\left(\log\left(1-\frac{O_{aa'}(Z)}{n_{aa'}(Z)}\right)-\log\left(1-\widetilde{B}_{aa'}\right)\right)}_{Err_2(Z,Z')},
\end{aligned}
\tag{70}
$$

where we write

$$P(Z,Z')=\sum_{a\le a'}\mathbb{E}\left[\boldsymbol{\Delta} O_{aa'}\right]\log\widetilde{B}_{aa'}+\left(\boldsymbol{\Delta} n_{aa'}-\mathbb{E}\left[\boldsymbol{\Delta} O_{aa'}\right]\right)\log\left(1-\widetilde{B}_{aa'}\right), \tag{71}$$

$$Err_1(Z,Z')=\sum_{a\le a'}\left(\boldsymbol{\Delta} O_{aa'}-\mathbb{E}\left[\boldsymbol{\Delta} O_{aa'}\right]\right)\log\frac{\widetilde{B}_{aa'}}{1-\widetilde{B}_{aa'}}. \tag{72}$$

Recall $g(Z)$ and $\mathcal{B}(Z)$ are defined in (32) and (35) respectively. Let $N=|B(Z)\cap S_\alpha|$. By Lemma 42, we have $N\ge\min\{cn,m\}$ for some constant c.

Step 1: bound $P(Z,Z')$. By Lemma 47, for $Z\in\mathcal{M}_l$ with $m=d(Z,Z^*)$ and any $Z'\in\mathcal{B}(Z)\cap S_\alpha$, we have

$$-P(Z,Z')\ge\frac{nI}{2\alpha^2}\varepsilon_m^3\max\{1-\beta+1/\alpha,\varepsilon_m\}(1-o(1))\gtrsim\varepsilon_m^3 nI,$$

where $\varepsilon_m=1-Km/n$. Note that $m<n/K\beta$ under the condition. The last inequality holds because $\varepsilon_m\to 0$ only if $\beta\to 1$, and $m/n\to 1/K$. Thus, $1-\beta+1/\alpha\gg\varepsilon_m$.

Step 2: bound $Err_1(Z,Z')$. Recall that $N=|\mathcal{B}(Z)\cap S_\alpha|$. Under the event $\mathcal{E}_6(\gamma,\theta)$ defined in (40), we have

$$\max_{Z\in S_\alpha:m>\gamma n}\frac{1}{N}\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha}\sum_{a\le a'}\left|\boldsymbol{\Delta} O_{aa'}-\mathbb{E}\left[\boldsymbol{\Delta} O_{aa'}\right]\right|\le\theta n(p-q),$$

for the positive sequence $\theta \to 0$ defined in the beginning of the proof. When $K = 2$, and Z' corrects one sample from group b to b', by Lemma 43, $Err_1(Z, Z')$ in (72) can be expressed as

$$Err_1(Z,Z') = (\boldsymbol{\Delta} O_{bb} - \mathbb{E}[\boldsymbol{\Delta} O_{bb}]) \log \frac{\widetilde{B}_{bb}(1-\widetilde{B}_{b'b})}{\widetilde{B}_{b'b}(1-\widetilde{B}_{bb})} + (\boldsymbol{\Delta} O_{b'b'} - \mathbb{E}[\boldsymbol{\Delta} O_{b'b'}]) \log \frac{\widetilde{B}_{b'b'}(1-\widetilde{B}_{bb'})}{\widetilde{B}_{bb'}(1-\widetilde{B}_{b'b'})}.$$

By $p \asymp q$ and $(n-2m)(n-2\beta m)/n \to \infty$, it follows by Lemma 44 that

$$\begin{aligned} \frac{1}{N} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} Err_1(Z,Z') &\le C_1 \frac{\theta n (p-q)}{p} \cdot \left(\left| \widetilde{B}_{b'b'} - \widetilde{B}_{bb'} \right| + \left| \widetilde{B}_{bb} - \widetilde{B}_{b'b} \right| \right) \\ &\le C_2 \frac{\theta n (p-q)^2 \det(R)}{p} \left(\frac{2\alpha}{n} \right)^3 (|R_{b'b'} - R_{b'b}| + |R_{bb} - R_{bb'}|) \\ &\le C_3 \frac{\theta I \det(R)(n-2m)}{n^2}, \end{aligned}$$

for some constants C_1, C_2, C_3. Hence, under the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$, where the sequences are defined in the beginning of the proof, if $(n-2m)(n-2\beta m)/n \to \infty$ and $(n-2m)/n \gg \theta$, then by Lemma 48 we have

$$\frac{1}{N} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} Err_1(Z,Z') \ll \frac{\theta I \det(R)(n-2m)}{n^2} \ll -\frac{1}{N} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} P(Z,Z'),$$

for any $Z \in \mathcal{M}_l$.

Step 3: bound $Err_2(Z, Z')$. Recall $N = |\mathcal{B}(Z) \cap S_\alpha|$. By (70), we have

$$\begin{aligned} \frac{1}{N} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} Err_2(Z,Z') &= \frac{1}{N} \sum_{a \le a'} \log \frac{\widehat{B}_{aa'}(1-\widetilde{B}_{aa'})}{\widetilde{B}_{aa'}(1-\widehat{B}_{aa'})} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} (\boldsymbol{\Delta} O_{aa'} - \lambda^*_{aa'} \boldsymbol{\Delta} n_{aa'}) \\ &\le \frac{C}{p} \|\widehat{B} - \widetilde{B}\|_\infty \sum_{a \le a'} \left| \frac{1}{N} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} (\boldsymbol{\Delta} O_{aa'} - \lambda^*_{aa'} \boldsymbol{\Delta} n_{aa'}) \right|, \end{aligned}$$

where

$$\lambda^*_{aa'} = \log \frac{1-\widetilde{B}_{aa'}}{1-\widehat{B}_{aa'}} \Big/ \log \frac{\widehat{B}_{aa'}(1-\widetilde{B}_{aa'})}{\widetilde{B}_{aa'}(1-\widehat{B}_{aa'})} \in \left[\widetilde{B}_{aa'} \wedge \widehat{B}_{aa'}, \widetilde{B}_{aa'} \vee \widehat{B}_{aa'} \right].$$

Under the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$ defined in (40), by Lemma 28, we have $\max_{Z \in S_\alpha} \|\widehat{B} - \widetilde{B}\|_\infty \lesssim \bar{\varepsilon}(p-q)$. By the same argument from (63) to (68), we have

$$\frac{1}{N}\sum_{a\leq a'}\left|\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha}\left(\boldsymbol{\Delta}O_{aa'} - \mathbb{E}\left[\boldsymbol{\Delta}O_{aa'}\right]\right)\right| \leq \theta n(p-q),$$

$$\frac{1}{N}\sum_{a\leq a'}\left|\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha}\left(\mathbb{E}\left[\boldsymbol{\Delta}O_{aa'}\right] - \widetilde{B}_{aa'}\boldsymbol{\Delta}n_{aa'}\right)\right| \leq C(n-2m)(p-q) \lesssim \varepsilon_m n(p-q),$$

$$\frac{1}{N}\sum_{a\leq a'}\left|\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha}(\widetilde{B}_{aa'} - \lambda_{aa'})\boldsymbol{\Delta}n_{aa'}\right| \leq K^2 n\|\widehat{B} - \widetilde{B}\|_\infty \lesssim \bar{\varepsilon}n(p-q).$$

It follows that for any $Z \in \mathcal{M}_l$ with $d(Z, Z^*) = m$,

$$\frac{1}{N}\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha} Err_2(Z, Z') \lesssim \frac{\bar{\varepsilon}n(p-q)^2}{p}\left(\theta + \varepsilon_m + \bar{\varepsilon}\right),$$

where $\varepsilon_m = 1 - 2m/n$. Hence, under the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$, where the sequences are defined in the beginning of the proof. If $\varepsilon_m^3 \gg \theta\bar{\varepsilon}$, $\varepsilon_m^2 \gg \bar{\varepsilon}$, and $\varepsilon_m^3 \gg \bar{\varepsilon}^2$, then we have

$$\frac{1}{N}\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha} Err_2(Z, Z') \ll -\frac{1}{N}\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha} P(Z, Z'),$$

for any $Z \in \mathcal{M}_l$.

By combining all three steps, we require that for all $Z \in \mathcal{G}(\gamma_0)$ with $d(Z, Z^*) = m$,

$$(1-2m/n)^4 nI \to \infty, \quad (n-2m)(n-2\beta m)/n \to \infty \text{ (for Lemma 47)}. \tag{73}$$

By the definition of $\mathcal{G}(\gamma_0)$ in (30) and by Lemma 15, it suffices to require

$$(1-2\gamma_0)^4 nI \to \infty, \quad (1-2\gamma_0)(1-2\beta\gamma_0)n \to \infty.$$

Recall that $\varepsilon_{\gamma_0} = 1 - 2\gamma_0$ in the beginning of the proof. Then, under the event $\mathcal{E}(\bar{\varepsilon}, \gamma, \theta)$ defined in (40), by the conclusions of three steps and by Lemma 35, we have

$$
\begin{aligned}
& \max_{Z \in \mathcal{M}_l} \min_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \log \frac{\Pi(Z|A)}{\Pi(Z'|A)} \\
\leq & \max_{Z \in \mathcal{M}_l} \frac{1}{N} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \log \frac{\Pi(Z|A)}{\Pi(Z'|A)} \\
\leq & \max_{Z \in \mathcal{M}_l} \frac{1}{N} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} (Q_{LM}(Z, A) - Q_{LM}(Z', A) + \varepsilon_{LM}) \\
= & \max_{Z \in \mathcal{M}_l} \frac{1}{N} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \left\{ P(Z, Z') + Err_1(Z, Z') + Err_2(Z, Z') \right\} + \varepsilon_{LM} \qquad (74) \\
= & \max_{Z \in \mathcal{M}_l} \frac{1}{N} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} P(Z, Z')(1 - o(1)) + \varepsilon_{LM} \\
\leq & - \frac{nI}{2\alpha^2} \varepsilon_{\gamma_0}^3 \max\{1 - \beta + 1/\alpha, \varepsilon_{\gamma_0}\}(1 - o(1)) \\
\leq & - \frac{nI}{2\alpha^2} \varepsilon_{\gamma_0}^4 (1 - o(1)).
\end{aligned}
$$

Combine the results of two regions, and Lemma 22 directly follows. ■

5.8 Proof of Lemma 19

For the simplicity of presentation, we first introduce some notations that will be used in the proof. Denote

$$
\widetilde{m} = n \max\{\gamma_0, n^{-\tau}\} + \log^2 n, \quad m^* = n^{1-\tau},
$$

where τ is a sufficiently small constant defined in (30). For any $Z \in \mathcal{G}(\gamma_0)$, we define the following set

$$
\mathcal{S}(Z, \eta) = \left\{ S' \subset \mathcal{N}(Z) \cap S_\alpha : |S'| \geq \eta \left| \mathcal{B}(Z) \cap S_\alpha \right| \right\}, \qquad (75)
$$

where η is a small constant satisfying $2\tau < \eta < 1/2$, and $\mathcal{A}(Z), \mathcal{B}(Z), \mathcal{N}(Z)$ are defined in (35). Then, we have the following lemma.

Lemma 24 *Suppose γ_0, ξ satisfy all conditions for Theorem 5 or those for Theorem 11, with probability at least $1 - \exp(-n^{1-\tau})$, we have*

$$
\max_{Z \in S_\alpha : m^* \leq m \leq \widetilde{m}} \max_{S' \in \mathcal{S}(Z, \eta)} \min \left\{ \min_{Z' \in S' \cap \mathcal{B}(Z)} \frac{\widetilde{\Pi}(Z|A)}{\widetilde{\Pi}(Z'|A)}, \min_{Z' \in S' \cap \mathcal{A}(Z)} \frac{\widetilde{\Pi}(Z'|A)}{\widetilde{\Pi}(Z|A)} \right\} \leq \exp(-C\bar{n}I),
$$

for some constant $C > 1 - \varepsilon_0$, and $2\tau < \eta < 1/2$.

To understand Lemma 24, we say that $Z' \in \mathcal{N}(Z)$ is *making a mistake* if $Z' \in \mathcal{B}(Z)$ but rejected, or $Z' \in \mathcal{A}(Z)$ but accepted. Lemma 24 implies that under all the conditions, for any current state Z with $m \in [m^*, \widetilde{m}]$, if we make at least $\eta|\mathcal{B}(Z) \cap S_\alpha|$ different choices of Z', then there is at least one Z' such that it is not making a mistake with high probability. In other words, it holds with high probability that Z' will make less than $\eta|\mathcal{B}(Z) \cap S_\alpha|$ mistakes among all possible choices.

Though Lemma 24 seems very similar to Lemma 21, Lemma 21 works for all $Z \in \mathcal{G}(\gamma_0)$, and focuses on the posterior ratio of the current state and the next possible state in set $\mathcal{B}(Z)$, while Lemma 24 works for $Z \in \mathcal{G}(\gamma_0)$ with $m \in [m^*, \widetilde{m}]$, and also bounds the probability of updating to $\mathcal{A}(Z)$.

Proof [Proof of Lemma 19] By Lemma 24, with probability at least $1 - \exp(-n^{1-\tau})$, for any $Z \in \mathcal{G}(\gamma_0)$ with $m \in [m^*, \widetilde{m}]$, by (36), we have that

$$p_m(Z) = p(Z, \mathcal{A}(Z)) = \frac{1}{2(K-1)n} \sum_{Z' \in \mathcal{A}(Z) \cap S_\alpha} \min\left\{1, \frac{\widetilde{\Pi}_g(Z'|A)}{\widetilde{\Pi}_g(Z|A)}\right\} \leq \frac{\eta|\mathcal{B}(Z) \cap S_\alpha| + \varepsilon}{2(K-1)n},$$

$$q_m(Z) = p(Z, \mathcal{B}(Z)) = \frac{1}{2(K-1)n} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \min\left\{1, \frac{\widetilde{\Pi}_g(Z'|A)}{\widetilde{\Pi}_g(Z|A)}\right\} \geq \frac{(1-\eta)|\mathcal{B}(Z) \cap S_\alpha|}{2(K-1)n},$$

where $\varepsilon = n\exp(-C\bar{n}I) \to 0$ for some $C > 1 - \varepsilon_0$. It follows that with probability at least $1 - \exp(-n^{1-\tau})$, $p_m(Z)/q_m(Z) \leq 2\eta$ holds for any $Z \in \mathcal{G}(\gamma_0)$ with $m \in [m^*, \widetilde{m}]$. ■

In order to prove Lemma 24, we first state two lemmas according to whether the connectivity probabilities are known or not. In the case of known connectivity probabilities, we use $\Pi_0(\cdot|A)$ to denote the posterior distribution.

Lemma 25 *When p, q are both known, given τ sufficiently small and η satisfying $2\tau < \eta < 1/2$, if $(1 - K\alpha\gamma_0)^2 nI \to \infty$, then we have*

$$\max_{S' \in \mathcal{S}(Z,\eta)} \min\left\{\min_{Z' \in S' \cap \mathcal{B}(Z)} \frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)}, \min_{Z' \in S' \cap \mathcal{A}(Z)} \frac{\Pi_0(Z'|A)}{\Pi_0(Z|A)}\right\}$$
$$\leq \begin{cases} \exp\left(-\varepsilon\bar{n}I\right), & \text{if } m^* \leq m \leq \gamma n, \\ \exp\left(-4(1/K\alpha - \gamma_0)nI(1-o(1))\right), & \text{if } \gamma n < m \leq \widetilde{m}, \end{cases}$$

holds uniformly for all $Z \in \mathcal{G}(\gamma_0)$ with $m \in [m^*, \widetilde{m}]$ and some sequence $\gamma \to 0$, with probability at least $1 - \exp\left(-n^{1-\tau}\right)$. Here, ε is any small constant satisfying $\varepsilon < 2\varepsilon_0$.

Lemma 26 *Given τ sufficiently small and η satisfying $2\tau < \eta < 1/2$. Suppose γ_0 satisfies Condition B or D, there exists some positive sequence $\gamma \to 0$ such that with probability at least $1 - \exp\left(-n^{1-\tau}\right)$,*

$$\max_{S' \in \mathcal{S}(Z,\eta)} \min\left\{\min_{Z' \in S' \cap \mathcal{B}(Z)} \frac{\Pi(Z|A)}{\Pi(Z'|A)}, \min_{Z' \in S' \cap \mathcal{A}(Z)} \frac{\Pi(Z'|A)}{\Pi(Z|A)}\right\}$$
$$\leq \begin{cases} \exp\left(-\varepsilon\bar{n}I(1-o(1))\right), & \text{if } m^* \leq m \leq \gamma n, \\ \exp\left(-\frac{(1-K\gamma_0)^4 nI}{2\alpha^3}(1-o(1))\right), & \text{if } \gamma n < m \leq \widetilde{m}, \end{cases}$$

holds uniformly for all $Z \in \mathcal{G}(\gamma_0)$ with $m \in [m^*, \widetilde{m}]$. Here, ε is any constant satisfying $\varepsilon < 2\varepsilon_0$.

The proofs of Lemma 25 and Lemma 26 will be presented in the sequel. We first proceed to prove Lemma 24 based on these two lemmas.

Proof [Proof of Lemma 24] The result directly follows Lemma 25 and Lemma 26 by choosing ξ properly in Theorem 5 or Theorem 11. Then, by choosing $\varepsilon \in ((1-\varepsilon_0)/\xi, 2\varepsilon_0)$, we have that

$$\max_{Z \in S_\alpha : m^* \le m \le \widetilde{m}} \max_{S' \in \mathcal{S}(Z,\eta)} \min \left\{ \min_{Z' \in S' \cap \mathcal{B}(Z)} \frac{\widetilde{\Pi}(Z|A)}{\widetilde{\Pi}(Z'|A)}, \min_{Z' \in S' \cap \mathcal{A}(Z)} \frac{\widetilde{\Pi}(Z'|A)}{\widetilde{\Pi}(Z|A)} \right\} \le \exp(-C\bar{n}I),$$

for some constant $C > 1-\varepsilon_0$ with probability at least $1-\exp(-n^{1-\tau})$ for the sufficiently small constant τ. ■

We finally present the proofs of Lemma 25 and Lemma 26 to complete this section.

Proof [Proof of Lemma 25] We consider any positive sequences γ, θ satisfying $\gamma, \theta \to 0$, $\theta^2 \gamma n I \to \infty$, and $\theta \ll 1 - K\alpha\gamma_0$. Suppose $\gamma n \in [m^*, \widetilde{m}]$, and we perform analyses for the following cases.

Case 1: $m^* \le m \le \gamma n$. Since the minimum is upper bounded by the average, it follows that

$$\begin{aligned} &\log \max_{Z: m^* \le m \le \gamma n} \max_{S' \in \mathcal{S}(Z,\eta)} \min \left\{ \min_{Z' \in S' \cap \mathcal{B}(Z)} \frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)}, \min_{Z' \in S' \cap \mathcal{A}(Z)} \frac{\Pi_0(Z'|A)}{\Pi_0(Z|A)} \right\} \qquad (76) \\ &\le \max_{Z: m^* \le m \le \gamma n} \max_{S' \in \mathcal{S}(Z,\eta)} \frac{1}{|S'|} \left(\sum_{Z' \in S' \cap \mathcal{B}(Z)} \log \frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)} + \sum_{Z' \in S' \cap \mathcal{A}(Z)} \log \frac{\Pi_0(Z'|A)}{\Pi_0(Z|A)} \right) \\ &= \max_{Z: m^* \le m \le \gamma n} \max_{S' \in \mathcal{S}(Z,\eta)} \frac{2t^*}{|S'|} \left(\sum_{Z' \in S' \cap \mathcal{B}(Z)} \Delta_n(Z,Z') + \sum_{Z' \in S' \cap \mathcal{A}(Z)} \Delta_n(Z',Z) \right), \end{aligned}$$

where $\Delta_n(Z,Z')$ is defined in (51). By a similar argument from (56) to (57), we have

$$\mathbb{E}\left[t^* \left(\sum_{Z' \in S' \cap \mathcal{B}(Z)} \Delta_n(Z,Z') + \sum_{Z' \in S' \cap \mathcal{A}(Z)} \Delta_n(Z',Z) \right) \right] \le \exp(-(1-C\gamma)|S'|\bar{n}I).$$

We also have $|\mathcal{B}(Z) \cap S_\alpha| = m$ by Lemma 42. For any small constant $\varepsilon < 2\varepsilon_0$, write $C_{\gamma,\varepsilon} = 1 - C\gamma - \varepsilon/2$ for simplicity. It follows that

$$\begin{aligned} \mathbb{P}\left\{(76) \ge -\varepsilon\bar{n}I\right\} &\le \underbrace{\sum_{m^* \le m \le \gamma n} \binom{n}{m} (K-1)^m}_{\text{all possible } Z} \underbrace{\sum_{|S'| \ge \eta m} \binom{Kn}{|S'|}}_{\text{all possible } S'} \underbrace{\exp\left(-C_{\gamma,\varepsilon}|S'|\bar{n}I\right)}_{\text{bound for each given } Z \text{ and } S'} \\ &\lesssim \sum_{m^* \le m \le \gamma n} \binom{n}{m} (K-1)^m \binom{Kn}{\eta m} \exp\left(-C_{\gamma,\varepsilon}\eta m \bar{n}I\right) \\ &\lesssim \binom{n}{m^*} (K-1)^{m^*} \binom{Kn}{\eta m^*} \exp\left(-C_{\gamma,\varepsilon}\eta m^* \bar{n}I\right) \\ &\lesssim \exp\left(-C_{\gamma,\varepsilon}\eta m^* \bar{n}I + (\eta+1)m^* \log \frac{eKn}{\eta m^*}\right). \qquad (77) \end{aligned}$$

For any sufficiently small τ, when $\eta > 2\tau$ and $m^* = n^{1-\tau}$, it is easy to check that

$$(77) \lesssim \exp\left(-m^*\left(C_{\gamma,\varepsilon}\eta\bar{n}I - (\eta+1)\log\frac{eKn^\tau}{\eta}\right)\right) \le e^{-m^*}.$$

Case 2: $\gamma n < m \le \widetilde{m}$**.** Recall that $\boldsymbol{\Delta} O_s(Z,Z') = O_s(Z) - O_s(Z')$ for any label assignments Z and Z'. By (52) and (58), we have that

$$\log\frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)} = 2t^*\Delta_n(Z,Z'),$$
$$\max_{Z\in\mathcal{G}(\gamma_0)}\max_{Z'\in\mathcal{B}(Z)\cap S_\alpha}\mathbb{E}\left[\Delta_n(Z,Z')\right] \le -\left(\frac{1}{K\alpha}-\gamma_0\right)(p-q)(1-o(1)).$$

Since $|S'| \ge \eta|\mathcal{B}(Z)\cap S_\alpha|$, by the same proof in Lemma 33, we have that with probability at least $1-\exp(-n)$,

$$\begin{aligned}
&\max_{Z\in S_\alpha:\gamma n<m\le\widetilde{m}}\max_{S'\in S(Z,\alpha)}\left\{\frac{1}{|S'|}\sum_{Z'\in S'}\left|\Delta_n(Z,Z')-\mathbb{E}\left[\Delta_n(Z,Z')\right]\right|\right\}\\
=&\max_{Z\in S_\alpha:\gamma n<m\le\widetilde{m}}\max_{S'\in S(Z,\alpha)}\left\{\frac{1}{|S'|}\sum_{Z'\in S'}\left|\boldsymbol{\Delta} O_s(Z,Z')-\mathbb{E}\left[\boldsymbol{\Delta} O_s(Z,Z')\right]\right|\right\}\\
\le&\theta n(p-q),
\end{aligned}$$

where the positive sequence θ is defined in the beginning of the proof. Thus, by a similar argument from (58) to (59), the result directly follows.

Combining two cases gives Lemma 25 directly. Note that in the case of $m^* \ge \gamma n$ or $\widetilde{m} < \gamma n$, the result trivially follows. ■

Proof [Proof of Lemma 26] Consider any positive sequences $\bar{\varepsilon},\gamma,\theta \to 0$ satisfying that $\bar{\varepsilon}^2 nI \to \infty$, $\theta^2\gamma nI \to \infty$, $(1-K\gamma_0)^2 \gg \bar{\varepsilon}$, and $(1-K\gamma_0)^3 \gg \theta\bar{\varepsilon}$. Note that the second case in Lemma 26 is only for the case of $K=2$. When $K\ge 3$, we require $\gamma_0 \to 0$, and thus there exists some $\gamma\to 0$, such that for all $Z\in\mathcal{G}(\gamma_0)$, $m\le\gamma n$ always holds.

The following proof is similar with those of Lemma 22 and Lemma 25. Denote $\Delta Q(Z,Z') = Q_{LM}(Z,A) - Q_{LM}(Z',A)$ for any two label assignments $Z,Z'\in S_\alpha$, where $Q_{LM}(Z,A)$ is the likelihood modularity function defined in (42). Under the event $\mathcal{E}_1(\bar{\varepsilon})$, by Lemma 35, there exists some sequence $\varepsilon_{LM}\to 0$ such that

$$\begin{aligned}
&\log\max_{S'\in\mathcal{S}(Z,\eta)}\min\left\{\min_{Z'\in S'\cap\mathcal{B}(Z)}\frac{\Pi(Z|A)}{\Pi(Z'|A)},\min_{Z'\in S'\cap\mathcal{A}(Z)}\frac{\Pi(Z'|A)}{\Pi(Z|A)}\right\}\\
\le&\max_{S'\in\mathcal{S}(Z,\eta)}\frac{1}{|S'|}\left\{\sum_{Z'\in S'\cap\mathcal{B}(Z)}\log\frac{\Pi(Z|A)}{\Pi(Z'|A)}+\sum_{Z'\in S'\cap\mathcal{A}(Z)}\log\frac{\Pi(Z'|A)}{\Pi(Z|A)}\right\}\\
\le&\max_{S'\in\mathcal{S}(Z,\eta)}\frac{1}{|S'|}\left\{\sum_{Z'\in S'\cap\mathcal{B}(Z)}\Delta Q(Z,Z')+\sum_{Z'\in S'\cap\mathcal{A}(Z)}\Delta Q(Z',Z)\right\}+\varepsilon_{LM}. \qquad (78)
\end{aligned}$$

The first inequality is because minimum is smaller than the average.

Case 1: $m^* \le m \le \gamma n$. In this case, $B(Z) \subset S_\alpha$ by Lemma 42. By (61), we have that under the event $\mathcal{E}_1(\bar{\varepsilon})$,

$$(78) \le \max_{S' \in \mathcal{S}(Z,\eta)} \frac{1}{|S'|} \left\{ \sum_{Z' \in S' \cap \mathcal{B}(Z)} \log \frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)} + \sum_{Z' \in S' \cap \mathcal{A}(Z)} \log \frac{\Pi_0(Z'|A)}{\Pi_0(Z|A)} + \sum_{Z' \in S'} \left| Err(Z, Z') \right| \right\} + \varepsilon_{LM}.$$

By a similar argument from (61) to (69), in order to prove Lemma 26, it suffices to show that

$$\max_{Z \in \mathcal{G}(\gamma_0): m^* \le m \le \gamma n} \max_{S' \in \mathcal{S}(Z,\eta)} \frac{1}{|S'|} \sum_{Z' \in S'} \left| Err(Z, Z') \right| = o(\bar{n} I). \tag{79}$$

Recall that $Err(Z, Z')$ is defined in (61). It follows by (65) that under the event $\mathcal{E}_1(\bar{\varepsilon})$, for any $Z \in \mathcal{G}(\gamma_0)$ with $m \in [m^*, \gamma n]$,

$$\begin{aligned}
&\max_{S' \in \mathcal{S}(Z,\eta)} \frac{1}{|S'|} \sum_{Z' \in S'} \left| Err(Z, Z') \right| \\
\lesssim &(\gamma + \bar{\varepsilon}) \frac{p-q}{p} \max_{S' \in \mathcal{S}(Z,\eta)} \frac{1}{|S'|} \sum_{Z' \in S'} \sum_{a \le a'} \left| \boldsymbol{\Delta} O_{aa'} - \lambda^*_{aa'} \boldsymbol{\Delta} n_{aa'} \right| \\
\lesssim &(\gamma + \bar{\varepsilon}) \frac{p-q}{p} \left\{ (A) + (B) + (C) \right\},
\end{aligned}$$

where

$$\begin{aligned}
(A) &= \max_{S' \in \mathcal{S}(Z,\eta)} \frac{1}{|S'|} \sum_{Z' \in S'} \sum_{a \le a'} \left| \boldsymbol{\Delta} O_{aa'} - \mathbb{E}\left[\boldsymbol{\Delta} O_{aa'} \right] \right| \lesssim n(p-q), \\
(B) &= \max_{S' \in \mathcal{S}(Z,\eta)} \frac{1}{|S'|} \sum_{Z' \in S'} \sum_{a \le a'} \left| \mathbb{E}\left[\boldsymbol{\Delta} O_{aa'} \right] - B_{aa'} \boldsymbol{\Delta} n_{aa'} \right| \lesssim n(p-q), \\
(C) &= \max_{S' \in \mathcal{S}(Z,\eta)} \frac{1}{|S'|} \sum_{Z' \in S'} \sum_{a \le a'} \left| (B_{aa'} - \lambda^*_{aa'}) \boldsymbol{\Delta} n_{aa'} \right| \ll n(p-q).
\end{aligned}$$

The first inequality holds with probability at least $1 - e^{-m^*}$ by the same proof of Lemma 25 and Lemma 32. The second and the third inequalities hold due to the same arguments for (67) and (68). Hence, the proof is complete for the small mistake region.

Case 2: $\gamma n < m \le \widetilde{m}$. We only analyze this case for $K = 2$. By (70), we have that under the event $\mathcal{E}_1(\bar{\varepsilon})$,

$$\begin{aligned}
(78) \le &\max_{S' \in \mathcal{S}(Z,\eta)} \frac{1}{|S'|} \left\{ \sum_{Z' \in S' \cap \mathcal{B}(Z)} P(Z, Z') + \sum_{Z' \in S' \cap \mathcal{A}(Z)} P(Z', Z) \right\} + \\
&\max_{S' \in \mathcal{S}(Z,\eta)} \frac{1}{|S'|} \sum_{Z' \in S'} \left(\left| Err_1(Z, Z') \right| + \left| Err_2(Z, Z') \right| \right) + \varepsilon_{LM},
\end{aligned}$$

where $Err_1(Z, Z')$ and $Err_2(Z, Z')$ are defined in (70) and (72). According to arguments in Lemma 17, we only need to bound $Err_1(Z, Z')$ and $Err_2(Z, Z')$ in order to upper bound

bound (78) as well as the posterior ratio. The only term inside $Err_1(Z,Z')$ and $Err_2(Z,Z')$ needed to be treated specially is

$$\max_{Z\in S_\alpha:\gamma n<m\leq\widetilde{m}}\max_{S'\in\mathcal{S}(Z,\eta)}\frac{1}{|S'|}\sum_{Z'\in S'}\sum_{a\leq a'}\left|\boldsymbol{\Delta}O_{aa'}(Z,Z')-\mathbb{E}\left[\boldsymbol{\Delta}O_{aa'}(Z,Z')\right]\right|,$$

denoted by D for simplicity. By the same proof of Lemma 33, we have that

$$\mathbb{P}\left\{D\geq\theta n(p-q)\right\}\leq e^{-n},$$

for the positive sequence θ defined in the beginning of the proof. Hence, following the same arguments from (70) to (74), the proof of Lemma 26 is completed for the large mistake region.

Combining two cases gives Lemma 26 directly. Note that in the case of $m^*\geq\gamma n$ or $\widetilde{m}<\gamma n$, the result trivially follows. ■

5.9 Proof of Lemma 7

In this section, we proceed to lower bound the posterior distribution.

5.9.1 When connectivity probabilities are known

Let $\{x_i\}_{i\geq 1},\{y_j\}_{j\geq 1}$ be i.i.d. copies of Bernoulli(q) and Bernoulli(p). According to (24) and (46), for any $Z\in S_\alpha$, we have that

$$\log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}=\log\frac{p(1-q)}{q(1-p)}\left(\boldsymbol{\Delta}\widetilde{O}_s-\mathbb{E}\left[\boldsymbol{\Delta}\widetilde{O}_s\right]+\mathbb{E}\left[\boldsymbol{\Delta}\widetilde{O}_s\right]-\lambda^*\boldsymbol{\Delta}\widetilde{n}_s\right),$$

where $\boldsymbol{\Delta}\widetilde{O}_s=\sum_{i=1}^{N_d}x_i-\sum_{i=1}^{N_s}y_i$, and

$$N_d=\sum_{i<j}\mathbb{I}\left\{Z_i=Z_j,Z_i^*\neq Z_j^*\right\},\quad N_s=\sum_{i<j}\mathbb{I}\left\{Z_i^*=Z_j^*,Z_i\neq Z_j\right\}. \tag{80}$$

Recall that $m_k=\sum_{i=1}^n\mathbb{I}\{Z_i=k,\ Z_i^*\neq k\}$, and it follows that $m=\sum_{k\in[K]}m_k$. Write $\beta_k=\sum_{i<j}\mathbb{I}\left\{Z_i=Z_j=k,Z_i^*\neq Z_j^*\right\}$ for simplicity, and we have

$$\beta_k=\sum_{a<b}R_{ka}R_{kb}\leq R_{kk}\cdot m_k+m_k^2=n_k'm_k\leq\frac{\alpha n m_k}{K}.$$

It follows that $N_d=\sum_{k=1}^K\beta_k\leq\alpha mn/K$. Similarly, we have $N_s\leq\beta mn/K$, and

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\Delta}\widetilde{O}_s\right]-\lambda^*\boldsymbol{\Delta}\widetilde{n}_s&=N_dq-N_sp-\lambda^*(N_d-N_s)\\&=-(N_d\cdot(\lambda^*-q)+N_s\cdot(p-\lambda^*))\\&\geq-(N_d+N_s)\cdot(p-q)\geq-(\alpha+\beta)mn(p-q)/K.\end{aligned}$$

Furthermore, by Lemma 29, with probability at least $1 - n\exp(-(1-o(1))\bar{n}I)$, for any $Z \in S_\alpha$ with $m = d(Z, Z^*)$, we have

$$\left|\boldsymbol{\Delta}\widetilde{O}_s - \mathbb{E}\left[\boldsymbol{\Delta}\widetilde{O}_s\right]\right| \le \sum_{a\in[K]} \left|\boldsymbol{\Delta}\widetilde{O}_{aa} - \mathbb{E}\left[\boldsymbol{\Delta}\widetilde{O}_{aa}\right]\right| = \sum_{a\in[K]} |X_{aa}(Z) - X_{aa}(Z^*)|$$
$$\le K\|X(Z) - X(Z^*)\|_\infty \le Cmn(p-q).$$

Hence, it follows that

$$\begin{aligned}\log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)} &\ge -\log\frac{p(1-q)}{q(1-p)}\cdot Cmn(p-q)\\ &\ge -Cmn\cdot\frac{(p-q)^2}{q(1-p)}\\ &\ge -C'nmI\cdot(1+o(1)),\end{aligned}$$

for some constants C and C' with probability at least $1 - n\exp(-(1-o(1))\bar{n}I)$.

5.9.2 When connectivity probabilities are unknown

In order to simplify the proof, we first define some events, and all the following analysis are conditioning on the given events. For any positive sequences $\gamma, \theta \to 0$ with $\gamma^2 nI \to \infty$, and $\theta^2\gamma nI \to \infty$, let $\bar{\varepsilon} = \gamma$ for simplicity. Consider events $\mathcal{E}_1(\bar{\varepsilon})$, $\mathcal{E}_2$, $\mathcal{E}_3(\bar{\varepsilon})$, $\mathcal{E}_4(\gamma,\theta)$ defined in (40). Under the events $\mathcal{E}_1(\bar{\varepsilon})$ and $\mathcal{E}_3(\bar{\varepsilon})$, by Lemma 34, we have that for any $Z \in S_\alpha$,

$$\log\frac{\Pi(Z|A)}{\Pi(Z^*|A)} - \Delta Q(Z, Z^*) \ge -C_{LM},$$

where $\Delta Q(Z, Z^*) = Q_{LM}(Z, A) - Q_{LM}(Z^*, A)$. Thus, it suffices to lower bound $\Delta Q(Z, Z^*)$.

Case 1: $m \le \gamma n$. By (45), we have

$$\begin{aligned}&\Delta Q(Z, Z^*)\\ &= \underbrace{\log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}}_{(A)} - \underbrace{\sum_{a\le b} n_{ab}(Z^*)\cdot D\left(\frac{O_{ab}}{n_{ab}}\middle\|\frac{O_{ab}(Z)}{n_{ab}(Z)}\right)}_{(B)}\\ &\quad + \underbrace{\sum_{a\le b}\boldsymbol{\Delta}\widetilde{O}_{ab}\left(\log\frac{O_{ab}(Z)}{n_{ab}(Z)} - \log B_{ab}\right) + (\boldsymbol{\Delta}\widetilde{n}_{ab} - \boldsymbol{\Delta}\widetilde{O}_{ab})\left(\log\frac{n_{ab}(Z) - O_{ab}(Z)}{n_{ab}(Z)} - \log(1 - B_{ab})\right)}_{(C)},\end{aligned}$$

where $\Pi_0(\cdot|A)$ is defined in (46). We proceed to bound each term above separately. Under the event $\mathcal{E}_2$ and by the same argument in Section 5.9.1, we have

$$A \gtrsim -mnI.$$

By Lemma 51, under the events $\mathcal{E}_1(\bar{\varepsilon})$ and $\mathcal{E}_2$, we have

$$B \lesssim m^2 I.$$

By Lemma 37, under the events $\mathcal{E}_1(\bar{\varepsilon})$ and $\mathcal{E}_2$, we have

$$C \gtrsim -(\bar{\varepsilon} + \gamma) mnI.$$

Hence, under the events $\mathcal{E}_1(\bar{\varepsilon}), \mathcal{E}_2, \mathcal{E}_3(\bar{\varepsilon})$, we have that for any $Z \in S_\alpha$ with $m \leq \gamma n$,

$$\log \frac{\Pi(Z|A)}{\Pi(Z^*|A)} \geq \Delta Q(Z, Z^*) - C_{LM} \geq -CmnI,$$

for some constant C.

Case 2: $m > \gamma n$**.** By (49), we have

$$\Delta Q(Z, Z^*) = G(Z) - G(Z^*) + \Delta(Z) - \Delta(Z^*),$$

where by (50) and Lemma 50, we have

$$G(Z) - G(Z^*) \geq -\frac{1}{2} \sum_{a,b,k,l} R_{ak} R_{bl} D\left(B_{kl} \middle\| \widetilde{B}ab\right) \gtrsim -mnI.$$

By Lemma 41, under the events $\mathcal{E}_1(\bar{\varepsilon})$ and $\mathcal{E}_4(\gamma, \theta)$, for any $Z \in S_\alpha$ with $m > \gamma n$, we have

$$\Delta(Z) - \Delta(Z^*) \geq -\varepsilon mnI$$

for some sequence $\varepsilon \to 0$. Hence, under the events $\mathcal{E}_1(\bar{\varepsilon})$, $\mathcal{E}_3(\bar{\varepsilon})$, and $\mathcal{E}_4(\gamma, \theta)$, we have that for any $Z \in S_\alpha$ with $m > \gamma n$,

$$\log \frac{\Pi(Z|A)}{\Pi(Z^*|A)} \geq \Delta Q(Z, Z^*) - C_{LM} \geq -C' mnI,$$

for some constant C'.

Combining two cases, we have that with probability at least $1 - n\exp(-(1 - o(1))\bar{n}I)$,

$$\min_{Z \in S_\alpha} \left(\log \frac{\Pi(Z|A)}{\Pi(Z^*|A)} + CmnI \right) \geq 0,$$

for some constant C. By Theorem 1, we have $\log \Pi(Z^*|A) \geq C$ for some constant C with high probability. To conclude, there exists some constants C_3, C_4 and C_5 such that for any $Z \in S_\alpha$,

$$\min_{Z \in S_\alpha} \left(\log \Pi(Z|A) + C_3 mnI \right) \geq 0.$$

5.10 Bounding probability of events

We first introduce some notations. Recall that

$$S_\alpha = \left\{ Z : \sum_{i=1}^{n} \mathbb{I}\left\{Z_i = k\right\} \in \left[\frac{n}{\alpha K}, \frac{\alpha n}{K}\right], \text{ for all } k \in [K] \right\}, \tag{81}$$

for some large constant α. For any $Z \in [K]^n$ and any $a \in [K]$, we use n'_a to denote $n_a(Z)$, m_a to denote $n_a(Z) - R_{aa}$ for simplicity.

Lemma 27 *Denote $X_{ab}(Z) = O_{ab}(Z) - \mathbb{E}\left[O_{ab}(Z)\right]$ for any $a, b \in [K]$, then for a general $K \geq 2$, we have*

$$\mathbb{P}\left\{ \max_{Z\in[K]^n} \|X(Z)\|_\infty \geq \bar{\varepsilon} n^2(p-q) \right\} \leq \exp(-n),$$

as long as $\bar{\varepsilon}^2 nI \to \infty$.

Proof For $K \geq 2$, we have $\mathrm{Var}\left(O_{ab}(Z)\right) \leq n_{ab}(Z)p \leq n^2 p$. Then, by a union bound and Bernstein inequality, it follows that

$$\begin{aligned}
\mathbb{P}\left\{ \max_{Z\in[K]^n} \|X(Z)\|_\infty \geq \bar{\varepsilon} n^2(p-q) \right\} &\leq 2K^2K^n \left(\exp\left(-\frac{\bar{\varepsilon}^2 n^2 (p-q)^2}{2p} \right) + \exp\left(-\frac{3\bar{\varepsilon} n^2(p-q)}{2} \right) \right) \\
&\leq 2K^{n+2} \exp\left(-(1-o(1)) \frac{\bar{\varepsilon}^2 n^2 I}{2} \right) \\
&\leq \exp(-n),
\end{aligned}$$

for any $\bar{\varepsilon}$ satisfying that $\bar{\varepsilon}^2 nI \to \infty$. ■

The conclusion of Lemma 27 directly leads to the following lemma.

Lemma 28 *For $K \geq 2$ and any $a, b \in [K]$, let $\widehat{B}_{ab} = O_{ab}(Z)/n_{ab}(Z)$, and $\widetilde{B}_{ab} = \mathbb{E}\left[O_{ab}(Z)\right]/n_{ab}(Z)$. Under the event $\mathcal{E}$ defined in* (41)*, we have*

$$\max_{Z\in S_\alpha} \|\widehat{B} - \widetilde{B}\|_\infty = \max_{Z\in S_\alpha} \max_{a,b\in[K]} \frac{X_{ab}(Z)}{n_{ab}(Z)} \leq \frac{\bar{\varepsilon} n^2(p-q)}{(K\alpha/n)^2} = C\bar{\varepsilon}(p-q),$$

for some constant C depending on K and α.

We state the following lemmas whose proofs will be given together.

Lemma 29

$$\mathbb{P}\left\{ \max_{Z\in S_\alpha} \|X(Z) - X(Z^*)\|_\infty - Cmn(p-q) \geq 0 \right\} \leq n\exp(-(1-o(1))\bar{n}I),$$

for some constant C.

Lemma 30

$$\mathbb{P}\left\{\max_{Z\in S_\alpha}\|X(Z)-X(Z^*)\|_\infty \geq \bar{\varepsilon}n^2(p-q)\right\} \leq \exp(-n),$$

as long as $\bar{\varepsilon}^2 nI \to \infty$.

Lemma 31 *For any positive sequences* $\gamma \to 0$, $\theta \to 0$ *satisfying* $\theta^2\gamma nI \to \infty$, *we have*

$$\mathbb{P}\left\{\max_{Z\in S_\alpha: m>\gamma n}\|X(Z)-X(Z^*)\|_\infty \geq \theta mn(p-q)\right\} \leq \exp(-n).$$

Proof [Proofs of Lemmas 29, 30, and 31] Recall that $X_{ab}(Z)-X_{ab}(Z^*) = \boldsymbol{\Delta}\widetilde{O}_{ab} - \mathbb{E}\left[\boldsymbol{\Delta}\widetilde{O}_{ab}\right]$ for $a,b\in[K]$. We first consider the case where $a\neq b$, and it follows that $\boldsymbol{\Delta}\widetilde{O}_{ab} = \sum_{i,j\in S_1} A_{ij} - \sum_{i,j\in S_2} A_{ij}$, where

$$|S_1| = \sum_{i,j\in[n]}\left(\mathbb{I}\left\{Z_i=a, Z_j=b\right\} - \mathbb{I}\left\{Z_i=Z_i^*=a, Z_j=Z_j^*=b\right\}\right),$$

$$|S_2| = \sum_{i,j\in[n]}\left(\mathbb{I}\left\{Z_i^*=a, Z_j^*=b\right\} - \mathbb{I}\left\{Z_i=Z_i^*=a, Z_j=Z_j^*=b\right\}\right).$$

Therefore, we have

$$|S_1| = n_a'n_b' - R_{aa}R_{bb} \leq m_a n_b' + m_b n_a' \leq \frac{\alpha mn}{K}.$$

Similarly, we have $|S_2|\leq \beta mn/K$. Thus, $\mathrm{Var}\left(\boldsymbol{\Delta}\widetilde{O}_{ab}\right) \leq (\alpha+\beta)mnp/K$. A similar argument gives that for any $a\in[K]$, $\mathrm{Var}\left(\boldsymbol{\Delta}\widetilde{O}_{ab}\right) \leq (\alpha+\beta)mnp/K$ also holds. Then, by a union bound and Bernstein inequality, we have

$$\begin{aligned}
&\mathbb{P}\left\{\max_{Z\in S_\alpha}\|X(Z)-X(Z^*)\|_\infty - (\alpha+\beta)mn(p-q)/K \geq 0\right\}\\
\leq & 2K^2\sum_{Z:m<cn/K}\binom{n}{m}(K-1)^m\exp\left(-(1-o(1))mnI/K\right)\\
\leq & n\exp(-(1-o(1))nI/K),
\end{aligned}$$

for some constant c, which leads to Lemma 29.

Since $\mathrm{Var}\left(\boldsymbol{\Delta}\widetilde{O}_{ab}\right) \leq (\alpha+\beta)n^2p/K$ for any $a,b\in[K]$, we also have

$$\begin{aligned}
&\mathbb{P}\left\{\max_{Z\in S_\alpha}\|X(Z)-X(Z^*)\|_\infty \geq \bar{\varepsilon}n^2(p-q)\right\}\\
\leq & 2K^2K^n\exp\left(-(1-o(1))\frac{\bar{\varepsilon}^2Kn^2(p-q)^2}{2(\alpha+\beta)^2p}\right)\\
\leq & \exp(-n),
\end{aligned}$$

as long as $\bar{\varepsilon}^2 nI \to \infty$, which leads to Lemma 30.

For any sequences $\gamma, \theta \to 0$ satisfying $\theta^2\gamma nI \to \infty$, we also have

$$\mathbb{P}\left\{\|X(Z) - X(Z^*)\|_\infty \geq \theta mn(p-q)\right\} \leq 2K^2\left(\exp\left(-\frac{\theta^2 m^2 n^2 (p-q)^2}{2(\alpha+\beta)mnp/K}\right) + \exp\left(-\frac{3\theta mn(p-q)}{2}\right)\right).$$

It follows that

$$\begin{aligned}
&\mathbb{P}\left\{\max_{Z\in S_\alpha : m>\gamma n} \|X(Z) - X(Z^*)\|_\infty \leq \theta mn(p-q)\right\} \\
\leq& C' \sum_{m>\gamma n} \binom{n}{m} K^m \exp(-C\theta^2 mnI) \\
\leq& C' \sum_{m>\gamma n} \left(\frac{Ke}{\gamma}\right)^m \exp(-C\theta^2 mnI) \\
\leq& C' \sum_{m>\gamma n} \exp\left(-m\left(C\theta^2 nI - \log\frac{Ke}{\gamma}\right)\right) \\
\leq& \exp(-n).
\end{aligned} \tag{82}$$

The last inequality holds since $\theta^2\gamma nI \to \infty$ and thus $\theta^2 nI \gg 1/\gamma \gg \log(1/\gamma)$. It directly leads to Lemma 31. ■

Recall that for any $a, a' \in [K]$,

$$X_{aa'}(Z) = O_{aa'}(Z) - \mathbb{E}\left[O_{aa'}(Z)\right], \quad \boldsymbol{\Delta} O_{aa'} = O_{aa'}(Z) - O_{aa'}(Z'),$$

it follows that

$$X_{aa'}(Z) - X_{aa'}(Z') = \boldsymbol{\Delta} O_{aa'} - \mathbb{E}\left[\boldsymbol{\Delta} O_{aa'}\right].$$

We state the following two lemmas and the proofs will be presented together.

Lemma 32 *Let $N = |\mathcal{B}(Z) \cap S_\alpha|$, and we have*

$$\mathbb{P}\left\{\max_{Z\in S_\alpha} \frac{1}{N} \sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha} \sum_{a\leq a'} \left|\boldsymbol{\Delta} O_{aa'} - \mathbb{E}\left[\boldsymbol{\Delta} O_{aa'}\right]\right| \geq 10n(p-q)\right\} \leq n\exp(-2\bar{n}I).$$

Lemma 33 *Let $N = |\mathcal{B}(Z) \cap S_\alpha|$. For any positive sequence $\gamma \to 0$, $\theta \to 0$, and $\theta^2\gamma nI \to \infty$, we have*

$$\mathbb{P}\left\{\max_{Z\in S_\alpha : m>\gamma n} \frac{1}{N} \sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha} \sum_{a\leq a'} \left|\boldsymbol{\Delta} O_{aa'} - \mathbb{E}\left[\boldsymbol{\Delta} O_{aa'}\right]\right| \geq \theta n(p-q)\right\} \leq \exp(-n).$$

Proof [Proofs of Lemma 32 and Lemma 33] In order to apply Bernstein inequality, we proceed to eliminate the absolute function. Note that $\boldsymbol{\Delta} O_{aa'}$ depends on both the current state Z and the next state Z'. For any $Z \in S_\alpha$, we can rewrite

$$\begin{aligned}
&\sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \sum_{a \leq a'} \left| \boldsymbol{\Delta} O_{aa'} - \mathbb{E}\left[\boldsymbol{\Delta} O_{aa'}\right] \right| \\
=& \max_{h \in \{-1,1\}^{(2K-1)\times N}} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \left\{ \sum_{a \in [K] \setminus \{b'\}} h_a(Z') \left(\boldsymbol{\Delta} O_{ab} - \mathbb{E}\left[\boldsymbol{\Delta} O_{ab}\right]\right) + \right. \\
& \left. \sum_{a \in [K] \setminus \{b\}}^{K} h_{a+K-1}(Z') \left(\boldsymbol{\Delta} O_{ab'} - \mathbb{E}\left[\boldsymbol{\Delta} O_{ab'}\right]\right) + h_{2K-1}(Z') \left(\boldsymbol{\Delta} O_{bb'} - \mathbb{E}\left[\boldsymbol{\Delta} O_{bb'}\right]\right) \right\} \\
\triangleq & \max_{h \in \{-1,1\}^{(2K-1)\times N}} S(h)
\end{aligned} \tag{83}$$

where $h \in \{-1, 1\}^{(2K-1)\times N}$ is a matrix whose elements are either 1 or -1. We use $h(Z')$ to denote a column vector of h corresponding to Z', and $h_a(Z')$ is the ath element of $h(Z')$. To prove the equality in (83) holds, we suppose the next state Z' updates one sample from a group b to another group b'. Then, $\boldsymbol{\Delta} O_{aa'} = 0$ for any $a, a' \in [K] \setminus \{b, b'\}$. Hence, for any possible choice of Z', $h(Z') \in \{-1, 1\}^{2K-1}$, and then the equality holds.

Claim: for any samples i, j, the random variable A_{ij} appears at most four times in (83). This is because A_{ij} can only appear when we update sample i or sample j. Suppose the sample i is corrected, then A_{ij} appears in $\boldsymbol{\Delta} O_{Z_i,Z_j}$ and $\boldsymbol{\Delta} O_{Z'_i,Z_j}$. Thus, the claim holds, which means the same Bernoulli variable appears at most four times in (83).

By the above claim, for any matrix h, it trivially follows that

$$V(h) = \mathrm{Var}\left(S(h)\right) \leq 16Nnp.$$

Hence, for any $Z \in S_\alpha$ and $w > 0$, by a union bound and Bernstein inequality, we have

$$\begin{aligned}
&\mathbb{P}\left\{ \frac{1}{N} \sum_{Z' \in \mathcal{B}(Z) \cap S_\alpha} \sum_{a \leq a'} \left| \boldsymbol{\Delta} O_{aa'} - \mathbb{E}\left[\boldsymbol{\Delta} O_{aa'}\right] \right| \geq w \right\} \\
=& \mathbb{P}\left\{ \max_{h \in \{-1,1\}^{(2K-1)\times N}} S(h) \geq Nw \right\} \\
\leq & \sum_{h \in \{-1,1\}^{(2K-1)\times N}} \mathbb{P}\left\{ S(h) \geq Nw \right\} \\
\leq & 2^{2KN} \left(\exp\left(-\frac{N^2 w^2}{2V} \right) + \exp\left(-\frac{3Nw}{2 \cdot 4} \right) \right),
\end{aligned}$$

where $V = 16Nnp$. Thus, by a union bound, there exists some constant C such that for $w = Cn(p-q)$, we have

$$
\begin{aligned}
&\mathbb{P}\left\{\max_{Z\in S_\alpha}\frac{1}{N}\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha}\sum_{a\le a'}\left|\boldsymbol{\Delta}O_{aa'}-\mathbb{E}\left[\boldsymbol{\Delta}O_{aa'}\right]\right|\ge w\right\}\\
\le&\sum_{m\ge1}\binom{n}{m}(K-1)^m\mathbb{P}\left\{\frac{1}{N}\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha}\sum_{a\le a'}\left|\boldsymbol{\Delta}O_{aa'}-\mathbb{E}\left[\boldsymbol{\Delta}O_{aa'}\right]\right|\ge w\right\}\\
\le&\sum_{m\ge1}\binom{n}{m}(K-1)^m 2^{2KN}\left(\exp\left(-\frac{N^2w^2}{2V}\right)+\exp\left(-\frac{3Nw}{2\cdot 4}\right)\right)\\
\le& n\exp(-2\bar{n}I).
\end{aligned}
$$

The last inequality holds since $N \ge \min\{m,\ c_{\alpha,\beta}n\}$ by Lemma 42. It is easy to verify that when $C = 10$, the result holds, which leads to Lemma 32.

For any positive sequences $\gamma, \theta \to 0$ satisfying $\gamma^2\theta nI \to \infty$, let $w = \theta n(p-q)$. Then, it follows that

$$
\begin{aligned}
&\mathbb{P}\left\{\max_{Z\in S_\alpha:m>\gamma n}\frac{1}{N}\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha}\sum_{a\le a'}\left|\boldsymbol{\Delta}O_{aa'}-\mathbb{E}\left[\boldsymbol{\Delta}O_{aa'}\right]\right|\ge w\right\}\\
\le&\sum_{m>\gamma n}\binom{n}{m}(K-1)^m\mathbb{P}\left\{\frac{1}{N}\sum_{Z'\in\mathcal{B}(Z)\cap S_\alpha}\sum_{a\le a'}\left|\boldsymbol{\Delta}O_{aa'}-\mathbb{E}\left[\boldsymbol{\Delta}O_{aa'}\right]\right|\ge w\right\}\\
\le& C'\sum_{m\ge1}\binom{n}{m}(K-1)^m 2^{2KN}\exp\left(-C\theta^2 NnI\right)\\
\le&\exp(-n),
\end{aligned}
$$

where the last inequality holds by the same argument for (82) and Lemma 42. Thus, the proof of Lemma 33 is complete. ■

5.11 Proofs of auxiliary lemmas

Lemma 34 *When the connectivity probabilities are unknown, under the events $\mathcal{E}_1(\bar{\varepsilon})$ and $\mathcal{E}_3(\bar{\varepsilon})$ defined in (40), if $p \asymp q$, then we have*

$$
\max_{Z\in S_\alpha}\left|\log\frac{\Pi(Z|A)}{\Pi(Z^*|A)}-\left(Q_{LM}(Z,A)-Q_{LM}(Z^*,A)\right)\right|\le C_{LM}
$$

for some constant C_{LM}.

Proof When the connectivity probabilities are unknown, by (5) we have

$$\log \Pi(Z|A) = \sum_{a\leq b} \log \mathrm{Beta}(O_{ab}(Z)+\kappa_1, n_{ab}(Z)-O_{ab}(Z)+\kappa_2) + Const.$$

By Lemma 53, we have

$$\begin{aligned}
&\log \frac{\Pi(Z|A)}{\Pi(Z^*|A)} \qquad (84)\\
\leq& Q_{LM}(Z,A) - Q_{LM}(Z^*,A) + \\
&(\kappa_1+\kappa_2)^2 \cdot \sum_{a\leq b}\left(\frac{1}{O_{ab}(Z^*)+\kappa_1} + \frac{1}{n_{ab}(Z^*)-O_{ab}(Z^*)+\kappa_2} + \frac{1}{n_{ab}(Z)+\kappa_1+\kappa_2}\right) + \qquad (85)\\
&(\kappa_1+\kappa_2+2)\sum_{a\leq b}\left(\left|\log\frac{O_{ab}(Z)+\kappa_1}{O_{ab}(Z^*)+\kappa_1}\right| + \left|\log\frac{n_{ab}(Z)-O_{ab}(Z)+\kappa_2}{n_{ab}(Z^*)-O_{ab}(Z^*)+\kappa_2}\right| + \left|\log\frac{n_{ab}(Z)+\kappa_1+\kappa_2}{n_{ab}(Z^*)+\kappa_1+\kappa_2}\right|\right). \qquad (86)
\end{aligned}$$

Recall that Z^* is the true label assignment, and $\boldsymbol{\Delta}\widetilde{O}_{ab} = O_{ab}(Z) - O_{ab}(Z^*)$ for any $a, b \in [K]$. Under the events $\mathcal{E}_1(\bar{\varepsilon})$ and $\mathcal{E}_3(\bar{\varepsilon})$, we have

$$\begin{aligned}
\max_{Z\in S_\alpha}\max_{a,b} |O_{ab}(Z) - \mathbb{E}[O_{ab}(Z)]| &\leq \bar{\varepsilon} n^2(p-q),\\
\max_{Z\in S_\alpha}\max_{a,b} \left|\boldsymbol{\Delta}\widetilde{O}_{ab} - \mathbb{E}\left[\boldsymbol{\Delta}\widetilde{O}_{ab}\right]\right| &\leq \bar{\varepsilon} n^2(p-q).
\end{aligned}$$

It is easy to check that $\sum_{a\leq b}\frac{1}{O_{ab}(Z^*)+\kappa_1}$ is the dominant term in (85). Thus, we have

$$(85) \leq CK^2 \cdot \frac{1}{pn^2/K^2\alpha^2} \asymp \frac{1}{n^2 p},$$

for some constant C. Since $|\log(a/b)| \leq |a-b|/|\min\{a,b\}|$, by a similar argument, we have

$$(86) \leq C'K^2 \cdot \left(\frac{n^2p + \bar{\varepsilon}n^2 p}{n^2p/K^2\alpha^2} + \frac{n^2\alpha^2/K^2}{n^2/K^2\alpha^2}\right) \asymp 1,$$

for some constant C'. By symmetry, the same argument also applies to upper bound $\log \Pi(Z^*|A) - \log \Pi(Z|A)$. Hence, the result of Lemma 20 holds. ■

Lemma 35 *When the connectivity probabilities are unknown, under the event $\mathcal{E}_1(\bar{\varepsilon})$ defined in* (40)*, if $p \asymp q$, then we have that*

$$\max_{Z\in S_\alpha}\max_{Z'\in\mathcal{N}(Z)}\left|\log\frac{\Pi(Z|A)}{\Pi(Z'|A)} - \big(Q_{LM}(Z,A) - Q_{LM}(Z',A)\big)\right| \leq \varepsilon_{LM}, \qquad (87)$$

for some positive sequence $\varepsilon_{LM} \to 0$.

Proof Recall the definition of $\mathcal{N}(Z)$ in (35). For any label assignment $Z \in S_\alpha$ and any $Z' \in \mathcal{N}(Z)$, by Lemma 53, we have

$$\log \frac{\Pi(Z|A)}{\Pi(Z'|A)}$$

$$\leq Q_{LM}(Z,A) - Q_{LM}(Z',A) +$$

$$(\kappa_1+\kappa_2)^2 \cdot \sum_{a\leq b} \left(\frac{1}{O_{ab}(Z')+\kappa_1} + \frac{1}{n_{ab}(Z') - O_{ab}(Z') + \kappa_2} + \frac{1}{n_{ab}(Z)+\kappa_1+\kappa_2} \right) + \tag{88}$$

$$(\kappa_1+\kappa_2+2) \sum_{a\leq b} \left(\left| \log \frac{O_{ab}(Z)+\kappa_1}{O_{ab}(Z')+\kappa_1} \right| + \left| \log \frac{n_{ab}(Z) - O_{ab}(Z)+\kappa_2}{n_{ab}(Z') - O_{ab}(Z')+\kappa_2} \right| + \left| \log \frac{n_{ab}(Z)+\kappa_1+\kappa_2}{n_{ab}(Z')+\kappa_1+\kappa_2} \right| \right). \tag{89}$$

Under the event $\mathcal{E}_1(\bar{\varepsilon})$ defined in (40), we have

$$\max_{Z\in S_\alpha} \max_{a,b\in[K]} \left| O_{ab}(Z) - \mathbb{E}\left[O_{ab}(Z)\right] \right| \leq \bar{\varepsilon} n^2 (p-q).$$

Since $\max_{a,b\in[K]} \mathbb{E}\left[O_{ab}(Z)\right] \leq n_{ab}(Z)p = \mathcal{O}\left(n^2p/K^2\right)$, it follows that under the event $\mathcal{E}_1(\bar{\varepsilon})$, $O_{ab}(Z) = \mathcal{O}\left(n^2 p\right)$. Since $\sum_{a\leq b} \frac{1}{O_{ab}(Z')+\kappa_1}$ is the dominant term in (88), there exists some constant C such that

$$(88) \leq \frac{C}{n^2/K^2 \cdot p} K^2 \lesssim \frac{1}{n^2 p}.$$

By Lemma 43, $|\boldsymbol{\Delta} O_{ab}| \leq 2n\alpha/K$ and $|\boldsymbol{\Delta} n_{ab}| \leq 2n\alpha/K$ always hold. Since $|\log(x/y)| \leq |x-y| / \min\{x,y\}$ for $x, y > 0$, we have that

$$(89) \leq C' K^2 \left(\frac{n}{n^2p/K^2} + \frac{n}{n^2/K^2} \right) \lesssim \frac{1}{np},$$

for some constant C'. Hence, under the event $\mathcal{E}_1(\bar{\varepsilon})$, we have that

$$\max_{Z\in S_\alpha} \max_{Z'\in\mathcal{N}(Z)} \left| \log \frac{\Pi(Z|A)}{\Pi(Z'|A)} - \left(Q_{LM}(Z,A) - Q_{LM}(Z',A) \right) \right| \leq \varepsilon_{LM},$$

for some positive sequence $\varepsilon_{LM} \to 0$. The absolute sign is due to the symmetry. ■

Lemma 36 *For a general $K \geq 2$, define $\widetilde{B}_{ab} = \mathbb{E}\left[O_{ab}(Z)\right]/n_{ab}(Z)$, and we have*

$$\frac{\left|\widetilde{B}_{aa'} - B_{aa'}\right|}{p-q} = \begin{cases} \dfrac{\sum_{k\neq l} R_{ak}R_{al}}{n'_a(n'_a-1)}, & \text{if } a = a', \\[2ex] \dfrac{\sum_k R_{ak}R_{a'k}}{n'_a n'_{a'}}, & \text{if } a \neq a'. \end{cases}$$

It follows that

$$\|\widetilde{B} - B\|_\infty = \max_{a,b\in[K]} \left|\widetilde{B}_{ab} - B_{ab}\right| \leq \frac{2K\alpha m}{n}(p-q).$$

Proof We split the proof of Lemma 36 into two cases and calculate the results based on (44).

Case 1: $a = b$**.** It follows that

$$\frac{\left|\widetilde{B}_{aa} - B_{aa}\right|}{p-q} = \left|\frac{(RBR^T)_{aa} - pn'_a}{n'_a(n'_a-1)} - p\right| \Big/ (p-q) = \frac{\sum_{k\neq l} R_{ak}R_{al}}{n'_a(n'_a-1)} \leq \frac{2m_a}{n'_a} \leq \frac{2K\alpha m_a}{n} \leq \frac{2K\alpha m}{n},$$

where the first inequality holds trivially by analyzing the cases with $m_a = 0$, $m_a = 1$, and $m_a \geq 2$, respectively. The second inequality is by the definition of S_α.

Case 2: $a \neq b$**.** In this case, we have

$$\begin{aligned}\frac{\left|\widetilde{B}_{ab} - B_{ab}\right|}{p-q} &= \left|\frac{(RPR^T)_{ab}}{n'_a \cdot n'_b} - q\right| \Big/ (p-q) = \frac{\sum_k R_{ak}R_{bk}}{n'_a \cdot n'_b} \leq \frac{R_{aa}R_{ba} + R_{bb}R_{ab} + m_a m_b}{n'_a \cdot n'_b} \\ &\leq \frac{n'_a \cdot m_b + n'_b \cdot m_a}{n'_a \cdot n'_b} \leq \frac{K\alpha m_a + K\alpha m_b}{n} \leq \frac{K\alpha m}{n}.\end{aligned}$$

The result simply follows by combining two cases. ∎

Lemma 37 *Let $\bar{\varepsilon}, \gamma$ be any two positive sequences satisfying that $\gamma, \bar{\varepsilon} \to 0$ and $\bar{\varepsilon}^2 nI \to \infty$. Denote $\widehat{B}_{ab} = O_{ab}(Z)/n_{ab}(Z)$. Under the events $\mathcal{E}_1(\bar{\varepsilon})$ and $\mathcal{E}_2$ defined in (40), for any $Z \in S_\alpha$ with $m \leq \gamma n$, we have*

$$\sum_{a\leq b}\left|\boldsymbol{\Delta}\widetilde{O}_{ab}\log\left(\frac{\widehat{B}_{ab}}{B_{ab}}\right) + (\boldsymbol{\Delta}\widetilde{n}_{ab} - \boldsymbol{\Delta}\widetilde{O}_{ab})\log\left(\frac{1-\widehat{B}_{ab}}{1-B_{ab}}\right)\right| \lesssim (\bar{\varepsilon}+\gamma)mnI. \tag{90}$$

Proof By Lemma 55, we have

$$(90) \leq \underbrace{\sum_{a\leq b}\left|\boldsymbol{\Delta}\widetilde{O}_{ab} - B_{ab}\boldsymbol{\Delta}\widetilde{n}_{ab}\right| \cdot \left|\log\frac{\widehat{B}_{ab}(1-B_{ab})}{B_{ab}(1-\widehat{B}_{ab})}\right|}_{(A)} + \underbrace{\sum_{a\leq b}|\boldsymbol{\Delta}\widetilde{n}_{ab}| \cdot D\left(B_{ab}\middle\|\widehat{B}_{ab}\right)}_{(B)}.$$

For any $a, b \in [K]$, we have

$$\begin{aligned}\left|\boldsymbol{\Delta}\widetilde{O}_{ab} - \boldsymbol{\Delta}\widetilde{n}_{ab}\cdot B_{ab}\right| &\leq \left|\boldsymbol{\Delta}\widetilde{O}_{ab} - \mathbb{E}\left[\boldsymbol{\Delta}\widetilde{O}_{ab}\right]\right| + \left|\mathbb{E}\left[\boldsymbol{\Delta}\widetilde{O}_{ab}\right] - \boldsymbol{\Delta}\widetilde{n}_{ab}\cdot B_{ab}\right| \\ &= |X_{ab}(Z) - X_{ab}(Z^*)| + n_{ab}(Z)\left|\widetilde{B}_{ab} - B_{ab}\right| \\ &\leq \|X(Z) - X(Z^*)\|_\infty + \left(\frac{\alpha n}{K}\right)^2 \cdot \|\widetilde{B} - B\|_\infty.\end{aligned}$$

Thus, under the events $\mathcal{E}_1(\bar{\varepsilon})$ and $\mathcal{E}_2$, by Lemma 36, it follows that

$$\max_{a,b}\left|\boldsymbol{\Delta}\widetilde{O}_{ab} - \boldsymbol{\Delta}\widetilde{n}_{ab}\cdot B_{ab}\right| \lesssim mn(p-q).$$

Under the events $\mathcal{E}_1(\bar{\varepsilon})$, by Lemma 28 and Lemma 36, we further have that for any $Z \in S_\alpha$,

$$\|\widehat{B} - B\|_\infty \le \|\widehat{B} - \widetilde{B}\|_\infty + \|\widetilde{B} - B\|_\infty \lesssim (\bar{\varepsilon} + \gamma)(p - q),$$

and thus it follows that

$$(A) \lesssim (\bar{\varepsilon} + \gamma) mnI.$$

For the term (B), under the events $\mathcal{E}_1(\bar{\varepsilon})$, since $|\boldsymbol{\Delta}\widetilde{n}_{ab}| = |n_{ab}(Z) - n_{ab}(Z^*)| \le 2mn$ trivially holds, by bounding the Kullback-Leibler divergence with χ^2-divergence, it follows that

$$(B) \le K^2 \max_{a,b\in[K]} \left\{ |\boldsymbol{\Delta}\widetilde{n}_{ab}| \cdot D\left(B_{ab} \middle\| \widehat{B}_{ab}\right) \right\} \lesssim mn \cdot \frac{\|\widehat{B} - B\|_\infty^2}{p} \lesssim (\bar{\varepsilon} + \gamma)^2 mnI.$$

By combining (A) and (B), the result directly follows. ■

Lemma 38 *Suppose $p \asymp q$. Then, we have*

$$\sum_a n_a(Z) D\left(p \middle\| \widetilde{B}_{aa}\right) \le CnI,$$

for some constant C depending on α.

Proof Recall that $n'_a = n_a(Z)$. By Lemma 36, since $\widetilde{B}_{aa} \le p$ for any $a \in [K]$, then we have

$$p - \widetilde{B}_{aa} = \frac{\sum_{k\neq l} R_{ak} R_{al}}{n'_a(n'_a - 1)}(p - q) \le \frac{2m_a}{n'_a}(p - q).$$

We upper bound the Kullback-Leibler divergence by χ^2-divergence, and it follows that

$$\begin{aligned} \sum_a n_a(Z) D\left(p \middle\| \widetilde{B}_{aa}\right) &\le \sum_a n'_a \frac{(p - \widetilde{B}_{aa})^2}{q(1-q)} \\ &\le \sum_a n'_a \frac{1}{q(1-q)} \frac{4m_a^2(p-q)^2}{(n'_a)^2} \\ &\le CnI, \end{aligned}$$

for some constant C depending on α. ■

Lemma 39 *Suppose $p, q \to 0$, $p \asymp q$. For any $x, y \in [q, p]$, we have*

$$D(x\|y) \ge \frac{(x-y)^2}{2p}.$$

Proof Suppose we fix x, and construct

$$f(y) = x \log \frac{x}{y} + (1 - x) \log \frac{1-x}{1-y} - \frac{(x-y)^2}{2p},$$

and

$$f'(y) = (y - x)\left(\frac{1}{y(1-y)} - \frac{1}{p}\right),$$

where $1/y(1-y) > 1/p$ always holds. Thus $f(y) \geq f(x) = 0$. ■

Lemma 40 *Let $\gamma \to 0$ be any positive sequence with $\gamma^2 nI \to \infty$. For $m \geq \gamma n$, we have that*

$$\sum_{a,b,k,l} R_{ak}R_{bl}D\left(B_{kl}\middle\|\widetilde{B}_{ab}\right) \geq C(\alpha,\beta,K)mnI,$$

for some constant C depending on α, β, K.

Proof By Lemma 39, $D(x\|y) \geq \frac{(x-y)^2}{2p}$. Then, we have

$$\begin{aligned}
&\sum_{a,b,k,l} R_{ak}R_{bl}D\left(B_{kl}\middle\|\widetilde{B}_{ab}\right) \\
\geq&\sum_{a}\sum_{k,l} R_{ak}R_{al}D\left(B_{kl}\middle\|\widetilde{B}_{aa}\right) \\
\geq&\frac{1}{2p}\sum_{a}\sum_{k} R_{ak}^2(p-\widetilde{B}_{aa})^2 + \frac{1}{2p}\sum_{a}\sum_{k\neq l} R_{ak}R_{al}[(p-q)-(p-\widetilde{B}_{aa})]^2. \qquad (91)
\end{aligned}$$

The first inequality holds since we only keep the terms with $b = a$. The second inequality is using Lemma 39. By Lemma 36, it follows that

$$p - \widetilde{B}_{aa} = \frac{\sum_{k\neq l} R_{ak}R_{al}}{n'_a(n'_a-1)}(p-q).$$

For simplicity, let

$$T_a = n'_a(n'_a-1), \quad B_a = \sum_{k\neq l} R_{ak}R_{al}, \quad \sum_k R_{ak}^2 = T_a + n'_a - B_a.$$

Then we have that

$$\begin{aligned}
(91) =&\frac{(p-q)^2}{2p}\sum_a \frac{1}{T_a^2}\left((T_a + n'_a - B_a)B_a^2 + B_a(T_a - B_a)^2\right) \\
=&\frac{(p-q)^2}{2p}\sum_a \frac{1}{T_a^2}\left(T_a(T_a - B_a)B_a + n'_a B_a^2\right) \\
\geq&\frac{(p-q)^2}{2p}\sum_a \frac{B_a(T_a - B_a)}{T_a}.
\end{aligned}$$

Let $x = \sum_{k \neq a} R_{ak}^2$, and thus $0 \leq x \leq m_a^2$. Then, we can write

$$\begin{aligned} B_a &= 2R_{aa}m_a + m_a^2 - x, \\ T_a - B_a &= R_{aa}^2 + x - n_a'. \end{aligned}$$

Since $B_a(T_a - B_a)$ is a quadratic function of x that is concave, it follows that

$$(91) \geq \frac{(p-q)^2}{2p} \sum_a \frac{1}{T_a} \min\left\{2R_{aa}m_a\left(R_{aa}^2 + m_a^2 - n_a'\right),\ (2R_{aa}m_a + m_a^2)(R_{aa}^2 - n_a')\right\}.$$

Claim: there exists an $a' \in [K]$ such that $R_{a'a'} \geq Cn$ and $m_{a'} \geq C'm$ for some constants C and C'. This is because of the following argument. Since $m = \sum_a m_a$, there must exist some a such that $m_a \geq m/K$. Without loss of generality, suppose $m_1 \geq m/K$. Then, there are two cases we need to consider next. Case 1: if $R_{11} \geq \frac{n}{2\beta K^2}$, then we take $a' = 1$. Case 2: if $R_{11} < \frac{n}{2\beta K^2}$, since $n_1 \geq \frac{n}{K\beta}$, it follows that $\sum_{i \neq 1} R_{i1} \geq \frac{(2K-1)n}{2K^2\beta}$. Then, there must exists some $a \neq 1$ such that $R_{a1} \geq \frac{(2K-1)n}{2K^2(K-1)\beta}$. Without loss of generality, suppose $R_{21} \geq \frac{(2K-1)n}{2K^2(K-1)\beta} \geq \frac{n}{K^2\beta}$. Then, we have $m_2 \geq R_{21} \geq \frac{n}{K^2\beta} \geq \frac{m}{K^2\beta}$, and thus $R_{22} \geq R_{21} - R_{11} > \frac{n}{2K^2\beta}$ by the definition of discrepancy matrix R. Then, we take $a' = 2$. Hence, the claim always holds.

Based on the above claim, we have that

$$\begin{aligned} (91) &\geq \frac{(p-q)^2}{2p} \frac{1}{T_{a'}} \min\left\{2R_{a'a'}m_a\left(R_{a'a'}^2 + m_{a'}^2 - n_{a'}'\right),\ (2R_{a'a'}m_{a'} + m_{a'}^2)(R_{a'a'}^2 - n_{a'}')\right\} \\ &\gtrsim mnI. \end{aligned}$$

The proof is complete. ■

Lemma 41 *Recall that*

$$\Delta(\cdot) = \sum_{a \leq b} n_{ab}(\cdot)\left(\tau\left(\frac{O_{ab}(\cdot)}{n_{ab}(\cdot)}\right) - \tau\left(\frac{\mathbb{E}\left[O_{ab}(\cdot)\right]}{n_{ab}(\cdot)}\right)\right),$$

where $\tau(x) = x\log x + (1-x)\log(1-x)$. *For any positive sequences* $\bar{\varepsilon} = \gamma \to 0$, $\theta \to 0$ *with* $\gamma^2 nI \to \infty$ *and* $\theta^2\gamma nI \to \infty$, *under the events* $\mathcal{E}_1(\bar{\varepsilon})$ *and* $\mathcal{E}_4(\gamma,\theta)$ *defined in* (40), *we have that for any* $Z \in S_\alpha$ *with* $m > \gamma n$,

$$|\Delta(Z) - \Delta(Z^*)| \leq \varepsilon mnI,$$

for some positive sequence $\varepsilon \to 0$.

Proof Recall that for any $a, b \in [K]$, $\widehat{B}_{ab}(Z) = O_{ab}(Z)/n_{ab}(Z)$ and $\widetilde{B}_{ab}(Z) = \mathbb{E}\left[O_{ab}(Z)\right]/n_{ab}(Z)$. Then, $\widetilde{B}_{ab}(Z^*) = B_{ab}$. Note that $\tau'(x) = \log\frac{x}{1-x}$, and $\tau''(x) = \frac{1}{x(1-x)}$. By Taylor expansion, it follows that for any $Z \in S_\alpha$ and any $a, b \in [K]$, there exists some $\xi_{ab}(Z) \in [\widehat{B}_{ab}(Z), \widetilde{B}_{ab}(Z)]$ such that

$$\Delta(Z) = \sum_{a \leq b} \tau'\left(\widetilde{B}_{ab}(Z)\right) \cdot X_{ab}(Z) + \underbrace{\sum_{a \leq b} \frac{X_{ab}(Z)^2}{2n_{ab}(Z)} \cdot \frac{1}{\xi_{ab}(Z)(1-\xi_{ab}(Z))}}_{Err(Z)}.$$

Similarly, since $\widetilde{B}_{ab}(Z^*) = B_{ab}$, we have $\xi_{ab}(Z^*) \in [\widehat{B}_{ab}(Z^*), B_{ab}]$ such that

$$\Delta(Z^*) = \sum_{a \leq b} \tau'(B_{ab}) \cdot X_{ab}(Z^*) + Err(Z^*).$$

We write $\widetilde{B}(Z) = \widetilde{B}$ for simplicity. Then, we have

$$\begin{aligned}
&|\Delta(Z) - \Delta(Z^*)| \\
=& \left| \sum_{a \leq b} \tau'(\widetilde{B}_{ab}) X_{ab}(Z) + \sum_{a \leq b} \tau'(B_{ab}) X_{ab}(Z^*) + Err(Z) - Err(Z^*) \right| \\
=& \underbrace{\sum_{a \leq b} \left| \tau'(\widetilde{B}_{ab}) - \tau'(B_{ab}) \right| |X_{ab}(Z)|}_{(A)} + \underbrace{\left| \sum_{a \leq b} \tau'(B_{ab}) \left(X_{ab}(Z) - X_{ab}(Z^*) \right) \right|}_{(B)} + \underbrace{|Err(Z) - Err(Z^*)|}_{(C)},
\end{aligned}$$

and we proceed to bound each term.

Under the event $\mathcal{E}_1(\gamma)$, by Lemma 36, we have that for any $Z \in S_\alpha$,

$$(A) \leq \sum_{a \leq b} \left| \log \frac{\widetilde{B}_{ab}(1 - B_{ab})}{B_{ab}(1 - \widetilde{B}_{ab})} \right| \cdot \|X(Z)\|_\infty \lesssim \frac{\|\widetilde{B} - B\|_\infty}{p} \cdot \|X(Z)\|_\infty \lesssim \gamma m n I.$$

Under the event $\mathcal{E}_4(\gamma, \theta)$, we have that for any $Z \in S_\alpha$ with $m > \gamma n$,

$$\begin{aligned}
(B) &= \left| \tau'(p) \sum_a (X_{aa}(Z) - X_{aa}(Z^*)) + \tau'(q) \sum_{a<b} (X_{ab}(Z) - X_{ab}(Z^*)) \right| \\
&= \left| (\tau'(p) - \tau'(q)) \cdot \sum_a (X_{aa}(Z) - X_{aa}(Z^*)) \right| \\
&\leq \log \frac{p(1-q)}{q(1-p)} \cdot K \|X(Z) - X(Z^*)\|_\infty \\
&\lesssim \theta m n I,
\end{aligned}$$

where the second equality holds since $\sum_{a \leq b} O_{ab}(Z) = \sum_{a \leq b} O_{ab}(Z^*)$, and thus $\sum_{a \leq b}(X_{ab}(Z) - X_{ab}(Z^*)) = 0$ always holds. Under the event $\mathcal{E}_1(\gamma)$, we have that for any $Z \in S_\alpha$,

$$\begin{aligned}
(C) &\leq 2 \max_{Z \in S_\alpha} |Err(Z)| \leq \max_{Z \in S_\alpha} \sum_{a \leq b} \left| \frac{X_{ab}(Z)^2}{n_{ab}(Z)} \cdot \frac{1}{\xi_{ab}(Z)(1 - \xi_{ab}(Z))} \right| \\
&\lesssim K^2 \frac{\|X(Z)\|_\infty}{n^2} \cdot \frac{1}{p} \lesssim \gamma^2 n^2 I.
\end{aligned}$$

Since $m > \gamma n$, it follows that under the events $\mathcal{E}_1(\gamma)$ and $\mathcal{E}_4(\gamma, \theta)$, there exists some sequence $\varepsilon \to 0$, such that for any $Z \in S_\alpha$ with $m > \gamma n$,

$$|\Delta(Z) - \Delta(Z^*)| \leq \varepsilon mnI.$$

The proof is complete. ∎

Lemma 42 *Recall that $\mathcal{B}(Z)$ is defined in* (35). *We have*

$$|S_\alpha \cap \mathcal{B}(Z)| \geq \min\left\{\frac{n}{\beta K} - \frac{n}{\alpha K} - 1, \frac{\alpha n - \beta n}{K} - 1, m\right\} \geq \min\{c_{\alpha,\beta} n, m\},$$

for some constant $c_{\alpha,\beta}$.

Proof We split the proof of Lemma 42 into three cases.

Case 1. Suppose there are totally k' groups with size $\lceil n/K\alpha \rceil$ (reaching the small size boundary), denoted as the set $\mathcal{K}'$, and $|\mathcal{K}'| = k'$. Then,

$$\begin{aligned}
k' \lceil \frac{n}{K\alpha} \rceil = \sum_{a \in \mathcal{K}'} n'_a &= \sum_{a \in \mathcal{K}'} R_{aa} + \sum_{a,b \in \mathcal{K}', a \neq b} R_{ab} + \sum_{a \in \mathcal{K}', b \notin \mathcal{K}'} R_{ab} \\
&= \sum_{a \in \mathcal{K}'} \left(n_a - \sum_{b \in \mathcal{K}' \setminus \{a\}} R_{ba} - \sum_{b \notin \mathcal{K}'} R_{ba} \right) + \sum_{a,b \in \mathcal{K}', a \neq b} R_{ab} + \sum_{a \in \mathcal{K}', b \notin \mathcal{K}'} R_{ab} \\
&= \sum_{a \in \mathcal{K}'} n_a - \sum_{a \in \mathcal{K}'} \sum_{b \notin \mathcal{K}'} R_{ba} + \sum_{a \in K', b \notin \mathcal{K}'} R_{ab} \\
&\geq k' \frac{n}{K\beta} - \sum_{a \notin \mathcal{K}', b \in \mathcal{K}'} R_{ab},
\end{aligned}$$

and thus $|\mathcal{B}(Z) \cap S_\alpha| \geq \sum_{a \notin \mathcal{K}', b \in \mathcal{K}'} R_{ab} \geq k' \frac{n}{K\beta} - \lceil \frac{n}{K\alpha} \rceil \geq \frac{n}{K\beta} - \frac{n}{K\alpha} - 1$.

Case 2. If there is at least one group with size $\lfloor \alpha n/K \rfloor$ (reaching the large size boundary), denoted as group a. It directly follows that

$$m_a = n'_a - R_{aa} \geq n'_a - n_a \geq \lfloor \frac{\alpha n}{K} \rfloor - \frac{\beta n}{K} \geq \frac{\alpha n}{K} - \frac{\beta n}{K} - 1.$$

Case 3. If $\mathcal{B}(Z) \subset S_\alpha$, then $|S_\alpha \cap \mathcal{B}(Z)| = |\mathcal{B}(Z)| = m$.

∎

Lemma 43 *Suppose the current label assignment is Z, and Z' corrects the kth sample from a misclassified group b to its true group b'. Then, for any $a, a' \in [K] \setminus \{b, b'\}$, we have*

$$
\begin{aligned}
\boldsymbol{\Delta} O_{ab} &= \sum_{i,j} A_{ij} \left(\mathbb{I}\left\{ Z_i = a, Z_j = b \right\} - \mathbb{I}\left\{ Z'_i = a, Z'_j = b \right\} \right) = \sum_i A_{ik} \mathbb{I}\left\{ Z_i = a \right\}, \\
\boldsymbol{\Delta} O_{ab'} &= \sum_{i,j} A_{ij} \left(\mathbb{I}\left\{ Z_i = a, Z_j = b' \right\} - \mathbb{I}\left\{ Z'_i = a, Z'_j = b' \right\} \right) = -\sum_i A_{ik} \mathbb{I}\left\{ Z_i = a \right\}, \\
\boldsymbol{\Delta} O_{bb} &= \sum_{i<j} A_{ij} \left(\mathbb{I}\left\{ Z_i = Z_j = b \right\} - \mathbb{I}\left\{ Z'_i = Z'_j = b \right\} \right) = \sum_i A_{ik} \mathbb{I}\left\{ Z'_i = b \right\}, \\
\boldsymbol{\Delta} O_{b'b'} &= \sum_{i<j} A_{ij} \left(\mathbb{I}\left\{ Z_i = Z_j = b' \right\} - \mathbb{I}\left\{ Z'_i = Z'_j = b' \right\} \right) = -\sum_i A_{ik} \mathbb{I}\left\{ Z_i = b' \right\}, \\
\boldsymbol{\Delta} O_{bb'} &= \sum_{i,j} A_{ij} \left(\mathbb{I}\left\{ Z_i = b, Z_j = b' \right\} - \mathbb{I}\left\{ Z'_i = b, Z'_j = b' \right\} \right) = -\boldsymbol{\Delta} O_{b'b'} - \boldsymbol{\Delta} O_{bb}, \\
\boldsymbol{\Delta} O_{aa'} &= 0.
\end{aligned}
$$

Lemma 44 *Recall that $\widetilde{B}_{aa'} = \mathbb{E}\left[O_{aa'}(Z) \right] / n_{aa'}(Z)$ for any $a, a' \in [K]$. When $K = 2$, if $(n - 2m)(n - 2\beta m)/n \to \infty$, we have that*

$$
\widetilde{B}_{11} - \widetilde{B}_{12} = \frac{\det(R)(R_{11} - R_{12})(p - q)}{n_1(n_1 - 1)n_2}(1 - o(1)),
$$

and by symmetry,

$$
\widetilde{B}_{22} - \widetilde{B}_{21} = \frac{\det(R)(R_{22} - R_{21})(p - q)}{n_2(n_2 - 1)n_1}(1 - o(1)).
$$

Proof We use b to denote one group label, and use b' to denote the other. By Lemma 36,

$$
\begin{aligned}
\widetilde{B}_{b'b'} - \widetilde{B}_{b'b} &= \left(p - \frac{\sum_{k \neq l} R_{b'k} R_{b'l}}{n'_{b'}(n'_{b'} - 1)}(p - q) \right) - \left(q + \frac{\sum_k R_{bk} R_{b'k}}{n'_b \cdot n'_{b'}}(p - q) \right) \\
&= \frac{\det(R)(R_{b'b'} - R_{b'b})(p - q)}{n'_{b'}(n'_{b'} - 1)n'_b} - \frac{\left(R_{bb} R_{b'b'} + R_{bb'} R_{b'b} \right)(p - q)}{n'_{b'}(n'_{b'} - 1)n'_b},
\end{aligned}
$$

where $R_{b'b'} - R_{b'b} = n'_{b'} - m \geq n/2\beta - m$. By Lemma 46, $\det(R) \gtrsim n(n - 2m)$. Since $R_{bb} R_{b'b'} + R_{bb'} R_{b'b} \lesssim n^2$, under the condition that $(n - 2m)(n - 2\beta m)/n \to \infty$, the second term is negligible compared to the first term. ■

Lemma 45 *For any $Z \in S_\alpha$, suppose Z' corrects one sample from a misclassified group b to its true group b'. Recall $P(Z, Z')$ is defined in (71). Then, we have*

$$
\begin{aligned}
& P(Z, Z') \\
=& -\sum_{a,l=1}^{K} \widetilde{R}_{al} \left(B_{lb'} \log \frac{\widetilde{B}_{ab'}}{\widetilde{B}_{ab}} + (1 - B_{lb'}) \log \frac{1 - \widetilde{B}_{ab'}}{1 - \widetilde{B}_{ab}} \right) \\
=& -\frac{p - q}{n_{b'}(Z)} \sum_{a \in [K]} \log \frac{\widetilde{B}_{ab'}(1 - \widetilde{B}_{ab})}{\widetilde{B}_{ab}(1 - \widetilde{B}_{ab'})} \left(\sum_{k \neq b'} R_{b'k}(R_{ab'} - R_{ak}) \right) - \sum_{a \in [K]} (n_a(Z) - \delta_{ab'}) D\left(\widetilde{B}_{ab'} \middle\| \widetilde{B}_{ab} \right),
\end{aligned}
$$

where $\widetilde{R}_{b'b'} = R_{b'b'} - 1$, *otherwise* $\widetilde{R}_{ab} = R_{ab}$. *Here,* $\delta_{ab'} = 1$ *when* $a = b'$, *otherwise* $\delta_{ab'} = 0$.

Proof We write $P(Z, Z') = \sum_{a,a'} P_{a,a'}(Z, Z')$. Since Z' updates one sample from a misclassified group b to its true group b', by Lemma 43, we have for $a \in [K] \setminus \{b, b'\}$,

$$P_{a,b}(Z, Z') + P_{a,b'}(Z, Z') = -\left(\mathbb{E}\left[\boldsymbol{\Delta} O_{ab}\right] \log \frac{\widetilde{B}_{ab'}}{\widetilde{B}_{ab}} + (\boldsymbol{\Delta} n_{ab} - \mathbb{E}\left[\boldsymbol{\Delta} O_{ab}\right]) \log \frac{1 - \widetilde{B}_{ab'}}{1 - \widetilde{B}_{ab}}\right),$$

where $\boldsymbol{\Delta} n_{ab} = n_a(Z) = n_a(Z')$, and $\mathbb{E}\left[\boldsymbol{\Delta} O_{ab}\right] = \sum_{l \in [K]} R_{al} B_{lb'}$. Then, it follows that

$$P_{a,b}(Z, Z') + P_{a,b'}(Z, Z') = -\left(\log \frac{\widetilde{B}_{ab'}}{\widetilde{B}_{ab}} \sum_{l \in [K]} R_{al} B_{lb'} + \log \frac{1 - \widetilde{B}_{ab'}}{1 - \widetilde{B}_{ab}} \sum_{l \in [K]} R_{al}(1 - B_{lb'})\right).$$

Furthermore, we have

$$\begin{aligned}P_{b,b}(Z, Z') + P_{b',b'}(Z, Z') + P_{b,b'}(Z, Z') = &-\left(\mathbb{E}\left[\boldsymbol{\Delta} O_{bb}\right] \log \frac{\widetilde{B}_{bb'}}{\widetilde{B}_{bb}} + (\boldsymbol{\Delta} n_{bb} - \mathbb{E}\left[\boldsymbol{\Delta} O_{bb}\right]) \log \frac{1 - \widetilde{B}_{bb'}}{1 - \widetilde{B}_{bb}}\right) \\ &-\left(\mathbb{E}\left[\boldsymbol{\Delta} O_{b'b'}\right] \log \frac{\widetilde{B}_{bb'}}{\widetilde{B}_{b'b'}} + (\boldsymbol{\Delta} n_{b'b'} - \mathbb{E}\left[\boldsymbol{\Delta} O_{b'b'}\right]) \log \frac{1 - \widetilde{B}_{bb'}}{1 - \widetilde{B}_{b'b'}}\right),\end{aligned}$$

where $\boldsymbol{\Delta} n_{bb} = n_b(Z') = n_b(Z) - 1$, $\boldsymbol{\Delta} n_{b'b'} = -n_{b'}(Z)$, $\mathbb{E}\left[\boldsymbol{\Delta} O_{bb}\right] = \sum_l R_{bl} B_{lb'} - B_{b'b'}$, and $\mathbb{E}\left[\boldsymbol{\Delta} O_{b'b'}\right] = \sum_l R_{b'l} B_{lb'}$. It follows that

$$\begin{aligned}&P_{b,b}(Z, Z') + P_{b',b'}(Z, Z') + P_{b,b'}(Z, Z') \\ &= -\sum_{a \in \{b,b'\}} \left(\log \frac{\widetilde{B}_{ab'}}{\widetilde{B}_{ab}} \sum_l R_{al} B_{lb'} + \log \frac{1 - \widetilde{B}_{ab'}}{1 - \widetilde{B}_{ab}} \sum_l R_{al}(1 - B_{lb'})\right) \qquad (92) \\ &\quad + B_{b'b'} \log \frac{\widetilde{B}_{b'b'}}{\widetilde{B}_{bb}} + (1 - B_{b'b'}) \log \frac{1 - \widetilde{B}_{b'b'}}{1 - \widetilde{B}_{bb}}. \qquad (93)\end{aligned}$$

Thus we have

$$P(Z, Z') = -\sum_{a,l=1}^{K} \widetilde{R}_{al} \left(B_{lb'} \log \frac{\widetilde{B}_{ab'}}{\widetilde{B}_{ab}} + (1 - B_{lb'}) \log \frac{1 - \widetilde{B}_{ab'}}{1 - \widetilde{B}_{ab}}\right),$$

where $\widetilde{R}_{b'b'} = R_{b'b'} - 1$, otherwise $\widetilde{R}_{aa'} = R_{aa'}$. It leads to the first equality of Lemma 45.

By some calculations, it follows that

$$-P(Z, Z') = \underbrace{\sum_{a,l \in [K]} \widetilde{R}_{al} \left((B_{lb'} - \widetilde{B}_{ab'}) \log \frac{\widetilde{B}_{ab'}}{\widetilde{B}_{ab}} + (\widetilde{B}_{ab'} - B_{lb'}) \log \frac{1 - \widetilde{B}_{ab'}}{1 - \widetilde{B}_{ab}}\right)}_{H_1} + \sum_{a,l \in [K]} \widetilde{R}_{al} D\left(\widetilde{B}_{ab'} \middle\| \widetilde{B}_{ab}\right),$$

where H_1 can be written as

$$H_1 = \sum_{a \in [K]} \log \frac{\widetilde{B}_{ab'}(1 - \widetilde{B}_{ab})}{\widetilde{B}_{ab}(1 - \widetilde{B}_{ab'})} \sum_{l \in [K]} \widetilde{R}_{al}(B_{lb'} - \widetilde{B}_{ab'}).$$

By Lemma 36, we have for $a \neq b'$,

$$\begin{aligned}
\sum_{l\in[K]} R_{al}(B_{lb'} - \widetilde{B}_{ab'}) &= \sum_{l\in[K]} R_{al}(B_{lb'} - q - \frac{\sum_{k\in[K]} R_{ak}R_{b'k}}{n'_a n'_{b'}}(p-q)) \\
&= R_{ab'}(p-q) - \sum_{l\in[K]} R_{al}\frac{\sum_{k\in[K]} R_{ak}R_{b'k}}{n'_a n'_{b'}}(p-q) \\
&= \frac{p-q}{n'_{b'}}\left(R_{ab'}\sum_{k\in[K]} R_{b'k} - \sum_{k\in[K]} R_{ak}R_{b'k}\right) \\
&= \frac{p-q}{n'_{b'}}\left(\sum_{k\neq b'} R_{b'k}(R_{ab'} - R_{ak})\right),
\end{aligned}$$

and for $a = b'$, we have

$$\begin{aligned}
\sum_{l\in[K]} \widetilde{R}_{b'l}(B_{lb'} - \widetilde{B}_{b'b'}) &= \sum_{l\in[K]} \widetilde{R}_{b'l}\left(B_{lb'} - q - \frac{\sum_{k\in[K]} R^2_{b'k} - n'_{b'}}{n'_{b'}(n'_{b'}-1)}\right) \\
&= \widetilde{R}_{b'b'}(p-q) - \sum_{l\in[K]} \widetilde{R}_{b'l}\frac{\sum_{k\in[K]} R^2_{b'k} - n'_{b'}}{n'_{b'}(n'_{b'}-1)} \\
&= \frac{p-q}{n'_{b'}}\left((R_{b'b'}-1)n'_{b'} - \sum_{k\in[K]} R^2_{b'k} + n'_{b'}\right) \\
&= \frac{p-q}{n'_{b'}}\left(\sum_{k\neq b'} R_{b'k}(R_{b'b'} - R_{b'k})\right).
\end{aligned}$$

Thus, the result follows by plugging the results into H_1. ■

Lemma 46 *For $K = 2$, when $m \le n/K\beta$, we have $\det(R) \ge n(n-2m)/4\alpha$.*

Proof Suppose $n_1 = n_1(Z^*)$, $n_2 = n_2(Z^*)$, and $m = d(Z, Z^*)$. Then, we can write R as

$$R = \begin{pmatrix} n_1 - \left(\frac{m}{2} - x\right) & \frac{m}{2} + x \\ \frac{m}{2} - x & n_2 - \left(\frac{m}{2} + x\right) \end{pmatrix},$$

for some $x \in [-m/2, m/2]$. Therefore,

$$\det(R) = n_1 n_2 - \frac{m}{2}n - x(n_1 - n_2) = (n - n_2)n_2 - x(n - 2n_2) - mn/2.$$

Without loss of generality, we assume $\frac{\beta n}{2} \ge n_1 \ge \frac{n}{2} \ge n_2 \ge \frac{n}{2\beta}$. Then, we have the following conditions that

$$n_1 + 2x = n - n_2 + 2x \le \frac{n\alpha}{2}, \quad n_2 - 2x \ge \frac{n}{2\alpha}.$$

Combine all these conditions, and it directly follows that

$$\begin{aligned}\det(R) &\geq \frac{n}{2}\left(1-\frac{1}{\alpha}\right)n_2 + \frac{n^2}{4\alpha} - \frac{mn}{2} \\ &\geq \frac{n}{2}\left(\frac{n}{2\beta} + \frac{n}{2\alpha}\left(1-\frac{1}{\beta}\right) - m\right) \\ &= \frac{n}{2}\left(\frac{n}{2\beta}\left(1-\frac{1}{\alpha}\right) + \frac{n}{2\alpha} - m\right) \\ &\geq \frac{n(n-2m)}{4\alpha},\end{aligned}$$

where the last inequality holds since $m \leq \frac{n}{2\beta}$. ■

Lemma 47 *When $K=2$, for any $Z \in S_\alpha$ with $d(Z, Z^*) = m$, denote $\varepsilon_m = 1 - 2m/n$ for simplicity. If $(n-2m)(n-2\beta m)/n \to \infty$ and $(1-2m/n)^3 nI \to \infty$, then we have*

$$\min_{Z' \in B(Z) \cap S_\alpha} -P(Z, Z') \geq \frac{\varepsilon_m^3 nI}{2\alpha^2} \max\left\{1 - \beta + 1/\alpha, \varepsilon_m\right\}(1 - o(1)).$$

Proof Suppose $Z' \in \mathcal{B}(Z) \cap S_\alpha$ is corrects one sample from a misclassified group b to its true group b'. Decompose $-P(Z, Z') = (A) + (B)$ by Lemma 45, where

$$\begin{aligned}(A) &= \frac{p-q}{n_{b'}(Z)} \sum_{a \in [K]} \log \frac{\widetilde{B}_{ab'}(1-\widetilde{B}_{ab})}{\widetilde{B}_{ab}(1-\widetilde{B}_{ab'})} \left(\sum_{k \neq b'} R_{b'k}(R_{ab'} - R_{ak})\right) \\ &= \frac{R_{b'b}(p-q)}{n'_{b'}} \sum_{a \in \{b, b'\}} \left[\log\left(\frac{\widetilde{B}_{ab'}(1-\widetilde{B}_{ab})}{\widetilde{B}_{ab}(1-\widetilde{B}_{ab'})}\right)(R_{ab'} - R_{ab})\right]\end{aligned}$$

and

$$(B) = \sum_{a \in [K]} (n_a(Z) - \delta_{ab'}) D\left(\widetilde{B}_{ab'} \middle\| \widetilde{B}_{ab}\right)$$

By Lemma 44, when $a = b'$, under the condition that $(n-2m)(n-2\beta m)/n \to \infty$, since $R_{b'b'} - R_{b'b} = n_{b'} - m > 0$, we have

$$\begin{aligned}\log\left(\frac{\widetilde{B}_{b'b'}(1-\widetilde{B}_{b'b})}{\widetilde{B}_{b'b}(1-\widetilde{B}_{b'b'})}\right)(R_{b'b'} - R_{b'b}) &\geq \frac{\widetilde{B}_{b'b'} - \widetilde{B}_{b'b}}{\widetilde{B}_{b'b'}(1-\widetilde{B}_{b'b'})} \cdot (R_{b'b'} - R_{b'b}) \\ &\geq (1-o(1))\frac{\det(R)(R_{b'b'} - R_{b'b})^2(p-q)}{n'_{b'}(n'_{b'}-1)n'_b} \cdot \frac{1}{\widetilde{B}_{b'b'}(1-\widetilde{B}_{b'b'})} \\ &\geq (1-o(1))\frac{\det(R)(R_{b'b'} - R_{b'b})^2(p-q)}{n'_{b'}(n'_{b'}-1)n'_b} \cdot \frac{1}{p},\end{aligned}$$

and a similar argument also applies to the case with $a = b$. Hence, it follows that

$$\begin{aligned}(A) &\geq \frac{R_{b'b}(p-q)^2 \det(R)}{n'_b}\left(\frac{(R_{b'b'}-R_{b'b})^2}{n'^2_{b'}n'_b}+\frac{(R_{bb}-R_{bb'})^2}{n'^2_b n'_{b'}}\right)(1-o(1)) \\ &\geq \frac{R_{b'b}(p-q)^2 \det(R)}{n'_b}\frac{(n-2m)^2}{n'_{b'}n'_b n}(1-o(1)) \\ &\geq \frac{8R_{b'b}(p-q)^2 \det(R)(n-2m)^2}{\alpha n^4}(1-o(1)),\end{aligned}$$

where the second inequality holds by Cauchy-Schwarz inequality,

$$\left(\frac{(R_{b'b'}-R_{b'b})^2}{n'_{b'}}+\frac{(R_{bb}-R_{bb'})^2}{n'_b}\right)\cdot(n'_{b'}+n'_b) \geq (R_{b'b'}-R_{b'b}+R_{bb}-R_{bb'})^2 = (n-2m)^2.$$

Third inequalities holds since $n'_{b'}n_{b'} \leq (n'_{b'}+n'_b)^2/4$ and $n'_b \leq \alpha n/2$. We proceed to lower bound the term (B). By Lemma 39, we have $D\left(x\|y\right) \geq (x-y)^2/2p$, and

$$\begin{aligned}&n'_{b'}D\left(\widetilde{B}_{b'b'}\middle\|\widetilde{B}_{b'b}\right)+n'_b D\left(\widetilde{B}_{bb'}\middle\|\widetilde{B}_{bb}\right) \\ &\geq \det(R)^2\frac{(p-q)^2}{2p}\left(\frac{(R_{b'b'}-R_{b'b})^2}{n'^3_{b'}n'^2_b}+\frac{(R_{bb}-R_{bb'})^2}{n'^3_b n'^2_{b'}}\right)(1-o(1)) \\ &\geq \frac{(p-q)^2}{2p}\det(R)^2\frac{(n-2m)^2}{n'^2_b n'^2_{b'}n}(1-o(1)) \\ &\geq \frac{8(p-q)^2\det(R)^2(n-2m)^2}{n^5 p}(1-o(1)).\end{aligned}$$

Upper bound Kullback-Leibler divergence by χ^2-divergence, and we have that $D\left(\widetilde{B}_{bb'}\middle\|\widetilde{B}_{bb}\right) \leq CI$ for some constant C. Under the condition that $(1-2m/n)^3 nI \to \infty$, we have

$$(B) \geq \frac{8(p-q)^2\det(R)^2(n-2m)^2}{n^5 p}(1-o(1)).$$

It directly follows that

$$-P(Z,Z') = (A)+(B) \geq \frac{8(p-q)^2(n-2m)^2\det(R)}{n^4 p}\left(\frac{R_{b'b}}{\alpha}+\frac{\det(R)}{n}\right)(1-o(1)). \tag{94}$$

Note that

$$\frac{n}{2\alpha} \leq R_{b'b}+R_{b'b'} \leq R_{b'b}+\frac{\beta n}{2}-(m-R_{b'b}),$$

and thus

$$R_{b'b} \geq \max\left\{\frac{m}{2}-\frac{\beta n}{4}+\frac{n}{4\alpha},\ 0\right\}. \tag{95}$$

By Lemma 46, we have

$$\begin{aligned} -P(Z,Z') &\geq \frac{(p-q)^2(n-2m)^3}{\alpha^2 n^3 p}\left[\max\left\{m-\frac{\beta n}{2}+\frac{n}{2\alpha},0\right\}+\frac{n-2m}{2}\right](1-o(1)) \\ &= \frac{(p-q)^2(n-2m)^3}{\alpha^2 n^3 p}\max\left\{\frac{n}{2}(1-\beta+1/\alpha),\frac{n-2m}{2}\right\}(1-o(1)), \end{aligned}$$

where $(p-q)^2/p \geq I$, and thus the result follows. ■

Lemma 48 *Under the conditions of Lemma 47, for $Z \in S_\alpha$ with $d(Z,Z') = m$, we have that*

$$-P(Z,Z') \gtrsim \frac{(n-2m)^2 \det(R) I}{n^3}.$$

for any $Z' \in \mathcal{B}(Z) \cap S_\alpha$.

Proof By (94), (95) and Lemma 46, we have

$$-P(Z,Z') \gtrsim \frac{(n-2m)^2 \det(R) I}{n^3}\max\{1-\beta+1/\alpha, 1-2m/n\}.$$

Since $m \leq n/2\beta$, then $1-2m/n \to 0$ only if $\beta \to 1$, and thus $1-\beta+1/\alpha$ is a constant. Hence, the result follows. ■

Lemma 49 *For $K = 2$, we have*

$$\max_{Z' \in \mathcal{B}(Z)\cap S_\alpha}\max_{a,a' \in \{1,2\}}\left|\mathbb{E}\left[\boldsymbol{\Delta} O_{aa'}\right] - \widetilde{B}_{aa'}\boldsymbol{\Delta} n_{b'b'}\right| \leq C(n-2m)(p-q)$$

for some constant C depending on α.

Proof Without loss of generality, suppose the current state is Z, and Z' corrects one sample from misclassified group 1 to true group 2. We write $n_1 = n_1(Z^*)$, $n_2 = n_2(Z^*)$ for simplicity. Then, we have

$$R_Z = \begin{pmatrix} n_1 - s & t \\ s & n_2 - t \end{pmatrix}, \quad R_{Z'} = \begin{pmatrix} n_1 - s & t-1 \\ s & n_2 - t + 1 \end{pmatrix},$$

where $t = \sum_{i=1}^n \mathbb{I}\{Z_i = 1,\ Z_i^* = 2\}$, and $s = \sum_{i=1}^n \mathbb{I}\{Z_i = 2,\ Z_i^* = 1\}$. Thus, $t + s = m$. By Lemma 43, we have

$$\begin{aligned} \boldsymbol{\Delta} O_{11} - \widetilde{B}_{11}\boldsymbol{\Delta} n_{11} &= \frac{-(n_1-s)(n_1-m)}{n_1'}, \\ \boldsymbol{\Delta} O_{22} - \widetilde{B}_{22}\boldsymbol{\Delta} n_{22} &= \frac{-s(n_2-m+1)}{n_2'-1}, \\ \boldsymbol{\Delta} O_{11} - \widetilde{B}_{12}\boldsymbol{\Delta} n_{11} &= \frac{-s(n_1-m)}{n_2'} - \frac{n_1 s + n_2 t - m}{n_1' n_2'}, \\ \boldsymbol{\Delta} O_{22} - \widetilde{B}_{12}\boldsymbol{\Delta} n_{22} &= \frac{-(n_1-s)(n_2-m)}{n_1'}, \end{aligned}$$

and it directly follows that

$$\max_{a,a'\in\{1,2\}}\left|\mathbb{E}\left[\boldsymbol{\Delta}O_{aa'}\right]-\widetilde{B}_{aa'}\boldsymbol{\Delta}n_{aa'}\right|\lesssim(n-2m)\,(p-q).$$

■

Lemma 50 *For any $Z\in S_\alpha$, write $R_Z(a,b)$ as R_{ab} for simplicity. Then, we have*

$$\sum_{a,b,k,l}R_{ak}R_{bl}D\left(B_{kl}\middle\|\widetilde{B}_{ab}\right)\lesssim mnI.$$

Proof Since $D\left(x\|y\right)\leq\frac{(x-y)^2}{y(1-y)}$ for any $x,y\in(0,1)$, we have

$$\begin{aligned}\sum_{a,b,k,l}R_{ak}R_{bl}D\left(B_{kl}\middle\|\widetilde{B}_{ab}\right)&\leq\frac{1}{q(1-q)}\sum_{a,b,k,l}R_{ak}R_{bl}(B_{kl}-\widetilde{B}_{ab})^2\\&=\frac{1}{q(1-q)}\sum_{a,b,k,l}R_{ak}R_{bl}((B_{kl}-B_{ab})+(B_{ab}-\widetilde{B}_{ab}))^2\\&\leq\frac{2}{q(1-q)}\sum_{a,b,k,l}R_{ak}R_{bl}\left((B_{kl}-B_{ab})^2+(B_{ab}-\widetilde{B}_{ab})^2\right)\\&=(A)+(B),\end{aligned}$$

where by Lemma 36,

$$(B)=\frac{1}{q(1-q)}\sum_{a,b,k,l}R_{ak}R_{bl}(\widetilde{B}_{ab}-B_{ab})^2\leq\frac{2(p-q)^2}{q(1-q)}n^2\left(\frac{2K\alpha m}{n}\right)^2\lesssim mnI,$$

and

$$(A)\leq\frac{2(p-q)^2}{q(1-q)}\left(\sum_a\sum_{k\neq l}R_{ak}R_{al}+\sum_{a\neq b}\sum_k R_{ak}R_{bk}\right)\lesssim mnI,$$

since $\sum_a\sum_{k\neq l}R_{ak}R_{al}\leq\sum_a n_a(Z)m_a\leq mn$, and a similar bound holds for $\sum_{a\neq b}\sum_k R_{ak}R_{bk}$. By combining (A) and (B), the proof is complete. ■

Lemma 51 *Let $\gamma\to0$ be any positive sequence. Under the events $\mathcal{E}_1(\bar{\varepsilon})$ and $\mathcal{E}_2$ defined in (40), for any $Z\in S_\alpha$ with $m\leq\gamma n$, we have*

$$\sum_{a<b}n_{ab}\cdot D\left(\frac{O_{ab}}{n_{ab}}\middle\|\frac{O_{ab}(Z)}{n_{ab}(Z)}\right)\leq Cm^2I$$

for some constant C only depending on K,β,α.

Proof For any $a, b \in [K]$, we have

$$\left|\frac{O_{ab}}{n_{ab}} - \frac{O_{ab}(Z)}{n_{ab}(Z)}\right| = \left|\frac{X_{ab}(Z^*)}{n_{ab}} - \frac{X_{ab}(Z)}{n_{ab}(Z)} + B_{ab} - \widetilde{B}_{ab}\right|$$
$$\le \underbrace{\left|B_{ab} - \widetilde{B}_{ab}\right|}_{(A)} + \underbrace{\left|\frac{X_{ab}(Z^*)\,(n_{ab}(Z) - n_{ab})}{n_{ab}(Z)\cdot n_{ab}}\right|}_{(B)} + \underbrace{\left|\frac{X_{ab}(Z^*) - X_{ab}(Z)}{n_{ab}(Z)}\right|}_{(C)}.$$

By Lemma 36, we have

$$(A) \le \frac{2K\alpha m}{n}(p-q) \asymp \frac{m}{n}(p-q).$$

Under the event $\mathcal{E}_1(\bar{\varepsilon})$, we have

$$(B) \le \frac{\bar{\varepsilon} n^2 (p-q)\cdot 2mn}{(n/K\alpha)^4} \asymp \bar{\varepsilon}\frac{m}{n}(p-q),$$

where $\bar{\varepsilon}$ is the positive sequence that defines the event $\mathcal{E}_1(\bar{\varepsilon})$. Under the event $\mathcal{E}_2$, we have that

$$(C) \le \frac{(\alpha+\beta)mn(p-q)/K}{(n/K\alpha)^2} \asymp \frac{m}{n}(p-q).$$

Hence, it follows that

$$\left|\frac{O_{ab}}{n_{ab}} - \frac{O_{ab}(Z)}{n_{ab}(Z)}\right| \lesssim \frac{m}{n}(p-q),$$

for all $Z \in S_\alpha$ with $m \le \gamma n$. Furthermore, under the event $\mathcal{E}_1(\bar{\varepsilon})$, we have

$$\left|\frac{O_{ab}(Z)}{n_{ab}(Z)} - B_{ab}\right| \le \left|\frac{X_{ab}(Z)}{n_{ab}(Z)}\right| + \left|\widetilde{B}_{ab} - B_{ab}\right| \lesssim (\bar{\varepsilon}+\gamma)(p-q),$$

and thus $\frac{O_{ab}(Z)}{n_{ab}(Z)} \gtrsim p$. Hence, by $D\left(x \| y\right) \le \frac{(x-y)^2}{y(1-y)}$ for any $x, y \in (0,1)$, we have that,

$$\sum_{a\le b} n_{ab}\cdot D\left(\frac{O_{ab}}{n_{ab}}\middle\|\frac{O_{ab}(Z)}{n_{ab}(Z)}\right) \lesssim K^2 n^2 \left(\frac{m}{n}\right)^2 \frac{(p-q)^2}{p} \lesssim m^2 I.$$

■

5.12 Proofs of technical lemmas

Lemma 52 *$P, \widetilde{P}$ are the probability measures defined in set Ω. Suppose there exists a subset $A \subset \Omega$ such that $\widetilde{P}(B) = P(B\cap A)/P(A)$ for any set $B \subset \Omega$. Then, we have*

$$\|\widetilde{P} - P\|_{\mathsf{TV}} \le 2P(A^c).$$

Proof It is obvious that $P(B) = P(B \cap A) + P(B \cap A^c)$. Then,

$$\begin{aligned}\|\widetilde{P} - P\|_{\mathsf{TV}} &= \max_B \left| \frac{P(B \cap A) - P(A)P(B)}{P(A)} \right| \\ &= \max_B \left| \frac{P(B \cap A)P(A^c) - P(A)P(B \cap A^c)}{P(A)} \right| \\ &\le 2P(A^c).\end{aligned}$$

■

Lemma 53 *For any positive integers x and y, for any constant $\beta > 0$, we have*

$$\log \frac{\Gamma(x+\beta)}{\Gamma(y+\beta)} \le x \log x - y \log y - (x-y) + \frac{\beta^2}{y+\beta} + (\beta+2) \left| \log \frac{y+\beta}{x+\beta} \right|.$$

Proof For any positive constants a and b, if $b-a$ is a positive integer, then it is easy to verify that

$$\begin{aligned}\log \frac{\Gamma(b)}{\Gamma(a)} &= \sum_{k=a}^{b-1} \log k \le b \log b - a \log a - (b-a), \\ \log \frac{\Gamma(b)}{\Gamma(a)} &\ge b \log b - a \log a - (b-a) - \left(\frac{1}{a} - \frac{1}{b} + \log \frac{b}{a} \right). \\ &\ge b \log b - a \log a - (b-a) - 2\log \frac{b}{a}\end{aligned}$$

Then, for any $a, b \ge 1$, if $a - b$ is an integer, then we have

$$\log \frac{\Gamma(a)}{\Gamma(b)} \le a \log a - b \log b - (a-b) + 2 \left(\log \frac{b}{a} \right)_+ .$$

Now, let $a = x + \beta$ and $b = y + \beta$. It follows that

$$\begin{aligned}&\log \frac{\Gamma(x+\beta)}{\Gamma(y+\beta)} \\ \le& x \log(x+\beta) - y \log(y+\beta) - (x-y) + \beta(\log(x+\beta) - \log(y+\beta)) + 2 \left(\log \frac{y+\beta}{x+\beta} \right)_+ \\ \le& x \log x - y \log y - (x-y) + (\beta+2) \left| \log \frac{y+\beta}{x+\beta} \right| + Err,\end{aligned}$$

where we write

$$\begin{aligned}Err =& (x \log(x+\beta) - y \log(y+\beta)) - (x \log x - y \log y) \\ =& x \log \left(\frac{x+\beta}{x} \right) + y \log \left(\frac{y}{y+\beta} \right) \\ \le& \beta - \frac{\beta y}{y+\beta} \\ =& \frac{\beta^2}{y+\beta}.\end{aligned}$$

Hence, the result follows. ■

Lemma 54 *Suppose* $x \sim \text{Bernoulli}(q)$ *and* $y \sim \text{Bernoulli}(p)$ *with* $p, q \to 0, p \asymp q$. *Then, for any constant* C, *we have that*

$$\max\left\{\mathbb{E}\left[e^{Ct^*(x-\lambda^*)}\right], \mathbb{E}\left[e^{Ct^*(y-\lambda^*)}\right]\right\} \le e^{C'I}$$

for some constant C', *and* t^*, λ^* *are defined in* (47).

Proof Since $p, q \to 0$ and $p \asymp q$, we have

$$\log\frac{1-q}{1-p} \le \log\frac{p}{q} \le \frac{p}{q} - 1, \quad \log\frac{1-q}{1-p} \ge \frac{p-q}{1-q} \ge p - q.$$

Suppose $C > 0$, and then it follows that

$$\begin{aligned}
\mathbb{E}\left[\exp\left(Ct^*(x-\lambda^*)\right)\right] &= \exp(-Ct^*\lambda^*)\left(q\exp(Ct^*) + 1 - q\right) \\
&\le \exp(-Ct^*\lambda^* + q\exp(Ct^*) - q) \\
&\le \exp\left(\frac{-\frac{C}{2}(p-q) + q(p/q)^C - q}{(\sqrt{p}-\sqrt{q})^2}(1-o(1))\cdot I\right) \\
&\le \exp\left(C' \cdot I\right),
\end{aligned}$$

for some constant C' depending on C. The other cases follow by a similar argument. ■

Lemma 55 *For any* $a, b \in (0,1)$ *and* $x, y \in \mathbb{R}$, *we have*

$$\left|x\log\frac{a}{b} + (y-x)\log\frac{1-a}{1-b}\right| \le |x - by| \cdot \left|\log\frac{a(1-b)}{b(1-a)}\right| + |y| \cdot D\left(b\|a\right).$$

Proof It follows that

$$\begin{aligned}
x\log\frac{a}{b} + (y-x)\log\frac{1-a}{1-b} &= (x - by + by)\log\frac{a}{b} + (y - by + by - x)\log\frac{1-a}{1-b} \\
&= (x - by)\log\frac{a(1-b)}{b(1-a)} - y \cdot D\left(b\|a\right).
\end{aligned}$$

Thus, the result directly follows. ■

Lemma 56 *Suppose* $X \sim \text{Bernoulli}(q)$ *and* $Y \sim \text{Bernoulli}(p)$ *for any* $p, q \in (0,1)$. *Let* $t = \frac{1}{2}\log\frac{p(1-q)}{q(1-p)}$, *and* $\lambda = \frac{1}{2t}\log\frac{1-q}{1-p}$. *Then, for any two constants* $\gamma, \alpha \ge 0$, *we have*

$$\mathbb{E}\left[e^{t(X-\lambda)}\right]^{\gamma}\mathbb{E}\left[e^{-t(Y-\lambda)}\right]^{\alpha} = \exp\left(-\frac{\alpha+\gamma}{2}I\right),$$

where $I = -2\log\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right)$.

Proof It is easy to verify that

$$e^{-2t\lambda} = \frac{1-p}{1-q} = \frac{pe^{-t}+1-p}{qe^{t}+1-q} = \frac{\mathbb{E}\left[e^{-tY}\right]}{\mathbb{E}\left[e^{tX}\right]}.$$

Hence,

$$\mathbb{E}\left[e^{t(X-\lambda)}\right]^{\gamma}\mathbb{E}\left[e^{-t(Y-\lambda)}\right]^{\alpha} = \mathbb{E}\left[e^{tX}\right]^{\gamma}\mathbb{E}\left[e^{-tY}\right]^{\alpha}e^{-t\lambda(\gamma-\alpha)} = \mathbb{E}\left[e^{t(X-Y)}\right]^{\frac{\alpha+\gamma}{2}} = \exp\left(-\frac{\alpha+\gamma}{2}I\right).$$

■

Acknowledgments

Research of CG is supported in part by NSF grant DMS-1712957, NSF grant CCF-1934931 and NSF CAREER award DMS-1847590.

Appendix A. Analysis of Heterogeneous SBM

In this appendix, we introduce heterogeneous SBM and present the results on the algorithmic complexity of the Metropolis-Hastings algorithm.

A.1 Main results

To begin with, let us first introduce the parameter space for the heterogeneous SBM, defined as

$$\begin{aligned}\Theta(n,K,p,q,\beta,C_1,C_2) = \Big\{&(B,Z^*) : Z^* : [n] \to [K], \quad \frac{n}{\beta K} \le \sum_{i=1}^{n}\mathbb{I}\left\{Z_i^* = k\right\} \le \frac{\beta n}{K}, \text{ for all } k \in [K],\\ &B = B^T = (B_{ab}) \in [0,1]^{K\times K},\\ &q - C_1(p-q) \le \max_{a\neq b} B_{ab} = q < p = \min_{a} B_{aa} \le p + C_2(p-q)\}.\end{aligned}$$

The quantity I is defined in the same way as before,

$$I = -2\log\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right).$$

Lemma 57 (Posterior strong consistency for heterogeneous SBM) *For any $\eta = o(1)$, define*

$$\widehat{\Pi}(Z|A) \propto \Pi(Z|A)\mathbb{I}\left\{\ell(Z,Z^*) \le \eta\right\}.$$

Suppose that $\liminf_{n\to\infty}\frac{\bar{n}I}{\log n} > 1$, *and $\alpha - \beta$ is a positive constant. Then, we have that*

$$\mathbb{E}\left[\widehat{\Pi}(Z \in \Gamma(Z^*)|A)\right] \ge 1 - n\exp\left(-(1-o(1))\bar{n}I\right).$$

Lemma 57 characterizes the posterior strong consistency in a constraint set under the optimal condition. Though it is not the exact posterior distribution, we will show that the Metropolis-Hastings algorithm will converge to $\widehat{\Pi}(\cdot|A)$ in polynomial time from any initial Z_0 that is is weakly consistent.

We first present a lemma that shows all label assignments sampled from the algorithm remain in a good region with high probability.

Lemma 58 (Stay in small mistake region) *Suppose we start at a fixed initializer Z_0 with $\ell(Z_0, Z^*) \leq \gamma_0$ where $\gamma_0 = o(1)$. Then, the number of misclassified nodes in any polynomial running time can be upper bounded by*

$$m = n \cdot \ell(Z, Z^*) \leq n \max\left\{\gamma_0, n^{-\tau}\right\} + \log^2 n, \tag{96}$$

with probability at least $1 - \exp(-\log^2 n)$, where $\tau > 0$ is a sufficiently small constant.

Before stating the main result, let us introduce the following notations. First, by Lemma 58, we define a good region in the clustering space as

$$\breve{\mathcal{G}}(\gamma_0) = \left\{\Gamma(Z) : \ell(Z, Z_0) \leq \max\left\{\gamma_0, n^{-\tau}\right\} + \log^2 n/n\right\},$$

and thus the clustering structure $\Gamma(Z)$ generated by the algorithm stays inside $\breve{\mathcal{G}}(\gamma_0)$ for any polynomial running time with high probability. Recall that the distribution $\breve{\Pi}(\Gamma|A) \propto \sum_{Z\in\Gamma} \Pi^{\xi}(Z|A)$, and let us define distribution constrained in the set $\breve{\mathcal{G}}(\gamma_0)$ that $\breve{\Pi}_g(\Gamma|A) \propto \breve{\Pi}(\Gamma|A)\mathbb{I}\left\{\Gamma \in \breve{\mathcal{G}}(\gamma_0)\right\}$. Here, $\breve{\Pi}_g(\Gamma|A)$ serves as our target distribution. The computational complexity is measured through ε-mixing time converging to the target distribution $\breve{\Pi}_g(\Gamma|A)$. To be specific, let $\Gamma_0 = \Gamma(Z_0)$ be the initial state of the Markov chain, and the total variation distance to the target distribution after t distributions is defined by

$$\Delta_{Z_0}(t) = \left\|\breve{P}^t(\Gamma_0, \cdot) - \breve{\Pi}_g(\cdot|A)\right\|_{\mathsf{TV}} = \frac{1}{2}\sum_{\Gamma\in\{\Gamma(Z):Z\in S_\alpha\}} \left|\breve{P}^t(\Gamma_0, \Gamma) - \breve{\Pi}_g(\Gamma|A)\right|.$$

Therefore the ε-mixing time of Algorithm 1 is defined by

$$\tau_\varepsilon(Z_0) = \min\left\{t \in \mathbb{N} : \Delta_{Z_0}(t') \leq \varepsilon \text{ for all } \ t' \geq t\right\}.$$

Now we are ready to present the main result.

Theorem 59 (Rapidly mixing for heterogeneous SBM) *Suppose* $\limsup_{n\to\infty} \log n/\bar{n}I = 1 - \varepsilon_0$. *If $\gamma_0 = o(1)$ and $\xi \geq \frac{1-\varepsilon_0}{2\varepsilon_0}$. Then, the ε-mixing time of the Metropolis-Hastings algorithm (Algorithm 1) for heterogeneous network is upper bounded by*

$$\tau_\varepsilon(Z_0) \lesssim n^2 \max\left\{\gamma_0, n^{-\tau}\right\} \cdot \left(\xi n^2 I \gamma_0 + \log(\varepsilon^{-1})\right),$$

for some sufficiently small constant τ with high probability.

By Lemma 57 and Theorem 59, we directly have the following corollary.

Corollary 60 *Under the condition of Theorem 59, for any iteration number T such that $T \geq Cn^2(n^2 I + \log(1/\varepsilon))$ for some constant C, the output Z_T of the Algorithm 1 satisfies that $Z_T \in \Gamma(Z^*)$ with high probability, or equivalently, $\ell(Z_T, Z^*) = 0$.*

A.2 Proof of Theorem 59

First of all, we are able to define all basic events as in (40) using different constants, and all events still hold with high probability. The reason is that in the proofs of Lemmas 27, 29, 30, 31, 32, and 33, we can apply Bernstein inequality bound, and the variance of each A_{ij} is still bounded by the order p by the definition of heterogeneous SBM.

Recall that B_{ab} is the connectivity probability between community a and community b for all $a, b \in [K]$. Let Π_0 denote the posterior distribution with known connectivity probabilities. Then we have

$$\log \Pi_0(Z|A) = \sum_{a \leq b} O_{ab}(Z) \log \frac{B_{ab}}{1 - B_{ab}} + n_{ab}(Z) \log(1 - B_{ab}) + Const, \quad \text{for } Z \in S_\alpha, \quad (97)$$

where $O_{ab}(Z), n_{ab}(Z)$ are defined in (5) in the paper. Under the assumption of Theorem 59, the initial misclassification error rate is $\gamma_0 = o(1)$, and by Lemma 15, it suffices to prove the result constrained in small mistake region $\breve{\mathcal{G}}(\gamma_0)$. Hence, we can construct canonical paths within $\breve{\mathcal{G}}(\gamma_0)$ following the same protocol in Section 5.2.4, and the following lemma provides an upper bound for posterior ratios along the canonical path.

Lemma 61 *Define $\mathcal{G}(\gamma_0) = \{Z : m \leq n \max\{\gamma_0, n^{-\tau}\} + \log^2 n\}$. Suppose $\gamma_0 = o(1)$, and $\xi \geq \frac{2\varepsilon_0}{1-\varepsilon_0}$. Then, we have*

$$\max_{Z \in \mathcal{G}(\gamma_0)} \min_{Z' \in \mathcal{B}(Z)} \frac{\Pi^\xi(Z|A)}{\Pi^\xi(Z'|A)} \leq \exp(-C\bar{n}I),$$

for some constant $C > 1 - \varepsilon_0$ with probability at least $1 - C_1 n^{-C_2}$, where $\mathcal{B}(Z)$ is defined in (35).

Proof We mainly follow the proof of Lemma 17. It is easy to check that Lemma 35 still holds, and thus $\frac{\Pi^\xi(Z|A)}{\Pi^\xi(Z'|A)}$ can still be bounded through the difference between likelihood modularity. Then, based on Lemma 64, the same argument in the proof of Lemma 17 from (60) to 69 still follows, and it suffices to upper bound $\mathbb{E} \exp\left(\frac{1}{2} \sum_{Z' \in \mathcal{B}(Z)} \log \frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)}\right)$, where Π_0 is defined in (97). Lemma 63 directly gives the result. ■

Lemma 61 replicates the result of Lemma 17. Thus, following the same argument in Section 5.2.4, for the constructed canonical paths ensemble $\mathcal{T}$, we have that $\rho(\mathcal{T}) \lesssim n$, and the length of the longest path $\ell(\mathcal{T}) \lesssim n \max\{\gamma_0, n^{-\tau}\}$ for some sufficiently small constant τ. At last, combined with canonical path technique in Section 5.2.4 and coupling technique in Section 5.2.5, we have that

$$\tau_\varepsilon(Z_0) \lesssim n^2 \max\left\{\gamma_0, n^{-\tau}\right\} \cdot \left(\xi \log \Pi(Z_0|A)^{-1} + \log(\varepsilon^{-1})\right),$$

for some sufficiently small constant τ. Finally, by Lemma 65, we have $\log \Pi(Z_0|A) \geq -Cn^2 I \gamma_0$ for some constant C, and it follows that

$$\tau_\varepsilon(Z_0) \lesssim n^2 \max\left\{\gamma_0, n^{-\tau}\right\} \cdot \left(n^2 I \gamma_0 + \log(\varepsilon^{-1})\right).$$

A.3 Proofs of Lemma 57 and Lemma 58

Proof [Proof of Lemma 57] It is easy to check that Lemma 34 still holds, and thus $\log \frac{\Pi(Z|A)}{\Pi(Z^*|A)}$ can be uniformly upper bounded by $Q_{LM}(Z,A) - Q_{LM}(Z^*,A) + C_{LM}$ for some constant C_{LM} and all Z with $m \leq \gamma n$, where $Q_{LM}(\cdot, A)$ is the likelihood modularity. Based on Lemma 64, the same arguments in the proof of Lemma 13 from (45) to (48) still follows, and we have

$$\mathbb{P}\left\{\max_{Z\in S_\alpha : m\leq \gamma n} \log\frac{\Pi(Z|A)}{\Pi(Z^*|A)} - \log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)} - C\gamma mnI > 0\right\} \leq n\exp\left(-(1-o(1))\bar{n}I\right).$$

for some constant C, which is the second claim in Lemma 13. By Lemma 62, the proof is complete following the same argument for the small mistake region in the proof of Theorem 1. ■

Proof [Proof of Lemma 58] The lemma replicate the result of Lemma 15 in the paper, and the only difference is the proof of Lemma 24. By using the same argument in the proof of Lemma 63 from (99) to (103), we can see that Lemma 24 still holds. Then, it follows that Lemma 19 also holds and leads to the final result, which can be proved using the same arguments in Section 5.8 and Section 15 in the paper. ■

A.4 Proofs of other lemmas

In this section, for any $a, b \in [0,1]$, we write $I(a,b) = -2\log(\sqrt{ab} + \sqrt{(1-a)(1-b)})$ for simplicity, which is the Rényi divergence of order 1/2 between Bernoulli(a) and Bernoulli(b).

Lemma 62 *For any $\alpha > \beta \geq 1$, let $Z \in S_\alpha$ be an arbitrary assignment satisfying that $d(Z,Z^*) = m$ with $0 < m < n$. Then, for the $\Pi_0(Z|A)$ defined in (97), when $m \leq \gamma n$ for any $\gamma = o(1)$, we have*

$$\mathbb{P}\left\{\Pi_0(Z|A) > \Pi_0(Z^*|A)\right\} \leq \mathbb{E}\left[\sqrt{\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}}\right] \leq \exp\left(-(1-o(1))(\bar{n}m - m^2)I\right).$$

Proof For simplicity, we write

$$O(a,b,c,d) = \begin{cases} \sum_{i<j} A_{ij}\mathbb{I}\left\{Z_i = a, Z_j = b, Z_i^* = c, Z_j^* = d\right\}, & a=b,\ c=d \\ \sum_{i,j} A_{ij}\mathbb{I}\left\{Z_i = a, Z_j = b, Z_i^* = c, Z_j^* = d\right\}, & o.w., \end{cases}$$

$$n(a,b,c,d) = \begin{cases} \sum_{i<j} \mathbb{I}\left\{Z_i = a, Z_j = b, Z_i^* = c, Z_j^* = d\right\}, & a=b,\ c=d \\ \sum_{i,j} \mathbb{I}\left\{Z_i = a, Z_j = b, Z_i^* = c, Z_j^* = d\right\}, & o.w. \end{cases}$$

Hence, $O(a,b,c,d)$ is summation of $n(a,b,c,d)$ many Bernoulli(B_{cd}) random variables. By (97), it follows that

$$\log \frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)} = \sum_{(a,b,c,d)\in\mathcal{E}} O(a,b,c,d)\log\frac{B_{ab}(1-B_{cd})}{B_{cd}(1-B_{ab})} - n(a,b,c,d)\log\frac{1-B_{cd}}{1-B_{ab}}, \tag{98}$$

where $\mathcal{E}$ contains all 6 cases of situations: $\{a=b, c<d\}$, $\{a=b, c=d\neq a\}$, $\{a<b, c=d\}$, $\{a<b, c<d, c=a, d\neq b\}$, $\{a<b, c<d, c\neq a, d=b\}$, $\{a<b, c<d, c\neq a, d\neq b\}$. All terms are independent with each other. By Lemma 66, we have

$$\mathbb{E}\left[\sqrt{\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}}\right] = \exp\left(-\frac{\sum_{(a,b,c,d)\in\mathcal{E}} I(B_{ab},B_{cd})}{2}\right),$$

We have $I(B_{ab},B_{cd})\geq 0$ for all $(a,b,c,d)\in\mathcal{E}$, and in case of $\{a=b,c<d\}$ and $\{a<b,c=d\}$, we have $I(B_{ab},B_{cd})\geq I(p,q):=I$ by definition of heterogeneous network. Then, we have

$$\mathbb{E}\left[\sqrt{\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}}\right] \leq \exp\left(-\frac{\sum_{a\in[K],c<d} n(a,a,c,d) + \sum_{a<b,c\in[K]} n(a,b,c,c)}{2} I\right).$$

By Lemma 5.3 in Zhang et al. (2016), we have that

$$\min\left\{\sum_{a\in[K],c<d} n(a,a,c,d), \sum_{a<b,c\in[K]} n(a,b,c,c)\right\} \geq (1-o(1))(\bar{n}m - m^2).$$

By Markov inequality, we have

$$\mathbb{P}\left\{\Pi_0(Z|A) > \Pi_0(Z^*|A)\right\} \leq \mathbb{E}\left[\sqrt{\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}}\right] \leq \exp\left(-(1-o(1))(\bar{n}m-m^2)I\right).$$

■

Lemma 63 *Suppose $\gamma_0 = o(1)$. Then, for any $\gamma\to 0$, we have*

$$\mathbb{E}\exp\left(\frac{1}{2}\sum_{Z'\in\mathcal{B}(Z)} \log\frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)}\right) \leq \exp(-(1-\mathcal{O}(\gamma))m\bar{n}I).$$

Proof Suppose the current state is Z, and we move one misclassified node from community a to its true community b to get Z'. Then we have

$$\log\frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)} = \sum_{k\in\{a,b\}} \boldsymbol{\Delta}O_k \log\frac{B_{kk}(1-B_{ab})}{B_{ab}(1-B_{kk})} - \boldsymbol{\Delta}n_{kk}\log\frac{1-B_{ab}}{1-B_{kk}},$$

where

$$\boldsymbol{\Delta} O_{aa} = O_{aa}(Z) - O_{aa}(Z') = \sum_{i=1}^{R_{ab}-1} y_{ib} + \sum_{i=1}^{R_{aa}} x_{ia} + \sum_{k \notin \{a,b\}} \sum_{i=1}^{R_{ak}} x_{ik},$$
$$\boldsymbol{\Delta} O_{bb} = O_{bb}(Z) - O_{bb}(Z') = -\sum_{i=1}^{R_{bb}} \tilde{y}_{ib} - \sum_{k \neq b} \sum_{i=1}^{R_{bk}} \tilde{x}_{ik},$$

where $\{x_{ik}\}_{i\geq 1}, \{\tilde{x}_{ik}\}_{i\geq 1}$ are iid draws from Bernoulli (B_{bk}) for any $k \neq b$, and $\{y_i\}_{i\geq 1}$ $\{\tilde{y}_i\}_{i\geq 1}$ are iid draws from Bernoulli (B_{bb}). For simplicity, write $t(p,q) = \frac{1}{2}\log\frac{p(1-q)}{q(1-p)}$ and $\lambda(p,q) = \frac{1}{2t}\log\frac{1-q}{1-p}$ for any $p, q \in (0,1)$. Hence, we have

$$\mathbb{E}\left[\sqrt{\frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)}}\right] = \left[e^{-t(B_{aa},B_{ab})\lambda(B_{aa},B_{ab})}\left(B_{ab}e^{t(B_{aa},B_{ab})} + 1 - B_{ab}\right)\right]^{\boldsymbol{\Delta} n_{aa}} \tag{99}$$
$$\times \left[e^{t(B_{bb},B_{ab})\lambda(B_{bb},B_{ab})}\left(B_{bb}e^{-t(B_{bb},B_{ab})} + 1 - B_{bb}\right)\right]^{\boldsymbol{\Delta} n_{bb}} \tag{100}$$
$$\times \left(\frac{B_{bb}e^{t(B_{aa},B_{ab})} + 1 - B_{bb}}{B_{ab}e^{t(B_{aa},B_{ab})} + 1 - B_{ab}}\right)^{R_{ab}-1} \times \prod_{k \notin \{a,b\}} \left(\frac{B_{bk}e^{t(B_{aa},B_{ab})} + 1 - B_{bk}}{B_{ab}e^{t(B_{aa},B_{ab})} + 1 - B_{ab}}\right)^{R_{ak}} \tag{101}$$
$$\times \prod_{k \neq b} \left(\frac{B_{bk}e^{-t(B_{bb},B_{ab})} + 1 - B_{bk}}{B_{bb}e^{-t(B_{bb},B_{ab})} + 1 - B_{bb}}\right)^{R_{bk}}. \tag{102}$$

By Lemma 66, it follows that

$$(99) = \exp\left(-\boldsymbol{\Delta} n_{aa} I(B_{aa}, B_{ab})/2\right) = \exp\left(\frac{-n_a(Z) - 1}{2} I(B_{aa}, B_{ab})\right),$$
$$(100) = \exp\left(-\boldsymbol{\Delta} n_{bb} I(B_{bb}, B_{ab})/2\right) = \exp\left(\frac{-n_b(Z)}{2} I(B_{bb}, B_{ab})\right).$$

By Lemma 68, we have

$$(101) \leq \exp\left(\mathcal{O}(1)\, m_a(Z) I(B_{aa}, B_{ab})\right), \qquad (102) \leq \exp\left(\mathcal{O}(1)\, m_b(Z) I(B_{bb}, B_{ab})\right).$$

Define $I = I(p,q)$. All combined together, by $I(B_{aa}, B_{ab}) \geq I$, $I(B_{bb}, B_{ab}) \geq I$, and we have

$$\mathbb{E}\left[\sqrt{\frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)}}\right] \leq \exp(-(1 - \mathcal{O}(\gamma))\bar{n} I).$$

Now we consider the summation over $Z' \in \mathcal{B}(Z)$. In the summation, it is easy to see that in $x_{ia}, \tilde{y}_{ib}, \tilde{x}_{ik}$ for $k \notin \{a,b\}$ are independent with all others in the summation over $Z' \in \mathcal{B}(Z)$, and $y_{ib}, \tilde{x}_{ia}, x_{ik}$ for $k \notin \{a,b\}$ appear twice in the summation. Since there are at most $\binom{m}{2}$ pairs, by Corollary 70, we directly have

$$\mathbb{E}\exp\left(\frac{1}{2}\sum_{Z'\in\mathcal{B}(Z)}\log\frac{\Pi_0(Z|A)}{\Pi_0(Z'|A)}\right)\leq\exp(-(1-\mathcal{O}(\gamma))m\bar{n}I).\tag{103}$$

■

Lemma 64 *For any label assignment* Z *such that* $m=o(n)$*, define* $\widetilde{B}_{ab}=\mathbb{E}\left[O_{ab}(Z)\right]/n_{ab}(Z)$*, and we have*

$$\|\widetilde{B}-B\|_\infty=\max_{a,b\in[K]}\left|\widetilde{B}_{ab}-B_{ab}\right|\leq\frac{Cm}{n}(p-q)$$

for some constant C*.*

Proof Recall that $n_a(Z)=\sum_k R_{ak}$, and $m_a=n_a(Z)-R_{aa}$. When $a=b$, it follows that

$$\begin{aligned}\left|\widetilde{B}_a-B_{aa}\right|&=\left|\frac{\sum_{kl}R_{ak}(B_{kl}-B_{aa})R_{al}-\sum_k R_{ak}(B_{kk}-B_{aa})}{n_a(Z)(n_a(Z)-1)}\right|\\&\lesssim(p-q)\left|\frac{\left(\sum_{kl}R_{ak}R_{al}-R_{aa}^2\right)-\left(\sum_k R_{ak}-R_{aa}\right)}{n_a(Z)(n_a(Z)-1)}\right|\\&=(p-q)\left|\frac{R_{aa}m_a+m_a(R_{aa}-1)+m_a^2}{n_a(Z)(n_a(Z)-1)}\right|\\&\lesssim(p-q)\frac{m_a}{n_a(Z)}\lesssim(p-q)\frac{m}{n}.\end{aligned}$$

The first inequality holds since $\max_a B_{aa}-\min_{a\neq b}B_{ab}\lesssim(p-q)$. By similar argument, we can get the result for the case $a\neq b$. ■

Lemma 65 *Suppose* $\gamma_0=o(1)$*, we have*

$$\log\Pi(Z_0|A)\geq -Cn^2I\gamma_0,\tag{104}$$

with probability at least $1-C_1n^{-C_2}$*, where* $\gamma_0=\ell(Z_0,Z^*)$*.*

Proof We begin with the case where connectivity probabilities are known. By (98), we have that

$$\begin{aligned}\mathbb{P}\left\{\log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}\leq -CmnI\right\}&=\mathbb{P}\left\{-\frac{1}{2}\log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}\geq C'mnI\right\}\\&\leq\mathbb{E}\left[\exp\left(-\frac{1}{2}\log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}\right)\right]\cdot\exp(-C'mnI).\end{aligned}$$

By Lemma 67 and by (98), we have

$$\mathbb{E}\left[\exp\left(-\frac{1}{2}\log\frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)}\right)\right]\leq\exp\left(C\sum_{(a,b,c,d)\in\mathcal{E}}n(a,b,c,d)I\right)=\exp(CnmI),$$

for some constant C, where $\mathcal{E}$ is defined in (98). It follows that

$$\mathbb{P}\left\{\log \frac{\Pi_0(Z|A)}{\Pi_0(Z^*|A)} \leq -CmnI\right\} \leq \exp(-C'mnI),$$

for some constants C, C'.

When connectivity probabilities are unknown, following the same argument in the proof of Lemma 7 (Section 5.9.2), the final result follows. ■

A.5 Proofs of Auxiliary Lemmas

Lemma 66 *For any $p, q \in (0,1)$, suppose $X \sim \text{Bernoulli}(q)$, and let $t = \frac{1}{2}\log\frac{p(1-q)}{q(1-p)}$, and $\lambda = \frac{1}{2t}\log\frac{1-q}{1-p}$. Then, we have*

$$\mathbb{E}\left[e^{t(X-\lambda)}\right] = \exp\left(-\frac{I}{2}\right),$$

where $I = -2\log\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right)$.

Proof

$$\begin{aligned}\mathbb{E}\left[e^{t(X-\lambda)}\right] &= e^{-t\lambda}(qe^t + 1 - q) \\ &= \sqrt{\frac{1-p}{1-q}}\left(q\sqrt{\frac{p(1-q)}{q(1-p)}} + (1-q)\right) = \exp(-I/2).\end{aligned}$$

■

Lemma 67 *For any $p, q \in (0,1)$ such that $p, q \to 0$, suppose $X \sim \text{Bernoulli}(q)$, and let $t = \frac{1}{2}\log\frac{p(1-q)}{q(1-p)}$, and $\lambda = \frac{1}{2t}\log\frac{1-q}{1-p}$. Then, we have*

$$\mathbb{E}\left[e^{-t(X-\lambda)}\right] \leq \exp\left(CI\right)$$

for some constant C, where $I = -2\log\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right)$.

Proof Suppose $Y \sim \text{Bernoulli}(p)$, and then by Lemma 66, $\mathbb{E}\left[e^{-t(Y-\lambda)}\right] = \exp(-\frac{I}{2})$. By Lemma 68, we further have

$$\frac{\mathbb{E}\left[e^{-t(X-\lambda)}\right]}{\mathbb{E}\left[e^{-t(Y-\lambda)}\right]} \leq \exp(\mathcal{O}\left(I\right)).$$

Hence, the result follows. ■

Lemma 68 *For any* $p, q, w \in (0,1)$*, suppose* $p, q \to 0, p \asymp q$*, and* $q - C_1 |p-q| \leq w \leq p + C_2 |p-q|$ *for some constants* C_1, C_2*. Let* $t = \frac{1}{2}\log\frac{p(1-q)}{q(1-p)}$*. Then, we have*

$$\frac{we^t + 1 - w}{qe^t + 1 - q} \leq \exp(\mathcal{O}(1)\, I),$$

where $I = -2\log\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right)$*.*

Proof We have

$$\left|e^t - 1\right| = \left|\sqrt{\frac{p(1-q)}{q(1-p)}} - 1\right| = \frac{(p-q)/q(1-p)}{1 + \sqrt{p(1-q)/q(1-p)}} \leq \mathcal{O}\left(\frac{|p-q|}{p}\right).$$

It follows that

$$\frac{we^t + 1 - w}{qe^t + 1 - q} = 1 + \frac{(w-q)(e^t - 1)}{q(e^t - 1) + 1} \leq 1 + \mathcal{O}\left(\frac{(p-q)^2}{p}\right) \leq \exp(CI),$$

for some constant C. ■

Lemma 69 *For any* $p, q \in (0,1)$*, suppose* $p > q$*,* $p, q \to 0, p \asymp q$*. Let* $L = q - C_1 |p-q|$*, and* $U = p + C_2 |p-q|$ *for simplicity for any constants* C_1, C_2*. Let* $t_1 = \frac{1}{2}\log\frac{p_1(1-q_1)}{q_1(1-p_1)}$*,* $t_2 = \frac{1}{2}\log\frac{p_2(1-q_2)}{q_2(1-p_2)}$*, for any* $p_1, p_2 \in [p, U]$*, and* $q_1, q_2 \in [L, q]$*. For any* $w, s \in [L, U]$*, we have*

$$\frac{\frac{w+s}{2}e^{t_1+t_2} + 1 - \frac{w+s}{2}}{(we^{t_1} + 1 - w)(se^{t_2} + 1 - s)} \leq \exp(\mathcal{O}(1)\, I),$$

where $I = -2\log\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right)$*.*

Proof By the similar argument in the proof of Lemma 68, we have

$$\left|e^{t_1} - 1\right| \vee \left|e^{t_2} - 1\right| \lesssim \frac{|p-q|}{p}.$$

We have

$$\begin{aligned}
&\left(\frac{w+s}{2}e^{t_1+t_2} + 1 - \frac{w+s}{2}\right) - (we^{t_1} + 1 - w)(se^{t_2} + 1 - s)\\
=&\frac{w+s-2ws}{2}(e^{t_1+t_2} - 1) - w(1-s)(e^{t_1} - 1) - s(1-w)(e^{t_2} - 1)\\
=&\frac{w(1-w) + s(1-s)}{2}(e^{t_1} - 1)(e^{t_2} - 1) + \frac{s(1-s) - w(1-w)}{2}(e^{t_1} - 1) + \frac{w(1-w) - s(1-s)}{2}(e^{t_2} - 1)\\
\lesssim&\frac{(p-q)^2}{p}.
\end{aligned}$$

The result directly follows. ■

Corollary 70 *Under the condition of Lemma 69, for any $w, s, v \in [L, U]$, we have*

$$\frac{ve^{t_1+t_1}+1-v}{(we^{t_1}+1-w)(se^{t_2}+1-s)} \leq \exp(\mathcal{O}(I)).$$

Proof We have

$$\frac{ve^{t_1+t_1}+1-v}{(we^{t_1}+1-w)(se^{t_2}+1-s)} = \frac{ve^{t_1+t_1}+1-v}{\frac{w+s}{2}e^{t_1+t_2}+1-\frac{w+s}{2}} \cdot \frac{\frac{w+s}{2}e^{t_1+t_2}+1-\frac{w+s}{2}}{(we^{t_1}+1-w)(se^{t_2}+1-s)},$$

and the result directly follows by Lemma 68 and Lemma 69. ∎

References

Emmanuel Abbe. Community detection and the stochastic block model: recent developments. *Manuscript*, 2016.

Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *arXiv preprint arXiv:1405.3267*, 2014.

Nathanaël Berestycki. Mixing times of markov chains: Techniques and examples. *Alea-Latin American Journal of Probability and Mathematical Statistics*, 2016.

Peter Bickel, David Choi, Xiangyu Chang, and Hai Zhang. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics*, pages 1922–1943, 2013.

Peter J Bickel and Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50): 21068–21073, 2009.

Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*, pages 223–231. IEEE, 1997.

Alain Celisse, Jean-Jacques Daudin, Laurent Pierre, et al. Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics*, 6:1847–1899, 2012.

Jingchun Chen and Bo Yuan. Detecting functional modules in the yeast protein–protein interaction network. *Bioinformatics*, 22(18):2283–2290, 2006.

Amin Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability and Computing*, 19(2):227–284, 2010.

BEAU DABBS. Markov chains and mixing times. *Manuscript*, 2009.

Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of markov chains. *The Annals of Applied Probability*, pages 36–61, 1991.

Donniell E Fishkind, Daniel L Sussman, Minh Tang, Joshua T Vogelstein, and Carey E Priebe. Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown. *SIAM Journal on Matrix Analysis and Applications*, 34(1):23–39, 2013.

Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

Junxian Geng, Anirban Bhattacharya, and Debdeep Pati. Probabilistic community detection with unknown number of communities. *Journal of the American Statistical Association*, 114(526):893–905, 2019.

Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

Anna Goldenberg, Alice X Zheng, Stephen E Fienberg, Edoardo M Airoldi, et al. A survey of statistical network models. *Foundations and Trends® in Machine Learning*, 2(2): 129–233, 2010.

Venkatesan Guruswami. Rapidly mixing markov chains: a comparison of techniques (a survey). *arXiv preprint arXiv:1603.01512*, 2016.

Mark S Handcock, Adrian E Raftery, and Jeremy M Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354, 2007.

Jake M Hofman and Chris H Wiggins. Bayesian approach to network modularity. *Physical review letters*, 100(25):258701, 2008.

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Adrian Hutter, James R Wootton, and Daniel Loss. Efficient markov chain monte carlo algorithm for the surface code. *Physical Review A*, 89(2):022326, 2014.

David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

Frank McSherry. Spectral partitioning of random graphs. In *focs*, page 529. IEEE, 2001.

Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

Sean P Meyn, Robert L Tweedie, et al. Computable bounds for geometric convergence rates of markov chains. *The Annals of Applied Probability*, 4(4):981–1011, 1994.

Jesper Møller, Anthony N Pettitt, Robert Reeves, and Kasper K Berthelsen. An efficient markov chain monte carlo method for distributions with intractable normalising constants. *Biometrika*, 93(2):451–458, 2006.

Ravi Montenegro, Prasad Tetali, et al. Mathematical aspects of mixing times in markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006.

Mark Newman. *Networks: an introduction.* Oxford university press, 2010.

Mark EJ Newman and Elizabeth A Leicht. Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, 104(23):9564–9569, 2007.

Krzysztof Nowicki and Tom A B Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American statistical association*, 96(455):1077–1087, 2001.

Dana Randall. Rapidly mixing markov chains with applications in computer science and physics. *Computing in Science & Engineering*, 8(2):30–41, 2006.

Gareth O Roberts, Jeffrey S Rosenthal, et al. Geometric ergodicity and hybrid markov chains. *Electron. Comm. Probab*, 2(2):13–25, 1997.

Karl Rohe, Sourav Chatterjee, Bin Yu, et al. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

Alistair Sinclair. Improved bounds for mixing rates of markov chains and multicommodity flow. *Combinatorics, probability and Computing*, 1(4):351–370, 1992.

Tom AB Snijders and Krzysztof Nowicki. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of classification*, 14(1):75–100, 1997.

Yuttapong Thawornwattana, Daniel Dalquen, Ziheng Yang, et al. Designing simple and efficient markov chain monte carlo proposal kernels. *Bayesian Analysis*, 2018.

Zhuowen Tu and Song-Chun Zhu. Image segmentation by data-driven markov chain monte carlo. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):657–673, 2002.

SL van der Pas, AW van der Vaart, et al. Bayesian community detection. *Bayesian Analysis*, 2017.

Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.

Dawn B Woodard, Jeffrey S Rosenthal, et al. Convergence rate of markov chain methods for genomic motif discovery. *The Annals of Statistics*, 41(1):91–124, 2013.

Keijan Wu, Naoise Nunan, John W Crawford, Iain M Young, and Karl Ritz. An efficient markov chain model for the simulation of heterogeneous soil structure. *Soil Science Society of America Journal*, 68(2):346–351, 2004.

Yun Yang, Martin J Wainwright, Michael I Jordan, et al. On the computational complexity of high-dimensional bayesian variable selection. *The Annals of Statistics*, 44(6):2497–2532, 2016.

Anderson Y Zhang, Harrison H Zhou, et al. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016.