MINIMAX RATES OF COMMUNITY DETECTION IN STOCHASTIC BLOCK MODELS

By Anderson Y. Zhang, and Harrison H. Zhou

Yale University

Recently network analysis has gained more and more attentions in statistics, as well as in computer science, probability, and applied mathematics. Community detection for the stochastic block model (SBM) is probably the most studied topic in network analysis. Many methodologies have been proposed. Some beautiful and significant phase transition results are obtained in various settings. In this paper, we provide a general minimax theory for community detection. It gives minimax rates of the mis-match ratio for a wide rage of settings including homogeneous and inhomogeneous SBMs, dense and sparse networks, finite and growing number of communities. The minimax rates are exponential, different from polynomial rates we often see in statistical literature. An immediate consequence of the result is to establish threshold phenomenon for strong consistency (exact recovery) as well as weak consistency (partial recovery). We obtain the upper bound by a range of penalized likelihood-type approaches. The lower bound is achieved by a novel reduction from a global mis-match ratio to a local clustering problem for one node through an exchangeability property.

1. Introduction. Network science [10, 23, 28, 17] has become one of the most active research areas over the past few years. It has applications in many disciplines, for example, physics [24], sociology [29], biology [4], and Internet [2]. Detecting and identifying communities is fundamentally important to understand the underlying structure of the network [12]. Many models and methodologies have been proposed for community detection from different perspectives, including RatioCut[13], Ncut [26], and spectral method [19, 25, 16] from computer science, Newman–Girvan Modularity [12] from physics, semi-definite programming [7, 14] from engineering, and maximum likelihood estimation [3, 6] from statistics.

Deep theoretical developments have been actively pursued as well. Recently, celebrated works of Mossel et al. [20, 21] and Massoulie [18] considered balanced two-community sparse networks, and discovered the threshold phenomenon for both weak and strong consistency of community detection. Fur-

1

ther extensions to slowly growing number of communities have been made in [14, 22, 8, 1]. Recently in statistical literature, theoretical properties of various methods had been investigated as well in [8, 31, 5, 9, 25, 16], usually under weaker conditions and better suited for real data applications, but the convergence rates may often be sub-optimal.

Despite recent active and significant developments in network analysis, assumptions and conclusions can be very different in different papers. There is not an integrated framework on optimal community detection. In this paper, we attempt to give a fundamental and unified understanding of the community detection problem for the Stochastic Block Model (SBM). Our framework is quite general, including homogeneous and inhomogeneous SBMs, dense and sparse networks, equal and non-equal community sizes, and finite and growing number of communities. For example, the connection probability can be as small as an order of $1/n$, or as large as a constant order, and the total number of communities can be as large as $n/\log n$. Under this framework, a sharp minimax result is obtained with an exponential rate. This result gives a clear and smooth transition from weak consistency (partial recovery) to strong consistency (exact recovery), i.e., clustering error rates from $o(1)$ to $o(n^{-1})$. As a consequence, we obtain phase transitions for non-consistency and strong consistency, under various settings, which recover the tight thresholds for phase transition in [20, 21, 22, 8].

The Stochastic Block Model, proposed by [15], is possibly the most studied model in community detection [6, 25, 16]. Consider an undirected network with totally n nodes, and K communities labeled as $\{1, 2 \ldots, K\}$. Each node is assigned to one community. Denote σ to be an assignment, and $\sigma(i)$ is the community assignment for the i-th node. Let $n_k = |\{i : \sigma(i) = k\}|$ be the size of the k-th community, for each $k \in \{1, 2, \ldots, K\}$. We observe the connectivity of the network, which is encoded into the adjacency matrix $\{A_{i,j}\}$ taking values in $\{0, 1\}^{n \times n}$. If there exists a connection between two nodes, $A_{i,j}$ is equal to 1, and 0 otherwise. We assume each $A_{i,j}$ for any $i \geq j$ to be an independent Bernoulli random variable with a success probability $\theta_{i,j}$. Let $A_{i,i} = 0$ (no self-loop) and $A_{i,j} = A_{j,i}$ (symmetry) for any i, j. In the SBM, $\{\theta_{i,j}\}$ is assumed to have a blockwise structure, in the sense that $\theta_{i,j} = \theta_{i',j'}$ if i and i' are from the same community, and so are j and j'. We require that the within-community probabilities are larger than the between-communities probabilities, as in reality individuals from the same community are often more likely to be connected.

We consider a general SBM with parameter space defined as follows,

$$\Theta(n, K, a, b, \beta) \triangleq \left\{ (\sigma, \{\theta_{i,j}\}) : \ \sigma : [n] \to [K]^n, n_k \in \left[\frac{n}{\beta K}, \frac{\beta n}{K}\right], \forall k \in [K], \ \{\theta_{i,j}\} \in [0, 1]^{n \times n}, \right.$$

$$\left. \theta_{i,j} \geq \frac{a}{n} \text{ if } \sigma(i) = \sigma(j) \text{ and } \theta_{i,j} \leq \frac{b}{n} \text{ if } \sigma(i) \neq \sigma(j), \ \theta_{i,i} = 0, \ \theta_{i,j} = \theta_{j,i}, \ \forall i \neq j \right\},$$

where $\beta \geq 1$ and is bounded. When $\beta = 1 + o(1)$, all communities have almost the same size. The parameters a/n and b/n have straightforward interpretation, with the former one as the smallest within-community probability and the later as the largest between-community probability. Throughout the paper, we assume $\epsilon < b < a$ and $a/n < 1 - \epsilon$ for a small constant $\epsilon > 0$, allowing the network to be from very sparse to very dense.

We use the mis-match ratio $r(\sigma, \hat{\sigma})$ to measure the performance of community detection. It is the proportion of nodes mis-clustered by $\hat{\sigma}$ against the truth σ. The exact definition is given in Section 2.1. The minimax rate for the parameter space $\Theta(n, K, a, b, \beta)$ in terms of the mis-match ratio loss is as follows.

THEOREM 1.1. *Assume $\frac{nI}{K \log K} \to \infty$, then*

$$(1.1) \qquad \inf_{\hat{\sigma}} \ \sup_{\Theta(n, K, a, b, \beta)} \mathbb{E} r(\sigma, \hat{\sigma}) = \begin{cases} \exp\left(-(1 + o(1))\frac{nI}{2}\right), K = 2, \\ \exp\left(-(1 + o(1))\frac{nI}{\beta K}\right), K \geq 3, \end{cases}$$

where $1 \leq \beta < \sqrt{5/3}$. In addition, if $nI/K = O(1)$, there are at least a constant proportion of nodes mis-clustered, i.e., $\inf_{\hat{\sigma}} \sup_{\Theta(n, K, a, b, \beta)} \mathbb{E} r(\sigma, \hat{\sigma}) \geq c$, for some fixed constant $c > 0$.

Note that when K is finite, $nI \to \infty$ is a sufficient condition to get Equation (1.1) since it is equivalent to $\frac{nI}{K \log K} \to \infty$. Here the key quantity I is defined as

$$(1.2) \qquad I = -2 \log\left(\sqrt{\frac{a}{n}\frac{b}{n}} + \sqrt{1 - \frac{a}{n}}\sqrt{1 - \frac{b}{n}}\right),$$

which is exactly $D_{1/2}(\text{Ber}(\frac{a}{n})\|\text{Ber}(\frac{b}{n}))$, the Rényi divergence of order $1/2$ between two Bernoulli distributions $\text{Ber}(\frac{a}{n})$ and $\text{Ber}(\frac{b}{n})$. The form of I is closely related to the Hellinger distance between those two Bernoulli probability measures. It is worth pointing out that I is equal to $(a - b)^2/(an)$, up to a constant factor, which can be interpreted as the signal-to-noise ratio,

as long as $a/n \leq 1 - \epsilon$ for some $\epsilon > 0$. In particular, when $a = o(n)$, I is equal to $(1 + o(1))(\sqrt{a} - \sqrt{b})^2/n$.

The lower bound of (1.1) is achieved by a novel reduction of the *global* minimax rate into a *local* testing problem. A range of new penalized likelihood-type methods are proposed for obtaining the upper bound. These ideas inspired the follow-up paper [11] to develop polynomial-time and rate-optimal algorithms.

Theorem 1.1 covers both dense and sparse networks. It holds for a wide range of possible values of a and b, from a constant order to an order of n. It implies that when the connectivity probability $\frac{a}{n}$ is $O(n^{-1})$, no consistent algorithm exists for community detection. The number of communities K is allowed to grow fast. It can be as large as in the order of $n/\log n$ when the connectivity probability is a constant order, in which each community contains an order of $\log n$ nodes. In addition, for finite number of communities, Theorem 1.1 shows $\frac{(a-b)^2}{a} \to \infty$ is a necessary and sufficient condition for consistent community detection, which implies consistency results in [20, 21]. It also recovers the strong consistency results in [22, 14], in which they additionally assume $a \asymp \log n$.

The minimax rate is of an exponential form, contrast to the polynomial rates in [25, 16]. The term $\frac{nI}{K}$ plays a dominating role in determining the rate. Consider the $\beta = 1$ case. Rewrite $\frac{nI}{K}$ in the form of $\rho \log n$, and then approximately we fail to recover essentially $n^{1-\rho}$ nodes. When $\rho > 1$, the network enjoys strong consistency property (exact recovery) since $n^{1-\rho} = o(1)$, i.e., every node is correctly clustered. While for $0 < \rho < 1$, it is impossible to recover the communities exactly.

Organization. The paper is organized as follows. The fundamental limits of community detection are discussed in Section 2. We present the penalized likelihood-type procedures in Section 3 to achieve the optimal rate. Some special cases of our result and the computational feasibility are discussed in Section 4. Section 5 gives the proofs of the main theorems, while Section 6 provides the proofs of key technical lemmas.

Notation. For any set B, we use $|B|$ to indicate its cardinality. For two arbitrary equal-length vectors $x = \{x_i\}$ and $y = \{y_i\}$, define the Hamming distance between x and y as $d_H(x, y) = |\{i : x_i \neq y_i\}|$, i.e., the number of coordinates with different values. For any positive integer m, we use $[m]$ to denote the set $\{1, 2, \ldots, m\}$. For any two random variables X and Y, we use $X \perp Y$ to indicate that they are independent. Denote $\mathrm{Ber}(q)$ as a Bernoulli distribution with success probability q, and $\mathrm{Bin}(m, q)$ as a binomial distri-

bution with m trials and success probability q. For two positive sequences x_n and y_n, $x_n \lesssim y_n$ means $x_n \leq c y_n$ for some constant c not depending on n. We adopt the notation $x \asymp y$ if $x_n \lesssim y_n$ and $y_n \lesssim x_n$. For any scalar z, let $\lfloor z \rfloor = \{m \in \mathbb{Z} : m \leq z\}$ and $\lceil z \rceil = \{m \in \mathbb{Z} : m \geq z\}$. We use Θ short for $\Theta(n, K, a, b, \beta)$ when there is no ambiguity to drop the index (n, K, a, b, β).

2. Fundamental Limits of Community Detection.

2.1. *Mis-match Ratio.* Before giving the exact definition of mis-match ratio, we need to introduce permutations $\Delta : [K] \to [K]$ to define equivalent partitions. For the community detection problem, there exists an identifiability issue involved with the community label. For instance, for a network with 4 nodes, assignments $(1, 1, 2, 2)$ and $(2, 2, 1, 1)$ give the same network partition. Define $\delta \circ \sigma$ as $\delta \in \Delta$ to be a new assignment with $(\delta \circ \sigma)(i) = \delta(\sigma(i))$ for each $i \in [n]$. This assignment is equivalent to σ. The mis-match ratio is used as the loss function, counting the proportion of nodes incorrectly clustered, minimizing over all the possible permutations as follows,

$$r(\sigma, \hat{\sigma}) = \inf_{\delta} d_H(\sigma, \delta \circ \hat{\sigma})/n.$$

The Hamming distance between σ and $\hat{\sigma}$ is just to count the number of entries having different values in two vectors. Thus $r(\sigma, \hat{\sigma})$ is the total number of errors divided by the total number of nodes.

2.2. *Homogeneous Stochastic Block Model.* The Stochastic Block Model assumes the network has an underlying blockwise structure. When all $\{\theta_{i,j}\}$ take two possible values a/n or b/n, depending on whether $\sigma(i) = \sigma(j)$ or not, we call the SBM *homogeneous*. In this case $\{\theta_{i,j}\}$ is unique for any given σ. The homogeneous SBM is the most studied model in computer science and probability [20, 21, 22, 14, 8]. Define

$$\Theta_1(n, K, a, b, \beta) \triangleq \left\{ (\sigma, \{\theta_{i,j}\}) \in \Theta(n, K, a, b, \beta) : \ \theta_{i,j} = \frac{a}{n} \text{ if } \sigma(i) = \sigma(j) \right.$$
$$\left. \text{and } \theta_{i,j} = \frac{b}{n} \text{ if } \sigma(i) \neq \sigma(j), \ \forall i \neq j \right\}.$$

This is a homogeneous SBM. In Θ_1, since $\{\theta_{i,j}\}$ is uniquely determined by any given σ, we may write $\sigma \in \Theta_1$ instead of $(\sigma, \{\theta_{i,j}\}) \in \Theta_1$ for simplicity. The same rule may be applied for any other homogeneous SBM.

Note that Θ_1 is *closed under permutation*. Let π be any permutation on $[n]$, then for any $\sigma \in \Theta_1$, a new assignment σ' defined as $\sigma'(i) = \sigma(\pi^{-1}(i))$ also belongs to Θ_1. This property is very helpful for us to show Θ_1 is a least

favorable subspace of Θ for community detection. A minimax lower bound over Θ_1 immediately gives a lower bound for a larger parameter space, such as Θ.

2.3. *From Global to Local.*   To establish a lower bound is challenging to work with the loss function $r(\sigma, \hat{\sigma})$ directly, as it takes infimum over an equivalent class. The mis-match ratio is a *global* property of the network. The key idea in this paper is to define a *local* loss, and to reduce the global minimax problem into a local classification for one node.

The local loss focuses only on one node. Given the truth σ and any procedure $\hat{\sigma}$, the loss of estimating the label for the i-th node is defined as follows. Let $S_\sigma(\hat{\sigma}) = \{\sigma' : \sigma' = \delta \circ \hat{\sigma}, \delta \in \Delta, \ d_H(\sigma', \sigma) = \inf_\delta d_H(\sigma, \delta \circ \hat{\sigma})\}$, and define

$$r(\sigma(i), \hat{\sigma}(i)) \triangleq \sum_{\sigma' \in S_\sigma(\hat{\sigma})} \frac{d_H(\sigma(i), \sigma'(i))}{|S_\sigma(\hat{\sigma})|},$$

for each $i \in [n]$. It is an average over all the possible $\sigma' \in S_\sigma(\hat{\sigma})$.

We will see later that it is relatively easy to study the local loss. Lemma 2.1 shows that the global loss is equal to the local one when the SBM is homogeneous and closed under permutation.

LEMMA 2.1 (Global to local).   *Let Λ be any homogeneous parameter space that is closed under permutation. Let τ be the uniform prior over all the elements in Λ. Define the global Bayesian risk as $B_\tau(\hat{\sigma}) = \frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \mathbb{E}r(\sigma, \hat{\sigma})$ and the local Bayesian risk $B_\tau(\hat{\sigma}(1)) = \frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \mathbb{E}r(\sigma(1), \hat{\sigma}(1))$ for the first node. Then*

$$\inf_{\hat{\sigma}} B_\tau(\hat{\sigma}) = \inf_{\hat{\sigma}} B_\tau(\hat{\sigma}(1)).$$

The proof of Lemma 2.1 is involved. It is established by exploiting the property of exchangeability of the parameter space Λ.

2.4. *Minimax Lower Bound.*   By constructing a least favorable case of Θ_1, we have the following lower bound for the minimax rate. We present the lower bound under milder conditions than what is stated in Theorem 1.1.

THEOREM 2.1.   *Under the assumption $\frac{nI}{K} \to \infty$, we have*

$$(2.1) \qquad \inf_{\hat{\sigma}} \sup_{\Theta_1(n,K,a,b,\beta)} \mathbb{E}r(\sigma, \hat{\sigma}) \geq \begin{cases} \exp\big(-(1+o(1))\frac{nI}{2}\big), K = 2, \\ \exp\big(-(1+o(1))\frac{nI}{\beta K}\big), K \geq 3. \end{cases}$$

If $\frac{nI}{K} = O(1)$, then $\inf_{\hat{\sigma}} \sup_{\Theta_1(n,K,a,b,\beta)} \mathbb{E}r(\sigma, \hat{\sigma}) \geq c$ for some positive constant c.

The forms of minimax rates are different for two cases $K \geq 3$ and $K = 2$. For $K \geq 3$, it is relatively more challenging to discover and distinguish small communities, rather than the communities with larger sizes. The least favorable case is the case for which at least a constant proportion of communities are of size $\frac{n}{\beta K}$. The hardness of the community detection in this setting is then determined by the ability to recover and distinguish such small communities. For $K = 2$, the least favorable setting in Θ_1 is when the two communities are of the same size. When there are only two communities, it is actually easier to recover the non-equal-sized communities, by identifying the larger one first and then labeling the remaining nodes as from the smaller one.

Approximately Equal-Sized Case: We are interested in the case with $\beta = 1 + o(1)$, where communities are almost of the same size. Networks of community sizes exactly equal to $\frac{n}{K}$ are the most studied settings [8, 21, 9]. Here we allow a small fluctuation of community sizes. Denote Θ^0 as follows,

$$\Theta^0(n, K, a, b) \triangleq \left\{ (\sigma, \{\theta_{i,j}\}) : \; \sigma : [n] \to [K]^n, n_k = (1 + o(1))\frac{nI}{K}, \forall k \in [K], \right.$$

$$\left. \theta_{i,i} = 0, \; \forall i \in [n], \; \theta_{i,j} = \frac{a}{n} \text{ if } \sigma(i) = \sigma(j) \text{ and } \theta_{i,j} = \frac{b}{n} \text{ if } \sigma(i) \neq \sigma(j), \; \forall i \neq j \right\}.$$

Note that $\Theta^0(n, K, a, b)$ is $\Theta_1(n, K, a, b, \beta)$ with $\beta = 1 + o(1)$, for which we have the following minimax lower bound.

THEOREM 2.2. *Under the assumption $\frac{nI}{K} \to \infty$, we have*

$$(2.2) \qquad \inf_{\hat{\sigma}} \sup_{\Theta^0(n,K,a,b)} \mathbb{E}r(\sigma, \hat{\sigma}) \geq \exp\left(-(1 + o(1))\frac{nI}{K} \right).$$

If $\frac{nI}{K} = O(1)$, then $\inf_{\hat{\sigma}} \sup_{\Theta^0(n,K,a,b)} \mathbb{E}r(\sigma, \hat{\sigma}) \geq c$ for some positive constant c.

Compared with Theorem 2.1, the forms of rates for $K = 2$ and $K \geq 3$ are the same in Θ^0. The proof of Theorem 2.2 is provided in Section 5. We defer the proof of Theorem 2.1 to the supplement material [30], since it is almost identical to that of Theorem 2.2.

3. Rate-optimal Procedure. We develop a range of penalized likelihood-type procedures to achieve the optimal mis-match ratio. Throughout the section σ_0 is denoted as the underlying truth.

3.1. *Penalized Likelihood-type Estimation.* The penalized procedure is based on the likelihood of a homogeneous network, although risk upper bounds are established for more general networks. If the network is homogeneous (Θ^0 and Θ_1), for which the within and between community probabilities are exactly equal to a/n and b/n respectively, the log-likelihood function is

$$L(\sigma; A) = \log(\frac{a}{n}) \sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}} + \log(1 - \frac{a}{n}) \sum_{i<j} (1 - A_{i,j}) 1_{\{\sigma(i)=\sigma(j)\}}$$

$$+ \log(\frac{b}{n}) \sum_{i<j} A_{i,j} 1_{\{\sigma(i)\neq\sigma(j)\}} + \log(1 - \frac{b}{n}) \sum_{i<j} (1 - A_{i,j}) 1_{\{\sigma(i)\neq\sigma(j)\}}.$$

Since $\sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}} + \sum_{i<j} A_{i,j} 1_{\{\sigma(i)\neq\sigma(j)\}} = \sum_{i<j} A_{i,j}$ for all σ, we can write $L(\sigma; A)$ as

$$L(\sigma; A) = \log \frac{a(1 - b/n)}{b(1 - a/n)} \sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}} - \log \frac{1 - b/n}{1 - a/n} \sum_{i<j} 1_{\{\sigma(i)=\sigma(j)\}} + f(A),$$

where $f(A)$ is a function not depending on σ. Then the maximum likelihood estimator $\hat{\sigma}^{MLE}$ is as follows,

(3.1)
$$\hat{\sigma}^{MLE} = \arg\max_{\sigma} L(\sigma; A)$$

$$= \arg\max_{\sigma} \log \frac{a(1 - b/n)}{b(1 - a/n)} \sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}} - \log \frac{1 - b/n}{1 - a/n} \sum_{i<j} 1_{\{\sigma(i)=\sigma(j)\}}.$$

The above maximum likelihood estimator can be decomposed into two terms. The first one is the sum of all $A_{i,j}$ for all i and j belonging to the same communities of σ. The second term is a penalty over the sum of sizes of all communities. There is a trade-off between these two terms. The first term is maximized when there is only one community, while the second term, a penalty term, is maximized when all community sizes are equal. However the second term is dropped when the community sizes are required to be exactly equal, i.e., the maximum likelihood estimator over all σ with a community size n/K for every community has a simpler form, $\hat{\sigma}^{MLE} = \arg\max_{\sigma} \sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}}$.

When the parameter space is not homogeneous (e.g. Θ), the maximum likelihood estimator may not have a simple form as Equation (3.1). However, we still propose to use the identical simple form of penalized likelihood estimator as Equation (3.1), i.e.,

$$\hat{\sigma} = \arg\max_{\sigma \in \Theta} T(\sigma) \text{ with } T(\sigma) \triangleq \sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}} - \lambda \sum_{i<j} 1_{\{\sigma(i)=\sigma(j)\}},$$

where we set

(3.2) $$\lambda = \log\Big(\frac{1-b/n}{1-a/n}\Big)\Big/\log\Big(\frac{a(1-b/n)}{b(1-a/n)}\Big),\ \forall K \geq 2.$$

When the parameter space is homogeneous, $\hat{\sigma}$ is identical to the maximum likelihood estimator. The optimality result will be obtained for the parameter space Θ, which allows the network to be inhomogeneous, and imbalanced in the sense that the community sizes may be different.

3.2. *Other Choices of λ.* In the previous section we provide a unified λ for the penalized likelihood-type estimation for both $K = 2$ and $K \geq 3$. It is worthwhile to point out that for $K \geq 3$ the optimality can be attained for a wide range of λ. Let $t^\star = \frac{1}{2}\log\frac{a(1-b/n)}{b(1-a/n)}$. It can be shown that $t^\star$ is the minimizer of the moment generating function for the difference of two Bernoulli variables, i.e., $t^\star = \arg\min_{t>0}\mathbb{E}e^{t(X-Y)}$, where $X \sim \text{Ber}(\frac{b}{n})$ and $Y \sim \text{Ber}(\frac{a}{n})$. It is equivalent to write λ in Equation (3.2) as follows,

$$\lambda = -\frac{1}{2t^\star}\log\Big(\frac{\frac{a}{n}\exp(-t^\star) + 1 - \frac{a}{n}}{\frac{b}{n}\exp(t^\star) + 1 - \frac{b}{n}}\Big)$$
$$= -\frac{1}{2t^\star}\log\Big(\frac{a}{n}\exp(-t^\star) + 1 - \frac{a}{n}\Big) + \frac{1}{2t^\star}\log\Big(\frac{b}{n}\exp(t^\star) + 1 - \frac{b}{n}\Big).$$

From the equation above, we can interpret λ as a weighted sum between two terms, with the first one more involving the within-community probability $\frac{a}{n}$, and the second more focusing on the between-community probability $\frac{b}{n}$. Define

(3.3) $$\lambda = \begin{cases} -\frac{1}{2t^\star}\log\big(\frac{a}{n}e^{-t^\star} + 1 - \frac{a}{n}\big) + \frac{1}{2t^\star}\log\big(\frac{b}{n}e^{t^\star} + 1 - \frac{b}{n}\big), & K = 2 \\ -\frac{w}{t^\star}\log\big(\frac{a}{n}e^{-t^\star} + 1 - \frac{a}{n}\big) + \frac{(1-w)}{t^\star}\log\big(\frac{b}{n}e^{t^\star} + 1 - \frac{b}{n}\big) & K \geq 3, \end{cases}$$

where w in any constant in $[0,1]$. We can clearly see that λ in Equation (3.2) is a special case of λ in (3.3) with $w = 1/2$. In Section 3.3, we give theoretical properties of penalized likelihood estimation for all λ in Equation (3.3).

3.3. *Minimax Upper Bound.* For the general SBM Θ, the risk upper bound of the penalized likelihood estimator, for every λ in Equation (3.3), defined in the previous section, matches the minimax lower bound given in Theorem 2.1.

THEOREM 3.1. *Assume $\frac{nI}{K\log K} \to \infty$ and $K \geq 2$. For the penalized maximum likelihood estimator $\hat{\sigma}$ with λ defined in (3.3), we have*

$$\sup_{\Theta(n,K,a,b,\beta)} \mathbb{E}r(\hat{\sigma},\sigma) \leq \begin{cases} \exp\big(-(1+o(1))\frac{nI}{2}\big), K = 2, \\ \exp\big(-(1+o(1))\frac{nI}{\beta K}\big), K \geq 3, \end{cases}$$

where $1 \leq \beta < \sqrt{5/3}$.

Approximately Equal-Sized Case: For the special parameter space Θ^0 for which community sizes are almost equal, we have the following result, a form analogous to Theorem 3.1.

THEOREM 3.2.   *Assume* $\frac{nI}{K \log K} \to \infty$ *and* $K \geq 2$. *For the penalized maximum likelihood estimator* $\hat{\sigma}$ *with* λ *defined in (3.3), we have*

$$\sup_{\Theta^0(n,K,a,b)} \mathbb{E}r(\hat{\sigma}, \sigma) \leq \exp\big(-(1+o(1))\frac{nI}{K}\big).$$

The proof of the above theorem is provided in Section 5. Due to the similarity, the proof of Theorem 3.1 is given in the supplement material [30].

4. Discussion.

4.1. *Implications on Sharp Thresholds.*   The minimax rates in Theorem 1.1 immediately imply various sharp thresholds in [20, 21, 22, 14]. By letting the rates equal to $o(1/n)$ or $o(1)$, we can get critical values for strong and weak consistency respectively, under various settings.

Special Case with $a = o(n)$ **and** $\frac{a-b}{a} = o(1)$**.** Under this scenario the difference of within-community probability and between-community probability is relatively small. Note that $I = (1 + o(1))(a - b)^2/(4an)$, which reduces the minimax result into the form of $\exp(-(1+o(1))(a-b)^2/(4aK))$. In the case of $K = 2$, Theorem 1.1 implies the results from [20, 21]. With the additional assumption $a, b = n^{o(1/\log\log n)}$, they show that $(a-b)^2/a \to \infty$ is the necessary and sufficient condition to get consistency. It also agrees with the sharp threshold for strong consistency in [22].

Special Case with Probability in the Order of $\log n$**.** Consider a more special setting where a and b are in the order of $\log n$. Denote $a = e_1 \log n$ and $b = e_2 \log n$, with $e_1 \geq e_2 > 0$. Note that I can be written as $I = (1 + o(1))(\sqrt{e_1} - \sqrt{e_2})^2 \log n/n$.

COROLLARY 4.1.   *Assume* $K = n^{o(1)}$. *There exists a strongly consistent estimator if* $\liminf_{n\to\infty} \frac{\sqrt{e_1} - \sqrt{e_2}}{\sqrt{K}} > 1$.

For any finite K, the recovery threshold is identical to the result in [14]. For the two-community case with e_1 and e_2 constants, $\sqrt{e_1} - \sqrt{e_2} > \sqrt{2}$ for exact recovery is proved in [22].

4.2. *Computational Feasibility.* The penalized likelihood estimator we propose searches all the possible assignments in the parameter space. It is computationally intractable due to the enormous cardinality of the assignments. However, the idea of reducing global estimation into local testing problem we developed in this paper establishes a guideline for constructing both efficient and optimal algorithms. Along with the *global to local* scheme, the penalized likelihood estimator can be further modified into an node-wise procedure, whose purpose is to assign the label node by node. In this way the exhaustive search over the parameter space is avoided and the computation complexity is dramatically reduced. By exploiting the local idea, in the subsequent paper [11] a two-stage algorithm is proposed to simultaneously achieve the optimal rate and computational feasibility.

5. Proofs of Main Theorems. In this section, we prove two main theorems, Theorem 2.2 and Theorem 3.2. The proofs of Theorem 2.1 and Theorem 3.1 are almost identical to those of Theorem 2.2 and Theorem 3.2. We put them in the supplement material [30].

5.1. *Proof of Theorem 2.2.* To get the lower bound for the parameter space Θ^0, we will first construct and analyze a least favorable case in term of the sizes of the communities. In particular the community sizes only take value in $\{\lfloor \frac{n}{K} \rfloor, \lfloor \frac{n}{K} \rfloor + 1, \lfloor \frac{n}{K} \rfloor - 1\}$, and the number of communities with size $\lfloor \frac{n}{K} \rfloor$ or $\lfloor \frac{n}{K} \rfloor + 1$ is of a constant proportion of K.

First consider the case with $K \geq 3$. For each pair of (n, K), the integer K can always be decomposed as the sum of three integers: $K = K_1 + K_2 + K_3$, satisfying *(1)* there exists a constant $\epsilon > 0$ such that $\epsilon K < \min(K_1, K_2) \leq \max(K_1, K_2) < (1 - \epsilon)K$; and *(2)* either of the following two conditions:

$$(5.1) \qquad \lfloor \frac{n}{K} \rfloor K_1 + (\lfloor \frac{n}{K} \rfloor + 1)K_2 + (\lfloor \frac{n}{K} \rfloor - 1)K_3 = n;$$

$$(5.2) \qquad \text{or } \lceil \frac{n}{K} \rceil K_1 + (\lceil \frac{n}{K} \rceil + 1)K_2 + (\lceil \frac{n}{K} \rceil - 1)K_3 = n;$$

When $K \geq 3$, it can be shown that such decomposition always exists. Write $n = \lfloor \frac{n}{K} \rfloor K + r$, where $0 \leq r \leq K - 1$ is an integer. If $r \geq 2\epsilon K$ and $r \leq (1 - 2\epsilon)K$ for a constant $\epsilon > 0$, we have $n = \lfloor \frac{n}{K} \rfloor (K - r) + (\lfloor \frac{n}{K} \rfloor + 1)r$, which satisfies Equation (5.1). Otherwise, if $r < 2\epsilon K$ for a small positive constant ϵ, write $n = \lfloor \frac{n}{K} \rfloor (K - 2\lfloor \frac{K}{3} \rfloor - r) + (\lfloor \frac{n}{K} \rfloor + 1)(\lfloor \frac{K}{3} \rfloor + r) + (\lfloor \frac{n}{K} \rfloor - 1)\lfloor \frac{K}{3} \rfloor$, which satisfies Equation (5.1) for ϵ sufficient small. If $K - r > 2\epsilon K$, we may argue similarly to get Equation (5.2).

Recall that we use n_k to denote the size of the k-th community for each $k \in [K]$. Without loss of generality, assume there exist $\{K_i\}_{1 \leq i \leq 3}$ satisfying

Equation (5.1) with $\epsilon K < \min(K_1, K_2) \le \max(K_1, K_2) < (1 - \epsilon)K$. Define a subparameter space of Θ^0 as follows,

$$\Theta^L(n, K, a, b, \{K_i\}) = \left\{ (\sigma, \{\theta_{i,j}\}) \in \Theta^0(n, K, a, b) : \left| \{k : n_k = \lfloor \frac{n}{K} \rfloor\} \right| = K_1, \right.$$
$$\left. \left| \{k : n_k = \lfloor \frac{n}{K} \rfloor + 1\} \right| = K_2, \left| \{k : n_k = \lfloor \frac{n}{K} \rfloor - 1\} \right| = K_3 \right\}.$$

For the case with $K = 2$, we can define the least favorable case in an analogous way. It has a slight different form depending on whether $n/2$ is an integer or not. If $\frac{n}{2} \neq \lfloor \frac{n}{2} \rfloor$, $\Theta^L(n, 2, a, b) \triangleq \left\{ (\sigma, \{\theta_{i,j}\}) \in \Theta^0(n, 2, a, b) : (n_1, n_2) = (\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil) \right\}$. Otherwise, $\Theta^L(n, 2, a, b) \triangleq \left\{ (\sigma, \{\theta_{i,j}\}) \in \Theta^0(n, 2, a, b) : (n_1, n_2) \in \{(\frac{n}{2}, \frac{n}{2}), (\frac{n}{2} + 1, \frac{n}{2} - 1)\} \right\}$.

Note that Θ^L is homogeneous and closed under permutation. Compared with Θ^0, Θ^L is quite small, enough for us to do some lower bound analysis. On the other hand, it is large enough to match the lower bound in Equation (2.1).

LEMMA 5.1.   *Let τ be the uniform prior over all the elements in Θ^L. For the first node, define the local Bayesian risk to be $B_\tau(\hat{\sigma}(1)) = \frac{1}{|\Theta^L|} \sum_{\sigma \in \Theta^L} \mathbb{E} r(\sigma(1), \hat{\sigma}(1))$. Then there exists a constant $\epsilon > 0$ such that*

$$B_\tau(\hat{\sigma}(1)) \ge \epsilon \mathbb{P}\left(\sum_{u=1}^{\lfloor n/K \rfloor} X_u \ge \sum_{u=1}^{\lfloor n/K \rfloor} Y_u \right),$$

where $X_i \overset{iid}{\sim} Ber(\frac{b}{n})$, $Y_i \overset{iid}{\sim} Ber(\frac{a}{n})$, for $i = 1, 2, \ldots, \lfloor \frac{n}{K} \rfloor$, and $\{X_i\}_{i=1}^{\lfloor \frac{n}{K} \rfloor} \perp \{Y_i\}_{i=1}^{\lfloor \frac{n}{K} \rfloor}$.

Lemma 5.1 shows the lower bound is only involved with $2\lfloor \frac{n}{K} \rfloor$ Bernoulli random variables, whose success probability is either a/n or b/n. Recall that a/n is the smallest within-community probability and b/n is the largest between-community probability. The lower bound here will be determined by testing two probability measures. In Θ^L, the most difficult case is testing two assignment vectors with Hamming distance 1. The difference of their probability measures is exactly the difference between probability measures of X and Y.

LEMMA 5.2.   *Let $n' = \lfloor \frac{n}{K} \rfloor$. Define $Z_i = X_i - Y_i$ with $\{X_i\} \overset{iid}{\sim} Ber(\frac{b}{n})$, $\{Y_i\} \overset{iid}{\sim} Ber(\frac{a}{n})$, and $\{X_i\} \perp \{Y_i\}$, for $i = 1, 2, \ldots, n'$. If $\frac{nI}{K} \to \infty$, we have*

$$\mathbb{P}\left(\frac{1}{n'} \sum_{i=1}^{n'} Z_i > 0 \right) \ge \exp\left(-(1 + o(1))nI/K \right).$$

In addition, if $nI/K = O(1)$, then $\mathbb{P}\left(\frac{1}{n'}\sum_{i=1}^{n'} Z_i > 0\right) \geq c$ for some positive constant $c > 0$.

Lemma 5.2 provides an explicit expression for the lower bound. The proof mainly follows the proof of Cramer-Chernoff Theorem [27]. The general Cramer-Chernoff Theorem gives a lower bound for the tail probability that the sum of random variables deviates from its mean. Usually it is for the case where these random variables are from a distribution independent of the sample size. In our setting we allow a and b to depend on n'.

PROOF OF THEOREM 2.2. Since $\Theta^L \subset \Theta^0$, we have $\inf_{\hat{\sigma}} \sup_{\Theta^0} \mathbb{E}r(\sigma, \hat{\sigma}) \geq \inf_{\hat{\sigma}} \sup_{\sigma \in \Theta^L} \mathbb{E}r(\sigma, \hat{\sigma})$. Due to the fact that Bayes risk always lower bounds the global risk we have $\inf_{\hat{\sigma}} \sup_{\sigma \in \Theta^L} \mathbb{E}r(\sigma, \hat{\sigma}) \geq \inf_{\hat{\sigma}} \sup_{\sigma \in \Theta^L} B_\tau(\hat{\sigma})$. By the fact that Θ^L is a homogeneous parameter space closed under permutation for both $K \geq 3$ and $K = 2$, Lemma 2.1 implies $\inf_{\hat{\sigma}} \sup_{\sigma \in \Theta^L} B_\tau(\hat{\sigma}) = \inf_{\hat{\sigma}} \sup_{\sigma \in \Theta^L} B_\tau(\hat{\sigma}(1))$. Thus

$$\inf_{\hat{\sigma}} \sup_{\Theta^0} \mathbb{E}r(\sigma, \hat{\sigma}) \geq \inf_{\hat{\sigma}} \sup_{\sigma \in \Theta^L} B_\tau(\hat{\sigma}(1)),$$

which, together with Lemma 5.1 and Lemma 5.2, implies Equation (2.2) of Theorem 2.2. $\qquad\square$

5.2. *Proof of Theorem 3.2.* Recall that Δ is the set of all permutations from $[K]$ to $[K]$. For an arbitrary $\sigma \in \Theta^0$, define $\Gamma(\sigma)$ as the equivalent class of σ with $\Gamma(\sigma) = \{\sigma' : \exists \delta \in \Delta, \text{ s.t. } \sigma' = \delta \circ \sigma\}$. We use the notation Γ as a general reference for equivalent class, and $\{\Gamma\}$ as the set consisting of all the possible equivalent classes with respect to Θ^0. For any $\sigma_1, \sigma_2 \in \Theta$, define the distance between σ_1 and σ_2 as

$$d(\sigma_1, \sigma_2) \triangleq \inf_{\sigma_2' \in \Gamma(\sigma_2)} d_H(\sigma_1, \sigma_2') = \inf_{\sigma_1' \in \Gamma(\sigma_1), \sigma_2' \in \Gamma(\sigma_2)} d_H(\sigma_1', \sigma_2').$$

Here we view $d(\cdot, \cdot)$ as a distance between the equivalent class $\Gamma(\sigma_1)$ and $\Gamma(\sigma_2)$. Accordingly the mis-match ratio $r(\sigma, \hat{\sigma})$ is exactly equal to

$$r(\sigma, \hat{\sigma}) = \frac{1}{n}d(\sigma, \hat{\sigma}).$$

In the following sections we denote the true assignment by σ_0. Define

(5.3) $$P_m = \mathbb{P}\big(\exists \sigma \in \Theta^0 : d(\sigma_0, \sigma) = m \text{ and } T(\sigma) \geq T(\sigma_0)\big)$$

for any integer m with $0 < m < n$. The key step is to get a tight bound of the probability $\mathbb{P}\big(T(\sigma) \geq T(\sigma_0)\big)$ for one fixed assignment σ satisfying $d(\sigma, \sigma_0) =$

m. Let $\{n_k\}$ to be the size of communities under the truth σ_0. Without loss of generality, assume $\sigma_0(i) = k$ for any $i \in [\sum_{j \leq k-1} n_j + 1, \sum_{j \leq k} n_j]$. Then the value of $2 \sum_{i<j} A_{i,j}\{\sigma_0(i) = \sigma_0(j)\}$ is just to add up all the entries in the K diagonal blocks of the adjacency matrix A. It is illustrated by color plates in the Figure 1. The gray parts represent the within-community connections, and blank parts represent the between-community connections. It is obvious to see that $2 \sum_{i<j} A_{i,j}\{\sigma_0(i) = \sigma_0(j)\}$ precisely includes all the gray parts, i.e., all the Bernoulli random variables with success probability $\frac{a}{n}$ in the adjacency matrix.
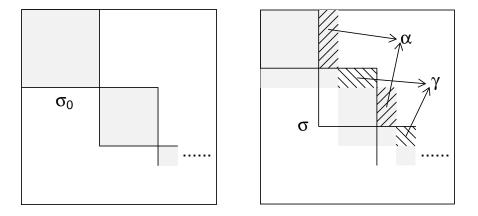

FIG 1. *Each gray block stands for all the within-community connnection in one single community. The areas inside the squares are all the $A_{i,j}$ entries summed up.* **Left:** *For $2 \sum_{i<j} A_{i,j}\{\sigma_0(i) = \sigma_0(j)\}$, the squares exactly overlap with the gray regions.* **Right:** *For $2 \sum_{i<j} A_{i,j}\{\sigma(i) = \sigma(j)\}$, there would be some differences between the squares and gray parts, which are labeled as α or γ according to their relative positions.*

When $d_H(\sigma, \sigma_0) = d(\sigma, \sigma_0) = m$, by comparing the two color plates in Figure 1, we can clearly see where the difference $\sum_{i<j} A_{i,j}\{\sigma_0(i) = \sigma_0(j)\} - \sum_{i<j} A_{i,j}\{\sigma(i) = \sigma(j)\}$ lies in. Note that

$$\sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}} - \sum_{i<j} A_{i,j} 1_{\{\sigma_0(i)=\sigma_0(j)\}} = \sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}} 1_{\{\sigma_0(i)\neq\sigma_0(j)\}}$$
$$- \sum_{i<j} A_{i,j} 1_{\{\sigma(i)\neq\sigma(j)\}} 1_{\{\sigma_0(i)=\sigma_0(j)\}}.$$

Define $\alpha(\sigma; \sigma_0) = |\{(i,j) : i < j, \ \sigma_0(i) = \sigma_0(j) \text{ and } \sigma(i) \neq \sigma(j)\}|$, and $\gamma(\sigma; \sigma_0) = |\{(i,j) : i < j, \ \sigma_0(i) \neq \sigma_0(j) \text{ and } \sigma(i) = \sigma(j)\}|$. We use the

notations α and γ for short when there is no ambiguity, then

$$\sum_{i<j} 1_{\{\sigma(i)=\sigma(j)\}} - \sum_{i<j} 1_{\{\sigma_0(i)=\sigma_0(j)\}} = \alpha - \gamma.$$

The following proposition is helpful to study P_m defined in Equation (5.3).

PROPOSITION 5.1. *Let $\sigma \in \Theta^0$ be an arbitrary assignment satisfying $d(\sigma, \sigma_0) = m$, where $0 < m < n$ is a positive integer. Then*

$$\mathbb{P}\big(T(\sigma) \geq T(\sigma_0)\big) \leq \mathbb{P}\Big(\sum_{i=1}^{\gamma} X_i - \sum_{i=1}^{\alpha} Y_i \geq \lambda(\gamma - \alpha)\Big| X_i \overset{iid}{\sim} Ber(\frac{b}{n}), Y_i \overset{iid}{\sim} Ber(\frac{a}{n})\Big)$$

$$\leq \exp(-\gamma I),$$

for λ defined in Equation (3.3).

Note that the value of γ depends on σ and σ_0. Lemma 5.3 provides a lower bound on γ for each m.

LEMMA 5.3. *Let $\sigma \in \Theta^0$ be an arbitrary assignment satisfying $d(\sigma, \sigma_0) = m$, where $0 < m < n$ is a positive integer. Then*

$$\alpha(\sigma; \sigma_0) \wedge \gamma(\sigma; \sigma_0) \geq \begin{cases} \frac{(1-\eta)nm}{K} - m^2, & \text{if } m \leq \frac{n}{2K}, \\ \frac{2(1-\eta)nm}{9K}, & \text{if } m > \frac{n}{2K}. \end{cases}$$

Lemma 5.3, together with Proposition 5.1, immediately implies an upper bound on $\mathbb{P}\big(T(\sigma) \geq T(\sigma_0)\big)$ for each given σ.

LEMMA 5.4. *Let $\sigma \in \Theta^0$ be an arbitrary assignment satisfying $d(\sigma, \sigma_0) = m$, where $0 < m < n$ is a positive integer. There exists a positive sequence $\eta \to 0$, independent of the choice of σ, such that*

$$\mathbb{P}\big(T(\sigma) \geq T(\sigma_0)\big) \leq \begin{cases} \exp\big(-\frac{(1-\eta)nmI}{K} + m^2 I\big), & \text{if } m \leq \frac{n}{2K}, \\ \exp\big(-\frac{2(1-\eta)nmI}{9K}\big), & \text{if } m \geq \frac{n}{2K}, \end{cases}$$

for λ defined in Equation (3.3).

We will apply a union bound to get an upper bound for P_m. It is worthwhile to point out that, in the union bound we should not use the cardinality of $\{\sigma \in \Theta^0 : d(\sigma, \sigma_0) = m\}$, which is too large due to counting the assignments from the same equivalent class repetitively. Proposition 5.2 gives an upper bound for cardinality of the equivalent class $\{\Gamma\}$.

PROPOSITION 5.2. *The cardinality of equivalent class that has distance m from σ_0 is upper bounded as follows,*

$$\left| \left\{ \Gamma : \exists \sigma \in \Gamma \ s.t. \ d(\sigma, \sigma_0) = m \right\} \right| \leq \min \left\{ \left(\frac{enK}{m} \right)^m, K^n \right\},$$

where $0 < m < n$ is a positive integer.

With Proposition 5.2 and the union bound we are able to get a satisfactory bound by

$$P_m \leq \left| \left\{ \Gamma : \exists \sigma \in \Gamma \ \text{s.t.} \ d(\sigma, \sigma_0) = m \right\} \right| \max_{\{\sigma: \ d(\sigma,\sigma_0)=m\}} \mathbb{P}\big(T(\sigma) \geq T(\sigma_0) \big).$$

PROOF OF THEOREM 3.2. We only prove the case with $K \to \infty$ and $\frac{nI}{K \log K} \to \infty$. Let $\eta \to 0$ be a universal positive sequence given in Lemma 5.4. We consider three scenarios as follows.

(1) If $\liminf_{n \to \infty} \frac{nI}{K \log n} > 1$, there exists a small constant $\epsilon > 0$ such that $\frac{(1-\eta)nI}{K \log n} > 1 + \epsilon$. Let η decay slowly such that both $\frac{\eta nI}{K \log K}$ and $\frac{\eta n}{K}$ go to infinity. We have $P_1 \leq nK \exp\big(- \big(\frac{(1-\eta)n}{K} - 1 \big)I \big) \leq R$, where $R \triangleq n \exp\big(- (1 - 2\eta)nI/K \big)$. Since

$$n\mathbb{E}r(\sigma, \hat{\sigma}) \leq P_1 + \sum_{m=2}^{n} mP_m,$$

it is sufficient to show $\sum_{i=2}^{n} mP_m$ is negligible compared with R. For $m \in [2, m']$, where $m' = \frac{\epsilon n}{3K}$, we have

$$
\begin{aligned}
P_m &\leq \big(\frac{enK}{2} \exp\big(- \frac{(1-\eta)nI}{K} + mI) \big)^m \\
&\leq \big(\frac{enK}{2} \exp\big(- \frac{(1-\eta)nI}{K} + mI) \big) \big(\frac{enK}{2} \exp\big(- \frac{(1-\eta)nI}{K} + m'I) \big)^{m-1} \\
&\leq n \exp\big(- \frac{(1-2\eta)nI}{K} \big) \exp(mI) n^{-\epsilon(m-1)/3} \\
&\leq R n^{-\epsilon(m-1)/6},
\end{aligned}
$$

where we use the fact that $I \lesssim 1$ in the fourth inequality to show $e^I n^{-\epsilon/6} < 1$ when n is large enough. As a consequence, $\sum_{i=2}^{m'} mP_m = o(R)$, as $\{mP_m\}_{i=2}^{m'}$ is dominated by a fast-decay geometric series.

For $m \in [m', n]$, we have

$$
P_m \leq \left(\frac{enK}{m'} \exp\left(-\frac{2(1-\eta)nI}{9K} \right) \right)^m
$$

$$
\leq n \exp\left(-\frac{(1-2\eta)nI}{K} \right) \left(\frac{enK}{m'} \exp\left(-\frac{2(1-\eta)nI}{9K} \right) \right)^{m-9}
$$

$$
\leq Rn^{-2(m-9)/9}.
$$

Since $m' \to \infty$, $\{mP_m\}_{m \geq m'}$ is dominated by a fast-decay geometric series, which leads to $\sum_{i>m'}^{n} mP_m = o(R)$.

(2) If $\limsup_{n \to \infty} \frac{nI}{K \log n} < 1$, there exists a small constant $\epsilon > 0$ such that $\frac{(1-\eta)nI}{K \log n} < 1 - \epsilon$. Let $m_0 = n \exp\left(-(1 - K^{-\epsilon/2})\frac{(1-\eta)nI}{K} \right)$, which satisfies both $m_0 \geq (nK)^{\epsilon/2}$ and $m_0 = o(\frac{n}{K^2})$. We are going to show that $\{P_m\}_{m \geq m_0}$ is upper bounded by a fast decaying series $\{Q_m\}_{m \geq m_0}$.

For any $m \in [m_0, m']$, where $m' = \frac{n}{K^{1+\epsilon}}$, we have

$$
P_m \leq \left(\left(\frac{enK}{m_0} \right) \exp\left(-\frac{(1-\eta)nI}{K} + m'I \right) \right)^m
$$

$$
\leq \left(\exp\left(\log(nK) + \left((1 - K^{-\epsilon/2}) - (1 - 2K^{-\epsilon}) \right) \frac{(1-\eta)nI}{K} \right) \right)^m
$$

$$
\leq \exp\left(-\frac{m}{2K^{\epsilon/2}} \frac{(1-\eta)nI}{K} \right),
$$

which is denoted as Q_m. Since $\frac{m_0}{K^{\epsilon/2}} \gg \log n$, we have $\sum_{m=m_0}^{m'} P_m \leq \sum_{m=m_0}^{m'} Q_m \leq m'Q_{m_0} \leq \exp\left(\log n - \frac{m_0}{2K^{\epsilon/2}} \frac{(1-\eta)nI}{K} \right) = o(\frac{m_0}{n})$. For $m' \leq m$, we have

$$
P_m \leq \left(\frac{enK}{m'} \exp\left(-\frac{2(1-\eta)nI}{9K} \right) \right)^m \leq \exp\left(-\frac{nmI}{9K} \right).
$$

Denote $Q_m = \exp\left(-\frac{nmI}{9K} \right)$, which decays geometrically fast, as $\frac{nI}{K} \to \infty$. Thus $\sum_{m=m'}^{n} P_m \leq \sum_{m=m'}^{n} Q_m \leq 2Q_{m'} = o(\frac{m_0}{n})$. Consequently,

$$
\mathbb{E}r(\sigma, \hat{\sigma}_0) \leq \frac{m_0}{n} + \mathbb{P}\left(\exists \sigma \in \Theta^0 : d(\sigma_0, \sigma) \geq m_0 \ \& \ l(\sigma) \geq l(\sigma_0) \right)
$$

$$
\leq \frac{m_0}{n} + \sum_{m>m_0}^{m'} P_{m'} + \sum_{m>m'}^{n} P_m
$$

$$
\leq \frac{m_0}{n} + m'Q_{m_0} + 2Q_{m'}
$$

$$
= \exp\left(-\frac{(1-o(1))nI}{K} \right).
$$

(3) If $\frac{nI}{K \log n} = 1 + o(1)$, there exists a positive sequence $w \to 0$ such that $|\frac{(1-\eta)nI}{K \log n} - 1| \ll w$, $\frac{1}{\sqrt{\log n}} \le w$ and $\frac{wnmI}{K \log K} \to \infty$. Define $m_0 = n \exp\left(-(1-w)\frac{(1-\eta)nI}{K}\right)$. Thus $m_0 \ge n^{w/2} \to \infty$, and $m_0 = o(m')$ for $m' = w^2 n/K$. We are going to find a fast decay series $\{Q_m\}$ to upper bound $\{P_m\}$. For $m \in [m_0, m']$,

$$P_m \le \left(\left(\frac{enK}{m_0}\right) \exp\left(-\frac{(1-\eta)nI}{K} + m'I\right)\right)^m$$

$$\le \left(\log(eK) + \frac{(1-w)(1-\eta)nI}{K} - \frac{(1-\eta)nI}{K} + \frac{w^2 nI}{K}\right)^m$$

$$\le \exp\left(-\frac{\omega(1-\eta)nmI}{4K}\right),$$

which is denoted as Q_m. Note that $\omega m_0 \ge w n^{w/2} \to \infty$. We have $Q_{m_0} < 1$, and furthermore

$$\sum_{m=m_0}^{m'} P_m \le \sum_{m=m_0}^{m'} Q_m \le m' Q_{m_0} \le \exp\left(\log n - \frac{\omega m_0 (1-\eta)nI}{4K}\right) = o\left(\frac{m_0}{n}\right).$$

For $m \in [m', n]$, we have

$$P_m \le \left(\frac{enK}{m'} \exp\left(-\frac{2(1-\eta)nI}{9K}\right)\right)^m \le \exp\left(-\frac{nmI}{9K}\right),$$

Let $Q_m = \exp\left(-\frac{nmI}{9K}\right)$, which decays geometrically fast. Then $\sum_{m=m'}^{n} P_m \le \sum_{m=m'}^{n} Q_m \le 2Q_{m'} = o\left(\frac{m_0}{n}\right)$. Hence

$$\mathbb{E}r(\sigma, \hat{\sigma}_0) \le \frac{m_0}{n} + \sum_{m>m_0}^{m'} P_{m'} + \sum_{m>m'}^{n} P_m \le \exp\left(-\frac{(1-o(1))nI}{K}\right).$$

When K is a fixed constant, the proof is nearly identical but with different m' under each scenario. The proof is thus omitted. $\square$

6. Proofs of Auxiliary Lemmas.   We prove Lemma 2.1, Lemma 5.1, Lemma 5.2, Lemma 5.3, Proposition 5.1 and Propostion 5.2 respectively in this section.

6.1. *Proof of Lemma 2.1.*   Before going directly into the proof we define another network operator: (element-wise) permutation. Let $\pi : [1, 2, \ldots, n] \to [1, 2, \ldots, n]$ be a permutation. Denote Π to be the set consisting of all such

permutations, whose cardinality is $n!$. Define σ_π to be a new assignment with

$$\sigma_\pi(i) \triangleq \sigma(\pi^{-1}(i)), \forall 1 \leq i \leq n.$$

It is obvious that for an arbitrary assignment $\sigma \in \Lambda$, each of its permutation σ_π is also in the parameter space Λ.

On the other hand, a permutation on the nodes leads to the change of the network. For a network G with an adjacency matrix A, define G_π as the network after permutation with a new adjacency matrix A_π, where

$$(A_\pi)_{i,j} = A_{\pi^{-1}(i),\pi^{-1}(j)}.$$

Note that G_π can be seen as a network sampled from the assignment σ_π, since $(A_\pi)_{i,j} \sim \text{Ber}(\theta_{\pi^{-1}(i),\pi^{-1}(j)})$.

The proof of Lemma 2.1 is mainly by exploring the exchangeability of the network. Any estimator $\hat{\sigma}$ is a mapping from a network to a length n vector. We use the square brackets $\hat{\sigma}[G]$ to indicate that the outcome of $\hat{\sigma}$ is implemented on the network G. And $\hat{\sigma}[G](i)$ is the value of the i-th component of $\hat{\sigma}[G]$, and when the meaning is clear, we write $\hat{\sigma}(i)$ for simplicity.

Based on $\hat{\sigma}$, we can always design a new (unless they are the same) procedure by permutation. Given a network G, we can either directly apply $\hat{\sigma}$ (to be more precise, it is $\hat{\sigma}[G]$), or first permute the network into G_π, then implement $\hat{\sigma}$ on it to get $\hat{\sigma}[G_\pi]$, and then finally "permute back" to get the estimation in the original order. To be more precise, define procedure $\hat{\sigma}^\pi$ as

$$\hat{\sigma}^\pi[G](i) = \hat{\sigma}[G_\pi](\pi(i)).$$

We use the notation $\hat{\sigma}^\pi(i)$ short for $\hat{\sigma}^\pi[G](i)$. See Figure 2 for the illustration on getting $\hat{\sigma}^\pi$.

Intuitively, due to the exchangeability of G, if $\hat{\sigma}$ is optimal, it should have the same risk as $\hat{\sigma}^\pi$ for any possible π. With this trick we are able to show the existence of a universal procedure $\bar{\sigma}$ which has the equal global risk for all $\sigma \in \Lambda$ and the equal local risk for all $i \in [n]$. Then the proof is completed by the fact that the minimax risk is lower bounded by the Bayes risk.

PROOF OF LEMMA 2.1. Denote the network to be G. Assume $\tilde{\sigma}$ be one of the estimators that achieve the global Bayes risk, i.e., $B_\tau(\tilde{\sigma}) = \inf_{\hat{\sigma}} B_\tau(\hat{\sigma})$. Based on $\tilde{\sigma}$, we can define a randomized procedure $\bar{\sigma}$ as $\mathbb{P}(\bar{\sigma} = \tilde{\sigma}^\pi) = 1/|\Pi|$, for each $\pi \in \Pi$. We will show $\bar{\sigma}$ is also a global Bayes estimator in terms of
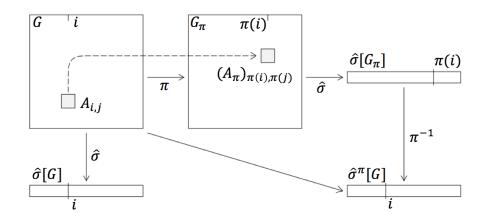
FIG 2. *Illustration on getting $\hat{\sigma}^\pi$ based on the original network G. All of $\hat{\sigma}[G]$, $\hat{\sigma}[G_\pi]$ and $\hat{\sigma}^\pi[G]$ are demonstrated as n-by-1 vectors. It shows $A_{i,j}$ becomes $(A_\pi)_{\pi_i,\pi_j}$ after the permutation π of the network. For any specific node i in G, its location is changed into $\pi(i)$ in G_π. The procedure $\hat{\sigma}[G_\pi]$ estimates the assignemnt of the permuted nodes $\{\pi(i)\}$, while $\hat{\sigma}^\pi[G]$ estimates the assignment of the original nodes.*

τ. For an arbitrary $\sigma \in \Lambda$, we have

$$\mathbb{E}r(\sigma, \bar{\sigma}) = \frac{1}{n!}\sum_{\pi \in \Pi}\mathbb{E}r(\sigma, \tilde{\sigma}^\pi).$$

Recall that $\mathbb{E}r(\sigma, \tilde{\sigma}^\pi) = \mathbb{E}\inf_{\sigma' \in \Gamma(\tilde{\sigma}^\pi)} d_H(\sigma, \sigma')$. There exists a one-by-one relation between $\Gamma(\tilde{\sigma}^\pi)$ and $\Gamma(\tilde{\sigma}[G_\pi])$, in a the sense that, for any σ' from the former set, there is σ'' in the latter set such that $\sigma''(i) = \sigma'(\pi^{-1}(i)), \forall i \in [n]$, and the reverse also holds. We have the following equation (we add subscript σ to explicitly indicate that the expectation is taken with respect to the assignment σ),

$$\mathbb{E}r(\sigma, \tilde{\sigma}^\pi) = \frac{1}{n}\mathbb{E}_\sigma \inf_{\sigma' \in \Gamma(\tilde{\sigma}^\pi)} \sum_{i=1}^n 1\{\sigma(i) \neq \sigma'(i)\}$$

$$= \frac{1}{n}\mathbb{E}_\sigma \inf_{\sigma'' \in \Gamma(\tilde{\sigma}[G_\pi])} \sum_{i=1}^n 1\{\sigma_\pi(\pi(i)) \neq \sigma''(\pi(i))\}$$

$$= \frac{1}{n}\mathbb{E}_\sigma \inf_{\sigma'' \in \Gamma(\tilde{\sigma}[G_\pi])} \sum_{i=1}^n 1\{\sigma_\pi(i) \neq \sigma''(i)\}.$$

The expectation can be further expanded into

$$\mathbb{E} r(\sigma, \tilde{\sigma}^{\pi}) = \frac{1}{n} \sum_{G \in \mathbb{G}} \left(\inf_{\sigma'' \in \Gamma(\tilde{\sigma}[G_{\pi}])} \sum_{i=1}^{n} 1\{\sigma_{\pi}(i) \neq \sigma''(i)\} \right) \mathbb{P}_{\sigma}(G),$$

where $\mathbb{G}$ contains all the possible realizations of the graph. Here the subscript of $\mathbb{P}_{\sigma}(G)$ emphasizes that the probability measure is associated with the assignment σ. Note that $\mathbb{P}_{\sigma}(G) = \mathbb{P}_{\sigma_{\pi}}(G_{\pi})$ for any G and that the set $\{G_{\pi} : G \in \mathbb{G}\}$ is exactly equal to $\mathbb{G}$, we have

$$\mathbb{E} r(\sigma, \tilde{\sigma}^{\pi}) = \frac{1}{n} \sum_{G \in \mathbb{G}} \left(\inf_{\sigma'' \in \Gamma(\tilde{\sigma}[G_{\pi}])} \sum_{i=1}^{n} 1\{\sigma_{\pi}(i) \neq \sigma''(i)\} \right) \mathbb{P}_{\sigma_{\pi}}(G_{\pi})$$

$$= \frac{1}{n} \sum_{G_{\pi} \in \mathbb{G}} \left(\inf_{\sigma'' \in \Gamma(\tilde{\sigma}[G_{\pi}])} \sum_{i=1}^{n} 1\{\sigma_{\pi}(i) \neq \sigma''(i)\} \right) \mathbb{P}_{\sigma_{\pi}}(G_{\pi})$$

$$= \frac{1}{n} \sum_{G \in \mathbb{G}} \left(\inf_{\sigma'' \in \Gamma(\tilde{\sigma}[G])} \sum_{i=1}^{n} 1\{\sigma_{\pi}(i) \neq \sigma''(i)\} \right) \mathbb{P}_{\sigma_{\pi}}(G),$$

which yields

$$\mathbb{E} r(\sigma, \tilde{\sigma}^{\pi}) = \frac{1}{n} \mathbb{E}_{\sigma_{\pi}} \inf_{\sigma'' \in \Gamma(\tilde{\sigma}[G])} \sum_{i=1}^{n} 1\{\sigma_{\pi}(i) \neq \sigma''(i)\}$$

$$= \mathbb{E} r(\sigma_{\pi}, \tilde{\sigma}).$$

Thus

$$B_{\tau}(\bar{\sigma}) = \frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \left(\frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E} r(\sigma_{\pi}, \tilde{\sigma}) \right) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \left(\frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \mathbb{E} r(\sigma_{\pi}, \tilde{\sigma}) \right).$$

Since $\{\sigma_{\pi} : \sigma \in \Lambda\}$ is exactly equal to Λ for any π, we have

$$B_{\tau}(\bar{\sigma}) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \left(\frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \mathbb{E} r(\sigma, \tilde{\sigma}) \right) = \frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \left(\frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E} r(\sigma, \tilde{\sigma}) \right) = B_{\tau}(\tilde{\sigma}).$$

Thus $\bar{\sigma}$ also achieves the minimum Bayes risk. We will show $B_{\tau}(\bar{\sigma}(i)) = B_{\tau}(\bar{\sigma}(j))$ for any $i, j \in [n]$. It is equivalent to define $\bar{\sigma}$ as

$$\mathbb{P}\left(\bar{\sigma}(i) = \tilde{\sigma}^{\pi}(i) \right) = \frac{1}{|\Pi|}, \forall i \in [n],$$

which implies

$$\mathbb{E} r(\sigma(i), \bar{\sigma}(i)) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E} r(\sigma(i), \tilde{\sigma}^{\pi}(i)).$$

Note that $\tilde{\sigma}^\pi(i) = \tilde{\sigma}[G_\pi](\pi(i))$, and $\sigma(i) = \sigma_\pi(\pi(i))$. Recall that the definition of the local risk is

$$\mathbb{E}r(\sigma(i), \tilde{\sigma}^\pi(i)) = \mathbb{E}_\sigma \sum_{\sigma' \in S_\sigma(\tilde{\sigma}^\pi)} \frac{1\{\sigma(i) \neq \sigma'(i)\}}{|S_\sigma(\tilde{\sigma}^\pi)|}.$$

Here recall $S_\sigma(\hat{\sigma}) \triangleq \{\sigma' \in \Gamma(\hat{\sigma}) : d_H(\sigma, \sigma') = d(\sigma, \hat{\sigma})\}$ for any estimator $\hat{\sigma}$. It is obvious that there exists a one-by-one relation between $S_\sigma(\tilde{\sigma}^\pi)$ and $S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])$. For any $\sigma' \in S_\sigma(\tilde{\sigma}^\pi)$, there is a unique corresponding $\sigma'' \in S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])$ defined as $\sigma''(i) = \sigma'(\pi^{-1}(i)), \forall i \in [n]$, and the reverse also holds. Thus the event $\{\sigma(i) \neq \sigma'(i)\}$ is equivalent to $\{\sigma_\pi(\pi(i)) \neq \sigma''(\pi(i))\}$, and $|S_\sigma(\tilde{\sigma}^\pi)| = |S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])|$. We have

$$\mathbb{E}r(\sigma(i), \tilde{\sigma}^\pi(i)) = \mathbb{E}_\sigma \sum_{\sigma'' \in S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])} \frac{1\{\sigma_\pi(\pi(i)) \neq \sigma''(\pi(i))\}}{|S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])|}.$$

By the same argument as the previous one, together with the fact that $\mathbb{P}_\sigma(G) = \mathbb{P}_{\sigma_\pi}(G_\pi)$, we expand the expectation and then have

$$\mathbb{E}r(\sigma(i), \tilde{\sigma}^\pi(i)) = \sum_{G \in \mathbb{G}} \Big(\sum_{\sigma'' \in S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])} \frac{1\{\sigma_\pi(\pi(i)) \neq \sigma''(\pi(i))\}}{|S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])|} \Big) \mathbb{P}_\sigma(G)$$

$$= \sum_{G \in \mathbb{G}} \Big(\sum_{\sigma'' \in S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])} \frac{1\{\sigma_\pi(\pi(i)) \neq \sigma''(\pi(i))\}}{|S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])|} \Big) \mathbb{P}_{\sigma_\pi}(G_\pi)$$

$$= \sum_{G \in \mathbb{G}} \Big(\sum_{\sigma'' \in S_{\sigma_\pi}(\tilde{\sigma}[G])} \frac{1\{\sigma_\pi(\pi(i)) \neq \sigma''(\pi(i))\}}{|S_{\sigma_\pi}(\tilde{\sigma}[G])|} \Big) \mathbb{P}_{\sigma_\pi}(G).$$

Thus

$$\mathbb{E}r(\sigma(i), \tilde{\sigma}^\pi(i)) = \mathbb{E}_{\sigma_\pi} \sum_{\sigma'' \in S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])} \frac{1\{\sigma_\pi(\pi(i)) \neq \sigma''(\pi(i))\}}{|S_{\sigma_\pi}(\tilde{\sigma}[G])|}$$

$$= \mathbb{E}r(\sigma_\pi(\pi(i)), \tilde{\sigma}(\pi(i))).$$

This gives

$$\mathbb{E}r(\sigma(i), \bar{\sigma}(i)) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E}r\big(\sigma_\pi(\pi(i)), \tilde{\sigma}(\pi(i))\big), \forall i \in [n].$$

Then for the local risk we have

$$
\begin{aligned}
B_\tau(\bar{\sigma}(i)) &= \frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \left(\frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E} r\big(\sigma_\pi(\pi(i)), \tilde{\sigma}(\pi(i))\big) \right) \\
&= \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \left(\frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \mathbb{E} r\big(\sigma_\pi(\pi(i)), \tilde{\sigma}(\pi(i))\big) \right) \\
&= \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \left(\frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \mathbb{E} r\big(\sigma(\pi(i)), \tilde{\sigma}(\pi(i))\big) \right) \\
&= \frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \left(\frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E} r\big(\sigma(\pi(i)), \tilde{\sigma}(\pi(i))\big) \right) \\
&= \frac{1}{|\Lambda|} \sum_{\sigma \in \Lambda} \left(\frac{1}{n} \sum_{l=1}^{n} \mathbb{E} r\big(\sigma(l), \tilde{\sigma}(l)\big) \right),
\end{aligned}
$$

where in the third equation we again use the fact that $\{\sigma_\pi : \sigma \in \Lambda\}$ is exactly equal to Λ for any π. So we conclude $B_\tau(\bar{\sigma}(i)) = B_\tau(\bar{\sigma}(j))$ for any $i, j \in [n]$. Due to the equality

$$
\begin{aligned}
\mathbb{E} r(\sigma, \hat{\sigma}) &= \mathbb{E} \inf_\delta \sum_{i=1}^{n} \frac{1\{(\delta \circ \hat{\sigma})(i) \neq \sigma(i)\}}{n} \\
&= \mathbb{E} \frac{1}{|S_\sigma(\hat{\sigma})|} \sum_{\sigma' \in S_\sigma(\hat{\sigma})} \sum_{i=1}^{n} \frac{1\{i : \sigma'(i) \neq \sigma(i)\}}{n} \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \sum_{\sigma' \in S_\sigma(\hat{\sigma})} \frac{1\{i : \sigma'(i) \neq \sigma(i)\}}{|S_\sigma(\hat{\sigma})|} \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} r(\sigma(i), \hat{\sigma}(i)),
\end{aligned}
$$

we have $B_\tau(\bar{\sigma}) = \sum_{i=1}^{n} B_\tau(\bar{\sigma}(i))/n$, which leads to $\inf_{\hat{\sigma}} B_\tau(\hat{\sigma}) = B_\tau(\bar{\sigma}) = B_\tau(\bar{\sigma}(1)) \geq \inf_{\hat{\sigma}} B_\tau(\hat{\sigma}(1))$. We omit the proof of the other direction of the equality stated in the lemma, which uses a nearly identical argument. The proof is complete. $\qquad\square$

6.2. *Proof of Lemma 5.1.* First consider the case with $K \geq 3$. Define $\Theta_1^L = \{(\sigma, \{\theta_{i,j}\}) \in \Theta^L : n_{\sigma(1)} = \lfloor \frac{n}{K} \rfloor + 1\}$. So for each $\sigma \in \Theta_1^L$, the community containing the first node always has size $\lfloor \frac{n}{K} \rfloor + 1$. We will show the ratio of the cardinality of Θ_1^L against that of Θ^L is a constant. Denote $x_1 = \lfloor n/K \rfloor K_1$ and $x_2 = (\lfloor n/K \rfloor + 1)K_2$, then

$$
|\Theta^L| = C \binom{n}{x_2} \binom{n - x_2}{x_1} \text{ and } |\Theta_1^L| = C \binom{n - 1}{x_2 - 1} \binom{n - x_2}{x_1},
$$

where C is the number of combinations to select x_1 balls into K_1 bins with size $\lfloor \frac{n}{K} \rfloor$, x_2 balls into K_2 bins with size $\lfloor \frac{n}{K} \rfloor + 1$, and another $n - x_1 - x_2$ balls into K_3 bins with size $\lfloor \frac{n}{K} \rfloor - 1$. Thus

$$\frac{|\Theta_1^L|}{|\Theta^L|} = \frac{\binom{n-1}{x_2-1}}{\binom{n}{x_2}} = \frac{x_2}{n} \geq \epsilon.$$

It is equivalent to the probability that the first node is assigned to the K_2 bins with size $\lfloor \frac{n}{K} \rfloor + 1$. Then

$$B_\tau(\hat{\sigma}(1)) \geq \frac{1}{|\Theta^L|} \sum_{\sigma \in \Theta_1^L} \mathbb{E}r(\sigma(1), \hat{\sigma}(1)) \geq \frac{\epsilon}{|\Theta_1^L|} \sum_{\sigma \in \Theta_1^L} \mathbb{E}r(\sigma(1), \hat{\sigma}(1)).$$

For each $\sigma_0 \in \Theta_1^L$, let $\kappa_0(\sigma_0) = \sigma_0(1)$ be the index of community that the first node belongs to. And let $\kappa(\sigma_0)$ be the indices of communities whose sizes are $\lfloor \frac{n}{K} \rfloor$, i.e.,

$$\kappa(\sigma_0) = \{k \in [K] : n_k = \lfloor \frac{n}{K} \rfloor\}.$$

Note that $\kappa_0(\sigma_0) \notin \kappa(\sigma_0)$. If we replace $\sigma_0(1)$ by any $k \in \kappa(\sigma_0)$ while keep the value of the rest of nodes, then we get a new assignment also contained in Θ_1^L and has distance 1 from σ_0. In particular, we use the following procedure to get a new assignment $\sigma[\sigma_0]$ based on σ_0:

$$\sigma[\sigma_0](1) = \begin{cases} \min\{k \in \kappa(\sigma_0) : k > \kappa_0(\sigma_0)\} \text{ if } \max \kappa(\sigma_0) > \kappa_0(\sigma_0); \\ \min \kappa(\sigma_0) \text{ if } \max \kappa(\sigma_0) < \kappa_0(\sigma_0), \end{cases}$$

and $\sigma[\sigma_0](i) = \sigma_0(i)$ for all $i \geq 2$. It is clear that $\sigma[\sigma_0] \in \Theta_1^L$ and $d_H(\sigma_0, \sigma[\sigma_0]) = 1$. It is also guaranteed that for any $\sigma_0, \sigma_1 \in \Theta_1^L$ and $\sigma_0 \neq \sigma_1$, the new assignments are also different $\sigma[\sigma_0] \neq \sigma[\sigma_1]$. This leads to that Θ^L is equal to the set $\{\sigma[\sigma_0] : \sigma_0 \in \Theta^L\}$, and hence

$$B_\tau(\hat{\sigma}(1)) \geq \frac{\epsilon}{2|\Theta_1^L|} \sum_{\sigma_0 \in \Theta_1^L} 2\mathbb{E}r(\sigma_0(1), \hat{\sigma}(1))$$

$$\geq \frac{\epsilon}{|\Theta_1^L|} \sum_{\sigma_0 \in \Theta_1^L} \frac{1}{2}\big(\mathbb{E}r(\sigma_0(1), \hat{\sigma}(1)) + \mathbb{E}r(\sigma[\sigma_0](1), \hat{\sigma}(1))\big).$$

We are going to derive the Bayes risk $\inf_{\hat{\sigma}} \frac{1}{2}\big(\mathbb{E}r(\sigma_0(1), \hat{\sigma}(1)) + \mathbb{E}r(\sigma[\sigma_0](1), \hat{\sigma}(1))\big)$ for a given $\sigma_0 \in \Theta^L$. Let $\tilde{\sigma}$ be any estimator achieving the infimum. Since $d_H(\sigma_0, \sigma[\sigma_0]) = 1$, we have $r(\sigma_0(1), \tilde{\sigma}(1)) = d_H(\sigma_0(1), \tilde{\sigma}(1))$ and a similar

equation holds for $\sigma[\sigma_0]$. The estimator $\tilde{\sigma}(1)$ can be interpreted as the Bayes estimator with respect to the zero-one loss. Then $\tilde{\sigma}(1)$ must be the mode of the posterior distribution. Let J_0 to be the set $\{u \in [n]\backslash\{1\} : \sigma_0(u) = \sigma_0(1)\}$, and $J_1 = \{u \in [n] : \sigma_0(u) = \sigma[\sigma_0](1)\}$. For a given adjacency matrix A, the conditional distributions are

$$\mathbb{P}(A|\sigma_0) = \prod_{u \in J_0} (\frac{a}{n})^{A_{1,u}}(1 - \frac{a}{n})^{1-A_{1,u}} \prod_{u \in J_1} (\frac{b}{n})^{A_{1,u}}(1 - \frac{b}{n})^{1-A_{1,u}} f(A^C),$$

and

$$\mathbb{P}(A|\sigma[\sigma_0]) = \prod_{u \in J_1} (\frac{a}{n})^{A_{1,u}}(1 - \frac{a}{n})^{1-A_{1,u}} \prod_{u \in J_0} (\frac{b}{n})^{A_{1,u}}(1 - \frac{b}{n})^{1-A_{1,u}} f(A^C).$$

Here A^C consists all the rest of the entries: $A^C = \{(u, v) : v > u \geq 2,$ or $u = 1$ and $v \notin J_0 \cup J_1\}$. It is obvious that $f(A^C)$ is invariant to the choice of σ_0 or $\sigma[\sigma_0]$. Thus

$$\tilde{\sigma}(1) = \begin{cases} \sigma_0(1), & \text{if } \sum_{u \in J_0} A_{1,u} \geq \sum_{u \in J_1} A_{1,u}, \\ \sigma[\sigma_0](1), & \text{if } \sum_{u \in J_0} A_{1,u} < \sum_{u \in J_1} A_{1,u}. \end{cases}$$

Thus $\mathbb{E}r(\sigma_0(1), \hat{\sigma}(1)) = \mathbb{P}_{\sigma_0}\left(\sum_{u \in J_0} A_{1,u} < \sum_{u \in J_1} A_{1,u}\right) \geq \mathbb{P}\left(\sum_{u=1}^{\lfloor n/K \rfloor} X_u \geq \sum_{u=1}^{\lfloor n/K \rfloor} Y_u\right)$, and $\mathbb{E}r(\sigma[\sigma_0](1), \hat{\sigma}(1))) = \mathbb{P}_{\sigma[\sigma_0]}\left(\sum_{u \in J_0} A_{1,u} \geq \sum_{u \in J_1} A_{1,u}\right) \geq \mathbb{P}\left(\sum_{u=1}^{\lfloor n/K \rfloor} X_u \geq \sum_{u=1}^{\lfloor n/K \rfloor} Y_u\right)$. Consequently,

$$\frac{1}{2}\left(\mathbb{E}r(\sigma_0(1), \hat{\sigma}(1)) + \mathbb{E}r(\sigma[\sigma_0](1), \hat{\sigma}(1)))\right) \geq \mathbb{P}\left(\sum_{u=1}^{\lfloor n/K \rfloor} X_u \geq \sum_{u=1}^{\lfloor n/K \rfloor} Y_u\right).$$

The above inequality holds for each $\sigma_0 \in \Theta^L$. Hence

$$\inf_{\hat{\sigma}} B_\tau(\hat{\sigma}(1)) \geq \frac{\epsilon}{|\Theta_1^L|} \sum_{\sigma_0 \in \Theta_1^L} \inf_{\hat{\sigma}} \frac{1}{2}\left(\mathbb{E}r(\sigma_0(1), \hat{\sigma}(1)) + \mathbb{E}r(\sigma[\sigma_0](1), \hat{\sigma}(1)))\right)$$

$$\geq \frac{\epsilon}{|\Theta_1^L|} \sum_{\sigma_0 \in \Theta_1^L} \frac{1}{2}\left(\mathbb{E}r(\sigma_0(1), \tilde{\sigma}(1)) + \mathbb{E}r(\sigma[\sigma_0](1), \tilde{\sigma}(1)))\right)$$

$$\geq \epsilon \mathbb{P}\left(\sum_{u=1}^{\lfloor n/K \rfloor} X_u \geq \sum_{u=1}^{\lfloor n/K \rfloor} Y_u\right).$$

For the case $K = 2$, we re-define Θ_1^L and show that its cardinality is same with that of Θ^L up to a constant factor. *(1)* If $\frac{n}{2} \neq \lfloor \frac{n}{2} \rfloor$, then define

$\Theta_1^L = \{(\sigma, \{\Theta_{i,j}^L\}) \in \Theta^L : n_{\sigma(1)} = \lceil \frac{n}{2} \rceil\}$. Then $|\Theta_1^L|/|\Theta^L| = 1/2$. (2) If $\frac{n}{2} = \lfloor \frac{n}{2} \rfloor$, then define $\Theta_1^L = \{(\sigma, \{\Theta_{i,j}^L\}) \in \Theta^L : n_{\sigma(1)} > \frac{n}{2}\}$. Then

$$\frac{|\Theta_1^L|}{\Theta^L} = 1 - \frac{|\Theta^L \setminus \Theta_1^L|}{|\Theta^L|} = 1 - \frac{\binom{n-1}{n/2-1}}{\binom{n}{n/2} + 2\binom{n}{n/2+1}} = 1 - \frac{(n/2-1)/n}{1 + \frac{n/2}{n/2+1}} > \frac{1}{2}.$$

Then with exactly the same argument used for $K \geq 3$ we finish the proof.

6.3. *Proof of Lemma 5.2.* (1) First consider the case when $\frac{nI}{K \log K} \to \infty$. Let $p(x)$ be the probability mass function of Z_i, and $M(t)$ be the moment generating function of Z_i. That is

$$M(t) = \mathbb{E}e^{tX_i}\mathbb{E}e^{-tY_i}$$
$$= \left(e^t\frac{b}{n} + 1 - \frac{b}{n}\right)\left(e^{-t}\frac{a}{n} + 1 - \frac{a}{n}\right).$$

The minimum of $M(t)$ is achieved at $t^\star = \frac{1}{2}\log\frac{a(1-b/n)}{b(1-a/n)}$, with $M(t^\star) = \left(\sqrt{\frac{a}{n}\frac{b}{n}} + \sqrt{(1-\frac{a}{n})(1-\frac{b}{n})}\right)^2$. This gives $I = -\log M(t^\star) = \max_t(-\log M(t))$. Let δ be a positive number which may depend on n. Denote $S_{n'} = \sum_{i=1}^{n'} Z_i$. Then

$$\mathbb{P}(S_{n'} \geq 0) \geq \sum_{n'\delta > S_{n'} \geq 0} \prod_{i=1}^{n'} p(z_i)$$
$$\geq \frac{M^{n'}(t^\star)}{\exp(n't^\star\delta)} \sum_{n'\delta > S_{n'} \geq 0} \prod_{i=1}^{n'} \frac{\exp(t^\star z_i)p(z_i)}{M(t^\star)},$$

where we use the fact that $\exp(n't^\star\delta) \geq \exp(t^\star \sum_i z_i) \geq \prod \exp(t^\star z_i)$ when $\sum_i z_i < n'\delta$. Denote $q(w) = \frac{\exp(t^\star w)p(w)}{M(t^\star)}$. Then

$$\mathbb{P}(S_{n'} \geq 0) \geq \frac{M^{n'}(t^\star)}{\exp(n't^\star\delta)} \sum_{n'\delta > S_{n'} \geq 0} \prod_{i=1}^{n'} q(z_i)$$
$$= \exp\left(-n'I\right)\exp(-n't^\star\delta) \sum_{n'\delta > S_{n'} \geq 0} \prod_{i=1}^{n'} q(z_i).$$

Note that $q(w)$ is a probability mass function, as $\sum_w \frac{\exp(t^\star w)p(w)}{M(t^\star)} = 1$. Let $W_1, W_2, \ldots, W_{n'}$ be i.i.d random variable with probability mass function

$q(w)$, then

$$\mathbb{P}(S_{n'} \geq 0) \geq \exp\left(-n'I\right)\exp(-n't^\star\delta)\mathbb{P}\left(\delta > \frac{1}{n'}\sum_{i=1}^{n'} W_i \geq 0\right).$$

A closer look on W_1 gives

$$\mathbb{P}(W_1 = 1) = \mathbb{P}(W_1 = -1) = \frac{1}{M(t^\star)}\sqrt{\frac{a}{n}\frac{b}{n}(1-\frac{a}{n})(1-\frac{b}{n})},$$

and $\mathbb{P}(W_1 = 0) = 1 - \mathbb{P}(W_1 = 1) - \mathbb{P}(W_1 = -1)$. Thus $\mathbb{E}W_1 = 0$ and

$$\mathrm{Var}(W_1) = \frac{2}{M(t^\star)}\sqrt{\frac{a}{n}\frac{b}{n}(1-\frac{a}{n})(1-\frac{b}{n})}.$$

Denote $V = \mathrm{Var}(\sum_{i=1}^{n'} W_i/n') = \mathrm{Var}(W_1)/n'$. We will later show that $I/(t^\star\sqrt{V}) \to \infty$. Now if it holds, then define $\delta = V^{\frac{1}{4}}I^{\frac{1}{2}}(t^\star)^{-\frac{1}{2}}$. It satisfies $\sqrt{V} = o(\delta)$. By the Central Limit Theorem, we have

$$\mathbb{P}\left(\delta > \frac{1}{n'}\sum_{i=1}^{n'} W_i \geq 0\right) \to \frac{1}{2}.$$

Together with $t^\star\delta = o(I)$ we have $\mathbb{P}(S_{n'} > 0) \geq \exp(-(1 + o(1))n'I)$.

To prove $I/(t^\star\sqrt{V}) \to \infty$, first consider the case with $a \asymp b$. Since $I \asymp (a-b)^2/na$, and $t^\star = \frac{1}{2}\log((1 + \frac{a-b}{b})(1 + \frac{a-b}{n(1-a/n)})) \asymp \frac{a-b}{a}$, and $\sqrt{V} \asymp \sqrt{aK}/n$, we have $\frac{I}{t^\star\sqrt{V}} \asymp \frac{a-b}{\sqrt{aK}} \to \infty$, implied by the fact $\frac{nI}{K} \asymp \frac{a-b}{\sqrt{aK}} \to \infty$. On the other hand, if $a/b \to \infty$ (recall we assume $b > \epsilon > 0$ and $a/n < 1 - \epsilon$), we have $I \asymp a/n$ and $M(t^\star) \asymp 1$. Note that $(\log\frac{a}{b})(\frac{b}{a})^{\frac{1}{4}}$ goes to 0, hence $t^\star\sqrt{V} = o(\sqrt{aK}/n)$. Then $\frac{I}{t^\star\sqrt{V}} \gg \sqrt{a/K}$. Since $nI/K \asymp a/K \to \infty$, $\frac{I}{t^\star\sqrt{V}}$ also goes to infinity.

(2) If $\frac{nI}{K} = O(1)$, we can choose δ such that $nt^\star\delta/K$ is also a constant. Then by considering the case $a \asymp b$ and $a/b \to \infty$ separately, we have $\frac{\delta}{\sqrt{V}} \asymp \frac{K}{nt^\star\sqrt{V}} \asymp \frac{K}{nI}$ with a similar argument used above. Thus $\mathbb{P}(S_{n'} > 0)$ is a constant.

6.4. *Proof of Lemma 5.3.* Due to the symmetry between σ and σ_0 (both are in the same parameter space), we have $\alpha(\sigma; \sigma_0) = \gamma(\sigma_0; \sigma)$ and $\gamma(\sigma; \sigma_0) = \alpha(\sigma_0; \sigma)$. It is sufficient to get the desired lower bound for $\gamma(\sigma; \sigma_0)$, as the same bound automatically holds for $\alpha(\sigma; \sigma_0)$.

By the definition of Θ^0 there must exist a $\eta_1 \to 0$ such that $|\frac{n_k}{n/K} - 1| \le \eta_1$ for every $k \in [K]$. First consider $m \le \frac{n}{2K}$. Without loss of generality, let σ satisfy

$$\sigma(i) = k, \forall i \in [\sum_{j=1}^{k-1} n'_j + 1, \sum_{j=1}^{k} n'_j],$$

where $\{n'_k\}_{k=1}^{K}$ are the sizes of communities in σ. Recall $\{n_k\}_{k=1}^{K}$ are the true community sizes in σ_0. Define $m_k = |\{i : \sigma(i) = k, \sigma_0(i) \ne k\}|$, then $m = \sum_k m_k$. For $k \in [K]$, define

$$\gamma_k(\sigma; \sigma_0) = |\{(i,j) : \sigma(i) = \sigma(j) = k, \sigma_0(i) \ne \sigma_0(j), i < j\}|$$

$$= \left| \left\{ (i,j) : \sigma_0(i) \ne \sigma_0(j), \sum_{j=1}^{k-1} n'_j + 1 \le i < j \le \sum_{j=1}^{k} n'_j \right\} \right|.$$

Obviously $\gamma(\sigma; \sigma_0) = \sum_{k=1}^{K} \gamma_k(\sigma; \sigma_0)$. We have $\gamma_k(\sigma; \sigma_0) \ge |\{i : \sigma(i) = k, \sigma_0(i) = k\}||\{i : \sigma(i) = k, \sigma_0(i) \ne k\}| = (n_k - m_k)m_k$. Then

$$\gamma(\sigma; \sigma_0) \ge \sum_k m_k(n_k - m_k) \ge \frac{(1 - o(1))mn}{K} - \sum_k m_k^2 \ge \frac{(1 - o(1))mn}{K} - m^2.$$

Now consider the case $m > \frac{n}{2K}$. Define $m_{k,k'} = |\{i : \sigma(i) = k, \sigma_0(i) = k'\}|$ for any $k, k' \in [K]$. It is obvious that equations $m_k = \sum_{k' \ne k} m_{k,k'}$, $n'_k = m_k + m_{k,k}$ and $n_{k'} = \sum_k m_{k,k'}$ hold for any k and k'.

It can be shown that we cannot find an pair of (k, k') such as $k \ne k'$ and $m_{k,k'} > \frac{2(1+\eta_1)n}{3K}$. Otherwise, if $m_{k,k'} > \frac{2(1+\eta_1)n}{3K}$, then $m_{k',k'} \le n_{k'} - m_{k,k'} < \frac{(1+\eta_1)n}{3K}$. Then we can exchange the label of k and k' to get a new estimation σ'. Compared with σ, this helps correctly recover at least $m_{k,k'} - (n'_k - m_{k,k'}) - m_{k',k'} > 0$ nodes. Since $\sigma' \in \Gamma(\sigma)$, then $m = d(\sigma_0, \sigma) \le d_H(\sigma_0, \sigma') < m$, which leads to a contradiction.

So we have $m_{k,k'} \le \frac{2(1+\eta_1)n}{3K}$ for all $k \ne k'$. For a given m_k, we have

$$\frac{\gamma_k(\sigma; \sigma_0)}{n'_k m_k} = \frac{\frac{1}{2}(n'^2_k - \sum_{k'} m^2_{k,k'})}{n'_k m_k},$$

with a constrain $m_k = \sum_{k' \ne k} m_{k,k'}$. When $m_k \le \frac{2(1+\eta_1)n}{3K}$, it can be shown that

$$\frac{\gamma_k(\sigma; \sigma_0)}{n'_k m_k} \ge \frac{\frac{1}{2}(n'^2_k - (n'_k - m_k)^2 - m^2_k)}{n'_k m_k} = \frac{n'_k - m_k}{n'_k} \ge \frac{(1 - 5\eta_1)\frac{n}{K}}{3n'_k}.$$

And when $m_k \geq \frac{2(1+\eta_1)n}{3K}$,

$$\frac{\gamma_k(\sigma;\sigma_0)}{n'_k m_k} \geq \frac{\frac{1}{2}(n'^2_k - (n'_k - m_k)^2 - (m_k - \frac{2(1+\eta_1)n}{3K})^2 - (\frac{2(1+\eta_1)n}{3K})^2)}{n'_k m_k}$$

$$\geq \frac{m_k(n'_k - m_k) + \frac{2(1+\eta_1)n}{3K}(m_k - \frac{2(1+\eta_1)n}{3K})}{n'_k m_k}$$

$$\geq \frac{2(1 - 5\eta_1)\frac{n}{K}}{9n'_k}.$$

Then sum up over all k and we get $\gamma(\sigma;\sigma_0) \geq \frac{2(1-5\eta_1)nm}{9K}$. By choosing $\eta = 5\eta_1$ the proof is complete.

6.5. *Proofs of Proposition 5.1 and 5.2.* We first present proposition 6.1, which is easy to be verified by coupling. It is helpful for the proof of Proposition 5.1.

PROPOSITION 6.1. *Let α and γ be arbitrary positive integers, and m take any value in $\mathbb{R}$. Define series of independent variables $\{X_i\}^\alpha_{i=1}$, $\{Y_i\}^\gamma_{i=1}$, $\{U_i\}^\alpha_{i=1}$ and $\{V_i\}^\gamma_{i=1}$. Let $U_i \sim Ber(p_i)$, $V_i \sim Ber(q_i)$, $X_i \sim Ber(p)$ and $Y_i \sim Ber(q)$ with $\min p_i \geq p$ and $\max q_i \leq q$. Then*

$$\mathbb{P}\Big(m + \sum_{i=1}^\alpha U_i \leq \sum_{i=1}^\gamma V_i\Big) \leq \mathbb{P}\Big(m + \sum_{i=1}^\alpha X_i \leq \sum_{i=1}^\gamma Y_i\Big).$$

PROOF OF PROPOSITION 5.1. Let $\{X_i\}_{1 \leq i \leq \gamma}$ be i.i.d $Ber(\frac{b}{n})$ random variables and $\{Y_i\}_{1 \leq i \leq \alpha}$ be i.i.d $Ber(\frac{a}{n})$ random variables, and $\{X_i\} \perp \{Y_i\}$. Then by Proposition 6.1, we have

$$\mathbb{P}(T(\sigma) \geq T(\sigma_0)) \leq \mathbb{P}\Big(\sum_{i=1}^\gamma X_i - \sum_{i=1}^\alpha Y_i \geq \lambda(\gamma - \alpha)\Big).$$

As an application of Markov inequality,

$$\mathbb{P}(T(\sigma) \geq T(\sigma_0)) \leq \mathbb{P}\Big(\exp\Big(t\sum^\gamma X_i - t\sum^\alpha Y_i\Big) \geq \exp(t\lambda(\gamma - \alpha))\Big)$$

$$\leq e^{-t\lambda(\gamma-\alpha)}\Big(\mathbb{E}e^{tX_1}\Big)^\gamma\Big(\mathbb{E}e^{-tY_1}\Big)^\alpha$$

$$= \Big(\mathbb{E}e^{tX_1}\mathbb{E}e^{-tY_1}\Big)^{(1-w)\alpha+w\gamma}\Big(\frac{(\mathbb{E}e^{tX_1})^{1-w}}{(\mathbb{E}e^{-tY_1})^w}e^{-t\lambda}\Big)^{\gamma-\alpha}$$

holds for any $t > 0$. Choose $t = t^\star$. Then $\mathbb{E}e^{t^\star X_1}\mathbb{E}e^{-t^\star Y_1} = e^{-I}$, and $\frac{(\mathbb{E}e^{t^\star X_1})^{1-w}}{(\mathbb{E}e^{-t^\star Y_1})^w}e^{-t^\star\lambda}$ is exactly equal to 1. Thus $\mathbb{P}(T(\sigma) \geq T(\sigma_0)) \leq e^{-\gamma I}$. □

PROOF OF PROPOSITION 5.2. Without loss of generality we assume that $d_H(\sigma, \sigma_0) = d(\sigma, \sigma_0)$. Then σ assigns m nodes with different values from σ_0, and there are K possible values for each node. Thus

$$\left| \left\{ \Gamma : \exists \sigma \in \Gamma \text{ s.t. } d(\sigma, \sigma_0) = m \right\} \right| \leq \binom{n}{m} K^m \leq \left(\frac{enK}{m} \right)^m.$$

In addition, since each node has at most K possible choices, we have a naive bound for the cardinality of Γ as $|\{\Gamma\}| \leq K^n$. $\square$

SUPPLEMENTARY MATERIAL

Supplement A: Supplement to "Mimimax Rates of Community Detection in Stochastic Block Models"

(url to be specified). In the supplement [30], we provide proofs for Theorems 2.1 and 3.1, which extend the minimax results of Theorems 2.2 and 3.2 to a larger parameter space Θ.

References.

[1] Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms. *arXiv preprint arXiv:1503.00609*, 2015.

[2] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.

[3] Arash A Amini, Aiyou Chen, Peter J Bickel, Elizaveta Levina, et al. Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41(4):2097–2122, 2013.

[4] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

[5] Peter Bickel, David Choi, Xiangyu Chang, Hai Zhang, et al. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics*, 41(4):1922–1943, 2013.

[6] 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.

[7] Tony Cai and Xiaodong Li. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *arXiv preprint arXiv:1404.6000*, 2014.

[8] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *arXiv preprint arXiv:1402.1267*, 2014.

[9] Peter Chin, Anup Rao, and Van Vu. Stochastic block model and community detection in the sparse graphs: A spectral algorithm with optimal rate of recovery. *arXiv preprint arXiv:1501.05021*, 2015.

[10] David Easley and Jon Kleinberg. Networks, crowds, and markets. *Cambridge University*.

[11] Chao Gao, Zongming Ma, Anderson Y. Zhang, and Harrison H. Zhou. Achieving optimal misclassification proportion in stochastic block model. *arXiv preprint arXiv:1312.1733*, 2013.

[12] 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.

[13] Lars Hagen and Andrew B Kahng. New spectral methods for ratio cut partitioning and clustering. *Computer-aided design of integrated circuits and systems, ieee transactions on*, 11(9):1074–1085, 1992.

[14] Bruce Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *arXiv preprint arXiv:1502.07738*, 2015.

[15] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[16] Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in sparse stochastic block models. *arXiv preprint arXiv:1312.2050*, 2013.

[17] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[18] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 694–703. ACM, 2014.

[19] Frank McSherry. Spectral partitioning of random graphs. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 529–537. IEEE, 2001.

[20] Elchanan Mossel, Joe Neeman, and Allan Sly. Stochastic block models and reconstruction. *arXiv preprint arXiv:1202.1499*, 2012.

[21] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *arXiv preprint arXiv:1311.4115*, 2013.

[22] Elchanan Mossel, Joe Neeman, and Allan Sly. Consistency thresholds for binary symmetric block models. *arXiv preprint arXiv:1407.1591*, 2014.

[23] Mark Newman. *Networks: an introduction*. Oxford University Press, 2010.

[24] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[25] 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.

[26] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.

[27] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[28] Piet Van Mieghem. *Performance analysis of communications networks and systems*. Cambridge University Press, 2006.

[29] Stanley Wasserman. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.

[30] Anderson Y. Zhang and Harrison H. Zhou. Supplement to "minimax rates of community detection in stochastic block models". 2015.

[31] Yunpeng Zhao, Elizaveta Levina, Ji Zhu, et al. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266–2292, 2012.

SUPPLEMENT TO "MIMIMAX RATES OF COMMUNITY DETECTION IN STOCHASTIC BLOCK MODELS"

BY Anderson Y. Zhang and Harrison H. Zhou

Yale University

APPENDIX A: ADDITIONAL PROOFS

In this appendix we provide the proofs of Theorem 2.1 and Theorem 3.1.

A.1. Proof of Theorem 2.1.

(1) For $K = 2$, the least favorable case for Θ is still Θ^0. The proof is identical to that of Theorem 2.2.

(2) For $K = 3$, it is always possible to have $\sigma \in \Theta$ such that a constant proportion of communities have size $\lfloor \frac{n}{\beta K} \rfloor$, and another constant proportion have the same size $\lceil \frac{n}{\beta K} \rceil$, with the rest communities have much larger size. Define Θ^L to contain all such σ. Then with identical arguments used to establish Lemma 5.1 and Lemma 5.2 we have

$$\inf_{\hat{\sigma}} \sup_{\Theta} \mathbb{E} r(\sigma, \hat{\sigma}) \geq \inf_{\hat{\sigma}} \sup_{\sigma \in \Theta^L} B_\tau(\hat{\sigma}(1))$$

$$\geq c\mathbb{P}\Big(\sum_{u=1}^{\lfloor n/\beta K \rfloor} X_u \geq \sum_{u=1}^{\lfloor n/\beta K \rfloor} Y_u \Big)$$

$$\geq \exp(-(1 + o(1))nI/\beta K).$$

A.2. Proof of Theorem 3.1 ($K = 2$).

Without loss of generality we assume $\frac{n}{2} = \lfloor \frac{n}{2} \rfloor$ throughout this section. For arbitrary $\sigma, \sigma_0 \in \Theta$ with $d(\sigma, \sigma_0) = m$, we can define $\alpha(\sigma; \sigma_0)$ and $\gamma(\sigma; \sigma_0)$ the same way as in Section 5.2. Note that $m \leq \frac{n}{2}$ since $d(\sigma, \sigma_0) = \min\{d_H(\sigma, \sigma_0), n - d_H(\sigma, \sigma_0)\}$. By Proposition 5.1, we have

$$\mathbb{P}(T(\sigma) \geq T(\sigma)) \leq \mathbb{P}\left(\sum_{i=1}^{\gamma} X_i - \sum_{i=1}^{\alpha} Y_i \geq \lambda(\gamma - \alpha) \Big| X_i \overset{iid}{\sim} \mathrm{Ber}(\frac{b}{n}), Y_i \overset{iid}{\sim} \mathrm{Ber}(\frac{a}{n}) \right).$$

Note that in $K = 2$ we have a specific equality as $\alpha + \gamma = m(n - m)$. Recall that $\lambda = -\frac{1}{2t^\star} \log \big(\frac{\frac{a}{n} \exp(-t^\star) + 1 - \frac{a}{n}}{\frac{b}{n} \exp(t^\star) + 1 - \frac{b}{n}} \big)$. By the Chernoff bound,

$$\mathbb{P}(T(\sigma) \geq T(\sigma_0)) \leq \big(\mathbb{E}e^{t^\star X_i} \big)^\gamma \big(\mathbb{E}e^{-t^\star Y_i} \big)^\alpha e^{-t^\star \lambda(\gamma - \alpha)}$$

$$= \big(\mathbb{E}e^{t^\star X_i} \mathbb{E}e^{-t^\star Y_i} \big)^{\frac{m(n-m)}{2}} \Big(\frac{\mathbb{E}e^{t^\star X_i}}{\mathbb{E}e^{-t^\star Y_i}} e^{-2t^\star \lambda'} \Big)^{\gamma - \frac{m(n-m)}{2}}$$

$$= \exp \Big(- \frac{m(n - m)I}{2} \Big),$$

where we use $\mathbb{E}e^{t^\star X_i}\mathbb{E}e^{-t^\star Y_i} = \exp(-I)$ and $e^{2t^\star\lambda'} = \frac{\mathbb{E}e^{t^\star X_i}}{\mathbb{E}e^{-t^\star Y_i}}$. The proof is similar to that of Theorem 3.2. Here we only include the key technique and omit the details. Assume $0 < \epsilon < 1/8$. Consider the following three cases:

(1) If $nI/2 > (1+\epsilon)\log n$, define $m_0 = 1$ and $m' = \epsilon n/2$. Then $P_1 \leq n\exp(-(n-1)I/2)$. Denote $R = n\exp(-(n-1)I/2)$. We have

$$P_m \leq \begin{cases} (\frac{2en}{2})^m \exp(-\frac{m(n-m)I}{2}) \leq Rn^{-\epsilon m/4}, & \text{for } m_0 < m \leq m' \\ (\frac{2en}{\epsilon n})^m \exp(-\frac{nmI}{4}) \leq R\exp(-\frac{n(m-4)I}{8}), & \text{for } m' < m \leq n/2. \end{cases}$$

Then $n\mathbb{E}r(\sigma, \hat{\sigma}) \leq \sum_{m=1}^{n/2} mP_m = (1+o(1))R$.

(2) If $nI/2 < (1-\epsilon)\log n$, define $m_0 = n\exp(-(1-e^{-\epsilon nI/2})nI/2)$ and $m' = n\exp(-nI/8)$. We have

$$P_m \leq \begin{cases} (\frac{2en}{m_0})^m \exp(-\frac{m(n-m')I}{2}) = \exp(-e^{-\frac{\epsilon nI}{2}}\frac{nmI}{4}), & \text{for } m_0 < m \leq m', \\ (\frac{2en}{m'})^m \exp(-\frac{nmI}{4}) \leq \exp(-\frac{nmI}{16}), & \text{for } m' < m \leq n/2. \end{cases}$$

Then $\mathbb{E}r(\sigma, \hat{\sigma}) \leq m_0/n + \sum_{m>m_0}^{n/2} P_m = (1+o(1))m_0/n$.

(3) If $\frac{nI}{2\log n} \to 1$, there exists a positive sequence $w \to 0$ such that $|\frac{nI}{2\log n} - 1| \ll w$ and $\frac{1}{\sqrt{\log n}} \leq w$. Define $m_0 = n\exp(-(1-w)nI/2)$ and $m' = w^2 n$.

$$P_m \leq \begin{cases} (\frac{2en}{m_0})^m \exp(-\frac{m(n-m')I}{2}) \leq \exp(-\frac{wnmI}{4}), & \text{for } m_0 < m \leq m' \\ (\frac{2en}{m'})^m \exp(-\frac{nmI}{4}) \leq \exp(-\frac{nmI}{8}), & \text{for } m' < m \leq n/2. \end{cases}$$

Then $\mathbb{E}r(\sigma, \hat{\sigma}) \leq m_0/n + \sum_{m>m_0}^{n/2} P_m = (1+o(1))m_0/n$.

A.3. Proof of Theorem 3.1 ($K \geq 3$). For the upper bound, we need the following lemma in replace of Lemma 5.3. Other than that, the proof is identical to that for Theorem 3.2 and thus omitted.

LEMMA A.1. *Assume $1 \leq \beta < \sqrt{\frac{5}{3}}$. Let $\sigma \in \Theta$ be an arbitrary assignment satisfying $d(\sigma, \sigma_0) = m$, where $0 < m < n$ is a positive integer. Then*

$$\alpha(\sigma; \sigma_0) \wedge \gamma(\sigma; \sigma_0) \geq \begin{cases} \frac{nm}{K\beta} - m^2, & \text{if } m \leq \frac{n}{2K}, \\ \frac{c_\beta nm}{K}, & \text{if } m > \frac{n}{2K}, \end{cases}$$

where $c_\beta = \frac{(5-3\beta^2)^2}{2\beta(1+3(5-3\beta^2)^2)}$.

PROOF OF LEMMA A.1. It is sufficient to show the equality for $\gamma(\sigma; \sigma_0)$. First consider the case $m \leq \frac{n}{2\beta K}$. Without loss of generosity, let σ satisfy

$$\sigma(i) = k, \forall i \in \left[\sum_{j=1}^{k-1} n'_j + 1, \sum_{j=1}^{k} n'_j\right].$$

Here $\{n'_k\}$ are sizes of all communities in σ. Assume $d_H(\sigma, \sigma_0) = m$, then $m = |\{i : \sigma(i) \neq \sigma_0(i)\}|$. Define $m_k = |\{i : \sigma(i) = k, \sigma_0(i) \neq k\}|$ then $m = \sum_k m_k$. For $k \in [K]$, define

$$\gamma_k(\sigma; \sigma_0) = |\{(i,j) : \sigma(i) = \sigma(j) = k, \sigma_0(i) \neq \sigma_0(j), i < j\}|$$

$$= \left|\left\{(i,j) : \sigma_0(i) \neq \sigma_0(j), \sum_{j=1}^{k-1} n'_j + 1 \leq i < j \leq \sum_{j=1}^{k} n'_j\right\}\right|.$$

We see that $\gamma(\sigma; \sigma_0) = \sum_{k=1}^{K} \gamma_k(\sigma; \sigma_0)$. We have $m_k \leq \frac{n}{2\beta K} \leq \frac{n'_k}{2}$, and also $\gamma_k(\sigma; \sigma_0) \geq |\{i : \sigma(i) = k, \sigma_0(i) = k\}||\{i : \sigma(i) = k, \sigma_0(i) \neq k\}| = m_k(n_k - m_k)$. Then

$$\gamma(\sigma; \sigma_0) \geq \sum_k m_k(n_k - m_k) \geq \frac{mn}{\beta K} - m^2.$$

Now consider the case that $m > \frac{n}{2\beta K}$. Define $m_{k,k'} = |\{i : \sigma(i) = k, \sigma_0(i) = k'\}|$ for any $k, k' \in [K]$. We see that equations $m_k = \sum_{k' \neq k} m_{k,k'}$ and $n'_k = m_k + m_{k,k}$ and $n_{k'} = \sum_k m_{k,k'}$ hold for all $k, k' \in [K]$.

For each $k \in [K]$, we want to get the value of $\gamma_k(\sigma; \sigma_0)$. We divide $k \in [K]$ into the following three categories:

(1) We say $k \in \mathcal{K}_1$ if for all $k' \neq k$, $m_{k,k'} \leq \frac{2}{3} n'_k$. For a given m_k, we have

$$\frac{\gamma_k(\sigma; \sigma_0)}{n'_k m_k} = \frac{\frac{1}{2}(n'^2_k - \sum_{k'} m^2_{k,k'})}{n'_k m_k},$$

with $m_k = \sum_{k' \neq k} m_{k,k'}$. When $m_k \leq \frac{2}{3} n'_k$, it is easy to check

$$\frac{\gamma_k(\sigma; \sigma_0)}{n'_k m_k} \geq \frac{\frac{1}{2}(n'^2_k - (n'_k - m_k)^2 - m^2_k)}{n'_k m_k} = \frac{n'_k - m_k}{n'_k} \geq \frac{1}{3}.$$

When $m_k > \frac{2}{3} n'_k$,

$$\frac{\gamma_k(\sigma; \sigma_0)}{n'_k m_k} \geq \frac{\frac{1}{2}(n'^2_k - (n'_k - m_k)^2 - (m_k - \frac{2}{3}n'_k)^2 - (\frac{2}{3}n'_k)^2)}{n'_k m_k}$$

$$\geq \frac{m_k(n'_k - m_k) + \frac{2}{3}n'_k(m_k - \frac{2}{3}n'_k)}{n'_k m_k}$$

$$\geq \frac{2}{9}.$$

Thus $\gamma_k(\sigma; \sigma_0) \geq \frac{2nm_k}{9\beta K}$ in both cases.

(2) We say $k \in \mathcal{K}_2$ if exists $k' \neq k$ such that $m_{k,k'} > \frac{2}{3}n'_k$. Claim $m_{k',k'} > \frac{1}{3}n'_k$. Otherwise, from σ we can exchange the labels k and k' to get a new estimator σ'. This helps to correctly recover at least $m_{k,k'} - m_{k,k} - m_{k',k'} > \frac{2}{3}n'_k - \frac{1}{3}n'_k - \frac{1}{3}n'_k > 0$ more nodes. Since $\sigma' \in \Gamma(\sigma)$, this implies $m = d(\sigma_0, \sigma) \leq d_H(\sigma_0, \sigma') < d_H(\sigma_0, \sigma) = m$, which leads to a contradiction.

On the other hand, we have $m_{k'} = n'_{k'} - m_{k',k'} \geq n'_{k'} - (n_{k'} - m_{k,k'}) \geq \frac{n}{\beta K} - \frac{\beta n}{K} + \frac{2n}{3\beta K} \geq \frac{(5-3\beta^2)n}{3\beta K} > 0$. This implies

$$\frac{\gamma_k(\sigma; \sigma_0) + \gamma_{k'}(\sigma; \sigma_0)}{m_k + m_{k'}} \geq \frac{\gamma_{k'}(\sigma; \sigma_0)}{m_k + m_{k'}} \geq \frac{m_{k',k'}m_{k'}}{m_k + m_{k'}} \geq \frac{\frac{1}{3}n'_k}{\frac{m_k}{m_{k'}} + 1} \geq \frac{\frac{n}{3\beta K}}{\frac{\beta}{(5-3\beta)^2/(3\beta)} + 1}.$$

Thus we have $\gamma_k(\sigma; \sigma_0) + \gamma_{k'}(\sigma; \sigma_0) \geq \frac{2c_\beta n(m_k + m_{k'})}{K} \geq \frac{2c_\beta n m_k}{K}$.

Apparently $[K] = \mathcal{K}_1 \cup \mathcal{K}_2$ and $\mathcal{K}_1 \cap \mathcal{K}_2 = \emptyset$. Claim for any $k \in \mathcal{K}_1$, there exists at most one $k' \neq k$ such that $m_{k',k} > \frac{2}{3}n'_{k'}$. Otherwise if there exists another $k'' \neq k'$ such that $k'' \neq k$ and $m_{k'',k} > \frac{2}{3}n'_{k''}$. Since $k', k'' \in \mathcal{K}_2$, this leads to $m_{k,k} \geq \frac{1}{3}(n'_k \vee n'_{k'})$. Then $n_k \geq m_{k',k} + m_{k'',k} + m_{k,k} > n'_{k'} + \frac{2}{3}n'_{k''} \geq \frac{5n}{3\beta K} > \frac{\beta n}{K}$ which leads to a contradiction. Note that $c_\beta \leq \frac{2}{9}$. Thus

$$\gamma(\sigma; \sigma_0) = \frac{1}{2} \sum_{k \in [K]} 2\gamma_k(\sigma; \sigma_0)$$

$$\geq \frac{1}{2} \left(\sum_{k \in \mathcal{K}_1} \frac{2nm_k}{9\beta K} + \sum_{k \in \mathcal{K}_2} \frac{2c_\beta n m_k}{K} \right)$$

$$\geq \frac{c_\beta n m}{K}.$$

$\square$

DEPARTMENT OF STATISTICS
YALE UNIVERSITY
NEW HAVEN, CT 06511
E-MAIL: ye.zhang@yale.edu
E-MAIL: huibin.zhou@yale.edu
URL: http://www.stat.yale.edu/~hz68/