The Annals of Statistics
2016, Vol. 44, No. 5, 2252–2280
DOI: 10.1214/15-AOS1428

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 attention 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, 18, 25, 29] 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 [30], biology [4] and the Internet [2]. Detecting and identifying communities is fundamentally important to understand the underlying structure of the network [13]. Many models and methodologies have been proposed for community detection from different perspectives, including RatioCut [14], Ncut [27], and spectral method [17, 20, 26] from computer sci-

---

Tribute: Harry heard the sad loss of Peter Hall on his way back home from an international trip on January 9, 2016. At that moment, all Harry wanted to do was to cry. Peter was not only a legendary scholar and a foremost leader of our field, but also one of the kindest people we had met. He was always thinking about what was the best for others, for the statistics community, and for our society. Four years ago, Peter was ranked as the top candidate for the *Annals of Statistics* editorship. Harry served on the IMS Committee to Select Editors then and had asked Peter, “Why did you agree to serve? We were thrilled and surprised by your decision.” Although Peter had done too much for our discipline already, he responded, “It would be better for my career to serve as an editor when I was younger. I need to pay back to our community.” Peter was a role model and a source of inspiration to us all. May he rest in peace.

Received July 2015; revised December 2015.
MSC2010 subject classifications. 60G05.
Key words and phrases. Network, community detection, stochastic block model, minimax rate.

ence, Newman–Girvan modularity [13] from physics, semi-definite programming [7, 15] 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. [21, 22] and Massoulie [19] considered balanced two-community sparse networks, and discovered the threshold phenomenon for both weak and strong consistency of community detection. Further extensions to slowly growing number of communities have been made in [1, 8, 15, 23]. Recently, in statistical literature, theoretical properties of various methods had been investigated as well in [5, 8, 9, 17, 26, 32], usually under weaker conditions and better suited for real data applications, but the convergence rates may often be suboptimal.

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 nonequal 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), that is, clustering error rates from $o(1)$ to $o(n^{-1})$. As a consequence, we obtain phase transitions for nonconsistency and strong consistency, under various settings, which recover the tight thresholds for phase transition in [8, 21–23].

The stochastic block model, proposed by [16], is possibly the most studied model in community detection [6, 17, 26]. Consider an undirected network with totally n nodes, and K communities labeled as $\{1, 2, \dots, K\}$. Each node is assigned to one community. Denote σ to be an assignment, and $\sigma(i)$ is the community assignment for the ith node. Let $n_k = |\{i : \sigma(i) = k\}|$ be the size of the kth community, for each $k \in \{1, 2, \dots, K\}$. 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'. In this paper, we focus on the case 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:

$$\Theta(n, K, a, b, \beta) \triangleq \Big\{(\sigma, \{\theta_{i,j}\}) : \sigma \in [K]^n, n_k \in \Big[\frac{n}{\beta K}, \frac{\beta n}{K}\Big],$$
$$\forall k \in [K], \{\theta_{i,j}\} \in [0,1]^{n\times n},$$
$$\theta_{i,j} \geq \frac{a}{n} \text{ if } \sigma(i) = \sigma(j) \text{ and } \theta_{i,j} \leq \frac{b}{n}$$
$$\text{if } \sigma(i) \neq \sigma(j), \theta_{i,i} = 0, \theta_{i,j} = \theta_{j,i}, \forall i \neq j\Big\},$$

where $\beta > 1$ 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. In this paper, we assume

$$0 < b < a < (1 - c_0)n, \tag{1.1}$$

where $c_0 > 0$ is any constant, allowing the network to be very sparse or 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.

Define I as

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

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 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*

$$\inf_{\hat{\sigma}} \sup_{\Theta(n,K,a,b,\beta)} \mathbb{E}r(\sigma, \hat{\sigma}) = \begin{cases} \exp\Big(-(1 + o(1))\dfrac{nI}{2}\Big), & K = 2, \\ \exp\Big(-(1 + o(1))\dfrac{nI}{\beta K}\Big), & K \geq 3, \end{cases} \tag{1.3}$$

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

Note that when K is finite, $nI \to \infty$ is a sufficient condition to yield equation (1.3) since it is equivalent to $\frac{nI}{K\log K} \to \infty$. The form of I is closely related to the Hellinger distance between those two Bernoulli probability measures. It is worth

TABLE 1
Summary of assumptions for lower and upper bounds

	Assumption on β	**Assumption on K**	**Theorems**
Lower bound	$\beta \geq 1 + \varepsilon_n$	$K \geq 2$	Theorems 2.1, 2.2
Upper bound	$\sqrt{5/3} > \beta \geq 1$	$nI/(K \log K) \to \infty$	Theorems 3.1, 3.2

pointing out that I is equal to $(a-b)^2/(an)$ up to a constant factor (cf. Lemma B.1 in the supplementary material [31]), which can be interpreted as the signal-to-noise ratio. In particular, when $a = o(n)$, I is equal to $(1 + o(1))(\sqrt{a} - \sqrt{b})^2/n$.

The lower bound of (1.3) 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 is $O(n^{-1})$ [i.e., when $a = O(1)$], 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 a finite number of communities, Theorem 1.1 shows $(a-b)^2/a \to \infty$ is a necessary and sufficient condition for consistent community detection, which implies consistency results in [21, 22]. It also recovers the strong consistency results in [15, 23], in which they additionally assume $a \asymp \log n$.

The minimax rate is of an exponential form, in contrast to the polynomial rates in [17, 26]. The term $\frac{nI}{K}$ plays a dominating role in determining the rate. Rewrite $\frac{nI}{\beta K}$ in the form of $\rho \log n$, and then we fail to recover approximately $n^{1-\rho}$ nodes. When $\rho > 1$, the network enjoys strong consistency property (exact recovery) since $n^{1-\rho} = o(1)$, that is, every node is correctly clustered. While for $0 < \rho < 1$, it is impossible to recover the communities exactly.

We provide Table 1 to summarize various assumptions for lower and upper bounds as follows.

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 $d_H(x, y) = |\{i : x_i \neq y_i\}|$, that is, the number of coordinates with different values. For any positive integer m, we use $[m]$ to denote the set $\{1, 2, \dots, m\}$. 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 distribution 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 $\lfloor z \rfloor = \max\{m \in \mathbb{Z} : m \leq z\}$ and $\lceil z \rceil = \min\{m \in \mathbb{Z} : m > z\}$. For any two scalars z_1 and z_2, denote $z_1 \wedge z_2$ to be $\min\{z_1, z_2\}$ and $z_1 \vee z_2$ to be $\max\{z_1, z_2\}$. 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 it homogeneous. A parameter space is called *homogeneous* if it only consists of homogeneous $\{\theta_{i,j}\}$. In such parameter space, $\{\theta_{i,j}\}$ is unique for any given σ. The homogeneous SBM is the most studied model in computer science and probability [8, 15, 21–23]. Define

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

The parameter space Θ_1 is homogeneous. 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 parameter space.

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 helpful 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 ith 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

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

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 parameter space of SBM that is homogeneous and 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*

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

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*

$$\inf_{\hat{\sigma}} \sup_{\Theta_1(n,K,a,b,\beta)} \mathbb{E}r(\sigma,\hat{\sigma}) \geq \begin{cases} \exp\Bigl(-\bigl(1+o(1)\bigr)\dfrac{nI}{2}\Bigr), & K=2, \\ \exp\Bigl(-\bigl(1+o(1)\bigr)\dfrac{nI}{\beta K}\Bigr), & K\geq 3, \end{cases} \tag{2.1}$$

where $\beta \geq 1+\varepsilon_n$ *for some* $\varepsilon_n = CK/n$ *with constant* C *large enough. 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 constant* $c > 0$.

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 nonequal-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 where communities are almost of the same size. Networks of community sizes exactly equal to n/K are the most studied settings [8, 9, 22]. Here, we allow a small fluctuation of community sizes. Denote Θ^0 as follows:

$$\begin{aligned} \Theta^0(n,K,a,b) \triangleq \Bigl\{\bigl(\sigma,\{\theta_{i,j}\}\bigr) : \sigma \in [K]^n, n_k \in \Bigl[\frac{n}{(1+\varepsilon_n)K}, \frac{(1+\varepsilon_n)n}{K}\Bigr], \forall k \in [K], \\ \theta_{i,i} = 0, \forall i \in [n], \theta_{i,j} = \frac{a}{n} \text{ if } \sigma(i) = \sigma(j) \\ \text{and } \theta_{i,j} = \frac{b}{n} \text{ if } \sigma(i) \neq \sigma(j), \forall i \neq j\Bigr\}, \end{aligned}$$

where $\varepsilon_n = o(1)$ is any positive sequence satisfying $\varepsilon_n \geq CK/n$ for some constant C large enough, such that the parameter space contains networks with fluctuation in community sizes. Note that $\Theta^0(n,K,a,b)$ is $\Theta_1(n,K,a,b,\beta)$ with $\beta = 1+\varepsilon_n$, for which we have the following minimax lower bound.

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

$$\inf_{\hat{\sigma}} \sup_{\Theta^0(n,K,a,b)} \mathbb{E}r(\sigma,\hat{\sigma}) \geq \exp\Bigl(-\bigl(1+o(1)\bigr)\frac{nI}{K}\Bigr). \tag{2.2}$$

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 constant* $c > 0$.

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 [31], 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

$$
\begin{aligned}
L(\sigma; A) = \log\left(\frac{a}{n}\right)\sum_{i<j} A_{i,j}1_{\{\sigma(i)=\sigma(j)\}} + \log\left(1-\frac{a}{n}\right)\sum_{i<j}(1-A_{i,j})1_{\{\sigma(i)=\sigma(j)\}} \\
+ \log\left(\frac{b}{n}\right)\sum_{i<j} A_{i,j}1_{\{\sigma(i)\neq\sigma(j)\}} \\
+ \log\left(1-\frac{b}{n}\right)\sum_{i<j}(1-A_{i,j})1_{\{\sigma(i)\neq\sigma(j)\}}.
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
\hat{\sigma}^{\mathrm{MLE}} &= \arg\max_{\sigma} L(\sigma; A) \\
&= \arg\max_{\sigma}\left(\log\frac{a(1-b/n)}{b(1-a/n)}\sum_{i<j} A_{i,j}1_{\{\sigma(i)=\sigma(j)\}}\right. \\
&\qquad \left.- \log\frac{1-b/n}{1-a/n}\sum_{i<j} 1_{\{\sigma(i)=\sigma(j)\}}\right). \qquad (3.1)
\end{aligned}
$$

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, that is, the maximum likelihood estimator over all σ with a community size n/K for every community has a simpler form, $\hat{\sigma}^{\mathrm{MLE}} = \arg\max_\sigma \sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}}$.

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), that is,

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

where we set

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

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

$$t^\star = \frac{1}{2} \log \frac{a(1-b/n)}{b(1-a/n)}. \tag{3.3}$$

It can be shown that $t^\star$ is the minimizer of the moment generating function for the difference of two Bernoulli variables, that is, $t^\star = \arg\min_{t>0} \mathbb{E}e^{t(X-Y)}$, where $X \sim \mathrm{Ber}(\frac{b}{n})$ and $Y \sim \mathrm{Ber}(\frac{a}{n})$. It is equivalent to write λ in equation (3.2) as follows:

$$\begin{aligned}
\lambda &= -\frac{1}{2t^\star} \log\left(\frac{(a/n)\exp(-t^\star) + 1 - a/n}{(b/n)\exp(t^\star) + 1 - b/n}\right) \\
&= -\frac{1}{2t^\star} \log\left(\frac{a}{n}\exp(-t^\star) + 1 - \frac{a}{n}\right) \\
&\quad + \frac{1}{2t^\star} \log\left(\frac{b}{n}\exp(t^\star) + 1 - \frac{b}{n}\right).
\end{aligned}$$

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

$$\lambda = \begin{cases} -\frac{1}{2t^\star}\log\left(\frac{a}{n}e^{-t^\star}+1-\frac{a}{n}\right) \\ \qquad +\frac{1}{2t^\star}\log\left(\frac{b}{n}e^{t^\star}+1-\frac{b}{n}\right), & K=2, \\ -\frac{w}{t^\star}\log\left(\frac{a}{n}e^{-t^\star}+1-\frac{a}{n}\right) \\ \qquad +\frac{(1-w)}{t^\star}\log\left(\frac{b}{n}e^{t^\star}+1-\frac{b}{n}\right), & K\geq 3, \end{cases} \tag{3.4}$$

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 (3.4) with $w=1/2$. In Section 3.3, we give theoretical properties of penalized likelihood estimation for all λ in equation (3.4).

3.3. *Minimax upper bound.* For the general SBM Θ, the risk upper bound of the penalized likelihood estimator, for every λ in equation (3.4), 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* (3.4), *we have*

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

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

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* (3.4), *we have*

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

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 [31].

4. Discussion.

4.1. *Comparison and connection with prior works.* We follow the notation and definition of strong/weak consistency as in Mossel et al. [21–23]. A global-to-local scheme was introduced in [23] to study phase transitions and thresholds. In this paper, we developed a different global-to-local scheme independently to build a strong connection between global and local rates by a Bayesian approach.

Mossel et al. [21–23] obtained thresholds for strong/weak consistency for $K = 2$ (i.e., two-community case), while in this paper we study minimax rates for arbitrary K under much weaker assumptions on a and b. As we will show next, the minimax rates in Theorem 1.1 immediately imply phase transitions and thresholds under various settings. When the minimax rates are equal to $o(1/n)$ or $o(1)$, we can obtain critical values for strong and weak consistency, respectively.

(1) *Let* $a = o(n)$ *and* $(a - b)/a = o(1)$. Under this scenario, the difference of within-community probability and between-community probability is relatively small. Note that $I = (1+o(1))(a-b)^2/(4an)$ (cf. Lemma B.1), which reduces the minimax result into the form of $\exp(-(1+o(1))(a-b)^2/(4aK))$. When $K = 2$, Theorem 1.1 implies the results from [21, 22]. 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 for consistency. In this paper, K can be as large as $n/\log n$, and a, b can take any value from 0 to n. Similarly, we also obtain the sharp threshold for strong consistency in [23] when $K = 2$.

(2) *Let a and b be an 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 $I = (1 + o(1))(\sqrt{e_1} - \sqrt{e_2})^2 \log n/n$. Under the assumption $K = n^{o(1)}$, Theorem 1.1 implies that there exists a strongly consistent estimator if $\liminf_{n\to\infty}(\sqrt{e_1} - \sqrt{e_2})/\sqrt{K} > 1$. This threshold is obtained in [15] but only for finite K. In particular, 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 [23].

It is worth mentioning that in [22] $(a-b)^2 = 2(a+b)$ is proved to be a critical threshold to do better than random guess, and an efficient algorithm is proposed to outperform random guess as long as $(a-b)^2 > 2(a+b)$.

4.2. *Comments on the minimax rates.* Theorem 1.1 implies the minimax rates are determined by sizes of smallest two communities, which are the most difficult parts of the network to be correctly recovered. In this paper, we consider $\Theta(n, K, a, b, \beta)$ with a bound on β^2, the ratio of the largest and smallest community size, due to the technical difficulties in deriving upper bounds. We think the same bounds hold for a wide range of β values.

Under the assumption $nI/(K \log K) \to \infty$, we establish a lower bound $\exp(-(1 + o(1))nI/(\beta K))$. However, we conjecture that the exact minimax rate is $(K - 1)\exp(-(1 + o(1))nI/(\beta K))$, which is equivalent to $\exp(-(1 + o(1))nI/(\beta K))$ when $nI/(K \log K) \to \infty$. We think the same proof scheme leads to the sharper lower bound with more sophisticated technical arguments, though we have not succeeded yet. If our goal is only to obtain a lower bound $\exp(-(1+o(1))nI/(\beta K))$, there exists a much simpler way [12] to replace the role of Lemma 2.1. The key idea there is to find a smaller parameter space with no identifiability issue of labels as follows. Consider the parameter space with assignments of $n/(4\beta K)$ nodes undecided. The labels for the remaining $n - n/(4\beta K)$ nodes are the same and known. The distance between any two assignments is then just the Hamming distance, since the distance is smaller than $n/(2\beta K)$, half of the smallest community size. Thus, the identifiablity issue is avoided. Eventually, we can show the global minimax risk is lower bounded by local minimax risk, up to a factor of K^{-1}.

4.3. *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 a 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 computational 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, Theorems 2.2 and 3.2. The proofs of Theorems 2.1 and 3.1 are almost identical to those of Theorems 2.2 and 3.2. We put them in the supplement material [31].

5.1. *Proof of Theorem* 2.2. To obtain the lower bound for the parameter space Θ^0, we will first construct and analyze a least favorable case in terms of the sizes of the communities. In particular, the community sizes only take values 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 a constant proportion of K. Note that $\lfloor\frac{n}{K}\rfloor-1$ or $\lceil\frac{n}{K}\rceil+1$ as community size is allowed in Θ^0 since we assume $\varepsilon_n \geq CK/n$, as long as $C \geq 3$.

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 $\varepsilon > 0$ such that $\varepsilon K < \min(K_1, K_2) \leq \max(K_1, K_2) < (1-\varepsilon)K$; and (2) either of the following two conditions:

$$\left\lfloor\frac{n}{K}\right\rfloor K_1 + \left(\left\lfloor\frac{n}{K}\right\rfloor+1\right)K_2 + \left(\left\lfloor\frac{n}{K}\right\rfloor-1\right)K_3 = n; \quad \text{or} \tag{5.1}$$

$$\left\lceil\frac{n}{K}\right\rceil K_1 + \left(\left\lceil\frac{n}{K}\right\rceil+1\right)K_2 + \left(\left\lceil\frac{n}{K}\right\rceil-1\right)K_3 = n. \tag{5.2}$$

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 $r \geq 2\varepsilon K$ and $r \leq (1-2\varepsilon)K$ for a constant $\varepsilon > 0$, we have $n = \lfloor \frac{n}{K} \rfloor (K-r) + (\lfloor \frac{n}{K} \rfloor + 1)r$, which satisfies equation (5.1). Otherwise, if $r < 2\varepsilon K$ 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 $K - r > 2\varepsilon K$, we may argue similarly to obtain equation (5.2).

Recall that we use n_k to denote the size of the kth 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 $\varepsilon K < \min(K_1, K_2) \leq \max(K_1, K_2) < (1-\varepsilon)K$. Define a subparameter space of Θ^0 as follows:

$$\begin{aligned}
&\Theta^L\big(n, K, a, b, \{K_i\}\big) \\
&\quad = \bigg\{\big(\sigma, \{\theta_{i,j}\}\big) \in \Theta^0(n, K, a, b) : \bigg|\bigg\{k : n_k = \bigg\lfloor \frac{n}{K} \bigg\rfloor\bigg\}\bigg| = K_1, \\
&\qquad \bigg|\bigg\{k : n_k = \bigg\lfloor \frac{n}{K} \bigg\rfloor + 1\bigg\}\bigg| = K_2, \bigg|\bigg\{k : n_k = \bigg\lfloor \frac{n}{K} \bigg\rfloor - 1\bigg\}\bigg| = K_3\bigg\}.
\end{aligned}$$

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

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 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 $\varepsilon > 0$ such that*

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

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

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 $\{X_i\} \overset{\text{i.i.d.}}{\sim} \text{Ber}(\frac{b}{n})$, $\{Y_i\} \overset{\text{i.i.d.}}{\sim} \text{Ber}(\frac{a}{n})$, and $\{X_i\} \perp \{Y_i\}$, for $i = 1, 2, \dots, n'$. If $\frac{nI}{K} \to \infty$, we have*

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

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

Lemma 5.2 provides an explicit expression for the lower bound. The proof mainly follows the proof of the Cramér–Chernoff theorem [28]. The general Cramér–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 minmax risk, we have $\inf_{\hat{\sigma}} \sup_{\sigma \in \Theta^L} \mathbb{E}r(\sigma, \hat{\sigma}) \geq \inf_{\hat{\sigma}} B_\tau(\hat{\sigma})$. By the fact that Θ^L is a homogeneous parameter space and also closed under permutation for both $K \geq 3$ and $K = 2$, Lemma 2.1 implies $\inf_{\hat{\sigma}} B_\tau(\hat{\sigma}) = \inf_{\hat{\sigma}} B_\tau(\hat{\sigma}(1))$. Thus,

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

which, together with Lemmas 5.1 and 5.2, implies Theorem 2.2. $\square$

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 equivalence 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 equivalence class, and $\{\Gamma\}$ as the set consisting of all the possible equivalence classes with respect to Θ^0. For any $\sigma_1, \sigma_2 \in \Theta$, define the distance between σ_1 and σ_2 as

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

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

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

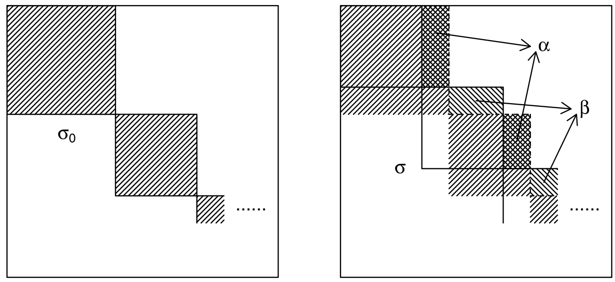

FIG. 1. *Each block filled by* 45 *degree lines stands for all the within-community connection in one single community. The areas inside the squares are all the* $A_{i,j}$ *entries summed up*. Left: *For* $2\sum_{i<j} A_{i,j}\{\sigma_0(i)=\sigma_0(j)\}$, *the squares exactly overlap with the within-community connection regions*. Right: *For* $2\sum_{i<j} A_{i,j}\{\sigma(i)=\sigma(j)\}$, *there are some differences between the squares and within-community connection parts*, *which are labeled as* α *or* γ *according to their positions*.

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

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

for any integer m with $0 < m < n$. The key step is to have a tight bound of the probability $\mathbb{P}(T(\sigma) \geq T(\sigma_0))$ 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 Figure 1. The parts filled with 45 degree lines 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 filled blocks, that is, all the Bernoulli random variables with success probability $\frac{a}{n}$ in the adjacency matrix.

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

$$\begin{aligned} &\sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}} - \sum_{i<j} A_{i,j} 1_{\{\sigma_0(i)=\sigma_0(j)\}} \\ &\quad = \sum_{i<j} A_{i,j} 1_{\{\sigma(i)=\sigma(j)\}} 1_{\{\sigma_0(i)\neq\sigma_0(j)\}} \\ &\qquad - \sum_{i<j} A_{i,j} 1_{\{\sigma(i)\neq\sigma(j)\}} 1_{\{\sigma_0(i)=\sigma_0(j)\}}. \end{aligned} \tag{5.4}$$

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

$$\sum_{i<j} 1_{\{\sigma(i)=\sigma(j)\}} - \sum_{i<j} 1_{\{\sigma_0(i)=\sigma_0(j)\}} = \alpha - \gamma. \tag{5.5}$$

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*

$$\mathbb{P}\big(T(\sigma) \geq T(\sigma_0)\big) \leq \exp\big(-(\alpha \wedge \gamma)I\big),$$

for λ defined in equation (3.4).

Note that the value of γ and α depends on σ and σ_0. Lemma 5.3 provides a lower bound on γ and α for each m. It implies that $\alpha \wedge \gamma$ is lower bounded by $(1 - o(1))nm/K$ if $m = o(n/K)$, where the factor $(1 - o(1))$ is tight and essential to our proof of Theorem 3.2.

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 there exists a positive sequence $\eta \to 0$, independent of the choice of σ, such that*

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

Lemma 5.3, together with Proposition 5.1, immediately implies an upper bound on $\mathbb{P}(T(\sigma) \geq T(\sigma_0))$ 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*

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

for λ defined in equation (3.4).

We will apply a union bound to obtain 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 equivalence class repetitively. Proposition 5.2 gives an upper bound for cardinality of the equivalence classes $\{\Gamma\}$.

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

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

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

Recall the definition of P_m in equation (5.3). With Proposition 5.2 and the union bound, we are able to have a satisfactory bound by

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

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 the 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 $\varepsilon > 0$ such that $\frac{(1-\eta)nI}{K\log n} > 1 + \varepsilon$. Note that Lemma 5.4 still holds for any positive sequence goes to 0 slower than η. Thus, we can assume η decay slowly such that both $\frac{\eta nI}{K\log K}$ and $\frac{\eta n}{K}$ go to infinity. We have $P_1 \leq nK\exp(-(\frac{(1-\eta)n}{K} - 1)I) \leq R$, where $R \triangleq n\exp(-(1-2\eta)nI/K)$. 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 $m' = \frac{\varepsilon n}{3K}$, we have

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

where in the fourth inequality we use $\frac{(1-\eta)nI}{K\log n} > 1+\varepsilon$, and in the last inequality we use the fact that $I \lesssim 1$ to show $e^I n^{-\varepsilon/6} < 1$ 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

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

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

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 $\varepsilon > 0$ such that $\frac{(1-\eta)nI}{K\log n} < 1-\varepsilon$. Let $m_0 = n\exp(-(1-K^{-\varepsilon/2})\frac{(1-\eta)nI}{K})$, which satisfies both $m_0 \geq (nK)^{\varepsilon/2}$ 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 $m' = \frac{n}{K^{1+\varepsilon}}$, we have

$$\begin{aligned}
P_m &\leq \left(\left(\frac{enK}{m_0}\right)\exp\left(-\frac{(1-\eta)nI}{K} + m'I\right)\right)^m \\
&\leq \left(\exp\left(\log(nK) + \left((1-K^{-\varepsilon/2}) - (1-2K^{-\varepsilon})\right)\frac{(1-\eta)nI}{K}\right)\right)^m \\
&\leq \exp\left(-\frac{m}{2K^{\varepsilon/2}}\frac{(1-\eta)nI}{K}\right),
\end{aligned}$$

which is denoted as Q_m. Since $\frac{m_0}{K^{\varepsilon/2}} \gg \log n$, we have $\sum_{m=m_0}^{m'} P_m \leq \sum_{m=m_0}^{m'} Q_m \leq m'Q_{m_0} \leq \exp(\log n - \frac{m_0}{2K^{\varepsilon/2}}\frac{(1-\eta)nI}{K}) = o(\frac{m_0}{n})$. 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 $Q_m = \exp(-\frac{nmI}{9K})$, 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,

$$\begin{aligned}
\mathbb{E}r(\sigma,\hat{\sigma}_0) &\leq \frac{m_0}{n} + \mathbb{P}\big(\exists\sigma\in\Theta^0 : d(\sigma_0,\sigma) > m_0 \text{ and } l(\sigma) \geq l(\sigma_0)\big) \\
&\leq \frac{m_0}{n} + \sum_{m>m_0}^{m'} P_{m'} + \sum_{m>m'}^{n} P_m \\
&\leq \frac{m_0}{n} + m'Q_{m_0} + 2Q_{m'} \\
&= \exp\left(-\frac{(1-o(1))nI}{K}\right).
\end{aligned}$$

(3) If $\frac{nI}{K\log n} = 1 + o(1)$, there exists a positive sequence $\omega \to 0$ such that $|\frac{(1-\eta)nI}{K\log n} - 1| \ll \omega$, $\frac{1}{\sqrt{\log n}} \leq \omega$ and $\frac{\omega nmI}{K\log K} \to \infty$. Define $m_0 = n\exp(-(1-\omega)\frac{(1-\eta)nI}{K})$. Thus, $m_0 \geq n^{\omega/2} \to \infty$, and $m_0 = o(m')$ for $m' = \omega^2 n/K$. We are

going to find a fast decay series $\{Q_m\}$ to upper bound $\{P_m\}$. For $m \in [m_0, m']$,

$$P_m \le \left(\left(\frac{enK}{m_0}\right)\exp\left(-\frac{(1-\eta)nI}{K} + m'I\right)\right)^m$$
$$\le \left(\log(eK) + \frac{(1-\omega)(1-\eta)nI}{K} - \frac{(1-\eta)nI}{K} + \frac{\omega^2 nI}{K}\right)^m$$
$$\le \exp\left(-\frac{\omega(1-\eta)nmI}{4K}\right),$$

which is denoted as Q_m. Note that $\omega m_0 \ge \omega n^{\omega/2} \to \infty$. 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 $Q_m = \exp(-\frac{nmI}{9K})$, which decays geometrically fast. Then $\sum_{m=m'}^{n} P_m \le \sum_{m=m'}^{n} Q_m \le 2Q_{m'} = o(\frac{m_0}{n})$. 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 Lemmas 2.1, 5.1 and 5.3, 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 $\pi : [1, 2, \dots, n] \to [1, 2, \dots, n]$ 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

$$\sigma_\pi(i) \triangleq \sigma\left(\pi^{-1}(i)\right) \qquad \forall 1 \le i \le n.$$

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

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

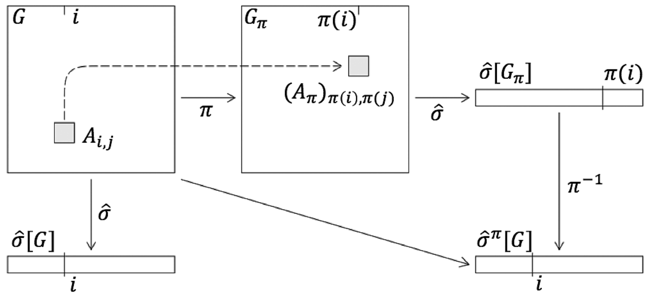

FIG. 2. *Illustration on $\hat{\sigma}^\pi$ based on the original network G. All of $\hat{\sigma}[G]$, $\hat{\sigma}[G_\pi]$ and $\hat{\sigma}^\pi[G]$ are demonstrated as n-by-1 vectors. It shows $A_{i,j}$ becomes $(A_\pi)_{\pi_i,\pi_j}$ after the permutation π of the network. For any specific node i in G, its location is changed into $\pi(i)$ in G_π. The procedure $\hat{\sigma}[G_\pi]$ estimates the assignment of the permuted nodes $\{\pi(i)\}$, while $\hat{\sigma}^\pi[G]$ estimates the assignment of the original nodes.*

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

We prove Lemma 2.1 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 ith 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 have $\hat{\sigma}[G_\pi]$, and then finally "permute back" to obtain the estimation in the original order. To be more precise, define procedure $\hat{\sigma}^\pi$ as

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

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

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 equal global risk for all $\sigma \in \Lambda$ and 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, that is, $B_\tau(\tilde{\sigma}) = \inf_{\hat{\sigma}} B_\tau(\hat{\sigma})$. 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 τ. For an arbitrary $\sigma \in \Lambda$, we have (we add subscript σ to explicitly indicate that the expectation is taken with respect to the assignment σ)

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

Recall that $\mathbb{E}_\sigma(\sigma, \tilde{\sigma}^\pi) = \frac{1}{n}\mathbb{E}_\sigma \inf_{\sigma' \in \Gamma(\tilde{\sigma}^\pi)} d_H(\sigma, \sigma')$. There exists a one-to-one relation between $\Gamma(\tilde{\sigma}^\pi)$ and $\Gamma(\tilde{\sigma}[G_\pi])$, in 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:

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

The expectation can be further expanded into

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

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

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

which yields

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

Thus,

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

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

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

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

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

which implies

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

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

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

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-to-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

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

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

$$\begin{aligned}\mathbb{E}_\sigma r\big(\sigma(i), \tilde{\sigma}^\pi(i)\big) &= \sum_{G\in\mathbb{G}}\bigg(\sum_{\sigma''\in S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])} \frac{1\{\sigma_\pi(\pi(i)) \neq \sigma''(\pi(i))\}}{|S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])|}\bigg)\mathbb{P}_\sigma(G)\\ &= \sum_{G\in\mathbb{G}}\bigg(\sum_{\sigma''\in S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])} \frac{1\{\sigma_\pi(\pi(i)) \neq \sigma''(\pi(i))\}}{|S_{\sigma_\pi}(\tilde{\sigma}[G_\pi])|}\bigg)\mathbb{P}_{\sigma_\pi}(G_\pi)\\ &= \sum_{G\in\mathbb{G}}\bigg(\sum_{\sigma''\in S_{\sigma_\pi}(\tilde{\sigma}[G])} \frac{1\{\sigma_\pi(\pi(i)) \neq \sigma''(\pi(i))\}}{|S_{\sigma_\pi}(\tilde{\sigma}[G])|}\bigg)\mathbb{P}_{\sigma_\pi}(G).\end{aligned}$$

Thus,

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

This gives

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

Then for the local risk we have

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

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

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

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. $\square$

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

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

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,

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

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

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

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

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

Note that $\sigma_0(1) \notin \kappa(\sigma_0)$. If we replace $\sigma_0(1)$ by any $k \in \kappa(\sigma_0)$ while keeping the labels of the rest of the nodes, then we have a new assignment also contained in Θ_1^L and which has distance 1 from σ_0. In particular, we use the following procedure to generate assignment $\sigma[\sigma_0] \in [K]^n$ based on σ_0. Let $\sigma[\sigma_0](i)$ be the index of the ith entry in $\sigma[\sigma_0]$. Define

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

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

$$\begin{aligned} B_\tau(\hat{\sigma}(1)) &\geq \frac{\varepsilon}{2|\Theta_1^L|} \sum_{\sigma_0 \in \Theta_1^L} 2\mathbb{E} r(\sigma_0(1), \hat{\sigma}(1)) \\ &\geq \frac{\varepsilon}{|\Theta_1^L|} \sum_{\sigma_0 \in \Theta_1^L} \frac{1}{2}\big(\mathbb{E}_{\sigma_0} r(\sigma_0(1), \hat{\sigma}(1)) + \mathbb{E}_{\sigma[\sigma_0]} r(\sigma[\sigma_0](1), \hat{\sigma}(1))\big). \end{aligned}$$

We are going to derive the Bayes risk $\inf_{\hat{\sigma}} \frac{1}{2}(\mathbb{E}_{\sigma_0} r(\sigma_0(1), \hat{\sigma}(1)) + \mathbb{E}_{\sigma[\sigma_0]} r(\sigma[\sigma_0](1), \hat{\sigma}(1)))$ for a given $\sigma_0 \in \Theta^L$. Let $\tilde{\sigma}$ be any estimator achieving the infimum. Since σ_0 and $\sigma[\sigma_0]$ only differ at the first node, $\tilde{\sigma}$ must satisfy $\tilde{\sigma}(i) = \sigma(i) = \sigma[\sigma_0](i)$ for $i \geq 2$ and $\tilde{\sigma}(1)$ must take value either $\sigma_0(1)$ or $\sigma[\sigma_0](1)$. Thus, $r(\sigma_0(1), \tilde{\sigma}(1)) = d_H(\sigma_0(1), \tilde{\sigma}(1))$ and a similar equation holds for $\sigma[\sigma_0]$. So now the problem is reduced into a testing problem between two distributions $\mathbb{P}_{\sigma_0}$ and $\mathbb{P}_{\sigma[\sigma_0]}$, which is just a test between two sequences of Bernoulli random variables, since the two corresponding matrices are different only at $2\lfloor n/K \rfloor$ entries of the first row/column.

We are going to show that majority voting (i.e., likelihood ratio) gives the optimal estimator. 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 be the set $\{u \in [n] \setminus \{1\} : \sigma_0(u) = \sigma_0(1)\}$, and $J_1 = \{u \in [n] : \sigma_0(u) = \sigma[\sigma_0](1)\}$. For a given adjacency matrix A, the conditional distributions are

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

and

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

Here, A^C consists of all the rest of the entries: $A^C = \{(u, v) : v > u \geq 2, \text{ or } u = 1 \text{ and } v \notin J_0 \cup J_1\}$. 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, $\mathbb{E}r(\sigma_0(1), \hat{\sigma}(1)) = \mathbb{P}_{\sigma_0}(\sum_{u \in J_0} A_{1,u} < \sum_{u \in J_1} A_{1,u}) \geq \mathbb{P}(\sum_{u=1}^{\lfloor n/K \rfloor} X_u \geq \sum_{u=1}^{\lfloor n/K \rfloor} Y_u)$, and $\mathbb{E}r(\sigma[\sigma_0](1), \hat{\sigma}(1)) = \mathbb{P}_{\sigma[\sigma_0]}(\sum_{u \in J_0} A_{1,u} \geq \sum_{u \in J_1} A_{1,u}) \geq \mathbb{P}(\sum_{u=1}^{\lfloor n/K \rfloor} X_u \geq \sum_{u=1}^{\lfloor n/K \rfloor} Y_u)$. Consequently,

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

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

$$\begin{aligned}
\inf_{\hat{\sigma}} B_\tau\big(\hat{\sigma}(1)\big) &\geq \frac{\varepsilon}{|\Theta_1^L|} \sum_{\sigma_0 \in \Theta_1^L} \inf_{\hat{\sigma}} \frac{1}{2}\big(\mathbb{E}_{\sigma_0} r\big(\sigma_0(1), \hat{\sigma}(1)\big) + \mathbb{E}_{\sigma[\sigma_0]} r\big(\sigma[\sigma_0](1), \hat{\sigma}(1)\big)\big) \\
&\geq \frac{\varepsilon}{|\Theta_1^L|} \sum_{\sigma_0 \in \Theta_1^L} \frac{1}{2}\big(\mathbb{E}_{\sigma_0} r\big(\sigma_0(1), \tilde{\sigma}(1)\big) + \mathbb{E}_{\sigma[\sigma_0]} r\big(\sigma[\sigma_0](1), \tilde{\sigma}(1)\big)\big) \\
&\geq \varepsilon \mathbb{P}\left(\sum_{u=1}^{\lfloor n/K \rfloor} X_u \geq \sum_{u=1}^{\lfloor n/K \rfloor} Y_u\right).
\end{aligned}$$

For the case $K = 2$, we re-define Θ_1^L and show that its cardinality is the same as 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

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

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

6.3. *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 $|\frac{n_k}{n/K} - 1| \leq \eta_1$ for every $k \in [K]$. First consider $m \leq \frac{n}{2K}$. Without loss of generality, let σ satisfy

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

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 $m_k = |\{i : \sigma(i) = k, \sigma_0(i) \neq k\}|$, then $m = \sum_k m_k$. For $k \in [K]$, define

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

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

$$\gamma(\sigma;\sigma_0) \geq \sum_k m_k(n_k - m_k) \geq \frac{(1-\eta_1)mn}{K} - \sum_k m_k^2 \geq \frac{(1-\eta_1)mn}{K} - m^2.$$

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 $m_k = \sum_{k' \neq k} m_{k,k'}$, $n'_k = m_k + m_{k,k}$ 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 $k \neq k'$ and $m_{k,k'} > \frac{2(1+\eta_1)n}{3K}$. Otherwise, if $m_{k,k'} > \frac{2(1+\eta_1)n}{3K}$, then $m_{k',k'} \leq n_{k'} - m_{k,k'} < \frac{(1+\eta_1)n}{3K}$. 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 $m = d(\sigma_0, \sigma) \leq d_H(\sigma_0, \sigma') < m$, which leads to a contradiction.

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

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

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

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

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

$$\begin{aligned}\frac{\gamma_k(\sigma;\sigma_0)}{n'_k m_k} &\geq \frac{1}{2}\left(n_k'^2 - (n'_k - m_k)^2 - \left(m_k - \frac{2(1+\eta_1)n}{3K}\right)^2 - \left(\frac{2(1+\eta_1)n}{3K}\right)^2\right) \\ &\quad \Big/ (n'_k m_k) \\ &\geq m_k(n'_k - m_k) + \frac{2(1+\eta_1)n}{3K}\left(m_k - \frac{2(1+\eta_1)n}{3K}\right) \Big/ (n'_k m_k) \\ &\geq \frac{2(1-5\eta_1)(n/K)}{9n'_k}.\end{aligned}$$

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.

Acknowledgements. We thank the Associate Editor and the referees for their constructive comments and suggestions that lead to the improvement of the paper.

SUPPLEMENTARY MATERIAL

Supplement to "Mimimax rates of community detection in stochastic block models" (DOI: 10.1214/15-AOS1428SUPP; .pdf). In the supplement [31], we provide proofs of Lemma 5.2, Propositions 5.1 and 5.2. We also 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. In addition, we state and prove the asymptotic equivalence of I.

REFERENCES

[1] ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient recovery algorithms. Preprint. Available at arXiv:1503.00609.

[2] ALBERT, R., JEONG, H. and BARABÁSI, A.-L. (1999). Hawoong Jeong, and Albert-László Barabási. Internet: Diameter of the world-wide web. *Nature* **401** 130–131.

[3] AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. MR3127859

[4] BARABÁSI, A. and OLTVAI, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5** 101–113.

[5] BICKEL, P., CHOI, D., CHANG, X. and ZHANG, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.* **41** 1922–1943. MR3127853

[6] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

[7] CAI, T. T. and LI, X. (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Ann. Statist.* **43** 1027–1059. MR3346696

[8] CHEN, Y. and XU, J. (2014). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. Preprint. Available at arXiv:1402.1267.

[9] CHIN, P., RAO, A. and VU, V. (2015). Stochastic block model and community detection in the sparse graphs: A spectral algorithm with optimal rate of recovery. Preprint. Available at arXiv:1501.05021.

[10] EASLEY, D. and KLEINBERG, J. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge Univ. Press, Cambridge. MR2677125

[11] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2015). Achieving optimal misclassification proportion in stochastic block model. Preprint. Available at arXiv:1505.03772.

[12] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2016). Optimal community detection in degree-corrected block model. Manuscript.

[13] GIRVAN, M. and NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99** 7821–7826 (electronic). MR1908073

[14] HAGEN, L. and KAHNG, A. B. (1992). New spectral methods for ratio cut partitioning and clustering. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **11** 1074–1085.

[15] HAJEK, B., WU, Y. and XU, J. (2015). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. Preprint. Available at arXiv:1502.07738.

[16] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5** 109–137. MR0718088

[17] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605

[18] LOVÁSZ, L. (2012). *Large Networks and Graph Limits. American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035

[19] MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC'*14—*Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 694–703. ACM, New York. MR3238997

[20] MCSHERRY, F. (2001). Spectral partitioning of random graphs. In 42*nd IEEE Symposium on Foundations of Computer Science* (*Las Vegas*, *NV*, 2001) 529–537. IEEE Computer Soc., Los Alamitos, CA. MR1948742

[21] MOSSEL, E., NEEMAN, J. and SLY, A. (2012). Stochastic block models and reconstruction. Preprint. Available at arXiv:1202.1499.

[22] MOSSEL, E., NEEMAN, J. and SLY, A. (2013). A proof of the block model threshold conjecture. Preprint. Available at arXiv:1311.4115.

[23] MOSSEL, E., NEEMAN, J. and SLY, A. (2014). Consistency thresholds for binary symmetric block models. Preprint. Available at arXiv:1407.1591.

[24] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45** 167–256 (electronic). MR2010377

[25] NEWMAN, M. E. J. (2010). *Networks*: *An Introduction*. Oxford Univ. Press, Oxford. MR2676073

[26] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856

[27] SHI, J. and MALIK, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22** 888–905.

[28] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

[29] VAN MIEGHEM, P. (2006). *Performance Analysis of Communications Networks and Systems*. Cambridge Univ. Press, Cambridge.

[30] WASSERMAN, S. (1994). *Social Network Analysis*: *Methods and Applications* **8**. Cambridge Univ. Press, Cambridge.

[31] ZHANG, A. Y. and ZHOU, H. H. (2016). Supplement to "Minimax rates of community detection in stochastic block models." DOI:10.1214/15-AOS1428SUPP.

[32] ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. MR3059083

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