Submitted to the Annals of Statistics

COMMUNITY DETECTION IN DEGREE-CORRECTED BLOCK MODELS

By Chao Gao,* Zongming Ma,† Anderson Y. Zhang,‡ and Harrison H. Zhou‡

University of Chicago, University of Pennsylvania, Yale University and Yale University

Community detection is a central problem of network data analysis. Given a network, the goal of community detection is to partition the network nodes into a small number of clusters, which could often help reveal interesting structures. The present paper studies community detection in Degree-Corrected Block Models (DCBMs). We first derive asymptotic minimax risks of the problem for a misclassification proportion loss under appropriate conditions. The minimax risks are shown to depend on degree-correction parameters, community sizes, and average within and between community connectivities in an intuitive and interpretable way. In addition, we propose a polynomial time algorithm to adaptively perform consistent and even asymptotically optimal community detection in DCBMs.

1. Introduction. In many fields such as social science, neuroscience and computer science, it has become increasingly important to process and make inference on relational data. The analysis of network data, a prevalent form of relational data, becomes an important topic for statistics and machine learning. One central problem of network data analysis is *community detection*: to partition the nodes in a network into subsets. A meaningful partition of nodes can often uncover interesting information that is not apparent in a complicated network.

An important line of research on community detection is based on Stochastic Block Models (SBMs) [15]. For any $p \in [0,1]$, let $\mathrm{Bern}(p)$ be the Bernoulli distribution with success probability p. Under an SBM with n nodes and k communities, given a symmetric probability matrix $B = (B_{uv}) = B^T \in [0,1]^{k\times k}$ and a label vector $z = (z(1),\ldots,z(n))^T \in [k]^n$, where $[k] = \{1,\ldots,k\}$ for any $k \in \mathbb{N}$, its adjacency matrix $A = (A_{ij}) \in \{0,1\}^{n\times n}$, with ones encoding edges, is assumed to be symmetric with zero diagonals and

*The research of C. Gao is supported in part by NSF DMS-1712957.

†The research of Z. Ma is supported in part by NSF Career Award DMS-1352060 and a Sloan Research Fellowship.

‡The research of A.Y. Zhang and H.H. Zhou is supported in part by NSF DMS-1209191 and NSF DMS-1507511.

AMS 2000 subject classifications: Primary 62H30,91D30; secondary 62C20,90B15

Keywords and phrases: Clustering, Minimax rates, Network analysis, Spectral clustering, Stochastic block model

$A_{ij} = A_{ji} \overset{ind.}{\sim} \text{Bern}(B_{z(i)z(j)})$ for all $i > j$. In other words, the probability of an edge connecting any pair of nodes only depends on their community memberships. To date, researchers in physics, computer science, probability theory and statistics have gained great understanding on community detection in SBMs. See, for instance, [5, 9, 20, 21, 19, 22, 2, 13, 7, 14, 1, 10, 26] and the references therein. Despite a rich literature dedicated to their theoretical properties, SBMs suffer significant drawbacks when it comes to modeling real world social and biological networks. In particular, due to the model assumption, all nodes within the same community in an SBM are exchangeable and hence have the same degree distribution. In comparison, nodes in real world networks often exhibit degree heterogeneity even when they belong to the same community [23]. For example, Bickel and Chen [5] showed that for a karate club network, SBM does not provide a good fit for the data set, and the resulting clustering analysis is qualitatively different from the truth.

One way to accommodate degree heterogeneity is to introduce a set of degree-correction parameters $\{\theta_i : i = 1, \ldots, n\}$, one for each node, which can be interpreted as the popularity or importance of a node in the network. Then one could revise the edge distributions to $A_{ij} = A_{ji} \overset{ind.}{\sim} \text{Bern}(\theta_i\theta_j B_{z(i)z(j)})$ for all $i > j$, and this gives rise to the Degree-Corrected Block Models (DCBMs) [8, 17]. In a DCBM, within the same community, a node with a larger value of degree-correction parameter is expected to have more connections than that with a smaller value. On the other hand, SBMs are special cases of DCBMs in which the degree-correction parameters are all equal. Empirically, the larger class of DCBMs is able to provide possibly much better fits to many real world network datasets [23]. Throughout the paper, we allow k and B to scale with n as n tends to infinity. Since the proposal of the model, there have been various methods proposed for community detection in DCBMs, including but not limited to spectral clustering [24, 18, 16, 12] and modularity based approaches [17, 27, 4, 6]. On the theoretical side, [11] provides an information-theoretic characterization of the impossibility region of community detection for DCBMs with two clusters, and sufficient conditions have been given in [27, 6] for strongly and weakly consistent community detection. However, two fundamental statistical questions remain unanswered:

- What are the fundamental limits of community detection in DCBMs?
- Once we know these limits, can we achieve them adaptively via some polynomial time algorithm?

The answer to the first question can provide important benchmarks for comparing existing approaches and for developing new procedures. The answer to the second question can lead to new practical methodologies with theoretically justified optimality. The present paper is dedicated to provide answers to these two questions.

Main contributions. Our main contributions are two-folded. First, we carefully formulate community detection in DCBMs as a decision-theoretic problem and then work out its asymptotic minimax risks with sharp constant in the exponent under certain regularity conditions. For example, let k be a fixed constant. Suppose there are k communities all about the same size n/k and the average within community and between community edge probabilities are p and q respectively with $p > q$ and $p/q = O(1)$, then under mild regularity conditions, the minimax risk under the loss function that counts the proportion of misclassified nodes takes the form

$$\left[\frac{1}{n}\sum_{i=1}^{n}\exp\left(-\theta_i\frac{n}{k}(\sqrt{p}-\sqrt{q})^2\right)\right]^{1+o(1)} \tag{1}$$

as $n \to \infty$ whenever it converges to zero and the maximum expected node degree scales at a sublinear rate with n. The general fundamental limits to be presented in Section 2 allow the community sizes to differ and the number of communities k to grow to infinity with n. To the best our knowledge, this is the first minimax risk result for community detection in DCBMs. The minimax risk (1) has an intuitive form. In particular, the i^{th} term in the summation can be understood as the probability of the i^{th} node being misclassified. When θ_i is larger, the chance of the node being misclassified gets smaller as it has more edges and hence more information of its community membership is available in the network. The term n/k is roughly the community size. Since the community detection problem can be reduced to a hypothesis testing problem with n/k as its effective sample size, a larger k implies a more difficult problem. Furthermore, $(\sqrt{p}-\sqrt{q})^2$ reflects the degree of separation among the k clusters. Note that p and q are the average within and between community edge probabilities and so $(\sqrt{p}-\sqrt{q})^2$ measures the difference of edge densities within and between communities. If the clusters are more separated in the sense that the within and between community edge densities differ more, the chance of each node being misclassified becomes smaller. When the degree-correction parameters are all equal to one and $p = o(1)$, the expression in (1) reduces to the minimax risk of community detection in SBMs in [26].

In addition, we investigate computationally feasible algorithms for adaptively achieving minimax optimal performance. In particular, we propose a polynomial time two-stage algorithm. In the first stage, we obtain a relatively crude community assignment via a weighted k-medians procedure on a low-rank approximation to the adjacency matrix. Working with a low-rank approximation (as opposed to the leading eigenvectors of the adjacency matrix) enables us to avoid common eigen-gap conditions needed to establish weak consistency for spectral clustering methods. Based on result of the first stage, the second stage applies a local optimization to improve on the community assignment of each network node. Theoretically, we show that it

can adaptively achieve asymptotic minimax optimal performance for a large collection of parameter spaces. The empirical effectiveness of the algorithm is illustrated by simulation.

Connection to previous work. The present paper is connected to a number of papers on community detection in DCBMs and SBMs.

It is connected to the authors' previous work on minimax community detection in SBMs [10, 26]. However, the involvement of degree-correction parameters poses significant new challenges. For the study of fundamental limits, especially minimax lower bounds, the fundamental two-point testing problem in DCBMs compares two product probability distributions with different marginals, while in SBMs, the two product distributions can be divided to two equal sized blocks within which the marginals are the same. Consequently, a much more refined Cramér–Chernoff argument is needed to establish the desired bound. In addition, to establish matching minimax upper bounds, the analysis of the maximum likelihood estimators is technically more challenging than that in [26] due to the presence of degree-correction parameters and the wide range in which they can take values. In particular, we use a new folding argument to obtain the desired bounds. For adaptive estimation, the degree-correction parameters further increase the number of nuisance parameters. As a result, although we still adopt a "global-to-local" two-stage strategy to construct the algorithm, neither stage of the proposed algorithm in the present paper can be borrowed from the algorithm proposed in [10]. We will give more detailed comments on the first stage below. For the second stage, the penalized neighbor voting approach in [10] requires estimation of degree-correction parameters with high accuracy and hence is infeasible. We propose a new *normalized* neighbor voting procedure to avoid estimating θ_i's.

The first stage of the proposed algorithm is connected to the literature on spectral clustering, especially [16]. The novelty in our proposal is that we cluster the rows of a low-rank approximation to the adjacency matrix directly as opposed to the rows of the matrix containing the leading eigenvectors of the adjacency matrix. As a result, the new spectral clustering algorithm does not require any eigen-gap condition to achieve consistency.

Organization. After a brief introduction to common notation, the rest of the paper is organized as follows. Section 2 presents the decision-theoretic formulation of community detection in DCBMs and derives matching asymptotic minimax lower and upper bounds under appropriate conditions. Given the fundamental limits obtained, we propose in Section 3 a polynomial time two-stage algorithm and study when a version of it can adaptively achieve minimax optimal rates of convergence. The finite sample performance of the proposed algorithm is examined in Section 4 on simulated data examples. Some proofs of the main results are given in Section 5 with additional proofs deferred to the appendices.

Notation. For an integer d, we use $[d]$ to denote the set $\{1, 2, ..., d\}$. For a positive real number x, $\lceil x \rceil$ is the smallest integer no smaller than x and $\lfloor x \rfloor$ is the largest integer no larger than x. For a set S, we use $\mathbf{1}_{\{S\}}$ to denote its indicator function and $|S|$ to denote its cardinality. For a vector $v \in \mathbb{R}^d$, its norms are defined by $\|v\|_1 = \sum_{i=1}^n |v_i|$, $\|v\|^2 = \sum_{i=1}^n v_i^2$ and $\|v\|_\infty = \max_{1\le i\le n} |v_i|$. For two matrices $A, B \in \mathbb{R}^{d_1\times d_2}$, their trace inner product is defined as $\langle A, B\rangle = \sum_{i=1}^{d_1}\sum_{j=1}^{d_2} A_{ij}B_{ij}$. The Frobenius norm and the operator norm of A are defined by $\|A\|_{\mathrm{F}} = \sqrt{\langle A, A\rangle}$ and $\|A\|_{\mathrm{op}} = s_{\max}(A)$, where $s_{\max}(\cdot)$ denotes the largest singular value.

2. Fundamental Limits. In this section, we present fundamental limits of community detection in DCBMs. We shall first define an appropriate parameter space and a loss function. A characterization of asymptotic minimax risks then follows.

2.1. *Parameter Space and Loss Function.* Recall that a random graph of size n generated by a DCBM has its adjacency matrix A satisfying $A_{ii} = 0$ for all $i \in [n]$ and

$$A_{ij} = A_{ji} \overset{ind}{\sim} \mathrm{Bern}(\theta_i\theta_j B_{z(i)z(j)}) \ \text{ for all } i \neq j \in [n]. \tag{2}$$

For each $u \in [k]$ and a given $z \in [k]^n$, we let $n_u = n_u(z) = \sum_{i=1}^n \mathbf{1}_{\{z(i)=u\}}$ be the size of the u^{th} community. Let $P = \mathbb{E}[A] \in [0,1]^{n\times n}$. We propose to consider the following parameter space for DCBMs of size n:

$$\begin{aligned}\mathcal{P}_n(\theta, p, q, k, \beta; \delta) = \big\{ & P \in [0,1]^{n\times n} : \exists z \in [k]^n \text{ and } B = B^T \in \mathbb{R}^{k\times k}, \\ & \text{s.t. } P_{ii} = 0, P_{ij} = \theta_i\theta_j B_{z(i)z(j)}, \forall i \neq j \in [n], \\ & \frac{1}{n_u}\sum_{z(i)=u} \theta_i \in [1-\delta, 1+\delta], \forall u \in [k], \\ & \max_{u\neq v} B_{uv} \le q < p \le \min_u B_{uu}, \\ & \frac{n}{\beta k} - 1 \le n_u \le \frac{\beta n}{k} + 1, \ \forall u \in [k]\big\}.\end{aligned} \tag{3}$$

We are mostly interested in the behavior of minimax risks over a sequence of such parameter spaces as n tends to infinity and the key model parameters θ, p, q, k scale with n in some appropriate way. On the other hand, we take $\beta \ge 1$ as an absolute constant and require the (slack) parameter δ to be an $o(1)$ sequence throughout the paper.

To see the rationale behind the definition in (3), let us examine each of the parameters used in the definition. The starting point is $\theta \in \mathbb{R}_+^n$, which we treat for now as a given sequence of degree-correction parameters. Given

θ, we consider all possible label vectors z such that the approximate normalization $\frac{1}{n_u}\sum_{z(i)=u}\theta_u = 1+o(1)$ holds for all communities. The introduction of the slack parameter $0 < \delta = o(1)$ rules out those parameter spaces in which community detection can be trivially achieved by only examining the normalization of the θ_i's. On the other hand, the proposed normalization ensures that for all $u \neq v \in [k]$,

$$B_{uu} \approx \frac{1}{n_u(n_u-1)} \sum_{i:z(i)=u} \sum_{j\neq i:z(j)=u} P_{ij} \quad \text{and} \quad B_{uv} \approx \frac{1}{n_u n_v} \sum_{i:z(i)=u} \sum_{j:z(j)=v} P_{ij}.$$

Therefore, B_{uu} and B_{uv} can be understood as the (approximate) average connectivity within the u^{th} community and between the u^{th} and the v^{th} communities, respectively. Under this interpretation, p can be seen as a lower bound on the within community connectivities and q an upper bound on the between community connectivities. We require the assumption $p > q$ to ensure that the model is "assortative" in an average sense. Finally, we also require the individual community sizes to be contained in the interval $[n/(\beta k) - 1, \beta n/k + 1]$. In other words, the community sizes are assumed to be of the same order. Although we have focused on the case of assortative networks, we expect the same expression of minimax rates to hold in the disassortative case, i.e., $\min_{u\neq v} B_{uv} \geq q > p \geq \max_u B_{uu}$.

REMARK 1. An interesting special case of the parameter space in (3) is when $\theta = 1_n$, where $1_n \in \mathbb{R}^n$ is the all one vector. In this case, the parameter space reduces to one for assortative stochastic block models.

As for the loss function, we use the following misclassification proportion that has been previously used in the investigation of community detection in stochastic block models [26, 10]:

$$\ell(\hat{z}, z) = \frac{1}{n} \min_{\pi\in\Pi_k} H(\hat{z}, \pi(z)), \tag{4}$$

where $H(\cdot,\cdot)$ is the Hamming distance defined as $H(z_1, z_2) = \sum_{i\in[n]} \mathbf{1}_{\{z_1(i)\neq z_2(i)\}}$ and Π_k is the set of all permutations on $[k]$. Here, the minimization over all permutations is necessary since we are only interested in comparing the partitions resulting from z and $\hat{z}$ and so the actual labels used in defining the partitions should be inconsequential.

2.2. *Minimax Risks.* Now we study the minimax risk of the problem

$$\inf_{\hat{z}} \sup_{\mathcal{P}_n(\theta,p,q,k,\beta;\delta)} \mathbb{E}\,\ell(\hat{z}, z). \tag{5}$$

In particular, we characterize the asymptotic behavior of (5) as a function of n, θ, p, q, k and β. The key information-theoretic quantity that governs the

minimax risk of community detection is I, which is defined through

$$\exp(-I) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} \exp\left(-\theta_i \frac{n}{2}(\sqrt{p}-\sqrt{q})^2\right), & k = 2, \\ \frac{1}{n}\sum_{i=1}^{n} \exp\left(-\theta_i \frac{n}{\beta k}(\sqrt{p}-\sqrt{q})^2\right), & k \geq 3. \end{cases} \tag{6}$$

Note that I depends on n not only directly but also through θ, p, q and k.

Minimax upper bounds. Given any parameter space $\mathcal{P}_n(\theta, p, q, k, \beta; \delta)$, we can define the following estimator:

$$\hat{z} = \underset{z' \in \mathcal{P}_n(\theta,p,q,k,\beta;\delta)}{\operatorname{argmax}} \prod_{1 \leq i < j \leq n} \big[(\theta_i\theta_j p)^{A_{ij}}(1-\theta_i\theta_j p)^{1-A_{ij}} \mathbf{1}_{\{z'(i)=z'(j)\}} + (\theta_i\theta_j q)^{A_{ij}}(1-\theta_i\theta_j q)^{1-A_{ij}} \mathbf{1}_{\{z'(i)\neq z'(j)\}}\big]. \tag{7}$$

If there is a tie, we break it arbitrarily. The estimator (7) is the maximum likelihood estimator for a special case of DCBM where $B_{uu} = p$ and $B_{uv} = q$ for all $u \neq v \in [k]$. In other cases, the objective function in (7) is a misspecified likelihood function. For any sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n = \Omega(b_n)$ if $a_n \geq C b_n$ for some absolute constant $C > 0$ for all $n \geq 1$. The following theorem characterizes the asymptotic behavior of the risk bounds for the estimator (7).

THEOREM 1 (Minimax Upper Bounds). *Consider any sequence* $\{\mathcal{P}_n(\theta, p, q, k, \beta; \delta)\}_{n=1}^{\infty}$ *such that as* $n \to \infty$, $I \to \infty$, $p > q$, $\|\theta\|_\infty = o(n/k)$, $\min_{i\in[n]} \theta_i = \Omega(1)$ *and* $\log k = o(\min(I, \log n))$. *When* $k \geq 3$, *further assume* $\beta \in [1, \sqrt{5/3})$. *Then the estimator in* (7) *satisfies*

$$\limsup_{n\to\infty} \frac{1}{I} \log\left(\sup_{\mathcal{P}_n(\theta,p,q,k,\beta;\delta)} \mathbb{E}\,\ell(\hat{z}, z)\right) \leq -1.$$

Before proceeding, we briefly discuss the conditions in Theorem 1. First, the condition $\min_{i\in[n]} \theta_i = \Omega(1)$ requires that all θ_i's are at least of constant order. One should note that this condition does not rule out the possibility that $\max_i \theta_i \gg \min_i \theta_i$, and so a great extent of degree variation, even within the same community, is allowed. Next, $\log k = o(\log n)$ requires that the number of communities k, if it diverges to infinity, grows at a sub-polynomial rate with the number of nodes n. Furthermore, $\beta \in [1, \sqrt{5/3})$ is a technical condition that we need for a combinatorial argument in the proof to go through when $k \geq 3$. When $k = O(1)$ and $\Omega(1) = \min_i \theta_i \leq \|\theta\|_\infty = O(1)$, Theorem 1 only requires $I \to \infty$, which is equivalent to $n(p-q)^2/p \to \infty$. Informed readers might find the result in Theorem 1 in parallel to that in [26]. However, due to the presence of degree-correction parameters, the proof of Theorem 1 is significantly different from that of the corresponding result in [26]. For example, a new folding argument is employed to deal with degree heterogeneity.

Minimax lower bounds. We now show that the rates in Theorem 1 are asymptotic minimax optimal by establishing matching minimax lower bounds. To this end, we require the following condition on the degree-correction parameters $\theta \in \mathbb{R}_+^n$. The condition guarantees that $\mathcal{P}_n(\theta, p, q, k, \beta; \delta)$ is nonempty. Moreover, it is only needed for establishing minimax lower bounds.

CONDITION N. *We say that $\theta \in \mathbb{R}_+^n$ satisfies Condition N if*

1. *When $k = 2$, there exists a disjoint partition $\mathcal{C}_1, \mathcal{C}_2$ of $[n]$, such that $|\mathcal{C}_1| = \lfloor n/2 \rfloor$, $|\mathcal{C}_2| \in \{\lfloor n/2 \rfloor, \lfloor n/2 \rfloor + 1\}$ and $|\mathcal{C}_u|^{-1} \sum_{i \in \mathcal{C}_u} \theta_i \in (1 - \delta/4, 1 + \delta/4)$ for $u = 1, 2$.*
2. *When $k \geq 3$, there exists a disjoint partition $\{\mathcal{C}_u\}_{u \in [k]}$ of $[n]$, such that $|\mathcal{C}_1| \leq |\mathcal{C}_2| \leq \ldots \leq |\mathcal{C}_k|$, $|\mathcal{C}_1| = |\mathcal{C}_2| = \lfloor n/(\beta k) \rfloor$ and $|\mathcal{C}_u|^{-1} \sum_{i \in \mathcal{C}_u} \theta_i \in (1 - \delta/4, 1 + \delta/4)$ for all $u \in [k]$.*

We note that the condition is only on θ (as opposed to the parameter space) and the actually communities in the data generating model need not coincide with the partition that occurs in the statement of the condition.

With the foregoing definition, we have the following result.

THEOREM 2 (Minimax Lower Bounds). *Consider any sequence $\{\mathcal{P}_n(\theta, p, q, k, \beta; \delta)\}_{n=1}^{\infty}$ such that as $n \to \infty$, $I \to \infty$, $1 < p/q = O(1)$, $p\|\theta\|_\infty^2 = o(1)$, $\log k = o(I)$, $\log(1/\delta) = o(I)$ and θ satisfies Condition N. Then*

$$\liminf_{n \to \infty} \frac{1}{I} \log \left(\inf_{\hat{z}} \sup_{\mathcal{P}_n(\theta, p, q, k, \beta; \delta)} \mathbb{E}\, \ell(\hat{z}, z) \right) \geq -1.$$

Compared with the conditions in Theorem 1, the conditions of Theorem 2 are slightly different. The condition $1 < p/q = O(1)$ ensures that the smallest average within community connectivity is of the same order as (albeit larger than) the largest average between community connectivity. Such an assumption is typical in the statistical literature on block models. The condition $\|\theta\|_\infty^2 p = o(1)$ ensures that the maximum expected node degree scales at a sublinear rate with the network size n. Furthermore, when $k = O(1)$, the condition $\log k = o(I)$ can be dropped because it is equivalent to $I \to \infty$, which in turn is necessary for the minimax risk to converge to zero.

Combining both theorems, we have the minimax risk of the problem.

COROLLARY 1. *Under the conditions of Theorems 1 and 2, we have*

$$\inf_{\hat{z}} \sup_{\mathcal{P}_n(\theta, p, q, k, \beta; \delta)} \mathbb{E}\ell(\hat{z}, z) = \exp(-(1 + o(1))I),$$

where $o(1)$ stands for a sequence whose absolute values tend to zero as n tends to infinity.

Setting $\beta = 1$ in Corollary 1 leads to the minimax result (1) in the introduction. We refer to Section 1 for the meanings of the terms in I.

REMARK 2. When $\theta = 1_n$, the foregoing minimax risk reduces to the corresponding result for stochastic block models [26] in the sparse regime where $q < p = o(1)$. In this case, (6) implies that the minimax risk is

$$\exp(-(1+o(1))I) = \begin{cases} \exp\left(-(1+o(1))\frac{n}{2}(\sqrt{p}-\sqrt{q})^2\right), & k = 2, \\ \exp\left(-(1+o(1))\frac{n}{\beta k}(\sqrt{p}-\sqrt{q})^2\right), & k \geq 3. \end{cases}$$

Note that when $q < p = o(1)$, the Rényi divergence of order $\frac{1}{2}$ used in the minimax risk expression in [26] is equal to $(1+o(1))(\sqrt{p}-\sqrt{q})^2$.

3. An Adaptive and Computationally Feasible Procedure. Theorem 1 shows that the minimax rate can be achieved by the estimator (7) obtained via combinatorial optimization which is not computationally feasible. Moreover, the procedure depends on the knowledge of the parameters θ, p and q. These features make it not applicable in practical situations. In this section, we introduce a two-stage algorithm for community detection in DCBMs which is not only computationally feasible but also adaptive over a wide range of unknown parameter values. We show that the procedure achieves minimax optimal rates under certain regularity conditions.

3.1. *A Two-Stage Algorithm.* The proposed algorithm consists of an initialization stage and a refinement stage.

Initialization: weighted k-medians clustering. To explain the rationale behind our proposal, with slight abuse of notation, let $P = (P_{ij}) \in [0,1]^{n\times n}$, where for all $i, j \in [n]$, $P_{ij} = P_{ji} = \theta_i\theta_j B_{z(i)z(j)}$. Except for the diagonal entries, P is the same as in (3). For any $i \in [n]$, let P_i denote the i^{th} row of P. Then for all i such that $z(i) = u$, we observe that

$$\theta_i^{-1}P_i = (\theta_1 B_{u,z(1)}, \ldots, \theta_n B_{u,z(n)})$$

are all equal. Thus, there are exactly k different vectors that the normalized row vectors $\{\theta_i^{-1}P_i\}_{i=1}^n$ can be. Moreover, which one of the k vectors the i^{th} normalized row vector equals is determined solely by its community label $z(i)$. This observation suggests one can design a reasonable community detection procedure by clustering the sample counterparts of the vectors $\{\theta_1^{-1}P_1, \theta_2^{-1}P_2, ..., \theta_n^{-1}P_n\}$, which leads us to the proposal of Algorithm 1.

In Algorithm 1, Steps 1 and 2 aim to find an estimator $\hat{P}$ of P by solving a low rank approximation problem. Then, in Step 3, we can use $\|\hat{P}_i\|_1^{-1}\hat{P}_i$ as a surrogate for $\theta_i^{-1}P_i$. Finally, Step 4 performs a weighted k-median clustering procedure applied on the row vectors of the $n \times k$ matrix $\begin{bmatrix} \|\hat{P}_1\|_1^{-1}\hat{P}_1 \\ \cdots \\ \|\hat{P}_n\|_1^{-1}\hat{P}_n \end{bmatrix}$.

The main novelty of the proposed Algorithm 1 lies in the first two steps. To improve the effect of denoising in the sparse regime, Step 1 removes the rows and the columns of A whose sums are too large. This idea was previously used in community detection in SBMs [7]. If one omits this step, the high probability error bound for the output of Algorithm 1 could suffer an extra multiplier of order $O(\log n)$. The choice of τ will be made clear in Lemma 1 and Remark 4 below. Note that the potential loss of information in Step 1 for those highly important nodes will be later recovered in the refinement state. The $\hat{P}$ matrix sought in Step 2 can be obtained by an eigen-decomposition of $T_\tau(A)$. That is, $\hat{P} = \hat{U}\hat{\Lambda}\hat{U}^T$, where $\hat{U} \in \mathbb{R}^{n\times k}$ collects the k leading eigenvectors, and $\hat{\Lambda}$ is a diagonal matrix of top k eigenvalues. A notable difference between Algorithm 1 and many existing spectral clustering algorithms (e.g., [24, 18, 16]) is that we work with the estimated probability matrix $\hat{P}$ directly rather than its leading eigenvectors $\hat{U}$. As we shall see later, such a difference allows us to avoid eigen-gap assumption required for performance guarantees in the aforementioned papers. Using weighted k-median in step 4 is mainly for technical reasons, as it allows us to establish the same error bound under weaker conditions. In a recent paper [6], a weighted k-medians algorithm was also used in community detection in DCBMs. A key difference is that we apply it on the matrix $\hat{P}$, while [6] applied it on an estimator of the membership matrix $(\mathbf{1}_{\{z(i)=z(j)\}}) \in \{0,1\}^{n\times n}$ obtained from a convex program.

Algorithm 1: Weighted k-medians Clustering

Data: Adjacency matrix $A \in \{0,1\}^{n\times n}$, number of clusters k, tuning parameter τ.
Result: Initial label estimator $\hat{z}^0$.

1 Define $T_\tau(A) \in \{0,1\}^{n\times n}$ by replacing the ith row and column of A whose row sum is larger than τ by zeroes for each $i \in [n]$;

2 Solve

$$\hat{P} = \underset{\mathrm{rank}(P)\le k}{\mathrm{argmin}} \|T_\tau(A) - P\|_{\mathrm{F}}^2;$$

3 Let $\hat{P}_i$ be the i^{th} row of $\hat{P}$. Define $S_0 = \{i \in [n] : \|\hat{P}_i\|_1 = 0\}$. Set $\hat{z}^0(i) = 0$ for $i \in S_0$, and define $\tilde{P}_i = \hat{P}_i/\|\hat{P}_i\|_1$ for $i \notin S_0$;

4 Solve a $(1+\epsilon)$-k-median optimization problem on S_0^c. That is, find $\{\hat{z}^0(i)\}_{i\in S_0^c}$ in $[k]^{|S_0^c|}$ that satisfies

(8)
$$\sum_{u=1}^{k} \min_{v_u\in\mathbb{R}^n} \sum_{\{i\in S_0^c:\hat{z}^0(i)=u\}} \|\hat{P}_i\|_1\|\tilde{P}_i - v_u\|_1 \le (1+\epsilon) \min_{z\in[k]^n} \sum_{u=1}^{k} \min_{v_u\in\mathbb{R}^n} \sum_{\{i\in S_0^c:z(i)=u\}} \|\hat{P}_i\|_1\|\tilde{P}_i - v_u\|_1.$$

Refinement: normalized network neighbor counts. As we shall show later, the error rate of Algorithm 1 decays polynomially with respect to the key

Algorithm 2: A Prototypical Refinement Procedure

Data: Adjacency matrix $A \in \{0,1\}^{n\times n}$, number of clusters k and a community label vector $\hat{z}^0$;

Result: A refined community label vector $\hat{z} \in [k]^n$;

1 For each $i \in [n]$, let

$$\hat{z}(i) = \operatorname*{argmax}_{u\in[k]} \frac{1}{|\{j : \hat{z}^0(j) = u\}|} \sum_{\{j:\hat{z}^0(j)=u\}} A_{ij}. \tag{9}$$

quantity I defined in (6). To achieve the desired exponential decay rate with respect to I as in the minimax rate, we need to further refine the community assignments obtained from Algorithm 1.

To this end, we propose a prototypical refinement procedure in Algorithm 2. The algorithm determines a possibly new community label for the i^{th} node by counting the number of neighbors that the i^{th} node has in each community normalized by the corresponding community size, and then picking the label of the community that maximizes the normalized counts. If there is a tie, we break it in an arbitrary way.

To see the rationale behind Algorithm 2, let us consider a simplified version of the problem. Suppose $k = 2$, $n = 2m + 1$ for some integer $m \geq 1$, $B_{11} = B_{22} = p$ and $B_{12} = B_{21} = q$. Moreover, let us assume that the community labels of the first $2m$ nodes are such that $z(i) = 1$ for $i = 1, \ldots, m$ and $z(i) = 2$ for $i = m+1, \ldots, 2m$. The label of the last node $z(n)$ remains to be determined from the data. When $\{z(i) : i = 1, \ldots, 2m\}$ are the truth, the determination of the label for the n^{th} node reduces to the following testing problem:

$$\begin{aligned} H_0 &: \{A_{n,i}\}_{i\in[n-1]} \sim \bigotimes_{i=1}^{m} \mathrm{Bern}(\theta_n\theta_i p) \otimes \bigotimes_{i=m+1}^{2m} \mathrm{Bern}(\theta_n\theta_i q), \quad \text{vs.} \\ H_1 &: \{A_{n,i}\}_{i\in[n-1]} \sim \bigotimes_{i=1}^{m} \mathrm{Bern}(\theta_n\theta_i q) \otimes \bigotimes_{i=m+1}^{2m} \mathrm{Bern}(\theta_n\theta_i p). \end{aligned} \tag{10}$$

The hypotheses H_0 and H_1 are joint distributions of $\{A_{n,i}\}_{i\in[n-1]}$ in the two cases $z(n) = 1$ and $z(n) = 2$, respectively. For this simple vs. simple testing problem, the Neyman–Pearson lemma dictates that the likelihood ratio test is optimal. However, it is not a satisfying answer for our goal, since the likelihood ratio test needs to use the values of the unknown parameters p, q and θ. While it is possible to obtain sufficiently accurate estimators for p and q, it is hard to do so for θ, especially when the network is sparse. In summary, the dependence of the likelihood ratio test on nuisance parameters makes it

impossible to apply in practice. To overcome this difficulty, we propose to consider a simple test which

$$\text{rejects } H_0 \text{ if } \sum_{i:z(i)=1} A_{n,i} < \sum_{i:z(i)=2} A_{n,i}. \tag{11}$$

As we shall show later in Lemma 2 and (21), this simple procedure achieves the optimal testing error exponent. It is worthwhile to point out that it does not require any knowledge of p, q or θ, and hence the procedure is adaptive. A detailed study of the testing problem (10) is given in Section 5.1.

Inspired by the foregoing discussion, when $k = 2$ and the two community sizes are different, we propose to normalize the counts in (11) by the community sizes. Moreover, when there are more than two communities, we propose to perform pairwise comparison based on the foregoing (normalized) test statistic for each pair of community labels, which becomes the procedure in (9) as long as we replace the unknown truth z with an initial estimator $\hat{z}^0$. For a good initial estimator such as the one output by Algorithm 1, the refinement can lead to minimax optimal errors in misclassification proportion for a large collection of parameter spaces.

In the disassortative case, i.e., when $\min_{u\neq v} B_{uv} \geq q > p \geq \max_u B_{uu}$, we may keep using Algorithm 1 while replacing the definition of $\hat{z}(i)$ in Step 1 of Algorithm 2 with $\hat{z}(i) = \operatorname{argmin}_{u\in[k]} \frac{\sum_{j:\hat{z}^0(j)=u} A_{ij}}{|\{j:\hat{z}^0(j)=u\}|}$. We expect the analysis in the next subsection to go through in the disassortative case with the foregoing modification.

3.2. *Performance Guarantees.* In this part, we state high probability performance guarantees for the proposed procedure. The theoretical property of the algorithms requires an extra bound on the maximal entry of $\mathbb{E}A$. We incorporate this condition into the following parameter space

$$\begin{aligned}&\mathcal{P}_n'(\theta, p, q, k, \beta; \delta, \alpha)\\ &= \big\{P = (\theta_i\theta_j B_{z(i)z(j)}\mathbf{1}_{\{i\neq j\}}) \in \mathcal{P}_n(\theta, p, q, k, \beta; \delta) : \max_{u\in[k]} B_{uu} \leq \alpha p\big\}.\end{aligned}$$

The parameter α is assumed to be a constant no smaller than 1 that does not change with n. By studying the proofs of Theorem 2 and Theorem 1, the minimax lower and upper bounds do not change for the slightly smaller parameter space $\mathcal{P}_n'(\theta, p, q, k, \beta; \delta, \alpha)$. Therefore, the rate $\exp(-(1+o(1))I)$ still serves as a benchmark for us to develop theoretically justifiable algorithms for the parameter space $\mathcal{P}_n'(\theta, p, q, k, \beta; \delta, \alpha)$.

Error rate for the initialization stage. As a first step, we provide the following high probability error bound for Algorithm 1.

LEMMA 1 (Error Bound for Algorithm 1). *Assume $\delta = o(1)$, $1 < p/q = O(1)$ and $\|\theta\|_\infty = o(n/k)$. Let $\tau = C_1(np\|\theta\|_\infty^2 + 1)$ for some sufficiently large constant $C_1 > 0$ in Algorithm 1. Then, there exist some constants $C', C > 0$, such that for any generative model in $\mathcal{P}_n'(\theta, p, q, k, \beta; \delta, \alpha)$, we have with probability at least $1 - n^{-(1+C')}$,*

$$\min_{\pi \in \Pi_k} \sum_{\{i:\hat{z}(i) \neq \pi(z(i))\}} \theta_i \leq C \frac{(1+\epsilon)k^{5/2}\sqrt{n\|\theta\|_\infty^2 p + 1}}{p - q}.$$

Lemma 1 provides a uniform high probability bound for the sum of θ_i's of the nodes which are assigned wrong labels. Before discussing the implication of this result, we give two remarks.

REMARK 3. Algorithm 1 applies a weighted k-medians procedure on the matrix $\hat{P}$ instead of its leading eigenvectors. This is the main difference between Algorithm 1 and many traditional spectral clustering algorithms. As a result, we avoid any eigengap assumption that is imposed to prove consistency results for spectral clustering algorithms [25, 24, 18, 16].

REMARK 4. Lemma 1 suggests that the thresholding parameter τ in Algorithm 1 should be set at the order of $np\|\theta\|_\infty^2 + 1$. Under the extra assumption $\frac{\max_{i \neq j} \mathbb{E}A_{ij}}{\min_{i \neq j} \mathbb{E}A_{ij}} = O(1)$, we can use a data-driven version $\tau = C_1 \frac{1}{n} \sum_{i \neq j} A_{ij}$ for some large constant $C_1 > 0$. The result of Lemma 1 stays the same.

REMARK 5. The extra $(1 + \epsilon)$ slack that we allow in Algorithm 1 is also reflected in the error bound.

The following corollary exemplifies how the result of Lemma 1 can be specialized into a high probability bound for the loss function (4) with a rate depending on I under some stronger conditions. These conditions, especially $\min_i \theta_i = \Omega(1)$, can be relaxed in Theorem 4 stated in the next paragraph.

COROLLARY 2. *Under the conditions of Lemma 1, if we further assume $p \geq n^{-1}$, $k = O(1)$ and $\Omega(1) = \min_i \theta_i \leq \|\theta\|_\infty = O(1)$, then there exist some constants $C', C > 0$, such that for any generative model in $\mathcal{P}_n'(\theta, p, q, k, \beta; \delta, \alpha)$, we have $\ell(\hat{z}, z) \leq C(1+\epsilon)I^{-1/2}$ with probability at least $1 - n^{-(1+C')}$.*

Error rate for the refinement stage. As exemplified in Corollary 2, the convergence rate for the initialization step is typically only polynomial in I as opposed to the exponential rate in the minimax rate. Thus, there is room for improvement. In what follows, we show that a specific way of applying Algorithm 2 on the output of Algorithm 1 leads to significant performance enhancement in terms of misclassification proportion. To this end, let us first state in Algorithm 3 the combined algorithm for which we are able to establish the improved error bounds. Here and after, for any $i \in [n]$, let

$A_{-i} \in \{0,1\}^{(n-1)\times(n-1)}$ be the submatrix of A obtained from removing the i^{th} row and column of A.

Algorithm 3: A Provable Version of Algorithm 1 + Algorithm 2

Data: Adjacency matrix $A \in \{0,1\}^{n\times n}$ and number of clusters k;
Result: Clustering label estimator $\hat{z} \in [k]^n$;

1 For each $i \in [n]$, apply Algorithm 1 to A_{-i}. The result, which is a vector of dimension $n-1$, is stored in $(\hat{z}^0_{-i}(1), ..., \hat{z}^0_{-i}(i-1), \hat{z}^0_{-i}(i+1), ..., \hat{z}^0_{-i}(n))$;

2 For each $i \in [n]$, the ith entry of $\hat{z}^0_{-i}$ is set as

$$\hat{z}^0_{-i}(i) = \underset{u\in[k]}{\operatorname{argmax}} \frac{1}{|\{j : \hat{z}^0_{-i}(j) = u\}|} \sum_{j:\hat{z}^0_{-i}(j)=u} A_{ij};$$

3 Set $\hat{z}(1) = \hat{z}^0_{-1}(1)$. For each $i \in \{2, ..., n\}$, set

$$\hat{z}(i) = \underset{u\in[k]}{\operatorname{argmax}} |\{j : \hat{z}^0_{-1}(j) = u\} \cap \{j : \hat{z}^0_{-i}(j) = \hat{z}^0_{-i}(i)\}|. \tag{12}$$

REMARK 6. The last step (12) of Algorithm 3 constructs a final label estimator $\hat{z}$ from $\hat{z}^0_{-1}, \hat{z}^0_{-2}, ..., \hat{z}^0_{-n}$. Since the labels given by $\hat{z}^0_{-1}, \hat{z}^0_{-2}, ..., \hat{z}^0_{-n}$ are only comparable after certain permutations in Π_k, we need this extra step to resolve this issue.

REMARK 7. Algorithm 3 is a theoretically justifiable version for combining Algorithm 1 and Algorithm 2. In order to obtain a rate-optimal label assignment for the i^{th} node, we first apply the initial clustering procedure in Algorithm 1 on the sub-network consisting of the remaining $n-1$ nodes and the edges among them. Then, one applies the refinement procedure in Algorithm 2 to assign a label for the i^{th} node. The independence between the initialization and the refinement stages facilitates the technical arguments in the proof. However, in practice, one can simply apply Algorithm 1 followed by Algorithm 2. The numerical difference from Algorithm 3 is negligible in all the data examples we have examined.

A special case: almost equal community sizes. In the special case where the community sizes are almost equal, we can show that $\hat{z}$ output by Algorithm 3 achieves the minimax rate.

THEOREM 3. *Under the conditions of Lemma 1, we further assume* $\beta = 1$, $k = O(1)$, $\min_i \theta_i = \Omega(1)$, $\delta = o(\frac{p-q}{p})$, $\|\theta\|^2_\infty p \geq n^{-1}$, *and* $\frac{(1+\epsilon)\|\theta\|_\infty p^{3/2}}{\sqrt{n}(p-q)^2} = o(1)$. *Then there is a sequence* $\eta = o(1)$ *such that the output* $\hat{z}$ *of Algorithm 3 satisfies*

$$\lim_{n\to\infty} \inf_{\mathcal{P}'_n(\theta,p,q,k,\beta;\delta,\alpha)} \mathbb{P}\left\{\ell(\hat{z}, z) \leq \exp\big(-(1-\eta)I\big)\right\} = 1.$$

Theorem 3 shows that when the community sizes are almost equal, the minimax rate $\exp(-(1+o(1))I)$ can be achieved within polynomial time. We note that the conditions that we need here are stronger than those of Theorem 1. When $k = O(1)$, $\epsilon = O(1)$ and $\Omega(1) = \min_i \theta_i \le \|\theta\|_\infty = O(1)$, Theorem 1 only requires $I \to \infty$, which is equivalent to $\frac{p^{1/2}}{\sqrt{n}(p-q)} = o(1)$, while Theorem 3 requires $\frac{p^{3/2}}{\sqrt{n}(p-q)^2} = o(1)$. Whether the extra factor $\frac{p}{p-q}$ can be removed from the assumptions is an interesting problem to investigate in the future. However, as long as $p \asymp p - q$, then all the other conditions in Theorem 1 and Theorem 3 match and the algorithm achieves the desired rate even when $\frac{1}{n} \ll p \ll \frac{\log n}{n}$.

General case. We now state a general high probability error bound for Algorithm 3. To introduce this result, we define another information-theoretic quantity. For any $t \in (0,1)$, define

$$J_t(p,q) = 2\left(tp + (1-t)q - p^t q^{1-t}\right). \tag{13}$$

By Jensen's inequality, it is straightforward to verify that $J_t(p,q) \ge 0$ and $J_t(p,q) = 0$ if and only if $p = q$. As a special case, when $t = \frac{1}{2}$, we have

$$J_{\frac{1}{2}}(p,q) = (\sqrt{p} - \sqrt{q})^2. \tag{14}$$

For a given $z \in [k]^n$, let $n_{(1)} \le ... \le n_{(k)}$ be the order statistics of community sizes $\{n_u(z) : u = 1, \dots, k\}$. Then, we define the quantity J by through

$$\exp(-J) = \frac{1}{n}\sum_{i=1}^{n} \exp\left(-\theta_i \left(\frac{n_{(1)} + n_{(2)}}{2}\right) J_{t^*}(p,q)\right) \tag{15}$$

with $t^* = \frac{n_{(1)}}{n_{(1)}+n_{(2)}}$. With the foregoing definitions, the following theorem gives a general error bound for Algorithm 3.

THEOREM 4. *Under the conditions of Lemma 1, we further assume that* $\delta = o(\frac{p-q}{p})$, $\|\theta\|_\infty^2 p \ge n^{-1}$,

$$\frac{(1+\epsilon)k^{5/2}\|\theta\|_\infty\sqrt{p}}{\sqrt{n}(p-q)} = o\left(\frac{p-q}{kp}\right), \quad \textit{and} \tag{16}$$

$$\min_{\gamma \ge 0}\left\{n^{-1}|\{i \in [n] : \theta_i \le \gamma\}| + \frac{(1+\epsilon)k^{5/2}\|\theta\|_\infty\sqrt{p}}{\gamma\sqrt{n}(p-q)}\right\} = o\left(\frac{p-q}{k^2 p}\right). \tag{17}$$

Then there is a sequence $\eta = o(1)$ *such that the output* $\hat{z}$ *of Algorithm 3 satisfies*

$$\lim_{n\to\infty} \inf_{\mathcal{P}_n'(\theta,p,q,k,\beta;\delta,\alpha)} \mathbb{P}\left\{\ell(\hat{z},z) \le \exp\left(-(1-\eta)J\right)\right\} = 1.$$

Theorem 4 gives a general error bound for the performance of Algorithm 3. It is easy to check that the conditions (16) and (17) are satisfied under the settings of Theorem 3. Therefore, Theorem 3 is a special case of Theorem 4. Theorem 4 shows that Algorithm 3 converges at the rate $\exp(-(1+o(1))J)$. According to the properties of $J_t(p,q)$ stated in Appendix C, one can show that when $n_{(1)} = (1+o(1))n_{(2)}$, $J = (1+o(1))I$, and that in general

$$n_{(1)}(\sqrt{p}-\sqrt{q})^2 \le \left(\frac{n_{(1)}+n_{(2)}}{2}\right) J_{t^*}(p,q) \le \left(\frac{n_{(1)}+n_{(2)}}{2}\right)(\sqrt{p}-\sqrt{q})^2.$$

Using this relation, we can state the convergence rate in Theorem 4 using the quantity I.

COROLLARY 3. *Under the conditions of Theorem 4, there is a sequence $\eta = o(1)$ such that the output $\hat{z}$ of Algorithm 3 satisfies*

$$\lim_{n\to\infty} \inf_{\mathcal{P}'_n(\theta,p,q,2,\beta;\delta,\alpha)} \mathbb{P}\left\{\ell(\hat{z},z) \le \exp\big(-(1-\eta)\beta^{-1}I\big)\right\} = 1,$$
$$\lim_{n\to\infty} \inf_{\mathcal{P}'_n(\theta,p,q,k,\beta;\delta,\alpha)} \mathbb{P}\left\{\ell(\hat{z},z) \le \exp\big(-(1-\eta)I\big)\right\} = 1, \text{ for } k \ge 3.$$

Therefore, when $k \ge 3$, the minimax rate $\exp(-(1+o(1))I)$ is achieved by Algorithm 3. The only situation where the minimax rate is not achieved by Algorithm 3 is when $k = 2$ and $\beta > 1$. For this case, there is an extra β^{-1} factor on the exponent of the convergence rate.

REMARK 8. If we further assume that $\min_{i\neq j} B_{ij} = \Omega(q)$, a careful examination of the proofs shows that we can improve the term $k^{5/2}$ in the conclusion of Lemma 1 and in the conditions (16) and (17) to $k^{3/2}$. Since it is unclear what the optimal power exponent for k is in these circumstances, we do not pursue it explicitly in this paper.

4. Numerical Results. In this section, we present numerical experiments on simulated datasets generated from DCBMs. In particular, we compare the performance of two versions of our algorithm with three state-of-the-art methods: SCORE [16], CMM [6] and RSC (Regularized Spectral Clustering) [24] in two different scenarios. On simulated data examples, both versions of our algorithm outperformed SCORE and RSC in terms of misclassification proportion. The performance of CMM was comparable to our algorithm. However, our algorithm not only gives slightly better accuracy on the simulated datasets, but it also enjoys the advantage of easy implementation, fast computation and scalability to networks of large sizes since it does not involve convex programming. We also apply our method to a real data set [3]. Our algorithms perform comparably with the best results in literature.

Scenario 1. We set $n = 300$ nodes and $k = 2$. The sizes of the two communities are set as 100 and 200, respectively. The off-diagonal entries of the adjacency matrix were generated as $A_{ij} = A_{ji} \overset{ind.}{\sim} \text{Bern}(\theta_i\theta_j p)$ if $z(i) = z(j)$ and $A_{ij} = A_{ji} \overset{ind.}{\sim} \text{Bern}(\theta_i\theta_j q)$ if $z(i) \neq z(j)$. We let $p = 0.1$ and $q = 3p/10$. The degree-correction parameters are set as $\theta_i = |Z_i| + 1 - (2\pi)^{-1/2}$ where $Z_i \overset{iid}{\sim} N(0, 0.25)$ for $i = 1, \ldots, n$. It is straightforward to verify that $\mathbb{E}\theta_i = 1$.

We compare misclassification proportions of the following six algorithms[1]:

1. The SCORE method in [16];
2. The weighted k-medians procedure in Algorithm 1;
3. Refinement of the output of Algorithm 1 by Algorithm 2;
4. Iterate Algorithm 2 10 times after initialization by Algorithm 1.
5. The CMM method in [6];
6. The RSC method in [24].

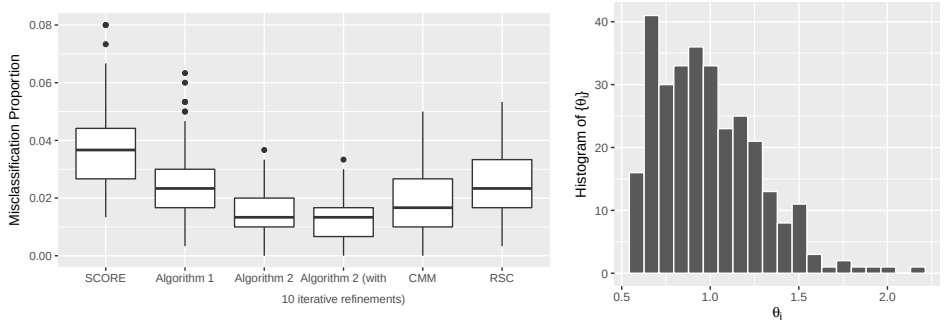

FIG 1. *Left panel: boxplots of misclassification proportions for the five algorithms over* 100 *independent repetitions. Right panel: histogram of* θ_i.

We conducted the experiments with 100 independent repetitions and summarize the performances of the six algorithms through boxplots of misclassification proportions. Fig. 1 shows that our refinement step (Algorithm 2) significantly improved the performance of the initialization step (Algorithm 1). Moreover, it helps to further reduce the error if we apply the refinement step for a few more iterations. Among the six algorithms, our proposed procedures give the best performance. The CMM algorithm is slightly worse than our procedures with refinement, but is better than Algorithm 1, RSC and SCORE. See Appendix A.1 for detailed comparison of running times.

Scenario 2. Here, we set $n = 800$ and $k = 4$ and all community sizes are set equal. The adjacency matrix is generated in the same way as in Scenario

[1]The numerical performance of Algorithm 3 was indistinguishable from that of the second algorithm in the list in all the experiments conducted.

1 except that θ_i's are independent draws from a Pareto distribution with density function $f(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}} 1_{\{x \geq \beta\}}$, with $\alpha = 5$ and $\beta = 4/5$. The choice of α and β ensures that $\mathbb{E}\theta_i = 1$.

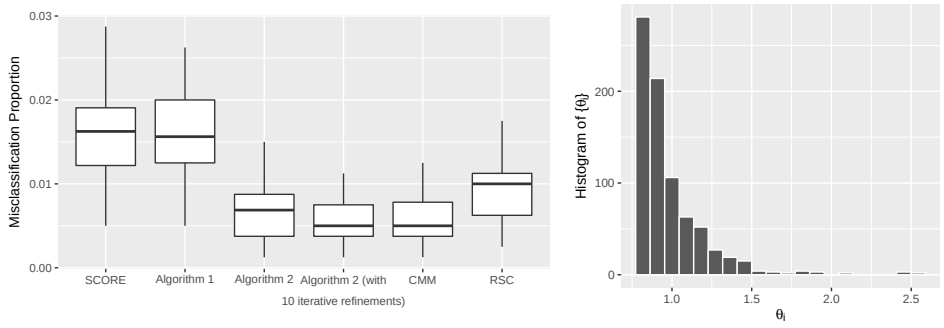

FIG 2. *Left panel: boxplots of misclassification proportions for the five algorithms over* 100 *independent repetitions. Right panel: histogram of* θ_i.

As in Scenario 1, we compare the performance of the six algorithms over 100 independent repetitions. Fig. 2 shows the boxplots of the misclassification proportions. The overall message is similar to Scenario 1, except that CMM is almost as good as our procedures with refinement, but all of them outperform Algorithm 1, RSC and SCORE. Despite its comparable performance on this task, CMM requires noticeably longer running time than our procedures on the simulated datasets due to the involvement of convex programming. Therefore, its scalability to large networks is more limited. See Appendix A.1 for detailed comparison of speed. For the implementations of Algorithm 1 in both scenarios, we let $\tau = c(\sum_{i,j} A_{i,j})/n$ with $c = 20$. However the choice of τ has minimal effect on the performance of Algorithm 1 as long as $c \geq 5$. See Appendix A.2.

The Political Blog Dataset. We compare our algorithm with the others on the political blog dataset [3]. It presents a network of connections among 1490 political blogs. After the pre-processing step as in [17], we consider 1222 nodes in the largest component. The network consists two communities, the liberal one with 586 blogs and the conservative one with 636 blogs. It is undirected and with no self-loop. The SCORE method has the best known error rate (58 nodes) in the literature. Our algorithms give comparable (59 nodes for Algorithm 2) and even slightly better results (57 nodes for Algorithm 2 with 10 iterations).

5. Proofs. This section presents proofs for some main results of the paper. In Section 5.1, we first study a fundamental testing problem for community detection in DCBMs. The theoretical results for this testing problem

TABLE 1
Comparison of Algorithms on the Political Blog Dataset

Method	SCORE	Algo. 1	Algo. 2	Algo. 2 (10 iter.)	CMM	RSC
# errors	58	65	59	57	61	63

are critical tools to prove minimax lower and upper bounds for community detection. We then state the proof of Theorem 1. Proofs of the other main results are deferred to the appendices.

5.1. *A Fundamental Testing Problem.* As a prelude to all proofs, we consider the hypothesis testing problem (10) that not only is fundamental to the study of minimax risk but also motivates the proposal of Algorithm 2. To paraphrase the problem, suppose $X = (X_1, \ldots, X_m, X_{m+1}, \ldots, X_{2m})$ have independent Bernoulli entries. Given $1 \geq p > q \geq 0$ and $\theta_0, \theta_1, \ldots, \theta_{2m} > 0$ such that

$$\sum_{i=1}^{m} \theta_i = \sum_{i=m+1}^{2m} \theta_i = m. \tag{18}$$

We are interested in understanding the minimum possible Type I+II error of testing

$$\begin{aligned} H_0 : X \sim \bigotimes_{i=1}^{m} \mathrm{Bern}\left(\theta_0\theta_i p\right) \otimes \bigotimes_{i=m+1}^{2m} \mathrm{Bern}\left(\theta_0\theta_i q\right) \\ \text{vs.} \quad H_1 : X \sim \bigotimes_{i=1}^{m} \mathrm{Bern}\left(\theta_0\theta_i q\right) \otimes \bigotimes_{i=m+1}^{2m} \mathrm{Bern}\left(\theta_0\theta_i p\right). \end{aligned} \tag{19}$$

In particular, we are interested in the asymptotic behavior of the error for a sequence of such testing problems in which p, q and the θ_i's scale with m as $m \to \infty$. First, we have the following lower bound result.

LEMMA 2. *Suppose that as* $m \to \infty$, $1 < p/q = O(1)$ *and* $p \max_{0 \leq i \leq 2m} \theta_i^2 = o(1)$. *If* $\theta_0 m(\sqrt{p} - \sqrt{q})^2 \to \infty$,

$$\inf_{\phi} \left(P_{H_0}\phi + P_{H_1}(1 - \phi)\right) \geq \exp\left(-(1 + o(1))\theta_0 m(\sqrt{p} - \sqrt{q})^2\right).$$

Otherwise if $\theta_0 m(\sqrt{p} - \sqrt{q})^2 = O(1)$, *there exists a constant* $c \in (0, 1)$ *such that* $\inf_{\phi} \left(P_{H_0}\phi + P_{H_1}(1 - \phi)\right) \geq c$.

According to the Neyman–Pearson lemma, the optimal testing procedure is the likelihood ratio test. However, such a test depends on the values of

$\{\theta_i\}_{i=1}^{2m}, p, q$, and is not appropriate in practice. On the other hand, the following simple test

$$\phi = \mathbf{1}_{\{\sum_{i=1}^{m} X_i < \sum_{i=m+1}^{2m} X_i\}}. \tag{20}$$

can be shown to achieve the optimal error bound:

$$P_{H_0}\phi + P_{H_1}(1-\phi) \le 2\exp\left(-\theta_0 m(\sqrt{p}-\sqrt{q})^2\right). \tag{21}$$

Combining Lemma 2 and (21), we find that the minimax testing error for the problem (19) is $e^{-(1+o(1))\theta_0 m(\sqrt{p}-\sqrt{q})^2}$. This explains why the minimax rate for community detection in DCBM takes the form of $e^{-(1+o(1))I}$ in Corollary 1. Moreover, the simple testing function (20) serves as a critical component in Algorithm 2. The fact that (20) can achieve the optimal testing error exponent in (21) explains why our algorithm can achieve the minimax rate when the community sizes are equal (Theorem 3).

5.2. *Proof of Theorem 1.* Throughout the proof, we let z denote the truth, $\hat{z}$ the estimator defined in (7) and $\tilde{z}$ a generic assignment vector. In addition, we let L denote the objective function in (7). In what follows, we focus on proving the upper bounds for $k \ge 3$ while the case of $k = 2$ is deferred to Appendix B.1. The proof for $k = 2$ is slightly different because in this case, when a node is mis-clustered, one is able to identify the exact wrong label the node is assigned which is not possible when $k \ge 3$.

Outline and additional notation. We have the following basic equality

$$\mathbb{E}n\ell(\hat{z}, z) = \sum_{m=1}^{n} m\mathbb{P}(n\ell(\hat{z}, z) = m). \tag{22}$$

Thus, to prove the desired upper bounds, we are to work out appropriate bounds for the individual probabilities $\mathbb{P}(n\ell(\hat{z}, z) = m)$. To this end, for any given m, our basic idea is to first bound $\mathbb{P}(L(\tilde{z}) > L(z))$ for any $\tilde{z}$ such that $n\ell(\hat{z}, z) = m$ and then apply the union bound. To carry out these calculations in details, we divide the entire proof into three major steps: (i) In the first step, we derive a generic upper bound expression for the quantity $\mathbb{P}(L(\tilde{z}) > L(z))$ for any deterministic $\tilde{z}$.

(ii) In the second step, we further materialize the upper bound expression according to different values of m where $m = n\ell(\tilde{z}, z)$. In particular, we shall use different arguments in three different regimes of m values. Together with the union bound, we shall obtain bounds for all probabilities $\mathbb{P}(n\ell(\hat{z}, z) = m)$.

(iii) In the last step, we supply the bounds obtained in the second step to (22) to establish the theorem. Indeed, the bounds we derive in the second

step decay geometrically once m is larger than some critical value which depends on the rate of the error bounds. Thus, we divide the final arguments here according to three different regimes of error rates.

After a brief introduction to some additional notation, we carry out these three steps in order. We denote $n_{\min} = \min_{u\in[k]} |\{i : z(i) = u\}|$, $n_{\max} = \max_{u\in[k]} |\{i : z(i) = u\}|$ and $\theta_{\min} = \min_{i\in[n]} \theta_i$. Note that $n_{\min} \geq n/(\beta k)$, $n_{\max} \leq \beta n/k$ and $\theta_{\min} = \Omega(1)$. For any $t \in (0, 1)$, we define

$$R_t = \frac{1}{n}\sum_{i=1}^{n} \exp\left(-(1-t)\theta_i n_{\min}(\sqrt{p}-\sqrt{q})^2\right). \tag{23}$$

In order to show $\mathbb{E}\ell(\hat{z}, z) \leq \exp(-(1 - o(1))I)$, it is sufficient to prove $\mathbb{E}\ell(\hat{z}, z) \leq R_t$ for some $t = o(1)$, since $R_t \leq \frac{1}{n}\sum_{i=1}^{n} \exp\left(-(1-t)\theta_i n/(\beta k)(\sqrt{p}-\sqrt{q})^2\right) \leq \exp(-(1-t)I)$, where the second inequality is by Jensen inequality.

Step 1: bounding $\mathbb{P}(L(\tilde{z}) > L(z))$. For any deterministic $\tilde{z}$, we have

$$\mathbb{P}(L(\tilde{z}) > L(z)) = \mathbb{P}\Bigg(\sum_{\substack{i<j, z(i)=z(j)\\ \tilde{z}(i)\neq\tilde{z}(j)}} \Big(A_{ij}\log\frac{q(1-\theta_i\theta_j p)}{p(1-\theta_i\theta_j q)} + \log\frac{1-\theta_i\theta_j q}{1-\theta_i\theta_j p}\Big) + \sum_{\substack{i<j, z(i)\neq z(j)\\ \tilde{z}(i)=\tilde{z}(j)}} \Big(A_{ij}\log\frac{p(1-\theta_i\theta_j q)}{q(1-\theta_i\theta_j p)} + \log\frac{1-\theta_i\theta_j p}{1-\theta_i\theta_j q}\Big) > 0\Bigg).$$

When $z(i) = z(j)$, we have $\mathbb{P}(A_{ij} = 1) = \theta_i\theta_j p'$ for some $p' \geq p$, and so

$$\begin{aligned}
&\mathbb{E}\exp\left(\frac{1}{2}\Big(A_{ij}\log\frac{q(1-\theta_i\theta_j p)}{p(1-\theta_i\theta_j q)} + \log\frac{1-\theta_i\theta_j q}{1-\theta_i\theta_j p}\Big)\right)\\
&= \theta_i\theta_j\sqrt{\frac{q}{p}}p' + \sqrt{\frac{1-\theta_i\theta_j q}{1-\theta_i\theta_j p}}(1-\theta_i\theta_j p')\\
&= \theta_i\theta_j\sqrt{qp} + \sqrt{(1-\theta_i\theta_j q)(1-\theta_i\theta_j p)} + \theta_i\theta_j(p'-p)\left(\sqrt{\frac{q}{p}} - \sqrt{\frac{1-\theta_i\theta_j q}{1-\theta_i\theta_j p}}\right)\\
&\leq \exp\Big(\log\Big(\sqrt{pq}\theta_i\theta_j + \sqrt{1-\theta_i\theta_j q}\sqrt{1-\theta_i\theta_j p}\Big)\Big) \leq \exp(-\frac{1}{2}\theta_i\theta_j(\sqrt{q}-\sqrt{p})^2).
\end{aligned}$$

Similarly, when $z(i) \neq z(j)$ we also have

$$\begin{aligned}
&\mathbb{E}\exp\left(\frac{1}{2}\Big(A_{ij}\log\frac{p(1-\theta_i\theta_j q)}{q(1-\theta_i\theta_j p)} + \log\frac{1-\theta_i\theta_j p}{1-\theta_i\theta_j q}\Big)\right)\\
&\leq \exp\Big(\log\Big(\sqrt{pq}\theta_i\theta_j + \sqrt{1-\theta_i\theta_j q}\sqrt{1-\theta_i\theta_j p}\Big)\Big) \leq \exp(-\frac{1}{2}\theta_i\theta_j(\sqrt{q}-\sqrt{p})^2).
\end{aligned}$$

For any assignment vector $\tilde{z}$, Define membership matrix $\tilde{Y} \in \{0,1\}^{n\times n}$ with $\tilde{Y}_{ij} = \mathbf{1}_{\{\tilde{z}(i)=\tilde{z}(j)\}}$ for all $i \neq j$ and zero otherwise. Let Y be the membership matrix associated with the truth z. Note that the membership matrix is invariant under permutation of the community labels. By applying the Chernoff bound with $t = \frac{1}{2}$ we have

$$\begin{aligned}\mathbb{P}\left(L(\tilde{z}) > L(z)\right) &\leq \prod_{\substack{i<j \\ \tilde{Y}_{ij}\neq Y_{ij}}} \exp\left(-\frac{1}{2}\theta_i\theta_j(\sqrt{p}-\sqrt{q})^2\right) \\ &= \prod_{\substack{i\neq j \\ \tilde{Y}_{ij}\neq Y_{ij}}} \exp\left(-\frac{1}{4}\theta_i\theta_j(\sqrt{p}-\sqrt{q})^2\right).\end{aligned} \tag{24}$$

Step 2: bounding $\mathbb{P}(n\ell(\hat{z},z)=m)$. To obtain the desired bounds on these probabilities, we introduce a way to partition each community according to the values of the degree-correction parameters. Given the truth z and any deterministic assignment vector $\tilde{z}$, let $\mathcal{C}_u = \{i \in [n] : z(i) = u\}$, $\Gamma_u = \{i \in [n] : z(i) = u, \tilde{z}(i) \neq u\}$ and $\Gamma = \cup_{u\in[k]}\Gamma_u$. Note that Γ_u and Γ depend on $\tilde{z}$.

Let $M \geq 2$ be a large enough constant integer to be determined later. For each $\mathcal{C}_u$ we decompose it as $\mathcal{C}_u = \mathcal{C}_u^+ \cup \mathcal{C}_u^-$ such that

$$\mathcal{C}_u^+ \cap \mathcal{C}_u^- = \emptyset, \quad |\mathcal{C}_u^-| = \left\lceil \frac{|\mathcal{C}_u|}{M} \right\rceil, \quad \min_{i\in\mathcal{C}_u^+}\theta_i \geq \max_{i\in\mathcal{C}_u^-}\theta_i. \tag{25}$$

Due to the approximate normalization of degree-correction parameters, for sufficiently large values of n,

$$\max_{i\in\mathcal{C}_u^-}\theta_i \leq 3/2. \tag{26}$$

Since $|\mathcal{C}_u^+| \leq (M-1)|\mathcal{C}_u^-|$, we can define a mapping $\tau_u : \mathcal{C}_u \to \mathcal{C}_u^-$ such that its restriction on $\mathcal{C}_u^-$ is identity. Moreover, we could require that for any $i \in \mathcal{C}_u^-$, $|\tau_u^{-1}(i)| \leq M$. Let τ be the mapping from $[n]$ to $\cup_{u=1}^k \mathcal{C}_u^-$ such that the restriction of τ on $\mathcal{C}_u$ is τ_u. The main reason for introducing τ is to deal with the range of values the degree-correction parameters can take. The right side of (24) shows that the desired bounds depend crucially on quantities of the form $\sum_{i\in S}\theta_i$ for some set S. For any set S, the sum $\sum_{i\in S}\theta_i$ is not necessarily upper bounded by a constant multiple of the size of the set $|S|$. However, by (26), we can always upper bound $\sum_{i\in S}\theta_{\tau(i)}$ by a constant multiple of $|S|$. This gives us a way to relate the probability bounds and the number of misclassified nodes. Such a point can be seen more clearly as we go to explicit calculation below.

Let

$$m' = \eta n/k \tag{27}$$

for some $\eta = o(1)$ with $\eta^{-1} = o(I)$ and $k \le n^{\eta}$. We now derive bounds for $\mathbb{P}(n\ell(\hat{z}, z) = m)$ for $m \in [1, M]$, $(M, m']$ and $(m', n]$ separately.

Case 1: $1 \le m \le M$. In this case, we have

$$\begin{aligned}
\mathbb{P}(n\ell(\hat{z}, z) = m) &\le \sum_{\tilde{z}:|\Gamma|=m} \exp\Big(-\frac{1}{2}\sum_{i\in\Gamma}\theta_i\Big((1-\delta)2n_{\min} - \sum_{i\in\Gamma}\theta_i\Big)(\sqrt{p}-\sqrt{q})^2\Big) \\
&\le \sum_{\tilde{z}:|\Gamma|=m} \exp\Big(-\frac{1}{2}\sum_{i\in\Gamma}\theta_{\tau(i)}\Big((1-\delta)2n_{\min} - \sum_{i\in\Gamma}\theta_{\tau(i)}\Big)(\sqrt{p}-\sqrt{q})^2\Big) \\
&\le \sum_{\tilde{z}:|\Gamma|=m}\prod_{i\in\Gamma} \exp\Big(-\frac{1}{2}\theta_{\tau(i)}\left((1-\delta)2n_{\min} - 2M\right)(\sqrt{p}-\sqrt{q})^2\Big).
\end{aligned}$$

Here, the first inequality comes from direct application of (24) and the union bound. Since $\|\theta\|_\infty = o(n/k) = o(n_{\min})$ and M is a constant, we have $\sum_{i\in\Gamma}\theta_i = o(n_{\min})$. This, together with the monotonicity of the function $x(1-x)$ when x is in the right neighborhood of zero, implies the second inequality. The third inequality is due to (26). Since M is a constant and $M/n_{\min}$ can be upper bounded by η for large values of n, we further have

$$\begin{aligned}
\mathbb{P}(n\ell(\hat{z}, z) = m) &\le \sum_{\tilde{z}:|\Gamma|=m}\prod_{i\in\Gamma} \exp\Big(-\theta_{\tau(i)}(1-\delta-2\eta)2n_{\min}(\sqrt{p}-\sqrt{q})^2\Big) \\
&\le k^m\Big(\sum_{i=1}^{n} \exp\Big(-\theta_{\tau(i)}(1-\delta-2\eta)n_{\min}(\sqrt{p}-\sqrt{q})^2\Big)\Big)^m \\
&\le k^m\Big(M\sum_{i=1}^{n} \exp\Big(-\theta_{i}(1-\delta-2\eta)n_{\min}(\sqrt{p}-\sqrt{q})^2\Big)\Big)^m \\
&= (knMR_{\delta+2\eta})^m.
\end{aligned}$$

Here and after, the notation $\sum_{\tilde{z}:|\Gamma|=m}$ means summing over all deterministic assignment vectors $\tilde{z}$ such that $|\Gamma| = n\ell(\tilde{z}, z) = m$. The last inequality holds since for any $i \in \mathcal{C}_u^-$, $|\tau_u^{-1}(i)| \le M$.

Case 2: $M < m \le m'$. In this case, we cannot directly apply the argument in case 1 since we can no longer guarantee that $\sum_{i\in\Gamma}\theta_i = o(n_{\min})$ and so the second inequality of the last display no longer holds. To proceed, we can further bound the rightmost side of (24) by $B_1 \times B_2$, where

$$\begin{aligned}
B_1 &= \prod_{\substack{(i,j):z(i)=z(j)\\ \tilde{z}(i)\neq\tilde{z}(j)}} \exp\left(-\frac{1}{4}\theta_i\theta_j(\sqrt{p}-\sqrt{q})^2\right), \\
B_2 &= \prod_{\substack{(i,j):z(i)\neq z(j)\\ \tilde{z}(i)=\tilde{z}(j)}} \exp\left(-\frac{1}{4}\theta_i\theta_j(\sqrt{p}-\sqrt{q})^2\right).
\end{aligned}$$

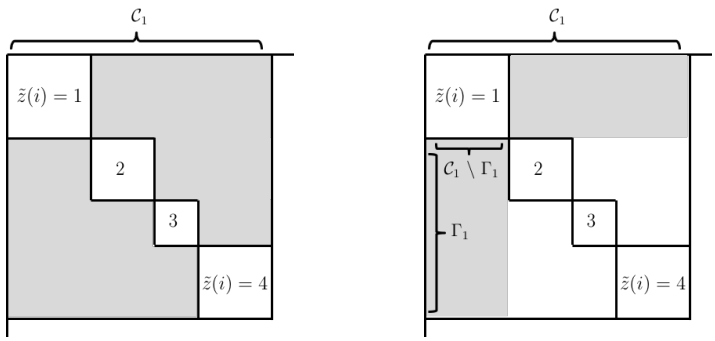

FIG 3. *Illustration of reduction to (29) when $k=4$. Only nodes from the first community are shown, which are rearranged according to $\tilde{z}$. In the left panel the gray regions correspond to terms in (28). In the right panel the gray regions correspond to terms in (29). Note that the area of the gray regions in the left panel is larger than the one in the right panel.*

In what follows, we focus on upper bounding B_1 and the same upper bound holds for B_2 by essentially repeating the arguments. For B_1, we have

$$
\begin{aligned}
B_1 &= \prod_{u=1}^{k}\prod_{u'=1}^{k}\prod_{\substack{\{i:z(i)=u,\\ \tilde{z}(i)=u'\}}}\prod_{\substack{\{j:z(j)=u,\\ \tilde{z}(j)\neq u'\}}} \exp\Big(-\frac{1}{4}\theta_i\theta_j(\sqrt{p}-\sqrt{q})^2\Big) \qquad (28)\\
&= \prod_{u=1}^{k}\prod_{u'\neq u}\exp\Big(-\frac{1}{4}\sum_{\substack{\{i:z(i)=u,\\ \tilde{z}(i)=u'\}}}\theta_i\sum_{\substack{\{j:z(j)=u,\\ \tilde{z}(j)\neq u'\}}}\theta_j(\sqrt{p}-\sqrt{q})^2\Big)\\
&\quad\times\prod_{u=1}^{k}\prod_{u'=u}\exp\Big(-\frac{1}{4}\sum_{\substack{\{i:z(i)=u,\\ \tilde{z}(i)=u'\}}}\theta_i\sum_{\substack{\{j:z(j)=u,\\ \tilde{z}(j)\neq u'\}}}\theta_j(\sqrt{p}-\sqrt{q})^2\Big).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
B_1 &\le \prod_{u=1}^{k}\prod_{u'\neq u}\exp\Big(-\frac{1}{4}\sum_{\substack{\{i:z(i)=u,\\ \tilde{z}(i)=u'\}}}\theta_i\sum_{\substack{\{j:z(j)=u,\\ \tilde{z}(j)=u\}}}\theta_j(\sqrt{p}-\sqrt{q})^2\Big)\\
&\quad\times\prod_{u=1}^{k}\prod_{u'\neq u}\exp\Big(-\frac{1}{4}\sum_{\substack{\{i:z(i)=u,\\ \tilde{z}(i)=u\}}}\theta_i\sum_{\substack{\{j:z(j)=u,\\ \tilde{z}(j)=u'\}}}\theta_j(\sqrt{p}-\sqrt{q})^2\Big). \qquad (29)
\end{aligned}
$$

Fig. 3 illustrates why the inequality in (29) holds. Furthermore, we notice

that

$$(29) = \prod_{u=1}^{k} \prod_{u' \neq u} \exp\Bigg(-\frac{1}{2} \sum_{\substack{\{i:z(i)=u,\\ \tilde{z}(i)=u'\}}} \theta_i \sum_{\substack{\{j:z(j)=u,\\ \tilde{z}(j)=u\}}} \theta_j (\sqrt{p}-\sqrt{q})^2 \Bigg)$$

$$= \prod_{u=1}^{k} \exp\Bigg(-\frac{1}{2} \sum_{i \in \Gamma_u} \theta_i \sum_{j \in \mathcal{C}_u \setminus \Gamma_u} \theta_j (\sqrt{p}-\sqrt{q})^2 \Bigg)$$

$$= \prod_{u=1}^{k} \exp\Bigg(-\frac{1}{2} \sum_{i \in \Gamma_u} \theta_i \Bigg(\sum_{i \in \mathcal{C}_u} \theta_i - \sum_{i \in \Gamma_u} \theta_i \Bigg) (\sqrt{p}-\sqrt{q})^2 \Bigg). \tag{30}$$

To further bound the right side of (30), recall that $\theta_{\min} = \min_i \theta_i = \Omega(1)$. Then

$$\sum_{i \in \mathcal{C}_u} \theta_i - \sum_{i \in \Gamma_u} \theta_i = \sum_{\mathcal{C}_u \setminus \Gamma_u} \theta_i \geq (|\mathcal{C}_u| - |\Gamma_u|)\theta_{\min} \geq \frac{|\mathcal{C}_u|\theta_{\min}}{2}.$$

Here the last inequality holds since $|\Gamma_u| \leq |\Gamma| \leq m' = o(n_{\min}) \leq \frac{1}{2}|\mathcal{C}_u|$. Together with the property of the function $x(1-x), x \in [0,1]$, when $M \geq \frac{5}{\theta_{\min}}$, we have

$$\begin{aligned} \sum_{i \in \Gamma_u} \theta_i \Bigg(\sum_{i \in \mathcal{C}_u} \theta_i - \sum_{i \in \Gamma_u} \theta_i \Bigg) &\geq \sum_{i \in \tau_u(\Gamma_u)} \theta_i \Bigg(\sum_{i \in \mathcal{C}_u} \theta_i - \sum_{i \in \tau_u(\Gamma_u)} \theta_i \Bigg) \\ &\geq \sum_{i \in \tau_u(\Gamma_u)} \theta_i \Bigg(\sum_{i \in \mathcal{C}_u} \theta_i - 2\eta n_{\min} \Bigg). \end{aligned} \tag{31}$$

Here, the first inequality holds since $\sum_{i \in \tau_u(\Gamma_u)} \theta_i \leq |\tau_u(\Gamma_u)| \max_{i \in \mathcal{C}_u^-} \theta_i \leq 2(M^{-1}|\mathcal{C}_u| + 1) \leq \frac{1}{2}|\mathcal{C}_u|\theta_{\min}$. The second inequality is due to (26), $n_{\min} \geq \frac{n}{\beta k} - 1$ and the fact $|\tau_u(\Gamma_u)| \leq |\Gamma_u| \leq \eta n/k$ in the current case. Thus

$$\begin{aligned} B_1 &\leq \prod_{u=1}^{k} \exp\left(-\frac{1}{2} \sum_{i \in \Gamma_u} \theta_i \left(\sum_{i \in \mathcal{C}_u} \theta_i - \sum_{i \in \Gamma_u} \theta_i \right) (\sqrt{p}-\sqrt{q})^2 \right) \\ &\leq \prod_{u=1}^{k} \exp\left(-\frac{1}{2} \sum_{i \in \tau_u(\Gamma_u)} \theta_i \left(\sum_{i \in \mathcal{C}_u} \theta_i - 2\eta n_{\min} \right) (\sqrt{p}-\sqrt{q})^2 \right) \\ &\leq \prod_{i \in \tau(\Gamma)} \exp\left(-\frac{1}{2} \theta_i \left(\sum_{j \in \mathcal{C}_{z(i)}} \theta_j - 2\eta n_{\min} \right) (\sqrt{p}-\sqrt{q})^2 \right) \\ &\leq \prod_{i \in \tau(\Gamma)} \exp\left(-\frac{1}{2} \theta_i (1 - \delta - 2\eta) n_{\min} (\sqrt{p}-\sqrt{q})^2 \right). \end{aligned}$$

Here, the last inequality is due to the approximate normalization constraint on the θ_i's. Thus with the same bound on B_2 we obtain that for any $\tilde{z}$ such that $M < n\ell(\tilde{z}, z) \le m'$,

$$\mathbb{P}\left(L(\tilde{z}) > L(z)\right) \le \prod_{i\in\tau(\Gamma)} \exp\left(-\theta_i\left(1-\delta-2\eta\right) n_{\min}(\sqrt{p}-\sqrt{q})^2\right).$$

Since for any $i \in \mathcal{C}_u^-$, $|\tau_u^{-1}(i)| \le M$, we obtain that $|\tau(\Gamma)| \ge m/M$, and so we have

$$\begin{aligned}
&\mathbb{P}(n\ell(\hat{z}, z) = m)\\
&\le \sum_{\tilde{z}:|\Gamma|=m} k^m \prod_{i\in\tau(\Gamma)} \exp\left(-\theta_i(1-\delta-2\eta) n_{\min}(\sqrt{p}-\sqrt{q})^2\right)\\
&\le \binom{mM}{m}\binom{m}{m/M} k^m \frac{1}{(m/M)!}\left(\sum_{i=1}^{n} \exp\left(-\theta_i(1-\delta-2\eta) n_{\min}(\sqrt{p}-\sqrt{q})^2\right)\right)^{m/M}\\
&\le (ekM)^m \left(\frac{e^2 MnR_{\delta+2\eta}}{m/M}\right)^{m/M}.
\end{aligned}$$

Here, the second inequality is based on counting and the details are as follows. Note that each term in $\prod_{i\in\tau(\Gamma)}$ is a product of at least m/M terms. First, there are at most $\binom{m}{m/M}$ of sets $\tau(\Gamma)$ that map to the same m/M-product. Then, there are at most $\binom{mM}{m}$ of sets Γ that map to the same $\tau(\Gamma)$ (recall that for any $i \in \mathcal{C}_u^-$, $|\tau_u^{-1}(i)| < M$). For each m/M-product, it appear at most $(m/M)!$ times from the expansion of $nR_{\delta+2\eta}$, and that explains the existence of $1/(m/M)!$. The last inequality holds since $\binom{n}{m} \le \left(\frac{en}{m}\right)^m$ and $n! \ge \sqrt{2\pi} n^{n+1/2} e^{-n}$.

Case 3: $m > m'$. In this case, we cannot use the same argument as in case 2 since (31) does not necessarily hold. To proceed, let $\Gamma_{u,u'} = \{i \in [n] : z(i) = u, \tilde{z}(i) = u'\}$ for any $u, u' \in [n]$. We have

$$\begin{aligned}
B_1 &= \prod_{u=1}^{k}\prod_{u'=1}^{k} \exp\left(-\frac{1}{4}\sum_{\substack{\{i:z(i)=u,\\ \tilde{z}(i)=u'\}}} \theta_i \sum_{\substack{\{j:z(j)=u,\\ \tilde{z}(j)\neq u'\}}} \theta_j(\sqrt{p}-\sqrt{q})^2\right)\\
&= \prod_{u=1}^{k}\prod_{u'=1}^{k} \exp\left(-\frac{1}{4}\sum_{i\in\Gamma_{u,u'}} \theta_i \sum_{j\in\mathcal{C}_u\setminus\Gamma_{u,u'}} \theta_j(\sqrt{p}-\sqrt{q})^2\right)\\
&= \prod_{u=1}^{k}\prod_{u'=1}^{k} \exp\left(-\frac{1}{4}\sum_{i\in\Gamma_{u,u'}} \theta_i\Big(\sum_{j\in\mathcal{C}_u} \theta_j - \sum_{j\in\Gamma_{u,u'}} \theta_j\Big)(\sqrt{p}-\sqrt{q})^2\right).
\end{aligned}$$

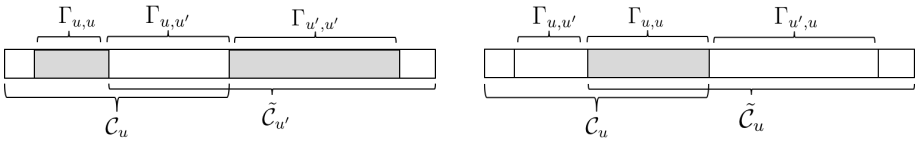

FIG 4. *In the left panel we display nodes in* $\mathcal{C}_u \cup \tilde{\mathcal{C}}_{u'}$, *where* $\tilde{\mathcal{C}}_{u'} = \{i : \tilde{z}(i) = u'\}$. *We also define* $\tilde{\mathcal{C}}_u$ *in the same way. The gray parts indicates nodes correctly clustered, i.e.,* $\{i \in \mathcal{C}_u \cup \tilde{\mathcal{C}}_{u'} : \tilde{z}(i) = z(i)\}$. *Note* $|\Gamma_{u,u}| \le |\mathcal{C}_u \setminus \Gamma_{u,u'}|$ *and* $|\Gamma_{u',u'}| \le |\tilde{\mathcal{C}}_{u'} \setminus \Gamma_{u',u'}|$. *The right panel displays the same nodes but after the labels* u *and* u' *flipped, and the gray part indicating nodes correctly clustered after flipping.*

To further proceed, we need to lower bound all $|\mathcal{C}_u \setminus \Gamma_{u,u'}|$ for all $u \neq u'$. To this end, we essentially follow the arguments leading to Lemma A.1 of [26]. Let $n_{\max}$ and $n_{\min}$ be the maximum and the minimum community sizes. We argue that we must have $|\mathcal{C}_u \setminus \Gamma_{u,u'}| \ge n_{\min}/9$ for all $u' \neq u$. Indeed, if this were not the case, we could switch the labels u and u' in z. This could reduce the Hamming distance between z and $\tilde{z}$ by at least (see Fig. 4 for illustration)

$$\begin{aligned}&|\Gamma_{u,u'}| - |\mathcal{C}_u \setminus \Gamma_{u,u'}| - |\left\{i : \tilde{z}(i) = u'\right\} \setminus \Gamma_{u,u'}| \\ &\ge n_{\min} - \frac{1}{9}n_{\min} - \frac{1}{9}n_{\min} - (n_{\max} - (n_{\min} - \frac{1}{9}n_{\min})) \ge \frac{n}{k}\Big(\frac{5}{3\beta} - \beta\Big) > 0.\end{aligned}$$

Here, the last inequality holds when $1 \le \beta < \sqrt{5/3}$. This leads to a contradiction since by definition, no permutation of the labels should be able to reduce $\ell(\tilde{z}, z)$.

In each $\Gamma_{u,u'}, \forall 1 \le u, u' \le k$ we can find an arbitrary subset $\Gamma'_{u,u'} \subset \Gamma_{u,u'}$ such that $|\Gamma'_{u,u'}| = \eta|\Gamma_{u,u'}|$. In this way $\Gamma' \triangleq \cup_{u\in[k]} \cup_{u'\neq u} \Gamma'_{u,u'}$ satisfies $|\Gamma'| = \eta|\Gamma| \le \eta m$ and $\sum_{i\in\Gamma'_{u,u'}} \theta_i \le 2|\Gamma'_{u,u'}| \le 2\eta|\mathcal{C}_u|$.

Note that for $\eta = o(1)$,

$$\sum_{i\in\mathcal{C}_u} \theta_i - \sum_{i\in\Gamma_{u,u'}} \theta_i \ge \theta_{\min}|\mathcal{C}_u \setminus \Gamma_{u,u'}| \ge \frac{n_{\min}\theta_{\min}}{9} \ge 2\eta\frac{\beta n}{k} \ge 2\eta|\mathcal{C}_u|.$$

Together with the property of the function $x(1-x), x \in [0,1]$, we have

$$\begin{aligned}\sum_{i\in\Gamma_{u,u'}} \theta_i\Big(\sum_{i\in\mathcal{C}_u} \theta_i - \sum_{i\in\Gamma_{u,u'}} \theta_i\Big) &\ge \sum_{i\in\Gamma'_{u,u'}} \theta_i\Big(\sum_{i\in\mathcal{C}_u} \theta_i - \sum_{i\in\Gamma'_{u,u'}} \theta_i\Big) \\ &\ge \sum_{i\in\Gamma'_{u,u'}} \theta_i\big((1-\delta)|\mathcal{C}_u| - 2\eta|\mathcal{C}_u|\big) \ge \sum_{i\in\Gamma'_{u,u'}} \theta_i(1-\delta-2\eta)n_{\min}.\end{aligned}$$

Then

$$\begin{aligned} B_1 &\le \prod_{u=1}^{k}\prod_{u'=1}^{k} \exp\Big(-\frac{1}{4}\sum_{i\in\Gamma'_{u,u'}} \theta_i(1-\delta-2\eta)n_{\min}(\sqrt{p}-\sqrt{q})^2\Big) \\ &= \prod_{u=1}^{k} \exp\Big(-\frac{1}{4}\sum_{i\in\Gamma'_{u}} \theta_i(1-\delta-2\eta)n_{\min}(\sqrt{p}-\sqrt{q})^2\Big) \\ &\le \prod_{i\in\Gamma'} \exp\Big(-\frac{1}{4}\theta_i(1-\delta-2\eta)n_{\min}(\sqrt{p}-\sqrt{q})^2\Big) \end{aligned}$$

Thus with the same bound on B_2 we get

$$\mathbb{P}\left(L(\tilde{z}) > L(z)\right) \le \prod_{i\in\Gamma'} \exp\left(-\frac{1}{2}\theta_i\left(1-\delta-2\eta\right) n_{\min}(\sqrt{p}-\sqrt{q})^2\right).$$

Note we have an extra $1/2$ factor inside the exponent compared with Case 2. Since for each Γ we can find a subset Γ' with $|\Gamma'| = \eta|\tau(\Gamma)| \ge \eta m/M$ satisfying the above inequality, we have

$$\begin{aligned} &\mathbb{P}(n\ell(\hat{z},z)=m) \\ &\le \sum_{\tilde{z}:|\Gamma|=m} k^m \prod_{i\in\Gamma'} \exp\Big(-\frac{1}{2}\theta_i(1-\delta-2\eta)n_{\min}(\sqrt{p}-\sqrt{q})^2\Big) \\ &\le k^m \binom{mM}{m}\binom{m}{\eta m/M}\frac{1}{(\eta m/M)!}\Big(\sum_{i=1}^{n}\exp\Big(-\frac{1}{2}\theta_i(1-\delta-2\eta)n_{\min}(\sqrt{p}-\sqrt{q})^2\Big)\Big)^{\eta m/M} \\ &\le (ekM)^m \left(\frac{e^2MnR_{\delta+2\eta}^{1/2}}{\eta^2 m/M}\right)^{\eta m/M}, \end{aligned}$$

where the last equality is due to Cauchy-Schwarz.

To sum up, till now we have derived the probability $\mathbb{P}(n\ell(\hat{z},z)=m)$ for each $1 \le m \le n$ as follows

$$\mathbb{P}(n\ell(\hat{z},z)=m) \le \begin{cases} (knMR_{\delta+2\eta})^m, & 1 \le m \le M \\ (ekM)^m\left(\frac{e^2MnR_{\delta+2\eta}}{m/M}\right)^{m/M}, & M < m \le \frac{\eta n}{k} \\ (ekM)^m\left(\frac{e^2MnR_{\delta+2\eta}^{1/2}}{\eta^2 m/M}\right)^{\eta m/M}, & m > \frac{\eta n}{k}. \end{cases} \tag{32}$$

Step 3: bounding $\mathbb{E}\ell(\hat{z},z)$. As we have pointed out in the proof outline, we shall combine (22) with (32) in this step to finish the proof. To this end, we divide the argument into three cases according to different possible growth rates of $R_{\delta+2\eta}$.

Case 1: $R_{\delta+2\eta} \le \frac{1}{2(ekM)^{M+2}n}$. Recall that $m' = \eta n/k$. Then

$$\mathbb{E}n\ell(\hat{z}, z) = \mathbb{P}(n\ell(\hat{z}, z) = 1) + \sum_{m=2}^{m'} m\mathbb{P}(n\ell(\hat{z}, z) = m) + \sum_{m=m'+1}^{n} m\mathbb{P}(n\ell(\hat{z}, z) = m).$$

We have $\mathbb{P}(n\ell(\hat{z}, z) = 1) = kMnR_{\delta+2\eta}$ which is upper bounded by $1/2$. Together with $(ekM)^M e^2 MnR_{\delta+2\eta} \le 1/2$, we have

$$\begin{aligned}\sum_{m=2}^{m'} m\mathbb{P}(n\ell(\hat{z}, z) = m) &= \sum_{m=2}^{M} m\mathbb{P}(n\ell(\hat{z}, z) = m) + \sum_{m=M+1}^{m'} m\mathbb{P}(n\ell(\hat{z}, z) = m) \\ &\le \sum_{m=2}^{M} m2^{-m} + \sum_{m=M+1}^{m'} (ekM)^M (e^2 MnR_{\delta+2\eta}) m2^{-(m-M)/M} \\ &\le C_1 (ekM)^M e^2 MnR_{\delta+2\eta},\end{aligned}$$

for some constant $C_1 > 1$ where the last inequality is due to the properties of power series. For $m > m'$ we have

$$\begin{aligned}\frac{\mathbb{P}(n\ell(\hat{z}, z) = m)}{nR_{\delta+2\eta}} &\le (ekM)^m \left(\frac{e^2 MnR_{\delta+2\eta}^{1/2}}{\eta^2 m/M}\right)^{\eta m/M-2} \\ &\le (ekM)^m \left(\left(\frac{e^2 Mn}{\eta^2 m/M}\right)^{\frac{1}{2}} \left(\frac{e^2 MnR_{\delta+2\eta}}{\eta^2 m/M}\right)^{\frac{1}{2}}\right)^{\eta m/M-2}.\end{aligned}$$

We are to show that the above ratio is upper bounded by e^{-m}. This is because $e^2 Mn/(\eta^2 m/M) \le \eta^{-3} e^2 M^2 k$ since $m \ge \eta n/k$ and $e^2 MnR_{\delta+2\eta}/(\eta^2 m/M) \le 1/(2(kM)^M \eta^3 n)$ since $nR_{\delta+2\eta} \le 1/(2(eM)^M)$. Then for some constant $C_2 > 0$ we have

$$\frac{\mathbb{P}(n\ell(\hat{z}, z) = m)}{nR_{\delta+2\eta}} \le (ekM)^m \left(\frac{C_2}{\eta^3 n}\right)^{\frac{\eta m}{2M}} = \exp\left(m\log(ekM) + \frac{\eta m}{2M}\log\left(\frac{C_2}{\eta^3 n}\right)\right) \le e^{-m},$$

where the last inequality is due to the fact that $k \le n^\eta$. By the property of power series we have

$$\sum_{m=m'+1}^{n} m\mathbb{P}(n\ell(\hat{z}, z) = m) \le nR_{\delta+2\eta} \sum_{m=m'+1}^{n} me^{-m} \le C_3 nR_{\delta+2\eta},$$

for some constant $C_3 > 0$. Finally by Jensen's inequality and the assumption $\log k = o(I)$,

$$\mathbb{E}n\ell(\hat{z}, z) \le kMnR_{\delta+2\eta} + C_1(ekM)^M e^2 MnR_{\delta+2\eta} + C_3 nR_{\delta+2\eta} = n\exp(-(1-o(1))I).$$

Case 2: $R_{\delta+2\eta} \geq \frac{M\log n}{(ekM)^{M+2}n}$. Let $m_0 = 2(ekM)^{M+2}nR_{\delta+2\eta}$. Recall that $I \to \infty$ and that $\log k = o(I)$. So $\eta^{-1} = o(I)$ and $m_0 \leq m'$. We have

$$\mathbb{E}\ell(\hat{z}, z) \leq \frac{m_0}{n} + \sum_{m=m_0+1}^{m'} \mathbb{P}\left(n\ell(\hat{z}, z) = m\right) + \sum_{m>m'} \mathbb{P}\left(n\ell(\hat{z}, z) = m\right).$$

To obtain the last display, we divide both sides of (22) by n, replace all the m's in front of the probabilities in the summation by n and then upper bound the first m_0 probabilities by one. To further bound the right side of the last display, we have

$$\begin{aligned}\sum_{m=m_0+1}^{m'} \mathbb{P}\left(n\ell(\hat{z}, z) = m\right) &\leq \sum_{m=m_0+1}^{m'} \left(\frac{(ekM)^{M+2}nR_{\delta+2\eta}}{m_0}\right)^{m/M} \\ &\leq \sum_{m=m_0+1}^{m'} 2^{-m/M} \leq M2^{-m_0/M}.\end{aligned}$$

Since $m_0 \geq 2M\log n$, we have $2^{-m/M} \leq 2^{-2\log n} \leq m_0/n$. Thus

$$\sum_{m=m_0+1}^{m'} \mathbb{P}\left(n\ell(\hat{z}, z) = m\right) \leq \frac{Mm_0}{n}.$$

For $m \geq m'$, we are going to show $\mathbb{P}\left(n\ell(\hat{z}, z) = m\right) \leq 2^{-\eta m/M}$. We have

$$\mathbb{P}\left(n\ell(\hat{z}, z) = m\right) \leq \left(\frac{(ekM)^{\frac{M}{\eta}}e^2M^2nR_{\delta+2\eta}^{1/2}}{\eta^2 m'}\right)^{\eta m/M} \leq \left(\frac{(ekM)^{\frac{M}{\eta}+3}R_{\delta+2\eta}^{1/2}}{\eta^3}\right)^{\eta m/M}.$$

Since $R_{\delta+2\eta} \leq \exp(-(1-\delta-2\eta)I)$ by Jensen's inequality, $\log k = o(I)$ and $\eta^{-1} = o(I)$, we have $\eta^{-3}(ekM)^{\frac{M}{\eta}+3}R_{\delta+2\eta}^{1/2} \leq 1/2$. Then

$$\sum_{m>m'} \mathbb{P}\left(n\ell(\hat{z}, z) = m\right) \leq \sum_{m>m'} 2^{-\eta m/M} \leq \eta^{-1}M2^{-\eta^2 n/M} \leq \frac{m_0}{n}.$$

Thus by Jensen's inequality $\mathbb{E}\ell(\hat{z}, z) \leq (2+M)m_0/n \leq \exp(-(1-o(1))I)$.

Case 3: $\frac{1}{2(ekM)^{M+2}n} < R_{\delta+2\eta} < \frac{M\log n}{(ekM)^{M+2}n}$. Let $m_0 = 2M\log n$. As we have shown in Case 2, $M\log n \leq m'$. Then

$$\mathbb{E}\ell(\hat{z}, z) \leq \frac{m_0}{n} + \sum_{m=m_0+1}^{m'} \mathbb{P}\left(n\ell(\hat{z}, z) = m\right) + \sum_{m>m'} \mathbb{P}\left(n\ell(\hat{z}, z) = m\right).$$

We have

$$\sum_{m=m_0+1}^{m'} \mathbb{P}\left(n\ell(\hat{z},z)=m\right) \le \sum_{m=m_0+1}^{m'} \left(\frac{(ekM)^{M+2} n R_{\delta+2\eta}}{m_0}\right)^{m/M}$$
$$\le \sum_{m=m_0+1}^{m'} 2^{-m/M} \le M 2^{-m_0} \le \frac{m_0}{n}.$$

For $m > m'$, in Case 2 we have shown $\sum_{m>m'} \mathbb{P}\left(n\ell(\hat{z},z)=m\right) \le \sum_{m>m'} 2^{-\eta m/M} \le \eta^{-1} M 2^{-\eta^2 n/M}$, which is also upper bounded by m_0/n. Together we have $\mathbb{E}\ell(\hat{z},z) \le 3m_0/n$. Since $2(ekM)^{M+2} R_{\delta+2\eta} \ge 1/n$ and $\log k = o(I)$, we have $n\exp(-(1-o(1))I) \ge M\log n$ for some positive sequence $o(1)$. Then $\mathbb{E}\ell(\hat{z},z) \le \exp(-(1-o(1))I)$.

6. Concluding Remarks. This paper studies community detection for DCBMs. We have derived the minimax rates of the problem for a wide collection of parameter spaces. An efficient two-stage algorithm has been proposed and proved to achieve the minimax rates in various scenarios. The results provide a solid foundation for future investigations of some interesting open problems in this area. For example, it is unknown whether the minimax rates in this paper can still be achieved if the number of clusters k is unknown. Moreover, it is of interest to see whether the signal-to-noise ratio condition $\frac{p^{3/2}}{n^{1/2}(p-q)^2} = o(1)$ in Theorem 3 can be improved by a polynomial-time algorithm. Finally, the minimax rate may exhibit a different form if the sizes of clusters are far from being comparable. In this case, it also makes sense to study a different loss function other than that used in this paper.

SUPPLEMENTARY MATERIAL

Supplement to "Community Detection in Degree-Corrected Block Models"
(; .pdf). The supplement presents additional numerical results, additional proofs of main results, properties of $J_t(p,q)$ and proofs of auxiliary results.

References.

[1] E. Abbe and C. Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 670–688. IEEE, 2015.

[2] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *Information Theory, IEEE Transactions on*, 62(1):471–487, 2016.

[3] L. A. Adamic and N. Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM, 2005.

[4] A. A. Amini, A. Chen, P. J. Bickel, and E. Levina. Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41(4):2097–2122, 2013.

[5] P. J. Bickel and A. 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.

[6] Y. Chen, X. Li, and J. Xu. Convexified modularity maximization for degree-corrected stochastic block models. *arXiv preprint arXiv:1512.08425*, 2015.

[7] P. Chin, A. Rao, and V. Vu. Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Proceedings of The 28th Conference on Learning Theory*, pages 391–423, 2015.

[8] A. Dasgupta, J. E. Hopcroft, and F. McSherry. Spectral analysis of random graphs with skewed degree distributions. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 602–610. IEEE, 2004.

[9] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

[10] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving optimal misclassification proportion in stochastic block model. *arXiv preprint arXiv:1505.03772*, 2015.

[11] L. Gulikers, M. Lelarge, and L. Massoulié. An impossibility result for reconstruction in a degree-corrected planted-partition model. *arXiv preprint arXiv:1511.00546*, 2015.

[12] L. Gulikers, M. Lelarge, and L. Massoulié. A spectral method for community detection in moderately-sparse degree-corrected stochastic block models. *arXiv preprint arXiv:1506.08621*, 2015.

[13] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *arXiv preprint arXiv:1412.6156*, 2014.

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

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

[16] J. Jin. Fast community detection by SCORE. *The Annals of Statistics*, 43(1):57–89, 2015.

[17] B. Karrer and M. E. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

[18] J. Lei, A. Rinaldo, et al. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

[19] L. Massoulié. Community detection thresholds and the weak ramanujan property. In *STOC*, pages 694–703. ACM, 2014.

[20] E. Mossel, J. Neeman, and A. Sly. Stochastic block models and reconstruction. *arXiv preprint arXiv:1202.1499*, 2012.

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

[22] E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for binary symmetric block models. *arXiv preprint arXiv:1407.1591*, 2014.

[23] T. P. Peixoto. Model selection and hypothesis testing for large-scale network models with overlapping groups. *Physical Review X*, 5(1):011033, 2015.

[24] T. Qin and K. Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems*, pages 3120–3128, 2013.

[25] K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

[26] A. Y. Zhang and H. H. Zhou. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, to appear, 2015.

[27] Y. Zhao, E. Levina, and J. Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266–2292, 2012.

Department of Statistics
University of Chicago
Chicago, IL 60615.
E-mail: chaogao@galton.uchicago.edu
URL: https://galton.uchicago.edu/~chaogao

Department of Statistics
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104.
E-mail: zongming@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/~zongming

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