Exact Recovery Threshold in the Binary Censored Block Model

Bruce Hajek \quad Yihong Wu \quad Jiaming Xu

Abstract—Given a background graph with \( n \) vertices, the binary censored block model assumes that vertices are partitioned into two clusters, and every edge is labeled independently at random with labels drawn from \( \text{Bern}(1 - \epsilon) \) if two endpoints are in the same cluster, or from \( \text{Bern}(\epsilon) \) otherwise, where \( \epsilon \in [0, 1/2] \) is a fixed constant. For Erdős-Rényi graphs with edge probability \( p = a \log n / n \) and fixed \( a \), we show that the semidefinite programming relaxation of the maximum likelihood estimator achieves the optimal threshold \( a(\sqrt{1 - \epsilon - \epsilon})^2 > 1 \) for exactly recovering the partition from the labeled graph with probability tending to one as \( n \to \infty \). For random regular graphs with degree scaling as \( a \log n \), we show that the semidefinite programming relaxation also achieves the optimal recovery threshold \( aD(\text{Bern}(1/2) \| \text{Bern}(\epsilon)) > 1 \), where \( D \) denotes the Kullback-Leibler divergence.

I. INTRODUCTION

This paper studies the problem of finding communities in a network based on edge labels, under the binary censored block model [1], [12]. This generative model is described as follows: given a background graph \( G = ([n], E) \) with \( n \) vertices and edge set \( E \) and an arbitrary partition of the vertices into two clusters, for every edge \((i, j) \in E\), a label \( L_{ij} \in \{\pm1\} \) is independently drawn according to the following distribution:

\[
\mathbb{P} \{ L_{ij} = \ell | \sigma^*_i, \sigma^*_j \} = \begin{cases} 1 - \epsilon & \ell = \sigma_i^* \sigma_j^* \\ \epsilon & \ell = -\sigma_i^* \sigma_j^* \end{cases},
\]

where \( \sigma^*_i = 1 \) if vertex \( i \) is in the first cluster and \( \sigma^*_i = -1 \) otherwise; \( \epsilon \in [0, 1/2] \) is a fixed constant. In other words, each label is a noisy observation about whether a pair of neighbors belong to the same cluster.

In this paper, we study the problem of exact cluster recovery, where the goal is to construct an estimator \( \hat{\sigma} \) based on the graph \( G \) and the edge labels \( \{L_{ij} \}_{(i,j) \in E} \) such that \( \hat{\sigma} \) coincides with \( \sigma^* \) up to a global flip of signs with high probability. We focus on two particular cases for the background graph \( G \):

- \( G \) is Erdős-Rényi with edge probability \( p \):
  \[
  p = a \log n / n \quad n \to \infty,
  \]

- \( G \) is regular with degree \( d \) (assume \( nd \) is even):
  \[
  d = [(a \log n) / n \to \infty].
  \]

where \( a > 0 \) is a fixed constant. Note that under the censored block model, the background graph \( G \) alone does not contain any information about the underlying cluster structure.

In the case where the background graph \( G \) is Erdős-Rényi, the binary censored block model is a special case of the labeled stochastic block model proposed in [9]. Under the scaling regime (1), denote by \( a^*(\epsilon) \) the optimal recovery threshold, namely, the infimum of \( a > 0 \) such that exact cluster recovery is possible with probability converging to one as \( n \to \infty \). The following result is obtained in [1] about the sharp threshold in the very noisy regime:

**Theorem 1.** As \( \epsilon \to 1/2 \), \( a^*(\epsilon) = \frac{2 + o(1)}{(1 - 2\epsilon)^2} \).

Apart from the above limiting case, the expression for the sharp threshold \( a^*(\epsilon) \) is unknown. Furthermore, the threshold in Theorem 1 is achieved by the maximum likelihood (ML) estimator, which involves an exhaustive search over \( \{ \pm1 \}^n \) and is computationally intractable. It remains open to find a procedure to achieve the exact recovery threshold in polynomial time. It was proved in [1] that a semidefinite programming (SDP) relaxation of the ML estimator succeeds when \( a \) exceeds a certain threshold \( a(\epsilon) \), which satisfies

\[
a(\epsilon) = \frac{4 + o(1)}{(1 - 2\epsilon)^2} \text{ as } \epsilon \to 1/2.
\]

In this paper, we show that for all \( \epsilon \in [0, 1/2] \), the optimal recovery threshold is given by

\[
a^*(\epsilon) = \frac{1}{(\sqrt{1 - \epsilon - \epsilon})^2},
\]

which recovers Theorem 1 as a special case. Furthermore, the optimal threshold (3) can be achieved by the SDP relaxations. The proof techniques are similar to those in [7]; we show that a necessary condition for the maximum likelihood estimator to succeed coincides with a sufficient condition for the correctness of the SDP procedure, thereby establishing both the optimal recovery threshold and the optimality of the SDP relaxation.

In the case where the background graph \( G \) is regular, the following converse bound is proved in [1]:

**Theorem 2.** Assume \( G \) is a \( d \)-regular graph. Under the scaling regime (2), exact cluster recovery with probability converging to 1 is impossible if \( ad(1/2)|\epsilon| < 1 \), where

\[
D(p||q) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{q} \text{ denotes the Kullback-Leibler divergence between } \text{Bern}(p) \text{ and } \text{Bern}(q).
\]

Sufficient conditions for the success of SDP relaxation given in [1] do not match the above necessary condition. It remains open whether the converse bound is tight and achievable in
polynomial time. In this paper, we improve the sufficient conditions derived in [1, Theorem 5.3], and show that the SDP relaxation succeeds if $aD(e\lambda_2 + \frac{1}{2}(1 - \lambda_2))||e > 1$, where $\lambda_2$ is the normalized second largest eigenvalue of the adjacency matrix of $G$ as defined in (7). Interestingly, if $G$ is a random $d$-regular graph, then $\lambda_2 = o_p(1)$ in the regime (2) and thus our sufficient condition matches the converse bound in Theorem 2, i.e., SDP relaxation achieves the optimal recovery threshold $aD(1/2) ||e > 1$.

An interesting open problem is to identify the sharp recovery thresholds in terms of graph characteristics when $G$ is any deterministic graph. Such attempt has been pursued in [1] and more recently in [4] in a more general setup.

A. Notation

Denote the identity matrix by $I$ and the all-one matrix by $J$. We write $X \succeq 0$ if $X$ is positive semidefinite and $X \succeq 0$ if all the entries of $X$ are non-negative. Let $S^n$ denote the set of all $n \times n$ symmetric matrices. For $X \in S^n$, let $\lambda_2(X)$ denote its second largest eigenvalue. For any matrix $Y$, let $\|Y\|$ denote its spectral norm. For any positive integer $n$, let $[n] = \{1, \ldots, n\}$. For any set $T \subset [n]$, let $|T|$ denote its cardinality and $T^c$ denote its complement. We use standard big $O$ notations, e.g., for any sequences $\{a_n\}$ and $\{b_n\}$, $a_n = O(b_n)$ if there is an absolute constant $c > 0$ such that $1/c \leq a_n/b_n \leq c$; $a_n = \Omega(b_n)$ or $b_n = O(a_n)$ if there exists an absolute constant $c > 0$ such that $a_n/b_n \geq c$. Let Bern$(p)$ denote the Bernoulli distribution with mean $p$ and Binom$(n, p)$ denote the binomial distribution with $n$ trials and success probability $p$. All logarithms are natural and we use the convention $0 \log 0 = 0$.

II. MAIN RESULTS

Under the binary censored block model, with possibly unequal cluster sizes, the cluster structure can be represented by a vector $\sigma \in \{\pm 1\}^n$ such that $\sigma_i = 1$ (resp. $-1$) if vertex $i$ is in the first (resp. second) cluster. Recall that $\sigma^* \in \{\pm 1\}^n$ correspond to the true clusters. Let $A$ denote the weighted adjacency matrix such that $A_{ij} = 0$ if $i, j$ are not connected and $A_{ij} = 1$ (resp. $-1$) if $i, j$ are connected by an edge with label $+1$ (resp. $-1$). Then the ML estimator of $\sigma^*$ can be simply stated as

$$\max_{\sigma} \sum_{i,j} A_{ij} \sigma_i \sigma_j \quad \text{s.t. } \sigma_i \in \{\pm 1\}, \ i \in [n],$$

which maximizes the number of in-cluster $+1$ edges minus that of in-cluster $-1$ edges, or equivalently, maximizes the number of cross-cluster $-1$ edges minus that of cross-cluster $+1$ edges. The NP-hard max-cut problem can be reduced to (4) by simply labeling all the edges in the input graph as $-1$ edges, and thus (4) is computationally intractable in the worst case. Instead, we consider the SDP studied in [1] obtained by convex relaxation. Let $Y = \sigma \sigma^\top$. Then $Y_{ii} = 1$ is equivalent to $\sigma_i = \pm 1$. Therefore, (4) can be recast as

$$\max_{Y} \langle A, Y \rangle \quad \text{s.t. } \text{rank}(Y) = 1, \ Y_{ii} = 1, \ i \in [n].$$

Replacing the rank-one constraint by positive semidefiniteness, we obtain the following convex relaxation of (5), which is an SDP:

$$\hat{Y}_{\text{SDP}} = \text{arg} \max_{Y} \langle A, Y \rangle \quad \text{s.t. } Y \succeq 0, \ Y_{ii} = 1, \ i \in [n].$$

We remark that (6) does not rely on any knowledge of the model parameters. Let $Y^* = \sigma^* \sigma^*^\top$ and $Y_n = \{\sigma \sigma^\top : \sigma \in \{\pm 1\}^n\}$.

A. Erdős-Rényi random background graph

The following result establishes the success condition of the SDP procedure:

**Theorem 3.** Assume $G$ is an Erdős-Rényi random graph. Under the scaling regime (1), if $a(\sqrt{1 - \epsilon} - \sqrt{\epsilon})^2 > 1$, then as $n \to \infty$, $\min_{Y \in Y_n} \mathbb{P}\{\hat{Y}_{\text{SDP}} = Y^*\} \geq 1 - n^{-O(1)}$.

Next we prove a converse for Theorem 3 which shows that the recovery threshold achieved by the SDP relaxation is in fact optimal.

**Theorem 4.** Assume $G$ is an Erdős-Rényi random graph. Under the scaling regime (1), if $a(\sqrt{1 - \epsilon} - \sqrt{\epsilon})^2 < 1$ and $\sigma^*$ is uniformly chosen from $\{\pm 1\}^n$, then for any sequence of estimators $\hat{Y}_n$, $\mathbb{P}\{\hat{Y}_n = Y^*\} \to 0$ as $n \to \infty$.

It can be shown that Theorem 4 continues to hold if the cluster sizes are proportional to $n$ and known a priori, i.e., the prior distribution of $\sigma^*$ is uniform over $\{\sigma \in \{\pm 1\}^n : \sigma^\top 1 = 2K - n\}$ for $K = \lfloor mp \rfloor$ with $p \in (0, 1/2]$. Together with Theorem 3, this implies that the recovery threshold $a(\sqrt{1 - \epsilon} - \sqrt{\epsilon})^2 > 1$ is insensitive to $p$.

The above exact recovery threshold in the regime $p = a \log n/n$ should be contrasted with the correlated recovery threshold in the sparse regime $p = a/n$ for constant $a$. In this sparse regime, there exists at least a constant fraction of vertices with no neighbor and exactly recovering the clusters is hopeless; instead, the goal is to find an estimator $\hat{\sigma}$ that is positively correlated with $\sigma^*$ up to a global flip of signs. It was conjectured in [9] that the positively correlated recovery is possible if and only if $a(1 - 2\epsilon)^2 > 1$; the converse part is shown in [10] and recently it is proved in [12] that spectral algorithms achieve the sharp threshold in polynomial-time.

B. Regular background graph

In this section, we consider background graph $G$ which is regular with degree $d = [a \log n]$. Let $A_G$ denote the
The adjacency matrix of $G$ and define $\lambda_2$ to be the second largest eigenvalue of $\frac{1}{d} A_G$, i.e.,

$$\lambda_2 \triangleq \max_{x \neq 0, \|x\|_2 = 1} x^T \left( \frac{1}{d} A_G \right) x.$$  

(7)

The following result establishes the success condition of the SDP procedure:

**Theorem 5.** Assume $G$ is a $d$-regular graph. Under the scaling regime (2), if $aD(\lambda_2 + \frac{1}{2}(1 - \lambda_2)) ||\epsilon|| > 1$, then $\min_{Y \in \mathbb{Y}_n} \mathbb{P}\{|Y_{\text{SDP}} = Y^*\} \geq 1 - \lambda_2 n^{-\Theta(1)}$ as $n \to \infty$.

Together with Theorem 2, Theorem 5 shows that if $\lambda_2 = o(p(1))$, then the SDP relaxation achieves the optimal approximation threshold $aD(\frac{1}{d} ||\epsilon||) > 1$. In particular, it is known that for the random $d$-regular graph, for any $\delta > 0$, $n^2 \lambda_2 \leq 2\sqrt{d - 1 + \delta}$ with high probability [6].

Theorem 5 improves the sufficient conditions in [1, Theorem 5.3]. In particular, by Pinsker’s inequality, $D(\lambda_2 + \frac{1}{2}(1 - \lambda_2)) ||\epsilon|| \geq (1 - \lambda_2)^2(1 - 2\epsilon)^2$. Thus Theorem 5 implies that the SDP succeeds if $a(1 - 2\epsilon)^2(1 - \lambda_2)^2 > 2$, while [1, Theorem 5.3] requires $a(1 - 2\epsilon)^2(1 - \lambda_2)^2 > 4(1 + |\lambda_2|)$ if $\epsilon \to 1/2,$ where $|\lambda_2|$ is the smallest eigenvalue of $\frac{1}{d} A_G$.

III. PROOFS

Our analysis of the SDP relies on the standard dual certificate argument, which amounts to constructing the dual variables so that the desired KKT conditions are satisfied for the primal variable corresponding to the true clusters. In particular, the following dual certificate lemma provides a deterministic sufficient condition for the success of SDP (6).

**Lemma 1.** Suppose there exist $D^* = \text{diag}\{d_i^*\}$ such that $S^* \triangleq D^* - A$ satisfies $S^* \succeq 0$, $\lambda_2(S^*) > 0$ and

$S^* \sigma^* = 0.$

Then $Y_{\text{SDP}} = Y^*$ is the unique solution to (6).

**Proof:** The Lagrangian function is given by

$$L(Y, S, D) = \langle A, Y \rangle + \langle S, Y \rangle - \langle D, Y - I \rangle,$$

where the Lagrangian multipliers are $S \succeq 0$ and $D = \text{diag}\{d_i\}$. Then for any $Y$ satisfying the constraints in (6),

$$\langle A, Y \rangle \overset{(a)}{\leq} L(Y, S^*, D^*) = \langle D^*, I \rangle = \langle D^*, Y^* \rangle = (D^*, Y^*) \overset{(b)}{=} (A + S^*, Y^*),$$

where $(a)$ holds because $\langle S^*, Y \rangle \geq 0$; $(b)$ holds because $(Y^*, S^*) = (\sigma^*)^T S^* \sigma^* = 0$ by (8). Hence, $Y^*$ is an optimal solution. It remains to establish its uniqueness. To this end, suppose $\hat{Y}$ is an optimal solution. Then,

$$\langle S^*, \hat{Y} \rangle = \langle D^* - A, \hat{Y} \rangle \overset{(a)}{=} \langle D^* - A, Y^* \rangle = (S^*, Y^*) = 0.$$

where $(a)$ holds because $(A, \hat{Y}) = (A, Y^*)$ and $\hat{Y}_{ii} = Y^*_{ii} = 1$ for all $i \in [n]$. In view of (8), since $\hat{Y} \succeq 0$, $S^* \succeq 0$ with $\lambda_2(S^*) > 0$, $\hat{Y}$ must be a multiple of $Y^* = \sigma^*(\sigma^*)^T$. Because $\hat{Y}_{ii} = 1$ for all $i \in [n]$, $\hat{Y} = Y^*$.

A. Proofs for Erdős-Rényi random background graph

One of the key ingredients of the proof is the spectrum of labeled Erdős-Rényi random graph. Recall that $A$ is a symmetric and zero-diagonal random matrix, where the entries $A_{ij} : i < j$ are independent and $A_{ij} \sim P(1 - c) \delta_{i,j} + P(p \delta_{i,j} + (1 - p)\delta_{i,j})$ if $i, j$ are in the same cluster and $A_{ij} \sim P(1 - c) \delta_{i,j} + P(p \delta_{i,j} + (1 - p)\delta_{i,j})$ otherwise. Assume $p \geq c_0 \log n$ for any constant $c_0 > 0$. The following theorem shows that $|A - E[A]|_2 \leq c^2 \sqrt{np}$ with high probability for some constant $c' > 0$. Its proof can be found in [8].

**Theorem 6.** For any $c > 0$, there exists $c' > 0$ such that for any $n \geq 1$, $\mathbb{P}\{|A - E[A]|_2 \leq c^2 \sqrt{np} \geq 1 - n^{-c'}$.

Let $X_1, X_2, \ldots, X_m \overset{i.i.d.}{\sim} P(1 - c) \delta_{i,j} + P(p \delta_{i,j} + (1 - p)\delta_{i,j})$ for $m \in \mathbb{N}$, $p \in [0, 1]$ and a fixed constant $\epsilon \in [0, 1/2]$, where $m = n + o(n)$ and $p = a \log n / n$ for some $a > 0$ as $n \to \infty$. The following upper tail bound for $\sum_{i=1}^m X_i$ follows from the Chernoff bound.

**Lemma 2.** Assume that $k \in m$ and $k_n = (1 + o(1)) \frac{\log n}{\log \log n}$. Then

$$\mathbb{P}\left\{ \sum_{i=1}^m X_i \leq k \right\} \leq n^{-a(\sqrt{1 - \epsilon} - \sqrt{2})^2 + o(1)}.$$

**Proof:** If $\epsilon = 0$, then $\sum_{i=1}^m X_i \sim \text{Binom}(m, p)$ and the lemma follows from [7, Lemma 2]. Next we focus on the case $\epsilon > 0$. It follows from the Chernoff bound that

$$\mathbb{P}\left\{ \sum_{i=1}^m X_i \leq k \right\} \leq \exp(-m \ell(k_n/m)).$$

where $\ell(x) = \sup_{\lambda > 0} -\lambda x - \log \mathbb{E}[e^{-\lambda X_1}]$. Since $X_1 \sim P(1 - c) \delta_{i,j} + P(p \delta_{i,j} + (1 - p)\delta_{i,j})$, $\mathbb{E}[e^{-\lambda X_1}] = 1 + p \left( e^{-\lambda(1 - c)} + e^{\lambda c - 1} \right)$. Notice that $-\lambda x - \log \mathbb{E}[e^{-\lambda X_1}]$ is concave in $\lambda$, whose maximum is attained at $\lambda^*$ such that

$$-x + \frac{p(e^{-\lambda^*}(1 - c) - e^{\lambda^*}c)}{1 + p(e^{-\lambda^*}(1 - c) + e^{\lambda^*}c - 1)} = 0.$$ 

Hence, for $x = k_n / m$ we obtain $\lambda^* = \frac{1}{2} \log \frac{1 - c}{c} + o(1)$ and

$$\ell(k_n / m) = -\lambda x k_n / m - \log \left( 1 + p \left( e^{-\lambda^*}(1 - c) + e^{\lambda^*}c - 1 \right) \right) = -\frac{k_n}{2m} \log \left( 1 - \frac{\epsilon}{\epsilon} \right) - \log \left( 1 - p(\sqrt{1 - \epsilon} - \sqrt{2})^2 \right) + o(k_n / m) = a(\sqrt{1 - \epsilon} - \sqrt{2})^2 \log n / n + o(\log n / n),$$

where the last equality holds due to the Taylor expansion of $\log(1 - x)$ at $x = 0$ and $p = a \log n / n$. Combining the last displayed equation with (10) gives the desired (9).

The following lemma establishes a lower tail bound for $\sum_{i=1}^m X_i$.

}
Lemma 3. Let $k_n$ be defined in Lemma 2. Then
\[
P \left\{ \sum_{i=1}^{m} X_i \leq -k_n \right\} \geq n^{a(\sqrt{n}-\sqrt{\lambda})^2 + o(1)}, \]
where (a) holds because conditioning on $\sum_{i=1}^{m} X_i = k^*$, $\sum_{i=1}^{m} X_i$ and $\sum_{i=1}^{m} Z_i$ have the same distribution. Next we lower bound the two terms in (11) separately.

We use the following non-asymptotic bound on the binomial tail probability [2, Lemma 4.7.2]: For $U \sim \text{Binom}(n,p)$,
\[
(8k(1-\lambda))^{-1/2}e^{-nD(\lambda||p)} \leq P \{ U \geq k \} \leq e^{-nD(\lambda||p)},
\]
where $\lambda = k/n \in (0, 1)$, $p \leq 0$. Let $W \sim \text{Binom}(k^*,\epsilon)$. Then,
\[
P \left\{ \sum_{i=1}^{k^*} Z_i \leq -k_n \right\} = P \left\{ W \geq \frac{k^* + k_n}{2} \right\} \geq \exp \left[ -k^* D \left( \frac{1}{2} + \frac{k_n}{2k^*} \right) \right] = \exp \left[ -k^* D(1/2|\epsilon) + o(\log n) \right],
\]
Moreover, using the following bound on binomial coefficients [2, Lemma 4.7.1]:
\[
\frac{\sqrt{\pi}}{2} \leq \left( \frac{n!}{(\lambda n)^{\lambda} \lambda^{n\lambda}} \right) \leq 1,
\]
where $n = k/n \in (0, 1)$ and $h(\lambda) = -\lambda \log(1-\lambda) \log(1-\lambda)$ is the binary entropy function, we have
\[
P \left\{ \sum_{i=1}^{m} X_i = k^* \right\} = \sum_{k^*} p^{k^*} (1-p)^{k^*} \geq 1 \frac{1}{\sqrt{2k^*}} \exp \left[ -m D(k^*/n||p) \right] = \exp \left[ -a \log n + k^* \log \frac{ea\log n}{k^*} + o(\log n) \right].
\]
Observe that by the definition of $k^*$, $\log \frac{a\log n}{k^*} = D(1/2|\epsilon) + o(1)$ and it follows from (11) that
\[
P \left\{ \sum_{i=1}^{m} X_i \leq -k_n \right\} \geq e^{-a \log n + 2a\sqrt{a(1-\epsilon)} \log n} + o(\log n) \geq n^{a(\sqrt{n}-\sqrt{\lambda})^2 + o(1)}.
\]

Proof of Theorem 3:
Let $D^* = \text{diag} \{ d_i^* \}$ with
\[
d_i^* = \sum_{j=1}^{m} A_{ij} \sigma_j^* \sigma_j^T.
\]
It suffices to show that $S^* = D^* - A$ satisfies the conditions in Lemma 1 with high probability. By definition, $d_i^* \sigma_j^* = \sum_j A_{ij} \sigma_j^*$ for all $i$, i.e., $D^* \sigma^* = \Lambda \sigma^*$. Thus (8) holds, that is, $S^* \sigma^* = 0$. It remains to verify that $S^* \geq 0$ and $\lambda_i(S^*) > 0$ with probability converging to one, which amounts to showing that
\[
P \left\{ \inf_{x \perp \sigma^*, \|x\|_2^2 = 1} x^T S^* x > 0 \right\} \to 1.
\]
Note that $E[A] = (1-2\epsilon)p(Y^* - I)$ and $Y^* = \sigma^*(\sigma^*)^T$. Thus for any $x$ such that $x \perp \sigma^*$ and $\|x\|_2 = 1$,
\[
x^T S^* x = x^T D^* x - x^T E[A] x - x^T (A - E[A]) x \geq \min d_i^* + (1-2\epsilon)p \|A - E[A]\| \geq \min d_i^* + (1-2\epsilon)p \|A - E[A]\|.
\]
where (a) holds since $\langle x, \sigma^* \rangle = 0.0.0$. It follows from Theorem 6 that $\|A - E[A]\| \leq c' \sqrt{\log n}$ with high probability for a constant $c'$ depending only on $a$. Moreover, note that $d_i$ is equal to distribution to $\sum_{i=1}^{n} X_i$, where $X_i \sim p(1-\epsilon)\delta_{x_i} + p\epsilon, \delta_{x_i} + 1 - p\delta_{0}$. Hence, Lemma 2 implies that
\[
P \left\{ \sum_{i=1}^{n} X_i \geq \frac{\log n}{\log \log n} \right\} \geq 1 - n^{-a(\sqrt{n}-\sqrt{\lambda})^2 + o(1)}.
\]
Applying the union bound yields that $\min_{i \in [n]} d_i^* \geq \frac{\log n}{\log \log n}$ holds with probability at least $1 - n^{-a(\sqrt{n}-\sqrt{\lambda})^2 + o(1)}$. It follows from the assumption $a(\lambda - \sqrt{\lambda}) > 1$ and (15) that the desired (14) holds, completing the proof.

Proof of Theorem 4:
Let $\sigma^*$ be uniformly distributed on $\{ \pm 1 \}^n$. First consider the case of $\epsilon = 0$. If $a < 1$, then the number of isolated vertices tends to infinity in probability [5]. Notice that for isolated vertices $i$, vertex $\sigma_i^*$ is equally likely to be $\pm 1$ conditional on the graph. Hence, the probability of exact recovery converges to 0.

Next we consider $\epsilon > 0$. Since the prior distribution of $\sigma^*$ is uniform, the ML estimator minimizes the average error probability among all estimators and thus we only need to show that the ML estimator fails with high probability. Let $\epsilon(i,T) : \{ \sum_{j \in T} |A_{ij}| \}$ denote the number of edges between vertex $i$ and vertices in set $T \subset [n]$. Let $s_i = \sum_{j : \sigma_j^* = \sigma_i^*} A_{ij}$ and $r_i = \sum_{j : \sigma_j^* \neq \sigma_i^*} A_{ij}$. Let $F$ denote the event that $\min_{i \in [n]} (s_i - r_i) \leq -1.0$. Notice that $F$ implies the existence of $i \in [n]$ such that $\sigma_i^*$ with $\sigma_i^* = -\sigma_i^*$ and $\sigma_j^* = \sigma_j^*$ for $j \neq i$ achieves a strictly higher likelihood than $\sigma^*$. Hence $\Pr \{ \text{ML fails} \} \geq \Pr \{ F \}$. Next we bound $\Pr \{ F \}$ from below.
Let $T$ denote the set of first $\lceil \frac{n}{\log \log n} \rceil$ vertices. Let $s'_i = \sum_{j \in T : \sigma_j = \sigma_i} A_{ij}$ and $r'_i = \sum_{j \in T : \sigma_j \neq \sigma_i} A_{ij}$. Then

$$\min_{i \in [n]} (s'_i - r'_i) \leq \min_{i \in T} (s_i - r_i) \leq \min_{i \in T} (s'_i - r'_i) + \max e(i, T).$$

Let $E_1, E_2$ denote the event that $\max_{i \in T} e(i, T) \leq \frac{\log n}{\log \log n} - 1$, $\min_{i \in T} (s'_i - r'_i) \geq -\frac{\log n}{\log \log n}$, respectively. In view of (16), we have $F \supseteq E_1 \cap E_2$ and hence it boils down to proving that $\mathbb{P}\{E_i\} \to 1$ for $i = 1, 2$.

Notice that $e(i, T) \sim \text{Binom}(|T|, a \log n/n)$. In view of the following Chernoff bound for binomial distributions [11, Theorem 4.4]: For $r \geq 1$ and $X \sim \text{Binom}(n, p)$, $\mathbb{P}\{X > np\} \leq (e/r)^{np}$, we have

$$\mathbb{P}\left\{ e(i, T) \geq \frac{\log n}{\log \log n} - 1 \right\} \leq \left( \frac{e \log^2 n}{a \log \log n} \right)^{\frac{\log n}{\log \log n} + 1} = n^{-2+o(1)}.$$

Applying the union bound yields $\mathbb{P}\{E_1\} \geq 1 - n^{-1+o(1)}$. Moreover,

$$\mathbb{P}\{E_2\} \leq 1 - \sum_{i \in T} \mathbb{P}\left\{ s'_i - r'_i > \frac{\log n}{\log \log n} \right\} \leq 1 - \left( 1 - n^{-a(\sqrt{T} - \epsilon)} + o(1) \right)^{|T|} \leq 1 - \left( 1 - n^{-a(\sqrt{T} - \epsilon)} + o(1) \right) \to 1,$$

where (a) holds because $\{s'_i - r'_i\}_{i \in T}$ are mutually independent; (b) follows from Lemma 3; (c) is due to $1 + x \leq e^x$ for all $x \in \mathbb{R}$; (d) follows from the assumption that $a(\sqrt{T} - \epsilon) > 1$. Thus $\mathbb{P}\{F\} \to 1$ and the theorem follows.

B. Proofs for regular background graph

Proof of Theorem 5: Let $D^* = \text{diag} \{d^*_i\}$ with $d^*_i$ defined as (13). It suffices to show that $S^* = D^* - A$ satisfies the conditions in Lemma 1 with probability tending to one. We have shown that $S^* \mathbf{1} = 0$ in the proof of Theorem 3. It remains to verify that $S^* \geq 0$ and $\lambda_2(S^*) > 0$ with probability converging to one, which amounts to showing that

$$\mathbb{P}\left\{ \inf_{x \perp \mathbf{1}, \|x\|_2 = 1} x^T S^* x > 0 \right\} \to 1. \tag{17}$$

Recall that $A_G$ is the adjacency matrix of $G$ and $Y^* = \mathbf{1}^T \sigma^* (\sigma^*)^T$. Then $\mathbb{E}\{A\} = (1 - 2\epsilon) A_G \circ Y^*$, where $\circ$ denotes the element-wise matrix product. Thus for any $x$ such that $x \perp \sigma^*$ and $\|x\|_2 = 1$, letting $y \in \mathbb{R}^n$ such that $y_i = x_i \sigma_i^*$, we have $x^T S^* x = x^T D^* x - x^T \mathbb{E}\{A\} x - x^T (A - \mathbb{E}\{A\}) x \\
\geq x^T D^* x - (1 - 2\epsilon) x^T (A_G \circ Y^*) x - \|A - \mathbb{E}\{A\}\| x \\
= x^T D^* x - (1 - 2\epsilon) y^T A_G y - \|A - \mathbb{E}\{A\}\| \\
\geq \min_{i \in [n]} d^*_i - (1 - 2\epsilon) \lambda_2 - \|A - \mathbb{E}\{A\}\|, \tag{18}$$

where (a) holds by the definition of $y$; (b) holds due to the definition of $\lambda_2$ in (7) and the fact that $\{y, 1\} = 0$ and $\|y\|_2 = 1$.

It follows from [3, Corollary 3.2] that $\|A - \mathbb{E}\{A\}\| \leq c' \sqrt{\log n}$ with probability tending to one for a positive constant $c'$ depending only on $a$. Let $X_1 \overset{i.i.d.}{\sim} (1 - \epsilon) \delta_{\lambda_1} + \epsilon \delta_{\lambda_2}$ and $Z \sim \text{Binom}(d, \epsilon)$. Then for all $i \in [n]$, $d_i^*$ has the same distribution as $\sum_{j=1}^d X_j$, or equivalently, $d - 2Z$. It follows from Chernoff’s bound that

$$\mathbb{P}\left\{ \sum_{i=1}^d X_i \leq \frac{\log n}{\log \log n} + (1 - 2\epsilon) d \lambda_2 \right\} = \mathbb{P}\left\{ Z \geq \frac{1}{2} \left( d - \frac{\log n}{\log \log n} - (1 - 2\epsilon) d \lambda_2 \right) \right\} \leq \exp \left( -dD \left( \frac{1 - \lambda_2}{2} + \epsilon \lambda_2 - \frac{a}{\log \log n} \epsilon \right) \right) = n^{-dD(1 - \lambda_2)/2 + \epsilon \lambda_2} \exp(-o(1)).$$

Applying the union bound implies that $\min_{i \in [n]} d_i^* \geq \log n / \log \log n + (1 - 2\epsilon) d \lambda_2$ holds with probability at least $1 - n^{-1 \cdot d} D(1 - \lambda_2)/2 + \epsilon \lambda_2 \exp(-o(1))$. It follows from the assumption $\epsilon D((1 - \lambda_2)/2 + \epsilon \lambda_2) \|\epsilon\| > 1$ and (18) that the desired (17) holds, completing the proof.

References


