Semidefinite Programming for Community Detection With Side Information

Mohammad Esmaeili, Hussein Metwaly Saad, *Member, IEEE*, and Aria Nosratinia, *Fellow, IEEE*

***Abstract*—This paper produces an efficient semidefinite programming (SDP) solution for community detection that incorporates non-graph data, which in this context is known as side information. SDP is an efficient solution for standard community detection on graphs. We formulate a semi-definite relaxation for the maximum likelihood estimation of node labels, subject to observing *both* graph and non-graph data. This formulation is distinct from the SDP solution of standard community detection, but maintains its desirable properties. We calculate the exact recovery threshold for three types of non-graph information, which in this paper are called side information: partially revealed labels, noisy labels, as well as multiple observations (features) per node with arbitrary but finite cardinality. We find that SDP has the same exact recovery threshold in the presence of side information as maximum likelihood with side information. Thus, the methods developed herein are computationally efficient as well as asymptotically accurate for the solution of community detection in the presence of side information. Simulations show that the asymptotic results of this paper can also shed light on the performance of SDP for graphs of modest size.**

***Index Terms*—Community Detection, SDP, Stochastic Block Model, Censored Block Model, Side Information.**

I. Introduction

DETECTING communities (or clusters) in graphs is a fundamental problem that has many applications, such as finding like-minded people in social networks [1], and improving recommendation systems [2]. Community detection is affiliated with various problems in network science such as network structure reconstruction [3], networks with dynamic interactions [4], and complex networks [5]. Random graph models [6], [7] are frequently used in the analysis of community detection, prominent examples of which include the stochastic block model [7]–[9] and the censored block model [10], [11]. In the context of these models, community detection recovers latent node labels (communities) by observing the edges of a graph.

Manuscript received March 25, 2020; revised January 5, 2021 and March 31, 2021; accepted May 2, 2021. Date of publication May 10, 2021; date of current version July 7, 2021. This work was supported by National Science Foundation under Grant 2008684. Recommended for acceptance by Dr. Zhen Wang. *(Corresponding author: Aria Nosratinia.)*

Mohammad Esmaeili and Aria Nosratinia are with the Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75083-0688 USA (e-mail: esmaeili@utdallas.edu; aria@utdallas.edu).

Hussein Metwaly Saad is with the Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: hussein.saad@utdallas.edu).

Digital Object Identifier 10.1109/TNSE.2021.3078612

Community detection utilizes several metrics for residual error as the size of the graph grows: correlated recovery [12]–[15] (recovering the hidden community better than random guessing), weak recovery [16]–[18] (the fraction of misclassified labels in the graph vanishes with probability converging to one), and exact recovery [9], [19], [20] (all nodes are classified correctly with high probability). Recovery techniques include spectral methods [9], [21], belief propagation [22], and SDP relaxation [23].

Semidefinite programming is a computationally efficient convex optimization technique that has shown its utility in solving signal processing problems [24], [25]. In the context of community detection, SDP was introduced in [26], where it was used for solving a minimum bisection problem, obtaining a sufficient condition that is not optimal. In [27], a SDP relaxation was considered for a maximum bisection problem. For the binary symmetric stochastic block model, [28] showed that the SDP relaxation of maximum likelihood can achieve the optimal exact recovery threshold with high probability. These results were later extended to more general models in [29]. Also, [30] showed the power of SDP for solving a community detection problem in graphs with a secondary latent variable for each node.

Community detection on graphs has been widely studied in part because the graph structure is amenable to analysis and admits efficient algorithms. In practice, however, the available information for inference is often not purely graphical. For instance, in a citation network, beside the names of authors, there are some additional *non-graph* information such as keywords and abstract that can be used and improve the performance of community detection algorithms. For illustration, consider public-domain libraries such as Citeseer and Pubmed. Citation networks in these libraries have been the subject of several community detection studies, which can be augmented by incorporating individual (non-graph) attributes of the documents that affect the likelihood of community memberships.

The non-graph data assisting in the solution of graph problems is called *side information.* In [31], [32], the effect of side information on the phase transition of the exact recovery was studied for the binary symmetric stochastic block model. In [33]–[35], the effect of side information was studied on the phase transition of the weak and exact recovery as well as the phase transition of belief propagation in the single community stochastic block model. The impact of side information on the performance of belief propagation was further studied in [34], [36].

The contribution of this paper is the analysis of the impact of side information on SDP solutions for community detection. More specifically, we study the behavior of the SDP detection threshold under the exact recovery metric. We consider graphs following the binary censored block model and the binary symmetric stochastic block model. We begin with the development of SDP for partially-revealed labels and noisy labels, which are easier to grasp and visualize. This builds intuition for the more general setting, in which we study side information with multiple features per node, each of which is a random variable with arbitrary but finite cardinality. The former results also facilitate the understanding and interpretation of the latter. Most categories of side information give rise to a complete quadratic form in the likelihood function, which presents challenges in the analysis of their semidefinite programming relaxation. Overcoming these challenges is one of the main technical contributions of the present work.

Simulation results show that the thresholds calculated in this paper can also shed light on the understanding of the behavior of SDP in graphs of modest size.

Notation: Matrices and vectors are denoted by capital letters, and their elements with small letters. $\mathbf{I}$ is the identity matrix and $\mathbf{J}$ the all-one matrix. $S \succeq 0$ indicates a positive semidefinite matrix and $S \geq 0$ a matrix with non-negative entries. $||S||$ is the spectral norm, $\lambda_2(S)$ the second smallest eigenvalue (for a symmetric matrix), and $\langle \cdot, \cdot \rangle$ is the inner product. We abbreviate $[n] \triangleq \{1, \ldots, n\}$. Probabilities are denoted by $\mathbb{P}(\cdot)$ and random variables with Bernoulli and Binomial distribution are indicated by $\mathrm{Bern}(p)$ and $\mathrm{Binom}(n, p)$, respectively.

II. System Model

This paper analyzes community detection in the presence of a graph observation as well as individual node attributes. The graphs in this paper follow the binary stochastic block model and the censored block model, and side information is in the form of either partially revealed labels, noisy labels, or an alphabet other than the labels.

This paper considers a fully connected regime, guaranteeing that exact recovery is possible. Throughout this paper, the graph adjacency matrix is denoted by G. Node labels are independent and identically distributed across n, with labels $+1$ and -1. The vector of node labels is denoted by X, and a corresponding vector of side information is denoted by Y. The log-likelihood of the graph and side information is

$$\log \mathbb{P}(G, Y|X) = \log \mathbb{P}(G|X) + \log \mathbb{P}(Y|X),$$

i.e., G and Y are independent given X.

A. Binary Censored Block Model

The model consists of an Erdős-Rényi graph with n nodes and edge probability $p = a\frac{\log n}{n}$ for a fixed $a > 0$. The nodes belong to two communities represented by the binary node labels, which are latent. The entries $G_{ij} \in \{-1, 0, 1\}$ of the weighted adjacency matrix of the graph have a distribution that depends on the community labels x_i and x_j as follows:

$$G_{ij} \sim \begin{cases} p(1-\xi)\delta_{+1} + p\xi\delta_{-1} + (1-p)\delta_0 & \text{when } x_i = x_j \\ p(1-\xi)\delta_{-1} + p\xi\delta_{+1} + (1-p)\delta_0 & \text{when } x_i \neq x_j \end{cases}$$

where δ is Dirac delta function and $\xi \in [0, \frac{1}{2}]$ is a constant. Further, $G_{ii} = 0$ and $G_{ij} = G_{ji}$. For all $j > i$, the edges G_{ij} are mutually independent conditioned on the node labels. The log-likelihood of G is

$$\log \mathbb{P}(G|X) = \frac{1}{4} T_1 X^T G X + C_1, \tag{1}$$

where $T_1 \triangleq \log\left(\frac{1-\xi}{\xi}\right)$ and C_1 is a deterministic scalar.

B. Binary Symmetric Stochastic Block Model

In this model, if nodes i, j belong to the same community, $G_{i,j} \sim \mathrm{Bern}(p)$, otherwise $G_{ij} \sim \mathrm{Bern}(q)$ with

$$p = a\frac{\log n}{n}, \qquad q = b\frac{\log n}{n},$$

and $a \geq b > 0$. Then the log-likelihood of G is

$$\log \mathbb{P}(G|X) = \frac{1}{4} T_1 X^T G X + C_2, \tag{2}$$

where $T_1 \triangleq \log\left(\frac{p(1-q)}{q(1-p)}\right)$ and C_2 is a deterministic scalar.

C. Side Information: Partially Revealed Labels

Partially-revealed side information vector Y consists of elements that with probability $1-\epsilon$ are equal to the true label and with probability ϵ take value 0, i.e., are erased.

Conditioned on each node label, the corresponding side information is assumed independent from other labels and from the graph edges. Thus, the log-likelihood of Y is

$$\log \mathbb{P}(Y|X) = Y^T Y \log\left(\frac{1-\epsilon}{\epsilon}\right) + n\log(\epsilon). \tag{3}$$

D. Side Information: Noisy Labels

Noisy-label side information vector Y consists of elements that with probability $1-\alpha$ agree with the true label ($y_i = x_i^*$) and with probability α are erroneous ($y_i = -x_i^*$), where $\alpha \in (0, 0.5)$. Then the log-likelihood of Y is

$$\log \mathbb{P}(Y|X) = \frac{1}{2} T_2 X^T Y + T_2 \frac{n}{2} + n\log\alpha, \tag{4}$$

where $T_2 \triangleq \log(\frac{1-\alpha}{\alpha})$.

E. Side Information: Multiple Variables & Larger Alphabets

In this model, we disengage the cardinality of side information alphabet from the node latent variable, and also allow for more side information random variables per node. This is motivated by practical conditions where the available non-graph information may be different from the node latent variable, and there may be multiple types of side information with varying utility for the inference.

Formally, $y_{i,k}$ is the random variable representing feature k at node i. Each feature has cardinality M_k that is finite and fixed across the graph. We group these variables into a vector y_i of dimension K, representing side information for node i, and group the vectors into a matrix Y representing all side information for the graph.[1]

Without loss of generality, the alphabet of each feature k is the set of integers $\{1, \dots, M_k\}$. The posterior probability of the features are denoted by

$$\alpha^k_{+,m_k} \triangleq \mathbb{P}(y_{i,k} = m_k | x_i = 1),$$
$$\alpha^k_{-,m_k} \triangleq \mathbb{P}(y_{i,k} = m_k | x_i = -1),$$

where m_k indexes the alphabet of feature k. Then the log-likelihood of Y is

$$\begin{aligned}\log \mathbb{P}(Y|X) &= \sum_{i=1}^{n} \log \mathbb{P}(y_i|x_i) \\ &= \frac{1}{2}\sum_{i=1}^{n} x_i \sum_{k=1}^{K}\sum_{m_k=1}^{M_k} \mathbb{1}_{y_{i,k}=m_k} \log\left(\frac{\alpha^k_{+,mk}}{\alpha^k_{-,m_k}}\right) \\ &\quad + \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{m_k=1}^{M_k} \mathbb{1}_{y_{i,k}=m_k} \log\left(\alpha^k_{+,m_k}\alpha^k_{-,m_k}\right),\end{aligned}$$

where $\mathbb{1}$ is the indicator function. Define

$$\tilde{y}_i \triangleq \sum_{k=1}^{K}\sum_{m_k=1}^{M_k} \mathbb{1}_{\{y_{i,k}=m_k\}} \log\left(\frac{\alpha^k_{+,m_k}}{\alpha^k_{-,m_k}}\right),$$

and $\tilde{Y} \triangleq [\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n]^T$. Then the log-likelihood of Y is

$$\log \mathbb{P}(Y|X) = \frac{1}{2} X^T \tilde{Y} + C_3, \tag{5}$$

for some constant C_3. In the remainder of this paper, side information thus defined is referred to as *general side information*.

III. Detection Via SDP

For organizational convenience, the main results of the paper are concentrated in this section.

For the formulation of SDP, we utilize the additional variables $Z \triangleq XX^T$ and $W \triangleq YY^T$. Also, let $Z^* \triangleq X^* X^{*T}$.

A. Censored Block Model With Partially Revealed Labels

Combining (1) and (3), the maximum likelihood detector is

$$\begin{aligned}\hat{X} = \arg\max_X \;& X^T G X \\ \text{subject to} \;& x_i \in \{\pm 1\}, \quad i \in [n] \\ & X^T Y = Y^T Y,\end{aligned} \tag{6}$$

where the constraint $X^T Y = Y^T Y$ ensures that detected values agree with available side information. This is a non-convex problem, therefore we consider a convex relaxation [19], [37]. Replacing $x_i \in \{\pm 1\}$ with $Z_{ii} = 1$, and $X^T Y = \pm Y^T Y$ with $\langle Z, W\rangle = (Y^T Y)^2$,

[1] If vectors y_i have unequal dimension, matrix Y will accommodate the largest vector, producing vacant entries that are defaulted to zero.

$$\begin{aligned}\widehat{Z} = \arg\max_Z \;& \langle Z, G\rangle \\ \text{subject to} \;& Z = XX^T \\ & Z_{ii} = 1, \quad i \in [n] \\ & \langle Z, W\rangle = (Y^T Y)^2.\end{aligned} \tag{7}$$

By relaxing the rank-one constraint introduced via Z, we obtain the following SDP relaxation:

$$\begin{aligned}\widehat{Z} = \arg\max_Z \;& \langle Z, G\rangle \\ \text{subject to} \;& Z \succeq 0 \\ & Z_{ii} = 1, \quad i \in [n] \\ & \langle Z, W\rangle = (Y^T Y)^2.\end{aligned} \tag{8}$$

Let $\beta \triangleq \lim_{n\to\infty} -\frac{\log \epsilon}{\log n}$, where $\beta \geq 0$.

Theorem 1: Under the binary censored block model and partially revealed labels, if

$$a(\sqrt{1-\xi} - \sqrt{\xi})^2 + \beta \, > \, 1,$$

then the SDP estimator is asymptotically optimal, i.e., $\mathbb{P}(\widehat{Z} = Z^*) \geq 1 - o(1)$.

Proof: See Appendix A. ■

Theorem 2: Under the binary censored block model and partially revealed labels, if

$$a(\sqrt{1-\xi} - \sqrt{\xi})^2 + \beta \, < \, 1,$$

then for any sequence of estimators $\widehat{Z}_n$, $\mathbb{P}(\widehat{Z}_n = Z^*) \to 0$ as $n \to \infty$.

Proof: See Appendix B. ■

B. Censored Block Model With Noisy Labels

Combining (1) and (4), the maximum likelihood detector is

$$\begin{aligned}\hat{X} = \arg\max_X \;& T_1 X^T G X + 2T_2 X^T Y \\ \text{subject to} \;& x_i \in \{\pm 1\}, \quad i \in [n].\end{aligned} \tag{9}$$

Then (9) is equivalent to

$$\begin{aligned}\widehat{Z} = \arg\max_{Z,X} \;& T_1\langle G, Z\rangle + 2T_2 X^T Y \\ \text{subject to} \;& Z = XX^T \\ & Z_{ii} = 1, \quad i \in [n].\end{aligned} \tag{10}$$

Relaxing the rank-one constraint, using

$$Z - XX^T \succeq 0 \Leftrightarrow \begin{bmatrix} 1 & X^T \\ X & Z \end{bmatrix} \succeq 0,$$

yields the SDP relaxation of (10):

$$\begin{aligned}\widehat{Z} = \arg\max_{Z,X} \; & T_1\langle G, Z\rangle + 2T_2 X^T Y \\ \text{subject to} \; & \begin{bmatrix} 1 & X^T \\ X & Z \end{bmatrix} \succeq 0 \\ & Z_{ii} = 1, \quad i \in [n]. \end{aligned} \tag{11}$$

Let $\beta \triangleq \lim_{n\to\infty} \frac{T_2}{\log n}$, where $\beta \geq 0$. Also, for convenience define

$$\eta(a,\beta) \triangleq a - \frac{\gamma}{T_1} + \frac{\beta}{2T_1}\log\left(\frac{(1-\xi)(\gamma+\beta)}{\xi(\gamma-\beta)}\right),$$

where $\gamma \triangleq \sqrt{\beta^2 + 4\xi(1-\xi)a^2T_1^2}$.

Theorem 3: Under the binary censored block model and noisy labels, if

$$\begin{cases} \eta(a,\beta) > 1 & \text{when } 0 \leq \beta < aT_1(1-2\xi) \\ \beta > 1 & \text{when } \beta \geq aT_1(1-2\xi) \end{cases}$$

then the SDP estimator is asymptotically optimal, i.e., $\mathbb{P}(\widehat{Z} = Z^*) \geq 1 - o(1)$.

Proof: See Appendix C. ■

Theorem 4: Under the binary censored block model and noisy labels, if

$$\begin{cases} \eta(a,\beta) < 1 & \text{when } 0 \leq \beta < aT_1(1-2\xi) \\ \beta < 1 & \text{when } \beta \geq aT_1(1-2\xi) \end{cases}$$

then for any sequence of estimators $\widehat{Z}_n$, $\mathbb{P}(\widehat{Z}_n = Z^*) \to 0$ as $n \to \infty$.

Proof: See Appendix D. ■

C. Censored Block Model With General Side Information

Combining (1) and (5), the SDP relaxation is

$$\begin{aligned}\widehat{Z} = \arg\max_{Z,X} \; & T_1\langle G, Z\rangle + 2X^T\tilde{Y} \\ \text{subject to} \; & \begin{bmatrix} 1 & X^T \\ X & Z \end{bmatrix} \succeq 0 \\ & Z_{ii} = 1, \quad i \in [n]. \end{aligned} \tag{12}$$

The log-likelihoods and the log-likelihood-ratio of side information, combined over all features, are as follows:

$$f_1(n) \triangleq \sum_{k=1}^{K} \log \frac{\alpha^k_{+,m_k}}{\alpha^k_{-,m_k}},$$

$$f_2(n) \triangleq \sum_{k=1}^{K} \log \alpha^k_{+,m_k},$$

$$f_3(n) \triangleq \sum_{k=1}^{K} \log \alpha^k_{-,m_k}.$$

Two exponential orders will feature prominently in the following results and proofs:

$$\beta_1 \triangleq \lim_{n\to\infty} \frac{f_1(n)}{\log n},$$

$$\beta \triangleq \lim_{n\to\infty} -\frac{\max(f_2(n), f_3(n))}{\log n}.$$

Although the definition of β varies in the context of different models, its role remains the same. In each case, β is a parameter representing the asymptotic quality of side information.[2]

Theorem 5: Under the binary censored block model and general side information, if

$$\begin{cases} \eta(a,|\beta_1|) + \beta > 1 & \text{when } |\beta_1| \leq aT_1(1-2\xi) \\ |\beta_1| + \beta > 1 & \text{when } |\beta_1| > aT_1(1-2\xi) \end{cases}$$

then the SDP estimator is asymptotically optimal, i.e., $\mathbb{P}(\widehat{Z} = Z^*) \geq 1 - o(1)$.

Proof: See Appendix E. ■

Theorem 6: Under the binary censored block model and general side information, if

$$\begin{cases} \eta(a,|\beta_1|) + \beta < 1 & \text{when } |\beta_1| \leq aT_1(1-2\xi) \\ |\beta_1| + \beta < 1 & \text{when } |\beta_1| > aT_1(1-2\xi) \end{cases}$$

then for any sequence of estimators $\widehat{Z}_n$, $\mathbb{P}(\widehat{Z}_n = Z^*) \to 0$.

Proof: See Appendix F. ■

D. Stochastic Block Model With Partially Revealed Labels

Similar to the binary censored block model with partially revealed labels, by combining (2) and (3), the SDP relaxation is

$$\begin{aligned}\widehat{Z} = \arg\max_{Z} \; & \langle Z, G\rangle \\ \text{subject to} \; & Z \succeq 0 \\ & Z_{ii} = 1, \quad i \in [n] \\ & \langle \mathbf{J}, Z\rangle = 0 \\ & \langle Z, W\rangle = (Y^TY)^2, \end{aligned} \tag{13}$$

where the constraint $\langle \mathbf{J}, Z\rangle = 0$ arises from two equal-sized communities.

Theorem 7: Under the binary symmetric stochastic block model and partially revealed labels, if

$$\left(\sqrt{a} - \sqrt{b}\right)^2 + 2\beta > 2,$$

then the SDP estimator is asymptotically optimal, i.e., $\mathbb{P}(\widehat{Z} = Z^*) \geq 1 - o(1)$.

Proof: See Appendix G. ■

Remark 1: The converse is given by [31, Theorem 3].

E. Stochastic Block Model With Noisy Labels

Similar to the binary censored block model with noisy labels, by combining (2) and (4), the SDP relaxation is

[2] In each case, β is proportional to the exponential order of the likelihood function.

$$\begin{aligned}\widehat{Z} = \arg\max_{Z,X} \; & T_1\langle G, Z\rangle + 2T_2 X^T Y \\ \text{subject to} \; & \begin{bmatrix} 1 & X^T \\ X & Z \end{bmatrix} \succeq 0 \\ & Z_{ii} = 1, \quad i \in [n] \\ & \langle \mathbf{J}, Z\rangle = 0. \end{aligned} \tag{14}$$

For convenience let

$$\eta(a,b,\beta) \triangleq \frac{a+b}{2} + \frac{\beta}{2} - \frac{\gamma}{T_1} + \frac{\beta}{2T_1}\log\left(\frac{\gamma+\beta}{\gamma-\beta}\right),$$

where $\gamma \triangleq \sqrt{\beta^2 + abT_1^2}$.

Theorem 8: Under the binary symmetric stochastic block model and noisy label side information, if

$$\begin{cases} \eta(a,b,\beta) > 1 & \text{when } 0 \le \beta < \frac{T_1}{2}(a-b) \\ \beta > 1 & \text{when } \beta \ge \frac{T_1}{2}(a-b) \end{cases}$$

then the SDP estimator is asymptotically optimal, i.e., $\mathbb{P}(\widehat{Z} = Z^*) \ge 1 - o(1)$.

Proof: See Appendix H. ∎

Remark 2: The converse is given by [31, Theorem 2].

F. Stochastic Block Model With General Side Information

Similar to the binary censored block model with general side information, by combining (2) and (5), the SDP relaxation is

$$\begin{aligned}\widehat{Z} = \arg\max_{Z,X} \; & T_1\langle G, Z\rangle + 2X^T \tilde{Y} \\ \text{subject to} \; & \begin{bmatrix} 1 & X^T \\ X & Z \end{bmatrix} \succeq 0 \\ & Z_{ii} = 1, \quad i \in [n] \\ & \langle \mathbf{J}, Z\rangle = 0. \end{aligned} \tag{15}$$

Theorem 9: Under the binary symmetric stochastic block model and general side information, if

$$\begin{cases} \eta(a,b,|\beta_1|) + \beta > 1 & \text{when } |\beta_1| \le T_1\frac{(a-b)}{2} \\ |\beta_1| + \beta > 1 & \text{when } |\beta_1| > T_1\frac{(a-b)}{2} \end{cases}$$

then the SDP estimator is asymptotically optimal, i.e., $\mathbb{P}(\widehat{Z} = Z^*) \ge 1 - o(1)$.

Proof: See Appendix I. ∎

Remark 3: The converse is given by [31, Theorem 5].

IV. Numerical Results

This section produces numerical simulations that shed light on the domain of applicability of the asymptotic results obtained earlier in the paper[3].

[3] The code is available online at https://github.com/mohammadesmaeili/Community-Detection-by-SDP

TABLE I
SDP With Partially Revealed Labels

a	b	ξ	β	n	Error Probability
3	1	-	0.2	100	4.1×10^{-2}
3	1	-	0.2	200	3.1×10^{-2}
3	1	-	0.2	300	2.5×10^{-2}
3	1	-	0.2	400	2.2×10^{-2}
3	1	-	0.2	500	1.9×10^{-2}
3	1	-	0.8	100	5.0×10^{-4}
3	1	-	0.8	200	3.2×10^{-4}
3	1	-	0.8	300	1.6×10^{-4}
3	1	-	0.8	400	1.2×10^{-4}
3	1	-	0.8	500	9.3×10^{-5}
1	-	0.2	0.3	100	4.1×10^{-2}
1	-	0.2	0.3	200	2.9×10^{-2}
1	-	0.2	0.3	300	2.2×10^{-2}
1	-	0.2	0.3	400	1.9×10^{-2}
1	-	0.2	0.3	500	1.7×10^{-2}
1	-	0.2	1	100	1.1×10^{-3}
1	-	0.2	1	200	4.2×10^{-4}
1	-	0.2	1	300	2.7×10^{-4}
1	-	0.2	1	400	2.1×10^{-4}
1	-	0.2	1	500	1.5×10^{-4}

Table I shows the misclassification error probability of the SDP estimators (8) and (13) with partially revealed side information. Under the binary stochastic block model with $a = 3$ and $b = 1$, when the side information $\beta = 0.8$, error probability diminishes with n as predicted by earlier asymptotic results. For these parameters, $\eta = 1.1 > 1$, and exact recovery is possible based on the theoretical results. When $\beta = 0.2$, then $\eta = 0.5 < 1$ which does not fall in the asymptotic perfect recovery regime, the misclassification error probability is much higher. Under the binary censored block model with $a = 1$ and $\xi = 0.2$, when the side information $\beta = 1$, error probability diminishes with n. For these values, $\eta = 1.2 > 1$, and exact recovery is possible based on the theoretical results. When $\beta = 0.3$, the misclassification error probability is much higher. For this value of β, $\eta = 0.5 < 1$ which means exact recovery is not asymptotically possible.

Table II shows the misclassification error probability of the SDP estimators (11) and (14) with noisy labels side information. Under the stochastic block model with $a = 4$ and $b = 1$, when the side information $\beta = 1$, then $\eta = 1.1 > 1$ and the error probability diminishes with n as predicted by earlier theoretical results. When $\beta = 0.2$, then $\eta = 0.6 < 1$ which does not fall in the asymptotic perfect recovery regime. For this case the misclassification error is much higher. Under the censored block model with $a = 4$ and $\xi = 0.25$, when the side information $\beta = 1.1$, then $\eta = 1.2 > 1$ and the error probability diminishes with n. When $\beta = 0.1$, then $\eta = 0.6 < 1$ which means that exact recovery is not possible asymptotically. For this value of β and a finite n, the misclassification error is not negligible.

For comparison, Table III shows the misclassification error probability of the SDP estimator *without* side information, i.e., $\beta = 0$. Under the binary stochastic block model, when $a = 3$ ($a = 4$) and $b = 1$, it is seen that the error probability increases in comparison with the corresponding error probability in Table I (Table II) where side information is available. Also,

TABLE II
SDP WITH NOISY LABELS

a	b	ξ	β	n	Error Probability
4	1	-	0.2	100	2.0×10^{-2}
4	1	-	0.2	200	1.5×10^{-2}
4	1	-	0.2	300	1.3×10^{-2}
4	1	-	0.2	400	1.1×10^{-2}
4	1	-	0.2	500	1.0×10^{-2}
4	1	-	1	100	1.1×10^{-3}
4	1	-	1	200	7.4×10^{-4}
4	1	-	1	300	3.0×10^{-5}
4	1	-	1	400	2.7×10^{-5}
4	1	-	1	500	2.2×10^{-5}
4	-	0.25	0.1	100	2.9×10^{-2}
4	-	0.25	0.1	200	1.8×10^{-2}
4	-	0.25	0.1	300	1.4×10^{-2}
4	-	0.25	0.1	400	1.2×10^{-2}
4	-	0.25	0.1	500	1.0×10^{-2}
4	-	0.25	1.1	100	2.7×10^{-3}
4	-	0.25	1.1	200	1.0×10^{-3}
4	-	0.25	1.1	300	6.2×10^{-4}
4	-	0.25	1.1	400	4.1×10^{-4}
4	-	0.25	1.1	500	3.3×10^{-4}

TABLE III
SDP WITHOUT SIDE INFORMATION

a	b	ξ	n	Error Probability
3	1	-	100	1.4×10^{-1}
3	1	-	200	1.2×10^{-1}
3	1	-	300	1.1×10^{-1}
3	1	-	400	9.8×10^{-2}
3	1	-	500	9.1×10^{-2}
4	1	-	100	2.3×10^{-2}
4	1	-	200	1.7×10^{-2}
4	1	-	300	1.6×10^{-2}
4	1	-	400	1.3×10^{-2}
4	1	-	500	1.2×10^{-2}
1	-	0.2	100	2.9×10^{-1}
1	-	0.2	200	2.5×10^{-1}
1	-	0.2	300	2.2×10^{-1}
1	-	0.2	400	2.1×10^{-1}
1	-	0.2	500	1.9×10^{-1}
4	-	0.25	100	3.0×10^{-2}
4	-	0.25	200	1.9×10^{-2}
4	-	0.25	300	1.5×10^{-2}
4	-	0.25	400	1.2×10^{-2}
4	-	0.25	500	1.1×10^{-2}

under the binary censored block model, when $a = 1$ and $\xi = 0.2$ ($a = 4$ and $\xi = 0.25$), it is seen that the error probability increases in comparison with the corresponding error probability in Table I (Table II) where side information is available.

Using standard arguments form numerical linear algebra, the computational complexity of the algorithms in this paper are on the order $O(mn^3 + m^2n^2)$, where n is the number of nodes in the graph, and m is a small constant, typically between 2 to 4, indicating assumptions of the problem that manifest as constraints in the optimization.

V. CONCLUSION

This paper calculated the exact recovery threshold for community detection under SDP with several types of side information. Among other insights, our results indicate that in the presence of side information, the exact recovery threshold for SDP and for maximum likelihood detection remain identical. We anticipate that models and methods of this paper may be further extended to better match the statistics of real-world graph data.

APPENDIX A
PROOF OF THEOREM 1

We begin by stating sufficient conditions for the optimum solution of (8) matching the true labels X^*.

Lemma 1: For the optimization problem (8), consider the Lagrange multipliers

$$\mu^*, \quad D^* = \mathrm{diag}(d_i^*), \quad S^*.$$

If we have

$$\begin{aligned} &S^* = D^* + \mu^* W - G, \\ &S^* \succeq 0, \\ &\lambda_2(S^*) > 0, \\ &S^* X^* = 0, \end{aligned}$$

then (μ^*, D^*, S^*) is the dual optimal solution and $\widehat{Z} = X^* X^{*T}$ is the unique primal optimal solution of (8).

Proof: The Lagrangian of (8) is given by

$$\begin{aligned} L(Z, S, D, \mu) = \langle G, Z\rangle + \langle S, Z\rangle - \langle D, Z - \mathbf{I}\rangle \\ - \mu(\langle W, Z\rangle - (Y^T Y)^2), \end{aligned}$$

where $S \succeq 0$, $D = \mathrm{diag}(d_i)$, and $\mu \in \mathbb{R}$ are Lagrange multipliers. For any Z that satisfies the constraints in (8),

$$\begin{aligned} \langle G, Z\rangle &\overset{(a)}{\leq} L(Z, S^*, D^*, \mu^*) \\ &= \langle D^*, \mathbf{I}\rangle + \mu^*(Y^T Y)^2 \\ &\overset{\text{(b)}}{=} \langle D^*, Z^*\rangle + \mu^*(Y^T Y)^2 \\ &= \langle G + S^* - \mu^* W, Z^*\rangle \\ &\quad + \mu^*(Y^T Y)^2 \\ &\overset{\text{(c)}}{=} \langle G, Z^*\rangle, \end{aligned}$$

where (a) holds because $\langle S^*, Z\rangle \geq 0$, (b) holds because $Z_{ii} = 1$ for all $i \in [n]$, and (c) holds because $\langle S^*, Z^*\rangle = X^{*T} S^* X^* = 0$ and $\langle W, Z^*\rangle = (Y^T Y)^2$. Therefore, Z^* is a primal optimal solution. Now, we will establish the uniqueness of the optimal solution. Assume $\tilde{Z}$ is another primal optimal solution. Then

$$\begin{aligned} \langle S^*, \tilde{Z}\rangle &= \langle D^* - G + \mu^* W, \tilde{Z}\rangle \\ &= \langle D^*, \tilde{Z}\rangle - \langle G, \tilde{Z}\rangle + \mu^*\langle W, \tilde{Z}\rangle \\ &\overset{\text{(a)}}{=} \langle D^*, Z^*\rangle - \langle G, Z^*\rangle + \mu^*\langle W, Z^*\rangle \\ &= \langle D^* - G + \mu^* W, Z^*\rangle \\ &= \langle S^*, Z^*\rangle = 0, \end{aligned}$$

where (a) holds because $\langle W, Z^*\rangle = \langle W, \tilde{Z}\rangle = (Y^T Y)^2$, $\langle G, Z^*\rangle = \langle G, \tilde{Z}\rangle$, and $Z^*_{ii} = \tilde{Z}_{ii} = 1$ for all $i \in [n]$. Since $\tilde{Z} \succeq 0$ and

$S^* \succeq 0$ while its second smallest eigenvalue $\lambda_2(S^*)$ is positive, $\tilde{Z}$ must be a multiple of Z^*. Also, since $\tilde{Z}_{ii} = Z^*_{ii} = 1$ for all $i \in [n]$, we have $\tilde{Z} = Z^*$. ■

We now show that $S^* = D^* + \mu^* W - G$ satisfies other conditions in Lemma 1 with probability $1 - o(1)$. Let

$$d_i^* = \sum_{j=1}^{n} G_{ij} x_j^* x_i^* - \mu^* \sum_{j=1}^{n} y_i y_j x_j^* x_i^*. \tag{16}$$

Then $D^* X^* = GX^* - \mu^* W X^*$ and based on the definition of S^* in Lemma 1, S^* satisfies the condition $S^* X^* = 0$. It remains to show that $S^* \succeq 0$ and $\lambda_2(S^*) > 0$ with probability $1 - o(1)$. In other words, we need to show that

$$\mathbb{P}\left(\inf_{V \perp X^*, \|V\|=1} V^T S^* V > 0 \right) \geq 1 - o(1), \tag{17}$$

where V is a vector of length n. Since for the binary censored block model

$$\mathbb{E}[G] = p(1-2\xi)(X^* X^{*T} - \mathbf{I}), \tag{18}$$

it follows that for any V such that $V^T X^* = 0$ and $\|V\| = 1$,

$$V^T S^* V = V^T D^* V + \mu^* V^T W V - V^T (G - \mathbb{E}[G]) V + p(1-2\xi).$$

Lemma 2: [29, Thoerem 9] For any $c > 0$, there exists $c' > 0$ such that for any $n \geq 1$, $\|G - \mathbb{E}[G]\| \leq c' \sqrt{\log n}$ with probability at least $1 - n^{-c}$.

Lemma 3: [38, Lemma 3]

$$\mathbb{P}\left(V^T W V \geq \sqrt{\log n} \right) \leq \frac{1-\epsilon}{\sqrt{\log n}} = n^{-\frac{1}{2} + o(1)}.$$

Since $V^T D^* V \geq \min_{i \in [n]} d_i^*$ and $V^T (G - \mathbb{E}[G]) V \leq \|G - \mathbb{E}[G]\|$, applying Lemmas 2 and 3 implies that with probability $1 - o(1)$,

$$V^T S^* V \geq \min_{i \in [n]} d_i^* + (\mu^* - c') \sqrt{\log n} + p(1-2\xi). \tag{19}$$

Lemma 4: Consider a sequence of i.i.d. random variables $\{S_1, \ldots, S_m\}$ with distribution $p(1-\xi)\delta_{+1} + p\xi\delta_{-1} + (1-p)\delta_0$. Let $U \sim \text{Binom}(n-1, 1-\epsilon)$, $\mu^* < 0$, and $\delta = \frac{\log n}{\log logn}$. Then

$$\mathbb{P}\left(\sum_{i=1}^{n-1} S_i \leq \delta \right) \leq n^{-a(\sqrt{1-\xi} - \sqrt{\xi})^2 + o(1)},$$

$$\mathbb{P}\left(\sum_{i=1}^{n-1} S_i - \mu^* U \leq \delta + \mu^* \right) \leq \epsilon^{n[\log \epsilon + o(1)]}.$$

Proof: It follows from Chernoff bound. ■

It can be shown that $\sum_{j=1}^{n} G_{ij} x_i^* x_j^*$ in (16) is equal in distribution to $\sum_{i=1}^{n-1} S_i$ in Lemma 4. Then

$$\begin{aligned} \mathbb{P}(d_i^* \leq \delta) &= \mathbb{P}\left(\sum_{j=1}^{n} G_{ij} x_i^* x_j^* \leq \delta \right) \epsilon \\ &\quad + \mathbb{P}\left(\sum_{j=1}^{n} G_{ij} x_i^* x_j^* - \mu^* Z_i \leq \delta + \mu^* \right) (1-\epsilon) \\ &\leq \epsilon n^{-a(\sqrt{1-\xi} - \sqrt{\xi})^2 + o(1)} + (1-\epsilon) \epsilon^{n(\log \epsilon + o(1))} \\ &= e^{\left(\frac{\log \epsilon}{\log n} - a(\sqrt{1-\xi} - \sqrt{\xi})^2 + o(1) \right) \log n}, \end{aligned}$$

where $Z_i \sim \text{Binom}(n-1, 1-\epsilon)$ and $(1-\epsilon)\epsilon^{n(\log \epsilon + o(1))}$ vanishes as $n \to \infty$. Recall that $\beta \triangleq \lim_{n \to \infty} -\frac{\log \epsilon}{\log n}$, where $\beta \geq 0$. Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\beta - a(\sqrt{1-\xi} - \sqrt{\xi})^2 + o(1)}.$$

Using the union bound,

$$\mathbb{P}\left(\min_{i \in [n]} d_i^* \geq \frac{\log n}{\log \log n} \right) \geq 1 - n^{1 - \beta - a(\sqrt{1-\xi} - \sqrt{\xi})^2 + o(1)}.$$

When $\beta + a(\sqrt{1-\xi} - \sqrt{\xi})^2 > 1$, $\min_{i \in [n]} d_i^* \geq \frac{\log n}{\log \log n}$ holds with probability $1 - o(1)$. Combining this result with (19), if $\beta + a(\sqrt{1-\xi} - \sqrt{\xi})^2 > 1$, then with probability $1 - o(1)$,

$$V^T S^* V \geq \frac{\log n}{\log \log n} + (\mu^* - c') \sqrt{\log n} + p(1-2\xi) > 0,$$

which concludes Theorem 1.

Appendix B
Proof of Theorem 2

Since the prior distribution of X^* is uniform, among all estimators, the maximum likelihood estimator minimizes the average error probability. Therefore, it suffices to show that with high probability the maximum likelihood estimator fails. Let

$$F \triangleq \left\{ \min_{i \in [n], y_i = 0} \sum_{j=1}^{n} G_{ij} x_j^* x_i^* \leq -1 \right\}.$$

Then $\mathbb{P}(\text{ML Fails}) \geq \mathbb{P}(F)$. If we show that $\mathbb{P}(\text{F}) \to 1$, the maximum likelihood estimator fails. Let H denote the set of first $\lfloor \frac{n}{\log^2 n} \rfloor$ nodes and $e(i, H)$ denote the number of edges between node i and nodes in the set H. Then

$$\begin{aligned} \min_{i \in [n], y_i = 0} \sum_{j=1}^{n} G_{ij} x_j^* x_i^* &\leq \min_{i \in H, y_i = 0} \sum_{j=1}^{n} G_{ij} x_j^* x_i^* \\ &\leq \min_{i \in H, y_i = 0} \sum_{j \in H^c} G_{ij} x_j^* x_i^* + \max_{i \in H, y_i = 0} e(i, H), \end{aligned}$$

Define the events

$$E_1 \triangleq \left\{ \max_{i \in H, y_i = 0} e(i, H) \leq \delta - 1 \right\},$$

$$E_2 \triangleq \left\{ \min_{i \in H, y_i = 0} \sum_{j \in H^c} G_{ij} x_j^* x_i^* \leq -\delta \right\}.$$

Notice that $F \supset E_1 \cap E_2$. Hence, to show that the maximum likelihood estimator fails, it suffices to show that $\mathbb{P}(E_1) \to 1$ and $\mathbb{P}(E_2) \to 1$.

Lemma 5: [39, Lemma 5] When $S \sim \text{Binom}(n, p)$, for any $r \geq 1$, $\mathbb{P}(S \geq rnp) \leq (\frac{e}{r})^{rnp} e^{-np}$.

Since $e(i, H) \sim \text{Binom}(|H|, a\frac{\log n}{n})$, it follows from Lemma 5 that

$$\begin{aligned}&\mathbb{P}(e(i,H) \geq \delta - 1, y_i = 0)\\ &\leq \epsilon\left(\frac{\log^2 n}{ae\log\log n} - \frac{\log n}{ae}\right)^{1-\frac{\log n}{\log\log n}} e^{-\frac{a}{\log n}} \leq \epsilon n^{-2+o(1)}.\end{aligned}$$

Using the union bound, $\mathbb{P}(E_1) \geq 1 - \epsilon n^{-1+o(1)}$. Thus, $\mathbb{P}(E_1) \to 1$.

Lemma 6: [29, Lemma 8] Consider a sequence of i.i.d. random variables $\{S_1, \ldots, S_m\}$ with distribution $p(1-\xi)\delta_{+1} + p\xi\delta_{-1} + (1-p)\delta_0$, where $m - n = o(n)$. Let $f(n) = \frac{\log n}{\log\log n}$. Then

$$\mathbb{P}\left(\sum_{i=1}^{m} S_i \leq -f(n)\right) \geq n^{-a(\sqrt{1-\xi}-\sqrt{\xi})^2+o(1)}.$$

Using Lemma 6 and since $\{\sum_{j\in H^c} G_{ij}x_j^* x_i^*\}_{i\in H}$ are mutually independent,

$$\begin{aligned}\mathbb{P}(E_2) &= 1 - \prod_{i\in H}\left[1 - \mathbb{P}\left(\sum_{j\in H^c} G_{ij}x_j^*x_i^* \leq -\delta, y_i = 0\right)\right]\\ &\geq 1 - \left[1 - \epsilon n^{-a(\sqrt{1-\xi}-\sqrt{\xi})^2+o(1)}\right]^{|H|}. \end{aligned} \tag{20}$$

Since $\beta = \lim_{n\to\infty} -\frac{\log \epsilon}{\log n}$, it follows from (20) that

$$\begin{aligned}\mathbb{P}(E_2) &\geq 1 - \left[1 - n^{-\beta-a(\sqrt{1-\xi}-\sqrt{\xi})^2+o(1)}\right]^{|H|}\\ &\geq 1 - \exp\left(-n^{1-\beta-a(\sqrt{1-\xi}-\sqrt{\xi})^2+o(1)}\right), \end{aligned} \tag{21}$$

using $1 + x \leq e^x$. From (21), if $a(\sqrt{1-\xi} - \sqrt{\xi})^2 + \beta < 1$, then $\mathbb{P}(E_2) \to 1$. Therefore, $\mathbb{P}(F) \to 1$ and Theorem 2 follows.

Appendix C
Proof of Theorem 3

We begin by deriving sufficient conditions for the SDP estimator to produce the true labels X^*.

Lemma 7: For the optimization problem (11), consider the Lagrange multipliers

$$D^* = \text{diag}(d_i^*), \qquad S^* \triangleq \begin{bmatrix} S_A^* & S_B^{*T} \\ S_B^* & S_C^* \end{bmatrix}.$$

If we have

$$\begin{aligned}&S_A^* = T_2 Y^T X^*,\\ &S_B^* = -T_2 Y,\\ &S_C^* = D^* - T_1 G,\\ &S^* \succeq 0,\\ &\lambda_2(S^*) > 0,\\ &S^*[1, X^{*T}]^T = 0\end{aligned}$$

then (D^*, S^*) is the dual optimal solution and $\widehat{Z} = X^* X^{*T}$ is the unique primal optimal solution of (11).

Proof: Define

$$H \triangleq \begin{bmatrix} 1 & X^T \\ X & Z \end{bmatrix}.$$

The Lagrangian of (11) is given by

$$L(Z, X, S, D) = T_1\langle G, Z\rangle + 2T_2\langle Y, X\rangle + \langle S, H\rangle - \langle D, Z - \mathbf{I}\rangle,$$

where $S \succeq 0$ and $D = \text{diag}(d_i)$ are Lagrange multipliers. For any Z that satisfies the constraints in (11),

$$\begin{aligned}T_1\langle G, Z\rangle + 2T_2\langle Y, X\rangle &\overset{(a)}{\leq} L(Z, X, S^*, D^*)\\ &= \langle D^*, \mathbf{I}\rangle + S_A^*\\ &\overset{(b)}{=} \langle D^*, Z^*\rangle - \langle S_B^*, X^*\rangle\\ &= \langle S_C^* + T_1 G, Z^*\rangle - \langle S_B^*, X^*\rangle\\ &\overset{(c)}{=} T_1\langle G, Z^*\rangle - 2\langle S_B^*, X^*\rangle\\ &\overset{(d)}{=} T_1\langle G, Z^*\rangle + 2T_2\langle Y, X^*\rangle,\end{aligned}$$

where (a) holds because $\langle S^*, H\rangle \geq 0$, (b) holds because $Z_{ii} = 1$ for all $i \in [n]$ and $S_A^* = -S_B^{*T} X^*$, (c) holds because $S_B^* = -S_C^* X^*$, and (d) holds because $S_B^* = -T_2 Y$. Therefore, $Z^* = X^* X^{*T}$ is a primal optimal solution. Now, assume $\tilde{Z}$ is another optimal solution.

$$\begin{aligned}\langle S^*, \tilde{H}\rangle &= S_A^* + 2\langle S_B^*, \tilde{X}\rangle + \langle D^* - T_1 G, \tilde{Z}\rangle\\ &\overset{(a)}{=} S_A^* + 2\langle S_B^*, X^*\rangle + \langle D^*, Z^*\rangle - T_1\langle G, Z^*\rangle\\ &= \langle S^*, H^*\rangle = 0\end{aligned}$$

where (a) holds because $\langle G, Z^*\rangle = \langle G, \tilde{Z}\rangle$, $Z_{ii}^* = \tilde{Z}_{ii} = 1$ for all $i \in [n]$, and $\langle S_B^*, X^*\rangle = \langle S_B^*, \tilde{X}\rangle$. Since $\tilde{H} \succeq 0$ and $S^* \succeq 0$ while its second smallest eigenvalue $\lambda_2(S^*)$ is positive, $\tilde{H}$ must be a multiple of H^*. Also, since $\tilde{Z}_{ii} = Z_{ii}^* = 1$ for all $i \in [n]$, we have $\tilde{H} = H^*$. ∎

We now show that S^* defined by S_A^*, S_B^*, and S_C^* satisfies other conditions in Lemma 7 with probability $1 - o(1)$. Let

$$d_i^* = T_1 \sum_{j=1}^{n} G_{ij} x_j^* x_i^* + T_2 y_i x_i^*. \tag{22}$$

Then $D^*X^* = T_1GX^* + T_2Y$ and based on the definitions of S_A^*, S_B^*, and S_C^* in Lemma 7, S^* satisfies the condition $S^*[1, X^{*T}]^T = 0$. It remains to show that $S^* \succeq 0$ and $\lambda_2(S^*) > 0$ with probability $1 - o(1)$. In other words, we need to show that

$$\mathbb{P}\left(\inf_{V\perp[1,X^{*T}]^T, \|V\|=1} V^T S^* V > 0\right) \geq 1 - o(1), \tag{23}$$

where V is a vector of length $n+1$. Let $V \triangleq [v, U^T]^T$, where v is a scalar and $U \triangleq [u_1, u_2, \ldots, u_n]^T$. For any V such that $V^T[1, X^{*T}]^T = 0$ and $\|V\| = 1$, we have

$$\begin{aligned} V^T S^* V &= v^2 S_A^* - 2T_2 v U^T Y + U^T D^* U - T_1 U^T G U \\ &\geq (1 - v^2)\left[\min_{i\in[n]} d_i^* - T_1\|G - \mathbb{E}[G]\| + T_1 p(1 - 2\xi)\right] \\ &\quad + v^2\left[T_2 Y^T X^* - 2T_2\frac{\sqrt{n(1 - v^2)}}{|v|} - T_1 p(1 - 2\xi)\right], \end{aligned} \tag{24}$$

where the last inequality holds because

$$U^T D^* U \geq (1 - v^2)\min_{i\in[n]} d_i^*,$$

$$U^T(G - \mathbb{E}[G])U \leq (1 - v^2)\|G - \mathbb{E}[G]\|,$$

$$vU^T Y \leq |v|\sqrt{n(1 - v^2)}.$$

Lemma 8: Under the noisy label side information with noise parameter α,

$$\mathbb{P}\left(\sum_{i=1}^n x_i^* y_i \leq \sqrt{n}\log n\right) \leq e^{n\left(\log\left(2\sqrt{\alpha(1-\alpha)}\right) + o(1)\right)}.$$

Proof: It follows from Chernoff bound. ■

Using Lemma 8, it can be shown that with probability converging to one, $\sum_{i=1}^n x_i^* y_i \geq \sqrt{n}\log n$. Thus,

$$v^2\left[T_2\sqrt{n}\log n - 2T_2\frac{\sqrt{n(1 - v^2)}}{|v|} - T_1 p(1 - 2\xi)\right] \geq 0,$$

as $n \to \infty$. Applying Lemma 2,

$$V^T S^* V \geq (1 - v^2)\left(\min_{i\in[n]} d_i^* - T_1 c'\sqrt{\log n} + T_1 p(1 - 2\xi)\right). \tag{25}$$

Lemma 9: Consider a sequence $f(n)$, and for each n a sequence of i.i.d. random variables $\{S_1, \ldots, S_m\}$ with distribution $p_1\delta_{+1} + p_2\delta_{-1} + (1 - p_1 - p_2)\delta_0$, where the parameters of the distribution depend on n via $p_1 = \rho_1\frac{\log n}{n}$, and $p_2 = \rho_2\frac{\log n}{n}$ for some positive constants ρ_1, ρ_2. We assume $m(n) - n = o(n)$, where in the sequel the dependence of m on n is implicit. Define $\omega \triangleq \lim_{n\to\infty}\frac{f(n)}{\log n}$. For sufficiently large n, when $\omega < \rho_1 - \rho_2$,

$$\mathbb{P}\left(\sum_{i=1}^m S_i \leq f(n)\right) \leq n^{-\eta^* + o(1)}, \tag{26}$$

and when $\omega > \rho_1 - \rho_2$,

$$\mathbb{P}\left(\sum_{i=1}^m S_i \geq f(n)\right) = n^{-\eta^* + o(1)}, \tag{27}$$

where $\eta^* = \rho_1 + \rho_2 - \gamma^* + \frac{\omega}{2}\log\left(\frac{\rho_2(\gamma^* + \omega)}{\rho_1(\gamma^* - \omega)}\right)$ and $\gamma^* = \sqrt{\omega^2 + 4\rho_1\rho_2}$.

Proof: Inequality (26) is derived by applying Chernoff bound. Equality (27) is obtained by a sandwich argument on the probability: an upper bound derived via Chernoff bound, and a lower bound from [31, Lemma 15]. ■

It follows from (22) that

$$\begin{aligned} \mathbb{P}(d_i^* \leq \delta) = {}& \mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^* x_j^* \leq \frac{\delta - T_2}{T_1}\right)(1 - \alpha) \\ &+ \mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^* x_j^* \leq \frac{\delta + T_2}{T_1}\right)\alpha, \end{aligned}$$

where $\sum_{j=1}^n G_{ij}x_i^* x_j^*$ is equal in distribution to $\sum_{i=1}^{n-1} S_i$ in Lemma 9 with $p_1 = p(1 - \xi)$ and $p_2 = p\xi$.

Recall that $\beta \triangleq \lim_{n\to\infty}\frac{T_2}{\log n}$, where $\beta \geq 0$. First, we bound $\min_{i\in[n]} d_i^*$ under the condition $0 \leq \beta < aT_1(1 - 2\xi)$. It follows from Lemma 9 that

$$\mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^* x_j^* \leq \frac{\delta - T_2}{T_1}\right) \leq n^{-\eta(a,\beta) + o(1)},$$

$$\mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^* x_j^* \leq \frac{\delta + T_2}{T_1}\right) \leq n^{-\eta(a,\beta) + \beta + o(1)}.$$

Then

$$\begin{aligned} \mathbb{P}(d_i^* \leq \delta) &\leq n^{-\eta(a,\beta) + o(1)}(1 - \alpha) + n^{-\eta(a,\beta) + \beta + o(1)}\alpha \\ &= n^{-\eta(a,\beta) + o(1)}. \end{aligned}$$

Using the union bound,

$$\mathbb{P}\left(\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}\right) \geq 1 - n^{1 - \eta(a,\beta) + o(1)}.$$

When $\eta(a, \beta) > 1$, it follows $\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}$ with probability $1 - o(1)$. Thus, as long as $\eta(a, \beta) > 1$, we can replace $\min d_i^*$ in (25) with $\frac{\log n}{\log\log n}$ and obtain, with probability $1 - o(1)$:

$$V^T S^* V \geq (1-v^2)\left(\frac{\log n}{\log\log n} - T_1 c'\sqrt{\log n} + T_1 p(1-2\xi)\right) > 0,$$

which concludes the first part of Theorem 3.

We now bound $\min_{i\in[n]} d_i^*$ under the condition $\beta > aT_1(1-2\xi)$. It follows from Lemma 9 that

$$\mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^* x_j^* \leq \frac{\delta - T_2}{T_1}\right) \leq n^{-\eta(a,\beta)+o(1)},$$
$$\mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^* x_j^* \leq \frac{\delta + T_2}{T_1}\right) \leq 1.$$

Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\eta(a,\beta)+o(1)} + n^{-\beta+o(1)},$$

where $\alpha = n^{-\beta+o(1)}$. Using the union bound,

$$\mathbb{P}\left(\min_{i\in[n]} d_i^* \geq \delta\right) \geq 1 - \left(n^{1-\eta(a,\beta)+o(1)} + n^{1-\beta+o(1)}\right).$$

Lemma 10: If $\beta > 1$, then $\eta(a,\beta) > 1$.

Proof: Define $\psi(a,\beta) \triangleq \eta(a,\beta) - \beta$. It can be shown that $\psi(a,\beta)$ is a convex function in β. At the optimal β^*, $\log(\frac{(1-\xi)(\gamma^*+\beta^*)}{\xi(\gamma^*-\beta^*)}) = 2T_1$. Then

$$\eta(a,\beta) - \beta \geq a - \frac{\gamma^*}{T_1}. \tag{28}$$

It can be shown that at the optimal β^*,

$$\frac{\gamma^*+\beta^*}{\gamma^*-\beta^*} = \frac{1-\xi}{\xi} = \frac{4\xi(1-\xi)a^2T_1^2}{(\gamma^*-\beta^*)^2}.$$

Then $\gamma^* = \beta^* + 2\xi a T_1$ and (28) is written as

$$\eta(a,\beta) - \beta \geq a - 2\xi a - \frac{\beta^*}{T_1}. \tag{29}$$

Also, it can be shown that at β^*, $\gamma^* = \frac{\beta^*}{1-2\xi}$. This implies that $\beta^* = (1-2\xi)aT_1$. Substituting in (29) leads to $\eta(a,\beta) - \beta \geq 0$, which implies that $\eta(a,\beta) > 1$ when $\beta > 1$. ■

When $\beta > 1$, using Lemma 10, it follows $\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}$ with probability $1 - o(1)$. Substituting in (25), if $\beta > 1$, with probability $1-o(1)$ we obtain:

$$V^T S^* V \geq (1-v^2)\left(\frac{\log n}{\log\log n} - T_1 c'\sqrt{\log n} + T_1 p(1-2\xi)\right) > 0,$$

which concludes the second part of Theorem 3.

Appendix D
Proof of Theorem 4

Since the prior distribution of X^* is uniform, among all estimators, the maximum likelihood estimator minimizes the average error probability. Therefore, we only need to show that with high probability the maximum likelihood estimator fails. Let

$$F \triangleq \left\{ \min_{i\in[n]} \left(T_1 \sum_{j=1}^n G_{ij}x_j^* x_i^* + T_2 x_i^* y_i\right) \leq -T_1 \right\}.$$

Then $\mathbb{P}(\text{ML Fails}) \geq \mathbb{P}(F)$. Let H denote the set of first $\lfloor \frac{n}{\log^2 n} \rfloor$ nodes and $e(i,H)$ denote the number of edges between node i and nodes in the set $H \subset [n]$. It can be shown that

$$\begin{aligned}
&\min_{i\in[n]} \left(T_1 \sum_{j\in[n]} G_{ij}x_j^* x_i^* + T_2 x_i^* y_i\right) \\
&\leq \min_{i\in H} \left(T_1 \sum_{j\in[n]} G_{ij}x_j^* x_i^* + T_2 x_i^* y_i\right) \\
&\leq \min_{i\in H} \left(T_1 \sum_{j\in H^c} G_{ij}x_j^* x_i^* + T_2 x_i^* y_i\right) + \max_{i\in H}\ e(i,H).
\end{aligned}$$

Define

$$E_1 \triangleq \left\{ \max_{i\in H}\ e(i,H) \leq \delta - T_1 \right\},$$
$$E_2 \triangleq \left\{ \min_{i\in H} \left(T_1 \sum_{j\in H^c} G_{ij}x_j^* x_i^* + T_2 x_i^* y_i\right) \leq -\delta \right\}.$$

Notice that $F \supset E_1 \cap E_2$ and it suffices to show $\mathbb{P}(E_1) \to 1$ and $\mathbb{P}(E_2) \to 1$ to prove that the maximum likelihood estimator fails. Since $e(i,H) \sim \text{Binom}(|H|, a\frac{\log n}{n})$, from Lemma 5,

$$\begin{aligned}
&\mathbb{P}(e(i,H) \geq \delta - T_1) \\
&\leq \left(\frac{\log^2 n}{ae\log\log n} - \frac{T_1 \log n}{ae}\right)^{T_1 - \frac{\log n}{\log\log n}} e^{-\frac{a}{\log n}} \leq n^{-2+o(1)}.
\end{aligned}$$

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

Let

$$E \triangleq \left\{ T_1 \sum_{j\in H^c} G_{ij}x_j^* x_i^* + T_2 x_i^* y_i \leq -\delta \right\},$$
$$E_\alpha \triangleq \left\{ \sum_{j\in H^c} G_{ij}x_j^* x_i^* \leq \frac{-\delta + T_2}{T_1} \right\},$$
$$E_{1-\alpha} \triangleq \left\{ \sum_{j\in H^c} G_{ij}x_j^* x_i^* \leq \frac{-\delta - T_2}{T_1} \right\}.$$

Then

$$\begin{aligned}
\mathbb{P}(E_2) &= 1 - \prod_{i\in H}[1 - \mathbb{P}(E)] \stackrel{(a)}{=} 1 - [1 - \mathbb{P}(E)]^{|H|} \\
&= 1 - [1 - \alpha\mathbb{P}(E_\alpha) - (1-\alpha)\mathbb{P}(E_{1-\alpha})]^{|H|},
\end{aligned} \tag{30}$$

where (a) holds because $\{T_1 \sum_{j\in H^c} G_{ij}x_j^* x_i^* + T_2 x_i^* y_i\}_{i\in H}$ are mutually independent.

First, we bound $\mathbb{P}(E_2)$ under the condition $0 \leq \beta < aT_1(1-2\xi)$. Using Lemma 9, $\mathbb{P}(E_\alpha) \geq n^{-\eta(a,\beta)+\beta+o(1)}$ and $\mathbb{P}(E_{1-\alpha}) \geq n^{-\eta(a,\beta)+o(1)}$. It follows from (30) that

$$\mathbb{P}(E_2) \overset{(a)}{\geq} 1 - \left[1 - n^{-\eta(a,\beta)+o(1)}\right]^{|H|} \overset{(b)}{\geq} 1 - \exp\Big(-n^{1-\eta(a,\beta)+o(1)}\Big),$$

where (a) holds because $\alpha = n^{-\beta+o(1)}$ and (b) is due to $1 + x \leq e^x$. Therefore, if $\eta(a,\beta) < 1$, then $\mathbb{P}(E_2) \to 1$ and the first part of Theorem 4 follows.

We now bound $\mathbb{P}(E_2)$ under the condition $\beta \geq aT_1(1-2\xi)$. Reorganizing (30),

$$\mathbb{P}(E_2) = 1 - [(1-\alpha)\mathbb{P}(E_{1-\alpha}^c) + \alpha\mathbb{P}(E_\alpha^c)]^{|H|}, \tag{31}$$

where

$$\mathbb{P}(E_\alpha^c) = \mathbb{P}\Bigg(\sum_{j\in H^c} G_{ij}x_j^* x_i^* \geq \frac{-\delta + T_2}{T_1}\Bigg),$$
$$\mathbb{P}(E_{1-\alpha}^c) = \mathbb{P}\Bigg(\sum_{j\in H^c} G_{ij}x_j^* x_i^* \geq \frac{-\delta - T_2}{T_1}\Bigg).$$

Also, $\sum_{j\in H^c} G_{ij}x_i^* x_j^*$ is equal in distribution to $\sum_{i=1}^{|H^c|-1} S_i$ in Lemma 9, where $p_1 = p(1-\xi)$ and $p_2 = p\xi$. Then $\mathbb{P}(E_\alpha^c) \leq n^{-\eta(a,\beta)+\beta+o(1)}$ and $\mathbb{P}(E_{1-\alpha}^c) \leq 1$. It follows from (31) that

$$\begin{aligned}\mathbb{P}(E_2) &\geq 1 - \left[(1-\alpha) + \alpha n^{-\eta(a,\beta)+\beta+o(1)}\right]^{|H|}\\ &\overset{(a)}{=} 1 - \left[1 - n^{-\beta+o(1)} + n^{-\eta(a,\beta)+o(1)}\right]^{|H|}\\ &\overset{(b)}{\geq} 1 - e^{-n^{1-\beta+o(1)}\left(1-n^{-\eta(a,\beta)+\beta+o(1)}\right)},\end{aligned}$$

where (a) holds because $\alpha = n^{-\beta+o(1)}$ and (b) is due to $1 + x < e^x$. Therefore, since $\beta \leq \eta(a,\beta)$, if $\beta < 1$, then $\mathbb{P}(E_2) \to 1$ and the second part of Theorem 4 follows.

Appendix E
Proof of Theorem 5

We begin by deriving sufficient conditions for the SDP estimator to produce the true labels X^*.

Lemma 11: The sufficient conditions of Lemma 7 apply to the general side information SDP (12) by replacing $S_A^* = \tilde{Y}^T X^*$ and $S_B^* = -\tilde{Y}$.

Proof: The proof is similar to the proof of Lemma 7. ■

It suffices to show that S^*, defined via its components S_A^*, S_B^*, and S_C^*, satisfies other conditions in Lemma 11 with probability $1 - o(1)$. Let

$$d_i^* = T_1 \sum_{j=1}^n G_{ij}x_j^* x_i^* + \tilde{y}_i x_i^*. \tag{32}$$

Then $D^*X^* = T_1 G X^* + \tilde{Y}$ and based on the definitions of S_A^*, S_B^*, and S_C^* in Lemma 11, S^* satisfies the condition $S^*[1,\ X^{*T}]^T = 0$. It remains to show that (23) holds, i.e., $S^* \succeq 0$ and $\lambda_2(S^*) > 0$ with probability $1 - o(1)$. Let

$$y_{max} \triangleq K \max_{k,m_k} \left|\log\left(\frac{\alpha_{+,mk}^k}{\alpha_{-,m_k}^k}\right)\right|, \tag{33}$$

where $k \in \{1, 2, \ldots, K\}$ and $m_k \in \{1, 2, \ldots, M_K\}$. For any V such that $V^T[1, X^{*T}]^T = 0$ and $\|V\| = 1$, we have

$$\begin{aligned}V^T S^* V &= v^2 S_A^* - 2vU^T\tilde{Y} + U^T D^* U - T_1 U^T G U\\ &\geq (1-v^2)\left[\min_{i\in[n]} d_i^* - T_1\|G - \mathbb{E}[G]\| + T_1 p(1-2\xi)\right]\\ &\quad + v^2\left[\tilde{Y}^T X^* - 2y_{max}\frac{\sqrt{n(1-v^2)}}{|v|} - T_1 p(1-2\xi)\right],\end{aligned} \tag{34}$$

where the last inequality holds in a manner similar to (24) with the difference that in the present case

$$vU^T\tilde{Y} \leq |v| y_{max}\sqrt{n(1-v^2)}.$$

Lemma 12: For feature k of general side information,

$$\mathbb{P}\Bigg(\sum_{i=1}^n x_i^* z_{i,k} \geq \sqrt{n}\log n\Bigg) \geq 1 - o(1),$$

where

$$z_{i,k} \triangleq \sum_{m_k=1}^{M_k} \mathbb{1}_{\{y_{i,k}=m_k\}} \log\left(\frac{\alpha_{+,m_k}^k}{\alpha_{-,m_k}^k}\right).$$

Proof: For feature k, let

$$\delta' \triangleq \sqrt{n}\log n, \rho_j \triangleq \frac{1}{n}|\{i \in [n] : y_{i,k} = j\}|,$$

where $j \in \{1, \ldots, M_k\}$ and $\sum_j \rho_j = 1$. Then

$$\mathbb{P}\Bigg(\sum_{i=1}^n x_i^* z_{i,k} \leq \delta'\Bigg) \leq \sum_{j=1}^{M_k} \mathbb{P}\Bigg(\sum_{i\in A_j} x_i^* z_{i,k} \leq \delta'\Bigg).$$

Applying Chernoff bound yields

$$\mathbb{P}\Bigg(\sum_{i\in A_j} x_i^* z_{i,k} \leq \delta'\Bigg) \leq e^{n(\psi_{k,j}+o(1))},$$

where

$$\psi_{k,j} \triangleq \rho_j \log\left(2\sqrt{\alpha_{+,j}^k \alpha_{-,j}^k \mathbb{P}(x_i^* = 1)\mathbb{P}(x_i^* = -1)}\right).$$

Since $\psi_{k,j} < 0$ for any values of $\alpha_{+,j}^k$ and $\alpha_{-,j}^k$, we have

$$\mathbb{P}\Bigg(\sum_{i=1}^n x_i^* z_{i,k} \geq \delta'\Bigg) \geq 1 - \sum_{j=1}^{M_k} e^{n(\psi_{k,j}+o(1))} = 1 - o(1).$$

Therefore, with probability $1-o(1)$, $\sum_{i=1}^{n} x_i^* z_{i,k} \geq \sqrt{n}\log n$ and Lemma 12 follows. ∎

Using Lemmas 2 and 12,

$$V^T S^* V \geq (1-v^2)\left(\min_{i\in[n]} d_i^* - T_1 c' \sqrt{\log n} + T_1 p(1-2\xi)\right). \tag{35}$$

It can be shown that $\sum_{j=1}^{n} G_{ij} x_i^* x_j^*$ in (32) is equal in distribution to $\sum_{i=1}^{n-1} S_i$ in Lemma 9, where $p_1 = p(1-\xi)$ and $p_2 = p\xi$. Then

$$\mathbb{P}(d_i^* \leq \delta) = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \cdots \sum_{m_K=1}^{M_K} P(m_1, \ldots, m_K), \tag{36}$$

where

$$P(m_1,\ldots,m_K) \triangleq \mathbb{P}(x_i=1) e^{f_2(n)} \mathbb{P}\left(\sum_{i=1}^{n-1} S_i \leq \frac{\delta - f_1(n)}{T_1}\right) + \mathbb{P}(x_i=-1) e^{f_3(n)} \mathbb{P}\left(\sum_{i=1}^{n-1} S_i \leq \frac{\delta + f_1(n)}{T_1}\right).$$

First, we bound $\min_{i\in[n]} d_i^*$ under the condition $|\beta_1| \leq aT_1(1-2\xi)$. It follows from Lemma 9 that

$$\mathbb{P}\left(\sum_{i=1}^{n-1} S_i \leq \frac{\delta - f_1(n)}{T_1}\right) \leq n^{-\eta(a,\beta_1)+o(1)},$$
$$\mathbb{P}\left(\sum_{i=1}^{n-1} S_i \leq \frac{\delta + f_1(n)}{T_1}\right) \leq n^{-\eta(a,\beta_1)-\beta_1+o(1)}.$$

Notice that

$$\beta \triangleq \lim_{n\to\infty} -\frac{\max(f_2(n), f_3(n))}{\log n}.$$

When $\beta_1 \geq 0$, $\lim_{n\to\infty} \frac{f_2(n)}{\log n} = -\beta$ and $\lim_{n\to\infty} \frac{f_3(n)}{\log n} = -\beta_1 - \beta$. Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\eta(a,\beta_1)-\beta+o(1)}.$$

When $\beta_1 < 0$, $\lim_{n\to\infty} \frac{f_3(n)}{\log n} = -\beta$ and $\lim_{n\to\infty} \frac{f_2(n)}{\log n} = \beta_1 - \beta$. Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\eta(a,\beta_1)+\beta_1-\beta+o(1)} = n^{-\eta(a,|\beta_1|)-\beta+o(1)}.$$

Using the union bound,

$$\mathbb{P}\left(\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}\right) \geq 1 - n^{1-\eta(a,|\beta_1|)-\beta+o(1)}.$$

When $\eta(a,|\beta_1|)+\beta > 1$, it follows that $\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}$ holds with probability $1-o(1)$. Substituting into (35), if $\eta(a,|\beta_1|)+\beta > 1$, then with probability $1-o(1)$,

$$V^T S^* V \geq (1-v^2)\left(\frac{\log n}{\log\log n} - T_1 c' \sqrt{\log n} + T_1 p(1-2\xi)\right) > 0,$$

which concludes the first part of Theorem 5.

We now bound $\min_{i\in[n]} d_i^*$ under the condition $|\beta_1| \geq aT_1(1-2\xi)$. When $\beta_1 \geq 0$, $\lim_{n\to\infty} \frac{f_2(n)}{\log n} = -\beta$ and $\lim_{n\to\infty} \frac{f_3(n)}{\log n} = -\beta_1 - \beta$. Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\beta+o(1)} + n^{-\beta-\beta_1+o(1)}.$$

When $\beta_1 < 0$, $\lim_{n\to\infty} \frac{f_3(n)}{\log n} = -\beta$ and $\lim_{n\to\infty} \frac{f_2(n)}{\log n} = \beta_1 - \beta$. Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\beta+\beta_1+o(1)} + n^{-\beta+o(1)}.$$

Using the union bound,

$$\mathbb{P}\left(\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}\right) \geq 1 - n^{1-|\beta_1|-\beta+o(1)}.$$

When $|\beta_1|+\beta > 1$, with probability $1-o(1)$, we have $\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}$. Substituting into (35), if $|\beta_1|+\beta > 1$, then with probability $1-o(1)$,

$$V^T S^* V \geq (1-v^2)\left(\frac{\log n}{\log\log n} - T_1 c' \sqrt{\log n} + T_1 p(1-2\xi)\right) > 0,$$

which concludes the second part of Theorem 5.

Appendix F
Proof of Theorem 6

Similar to the proof of Theorem 4, let

$$F \triangleq \left\{ \min_{i\in[n]} \left(T_1 \sum_{j=1}^{n} G_{ij} x_j^* x_i^* + x_i^* \tilde{y}_i\right) \leq -T_1 \right\}.$$

Then $\mathbb{P}(\text{ML Fails}) \geq \mathbb{P}(F)$ and if we show that $\mathbb{P}(\mathrm{F}) \to 1$, the maximum likelihood estimator fails. Let H be the set of first $\lfloor \frac{n}{\log^2 n} \rfloor$ nodes and $e(i,H)$ denote the number of edges between node i and other nodes in the set H. It can be shown that

$$\begin{aligned}
&\min_{i\in[n]} \left(T_1 \sum_{j\in[n]} G_{ij} x_j^* x_i^* + x_i^* \tilde{y}_i\right) \\
&\leq \min_{i\in H} \left(T_1 \sum_{j\in[n]} G_{ij} x_j^* x_i^* + x_i^* \tilde{y}_i\right) \\
&\leq \min_{i\in H} \left(T_1 \sum_{j\in H^c} G_{ij} x_j^* x_i^* + x_i^* \tilde{y}_i\right) + \max_{i\in H} e(i,H).
\end{aligned}$$

Let

$$E_1 \triangleq \left\{ \max_{i\in H} \; e(i,H) \le \delta - T_1 \right\},$$
$$E_2 \triangleq \left\{ \min_{i\in H} \; \left(T_1 \sum_{j\in H^c} G_{ij}x_j^* x_i^* + x_i^* \tilde{y}_i \right) \le -\delta \right\}.$$

Notice that $F \supset E_1 \cap E_2$. Then the maximum likelihood estimator fails if we show that $\mathbb{P}(E_1) \to 1$ and $\mathbb{P}(E_2) \to 1$. Since $e(i,H) \sim \text{Binom}(|H|, a\frac{\log n}{n})$, from Lemma 5,

$$\mathbb{P}(e(i,H) \ge \delta - T_1)$$
$$\le \left(\frac{\log^2 n}{ae \log\log n} - \frac{T_1 \log n}{ae} \right)^{T_1 - \frac{\log n}{\log\log n}} e^{-\frac{a}{\log n}} \le n^{-2+o(1)}.$$

Using the union bound, $\mathbb{P}(E_1) \ge 1 - n^{-1+o(1)}$.

Let

$$E \triangleq \left\{ T_1 \sum_{j\in H^c} G_{ij}x_j^* x_i^* + x_i^* \tilde{y}_i \le -\delta \right\},$$
$$E_+ \triangleq \left\{ \sum_{j\in H^c} G_{ij}x_j^* x_i^* \le \frac{-\delta - f_1(n)}{T_1} \right\},$$
$$E_- \triangleq \left\{ \sum_{j\in H^c} G_{ij}x_j^* x_i^* \le \frac{-\delta + f_1(n)}{T_1} \right\}.$$

Define

$$P(m_1,\ldots,m_K) \triangleq \mathbb{P}(x_i^*=1)e^{f_2(n)}\mathbb{P}(E_+) + \mathbb{P}(x_i^*=-1)e^{f_3(n)}\mathbb{P}(E_-).$$

Then

$$\mathbb{P}(E_2) = 1 - \prod_{i\in H}[1-\mathbb{P}(E)] \stackrel{(a)}{=} 1 - [1-\mathbb{P}(E)]^{|H|}$$
$$= 1 - \left[1 - \sum_{m_1=1}^{M_1} \cdots \sum_{m_K=1}^{M_K} P(m_1,\ldots,m_K) \right]^{|H|},$$

where (a) holds because $\{T_1 \sum_{j\in H^c} G_{ij}x_j^*x_i^* + x_i^*\tilde{y}_i\}_{i\in H}$ are mutually independent.

First, we bound $\mathbb{P}(E_2)$ under the condition $|\beta_1| \le aT_1(1-2\xi)$. Using Lemma 9, $\mathbb{P}(E_+) \ge n^{-\eta(a,\beta_1)+o(1)}$ and $\mathbb{P}(E_-) \ge n^{-\eta(a,\beta_1)+\beta_1+o(1)}$. When $\beta_1 \ge 0$, $\lim_{n\to\infty}\frac{f_2(n)}{\log n} = -\beta$ and $\lim_{n\to\infty}\frac{f_3(n)}{\log n} = -\beta_1 - \beta$. Then

$$\mathbb{P}(E_2) = 1 - \left[1 - n^{-\eta(a,\beta_1)-\beta+o(1)} \right]^{|H|}$$
$$\ge 1 - \exp\left(-n^{1-\eta(a,\beta_1)-\beta+o(1)} \right),$$

using $1+x \le e^x$. When $\beta_1 < 0$, $\lim_{n\to\infty}\frac{f_3(n)}{\log n} = -\beta$ and $\lim_{n\to\infty}\frac{f_2(n)}{\log n} = \beta_1 - \beta$. Then

$$\mathbb{P}(E_2) = 1 - \left[1 - n^{-\eta(a,\beta_1)+\beta_1-\beta+o(1)} \right]^{|H|}$$
$$\ge 1 - \exp\left(-n^{1-\eta(a,|\beta_1|)-\beta+o(1)} \right),$$

using $1+x \le e^x$ and $\eta(a,\beta_1) - \beta_1 = \eta(a,|\beta_1|)$. Therefore, if $\eta(a,|\beta_1|) + \beta < 1$, then $\mathbb{P}(E_2) \to 1$ and the first part of Theorem 6 follows.

We now bound $\mathbb{P}(E_2)$ under the condition $|\beta_1| \ge aT_1(1-2\xi)$. When $\beta_1 \ge 0$, $\lim_{n\to\infty}\frac{f_2(n)}{\log n} = -\beta$ and $\lim_{n\to\infty}\frac{f_3(n)}{\log n} = -\beta_1 - \beta$. Using Lemma 9, $\mathbb{P}(E_+) \ge n^{-\eta(a,\beta_1)+o(1)}$ and $\mathbb{P}(E_-) \ge 1 - o(1)$. Then

$$\mathbb{P}(E_2) \ge 1 - \left[1 - n^{-\eta(a,\beta_1)-\beta+o(1)} - n^{-\beta_1-\beta+o(1)} \right]^{|H|}$$
$$\ge 1 - \exp\left(-n^{1-\eta(a,\beta_1)-\beta+o(1)} - n^{1-\beta_1-\beta+o(1)} \right),$$

using $1+x \le e^x$. When $\beta_1 < 0$, $\lim_{n\to\infty}\frac{f_3(n)}{\log n} = -\beta$ and $\lim_{n\to\infty}\frac{f_2(n)}{\log n} = \beta_1 - \beta$. Using Lemma 9, $\mathbb{P}(E_+) \ge 1 - o(1)$ and $\mathbb{P}(E_-) \ge n^{-\eta(a,|\beta_1|)+o(1)}$. Then

$$\mathbb{P}(E_2) = 1 - \left[1 - n^{\beta_1-\beta+o(1)} - n^{-\eta(a,|\beta_1|)-\beta+o(1)} \right]^{|H|}$$
$$\ge 1 - \exp\left(-n^{1+\beta_1-\beta+o(1)} - n^{1-\eta(a,|\beta_1|)-\beta+o(1)} \right),$$

using $1+x \le e^x$. Therefore, since $|\beta_1| \le \eta(a,|\beta_1|)$, if $|\beta_1| + \beta < 1$, then $\mathbb{P}(E_2) \to 1$ and the second part of Theorem 6 follows.

Appendix G
Proof of Theorem 7

We begin by deriving sufficient conditions for the solution of SDP (13) to match the true labels.

Lemma 13: For the optimization problem (13), consider the Lagrange multipliers

$$\lambda^*, \quad \mu^*, \quad D^* = \text{diag}(d_i^*), \quad S^*.$$

If we have

$$S^* = D^* + \lambda^*\mathbf{J} + \mu^* W - G, S^* \succeq 0, \lambda_2(S^*) > 0, S^*X^* = 0,$$

then $(\lambda^*, \mu^*, D^*, S^*)$ is the dual optimal solution and $\widehat{Z} = X^*X^{*T}$ is the unique primal optimal solution of (13).

Proof: The proof is similar to the proof of Lemma 1. The Lagrangian of (13) is given by

$$L(Z,S,D,\lambda,\mu) = \langle G,Z\rangle + \langle S,Z\rangle - \langle D, Z-\mathbf{I}\rangle - \lambda\langle \mathbf{J},Z\rangle - \mu(\langle W,Z\rangle - (Y^TY)^2),$$

where $S \succeq 0$, $D = \text{diag}(d_i)$, λ, μ are Lagrange multipliers. Since $\langle \mathbf{J}, Z\rangle = 0$, for any Z that satisfies the constraints in (13), it can be shown that $\langle G,Z\rangle \le \langle G,Z^*\rangle$. Also, similar to the proof of Lemma 1, it can be shown that the optimum solution is unique. ∎

It suffices to show that $S^* = D^* + \lambda^*\mathbf{J} + \mu^* W - G$ satisfies other conditions in Lemma 13 with probability $1 - o(1)$. Let

$$d_i^* = \sum_{j=1}^n G_{ij} x_j^* x_i^* - \mu^* \sum_{j=1}^n y_i y_j x_j^* x_i^*. \tag{37}$$

Then $D^* X^* = GX^* - \mu^* W X^*$ and based on the definition of S^* in Lemma 13, S^* satisfies the condition $S^* X^* = 0$. It remains to show that (17) holds, i.e., $S^* \succeq 0$ and $\lambda_2(S^*) > 0$ with probability $1 - o(1)$. Under the binary stochastic block model,

$$\mathbb{E}[G] = \frac{p-q}{2} X^* X^{*T} + \frac{p+q}{2}\mathbf{J} - p\mathbf{I}. \tag{38}$$

It follows that for any V such that $V^T X^* = 0$ and $\|V\| = 1$,

$$\begin{aligned} V^T S^* V = V^T D^* V + \Big(\lambda^* - \frac{p+q}{2}\Big) V^* \mathbf{J} V + p \\ - V^T (G - \mathbb{E}[G]) V + \mu^* V^T W V. \end{aligned}$$

Let $\lambda^* \geq \frac{p+q}{2}$. Since $V^T D^* V \geq \min_{i\in[n]} d_i^*$ and $V^T(G - \mathbb{E}[G])V \leq \|G - \mathbb{E}[G]\|$,

$$V^T S^* V \geq \min_{i\in[n]} d_i^* + p - \|G - \mathbb{E}[G]\| + \mu^* V^T W V.$$

Lemma 14: [28, Thoerem 5] For any $c > 0$, there exists $c' > 0$ such that for any $n \geq 1$, $\|G - \mathbb{E}[G]\| \leq c' \sqrt{\log n}$ with probability at least $1 - n^{-c}$.

Also, it can be shown that Lemma 3 holds here. Choose $\mu^* < 0$, then in view of Lemmas 14 and 3, with probability $1 - o(1)$,

$$V^T S^* V \geq \min_{i\in[n]} d_i^* + p + (\mu^* - c')\sqrt{\log n}. \tag{39}$$

Lemma 15: When $\delta = \frac{\log n}{\log logn}$, then

$$\mathbb{P}(d_i^* \leq \delta) \leq \epsilon n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2 + o(1)} + (1-\epsilon)\epsilon^n.$$

Proof: It follows from Chernoff bound. ■

Recall that $\beta \triangleq \lim_{n\to\infty} -\frac{\log \epsilon}{\log n}$, where $\beta \geq 0$. It follows from Lemma 15 that

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2 - \beta + o(1)}.$$

Then using the union bound,

$$\mathbb{P}\bigg(\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}\bigg) \geq 1 - n^{1-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2 - \beta + o(1)}.$$

When $(\sqrt{a} - \sqrt{b})^2 + 2\beta > 2$, it follows that $\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}$ holds with probability $1 - o(1)$. Combining this result with (39), if $(\sqrt{a} - \sqrt{b})^2 + 2\beta > 2$, then with probability $1 - o(1)$,

$$V^T S^* V \geq \frac{\log n}{\log\log n} + p + (\mu^* - c')\sqrt{\log n} > 0,$$

which completes the proof of Theorem 7.

APPENDIX H
PROOF OF THEOREM 8

We begin by deriving sufficient conditions for the solution of SDP (14) to match the true labels.

Lemma 16: For the optimization problem (14), consider the Lagrange multipliers

$$\lambda^*, \quad D^* = \mathrm{diag}(d_i^*), \quad S^* \triangleq \begin{bmatrix} S_A^* & S_B^{*T} \\ S_B^* & S_C^* \end{bmatrix}.$$

If we have

$$\begin{aligned} & S_A^* = T_2 Y^T X^*, \\ & S_B^* = -T_2 Y, \\ & S_C^* = D^* + \lambda^* \mathbf{J} - T_1 G, \\ & S^* \succeq 0, \\ & \lambda_2(S^*) > 0, \\ & S^*[1, X^{*T}]^T = 0, \end{aligned}$$

then (λ^*, D^*, S^*) is the dual optimal solution and $\widehat{Z} = X^* X^{*T}$ is the unique primal optimal solution of (14).

Proof: The proof is similar to the proof of Lemma 7. The Lagrangian of (14) is given by

$$\begin{aligned} L(Z, X, S, D, \lambda) = T_1\langle G, Z\rangle + T_2\langle Y, X\rangle + \langle S, H\rangle \\ - \langle D, Z - \mathbf{I}\rangle - \lambda\langle \mathbf{J}, Z\rangle, \end{aligned}$$

where $S \succeq 0$, $D = \mathrm{diag}(d_i)$, and $\lambda \in \mathbb{R}$ are Lagrange multipliers. Since $\langle \mathbf{J}, Z^*\rangle = 0$, for any Z that satisfies the constraints in (14), it can be shown that $T_1\langle G, Z\rangle + T_2\langle Y, X\rangle \leq T_1\langle G, Z^*\rangle + T_2\langle Y, X^*\rangle$. Also, the uniqueness of optimum solution is proved similarly. ■

We now show that S^* defined by S_A^*, S_B^*, and S_C^* satisfies the remaining conditions in Lemma 16 with probability $1 - o(1)$. Let

$$d_i^* = T_1 \sum_{j=1}^n G_{ij} x_j^* x_i^* + T_2 y_i x_i^*. \tag{40}$$

Then $D^* X^* = T_1 G X^* + T_2 Y$ and based on the definitions of S_A^*, S_B^*, and S_C^* in Lemma 16, S^* satisfies the condition $S^*[1, X^{*T}]^T = 0$. It remains to show that (23) holds, i.e., $S^* \succeq 0$ and $\lambda_2(S^*) > 0$ with probability $1 - o(1)$.

For any V such that $V^T[1, X^{*T}]^T = 0$ and $\|V\| = 1$, we have

$$\begin{aligned}V^TS^*V &= v^2S_A^* - 2vT_2U^TY + U^TD^*U - T_1U^TGU\\ &\geq (1-v^2)\left[\min_{i\in[n]} d_i^* - T_1\|G-\mathbb{E}[G]\| + T_1p\right]\\ &\quad + v^2\left[Y^TX^* - 2T_2\frac{\sqrt{n(1-v^2)}}{|v|} - T_1\frac{p-q}{2}\right],\end{aligned}\tag{41}$$

where the last inequality holds in a manner similar to (24). Using Lemma 8,

$$V^TS^*V \geq (1-v^2)\left(\min_{i\in[n]} d_i^* - T_1c'\sqrt{\log n} + T_1p\right).\tag{42}$$

Lemma 17: Consider a sequence $f(n)$, and for each n, let $S_1 \sim \text{Binom}(\frac{n}{2}-1, p)$ and $S_2 \sim \text{Binom}(\frac{n}{2}, q)$, where $p = a\frac{\log n}{n}$, and $q = b\frac{\log n}{n}$ for some $a \geq b > 0$. Define $\omega \triangleq \lim_{n\to\infty}\frac{f(n)}{\log n}$. For sufficiently large n, when $\omega < \frac{a-b}{2}$,

$$\mathbb{P}(S_1 - S_2 \leq f(n)) \leq n^{-\eta^*+o(1)},$$

where $\eta^* = \frac{a+b}{2} - \gamma^* - \frac{\omega}{2}\log\left(\frac{a}{b}\right) + \frac{\omega}{2}\log\left(\frac{\gamma^*+\omega}{\gamma^*-\omega}\right)$ and $\gamma^* = \sqrt{\omega^2+ab}$.

Proof: It follows from Chernoff bound. ∎

It follows from (40) that

$$\begin{aligned}\mathbb{P}(d_i^* \leq \delta) &= \mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^*x_j^* \leq \frac{\delta-T_2}{T_1}\right)(1-\alpha)\\ &\quad + \mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^*x_j^* \leq \frac{\delta+T_2}{T_1}\right)\alpha,\end{aligned}$$

where $\sum_{j=1}^n G_{ij}x_i^*x_j^*$ is equal in distribution to $S_1 - S_2$ in Lemma 17.

Recall that $\beta \triangleq \lim_{n\to\infty}\frac{T_2}{\log n}$, where $\beta \geq 0$. First, we bound $\min_{i\in[n]} d_i^*$ under the condition $0 \leq \beta < \frac{T_1}{2}(a-b)$. It follows from Lemma 17 that

$$\begin{aligned}\mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^*x_j^* \leq \frac{\delta-T_2}{T_1}\right) &\leq n^{-\eta(a,b,\beta)+o(1)},\\ \mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^*x_j^* \leq \frac{\delta+T_2}{T_1}\right) &\leq n^{-\eta(a,b,\beta)+\beta+o(1)}.\end{aligned}$$

Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\eta(a,b,\beta)+o(1)}.$$

Using the union bound,

$$\mathbb{P}\left(\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}\right) \geq 1 - n^{1-\eta(a,b,\beta)+o(1)}.$$

When $\eta(a,b,\beta) > 1$, it follows that $\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}$ holds with probability $1-o(1)$. Substituting into (42), if $\eta(a,b,\beta) > 1$, then with probability $1-o(1)$,

$$V^TS^*V \geq (1-v^2)\left(\frac{\log n}{\log\log n} - T_1c'\sqrt{\log n} + T_1p\right) > 0,$$

which concludes the first part of Theorem 8.

We now bound $\min_{i\in[n]} d_i^*$ under the condition $\beta > \frac{T_1}{2}(a-b)$. It follows from Lemma 17 that

$$\begin{aligned}\mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^*x_j^* \leq \frac{\delta-T_2}{T_1}\right) &\leq n^{-\eta(a,b,\beta)+o(1)},\\ \mathbb{P}\left(\sum_{j=1}^n G_{ij}x_i^*x_j^* \leq \frac{\delta+T_2}{T_1}\right) &\leq 1.\end{aligned}$$

Then

$$\begin{aligned}\mathbb{P}(d_i^* \leq \delta) &\leq n^{-\eta(a,b,\beta)+o(1)}(1-\alpha) + \alpha\\ &= n^{-\eta(a,b,\beta)+o(1)} + n^{-\beta+o(1)},\end{aligned}$$

where $\alpha = n^{-\beta+o(1)}$. Using the union bound,

$$\mathbb{P}\left(\min_{i\in[n]} d_i^* \geq \delta\right) \geq 1 - n^{1-\eta(a,b,\beta)+o(1)} - n^{1-\beta+o(1)}.$$

Lemma 18: [31, Lemma 8] When $\beta > 1$, then $\eta(a,b,\beta) > 1$.

When $\beta > 1$, using Lemma 18, it follows that $\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}$ holds with probability $1-o(1)$. Substituting into (42), if $\beta > 1$, then with probability $1-o(1)$,

$$V^TS^*V \geq (1-v^2)\left(\frac{\log n}{\log\log n} - T_1c'\sqrt{\log n} + T_1p\right) > 0,$$

which concludes the second part of Theorem 8.

Appendix I
Proof of Theorem 9

We begin by deriving sufficient conditions for the solution of SDP (14) to match the true labels.

Lemma 19: The sufficient conditions of Lemma 16 apply to the general side information SDP (15) by replacing $S_A^* = \tilde{Y}^TX^*$ and $S_B^* = -\tilde{Y}$.

Proof: The proof is similar to the proof of Lemma 16. ∎

It suffices to show that S^* defined by S_A^*, S_B^*, and S_C^* satisfies other conditions in Lemma 19 with probability $1-o(1)$. Let

$$d_i^* = T_1\sum_{j=1}^n G_{ij}x_j^*x_i^* + \tilde{y}_ix_i^*.\tag{43}$$

Then $D^*X^* = T_1GX^* + \tilde{Y}$ and based on the definitions of S_A^*, S_B^*, and S_C^* in Lemma 19, S^* satisfies the condition $S^*[1, X^{*T}]^T = 0$. It remains to show that (23) holds, i.e., $S^* \succeq 0$ and $\lambda_2(S^*) > 0$ with probability $1-o(1)$. For any V such that $V^T[1, X^{*T}]^T = 0$ and $\|V\| = 1$, we have

$$\begin{aligned} V^T S^* V &= v^2 S_A^* - 2vT_2 U^T Y + U^T D^* U - T_1 U^T G U \\ &\overset{(a)}{\geq} (1-v^2)\left[\min_{i\in[n]} d_i^* - T_1\|G - \mathbb{E}[G]\| + T_1 p\right] \\ &\quad + v^2\left[\tilde{Y}^T X^* - 2y_{max}\frac{\sqrt{n(1-v^2)}}{|v|} - T_1\frac{p-q}{2}\right] \\ &\overset{(b)}{=} (1-v^2)\left[\min_{i\in[n]} d_i^* - T_1\|G - \mathbb{E}[G]\| + T_1 p\right], \end{aligned}$$

where (a) holds in a manner similar to (24) and (41), and (b) holds by applying Lemma 12. Then using Lemma 14,

$$V^T S^* V \geq (1-v^2)\left(\min_{i\in[n]} d_i^* - T_1 c'\sqrt{\log n} + T_1 p\right). \quad (44)$$

It can be shown that $\sum_{j=1}^n G_{ij} x_i^* x_j^*$ in (43) is equal in distribution to $S_1 - S_2$ in Lemma 17. Then

$$\mathbb{P}(d_i^* \leq \delta) = \sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}\cdots\sum_{m_K=1}^{M_K} P(m_1,\ldots,m_K),$$

where

$$\begin{aligned} P(m_1,\ldots,m_K) &\triangleq \mathbb{P}(x_i^*=1)e^{f_2(n)}\mathbb{P}\left(S_1 - S_2 \leq \frac{\delta - f_1(n)}{T_1}\right) \\ &+ \mathbb{P}(x_i^*=-1)e^{f_3(n)}\mathbb{P}\left(S_1 - S_2 \leq \frac{\delta + f_1(n)}{T_1}\right). \end{aligned}$$

First, we bound $\min_{i\in[n]} d_i^*$ under the condition $|\beta_1| \leq \frac{T_1}{2}(a-b)$. It follows from Lemma 17 that

$$\mathbb{P}\left(S_1 - S_2 \leq \frac{\delta - f_1(n)}{T_1}\right) \leq n^{-\eta(a,b,\beta_1)+o(1)},$$
$$\mathbb{P}\left(S_1 - S_2 \leq \frac{\delta + f_1(n)}{T_1}\right) \leq n^{-\eta(a,b,\beta_1)+\beta_1+o(1)}.$$

Notice that

$$\beta \triangleq \lim_{n\to\infty} -\frac{\max(f_2(n), f_3(n))}{\log n}.$$

When $\beta_1 \geq 0$, $\lim_{n\to\infty}\frac{f_2(n)}{\log n} = -\beta$ and $\lim_{n\to\infty}\frac{f_3(n)}{\log n} = -\beta_1 - \beta$. Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\eta(a,b,\beta_1)-\beta+o(1)}.$$

When $\beta_1 < 0$, $\lim_{n\to\infty}\frac{f_3(n)}{\log n} = -\beta$ and $\lim_{n\to\infty}\frac{f_2(n)}{\log n} = \beta_1 - \beta$. Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\eta(a,b,\beta_1)+\beta_1-\beta+o(1)} = n^{-\eta(a,b,|\beta_1|)-\beta+o(1)}.$$

Using the union bound,

$$\mathbb{P}\left(\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}\right) \geq 1 - n^{1-\eta(a,b,|\beta_1|)-\beta+o(1)}.$$

When $\eta(a,b,|\beta_1|) + \beta > 1$, it follows that $\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}$ holds with probability $1-o(1)$. Substituting into (44), if $\eta(a,b,|\beta_1|)+\beta > 1$, then with probability $1-o(1)$,

$$V^T S^* V \geq (1-v^2)\left(\frac{\log n}{\log\log n} - T_1 c'\sqrt{\log n} + T_1 p\right) > 0,$$

which concludes the first part of Theorem 9.

We now bound $\min_{i\in[n]} d_i^*$ under the condition $|\beta_1| \geq \frac{T_1}{2}(a-b)$. When $\beta_1 > 0$, $\lim_{n\to\infty}\frac{f_2(n)}{\log n} = -\beta$ and $\lim_{n\to\infty}\frac{f_3(n)}{\log n} = -\beta_1 - \beta$. Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\beta+o(1)} + n^{-\beta-\beta_1+o(1)}.$$

When $\beta_1 < 0$, $\lim_{n\to\infty}\frac{f_3(n)}{\log n} = -\beta$ and $\lim_{n\to\infty}\frac{f_2(n)}{\log n} = \beta_1 - \beta$. Then

$$\mathbb{P}(d_i^* \leq \delta) \leq n^{-\beta+\beta_1+o(1)} + n^{-\beta+o(1)}.$$

Using the union bound,

$$\mathbb{P}\left(\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}\right) \geq 1 - n^{1-|\beta_1|-\beta+o(1)}.$$

When $|\beta_1| + \beta > 1$, it follows that $\min_{i\in[n]} d_i^* \geq \frac{\log n}{\log\log n}$ holds with probability $1-o(1)$. Substituting into (44), if $|\beta_1| + \beta > 1$, then with probability $1-o(1)$,

$$V^T S^* V \geq (1-v^2)\left(\frac{\log n}{\log\log n} - T_1 c'\sqrt{\log n} + T_1 p\right) > 0,$$

which concludes the second part of Theorem 9.

References

[1] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, 2002.

[2] J. Xu, R. Wu, K. Zhu, B. Hajek, R. Srikant, and L. Ying, "Jointly clustering rows and columns of binary matrices, Algorithms and trade-offs," *SIGMETRICS Perform, Eval. Rev.*, vol. 42, no. 1, pp. 29–41, Jun. 2014.

[3] K. Huang, W. Deng, Y. Zhang, and H. Zhu, "Sparse bayesian learning for network structure reconstruction based on evolutionary game data," *Phys. A, Statist. Mech. Appl.*, vol. 541, 2020, Art. no. 123605. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437119320102

[4] Z. Wang, C. Wang, X. Li, C. Gao, X. Li, and J. Zhu, "Evolutionary Markov dynamics for network community detection," *IEEE Trans. Knowl. Data Eng.*, to be published, doi: 10.1109/TKDE.2020.2997043.

[5] K. Huang, Z. Wang, and M. Jusup, "Incorporating latent constraints to enhance inference of network structure," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 466–475, Jan.-Mar. 2020.

[6] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3, pp. 75–174, 2010.

[7] M. Esmaeili and A. Nosratinia, "Community detection, Exact recovery in weighted graphs," 2021, *arXiv:2102.04439*.

[8] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Netw.*, vol. 5, no. 2, pp. 109–137, 1983.

[9] E. Abbe and C. Sandon, "Community detection in general stochastic block models, Fundamental limits and efficient algorithms for recovery," in *Proc. IEEE 56th Annu. Symp. Foundations Comput. Sci.*, Oct. 2015, pp. 670–688.

[10] A. Saade, M. Lelarge, F. Krzakala, and L. Zdeborová, "Spectral detection in the censored block model," in *Proc. IEEE Int. Symp. Inf. Theory*, 2015, pp. 1184–1188.

[11] B. Hajek, Y. Wu, and J. Xu, "Exact recovery threshold in the binary censored block model," in *Proc. IEEE Inf. Theory Workshop-Fall*, 2015, pp. 99–103.

[12] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Inference and phase transitions in the detection of modules in sparse networks," *Phys. Rev. Lett.*, vol. 107, Aug. 2011, Art. no. 065701.

[13] E. Mossel, J. Neeman, and A. Sly, "Reconstruction and estimation in the planted partition model," *Probability Theory Related Fields*, vol. 162, no. 3, pp. 431–461, Aug. 2015.

[14] L. Massoulié, "Community detection thresholds and the weak ramanujan property," in *Proc. 46th Annu. ACM Symp. Theory Comput.*, 2014, pp. 694–703.

[15] E. Mossel, J. Neeman, and A. Sly, "A proof of the block model threshold conjecture," *Combinatorica*, vol. 38, no. 3, pp. 665–708, 2018.

[16] S.-Y. Yun and A. Proutiere, "Community detection via random and adaptive sampling," in *Proc. Conf. Learn. Theory*, 2014, pp. 138–175.

[17] E. Mossel and J. Xu, "Density evolution in the degree-correlated stochastic block model," in *Proc. Conf. Learn. Theory*, 2016, pp. 1319–1356.

[18] H. Saad, A. Abotabl, and A. Nosratinia, "Exit analysis for belief propagation in degree-correlated stochastic block models," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2016, pp. 775–779.

[19] E. Abbe, A. S. Bandeira, and G. Hall, "Exact recovery in the stochastic block model," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 471–487, Jan. 2016.

[20] E. Mossel, J. Neeman, and A. Sly, "Consistency thresholds for the planted bisection model," *Electron. J. Probab.*, vol. 21, pp. 1–24, 2016.

[21] Y. Chen and J. Xu, "Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 882–938, Jan. 2016.

[22] E. Mossel, J. Neeman, and A. Sly, "Belief propagation, robust reconstruction and optimal recovery of block models," in *Proc. Conf. Learn. Theory*, 2014, pp. 356–370.

[23] A. A. Amini *et al.*, "On semidefinite relaxations for the block model," *Ann. Statist.*, vol. 46, no. 1, pp. 149–179, 2018.

[24] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug. 2006.

[25] J. Mattingley and S. Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 50–61, May 2010.

[26] U. Feige and J. Kilian, "Heuristics for semirandom graph problems," *J. Comput. Syst. Sci.*, vol. 63, no. 4, pp. 639–671, 2001.

[27] A. Frieze and M. Jerrum, "Improved approximation algorithms for MAX k-CUT and MAX BISECTION," *Algorithmica*, vol. 18, no. 1, pp. 67–81, May 1997.

[28] B. Hajek, Y. Wu, and J. Xu, "Achieving exact cluster recovery threshold via semidefinite programming," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2788–2797, May 2016.

[29] B. Hajek, Y. Wu, and J. Xu, "Achieving exact cluster recovery threshold via semidefinite programming: Extensions," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5918–5937, Oct. 2016.

[30] M. Esmaeili and A. Nosratinia, "Community detection with secondary latent variables," in *Proc. IEEE Int. Symp. Inf. Theory*, 2020, pp. 1355–1360.

[31] H. Saad and A. Nosratinia, "Community detection with side information: Exact recovery under the stochastic block model," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 5, pp. 944–958, Oct. 2018.

[32] H. Saad, A. Abotabl, and A. Nosratinia, "Exact recovery in the binary stochastic block model with binary side information," in *Proc. Allerton Conf. Commun., Control, Comput.*, Oct. 2017, pp. 822–829.

[33] H. Saad and A. Nosratinia, "Recovering a single community with side information," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7939–7966, Dec. 2020.

[34] H. Saad and A. Nosratinia, "Belief propagation with side information for recovering a single community," in *Proc. IEEE Int. Symp. Inf. Theory*, 2018, pp. 1271–1275.

[35] H. Saad and A. Nosratinia, "Side information in recovering a single community: Information theoretic limits," in *Proc. IEEE Int. Symp. Inf. Theory*, 2018, pp. 2107–2111.

[36] E. Mossel and J. Xu, "Local algorithms for block models with side information," in *Proc. ACM Conf. Innovations Theoretical Comput. Sci.*, 2016, pp. 71–80.

[37] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, Nov. 1995.

[38] M. Esmaeili, H. Saad, and A. Nosratinia, "Exact recovery by semidefinite programming in the binary stochastic block model with partially revealed side information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2019, pp. 3477–3481.

[39] M. Esmaeili, H. Saad, and A. Nosratinia, "Community detection with side information via semidefinite programming," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 420–424.

Mohammad Esmaeili received the B.Sc. degree in electrical engineering from the K. N. Toosi University of Technology, Tehran, Iran, in 2013 and the M.S. degree in electrical engineering from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2015. He is currently working toward the Ph.D. degree in electrical engineering with the University of Texas at Dallas, Richardson, TX, USA. His research interests include machine learning, information theory, and its applications in machine learning and data science, signal processing, and wireless communications. He was awarded the Louis Beecherl, Jr. Graduate Fellowship for the 2019–2020 academic year and Jan Van der Ziel Fellowship for the 2020–2021 academic year.

Hussein Metwaly Saad (Member, IEEE) received the B.S. degree in electrical engineering from the German University, Cairo, Egypt, the M.S. degree in electrical engineering from Nile University, Giza, Egypt, and the Ph.D. degree in electrical engineering with the University of Texas at Dallas, Richardson, TX, USA. He is currently a Research Assistant Professor with Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech-Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA. His research interests include machine leanring, graph thoery, information theory, and wireless communications. In 2014, he was the recipient of the Erik Jonsson Graduate Fellowship from the University of Texas at Dallas.

Aria Nosratinia (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1996. He has held visiting appointments with Princeton University, Princeton, NJ, USA, Rice University, Houston, TX, USA, and University of California, Los Angeles, Los Angeles, CA, USA. He is currently the Erik Jonsson Distinguished Professor and Associate Head of the Electrical and computer Engineering Department, The University of Texas at Dallas, Richardson, TX, USA. His interests include information theory and signal processing. He was the Editor of the IEEE Transactions on Information Theory, IEEE Transactions on Communications, IEEE Signal Processing Letters, IEEE Transactions on Image Processing, and IEEE Wireless Communications. He was named a Highly Cited Researcher by Clarivate Analytics (formerly Thompson Reuters).