

Community Detection with Known, Unknown, or Partially Known Auxiliary Latent Variables

Mohammad Esmaeili and Aria Nosratinia, *Fellow, IEEE*

Abstract—Empirical observations suggest that in practice, community membership does not completely explain the dependence between the edges of an observation graph. The residual dependence of the graph edges are modeled in this paper, to first order, by auxiliary node latent variables that affect the statistics of the graph edges but *carry no information about the communities of interest*. We then study community detection in graphs obeying the stochastic block model and censored block model with auxiliary latent variables. We analyze the conditions for exact recovery when these auxiliary latent variables are unknown, representing unknown nuisance parameters or model mismatch. We also analyze exact recovery when these secondary latent variables have been either fully or partially revealed. Finally, we propose a semidefinite programming algorithm for recovering the desired labels when the secondary labels are either known or unknown. We show that exact recovery is possible by semidefinite programming down to the respective maximum likelihood exact recovery threshold.

Index Terms—Community Detection, Latent Variables, Stochastic Block Model (SBM), Censored Block Model (CBM), Graph Inference, Exact Recovery, Semidefinite Programming (SDP), Chernoff-Hellinger Divergence.

I. INTRODUCTION

Community detection refers to a clustering of the nodes of a graph based on the observation of the edges. In many applications, this involves identifying groups of nodes that are more densely connected within the group than to nodes outside the group. Community detection has many applications such as finding like-minded people in social networks [1], exploration of biomedical networks [2], improving link predictors and recommendation systems [3]–[5], and is also relevant to network reconstruction problems [6]–[9]. Community detection has been widely investigated in the literature from both theoretical and algorithmic perspectives. Community detection is based on graph models such as the stochastic block model and the censored block model [10]–[16]. Several metrics are used in this field to characterize the asymptotic behavior of the residual errors as the size of the graph grows, including correlated recovery, weak recovery, almost exact recovery, and exact recovery [17]–[26]. Among the various detection techniques one can name spectral methods, belief propagation, and semidefinite programming [27]–[32].

In the graph models that have so far been studied for community detection, the graph edges are generated independently conditioned on the community labels. A brief survey of models that are most closely related to the present work will

be presented shortly. However, in many practical community detection problems, the community labels do not fully explain the dependence between the graph edges. In other words, in many graphs encountered in practice, the graph edges conditioned on the desired community labels are not statistically independent. This happens when the structure of the graph is also influenced by factors other than the community of interest. For example, one may consider political affiliation communities on a social network in a university campus, where the social network graph is also influenced by other variables that may be unrelated to the community label of interest, such as membership in intramural and extramural activities. The nature and magnitude of the dependence of the graph on these secondary or auxiliary factors can have an effect on the performance of the community detection algorithm for the community label of interest. The present study models and analyzes community detection in this scenario.

Toward that goal, this paper introduces secondary or auxiliary latent variables in the graph model that are not subject to community detection themselves, but influence the structure of the graph. More specifically, we propose and employ a more general version of the stochastic block model and censored block model in which edges are independent conditioned on both the community labels and a set of secondary latent variables. The secondary or auxiliary latent variables represent a first-order model for the residual dependence of the edges of the graph once the effect of the community labels has been removed. Auxiliary variables are independent of community memberships and may or may not be observable. The auxiliary latent variable model is distinct from side-information model [33], [34] where the side information variables are directly observed and carry information about the communities. Side information represents non-graph information about communities, while auxiliary variables model the *graph connectivity patterns* that are *unrelated* to the communities.

We investigate the exact recovery threshold for community detection in the graphs with secondary latent variables. We also analyze the effect on the performance of community detection when this secondary latent variable is fully or partially known. We also propose and investigate a semidefinite programming algorithm for community detection with secondary latent variables. Our analysis shows that exact recovery via semidefinite programming is possible down to the respective maximum likelihood exact recovery threshold, for both unknown or known secondary latent variables.

In addition to addressing a novel problem, this paper also provides a novel proof for bounding the summation of the minimums of Poisson-distributed values from above and below via Chernoff-Hellinger divergence. Our result (Lemma 1)

M. Esmaeili and A. Nosratinia are with the Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75083-0688, USA, Email: esmaeili@utdallas.edu, aria@utdallas.edu. This work was supported in part by the NSF grant CIF-2008684.

eliminates certain technical difficulties that existed in earlier proofs, e.g., does not impose restrictions on the domain of Poisson distributions. This result is extended (Lemma 2) for the general censored block model. Also, the analysis of exact recovery for a graph generated based on two latent variables involves subtleties in extracting the maximum likelihood estimator and analyzing its semidefinite programming relaxation, which go beyond earlier works.

To put the model of this paper in perspective, we review several community detection graph models whose nodes are associated, beyond a scalar community detection label, with some other variables too. The latent space model [35]–[37] associates with each node a vector, often with small dimension, containing variables that are latent in the model. The graph edges are generated from a distribution that is parameterized based on the distance between the latent vectors of pairs of nodes, and the community is a scalar generated as a function of each latent vector. The overlapping stochastic block model [11], [38] recovers multiple independent, identically distributed, binary communities via observing a graph whose edges are drawn independently conditioned on all the community labels of the terminating nodes. An important distinction of overlapped communities from the present work is that all communities must be recovered in the overlapped model, therefore the overlapped model has significant similarity with a multi-community model. In the overlapped model, the multiple communities possess a structure that can be exploited, compared with a general multi-community model. Finally, there exists some work on combining non-graph observation with graph observations [33], [34]; these works have a superficial resemblance to the subsection in this paper where the secondary latent variable is revealed. However, the graph and the side information in [33], [34] are assumed independent of each other conditioned on community labels, therefore the revealed side information in [33], [34] has no direct influence on the graph. Thus, [33], [34] model a different phenomenon and also have a different mathematical structure, compared with the present work. In the interest of brevity, our coverage of various community detection models is limited, and the interested reader is referred to more comprehensive coverage available, e.g., in [11].

Notation: \mathbf{I} is the identity matrix and \mathbf{J} the all-one matrix. $S \succeq 0$ indicates a positive semidefinite matrix and $S \geq 0$ denotes a matrix with non-negative entries. $\|S\|$ is the spectral norm and $\lambda_2(S)$ is the second smallest eigenvalue (for a symmetric matrix). $[a, b]$ is a vector that is obtained by stacking vectors a and b . $\langle \cdot, \cdot \rangle$ is the inner product and $*$ is the element-wise product. We abbreviate $[n] \triangleq \{1, \dots, n\}$. $\mathbb{P}(\cdot)$ indicates the probability operator and $P(\cdot)$ a probability distribution which is identified by the choice of its variables whenever there is no confusion. Random variables with Bernoulli and Binomial distributions are indicated by $\text{Bern}(p)$ and $\text{Bin}(n, p)$, respectively, with n trials and success probability p . Also, random variables with Poisson distribution are indicated by $\mathcal{P}_\lambda(n)$ with n trials and parameter λ .

II. SYSTEM MODEL

We start by considering a two-latent variable model, and assume the cardinality of both is finite. For notational convenience throughout the paper, x, y are length- n vectors holding latent variable values for the whole graph, while the latent variables for any node v are represented with x_v, y_v . In our model, we aim to discover x , therefore nodes that share the same value for x are called a *community*. By *micro-community*, we refer to the set of nodes in the graph that share the same value for both latent variables x, y . The matrix P denotes prior probabilities

$$P_{i,j} = \mathbb{P}(x_v = i, y_v = j).$$

For convenience and for avoiding tensor calculations, we further define:

$$p \triangleq \text{vec}(P).$$

For both the two-latent variable stochastic block model and two-latent variable censored block model, the graph edges are Bernoulli distributed, conditioned on the latent variables of the two nodes terminating the edge. The conditional Bernoulli parameters for an arbitrary edge are organized in a symmetric matrix \bar{Q} , whose rows and columns are ordered in a manner compatible with vector p . In other words, assuming the latent variable x_v has m_x outcomes, then the probability of an edge between two nodes with latent variable pairs taking values (i, j) and (i', j') is given by the element of \bar{Q} in row $jm_x + i$ and column $j'm_x + i'$.

We are interested in a regime where edge probabilities diminish with the size of the graph n , in particular, in the context of our model there exist a constant matrix Q such that:

$$\bar{Q} = \frac{\log n}{n} Q.$$

This assumption asymptotically guarantees a fully connected graph.

Example 1. Consider a two-latent variable stochastic block model with $m_x = 2$ and $m_y = 3$. Then

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} \\ P_{1,0} & P_{1,1} & P_{1,2} \end{bmatrix},$$

$$p = [P_{0,0} \ P_{0,1} \ P_{0,2} \ P_{1,0} \ P_{1,1} \ P_{1,2}],$$

$$\bar{Q} = \frac{\log n}{n} \begin{bmatrix} Q_{0,0} & Q_{0,1} & Q_{0,2} & Q_{0,3} & Q_{0,4} & Q_{0,5} \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & Q_{1,3} & Q_{1,4} & Q_{1,5} \\ Q_{2,0} & Q_{2,1} & Q_{2,2} & Q_{2,3} & Q_{2,4} & Q_{2,5} \\ Q_{3,0} & Q_{3,1} & Q_{3,2} & Q_{3,3} & Q_{3,4} & Q_{3,5} \\ Q_{4,0} & Q_{4,1} & Q_{4,2} & Q_{4,3} & Q_{4,4} & Q_{4,5} \\ Q_{5,0} & Q_{5,1} & Q_{5,2} & Q_{5,3} & Q_{5,4} & Q_{5,5} \end{bmatrix}.$$

In addition, we define the columns of weighted versions of the matrix Q as

$$q^{(i,j)} \triangleq \text{diag}(p)Q e_{jm_x + i},$$

where e_k is the k -th canonical coordinate vector, and for convenience our notation of $q^{(i,j)}$ emphasizes dependence on

the latent variable outcomes rather than matrix coordinates. Thus, $q^{(i,j)}$ is the column of $\text{diag}(p)Q$. This vector represents the relative frequency of edges connecting a node from the micro-community (i, j) to all nodes of each micro-community (including the same micro-community). Also, we define the vector $\tilde{q}^{(i,j)}$ of size m_x with entries

$$\tilde{q}_{i'}^{(i,j)} \triangleq \sum_{j'} P_{i',j'} Q_{j'm_x+i',jm_x+i},$$

representing the relative frequency of edges, connecting a node from the micro-community (i, j) to all nodes of micro-communities with similar community latent variable.

For the two-latent variable censored block model, if an edge exists between a pair of nodes, the sign of the edge (positive or negative) is determined by a random variable drawn from a Bernoulli distribution with a certain parameter. The Bernoulli parameters for the positive sign of an edge are organized in a symmetric matrix Ξ , whose rows and columns are also ordered in a manner compatible with vector p . Finally, for the censored block model, we define similarly

$$g^{(i,j)} \triangleq \text{diag}(p)(\Xi * Q) e_{jm_x+i},$$

$$h^{(i,j)} \triangleq \text{diag}(p)((1 - \Xi) * Q) e_{jm_x+i},$$

and

$$\tilde{g}_{i'}^{(i,j)} \triangleq \sum_{j'} P_{i',j'} (\Xi * Q)_{j'm_x+i',jm_x+i},$$

$$\tilde{h}_{i'}^{(i,j)} \triangleq \sum_{j'} P_{i',j'} ((1 - \Xi) * Q)_{j'm_x+i',jm_x+i}.$$

Remark 1. The censored block model in [28], [39] with parameters a and ξ is a special case of the general censored model represented in this paper with

$$Q = \begin{bmatrix} a & a \\ a & a \end{bmatrix}, \quad \Xi = \begin{bmatrix} 1 - \xi & \xi \\ \xi & 1 - \xi \end{bmatrix}.$$

III. EXACT RECOVERY UNDER OPTIMAL DETECTION

The main results of this part are represented in the context of three scenarios, where the latent variable x is unknown and the latent variable y is either known or unknown (for all nodes in the graph) or partially known (for some nodes in the graph). Figure 1 shows graph realizations of a two-latent variable stochastic block model with $m_x = 2$ and $m_y = 2$. In each node, the community latent variable is indicated by the color of the inner circle, and the auxiliary latent variable is represented by the color of a ring around the inner circle.

The Chernoff-Hellinger divergence is due to Abbe [24] and is defined for two non-negative vectors a, b of the same dimension:

$$\text{Div}(a, b) \triangleq \max_{t \in [0,1]} \sum_i [ta_i + (1-t)b_i - a_i^t b_i^{1-t}]. \quad (1)$$

This is a generalization of the Hellinger divergence and the Chernoff divergence [11], [24]. In a manner similar to [11] we present a lemma that bounds a summation of the minimums of Poisson-distributed values.

Lemma 1. Let $a, b \in \mathbb{R}_+^m$, with $a \neq b$, and two positive scalars p, \hat{p} . For any Poisson multivariate distributions $\mathcal{P}_a(d)$ and $\mathcal{P}_b(d)$, define

$$I(a, b) \triangleq \sum_{d \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d)p, \mathcal{P}_b(d)\hat{p}\}.$$

Then

$$I(a, b) \leq \max\{p, \hat{p}\} e^{-\text{Div}(a,b)},$$

$$I(a, b) \geq \min\{p, \hat{p}\} e^{-\text{Div}(a,b)} \prod_{i=1}^m \frac{1}{e} (a_i^{t^*} b_i^{1-t^*})^{-\frac{1}{2}},$$

where t^* is the optimal parameter in the definition of Chernoff-Hellinger divergence $\text{Div}(a, b)$.

Proof. See Appendix A. \square

Let D be a random variable vector representing the number of edges that connect the node v to each micro-community. More specifically, $D^{(i',j')}$ is an element of the D indicating the number of edges connecting the node v to the micro-community (i', j') . For each node v , the proposed detection tests hypotheses

$$H_i : x_v = i.$$

If v belongs to micro-community (i, j) , then

$$D^{(i',j')} \sim \text{Bin}(nP_{i',j'}, \bar{Q}_{j'm_x+i',jm_x+i}).$$

In the regime where $\bar{Q} = Q \frac{\log n}{n}$, the Binomial distribution can be approximated by a Poisson distribution with the same mean, denoted $\lambda_{i',j'}^{(i',j')}$. Indeed, using Le Cam's inequality, the total variation distance between $\text{Bin}(nP_{i',j'}, \frac{\log n}{n} Q_{j'm_x+i',jm_x+i})$ and $\mathcal{P}(P_{i',j'} Q_{j'm_x+i',jm_x+i} \log n)$ asymptotically goes to zero. Then

$$\mathbb{P}(D = d | H_i, y_v = j) = \prod_{i'} \prod_{j'} \mathcal{P}_{\lambda_{i',j'}^{(i',j')}}(d^{(i',j')}),$$

where $\lambda_{i',j'}^{(i',j')} = P_{i',j'} Q_{j'm_x+i',jm_x+i} \log n$.

Theorem 1. Under the two-latent variable stochastic block model, all micro-communities are exactly recovered if and only if

$$\min_{(i,j) \neq (k,l)} \text{Div}(q^{(i,j)}, q^{(k,l)}) > 1.$$

Proof. It follows from the exact recovery under the general stochastic block model or the general overlapping stochastic block model. \square

Theorem 2. Under the two-latent variable stochastic block model, when the latent variable y is revealed, exact recovery of x is possible if and only if

$$\gamma_1 \triangleq \min_j \min_{i \neq k} \text{Div}(q^{(i,j)}, q^{(k,j)}) > 1.$$

Proof. See Appendix B. \square

Theorem 3. Under the two-latent variable stochastic block model, when both latent variables are unknown, exact recovery of x is possible if and only if

$$\gamma_2 \triangleq \min_j \min_{i \neq k} \text{Div}(\tilde{q}^{(i,j)}, \tilde{q}^{(k,j)}) > 1.$$

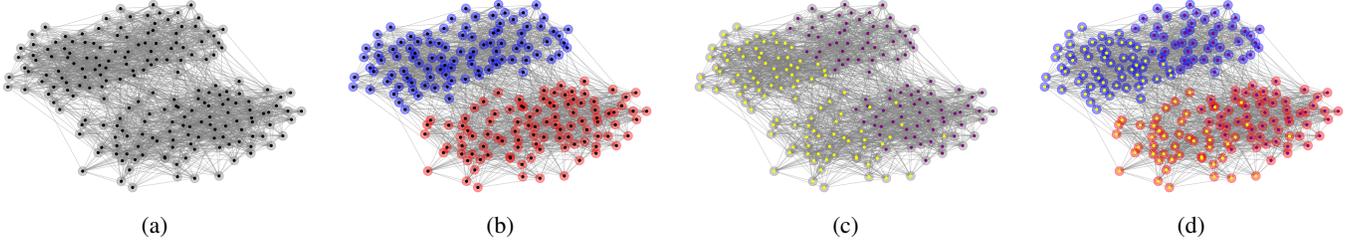


Fig. 1: For each node, (a) both latent variables are unknown, (b) knowing the statistics of the graph, the community latent variable is recovered while the auxiliary latent variable is unknown, (c) the auxiliary latent variable is known while the first one is unknown, (d) knowing the statistics of the graph, the community latent variable is recovered while the auxiliary latent variable is known.

Proof. See Appendix C. \square

Now we present the following Lemma which is similar to Lemma 1 and is crucial for the analysis of the censored block model.

Lemma 2. Let $a, b, \hat{a}, \hat{b} \in \mathbb{R}_+^m$, with $a \neq b$ or $\hat{a} \neq \hat{b}$, and two positive scalars p, \hat{p} . For any Poisson multivariate distributions $\mathcal{P}_a(d)$, $\mathcal{P}_b(d)$, $\mathcal{P}_{\hat{a}}(w)$, and $\mathcal{P}_{\hat{b}}(w)$, define

$$I(a, b, \hat{a}, \hat{b}) \triangleq \sum_{d, w \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d)\mathcal{P}_{\hat{a}}(w)p, \mathcal{P}_b(d)\mathcal{P}_{\hat{b}}(w)\hat{p}\}.$$

Then

$$\begin{aligned} I(a, b, \hat{a}, \hat{b}) &\leq \max\{p, \hat{p}\} e^{-\text{Div}([a, \hat{a}], [b, \hat{b}])}, \\ I(a, b, \hat{a}, \hat{b}) &\geq \min\{p, \hat{p}\} e^{-\text{Div}([a, \hat{a}], [b, \hat{b}])} \\ &\quad \times \prod_i \frac{1}{e^2} [(a_i \hat{a}_i)^{t^*} (b_i \hat{b}_i)^{1-t^*}]^{-\frac{1}{2}}, \end{aligned}$$

where t^* is the optimal parameter in the definition of Chernoff-Hellinger divergence $\text{Div}([a, \hat{a}], [b, \hat{b}])$.

Proof. See Appendix D. \square

Let D and W be random vectors representing the positive and negative edges that connect the node v to each micro-community, respectively. More specifically, $D^{(i', j')}$ and $W^{(i', j')}$ are elements of D and W indicating the number of positive and negative edges connecting the node v to the micro-community (i', j') , respectively. For each node v , the proposed detection tests hypotheses

$$H_i : x_v = i.$$

If v belongs to micro-community (i, j) , then

$$\begin{aligned} D^{(i', j')} &\sim \text{Bin}(nP_{i', j'}, (\Xi * \bar{Q})_{j' m_x + i', j' m_x + i}), \\ W^{(i', j')} &\sim \text{Bin}(nP_{i', j'}, ((1 - \Xi) * \bar{Q})_{j' m_x + i', j' m_x + i}). \end{aligned}$$

In the regime where $\bar{Q} = Q \frac{\log n}{n}$, the Binomial distribution can be approximated by a Poisson distribution with the same mean. The distributions of D and W can be approximated

by multivariate Poisson distributions $\mathcal{P}_{\lambda_{i,j}}$ and $\mathcal{P}_{\hat{\lambda}_{i,j}}$ with the vector means $\lambda_{i,j}$ and $\hat{\lambda}_{i,j}$, respectively. Therefore

$$\begin{aligned} \mathbb{P}(D = d, W = w | H_i, y_v = j) &= \mathbb{P}(D = d | H_i, y_v = j) \mathbb{P}(W = w | H_i, y_v = j) \\ &= \prod_{i'} \prod_{j'} \mathcal{P}_{\lambda_{i', j'}}(d^{(i', j')}) \mathcal{P}_{\hat{\lambda}_{i', j'}}(w^{(i', j')}), \end{aligned}$$

where

$$\begin{aligned} \lambda_{i', j'}^{(i', j')} &= P_{i', j'}(\Xi * Q)_{j' m_x + i', j' m_x + i} \log n, \\ \hat{\lambda}_{i', j'}^{(i', j')} &= P_{i', j'}((1 - \Xi) * Q)_{j' m_x + i', j' m_x + i} \log n. \end{aligned}$$

Theorem 4. Under two-latent variable censored block model, all micro-communities are exactly recovered if and only if

$$\min_{(i,j) \neq (k,l)} \text{Div}([g^{(i,j)}, h^{(i,j)}], [g^{(k,l)}, h^{(k,l)}]) > 1.$$

Proof. See Appendix E. \square

Theorem 5. Under the two-latent variable censored block model, when the latent variable y is revealed, exact recovery of x is possible if and only if

$$\gamma_3 \triangleq \min_j \min_{i \neq k} \text{Div}([g^{(i,j)}, h^{(i,j)}], [g^{(k,j)}, h^{(k,j)}]) > 1.$$

Proof. See Appendix F. \square

Theorem 6. Under the two-latent variable censored block model, when both latent variables are unknown, exact recovery of x is possible if and only if

$$\gamma_4 \triangleq \min_j \min_{i \neq k} \text{Div}([\tilde{g}^{(i,j)}, \tilde{h}^{(i,j)}], [\tilde{g}^{(k,j)}, \tilde{h}^{(k,j)}]) > 1.$$

Proof. See Appendix G. \square

Corollary 1. Assume x and y are unknown latent variables for all nodes. We randomly reveal the latent variable y for $(1 - \epsilon)n$ nodes, where $\epsilon \in (0, 1)$. This is equivalent to erasing the latent variable y which is a known latent variable from a node with erasure probability ϵ . Define

$$\beta_1 \triangleq - \lim_{n \rightarrow \infty} \frac{\log(1 - \epsilon)}{\log n}, \quad \beta_2 \triangleq - \lim_{n \rightarrow \infty} \frac{\log \epsilon}{\log n}.$$

- Under the two-latent variable stochastic block model exact recovery is asymptotically possible for latent variable x if and only if

$$\min(\gamma_1 + \beta_1, \gamma_2 + \beta_2) > 1.$$

- Under the two-latent variable censored block model exact recovery is asymptotically possible for latent variable x if and only if

$$\min(\gamma_3 + \beta_1, \gamma_4 + \beta_2) > 1.$$

The results of this part generalize to M latent variables without difficulty.

Remark 2. To prove the “if” part of all theorems in Section III, a partial recovery algorithm is required before applying a MAP estimator. For that purpose, the partial recovery algorithm in [11] is adopted and modified to match the scenarios in this paper. Please see Appendix I.

IV. SEMIDEFINITE PROGRAMMING RESULTS

This section describes a semidefinite programming algorithm for recovering the desired latent variable. The main results of this part are represented in the context of two scenarios, where the latent variable x is unknown and the latent variable y is either known or unknown (for all nodes in the graph). We consider $x, y \in \{\pm 1\}^n$ such that $x^T \mathbf{1} = 0$. Thus, the latent variable x represents two equal-sized communities. The sample size of the latent variable y , represented by $\rho \triangleq \frac{1}{n} |\{v \in [n] : y_v = 1\}|$, is an unknown quantity.¹

A. Two-latent variable stochastic block model

We highlight the specifics of a two-latent variable stochastic block model for the purposes of upcoming calculations. The probability of an edge drawn between two nodes v, u is characterized by four constants, q_0, q_1, q_2, q_3 such that:

$$A_{ij} \sim \begin{cases} \text{Bern}(q_0 \frac{\log n}{n}) & \text{if } x_v = x_u, y_v = y_u \\ \text{Bern}(q_1 \frac{\log n}{n}) & \text{if } x_v \neq x_u, y_v = y_u \\ \text{Bern}(q_2 \frac{\log n}{n}) & \text{if } x_v = x_u, y_v \neq y_u \\ \text{Bern}(q_3 \frac{\log n}{n}) & \text{if } x_v \neq x_u, y_v \neq y_u \end{cases}.$$

The corresponding matrix Q , as defined earlier, in this case will be:

$$Q = \begin{bmatrix} q_0 & q_1 & q_2 & q_3 \\ q_1 & q_0 & q_3 & q_2 \\ q_2 & q_3 & q_0 & q_1 \\ q_3 & q_2 & q_1 & q_0 \end{bmatrix}. \quad (2)$$

1) *Recovering x when y is known:* In the first scenario, given an observation of the graph A and y which corresponds to the observed graph, the latent variable x_v is recovered exactly for each node $v \in [n]$. In this part, y is considered as an observation which helps the estimator to recover the desired latent variable x . Let $W \triangleq yy^T$ and $B \triangleq W * A$. Since x is chosen uniformly over $\{x \in \{\pm 1\}^n : x^T \mathbf{1} = 0\}$, the maximum likelihood estimator gives the optimal solution. For this configuration, the log-likelihood is

$$\log \mathbb{P}(A|x, y) = \frac{T_1}{8} x^T B x + \frac{T_2}{8} x^T A x + c,$$

¹Note that semidefinite programming results in this section are obtained for binary equal-sized communities, while the results of Section III were more general.

where $T_1 \triangleq \log(\frac{q_0 q_3}{q_2 q_1})$ and $T_2 \triangleq \log(\frac{q_0 q_2}{q_1 q_3})$, as $n \rightarrow \infty$ and c is a constant. Considering the constraints, the maximum likelihood estimator is,

$$\begin{aligned} \hat{x} = \arg \max_x & T_1 x^T B x + T_2 x^T A x \\ \text{subject to} & x_i \in \{\pm 1\}, \quad i \in [n] \\ & x^T \mathbf{1} = 0, \end{aligned} \quad (3)$$

which is a non-convex optimization problem. Let $Z = xx^T$. Reorganizing (3),

$$\begin{aligned} \hat{Z} = \arg \max_Z & \langle Z, T_1 B + T_2 A \rangle \\ \text{subject to} & Z = xx^T \\ & Z_{ii} = 1, \quad i \in [n] \\ & \langle Z, \mathbf{J} \rangle = 0. \end{aligned} \quad (4)$$

By relaxing the rank-one constraint on Z , we obtain the following semidefinite programming relaxation of (4):

$$\begin{aligned} \hat{Z} = \arg \max_Z & \langle Z, T_1 B + T_2 A \rangle \\ \text{subject to} & Z \succeq 0 \\ & Z_{ii} = 1, \quad i \in [n] \\ & \langle Z, \mathbf{J} \rangle = 0. \end{aligned} \quad (5)$$

For convenience define

$$\eta_1(\mathbf{q}, \rho) \triangleq \frac{\rho}{2} (\sqrt{q_0} - \sqrt{q_1})^2 + \frac{1-\rho}{2} (\sqrt{q_2} - \sqrt{q_3})^2,$$

where $\mathbf{q} \triangleq [q_0, q_1, q_2, q_3]$.

Theorem 7. Under the two-latent variable stochastic block model with binary alphabet where the latent variable y has been revealed, if

$$\begin{cases} \eta_1(\mathbf{q}, \rho) > 1 & \text{when } \rho \leq 0.5 \\ \eta_1(\mathbf{q}, 1 - \rho) > 1 & \text{when } \rho > 0.5 \end{cases}$$

then the semidefinite programming estimator is asymptotically optimal, i.e., $\mathbb{P}(\hat{Z} = Z^*) \geq 1 - o(1)$. Also, if

$$\begin{cases} \eta_1(\mathbf{q}, \rho) < 1 & \text{when } \rho \leq 0.5 \\ \eta_1(\mathbf{q}, 1 - \rho) < 1 & \text{when } \rho > 0.5 \end{cases}$$

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

Proof. See Appendix H. \square

2) *Recovering x when y is unknown:* Given an observation of the graph A , the aim is to exactly recover x while both latent variables x and y are unknown latent variables. It is assumed that the estimator does not know anything about the auxiliary latent variable y , which its prior distribution is uniform over $\{y : y \in \{\pm 1\}^n\}$. Notice that x is drawn uniformly from $\{x \in \{\pm 1\}^n : x^T \mathbf{1} = 0\}$. The log-likelihood of A given x and y is

$$\log \mathbb{P}(A|x, y) = \frac{T_1}{8} y^T (A * xx^T) y + \frac{T_2}{8} x^T A x + \frac{T_3}{8} y^T A y + c,$$

where $T_1 \triangleq \log(\frac{q_0 q_3}{q_2 q_1})$, $T_2 \triangleq \log(\frac{q_0 q_2}{q_1 q_3})$, and $T_3 \triangleq \log(\frac{q_0 q_1}{q_2 q_3})$, as $n \rightarrow \infty$ and c is a constant. Then

$$\begin{aligned} \log \mathbb{P}(A|x) &\propto \log \sum_{\mathcal{Y}} \mathbb{P}(A|x, y) \\ &\propto \log \sum_{\mathcal{Y}} e^{\frac{T_1}{T_3} y^T (A * x x^T) y + \frac{T_2}{T_3} x^T A x + y^T A y} \\ &= \frac{T_1 + T_2}{T_3} x^T A x + \sum_i \sum_j A_{ij} \\ &\quad + \log \sum_{\mathcal{Y}} e^{\frac{T_1}{T_3} y^T (A * x x^T) y + y^T A y - \frac{T_1}{T_3} x^T A x - \sum_i \sum_j A_{ij}}. \end{aligned}$$

Applying the log-sum-exp approximation, the maximum likelihood estimator is

$$\begin{aligned} \hat{x} &= \arg \max_x x^T A x \\ \text{subject to } &x_i \in \{\pm 1\}, \quad i \in [n] \\ &x^T \mathbf{1} = 0, \end{aligned} \quad (6)$$

that is a non-convex optimization problem. Let $Z = x x^T$. Reorganizing (6) yields

$$\begin{aligned} \hat{Z} &= \arg \max_Z \langle Z, A \rangle \\ \text{subject to } &Z = x x^T \\ &Z_{ii} = 1, \quad i \in [n] \\ &\langle Z, \mathbf{J} \rangle = 0. \end{aligned} \quad (7)$$

Relaxing the rank-one constraint on Z , we obtain the following semidefinite programming relaxation of (7):

$$\begin{aligned} \hat{Z} &= \arg \max_Z \langle Z, A \rangle \\ \text{subject to } &Z \succeq 0 \\ &Z_{ii} = 1, \quad i \in [n] \\ &\langle Z, \mathbf{J} \rangle = 0. \end{aligned} \quad (8)$$

For convenience define

$$\eta_2(\mathbf{q}, \rho) \triangleq \frac{1}{2} \left(\sqrt{q_0 \rho + q_2(1-\rho)} - \sqrt{q_1 \rho + q_3(1-\rho)} \right)^2.$$

Theorem 8. *Under the two-latent variable stochastic block model with binary alphabet, if*

$$\min \{ \eta_2(\mathbf{q}, \rho), \eta_2(\mathbf{q}, 1-\rho) \} > 1,$$

then the semidefinite programming estimator is asymptotically optimal, i.e., $\mathbb{P}(\hat{Z} = Z^) \geq 1 - o(1)$. Also, if*

$$\min \{ \eta_2(\mathbf{q}, \rho), \eta_2(\mathbf{q}, 1-\rho) \} < 1,$$

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

Proof. See Appendix J. \square

Remark 3. *The results of Theorems 7 and 8 are consistent with Theorems 2 and 3, respectively.*

Remark 4. *The constraint $x^T \mathbf{1} = 0$ that has been considered for this part results in a well-defined phase transition threshold for exact recovery of latent variable x . In general, x may be a random variable which is drawn uniformly from $\{x \in \{\pm 1\}^n :$*

$x^T \mathbf{1} = (2\rho_x - 1)n\}$, where $\rho_x \triangleq \frac{1}{n} |\{v \in [n] : x_v = 1\}|$. Then $x^T \mathbf{1} = 0$ is substituted by $x^T \mathbf{1} = (2\rho_x - 1)n$ in semidefinite programming relaxations (5) and (8). Also, due to the robustness of semidefinite programming, an approximation of ρ_x can be replaced for recovering the latent variable x . Investigating the constraint $x^T \mathbf{1} = (2\rho_x - 1)n$ and the robustness of semidefinite programming are beyond the scope of this paper.

B. Two-latent variable censored block model

We highlight the specifics of a two-latent variable censored block model for the purposes of upcoming calculations. Let $P(k; q_0, \xi)$ be a discrete probability density function with parameters $q_0 > 0$ and $\xi \in [0, 1]$ as,

$$\begin{aligned} P(k; q_0, \xi) &\triangleq \xi q_0 \frac{\log n}{n} \delta[k-1] + (1-\xi) q_0 \frac{\log n}{n} \delta[k+1] \\ &\quad + \left(1 - q_0 \frac{\log n}{n}\right) \delta[k], \end{aligned}$$

where δ is Dirac delta function. The probability of an edge drawn between two nodes v, u is characterized by constants q_0, q_1, q_2, q_3 and ξ such that:

$$A_{ij} \sim \begin{cases} P(k; q_0, 1-\xi) & \text{if } x_v = x_u, y_v = y_u \\ P(k; q_1, \xi) & \text{if } x_v \neq x_u, y_v = y_u \\ P(k; q_2, \xi) & \text{if } x_v = x_u, y_v \neq y_u \\ P(k; q_3, \xi) & \text{if } x_v \neq x_u, y_v \neq y_u \end{cases}.$$

The corresponding matrix Q , as defined earlier, is the same as (2). Also, in this case, the corresponding matrix Ξ will be:

$$\Xi = \begin{bmatrix} (1-\xi) & \xi & \xi & \xi \\ \xi & (1-\xi) & \xi & \xi \\ \xi & \xi & (1-\xi) & \xi \\ \xi & \xi & \xi & (1-\xi) \end{bmatrix}. \quad (9)$$

1) *Recovering x when y is known:* Given an observation of the graph A and y which corresponds to the observed graph, the latent variable x_v is recovered exactly for each node $v \in [n]$. In this part, y is considered as an observation which helps the estimator to recover the desired latent variable x . Let

$$R \triangleq T A + T(A * W) + T_1(A * A * W) + T_2(A * A),$$

where $T \triangleq \log(\frac{1-\xi}{\xi})$ and $W \triangleq y y^T$. Since x is chosen uniformly over $\{x \in \{\pm 1\}^n : x^T \mathbf{1} = 0\}$, the maximum likelihood estimator gives the optimal solution. Similar to Section IV-A1, it can be shown that the semidefinite programming relaxation of maximum likelihood estimator for this configuration is

$$\begin{aligned} \hat{Z} &= \arg \max_Z \langle Z, R \rangle \\ \text{subject to } &Z \succeq 0 \\ &Z_{ii} = 1, \quad i \in [n] \\ &\langle Z, \mathbf{J} \rangle = 0. \end{aligned} \quad (10)$$

For convenience define

$$\begin{aligned} \mathbf{g} &\triangleq [(1-\xi)q_0, \xi q_1, \xi q_2, \xi q_3], \\ \mathbf{h} &\triangleq [\xi q_0, (1-\xi)q_1, (1-\xi)q_2, (1-\xi)q_3]. \end{aligned}$$

Theorem 9. Under the two-latent variable censored block model with binary alphabet where the latent variable y has been revealed, if

$$\begin{cases} \eta_1(\mathbf{g}, \rho) + \eta_1(\mathbf{h}, \rho) > 1 & \text{when } \rho \leq 0.5 \\ \eta_1(\mathbf{g}, 1 - \rho) + \eta_1(\mathbf{h}, 1 - \rho) > 1 & \text{when } \rho > 0.5 \end{cases}$$

then the semidefinite programming estimator is asymptotically optimal, i.e., $\mathbb{P}(\hat{Z} = Z^*) \geq 1 - o(1)$. Also, if

$$\begin{cases} \eta_1(\mathbf{g}, \rho) + \eta_1(\mathbf{h}, \rho) < 1 & \text{when } \rho \leq 0.5 \\ \eta_1(\mathbf{g}, 1 - \rho) + \eta_1(\mathbf{h}, 1 - \rho) < 1 & \text{when } \rho > 0.5 \end{cases}$$

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

Proof. See Appendix K. \square

2) *Recovering x when y is unknown:* Given an observation of the graph A , the aim is to exactly recover x while both latent variables x and y are unknown. It is assumed that the estimator does not know anything about the auxiliary latent variable y , which its prior distribution is uniform over $\{y : y \in \{\pm 1\}^n\}$. Notice that x is drawn uniformly from $\{x \in \{\pm 1\}^n : x^T \mathbf{1} = 0\}$. Similar to Section IV-A2, it can be shown that for this configuration the semidefinite programming relaxation of the maximum likelihood estimator is

$$\begin{aligned} \hat{Z} = \arg \max_Z & \langle Z, T A + T_2(A * A) \rangle \\ \text{subject to} & \quad Z \succeq 0 \\ & \quad Z_{ii} = 1, \quad i \in [n] \\ & \quad \langle Z, \mathbf{J} \rangle = 0. \end{aligned} \quad (11)$$

Theorem 10. Under the two-latent variable censored block model with binary alphabet, if

$$\min \{ \eta_2(\mathbf{g}, \rho) + \eta_2(\mathbf{h}, \rho), \eta_2(\mathbf{g}, 1 - \rho) + \eta_2(\mathbf{h}, 1 - \rho) \} > 1,$$

then the semidefinite programming estimator is asymptotically optimal, i.e., $\mathbb{P}(\hat{Z} = Z^*) \geq 1 - o(1)$. Also, if

$$\min \{ \eta_2(\mathbf{g}, \rho) + \eta_2(\mathbf{h}, \rho), \eta_2(\mathbf{g}, 1 - \rho) + \eta_2(\mathbf{h}, 1 - \rho) \} < 1,$$

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

Proof. See Appendix L. \square

Remark 5. The results of Theorems 9 and 10 are consistent with Theorems 5 and 6, respectively.

V. DISCUSSION & NUMERICAL RESULTS

It is illuminating to review the flow of the development of the achievability results throughtout this paper:

- 1) Calculate the Lagrangian of the corresponding optimization
- 2) Extract the dual optimal solution based on the Lagrange multipliers
- 3) Show that $\hat{Z} = Z^*$ is primal optimal solution
- 4) Show that $\hat{Z} = Z^*$ is unique
- 5) Extract the conditions under which the dual optimal solution holds

The converses follow the following sequence:

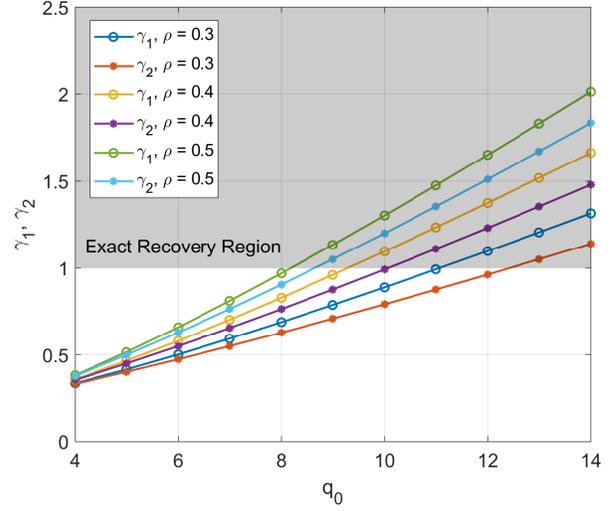


Fig. 2: Exact recovery region of x in the context of Eq. (2), with $q_2 = 3, q_1 = q_3 = 1$.

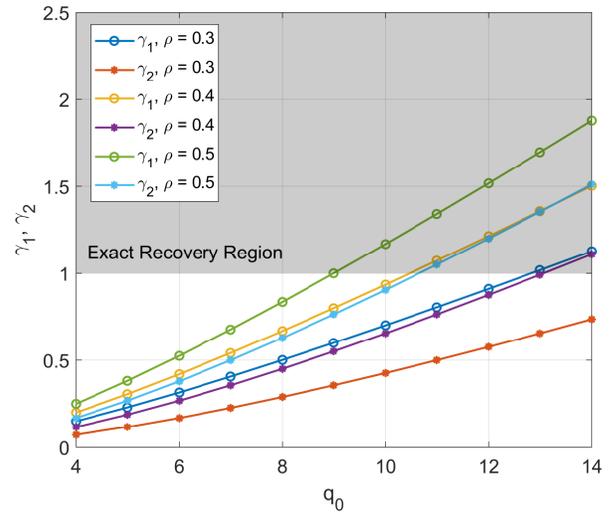


Fig. 3: Exact recovery region of x in the context of Eq. (2), with $q_1 = q_2 = q_3 = 1$.

- 1) Extract the maximum likelihood estimator
- 2) Extract the conditions under which the maximum likelihood estimator fails

To give a pictorial view of some results of the paper, we plot some results in the context of the two-latent variable stochastic block model represented by (2) and two-latent variable censored block model represented by (2) and (9). For ease of notation, we define

$$\begin{aligned} \gamma_1 &\triangleq \min\{\eta_1(\mathbf{q}, \rho), \eta_1(\mathbf{q}, 1 - \rho)\}, \\ \gamma_2 &\triangleq \min\{\eta_2(\mathbf{q}, \rho), \eta_2(\mathbf{q}, 1 - \rho)\}, \\ \gamma_3 &\triangleq \min\{\eta_1(\mathbf{g}, \rho) + \eta_1(\mathbf{h}, \rho), \eta_1(\mathbf{g}, 1 - \rho) + \eta_1(\mathbf{h}, 1 - \rho)\}, \\ \gamma_4 &\triangleq \min\{\eta_2(\mathbf{g}, \rho) + \eta_2(\mathbf{h}, \rho), \eta_2(\mathbf{g}, 1 - \rho) + \eta_2(\mathbf{h}, 1 - \rho)\}. \end{aligned}$$

For the two-latent variable stochastic block model, Figures 2 and 3 show the exact recovery region for recovering the latent

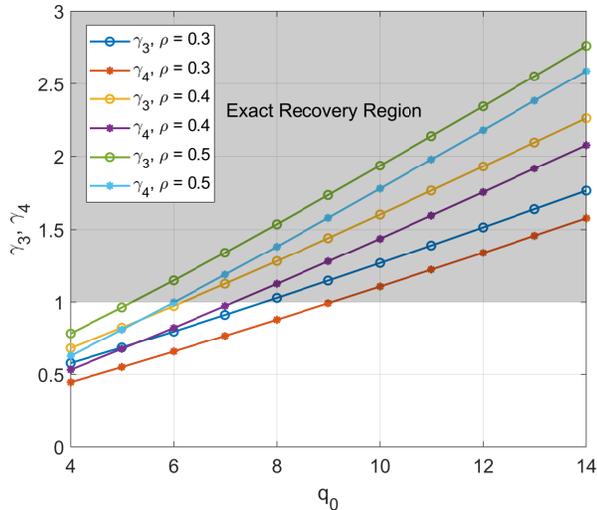


Fig. 4: Exact recovery region of x in the context of Eq. (2) and Eq. (9), with $\xi = 0.1$, $q_2 = 3$, and $q_1 = q_3 = 1$.

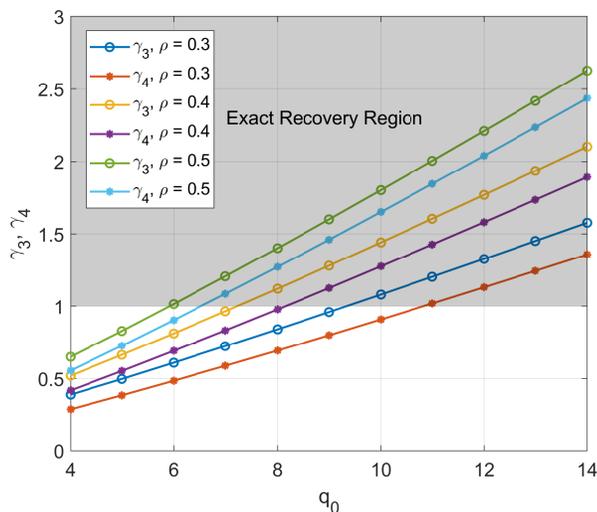


Fig. 5: Exact recovery region of x in the context of Eq. (2) and Eq. (9), with $\xi = 0.1$, and $q_1 = q_2 = q_3 = 1$.

variable x when the secondary latent variable y is either known or unknown. The curves in these figures are based on the obtained results in Theorem 7 and Theorem 8. These figures encompass several curves plotted for different values of q_0 , q_1 , q_2 , q_3 in (2), and ρ . At each figure, we consider fixed values for q_1 , q_2 , q_3 and vary the values of q_0 and ρ . A comparison between the curves in Figures 2 and 3 clarifies the role of the revealed latent variable y for recovering the desired latent variable x .

For the two-latent variable censored block model, Figures 4 and 5 show the exact recovery region for recovering the latent variable x when the secondary latent variable y is either known or unknown. The curves in these figures are based on the obtained results in Theorem 9 and Theorem 10. These figures consist of several curves plotted for different values of q_0 , q_1 ,

y	n	BSBM		BCBM		
		q_0	AEP	q_0	ξ	AEP
Known	100	7	3.8×10^{-2}	4	0.1	6.7×10^{-2}
Known	200	7	2.4×10^{-2}	4	0.1	4.9×10^{-2}
Known	300	7	1.9×10^{-2}	4	0.1	3.6×10^{-2}
Known	400	7	1.5×10^{-2}	4	0.1	2.5×10^{-2}
Known	500	7	1.1×10^{-2}	4	0.1	1.6×10^{-2}
Known	100	9	8.1×10^{-5}	6	0.1	4.6×10^{-5}
Known	200	9	5.9×10^{-5}	6	0.1	3.2×10^{-5}
Known	300	9	4.2×10^{-5}	6	0.1	2.4×10^{-5}
Known	400	9	2.8×10^{-5}	6	0.1	1.7×10^{-5}
Known	500	9	1.8×10^{-5}	6	0.1	1.2×10^{-5}
Unknown	100	8	5.7×10^{-2}	5	0.1	5.6×10^{-2}
Unknown	200	8	4.1×10^{-2}	5	0.1	3.9×10^{-2}
Unknown	300	8	2.7×10^{-2}	5	0.1	2.5×10^{-2}
Unknown	400	8	1.8×10^{-2}	5	0.1	1.6×10^{-2}
Unknown	500	8	1.3×10^{-2}	5	0.1	1.1×10^{-2}
Unknown	100	10	6.2×10^{-5}	7	0.1	6.3×10^{-5}
Unknown	200	10	4.4×10^{-5}	7	0.1	4.0×10^{-5}
Unknown	300	10	3.3×10^{-5}	7	0.1	2.3×10^{-5}
Unknown	400	10	2.3×10^{-5}	7	0.1	1.7×10^{-5}
Unknown	500	10	1.4×10^{-5}	7	0.1	1.3×10^{-5}

TABLE I: Semidefinite programming optimization of (8) and (10), with $q_2 = 3$, $q_1 = q_3 = 1$, and $\rho = 0.5$.

q_2 , q_3 in (2) and ρ , while $\xi = 0.1$ in (9). At each figure, we consider fixed values for ξ , q_1 , q_2 , q_3 and vary the values of q_0 and ρ . A comparison between the curves in Figures 4 and 5 clarifies the role of the revealed latent variable y for recovering the desired latent variable x .

To gain an understanding of the scope of our asymptotic results, under the conditions of Figures 2 and 4, we performed several simulations on 10^4 graph realizations with various graph sizes obtained from the proposed models in Section II. The obtained average error probability (AEP) is around 10^{-5} in the regimes just inside the region of exact recovery, and around 10^{-2} in the regimes just outside the region of exact recovery. The details of these simulations are represented in Tables I and II. At each simulation, we consider fixed values for q_1 , q_2 , q_3 and vary the values of q_0 , ρ , and n .

VI. CONCLUSION

This paper presents and analyzes a new generalization of the stochastic and censored block models in which, in addition to the latent variable representing community labels, there exists another (secondary) latent variables that are not part of community detection. These secondary latent variables may be known, unknown, or partially known. This model represents community detection problems where the community labels alone does not explain all the dependencies between the graph edges.

We investigate the exact recovery threshold for these models under maximum likelihood detection, and also analyze a semidefinite programming algorithm for recovering the desired latent variable under the two-latent variable stochastic block model and the two-latent variable censored block model for both scenarios.

y	n	BSBM		BCBM		
		q_0	AEP	q_0	ξ	AEP
Known	100	10	7.6×10^{-2}	7	0.1	4.1×10^{-2}
Known	200	10	5.1×10^{-2}	7	0.1	3.1×10^{-2}
Known	300	10	3.0×10^{-2}	7	0.1	2.3×10^{-2}
Known	400	10	2.1×10^{-2}	7	0.1	1.8×10^{-2}
Known	500	10	1.3×10^{-2}	7	0.1	1.3×10^{-2}
Known	100	12	6.7×10^{-5}	9	0.1	3.9×10^{-5}
Known	200	12	5.1×10^{-5}	9	0.1	2.5×10^{-5}
Known	300	12	3.6×10^{-5}	9	0.1	1.8×10^{-5}
Known	400	12	2.5×10^{-5}	9	0.1	1.2×10^{-5}
Known	500	12	1.6×10^{-5}	9	0.1	1.0×10^{-5}
Unknown	100	11	4.3×10^{-2}	8	0.1	4.2×10^{-2}
Unknown	200	11	3.3×10^{-2}	8	0.1	2.9×10^{-2}
Unknown	300	11	2.4×10^{-2}	8	0.1	2.0×10^{-2}
Unknown	400	11	1.7×10^{-2}	8	0.1	1.3×10^{-2}
Unknown	500	11	1.2×10^{-2}	8	0.1	1.0×10^{-2}
Unknown	100	13	4.2×10^{-5}	10	0.1	4.8×10^{-5}
Unknown	200	13	2.6×10^{-5}	10	0.1	3.3×10^{-5}
Unknown	300	13	1.7×10^{-5}	10	0.1	2.2×10^{-5}
Unknown	400	13	1.3×10^{-5}	10	0.1	1.5×10^{-5}
Unknown	500	13	1.1×10^{-5}	10	0.1	1.0×10^{-5}

TABLE II: Semidefinite programming optimization of (8) and (10), with $q_2 = 3, q_1 = q_3 = 1$, and $\rho = 0.3$.

APPENDIX A PROOF OF LEMMA 1

Define

$$f_1(t) \triangleq \prod_{i=1}^m \left(\frac{b_i}{a_i} \right)^{(t-1)d_i} e^{(t-1)(a_i-b_i)},$$

$$f_2(t) \triangleq \prod_{i=1}^m \left(\frac{b_i}{a_i} \right)^{td_i} e^{t(a_i-b_i)}.$$

For any $t \in [0, 1]$,

$$\begin{aligned} & \sum_{d \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d)p, \mathcal{P}_b(d)\hat{p}\} \\ & \leq \max\{p, \hat{p}\} \sum_{d \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d), \mathcal{P}_b(d)\} \\ & = \max\{p, \hat{p}\} \exp\left(-\sum_i [ta_i + (1-t)b_i - a_i^t b_i^{1-t}]\right) \\ & \quad \times \sum_{d \in \mathbb{Z}_+^m} \prod_i \frac{(a_i^t b_i^{1-t})^{d_i}}{d_i!} e^{-a_i^t b_i^{1-t}} \min\{f_1(t), f_2(t)\}. \end{aligned}$$

Both $f_1(t)$ and $f_2(t)$ are monotonic and $\frac{f_2(t)}{f_1(t)}$ is a positive constant (does not depend on t), thus $\min\{f_1, f_2\}$ is also monotonic in t . Since $f_1(1) = f_2(0) = 1$, for all t we have:

$$\min\{f_1(t), f_2(t)\} \leq 1.$$

Notice that

$$\sum_{d \in \mathbb{Z}_+^m} \prod_i \frac{(a_i^t b_i^{1-t})^{d_i}}{d_i!} e^{-a_i^t b_i^{1-t}} = 1.$$

Then

$$I(a, b) \leq \max\{p, \hat{p}\} e^{-\sum_{i=1}^m [ta_i + (1-t)b_i - a_i^t b_i^{1-t}]}. \quad (12)$$

For the value of t that maximizes the right-hand side of inequality (12), we have

$$\sum_{d \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d)p, \mathcal{P}_b(d)\hat{p}\} \leq \max\{p, \hat{p}\} e^{-\text{Div}(a, b)}.$$

Notice that t^* satisfies

$$\prod_{i=1}^m \left(\frac{b_i}{a_i} \right)^{a_i^{t^*} b_i^{1-t^*}} e^{a_i - b_i} = 1.$$

Then at the optimal t^* ,

$$\begin{aligned} & \sum_{d \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d)p, \mathcal{P}_b(d)\hat{p}\} \\ & \geq \min\{p, \hat{p}\} \sum_{d \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d), \mathcal{P}_b(d)\} \\ & \stackrel{(a)}{\geq} \min\{p, \hat{p}\} e^{-\text{Div}(a, b)} \prod_i \frac{(a_i^{t^*} b_i^{1-t^*})^{a_i^{t^*} b_i^{1-t^*}}}{a_i^{t^*} b_i^{1-t^*}!} e^{-a_i^{t^*} b_i^{1-t^*}} \\ & \stackrel{(b)}{\geq} \min\{p, \hat{p}\} e^{-\text{Div}(a, b)} \prod_i \frac{1}{e} (a_i^{t^*} b_i^{1-t^*})^{-\frac{1}{2}}, \end{aligned}$$

where (a) holds because

$$\sum_{d \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d), \mathcal{P}_b(d)\} \geq \min\{\mathcal{P}_a(d^*), \mathcal{P}_b(d^*)\},$$

where d^* is defined by $d_i^* \triangleq a_i^{t^*} b_i^{1-t^*}$, and (b) is due to Stirling's approximation $n! \leq n^{n+\frac{1}{2}} e^{-n+1}$ for any $n \geq 1$.

APPENDIX B PROOF OF THEOREM 2

We aim to recover x_v when y_v is known. Given a realization of D and y_v , our goal is to minimize the error probability by selecting the most likely hypothesis, i.e.,

$$\arg\max_i \mathbb{P}\{H_i | D = d, y_v\},$$

or equivalently, since d, y_v are known observations,

$$\arg\max_i P(d|H_i, y_v) \mathbb{P}\{H_i, y_v\},$$

which is the maximum a posteriori (MAP) detector, which we rewrite:

$$\arg\max_i P(d|H_i, y_v) P_{i, y_v}. \quad (13)$$

Solving (13) requires $m_x - 1$ pairwise comparisons of the hypotheses. From this viewpoint, if

$$P(d|H_i, y_v) P_{i, y_v} \leq P(d|H_k, y_v) P_{k, y_v}, \quad (14)$$

then a pairwise comparison will choose H_k over H_i . Now assume the correct hypothesis is H_i , and denote by \mathcal{B}_{ik} the region of D for which (14) is satisfied, i.e., H_i has a *worse* metric compared with H_k . Also denote by \mathcal{B}_i the region for D where the overall MAP decoder is in error. The dependence of error regions \mathcal{B}_{ik} and \mathcal{B}_i on y_v is implicit. Then the probability of error

$$P_e = \sum_i \mathbb{P}\{D \in \mathcal{B}_i | H_i, y_v\} P_{i, y_v}. \quad (15)$$

Since $\mathcal{B}_i \subset \cup_k \mathcal{B}_{ik}$,

$$P_e \leq \sum_i \sum_{k \neq i} \mathbb{P}\{D \in \mathcal{B}_{ik} | H_i, y_v\} P_{i, y_v}.$$

From the earlier Poisson assumption $P(d|H_i, y_v) = \mathcal{P}_{\lambda_i, y_v}(d)$ it follows that:

$$\min\{\mathcal{P}_{\lambda_i, y_v}(d) P_{i, y_v}, \mathcal{P}_{\lambda_k, y_v}(d) P_{k, y_v}\} = \begin{cases} \mathcal{P}_{\lambda_i, y_v}(d) P_{i, y_v} & \text{when } D \in \mathcal{B}_{ik} \\ \mathcal{P}_{\lambda_k, y_v}(d) P_{k, y_v} & \text{when } D \in \mathcal{B}_{ik}^c \end{cases}.$$

Therefore, substituting into the union bound:

$$P_e \leq \sum_d \sum_i \sum_{k > i} \min\{\mathcal{P}_{\lambda_i, y_v}(d) P_{i, y_v}, \mathcal{P}_{\lambda_k, y_v}(d) P_{k, y_v}\}. \quad (16)$$

For bounding the error probability (16), it suffices to find an upper bound for

$$\sum_d \min\{\mathcal{P}_{\lambda_i, y_v}(d) P_{i, y_v}, \mathcal{P}_{\lambda_k, y_v}(d) P_{k, y_v}\}. \quad (17)$$

It follows from Lemma 1 that

$$\begin{aligned} P_e &\leq \sum_i \sum_{k > i} \max\{P_{i, y_v}, P_{k, y_v}\} e^{-\text{Div}(\lambda_{i, y_v}, \lambda_{k, y_v})} \\ &= \sum_i \sum_{k > i} n^{-\text{Div}(q^{(i, y_v)}, q^{(k, y_v)}) + o(1)}. \end{aligned} \quad (18)$$

We now bound the error probability of decoding rule (13) from below. Since

$$\sum_{k \neq i} \mathbb{P}\{D \in \mathcal{B}_{ik} | H_i, y_k\} \leq (m_x - 1) \mathbb{P}\{D \in \mathcal{B}_i | H_i, y_v\}, \quad (19)$$

substituting (19) into (15) yields

$$\begin{aligned} P_e &\geq \frac{1}{m_x - 1} \sum_i \sum_{k \neq i} \mathbb{P}\{D \in \mathcal{B}_{ik} | H_i, y_v\} P_{i, y_v} = \frac{1}{m_x - 1} \\ &\times \sum_i \sum_{k > i} \sum_d \min\{\mathcal{P}_{\lambda_i, y_v}(d) P_{i, y_v}, \mathcal{P}_{\lambda_k, y_v}(d) P_{k, y_v}\}. \end{aligned}$$

Then it suffices to find a lower bound for (17) to bound the error probability from below. It follows from Lemma 1 that

$$\begin{aligned} P_e &\geq \sum_i \sum_{k > i} c' \min\{P_{i, y_v}, P_{k, y_v}\} (\log n)^{-\frac{m}{2}} e^{-\text{Div}(\lambda_{i, y_v}, \lambda_{k, y_v})} \\ &= \sum_i \sum_{k > i} n^{-\text{Div}(q^{(i, y_v)}, q^{(k, y_v)}) + o(1)}, \end{aligned} \quad (20)$$

where c' is a constant and m is the number of elements in vector d , i.e., the product of alphabet sizes of x_v and y_v . The lower and upper bounds (18) and (20) imply that the true hypothesis is recovered correctly if $\text{Div}(q^{(i, y_v)}, q^{(k, y_v)}) > 1$, for a given y_v and any $i \neq k$. This means that a known latent variable restricts the number of pairwise comparisons. Then under the two-latent variable stochastic block model in which the latent variable y is known, and the latent variable x is unknown, exact recovery is possible for x if and only if

$$\min_j \min_{i \neq k} \text{Div}(q^{(i, j)}, q^{(k, j)}) > 1. \quad (21)$$

APPENDIX C PROOF OF THEOREM 3

We aim to recover x_v when y_v is unknown, given a realization of D for node v . For this setting the MAP detector is

$$\operatorname{argmax}_i \mathbb{P}\{H_i | D = d\},$$

or equivalently,

$$\operatorname{argmax}_i \sum_{y_v} \prod_l P\left(\sum_j d^{(l, j)} | H_i, y_v\right) P_{i, y_v}. \quad (22)$$

Solving (22) requires $m_x - 1$ pairwise comparisons. In these comparisons, if

$$\begin{aligned} \sum_{y_v} \prod_l P\left(\sum_j d^{(l, j)} | H_i, y_v\right) P_{i, y_v} \\ < \sum_{y_v} \prod_l P\left(\sum_j d^{(l, j)} | H_k, y_v\right) P_{k, y_v}, \end{aligned} \quad (23)$$

then we conclude hypothesis H_i is ruled out, i.e., $x_v \neq i$, because another hypothesis H_k has a better metric. Denote by \mathcal{B}_{ik} the region of D for which H_i has a *worse* metric compared with H_k , i.e., the region for D in which (23) is satisfied. Also denote by \mathcal{B}_i the region for D where the overall MAP decoder is in error. The error probability of MAP decoder (22) is given by

$$P_e = \sum_i \sum_{y_v} \mathbb{P}\{D \in \mathcal{B}_i | H_i, y_v\} P_{i, y_v}. \quad (24)$$

Since $\mathcal{B}_i \subset \cup_k \mathcal{B}_{ik}$, via the union bound,

$$\begin{aligned} \sum_{y_v} \mathbb{P}\{D \in \mathcal{B}_i | H_i, y_v\} P_{i, y_v} \\ \leq \sum_{y_v} \sum_{k \neq i} \mathbb{P}\{D \in \mathcal{B}_{ik} | H_i, y_v\} P_{i, y_v}. \end{aligned} \quad (25)$$

Using the Poisson approximation and the additive property of Poisson distribution:

$$\begin{aligned} I(d, i, y_v) &\triangleq \prod_l P\left(\sum_j d^{(l, j)} | H_i, y_v\right) \\ &= \prod_l \mathcal{P}_{\sum_j \lambda_{i, y_v}^{(l, j)}}\left(\sum_j d^{(l, j)}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \min\{I(d, i, y_v) P_{i, y_v}, I(d, k, y_v) P_{k, y_v}\} = \\ \begin{cases} I(d, i, y_v) P_{i, y_v} & \text{when } D \in \mathcal{B}_{ik} \\ I(d, k, y_v) P_{k, y_v} & \text{when } D \in \mathcal{B}_{ik}^c \end{cases}. \end{aligned}$$

Substituting (25) into (24) yields

$$\begin{aligned} P_e &\leq \sum_i \sum_{k \neq i} \sum_{y_v} \mathbb{P}\{D \in \mathcal{B}_{ik} | H_i, y_v\} P_{i, y_v} \\ &= \sum_i \sum_{k > i} \sum_{y_v} \sum_d \min\{I(d, i, y_v) P_{i, y_v}, I(d, k, y_v) P_{k, y_v}\}. \end{aligned} \quad (26)$$

For bounding the error probability (26) from above, it suffices to find an upper bound for

$$\sum_{d \in \mathbb{Z}_+^m} \min \left\{ I(d, i, y_v) P_{i, y_v}, I(d, k, y_v) P_{k, y_v} \right\}. \quad (27)$$

Applying Lemma 1 yields

$$P_e \leq \sum_i \sum_{k>i} \sum_{y_v} n^{-\text{Div}(\tilde{q}^{(i, y_v)}, \tilde{q}^{(k, y_v)}) + o(1)}. \quad (28)$$

We now bound the error probability of decoding rule (22) from below. Notice that

$$\sum_{k \neq i} \mathbb{P}\{D \in \mathcal{B}_{ik} | H_i, y_v\} \leq (m_x - 1) \mathbb{P}\{D \in \mathcal{B}_i | H_i, y_v\}. \quad (29)$$

Substituting (29) into (24) yields

$$P_e \geq \frac{1}{m_x - 1} \sum_i \sum_{k \neq i} \sum_{y_v} \mathbb{P}\{D \in \mathcal{B}_{ik} | H_i, y_v\} P_{i, y_v} = \frac{1}{m_x - 1} \times \sum_i \sum_{k>i} \sum_{y_v} \sum_d \min \left\{ I(d, i, y_v) P_{i, y_v}, I(d, k, y_v) P_{k, y_v} \right\}.$$

Then it suffices to find a lower bound for (27). Applying Lemma 1 yields

$$P_e \geq \sum_i \sum_{k>i} \sum_{y_v} n^{-\text{Div}(\tilde{q}^{(i, y_v)}, \tilde{q}^{(k, y_v)}) + o(1)}. \quad (30)$$

The lower and upper bounds (28) and (30) imply that the true hypothesis is recovered correctly if $\text{Div}(\tilde{q}^{(i, y_v)}, \tilde{q}^{(k, y_v)}) > 1$ for any $i \neq k$ and any y_v . Then under two-latent variable stochastic block model in which both latent variables x, y are unknown, exact recovery is solvable for x if and only if

$$\min_j \min_{i \neq k} \text{Div}(\tilde{q}^{(i, j)}, \tilde{q}^{(k, j)}) > 1. \quad (31)$$

APPENDIX D PROOF OF LEMMA 2

Define

$$f_1(t) \triangleq \left(\frac{\mathcal{P}_a(d) \mathcal{P}_{\hat{a}}(w)}{\mathcal{P}_b(d) \mathcal{P}_{\hat{b}}(w)} \right)^{1-t},$$

$$f_2(t) \triangleq \left(\frac{\mathcal{P}_b(d) \mathcal{P}_{\hat{b}}(w)}{\mathcal{P}_a(d) \mathcal{P}_{\hat{a}}(w)} \right)^t,$$

$$f(t) \triangleq \mathcal{P}_a(d)^t \mathcal{P}_b(d)^{1-t} \mathcal{P}_{\hat{a}}(w)^t \mathcal{P}_{\hat{b}}(w)^{1-t}.$$

For any $t \in [0, 1]$,

$$\begin{aligned} & \sum_{d, w \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d) \mathcal{P}_{\hat{a}}(w) p, \mathcal{P}_b(d) \mathcal{P}_{\hat{b}}(w) \hat{p}\} \\ & \leq \max\{p, \hat{p}\} \sum_{d, w \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d) \mathcal{P}_{\hat{a}}(w), \mathcal{P}_b(d) \mathcal{P}_{\hat{b}}(w)\} \\ & \leq \max\{p, \hat{p}\} \exp \left(- \sum_i [t a_i + (1-t) b_i - a_i^t b_i^{1-t}] \right) \\ & \quad \times \exp \left(- \sum_i [t \hat{a}_i + (1-t) \hat{b}_i - \hat{a}_i^t \hat{b}_i^{1-t}] \right), \quad (32) \end{aligned}$$

where the last inequality holds because $\min\{f_1(t), f_2(t)\} \leq 1$, and

$$\sum_{d, w \in \mathbb{Z}_+^m} \prod_i \frac{(a_i^t b_i^{1-t})^{d_i}}{d_i!} e^{-a_i^t b_i^{1-t}} \frac{(\hat{a}_i^t \hat{b}_i^{1-t})^{w_i}}{w_i!} e^{-\hat{a}_i^t \hat{b}_i^{1-t}} = 1.$$

For the value of t that minimizes the upper bound of (32), we have

$$I(a, b, \hat{a}, \hat{b}) \leq \max\{p, \hat{p}\} e^{-\text{Div}([a, \hat{a}], [b, \hat{b}])}.$$

Notice that t^* satisfies

$$\prod_{i=1}^m \left(\frac{b_i}{a_i} \right)^{a_i^{t^*} b_i^{1-t^*}} \left(\frac{\hat{b}_i}{\hat{a}_i} \right)^{\hat{a}_i^{t^*} \hat{b}_i^{1-t^*}} e^{a_i - b_i + \hat{a}_i - \hat{b}_i} = 1.$$

Then at the optimal t^* ,

$$\begin{aligned} & \sum_{d, w \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d) \mathcal{P}_{\hat{a}}(w) p, \mathcal{P}_b(d) \mathcal{P}_{\hat{b}}(w) \hat{p}\} \\ & \geq \min\{p, \hat{p}\} \sum_{d, w \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d) \mathcal{P}_{\hat{a}}(w), \mathcal{P}_b(d) \mathcal{P}_{\hat{b}}(w)\} \\ & \stackrel{(a)}{\geq} \min\{p, \hat{p}\} e^{-\text{Div}([a, \hat{a}], [b, \hat{b}])} \\ & \quad \times \prod_i \frac{(a_i^{t^*} b_i^{1-t^*})^{a_i^{t^*} b_i^{1-t^*}}}{a_i^{t^*} b_i^{1-t^*}!} e^{-a_i^{t^*} b_i^{1-t^*}} \\ & \quad \times \prod_i \frac{(\hat{a}_i^{t^*} \hat{b}_i^{1-t^*})^{\hat{a}_i^{t^*} \hat{b}_i^{1-t^*}}}{\hat{a}_i^{t^*} \hat{b}_i^{1-t^*}!} e^{-\hat{a}_i^{t^*} \hat{b}_i^{1-t^*}} \\ & \stackrel{(b)}{\geq} \min\{p, \hat{p}\} e^{-\text{Div}([a, \hat{a}], [b, \hat{b}])} \prod_i \frac{1}{e^2} [(a_i \hat{a}_i)^{t^*} (b_i \hat{b}_i)^{1-t^*}]^{-\frac{1}{2}}, \end{aligned}$$

where (a) holds because

$$\begin{aligned} & \sum_{d, w \in \mathbb{Z}_+^m} \min\{\mathcal{P}_a(d) \mathcal{P}_{\hat{a}}(w), \mathcal{P}_b(d) \mathcal{P}_{\hat{b}}(w)\} \\ & \geq \min\{\mathcal{P}_a(d^*) \mathcal{P}_{\hat{a}}(w^*), \mathcal{P}_b(d^*) \mathcal{P}_{\hat{b}}(w^*)\}, \end{aligned}$$

where d^* is defined by $d_i^* \triangleq a_i^{t^*} b_i^{1-t^*}$ and w^* is defined by $w_i^* \triangleq \hat{a}_i^{t^*} \hat{b}_i^{1-t^*}$, and (b) is due to Stirling's approximation $n! \leq n^{n+\frac{1}{2}} e^{-n+1}$ for any $n \geq 1$.

APPENDIX E PROOF OF THEOREM 4

We aim to recover both x_v and y_v for node v , given a realization of D and a realization of W . Our goal is to minimize the error probability by selecting the most likely hypothesis, i.e.,

$$\underset{i, j}{\text{argmax}} \mathbb{P}\{H_{i, j} | D = d, W = w\},$$

where

$$H_{i, j} : x_v = i, y_v = j.$$

The maximum a posteriori (MAP) detector is rewrite as

$$\underset{i, j}{\text{argmax}} P(d, w | H_{i, j}) P_{i, j}. \quad (33)$$

Solving (33) requires $m_x m_y - 1$ pairwise comparisons of the hypotheses. From this viewpoint, if

$$P(d, w|H_{i,j})P_{i,j} \leq P(d, w|H_{k,l})P_{k,l},$$

then a pairwise comparison will choose $H_{k,l}$ over $H_{i,j}$. Now assume the correct hypothesis is $H_{i,j}$. Similar to the proof of Theorems 2 and 3, it can be shown that the probability of error for recovering the true hypothesis is bounded from above and below by controlling

$$\sum_{d,w} \min\{\mathcal{P}_{\lambda_{i,j}}(d)\mathcal{P}_{\hat{\lambda}_{i,j}}(w)P_{i,j}, \mathcal{P}_{\lambda_{k,l}}(d)\mathcal{P}_{\hat{\lambda}_{k,l}}(w)P_{k,l}\}.$$

It follows from Lemma 2 that

$$\begin{aligned} P_e &\leq \sum_{i,k>i} \sum_{j,l>j} \max\{P_{i,j}, P_{k,l}\} e^{-\text{Div}([\lambda_{i,j}, \hat{\lambda}_{i,j}], [\lambda_{k,l}, \hat{\lambda}_{k,l}])} \\ &= \sum_{i,k>i} \sum_{j,l>j} n^{-\text{Div}([g^{(i,j)}, g^{(i,j)}], [g^{(k,l)}, g^{(k,l)}]) + o(1)}, \end{aligned} \quad (34)$$

and

$$\begin{aligned} P_e &\geq \sum_{i,k>i} \sum_{j,l>j} \frac{c' \min\{P_{i,j}, P_{k,l}\}}{(\log n)^m} e^{-\text{Div}([\lambda_{i,j}, \hat{\lambda}_{i,j}], [\lambda_{k,l}, \hat{\lambda}_{k,l}])} \\ &= \sum_{i,k>i} \sum_{j,l>j} n^{-\text{Div}([g^{(i,j)}, h^{(i,j)}], [g^{(k,l)}, h^{(k,l)}]) + o(1)}, \end{aligned} \quad (35)$$

where c' is a constant and m is the number of elements in vector d , i.e., the product of alphabet sizes of x_v and y_v . The lower and upper bounds (34) and (35) imply that the true hypothesis is recovered correctly if $\text{Div}([g^{(i,j)}, h^{(i,j)}], [g^{(k,l)}, h^{(k,l)}]) > 1$, for any $(i, j) \neq (k, l)$. This means that under the two-latent variable censored block model all micro-communities are exactly recovered if and only if

$$\min_{(i,j) \neq (k,l)} \text{Div}([g^{(i,j)}, h^{(i,j)}], [g^{(k,l)}, h^{(k,l)}]) > 1.$$

APPENDIX F PROOF OF THEOREM 5

We aim to recover x_v when y_v is known. Given a realization of D , a realization of W , and y_v , our goal is to minimize the error probability by selecting the most likely hypothesis, i.e.,

$$\arg\max_i \mathbb{P}\{H_i | D = d, W = w, y_v\},$$

or equivalently,

$$\arg\max_i P(d|H_i, y_v)P(w|H_i, y_v)P_{i,y_v}, \quad (36)$$

which is the MAP detector. Solving (36) requires $m_x - 1$ pairwise comparisons of the hypotheses. Similar to the proof of Theorem 2, it can be shown that the error probability of finding true hypothesis is bounded from above and below by controlling

$$\sum_{d,w} \min\{\mathcal{P}_{\lambda_{i,y_v}}(d)\mathcal{P}_{\hat{\lambda}_{i,y_v}}(w)P_{i,y_v}, \mathcal{P}_{\lambda_{k,y_v}}(d)\mathcal{P}_{\hat{\lambda}_{k,y_v}}(w)P_{k,y_v}\}.$$

It follows from Lemma 2 that

$$\begin{aligned} P_e &\leq \sum_i \sum_{k>i} \max\{P_{i,y_v}, P_{k,y_v}\} e^{-\text{Div}([\lambda_{i,y_v}, \hat{\lambda}_{i,y_v}], [\lambda_{k,y_v}, \hat{\lambda}_{k,y_v}])} \\ &= \sum_i \sum_{k>i} n^{-\text{Div}([g^{(i,y_v)}, h^{(i,y_v)}], [g^{(k,y_v)}, h^{(k,y_v)}]) + o(1)}. \end{aligned} \quad (37)$$

and

$$\begin{aligned} P_e &\geq \sum_i \sum_{k>i} \frac{c}{(\log n)^{\frac{m}{2}}} e^{-\text{Div}([\lambda_{i,y_v}, \hat{\lambda}_{i,y_v}], [\lambda_{k,y_v}, \hat{\lambda}_{k,y_v}])} \\ &= \sum_i \sum_{k>i} n^{-\text{Div}([g^{(i,y_v)}, h^{(i,y_v)}], [g^{(k,y_v)}, h^{(k,y_v)}]) + o(1)}, \end{aligned} \quad (38)$$

where $c \triangleq c' \min\{P_{i,y_v}, P_{k,y_v}\}$ is a constant and m is the number of elements in vector d . The lower and upper bounds (37) and (38) imply that the true hypothesis is recovered correctly if $\text{Div}([g^{(i,y_v)}, h^{(i,y_v)}], [g^{(k,y_v)}, h^{(k,y_v)}]) > 1$, for a given y_v and any $i \neq k$. This means that a known latent variable restricts the number of pairwise comparisons. Then under the two-latent variable censored block model in which the latent variable y is known, and the latent variable x is unknown, exact recovery is possible for x if and only if

$$\min_j \min_{i \neq k} \text{Div}([g^{(i,j)}, h^{(i,j)}], [g^{(k,j)}, h^{(k,j)}]) > 1.$$

APPENDIX G PROOF OF THEOREM 6

We aim to recover x_v when y_v is unknown, given a realization of D and a realization of W for node v . For this setting the MAP detector is

$$\arg\max_i \mathbb{P}\{H_i | D = d, W = w\}.$$

For convenience define

$$I(d, w, i, y_v) \triangleq \prod_l P\left(\sum_j d^{(l,j)}, \sum_j w^{(l,j)} | H_i, y_v\right),$$

where $\sum_j w^{(l,j)}$ and $\sum_j d^{(l,j)}$ are independent given H_i and y_v . Then the MAP detector rewrite as

$$\arg\max_i \sum_{y_v} I(d, w, i, y_v) P_{i,y_v}. \quad (39)$$

Solving (39) requires $m_x - 1$ pairwise comparisons. In these comparisons, if

$$\sum_{y_v} I(d, w, i, y_v) P_{i,y_v} < \sum_{y_v} I(d, w, k, y_v) P_{k,y_v},$$

then we conclude hypothesis H_i is ruled out, i.e., $x_v \neq i$, because another hypothesis H_k has a better metric. Notice that using the Poisson approximation and the additive property of Poisson distribution, $I(d, w, i, y_v)$ can be reorganized as

$$\begin{aligned} I(d, w, i, y_v) &= \prod_l \mathcal{P}_{\sum_j \lambda_{i,y_v}^{(l,j)}}\left(\sum_j d^{(l,j)}\right) \\ &\quad \times \prod_l \mathcal{P}_{\sum_j \lambda_{i,y_v}^{(l,j)}}\left(\sum_j w^{(l,j)}\right). \end{aligned}$$

Similar to the proof of Theorem 3, it can be shown that the error probability of recovering the true hypothesis is bounded from above and below by controlling

$$\sum_{d,w \in \mathbb{Z}_+^m} \min \left\{ I(d, w, i, y_v) P_{i, y_v}, I(d, w, k, y_v) P_{k, y_v} \right\}.$$

Applying Lemma 2 yields

$$P_e \leq \sum_i \sum_{k>i} \sum_{y_v} n^{-\text{Div}([\tilde{g}^{(i, y_v)}, \tilde{h}^{(i, y_v)}], [\tilde{g}^{(k, y_v)}, \tilde{h}^{(k, y_v)}]) + o(1)}, \quad (40)$$

and

$$P_e \geq \sum_i \sum_{k>i} \sum_{y_v} n^{-\text{Div}([\tilde{g}^{(i, y_v)}, \tilde{h}^{(i, y_v)}], [\tilde{g}^{(k, y_v)}, \tilde{h}^{(k, y_v)}]) + o(1)}. \quad (41)$$

The lower and upper bounds (40) and (41) imply that the true hypothesis is recovered correctly if $\text{Div}([\tilde{g}^{(i, y_v)}, \tilde{h}^{(i, y_v)}], [\tilde{g}^{(k, y_v)}, \tilde{h}^{(k, y_v)}]) > 1$ for any $i \neq k$ and any y_v . Then under two-latent variable censored block model in which both latent variables x, y are unknown, exact recovery is solvable for x if and only if

$$\min_j \min_{i \neq k} \text{Div}([\tilde{g}^{(i, j)}, \tilde{h}^{(i, j)}], [\tilde{g}^{(k, j)}, \tilde{h}^{(k, j)}]) > 1.$$

APPENDIX H PROOF OF THEOREM 7

We begin by stating sufficient conditions for the optimum solution of (5) matching the true labels x^* .

Lemma 3. *For the optimization problem (5), consider the Lagrange multipliers*

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

If we have

$$\begin{aligned} S^* &= D^* + \lambda^* \mathbf{J} - T_1 B - T_2 A, \\ S^* &\succeq 0, \\ \lambda_2(S^*) &> 0, \\ S^* x^* &= 0, \end{aligned}$$

then (λ^*, D^*, S^*) is the dual optimal solution and $\hat{Z} = x^* x^{*T}$ is the unique primal optimal solution of (5).

Proof. Let $D = \text{diag}(d_i)$, $\lambda \in \mathbb{R}$, and $S \succeq 0$ denote the Lagrangian of (5). For any Z that satisfies the constraints in (5), we have

$$\begin{aligned} T_1 \langle B, Z \rangle + T_2 \langle A, Z \rangle &\stackrel{(a)}{\leq} L(Z, S^*, D^*, \lambda^*) = \langle D^*, \mathbf{I} \rangle \\ &\stackrel{(b)}{\leq} \langle S^* - \lambda^* \mathbf{J} + T_1 B + T_2 A, Z^* \rangle \\ &\stackrel{(c)}{=} T_1 \langle B, Z^* \rangle + T_2 \langle A, Z^* \rangle, \end{aligned}$$

where (a) holds because $\langle S^*, Z \rangle \geq 0$, (b) holds because $Z_{ii} = 1$ for all $i \in [n]$ and $S^* = D^* + \lambda^* \mathbf{J} - T_1 B - T_2 A$, and (c) holds because $S^* x^* = \mathbf{0}$ and $x^{*T} \mathbf{1} = 0$. Therefore, $Z^* = x^* x^{*T}$ is

an optimal solution of (5). Now, assume \tilde{Z} is another optimal solution. Then

$$\begin{aligned} \langle S^*, \tilde{Z} \rangle &= \langle D^* + \lambda^* \mathbf{J} - T_1 B - T_2 A, \tilde{Z} \rangle \\ &\stackrel{(a)}{=} \langle D^* + \lambda^* \mathbf{J} - T_1 B - T_2 A, Z^* \rangle = \langle S^*, Z^* \rangle = 0, \end{aligned}$$

where (a) holds because $\langle T_1 B + T_2 A, Z^* \rangle = \langle T_1 B + T_2 A, \tilde{Z} \rangle$, $Z_{ii}^* = \tilde{Z}_{ii} = 1$ for all $i \in [n]$, and $\langle \mathbf{J}, Z^* \rangle = \langle \mathbf{J}, \tilde{Z} \rangle = 0$. Since $\tilde{Z} \succeq 0$, and $S^* \succeq 0$ while its second smallest eigenvalue $\lambda_2(S^*)$ is positive (since $S^* \hat{x}^* = \mathbf{0}$), \tilde{Z} must be a multiple of Z^* . Also, since $\tilde{Z}_{ii} = Z_{ii}^* = 1$ for all $i \in [n]$, we have $\tilde{Z} = Z^*$. \square

We now show that $S^* = D^* + \lambda^* \mathbf{J} - T_1 B - T_2 A$ satisfies other conditions in Lemma 3 with probability $1 - o(1)$. Let

$$d_i^* = T_1 \sum_{j=1}^n B_{ij} x_j^* x_i^* + T_2 \sum_{j=1}^n A_{ij} x_j^* x_i^*. \quad (42)$$

Then $D^* x^* = T_1 B x^* + T_2 A x^*$ and based on the definition of S^* in Lemma 3, 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

$$\mathbb{P} \left\{ \inf_{v \perp x^*, \|v\|=1} v^T S^* v > 0 \right\} \geq 1 - o(1), \quad (43)$$

where v is a $n \times 1$ vector. Then for any v such that $v^T x^* = 0$ and $\|v\| = 1$,

$$\begin{aligned} v^T S^* v &= v^T D^* v + \lambda^* v^T \mathbf{J} v - T_1 v^T (B - \mathbb{E}[B]) v \\ &\quad - T_2 v^T (A - \mathbb{E}[A]) v - T_1 v^T \mathbb{E}[B] v - T_2 v^T \mathbb{E}[A] v \\ &\geq \min_i d_i^* + \lambda^* v^T \mathbf{J} v - T_1 \|B - \mathbb{E}[B]\| \\ &\quad - T_2 \|A - \mathbb{E}[A]\| - T_1 v^T \mathbb{E}[B] v - T_2 v^T \mathbb{E}[A] v. \end{aligned}$$

Notice that

$$\begin{aligned} T_1 v^T \mathbb{E}[B] v + T_2 v^T \mathbb{E}[A] v &= \frac{1}{4} [T_1 c_1 + T_2 c_2] v^T W v \\ &\quad + \frac{1}{4} [T_1 c_3 + T_2 c_4] v^T (Z * W) v \\ &\quad + \frac{1}{4} [T_1 c_1 + T_2 c_2] v^T \mathbf{J} v \\ &\quad - (T_1 + T_2) q_0 \frac{\log n}{n}, \end{aligned}$$

where

$$\begin{aligned} c_1 &\triangleq \frac{\log n}{n} (q_0 - q_2 + q_1 - q_3), \\ c_2 &\triangleq \frac{\log n}{n} (q_0 + q_2 + q_1 + q_3), \\ c_3 &\triangleq \frac{\log n}{n} (q_0 - q_2 - q_1 + q_3), \\ c_4 &\triangleq \frac{\log n}{n} (q_0 + q_2 - q_1 + q_3). \end{aligned}$$

Lemma 4. *For any $c > 0$, there exists $c', c'' > 0$ such that for any $n \geq 1$, $\|A - \mathbb{E}[A]\| \leq c' \sqrt{\log n}$ and $\|B - \mathbb{E}[B]\| \leq c' \sqrt{\log n}$ with probability at least $1 - n^{-c}$.*

Proof. The proof is similar to the proofs [39, Theorem 9] and [32, Theorem 5]. \square

Lemma 5. *With probability at least $1 - n^{-\frac{1}{2}}$,*

$$\begin{aligned} v^T(Z * W)v &\leq \sqrt{\log n}, \\ v^T W v &\leq \sqrt{\log n} + (2\rho - 1)^2 v^T \mathbf{J} v + 2|2\rho - 1| \sqrt{n \log n}. \end{aligned}$$

Proof. Since $-|v_i| \leq v_i y_i \leq |v_i|$, by applying the Chernoff bound we have

$$\mathbb{P}(v^T y - \mathbb{E}[v^T y] \geq \sqrt{\log n}) \leq n^{-\frac{1}{2}}.$$

Since $\mathbb{E}[v^T y] = (2\rho - 1)v^T \mathbf{1}$ and $|v^T \mathbf{1}| \leq \|v\|_2 \|\mathbf{1}\|_2 = \sqrt{n}$, with probability converging to one,

$$\begin{aligned} (v^T y)^2 &\leq \log n + (2\rho - 1)^2 v^T \mathbf{J} v + 2|v^T \mathbf{1}| |2\rho - 1| \sqrt{\log n} \\ &\leq \log n + (2\rho - 1)^2 v^T \mathbf{J} v + 2|2\rho - 1| \sqrt{n \log n}. \end{aligned}$$

Similarly, since $\mathbb{E}[\sum_i x_i y_i v_i] = 0$ and $-|v_i| \leq x_i y_i v_i \leq |v_i|$, applying the Chernoff bound yields $v^T(Z * W)v \leq \sqrt{\log n}$ with probability converging to one. \square

Lemma 6. *For $\delta = \frac{\log n}{\log \log n}$,*

$$\mathbb{P}\left(\min_{i \in [n]} d_i^* \geq \delta\right) \geq 1 - n^{1-\eta_1(\mathbf{q}, \rho)+o(1)} - n^{1-\eta_1(\mathbf{q}, 1-\rho)+o(1)}.$$

Proof. The proof is achieved by applying the Chernoff bound and taking the union bound. \square

Notice that $\rho \leq 0.5$ implies $\eta_1(\mathbf{q}, \rho) \leq \eta_1(\mathbf{q}, 1-\rho)$ and $\rho > 0.5$ implies $\eta_1(\mathbf{q}, \rho) \geq \eta_1(\mathbf{q}, 1-\rho)$. Then $\min_i d_i^* \geq \frac{\log n}{\log \log n}$ if

$$\begin{cases} \eta_1(\mathbf{q}, \rho) > 1 & \text{when } \rho \leq 0.5 \\ \eta_1(\mathbf{q}, 1-\rho) > 1 & \text{when } \rho > 0.5 \end{cases}. \quad (44)$$

Let $\lambda^* \geq \frac{1}{4}[T_1 c_1 + T_2 c_2](2\rho - 1)^2$. Therefore, applying Lemmas 4, 5, and 6, we get that if (44) holds, then

$$\begin{aligned} v^T S^* v &\geq \frac{\log n}{\log \log n} - (T_1 c' + T_2 c'') \sqrt{\log n} \\ &\quad + (T_1 + T_2) q_0 \frac{\log n}{n} > 0, \end{aligned}$$

and the first part of Theorem 7 follows.

To prove the second part, since x^* has a uniform distribution over $\{x \in \{\pm 1\}^n : x^T \mathbf{1} = 0\}$, maximum likelihood estimator minimizes the error probability among all estimators. Then we need to find when the maximum likelihood estimator fails. Let $e(i, \mathcal{H}) \triangleq \sum_{j \in \mathcal{H}} A_{ij}(T_1 y_i y_j + T_2)$. Define the events

$$\begin{aligned} F_1 &\triangleq \left\{ \min_{i \in C_1^*} (e(i, C_1^*) - e(i, C_2^*)) \leq -2 \right\}, \\ F_2 &\triangleq \left\{ \min_{i \in C_2^*} (e(i, C_2^*) - e(i, C_1^*)) \leq -2 \right\}, \end{aligned}$$

where $C_1^* = \{v \in [n] : x_v^* = 1\}$ and $C_2^* = \{v \in [n] : x_v^* = -1\}$. Then $\mathbb{P}(\text{ML fails}) \geq \mathbb{P}(F_1 \cap F_2)$. Thus, it suffices to show that with high probability $\mathbb{P}(F_1) \rightarrow 1$ and $\mathbb{P}(F_2) \rightarrow 1$. Here we just prove that $\mathbb{P}(F_1) \rightarrow 1$, while $\mathbb{P}(F_2) \rightarrow 1$ is proved similarly. By symmetry, we can condition on C_1^* being the

first $\frac{n}{2}$ nodes. Let \mathcal{T} denote the set of first $\lfloor \frac{n}{\log^2 n} \rfloor$ nodes of C_1^* . Then

$$\begin{aligned} \min_{i \in C_1^*} (e(i, C_1^*) - e(i, C_2^*)) &\leq \min_{i \in \mathcal{T}} (e(i, C_1^*) - e(i, C_2^*)) \\ &\leq \min_{i \in \mathcal{T}} (e(i, C_1^* \setminus \mathcal{T}) - e(i, C_2^*)) \\ &\quad + \max_{i \in \mathcal{T}} e(i, \mathcal{T}). \end{aligned}$$

Define the events

$$\begin{aligned} E_1 &\triangleq \left\{ \max_{i \in \mathcal{T}} e(i, \mathcal{T}) \leq \delta - 2 \right\}, \\ E_2 &\triangleq \left\{ \min_{i \in \mathcal{T}} (e(i, C_1^* \setminus \mathcal{T}) - e(i, C_2^*)) \leq -\delta \right\}. \end{aligned}$$

It suffices to show that $\mathbb{P}(E_1) \rightarrow 1$ and $\mathbb{P}(E_2) \rightarrow 1$, to have $\mathbb{P}(F_1) \rightarrow 1$. For any $i \in \mathcal{T}$,

$$e(i, \mathcal{T}) = (T_2 + T_1)X_1 + (T_2 - T_1)X_2,$$

where $X_1 \sim \text{Binom}(|\mathcal{T}|, q_0 \log n/n)$ and $X_2 \sim \text{Binom}(|\mathcal{T}|, q_2 \log n/n)$.

Lemma 7. [28, Lemma 5] *When $S \sim \text{Bin}(n, p)$, for any $r \geq 1$,*

$$\mathbb{P}(S \geq rnp) \leq \left(\frac{e}{r}\right)^{rnp} e^{-rnp}.$$

From Lemma 7,

$$\begin{aligned} \mathbb{P}\left(X_1 \geq \frac{\delta - 2}{2(T_1 + T_2)}\right) &\leq \left(\frac{(\delta - 2) \log n}{4(T_1 + T_2) e q_0}\right)^{\frac{2-\delta}{2(T_1 + T_2)}} \\ &\leq n^{-2+o(1)}, \\ \mathbb{P}\left(X_2 \geq \frac{\delta - 2}{2|T_2 - T_1|}\right) &\leq \left(\frac{(\delta - 2) \log n}{4|T_2 - T_1| e q_2}\right)^{\frac{2-\delta}{2|T_2 - T_1|}} \\ &\leq n^{-2+o(1)}. \end{aligned}$$

Since $|T_2 - T_1| > 0$ and $T_1 + T_2 > 0$,

$$\begin{aligned} \mathbb{P}(e(i, \mathcal{T}) \geq \delta - 2) \\ \leq \mathbb{P}((T_1 + T_2)X_1 + |T_2 - T_1|X_2 \geq \delta - 2) &\leq n^{-2+o(1)}. \end{aligned}$$

Using the union bound yields $\mathbb{P}(E_1) \geq 1 - n^{-1+o(1)}$. Therefore, $\mathbb{P}(E_1) \rightarrow 1$ with high probability.

Lemma 8. [34, Lemma 15] *Let $\{S_1, \dots, S_m\}$ be a sequence of i.i.d. random variables, where $m - n = o(n)$. Then for any $\mu \in \mathbb{R}$ and $\nu \geq 0$ we have*

$$\mathbb{P}\left(\sum_{i=1}^m S_i \geq \mu - \nu\right) \geq \min_{t>0} e^{-t\mu - |t|\nu} M(t) \left(1 - \frac{\sigma_Z^2}{\nu^2}\right),$$

where $M(t)$ is the moment generating function of $Z = \sum_{i=1}^m S_i$ and \hat{Z} is a random variable distributed according to $\frac{e^{t\hat{Z}} \mathbb{P}(\hat{Z})}{\mathbb{E}_Z[e^{tZ}]}$ with variance σ_Z^2 .

Lemma 9. *Let $e(i, \mathcal{H}) \triangleq \sum_{j \in \mathcal{H}} A_{ij}(T_1 y_i y_j + T_2)$. Define*

$$E'_2 \triangleq \left\{ e(i, C_1^* \setminus \mathcal{T}) - e(i, C_2^*) \leq -\delta \right\}.$$

Then

$$\mathbb{P}(E'_2) \geq n^{-\eta_1(\mathbf{q}, \rho)+o(1)} + n^{-\eta_1(\mathbf{q}, 1-\rho)+o(1)}.$$

Proof. The proof is achieved by applying Lemma 8 and the Chernoff bound. \square

Applying Lemma 9 yields

$$\begin{aligned} \mathbb{P}(E_2) &= 1 - \prod_{i \in \mathcal{T}} [1 - \mathbb{P}(E'_2)] \\ &\geq 1 - \left[1 - n^{-\eta_1(\mathbf{q}, 1-\rho)+o(1)} - n^{-\eta_1(\mathbf{q}, \rho)+o(1)} \right]^{|\mathcal{T}|} \\ &\geq 1 - e^{-n^{1-\eta_1(\mathbf{q}, 1-\rho)+o(1)} - n^{1-\eta_1(\mathbf{q}, \rho)+o(1)}}. \end{aligned}$$

Recall that $\rho \leq 0.5$ implies $\eta_1(\mathbf{q}, \rho) \leq \eta_1(\mathbf{q}, 1 - \rho)$ and $\rho > 0.5$ implies $\eta_1(\mathbf{q}, \rho) \geq \eta_1(\mathbf{q}, 1 - \rho)$. When $\rho \leq 0.5$, if $\eta_1(\mathbf{q}, \rho) < 1$ then $\mathbb{P}(E_2) \rightarrow 1$. When $\rho \geq 0.5$, if $\eta_1(\mathbf{q}, 1 - \rho) < 1$ then $\mathbb{P}(E_2) \rightarrow 1$ and the second part of Theorem 7 follows.

APPENDIX I PARTIAL RECOVERY ALGORITHM

In this paper, the partial recovery algorithm in [11] is employed with few changes to make it compatible for each Scenario. For the two-latent variable stochastic block model we can directly use the partial recovery algorithm in [11]:

A. The two-latent variable stochastic block model with known auxiliary latent variable y :

- 1) Cluster nodes according to the value of the auxiliary latent variable y , call them auxiliary clusters.
- 2) Extract submatrices of P and \bar{Q} representing each value of y , call them $P^{(k)}$ and $\bar{Q}^{(k)}$.
- 3) Separately in each auxiliary cluster, use respective submatrices $P^{(k)}$ and $\bar{Q}^{(k)}$ to construct a partial recovery estimator of communities x , and find the community estimate for all members of each cluster.

B. The two-latent variable stochastic block model with unknown latent variable y :

- 1) Use matrices P and \bar{Q} to construct a partial recovery estimator of all micro-communities.
- 2) Cluster nodes with the same community variable representing each value of x .

For the two-latent variable censored block model, we need a new variant of the partial recovery algorithm in [11]. In the new variant, the *vertex comparison algorithm* in [11] is used twice for each pair of nodes. First, the algorithm is employed using the eigenvalues of $\text{diag}(p)(\Xi * Q)$. For this case, if the two nodes belong to the same community, the output of the algorithm is 1; otherwise it returns 0. Then, the algorithm is employed using the eigenvalues of $\text{diag}(p)((1 - \Xi) * Q)$. For this case, if the two nodes belong to the same community, the output of the algorithm is 0; otherwise it returns 1. If the outputs are not equal, we are able to determine reliably whether the two nodes belong to the same community. If the outputs are equal, another pair of nodes are selected to repeat the partial recovery algorithm.

C. The two-latent variable censored block model with known latent variable y :

- 1) Cluster nodes according to the value of the auxiliary latent variable y , call them auxiliary clusters.
- 2) Extract submatrices of P , \bar{Q} , and Ξ representing each value of y , call them $P^{(k)}$ and $\bar{Q}^{(k)}$, and $\Xi^{(k)}$.
- 3) Separately in each auxiliary cluster, use respective submatrices $P^{(k)}$, $\bar{Q}^{(k)}$, and $\Xi^{(k)}$ to construct a partial recovery estimator of communities x , and find the community estimate for all members of each cluster.

D. The two-latent variable censored block model with unknown latent variable y :

- 1) Use matrices P , \bar{Q} , and Ξ to construct a partial recovery estimator of all micro-communities.
- 2) Cluster nodes with the same community variable representing each value of x .

Remark 6. When y is known, for each auxiliary latent variable y , definitions 4 and 5 in [11] are restated based on the new matrices $P^{(k)}$, $\bar{Q}^{(k)}$, and $\Xi^{(k)}$. Using these new matrices, **the vertex comparison algorithm, the vertex classification algorithm, the unreliable graph classification algorithm, and the reliable graph classification algorithm** in [11] are exploited separately. When y is unknown, these definitions and algorithms are followed from matrices P , \bar{Q} , and Ξ .

APPENDIX J PROOF OF THEOREM 8

We begin by deriving sufficient conditions for the semidefinite programming estimator (8) to produce the true labels x^* .

Lemma 10. *The sufficient conditions of Lemma 3 apply to semidefinite programming (8) by replacing*

$$S^* = D^* + \lambda^* \mathbf{J} - A.$$

Proof. The proof is similar to the proof of Lemma 3. \square

It suffices to show that $S^* = D^* + \lambda^* \mathbf{J} - A$ satisfies other conditions in Lemma 10 with probability $1 - o(1)$. Let

$$d_i^* = \sum_{j=1}^n A_{ij} x_j^* x_i^*.$$

Then $D^* x^* = A x^*$ and based on the definition of S^* in Lemma 10, S^* satisfies the condition $S^* x^* = 0$. It remains to show that $S^* \succeq 0$ and $\lambda_2(S^*) > 0$ with probability $1 - o(1)$, i.e., (43) holds. For any v such that $v^T x^* = 0$ and $\|v\| = 1$,

$$\begin{aligned} v^T S^* v &= v^T D^* v + \lambda^* v^T \mathbf{J} v - v^T (A - \mathbb{E}[A]) v - v^T \mathbb{E}[A] v \\ &\geq \min_i d_i^* + \lambda^* v^T \mathbf{J} v - \|A - \mathbb{E}[A]\| - v^T \mathbb{E}[A] v. \end{aligned}$$

Notice that

$$\begin{aligned} -v^T \mathbb{E}[A] v &= -\frac{1}{4} [c_1 v^T W v - c_2 v^T \mathbf{J} v - c_3 v^T (Z * W) v] \\ &\quad + q_0 \frac{\log n}{n}. \end{aligned}$$

Lemma 11. For $\delta = \frac{\log n}{\log \log n}$,

$$\mathbb{P}\left(\min_i d_i^* \geq \delta\right) \geq 1 - n^{1-\eta_2(\mathbf{q}, \rho)+o(1)} - n^{1-\eta_2(\mathbf{q}, 1-\rho)+o(1)}.$$

Proof. The proof is achieved by applying the Chernoff bound and the union bound. \square

Using Lemma 11, $\min_i d_i^* \geq \frac{\log n}{\log \log n}$ with probability converging to one, if $\min\{\eta_2(\mathbf{q}, \rho), \eta_2(\mathbf{q}, 1-\rho)\} > 1$. Let $\lambda^* \geq \frac{1}{4}[c_1(2\rho-1)^2 + c_2]$. Applying Lemmas 4, 5, and 11, we get that when $\min\{\eta_2(\mathbf{q}, \rho), \eta_2(\mathbf{q}, 1-\rho)\} > 1$,

$$v^T S^* v \geq \frac{\log n}{\log \log n} - c' \sqrt{\log n} + q_0 \frac{\log n}{n} > 0,$$

and the first part of Theorem 8 follows.

To prove the second part, it suffices to find when the maximum likelihood detector fails. The events F_1 , F_2 , E_1 , E_2 , and E'_2 are the same as we defined them in the proof of Theorem 7. Also, the definitions for C_1^* , C_2^* , and \mathcal{T} remain valid for this part. Then $\mathbb{P}(\text{ML fails}) \geq \mathbb{P}(F_1 \cap F_2)$. Here we just prove that $\mathbb{P}(F_1) \rightarrow 1$, while $\mathbb{P}(F_2) \rightarrow 1$ is proved similarly. By symmetry, we can condition on C_1^* being the first $\frac{n}{2}$ nodes. Then

$$\begin{aligned} \min_{i \in C_1^*} (e(i, C_1^*) - e(i, C_2^*)) &\leq \min_{i \in \mathcal{T}} (e(i, C_1^*) - e(i, C_2^*)) \\ &\leq \min_{i \in \mathcal{T}} (e(i, C_1^* \setminus \mathcal{T}) - e(i, C_2^*)) \\ &\quad + \max_{i \in \mathcal{T}} e(i, \mathcal{T}), \end{aligned}$$

where $e(i, \mathcal{H}) \triangleq \sum_{j \in \mathcal{H}} A_{ij}$. For $i \in \mathcal{T}$, $e(i, \mathcal{T}) = X_1 + X_2$, where $X_1 \sim \text{Binom}(|\mathcal{T}|, q_0 \log n/n)$ and $X_2 \sim \text{Binom}(|\mathcal{T}|, q_2 \log n/n)$. It follows from Lemma 7 that

$$\mathbb{P}\left(X_1 \geq \frac{\delta}{2} - 1\right) \leq \left(\frac{\log n}{2eq_0} \left(\frac{\delta}{2} - 1\right)\right)^{1-\frac{\delta}{2}} \leq n^{-2+o(1)},$$

$$\mathbb{P}\left(X_2 \geq \frac{\delta}{2} - 1\right) \leq \left(\frac{\log n}{2eq_2} \left(\frac{\delta}{2} - 1\right)\right)^{1-\frac{\delta}{2}} \leq n^{-2+o(1)}.$$

Then $\mathbb{P}(e(i, \mathcal{T}) \geq \delta - 2) \leq n^{-2+o(1)}$. Using the union bound, $\mathbb{P}(E_1) \geq 1 - n^{-1+o(1)}$. Therefore, $\mathbb{P}(E_1) \rightarrow 1$ with high probability.

Lemma 12. When $e(i, \mathcal{H}) \triangleq \sum_{j \in \mathcal{H}} A_{ij}$,

$$\mathbb{P}(E'_2) \geq n^{-\eta_2(\mathbf{q}, \rho)+o(1)} + n^{-\eta_2(\mathbf{q}, 1-\rho)+o(1)}.$$

Proof. The proof is achieved by applying Lemma 8 and the Chernoff bound. \square

Applying Lemma 12 yields

$$\begin{aligned} \mathbb{P}(E_2) &= 1 - \prod_{i \in \mathcal{T}} [1 - \mathbb{P}(E'_2)] \\ &\geq 1 - \left[1 - n^{-\eta_2(\mathbf{q}, \rho)+o(1)} - n^{-\eta_2(\mathbf{q}, 1-\rho)+o(1)}\right]^{|\mathcal{T}|} \\ &\geq 1 - e^{-n^{1-\eta_2(\mathbf{q}, \rho)+o(1)} - n^{1-\eta_2(\mathbf{q}, 1-\rho)+o(1)}}. \end{aligned}$$

Therefore, if $\min\{\eta_2(\mathbf{q}, \rho), \eta_2(\mathbf{q}, 1-\rho)\} < 1$ then $\mathbb{P}(E_2) \rightarrow 1$ and the second part of Theorem 8 follows.

APPENDIX K PROOF OF THEOREM 9

The proof is similar to the proof of Theorem 7. Here we just mention the proof outlines and important Lemmas for brevity. The following Lemma declares the sufficient conditions for the optimum solution of (10) matching the true labels x^* .

Lemma 13. For the optimization problem (10), consider the Lagrange multipliers

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

If we have

$$\begin{aligned} S^* &= D^* + \lambda^* \mathbf{J} - R, \\ S^* &\succeq 0, \\ \lambda_2(S^*) &> 0, \\ S^* x^* &= 0, \end{aligned}$$

then (λ^*, D^*, S^*) is the dual optimal solution and $\hat{Z} = x^* x^{*T}$ is the unique primal optimal solution of (10).

Proof. The proof is similar to the proof of Lemma 3. \square

Let

$$\begin{aligned} d_i^* &= T \sum_{j=1}^n A_{ij} x_j^* x_i^* + T \sum_{j=1}^n A_{ij} y_i y_j x_j^* x_i^* \\ &\quad + T_1 \sum_{j=1}^n A_{ij}^2 y_i y_j x_j^* x_i^* + T_2 \sum_{j=1}^n A_{ij}^2 x_j^* x_i^*. \end{aligned}$$

Then $D^* x^* = T A + T(A * W) + T_1(A * A * W) + T_2(A * A)$ and based on the definition of S^* in Lemma 13, S^* satisfies the condition $S^* x^* = 0$.

Lemma 14. For $\delta = \frac{\log n}{\log \log n}$,

$$\begin{aligned} \mathbb{P}\left(\min_{i \in [n]} d_i^* \geq \delta\right) &\geq 1 - n^{1-\eta_1(\mathbf{g}, \rho) - \eta_1(\mathbf{h}, \rho) + o(1)} \\ &\quad - n^{1-\eta_1(\mathbf{g}, 1-\rho) - \eta_1(\mathbf{h}, 1-\rho) + o(1)}. \end{aligned}$$

Proof. The proof is achieved by applying the Chernoff bound and taking the union bound. \square

Similar to the proof of Theorem 7, using Lemma 14, it can be shown that $S^* \succeq 0$ and $\lambda_2(S^*) > 0$ with probability $1 - o(1)$ if

$$\begin{cases} \eta_1(\mathbf{g}, \rho) + \eta_1(\mathbf{h}, \rho) > 1 & \text{when } \rho \leq 0.5 \\ \eta_1(\mathbf{g}, 1-\rho) + \eta_1(\mathbf{h}, 1-\rho) > 1 & \text{when } \rho > 0.5 \end{cases}.$$

To prove the second part, we start to find when the maximum likelihood estimator fails. To this end, let

$$e(i, \mathcal{H}) \triangleq \sum_{j \in \mathcal{H}} A_{ij} (T y_i y_j + T) + A_{ij}^2 (T_1 y_i y_j + T_2).$$

The definition of events F_1 , F_2 , E_1 , and E_2 in the proof of Theorem 7 are used to show that with high probability $\mathbb{P}(F_1) \rightarrow 1$ and $\mathbb{P}(F_2) \rightarrow 1$. Also, the definitions for C_1^* , C_2^* , and \mathcal{T} remain valid for this part. We prove that $\mathbb{P}(F_1) \rightarrow 1$, while $\mathbb{P}(F_2) \rightarrow 1$ is proved similarly. To show that $\mathbb{P}(F_1) \rightarrow 1$,

we must have $\mathbb{P}(E_1) \rightarrow 1$ and $\mathbb{P}(E_2) \rightarrow 1$. It can be shown that $\mathbb{P}(E_1) \geq 1 - n^{-1+o(1)}$ without difficulty.

Lemma 15. *Let $E'_2 \triangleq \{e(i, C_1^* \setminus \mathcal{T}) - e(i, C_2^*) \leq -\delta\}$. Then*

$$\mathbb{P}(E'_2) \geq n^{-\eta_1(\mathbf{g}, \rho) - \eta_1(\mathbf{h}, \rho) + o(1)} + n^{-\eta_1(\mathbf{g}, 1-\rho) - \eta_1(\mathbf{h}, 1-\rho) + o(1)}.$$

Proof. The proof is achieved by applying Lemma 8 and the Chernoff bound. \square

Applying Lemma 15 yields

$$\begin{aligned} \mathbb{P}(E_2) &= 1 - \prod_{i \in \mathcal{T}} [1 - \mathbb{P}(E'_2)] \\ &\geq 1 - e^{-n^{-\eta_1(\mathbf{g}, \rho) - \eta_1(\mathbf{h}, \rho) + o(1)} - n^{-\eta_1(\mathbf{g}, 1-\rho) - \eta_1(\mathbf{h}, 1-\rho) + o(1)}}. \end{aligned}$$

Recall that $\rho \leq 0.5$ implies

$$\eta_1(\mathbf{g}, \rho) + \eta_1(\mathbf{h}, \rho) \leq \eta_1(\mathbf{g}, 1-\rho) - \eta_1(\mathbf{h}, 1-\rho),$$

and $\rho > 0.5$ implies

$$\eta_1(\mathbf{g}, \rho) + \eta_1(\mathbf{h}, \rho) \geq \eta_1(\mathbf{g}, 1-\rho) + \eta_1(\mathbf{h}, 1-\rho).$$

When $\rho \leq 0.5$, if $\eta_1(\mathbf{g}, \rho) + \eta_1(\mathbf{h}, \rho) < 1$, then $\mathbb{P}(E_2) \rightarrow 1$. When $\rho \geq 0.5$, if $\eta_1(\mathbf{g}, 1-\rho) + \eta_1(\mathbf{h}, 1-\rho) < 1$, then $\mathbb{P}(E_2) \rightarrow 1$ and the second part of Theorem 9 follows.

APPENDIX L PROOF OF THEOREM 10

The proof is similar to the proof of Theorem 8. Here we just mention the proof outlines and important Lemmas for brevity. The following Lemma declares the sufficient conditions for the optimum solution of (11) matching the true labels x^* .

Lemma 16. *For the optimization problem (11), consider the Lagrange multipliers*

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

If we have

$$\begin{aligned} S^* &= D^* + \lambda^* \mathbf{J} - TA - T_2(A * A), \\ S^* &\succeq 0, \\ \lambda_2(S^*) &> 0, \\ S^* x^* &= 0, \end{aligned}$$

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

Proof. The proof is similar to the proof of Lemma 3. \square

Let

$$d_i^* = T \sum_{j=1}^n A_{ij} x_j^* x_i^* + T_2 \sum_{j=1}^n A_{ij}^2 x_j^* x_i^*.$$

Then $D^* x^* = TA + T_2(A * A)$ and based on the definition of S^* in Lemma 16, S^* satisfies the condition $S^* x^* = 0$.

Lemma 17. *For $\delta = \frac{\log n}{\log \log n}$,*

$$\begin{aligned} \mathbb{P}\left(\min_{i \in [n]} d_i^* \geq \delta\right) &\geq 1 - n^{-1 - \eta_2(\mathbf{g}, \rho) - \eta_2(\mathbf{h}, \rho) + o(1)} \\ &\quad - n^{-1 - \eta_2(\mathbf{g}, 1-\rho) - \eta_2(\mathbf{h}, 1-\rho) + o(1)}. \end{aligned}$$

Proof. The proof is achieved by applying the Chernoff bound and taking the union bound. \square

Similar to the proof of Theorem 8, using Lemma 17, it can be shown that $S^* \succeq 0$ and $\lambda_2(S^*) > 0$ with probability $1 - o(1)$ if

$$\min \{\eta_2(\mathbf{g}, \rho) + \eta_2(\mathbf{h}, \rho), \eta_2(\mathbf{g}, 1-\rho) + \eta_2(\mathbf{h}, 1-\rho)\} > 1.$$

To prove the second part, we start to find when the maximum likelihood estimator fails. To this end, let

$$e(i, \mathcal{H}) \triangleq \sum_{j \in \mathcal{H}} TA_{ij} + T_2 A_{ij}^2.$$

The definition of events F_1 , F_2 , E_1 , and E_2 in Theorem 7 are used to show that with high probability $\mathbb{P}(F_1) \rightarrow 1$ and $\mathbb{P}(F_2) \rightarrow 1$. Also, the definitions for C_1^* , C_2^* , and \mathcal{T} remain valid for this part. We prove that $\mathbb{P}(F_1) \rightarrow 1$, while $\mathbb{P}(F_2) \rightarrow 1$ is proved similarly. To show that $\mathbb{P}(F_1) \rightarrow 1$, we must have $\mathbb{P}(E_1) \rightarrow 1$ and $\mathbb{P}(E_2) \rightarrow 1$. It can be shown that $\mathbb{P}(E_1) \geq 1 - n^{-1+o(1)}$ without difficulty.

Lemma 18. *Let $E'_2 \triangleq \{e(i, C_1^* \setminus \mathcal{T}) - e(i, C_2^*) \leq -\delta\}$. Then*

$$\begin{aligned} \mathbb{P}(E'_2) &\geq n^{-\eta_2(\mathbf{g}, \rho) - \eta_2(\mathbf{h}, \rho) + o(1)} \\ &\quad + n^{-\eta_2(\mathbf{g}, 1-\rho) - \eta_2(\mathbf{h}, 1-\rho) + o(1)}. \end{aligned}$$

Proof. The proof is achieved by applying Lemma 8 and the Chernoff bound. \square

Applying Lemma 18 yields

$$\begin{aligned} \mathbb{P}(E_2) &= 1 - \prod_{i \in \mathcal{T}} [1 - \mathbb{P}(E'_2)] \\ &\geq 1 - e^{-n^{-\eta_2(\mathbf{g}, \rho) - \eta_2(\mathbf{h}, \rho) + o(1)} - n^{-\eta_2(\mathbf{g}, 1-\rho) - \eta_2(\mathbf{h}, 1-\rho) + o(1)}}. \end{aligned}$$

If

$$\min \{\eta_2(\mathbf{g}, \rho) + \eta_2(\mathbf{h}, \rho), \eta_2(\mathbf{g}, 1-\rho) + \eta_2(\mathbf{h}, 1-\rho)\} < 1,$$

then $\mathbb{P}(E_2) \rightarrow 1$ and the second part of Theorem 10 follows.

REFERENCES

- [1] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [2] J. Chen and B. Yuan, "Detecting functional modules in the yeast protein-protein interaction network," *Bioinformatics*, vol. 22, no. 18, pp. 2283–2290, 2006.
- [3] K. Berahmand, E. Nasiri, S. Forouzandeh, and Y. Li, "A preference random walk algorithm for link prediction through mutual influence nodes in complex networks," *Journal of King Saud University-Computer and Information Sciences*, 2021.
- [4] K. Berahmand, E. Nasiri, M. Rostami, and S. Forouzandeh, "A modified deepwalk method for link prediction in attributed social network," *Computing*, pp. 1–23, 2021.
- [5] 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.
- [6] L. Li, D. Xu, H. Peng, J. Kurths, and Y. Yang, "Reconstruction of complex network based on the noise via qr decomposition and compressed sensing," *Scientific reports*, vol. 7, no. 1, pp. 1–13, 2017.
- [7] K. Huang, W. Deng, Y. Zhang, and H. Zhu, "Sparse bayesian learning for network structure reconstruction based on evolutionary game data," *Physica A: Statistical Mechanics and its Applications*, vol. 541, p. 123605, 2020.

- [8] T. P. Peixoto, "Network reconstruction and community detection from dynamics," *Physical review letters*, vol. 123, no. 12, p. 128301, 2019.
- [9] W.-X. Wang, Y.-C. Lai, C. Grebogi, and J. Ye, "Network reconstruction based on evolutionary-game data via compressive sensing," *Physical Review X*, vol. 1, no. 2, p. 021021, 2011.
- [10] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.
- [11] E. Abbe and C. Sandon, "Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery," in *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE, 2015, pp. 670–688.
- [12] B. Hajek, Y. Wu, and J. Xu, "Exact recovery threshold in the binary censored block model," in *2015 IEEE Information Theory Workshop-Fall (ITW)*. IEEE, 2015, pp. 99–103.
- [13] M. Esmaeili and A. Nosratinia, "Community detection with secondary latent variables," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 1355–1360.
- [14] A. Saade, M. Lelarge, F. Krzakala, and L. Zdeborová, "Spectral detection in the censored block model," in *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 1184–1188.
- [15] P. Fronczak, A. Fronczak, and M. Bujok, "Exponential random graph models for networks with community structure," *Physical Review E*, vol. 88, no. 3, p. 032810, 2013.
- [16] M. Esmaeili and A. Nosratinia, "Community detection: Exact recovery in weighted graphs," *arXiv preprint arXiv:2102.04439*, 2021.
- [17] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Inference and phase transitions in the detection of modules in sparse networks," *Physical Review Letters*, vol. 107, no. 6, p. 065701, 2011.
- [18] E. Mossel, J. Neeman, and A. Sly, "Reconstruction and estimation in the planted partition model," *Probability Theory and Related Fields*, vol. 162, no. 3–4, pp. 431–461, 2015.
- [19] L. Massoulié, "Community detection thresholds and the weak ramanujan property," in *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, 2014, pp. 694–703.
- [20] E. Mossel, J. Neeman, and A. Sly, "A proof of the block model threshold conjecture," *Combinatorica*, vol. 38, no. 3, pp. 665–708, 2018.
- [21] M. Esmaeili, H. Saad, and A. Nosratinia, "Exact recovery by semidefinite programming in the binary stochastic block model with partially revealed side information," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 3477–3481.
- [22] E. Mossel and J. Xu, "Density evolution in the degree-correlated stochastic block model," in *Conference on Learning Theory*, 2016, pp. 1319–1356.
- [23] S.-Y. Yun and A. Proutiere, "Community detection via random and adaptive sampling," in *Conference on learning theory*, 2014, pp. 138–175.
- [24] E. Abbe, "Community detection and stochastic block models: recent developments," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6446–6531, 2017.
- [25] E. Abbe, A. S. Bandeira, and G. Hall, "Exact recovery in the stochastic block model," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 471–487, 2015.
- [26] E. Mossel, J. Neeman, and A. Sly, "Consistency thresholds for the planted bisection model," in *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, 2015, pp. 69–75.
- [27] Y. Chen and J. Xu, "Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 882–938, 2016.
- [28] M. Esmaeili, H. Saad, and A. Nosratinia, "Community detection with side information via semidefinite programming," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 420–424.
- [29] E. Mossel, J. Neeman, and A. Sly, "Belief propagation, robust reconstruction and optimal recovery of block models," in *Conference on Learning Theory*, 2014, pp. 356–370.
- [30] J. Z. Moghaddam, M. Esmaeili, and A. Nosratinia, "Exact recovery threshold in dynamic binary censored block model," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 1088–1093.
- [31] A. A. Amini, E. Levina *et al.*, "On semidefinite relaxations for the block model," *The Annals of Statistics*, vol. 46, no. 1, pp. 149–179, 2018.
- [32] B. Hajek, Y. Wu, and J. Xu, "Achieving exact cluster recovery threshold via semidefinite programming," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2788–2797, 2016.
- [33] M. Esmaeili, H. M. Saad, and A. Nosratinia, "Semidefinite programming for community detection with side information," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1957–1973, 2021.
- [34] H. Saad and A. Nosratinia, "Community detection with side information: Exact recovery under the stochastic block model," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 5, pp. 944–958, 2018.
- [35] P. D. Hoff, A. E. Raftery, and M. S. Handcock, "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, vol. 97, no. 460, pp. 1090–1098, 2002.
- [36] C. Ke and J. Honorio, "Information-theoretic limits for community detection in network models," in *Advances in Neural Information Processing Systems*, 2018, pp. 8324–8333.
- [37] P. Sarkar and A. W. Moore, "Dynamic social network analysis using latent space models," in *Advances in Neural Information Processing Systems*, 2006, pp. 1145–1152.
- [38] P. Latouche, E. Birmelé, C. Ambroise *et al.*, "Overlapping stochastic block models with application to the french political blogosphere," *The Annals of Applied Statistics*, vol. 5, no. 1, pp. 309–336, 2011.
- [39] B. Hajek, Y. Wu, and J. Xu, "Achieving exact cluster recovery threshold via semidefinite programming: Extensions," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5918–5937, 2016.