Recovering a Single Community with Side Information

Hussein Saad, and Aria Nosratinia, *Fellow, IEEE*

***Abstract*—We study the effect of the quality and quantity of side information on the recovery of a hidden community of size $K = o(n)$ in a graph of size n. Side information for each node in the graph is modeled by a random vector in which either the vector dimension or the LLR of each component with respect to node labels is independent of n. These two models represent the variation in quality and quantity of side information. Under maximum likelihood detection, we calculate tight necessary and sufficient conditions for exact recovery of the labels. We demonstrate how side information needs to evolve with n in terms of either its quantity, or quality, to improve the exact recovery threshold. A similar set of results are obtained for weak recovery. Under belief propagation, tight necessary and sufficient conditions for weak recovery are calculated when the LLRs are constant, and sufficient conditions when the LLRs vary with n. Moreover, we design and analyze a local voting procedure using side information that can achieve exact recovery when applied after belief propagation.**

***Index Terms*—Community detection, stochastic block model, side information**

A. Introduction

Detecting communities (or clusters) in graphs is a fundamental problem that has been studied in various fields, statistics [3]–[7], computer science [8]–[12] and theoretical statistical physics [13], [14]. It has many applications: finding like-minded people in social networks [15], improving recommendation systems [16], detecting protein complexes [17]. In this paper, we consider the problem of finding a single sub-graph (community) hidden in a large graph, where the community size is much smaller than the graph size. Applications of finding a hidden community include fraud activity detection [18], [19] and correlation mining [20].

Several models have been studied for random graphs that exhibit a community structure [21]. A widely used model in the context of community detection is the stochastic block model (SBM) [22]. In this paper, the stochastic block model for one community is considered [23]–[26]. The stochastic block model for one community consists of a graph of size n with a community of size K, where any two nodes are connected with probability p if they are both within the community, and with probability q otherwise.

The problem of finding a hidden community upon observing *only* the graph has been studied in [23]–[25]. The information limits[1] of weak recovery and exact recovery have been studied in [24]. Weak recovery is achieved when the expected number of misclassified nodes is $o(K)$, and exact recovery when all labels are recovered with probability approaching one. The limits of belief propagation for weak recovery have been characterized [23], [25] in terms of a signal-to-noise ratio parameter $\lambda = \frac{K^2(p-q)^2}{(n-k)q}$. The utility of a voting procedure after belief propagation to achieve exact recovery was pointed out in [25].

In many practical applications, non-graphical relevant information is available that can aid the inference. For example, social networks such as Facebook and Twitter have access to other information other than the graph edges such as date of birth, nationality, school. A citation network has the authors names, keywords, and therefore may provide significant additional information beyond the co-authoring relationships. This paper investigates *when and by how much can side information improve the information limit, as well as the phase transition of belief propagation, in single-community detection.*

We model a varying quantity and quality of side information by associating with each node a vector (i.e., non-graphical) observation whose dimension represents the quantity of side information and whose (element-wise) log-likelihood ratios (LLRs) with respect to node labels represents the quality of side information. The contributions of this paper can be summarized as follows:

- We calculate tight necessary and sufficient conditions for both weak and exact recovery. We show that weak recovery is achievable even when the size of the community is random and unknown. We find conditions under which achievability of weak recovery implies exact recovery. Subject to some mild conditions on the exponential moments of LLR, the results apply to both discrete as well as continuous-valued side information. When feature LLRs vary with n, our results apply to finite side information alphabet, and necessary and sufficient conditions for weak recovery are not tight.
- The phase transition of belief propagation is characterized with side information that has fixed dimension per node. When the LLRs are fixed across n, tight necessary and sufficient conditions are calculated for weak recovery. We show that when belief propagation fails, no local algorithm can achieve weak recovery, and also that belief propagation is *strictly* inferior to the maximum likelihood detector. We also calculate conditions under which

The authors are with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75083-0688 USA, E-mail: hussein.saad@utdallas.edu; aria@utdallas.edu.

This work was presented in part in ISIT 2018, Vail, CO. [1], [2].

This work was supported in part by grants 1711689 and 2008684 from the National Science Foundation.

[1]The extremal phase transition threshold is also known as *information theoretic limit* [22] or *information limit* [24]. We use the latter term throughout this paper.

belief propagation followed by a local voting procedure achieves exact recovery. When the side information LLR varies with n, the belief propagation error rate is calculated using density evolution. Our results generalize [26], which only considered binary side information consisting of noisy labels with vanishing noise.

We now present a brief review of the literature in the area of side information for community detection and highlight the distinctions of the present work. In the context of detecting two or more communities: Mossel and Xu [27] showed that, under certain condition, belief propagation with noisy label information has the same residual error as the maximum a-posteriori estimator for two symmetric communities. Cai *et. al* [28] studied weak recovery of two symmetric communities under belief propagation upon observing a vanishing fraction of labels. Neither [27] nor [28] establishes a converse. For two symmetric communities, Saad and Nosratinia [29], [30] studied exact recovery under side information. Asadi [31] studied the effect of i.i.d. vectors of side information on the phase transition of exact recovery for more than two communities. Kanade *et. al* [32] showed that observation of a vanishing number of labels is unhelpful to *correlated recovery*[2] phase transition. Deshpande*et. al* [33] studied the effect of Gaussian distributed side information on the information limits of correlated recovery. For single community detection, Kadavankandy *et al.* [26] studied belief propagation with noisy label information with vanishing noise (unbounded LLRs). Finally, other works focused on the information-computation gap [34]–[36].

The issue of side information in the context of single-community detection has not been addressed in the literature except for [26], whose results are generalized in this paper. Analyzing the effect of side information on information limit of weak recovery is a novel contribution of this work. A converse for the local algorithms such as belief propagation with side information has not been available prior to this work. The study of side information whose LLRs vary with n is largely novel. Whenever the side information LLR does not grow with n, we are able to find a sufficient statistic that bears a resemblance to the single-community detection problem without side information, therefore tools and techniques can be shared from, e.g., [24], [25]. When the LLR of each side information features varies with n, then the statistic used for detection will be heavily dependent on n, and in these cases the proof techniques are distinct from earlier work.

I. System Model and Definitions

Let $\boldsymbol{G}$ be a realization from a random ensemble of graphs $\mathcal{G}(n, K, p, q)$, where each graph has n nodes and contains a hidden community C^* with size $|C^*| = K$. The underlying distribution of the graph is as follows: an edge connects a pair of nodes with probability p if both nodes are in C^* and with probability q otherwise. G_{ij} is the indicator of an edge between nodes i, j. For each node i, a vector of dimension M is observed consisting of side information, whose distribution depends on the label x_i of the node. By convention $x_i = 1$ if $i \in C^*$ and $x_i = 0$ if $i \notin C^*$. For node i, the entries of the side information vector are each denoted $y_{i,m}$ and can be interpreted as different features of the side information. The side information for the entire graph is collected into the matrix $\boldsymbol{Y}_{n\times M}$. The column vector $\boldsymbol{y}_m = [y_{1,m}, \dots, y_{n,m}]^t$ collects the side information feature m for all nodes i.

[2]Correlated recovery denotes probability of error that is strictly better than a random guess, and is not a subject of this paper.

The vector of true labels is denoted $\boldsymbol{x}^* \in \{0, 1\}^n$. P and Q are Bernoulli distributions with parameters p, q, respectively, and

$$L_G(i,j) = \log\Big(\frac{P(G_{ij})}{Q(G_{ij})}\Big)$$

is the log-likelihood ratio of edge G_{ij} with respect to P and Q.

In this paper, we address the problem of *single-community detection*, i.e., recovering $\boldsymbol{x}^*$ from $\boldsymbol{G}$ and $\boldsymbol{Y}$, under the following conditions: $K = o(n)$ while $\lim_{n\to\infty} K = \infty$, $p \geq q$, $\frac{p}{q} = \theta(1)$ and $\limsup_{n\to\infty} p < 1$.

An estimator $\hat{\boldsymbol{x}}(\boldsymbol{G}, \boldsymbol{Y})$ is said to achieve exact recovery of $\boldsymbol{x}^*$ if, as $n \to \infty$, $\mathbb{P}(\hat{\boldsymbol{x}} = \boldsymbol{x}^*) \to 1$. An estimator $\hat{\boldsymbol{x}}(\boldsymbol{G}, \boldsymbol{Y})$ is said to achieve weak recovery if, as $n \to \infty$, $\frac{d(\hat{\boldsymbol{x}}, \boldsymbol{x}^*)}{K} \to 0$ in probability, where $d(\cdot, \cdot)$ denotes the Hamming distance. It was shown in [24] that the latter definition is equivalent to the existence of an estimator $\hat{\boldsymbol{x}}$ such that $\mathbb{E}[d(\hat{\boldsymbol{x}}, \boldsymbol{x}^*)] = o(K)$. This equivalence will be used throughout our paper.

II. Information Limits

A. Weak Recovery

In this subsection, the information limits of weak recovery under two models of side information are presented. When the dimension of side information, M, for each node varies but its LLR is fixed across n, tight necessary and sufficient conditions are presented. When the side information for each node has fixed dimension but varying LLR, we derive necessary conditions for weak recovery.

1) Fixed-Quality Features:

In this subsection, the side information for each node is allowed to evolve with n by having a varying number of independent and identically distributed scalar observations, each of which has a finite (imperfect) amount of information about the node label. By allowing the dimension of the side information per-node to vary and its scalar components to be identically distributed, the side information is represented with fixed-quality quanta. The results of this section demonstrate that as n grows, the number of these side information quanta per-node must increase in a prescribed fashion in order to have a positive effect on the threshold for recovery.

For all n, for all $i = 1, \dots, n$, define the distributions:

$$V(v) \triangleq \mathbb{P}(y_{i,m} = v | x_i = 1) \qquad U(v) \triangleq \mathbb{P}(y_{i,m} = v | x_i = 0)$$

Thus, the components of the side information for each node (features) are identically distributed for all nodes and all graph sizes n; we also assume all features are independent conditioned on the node labels $\boldsymbol{x}^*$. The dimension M of the side information per node is allowed to vary as the size of the graph n changes.

In addition, we assume U, V are such that the resulting LLR random variable, defined below, has bounded support:

$$L_S(i,m) = \log\Big(\frac{V(y_{i,m})}{U(y_{i,m})}\Big)$$

Throughout the paper, L_S will continue to denote the LLR random variable of one side information feature, and L_G denotes the random variable of the LLR of a graph edge.

Theorem 1. *For single community detection under bounded-LLR side information, weak recovery is achieved if and only if:*

$$(K-1)D(P||Q) + MD(V||U) \to \infty\ , \quad \liminf_{n\to\infty} \frac{(K-1)D(P||Q) + 2MD(V||U)}{\log(\frac{n}{K})} > 2 \tag{1}$$

Proof. For necessity please see Appendix B. For sufficiency, please see Appendix C. □

Remark 1. *If the features are conditionally independent but* not *identically distributed, it is easy to show the necessary and sufficient conditions are:*

$$(K-1)D(P||Q) + \sum_{m=1}^{M} D(V_m||U_m) \to \infty\ ,$$
$$\liminf_{n\to\infty} \frac{(K-1)D(P||Q) + 2\sum_{m=1}^{M} D(V_m||U_m)}{\log(\frac{n}{K})} > 2$$

where V_m and U_m are analogous to U and V earlier, except specialized to each feature.

Remark 2. *When the number of features, M, is assumed to be constant but the LLR of each feature is allowed to vary with n, the necessary conditions provided earlier hold. In other words, for single community detection under varying-LLR side information, weak recovery is achieved only if:*

$$(K-1)D(P||Q) + \sum_{m=1}^{M} (D(V_m||U_m) + D(U_m||V_m)) \to \infty \quad \liminf_{n\to\infty} \frac{(K-1)D(P||Q) + 2\sum_{m=1}^{M} D(V_m||U_m)}{\log(\frac{n}{K})} > 2 \tag{2}$$

Remark 3. *Theorem 1 shows how fast the number of side information features must grow with n so that the information limit of weak recovery is improved.*

Remark 4. *The condition of bounded support for the LLRs can be somewhat weakened to Eqs.* (65) *and* (68)*. As an example $U \sim \mathcal{N}(0,1)$ and $V \sim \mathcal{N}(\mu,1)$ with $\mu \neq 0$ satisfies* (65), (68) *and the theorem continues to hold even though the LLR is not bounded.*

The assumption that the size of the community $|C^*|$ is known a-priori is not always reasonable: we might need to detect a small community whose size is not known in advance. In that case, the performance is characterized by the following lemma.

Lemma 1. *For single-community detection under bounded-LLR side information, if the size of the community is not known in advance but obeys a probability distribution satisfying:*

$$\mathbb{P}\Big(\Big|\ |C^*| - K\Big| \leq \frac{K}{\log(K)}\Big) \geq 1 - o(1) \tag{3}$$

for some known $K = o(n)$. If conditions (1) *hold, then:*

$$\mathbb{P}\Big(\frac{|\hat{C}\triangle C^*|}{K} \leq 2\epsilon + \frac{1}{\log(K)}\Big) \geq 1 - o(1) \tag{4}$$

where

$$\epsilon = \Big(\min(\log(K), (K-1)D(P||Q) + MD(V||U))\Big)^{-\frac{1}{2}} = o(1).$$

Proof. Please see Appendix D □

Lemma 1 will be used in the sequel for characterizing the sufficient conditions for exact recovery.

B. Exact Recovery

In this subsection, the information limits of exact recovery under two models of side information are presented. However, unlike Section II-A, tight necessary and sufficient conditions are provided for both models of side information.

1) Fixed-Quality Features:

Recall the definitions of the random variables L_S and L_G, as well as the distributions V, U, P, and Q.

Definition 1.

$$\psi_{QU}(t, m_1, m_2) \triangleq m_1 \log(\mathbb{E}_Q[e^{tL_G}]) + m_2 \log(\mathbb{E}_U[e^{tL_S}]) \tag{5}$$

$$\psi_{PV}(t, m_1, m_2) \triangleq m_1 \log(\mathbb{E}_P[e^{tL_G}]) + m_2 \log(\mathbb{E}_V[e^{tL_S}]) \tag{6}$$

$$E_{QU}(\theta, m_1, m_2) \triangleq \sup_{t\in[0,1]} t\theta - \psi_{QU}(t, m_1, m_2) \tag{7}$$

$$E_{PV}(\theta, m_1, m_2) \triangleq \sup_{t\in[-1,0]} t\theta - \psi_{PV}(t, m_1, m_2) \tag{8}$$

where θ, m_1 and $m_2 \in \mathbb{R}$.

The sufficient conditions for exact recovery are derived using a two-step algorithm (see Table I). Its first step consists of any algorithm achieving weak recovery, e.g. maximum likelihood (see Lemma 1). The second step applies a local voting procedure.

Lemma 2. *Let S_k, $\tilde{C}$, and $\hat{C}_k$ be defined as in Table I. Define $C_k^* = C^* \cap {S_k}^c$, and assume $\hat{C}_k$ achieves weak recovery, i.e.*

$$\mathbb{P}\big(|\hat{C}_k \triangle C_k^*| \leq \delta K \text{ for } 1 \leq k \leq \frac{1}{\delta}\big) \to 1\ . \tag{9}$$

If

$$\liminf_{n\to\infty} \frac{E_{QU}\big(\log(\frac{n}{K}), K, M\big)}{\log(n)} > 1 \tag{10}$$

then $\mathbb{P}(\tilde{C} = C^*) \to 1$.

Proof. Please see Appendix E. □

Then the main result of this section follows:

TABLE I
ALGORITHM FOR EXACT RECOVERY.

Algorithm 1
1) Input: n, K, $\boldsymbol{G}$, $\boldsymbol{Y}$, $\delta \in (0,1) : n\delta, \frac{1}{\delta} \in \mathbb{N}$. 2) Consider a partition of the nodes $\{S_k\}$ with $\|S_k\| = n\delta$. $\boldsymbol{G}_k$ and $\boldsymbol{Y}_k$ are the subgraph and side information corresponding to ${S_k}^c$, i.e., after each member of partition has been withheld. 3) Consider estimator $\hat{C}_k(\boldsymbol{G}_k, \boldsymbol{Y}_k)$ that produces $\|\hat{C}_k\| = \lceil K(1-\delta) \rceil$ and further assume it achieves weak recovery. 4) For all S_k and all $i \in S_k$ calculate $r_i = (\sum_{j \in \hat{C}_k} L_G(ij)) + \sum_{m=1}^{M} L_S(i,m)$ 5) Output: $\tilde{C} = \{$Nodes corresponding to K largest $r_i\}$.

Theorem 2. *In single community detection under bounded-LLR side information, assume* (1) *holds, then exact recovery is achieved if and only if:*

$$\liminf_{n\to\infty} \frac{E_{QU}\big(\log(\frac{n}{K}), K, M\big)}{\log(n)} > 1 \tag{11}$$

Proof. For sufficiency, please see Appendix F. For necessity see Appendix G. □

Remark 5. *The assumption that* (1) *holds is necessary because otherwise weak recovery is not achievable, and by extension, exact recovery.*

Remark 6. *Theorem 2 shows how fast the number of side information features must grow with n so that the information limit of exact recovery is improved.*

To illustrate the effect of side information on information limits, consider the following example:

$$K = \frac{cn}{\log(n)}, \quad q = \frac{b\log^2(n)}{n}, \quad p = \frac{a\log^2(n)}{n} \tag{12}$$

for positive constants $c, a \geq b$. Then, $KD(P||Q) = O(\log(n))$, and hence, weak recovery is achieved without side information, and by extension, with side information. Moreover, *exact* recovery without side information is achieved if and only if:

$$\sup_{t\in[0,1]} \; tc(a-b) + bc - bc(\frac{a}{b})^t > 1 \tag{13}$$

Assume noisy label side information with error probability $\alpha \in (0, 0.5)$. By Theorem 2, exact recovery is achieved if and only if:

$$\sup_{t\in[0,1]} \; tc(a-b) + bc - bc(\frac{a}{b})^t - \frac{M}{\log(n)} \log((1-\alpha)^t\alpha^{(1-t)} + (1-\alpha)^{(1-t)}\alpha^t) > 1 \tag{14}$$

If $M = o(\log(n))$, then (14) reduces to (13), thus side information does not improve the information limits of exact recovery. If $M > o(\log(n))$, then $\log((1-\alpha)^t\alpha^{(1-t)} + (1-\alpha)^{(1-t)}\alpha^t) < 0$ since $t \in [0,1]$. It follows that (14) is less restrictive than (13), thus improving the information limit.

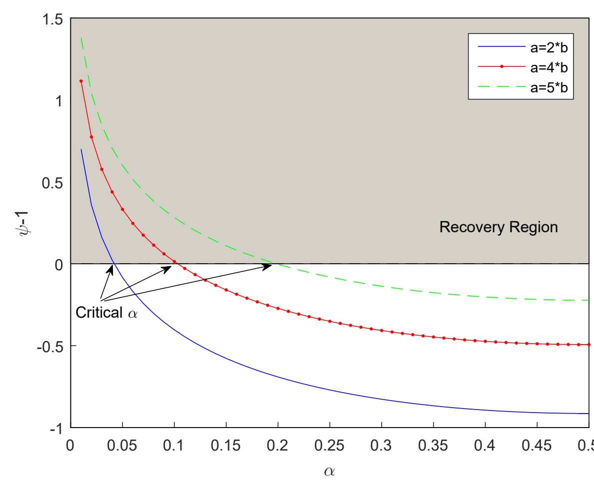

Fig. 1. Exact recovery threshold, $\psi - 1$ for different values of α at $c = b = 1$.

Let ψ denote the left hand side of (14) with $M = \log(n)$, i.e.,

$$\psi = \sup_{t\in[0,1]} \; tc(a-b) + bc - bc(\frac{a}{b})^t - \log((1-\alpha)^t\alpha^{(1-t)} + (1-\alpha)^{(1-t)}\alpha^t) \tag{15}$$

The behavior of ψ against α describes the influence of side information on exact recovery and is depicted in Fig. 1.

2) Variable-Quality Features:

In this section, the number of features, M, is assumed to be constant but the LLR of each feature is allowed to vary with n. We begin by concentrating on the following regime, and will subsequently show its relation to the set of problems that are both feasible and interesting.

$$K = \rho\frac{n}{\log(n)}, \qquad p = a\frac{\log(n)^2}{n} \qquad q = b\frac{\log(n)^2}{n} \tag{16}$$

with constants $\rho \in (0,1)$ and $a \geq b > 0$.

The alphabet for each feature m is denoted with $\{u_1^m, u_2^m, \cdots, u_{L_m}^m\}$, where L_m is the cardinality of feature m which, in this section, is assumed to be bounded and constant across n. The likelihoods of the features are defined as follows:

$$\alpha_{+,\ell_m}^m \triangleq \mathbb{P}(y_{i,m} = u_{\ell_m}^m | x_i = 1) \tag{17}$$

$$\alpha^m_{-,\ell_m} \triangleq \mathbb{P}(y_{i,m} = u^m_{\ell_m} | x_i = 0) \tag{18}$$

Recall that in our side information model, all features are independent conditioned on the labels. To ensure that the quality of the side information is increasing with n, both α^m_{+,ℓ_m} and α^m_{-,ℓ_m} are assumed to be either constant or monotonic in n.

To better understand the behavior of information limits, we categorize side information outcomes based on the trends of LLR and likelihoods. For simplicity we speak of trends for one feature of cardinality L; extension to multiple features is straightforward. Define h_ℓ to be the LLR of outcome ℓ, where $\ell \in \{1, \cdots, L\}$. An outcome is called *informative* if $h_\ell = O(\log(n))$ and *non-informative* if $h_\ell = o(\log(n))$. An outcome is called *rare* if $\log(\alpha_{\pm,\ell}) = O(\log(n))$ and *not rare* if $\log(\alpha_{\pm,\ell}) = o(\log(n))$. Among the four different combinations, the *worst* case is when the outcome is both *non-informative* and *not rare* for nodes inside and outside the community. We will show that if such an outcome exists, then side information will not improve the information limit. The *best* case is when the outcome is *informative* and *rare* for the nodes inside the community, or for the nodes outside the community, but not both. Two cases are in between: (1) an outcome that is *non-informative* and *rare* for nodes inside and outside the community and (2) an outcome that is *informative* and *not rare* for nodes inside and outside the community. It will be shown that the last three cases can affect the information limit under certain conditions.

For convenience we define:

$$T \triangleq \log\big(\frac{a}{b}\big) \tag{19}$$

We introduce the following functions whose value, as shown in the sequel, characterizes the exact recovery threshold:

$$\eta_1(\rho, a, b) \triangleq \rho\Big(b + \frac{a-b}{T}\log\big(\frac{a-b}{ebT}\big)\Big) \tag{20}$$

$$\eta_2(\rho, a, b, \beta) \triangleq \rho b + \frac{\rho(a-b)-\beta}{T}\log\big(\frac{\rho(a-b)-\beta}{\rho ebT}\big) + \beta \tag{21}$$

$$\eta_3(\rho, a, b, \beta) \triangleq \rho b + \frac{\rho(a-b)+\beta}{T}\log\big(\frac{\rho(a-b)+\beta}{\rho ebT}\big) \tag{22}$$

For example in the regime (16), one can conclude using (11) that exact recovery without side information is achieved if and only if $\eta_1 > 1$.

The LLR of each feature is denoted:

$$h^m_{\ell_m} \triangleq \log\Big(\frac{\alpha^m_{+,\ell_m}}{\alpha^m_{-,\ell_m}}\Big) \tag{23}$$

We also define the following functions of the likelihood and LLR of side information, whose evolution with n is critical to the phase transition of exact recovery [30].

$$f_1(n) \triangleq \sum_{m=1}^{M} h^m_{\ell_m}, \tag{24}$$

$$f_2(n) \triangleq \sum_{m=1}^{M} \log(\alpha^m_{+,\ell_m}), \tag{25}$$

$$f_3(n) \triangleq \sum_{m=1}^{M} \log(\alpha^m_{-,\ell_m}) \tag{26}$$

In the following, the side information outcomes $[u^1_{\ell_1}, \ldots, u^M_{\ell_M}]$ are represented by their index $[\ell_1, \ldots, \ell_M]$ without loss of generality. Throughout, dependence on n of outcomes and their likelihood is implicit.

Theorem 3. *In the regime characterized by* (16)*, assume M is constant and α^m_{+,ℓ_m} and α^m_{-,ℓ_m} are either constant or monotonic in n. Define $\xi_i \triangleq \lim_{n\to\infty} \frac{f_i(n)}{\log(n)}, i \in \{1, 2, 3\}$. Then, exact recovery occurs if and only if the following conditions are satisfied:*
Whenever for any sequence (over n) of side information outcomes $[\ell_1, \ldots, \ell_M]$

(i) $\xi_1 = \xi_2 = \xi_3 = 0$*, then* $\eta_1(\rho, a, b) > 1$ *holds.*
(ii) $\xi_1 = 0$ *and* $\xi_2 = \xi_3 < 0$*, then* $\eta_1(\rho, a, b) + \beta > 1$ *holds with* $\beta \triangleq -\xi_2 = -\xi_3$.
(iii) $0 < \xi_1 < \rho(a-b-bT)$ *and* $\xi_2 = 0$*, then* $\eta_2(\rho, a, b, \beta_1) > 1$ *holds with* $\beta_1 = \xi_1$.
(iv) $0 < \xi_1 < \rho(a-b-bT)$ *and* $\xi_3 = 0$*, then* $\eta_3(\rho, a, b, \beta_2) > 1$ *holds with* $\beta_2 \triangleq \xi_1$.
(v) $0 < \xi_1 < \rho(a-b-bT)$ *and* $\xi_2 < 0$*, then* $\eta_2(\rho, a, b, \beta_3) + \beta'_3 > 1$ *holds with* $\beta_3 \triangleq \xi_1$ *and* $\beta'_3 \triangleq -\xi_3$.
(vi) $0 < \xi_1 < \rho(a-b-bT)$ *and* $\xi_3 < 0$*, then* $\eta_3(\rho, a, b, \beta_4) + \beta'_4 > 1$ *holds with* $\beta_4 \triangleq \xi_1$ *and* $\beta'_4 \triangleq -\xi_3$.

Proof. For necessity, see Appendix H. For sufficiency, see Appendix I. □

Remark 7. *Each of Theorem 3 requirements could be triggered by just one sequence of outcomes, therefore the necessary and sufficient conditions might be an intersection of two or more conditions above. For example, if some outcome sequences satisfy $\xi_1 = 0$, $\xi_2 = \xi_3 < 0$ (Item (ii)) and some others satisfy $0 < \xi_1 < \rho(a-b-bT)$ and $\xi_2 = 0$ (Item (iii)), then the necessary and sufficient condition for exact recovery is* $\min\{\eta_1(\rho, a, b) + \beta\ ,\ \eta_2(\rho, a, b, \beta_1)\} > 1$.

Remark 8. *Theorem 3 does not address $f_1(n) = \omega(\log(n))$ because it leads to a trivial problem. For example, for noisy label side information, if the noise parameter $\alpha = e^{-n}$, then side information alone is sufficient for exact recovery. Also, when $f_1(n) = \beta\log(n)$ with $|\beta| \geq \rho(a-b-bT)$, a necessary condition is easily obtained but a matching sufficient condition for this case remains unavailable.*

In the following, we specialize the results of Theorem 3 to noisy-labels and partially-revealed-label side information.

Corollary 1. *For side information consisting of noisy labels with error probability $\alpha \in (0, 0.5)$, Theorem 3 combined with Lemma 17 state that exact recovery is achieved if and only if:*

$$\begin{cases} \eta_1(\rho, a, b) > 1, & \text{when } \log(\frac{1-\alpha}{\alpha}) = o(\log(n)) \\ \eta_2(\rho, a, b, \beta) > 1, & \text{when } \log(\frac{1-\alpha}{\alpha}) = (\beta + o(1))\log(n), \\ & \quad 0 < \beta < \rho(a-b-bT) \end{cases}$$

Figure 2 shows the error exponent for the noisy label side information as a function of β.

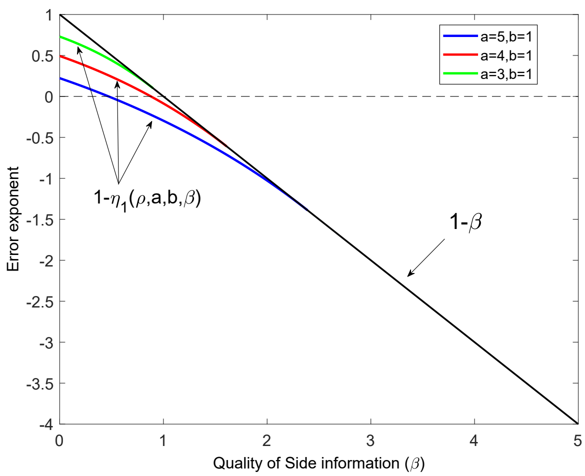

Fig. 2. Error exponent for noisy side information.

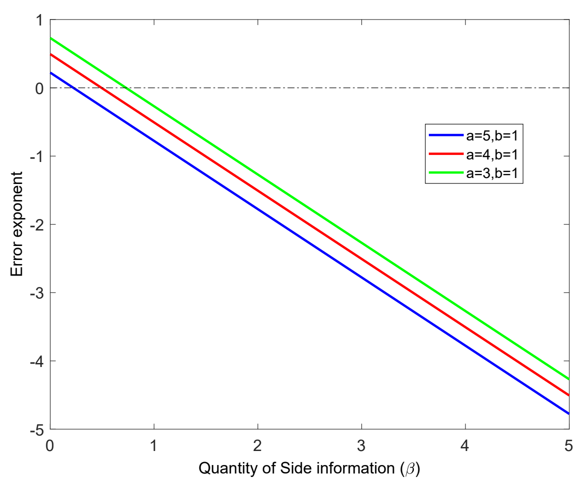

Fig. 3. Error exponent for partially revealed side information.

Corollary 2. *For side information consisting of a fraction* $1-\epsilon$ *of the labels revealed, Theorem 3 states that exact recovery is achieved if and only if:*

$$\begin{cases} \eta_1(\rho,a,b) > 1, & \textit{when } \log(\epsilon) = o(\log(n)) \\ \eta_1(\rho,a,b) + \beta > 1, & \textit{when } \log(\epsilon) = (-\beta + o(1))\log(n), \\ & \qquad \beta > 0 \end{cases}$$

Figure 3 shows the error exponent for partially revealed labels, as a function of β.

We now comment on the coverage of the regime (16). If the average degree of a node is $o(\log n)$, then the graph will have isolated nodes and exact recovery is impossible. If the average degree of the node is $\omega(\log n)$, then the problem is trivial. Therefore the regime of interest is when the average degree is $\Theta(\log n)$. This restricts Kp and Kq in a manner that is reflected in (16). Beyond that, in the system model of this paper $K = o(n)$, so $\frac{\log(\frac{n}{K})}{\log(n)}$ is either $o(1)$ or approaching a constant $C \in (0, 1]$. The regime (16) focuses on the former, but the proofs are easily modified to cover the latter. For the convenience of the reader, we highlight the places in the proof where a modification is necessary to cover the latter case.

III. Belief Propagation

Belief propagation for recovering a single community was studied *without* side information in [23], [25] in terms of a signal-to-noise ratio parameter $\lambda = \frac{K^2(p-q)^2}{(n-k)q}$, showing that *weak recovery* is achieved if and only if $\lambda > \frac{1}{e}$. Moreover, belief propagation followed by a local voting procedure was shown to achieve *exact recovery* if $\lambda > \frac{1}{e}$, as long as information limits allow exact recovery.

In this section $M = 1$, i.e. we consider scalar side information random variables that are discrete and take value from an alphabet size L. Extension to a vector side information is straightforward as long as dimensionality is constant across n; the extension is outlined in Corollary 3.

Denote the expectation of the likelihood ratio of the side information conditioned on $x = 1$ by:

$$\Lambda \triangleq \sum_{\ell=1}^{L} \frac{\alpha_{+,\ell}^2}{\alpha_{-,\ell}} \tag{27}$$

By definition, $\Lambda = \tilde{\chi}^2 + 1$, where $\tilde{\chi}^2$ is the chi-squared divergence between the conditional distributions of side information. Thus, $\Lambda \geq 1$.

A. Bounded LLR

We begin by presenting the belief propagation algorithm for community recovery with bounded side information. Define the message transmitted from node i to its neighboring node j at iteration $t+1$ as:

$$R_{i\to j}^{t+1} = h_i - K(p-q) + \sum_{k \in \mathcal{N}_i \setminus j} M(R_{k\to i}^t) \tag{28}$$

where $h_i = \log(\frac{\mathbb{P}(y_i|x_i=1)}{\mathbb{P}(y_i|x_i=0)})$, $\mathcal{N}_i$ is the set of neighbors of node i, $M(x) = \log(\frac{\frac{p}{q}e^{x-\nu}+1}{e^{x-\nu}+1})$ and $\nu = \log(\frac{n-K}{K})$.

The messages are initialized to zero for all nodes, i.e., $R_{i\to j}^0 = 0$ for all $i \in \{1, \cdots, n\}$ and $j \in \mathcal{N}_i$. The algorithm executes for $t-1$ iterations according to (28). At iteration t, the belief of each node is:

$$R_i^t = h_i - K(p-q) + \sum_{k \in \mathcal{N}_i} M(R_{k\to i}^{t-1}) \tag{29}$$

The algorithm then outputs the set of nodes corresponding to the K largest R_i^t. This algorithm is shown in Table II.

To analyze this algorithm, we begin by demonstrating its performance on a random *tree* with side information. Then, we show that the same performance is achieved on a random *graph* drawn from $\mathcal{G}(n, K, p, q)$ with side information. This is made possible via a coupling lemma [25] expressing local approximation of random graphs by trees.

TABLE II
BELIEF PROPAGATION ALGORITHM FOR COMMUNITY RECOVERY WITH SIDE INFORMATION.

Belief Propagation Algorithm
1) Input: $n, K, t \in \mathbb{N}$, $\boldsymbol{G}$ and $\boldsymbol{Y}$. 2) For all nodes i and $j \in \mathcal{N}_i$, set $R^0_{i\to j} = 0$. 3) For all nodes i and $j \in \mathcal{N}_i$, run $t-1$ iterations of belief propagation as in (28). 4) For all nodes i, compute its belief R^t_i based on (29). 5) Output $\tilde{C} = \{$Nodes corresponding to K largest $R^t_i\}$.

1) Belief Propagation on a Random Tree with Side Information:

We model random trees with side information in a manner roughly parallel to random graphs. Let T be an infinite tree with nodes i, each of them possessing a label $\tau_i \in \{0,1\}$. The root is node $i=0$. The subtree of depth t rooted at node i is denoted T^t_i. For brevity, the subtree rooted at $i=0$ with depth t is denoted T^t. Unlike the random graph counterpart, the tree and its node labels are generated together as follows: τ_0 is a Bernoulli-$\frac{K}{n}$ random variable. For any $i \in T$, the number of its children with label 1 is a random variable H_i that is Poisson with parameter Kp if $\tau_i = 1$, and Poisson with parameter Kq if $\tau_i = 0$. The number of children of node i with label 0 is a random variable F_i which is Poisson with parameter $(n-K)q$, regardless of the label of node i. The side information $\tilde{\tau}_i$ takes value in a finite alphabet $\{u_1, \cdots, u_L\}$. The set of all labels in T is denoted with $\boldsymbol{\tau}$, all side information with $\tilde{\boldsymbol{\tau}}$, and the labels and side information of T^t with $\boldsymbol{\tau}^t$ and $\tilde{\boldsymbol{\tau}}^t$ respectively. The likelihood of side information continues to be denoted by $\alpha_{+,\ell}, \alpha_{-,\ell}$, as earlier.

The problem of interest is to infer the label τ_0 given observations T^t and $\tilde{\boldsymbol{\tau}}^t$. The error probability of an estimator $\hat{\tau}_0(T^t, \tilde{\boldsymbol{\tau}}^t)$ can be written as:

$$p^t_e \triangleq \frac{K}{n}\mathbb{P}(\hat{\tau}_0 = 0|\tau_0 = 1) + \frac{n-K}{n}\mathbb{P}(\hat{\tau}_0 = 1|\tau_0 = 0) \quad (30)$$

The maximum a posteriori (MAP) detector minimizes p^t_e and can be written in terms of the log-likelihood ratio as $\hat{\tau}_{MAP} = \mathbb{1}_{\{\Gamma^t_0 \geq \nu\}}$, where $\nu = \log(\frac{n-K}{K})$ and:

$$\Gamma^t_0 = \log\left(\frac{\mathbb{P}(T^t, \tilde{\boldsymbol{\tau}}^t|\tau_0 = 1)}{\mathbb{P}(T^t, \tilde{\boldsymbol{\tau}}^t|\tau_0 = 0)}\right) \quad (31)$$

The probability of error of the MAP estimator can be bounded as follows [37]:

$$\frac{K(n-K)}{n^2}\rho^2 \leq p^t_e \leq \frac{\sqrt{K(n-K)}}{n}\rho \quad (32)$$

where $\rho = \mathbb{E}\big[e^{\frac{\Gamma^t_0}{2}}|\tau_0 = 0\big]$.

Lemma 3. *Let $\mathcal{N}_i$ denote the children of node i, $N_i \triangleq |\mathcal{N}_i|$ and $h_i \triangleq \log\big(\frac{\mathbb{P}(\tilde{\tau}_i|\tau_i=1)}{\mathbb{P}(\tilde{\tau}_i|\tau_i=0)}\big)$. Then,*

$$\Gamma^{t+1}_i = -K(p-q) + h_i + \sum_{k\in\mathcal{N}_i}\log\left(\frac{\frac{p}{q}e^{\Gamma^t_k-\nu}+1}{e^{\Gamma^t_k-\nu}+1}\right) \quad (33)$$

Proof. See Appendix K ☐

a) Lower and Upper Bounds on ρ:
Define for $t \geq 1$ and any node i:

$$\psi^t_i = -K(p-q) + \sum_{j\in\mathcal{N}_i} M(h_j + \psi^{t-1}_j) \quad (34)$$

where

$$M(x) \triangleq \log\Big(\frac{\frac{p}{q}e^{x-\nu}+1}{e^{x-\nu}+1}\Big) = \log\Big(1 + \frac{\frac{p}{q}-1}{1+e^{-(x-\nu)}}\Big).$$

Then, $\Gamma^{t+1}_i = h_i + \psi^{t+1}_i$ and $\psi^0_i = 0\ \forall i \in T^t$. Let Z^t_0 and Z^t_1 denote random variables drawn according to the distribution of ψ^t_i conditioned on $\tau_i = 0$ and $\tau_i = 1$, respectively. Similarly, let U_0 and U_1 denote random variables drawn according to the distribution of h_i conditioned on $\tau_i = 0$ and $\tau_i = 1$, respectively. Thus, $\rho = \mathbb{E}\big[e^{\frac{1}{2}(Z^t_0+U_0)}\big] = \mathbb{E}\big[e^{\frac{U_0}{2}}\big]\mathbb{E}\big[e^{\frac{Z^t_0}{2}}\big]$. Define:

$$b_t \triangleq \mathbb{E}\Big[\frac{e^{Z^t_1+U_1}}{1+e^{Z^t_1+U_1-\nu}}\Big] \quad (35)$$

$$a_t \triangleq \mathbb{E}\big[e^{Z^t_1+U_1}\big] \quad (36)$$

Lemma 4. *Let $B = (\frac{p}{q})^{1.5}$. Then:*

$$\mathbb{E}[e^{\frac{U_0}{2}}]e^{\frac{-\lambda}{8}b_t} \leq \rho \leq \mathbb{E}[e^{\frac{U_0}{2}}]e^{\frac{-\lambda}{8B}b_t} \quad (37)$$

Proof. See Appendix L. ☐

Thus to bound ρ, lower and upper bounds on b_t are needed.

Lemma 5. *For all $t \geq 0$, if $\lambda \leq \frac{1}{\Lambda e}$, then $b_t \leq \Lambda e$.*

Proof. See Appendix M. ☐

Lemma 6. *Define $C = \lambda(2+\frac{p}{q})$ and $\Lambda' = \mathbb{E}[e^{3U_0}]$. Assume that $b_t \leq \frac{\nu}{2(C-\lambda)}$. Then,*

$$b_{t+1} \geq \Lambda e^{\lambda b_t}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu}{2}}) \quad (38)$$

Proof. See Appendix N. ☐

Lemma 7. *The sequences a_t and b_t are non-decreasing in t.*

Proof. The proof follows directly from [25, Lemma 5]. ☐

Lemma 8. *Define $\log^*(\nu)$ to be the number of times the logarithm function must be iteratively applied to ν to get a result less than or equal to one. Let $C = \lambda(2+\frac{p}{q})$ and*

$\Lambda' = \mathbb{E}[e^{3U_0}]$. *Suppose* $\lambda > \frac{1}{\Lambda e}$. *Then there are constants* $\bar{t}_o$ *and* ν_o *depending only on* λ *and* Λ *such that:*

$$b_{\bar{t}_o+\log^*(\nu)+2} \geq \Lambda e^{\frac{\lambda\nu}{2(C-\lambda)}}(1-\frac{\Lambda'}{\Lambda}e^{\frac{-\nu}{2}}) \tag{39}$$

whenever $\nu \geq \nu_o$ *and* $\nu \geq 2\Lambda(C-\lambda)$.

Proof. See Appendix O. □

b) Achievability and Converse for the MAP Detector:

Lemma 9. *Let* $\Lambda' = \mathbb{E}[e^{3U_0}]$, $C = \lambda(2+\frac{p}{q})$ *and* $B = (\frac{p}{q})^{1.5}$. *If* $0 < \lambda \leq \frac{1}{\Lambda e}$, *then:*

$$p_e^t \geq \frac{K(n-K)}{n^2}\mathbb{E}^2[e^{\frac{U_0}{2}}]e^{\frac{-\lambda\Lambda e}{4}} \tag{40}$$

If $\lambda > \frac{1}{\Lambda e}$, *then:*

$$p_e^t \leq \sqrt{\frac{K(n-K)}{n^2}}\mathbb{E}[e^{\frac{U_0}{2}}]e^{\frac{-\lambda\Lambda}{8B}e^{\frac{\lambda\nu}{2(C-\lambda)}}(1-\frac{\Lambda'}{\Lambda}e^{\frac{-\nu}{2}})} \tag{41}$$

Moreover, since $\nu \to \infty$*:*

$$p_e^t \leq \sqrt{\frac{K(n-K)}{n^2}}\mathbb{E}[e^{\frac{U_0}{2}}]e^{-\nu(r+\frac{1}{2})} = \frac{K}{n}e^{-\nu(r+o(1))} \tag{42}$$

for some $r > 0$.

Proof. The proof follows directly from (32) and Lemmas 5 and 8. □

2) Performance of Belief Propagation for Community Recovery with Side Information:

In this section, the inference problem defined on the random tree is coupled to the problem of recovering a hidden community with side information. This can be done via a coupling lemma [25] that shows that under certain conditions, the neighborhood of a fixed node i in the graph is locally a tree with probability converging to one, and hence, the belief propagation algorithm defined for random trees in Section III-A1 can be used on the graph as well. The proof of the coupling lemma depends only on the tree structure, implying that it also holds for our system model, where the side information is independent of the tree structure given the labels.

Define $\boldsymbol{G}_u^{\hat{t}}$ to be the subgraph containing all nodes that are at a distance at most $\hat{t}$ from node u and define $\boldsymbol{x}_u^{\hat{t}}$ and $\boldsymbol{Y}_u^{\hat{t}}$ to be the set of labels and side information of all nodes in $\boldsymbol{G}_u^{\hat{t}}$, respectively.

Lemma 10 (Coupling Lemma [25])**.** *Suppose that* $\hat{t}(n)$ *are positive integers such that* $(2+np)^{\hat{t}(n)} = n^{o(1)}$. *Then:*

- *If the size of community is deterministic and known, i.e.,* $|C^*| = K$, *then for any node* u *in the graph, there exists a coupling between* $(\boldsymbol{G}, \boldsymbol{x}, \boldsymbol{Y})$ *and* $(T, \boldsymbol{\tau}, \tilde{\boldsymbol{\tau}})$ *such that:*

$$\mathbb{P}((\boldsymbol{G}_u^{\hat{t}}, \boldsymbol{x}_u^{\hat{t}}, \boldsymbol{Y}_u^{\hat{t}}) = (T^{\hat{t}}, \boldsymbol{\tau}^{\hat{t}}, \tilde{\boldsymbol{\tau}}^{\hat{t}})) \geq 1 - n^{-1+o(1)} \tag{43}$$

  where for convenience of notation, the dependence of $\hat{t}$ *on* n *is made implicit.*
- *If* $|C^*|$ *obeys a probability distribution so that* $\mathbb{P}(||C^*| - K| \geq \sqrt{3K\log(n)}) \leq n^{\frac{-1}{2}+o(1)}$ *with* $K \geq 3\log(n)$, *then for any node* u, *there exists a coupling between* $(\boldsymbol{G}, \boldsymbol{x}, \boldsymbol{y})$ *and* $(T, \boldsymbol{\tau}, \tilde{\boldsymbol{\tau}})$ *such that:*

$$\mathbb{P}((\boldsymbol{G}_u^{\hat{t}}, \boldsymbol{x}_u^{\hat{t}}, \boldsymbol{Y}_u^{\hat{t}}) = (T^{\hat{t}}, \boldsymbol{\tau}^{\hat{t}}, \tilde{\boldsymbol{\tau}}^{\hat{t}})) \geq 1 - n^{\frac{-1}{2}+o(1)} \tag{44}$$

We now return to the belief propagation algorithm highlighted in Table II. If we have $t = \hat{t}(n)$, according to Lemma 10 with probability converging to one $R_i^t = \Gamma_i^t$, where Γ_i^t was the log-likelihood defined for the random tree. Hence, the performance of belief propagation is the same as the MAP estimator defined as $\hat{\tau}_{MAP} = \mathbb{1}_{\{\Gamma_i^t \geq \nu\}}$, where $\nu = \log(\frac{n-K}{K})$. The only difference is that the MAP estimator decides based on $\Gamma_i^t \geq \nu$ while Algorithm II selects the K largest R_i^t. To manage this difference, let $\hat{C}$ define the community recovered by the MAP estimator, i.e. $\hat{C} = \{i : R_i^t \geq \nu\}$. Since $\tilde{C}$ is the set of nodes with the K largest R_i^t. Then,

$$\begin{aligned}|C^*\triangle\tilde{C}| &\leq |C^*\triangle\hat{C}| + |\hat{C}\triangle\tilde{C}| \\ &= |C^*\triangle\hat{C}| + ||\hat{C}| - K|\end{aligned} \tag{45}$$

Moreover,

$$\begin{aligned}||\hat{C}| - K| &\leq ||\hat{C}| - |C^*|| + ||C^*| - K| \\ &\leq |C^*\triangle\hat{C}| + ||C^*| - K|\end{aligned} \tag{46}$$

Using (46) and substituting in (45):

$$|C^*\triangle\tilde{C}| \leq 2|C^*\triangle\hat{C}| + ||C^*| - K| \tag{47}$$

We will use (47) to prove weak recovery.

a) Weak Recovery:

Theorem 4.
Achievability: *Suppose that* $(np)^{\log^*(\nu)} = n^{o(1)}$ *and* $\lambda > \frac{1}{\Lambda e}$. *Let* $\hat{t}(n) = \bar{t}_o + \log^*(\nu) + 2$, *where* $\bar{t}_o$ *is a constant depending only on* λ *and* Λ. *Apply Algorithm II with* $t = \hat{t}(n)$ *resulting in estimated community* $\tilde{C}$. *Then:*

$$\frac{\mathbb{E}[|C^*\triangle\tilde{C}|]}{K} \to 0 \tag{48}$$

for either $|C^*| = K$ *or random* $|C^*|$ *such that* $K \geq 3\log(n)$ *and* $\mathbb{P}(||C^*| - K| \geq \sqrt{3K\log(n)}) \leq n^{\frac{-1}{2}+o(1)}$.

Converse: *Suppose that* $\lambda \leq \frac{1}{\Lambda e}$. *Let* $\hat{t} \in \mathbb{N}$ *depend on* n *such that* $(2+np)^{\hat{t}} = n^{o(1)}$. *Then, for any local estimator* $\hat{C}$ *of* x_u^* *that has access to observations of the graph and side information limited to a neighborhood of radius* $\hat{t}$ *from* u,

$$\frac{\mathbb{E}[|C^*\triangle\hat{C}|]}{K} \geq (1-\frac{K}{n})\mathbb{E}^2[e^{\frac{U_0}{2}}]e^{\frac{-\lambda\Lambda e}{4}} - o(1) \tag{49}$$

Proof. For achievability, see Appendix P. For necessity, see Appendix Q. □

Corollary 3. *The same result holds for side information consisting of multiple features, i.e., constant* $M \geq 1$. *In other words, using the same notation as in Section II-B2, weak recovery is possible if and only if* $\lambda > \frac{1}{\Lambda e}$ *where* $\Lambda = \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_M=1}^{L_M}(\prod_{m=1}^{M}\frac{(\alpha_{+,\ell_m}^m)^2}{\alpha_{-,\ell_m}^m})$.

Remark 9. *It was shown in [25] that belief propagation achieves weak recovery without side information for all* $\lambda > C\frac{K}{n}\log(\frac{n}{K})$, *for some positive constant* C. *Building* *on this observation, Theorem 4 shows that belief propagation (with side information) achieves the maximum likelihood performance (without side information) when $\Lambda \to \infty$ fast enough, i.e., $\Omega(\frac{n}{K \log(\frac{n}{K})})$. In Section III-B, we show that belief propagation could achieve weak recovery for any fixed $\lambda > 0$ when $\Lambda \to \infty$ at a specific rate.*

b) Exact Recovery:

In Section II-B, it was shown that under certain conditions any estimator that achieves weak recovery on a random cluster size will also achieve exact recovery if followed by a local voting process. This can be used to demonstrate sufficient conditions for exact recovery under belief propagation. To do so, we employ a modified form of the algorithm in Table I, where in Step 3 for weak recovery we use the belief propagation algorithm presented in Table II.

Theorem 5. *Suppose that $(np)^{\log^*(\nu)} = n^{o(1)}$ and $\lambda > \frac{1}{\Lambda e}$. Let $\delta \in (0,1)$ such that $\frac{1}{\delta} \in \mathbb{N}$, $n\delta \in \mathbb{N}$ and $\lambda(1-\delta) > \frac{1}{\Lambda e}$. Let $\hat{t} = \bar{t}_o + \log^*(n) + 2$, where $\bar{t}_o$ is a constant depending only on $\lambda(1-\delta)$ and Λ as described in Lemma 8. Assume that (11) holds. Let $\tilde{C}$ be the estimated community produced by the modified version of Algorithm I with $t = \hat{t}(n)$. Then $\mathbb{P}(\tilde{C} = C^*) \to 1$ as $n \to \infty$.*

Proof. See Appendix R. □

c) Comparison with Information Limits:

Since $K \to \infty$ and the LLRs are bounded, the weak recovery result in Theorem 1 reduces to $\liminf_{n\to\infty} \frac{KD(P||Q)}{2\log(\frac{n}{K})} > 1$. This condition can be written as [25]:

$$\lambda > C\frac{K}{n}\log(\frac{n}{K}) \tag{50}$$

for some positive constant C. Thus, weak recovery only demands a vanishing λ. On the other hand, belief propagation achieves weak recovery for $\lambda > \frac{1}{\Lambda e}$, where Λ is greater than one and bounded as long as LLR is bounded. This implies a gap between the information limits and belief propagation limits for weak recovery. Since $\Lambda \geq 1$, side information diminishes the gap.

For exact recovery, the following regime is considered:

$$K = \frac{cn}{\log(n)}, \; q = \frac{b\log^2(n)}{n}, \; p = 2q \tag{51}$$

for fixed positive b, c as $n \to \infty$. In this regime, $KD(P||Q) = O(\log(n))$, and hence, weak recovery is always asymptotically possible. Also, $\lambda = c^2 b$. Moreover, exact recovery is asymptotically possible if $cb(1 - \frac{1+\log\log(2)}{\log(2)}) > 1$. For belief propagation, we showed that exact recovery is possible if $cb(1 - \frac{1+\log\log(2)}{\log(2)}) > 1$ and $\lambda > \frac{1}{\Lambda e}$.

Figure 4 compares the regions where weak recovery is achieved for belief propagation with and without side information, as well as exact recovery with bounded-LLR side information. Side information with $L = 2$ is considered, where each node observes a noisy label with cross-over probability $\alpha = 0.3$. In Region 1, the belief propagation algorithm followed by voting achieves exact recovery with no need for side information. In Region 2, belief propagation followed by voting achieves exact recovery with side information, but not without. In Region 3, weak recovery is achieved by belief propagation with no need for side information, but exact recovery is not asymptotically possible. In Region 4, weak recovery is achieved by the belief propagation as long as side information is available; exact recovery is not asymptotically possible. In Region 5, exact recovery is asymptotically possible, but belief propagation without side information or with side information whose $\alpha = 0.3$ cannot achieve even weak recovery (needs smaller α, i.e., better side information). In Region 6, weak recovery, but not exact recovery, is asymptotically possible via optimal algorithms, but belief propagation without side information or with side information whose $\alpha = 0.3$ cannot achieve even weak recovery.

Figure 5 explores the effect of different values of α, showing that as quality of side information improves (smaller α), the gap between the belief propagation limit and the information limit decreases.

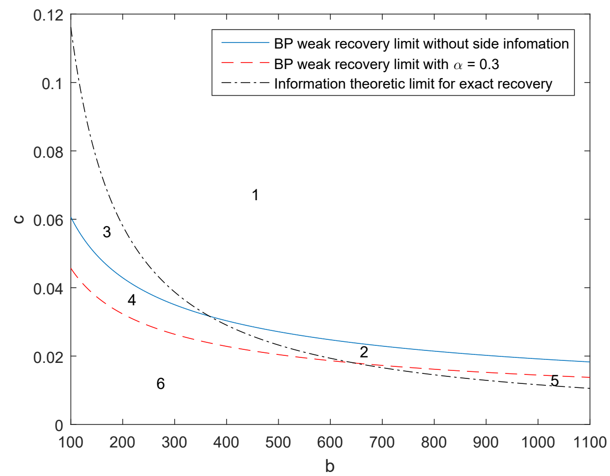

Fig. 4. Phase diagram with $K = c\frac{n}{\log(n)}$, $q = \frac{b\log^2(n)}{n}$, $p = 2q$ and $\alpha = 0.3$ for b, c fixed as $n \to \infty$.

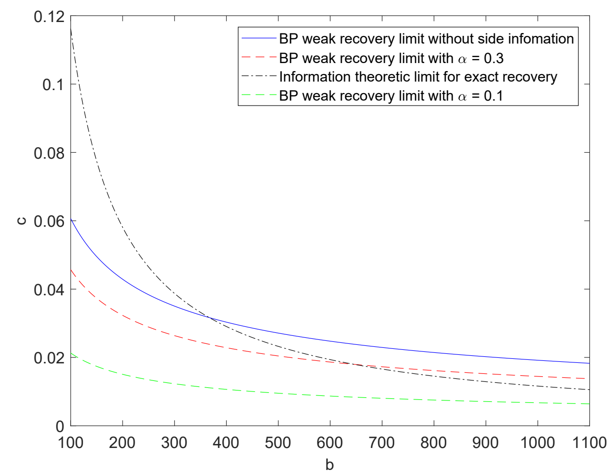

Fig. 5. Phase diagram with $K = c\frac{n}{\log(n)}$, $q = \frac{b\log^2(n)}{n}$, $p = 2q$ and $\alpha = 0.3, 0.1$ for b, c fixed as $n \to \infty$.

B. Unbounded LLR

The results of the previous section suggest that when $\Lambda \to \infty$ arbitrarily slowly, belief propagation achieves weak recovery for any fixed $\lambda > 0$. In this section we prove this result for scalar side information with finite cardinality and Λ that grows at a specific rate.

The proof technique uses density evolution of Γ_i^t. More precisely, we assume that ν, $\frac{\alpha_{+,\ell}}{\alpha_{-,\ell}}$, and λ are constants independent of n, while $nq, Kq \xrightarrow{n\to\infty} \infty$, which implies that $\frac{p}{q} \xrightarrow{n\to\infty} 1$. This assumption allows us to precisely characterize the conditional probability density function of Γ_i^t (asymptotically Gaussian), and hence, calculate the fraction of misclassified labels via the Q-function. Then, $\frac{n}{K}$ is allowed to grow and the behavior of the fraction of misclassified labels is studied as ν and the LLR of the side information grow.

Recall the definition of ψ_i^t from (34) and Γ_i^t from (31) as well as the definitions of Z_0^t, Z_1^t, U_0 and U_1 defined directly afterward.

Lemma 11. *Assume λ, $\frac{\alpha_{+,\ell}}{\alpha_{-,\ell}}$ and ν are constants independent of n while $nq, Kq \xrightarrow{n\to\infty} \infty$. Then, for all $t \geq 0$:*

$$\mathbb{E}[Z_0^{t+1}] = \frac{-\lambda}{2} b_t + o(1) \tag{52}$$

$$\mathbb{E}[Z_1^{t+1}] = \frac{\lambda}{2} b_t + o(1) \tag{53}$$

$$var(Z_0^{t+1}) = var(Z_1^{t+1}) = \lambda b_t + o(1) \tag{54}$$

Proof. See Appendix S. □

The following lemma shows that the distributions of Z_1^t and Z_0^t are asymptotically Gaussian.

Lemma 12. *Assume λ, $\frac{\alpha_{+,\ell}}{\alpha_{-,\ell}}$ and ν are constants independent of n while $nq, Kq \xrightarrow{n\to\infty} \infty$. Let $\phi(x)$ be the cumulative distribution function (CDF) of a standard normal distribution. Define $v_0 = 0$ and $v_{t+1} = \lambda \mathbb{E}_{Z,U_1}[\frac{1}{e^{-\nu} + e^{-(\frac{v_t}{2} + \sqrt{v_t}Z) - U_1}}]$, where $Z \sim \mathcal{N}(0,1)$. Then, for all $t \geq 0$:*

$$\sup_x \left|\mathbb{P}(\frac{Z_0^{t+1} + \frac{v_{t+1}}{2}}{\sqrt{v_{t+1}}} \leq x) - \phi(x)\right| \to 0 \tag{55}$$

$$\sup_x \left|\mathbb{P}(\frac{Z_1^{t+1} - \frac{v_{t+1}}{2}}{\sqrt{v_{t+1}}} \leq x) - \phi(x)\right| \to 0 \tag{56}$$

Proof. See Appendix T. □

Lemma 13. *Assume λ, $\frac{\alpha_{+,\ell}}{\alpha_{-,\ell}}$ and ν are constants independent of n while $nq, Kq \xrightarrow{n\to\infty} \infty$. Let $\hat{C}$ define the community recovered by the MAP estimator, i.e. $\hat{C} = \{i : \Gamma_i^t \geq \nu\}$. Then,*

$$\lim_{nq,Kq\to\infty} \lim_{n\to\infty} \frac{\mathbb{E}[\hat{C}\triangle C^*]}{K} = \frac{n-K}{K}\mathbb{E}_{U_0}[Q(\frac{\nu + \frac{v_t}{2} - U_0}{\sqrt{v_t}})] + \mathbb{E}_{U_1}[Q(\frac{-\nu + \frac{v_t}{2} + U_1}{\sqrt{v_t}})] \tag{57}$$

where $v_0 = 0$ and $v_{t+1} = \lambda \mathbb{E}_{Z,U_1}[\frac{1}{e^{-\nu} + e^{-(\frac{v_t}{2} + \sqrt{v_t}Z) - U_1}}]$, and $Z \sim \mathcal{N}(0,1)$.

Proof. Let $p_{e,0}, p_{e,1}$ denote Type I and Type II errors for recovering τ_0. Then the proof follows from Lemmas 11 and 12, and because

$$\frac{\mathbb{E}[\hat{C}\triangle C^*]}{K} = \frac{n}{K}p_e^t = \frac{n-K}{K}p_{e,0} + p_{e,1}.$$

□

Lemma 13 applies for side information with cardinality $L \geq 1$, and hence, generalizes [26] which was limited to $L = 2$. Now $\frac{n}{K}$ is allowed to grow and the behavior of the fraction of misclassified labels is studied as ν and the LLR of the side information grows without bound. The following lemma shows that if $\Lambda \to \infty$ such that $|h_\ell| = |\log(\frac{\alpha_{+,\ell}}{\alpha_{-,\ell}})| < \nu$, belief propagation achieves weak recovery for any fixed $\lambda > 0$ upon observing the tree structure of depth t^*+2 and side information with finite L, where $t^* = \log^*(\nu)$ is the number of times the logarithm function must be iteratively applied to ν to get a result less than or equal to one.

Lemma 14. *Let $\hat{C}$ be the output of the MAP estimator for the root of a random tree of depth $t^* + 2$ upon observing the tree structure and side information with cardinality $L < \infty$. Assume as $\frac{n}{K} \to \infty$, $\Lambda \to \infty$ such that $|h_\ell| < \nu$. Then for any fixed $\lambda > 0$:*

$$\lim_{\frac{n}{K}\to\infty} \lim_{nq,Kq\to\infty} \lim_{n\to\infty} \frac{\mathbb{E}[\hat{C}\triangle C^*]}{K} = 0 \tag{58}$$

Proof. See Appendix U. □

Although Lemma 14 is for L-ary side information, it focuses on one asymptotic regime of side information where $|h_\ell| < \nu$. To study other asymptotic regimes of side information, one example is considered for $L = 2$, i.e., side information takes values in $\{0, 1\}$. For constants $\eta, \beta \in (0, 1)$ and $\gamma > 0$, define:

$$\alpha_{+,1} = \mathbb{P}(y = 1|x^* = 1) = \eta\beta$$

$$\alpha_{-,1} = \mathbb{P}(y = 1|x^* = 0) = \frac{\eta(1-\beta)}{(\frac{n-K}{K})^\gamma} \tag{59}$$

Thus, $\Lambda \to \infty$ and $h_1 = (1 + o(1))\gamma \log(\frac{n-K}{K})$ and $h_2 = (1 + o(1)) \log(1 - \eta\beta)$. For $0 < \gamma < 1$, Lemma 14 shows that belief propagation achieves weak recovery for any fixed $\lambda > 0$. This implies that belief propagation achieves weak recovery also for $\gamma \geq 1$ because $\gamma \geq 1$ implies higher-quality side information. This generalizes the results obtained in [26] which was only for $\gamma = 1$.

1) Belief Propagation Algorithm for Community Recovery with Unbounded Side Information:

Lemma 13 characterizes the performance of the optimal estimator of the root of a random tree upon observing the tree of depth t and the side information. Similar to Section III-A2, the inference problem defined on the random tree is coupled to the problem of recovering a hidden community with side information. This is done via Lemma 10, which together with Equation (47) allow us to use Algorithm II (as long as $(np)^t = n^{o(1)}$). Let $\tilde{C}$ be the output of Algorithm II, i.e., the set of nodes with the K largest R_i^t. Then, using Equation (47) we have: $\frac{\mathbb{E}[\tilde{C}\triangle C^*]}{K} \leq 2\frac{\mathbb{E}[\hat{C}\triangle C^*]}{K}$. Thus, the results of Lemma 14 and the special case (59) hold. This also suggests that belief propagation (Algorithm II) achieves weak recovery for any $\lambda > 0$ when Λ grows with $\frac{n}{K}$ arbitrarily slowly.

IV. Conclusion

This paper studies the effect of the quality and quantity of side information on the recovery of a hidden community of size $K = o(n)$. Under maximum likelihood detection, tight necessary and sufficient conditions are calculated for exact recovery, where we demonstrate how side information must evolve with n in terms of either quantity or quality to improve the exact recovery threshold. A similar set of results are obtained for weak recovery. Under belief propagation, tight necessary and sufficient conditions for weak recovery are calculated when the LLRs are constant, and sufficient conditions when the LLRs vary with n. It is established that belief propagation followed by a local voting procedure achieves exact recovery, and its performance gap with respect to ML is reduced by side information. Simulations on finite synthetic data-sets show that the asymptotic results of this paper are relevant in assessing the performance of belief propagation at finite n.

Appendix

A. Auxiliary Lemmas For Information Limits

Lemma 15. *Define*

$$\hat{E}_{QU}(\theta, m_1, m_2) \triangleq \sup_{t\in\mathbb{R}} t\theta - m_1 \log_Q(\mathbb{E}[e^{tL_G}]) - m_2 \log_U(\mathbb{E}[e^{tL_S}])$$
$$\hat{E}_{PV}(\theta, m_1, m_2) \triangleq \sup_{t\in\mathbb{R}} t\theta - m_1 \log_P(\mathbb{E}[e^{tL_G}]) - m_2 \log_V(\mathbb{E}[e^{tL_S}])$$

For $\theta \in [-m_1 D(Q||P) - m_2 D(U||V), m_1 D(P||Q) + m_2 D(V||U)]$, *the following holds:*

$$\hat{E}_{QU}(\theta, m_1, m_2) = E_{QU}(\theta, m_1, m_2) \tag{60}$$
$$\hat{E}_{PV}(\theta, m_1, m_2) = E_{PV}(\theta, m_1, m_2) \tag{61}$$

Moreover, for $\delta : -m_1 D(Q||P) - m_2 D(U||V) \le \theta \le \theta + \delta \le m_1 D(P||Q) + m_2 D(V||U)]$, *the following holds:*

$$E_{QU}(\theta, m_1, m_2) \le E_{QU}(\theta+\delta, m_1, m_2) \le E_{QU}(\theta, m_1, m_2) + \delta \tag{62}$$
$$E_{PV}(\theta, m_1, m_2) \ge E_{PV}(\theta+\delta, m_1, m_2) \ge E_{PV}(\theta, m_1, m_2) - \delta \tag{63}$$

Proof. Equations (60) and (61) follow since $E_{PV}(\theta, m_1, m_2) = E_{QU}(\theta, m_1, m_2) - \theta$ and because:

$$E_{QU}(-m_1 D(Q||P) - m_2 D(U||V), m_1, m_2) = 0$$
$$E_{PV}(m_1 D(P||Q) + m_2 D(V||U), m_1, m_2) = 0$$
$$\psi'_{QU}(m_1, m_2, 0) = \psi'_{PV}(m_1, m_2, -1) = - m_1 D(Q||P) - m_2 D(U||V)$$
$$\psi'_{QU}(m_1, m_2, 1) = \psi'_{PV}(m_1, m_2, 0) = m_1 D(P||Q) + m_2 D(V||U) \tag{64}$$

Equations (62) and (63) follow since $E_{PV}(E_{QU})$ is decreasing (increasing) for $\theta \in [-m_1 D(Q||P) - m_2 D(U||V), m_1 D(P||Q) + m_2 D(V||U)]$. □

Lemma 16. *Assume* $|L_G| \le B$ *and* $|L_S| \le B'$ *for some positive constants* B *and* B'. *Define* $B'' = \max\{B, B'\}$. *Then, for* $t \in [-1, 1]$ *and* $\eta \in [0, 1]$,

$$\psi''_{QU}(m_1, m_2, t) \le 2e^{5B''}\Big(\min\big\{m_1 D(Q||P) + m_2 D(U||V), m_1 D(P||Q) + m_2 D(V||U)\big\}\Big) \tag{65}$$
$$\psi_{QU}(m_1, m_2, t) \le (m_1 D(Q||P) + m_2 D(U||V)) \times (-t + e^{5B''} t^2) \tag{66}$$
$$E_{QU}\Big(m_1, m_2, -(1-\eta)(m_1 D(Q||P) + m_2 D(U||V))\Big) \ge \frac{\eta^2}{4e^{5B''}}(m_1 D(Q||P) + m_2 D(U||V)) \tag{67}$$
$$\psi''_{PV}(m_1, m_2, t) \le 2e^{5B''}\Big(\min\big\{m_1 D(Q||P) + m_2 D(U||V), m_1 D(P||Q) + m_2 D(V||U)\big\}\Big) \tag{68}$$
$$\psi_{PV}(m_1, m_2, t) \le (m_1 D(P||Q) + m_2 D(V||U)) \times (t + e^{5B''} t^2) \tag{69}$$
$$E_{PV}(m_1, m_2, (1-\eta)(m_1 D(P||Q) + m_2 D(V||U))) \ge \frac{\eta^2}{4e^{5B''}}(m_1 D(P||Q) + m_2 D(V||U)) \tag{70}$$

where $\psi''_{QU}(m_1, m_2, t)$ *and* $\psi''_{PV}(m_1, m_2, t)$ *denote the second derivatives with respect to* t.

Proof. By direct computation of the second derivative,

$$\psi''_{QU}(m_1, m_2, t) \le m_1 \frac{\mathbb{E}_Q[L_G^2 e^{tL_G}]}{\mathbb{E}_Q[e^{tL_G}]} + m_2 \frac{\mathbb{E}_U[L_S^2 e^{tL_S}]}{\mathbb{E}_U[e^{tL_S}]} \overset{(a)}{\le} m_1 e^{2B} \mathbb{E}_Q[L_G^2] + m_2 e^{2B'} \mathbb{E}_U[L_S^2] \tag{71}$$

where (a) follows by the assumption that $|L_G| \le B$, $|L_S| \le B'$ and holds for all $t \in [-1, 1]$.

Now consider the following function: $\phi(x) = e^x - 1 - x$ restricted to $|x| \le B$. It is easy to see that $\phi(x)$ is non-negative, convex with $\phi(0) = \phi'(0) = 0$ and $\phi''(x) = e^x$. Hence, $e^{-B} \le \phi''(x) \le e^B$. From Taylor's theorem with integral remainder [38], we get: $\frac{e^{-B}x^2}{2} \le \phi(x) \le \frac{e^B x^2}{2}$, which implies $x^2 \le 2e^B \phi(x)$. Using this result for $x = L_G$ and $x = L_S$:

$$\mathbb{E}_Q[L_G^2] \le 2e^B \mathbb{E}_Q[\phi(L_G)] = 2e^B D(Q||P) \tag{72}$$
$$\mathbb{E}_U[L_S^2] \le 2e^{B'} \mathbb{E}_U[\phi(L_S)] = 2e^{B'} D(U||V) \tag{73}$$

Combining (71), (72), (73) lead to $\psi''_{QU}(m_1, m_2, t) \le 2m_1 e^{3B} D(Q||P) + 2m_2 e^{3B'} D(U||V)$ for $t \in [-1, 1]$. Similarly, it can shown for $t \in [0, 2]$: $\psi''_{QU}(m_1, m_2, t) \le 2m_1 e^{5B} D(Q||P) + 2m_2 e^{5B'} D(U||V)$.

On the other hand, using $\phi(x) = e^{-x} - 1 + x$ with $|x| \le B$, it can be shown that $\psi''_{PV}(m_1, m_2, t) \le 2m_1 e^{5B} D(P||Q) + 2m_2 e^{5B'} D(V||U)$, for $t \in [0, 2]$. By definition, $\psi_{QU}(m_1, m_2, t) = \psi_{PV}(m_1, m_2, t-1)$, and hence, $\psi''_{QU}(m_1, m_2, t) \le 2m_1 e^{5B} D(P||Q) + 2m_2 e^{5B'} D(V||U)$, for $t \in [-1, 1]$, which concludes the proof of (65). The proof of (68) follows similarly.

Now since $\psi_{QU}(m_1,m_2,0) = 0$ and $\psi'_{QU}(m_1,m_2,0) = -m_1 D(Q||P) - m_2 D(U||V)$, then using Taylor's theorem with integral remainder, we have for $t \in [-1,1]$:

$$\begin{aligned}
&\psi_{QU}(m_1,m_2,t)\\
&= \psi_{QU}(m_1,m_2,0) + t\psi'_{QU}(m_1,m_2,0)\\
&\quad + \int_t^0 (\lambda - t)\psi''_{QU}(m_1,m_2,t)d\lambda\\
&\overset{(a)}{\le} -t(m_1 D(Q||P) + m_2 D(U||V))\\
&\quad + e^{5B''}(m_1 D(Q||P) + m_2 D(U||V))t^2 \qquad (74)
\end{aligned}$$

where (a) follows using (65). Similarly, it can be shown that:

$$\begin{aligned}
&\psi_{PV}(m_1,m_2,t) \le t(m_1 D(P||Q) + m_2 D(V||U))\\
&+ e^{5B''}(m_1 D(P||Q) + m_2 D(V||U))t^2 \qquad (75)
\end{aligned}$$

Combining (74) and (75) concludes the proof of (66), (69). Using (66) and (69), we get:

$$\begin{aligned}
&E_{QU}\Big(m_1,m_2,-(1-\eta)(m_1 D(Q||P) + m_2 D(U||V))\Big)\\
&\ge \sup_{t\in[0,1]} t(-(1-\eta)(m_1 D(Q||P) + m_2 D(U||V)))\\
&\quad + t(m_1 D(Q||P) + m_2 D(U||V)) \qquad (76)\\
&\quad - e^{5B''}(m_1 D(Q||P) + m_2 D(U||V))t^2\\
&= \frac{\eta^2}{4e^{5B''}}(m_1 D(Q||P) + m_2 D(U||V)) \qquad (77)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&E_{PV}\Big(m_1,m_2,(1-\eta)(m_1 D(P||Q) + m_2 D(V||U))\Big)\\
&\ge \frac{\eta^2}{4e^{5B''}}(m_1 D(P||Q) + m_2 D(V||U)) \qquad (78)
\end{aligned}$$

Combining (76) and (78) concludes the proof of (67), (70). ☐

Lemma 17. *$\eta_3(\rho,a,b,\beta) \ge \eta_2(\rho,a,b,\beta)$, for $0 < \beta < \rho(a - b - bT)$.*

Proof. It is easy to show that $\eta_3(\rho,a,b,\beta) - \beta$ is convex in $\beta > 0$. Thus, the optimal β can be calculated as $\beta^* = \rho(aT - a + b)$ at which $\eta_3(\rho,a,b,\beta^*) - \beta^* = 0$. Thus, $\eta_3(\rho,a,b,\beta) \ge \beta$ for all $a \ge b > 0$.

Furthermore, note that $\eta_2(\rho,a,b,\beta)$ is convex and increasing in $0 < \beta < \rho(a-b-bT)$. By direct substitution, it can be shown that at $\beta = \rho(a-b-bT)$: $\eta_2(\rho,a,b,\beta) = \beta$. This implies that at $\beta = \rho(a-b-bT)$:

$$\eta_3(\rho,a,b,\beta) - \eta_2(\rho,a,b,\beta) = \eta_3(\rho,a,b,\beta) - \beta \ge 0 \qquad (79)$$

Using (79) together with the fact that $\eta_3(\rho,a,b,\beta) - \eta_2(\rho,a,b,\beta)$ is convex in $\beta > 0$, leads to the conclusion that $\eta_3(\rho,a,b,\beta) \ge \eta_2(\rho,a,b,\beta)$ for $0 < \beta < \rho(a-b-bT)$. ☐

Lemma 18. *Let $X_1, \cdots, X_n$ be a sequence of i.i.d random variables. Define $\Gamma(t) = \log(\mathbb{E}[e^{tX}])$. Define $S = \sum_{i=1}^n X_i$, then for any $\epsilon > 0$ and $a \in \mathbb{R}$:*

$$\mathbb{P}\big(S \ge a - \epsilon\big) \ge e^{-\left(t^*a - n\Gamma(t^*) + |t^*|\epsilon\right)}\Big(1 - \frac{n\sigma^2_{\hat{X}}}{\epsilon^2}\Big) \qquad (80)$$

$$\mathbb{P}\big(S \le a + \epsilon\big) \ge e^{-\left(t^*a - n\Gamma(t^*) + |t^*|\epsilon\right)}\Big(1 - \frac{n\sigma^2_{\hat{X}}}{\epsilon^2}\Big) \qquad (81)$$

where $t^* = \arg\sup_{t\in\mathbb{R}} ta - \Gamma(t)$, $\hat{X}$ is a random variable with the same alphabet as X but distributed according to $\frac{e^{t^*x}\mathbb{P}(x)}{\mathbb{E}_X[e^{t^*x}]}$ and $\mu_{\hat{X}}, \sigma^2_{\hat{X}}$ are the mean and variance of $\hat{X}$, respectively.

Proof.

$$\begin{aligned}
&\mathbb{P}\big(S \ge a - \epsilon\big) \ge \mathbb{P}\big(a - \epsilon \le S \le a + \epsilon\big)\\
&= \int_{a-\epsilon\le S\le a+\epsilon} \mathbb{P}(x_1)\cdots\mathbb{P}(x_n)dx_1\cdots dx_n\\
&\overset{(a)}{\ge} e^{-(ta - n\Gamma(t)) - |t|\epsilon}\int_{a-\epsilon\le S\le a+\epsilon}\prod_{i=1}^n\Big(\frac{e^{tx_i}\mathbb{P}(x_i)}{\mathbb{E}_X[e^{tx}]}dx_i\Big)\\
&\overset{(b)}{=} e^{-(ta - n\Gamma(t)) - |t|\epsilon}\mathbb{P}_{\hat{X}_n}\Big(a - \epsilon \le S \le a + \epsilon\Big)\\
&\overset{(c)}{\ge} e^{-(ta - n\Gamma(t)) - |t|\epsilon}\Big(1 - \frac{n\sigma^2_{\hat{X}} + (n\mu_{\hat{X}} - a)^2}{\epsilon^2}\Big) \qquad (82)
\end{aligned}$$

where, for all finite $\mathbb{E}[e^{tX}]$, (a) is true because $e^{t\sum x_i} \le e^{n(ta + |t|\epsilon)}$ over the range of integration, (b) holds because $\frac{e^{tx}\mathbb{P}_X(x)}{\mathbb{E}_X[e^{tX}]}$ is a valid distribution [39], and (c) holds by Chebyshev inequality and by defining $\mu_{\hat{X}}, \sigma^2_{\hat{X}}$ to be the mean and variance of $\hat{X}$, respectively. Since $ta - n\Gamma(t)$ is concave in t, to find $t^* = \arg\sup_t(ta - n\Gamma(t))$ we set the derivative to zero, finding $a = n\frac{\mathbb{E}_X[xe^{t^*x}]}{\mathbb{E}[e^{t^*x}]}$. Also, by direct computation of $\mu_{\hat{X}}$, it can be shown that $\mu_{\hat{X}} = \frac{\mathbb{E}_X[xe^{tx}]}{\mathbb{E}[e^{tx}]}$. This means that at $t = t^*$, $n\mu_{\hat{X}} = a$. Thus, substituting back in (82) leads to:

$$\mathbb{P}\big(S \ge a - \epsilon\big) \ge e^{-(t^*a - n\Gamma(t^*)) - |t^*|\epsilon}\big(1 - \frac{n\sigma^2_{\hat{X}}}{\epsilon^2}\big)$$

This concludes the proof of (80). The proof of (81) follows similarly.

In our model $\epsilon = \log^{\frac{2}{3}}(n)$ and $n\sigma^2_{\hat{X}}$ is $O(\log(n))$, and hence,

$$\mathbb{P}\big(S \ge a - \epsilon\big) \ge e^{-(t^*a - n\Gamma(t^*)) - |t^*|\epsilon}\big(1 - o(1)\big)$$

which concludes the proof. ☐

B. Necessity of Theorem 1

Let $\boldsymbol{x}^*_{\backslash i,j}$ represent the vector $\boldsymbol{x}^*$ with two coordinates i,j removed. We wish to determine x_i^* via an observation of $\boldsymbol{G},\boldsymbol{Y}$, as well as a node index J and the expurgated vector of labels $\boldsymbol{x}^*_{\backslash i,J}$, where node J is randomly and uniformly chosen from inside (outside) the community if node i is outside (inside) the community, i.e., $\{j : x_j^* \ne x_i^*\}$. Then:

$$\begin{aligned}
&\frac{\mathbb{P}(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\backslash i,J}|x_i^* = 0)}{\mathbb{P}(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\backslash i,J}|x_i^* = 1)}\\
&= \frac{\mathbb{P}(\boldsymbol{G}|\boldsymbol{Y},J,\boldsymbol{x}^*_{\backslash i,J},x_i^* = 0)}{\mathbb{P}(\boldsymbol{G}|\boldsymbol{Y},J,\boldsymbol{x}^*_{\backslash i,J},x_i^* = 1)}\\
&\quad\times\frac{\mathbb{P}(\boldsymbol{x}^*_{\backslash i,J}|J,x_i^* = 0,\boldsymbol{Y})}{\mathbb{P}(\boldsymbol{x}^*_{\backslash i,J}|J,x_i^* = 1,\boldsymbol{Y})}\frac{\mathbb{P}(\boldsymbol{Y},J|x_i^* = 0)}{\mathbb{P}(\boldsymbol{Y},J|x_i^* = 1)}\\
&\overset{(a)}{=}\frac{\mathbb{P}(\boldsymbol{G}|J,\boldsymbol{x}^*_{\backslash i,J},x_i^* = 0)}{\mathbb{P}(\boldsymbol{G}|J,\boldsymbol{x}^*_{\backslash i,J},x_i^* = 1)}\frac{\mathbb{P}(y_{i,1}\cdots,y_{i,M}|x_i^* = 0)}{\mathbb{P}(y_{i,1}\cdots,y_{i,M}|x_i^* = 1)}\\
&\quad\times\frac{\mathbb{P}(y_{J,1},\cdots,y_{J,M}|J,x_i^* = 0)}{\mathbb{P}(y_{J,1},\cdots,y_{J,M}|J,x_i^* = 1)}
\end{aligned}$$

$$=\Big(\prod_{\substack{k\neq i,J\\ x_k^*=1}}\frac{Q(G_{ik})P(G_{Jk})}{P(G_{ik})Q(G_{Jk})}\Big)\Big(\prod_{m=1}^{M}\frac{U(y_{i,m})V(y_{J,m})}{V(y_{i,m})U(y_{J,m})}\Big) \quad (83)$$

where (a) holds because $\boldsymbol{G}$ and $\boldsymbol{Y}$ are independent given the labels, $\mathbb{P}(J|x_i^*=0)=\mathbb{P}(J|x_i^*=1)$ and $\mathbb{P}(\boldsymbol{x}^*_{\setminus i,J}|J,x_i^*=0,\boldsymbol{Y})=\mathbb{P}(\boldsymbol{x}^*_{\setminus i,J}|J,x_i^*=1,\boldsymbol{Y})$.

Denote the set of nodes inside the community, excluding i,J, with $\mathcal{K}=\{k\neq i,J\ :\ x_k^*=1\}$, and construct a vector from four sets of random variables as follows:

$$T\triangleq\Big[\{y_{i,m}\}_{m=1}^M,\{y_{J,m}\}_{m=1}^M,\{G_{ik}\}_{k\in\mathcal{K}},\{G_{Jk}\}_{k\in\mathcal{K}}\Big].$$

where the members of each set appear in the vector in increasing order of their varying index. From (83), T is a sufficient statistic of $(G,\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})$ for testing $x_i^*\in\{0,1\}$. Moreover, conditioned on $x_i^*=0$, T is distributed according to $U^{\otimes M}V^{\otimes M}Q^{\otimes(K-1)}P^{\otimes(K-1)}$ and conditioned on $x_i^*=1$, T is distributed according to $V^{\otimes M}U^{\otimes M}P^{\otimes(K-1)}Q^{\otimes(K-1)}$. Then, for any estimator $\hat{\boldsymbol{x}}(\boldsymbol{G},\boldsymbol{Y})$ achieving weak recovery:

$$\begin{aligned}
&\mathbb{E}[d(\hat{\boldsymbol{x}},\boldsymbol{x}^*)]\\
&=\sum_{i=1}^n\mathbb{P}(x_i^*\neq\hat{x}_i)\\
&\geq\sum_{i=1}^n\min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y})}\mathbb{P}(x_i^*\neq\tilde{x}_i)\\
&\geq\sum_{i=1}^n\min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})}\mathbb{P}(x_i^*\neq\tilde{x}_i)\\
&=n\min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})}\mathbb{P}(x_i^*\neq\tilde{x}_i)\\
&=n\min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})}\Big(\frac{K}{n}\mathbb{P}(x_i^*\neq\tilde{x}_i|x_i^*=1)\\
&\qquad\qquad+\frac{n-K}{n}\mathbb{P}(x_i^*\neq\tilde{x}_i|x_i^*=0)\Big)\\
&\geq n\min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})}\Big(\frac{K}{n}\mathbb{P}(x_i^*\neq\tilde{x}_i|x_i^*=1)\\
&\qquad\qquad+\frac{K}{n}\mathbb{P}(x_i^*\neq\tilde{x}_i|x_i^*=0)\Big)\\
&=K\min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})}\big(\mathbb{P}(x_i^*\neq\tilde{x}_i|x_i^*=1)\\
&\qquad\qquad+\mathbb{P}(x_i^*\neq\tilde{x}_i|x_i^*=0)\big)
\end{aligned}\quad (84)$$

Since by assumption, $\mathbb{E}[d(\hat{\boldsymbol{x}},\boldsymbol{x}^*)]=o(K)$, then by (84), the sum of Type-I and II probabilities of error is $o(1)$, which implies that as $n\to\infty$ [40]:

$$\begin{aligned}
&TV\Big(U^{\otimes M}V^{\otimes M}Q^{\otimes(K-1)}P^{\otimes(K-1)},\\
&\qquad V^{\otimes M}U^{\otimes M}P^{\otimes(K-1)}Q^{\otimes(K-1)}\Big)\to 1
\end{aligned}\quad (85)$$

where $TV(\cdot,\cdot)$ is the total variational distance between probability distributions. By properties of the total variational distance and KL divergence [40], for any two distributions $\tilde{P},\tilde{Q}$: $D(\tilde{P}||\tilde{Q})\geq\log(\frac{1}{2(1-TV(\tilde{P}||\tilde{Q}))})$. Hence, using (85):

$$\begin{aligned}
&D\Big(U^{\otimes M}V^{\otimes M}Q^{\otimes(K-1)}P^{\otimes(K-1)}\Big|\Big|\\
&\qquad V^{\otimes M}U^{\otimes M}P^{\otimes(K-1)}Q^{\otimes(K-1)}\Big)\\
&=M\Big(D(U||V)+D(V||U)\Big)\\
&\qquad+(K-1)\Big(D(P||Q)+D(Q||P)\Big)\to\infty
\end{aligned}\quad (86)$$

Since the LLRs are bounded by assumption, using Lemma 16 in Appendix A,

$$\begin{aligned}
&(K-1)D(P||Q)+MD(V||U)\\
&=E_{QU}\Big((K-1)D(P||Q)+MD(V||U),K-1,M\Big)\\
&\geq E_{QU}\Big(-\frac{(K-1)D(Q||P)+MD(U||V)}{2},K-1,M\Big)\\
&\geq C\Big((K-1)D(Q||P)+MD(U||V)\Big)
\end{aligned}\quad (87)$$

for some positive constant C. Substituting in (86) leads to:

$$MD(V||U)+(K-1)D(P||Q)\to\infty \quad (88)$$

which proves the first condition in (1).

$\boldsymbol{x}^*$ is drawn uniformly from the set $\{\boldsymbol{x}\in\{0,1\}^n:w(\boldsymbol{x})=K\}$ and $w(\boldsymbol{x})=\sum_{j=1}^n x_j$; therefore x_i's are individually Bernoulli-$\frac{K}{n}$. Then, for any estimator $\hat{\boldsymbol{x}}(\boldsymbol{G},\boldsymbol{Y})$ achieving weak recovery we have the following, where $H(\cdot)$ and $I(\cdot;\cdot)$ are the entropy and mutual information of their respective arguments.

$$\begin{aligned}
I(\boldsymbol{G},\boldsymbol{Y};\boldsymbol{x}^*)&\overset{(a)}{\geq}I(\hat{\boldsymbol{x}}(\boldsymbol{G},\boldsymbol{Y});\boldsymbol{x}^*)\\
&\overset{(b)}{\geq}\min_{\mathbb{E}[d(\tilde{\boldsymbol{x}},\boldsymbol{x}^*)]\leq\epsilon_nK}I(\tilde{\boldsymbol{x}}(\boldsymbol{G},\boldsymbol{Y});\boldsymbol{x}^*) \quad (89)\\
&\geq H(\boldsymbol{x}^*)-\max_{\mathbb{E}[d(\tilde{\boldsymbol{x}},\boldsymbol{x}^*)]\leq\epsilon_nK}H(d(\tilde{\boldsymbol{x}},\boldsymbol{x}^*))\\
&\overset{(c)}{=}\log\big(\binom{n}{K}\big)-nh(\frac{\epsilon_nK}{n})\\
&\overset{(d)}{\geq}K\log(\frac{n}{k})(1+o(1)) \quad (90)
\end{aligned}$$

where (a) is due to the data processing inequality [40], in (b) we defined $\epsilon_n=o(1)$, (c) is due to the fact that $\max_{\mathbb{E}(w(X))\leq pn}H(X)=nh(p)$ for any $p\leq\frac{1}{2}$ [24], where $h(p)\triangleq-p\log(p)-(1-p)\log(1-p)$, and (d) holds because $\binom{n}{K}\geq(\frac{n}{K})^K$, the assumption $K=o(n)$ and the bound $h(p)\leq-p\log(p)+p$ for $p\in[0,1]$. Denoting by $P(\boldsymbol{G},\boldsymbol{Y},\boldsymbol{x}^*)$ the joint distribution of the graph, side information, and node labels, and using [40]:

$$\begin{aligned}
&I(\boldsymbol{G},\boldsymbol{Y};\boldsymbol{x}^*)\\
&=\min_{\tilde{Q}}D\Big(\mathrm{P}(\boldsymbol{G},\boldsymbol{Y}|\boldsymbol{x}^*)\ ||\ \tilde{\mathrm{Q}}\ |\ \mathrm{P}(\boldsymbol{x}^*)\Big)\\
&\leq D\Big(\mathrm{P}(\boldsymbol{G}|\boldsymbol{x}^*)\prod_{m=1}^M(\mathrm{P}(\boldsymbol{y}_m|\boldsymbol{x}^*))\Big|\Big|Q^{\otimes\binom{n}{2}}\prod_{m=1}^M(U^{\otimes n})\ |\mathrm{P}(\boldsymbol{x}^*)\Big)\\
&=\binom{K}{2}D(P||Q)+KMD(V||U)
\end{aligned}\quad (91)$$

Combining (90) and (91):

$$\liminf_{n\to\infty}(K-1)D(P||Q)+2MD(V||U)\geq 2\log(\frac{n}{K}) \quad (92)$$

which proves the second condition in (1).

C. Sufficiency of Theorem 1

The sufficient conditions for weak recovery is derived for the maximum likelihood (ML) detector. Define:

$$e_1(S,T) \triangleq \sum_{i\in S}\sum_{j\in T} L_G(i,j) \tag{93}$$

$$e_2(S) \triangleq \sum_{i\in S}\sum_{m=1}^{M} L_S(i,m) \tag{94}$$

for any subsets $S,T \subset \{1,\cdots,n\}$. Using these definitions, the maximum likelihood detection can be characterized as follows:

$$\hat{C} = \hat{C}_{ML} = \underset{\substack{C\subset\{1,\cdots,n\}\\ |C|=K}}{\arg\max}\ \big(e_1(C,C)+e_2(C)\big) \tag{95}$$

Let $R \triangleq |\hat{C}\cap C^*|$, then $|\hat{C}\triangle C^*| = 2(K-R)$, and hence, to show that maximum likelihood achieves weak recovery, it is sufficient to show that there exists positive $\epsilon = o(1)$, such that $\mathbb{P}\big(R \le (1-\epsilon)K\big) = o(1)$.

To bound the error probability of ML, we characterize the separation of its likelihood from the likelihood of the community C^*.

$$\begin{aligned} &e_1(\hat{C},\hat{C}) + e_2(\hat{C}) - \big(e_1(C^*,C^*)+e_2(C^*)\big)\\ &= e_1(\hat{C}\backslash C^*,\hat{C}\backslash C^*) + e_1(\hat{C}\backslash C^*,\hat{C}\cap C^*) - e_1(C^*\backslash\hat{C},C^*) +\\ &\quad e_2(\hat{C}\backslash C^*) - e_2(C^*\backslash\hat{C}) \end{aligned} \tag{96}$$

By definition $|C^*\backslash\hat{C}| = |\hat{C}\backslash C^*| = K-R$. Thus, for any $0\le r\le K-1$,

$$\begin{aligned} &\mathbb{P}(R=r)\\ &\le \mathbb{P}\Big(\big\{\hat{C}: |\hat{C}|=K, |\hat{C}\cap C^*| = r,\\ &\qquad e_1(\hat{C},\hat{C}) + e_2(\hat{C}) - e_1(C^*,C^*) - e_2(C^*) \ge 0\big\}\Big)\\ &= \mathbb{P}\Big(\big\{S\subset C^*, T\subset (C^*)^c : |S|=|T|=K-r,\\ &\qquad e_1(S,C^*)+e_2(S) \le e_1(T,T) + e_1(T,C^*\backslash S) + e_2(T)\big\}\Big)\\ &\le \mathbb{P}\Big(\big\{S\subset C^* : |S| = K-r, e_1(S,C^*)+e_2(S)\le\theta\big\}\\ &\qquad \cup \big\{S\subset C^*, T\subset (C^*)^c : |S|=|T|=K-r,\\ &\qquad\qquad e_1(T,T)+e_1(T,C^*\backslash S)+e_2(T)\ge\theta\big\}\Big) \end{aligned} \tag{97}$$

where $\theta = (1-\eta)(aD(P||Q) + (K-r)MD(V||U))$, for some $\eta\in(0,1)$ and $a = \binom{K}{2} - \binom{r}{2}$. We further assume random variables $L_{G,i}$ are drawn i.i.d. according to the distribution of L_G, and $L_{S,m,j}$ are similarly i.i.d. copies of L_S. Then, using (97) and a union bound:

$$\begin{aligned} &\mathbb{P}(R=r)\\ &\le \binom{K}{K-r}\mathbb{P}\Big(\sum_{i=1}^{a} L_{G,i} + \sum_{j=1}^{K-r}\sum_{m=1}^{M} L_{S,m,j} \le \theta\Big)\\ &\quad + \binom{K}{K-r}\binom{n-K}{K-r}\mathbb{P}\Big(\sum_{i=1}^{a} L_{G,i} + \sum_{j=1}^{K-r}\sum_{m=1}^{M} L_{S,m,j} \ge \theta\Big)\\ &\overset{(a)}{\le} e^{(K-r)\log(\frac{Ke}{K-r})}\\ &\quad\times e^{-\sup_{t\ge0} -t\theta - a\log_P(\mathbb{E}[e^{-tL_G}]) - (K-r)M\log_V(\mathbb{E}[e^{-tL_S}])}\\ &\quad + e^{(K-r)\log(\frac{(n-K)Ke^2}{(K-r)^2})}\\ &\quad\times e^{-\sup_{t\ge0} t\theta - a\log_Q(\mathbb{E}[e^{tL_G}]) - (K-r)M\log_U(\mathbb{E}[e^{tL_S}])}\\ &\overset{(b)}{\le} e^{(K-r)\log(\frac{Ke}{K-r}) - E_{PV}(\theta,a,M(K-r))}\\ &\quad + e^{(K-r)\log(\frac{(n-K)Ke^2}{(K-r)^2}) - E_{QU}(\theta,a,M(K-r))}\\ &\overset{(c)}{=} e^{(K-r)\log(\frac{Ke}{K-r}) - E_{PV}(\theta,a,M(K-r))}\\ &\quad + e^{(K-r)\log(\frac{(n-K)Ke^2}{(K-r)^2}) - E_{PV}(\theta,a,M(K-r)) - \theta}\\ &\overset{(d)}{\le} e^{(K-r)\log(\frac{Ke}{K-r}) - E_{PV}(\theta,a,M(K-r))}\\ &\quad + e^{-(K-r)\big((1-\eta)((\frac{K-1}{2})D(P||Q)+MD(V||U)) - \log(\frac{n-K}{K})\big)}\\ &\quad\times e^{2(K-r)\log(\frac{e}{\epsilon}) - E_{PV}(\theta,a,M(K-r))}\\ &\overset{(e)}{\le} 2e^{2(K-r)\log(\frac{e}{\epsilon}) - E_{PV}(\theta,a,M(K-r))} \end{aligned} \tag{98}$$

where (a) holds by Chernoff bound and because $\binom{a}{b}\le(\frac{ea}{b})^b$, (b) holds from Lemma 15 in Appendix A, (c) holds because $E_{PV}(\theta,a,M(K-r)) = E_{QU}(\theta,a,M(K-r)) - \theta$, (d) holds because $a\ge\frac{(K-r)(K-1)}{2}$, $r\le(1-\epsilon)K$ and (e) holds by assuming that $\liminf_{n\to\infty}(K-1)D(P||Q) + 2MD(V||U) > 2\log(\frac{n}{K})$, which implies that

$$(1-\eta)((\frac{K-1}{2})D(P||Q) + MD(V||U)) - \log(\frac{n-K}{K}) \ge 0.$$

Lemma 15 in Appendix A shows that

$$E_{PV}(\theta,a,M(K-r)) \ge C(aD(P||Q) + (K-r)MD(V||U)]).$$

Using $a\ge\frac{(K-r)(K-1)}{2}$ and substituting in (98),

$$\begin{aligned} \mathbb{P}(R=r) &\le 2e^{-(K-r)\big(C(\frac{K-1}{2}D(P||Q)+MD(V||U)) - 2\log(\frac{e}{\epsilon})\big)}\\ &\le 2e^{-(K-r)\big(\frac{C}{2}((K-1)D(P||Q)+MD(V||U)) - 2\log(\frac{e}{\epsilon})\big)} \end{aligned} \tag{99}$$

Choose $\epsilon = \big((K-1)D(P||Q) + MD(V||U)\big)^{-\frac{1}{2}}$ and let $E = \big(\frac{C}{2}((K-1)D(P||Q)+MD(V||U)) - 2\log(\frac{e}{\epsilon})\big)$. Thus,

$$\begin{aligned} \mathbb{P}(R\le(1-\epsilon)K) &= \sum_{r=0}^{(1-\epsilon)K}\mathbb{P}(R=r) \le \sum_{r=0}^{(1-\epsilon)K} 2e^{-(K-r)E}\\ &\overset{(a)}{\le} 2\sum_{r'=\epsilon K}^{\infty} e^{-r'E} \le 2\frac{e^{-\epsilon KE}}{1-e^{-E}} \overset{(b)}{\le} o(1) \end{aligned} \tag{100}$$

where (a) holds by defining $r' = K-r$ and (b) holds by assuming that $(K-1)D(P||Q) + MD(V||U)\to\infty$ and by the choice of ϵ. This concludes the proof of Theorem 2.

D. Proof of Lemma 1

The proof has similarities with the sufficiency of Theorem 1. For brevity, we only provide a sketch. The complete proof is provided online [41].

Recall the definition of $\hat{C}$ from (95). Note that under the conditions of this Lemma, $\hat{C}$ may no longer be the maximum likelihood solution because $|C^*|$ need not be K. Let $|C^*| = K'$. Then, by assumption, with probability converging to one, $|K'-K| \le \frac{K}{\log(K)}$. Let $R = |\hat{C}\cap C^*|$. Thus, $|\hat{C}\triangle C^*| =$

$K+K'-2R$. Hence, it is sufficient to show that $\mathbb{P}(R \leq (1-\epsilon)K - |K'-K|) = o(1)$, where ϵ is defined in the statement of the Lemma. Let $a = \binom{K}{2} - \binom{r}{2}$ and $a' = \binom{K'}{2} - \binom{r}{2}$, then for any $r \leq (1-\epsilon)K - |K'-K|$ and by the choice of ϵ, the following holds as $n \to \infty$:

$$\frac{K}{K'} \to 1 \ , \ \frac{K-r}{K'-r} \to 1 \ , \ \frac{a}{a'} \to 1 \tag{101}$$

The proof then follows along the ideas of the proof of sufficiency of Theorem 2.

E. Proof of Lemma 2

Lemma 19. *Suppose that* (11) *holds. Let* $\{W_\ell\}$ *and* $\{\tilde{W}_\ell\}$ *denote sequences of i.i.d. copies of* L_G *under* P *and* Q*, respectively. Also, for any node* i*, let* Z *and* $\tilde{Z}$ *denote* $\sum_{m=1}^{M} L_S(i,m)$ *under* V *and* U*, respectively. Then, for sufficiently small, but constant,* δ *and* $\gamma = \frac{\log(\frac{n}{K})}{K}$:

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-\delta)} \tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big) = o(\frac{1}{n}) \tag{102}$$

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \leq K(1-\delta)\gamma\Big) = o(\frac{1}{K}) \tag{103}$$

Proof. By Chernoff bound:

$$\begin{aligned}&\mathbb{P}\Big(\sum_{\ell=1}^{K(1-\delta)} \tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big)\\ &\leq e^{-(1-\delta)\sup_{t\geq 0} tK\gamma - K\log(\mathbb{E}_Q[e^{tL_G}]) - \frac{M}{1-\delta}\log(\mathbb{E}_U[e^{tL_S}])}\end{aligned} \tag{104}$$

From (1) it follows that for some positive ϵ_o:

$$\begin{aligned}K\gamma &\leq \frac{KD(P||Q)}{2+\epsilon_o} + \frac{MD(V||U)}{1+\frac{\epsilon_o}{2}}\\ &\leq KD(P||Q) + MD(V||U)\\ &\leq KD(P||Q) + \frac{M}{1-\delta}D(V||U)\end{aligned} \tag{105}$$

Hence, using Lemma 15 in Appendix A, $\sup_{t\geq 0}$ is replaced by $\sup_{t\in[0,1]}$. Also, $\log(\mathbb{E}_U[e^{tL_S}]) = (t-1)D_t(V||U) \leq 0$ where the first equality holds by the definition of the Rényi-divergence between distributions V and U [40] and the second inequality because $t \in [0,1]$. This implies that $\frac{M}{1-\delta}\log(\mathbb{E}_U[e^{tL_S}]) \leq M\log(\mathbb{E}_U[e^{tL_S}])$. Substituting in (104):

$$\begin{aligned}\mathbb{P}\Big(\sum_{\ell=1}^{K(1-\delta)} \tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big) &\leq e^{-(1-\delta)E_{QU}(K\gamma,K,M)}\\ &\leq e^{-(1-\delta)(1+\epsilon)\log(n)}\end{aligned} \tag{106}$$

where (106) follows since (11) holds by assumption, i.e., there exists $\epsilon \in (0,1) : E_{QU}(K\gamma,K,M) \geq (1+\epsilon)\log(n)$. Equation (106) implies that (102) holds for sufficiently small δ.

To show (103), Chernoff bound is used:

$$\begin{aligned}&\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \leq K(1-\delta)\gamma\Big)\\ &\overset{(a)}{\leq} e^{tK\gamma(1-\delta)+K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])+K\delta\log(\mathbb{E}_Q[e^{-tL_G}])}\\ &\quad\times e^{M(1-\delta)\log(\mathbb{E}_V[e^{-tL_S}])+M\delta\log(\mathbb{E}_U[e^{-tL_S}])}\\ &= e^{(1-2\delta)(tK\gamma+K\log(\mathbb{E}_P[e^{-tL_G}])+M\frac{1-\delta}{1-2\delta}\log(\mathbb{E}_V[e^{-tL_S}]))}\\ &\quad\times e^{\delta(tK\gamma+K\log(\mathbb{E}_Q[e^{-tL_G}])+M\log(\mathbb{E}_U[e^{-tL_S}]))}\\ &\overset{(b)}{\leq} e^{(1-2\delta)(tK\gamma+K\log(\mathbb{E}_P[e^{-tL_G}])+M\log(\mathbb{E}_V[e^{-tL_S}]))}\\ &\quad\times e^{\delta(tK\gamma+K\log(\mathbb{E}_Q[e^{-tL_G}])+M\log(\mathbb{E}_U[e^{-tL_S}]))}\end{aligned} \tag{107}$$

where (a) and (b) hold because $\frac{1-\delta}{1-2\delta} \geq 1$ for sufficiently small δ and $\log(\mathbb{E}_V[e^{-tL_S}]) = (t-1)D_t(U||V) \leq tD_{t+1}(U||V) = \log(\mathbb{E}_U[e^{-tL_S}])$, where $D_t(V||U)$ is the Rényi-divergence between distributions V and U, which is non-decreasing in $t \geq 0$ [40].

By definition $-E_{PV}(K\gamma,K,M) = -\sup_{\lambda\in[-1,0]} \lambda K\gamma - K\log(\mathbb{E}_P[e^{\lambda L_G}]) - M\log(\mathbb{E}_V[e^{\lambda L_S}]) = -\lambda^* K\gamma + K\log(\mathbb{E}_P[e^{\lambda^* L_G}]) + M\log(\mathbb{E}_V[e^{\lambda^* L_S}])$. Hence, by choosing $t = -\lambda^* \in [0,1]$ and substituting in (107),

$$\begin{aligned}&\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \leq K(1-\delta)\gamma\Big)\\ &\leq e^{-(1-2\delta)E_{PV}(K\gamma,K,M)}\\ &\quad\times e^{\delta(K\gamma+K\log(\mathbb{E}_Q[e^{-tL_G}])+M\log(\mathbb{E}_U[e^{-tL_S}]))}\end{aligned} \tag{108}$$

By Lemma 16 and convexity of $\psi_{QU}(t,m_1,m_2)$:

$$\begin{aligned}\psi_{QU}(-t,K,M) &\leq \psi_{QU}(-1,K,M)\\ &\leq A(KD(Q||P) + MD(U||V))\end{aligned} \tag{109}$$

for some positive constant A. Moreover, by Lemma 16, $E_{QU}(K\gamma,K,M) \geq E_{QU}(0,K,M) \geq A_1(KD(Q||P) + MD(U||V))$, for some positive constant A_1. Hence, by substituting in (108), for some positive constant A_2:

$$\begin{aligned}&\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \leq K(1-\delta)\gamma\Big)\\ &\leq e^{-(1-2\delta)E_{PV}(K\gamma,K,M)+\delta(K\gamma+A_2E_{QU}(K\gamma,K,M))}\\ &\overset{(a)}{\leq} e^{-E_{QU}(K\gamma,K,M)(1-2\delta-\delta A_2)+(1-\delta)K\gamma}\\ &\overset{(b)}{=} e^{-\log(n)((1+\epsilon)(1-2\delta-\delta A_2)+\delta-1)-\log(K)(1-\delta)}\\ &\overset{(c)}{=} o(\frac{1}{K})\end{aligned} \tag{110}$$

where (a) holds because $E_{PV}(K\gamma,K,M) = E_{QU}(K\gamma,K,M) - K\gamma$ from Lemma 16, (b) holds by the assumption that (11) holds, which implies that there exists $\epsilon \in (0,1) : E_{QU}(K\gamma,K,M) \geq (1+\epsilon)\log(n)$ and (c) holds for sufficiently small δ.

Equations (106) and (110) concludes the proof of Lemma 19. □

Now we use Lemma 19 to prove Lemma 2. Define the event $E \triangleq \{(\hat{C}_k, C_k^*) : |\hat{C}_k \triangle C_k^*| \leq \delta K \quad \forall k\}$; then conditioned on E we have:

$$\begin{aligned}|\hat{C}_k \cap C_k^*| &\geq |\hat{C}_k| - |\hat{C}_k \triangle C_k^*| \\ &= \lceil K(1-\delta) \rceil - |\hat{C}_k \triangle C_k^*| \\ &\geq K(1-2\delta)\end{aligned}$$

Thus, in Algorithm I, for nodes i within the community C^*, r_i is stochastically greater than or equal to $(\sum_{\ell=1}^{K(1-2\delta)} W_\ell) + (\sum_{\ell=1}^{K\delta} \tilde{W}_\ell) + Z$ by Lemma 19 and (107). For $i \notin C^*$, r_i has the same distribution as $(\sum_{\ell=1}^{K(1-\delta)} \tilde{W}_\ell) + \tilde{Z}$. Thus, by Lemma 19, with probability converging to 1,

$$\begin{aligned}r_i &> K(1-\delta)\gamma, \qquad i \in C^* \\ r_i &< K(1-\delta)\gamma, \qquad i \notin C^*\end{aligned}$$

Hence, $\mathbb{P}(\tilde{C} = C^*) \to 1$ as $n \to \infty$.

F. Sufficiency of Theorem 2

The cardinality $|C_k^*|$ is a random variable that corresponds to sampling, without replacement, from the nodes of the original graph. Let Z be a binomial random variable $\text{Bin}(n(1-\delta), \frac{K}{n})$. The Chernoff bound for Z:

$$\mathbb{P}\left(\left|Z - (1-\delta)K\right| \geq \frac{K}{\log(K)}\right) \leq e^{-\Omega(\frac{K}{\log^2(K)})} \tag{111}$$

A result of Hoeffding [42, Theorem 4] for sampling with and without replacement indicates that $\mathbb{E}[\phi(|C_k^*|)] \leq \mathbb{E}[\phi(Z)]$ for any convex ϕ. This can be applied to (111) on the negative and positive side, individually. Putting them back together, we get a bound on the tails of $|C_k^*|$:

$$\begin{aligned}\mathbb{P}\left(\left||C_k^*| - (1-\delta)K\right| \geq \frac{K}{\log(K)}\right) &\leq e^{-\Omega(\frac{K}{\log^2(K)})} \\ &\leq o(1)\end{aligned} \tag{112}$$

Since (1) holds, for sufficiently small δ,

$$\liminf_{n\to\infty} \lceil (1-\delta)K \rceil D(P||Q) + 2MD(V||U) > 2\log(\frac{n}{K})$$

which together with (112) indicates, via Lemma 1, that ML achieves weak recovery. The idea of a two-step procedure for this proof has a precedent in [24].

Thus, for any $1 \leq k \leq \frac{1}{\delta}$:

$$\mathbb{P}\Big(\frac{|\hat{C}_k \triangle C_k^*|}{K} \leq 2\epsilon + \frac{1}{\log(K)}\Big) \geq 1 - o(1) \tag{113}$$

with $\epsilon = o(1)$. Since δ is constant, by the union bound

$$\mathbb{P}\Big(\frac{|\hat{C}_k \triangle C_k^*|}{K} \leq 2\epsilon + \frac{1}{\log(K)}, \quad \forall k\Big) \geq 1 - o(1) \tag{114}$$

Since $\epsilon = o(1)$, the desired (9) holds.

G. Necessity of Theorem 2

The following Lemma characterizes necessary conditions that are weaker than needed for Theorem 2, i.e., the Lemma is stronger than needed at this point, but will subsequently be used for unbounded LLR as well.

Lemma 20. *Let $\{W_\ell\}$ and $\{\tilde{W}_\ell\}$ denote sequences of i.i.d. copies of L_G under P and Q, respectively. For any node i inside the community, let Z denote a random variable drawn according to the distribution of $\sum_{m=1}^M L_S(i,m)$. Let $\tilde{Z}$ be the corresponding random variable when i is* outside *the community. Let $K_o \to \infty$ such that $K_o = o(K)$. Then, for any estimator $\hat{C}$ achieving exact recovery, there exists a sequence θ_n such that for sufficiently large n:*

$$\mathbb{P}\Big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\theta_n - \tilde{\theta}_n\Big) \leq \frac{2}{K_o} \tag{115}$$

$$\mathbb{P}\Big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\theta_n\Big) \leq \frac{1}{n-K} \tag{116}$$

where

$$\tilde{\theta}_n \triangleq (K_o - 1)D(P||Q) + 6\sqrt{K_o}\sigma \tag{117}$$

and σ^2 is the variance of L_G under P.

Proof. Recall that ML is optimal for exact recovery since C^* is chosen uniformly. Assume $\mathbb{P}(\text{ML fails}) = o(1)$. Define

$$\begin{aligned}i_o &\triangleq \arg\min_{i\in C^*} e_1(i, C^*) + \sum_{m=1}^M L_S(i,m) \\ \tilde{C} &\triangleq C^* \backslash \{i_o\} \cup \{j\} \text{ for } j \notin C^*\end{aligned} \tag{118}$$

Also, define the following event:

$$\begin{aligned}F_M \triangleq \Big\{(\boldsymbol{G}, \boldsymbol{Y}) : &\min_{i\in C^*} e_1(i, C^*) + \sum_{m=1}^M L_S(i,m) \\ &\leq \max_{j\notin C^*} e(j, C^*\backslash\{i_o\}) + \sum_{m=1}^M L_S(j,m)\Big\}\end{aligned} \tag{119}$$

Since $\mathbb{P}(\text{ML fails}) = o(1)$, using (95):

$$\begin{aligned}&e_1(\tilde{C}, \tilde{C}) + e_2(\tilde{C}) - e_1(C^*, C^*) - e_2(C^*) \\ &= \Big(e(j, C^*\backslash\{i_o\}) + \sum_{m=1}^M L_S(j,m)\Big) \\ &\quad - \Big(e_1(i, C^*) + \sum_{m=1}^M L_S(i,m)\Big)\end{aligned} \tag{120}$$

For observations belonging to F_M, the expression (120) is non-negative, implying ML fails with non-zero probability. Then,

$$\mathbb{P}(F_M) \leq \mathbb{P}(\text{ML fails}) = o(1) \tag{121}$$

since ML achieves exact recovery.

Define θ_n', θ_n'' and the events E_1 and E_2 as follows:

$$\theta_n' \triangleq \inf\Big\{x \in \mathbb{R} : \mathbb{P}\Big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)x - \tilde{\theta}_n\Big) \geq \frac{2}{K_o}\Big\} \tag{122}$$

$$\theta''_n \triangleq \sup\Big\{x \in \mathbb{R} : \mathbb{P}\Big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)x\Big) \geq \frac{1}{n-K}\Big\} \quad (123)$$

$$E_1 \triangleq \Big\{(\boldsymbol{G}, \boldsymbol{Y}) : \max_{j \notin C^*} \Big(e(j, C^* \backslash \{i_o\}) + \sum_{m=1}^{M} L_S(j,m)\Big) \geq (K-1)\theta''_n\Big\} \quad (124)$$

$$E_2 \triangleq \Big\{(\boldsymbol{G}, \boldsymbol{Y}) : \min_{i \in C^*} \Big(e_1(i, C^*) + \sum_{m=1}^{M} L_S(i,m)\Big) \leq (K-1)\theta'_n\Big\} \quad (125)$$

where $\tilde{\theta}_n$ is defined in (117).

By definition, E_1 and E_2 are independent. Since $\mathbb{P}(\text{ML fails}) = o(1)$ implies that $\mathbb{P}(F_M) = o(1)$, it follows that:

$$\begin{aligned}\mathbb{P}(E_1 \cap E_2 \cap F_M^c) &\geq \mathbb{P}(E_1 \cap E_2) - \mathbb{P}(F_M) \\ &= \mathbb{P}(E_1)\mathbb{P}(E_2) - o(1) \\ &= \Omega(1) \end{aligned} \quad (126)$$

where (126) holds since by Lemma 21, $\mathbb{P}(E_1) = \Omega(1)$ and $\mathbb{P}(E_2) = \Omega(1)$.

It is easy to see that $E_1 \cap E_2 \cap F_M^c \subset \{\theta'_n > \theta''_n\}$. It follows $\mathbb{P}(\theta'_n > \theta''_n) = \Omega(1)$ for sufficiently large n. Let $\theta_n = \frac{\theta'_n + \theta''_n}{2}$. For sufficiently large n, $\theta_n < \theta'_n$ and $\theta_n > \theta'_n$. Combining this with the definitions of θ'_n and θ''_n, implies that (115) and (116) hold simultaneously.

□

Lemma 21. *$\mathbb{P}(E_1) = \Omega(1)$ and $\mathbb{P}(E_2) = \Omega(1)$.*

Proof. Available online [41], but omitted here for brevity.

□

We now use Lemma 20 to prove the necessity of Theorem 2, which has similarities with [24], and hence, we provide only a sketch. The complete proof is provided online [41].

The proof sketch expresses the following: subject to conditions (1), exact recovery implies (11). Lemma 20 shows that exact recovery implies (115) and (116). It remains to be shown that (115) and (116) imply (11). We show that by contraposition.

Assume (11) does not hold, then for arbitrarily small $\epsilon > 0$ and sufficiently large n

$$E_{QU}\big(\log(\frac{n}{K}), K, M\big) \leq (1-\epsilon)\log(n) \quad (127)$$

Let

$$\gamma \triangleq \frac{\log(\frac{n}{K})}{K}$$

and define $S \triangleq \sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z}$ and $a \triangleq (K-1)\gamma + \delta$, for some $\delta > 0$. Since (1) holds, for sufficiently large n and arbitrary small $\epsilon_o > 0$:

$$\begin{aligned}K\gamma &\leq \frac{KD(P||Q)}{2+\epsilon_o} + \frac{MD(V||U)}{(1+\frac{\epsilon_o}{2})} \\ &\leq KD(P||Q) + MD(V||U) \end{aligned} \quad (128)$$

Following a variant of Lemma 18, it follows that at $\theta_n = \gamma$:

$$\begin{aligned}&\mathbb{P}\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma\big) \\ &= \int_{S \geq (K-1)\gamma} \mathbb{P}\big(\tilde{w}_1, \cdots, \tilde{w}_{K-1}, \tilde{z}\big) \\ &\overset{(a)}{\geq} e^{-\left(ta - \psi_{QU}(K-1,M,t)\right) - |t|\delta} \\ &\times \Bigg(1 - \frac{\big((K-1)\tilde{\sigma}^2_{L_G} + M\tilde{\sigma}^2_{L_S}\big) + \big((K-1)\tilde{\mu}_{L_G} + M\tilde{\mu}_{L_S} - a\big)^2}{\delta^2}\Bigg) \end{aligned} \quad (129)$$

where in (a): $t \in \mathbb{R}$, $\frac{\mathbb{P}(\tilde{W}_\ell)e^{t\tilde{W}_\ell}}{\mathbb{E}[e^{t\tilde{W}_\ell}]}$ and $\frac{\mathbb{P}(\tilde{Z})e^{t\tilde{Z}}}{\mathbb{E}[e^{t\tilde{Z}}]}$ define two new probability distributions $\tilde{Q}$ and $\tilde{U}$ over the same support of Q and U, respectively, and $\tilde{\sigma}^2_{L_G}$, $\tilde{\mu}_{L_G}$, $\tilde{\sigma}^2_{L_S}$ and $\tilde{\mu}_{L_S}$ are the variances and means of L_G and L_S under $\tilde{Q}$ and $\tilde{U}$, respectively.

Since $ta - \psi_{QU}(K-1,M,t)$ is concave in t, finding $t^* = \arg\sup_{t \in \mathbb{R}} ta - \psi_{QU}(K-1,M,t)$, and by the definition of $\tilde{Q}$ and $\tilde{U}$, it follows that:

$$\begin{aligned}(K-1)\tilde{\mu}_{L_G} + M\tilde{\mu}_{L_S} &= \frac{(K-1)\mathbb{E}_Q[L_G e^{tL_G}]}{\mathbb{E}_Q[e^{tL_G}]} + M\frac{\mathbb{E}_U[L_S e^{tL_S}]}{\mathbb{E}_U[e^{tL_S}]} \\ &= a. \end{aligned}$$

Thus, by substituting in (129):

$$\begin{aligned}&\mathbb{P}_{QU}\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma\big) \\ &\geq e^{-\left(t^*a - \psi_{QU}(K-1,M,t^*)\right) - |t^*|\delta}\big(1 - \frac{(K-1)\tilde{\sigma}^2_{L_G} + M\tilde{\sigma}^2_{L_S}}{\delta^2}\big) \end{aligned} \quad (130)$$

Choose $\delta = \big((K-1)D(P||Q) + MD(V||U)\big)^{\frac{2}{3}}$. Then, for sufficiently large n:

$$\mathbb{P}_{QU}\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z}\big) \geq e^{-E_{QU}(a,K,M) - \delta} \quad (131)$$

where (131) holds because for sufficiently large n:

$$\begin{aligned}a &= (K-1)\gamma + \delta \\ &\overset{(a)}{\leq} KD(P||Q) + MD(V||U)\big(\frac{1}{1+\frac{\epsilon_o}{2}} + o(1)\big) \end{aligned} \quad (132)$$

where (a) holds from (128).

Moreover,

$$E_{QU}(a,K,M) \leq E_{QU}(K\gamma, K, M) + \delta \quad (133)$$

where (133) holds because $t \in [0,1]$ and by (132).

Also, by Lemma 16, for some positive constant B:

$$E_{QU}(0, K-1, M) \geq B'((K-1)D(P||Q) + MD(V||U)) \quad (134)$$

Thus, for sufficiently large n, and for some positive constant B'':

$$\delta \leq (B'' E_{QU}(K\gamma, K, M))^{\frac{2}{3}} \tag{135}$$

Combining Equations (133), (134), (135):

$$\begin{aligned} & E_{QU}(a, K, M) + \delta \\ & \leq E_{QU}(K\gamma, K, M) + 2(B'' E_{QU}(K\gamma, K, M))^{\frac{2}{3}} \end{aligned} \tag{136}$$

Substituting in (131):

$$\begin{aligned} & \mathbb{P}_{QU}\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma\big) \\ & \qquad \geq e^{-(E_{QU}(K\gamma,K,M)+2(B''E_{QU}(K\gamma,K,M))^{\frac{2}{3}})} \\ & \qquad \overset{(a)}{\geq} e^{-((1-\epsilon)\log(n)+2(B''(1-\epsilon)\log(n))^{\frac{2}{3}})} \\ & \qquad \geq e^{-(1-\epsilon)\log(n)(1+o(1))} \end{aligned} \tag{137}$$

where (a) comes from the contraposition assumption that (11) does not hold, i.e., $E_{QU}(K\gamma, K, M) \leq (1-\epsilon)\log(n)$ for arbitrary small $\epsilon > 0$. Equation (137) shows that

$$n\mathbb{P}_{QU}(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma) \geq n^{\epsilon(1+o(1))}$$

which implies that (116) does not hold for $\theta_n = \gamma$.

Similarly, we will show that (115) does not hold for $\theta_n = \gamma$. Define

$$\begin{aligned} K_o &= \frac{K}{\log(K)} = o(K) \\ \delta' &= \frac{(K_o - 1)(D(P||Q) - \gamma) + 6\sqrt{K_o}\sigma}{(K - K_o)D(P||Q) + MD(V||U)} \end{aligned} \tag{138}$$

Note that $\delta' = o(1)$, which holds because $K\gamma \leq KD(P||Q) + MD(V||U)$, $K_o = o(K)$ and $K_o\sigma^2 = K_o \frac{d^2(\log(\mathbb{E}_Q[e^{tL_G}]))}{dt^2}|_{t=1} \leq BK_oD(P||Q)$ by Lemma 16 for some positive constant B. Let $a = (K-K_o)(\gamma - \delta' D(P||Q) - \frac{\delta'}{K-K_o}MD(V||U)) - \delta$, for some $\delta > 0$. Then, by a similar analysis as in (129):

$$\begin{aligned} & \mathbb{P}_{PV}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\gamma + \tilde{\theta}_n\big) \\ & \quad \geq e^{-(t^* a - \psi_{PV}(K-K_o, M, t^*)) - |t^*|\delta}(1 - o(1)) \end{aligned} \tag{139}$$

in which the inequality holds for $\delta = ((K-K_o)D(P||Q) + MD(V||U))^{\frac{2}{3}}$, $t^* = \arg\sup_{t\in\mathbb{R}} ta - \psi_{PV}(K-K_o, M, t)$.

Following similar analysis as in (131), and (128), it follows that:

$$\mathbb{P}_{PV}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\gamma + \tilde{\theta}_n\big) \geq e^{-\log(K)(1-\epsilon+o(1))} \tag{140}$$

for arbitrary small $\epsilon > 0$. Equation (140) shows:

$$K_o\mathbb{P}_{PV}(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\gamma + \tilde{\theta}_n) \geq K^{\epsilon(1+o(1))}$$

which implies that (115) does not hold for $\theta_n = \gamma$.

Thus, if (11) does not hold, both (137) and (140) show that (115) and (116) does not hold simultaneously at $\theta_n = \gamma$. Thus, for any $\theta_n > \gamma$, (115) will not hold and for any $\theta_n < \gamma$, (116) will not hold, and hence, if (11) does not hold, then there does not exist θ_n such that (115) and (116) hold simultaneously. This concludes the proof.

H. *Necessity of Theorem 3*

Recall that Definition 1 introduced Chernoff-information-type functions for the LLR of the graph plus side information; for convenience we now introduce a narrowed version of the same functions that focus on graph information only.

Definition 2.

$$\psi_Q(t, m_1) \triangleq m_1 \log(\mathbb{E}_Q[e^{tL_G}]) \tag{141}$$
$$\psi_P(t, m_1) \triangleq m_1 \log(\mathbb{E}_P[e^{tL_G}]) \tag{142}$$
$$E_Q(\theta, m_1) \triangleq \sup_{t\in[0,1]} t\theta - \psi_Q(t, m_1) \tag{143}$$
$$E_P(\theta, m_1) \triangleq \sup_{t\in[-1,0]} t\theta - \psi_P(t, m_1) \tag{144}$$

The quantities introduced in Definition 1 reduce to Definition 2 by setting $m_2 = 0$, therefore Lemma 15 continues to hold.

In view of Lemma 20, it suffices to test whether there exists θ_n such that both (115) and (116) hold. We will show that if one of the conditions (1)-(6) of Theorem 3 is not satisfied, then there does not exist θ_n such that (115) and (116) hold simultaneously. For brevity, we provide only a sketch of the proof. The complete proof is provided online [41].

Let $\theta_n = \gamma = \frac{\log(\frac{n}{K})}{K}$, and $a = (K-1)\gamma - \sum_{m=1}^{M} h_{\ell_m}^m + \delta$ for $\delta = \log(n)^{\frac{2}{3}}$.

$$\begin{aligned} & \mathbb{P}\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma\big) \\ & = \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_M=1}^{L_M} \Big[(\prod_{m=1}^{M} \alpha_{-,\ell_m}^m) \\ & \qquad \times \mathbb{P}_Q\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell \geq (K-1)\gamma - \sum_{m=1}^{M} h_{\ell_m}^m\big)\Big] \\ & \overset{(a)}{\geq} \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_M=1}^{L_M} \Big[(\prod_{m=1}^{M} \alpha_{-,\ell_m}^m) \\ & \qquad \times e^{-(t^* a - (K-1)\log(\mathbb{E}_Q[e^{t^* L_G}])) - |t^*|\delta}(1 - o(1))\Big] \end{aligned} \tag{145}$$

where (a) holds by Lemma 18, where $t^* = \arg\sup_{t\in\mathbb{R}}(ta - (K-1)\log(\mathbb{E}_Q[e^{tL_G}]))$[3].

Under (16):

$$\begin{aligned} KD(Q||P) &= \rho(a - b - bT)(1 + o(1))\log(n) \\ KD(P||Q) &= \rho(aT + b - a)(1 + o(1))\log(n) \end{aligned}$$

Thus, according to conditions of Theorem 3,

$$a \in \big[-KD(Q||P), KD(P||Q)\big].$$

[3]For ease of notation, we omit any subscript for both a and t^*. However, both depend on the outcomes of the features as shown in their definitions.

So, by Lemma 15,

$$t^* = \arg\sup_{t\in\mathbb{R}}(ta - (K-1)\log(\mathbb{E}_Q[e^{tL_G}])) = \arg\sup_{t\in[0,1]}(ta - (K-1)\log(\mathbb{E}_Q[e^{tL_G}]))$$

Without loss of generality, we focus on one term of the nested sum in (145). Assume $\sum_{m=1}^M h^m_{\ell_m} = o(\log(n))$ and both $\sum_{m=1}^M \log(\alpha^m_{+,\ell_m})$ and $\sum_{m=1}^M \log(\alpha^m_{-,\ell_m})$ are $o(\log(n))$, then by evaluating the supremum and by substituting in (145),

$$\mathbb{P}\big(\sum_{\ell=1}^{K-1}\tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma\big) \geq n^{-\eta_1(\rho,a,b)+o(1)} \quad (146)$$

Thus, if $\eta_1(\rho,a,b) \leq 1-\varepsilon$ for some $0<\varepsilon<1$, then $(n-K)\mathbb{P}(\sum_{\ell=1}^{K-1}\tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma) \geq n^{\varepsilon+o(1)}$ which shows that (116) does not hold for $\theta_n = \gamma$. This establishes the first case in Theorem 3. Other cases are derived in a somewhat similar manner, whose proof is omitted here for brevity but provided online [41].

Now we show that (115) does not hold for $\theta_n = \gamma$. Let $K_o = \frac{K}{\log(K)} = o(K)$. Also, let $a = (K-1)\gamma + \tilde{\theta}_n - \sum_{m=1}^M h^m_{\ell_m} - \delta$ for $\delta = \log(n)^{\frac{2}{3}}$. Then,

$$\begin{aligned}
&\mathbb{P}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\gamma + \tilde{\theta}_n\big)\\
&\overset{(a)}{\geq} \sum_{\ell_1=1}^{L_1}\cdots\sum_{\ell_L=1}^{L_M}\big(\prod_{m=1}^M \alpha^m_{+,\ell_m}\big)\\
&\quad\times e^{-(t^*a-(K-K_o)\log(\mathbb{E}_P[e^{t^*L_G}]))-|t^*|\delta}(1-o(1))\\
&\overset{(b)}{=} \sum_{\ell_1=1}^{L_1}\cdots\sum_{\ell_L=1}^{L_M}\big(\prod_{m=1}^M \alpha^m_{+,\ell_m}\big)\\
&\quad\times e^{-(\lambda^*a-(K-K_o)\log(\mathbb{E}_Q[e^{\lambda^*L_G}]))+a-|\lambda^*-1|\delta}(1-o(1))
\end{aligned} \quad (147)$$

where (a) holds by Lemma 18, where $t^* = \arg\sup_{t\in\mathbb{R}}(ta - (K-K_o)\log(\mathbb{E}_P[e^{tL_G}]))$ and (b) holds for $\lambda^* = 1+t^*$ and by Lemma 15.

Thus, according to conditions of Theorem 3,

$$a \in \big[-KD(Q||P), KD(P||Q)\big]. \quad (148)$$

Hence, by Lemma 15, $\arg\sup_{t\in\mathbb{R}}$ is replaced by $\arg\sup_{t\in[-1,0]}$.

Thus, if $\sum_{m=1}^M h^m_{\ell_m} = o(\log(n))$ and both $\sum_{m=1}^M \log(\alpha^m_{+,\ell_m})$ and $\sum_{m=1}^M \log(\alpha^m_{-,\ell_m})$ are $o(\log(n))$, then by evaluating the supremum and by substituting in (147),

$$\mathbb{P}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\gamma + \tilde{\theta}_n\big) \geq n^{-\eta_1(\rho,a,b)+o(1)}$$

Thus, if $\eta_1(\rho,a,b) \leq 1-\varepsilon$ for some $0<\varepsilon<1$, then $K\mathbb{P}(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\gamma + \tilde{\theta}_n) \geq n^{\varepsilon+o(1)}$ which shows that (115) does not hold for $\theta_n = \gamma$. This establishes the first case in Theorem 3. Other cases are derived in a somewhat similar manner, whose proof is omitted here for brevity but provided online [41].

To summarize, when $\theta_n = \gamma$, if one of the conditions (1)-(6) of Theorem 3 does not hold, then (115) and (116) cannot hold simultaneously. Thus, for any $\theta_n > \gamma$, (115) will not hold and for any $\theta_n < \gamma$, (116) will not hold, and hence, if one of the conditions (1)-(6) of Theorem 3 does not hold, then there does not exist θ_n such that (115) and (116) hold simultaneously. This concludes the proof of the necessary conditions.

I. Sufficiency of Theorem 3

The sufficient conditions are derived via Algorithm I provided in Section II-B with only one modification in the weak recovery step. Since the LLRs of the side information may not be bounded, the maximum likelihood detector with side information presented in Lemma 1 cannot be used for the weak recovery step. Instead the maximum likelihood detector without side information provided in [24] will be used.

The following lemma gives sufficient conditions for Algorithm I to achieve exact recovery.

Lemma 22. *Define $C^*_k = C^* \cap {S_k}^c$ and assume $\hat{C}_k$ achieves weak recovery, i.e.*

$$\mathbb{P}\big(|\hat{C}_k \triangle C^*_k| \leq \delta K \text{ for } 1 \leq k \leq \frac{1}{\delta}\big) \to 1 \quad (149)$$

Under conditions (16)*, if conditions* (1)-(6) *of Theorem 3 hold, then* $\mathbb{P}(\tilde{C} = C^*) \to 1$.

Proof. To prove Lemma 22, we follow essentially the same strategy used for Lemma 2 in Appendix E. Namely, we intend to show that the total LLR for nodes inside and outside the community are, asymptotically, stochastically dominated by a certain constant. Since the strategy is essentially similar to an earlier result, we only provide a sketch in this appendix.

Lemma 23. *In the regime* (16)*, suppose conditions* (1)-(6) *of Theorem 3 hold. Let $\{W_\ell\}$ and $\{\tilde{W}_\ell\}$ denote two sequences of i.i.d copies of L_G under P and Q, respectively. Also, let Z be a random variable whose distribution is identical to $\sum_{m=1}^M h_{i,m}$ conditioned on $i \in C^*$, and $\tilde{Z}$ drawn according to the same distribution conditioned on $i \notin C^*$. Then, for sufficiently small constant δ and $\gamma = \frac{\log(\frac{n}{K})}{K}$:*

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-\delta)}\tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big) = o(\frac{1}{n}) \quad (150)$$

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K}\tilde{W}_\ell + Z \leq K(1-\delta)\gamma\Big) = o(\frac{1}{K}) \quad (151)$$

Proof. Using the Chernoff bound:

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{\ell=1}^{K(1-\delta)}\tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big)\\
&\leq \mathbb{P}\Big(\sum_{\ell=1}^{K}\tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big)
\end{aligned}$$

$$\leq \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_M=1}^{L_M} \Big(\prod_{m=1}^{M} \alpha^m_{-,\ell_m} \Big) \times e^{-\sup_{t\geq 0} t(K(1-\delta)\gamma - \sum_{m=1}^M h^m_{\ell_m}) - K\log(\mathbb{E}_Q[e^{tL_G}])} \quad (152)$$

The terms inside the nested sum in (152) are upper bounded by:

- $n^{-\eta_1(\rho,a,b)+o(1)}$, if $\sum_{m=1}^M h^m_{\ell_m} = o(\log(n))$ and both $\sum_{m=1}^M \log(\alpha^m_{+,\ell_m})$ and $\sum_{m=1}^M \log(\alpha^m_{-,\ell_m})$ are $o(\log(n))$.
- $n^{-\eta_1(\rho,a,b)-\beta+o(1)}$, if $\sum_{m=1}^M h^m_{\ell_m} = o(\log(n))$ and $\sum_{m=1}^M \log(\alpha^m_{+,\ell_m}) = \sum_{m=1}^M \log(\alpha^m_{-,\ell_m}) = -\beta\log(n) + o(\log(n)), \beta > 0$.
- $n^{-\eta_2(\rho,a,b,\beta)+o(1)}$, if $\sum_{m=1}^M h^m_{\ell_m} = \beta\log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT), \sum_{m=1}^M \log(\alpha^m_{+,\ell_m}) = o(\log(n))$.
- $n^{-\eta_3(\rho,a,b,\beta)+o(1)}$, if $\sum_{m=1}^M h^m_{\ell_m} = -\beta\log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT), \sum_{m=1}^M \log(\alpha^m_{-,\ell_m}) = o(\log(n))$.
- $n^{-\eta_2(\rho,a,b,\beta)-\beta'+o(1)}$, if $\sum_{m=1}^M h^m_{\ell_m} = \beta\log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT), \sum_{m=1}^M \log(\alpha^m_{+,\ell_m}) = -\beta'\log(n) + o(\log(n))$.
- $n^{-\eta_3(\rho,a,b,\beta)-\beta'+o(1)}$, if $\sum_{m=1}^M h^m_{\ell_m} = -\beta\log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT), \sum_{m=1}^M \log(\alpha^m_{-,\ell_m}) = -\beta'\log(n) + o(\log(n))$.

Since M and L_m are independent of n and finite, it follows that if items (1)-(6) of Theorem 3 are satisfied, then Equation (150) holds.

To show (151), Chernoff bound is used.

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \leq K(1-\delta)\gamma \Big)$$
$$\leq \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_M=1}^{L_M} (\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) \times e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h^m_{\ell_m}) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} \times e^{K\delta\log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \quad (153)$$

Without loss of generality, we focus on one term inside the nested sum in(153):

If $\sum_{m=1}^M h^m_{\ell_m} = o(\log(n))$ and both $\sum_{m=1}^M \log(\alpha^m_{+,\ell_m})$ and $\sum_{m=1}^M \log(\alpha^m_{-,\ell_m})$ are $o(\log(n))$, then:

$$(\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h^m_{\ell_m}) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} \times e^{K\delta\log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta}$$
$$\leq (\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) e^{(1-2\delta)\left(t(k\gamma - \frac{\sum_{m=1}^M h^m_{\ell_m}}{1-2\delta}) + K\log(\mathbb{E}_P[e^{-tL_G}])\right)} \times e^{\delta\left(tK\gamma + K\log(\mathbb{E}_Q[e^{-tL_G}])\right)} \quad (154)$$

Since $\sum_{m=1}^M h^m_{\ell_m} = o(\log(n))$, it is easy to show that

$$K\gamma - \frac{\sum_{m=1}^M h^m_{\ell_m}}{1-2\delta} \in [-KD(Q||P)\ ,\ KD(P||Q)].$$

Define $\theta \triangleq K\gamma - \frac{\sum_{m=1}^M h^m_{\ell_m}}{1-2\delta}$ and choose $t^* \in [0,1]$, such that $t^*\theta + K\log(\mathbb{E}[e^{-t^*L_G}]) = -E_P(\theta, K)$. Substituting in (154):

$$(\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h^m_{\ell_m}) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} \times e^{K\delta\log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta}$$
$$\leq (\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) e^{-(1-2\delta)E_P(\theta,K) + \delta\left(t^*K\gamma + K\log(\mathbb{E}_Q[e^{-t^*L_G}])\right)}$$
$$\leq (\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) e^{-(1-2\delta)E_P(\theta,K) + \delta\left(K\gamma + K\log(\mathbb{E}_Q[e^{-t^*L_G}])\right)} \quad (155)$$

where the last inequality holds because $t^* \in [0,1]$. Also, by Lemma 16 and convexity of $\log(\mathbb{E}_Q[e^{-tL_G}])$, the following holds for some positive constant A:

$$K\log(\mathbb{E}_Q[e^{-t^*L_G}]) \leq K\log(\mathbb{E}_Q[e^{-L_G}]) \leq AKD(Q||P) \quad (156)$$

Moreover, by Lemma 16, $E_P[\theta, K] = E_Q[\theta, K] - \theta$ and $E_Q[\theta, K] \geq E_Q[0, K] \geq A_1 KD(Q||P)$. Combining the last observation with (156), for some positive constant A_2,

$$(\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h^m_{\ell_m}) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} \times e^{K\delta\log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta}$$
$$\leq (\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) e^{-(1-2\delta)(E_Q(\theta,K)-\theta) + \delta K\gamma + \delta A_2 E_Q(\theta,K)}$$
$$= (\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) e^{-E_Q(\theta,K)(1-2\delta-\delta A_2) + (1-2\delta)\theta + \delta K\gamma} \quad (157)$$

Since $\sum_{m=1}^M \log(\alpha^m_{+,\ell_m}) = o(\log(n))$, evaluating the supremum in $E_Q[\theta, K]$ and substituting in (157) leads to:

$$(\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h^m_{\ell_m}) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} \times e^{K\delta\log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta}$$
$$\leq e^{-\log(n)(1-2\delta-\delta A_2)(\eta_1 + o(1))}$$
$$\leq n^{-(1+\varepsilon)(1-2\delta-\delta A_2) + o(1))} \quad (158)$$

where (158) holds by assuming $\eta_1 \geq 1 + \varepsilon$ for some $\varepsilon > 0$. Multiplying (158) by K:

$$K(\prod_{m=1}^{M} \alpha^m_{+,\ell_m}) \times e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h^m_{\ell_m}) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} \times e^{K\delta\log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta}$$
$$\leq n^{1-(1+\varepsilon)(1-2\delta-\delta A_2) + o(1))} \quad (159)$$

Thus, for any $\varepsilon > 0$, there exists a sufficiently small δ such that $(1+\varepsilon)(1-2\delta-\delta A_2) > 1$. This concludes the proof of the first case of Theorem 3. This establishes the first case in Theorem 3. Other cases are derived in a somewhat similar manner, whose proof is omitted here for brevity but provided online [41]. □

In view of Lemma 23, the proof of Lemma 22 then follows similarly as the proof of Lemma 2. □

In view of Lemma 22, it suffices to show that there exists an estimator that achieves weak recovery for a random cluster size and satisfies (149). We use the estimator presented in [24, Lemma 4], where it was shown that the maximum likelihood estimator can achieve weak recovery for a random cluster size upon observing only the graph if:

$$KD(P||Q) \to \infty \tag{160}$$

$$\liminf_{n\to\infty}(K-1)D(P||Q) \geq 2\log(\frac{n}{K}) \tag{161}$$

$$\mathbb{P}\left(\left||C_k^*|-(1-\delta)K\right| \geq \frac{K}{\log(K)}\right) \leq o(1) \tag{162}$$

It is obvious that in the regime (16), both (160) and (161) are satisfied. Thus, it remains to show that (162) holds too. Let $\hat{C}_k$ be the ML estimator for C_k^* based on observing $\boldsymbol{G}_k$ defined in Algorithm I. The distribution of $|C_k^*|$ is obtained by sampling the indices of the original graph without replacement. Hence, for any convex function ϕ: $\mathbb{E}[\phi(|C_k^*|)] \leq \mathbb{E}[\phi(Z)]$, where Z is a binomial random variable $\text{Bin}(n(1-\delta), \frac{K}{n})$. Therefore, the Chernoff bound for Z also holds for $|C_k^*|$. Thus,

$$\mathbb{P}\left(\left||C_k^*|-(1-\delta)K\right| \geq \frac{K}{\log(K)}\right) \leq o(1) \tag{163}$$

Thus, (162) holds, which implies that ML achieves weak recovery with K replaced with $\lceil(1-\delta)K\rceil$ in [24, Lemma 4]. Thus, from [24, Lemma 4], for any $1 \leq k \leq \frac{1}{\delta}$:

$$\mathbb{P}\Big(\frac{|\hat{C}_k \triangle C_k^*|}{K} \leq 2\epsilon + \frac{1}{\log(K)}\Big) \geq 1-o(1) \tag{164}$$

with $\epsilon = o(1)$. Since δ is constant, by the union bound over all $1 \leq k \leq \frac{1}{\delta}$, we have:

$$\mathbb{P}\Big(\frac{|\hat{C}_k \triangle C_k^*|}{K} \leq 2\epsilon + \frac{1}{\log(K)} \quad \forall 1 \leq k \leq \frac{1}{\delta}\Big) \geq 1-o(1) \tag{165}$$

Since $\epsilon = o(1)$, the desired (149) holds.

J. Auxiliary Lemmas For Belief Propagation

Lemma 24. *Recall the definition of Γ_0^t from (31). For any measurable function $g(.)$:*

$$\mathbb{E}[g(\Gamma_0^t)|\tau_0=0] = \mathbb{E}[g(\Gamma_0^t)e^{-\Gamma_0^t}|\tau_0=1] \tag{166}$$

Proof. Let $Y = (T^t, \tilde{\boldsymbol{\tau}}^t)$ denote the observed tree and side information. Then,

$$\begin{aligned}\mathbb{E}[g(\Gamma_0^t)|\tau_0=0] &= \mathbb{E}_{Y|\tau_0=0}[g(\Gamma_0^t)] \\ &= \int_Y g(\Gamma_0^t)\frac{\mathbb{P}(Y|\tau_0=0)}{\mathbb{P}(Y|\tau_0=1)}\mathbb{P}(Y|\tau_0=1) \\ &= \int_Y g(\Gamma_0^t)e^{-\Gamma_0^t}\mathbb{P}(Y|\tau_0=1) \\ &= \mathbb{E}_{Y|\tau_0=1}[g(\Gamma_0^t)e^{-\Gamma_0^t}] \\ &= \mathbb{E}[g(\Gamma_0^t)e^{-\Gamma_0^t}|\tau_0=1] \end{aligned} \tag{167}$$

□

Lemma 25. *Let $b_t = \mathbb{E}[\frac{e^{Z_1^t+U_1}}{1+e^{Z_1^t+U_1-\nu}}]$ and $a_t = \mathbb{E}[e^{2(Z_0^t+U_0)}]$. Let $\Lambda = \mathbb{E}[e^{U_1}] = \mathbb{E}[e^{2U_0}]$. Then, for any $t \geq 0$*

$$a_{t+1} = \mathbb{E}[e^{Z_1^t+U_1}] = \Lambda e^{\lambda b_t} \tag{168}$$

$$\begin{aligned}\mathbb{E}[e^{3(Z_0^t+U_0)}] &= \mathbb{E}[e^{2(Z_1^t+U_1)}] \\ &= \mathbb{E}[e^{3U_0}]e^{3\lambda b_t + \frac{\lambda^2}{K(p-q)}\mathbb{E}[(\frac{e^{Z_1^t+U_1}}{1+e^{Z_1^t+U_1-\nu}})^2]}\end{aligned} \tag{169}$$

Proof. The first equality in (168) holds by Lemma 24 for $g(x) = e^{2x}$. Similarly, the first equality in (169) holds by Lemma 24 for $g(x) = e^{3x}$.

Let $f(x) = \frac{1+\frac{p}{q}x}{1+x} = 1 + \frac{\frac{p}{q}-1}{1+x^{-1}}$. Then:

$$\begin{aligned}a_{t+1} &= \mathbb{E}[e^{2(Z_0^t+U_0)}] \\ &\overset{(a)}{=} e^{-2K(p-q)}\mathbb{E}[e^{2U_0}]\mathbb{E}[(\mathbb{E}[f^2(e^{Z_1^t+U_1-\nu})])^{H_u}] \\ &\qquad \times \mathbb{E}[(\mathbb{E}[f^2(e^{Z_0^t+U_0-\nu})])^{F_u}] \\ &\overset{(b)}{=} \Lambda e^{-2K(p-q)}e^{Kq(\mathbb{E}[f^2(e^{Z_1^t+U_1-\nu})]-1)} \\ &\qquad \times e^{(n-K)q(\mathbb{E}[f^2(e^{Z_0^t+U_0-\nu})]-1)}\end{aligned} \tag{170}$$

where (a) holds by the definition of Z_0^t and U_0, (b) holds by the definition of Λ and by using the fact that $\mathbb{E}[c^X] = e^{\lambda(c-1)}$ for $X \sim \text{Poi}(\lambda)$ and $c > 0$. By the definition of $f(x)$:

$$\begin{aligned}&Kq\big(\mathbb{E}\big[f^2(e^{Z_1^t+U_1-\nu})\big]-1\big) + (n-K)q\big(\mathbb{E}\big[f^2(e^{Z_0^t+U_0-\nu})\big]-1\big) \\ &= Kq\mathbb{E}\left[\frac{2(\frac{p}{q}-1)}{1+e^{-(Z_1^t+U_1-\nu)}} + \frac{(\frac{p}{q}-1)^2}{(1+e^{-(Z_1^t+U_1-\nu)})^2}\right] \\ &\quad + (n-K)q\mathbb{E}\left[\frac{2(\frac{p}{q}-1)}{1+e^{-(Z_0^t+U_0-\nu)}} + \frac{(\frac{p}{q}-1)^2}{(1+e^{-(Z_0^t+U_0-\nu)})^2}\right] \\ &\overset{(a)}{=} 2K(p-q) + Kq(\frac{p}{q}-1)^2\mathbb{E}\left[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\right] \\ &\overset{(b)}{=} 2K(p-q) + \lambda b_t\end{aligned} \tag{171}$$

where (a) holds by Lemma 24 and (b) holds by the definition of λ and b_t.

Using (171) and substituting in (170) concludes the proof of (168). The proof of (169) follows similarly using $f^3(x)$ instead of $f^2(x)$. □

K. Proof of Lemma 3

The independent splitting property of the Poisson distribution is used to give an equivalent description of the numbers of children having a given label for any vertex in the tree. An equivalent description of the generation of the tree is as follows: for each node i, generate a set $\mathcal{N}_i$ of children with $N_i = |\mathcal{N}_i|$. If $\tau_i = 1$, we generate $N_i \sim \text{Poi}(Kp+(n-K)q)$ children. Then for each child j, independent from everything else, let $\tau_j = 1$ with probability $\frac{Kp}{Kp+(n-K)q}$ and $\tau_j = 0$ with probability $\frac{(n-K)q}{Kp+(n-K)q}$. If $\tau_i = 0$ generate $N_i \sim \text{Poi}(nq)$, then for each child j, independent from everything else, let $\tau_j = 1$ with probability $\frac{K}{n}$ and $\tau_j = 0$ with probability $\frac{(n-K)}{n}$. Finally, for each node i in the tree, $\tilde{\tau}_i$ is observed according to $\alpha_{+,\ell}, \alpha_{-,\ell}$. Then:

$$\Gamma_0^{t+1} = \log\left(\frac{\mathbb{P}(T^{t+1}, \tilde{\tau}^{t+1}|\tau_0=1)}{\mathbb{P}(T^{t+1}, \tilde{\tau}^{t+1}|\tau_0=0)}\right)$$

$$= \log\left(\frac{\mathbb{P}(N_0, \tilde{\tau}_0, \{T_k^t\}_{k\in\mathcal{N}_0}, \{\tilde{\tau}_k^t\}_{k\in\mathcal{N}_0}|\tau_0 = 1)}{\mathbb{P}(N_0, \tilde{\tau}_0, \{T_k^t\}_{k\in\mathcal{N}_0}, \{\tilde{\tau}_k^t\}_{k\in\mathcal{N}_0}|\tau_0 = 0)}\right)$$
$$\stackrel{(a)}{=} \log\left(\frac{\mathbb{P}(N_0, \tilde{\tau}_0|\tau_0 = 1)}{\mathbb{P}(N_0, \tilde{\tau}_0|\tau_0 = 0)}\right) + \log\left(\frac{\prod_{k\in\mathcal{N}_0}\mathbb{P}(T_k^t, \tilde{\tau}_k^t|\tau_0 = 1)}{\prod_{k\in\mathcal{N}_0}\mathbb{P}(T_k^t, \tilde{\tau}_k^t|\tau_0 = 0)}\right)$$
$$\stackrel{(b)}{=} \log\left(\frac{\mathbb{P}(N_0|\tau_0 = 1)}{\mathbb{P}(N_0|\tau_0 = 0)}\right) + \log\left(\frac{\mathbb{P}(\tilde{\tau}_0|\tau_0 = 1)}{\mathbb{P}(\tilde{\tau}_0|\tau_0 = 0)}\right) + \sum_{k\in\mathcal{N}_0}\log\left(\frac{\sum_{\tau_k\in\{0,1\}}\mathbb{P}(T_k^t, \tilde{\tau}_k^t|\tau_k)\mathbb{P}(\tau_k|\tau_0 = 1)}{\sum_{\tau_k\in\{0,1\}}\mathbb{P}(T_k^t, \tilde{\tau}_k^t|\tau_k)\mathbb{P}(\tau_k|\tau_0 = 0)}\right)$$
$$\stackrel{(c)}{=} -K(p-q) + h_0 + \sum_{k\in\mathcal{N}_0}\log\left(\frac{\frac{p}{q}e^{\Gamma_k^t-\nu}+1}{e^{\Gamma_k^t-\nu}+1}\right) \tag{172}$$

where (a) holds because conditioned on τ_0: 1) $(N_0, \tilde{\tau}_0)$ are independent of the rest of the tree and 2) $(T_k^t, \tilde{\tau}_k^t)$ are independent random variables $\forall k \in \mathcal{N}_0$, (b) holds because conditioned on τ_0, N_0 and $\tilde{\tau}_0$ are independent, (c) holds by the definition of N_0 and h_0 and because τ_k is Bernoulli-$\frac{Kp}{Kp+(n-K)q}$ if $\tau_0 = 1$ and is Bernoulli-$\frac{K}{n}$ if $\tau_0 = 0$.

L. Proof of Lemma 4

Let $f(x) \triangleq \frac{1+\frac{p}{q}x}{1+x}$, then:

$$\mathbb{E}\left[e^{\frac{Z_0^t}{2}}\right] = e^{\frac{-K(p-q)}{2}}\mathbb{E}_{H_0}\left[(\mathbb{E}_{Z_1U_1}[f^{\frac{1}{2}}(e^{Z_1^t+U_1-\nu})])^{H_0}\right] \times \mathbb{E}_{F_0}\left[(\mathbb{E}_{Z_0U_0}[f^{\frac{1}{2}}(e^{Z_0^t+U_0-\nu})])^{F_0}\right]$$
$$\stackrel{(a)}{=} e^{\frac{-K(p-q)}{2}}e^{Kq(\mathbb{E}[f^{\frac{1}{2}}(e^{Z_1^t+U_1-\nu})]-1)} \times e^{(n-K)q(\mathbb{E}[f^{\frac{1}{2}}(e^{Z_0^t+U_0-\nu})]-1)} \tag{173}$$

where (a) holds using $\mathbb{E}[c^X] = e^{\lambda(c-1)}$ for $X \sim \text{Poi}(\lambda)$ and $c > 0$.

By the intermediate value form of Taylor's theorem, for any $x \geq 0$ there exists y with $1 \leq y \leq x$ such that $\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8(1+y)^{1.5}}$. Therefore,

$$\sqrt{1+x} \leq 1 + \frac{x}{2} - \frac{x^2}{8(1+A)^{1.5}}, \qquad 0 \leq x \leq A \tag{174}$$

Let $A = \frac{p}{q} - 1$ and $B = (1+A)^{1.5}$. By assumption, B is bounded. Then,

$$\left(\frac{1+\frac{p}{q}e^{Z_0^t+U_0-\nu}}{1+e^{Z_0^t+U_0-\nu}}\right)^{\frac{1}{2}} = \left(1 + \frac{\frac{p}{q}-1}{1+e^{-(Z_0^t+U_0-\nu)}}\right)^{\frac{1}{2}}$$
$$\leq 1 + \frac{1}{2}\frac{\frac{p}{q}-1}{1+e^{-(Z_0^t+U_0-\nu)}} - \frac{1}{8B}\frac{(\frac{p}{q}-1)^2}{(1+e^{-(Z_0^t+U_0-\nu)})^2} \tag{175}$$

It follows that:

$$Kq\left(\mathbb{E}[f^{\frac{1}{2}}(e^{Z_1^{t}+U_1-\nu})] - 1\right) + (n-K)q\left(\mathbb{E}[f^{\frac{1}{2}}(e^{Z_0^t+U_0-\nu})] - 1\right)$$
$$\leq \frac{Kq(\frac{p}{q}-1)}{2}\left(\mathbb{E}\left[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\right] + e^{\nu}\mathbb{E}\left[\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\right]\right) - \frac{Kq(\frac{p}{q}-1)^2}{8B}\left(\mathbb{E}\left[\frac{1}{(1+e^{-(Z_1^t+U_1-\nu)})^2}\right] + e^{\nu}\mathbb{E}\left[\frac{1}{(1+e^{-(Z_0^t+U_0-\nu)})^2}\right]\right)$$
$$\stackrel{(a)}{=} \frac{K(p-q)}{2} - \frac{K(p-q)^2}{8Bq}\mathbb{E}\left[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\right] \tag{176}$$
$$= \frac{K(p-q)}{2} - \frac{\lambda}{8B}b_t \tag{177}$$

where (a) holds by the following consequence of Lemma 24 (from Appendix J):

$$\mathbb{E}\left[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\right] + e^{\nu}\mathbb{E}\left[\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\right] = 1$$
$$\mathbb{E}\left[\frac{1}{(1+e^{-(Z_1^t+U_1-\nu)})^2}\right] + e^{\nu}\mathbb{E}\left[\frac{1}{(1+e^{-(Z_0^t+U_0-\nu)})^2}\right] = \mathbb{E}\left[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\right] \tag{178}$$

Using (173) and (177):

$$\mathbb{E}\left[e^{\frac{Z_0^t+U_0}{2}}\right] \leq \mathbb{E}\left[e^{\frac{U_0}{2}}\right]e^{\frac{-\lambda}{8B}b_t} \tag{179}$$

Similarly, using the fact that $\sqrt{1+x} \geq 1 + \frac{x}{2} - \frac{x^2}{8}$ for all $x \geq 0$:

$$\mathbb{E}\left[e^{\frac{Z_0^t+U_0}{2}}\right] \geq \mathbb{E}\left[e^{\frac{U_0}{2}}\right]e^{\frac{-\lambda}{8}b_t} \tag{180}$$

M. Proof of Lemma 5

Fix $\lambda > 0$ and define $(v_t : t \geq 0)$ recursively by $v_0 = 0$ and $v_{t+1} = \lambda\Lambda e^{v_t}$. From Lemma 25 in Appendix J, $a_{t+1} = \Lambda e^{\lambda b_t}$.

We first prove by induction that $\lambda b_t \leq \lambda a_t \leq v_{t+1}$ for all $t \geq 0$. $a_0 = \mathbb{E}[e^{U_1}] = \Lambda$ and $\lambda b_0 = \lambda\mathbb{E}[\frac{e^{U_1}}{1+e^{U_1-\nu}}] \leq \lambda\mathbb{E}[e^{U_1}] = \lambda a_0$. Thus, $\lambda b_0 \leq \lambda a_0 = \lambda\Lambda = v_1$. Assume that $\lambda b_{t-1} \leq \lambda a_{t-1} \leq v_t$. Then, $\lambda b_t \leq \lambda a_t = \lambda\Lambda e^{\lambda b_{t-1}} \leq \lambda\Lambda e^{v_t} = v_{t+1}$, where the first inequality holds by the definition of a_t and b_t and the second inequality holds by the induction assumption. Thus, $\lambda b_t \leq \lambda a_t \leq v_{t+1}$ for all $t \geq 0$.

Next we prove by induction that $\frac{v_t}{\lambda}$ is increasing in $t \geq 0$. We have $\frac{v_{t+1}}{\lambda} = \Lambda e^{v_t}$. Then, $\frac{v_1}{\lambda} = \Lambda \geq 0 = \frac{v_o}{\lambda}$. Now assume that $\frac{v_t}{\lambda} > \frac{v_{t-1}}{\lambda}$. Then, $\frac{v_{t+1}}{\lambda} = \Lambda e^{v_t} = \Lambda e^{\lambda(\frac{v_t}{\lambda})} > \Lambda e^{v_{t-1}} = \frac{v_t}{\lambda}$. Thus, we have: $\frac{v_{t+1}}{\lambda} > \frac{v_t}{\lambda}$ for all $t \geq 0$.

Note that $\frac{v_{t+1}}{\lambda} = \Lambda e^{\lambda(\frac{v_t}{\lambda})}$ has the form of $x = \Lambda e^{\lambda x}$, which has no solutions for $\lambda > \frac{1}{\Lambda e}$ and has two solutions for $\lambda \leq \frac{1}{\Lambda e}$, where the largest solution is Λe. Thus, for $\lambda \leq \frac{1}{\Lambda e}$, $b_t \leq \frac{v_{t+1}}{\lambda} \leq \Lambda e$.

N. Proof of Lemma 6

By definition of a_t, we have:

$$a_{t+1} - \mathbb{E}\left[e^{-\nu+2(Z_1^{t+1}+U_1)}\right] = \mathbb{E}\left[e^{Z_1^{t+1}+U_1}(1-e^{Z_1^{t+1}+U_1-\nu})\right] \leq \mathbb{E}\left[\frac{e^{Z_1^t+U_1}}{1+e^{Z_1^t+U_1-\nu}}\right]$$

$$= b_{t+1}$$

where the first inequality holds because $1-x \leq \frac{1}{1+x}$. Then,

$$\begin{aligned}
b_{t+1} &\geq a_{t+1} - \mathbb{E}[e^{-\nu+2(Z_1^{t+1}+U_1)}] \\
&\overset{(a)}{=} \Lambda e^{\lambda b_t} - e^{-\nu}\Lambda' e^{3\lambda b_t + \frac{\lambda^2}{K(p-q)}\mathbb{E}\left[(\frac{e^{Z_1^t+U_1}}{1+e^{Z_1^t+U_1-\nu}})^2\right]} \\
&\overset{(b)}{\geq} \Lambda e^{\lambda b_t} - \Lambda' e^{Cb_t-\nu} \\
&= \Lambda e^{\lambda b_t}\big(1 - \frac{\Lambda'}{\Lambda} e^{-\nu+(C-\lambda)b_t}\big) \\
&\overset{(c)}{\geq} \Lambda e^{\lambda b_t}\big(1 - \frac{\Lambda'}{\Lambda} e^{\frac{-\nu}{2}}\big) \qquad (181)
\end{aligned}$$

where (a) holds from Lemma 25, (b) holds because $(\frac{e^x}{1+e^{x-\nu}})^2 \leq e^\nu(\frac{e^x}{1+e^{x-\nu}})$, which holds because $e^\nu \geq \frac{e^x}{1+e^{x-\nu}}$ for all x, and (c) holds by the assumption that $b_t \leq \frac{\nu}{2(C-\lambda)}$.

O. Proof of Lemma 8

Given λ with $\lambda > \frac{1}{\Lambda e}$, assume $\nu \geq \nu_o$ and $\nu \geq 2\Lambda(C-\lambda)$ for some positive ν_o. Moreover, select the following constants depending only on λ and the LLR of side information:

- D and ν_o large enough such that $\lambda\Lambda e^{\lambda D}(1 - \frac{\Lambda'}{\Lambda}e^{-\nu_o}) > 1$ and $\Lambda\lambda e(1 - \frac{\Lambda'}{\Lambda}e^{-\nu_o}) \geq \sqrt{\lambda\Lambda e}$.
- $w_o > 0$ so large that

$$w_o\lambda\Lambda e^{\lambda D}(1 - \frac{\Lambda'}{\Lambda}e^{-\nu_o}) - \lambda D \geq w_o. \qquad (182)$$

- A positive integer $\bar{t}_o$ large enough such that $\lambda(\Lambda(\lambda\Lambda e)^{\frac{\bar{t}_o}{2}-1} - D) \geq w_o$

The goal is to show that there exists some $\tilde{t}$ after which $\nu = o(b_t)$.

Let $t^* = \max\{t > 0 : b_t < \frac{\nu}{2(C-\lambda)}\}$ and $\bar{t}_1 = \log^*(\nu)$. The first step is to show that $t^* \leq \bar{t}_o + \bar{t}_1$.

By the definition of b_t,

$$\begin{aligned}
b_0 &= \mathbb{E}\big[\frac{e^{U_1}}{1+e^{U_1-\nu}}\big] \\
&< \mathbb{E}[e^{U_1}] = \Lambda
\end{aligned}$$

Since $\nu \geq 2\Lambda(C-\lambda)$, we get $b_0 < \frac{\nu}{2(C-\lambda)}$.

Since for all $t \leq t^*$, $b_t < \frac{\nu}{2(C-\lambda)}$, then by Lemma 6:

$$\begin{aligned}
b_{t+1} &\geq \Lambda e^{\lambda b_t}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu}{2}}) \\
&\geq \Lambda e^{\lambda b_t}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu_o}{2}}) \qquad (183)
\end{aligned}$$

where the last inequality holds since $\nu \geq \nu_o$. Thus,

$$\begin{aligned}
b_1 &\geq \Lambda e^{\lambda b_0}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu_o}{2}}) \\
&\geq \Lambda(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu_o}{2}}) \\
&\geq \sqrt{\frac{\Lambda}{\lambda e}} \qquad (184)
\end{aligned}$$

where the last inequality holds by the choice of ν_o. Moreover,

$$\begin{aligned}
b_{t+1} &\geq \Lambda e^{\lambda b_t}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu_o}{2}}) \\
&\overset{(a)}{\geq} \Lambda e\lambda b_t(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu_o}{2}}) \\
&\overset{(b)}{\geq} \sqrt{\Lambda\lambda e}\, b_t \qquad (185)
\end{aligned}$$

where (a) holds because $e^u \geq eu$ for all $u > 0$ and (b) holds by choice of ν_0. Thus, for all $1 \leq t \leq t^*+1$: $b_t \geq \sqrt{\Lambda\lambda e}\, b_{t-1}$. Since $b_1 \geq \sqrt{\frac{\Lambda}{\lambda e}}$, it follows by induction that:

$$b_t \geq \Lambda(\lambda\Lambda e)^{\frac{t}{2}-1} \text{ for all } 1 \leq t \leq t^*+1 \qquad (186)$$

We now divide the analysis into two cases. First, if $\bar{t}_o$ is such that $b_{\bar{t}_o-1} \geq \frac{\nu}{2(C-\lambda)}$. This implies that $\bar{t}_o - 1 \geq t^*+1$ by the definition of t^*. Thus, $t^* \leq \bar{t}_o - 2 \leq \bar{t}_o + \bar{t}_1$, which proves our claim for the first case.

If $\bar{t}_o$ is such that $b_{\bar{t}_o-1} < \frac{\nu}{2(C-\lambda)}$. Then, $\bar{t}_o \leq t^*+1$. Thus, $b_{\bar{t}_o} \geq \Lambda(\lambda Le)^{\frac{\bar{t}_o}{2}-1}$. Let $t_o = \min\{t : b_t \geq \Lambda(\lambda\Lambda e)^{\frac{\bar{t}_o}{2}-1}\}$. Thus, by Lemma 7, we get $t_o \leq \bar{t}_o$. Moreover, by the choice of t_o and w_o:

$$w_o \leq \lambda(\Lambda(\lambda\Lambda e)^{\frac{\bar{t}_o}{2}-1} - D) \leq \lambda(b_{t_o} - D) \qquad (187)$$

Now define sequence $(w_t : t \geq 0)$: $w_{t+1} = e^{w_t}$, where w_o was chosen according to (182). We already showed that $w_o \leq \lambda(b_{t_o} - D)$. Assume that $w_{t-1} \leq \lambda(b_{t_o+t-1} - D)$ for $t_o + t - 1 \leq t^*$. Then,

$$\begin{aligned}
\lambda(b_{t_o+t} - D) &\overset{(a)}{\geq} \lambda(\Lambda e^{\lambda b_{t_o+t-1}}(1 - \frac{\Lambda'}{\Lambda}e^{-\nu_o}) - D) \\
&\overset{(b)}{\geq} \lambda(\Lambda e^{\lambda D + w_{t-1}}(1 - \frac{\Lambda'}{\Lambda}e^{-\nu_o}) - D) \\
&\overset{(c)}{=} \lambda\Lambda e^{\lambda D} w_t(1 - \frac{\Lambda'}{\Lambda}e^{-\nu_o}) - \lambda D \\
&\overset{(d)}{\geq} w_t
\end{aligned}$$

where (a) holds by Lemma 6, (b) holds by the assumption that $w_{t-1} \leq \lambda(b_{t_o+t-1} - D)$, (c) holds by the definition of the sequence w_t and (d) holds by the choice of w_o and the fact that $w_t \geq w_o$. Thus, we showed by induction that

$$w_t \leq \lambda(b_{t_o+t} - D) \text{ for } 0 \leq t \leq t^* - t_o + 1. \qquad (188)$$

By the definition of $\bar{t}_1$ and since $w_1 \geq 1$, we have $\nu \leq w_{\bar{t}_1+1}$. Thus, $w_{\bar{t}_1+1} \geq \nu - \lambda D$. Since, by the definition of C, $\lambda \leq 2(C-\lambda)$. Therefore, $w_{\bar{t}_1+1} \geq \frac{\nu\lambda}{2(C-\lambda)} - \lambda D$. We will show that $t^* \leq \bar{t}_o + \bar{t}_1$ by contradiction. Let $t^* > \bar{t}_o + \bar{t}_1$. Thus, from (188), for $t = t_o + \bar{t}_1 + 1$:

$$b_{t_o+\bar{t}_1+1} \geq \frac{w_{\bar{t}_1+1}}{\lambda} + D \geq \frac{\nu}{2(C-\lambda)} \qquad (189)$$

which implies that $t_o + \bar{t}_1 + 1 \geq t^* + 1$, i.e., $t_o + \bar{t}_1 \geq t^*$, which contradicts the assumption that $t^* > \bar{t}_o + \bar{t}_1$.

To sum up, we have shown so far that if $\lambda > \frac{1}{\Lambda e}$, then $t^* \leq \bar{t}_o + \bar{t}_1$.

Since t^* is the last iteration for $b_t < \frac{\nu}{2(C-\lambda)}$. Then, $b_{t^*+1} \geq \frac{\nu}{2(C-\lambda)}$. We begin with $b_{t^*+1} = \frac{\nu}{2(C-\lambda)}$. Then by Lemma 6:

$$b_{t^*+2} \geq \Lambda e^{\lambda b_{t^*+1}}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu}{2}}) \qquad (190)$$

By Lemma 7, the sequence b_t is non-decreasing in t. We also known $t^*+2 \le \bar{t}_o + \bar{t}_1 + 2$. Using (190):

$$b_{\bar{t}_o+\log^*(\nu)+2} \ge \Lambda e^{\frac{\lambda\nu}{2(C-\lambda)}}\big(1-\frac{\Lambda'}{\Lambda}e^{\frac{-\nu}{2}}\big) \tag{191}$$

which concludes one case of the proof.

When $b_{t^*+1} > \frac{\nu}{2(C-\lambda)}$, we use the truncation process [25, Lemma 6], which depends only on the tree structure. Applying this truncation process, it can directly be shown that the tree can be truncated such that with probability one the value of b_{t^*+1} in the truncated tree is $\frac{\nu}{2(C-\lambda)}$. The truncation process [25, Lemma 6] depends only on the structure of the tree. In this paper, the side information is independent of the tree structure given the labels, therefore the same truncation process holds for our case, which concludes the proof using (190) and (191).

P. Proof of Sufficiency of Theorem 4

The assumption $(np)^{\log^*(\nu)} = n^{o(1)}$ ensures that $(np)^{\hat{t}} = n^{o(1)}$. Since $\frac{K^2(p-q)^2}{q(n-K)} \to \lambda$, $p \ge q$ and $\frac{p}{q} = \theta(1)$, then $(\frac{n-K}{K})^2 = O(np)$. Since $K = o(n)$, then $np \to \infty$. Thus, $(np)^{\hat{t}} = n^{o(1)}$ can be replaced by $(np+2)^{\hat{t}} = n^{o(1)}$, and hence, the coupling Lemma 10 holds. Moreover, since $(\frac{n-K}{K})^2 = O(np)$ and $np = n^{o(1)}$, $K = n^{1-o(1)}$.

Consider a modified form of Algorithm II whose output is $\hat{C} = \{i : R_i^{\hat{t}} \ge \nu\}$. Then for deterministic $|C^*| = K$, the following holds:

$$\begin{aligned} p_e &= \mathbb{P}(\text{No coupling})p_{e|\text{no coupling}} + \mathbb{P}(\text{coupling})p_{e|\text{coupling}} \\ &\le n^{-1+o(1)} + \frac{K}{n}e^{-\nu(r+o(1))} \end{aligned} \tag{192}$$

where the last inequality holds by Lemmas 10 and 9 for some positive constant r. Multiplying (192) by $\frac{n}{K}$:

$$\frac{\mathbb{E}[|C^*\triangle\hat{C}|]}{K} \le \frac{n^{o(1)}}{K} + e^{-\nu(r+o(1))} \to 0 \tag{193}$$

where the last inequality holds because $K = n^{1-o(1)}$ and $\nu \to \infty$.

Now going back to Algorithm II and its output $\tilde{C}$, using Equation (47):

$$\frac{\mathbb{E}[|C^*\triangle\tilde{C}|]}{K} \le 2\frac{\mathbb{E}[|C^*\triangle\hat{C}|]}{K} \to 0 \tag{194}$$

which concludes the proof under deterministic $|C^*| = K$.

When $|C^*|$ is random such that $K \ge 3\log(n)$ and $\mathbb{P}(||C^*| - K| \ge \sqrt{3K\log(n)}) \le n^{\frac{-1}{2}+o(1)}$, we have $\mathbb{E}[||C^*|-K|] \le n^{\frac{1}{2}+o(1)}$. Thus, for $\tilde{C}$, using Equation (47):

$$\frac{\mathbb{E}[|C^*\triangle\tilde{C}|]}{K} \le 2\frac{\mathbb{E}[|C^*\triangle\hat{C}|]}{K} + \frac{\mathbb{E}[||C^*|-K|]}{K} \to 0 \tag{195}$$

which concludes the proof.

Q. Proof of Necessity of Theorem 4

Since $(np+2)^{\hat{t}} = n^{o(1)}$, the coupling Lemma 10 holds. Moreover, since $(\frac{n-K}{K})^2 = O(np)$ and $np = n^{o(1)}$, $K = n^{1-o(1)}$. Consider a deterministic $|C^*| = K$. Then, for any local estimator $\hat{C}$:

$$\begin{aligned} p_e &= \mathbb{P}(\text{No coupling})p_{e|\text{no coupling}} + \mathbb{P}(\text{coupling})p_{e|\text{coupling}} \\ &\ge \frac{K(n-K)}{n^2}\mathbb{E}^2[e^{\frac{U_0}{2}}]e^{\frac{-\lambda\Lambda e}{4}} - n^{-1+o(1)} \end{aligned} \tag{196}$$

where the last inequality holds by Lemmas 10 and 9. Multiplying (196) by $\frac{n}{K}$:

$$\frac{\mathbb{E}[|C^*\triangle\hat{C}|]}{K} \ge \Big(1-\frac{K}{n}\Big)\mathbb{E}^2[e^{\frac{U_0}{2}}]e^{\frac{-\lambda\Lambda e}{4}} - o(1) \tag{197}$$

where the last inequality holds because $K = n^{1-o(1)}$. Thus, for $\lambda \le \frac{1}{\Lambda e}$, $\frac{\mathbb{E}[|C^*\triangle\hat{C}|]}{K}$ is bounded away from zero for any local estimator $\hat{C}$.

It can be shown that under a non-deterministic $|C^*|$ that obeys a distribution in the class of distributions mentioned earlier, the local estimator will do no better, therefore the same converse will hold.

R. Proof of Theorem 5

Let Z be a binomial random variable $\text{Bin}(n(1-\delta), \frac{K}{n})$. In view of Lemma 2, it suffices to verify (9) when $\hat{C}_k$ for each k is the output of belief propagation for estimating C_k^* based on observing $\boldsymbol{G}_k$ and $\boldsymbol{Y}_k$. The distribution of $|C_k^*|$ is obtained by sampling the indices of the original graph without replacement. Thus, for any convex function ϕ: $\mathbb{E}[\phi(|C_k^*|)] \le \mathbb{E}[\phi(Z)]$. Therefore, Chernoff bound for Z also holds for $|C_k^*|$. This leads to:

$$\begin{aligned} \mathbb{P}\Big(\big||C_k^*| - (1-\delta)K\big| \ge \sqrt{3K(1-\delta)\log(n)}\Big) &\le n^{-1.5+o(1)} \\ &\le n^{\frac{-1}{2}+o(1)} \end{aligned} \tag{198}$$

Thus, by Theorem 4, belief propagation achieves weak recovery for recovering C_k^* for each k. Thus:

$$\mathbb{P}\big(|\hat{C}_k\triangle C_k^*| \le \delta K \quad \text{for } 1 \le k \le \frac{1}{\delta}\big) \to 1 \tag{199}$$

which together with Lemma 2 conclude the proof.

S. Proof of Lemma 11

First, we expand $M(x)$ using Taylor series:

$$\begin{aligned} M(x) = &\frac{\frac{p}{q}-1}{1+e^{-(x-\nu)}} - \frac{1}{2}\big(\frac{\frac{p}{q}-1}{1+e^{-(x-\nu)}}\big)^2 \\ &+ O\Big(\big(\frac{\frac{p}{q}-1}{1+e^{-(x-\nu)}}\big)^3\Big) \end{aligned} \tag{200}$$

Thus:

$$\begin{aligned} \mathbb{E}[Z_0^{t+1}] = &-K(p-q) + Kq\mathbb{E}[M(Z_1^t+U_1)] \\ &+ (n-K)q\mathbb{E}[M(Z_0^t+U_0)] \\ = &-K(p-q) + K(p-q)\mathbb{E}\Big[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\Big] \\ &+ (n-K)(p-q)\mathbb{E}\Big[\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\Big] \\ &- \frac{K(p-q)^2}{2q}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big] \end{aligned}$$

$$-\frac{(n-K)(p-q)^2}{2q}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^0+U_0-\nu)}}\big)^2\Big]$$
$$+O\Bigg(\frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^3\Big]$$
$$+\frac{(n-K)(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\big)^3\Big]\Bigg) \tag{201}$$

Using Lemma 24 for $g(x)=\frac{1}{1+e^{-(x-\nu)}}$,

$$K(p-q)\mathbb{E}\Big[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\Big] + (n-K)(p-q)\mathbb{E}\Big[\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\Big] = K(p-q) \tag{202}$$

Similarly:

$$\frac{K(p-q)^2}{2q}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big] + \frac{(n-K)(p-q)^2}{2q}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^0+U_0-\nu)}}\big)^2\Big] = \frac{K(p-q)^2}{2q}\mathbb{E}\Big[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\Big] \tag{203}$$

and,

$$\frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^3\Big] + \frac{(n-K)(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\big)^3\Big] = \frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big] \tag{204}$$

Using (202), (203) and (204) and substituting in (201):

$$\mathbb{E}[Z_0^{t+1}] = -\frac{\lambda}{2}b_t + O\Bigg(\frac{K(p-q)^3}{q^2}\mathbb{E}[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2]\Bigg) = -\frac{\lambda}{2}b_t + o(1) \tag{205}$$

where the last equality holds by the definition of λ and b_t and because $\frac{K(p-q)^3}{q^2} = \lambda\frac{n}{K}(1-\frac{K}{n})(\frac{p}{q}-1)$ which is $o(1)$ because of the assumptions of the lemma which also implies that $\frac{p}{q}\to 1$.

To show (53), we use Taylor series: $M(x)=\frac{\frac{p}{q}-1}{1+e^{-(x-\nu)}} + O((\frac{\frac{p}{q}-1}{1+e^{-(x-\nu)}})^2)$. Then,

$$\begin{aligned}\mathbb{E}[Z_1^{t+1}] &= \mathbb{E}[Z_0^{t+1}] + K(p-q)\mathbb{E}[M(Z_1^t+U_1)]\\ &= \mathbb{E}[Z_0^{t+1}] + \frac{K(p-q)^2}{q}\mathbb{E}\Big[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\Big]\\ &\quad + O\Bigg(\frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big]\Bigg)\\ &= \mathbb{E}[Z_0^{t+1}] + \lambda b_t + o(1) = \frac{\lambda}{2}b_t + o(1)\end{aligned} \tag{206}$$

We now calculate the variance. For $Y=\sum_{i=1}^L X_i$, where L is Poisson distributed and $\{X_i\}$ are independent of Y and are i.i.d., it is well-known that $\text{var}(Y)=\mathbb{E}[L]\mathbb{E}[X_1^2]$. Thus,

$$\begin{aligned}\text{var}(Z_0^{t+1}) &= Kq\,\mathbb{E}[M^2(Z_1^t+U_1)] + (n-K)q\,\mathbb{E}[M^2(Z_0^t+U_0)]\\ &\overset{(a)}{=} \frac{K(p-q)^2}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big]\\ &\quad + \frac{(n-K)(p-q)^2}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\big)^2\Big]\\ &\quad + O\Bigg(\frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^3\Big]\\ &\quad + \frac{(n-K)(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\big)^3\Big]\Bigg)\\ &\overset{(b)}{=} \lambda b_t + o(1)\end{aligned} \tag{207}$$

where (a) holds because $\log^2(1+x)=x^2+O(x^3)$ for all $x\geq 0$ and (b) holds by similar analysis as in (205).

Similarly,

$$\begin{aligned}\text{var}(Z_1^{t+1}) &= \text{var}(Z_0^{t+1}) + O\Bigg(\frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big]\Bigg)\\ &= \lambda b_t + o(1)\end{aligned} \tag{208}$$

T. Proof of Lemma 12

Before we prove the lemma, we need the following lemma from [43, Theorem 3].

Lemma 26. *Let $S_\gamma = X_1+\cdots+X_{N_\gamma}$, where $X_i : i\geq 1$ are i.i.d. random variables with mean μ, variance σ^2 and $\mathbb{E}[|X_i|^3]\leq\rho^3$, and for some $\gamma>0$, N_γ is a Poi(γ) random variable independent of $(X_i : i\geq 1)$. Then,*

$$\sup_x \Big|\mathbb{P}\big(\frac{S_\gamma-\gamma\mu}{\sqrt{\gamma(\mu^2+\sigma^2)}}\leq x\big)-\phi(x)\Big| \leq \frac{0.3041\rho^3}{\sqrt{\gamma(\mu^2+\sigma^2)^3}} \tag{209}$$

For $t\geq 0$, Z_0^{t+1} can be represented as follows:

$$Z_0^{t+1} = -K(p-q) + \sum_{i=1}^{N_{nq}} X_i \tag{210}$$

where N_{nq} is distributed according to Poi(nq), the random variables $X_i, i\geq 1$ are mutually independent and independent of N_{nq} and X_i is a mixture:

$$X_i = \frac{(n-K)q}{nq}M(Z_0^t+U_0) + \frac{Kq}{nq}M(Z_1^t+U_1).$$

Starting with (210), using the properties of compound Poisson distribution, and then applying Lemma 11:

$$nq\mathbb{E}[X_i^2] = \text{var}(Z_0^{t+1}) = \lambda b_t + o(1) \tag{211}$$

Also, using $\log^3(1+x)\leq x^3$ for all $x\geq 0$:

$$\begin{aligned}nq\mathbb{E}[|X_i^3|] &\leq \frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1)+\nu}}\big)^3\Big]\\ &\quad + \frac{(n-K)(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_0^t+U_0)+\nu}}\big)^3\Big]\\ &\overset{(a)}{\leq} \frac{K(p-q)^3}{q^2}\\ &\overset{(b)}{=} o(1)\end{aligned} \tag{212}$$

where (a) holds by Lemma 24 for $g(x) = \frac{1}{1+e^{-(x-\nu)}}$ and (b) holds since $\frac{p}{q} \to 1$.

Combining (211) and (212) yields $\frac{\mathbb{E}[|X_i^3|]}{\sqrt{nq\mathbb{E}^3[X_i^2]}} = \frac{nq\mathbb{E}[|X_i^3|]}{\sqrt{(nq\mathbb{E}[X_i^2])^3}} \to 0$, which together with Lemma 26 yields:

$$\sup_x \left|\mathbb{P}\Big(\frac{Z_0^{t+1} + \frac{\lambda b_t}{2}}{\sqrt{\lambda b_t}} \le x\Big) - \phi(x)\right| \to 0 \tag{213}$$

Similarly, for $t \ge 0$, Z_1^{t+1} can be represented as follows:

$$Z_1^{t+1} = -K(p-q) + \frac{1}{\sqrt{(n-K)q}} \sum_{i=1}^{N_{(n-K)q+Kp}} Y_i \tag{214}$$

where $N_{(n-K)q+Kp}$ is distributed according to Poi$((n-K)q+Kp)$, the random variables $Y_i, i \ge 1$ are mutually independent and independent of $N_{(n-K)q+Kp}$ and Y_i is a mixture:

$$\begin{aligned} Y_i =& \frac{(n-K)q}{(n-K)q+Kp} M(Z_0^t + U_0) \\ &+ \frac{Kp}{(n-K)q+Kp} M(Z_1^t + U_1). \end{aligned}$$

Starting with (214), using the properties of compound Poisson distribution, and then applying Lemma 11:

$$((n-K)q+Kp)\mathbb{E}[Y_i^2] = \text{var}(Z_1^{t+1}) = \lambda b_t + o(1) \tag{215}$$

Also, using $\log^3(1+x) \le x^3$ for all $x \ge 0$:

$$\begin{aligned} ((n-K)q+Kp)\mathbb{E}[|Y_i^3|] &= nq\mathbb{E}[|X_i|^3] \\ &\quad + K(p-q)\mathbb{E}\left[\Big(\frac{\frac{p}{q}-1}{1+e^{-(Z_1^t+U_1)+\nu}}\Big)^3\right] \\ &\le o(1) \end{aligned} \tag{216}$$

where (216) holds since $\frac{p}{q} \to 1$.

Combining (215) and (216) yields $\frac{\mathbb{E}[|Y_i^3|]}{\sqrt{(n-K)q+Kp)\mathbb{E}^3[Y_i^2]}} \to 0$, which together with Lemma 26 yields:

$$\sup_x \left|\mathbb{P}\Big(\frac{Z_1^{t+1} - \frac{\lambda b_t}{2}}{\sqrt{\lambda b_t}} \le x\Big) - \phi(x)\right| \to 0 \tag{217}$$

Hence, using (213) and (217), it suffices to show that $\lambda b_t \to v_{t+1}$, which implies that (55) and (56) are satisfied. We use induction to prove that $\lambda b_t \to v_{t+1}$. At $t = 0$, we have: $v_1 = \lambda\mathbb{E}[\frac{1}{e^{-\nu}+e^{-U_1}}] = \lambda b_0$. Hence, our claim is satisfied for $t = 0$. Assume that $\lambda b_t \to v_{t+1}$. Then,

$$\begin{aligned} b_{t+1} &= \mathbb{E}[\frac{1}{e^{-\nu}+e^{-(Z_1^{t+1}+U_1)}}] = \mathbb{E}_{U_1}[\mathbb{E}_{Z_1}[\frac{1}{e^{-\nu}+e^{-(Z_1^t+u)}}]] \\ &= \mathbb{E}_{U_1}[\mathbb{E}_{Z_1}[f(Z_1^{t+1};u,\nu)]] = \mathbb{E}_{U_1}[\mathbb{E}_{Z_1}[\mathcal{E}_n]] \end{aligned} \tag{218}$$

where $f(z;u,\nu) = \frac{1}{e^{-\nu}+e^{-(z+u)}}$ and $\mathcal{E}_n$ is a sequence of random variables representing $f(Z;u,\nu)$ as it evolves with n. Let $G(s)$ denote a Gaussian random variable with mean $\frac{s}{2}$ and variance s.

From (217), we have $\text{Kolm}\big(Z_1^{t+1}, G(\lambda b_t)\big) \to 0$ where $\text{Kolm}(\cdot,\cdot)$ is the Kolmogorov distance (supremum of absolute difference of CDFs). Since $f(z;u,\nu)$ is non-negative and monotonically increasing in z and since the Kolmogorov distance is preserved under monotone transformation of random variables, it follows that $\text{Kolm}\big(f(Z_1^{t+1};u,\nu), f(G(\lambda b_t);u,\nu)\big) \to 0$. Since $\lim_{z\to\infty} f(z;u\nu) = e^\nu$, using the definition of Kolmogorov distance and by expressing the CDF of $f(G(\lambda b_t);u,\nu)$ in terms of the CDF of $G(\lambda b_t)$ and the inverse of $f(z;u,\nu)$, we get:

$$\sup_{0<c<e^\nu} \left|F_{\mathcal{E}_n}(c) - F_{G(\lambda b_t)}\Big(\log\big(\frac{ce^{-u}}{1-ce^{-\nu}}\big)\Big)\right| \to 0 \tag{219}$$

From the induction hypothesis, $\lambda b_t \to v_{t+1}$. Thus,

$$\sup_{0<c<e^\nu} \left|F_{\mathcal{E}_n}(c) - F_{G(v_{t+1})}\Big(\log\big(\frac{ce^{-u}}{1-ce^{-\nu}}\big)\Big)\right| \to 0 \tag{220}$$

which implies that the sequence of random variables $\mathcal{E}_n$ converges in Kolmogorov distance to a random variable $\frac{1}{e^{-\nu}+e^{-(G(v_{t+1})+u)}}$ as $n \to \infty$. This implies the following convergence in distribution:

$$\mathcal{E}_n \overset{i.d.}{\to} \frac{1}{e^{-\nu}+e^{-(G(v_{t+1})+u)}} \tag{221}$$

Moreover, the second moment of $\mathcal{E}_n$ is bounded from above independently of n:

$$\mathbb{E}[\mathcal{E}_n^2] \overset{(a)}{\le} e^{2\nu} \overset{(b)}{\le} A \tag{222}$$

where (a) holds by the definition of $\mathcal{E}_n$, and (b) holds for positive constant A since based on the assumptions of the lemma, ν is constant as $n \to \infty$.

By (220), (221) and (222), the dominated convergence theorem implies that, as $n \to \infty$, the mean of $\mathcal{E}_n$ converges to the mean of the random variable $\frac{1}{e^{-\nu}+e^{-(G(v_{t+1})+u)}}$. Since the cardinality of side information is finite and independent of n, it follows that:

$$\begin{aligned} b_{t+1} &= \mathbb{E}_{U_1}\big[\mathbb{E}[\mathcal{E}_n]\big] \\ &\overset{(a)}{\to} \mathbb{E}_{U_1}\left[\mathbb{E}_Z\left[\frac{1}{e^{-\nu}+e^{-(\frac{v_{t+1}}{2}+\sqrt{v_{t+1}}Z)-u}}\right]\right] \\ &= \frac{v_{t+2}}{\lambda} \end{aligned} \tag{223}$$

where in (a) we define $Z \sim \mathcal{N}(0,1)$. Equation (223) implies that $\lambda b_{t+1} \to v_{t+2}$, which concludes the proof of the lemma.

U. Proof of Lemma 14

Let $\kappa = \frac{n}{K}$. Since for all ℓ: $|h_\ell| < \nu$, it follows that for any $t \ge 0$ and for sufficiently large κ:

$$\begin{aligned} v_{t+1} &= \lambda\, \mathbb{E}_{Z,U_1}\left[\frac{1}{e^{-\nu}+e^{-(\frac{v_t}{2}+\sqrt{v_t}Z)-U_1}}\right] \\ &= \lambda \sum_{\ell=1}^{L} \frac{\alpha_{+,\ell}^2}{\alpha_{-,\ell}}\, \mathbb{E}_Z\left[\frac{1}{e^{-\nu(1-\frac{h_\ell}{\nu})}+e^{-(\frac{v_t}{2}+\sqrt{v_t}Z)}}\right] \\ &\overset{(a)}{=} \lambda \sum_{\ell=1}^{L} \frac{\alpha_{+,\ell}^2}{\alpha_{-,\ell}}\, \mathbb{E}_Z\left[\frac{1}{e^{-C_l\nu}+e^{-(\frac{v_t}{2}+\sqrt{v_t}Z)}}\right] \\ &\overset{(b)}{=} \lambda\Lambda e^{v_t}(1+o(1)) \end{aligned} \tag{224}$$

where (a) holds for positive constants $C_\ell, \ell \in \{1,\cdots,L\}$ and (b) holds because $\mathbb{E}_Z[e^{\frac{v_t}{2}+\sqrt{v_t}Z}] = e^{v_t}$.

Consider the sequence $w_{t+1} = e^{w_t}$ with $w_0 = 0$. Define $t^* = \log^*(\nu)$ to be the number of times the logarithm function must be iteratively applied to ν to get a result less than or equal to one. Since $w_1 = 1$ and w_t is increasing in t, we have $w_{t^*+1} \geq \nu$ (check by applying the $\log$ function t^* times to both sides). Thus, as κ grows, we have $\nu = o(w_{t^*+2})$.

Since $\Lambda \to \infty$ as κ grows, it follows by induction that for any fixed $\lambda > 0$:

$$v_t \geq w_t \tag{225}$$

for all $t \geq 0$ and for all sufficiently large κ. Thus,

$$v_{t^*+2} \geq w_{t^*+2} \tag{226}$$

which implies that as κ grows, $\nu = o(v_{t^*+2})$ and $h_\ell = o(v_{t^*+2})$ for all ℓ. Since v_t is increasing in t, using (224) and (226), we get for all sufficiently large κ and after $t^* + 2$ iterations of belief propagation (or for a tree of depth $t^* + 2$):

$$\mathbb{E}_{U_0}\Big[Q(\frac{\nu + \frac{v_{t^*+2}}{2} - U_0}{\sqrt{v_{t^*+2}}})\Big] = Q\Big(\frac{1}{2}\sqrt{v_{t^*+2}}(1 + o(1))\Big) \tag{227}$$

$$\mathbb{E}_{U_1}\Big[Q(\frac{-\nu + \frac{v_{t^*+2}}{2} + U_1}{\sqrt{v_{t^*+2}}})\Big] = Q\Big(\frac{1}{2}\sqrt{v_{t^*+2}}(1 + o(1))\Big) \tag{228}$$

Since $Q(x) \leq e^{-\frac{1}{2}x^2}$ for $x \geq 0$, then using (226), (227) and (228):

$$\frac{n-K}{K} Q\Big(\frac{1}{2}\sqrt{v_{t^*+2}}(1 + o(1))\Big) \to 0 \tag{229}$$

$$Q\Big(\frac{1}{2}\sqrt{v_{t^*+2}}(1 + o(1))\Big) \to 0 \tag{230}$$

Using (229) and (230) and Lemma 13, we get:

$$\lim_{\frac{n}{K}\to\infty} \lim_{nq, Kq \to\infty} \lim_{n\to\infty} \frac{\mathbb{E}[\hat{C} \triangle C^*]}{K} = 0 \tag{231}$$

REFERENCES

[1] H. Saad and A. Nosratinia, "Belief propagation with side information for recovering a single community," in *IEEE International Symposium on Information Theory*, June 2018.

[2] ——, "Side information in recovering a single community: Information theoretic limits," in *IEEE International Symposium on Information Theory*, June 2018.

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

[4] A. Zhang and H. Zhou, "Minimax rates of community detection in stochastic block models," *The Annals of Statistics*, vol. 44, no. 5, pp. 2252–2280, Oct. 2016.

[5] P. J. Bickel and A. Chen, "A nonparametric view of network models and Newman-Girvan and other modularities," *National Academy of Sciences*, vol. 106, no. 50, pp. 21 068–21 073, 2009.

[6] T. T. Cai and X. Li, "Robust and computationally feasible community detection in the presence of arbitrary outlier nodes," *The Annals of Statistics*, vol. 43, no. 3, pp. 1027–1059, June 2015.

[7] T. A. B. Snijders and K. Nowicki, "Estimation and prediction for stochastic blockmodels for graphs with latent block structure," *Journal of Classification*, vol. 14, pp. 75–100, 1997.

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

[9] A. Coja-Oghlan, "Graph partitioning via adaptive spectral techniques," *Comb. Probab. Comput.*, vol. 19, no. 2, pp. 227–284, Mar. 2010.

[10] ——, "A spectral heuristic for bisecting random graphs," in *ACM-SIAM Symposium on Discrete Algorithms*, 2005, pp. 850–859.

[11] A. Anandkumar, R. Ge, D. Hsu, and S. M. Kakade, "A tensor approach to learning mixed membership community models," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 2239–2312, Jan. 2014.

[12] Y. Chen, S. Sanghavi, and H. Xu, "Improved graph clustering," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6440–6455, Oct. 2014.

[13] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Phys. Rev. E*, vol. 84, p. 066106, Dec. 2011.

[14] P. Zhang, F. Krzakala, J. Reichardt, and L. Zdeborová, "Comparative study for inference of hidden classes in stochastic block models," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 12, p. P12021, 2012.

[15] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[16] 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, June 2014.

[17] J. Chen and B. Yuan, "Detecting functional modules in the yeast protein-protein interaction network," *Bioinformatics*, vol. 22, no. 18, pp. 2283–2290, Sept. 2006.

[18] A. Beutel, W. Xu, V. Guruswami, C. Palow, and C. Faloutsos, "Copy-catch: Stopping group attacks by spotting lockstep behavior in social networks," in *Proceedings of the 22nd International Conference on World Wide Web*, 05 2013, pp. 119–130.

[19] D. H. Chau, S. Pandit, and C. Faloutsos, "Detecting fraudulent personalities in networks of online auctioneers," in *Proceedings of the 10th European Conference on Principle and Practice of Knowledge Discovery in Databases*, ser. PKDD'06, 2006, pp. 103–114.

[20] H. Firouzi, B. Rajaratnam, and A. Hero, III, "Predictive correlation screening: Application to two-stage predictor design in high dimension," in *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, vol. 31, 29 Apr–01 May 2013, pp. 274–288.

[21] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75 – 174, Jan. 2010.

[22] E. Abbe and C. Sandon, "Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery," in *Symposium on Foundations of Computer Science (FOCS)*, ser. FOCS '15, 2015, pp. 670–688.

[23] A. Montanari, "Finding one community in a sparse graph," *Journal of Statistical Physics*, vol. 161, no. 2, pp. 273–299, Oct. 2015.

[24] B. Hajek, Y. Wu, and J. Xu, "Information limits for recovering a hidden community," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 4729–4745, Aug 2017.

[25] ——, "Recovering a Hidden Community Beyond the Kesten-Stigum Threshold in $O(|E| \log^* |V|)$ Time," *ArXiv e-prints*, Oct. 2015.

[26] A. Kadavankandy, K. Avrachenkov, L. Cottatellucci, and R. Sundaresan, "The power of side-information in subgraph detection," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1905–1919, April 2018.

[27] E. Mossel and J. Xu, "Local algorithms for block models with side information," in *ACM Conference on Innovations in Theoretical Computer Science*, 2016, pp. 71–80.

[28] T. T. Cai, T. Liang, and A. Rakhlin, "Inference via message passing on partially labeled stochastic block models," *ArXiv e-prints*, Mar. 2016.

[29] H. Saad, A. Abotabl, and A. Nosratinia, "Exact recovery in the binary stochastic block model with binary side information," in *Allerton Conference on Communication, Control, and Computing*, Oct. 2017, pp. 822–829.

[30] 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*, 2018.

[31] A. R. Asadi, E. Abbe, and S. Verdú, "Compressing data on graphs with clusters," in *IEEE International Symposium on Information Theory*, June 2017, pp. 1583–1587.

[32] V. Kanade, E. Mossel, and T. Schramm, "Global and local information in clustering labeled block models," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5906–5917, Oct. 2016.

[33] Y. Deshpande, S. Sen, A. Montanari, and E. Mossel, "Contextual stochastic block models," in *Advances in Neural Information Processing Systems 31*, 2018, pp. 8581–8593.

[34] B. Hajek, Y. Wu, and J. Xu, "Computational lower bounds for community detection on random graphs," in *Proceedings of The 28th Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, vol. 40, 03–06 Jul 2015, pp. 899–928.

[35] Z. Ma and Y. Wu, "Computational barriers in minimax submatrix detection," *The Annals of Statistics*, vol. 43, no. 3, pp. 1089–1116, 06 2015.
[36] M. Brennan, G. Bresler, and W. Huleihel, "Reducibility and computational lower bounds for problems with planted sparse structure," in *Proceedings of the 31st Conference On Learning Theory*, ser. Proceedings of Machine Learning Research, vol. 75, 06–09 Jul 2018, pp. 48–166.
[37] H. Kobayashi and J. Thomas, "distance measures and related criteria," in *Allerton Conference Circuits and System Theory*, 1967.
[38] T. Apostol, *Calculus*. Blaisdell Pub. Co., 1962, vol. 2.
[39] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Berlin; New York: Springer-Verlag Inc, 2010.
[40] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," Jan. 2017.
[41] H. Saad and A. Nosratinia, "Recovering a Single Community with Side Information," *arXiv e-prints*, p. arXiv:1809.01738, Sep 2018.
[42] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American statistical association*, vol. 58, no. 301, pp. 13–30, 1963.
[43] V. Korolev and I. Shevtsova, "An improvement of the berry–esseen inequality with applications to poisson and mixed poisson random sums," *arXiv:0912.2795*, Dec. 2009.

Hussein Saad received his Ph.D. degree in electrical engineering at the University of Texas at Dallas, Richardson, TX, USA; his M.S. degree in electrical engineering from Nile University, Giza, Egypt; and his B.S. degree in electrical engineering from the German University, Cairo, Egypt. He worked as a research assistant professor at Wireless@VT, Bradley Dept. of Electrical and Computer Engineering, Virginia TechWireless@VT, Bradley Dept. of Electrical and Computer Engineering, Virginia Tech. His research interests include machine learning, graph theory, information theory and wireless communications. He is currently a Senior Engineer at Qualcomm Inc. Technologies.

Aria Nosratinia (S'87, M'97, SM'04, F'10) is Erik Jonsson distinguished professor and associate head of the Electrical and Computer Engineering department at the University of Texas at Dallas. He received his Ph.D. in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 1996. He has held visiting appointments at Princeton University, Rice University, and UCLA. His interests lie in information theory and signal processing, with application in wireless communications. He is a fellow of IEEE for contributions to multimedia and wireless communications. He has served as editor and area editor for the IEEE Transactions on Wireless Communications, and as editor for the IEEE Transactions on Information Theory, IEEE Transactions on Image Processing, IEEE Signal Processing Letters, IEEE Wireless Communications (Magazine), and Journal of Circuits, Systems, and Computers. He has received the National Science Foundation career award, and the outstanding service award from the IEEE Signal Processing Society, Dallas Chapter. He has served in the organizing and technical program committees of a number of conferences, including as general co-chair of IEEE Information Theory Workshop 2018. He was named a highly cited researcher by Clarivate Analytics (formerly Thomson Reuters).