A GENERALIZATION OF THE COMMUNITY DETECTION PROBLEM VIA SIDE INFORMATION

by

Hussein Metwaly Saad

APPROVED BY SUPERVISORY COMMITTEE:

Aria Nosratinia, Chair

Carlos A. Busso-Recabarren

Nasser Kehtarnavaz

Nicholas Ruozzi

Dedicated to my family, I would not be anything without your support.

A GENERALIZATION OF THE COMMUNITY DETECTION PROBLEM VIA SIDE INFORMATION

by

HUSSEIN METWALY SAAD, BS, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

August 2019

ACKNOWLEDGMENTS

First of all, thanks are due to the Almighty God for giving me the strength and patience to complete my studies and write this dissertation, which I sincerely hope will contribute to the welfare of people.

I owe my sincere gratitude to my PhD adviser Dr. Aria Nosratinia for his endless support and guidance though my PhD. I was lucky to learn from Dr. Aria Nosratinia on all levels, technical, personal, and professional. I also appreciate the valuable remarks of my PhD committee members, Dr. Carlos Busso, Dr. Nasser Kehtarnavaz, and Dr. Nicholas Ruozzi.

I am thankful to my colleagues who encouraged and supported me during my studies. First, thanks to Mostafa Ibrahim for the good memories, company, and amazing food. Second, thanks to Noha Helal, Ahmed Hindy, Mohamed Fadel, Ahmed Abotabl, Mahmoud ElGenedy, Ahmed Helmy, Ahmed Omar, Ahmed El-Samadouny, Mohamed Mokhtar, Mohamed Marzban, Ahmed El Shafie, Sherief Helwa, Yahia Ramdan, Mahmoud Abdelgalil, Ahmed Magdy, Rana Ahmed, Doaa Mahmoud, Fan Zhang, Hassan Zivari-Fard, Mohammad Esmaeili, and Heping Wang.

Finally, I am grateful to my family: my father, mother and sisters, who have been a constant source of motivation.

July 2019

A GENERALIZATION OF THE COMMUNITY DETECTION PROBLEM VIA SIDE INFORMATION

Hussein Metwaly Saad, PhD
The University of Texas at Dallas, 2019

Supervising Professor: Aria Nosratinia, Chair

The standard community detection consists of observing a graph and detecting its node labels. However, many practical problems offer further information about the individual node labels, which we denote side information. For example, social networks such as Facebook and Twitter have access to information other than the graph edges, such as age, gender, etc. This dissertation aims to understand when and by how much can side information change the fundamental limits of the community detection problem, devise efficient algorithms for it, and study the asymptotic performance of these algorithms.

In the context of the binary symmetric and single-community stochastic block models for the graph, we introduce a model for side information that conveniently captures the variation of its quantity and quality as the size of the graph grows.

For the binary symmetric stochastic block model under side information, we characterize tight necessary and sufficient conditions for exact recovery. An efficient, asymptotically optimal two-stage algorithm is introduced for recovery. Furthermore, the analysis of phase transition is extended to continuous-valued side information.

For the single community stochastic block model, we characterize tight necessary and sufficient conditions for weak and exact recovery. An efficient belief propagation algorithm under side information is proposed. A weak recovery phase transition for belief propagation is characterized when both quality and quantity are fixed. When quality of side information varies with graph size, sufficient conditions for weak recovery are established. Sufficient conditions for belief propagation to achieve exact recovery are derived.

The extrinsic information transfer (EXIT) method is applied to the analysis of belief propagation in community detection with side information, providing insights such as the asymptotic threshold for weak recovery, the performance of belief propagation near the optimal threshold, and the performance of belief propagation through the first few iterations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Community Structure in Graphs

The field of graph theory dates back to Euler's solution of the Konigsberg's bridges puzzle in 1736 (Euler, 1736). Since then a lot has been discovered and learned about graphs, their properties, and their mathematical analysis (Bollobs, 1998). Graphs represent a wide variety of systems and data sets that can be found in different areas of study. Social, biological, and technological networks can be represented as graphs, and graph analysis has been shown to be a crucial tool in understanding the properties and features of these networks, e.g., the degree distribution of the nodes of the graph. Community structure is one of the important features of these networks, and has been shown to be found in many real data sets and applications (Fortunato, 2010a). Basically, communities (called clusters as well) are groups of nodes which probably share common properties and/or play similar roles within the graph. Figure 1.1 shows a toy example of a graph with communities.

Fundamentally, the goal of the community detection problem is to learn/detect the hidden community structure upon observing the graph, see Figure 1.2 for a toy example. Community detection has many applications: detecting web clients who have similar interests (Krishnamurthy and Wang, 2000), finding like-minded people in social networks (Girvan and Newman, 2002), improving recommendation systems (Xu et al., 2014), detecting protein complexes (Chen and Yuan, 2006), detecting clusters of customers with similar interests in a purchase interaction network (Reddy et al., 2002). Grouping the nodes into communities can be used as a compression mechanism in order to efficiently store the graph (Wu et al., 2004). Moreover, identifying communities allows for classifying the nodes of the graph based on their structural position in the graph (Csermely, 2008).

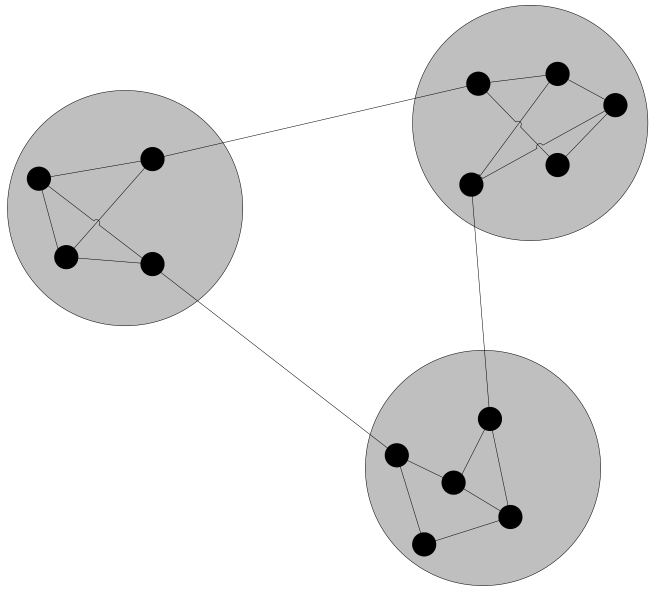

Figure 1.1. A simple graph with three communities.

1.2 Algorithms and Methods for Community Detection

Many algorithms and methods have been proposed for detecting communities in graphs, such as traditional clustering algorithms, e.g., K-means and hierarchical clustering, optimization based algorithms, spectral algorithms, and dynamic algorithms, see (Fortunato, 2010a) and reference therein.

Methods based on statistical inference are the main focus of this dissertation . These methods are based mainly on devising generative models (random graphs) whose parameters are then fit (tuned) to some real network using a statistical inference algorithm, e.g., maximum likelihood, expectation-maximization. Community detection on random graphs has been studied in statistics (Holland et al., 1983; Zhang and Zhou, 2016; Bickel and Chen, 2009; Cai and Li, 2015; Snijders and Nowicki, 1997), computer science (Chen and Xu, 2016; Coja-oghlan, 2010; Coja-Oghlan, 2005; Anandkumar et al., 2014; Chen et al., 2014) and theoretical statistical physics (Decelle et al., 2011a; Zhang et al., 2012). Among the dif-

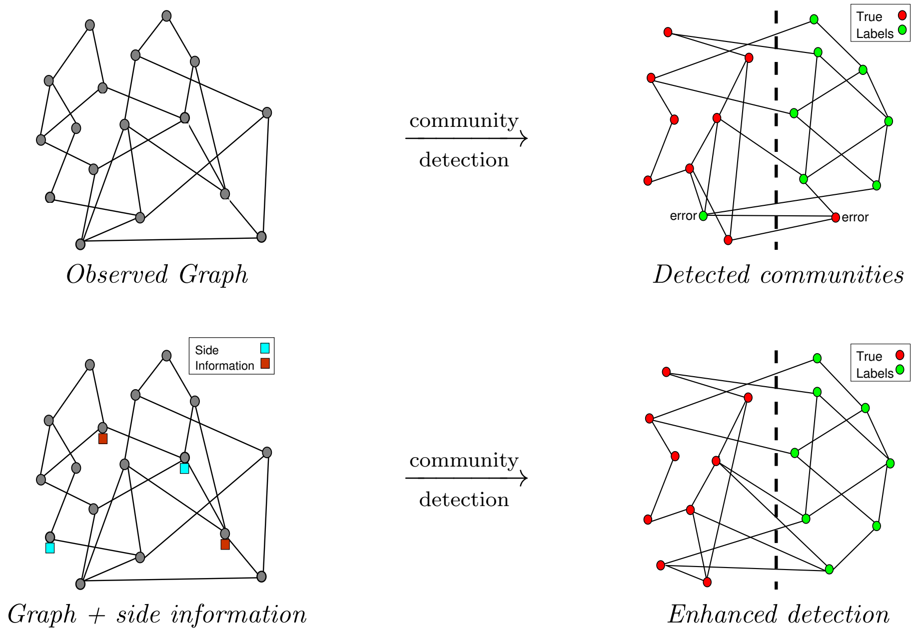

Figure 1.2. (top) standard community detection (bottom) Community detection with side information

ferent random graph (generative) models (Lancichinetti and Fortunato, 2009; Fortunato, 2010b), the stochastic block model (SBM) is widely used in the context of community detection (Abbe and Sandon, 2015). *The stochastic block model and its variants will be used as the main model through out this dissertation.*

1.3 Community Detection with Side Information

Despite the long history of the community detection problem, most of the literature focused on the graph as the only source of information. The graphical structure of the problem has lead to devising many community detection algorithms as well as well-characterized asymptotic results (e.g. phase transitions) that can give insights on the performance of inference algorithms on large data sets. But considering only graphical (pair-wise) observations also unfortunately limits the scope of the applicability of the model, since in many practical application non-graphical (per node) relevant information is available that can aid the inference. For example, social networks such as Facebook and Twitter have access to much information other than the graph edges, such as age, gender, school, etc. A citation network that has the authors names, keywords, and abstracts of papers, and therefore may provide significant additional information beyond the co-authoring relationships. Throughout this dissertation, we call such non-graphical observations *side information*.

1.4 Objectives

This dissertation answers the following question: *When and by how much does side information help in recovering the communities?* We answer this question in the following contexts:

- Studying the effect of side information on the fundamental limits of the community detection problem.
- Devising efficient algorithms that combine the graph and side information, and studying the asymptotic performance of these algorithms.

1.5 Related Work

In statistics, the problem of community detection with additional information such as "annotation" (Newman and Clauset, 2016), "attributes" (Yang et al., 2013), or "features" (Zhang et al., 2016) has been broached, wherein for matching to real (finite) data sets a parametric model is proposed that expresses the joint probability distribution of the graphical and non-graphical (attribute/feature) observations conditioned on the true label and a modeling parameter. These works concentrate on model-matching and inference using graphical and non-graphical observations. However, none of these works focus on the effect of side information on the fundamental limits of the community detection problem, nor they focus on the asymptotic limits of efficient algorithms.

Some works appeared that considered asymptotics when side information is available. Before stating them, a few basic definitions need to be highlighted.

- Correlated recovery refers to community detection that performs better than random guessing (Decelle et al., 2011b; Mossel et al., 2014; Massoulié, 2014; Mossel et al., 2018; Abbe and Sandon, 2018).

- Weak recovery refers to a vanishing fraction of misclassified labels (Yun and Proutiere, 2014; Mossel and Xu, 2016a; Saad et al., 2016).

- Exact recovery refers to recovering all community labels with probability converging to one as $n \to \infty$[1] (Abbe et al., 2016; Elchanan et al., 2015; Abbe and Sandon, 2015).

- Phase transition refers to a threshold on the random graph parameters such that on one side of the threshold no algorithm can achieve a certain form of recovery, and on the other side *some* algorithm exists to achieve recovery.

- A sparse regime is in place when the average degree of the graph is $\Omega(1)$, and a graph is dense if the average degree is $\Omega(\log n)$.

The asymptotic behavior of belief propagation with side information has been studied in binary community detection in the sparse regime. Mosel and Xu (Mossel and Xu, 2016b) considered side information in the form of noisy version of the true community labels, showing that subject to such side information, belief propagation under certain condition has the same residual error as the MAP estimator. Cai *et. al* (Cai et al., 2016) considered side information in the form of a fraction of the true labels being revealed, demonstrating regimes for correlated recovery and weak recovery. Both (Mossel and Xu, 2016b; Cai et al., 2016) present sufficient (but not necessary) conditions. Kadavankandy *et al.* (Kadavankandy et al., 2018) studies the single-community problem under side information consisting of noisy version of the true labels, where they showed weak recovery in the sparse regime. Kanade *et. al* (Kanade et al., 2016) showed that for symmetric communities, the phase transition of

[1] n is the number of nodes in the graph

correlated recovery is not affected if a vanishing fraction of the labels are revealed. The same side information was studied in (Caltagirone et al., 2018) under binary asymmetric communities, showing that local algorithms achieve correlated recovery up to the phase transition threshold.

1.6 Contributions and Organization

The contributions of this dissertation are motivated and directed by several observations. First, while the effect of side information on correlated recovery and weak recovery has been studied, its effect on exact recovery has been unknown. Second, for efficient algorithms, the literature has concentrated on the effect of side information only on belief propagation. Third, only binary side information about binary labels has been studied (or binary side information with erasures). The more general case where side information consists of several features each with an arbitrary alphabet is motivated by many practical applications has not been thoroughly studied either in the context of fundamental limits or efficient algorithms. Finally, in many cases (even for correlated and weak recovery) either necessary or sufficient conditions for recovery is known, but not both.

The contributions of this dissertation (divided by chapters) can be summarized as follows. Chapter 2 considers the binary symmetric stochastic block model with n nodes, and studies the effect of the quality and quantity of side information on the phase transition of exact recovery in the dense regime. To study the effect of quality, we consider three models of discrete-valued side information, where in all of them the log-likelihood ratio (LLR) of side information (representing quality) is allowed to vary with n. First, we consider noisy-label side information with noise parameter $\alpha \in (0, 0.5)$. Second, we consider partially revealed side information with parameter $\epsilon \in (0, 1)$. Finally, we consider side information consisting of multiple features each with arbitrary finite alphabet. For all three models, we characterize necessary and sufficient conditions that are tight, except in one special case. Moreover, for all three models we use the maximum likelihood detector to characterize the necessary conditions. For the sufficient conditions, we propose a two-stage efficient algorithm, where in the first stage, a weak recovery algorithm (already proposed in the literature) is used on the graph alone, then in the second stage a local improvement is proposed that uses both the graph and the side information. To the study of the quantity, we consider side information in the form of a vector of independent, and identically distributed (i.i.d) observations. In this model, the LLR of side information is fixed while the size of the vector (representing the quantity) is allowed to vary with n. We characterize tight necessary and sufficient conditions for exact recovery. The results of this chapter were published in (Saad et al., 2017; Saad and Nosratinia, 2018).

Chapter 3 refines and improves the analysis of the sharp threshold for exact recovery subject to side information that was characterize in chapter two. We provide a new analysis for the necessary conditions which achieves two goals: closing the gap between necessary and sufficient conditions in the special case left in chapter two, and generalizing the results to infinite cardinality, including continuous-valued side information. The results of this chapter were published in (Saad and Nosratinia, 2019).

Chapter 4 characterizes the utility of side information in single-community detection, in particular exploring when and by how much can side information improve the information limit (phase transition), as well as the phase transition of belief propagation, in single-community detection. Unlike the previous chapters, we consider weak and exact recovery. 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.

First, the information limits in the presence of side information are characterized. When the dimension of side information for each node varies but its LLR is fixed across n, tight necessary and sufficient conditions are calculated for both weak and exact recovery. Also, it is shown that under the same sufficient conditions, weak recovery is achievable even when the size of the community is random and unknown. We also find conditions on the graph and side information where achievability of weak recovery implies achievability of 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 the side information for each node has fixed dimension but varying LLR, we find tight necessary and sufficient conditions for exact recovery, and necessary conditions for weak recovery. Under varying LLR, our results apply to side information with finite alphabet. Second, the phase transition of belief propagation in the presence of side information is characterized, where we assume the side information per node has a fixed dimension. When the LLRs are fixed across n, tight necessary and sufficient conditions are calculated for weak recovery. Furthermore, it is shown that when belief propagation fails, no local algorithm can achieve weak recovery. It is also shown than belief propagation is strictly inferior to the maximum likelihood detector. Numerical results on finite synthetic data-sets validate our asymptotic analysis and show the relevance of our asymptotic results to even graphs of moderate size. We also calculate conditions under which belief propagation followed by a local voting procedure achieves exact recovery. When the side information has variable LLR across n, the belief propagation misclassification rate was calculated using density evolution. The results of this chapter were published in (Saad and Nosratinia, 2018,b,a).

Chapter 5 proposes a new tool, namely EXIT, for the analysis of the performance of local message passing algorithms, e.g., belief propagation, for community detection with side information. EXIT analysis has been used to understand the behavior of iterative algorithms (Ten Brink, 2001) in the context of error control and communication systems. By observing the EXIT chart, one can predict whether the decoder will fail, deduce the approximate number of iterations needed to decode, as well as approximate error probability after decoding. EXIT charts also have the additional benefit of an information theoretic interpretation (Ten Brink, 2001). We apply EXIT analysis to single-community detection as well as to binary symmetric community detection, each with side information, and leverage this technique to provide insights on: 1) The effect of the quality and quantity of side information on the performance of belief propagation, e.g. probability of error, 2) The asymptotic threshold for weak recovery, achieving a vanishing residual error, 3) The performance of belief propagation near the optimal threshold, 4) The performance of belief propagation through the first few iterations, and 5) Approximating the number of iterations needed for convergence. The results of this chapter were published in (Saad et al., 2016; Saad and Nosratinia, 2019). We conclude our contributions in Chapter 6.

CHAPTER 2

TWO SYMMETRIC COMMUNITIES WITH DISCRETE-VALUED SIDE INFORMATION [1] [2]

2.1 System Model and Assumptions

In this chapter, we consider the binary symmetric stochastic block model, with community labels denoted 1 and -1. The number of nodes in the graph is denoted with n. The node labels are independent and identically distributed across n, with 1 and -1 labels having equal probability. If two nodes belong to the same community, there is an edge between them with probability $p = a\frac{\log(n)}{n}$, and if they are from different communities, there is an edge between them with probability $q = b\frac{\log(n)}{n}$. Finally, for each node one or more scalar random variables are observed containing side information. Three models for this side information are considered.

First, for each node, a scalar side information is independently observed which is the true label with probability $(1-\alpha)$ or the negative of the true label with probability α, for $\alpha \in (0, 0.5)$. For the second model, for each node, a scalar side information is independently observed which is the true label with probability $1-\epsilon$ or 0 (erased) with probability ϵ, for $\epsilon \in (0, 1)$. For the third model, we consider side information consisting of K random variables (features) each has arbitrary finite and fixed cardinality M_k, $k \in \{1, \cdots, K\}$.

---

[1] © 2017 IEEE H. Saad and A. Abotabl and A. Nosratinia, "Exact recovery in the binary stochastic block model with binary side information," 2017 IEEE 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 822-82, 2017.

[2] © 2018 IEEE H. Saad and A. Nosratinia, "Community Detection With Side Information: Exact Recovery Under the Stochastic Block Model," 2018 IEEE Journal of Selected Topics in Signal Processing, vol. 12, pp. 944-958, 2018.

The observed graph is denoted by $G = (V, E)$, the vector of nodes' true assignment by $\boldsymbol{x}^*$, and the vector of nodes side information by $\boldsymbol{y_k}$ for feature $k \in \{1, \cdots, K\}$. The goal is to recover the node assignment $\boldsymbol{x}^*$ from the observation of $(G, \boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_K})$[3].

In this chapter, exact recovery is considered in the dense regime, i.e., when $p = a\frac{\log n}{n}$ and $q = b\frac{\log n}{n}$ with constants $a \geq b > 0$. We investigate the question: when and by how much can side information affect the phase transition threshold of exact recovery?[4]

2.2 Summary of Results

- When side information consists of observing node labels with erasure probability $\epsilon \in (0, 1)$, we show that if $\log(\epsilon) = o(\log(n))$, the phase transition is not improved by side information. On the other hand, if $\log(\epsilon) = -\beta \log(n) + o(\log(n))$ for some $\beta > 0$, i.e., $O(\log(n))$, a necessary and sufficient condition for exact recovery is $(\sqrt{a}-\sqrt{b})^2+2\beta > 2$.

- When side information consists of observing node labels with error probability $\alpha \in (0, 0.5)$, if $c = \log(\frac{1-\alpha}{\alpha})$ is $o(\log(n))$, then the phase transition is not improved by side information. On the other hand, if $c = \beta \log(n) + o(\log(n)), \beta > 0$, i.e., $O(\log(n))$, necessary and sufficient conditions for exact recovery are derived as follows:

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

where the following parameters are defined for convenience:

$$\eta(a, b, \beta) \triangleq a + b + \beta - \frac{2\gamma}{T} + \frac{\beta}{T}\log(\frac{\gamma+\beta}{\gamma-\beta}) \tag{2.1}$$

$$T \triangleq \log(\frac{a}{b}), \quad \gamma \triangleq \sqrt{\beta^2 + abT^2} \tag{2.2}$$

[3]For the special case where side information consists of noisy labels or partially revealed labels, $K = 1$ and hence, the subscript is omitted.

[4]The exact recovery phase transition without side information is $(\sqrt{a} - \sqrt{b})^2 > 2$ (Abbe et al., 2016).

- When side information consists of K features each with finite and fixed cardinality, two scenarios are considered: (1) K is fixed while the conditional distribution of each feature varies with n. In this scenario, we study how the quality of each feature must evolve as the size of the graph grows, so that phase transition can be improved. (2) K varies with n while the conditional distribution of features is fixed. In this scenario, the quality of the features is independent of n, and we study how many features are needed in addition to the graphical information, so that the phase transition can be improved.

Remark 1. *In earlier community detection problems (Abbe et al., 2016; Abbe and Sandon, 2015), LLRs do not depend on n even though individual likelihoods (obviously) do. This was very fortunate for calculating asymptotics. In the presence of side information, this convenience disappears and LLRs will now depend on n, creating complications in bounding error event probabilities en route to finding a threshold that must be independent of n. Overcoming this technical difficulty is part of the contributions.*

To illustrate the results of this chapter, Figures 2.1, 2.2 show the error exponent for the side information consisting of partially revealed labels or noisy label observation, as a function of β. It is observed that the value of β needed for recovery depends on a, b. For the partially revealed labels, when $(\sqrt{a} - \sqrt{b})^2 < 2$, the critical β is $1 - \frac{1}{2}(\sqrt{a} - \sqrt{b})^2$. For noisy label observations, when $(\sqrt{a} - \sqrt{b})^2 < 2$, the value of critical β can be determined as follows: if $\eta(a, b, \frac{T(a-b)}{2}) > 2$, then the critical β is the solution to $\eta = 2$. On the other hand, if $\eta(a, b, \frac{T(a-b)}{2}) < 2$, then the critical β is one.

2.3 Noisy-label Side Information

In this section, side information consisting of noisy labels is considered, where side information is incorrect with probability $\alpha \in (0, 0.5)$. Tight necessary and sufficient conditions are provided as a function of α.

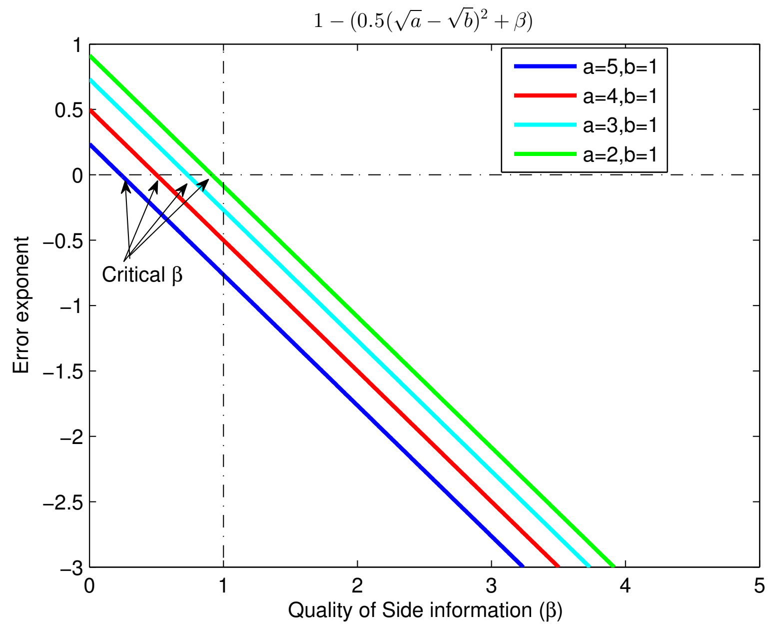

Figure 2.1. Error exponent for noisy label observations as a function of β.

First, the maximum likelihood rule for detecting the communities under side information is presented. It is known that without side information, in the binary symmetric stochastic block model, the maximum likelihood detector will find two communities by minimizing the number of edges between communities (min-cut) subject to each being of size $\frac{n}{2}$ (Abbe et al., 2016). The set of nodes belonging to the two communities are denoted with A and B, i.e., $A \triangleq \{i : x_i = 1\}$ and $B \triangleq \{i : x_i = -1\}$. $E(A)$ denotes the subset of edges whose two vertices belong to community A, and $E(B)$ the subset of edges whose two vertices belong to community B. The total number of edges in the graph is denoted E_t. Also, define:

$$J_+(A) \triangleq \big|\{i \in A : y_i = 1\}\big|$$
$$J_-(B) \triangleq \big|\{i \in B : y_i = -1\}\big|$$

Then, the log-likelihood function can be written as:

$$\log\big(\mathbb{P}(G, \boldsymbol{y}|\boldsymbol{x})\big) \overset{(a)}{=} \log\big(\mathbb{P}(G|\boldsymbol{x})\big) + \log\big(\mathbb{P}(\boldsymbol{y}|\boldsymbol{x})\big)$$

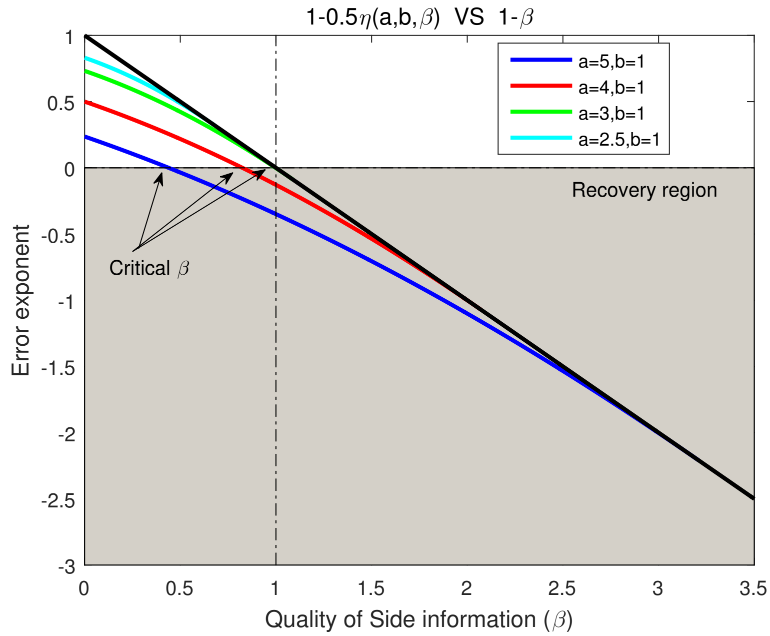

Figure 2.2. Error exponent of partial label observation as a function of β.

$$
\begin{aligned}
&= \log \Bigg(p^{E(A)+E(B)} q^{E_t - E(A) - E(B)} (1-p)^{2\binom{\frac{n}{2}}{2} - E(A) - E(B)} \\
&\quad (1-q)^{\frac{n^2}{4} - E_t + E(A) + E(B)} \Bigg) + \log \Big((1-\alpha)^{J_+(A)+J_-(B)} \alpha^{n - J_+(A) - J_-(B)} \Big) \\
&\overset{(b)}{=} R + T\big(E(A) + E(B)\big)(1 + o(1)) + c\big(J_+(A) + J_-(B)\big) \qquad (2.3)
\end{aligned}
$$

where (a) holds because $G, \boldsymbol{y}$ are independent given $\boldsymbol{x}$. In (b), all terms that are independent of $\boldsymbol{x}$ have been collected into a constant R, and $\log(\frac{p(1-q)}{q(1-p)})$ has been approximated by $(1+o(1))T$, which is made possible because $(1-p), (1-q)$ both approach 1 as $n \to \infty$. The difference between Eq. (2.3) and the likelihood function without side information is the term $c\big(J_+(A) + J_-(B)\big)$ and a constant $n \log \alpha$ that is hidden inside R.

The following lemma characterizes a lower bound on the probability of failure of the maximum likelihood detector. Let $E[\cdot, \cdot]$ denote the set of edges between two sets of nodes.[5]

[5]For economy of notation, in the arguments of $E[\cdot, \cdot]$ we allow singleton sets to be represented by their singular member.

Lemma 1. *Let A and B denote the true communities. Define the following events:*

$$\begin{aligned} F &\triangleq \{\textit{Maximum Likelihood fails}\} \\ F_A &\triangleq \{\exists i \in A : T(E[i,B] - E[i,A]) - cy_i \geq T\} \\ F_B &\triangleq \{\exists j \in B : T(E[j,A] - E[j,B]) + cy_j \geq T\} \end{aligned} \tag{2.4}$$

Then, $F_A \cap F_B \Rightarrow F$.

Proof. Define two new communities $\hat{A} = A \backslash \{i\} \cup \{j\}$ and $\hat{B} = B \backslash \{j\} \cup \{i\}$. If $\log\big(\mathbb{P}(G, \boldsymbol{y}|\hat{A}, \hat{B})\big) \geq \log\big(\mathbb{P}(G, \boldsymbol{y}|A, B)\big)$ it means maximum likelihood chooses incorrectly and therefore fails. We show that this happens under $F_A \cap F_B$.

Let $A_{ij} \sim Bern(q)$ be a random variable representing the existence of the edge between nodes i and j. Then, using (2.3):

$$\begin{aligned} \log\big(\mathbb{P}(G, \boldsymbol{y}|\hat{A}, \hat{B})\big) &= R + T\big(E(\hat{A}) + E(\hat{B})\big) + c\big(J_+(\hat{A}) + J_-(\hat{B})\big) \\ &= R + T\big(E(A) + E(B)\big) + c\big(J_+(A) + J_-(B)\big) - 2TA_{ij} \\ &\quad + T\big(E[j,A] - E[j,B] + E[i,B] - E[i,A]\big) + c(y_j - y_i) \\ &\overset{(a)}{\geq} \log\big(\mathbb{P}(G, \boldsymbol{y}|A, B)\big) + 2T(1 - A_{ij}) \\ &\overset{(b)}{\geq} \log\big(\mathbb{P}(G, \boldsymbol{y}|A, B)\big) \end{aligned} \tag{2.5}$$

where (a) holds by the assumption that $F_A \cap F_B$ happened and (b) holds because $(1 - A_{ij}) \geq 0$ and $T \geq 0$. The inequality (b) implies the failure of maximum likelihood. □

2.3.1 Necessary Conditions

Theorem 1. *Define $c \triangleq \log(\frac{1-\alpha}{\alpha})$. The maximum likelihood failure probability is bounded away from zero if:*

$$\begin{cases} (\sqrt{a} - \sqrt{b})^2 < 2, \;\; \textit{when } c = o(\log(n)) \\ \eta(a, b, \beta) < 2 \;\; \textit{when } c = (\beta + o(1))\log(n), 0 < \beta < \frac{T(a-b)}{2} \\ \beta < 1 \;\; \textit{when } c = (\beta + o(1))\log(n), \beta > \frac{T(a-b)}{2} \end{cases}$$

Proof. Since $\boldsymbol{x}^*$ is generated uniformly, then the ML detector minimizes the probability of error over all possible estimators. Hence, if the probability of failure of ML is bounded away from zero, then every other estimator has probability of failure bounded away from zero. The main difficulty in bounding the error probability of the ML is the dependency between the graph edges. To overcome this dependency, we follow similar steps as in (Abbe et al., 2016). However, our bounding techniques, unlike (Abbe et al., 2016), involve Chernoff type arguments and Cramer and Sanov large deviation principles that are more compact than combinatorial techniques used in (Abbe et al., 2016). Let H be a subset of A with $|H| = \frac{n}{\log^3(n)}$ and define the following auxiliary events:

$$\triangle_i = \big\{ i \in H : E[i, H] \leq \frac{\log(n)}{\log\log(n)} \big\}$$

$$F_i^H = \big\{ i \in H : T * E[i, A \backslash H] + c y_i + T + T\frac{\log(n)}{\log\log(n)} \leq T * E[i, B] \big\}$$

$$\triangle = \big\{ \forall i \in H : \triangle_i \text{ is true} \big\}$$

$$F^H = \big\{ \cup_{i \in H} F_i^H \big\}$$

Lemma 2. *If* $\mathbb{P}(F^H) \geq 1 - \delta$ *and* $\mathbb{P}(\triangle) \geq 1 - \delta$ *for* $\delta < \frac{1}{4}$, *then there exists a positive* δ' *so that* $\mathbb{P}(F) \geq \delta'$.

Proof. Clearly $\triangle \cap F^H \Rightarrow F_A$. Hence,

$$\mathbb{P}(F_A) \geq \mathbb{P}(F^H) + \mathbb{P}(\triangle) - 1 \geq 1 - 2\delta$$

By the symmetry of the graph and the side information, $\mathbb{P}(F_B) \geq 1 - 2\delta$ as well. Also, by Lemma 1 $F_A \cap F_B \Rightarrow F$. Then, the following holds:

$$\mathbb{P}(F) \geq \mathbb{P}(F_A) + \mathbb{P}(F_B) - 1 \geq 1 - 4\delta$$

For $\delta < \frac{1}{4}$, $\mathbb{P}(F)$ is bounded away from zero. $\square$

Lemma 3. $\lim_{n \to \infty} \mathbb{P}(\triangle) = 1$

Proof. Let $W_i \sim Bern(p)$. Then, the following holds:

$$\begin{aligned}
\mathbb{P}(\triangle_i^c) &= \mathbb{P}\Big(\sum_{j=1}^{i-1} W_j + \sum_{j=i+1}^{\frac{n}{\log^3(n)}} W_j \geq \frac{\log(n)}{\log(\log(n))}\Big) \\
&\leq \mathbb{P}\Big(\sum_{j=1}^{\frac{n}{\log^3(n)}} W_j \geq \frac{\log(n)}{\log(\log(n))}\Big) \\
&\overset{(a)}{\leq} \Big(\frac{1}{e}\frac{\log^3(n)}{a\log(\log(n))}\Big)^{\frac{-\log(n)}{\log(\log(n))}}
\end{aligned}$$

where (a) holds from a multiplicative form of Chernoff bound, which states that for a sequence of n i.i.d random variables X_i, $\mathbb{P}(\sum_{i=1}^n X_i \geq t\mu) \leq (\frac{t}{e})^{-t\mu}$, where $\mu = n\mathbb{E}[X]$. Thus, by union bound:

$$\begin{aligned}
\mathbb{P}(\triangle) &\geq 1 - \frac{n}{\log^3(n)}\Big(\frac{1}{e}\frac{\log^3(n)}{a\log(\log(n))}\Big)^{\frac{-\log(n)}{\log(\log(n))}} \\
&= 1 - e^{\log(n) - 3\log(\log(n))} e^{\Big[\frac{\log(n)\log(ae)}{\log(\log(n))} - \frac{\log(n)}{\log(\log(n))}\Big(3\log(\log(n)) - \log(\log(\log(n)))\Big)\Big]} \\
&= 1 - e^{-2\log(n) + o(\log(n))}
\end{aligned}$$

$\square$

Lemma 4. *For any $\delta \in (0,1)$ and for sufficiently large n, if $\mathbb{P}(F_i^H) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$, then $\mathbb{P}(F^H) \geq 1 - \delta$.*

Proof. Because F_i^H are i.i.d.:

$$\begin{aligned}
\mathbb{P}(F^H) &= \mathbb{P}(\cup_{i\in H} F_i^H) = 1 - \mathbb{P}(\cap_{i\in H}(F_i^H)^c) \\
&= 1 - \Big[\big(1 - \mathbb{P}(F_i^H)\big)^{\frac{1}{\mathbb{P}(F_i^H)}}\Big]^{(\frac{n\mathbb{P}(F_i^H)}{\log^3(n)})} \qquad (2.6) \\
&> 1 - \Big[\big(1 - \mathbb{P}(F_i^H)\big)^{\frac{1}{\mathbb{P}(F_i^H)}}\Big]^{-\log\delta}
\end{aligned}$$

where the last inequality holds by the statement of the Lemma. If $\mathbb{P}(F_i^H)$ is $o(1)$, then the quantity inside the bracket tends to e^{-1} and the result follows. If $\mathbb{P}(F_i^H)$ is not $o(1)$, then from Eq. (2.6) it follows that $\mathbb{P}(F^H) \to 1$ and again the result of the Lemma holds. $\square$

The following lemma completes the proof of Theorem 1.

Lemma 5. *For sufficiently large n, $\mathbb{P}(F_i^H) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ for $\delta \in (0,1)$, if one of the following is satisfied:*

$$\begin{cases} (\sqrt{a}-\sqrt{b})^2 < 2 & \textit{when } c = o(\log(n)) \\ \eta(a,b,\beta) < 2 & \textit{when } c = (\beta + o(1))\log(n), 0 < \beta < \frac{T(a-b)}{2} \\ \beta < 1 & \textit{when } c = (\beta + o(1))\log(n), \beta > \frac{T(a-b)}{2} \end{cases}$$

Proof. See Appendix 2.6.1. ☐

Combining Lemmas 2, 3, 4, 5 concludes the proof of the theorem. ☐

2.3.2 Sufficient Conditions

This section shows sufficient conditions for exact recovery by introducing an algorithm whose exact recovery conditions are identical to Section 2.3.1. The first stage of the proposed algorithm uses a component from (Massoulié, 2014), a method based on spectral properties of the graph that achieves weak recovery. We start with a random graph H_1 built on the same n nodes where each candidate edge has probability $\frac{D}{\log(n)}$. The complement of H_1 is denoted H_2. Then G is partitioned as follows: $G_1 = G \cap H_1$ and $G_2 = G \cap H_2$. G_1 will be used for the weak recovery step, G_2 for local modification. The partitioning of G allows the two steps to remain independent.

We perform a weak recovery algorithm (Massoulié, 2014) on G_1. Since G_1 is a graph with connectivity parameters $(\frac{Da}{n}, \frac{Db}{n})$, the weak recovery algorithm is guaranteed to return two communities A', B' that agree with the true communities A, B on at least $(1-\delta(D))n$ nodes so that $\lim_{D\to\infty}\delta(D) = 0$ (i.e., weak recovery). A sufficient condition for that to happen (Massoulié, 2014), e.g., is $D = O(\log\log n)$.

The community assignments are locally modified as follows: for a node $i \in A'$, flip its membership if the number of G_2 edges between i and B' is greater than or equal the number

Table 2.1. Algorithm for exact recovery.

Algorithm 1
1: Start with graph G and side information y
2: Generate an Erdös-Renyi graph H_1 with edge probability $\frac{D}{\log(n)}$. Use it to partition G into $G_1 = G \cap H_1$ and $G_2 = G \cap H_1^c$.
3: Apply weak recovery algorithm (Massoulié, 2014) on G_1, calling the resulting communities A'/B'.
4: Initialize $\tilde{A} \leftarrow A'$ and $\tilde{B} \leftarrow B'$.
5: For each node i modify $\tilde{A}$ and $\tilde{B}$ as follows: Flip membership if $i \in \tilde{A}$ and $E_{G_2}[i, \tilde{B}] \geq E_{G_2}[i, \tilde{A}] + \frac{c}{T} y_i$ Flip membership if $i \in \tilde{B}$ and $E_{G_2}[i, \tilde{A}] \geq E_{G_2}[i, \tilde{B}] - \frac{c}{T} y_i$
6: Check size of communities. If $\lvert A' \rvert \neq \lvert \tilde{A} \rvert$ or equivalently $\lvert B' \rvert \neq \lvert \tilde{B} \rvert$, discard changes via $\tilde{A} \leftarrow A'$ and $\tilde{B} \leftarrow B'$.

of G_2 edges between i and $A^{'}$ plus $\frac{c}{T} y_i$. For node $j \in B^{'}$, flip its membership if the number of G_2 edges between j and $A^{'}$ is greater than or equal the number of G_2 edges between j and $B^{'}$ minus $\frac{c}{T} y_j$. If the number of flips in the two clusters are not the same, keep the clusters unchanged. The detailed algorithm is shown in Table 2.1.

Theorem 2. *With probability approaching one as n grows, the algorithm above successfully recovers the communities if:*

$$\begin{cases} (\sqrt{a} - \sqrt{b})^2 > 2, \ \ \textit{when } c = o(\log(n)) \\ \eta(a, b, \beta) > 2 \ \textit{when } c = (\beta + o(1)) \log(n), 0 < \beta < \frac{T(a-b)}{2} \\ \beta > 1 \ \textit{when } c = (\beta + o(1)) \log(n), \beta > \frac{T(a-b)}{2} \end{cases}$$

Proof. We first upper bound the misclassification probability of a node assuming H_2 is a complete graph, then adjust the bound to reflect the departure of H_2 from a complete graph.

Figure 2.3 shows the mis-classification conditions: an error happens either when the weak recovery was correct and is overturned by the local modification, or when the weak recovery is incorrect and is *not* corrected by local modification. Let $W \sim Bern(p)$ and $Z \sim Bern(q)$

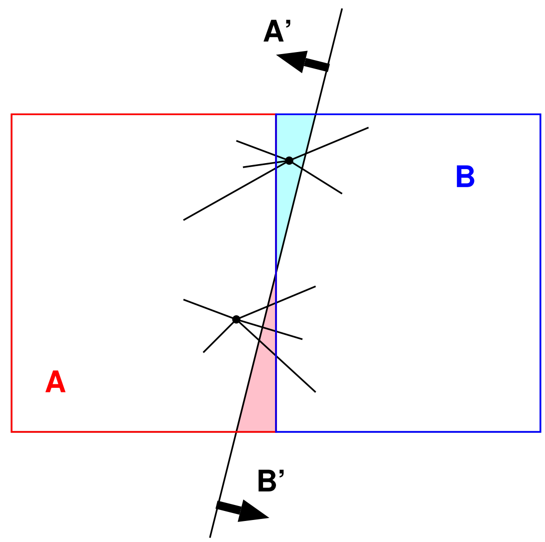

Figure 2.3. Two types of error events for the two-stage algorithm. The node in the top half of the figure is misclassified in weak recovery, and remains uncorrected via local modification. The node at the bottom half is correctly classified in weak recovery, but is mistakenly flipped by local modification.

represent edges inside a community and across communities, respectively. Let $y_i \in \{1, -1\}$ with probabilities $(1-\alpha), \alpha$, respectively. For simplicity, we will write δ instead of $\delta(D)$. Then, the mis-classification probability is:

$$\begin{aligned} P_e &= \mathbb{P}\big(\text{node } i \text{ is mislabeled}\big) \\ &= \mathbb{P}\bigg(\sum_{k=1}^{(1-\delta)\frac{n}{2}} Z_k + \sum_{k=1}^{\delta\frac{n}{2}} W_k \geq \sum_{j=1}^{(1-\delta)\frac{n}{2}} W_j + \sum_{j=1}^{\delta\frac{n}{2}} Z_j + \frac{c}{T} y_i\bigg) \end{aligned} \tag{2.7}$$

To adjust for the fact that H_2 is not complete, the following Lemma is used, noting that $H_2 = H_1^c$.

Lemma 6. *With high probability, the degree of any node in H_1 is at most $\frac{2Dn}{\log(n)}$.*

Proof. Let $\{Y_i\}_{i=1,\cdots,n}$ be a sequence of i.i.d. Bernoulli random variables with parameter $\frac{D}{\log(n)}$. Define $Y = \sum_{i=1}^{n-1} Y_i$. Then, $\mathbb{E}[Y] = \frac{Dn}{\log(n)}$ and hence, by Chernoff bound:

$$\mathbb{P}(Y \geq \frac{2Dn}{\log(n)}) \leq e^{-\frac{1}{4}\frac{D}{\log(n)}n} \tag{2.8}$$

Thus, by using a union bound:

$$\begin{aligned}\mathbb{P}\left(\exists \text{ a node degree} > \frac{2Dn}{\log(n)}\right) &\leq n\mathbb{P}\left(Y \geq \frac{2Dn}{\log(n)}\right) \\ &\leq e^{-\frac{1}{4}\frac{D}{\log(n)}n+\log(n)} \to 0 \end{aligned} \tag{2.9}$$

where the last statement holds as $n \to \infty$. □

Having bounded from below the degree of H_2, the correct error probability (for the incomplete H_2) can be arrived at by removing no more than $\frac{2D}{\log(n)}n$ terms from the summations on the right hand side of (2.7). If we remove exactly $\frac{2D}{\log(n)}n$ terms, the following upper bound on the error probability holds:

$$P_e \leq \mathbb{P}\left(\sum_{k=1}^{(1-\delta)\frac{n}{2}} Z_k + \sum_{k=1}^{\delta\frac{n}{2}} W_k \geq \sum_{j=1}^{(1-\delta)\frac{n}{2}-\frac{2D}{\log(n)}n} W_j + \sum_{j=1}^{\delta\frac{n}{2}-\frac{2D}{\log(n)}n} Z_j + \frac{c}{T}y_i\right) \tag{2.10}$$

The following lemma shows an upper bound on P_e.

Lemma 7.

$$P_e \leq \begin{cases} n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2+o(1)} + n^{-(1+\Omega(1))} & \textit{when } c = o(\log(n)) \\ n^{-\frac{1}{2}\eta(a,b,\beta)+o(1)} + n^{-(1+\Omega(1))} & \textit{when } c = (\beta+o(1))\log(n)\ ,\ 0 < \beta < \frac{T(a-b)}{2} \\ n^{-\frac{1}{2}\eta(a,b,\beta)+o(1)} + n^{-\beta} + n^{-(1+\Omega(1))} & \textit{when } c = (\beta+o(1))\log(n)\ ,\ \beta > \frac{T(a-b)}{2} \end{cases}$$

Proof. See Appendix 2.6.2. □

A simple union bound yields:

$$\mathbb{P}(\text{failure}) \leq \begin{cases} n^{1-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2+o(1)}, & \text{when } c = o(\log(n)) \\ n^{1-\frac{1}{2}\eta(a,b,\beta)+o(1)}, & \text{when } c = \beta\log(n)\ ,\ 0 < \beta < \frac{T(a-b)}{2} \\ n^{1-\beta+o(1)}, & \text{when } c = \beta\log(n)\ ,\ \beta > \frac{T(a-b)}{2} \end{cases} \tag{2.11}$$

For the last case, $\beta > 1$ remains sufficient because of the following lemma.

Lemma 8. $\beta > 1 \Rightarrow \eta > 2$.

Proof. Let $a + b - \beta - 2\frac{\gamma}{T} + \frac{\beta}{T}\log(\frac{\gamma+\beta}{\gamma-\beta}) = \psi(a,b,\beta)$. Then, from the definition of η:

$$\eta(a,b,\beta) - 2\beta = \psi(a,b,\beta) \tag{2.12}$$

Since $\psi(a,b,\beta)$ is convex in β, it can be shown that at the optimal β^*, $\log(\frac{\gamma^*+\beta^*}{\gamma^*-\beta^*}) = T$. Using this fact and by substituting in (2.12), the following holds:

$$\eta(a,b,\beta) - 2\beta \geq a + b - 2\frac{\gamma^*}{T} \tag{2.13}$$

By the definition of γ: $\gamma + \beta = \frac{abT^2}{\gamma-\beta}$. Using the fact that $\frac{\gamma^*+\beta^*}{\gamma^*-\beta^*} = \frac{a}{b}$ leads to $\frac{a}{b} = \frac{abT^2}{(\gamma^*-\beta^*)^2}$, which implies that $\gamma^* = bT + \beta^*$. Hence, by substituting in (2.13), the following holds:

$$\eta(a,b,\beta) - 2\beta \geq a - b - 2\frac{\beta^*}{T} \tag{2.14}$$

Also, it can be shown that at β^*, $\gamma^* = \beta^*(\frac{a+b}{a-b})$. This implies that $\beta^* = \frac{T(a-b)}{2}$. Substituting in (2.14) leads to: $\eta(a,b,\beta) - 2\beta \geq 0$, which implies that $\eta > 2$ when $\beta > 1$. $\square$

Combining the last lemma with (2.11) concludes the proof. $\square$

2.4 Partially Revealed Labels

In this section, we consider side information consisting of partially revealed labels, where $\epsilon \in (0,1)$ is the proportion of labels that remains unknown despite the side information. Tight necessary and sufficient conditions are presented for exact recovery under this type of side information. Similar to the noisy label side information, we begin by expressing the log-likelihood function. For a given side information vector $\boldsymbol{y}$, $\mathbb{P}(\boldsymbol{y}|\boldsymbol{x}) = 0$ if a label contradicts the side information.[6] All label vectors $\boldsymbol{x}$ that do not contradict side information and satisfy

[6] We say a label contradicts the side information if the side information is not an erasure and it disagrees with the label.

the balanced prior, have the same conditional probability. Thus, for all $\boldsymbol{x}$ that have non-zero conditional probability, the log-likelihood function can be written as:

$$\begin{aligned}\log\big(\mathbb{P}(G,\boldsymbol{y}|\boldsymbol{x})\big) &\stackrel{(a)}{=} \log\big(\mathbb{P}(G|\boldsymbol{x})\big) + \log\big(\mathbb{P}(\boldsymbol{y}|\boldsymbol{x})\big) \\ &\stackrel{(b)}{=} R + T\big(E(A)+E(B)\big)(1+o(1)) \end{aligned} \tag{2.15}$$

where (a) holds because $G, \boldsymbol{y}$ are independent given $\boldsymbol{x}$. In (b), all terms that are independent of $\boldsymbol{x}$ have been collected into a constant R, and $\log(\frac{p(1-q)}{q(1-p)})$ has been approximated by $(1+o(1))T$, which is made possible because $(1-p), (1-q)$ both approach 1 as $n \to \infty$.

The following lemma shows that if the graph includes at least one pair of nodes that have more connections to the opposite-labels than similar-labels *and* if their side information is an erasure, the maximum likelihood detector will fail.

Lemma 9. *Define the following events:*

$$F_A = \{\exists i \in A : (E[i,B] - E[i,A]) \geq 1 \textit{ and } y_i = 0\}$$

$$F_B = \{\exists j \in B : (E[j,A] - E[j,B]) \geq 1 \textit{ and } y_j = 0\}$$

Then, $F_A \cap F_B \Rightarrow F$.

Proof. From the sets A, B, we swap the nodes i, j, producing $\hat{A} = A\backslash\{i\} \cup \{j\}$ and $\hat{B} = B\backslash\{j\} \cup \{i\}$. We intend to show that subject to observing the graph G and the side information $\boldsymbol{y}$, the likelihood of $\hat{A}, \hat{B}$ is larger than the likelihood of A, B, therefore under the condition $F_A \cap F_B$, maximum likelihood will fail.

Let $A_{ij} \sim Bern(q)$ be a random variable representing the existence of the edge between nodes i and j. Then, from (2.15) the following holds:

$$\begin{aligned}\log\big(\mathbb{P}(G,\boldsymbol{y}|\hat{A},\hat{B})\big) &= R + T(1+o(1))\big(E(\hat{A})+E(\hat{B})\big) \\ &= R + T(1+o(1))\big(E(A)+E(B)\big) + T(1+o(1))\end{aligned}$$

$$
\begin{aligned}
&\times \big(E[j,A] - E[i,A] - E[j,B] + E[i,B] - 2A_{ij}\big) \\
&\overset{(a)}{\geq} \log\big(\mathbb{P}(G,\boldsymbol{y}|A,B)\big) + 2T(1+o(1))(1-A_{ij}) \\
&\overset{(b)}{\geq} \log\big(\mathbb{P}(G,\boldsymbol{y}|A,B)\big) \qquad (2.16)
\end{aligned}
$$

where (a) holds by the assumption that $F_A \cap F_B$ happened and (b) holds because $(1-A_{ij}) \geq 0$ and $T \geq 0$. The inequality (b) implies the failure of maximum likelihood. $\square$

2.4.1 Necessary Conditions

Theorem 3. *The maximum likelihood failure probability is bounded away from zero if:*

- $\log(\epsilon) = o(\log(n))$ *and* $(\sqrt{a} - \sqrt{b})^2 < 2$
- $\log(\epsilon) = -(\beta + o(1))\log(n)$, $\beta > 0$, *and* $\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + \beta < 1$

Proof. Let H be a subset of A with $|H| = \frac{n}{\log^3(n)}$. Define the following events:

$$
\begin{aligned}
&\triangle_i = \Big\{ i \in H : E[i,H] \leq \frac{\log(n)}{\log\log(n)} \Big\} \\
&F_i^H = \big\{ i \in H : y_i = 0 \text{ and} \\
&E[i, A\backslash H] + 1 + \frac{\log(n)}{\log\log(n)} \leq E[i,B] \big\} \\
&\triangle = \big\{ \forall i \in H : \triangle_i \text{ is true} \big\} \\
&F^H = \big\{ \cup_{i\in H} F_i^H \big\}
\end{aligned}
$$

Lemmas 2, 3, 4 directly apply with the above definitions. To complete the proof, it is sufficient to show when $\mathbb{P}(F_i^H) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ for $\delta \in (0,1)$ holds, which is shown in the following lemma.

Lemma 10. *For sufficiently large* n, $\mathbb{P}(F_i^H) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ *for* $\delta \in (0,1)$, *if one of the following is satisfied:*

$$
\begin{cases}
(\sqrt{a} - \sqrt{b})^2 < 2, & \textit{when } \log(\epsilon) = o(\log(n)) \\
(\sqrt{a} - \sqrt{b})^2 + 2\beta < 2, & \textit{when } \log(\epsilon) = -(\beta + o(1))\log(n), \beta > 0
\end{cases}
$$

Proof. See Appendix 2.6.3. □

Combining Lemmas 2, 3, 4, 10 concludes the proof of the theorem. □

2.4.2 Sufficient Conditions

This section shows sufficient conditions for exact recovery by introducing an algorithm whose exact recovery conditions are identical to Section 2.4.1. The first stage of the algorithm is the same as Section 2.3.2. The second stage involving local modification is new and is described below.

The community assignments are locally modified for each node i as follows: (a) if A'/B' membership contradicts side information y_i, flip node membership or (b) if $y_i = 0$, re-assign membership of i to the community A'/B' to which it is connected with more edges. After going through all nodes, if the the number of flips in two communities A', B' are not the same, void all local modifications.

Theorem 4. *The algorithm described above successfully recovers the communities with high probability if:*

$$\begin{cases} (\sqrt{a}-\sqrt{b})^2 > 2, \ \textit{when} \ \log(\epsilon) = o(\log(n)) \\ (\sqrt{a}-\sqrt{b})^2 + 2\beta > 2, \qquad \textit{when} \ \log(\epsilon) = -(\beta + o(1))\log(n), \beta > 0 \end{cases}$$

Proof. Let $P_e = \mathbb{P}(\text{node } i \text{ to be misclassified})$. Following the same analysis as in the proof of Theorem 2 leads to:

$$P_e \leq \epsilon \mathbb{P}\Bigg(\sum_{k=1}^{(1-\delta)\frac{n}{2}} Z_k + \sum_{k=1}^{\delta\frac{n}{2}} W_k \geq \sum_{j=1}^{(1-\delta)\frac{n}{2} - \frac{2D}{\log(n)}n} W_j + \sum_{j=1}^{\delta\frac{n}{2} - \frac{2D}{\log(n)}n} Z_j \Bigg) \tag{2.17}$$

Using Lemma 7 and strengthening $c = o(\log(n))$ to $c = 0$, equation (2.17) can be upper bounded as follows:

$$P_e \leq \epsilon n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2} + n^{-(1+\Omega(1))} \tag{2.18}$$

Thus, the following holds for different asymptotic regimes of ϵ:

$$P_e \leq \begin{cases} n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2+o(1)} + n^{-(1+\Omega(1))}, & \text{when } \log(\epsilon) = o(\log(n)) \\ n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2-\beta} + n^{-(1+\Omega(1))}, & \text{when } \log(\epsilon) = -(\beta+o(1))\log(n)\ ,\ \beta > 0 \end{cases}$$

A simple union bound yields:

$$\mathbb{P}(\text{failure}) \leq \begin{cases} n^{1-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2+o(1)}, & \text{when } \log(\epsilon) = o(\log(n)) \\ n^{1-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2-\beta+o(1)} & \text{when } \log(\epsilon) = -(\beta+o(1))\log(n)\ ,\ \beta > 0 \end{cases}$$

□

2.5 More General Discrete Side Information

We now generalize the side information random variable such that each node observes K features (side information) each has arbitrary fixed and finite cardinality $M_k, k \in \{1, \cdots, K\}$. The alphabet for each feature k is denoted with $\{u_1^k, u_2^k, \cdots, u_{M_k}^k\}$. Denote, for each node i and feature k, $\mathbb{P}(y_{i,k} = u_{m_k}^k | x_i = 1) = \alpha_{+,m_k}^k$ and $\mathbb{P}(y_{i,k} = u_{m_k}^k | x_i = -1) = \alpha_{-,m_k}^k, m_k \in \{1, \cdots, M_k\}$, where $\alpha_{+,m_k}^k \geq 0$, $\alpha_{-,m_k}^k \geq 0$ and $\sum_{m_k=1}^{M_k} \alpha_{+,m_k}^k = \sum_{m_k=1}^{M_k} \alpha_{-,m_k}^k = 1$ for all $k \in \{1, \cdots, K\}$. All features are assumed to be independent conditioned on the labels.

We first consider the case where K is fixed while α_{+,m_k}^k and α_{-,m_k}^k are varying with n for $m_k \in \{1, \cdots, M_k\}$ and $k \in \{1, \cdots, K\}$. To ensure that the quality of the side information is increasing with n, assume that α_{+,m_k}^k and α_{-,m_k}^k for $m_k \in \{1, \cdots, M_k\}$ and $k \in \{1, \cdots, K\}$ are constant or monotonic in n. Second, we consider the case where K is varying with n while α_{+,m_k}^k and α_{-,m_k}^k are fixed for $m_k \in \{1, \cdots, M_k\}$ and $k \in \{1, \cdots, K\}$. To ensure that the quality of the side information is increasing with n, assume that K is non-decreasing with n. Necessary and sufficient conditions for exact recovery that are tight except for one special case are provided. Due to space limitation and similarity with some results in previous sections, several proofs are provided as a sketch.

First the log-likelihood function is presented. For feature k, let the number of $\{i \in A : y_{i,k} = u^k_{m_k}\}$ and $\{i \in B : y_{i,k} = u^k_{m_k}\}$ be $J_{u^k_{m_k}}(A)$ and $J_{u^k_{m_k}}(B)$, respectively. Then, by using similar ideas as in (2.3), the following holds:

$$\log\Big(\mathbb{P}(G, \boldsymbol{y_1}, \boldsymbol{y_2}, \cdots, \boldsymbol{y_K}|\boldsymbol{x})\Big) = R + T\Big(E(A) + E(B)\Big)(1 + o(1)) + \sum_{k=1}^{K}\sum_{m_k=1}^{M_k} J_{u^k_{m_k}}(A)\log\left(\alpha^k_{+,m_k}\right) + J_{u^k_{m_k}}(B)\log\left(\alpha^k_{-,m_k}\right) \tag{2.19}$$

For convenience, define the following two quantities:

- $h^m_{\ell_m} = \log(\frac{\alpha^k_{+,m_k}}{\alpha^k_{-,m_k}})$ to be the log-likelihood ratio for the side information outcome $u^k_{m_k}$ of feature k.

- $h_{i,m} = h^k_{m_k}$ if $y_{i,k} = u^k_{m_k}$, i.e., $h_{i,k}$ is a random variable representing the LLR of the observation of node i for feature k.

Lemma 11. *Define the following events:*

$$F_A = \{\exists i \in A : T(E[i, B] - E[i, A]) - \sum_{k=1}^{K} h_{i,k} \geq T\}$$
$$F_B = \{\exists j \in B : T(E[j, A] - E[j, B]) + \sum_{k=1}^{K} h_{j,k} \geq T\}$$

Then, $F_A \cap F_B \Rightarrow F$.

Proof. The proof follows similarly as in Lemmas 1 and 9. □

2.5.1 Fixed Number of Features, Variable Quality

In this section, the number of features K is assumed to be fixed and we show how noisy the outcomes of the features should be so that side information changes the phase transition threshold of exact recovery. First the intuition behind the results are presented for the case when $K = 1$, i.e. one feature with M outcomes. To understand how side information will affect the phase transition of exact recovery, two main quantities have to be considered for each outcome $m \in \{1, \cdots, M\}$. The first quantity is the log-likelihood ratio $h_m = \log(\frac{\alpha_{+,m}}{\alpha_{-,m}})$ and the second is the conditional probability $\alpha_{\pm,m}$. An outcome is called *informative* if $h_m = O(\log(n))$[7] and *non-informative* if $h_m = o(\log(n))$. Also, an outcome is called *rare* if $\log(\alpha_{\pm,m}) = O(\log(n))$ and *not rare* if $\log(\alpha_{\pm,m}) = o(\log(n))$. Hence, four different combinations are possible. The *worst* case is when the outcome is both *non-informative* and *not rare* for both communities, e.g. noisy labels with $\alpha = \frac{1}{\log(n)}$. We will show that if such an outcome exists, then side information will not improve the phase transition threshold. The *best* case is when the outcome is *informative*, and *rare* for one community but *not rare* for the other. This happens, e.g., under noisy label side information with $\alpha = n^{-\beta+o(1)}$. We have two cases in between: (1) an outcome that is *non-informative* and *rare* for both communities, e.g. partial label reveal side information with $\epsilon = n^{-\beta+o(1)}$ and (2) an outcome that is *informative* and *not rare* for both communities. The last three cases can affect the phase transition threshold under certain conditions which is stated in the following theorem.

Theorem 5. *Assume α^k_{+,m_k} and α^k_{-,m_k} are either constant or monotonically increasing or decreasing in n. Then, necessary and sufficient conditions for exact recovery are as follows.*[8]

1. *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^K h^m_{\ell_m} = o(\log(n))$, $\sum_{k=1}^K \log(\alpha^k_{+,m_k}) = o(\log(n))$ and $\sum_{k=1}^K \log(\alpha^k_{-,m_k}) = o(\log(n))$, then $(\sqrt{a} - \sqrt{b})^2 > 2$ must hold.*

2. *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^K h^m_{\ell_m} = o(\log(n))$ and $\sum_{k=1}^K \log(\alpha^k_{+,m_k}) = \sum_{k=1}^K \log(\alpha^k_{-,m_k}) = -\beta \log(n) + o(\log(n)), \beta > 0$, then $(\sqrt{a} - \sqrt{b})^2 + 2\beta > 2$ must hold.*

[7] We say $h_m = O(\log n)$ when there exists a strictly positive constant C such that for all sufficiently large n, the following holds: $h_m < C \log(n)$.

[8] For clarity, in this theorem the side information outcomes $[u^1_{m_1}, \ldots, u^K_{m_K}]$ are represented by their index $[m_1, \ldots, m_K]$. Dependence on n is implicit.

3. *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^{K} h_{\ell_m}^{m} = \beta_1 \log(n) + o(\log(n))$, $|\beta_1| < T\frac{(a-b)}{2}$, $\sum_{k=1}^{K} \log(\alpha_{sgn(\beta_1),m_k}^{k}) = o(\log(n))$, then $\eta(a, b, |\beta_1|) > 2$ must hold.*

4. *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^{K} h_{\ell_m}^{m} = \beta_2 \log(n) + o(\log(n))$, $|\beta_2| < T\frac{(a-b)}{2}$, $\sum_{k=1}^{K} \log(\alpha_{sgn(\beta_2),m_k}^{k}) = -\beta_2' \log(n) + o(\log(n))$, then $\eta(a, b, |\beta_2|) + 2\beta_2' > 2$ must hold.*

Remark 2. *The four items in Theorem 5 must all hold. For example, if some outcomes fall under item 3 and some fall under item 4, then $\min(\eta(a, b, |\beta_1|), \eta(a, b, |\beta_2|) + 2\beta_2') > 2$ must hold.*

Remark 3. *When there is any sequence of side information outcomes that satisfies $\sum_{k=1}^{K} h_{\ell_m}^{m} = \beta \log(n) + o(\log(n))$ with $T\frac{(a-b)}{2} < |\beta|$, no matching necessary and sufficient conditions have been provided. A sufficient condition in this case easily follows other achievability proofs for Theorem 5, but a matching converse at this point remains unavailable.*

Proof. **Converse:** Unlike previous sections, the side information might not be symmetric. Hence, we need to define the events of Section 2.3.1 for both communities A and B. Let H_1 and H_2 be subsets of the true communities A and B, respectively, with $|H_1| = |H_2| = \frac{n}{\log^3(n)}$. Define the following events:

$$\begin{aligned}
\triangle_i^j &= \Big\{ i \in H_j : E[i, H_j] \leq \frac{\log(n)}{\log\log(n)} \Big\} \\
\triangle_j &= \Big\{ \forall i \in H_j : \triangle_i^j \text{ is true} \Big\} \\
F_i^{H_1} &= \Big\{ i \in H_1 : TE[i, A \backslash H_1] + \sum_{k=1}^{K} h_{i,k} + T + T\frac{\log(n)}{\log\log(n)} \leq TE[i, B] \Big\} \\
F_i^{H_2} &= \Big\{ i \in H_2 : TE[i, B \backslash H_2] - \sum_{k=1}^{K} h_{i,k} + T + T\frac{\log(n)}{\log\log(n)} \leq TE[i, A] \Big\} \\
F^{H_j} &= \Big\{ \cup_{i \in H_j} F_i^{H_j} \Big\}
\end{aligned} \tag{2.20}$$

where $j = 1, 2$ and $h_{i,k}$ is distributed according to α^k_{+,m_k} and α^k_{-,m_k} if node $i \in A, B$, respectively. Lemmas 2, 3, 4 extend directly to our case here using the above definitions for both communities A and B. To complete the proof, it is sufficient to show show when $\mathbb{P}(F_i^{H_1}) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ and $\mathbb{P}(F_i^{H_2}) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ hold for $\delta \in (0, 1)$.

Lemma 12. *Both $\mathbb{P}(F_i^{H_1})$ and $\mathbb{P}(F_i^{H_2})$ are greater than $\frac{\log^3(n)}{n}\log(\frac{1}{\delta})$, $\delta \in (0, 1)$ for sufficiently large n if at least one of the following conditions holds:*

- *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^K h^m_{\ell_m} = o(\log(n))$ and $\sum_{k=1}^K \log(\alpha^k_{+,m_k})$ and $\sum_{k=1}^K \log(\alpha^k_{-,m_k})$ are $o(\log(n))$ and concurrently $(\sqrt{a} - \sqrt{b})^2 < 2$.*

- *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^K h^m_{\ell_m} = o(\log(n)), \sum_{k=1}^K \log(\alpha^k_{+,m_k}) = \sum_{k=1}^K \log(\alpha^k_{-,m_k}) = -\beta\log(n) + o(\log(n)), \beta > 0$ and concurrently $(\sqrt{a} - \sqrt{b})^2 + 2\beta < 2$.*

- *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^K h^m_{\ell_m} = \beta\log(n) + o(\log(n)), |\beta| < T\frac{(a-b)}{2}, \sum_{k=1}^K \log(\alpha^k_{sgn(\beta),m_k}) = o(\log(n))$ and concurrently $\eta(a, b, \beta) < 2$.*

- *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^K h^m_{\ell_m} = \beta\log(n) + o(\log(n)), |\beta_2| < T\frac{(a-b)}{2}, \sum_{k=1}^K \log(\alpha^k_{sgn(\beta),m_k}) = -\beta'\log(n) + o(\log(n)), \beta' > 0$ and concurrently $\eta(a, b, \beta) + \beta' < 2$.*

Proof. Please see Appendix 2.6.4 □

Combining Lemmas 2, 3, 4, 12 concludes the proof of converse.

Achievability: Achievability of Theorem 5 is proven via an algorithm whose exact recovery conditions are identical to the necessary conditions provided in Lemma 12. The first stage of the algorithm is the same as Section 2.3.2. After the first stage, we have G_2, the side information $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_K$, $A^{'}$ and $B^{'}$. Locally modify the community assignment as follows: for a node $i \in A^{'}$, flip its membership if the number of edges between i and $B^{'}$ is greater than or equal the number of edges between i and $A^{'}$ plus $\sum_{k=1}^{K} \frac{h_{i,k}}{T}$ and for node $j \in B^{'}$, flip its membership if the number of of edges between j and $A^{'}$ is greater than or equal the number of of edges between j and $B^{'}$ minus $\sum_{k=1}^{K} \frac{h_{j,k}}{T}$. If the the number of flips in each cluster is not the same, keep the clusters unchanged.

Lemma 13. *The algorithm described above successfully recovers the communities with high probability if the following are satisfied simultaneously:*

- *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^{K} h_{\ell_m}^{m} = o(\log(n))$ and $\sum_{k=1}^{K} \log(\alpha_{+,m_k}^{k})$ and $\sum_{k=1}^{K} \log(\alpha_{-,m_k}^{k})$ are $o(\log(n))$ and concurrently $(\sqrt{a} - \sqrt{b})^2 > 2$.*

- *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^{K} h_{\ell_m}^{m} = o(\log(n)), \sum_{k=1}^{K} \log(\alpha_{+,m_k}^{k}) = \sum_{k=1}^{K} \log(\alpha_{-,m_k}^{k}) = -\beta \log(n) + o(\log(n)), \beta > 0$ and concurrently $(\sqrt{a} - \sqrt{b})^2 + 2\beta > 2$.*

- *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^{K} h_{\ell_m}^{m} = \beta \log(n) + o(\log(n)), |\beta| < T\frac{(a-b)}{2}, \sum_{k=1}^{K} \log(\alpha_{sgn(\beta),m_k}^{k}) = o(\log(n))$ and concurrently $\eta(a, b, \beta) > 2$.*

- *If there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^{K} h_{\ell_m}^{m} = \beta \log(n) + o(\log(n)), |\beta_2| < T\frac{(a-b)}{2}, \sum_{k=1}^{K} \log(\alpha_{sgn(\beta),m_k}^{k}) = -\beta^{'} \log(n) + o(\log(n)), \beta^{'} > 0$ and concurrently $\eta(a, b, \beta) + \beta^{'} > 2$.*

Proof. Define $P_e = \mathbb{P}(\text{node } i \text{ to be misclassified})$. Following similar analysis as in the proof of Lemma 2 leads to:

$$P_e \leq \frac{1}{2}\Bigg(n^{-1-\Omega(1)} + \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \cdots \sum_{m_K=1}^{M_K} \prod_{k=1}^{K} (\alpha_{+,m_k}^{k}) \times$$

$$\mathbb{P}\Bigg(\sum_{l=1}^{\frac{n}{2}}(Z_l - W_l) \geq \sum_{k=1}^{K}\frac{h^m_{\ell_m}}{T} + \psi_n \log(n)\Bigg)\Bigg) + \frac{1}{2}\Bigg(n^{-1-\Omega(1)}+$$
$$\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}\cdots\sum_{m_K=1}^{M_K}\prod_{k=1}^{K}(\alpha^k_{-,m_k}) \times \mathbb{P}\Bigg(\sum_{l=1}^{\frac{n}{2}}(Z_l - W_l) \geq -\sum_{k=1}^{K}\frac{h^m_{\ell_m}}{T} + \psi_n \log(n)\Bigg)\Bigg) \tag{2.21}$$

where $\psi_n = o(1)$.

Similar to Lemma 7, it can be shown that any term inside the nested sum in the last displayed equation can be upper bounded by:

- $n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2+o(1)}$ if there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^{K} h^m_{\ell_m} = o(\log(n))$ and $\sum_{k=1}^{K}\log(\alpha^k_{+,m_k})$ and $\sum_{k=1}^{K}\log(\alpha^k_{-,m_k})$ are $o(\log(n))$.

- $n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2-\beta+o(1)}$ if there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^{K} h^m_{\ell_m} = o(\log(n)), \sum_{k=1}^{K}\log(\alpha^k_{+,m_k}) = \sum_{k=1}^{K}\log(\alpha^k_{-,m_k}) = -\beta\log(n) + o(\log(n)), \beta > 0$

- $n^{-\frac{1}{2}\eta(a,b,\beta)+o(1)}$ if there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^{K} h^m_{\ell_m} = \beta\log(n) + o(\log(n)), |\beta| < T\frac{(a-b)}{2}$, $\sum_{k=1}^{K}\log(\alpha^k_{\mathrm{sgn}(\beta),m_k}) = o(\log(n))$

- $n^{-\frac{1}{2}\eta(a,b,\beta)-\beta'+o(1)}$ if there exists a sequence (over n) of side information outcomes $[m_1, \ldots, m_K]$ such that $\sum_{k=1}^{K} h^m_{\ell_m} = \beta\log(n) + o(\log(n)), |\beta_2| < T\frac{(a-b)}{2}$, $\sum_{k=1}^{K}\log(\alpha^k_{\mathrm{sgn}(\beta),m_k}) = -\beta'\log(n) + o(\log(n)), \beta' > 0$

Since K and $M_k, k \in \{1, \cdots, K\}$ are fixed, a union bound over nodes concludes the proof of Lemma 13. $\square$

This concludes the proof of achievability. $\square$

We now give an example of side information with $K = 1$ and fixed cardinality and analyze the effect of the evolution of the distribution of side information with growing n.

Consider the weakly symmetric side information whose transition probability matrix $\mathbb{P}(y|x)$ is defined as follows: every row of the transition matrix $\mathbb{P}(\cdot|x)$ is a permutation of every other row, and all the column sums $\sum_x \mathbb{P}(y|x)$ are equal. Since the labels are either 1 or -1, all the column sums is $\frac{2}{M}$. Without loss of generality, assume the first row $\mathbb{P}(y|x = +1)$ is arranged in descending order, i.e. $\mathbb{P}(y_{l+1}|x = +1) \geq \mathbb{P}(y_l|x = +1)$, $1 \leq l \leq M - 1$. Thus, for even M (odd M follow similarly), by the weakly symmetry property of $\mathbb{P}(y|x)$: $\alpha_{\pm,l} + \alpha_{\pm,M-l+1} = \frac{2}{M}$ and $h_l = -h_{M-l+1}$, $1 \leq l \leq \frac{M}{2}$. Thus, if $h_{\frac{M}{2}} = \beta \log(n) + o(\log(n))$, i.e., $h_{\frac{M}{2}} = O(\log(n))$, this implies that $h_l = O(\log(n))$ for all $1 \leq l \leq M$, and hence, this maps to the third case of Theorem 5. In other words, $\eta(a, b, |\beta|) > 2$ is necessary and sufficient for exact recovery (assuming $|\beta| < \frac{T(a-b)}{2}$). On the other hand, if $h_{\frac{M}{2}}$ is in the order of $o(\log(n))$, this maps to the first case of Theorem 5, and hence, side information does not change the exact recovery phase transition.

2.5.2 Varying Number of Fixed-Quality Features

In this section, α^k_{+,m_k} and α^k_{-,m_k} are independent of n. We study how many features K are needed so that side information can improve the phase transition threshold of exact recovery. We show that when $K = o(\log(n))$, side information will not improve the phase transition of exact recovery. A direct extension of our result shows that with $K = O(\log(n))$, side information can improve the phase transition, but this result is omitted here both in the interest of brevity and in part because it can be considered a straight forward extension of (Asadi et al., 2017, Theorem 4) which showed the result in the special case of $K = \log(n)$.

Theorem 6. *Assume that $M_k = M$ and all features are i.i.d. conditioned on the labels. Let α^k_{+,m_k} and α^k_{-,m_k} be non-zero and independent of n. Then, if $K = o(\log(n))$, $(\sqrt{a} - \sqrt{b})^2 > 2$ is necessary and sufficient for exact recovery.*

Proof. **Converse:** Using the same definitions as in (2.20), it remains to show when $\mathbb{P}(F_i^{H_1}) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ and $\mathbb{P}(F_i^{H_2}) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ hold for $\delta \in (0,1)$.

Lemma 14. *For $K = o(\log(n))$, both $\mathbb{P}(F_i^{H_1})$ and $\mathbb{P}(F_i^{H_2})$ are greater than $\frac{\log^3(n)}{n}\log(\frac{1}{\delta})$, $\delta \in (0,1)$ for sufficiently large n if $(\sqrt{a}-\sqrt{b})^2 < 2$.*

Proof. Let $W_i \sim Bern(p)$, $Z_i \sim Bern(q)$. Then, by following similar analysis as in Lemmas 15, 12, the following holds:

$$\begin{aligned}
&\mathbb{P}(F_i^{H_1})\\
\geq&\mathbb{P}\bigg(\sum_{l=1}^{\frac{n}{2}}[Z_l - W_l] \geq \sum_{k=1}^{K}\frac{h_{i,k}}{T} + 1 + \frac{\log(n)}{\log\log(n)}\bigg)\\
\geq& e^{-\log(n)(1+o(1))(\sup_{t\in\mathbb{R}}\frac{a+b}{2}-\frac{b}{2}(\frac{a}{b})^t-\frac{a}{2}(\frac{a}{b})^{-t}-\frac{K\log(\mathbb{E}_+[e^{-th_i}])}{\log(n)})}
\end{aligned} \tag{2.22}$$

where $\mathbb{E}_+[e^{-th_i}]$ is the moment generating function of the side information of any node i conditioned on $x_i^* = +1$. Since $K = o(\log(n)$, substituting in (2.22) leads to:

$$\begin{aligned}
\mathbb{P}(F_i^{H_1}) &\geq e^{-\log(n)(1+o(1))(\sup_{t\in\mathbb{R}}\frac{a+b}{2}-\frac{b}{2}(\frac{a}{b})^t-\frac{a}{2}(\frac{a}{b})^{-t})}\\
&\geq n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2+o(1)}
\end{aligned} \tag{2.23}$$

where the last inequality holds by evaluating the supremum. Thus, if $\frac{1}{2}(\sqrt{a}-\sqrt{b})^2 < 1$, $\mathbb{P}(F_i^{H_1}) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$, $\delta \in (0,1)$ for sufficiently large n.

Similarly,

$$\begin{aligned}
&\mathbb{P}(F_i^{H_2})\\
\geq& e^{-\log(n)(1+o(1))(\sup_{t\in\mathbb{R}}\frac{a+b}{2}-\frac{b}{2}(\frac{a}{b})^t-\frac{a}{2}(\frac{a}{b})^{-t}-\frac{K\log(\mathbb{E}_-[e^{th_i}])}{\log(n)})}\\
\geq& n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2+o(1)}
\end{aligned} \tag{2.24}$$

Thus, if $\frac{1}{2}(\sqrt{a}-\sqrt{b})^2 < 1$, $\mathbb{P}(F_i^{H_2}) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$, $\delta \in (0,1)$ for sufficiently large n. □

Combining Lemmas 2, 3, 4, 14 concludes the proof of converse.

Achievability: It is known that $\frac{1}{2}(\sqrt{a}-\sqrt{b})^2 > 1$ is sufficient if the only observation was the graph. Combining this with the converse completes the proof. □

2.6 Appendix

2.6.1 Proof of Lemma 5

Define $l = \frac{n}{2}$ and $\Gamma(t) = \log(\mathbb{E}_X[e^{tx}])$ for a random variable X. Then,

$$\begin{aligned}
\mathbb{P}(F_i^H) &= \mathbb{P}\Bigg(\sum_{k=1}^{\frac{n}{2}} Z_k - \sum_{k=1}^{\frac{n}{2}-\frac{n}{\log^3(n)}} W_k - cy_i \geq T + T\frac{\log(n)}{\log\log(n)}\Bigg) \\
&\geq \mathbb{P}\Bigg(\sum_{k=1}^{\frac{n}{2}} [Z_k - W_k] \geq cy_i + T + T\frac{\log(n)}{\log\log(n)}\Bigg) \\
&= (1-\alpha)\mathbb{P}\Bigg(\frac{1}{l}\sum_{k=1}^{\frac{n}{2}} [Z_k - W_k] \geq \frac{1}{l}(c + T + T\frac{\log(n)}{\log\log(n)})\Bigg) \\
&\quad + \alpha\mathbb{P}\Bigg(\frac{1}{l}\sum_{k=1}^{\frac{n}{2}} [Z_k - W_k] \geq \frac{1}{l}(-c + T + T\frac{\log(n)}{\log\log(n)})\Bigg) \\
&\overset{(a)}{\geq} (1-\alpha)e^{-l\left(t_1^* a_1 - \Gamma(t_1^*) + |t_1^*|\delta\right)}(1 - o(1)) + \alpha e^{-l\left(t_2^* a_2 - \Gamma(t_2^*) + |t_2^*|\delta\right)}(1 - o(1))
\end{aligned} \tag{2.25}$$

where (a) holds by defining $\delta = \frac{\log^{\frac{2}{3}}(n)}{l}$, $a_1 = \frac{1}{l}(c + T + T\frac{\log(n)}{\log\log(n)}) + \delta$, $a_2 = \frac{1}{l}(-c + T + T\frac{\log(n)}{\log\log(n)}) + \delta$, $t_1^* = \arg\sup_{t\in\mathbb{R}} ta_1 - \Gamma(t)$, $t_2^* = \arg\sup_{t\in\mathbb{R}} ta_2 - \Gamma(t)$ and by using Lemma 15 in Appendix 2.6.5.

Note that both supremums in (2.25) are very similar, thus the analysis for only one of them will be presented and the second should follow similarity.

$$ta_1 - \Gamma(t) = ta_1 - \log\left(1 - q(1 - (\frac{a}{b})^t)\right) - \log\left(1 - p(1 - (\frac{a}{b})^{-t})\right)$$

It is easy to check that the right hand side is concave in $t \in \mathbb{R}$. Hence, taking the derivative with respect to t leads to:

$$\begin{aligned}
&a_1 - \frac{Tq(\frac{a}{b})^t}{1 - q(1 - (\frac{a}{b})^t)} + \frac{Tp(\frac{a}{b})^{-t}}{1 - p(1 - (\frac{a}{b})^{-t})} \\
&= \frac{\log(n)}{n}\Bigg(\frac{2c}{\log(n)} + \frac{2T}{\log(n)} + \frac{2T}{\log\log(n)} + \frac{2}{\log^{\frac{1}{3}}(n)} - \frac{Tb(\frac{a}{b})^t}{1 - q(1 - (\frac{a}{b})^t)} + \frac{Ta(\frac{a}{b})^{-t}}{1 - p(1 - (\frac{a}{b})^{-t})}\Bigg)
\end{aligned}$$

$$= 0 \tag{2.26}$$

We consider two asymptotic regimes for α:

- $c = o(\log(n))$. Then, the first four terms in (2.26) is $o(1)$. This suggests that $t^* = \frac{1}{2}$. Hence, substituting back in (2.26) leads to:

$$\begin{aligned} ta_1 - \Gamma(t) &= \frac{1}{2}a_1 - \log\big(1 - q(1 - (\sqrt{\frac{a}{b}}))\big) - \log\big(1 - p(1 - (\sqrt{\frac{b}{a}})\big) \\ &\overset{(a)}{\leq} \frac{1}{2}a_1 + \frac{q(1-(\sqrt{\frac{a}{b}}))}{1 - q(1-(\sqrt{\frac{a}{b}}))} + \frac{p(1-(\sqrt{\frac{b}{a}}))}{1-p(1-(\sqrt{\frac{b}{a}}))} \\ &\overset{(b)}{=} \frac{\log(n)}{n}\Big((\sqrt{a}-\sqrt{b})^2 + o(1)\Big) \end{aligned} \tag{2.27}$$

  where (a) holds because $\log(1-x) \geq \frac{-x}{1-x}$ and (b) holds because both $(1 - q(1-(\sqrt{\frac{a}{b}})))$ and $(1 - q(1-(\sqrt{\frac{a}{b}}))) \to 1$ as $n \to \infty$. Hence, substituting in one of the supremums of (2.25) leads to:

$$e^{-l\left(t_1^* a_1 - \Gamma(t_1^*) + |t_1^*|\delta\right)} \geq e^{-\log(n)\left(\frac{1}{2}(\sqrt{a}-\sqrt{b})^2 + o(1)\right)}$$

  Finally, following the same steps for the second supremum and substituting in (2.25) lead to:

$$\mathbb{P}(F_i^H) \geq n^{-0.5(\sqrt{a}-\sqrt{b})^2 + o(1)}$$

  Thus, if $(\sqrt{a}-\sqrt{b})^2 \leq 2 - \varepsilon$ for some $0 < \varepsilon < 2$, then $\mathbb{P}(F_i^H) \geq n^{-1+\frac{\varepsilon}{2}} > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ for $\delta \in (0,1)$ for sufficiently large n. This proves the first case of Lemma 5.

- $c = \beta\log(n) + o(\log(n))$, $\beta > 0$. Substituting in (2.26), this suggests that $t_1^* = \frac{1}{T}\log(\frac{\gamma+\beta}{bT})$ and $t_2^* = \frac{1}{T}\log(\frac{\gamma-\beta}{bT})$, where $\gamma = \sqrt{\beta^2 + abT^2}$. Hence, by substituting back in (2.26) and following the same ideas as in (2.27), the following holds:

$$ta_1 - \Gamma(t) \leq \frac{\log(n)}{n}\big(2\beta t^* + b(1-(\frac{a}{b})^{t^*}) + a(1-(\frac{a}{b})^{-t^*}) + o(1)\big)$$

$$= \frac{\log(n)}{n}\big(a+b+\beta-\frac{2\gamma}{T}+\frac{\beta}{T}\log(\frac{\gamma+\beta}{\gamma-\beta})+o(1)\big)$$
$$= \frac{\log(n)}{n}(\eta(a,b,\beta)+o(1)) \tag{2.28}$$

Hence, substituting in one of the supremums of (2.25) leads to:

$$e^{-l\left(t_1^* a_1-\Gamma(t_1^*)+|t_1^*|\delta\right)} \geq e^{-\frac{\log(n)}{2}\left(\eta(a,b,\beta)+o(1)\right)}$$

Finally, by following the same steps for the second supremum and substituting in (2.25), the following holds:

$$\mathbb{P}(F_i^H) \geq n^{-0.5\eta(a,b,\beta)+o(1)} + \alpha n^{-0.5\eta(a,b,\beta)+\beta+o(1)}$$
$$= n^{-0.5\eta(a,b,\beta)+o(1)}$$

Thus, if $\eta(a,b,\beta) \leq 2-\varepsilon$ for some $0<\varepsilon<2$, then $\mathbb{P}(F_i^H) \geq n^{-1+\frac{\epsilon}{2}} > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ for $\delta \in (0,1)$ for sufficiently large n. This proves the second case of Lemma 5.

For the last case of Lemma 5, we begin as in (2.25) but take a different approach:

$$\mathbb{P}(F_i^H) \geq \mathbb{P}\bigg(\sum_{k=1}^{\frac{n}{2}}[Z_k-W_k] \geq cy_i+T+T\frac{\log(n)}{\log\log(n)}\bigg)$$
$$=(1-\alpha)\bigg(1-\mathbb{P}\bigg(\sum_{k=1}^{\frac{n}{2}}[Z_k-W_k] \leq c+T+T\frac{\log(n)}{\log\log(n)}\bigg)\bigg)$$
$$+\alpha\bigg(1-\mathbb{P}\bigg(\sum_{k=1}^{\frac{n}{2}}[Z_k-W_k] \leq -c+T+T\frac{\log(n)}{\log\log(n)}\bigg)\bigg)$$
$$\overset{(a)}{\geq} 1-(1-\alpha)\times e^{-n\sup_{t>0}\frac{-t}{n}\left(c+T+T\frac{\log(n)}{\log\log(n)}\right)-\frac{1}{2}\log\left(\mathbb{E}(e^{-t[Z-W]})\right)}$$
$$-\alpha e^{-n\sup_{t>0}\frac{-t}{n}\left(-c+T+T\frac{\log(n)}{\log\log(n)}\right)-\frac{1}{2}\log\left(\mathbb{E}(e^{-t[Z-W]})\right)} \tag{2.29}$$

where (a) holds by Chernoff bound. Note that unlike the previous cases, here the supremum is only on $t>0$. A direct computation of the logarithmic term leads to:

$$\log\left(\mathbb{E}\big[e^{-t\sum_{i=1}^{\frac{n}{2}}[Z_i-W_j]}\big]\right) \overset{(a)}{=} \frac{n}{2}\Big(\log\big(1-q(1-(\frac{p}{q})^{-t})\big)+\log\big(1-p(1-(\frac{p}{q})^{t})\big)\Big)$$

$$\stackrel{(b)}{\leq} -(\frac{n}{2})q(1-(\frac{p}{q})^{-t}) - (\frac{n}{2})p(1-(\frac{p}{q})^{t})$$

where (a) follows from the fact that W_i, Z_i are independent random variables $\forall i$, and (b) holds because $\log(1-x) \leq -x$. Thus, the first supremum in (2.29) can be lower bounded by:

$$\frac{\log(n)}{n} \sup_{t>0} -t(\beta + o(1)) + \frac{1}{2}\big(a+b-b(\frac{a}{b})^{-t}-a(\frac{a}{b})^{t}\big) \tag{2.30}$$

Again, by concavity of the last equation in t, it is easy to calculate the first derivative to get:

$$-\beta - \frac{aT}{2}(\frac{a}{b})^{t} + \frac{bT}{2}(\frac{a}{b})^{-t} = 0 \tag{2.31}$$

Hence, following the same analysis as before, it can be shown that t^* for the first and second supremums can be calculated as: $\frac{1}{T}\log(\frac{\gamma-\beta}{aT})$ and $\frac{1}{T}\log(\frac{\gamma+\beta}{aT})$, respectively. Since t has to be greater than zero, $\beta < \frac{T(b-a)}{2}$ is needed for for the first supremum, which can not be true, since β is positive and $b < a$. Hence, by the concavity of the function and the fact that it approaches $-\infty$ as $t \to \infty$, the optimal t for the first supremum is $t_1^* = 0$. On the other hand, for the second supremum, $\beta > \frac{T(a-b)}{2}$ is needed for t to be positive.

Thus, assume $\beta > \frac{T(a-b)}{2}$ and by substituting in (2.29), the following holds:

$$\mathbb{P}(F_i^H) \geq 1 - (1-\alpha)e^0 - \alpha n^{-\frac{1}{2}\eta(a,b,\beta)+\beta}$$
$$\stackrel{(a)}{=} n^{-\beta} - n^{-\frac{1}{2}\eta(a,b,\beta)}$$

where (a) holds by using the fact that $\alpha = n^{-\beta}$. Hence, if $\beta \leq 1-\varepsilon_1$ and $\frac{1}{2}\eta \geq 1+\varepsilon_2$, then $\mathbb{P}(F_i^H) \geq n^{-1}(n^{\varepsilon_1} - n^{-\varepsilon_2}) > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ for $\delta \in (0,1)$ for sufficiently large n. This proves the third and last case of Lemma 5.

2.6.2 Proof of Lemma 7

By upper bounding P_e, we get:

$$
\begin{aligned}
P_e \leq & \mathbb{P}\bigg(\sum_{k=1}^{(1-\delta)\frac{n}{2}} Z_k + \sum_{k=1}^{\delta\frac{n}{2}} W_k \geq \sum_{j=1}^{(1-\delta)\frac{n}{2} - \frac{2D}{\log(n)}n} W_j + \frac{c}{T} y_i \bigg) \\
\leq & \mathbb{P}\bigg(\sum_{k=1}^{\frac{n}{2}} Z_k + \sum_{k=1}^{\delta\frac{n}{2}} W_k \geq \sum_{j=1}^{(1-\delta)\frac{n}{2} - \frac{2D}{\log(n)}n} W_i + \frac{c}{T} y_i \bigg) \\
\leq & \mathbb{P}\bigg(\sum_{k=1}^{\frac{n}{2}} Z_k - \sum_{k=1}^{\frac{n}{2}} W_k + \sum_{j=1}^{\delta n + \frac{2D}{\log(n)}n} W_i \geq \frac{c}{T} y_i \bigg) \\
\overset{(a)}{\leq} & \mathbb{P}\bigg(\sum_{k=1}^{\frac{n}{2}} (Z_k - W_k) \geq \frac{c}{T} y_i - \psi\delta\log(n) \bigg) + \mathbb{P}\bigg(\sum_{j=1}^{\delta n + \frac{2D}{\log(n)}n} W_j \geq \psi\delta\log(n) \bigg) \\
= & (1-\alpha)\mathbb{P}\bigg(\sum_{k=1}^{\frac{n}{2}} (Z_k - W_k) \geq \frac{c}{T} - \psi\delta\log(n) \bigg) + \\
& \alpha\mathbb{P}\bigg(\sum_{k=1}^{\frac{n}{2}} (Z_k - W_k) \geq -\frac{c}{T} - \psi\delta\log(n) \bigg) + \mathbb{P}\bigg(\sum_{j=1}^{\delta n + \frac{2D}{\log(n)}n} W_j \geq \psi\delta\log(n) \bigg)
\end{aligned}
\tag{2.32}
$$

where (a) holds by defining $\psi = \frac{1}{\delta\sqrt{\log(\frac{1}{\delta})}}$.

Now we bound the second term. A multiplicative Chernoff bound that states that for a sequence of n i.i.d random variables X_i: $\mathbb{P}(\sum_{i=1}^{n} \geq t\mu) \leq (\frac{t}{e})^{-t\mu}$, where $\mu = n\mathbb{E}[X]$. Applying this bound to the second term with $\mu = a(\delta\log(n) + 2D)$ and $t = \frac{\psi\delta\log(n)}{a(\delta\log(n)+2D)}$ leads to:

$$
\mathbb{P}\bigg(\sum_{j=1}^{\delta n + \frac{2D}{\log(n)}n} W_j \geq \psi\delta\log(n) \bigg) \leq \bigg(\frac{\psi\delta\log(n)}{ae(\delta\log(n) + 2D)} \bigg)^{-\psi\delta\log(n)} \tag{2.33}
$$

$$
\begin{aligned}
&= \bigg(\frac{\psi}{ae(1 + \frac{2D}{\delta\log(n)})} \bigg)^{-\frac{\log(n)}{\sqrt{\log(\frac{1}{\delta})}}} \\
&= e^{\log(n)\left(\frac{1+\log(a)}{\sqrt{\log(\frac{1}{\delta})}} + \frac{\log(1+\frac{2D}{\delta\log(n)})}{\sqrt{\log(\frac{1}{\delta})}} + \frac{\log(\delta) + \frac{1}{2}\log\log(\frac{1}{\delta})}{\sqrt{\log(\frac{1}{\delta})}} \right)} \\
&\overset{(a)}{=} n^{-\sqrt{\log(\frac{1}{\delta})}\left(1 - \frac{\log(1+\frac{2D}{\delta\log(n)})}{\log(\frac{1}{\delta})} + o(1) \right)}
\end{aligned}
\tag{2.34}
$$

where (a) holds because $\delta \to 0$ as $D \to \infty$. Note that we can find D large enough such that $\frac{\log(1+\frac{2D}{\delta \log(n)})}{\log(\frac{1}{\delta})} < 1$. Hence,

$$\mathbb{P}\Bigg(\sum_{j=1}^{\delta n + \frac{2D}{\log(n)} n} W_j \geq \psi\delta \log(n) \Bigg) \leq n^{-(1+\Omega(1))} \tag{2.35}$$

Now for the first term in (2.32), Chernoff bound can be used as follows:

$$\begin{aligned}
&(1-\alpha)\mathbb{P}\Bigg(\sum_{k=1}^{\frac{n}{2}} (Z_k - W_k) \geq \frac{c}{T} - \psi\delta\log(n) \Bigg) + \alpha\mathbb{P}\Bigg(\sum_{k=1}^{\frac{n}{2}} (Z_k - W_k) \geq -\frac{c}{T} - \psi\delta\log(n) \Bigg) \\
&\overset{(a)}{\leq} (1-\alpha) e^{-\frac{\log(n)}{2} \sup_{t_1>0} 2t(\frac{c}{T\log(n)} - \psi\delta) + a + b - be^{t_1} - ae^{-t_1}} + \\
&\qquad \alpha e^{-\frac{\log(n)}{2} \sup_{t_2>0} 2t_2(-\frac{c}{T\log(n)} - \psi\delta) + a + b - be^{t_2} - ae^{-t_2}}
\end{aligned} \tag{2.36}$$

where (a) holds because $\log(1-x) \leq -x$. Since $\psi\delta \to 0$ as $D \to \infty$, $\psi\delta$ can be replaced by $o(1)$ for sufficiently large D. We consider the following asymptotic regimes for α.

- If $c = o(\log(n))$, this suggests that $t_1^* = t_2^* = \frac{1}{2}T$. Hence, (2.36) can be upper bounded by:

$$n^{-\frac{1}{2}(\sqrt{a}-\sqrt{b})^2 + o(1)} \tag{2.37}$$

- If $c = \beta\log(n) + o(\log(n))$, for $0 < \beta < \frac{T(a-b)}{2}$, then it can be shown that $t_1^* = \log(\frac{\gamma+\beta}{bT})$ and $t_2^* = \log(\frac{\gamma-\beta}{bT})$, where $\gamma = \sqrt{\beta^2 + abT^2}$. Hence, (2.36) can be upper bounded by:

$$(2-\alpha)n^{-\frac{1}{2}\eta(a,b,\beta) + o(1)} \tag{2.38}$$

- If $c = \beta\log(n)$, for $\beta > \frac{T(a-b)}{2}$, then it can be shown that $t_1^* = \log(\frac{\gamma+\beta}{bT})$ and $t_2^* = 0$. Hence, (2.36) can be upper bounded by:

$$(1-\alpha)n^{-\frac{1}{2}\eta(a,b,\beta) + o(1)} + n^{-\beta} \tag{2.39}$$

The last three equations and (2.35), substituting in (2.32), concludes the proof of the lemma.

2.6.3 Proof of Lemma 10

Define $l = \frac{n}{2}$ and let $\Gamma(t) = \log(\mathbb{E}_X[e^{tx}])$ for a random variable X. Then,

$$\begin{aligned}
\mathbb{P}(F_i^H) &= \epsilon\mathbb{P}\Bigg(\sum_{k=1}^{\frac{n}{2}}(Z_k) - \sum_{k=1}^{\frac{n}{2}-\frac{n}{\log^3(n)}}(W_k) \geq 1 + \frac{\log(n)}{\log\log(n)}\Bigg) \\
&\geq \epsilon\mathbb{P}\Bigg(\sum_{k=1}^{\frac{n}{2}}[Z_k - W_k] \geq 1 + \frac{\log(n)}{\log\log(n)}\Bigg) \\
&\overset{(a)}{\geq} \epsilon e^{-l\left(t^*a - \Gamma(t_1^*) + |t_1^*|\delta\right)}(1 - o(1)) \\
&= e^{-l\left(t^*a - \Gamma(t_1^*) + |t_1^*|\delta\right) + \log(\epsilon)}(1 - o(1))
\end{aligned} \tag{2.40}$$

where (a) holds by defining $\delta = \frac{\log^{\frac{2}{3}}(n)}{l}$, $a = \frac{1}{l}(1 + \frac{\log(n)}{\log\log(n)}) + \delta$, $t_1^* = \arg\sup_{t\in\mathbb{R}} at - \Gamma(t)$ and by using Lemma 15 in Appendix 2.6.5.

Following similar analysis as in (2.26) and (2.26), it can be shown that $t^* = \frac{1}{2}T$. Thus, by substituting back in (2.40) and using the fact that $\log(1-x) \geq \frac{-x}{1-x}$, we get:

$$\mathbb{P}(F_i^H) \geq \epsilon n^{-0.5(\sqrt{a}-\sqrt{b})^2 + o(1)} \tag{2.41}$$

Thus, if $\log(\epsilon) = o(\log(n))$, then, it is clear that if $(\sqrt{a}-\sqrt{b})^2 \leq 2-\varepsilon$ for some $0 < \varepsilon < 2$, then $\mathbb{P}(F_i^H) \geq n^{-1+\frac{\varepsilon}{2}} > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ for $\delta \in (0,1)$ for sufficiently large n. This proves the first case of Lemma 10. On the other hand, if $\log(\epsilon) = -\beta\log(n) + o(\log(n))$, for some $\beta > 0$, then, it is clear that if $(\sqrt{a} - \sqrt{b})^2 + 2\beta \leq 2 - \varepsilon$ for some $0 < \varepsilon < 2$, then $\mathbb{P}(F_i^H) \geq n^{-1+\frac{\varepsilon}{2}}\frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ for $\delta \in (0,1)$ for sufficiently large n. This proves the second and last case of Lemma 10.

2.6.4 Proof of Lemma 12

Let $W_i \sim Bern(p)$, $Z_i \sim Bern(q)$ and define $l = \frac{n}{2}$ and $\Gamma(t) = \log(\mathbb{E}_X[e^{tx}])$ for a random variable X. Then, we have the following:

$$\mathbb{P}(F_i^{H_1}) = \mathbb{P}\Bigg(\sum_{j=1}^{\frac{n}{2}}(Z_j) - \sum_{j=1}^{\frac{n}{2}-\frac{n}{\log^3(n)}}(W_j) \geq \sum_{k=1}^{K}\frac{h_{i,k}}{T} + 1 + \frac{\log(n)}{\log\log(n)}\Bigg)$$

$$\geq \sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}\cdots\sum_{m_K=1}^{M_K}(\prod_{k=1}^{K}\alpha^k_{+,m_k})\mathbb{P}\Big(\sum_{j=1}^{\frac{n}{2}}[Z_j-W_j]\geq\sum_{k=1}^{K}\frac{h^m_{\ell_m}}{T}+1+\frac{\log(n)}{\log\log(n)}\Big)$$

$$\stackrel{(a)}{\geq} \sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}\cdots\sum_{m_K=1}^{M_K}(\prod_{k=1}^{K}\alpha^k_{+,m_k})e^{-l\big(t^*a-\Gamma(t^*)+|t^*|\delta\big)} \tag{2.42}$$

where (a) holds by defining $\delta=\frac{\log^{\frac{2}{3}}(n)}{l}$, $a=\frac{1}{l}(\sum_{k=1}^{K}\frac{h^m_{\ell_m}}{T}+1+\frac{\log(n)}{\log\log(n)})+\delta$, $t^*=\arg\sup_{t\in\mathbb{R}} at-\Gamma(t)$ and by using Lemma 15 in the Appendix[9]. Similarly,

$$\mathbb{P}(F_i^{H_2})\geq \sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}\cdots\sum_{m_K=1}^{M_K}(\prod_{k=1}^{K}\alpha_{+,m_k})e^{-l\big(t^*a-\Gamma(t^*)+|^*|\delta\big)} \tag{2.43}$$

where $a=\frac{1}{l}(-\sum_{k=1}^{K}\frac{h^m_{\ell_m}}{T}+1+\frac{\log(n)}{\log\log(n)})+\delta$.

Without loss of generality, we focus on one term of the nested sum in (2.42) and (2.43). Then,

- If $\sum_{k=1}^{K}h^m_{\ell_m}=o(\log(n))$ and both $\sum_{k=1}^{K}\log(\alpha^k_{+,m_k}$ and $\sum_{k=1}^{K}\log(\alpha^k_{-,m_k})$ are $o(\log(n))$, then the optimal t for that term is $t^*=\frac{1}{2}T$ for both (2.42), (2.43). Hence, substituting in (2.42), (2.43) leads to:

$$\mathbb{P}(F_i^{H_1})\geq n^{-0.5(\sqrt{a}-\sqrt{b})^2+o(1)} \tag{2.44}$$

$$\mathbb{P}(F_i^{H_2})\geq n^{-0.5(\sqrt{a}-\sqrt{b})^2+o(1)} \tag{2.45}$$

  Thus, it is clear that if $(\sqrt{a}-\sqrt{b})^2\leq 2-\varepsilon$ for some $0<\varepsilon<2$, then $\mathbb{P}(F_i^{H_1})$ and $\mathbb{P}(F_i^{H_2})$ are both greater than $n^{-1+\frac{\varepsilon}{2}}>\frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ for $\delta\in(0,1)$ for sufficiently large n.

- If $\sum_{k=1}^{K}h^m_{\ell_m}=o(\log(n)),\sum_{k=1}^{K}\log(\alpha^k_{+,m_k})=\sum_{k=1}^{K}\log(\alpha^k_{-,m_k})=-\beta\log(n)+o(\log(n))$, $\beta>0$, then $t^*=\frac{1}{2}T$ for both (2.42), (2.43). Hence, by substituting in (2.42), (2.43), the following holds:

$$\mathbb{P}(F_i^{H_1})\geq n^{-0.5(\sqrt{a}-\sqrt{b})^2-\beta+o(1)} \tag{2.46}$$

[9]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.

$$\mathbb{P}(F_i^{H_2}) \geq n^{-0.5(\sqrt{a}-\sqrt{b})^2-\beta+o(1)} \tag{2.47}$$

Thus, it is clear that if $(\sqrt{a}-\sqrt{b})^2+2\beta \leq 2-\varepsilon$ for some $0<\varepsilon<2$, then $\mathbb{P}(F_i^{H_1})$ and $\mathbb{P}(F_i^{H_2})$ are both greater than $n^{-1+\frac{\varepsilon}{2}} > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ for $\delta \in (0,1)$ for sufficiently large n.

- If $\sum_{k=1}^{K} h_{\ell_m}^m = \beta\log(n) + o(\log(n)), 0 < \beta < T\frac{(a-b)}{2}$, then $t^* = \log(\frac{\gamma+\beta}{bT})$ for (2.42) and $t^* = \frac{1}{T}\log(\frac{\gamma-\beta}{bT})$ for (2.43). Hence, by substituting in (2.42), (2.43), the following holds:

$$\mathbb{P}(F_i^{H_1}) \geq e^{-\log(n)\left(0.5\eta(a,b,\beta)-\sum_{k=1}^{K}\frac{\log(\alpha_{+,m_k}^k)}{\log(n)}+o(1)\right)} \tag{2.48}$$

$$\mathbb{P}(F_i^{H_2}) \geq e^{-\log(n)\left(0.5\eta(a,b,\beta)-\beta-\sum_{k=1}^{K}\frac{\log(\alpha_{-,m_k}^k)}{\log(n)}+o(1)\right)} \tag{2.49}$$

Then, if $\sum_{k=1}^{K}\log(\alpha_{+,m_k}^k) = o(\log(n))$, this implies that $\sum_{k=1}^{K}\frac{\log(\alpha_{-,m_k}^k)}{\log(n)} = -\beta + o(1)$. Hence,

$$\mathbb{P}(F_i^{H_1}) \geq n^{-0.5\eta(a,b,\beta)+o(1)} \tag{2.50}$$

$$\mathbb{P}(F_i^{H_2}) \geq n^{-0.5\eta(a,b,\beta)+o(1)} \tag{2.51}$$

Thus, it is clear that if $\eta(a,b,\beta) \leq 2-\varepsilon$ for some $0<\varepsilon<2$, then $\mathbb{P}(F_i^{H_1})$ and $\mathbb{P}(F_i^{H_2})$ are both greater than $n^{-1+\frac{\varepsilon}{2}} > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ for $\delta \in (0,1)$ for sufficiently large n.

On the other hand, if $\sum_{k=1}^{K}\log(\alpha_{+,m_k}^k) = -\beta^{'}\log(n) + o(\log(n))$, this implies that $\sum_{k=1}^{K}\frac{\log(\alpha_{-,m_k}^k)}{\log(n)} = -\beta''$, for some $\beta''>0$ and $\beta = \beta''-\beta^{'}$. Hence,

$$\mathbb{P}(F_i^{H_1}) \geq n^{-0.5\eta(a,b,\beta)-\beta^{'}+o(1)} \tag{2.52}$$

$$\begin{aligned}\mathbb{P}(F_i^{H_2}) &\geq n^{-0.5\eta(a,b,\beta)+\beta-\beta^{''}+o(1)} \\ &= n^{-0.5\eta(a,b,\beta)-\beta^{'}+o(1)}\end{aligned} \tag{2.53}$$

Thus, it is clear that if $\eta(a,b,\beta)+2\beta^{'} \leq 2-\varepsilon$ for some $0<\varepsilon<2$, then $\mathbb{P}(F_i^{H_1})$ and $\mathbb{P}(F_i^{H_2})$ are both greater than $n^{-1+\frac{\varepsilon}{2}} > \frac{\log^3(n)}{n}\log(\frac{1}{\delta})$ $\delta \in (0,1)$ for sufficiently large n. The case when $T\frac{(a-b)}{2} < \beta < 0$ holds similarly.

2.6.5 Proof of Lemma 15

Lemma 15. *Let $X_1, \cdots, X_n$ be a sequence of i.i.d random variables. Define $\Gamma(t) = \log(\mathbb{E}[e^{tX}])$. Then, for any $a, \epsilon \in \mathbb{R}$:*

$$\mathbb{P}\big(\frac{1}{n}\sum_{i=1}^{n} X_i \geq a-\epsilon\big) \geq e^{-n\left(t^*a-\Gamma(t^*)+|t^*|\epsilon\right)}\left(1-\frac{\sigma_{\hat{X}}^2}{n\epsilon^2}\right)$$

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 $\mu_{\hat{X}}, \sigma_{\hat{X}}^2$ are the mean and variance of $\hat{X}$, respectively.

Proof.

$$\begin{aligned}
&\mathbb{P}\big(\frac{1}{n}\sum_{i=1}^{n} X_i \geq a-\epsilon\big) \geq \mathbb{P}\big(a-\epsilon \leq \frac{1}{n}\sum_{i=1}^{n} X_i \leq a+\epsilon\big)\\
&= \int_{a-\epsilon\leq\frac{1}{n}\sum_{i=1}^n X_i\leq a+\epsilon} \mathbb{P}(x_1)\cdots\mathbb{P}(x_n)dx_1\cdots dx_n\\
&\overset{(a)}{=} \int_{a-\epsilon\leq\frac{1}{n}\sum_{i=1}^n X_i\leq a+\epsilon} \frac{(\mathbb{E}_X[e^{t\sum_{i=1}^n x_i}])(e^{t\sum_{i=1}^n x_i})}{(\mathbb{E}_X[e^{t\sum_{i=1}^n x_i}])(e^{t\sum_{i=1}^n x_i})}\mathbb{P}(x_1)\cdots\mathbb{P}(x_n)dx_1\cdots dx_n\\
&\overset{(b)}{\geq} e^{-n(ta-\Gamma(t)+|t|\epsilon)}\int_{a-\epsilon\leq\frac{1}{n}\sum_{i=1}^n X_i\leq a+\epsilon}\prod_{i=1}^{n}\left(\frac{e^{tx_i}\mathbb{P}(x_i)}{\mathbb{E}_X[e^{tx}]}dx_i\right)\\
&\overset{(c)}{=} e^{-n(ta-\Gamma(t)+|t|\epsilon)}\mathbb{P}_{\hat{X}_n}\left(a-\epsilon\leq\frac{1}{n}\sum_{i=1}^{n}\hat{X}_i\leq a+\epsilon\right)\\
&\overset{(d)}{\geq} e^{-n(ta-\Gamma(t)+|t|\epsilon)}\left(1-\frac{n\sigma_{\hat{X}}^2+(n\mu_{\hat{X}}-na)^2}{n^2\epsilon^2}\right)
\end{aligned} \tag{2.54}$$

where (a) holds for any $t \in \mathbb{R}$ such that the expectation holds, (b) holds because $a-\epsilon \leq \frac{1}{n}\sum_{i=1}^n \hat{X}_i \leq a+\epsilon$, (c) holds because $\frac{e^{tx}\mathbb{P}(x)}{\mathbb{E}_X[e^{tx}]}$ defines a probability distribution (Dembo and Zeitouni, 2010) on a new random variable $\hat{X}$ with the same alphabet as X, and (d) holds by Chebyshev's inequality and defining $\mu_{\hat{X}}, \sigma_{\hat{X}}^2$ to be the mean and variance of $\hat{X}$, respectively.

Now, choose $t = t^* = \arg\sup_{t\in\mathbb{R}} ta-\Gamma(t)$. Since this function is concave in $t \in \mathbb{R}$ (Dembo and Zeitouni, 2010), then by setting the first derivative to zero, we have $a = \frac{\mathbb{E}_X[xe^{t^*x}]}{\mathbb{E}[e^{t^*x}]}$. 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 $t = t^*$, we have $\mu_{\hat{X}} = a$. Thus, substituting back in (2.54) leads to:

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq a - \epsilon\right) \geq e^{-n(t^*a - \Gamma(t^*) + |t^*|\epsilon)}\big(1 - \frac{\sigma_{\hat{X}}^2}{n\epsilon^2}\big)$$

Now, in our model $\epsilon = \frac{\log^{\frac{2}{3}}(n)}{n}$ and $X = Z - W$, where $Z \sim$ T*Bern(q) and $W \sim$ T*Bern(p), where $T = \log(\frac{a}{b})$. Hence, it can be easily shown that $\sigma_{\hat{X}}^2$ is in the order of $\frac{\log(n)}{n}$, and hence,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq a - \epsilon\right) \geq e^{-n(t^*a - \Gamma(t^*) + |t^*|\epsilon)}\big(1 - o(1)\big)$$

which concludes our proof. $\square$

CHAPTER 3

TWO SYMMETRIC COMMUNITIES WITH CONTINUOUS-VALUED SIDE INFORMATION [1]

3.1 System Model and Assumptions

We use the same system model used in chapter two, which we restate here with slightly different notations that fit the needs of this chapter. The binary symmetric stochastic block model includes a graph with nodes $u \in \{1, \ldots, n\}$ each labeled $x_u = 1$ or -1. Any pair of nodes are connected by an edge with probability $p = a\frac{\log(n)}{n}$ if the nodes belong to the same community, and with probability $q = b\frac{\log(n)}{n}$ otherwise, where $a \geq b$. For each node a side information $\mathbf{Y}_u$ is observed, collectively according to the distribution $\prod_u P(\mathbf{Y}_u|x_u)$. Conditioned on node labels, the side information of different nodes are assumed to be independent of the graph edges.

We denote the observed graph by G, the vector of nodes' true labels by $\mathbf{x}^*$, and the vector of nodes' side information by $\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_n]$. The goal is to recover the node labels $\mathbf{x}^*$ from the observation of $(G, \mathbf{Y})$.

Definition 1. *Let $E(\cdot, \cdot)$ denote the number of edges between two sets of nodes. Define $A \triangleq \{u : x_u = 1\}$ and $B \triangleq \{u : x_u = -1\}$. For a node u, define*

$$h_u \triangleq \log\left(\frac{P(\mathbf{Y}_u|x_u = 1)}{P(\mathbf{Y}_u|x_u = -1)}\right). \tag{3.1}$$

Therefore the distribution of h_u is the same for all u. Define $T \triangleq \log(\frac{a}{b})$. We say a node u is irregular *if one of the following holds:*

$$x_u = +1 \quad and \quad E(u, B) \geq E(u, A) + \frac{1}{T}h_u + 1$$

[1] © 2019 IEEE H. Saad and A. Nosratinia, "Exact Recovery in Community Detection With Continuous-Valued Side Information," 2019 IEEE Signal Processing Letters, vol. 26, pp. 332-336, 2019.

$$x_u = -1 \quad and \quad E(u, A) \geq E(u, B) - \frac{1}{T} h_u + 1$$

otherwise we call the node regular*.*

Definition 2. *Consider binomial random variables* $W \sim Bin(\frac{n}{2} - 1, p)$*,* $Z \sim Bin(\frac{n}{2}, q)$*, independent of each other and of* h_u*. Define:*

$$\begin{aligned}
\zeta_{n,+} &\triangleq \mathbb{P}(Z \geq W + \frac{1}{T} h_u + 1) \\
\zeta_{n,-} &\triangleq \mathbb{P}(Z \geq W - \frac{1}{T} h_u + 1) \\
\zeta_n &\triangleq \frac{1}{2}(\zeta_{n,+} + \zeta_{n,-})
\end{aligned}$$

Thus ζ_n *represents the probability of a node being irregular in a graph of size* n*.*

3.2 Exact Recovery Phase Transition

Theorem 7. *Exact recovery is possible if and only if every node* u *is almost-surely regular, i.e.,*

$$\zeta_n = o(n^{-1})$$

Remark 4. *The significance of this result is to show the critical nature of regularity of nodes. Theorem 7 admits arbitrary sequences* $P(\mathbf{Y}_u | x_u = 1)$ *and* $P(\mathbf{Y}_u | x_u = -1)$*, and hence generalizes (Saad and Nosratinia, 2018, Theorem 5) which was only for discrete distributions with finite cardinality.*

3.2.1 Sufficiency of Theorem 7

A modification of a two-step algorithm from (Saad and Nosratinia, 2018, Achievability of Theorem 5) is used. The first step of the original algorithm, which we use without change, produces detected communities A', B' with no more than δn errors with $\delta \to 0$ as $n \to \infty$.

Our method departs from (Saad and Nosratinia, 2018, Theorem 5) in the second step: for a node $u \in A'$, we flip its membership if $E(u, B') \geq E(u, A') + \frac{h_u}{T} + 1 + \gamma \log(n)$, where $\gamma = \frac{1}{\sqrt{-\log(\delta)}}$. For node $v \in B'$, we flip its membership if $E(v, A') \geq E(v, B') - \frac{h_v}{T} + 1 + \gamma \log(n)$. In the end, if the number of flips in A', B' are unequal, undo all flips.

Lemma 16. *The modified two-step algorithm recovers the community labels with probability one if $\zeta_n = o(n^{-1})$.*

Proof. Define two Bernoulli random variables $\tilde{W} \sim Bern(p)$ and $\tilde{Z} \sim Bern(q)$. Define $S_z^L \triangleq \sum_{i=1}^{L} \tilde{Z}_i$ and $S_w^L \triangleq \sum_{i=1}^{L} \tilde{W}_i$. Following (Saad and Nosratinia, 2018, Theorem 5), the misclassification probability is:

$$\begin{aligned}
P_e =& \mathbb{P}\big(\text{node } u \text{ is mislabeled}\big) \\
=& \frac{1}{2}\mathbb{P}\left(S_z^{(1-\delta)\frac{n}{2}} + S_w^{\delta\frac{n}{2}} \geq S_w^{(1-\delta)\frac{n}{2}} + S_z^{\delta\frac{n}{2}} + \frac{1}{T}h_u + 1 + \gamma\log(n)\right) + \\
& \frac{1}{2}\mathbb{P}\left(S_z^{(1-\delta)\frac{n}{2}} + S_w^{\delta\frac{n}{2}} \geq S_w^{(1-\delta)\frac{n}{2}} + S_z^{\delta\frac{n}{2}} - \frac{1}{T}h_u + 1 + \gamma\log(n)\right)
\end{aligned}$$

where $\delta \to 0$ as $n \to \infty$. Using (Saad and Nosratinia, 2018, Lemma 6):

$$\begin{aligned}
P_e \leq& \frac{1}{2}\mathbb{P}\big(S_z^{(1-\delta)\frac{n}{2}} + S_w^{\delta\frac{n}{2}} \geq S_w^{(1-\delta)\frac{n}{2} - \frac{2D}{\log(n)}n} + \quad S_z^{\delta\frac{n}{2} - \frac{2D}{\log(n)}n} + \frac{1}{T}h_u + 1 + \gamma\log(n)\big) + \\
& \frac{1}{2}\mathbb{P}\big(S_z^{(1-\delta)\frac{n}{2}} + S_w^{\delta\frac{n}{2}} \geq S_w^{(1-\delta)\frac{n}{2} - \frac{2D}{\log(n)}n} + \quad S_z^{\delta\frac{n}{2} - \frac{2D}{\log(n)}n} - \frac{1}{T}h_u + 1 + \gamma\log(n)\big) \qquad (3.2)
\end{aligned}$$

where as n grows $D \to \infty$ and $\frac{D}{\log(n)} \to 0$. We now bound the first term of (3.2); the second term follows similarly.

$$\begin{aligned}
& \mathbb{P}\big(S_z^{(1-\delta)\frac{n}{2}} + S_w^{\delta\frac{n}{2}} \geq S_w^{(1-\delta)\frac{n}{2} - \frac{2D}{\log(n)}n} + S_z^{\delta\frac{n}{2} - \frac{2D}{\log(n)}n} + \frac{1}{T}h_u + 1 + \gamma\log(n)\big) \\
\leq& \mathbb{P}\big(S_z^{\frac{n}{2}} + S_w^{\delta\frac{n}{2}} \geq S_w^{(1-\delta)\frac{n}{2} - \frac{2D}{\log(n)}n} + \frac{1}{T}h_u + 1 + \gamma\log(n)\big) \\
\leq& \mathbb{P}\big(S_z^{\frac{n}{2}} - S_w^{\frac{n}{2}} + S_w^{\delta n + \frac{2D}{\log(n)}n} \geq \frac{1}{T}h_u + 1 + \gamma\log(n)\big) \\
\leq& \mathbb{P}\big(S_z^{\frac{n}{2}} - S_w^{\frac{n}{2}} \geq \frac{1}{T}h_u + 1\big) + \mathbb{P}\big(S_w^{\delta n + \frac{2D}{\log(n)}n} \geq \gamma\log(n)\big)
\end{aligned}$$

$$\leq \zeta_{n,+} + \mathbb{P}\big(S_w^{\delta n + \frac{2D}{\log(n)} n} \geq \gamma \log(n)\big) \tag{3.3}$$

where (3.3) uses the definition of $\zeta_{n,+}$. From (Saad and Nosratinia, 2018, Lemma 7),

$$\mathbb{P}\big(S_w^{\delta n + \frac{2D}{\log(n)} n} \geq \gamma \log(n)\big) \leq n^{-(1+\Omega(1))} \tag{3.4}$$

Using (3.3) and (3.4) and substituting in (3.2):

$$P_e \leq \frac{1}{2}\big(\zeta_{n,+} + \zeta_{n,-}\big) + n^{-(1+\Omega(1))} = \zeta_n + n^{-(1+\Omega(1))} \tag{3.5}$$

A simple union bound yields:

$$\mathbb{P}(\text{failure}) \leq n\zeta_n + n^{-\Omega(1)} \tag{3.6}$$

Thus, if $\zeta_n = o(n^{-1})$, then as $n \to \infty$:

$$\mathbb{P}(\text{failure}) \leq o(1) \tag{3.7}$$

which concludes the achievability proof of Theorem 7. □

3.2.2 Necessity of Theorem 7

Lemma 17. *Define Bernoulli random variable M_u indicating the irregularity of node u. Let $Cov(M_u, M_v)$ denote the covariance between M_u and M_v. Then,*

$$|Cov(M_u, M_v)| \leq Cn^{-\frac{2}{3}}\zeta_n^2 + n^{-4} \tag{3.8}$$

for some positive constant C.

Proof. Define the Bernoulli random variable indicating an edge between u, v, i.e., $\delta = E(u, v)$ and note that M_u and M_v are independent conditioned on δ. Therefore, using the law of total covariance,

$$|Cov(M_u, M_v)| = |Cov(\mathbb{E}(M_u|\delta), \mathbb{E}(M_v|\delta))|$$

$$\leq \sqrt{Var(\mathbb{E}(M_u|\delta))Var(\mathbb{E}(M_v|\delta))}$$
$$= Var(\mathbb{E}(M_u|\delta)) \tag{3.9}$$

where (3.9) holds because M_u and M_v have the same distribution conditioned on δ. We now bound one-by-one the terms required for the calculation of $\mathbb{E}(M_u|\delta)$. First, by definition of irregularity,

$$\begin{aligned}\mathbb{P}(M_u = 1|\delta = 0) =&\frac{1}{4}\big(\mathbb{P}(S_z^{\frac{n}{2}} \geq S_w^{\frac{n}{2}-2} + \frac{1}{T}h_u + 1) + \mathbb{P}(S_z^{\frac{n}{2}} \geq S_w^{\frac{n}{2}-2} - \frac{1}{T}h_u + 1)\big)+\\ &\frac{1}{4}\big(\mathbb{P}(S_z^{\frac{n}{2}-1} \geq S_w^{\frac{n}{2}-1} + \frac{1}{T}h_u + 1) + \mathbb{P}(S_z^{\frac{n}{2}-1} \geq S_w^{\frac{n}{2}-1} - \frac{1}{T}h_u + 1)\big)\\ \leq&\frac{1}{4}\big(\mathbb{P}(S_z^{\frac{n}{2}} \geq S_w^{\frac{n}{2}-2} + \frac{1}{T}h_u + 1) + \mathbb{P}(S_z^{\frac{n}{2}} \geq S_w^{\frac{n}{2}-2} - \frac{1}{T}h_u + 1)\big)+\\ &\frac{1}{4}\big(\zeta_{n,+} + \zeta_{n,-}\big)\end{aligned} \tag{3.10}$$

The first two terms in (3.10) can be bounded as follows:

$$\begin{aligned}\zeta_{n,+} =&p\mathbb{P}(S_z^{\frac{n}{2}} \geq 1 + S_w^{\frac{n}{2}-2} + \frac{1}{T}h_u + 1) + (1-p)\mathbb{P}(S_z^{\frac{n}{2}} \geq S_w^{\frac{n}{2}-2} + \frac{1}{T}h_u + 1)\\ \geq&(1-p)\mathbb{P}(S_z^{\frac{n}{2}} \geq S_w^{\frac{n}{2}-2} + \frac{1}{T}h_u + 1)\\ \geq&(1-n^{-\frac{1}{3}})\mathbb{P}(S_z^{\frac{n}{2}} \geq S_w^{\frac{n}{2}-2} + \frac{1}{T}h_u + 1)\end{aligned}$$

resulting in:

$$\mathbb{P}(S_z^{\frac{n}{2}} \geq S_w^{\frac{n}{2}-2} + \frac{1}{T}h_u + 1) \leq (1 + Cn^{\frac{-1}{3}})\zeta_{n,+} \tag{3.11}$$

for some positive constant C. Using (3.11) and substituting in (3.10),

$$\mathbb{P}(M_u = 1|\delta = 0) \leq (1 + Cn^{-\frac{1}{3}})\zeta_n \tag{3.12}$$

Similar to (3.10),

$$\begin{aligned}\mathbb{P}(M_u = 1|\delta = 1) \leq&\frac{1}{4}(1 + Cn^{\frac{-1}{3}})\big(\zeta_{n,+} + \zeta_{n,-}\big) + \frac{1}{4}\big(\mathbb{P}(S_z^{\frac{n}{2}-1} \geq -1 + S_w^{\frac{n}{2}-1} + \frac{1}{T}h_u + 1)+\\ &\mathbb{P}(S_z^{\frac{n}{2}-1} \geq -1 + S_w^{\frac{n}{2}-1} - \frac{1}{T}h_u + 1)\big)\end{aligned} \tag{3.13}$$

In a manner similar to (3.11), and using Lemma 19 from the Appendix:

$$\mathbb{P}(M_u = 1|\delta = 1) \leq C_1 \log(n)\zeta_n + n^{-2} \tag{3.14}$$

for some positive constant C_1. Using (3.12) and (3.14),

$$\begin{aligned} Var(\mathbb{E}(M_u|\delta)) =& \frac{p+q}{2}(\mathbb{P}(M_u = 1|\delta = 1) - \zeta_n)^2 + (1 - \frac{p+q}{2})(\mathbb{P}(M_u = 1|\delta = 0) - \zeta_n)^2 \\ \leq& C_2 n^{-\frac{2}{3}}\zeta_n^2 + n^{-4} \end{aligned} \tag{3.15}$$

for some positive constant C_2. Substituting (3.15) into (3.9):

$$|Cov(M_u, M_v)| \leq Cn^{-\frac{2}{3}}\zeta_n^2 + n^{-4} \tag{3.16}$$

for some positive constant C. This concludes the proof of the lemma. □

Now we prove the converse part of Theorem 7. Let N be the number of nodes that are irregular. Thus, $\mathbb{E}[N] = n\zeta_n$. If ζ_n is not $o(n^{-1})$, then there exists $\epsilon > 0$ such that for infinitely many n, $\mathbb{E}[N] \geq \epsilon$. Moreover,

$$\begin{aligned} Var(N) &= \sum_u Var(M_u) + \sum_{u \neq v} Cov(M_u, M_v) \\ &\leq n\zeta_n + Cn^{-\frac{2}{3}}n^2\zeta_n^2 + n^{-2} \end{aligned} \tag{3.17}$$

where (3.17) holds by Lemma 17. Thus, by the Paley-Zygmund inequality (Paley and Zygmund, 1932), there exists $\delta > 0$ such that $\mathbb{P}(N \geq \delta) \geq \delta$. Thus,

$$\mathbb{P}(\exists u : \text{u is irregular}) = \mathbb{P}(N > 0) \geq \mathbb{P}(N \geq \delta) \geq \delta \tag{3.18}$$

By the symmetry of the labels of the nodes:

$$\begin{aligned} &\mathbb{P}(\exists u : x_u = 1, \text{u is irregular}) \geq \frac{\delta}{2} \\ &\mathbb{P}(\exists u : x_u = -1, \text{u is irregular}) \geq \frac{\delta}{2} \end{aligned} \tag{3.19}$$

Moreover, by Lemma 17, $|Cov(M_u, M_v)| = |\mathbb{P}(M_u = 1, M_v = 1) - \mathbb{P}(M_u = 1)\mathbb{P}(M_v = 1)| \leq Cn^{-\frac{2}{3}}\zeta_n^2$. Thus, $\mathbb{P}(M_u = 1, M_v = 1) \geq \mathbb{P}(M_u = 1)\mathbb{P}(M_v = 1)(1 - o(1))$. This implies that if ζ_n is not $o(n^{-1})$, then for infinitely many n:

$$\mathbb{P}(\exists u, v : x_u \neq x_v \text{ and } u, v \text{ are irregular }) \geq \frac{\delta^2}{4} \tag{3.20}$$

We have shown that if ζ_n is not $o(n^{-1})$, then with probability bounded away from zero, there exist irregular nodes in both communities. This implies that exact recovery fails with high probability, as shown below.

Lemma 18. *Let A and B denote the true communities. Define the following events:*

$$F \triangleq \{\textit{Maximum Likelihood fails}\}$$
$$F_A \triangleq \{\exists u \in A : u \textit{ is irregular}\}$$
$$F_B \triangleq \{\exists v \in B : v \textit{ is irregular}\}$$

Then, $F_A \cap F_B \Rightarrow F$.

Proof. Note that $F_A \cap F_B$ denote an event where there exist nodes with different labels that are irregular. Define two new communities $\hat{A} = A\backslash\{u\} \cup \{v\}$ and $\hat{B} = B\backslash\{v\} \cup \{u\}$. Let $A_{uv} \sim Bern(q)$ be a random variable representing the existence of the edge between nodes u and v. Then, by a direct computation of the likelihood function:

$$\begin{aligned}
\log\big(\mathbb{P}(G, \mathbf{Y}|\hat{A}, \hat{B})\big) \stackrel{(a)}{=} & R + T\big(E(\hat{A}) + E(\hat{B})\big) + \sum_{i \in \hat{A}\backslash\{v\}} \log(\mathbb{P}(\mathbf{Y}_i|x_i = 1)) + \\
& \sum_{j \in \hat{B}\backslash\{u\}} \log(\mathbb{P}(\mathbf{Y}_j|x_j = -1)) + \log(\mathbb{P}(\mathbf{Y}_u|x_u = 1)) + \\
& \log(\mathbb{P}(\mathbf{Y}_v|x_v = -1)) + h_v - h_u \\
= & R + T\big(E(A) + E(B)\big) + \log\big(\mathbb{P}(\mathbf{Y}|A, B)\big) - 2TA_{uv} + \\
& T\big(E[v, A] - E[v, B] + E[u, B] - E[u, A]\big) + h_v - h_u
\end{aligned}$$

$$\overset{(b)}{\geq} \log\big(\mathbb{P}(G, \mathbf{Y}|A, B)\big) + 2T(1 - A_{uv}) \overset{(c)}{\geq} \log\big(\mathbb{P}(G, \mathbf{Y}|A, B)\big)$$

where (a) holds for some constant R, (b) holds by the assumption that $F_A \cap F_B$ happened and (c) holds because $(1 - A_{ij}) \geq 0$ and $T \geq 0$. The inequality (c) implies the failure of maximum likelihood, which concludes the proof of the necessary condition of Theorem 7. $\square$

3.3 Closing the Gap in (Saad and Nosratinia, 2018, Theorem 5)

Consider a system model where each node observes K features (random variables) each with cardinality $M_k < \infty$. The alphabet for each feature k is denoted with $\{u_1^k, u_2^k, \cdots, u_{M_k}^k\}$. For each node i and feature k, define

$$\begin{aligned} \alpha_{+,m_k}^k &\triangleq \mathbb{P}(y_{i,k} = u_{m_k}^k | x_i = 1) \\ \alpha_{-,m_k}^k &\triangleq \mathbb{P}(y_{i,k} = u_{m_k}^k | x_i = -1) \\ h_{\ell_m}^m &\triangleq \log \alpha_{+,m_k}^k - \log \alpha_{-,m_k}^k \end{aligned}$$

Features are mutually independent conditioned on labels. Define the following functions of statistics of side information:

$$f_1(n) \triangleq \textstyle\sum_{k=1}^K h_{\ell_m}^m$$

$$f_2(n) \triangleq \textstyle\sum_{k=1}^K \log(\alpha_{+,m_k}^k) \qquad f_3(n) \triangleq \textstyle\sum_{k=1}^K \log(\alpha_{-,m_k}^k)$$

Theorem 8. *Assume α_{+,m_k}^k and α_{-,m_k}^k are either constant or monotonically increasing or decreasing in n. Then, necessary and sufficient conditions for exact recovery are as follows.*[2]

1. *Items* $(1) - (4)$ *in (Saad and Nosratinia, 2018, Theorem 5) must hold.*

[2]In this theorem the side information outcomes $[u_{m_1}^1, \ldots, u_{m_K}^K]$ are represented by their index $[m_1, \ldots, m_K]$. Dependence on n is implicit.

2. *If there exists a sequence (over n) of side information outcomes* $[m_1, \ldots, m_K]$ *such that* $f_1(n) = \beta \log(n) + o(\log(n)), |\beta| > T\frac{(a-b)}{2}$, *and furthermore* $f_2(n) = o(\log(n))$ *if* $\beta > 0$ *and* $f_3(n) = o(\log(n))$ *if* $\beta < 0$, *then* $\beta > 1$ *must hold.*

3. *If there exists a sequence (over n) of side information outcomes* $[m_1, \ldots, m_K]$ *such that* $f_1(n) = \beta_1 \log(n) + o(\log(n)), |\beta_1| > T\frac{(a-b)}{2}$, *and furthermore* $f_3(n) = -\beta_1' \log(n) + o(\log(n))$ *if* $\beta_1 > 0$ *and* $f_2(n) = -\beta_1' \log(n) + o(\log(n))$ *if* $\beta_1 < 0$, *then* $\beta_1' > 1$ *must hold.*

Proof. Using Theorem 7, it is sufficient to show that $\zeta_n = o(n^{-1})$ if and only if the three items in Theorem 8 hold. The first item was proved in (Saad and Nosratinia, 2018, Theorem 5), and we consider the last two items. By definition:

$$\zeta_n = \frac{1}{2}\Bigg(\sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \cdots \sum_{m_K=1}^{M_K} \Bigg(\prod_{k=1}^{K} (\alpha_{+,m_k}^k) \times \mathbb{P}\Bigg(\sum_{l=1}^{\frac{n}{2}} (Z_l - W_l) \geq \sum_{k=1}^{K} \frac{h_{\ell_m}^m}{T} + 1 \Bigg) + \prod_{k=1}^{K} (\alpha_{-,m_k}^k) \times \mathbb{P}\Bigg(\sum_{l=1}^{\frac{n}{2}} (Z_l - W_l) \geq -\sum_{k=1}^{K} \frac{h_{\ell_m}^m}{T} + 1 \Bigg)\Bigg)\Bigg) \tag{3.21}$$

Assume $\sum_{k=1}^{K} h_{\ell_m}^m = \beta \log(n) + o(\log(n)), \beta > T\frac{(a-b)}{2}$. The case when $\beta < -T\frac{(a-b)}{2}$ holds similarly. Without loss of generality, we focus on one term inside the nested sum in (3.21). Using Chernoff bound, ζ_n can be upper bounded as follows:

$$\begin{aligned} P_1 =& \prod_{k=1}^{K} (\alpha_{+,m_k}^k) \mathbb{P}\Bigg(\sum_{l=1}^{\frac{n}{2}} (Z_l - W_l) \geq \sum_{k=1}^{K} \frac{h_{\ell_m}^m}{T} + 1 \Bigg) + \\ & \prod_{k=1}^{K} (\alpha_{-,m_k}^k) \mathbb{P}\Bigg(\sum_{l=1}^{\frac{n}{2}} (Z_l - W_l) \geq -\sum_{k=1}^{K} \frac{h_{\ell_m}^m}{T} + 1 \Bigg) \\ \leq & \prod_{k=1}^{K} (\alpha_{+,m_k}^k) e^{-\sup_{t_1>0} \frac{t_1(\beta+o(1))}{T} \log(n) - \frac{n}{2} \log(\mathbb{E}[e^{t_1(Z-W)}])} + \\ & \prod_{k=1}^{K} (\alpha_{-,m_k}^k) e^{-\sup_{t_2>0} \frac{-t_2(\beta+o(1))}{T} \log(n) - \frac{n}{2} \log(\mathbb{E}[e^{t_2(Z-W)}])} \\ \stackrel{(a)}{\leq} & \prod_{k=1}^{K} (\alpha_{+,m_k}^k) e^{-\log(n) \sup_{t_1>0} \frac{t_1(\beta+o(1))}{T} + \frac{1}{2}(a+b-ae^{-t_1}-be^{t_1})} + \end{aligned}$$

$$\prod_{k=1}^{K}(\alpha_{-,m_k}^{k})e^{-\log(n)\sup_{t_2>0}\frac{-t_2(\beta+o(1))}{T}+\frac{1}{2}(a+b-ae^{-t_2}-be^{t_2})}$$

where (a) follows because $\log(1-x) \leq -x$. Since $\beta > \frac{T(a-b)}{2}$, supremum is achieved at $t_1^* = \log(\frac{\gamma+\beta}{bT})$ and $t_2^* = 0$, where $\gamma = \sqrt{\beta^2 + abT^2}$. Thus,

$$P_1 \leq \prod_{k=1}^{K}(\alpha_{+,m_k}^{k})n^{-(1+o(1))\eta(a,b,\beta)} + \prod_{k=1}^{K}(\alpha_{-,m_k}^{k}) \tag{3.22}$$

Consider two scenarios for $\prod_{k=1}^{K}(\alpha_{+,m_k}^{k})$. First, when $\log(\prod_{k=1}^{K}(\alpha_{+,m_k}^{k})) = o(\log(n))$, then $\log(\prod_{k=1}^{K}(\alpha_{-,m_k}^{k})) = -(\beta + o(1))\log(n)$. Thus,

$$P_1 \leq n^{-(\frac{1}{2}+o(1))\eta(a,b,\beta)} + n^{-\beta+o(1)} \leq n^{-\beta+o(1)} \tag{3.23}$$

where (3.23) holds by (Saad and Nosratinia, 2018, Lemma 8). For the lower bound, we use (Saad and Nosratinia, 2018, Lemma 15) resulting in:

$$P_1 \geq \prod_{k=1}^{K}(\alpha_{+,m_k}^{k})e^{-(1+o(1))\log(n)\sup_{t>0}\frac{t\beta}{T}+\frac{1}{2}(a+b-ae^{-t}-be^{t})} +$$
$$\prod_{k=1}^{K}(\alpha_{-,m_k}^{k})e^{-(1+o(1))\log(n)\sup_{t>0}\frac{-t\beta}{T}+\frac{1}{2}(a+b-ae^{-t}-be^{t})}$$

In a manner similar to (3.22):

$$P_1 \geq n^{-(\frac{1}{2}+o(1))\eta(a,b,\beta)} + n^{-\beta+o(1)} \geq n^{-\beta+o(1)} \tag{3.24}$$

Combining (3.23) and (3.24) concludes the proof for the case when $\log(\prod_{k=1}^{K}(\alpha_{+,m_k}^{k})) = o(\log(n))$. When $\log(\prod_{k=1}^{K}(\alpha_{+,m_k}^{k})) = -(\beta' + o(1))\log(n)$ for some positive β'. Then, $\log(\prod_{k=1}^{K}(\alpha_{-,m_k}^{k})) = -(\beta''+o(1))\log(n)$ such that $\beta = \beta'' - \beta'$. In a manner similar to (3.23):

$$P_1 \leq n^{-(\frac{1}{2}+o(1))\eta(a,b,\beta)-\beta'} + n^{-\beta''+o(1)} \leq n^{-\beta''+o(1)} \tag{3.25}$$

where (3.25) holds by (Saad and Nosratinia, 2018, Lemma 8). For the lower bound, in a manner similar to (3.24):

$$P_1 \geq n^{-(\frac{1}{2}+o(1))\eta(a,b,\beta)-\beta'} + n^{-\beta''+o(1)} \geq n^{-\beta''+o(1)} \tag{3.26}$$

Combining (3.25) and (3.26) concludes the proof. □

3.4 Appendix

Lemma 19. *Let $Z \sim Bin(n,q)$ and $W \sim Bin(n,p)$ be independent. Let Y be another random variable independent of Z and W. Also, define ψ to be a constant. Then,*

$$\mathbb{P}(Z \geq W - 1 + Y + \psi) \leq C \log(n) \mathbb{P}(Z \geq W + Y + \psi) + n^{-2} \tag{3.27}$$

for some positive constant C.

Proof. By rewriting the left hand side of (3.27),

$$\begin{aligned}
\mathbb{P}(W \leq Z + 1 - Y - \psi) &= \sum_{k=0}^{n} \mathbb{P}(W \leq k + 1 - Y - \psi)\mathbb{P}(Z = k) \\
&= \sum_{k'=0}^{\log^2(n)} \mathbb{P}(W \leq k' - Y - \psi)\mathbb{P}(Z = k' - 1) + \\
&\qquad \sum_{k'=\log^2(n)+1}^{n+1} \mathbb{P}(W \leq k' - Y - \psi)\mathbb{P}(Z = k' - 1)
\end{aligned} \tag{3.28}$$

Now, for $k' = o(n)$, we have:

$$\begin{aligned}
\log\left(\frac{\mathbb{P}(Z = k' - 1)}{\mathbb{P}(Z = k')}\right) &= \log\left(\frac{\binom{n}{k'-1} q^{k'-1}(1-q)^{n-k'+1}}{\binom{n}{k'} q^{k'}(1-q)^{n-k'}}\right) \\
&= \log\left(\frac{n(1-q)}{(n-k'+1)}\right) + \log\left(\frac{k'}{nq}\right) \\
&\leq \log\left(\frac{k'}{np}\right)
\end{aligned} \tag{3.29}$$

where (3.29) holds for sufficiently large n, since $k' = o(n)$. Thus, substituting in (3.28),

$$\begin{aligned}
\mathbb{P}(W \leq Z + 1 - Y - \psi) \leq &\sum_{k'=0}^{\log^2(n)} \frac{k'}{np} \mathbb{P}(W \leq k' - Y - \psi)\mathbb{P}(Z = k') + \\
&\sum_{k'=\log^2(n)+1}^{n+1} \mathbb{P}(W \leq k' - Y - \psi)\mathbb{P}(Z = k' - 1) \\
\leq &\sum_{k'=0}^{\log^2(n)} \frac{\log^2(n)}{np} \mathbb{P}(W \leq k' - Y - \psi)\mathbb{P}(Z = k') +
\end{aligned}$$

$$
\begin{aligned}
&\sum_{k'=\log^2(n)+1}^{n+1} \mathbb{P}(Z = k' - 1) \\
&\overset{(a)}{\leq} C \log(n) \sum_{k'=0}^{n} \mathbb{P}(W \leq k' - Y - \psi)\mathbb{P}(Z = k') + \\
&\quad \mathbb{P}(Z \geq \log^2(n)) \\
&\overset{(b)}{\leq} C \log(n)\mathbb{P}(W \leq Z - Y - \psi) + n^{-2} \qquad (3.30)
\end{aligned}
$$

where (a) holds because $np \approx \log(n)$ and (b) holds by Chernoff bound. $\square$

CHAPTER 4

SINGLE COMMUNITY DETECTION[1] [2]

In this chapter, the stochastic block model for one community is considered (Montanari, 2015; Hajek et al., 2017, 2018; Kadavankandy et al., 2018). The stochastic block model for one community consists of a graph of size n with a community of size K, where $K = o(n)$). The problem of finding a hidden community upon observing *only* the graph has been studied in (Montanari, 2015; Hajek et al., 2017, 2018). The information limits[3] of weak recovery and exact recovery have been studied in (Hajek et al., 2017). 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 (Hajek et al., 2018; Montanari, 2015) in terms of a signal-to-noise ratio parameter. The utility of a voting procedure after belief propagation to achieve exact recovery was pointed out in (Hajek et al., 2018).

4.1 System Model and Assumptions

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

[3]The extremal phase transition threshold is also known as *information theoretic limit* (Abbe and Sandon, 2015) or *information limit* (Hajek et al., 2017). We use the latter term throughout this chapter.

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 $\boldsymbol{y}_m = [y_{1,m}, \ldots, y_{n,m}]^t$ collects the side information feature m for all nodes i.

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

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

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

In this chapter, 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 (Hajek et al., 2017) 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 this chapter.

4.2 Information Limits

4.2.1 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 $i = 1, \ldots, n$, define the distributions:

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

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:

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

Throughout the chapter, 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.

Definition 3.

$$\psi_{QU}(t, m_1, m_2) \triangleq m_1 \log(\mathbb{E}_Q[e^{tL_G}]) + m_2 \log(\mathbb{E}_U[e^{tL_S}]) \tag{4.1}$$

$$\psi_{PV}(t, m_1, m_2) \triangleq m_1 \log(\mathbb{E}_P[e^{tL_G}]) + m_2 \log(\mathbb{E}_V[e^{tL_S}]) \tag{4.2}$$

$$E_{QU}(\theta, m_1, m_2) \triangleq \sup_{t\in[0,1]} t\theta - \psi_{QU}(t, m_1, m_2) \tag{4.3}$$

$$E_{PV}(\theta, m_1, m_2) \triangleq \sup_{t\in[-1,0]} t\theta - \psi_{PV}(t, m_1, m_2) \tag{4.4}$$

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

Weak Recovery

Theorem 1. *For single community detection under bounded-LLR side information, if:*

$$\begin{gathered}(K-1)D(P||Q)+MD(V||U)\rightarrow\infty\ ,\\ \liminf_{n\rightarrow\infty}\frac{(K-1)D(P||Q)+2MD(V||U)}{\log(\frac{n}{K})}>2\end{gathered}\tag{4.5}$$

then weak recovery is achieved and if weak recovery is achieved, then:

$$\begin{gathered}(K-1)D(P||Q)+MD(V||U)\rightarrow\infty\ ,\\ \liminf_{n\rightarrow\infty}\frac{(K-1)D(P||Q)+2MD(V||U)}{\log(\frac{n}{K})}\geq 2\end{gathered}\tag{4.6}$$

Proof. For necessity please see Appendix 4.4.2. For sufficiency, please see Appendix 4.4.3. □

Remark 1. *The condition of bounded support for the LLRs can be somewhat weakened to Eqs.* (4.67) *and* (4.70). *As an example* $U\sim\mathcal{N}(0,1)$ *and* $V\sim\mathcal{N}(\mu,1)$ *with* $\mu\neq 0$ *satisfies* (4.67), (4.70) *and the theorem continues to hold even though the LLR is not bounded.*

Remark 2. *Theorem 1 shows that if M grows with n slowly enough, e.g., if M is fixed and independent of n, or if $M=o(\log(\frac{n}{K}))$, side information does not affect the information limits.*

Remark 3. *If the features are conditionally independent but* not *identically distributed, it is easy to show the necessary and sufficient conditions are:*

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

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

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 20. *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:*

$$\mathbb{P}\Big(\Big|\ |C^*| - K\Big| \leq \frac{K}{\log(K)}\Big) \geq 1 - o(1) \tag{4.7}$$

for some known $K = o(n)$. If conditions (4.5) *hold, then:*

$$\mathbb{P}\Big(\frac{|\hat{C} \triangle C^*|}{K} \leq 2\epsilon + \frac{1}{\log(K)}\Big) \geq 1 - o(1) \tag{4.8}$$

where

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

Proof. Please see Appendix 4.4.4 □

Exact Recovery

The sufficient conditions for exact recovery are derived using a two-step algorithm (see Table 4.1). Its first step consists of any algorithm achieving weak recovery, e.g. maximum likelihood (see Lemma 20). The second step applies a local voting procedure.

Lemma 21. *Define $C_k^* = C^* \cap {S_k}^c$ and assume $\hat{C}_k$ achieves weak recovery, i.e.*

$$\mathbb{P}\big(|\hat{C}_k \triangle C_k^*| \leq \delta K \ \textit{for}\ 1 \leq k \leq \frac{1}{\delta}\big) \to 1\ . \tag{4.9}$$

If

$$\liminf_{n\to\infty} \frac{E_{QU}\big(\log(\frac{n}{K}), K, M\big)}{\log(n)} > 1 \tag{4.10}$$

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

Proof. Please see Appendix 4.4.5. □

Table 4.1. 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 $\lvert S_k\rvert = n\delta$. $\boldsymbol{G}_k$ 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 $\lvert\hat{C}_k\rvert = \lceil K(1-\delta)\rceil$ 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 $r_i\}$.

Then the main result of this section follows:

Theorem 2. *In single community detection under bounded-LLR side information, assume* (4.5) *holds, if*

$$\liminf_{n\to\infty} \frac{E_{QU}\big(\log(\frac{n}{K}), K, M\big)}{\log(n)} > 1 \tag{4.11}$$

then exact recovery is achieved, and if exact recovery is achieved, then:

$$\liminf_{n\to\infty} \frac{E_{QU}\big(\log(\frac{n}{K}), K, M\big)}{\log(n)} \geq 1 \tag{4.12}$$

Proof. For sufficiency, please see Appendix 4.4.6. For necessity see Appendix 4.4.7. □

Remark 4. *The assumption that* (4.5) *holds is necessary because otherwise weak recovery is not achievable, and by extension, exact recovery.*

Remark 5. *Theorem 2 shows if M grows with n slowly enough, e.g., M is fixed and independent of n or $M = o(K)$, side information will not affect the information limits of exact recovery.*

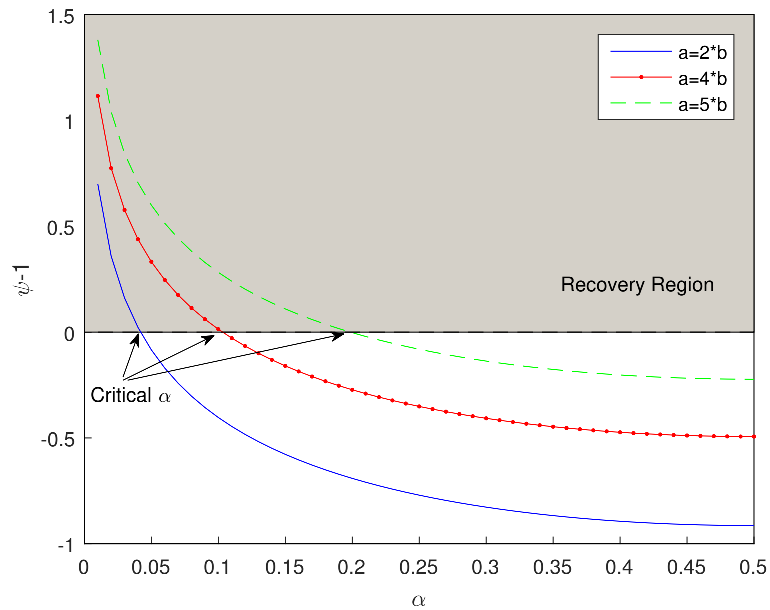

Figure 4.1. Exact recovery threshold, $\psi - 1$ for different values of α at $c = b = 1$.

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

$$K = \frac{cn}{\log(n)}, \quad q = \frac{b\log^2(n)}{n}, \quad p = \frac{a\log^2(n)}{n} \tag{4.13}$$

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:

$$\sup_{t\in[0,1]} tc(a-b) + bc - bc(\frac{a}{b})^t > 1 \tag{4.14}$$

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

$$\sup_{t\in[0,1]} tc(a-b) + bc - bc(\frac{a}{b})^t - \frac{M}{\log(n)}\log((1-\alpha)^t\alpha^{(1-t)} + (1-\alpha)^{(1-t)}\alpha^t) > 1 \tag{4.15}$$

If $M = o(\log(n))$, then (4.15) reduces to (4.14), 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 (4.15) is less restrictive than (4.14), thus improving the information limit.

Let ψ denote the left hand side of (4.15) with $M = \log(n)$, i.e.,

$$\psi = \sup_{t\in[0,1]} tc(a-b) + bc - bc(\frac{a}{b})^t - \log((1-\alpha)^t\alpha^{(1-t)} + (1-\alpha)^{(1-t)}\alpha^t) \tag{4.16}$$

The behavior of ψ against α describes the influence of side information on exact recovery and is depicted in Fig. 4.1.

4.2.2 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.

Weak Recovery

Recall that the probability distribution side information feature m is V_m when the node is inside and outside the community, and U_m when the node is outside the community.

Theorem 3 (Necessary Conditions for Weak Recovery)**.** *For single community detection under bounded-LLR side information, weak recovery is achieved only if:*

$$\begin{gathered}(K-1)D(P||Q) + \sum_{m=1}^{M}(D(V_m||U_m) + D(U_m||V_m)) \to \infty \\ \liminf_{n\to\infty} \frac{(K-1)D(P||Q) + 2\sum_{m=1}^{M} D(V_m||U_m)}{\log(\frac{n}{K})} \geq 2\end{gathered} \tag{4.17}$$

Proof. The proof follows similar to Theorem 1. □

Exact Recovery

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.

$$K = \rho \frac{n}{\log(n)}, \qquad p = a \frac{\log(n)^2}{n} \qquad q = b \frac{\log(n)^2}{n} \tag{4.18}$$

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:

$$\alpha_{+,\ell_m}^m \triangleq \mathbb{P}(y_{i,m} = u_{\ell_m}^m | x_i = 1) \tag{4.19}$$

$$\alpha_{-,\ell_m}^m \triangleq \mathbb{P}(y_{i,m} = u_{\ell_m}^m | x_i = 0) \tag{4.20}$$

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_{+,ℓ_m}^m and α_{-,ℓ_m}^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; extension to multiple features is straight forward. 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:

$$T \triangleq \log\big(\frac{a}{b}\big) \tag{4.21}$$

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

$$\eta_1(\rho, a, b) \triangleq \rho\Big(b + \frac{a-b}{T}\log\big(\frac{a-b}{ebT}\big)\Big) \tag{4.22}$$

$$\eta_2(\rho, a, b, \beta) \triangleq \rho b + \frac{\rho(a-b)-\beta}{T}\log\big(\frac{\rho(a-b)-\beta}{\rho ebT}\big) + \beta \tag{4.23}$$

$$\eta_3(\rho, a, b, \beta) \triangleq \rho b + \frac{\rho(a-b)+\beta}{T}\log\big(\frac{\rho(a-b)+\beta}{\rho ebT}\big) \tag{4.24}$$

For example in the regime (4.18), one can conclude using (4.11) that exact recovery without side information is achieved if and only if $\eta_1 > 1$.

The LLR of each feature is denoted:

$$h^m_{\ell_m} \triangleq \log\Big(\frac{\alpha^m_{+,\ell_m}}{\alpha^m_{-,\ell_m}}\Big) \tag{4.25}$$

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 (Saad and Nosratinia, 2018).

$$f_1(n) \triangleq \sum_{m=1}^{M} h^m_{\ell_m}, \tag{4.26}$$

$$f_2(n) \triangleq \sum_{m=1}^{M} \log(\alpha^m_{+,\ell_m}), \tag{4.27}$$

$$f_3(n) \triangleq \sum_{m=1}^{M} \log(\alpha^m_{-,\ell_m}) \tag{4.28}$$

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 4. *In the regime characterized by* (4.18)*, assume M is constant and α^m_{+,ℓ_m} and α^m_{-,ℓ_m} are either constant or monotonic in n. Then, necessary and sufficient conditions for exact recovery depend on side information statistics in the following manner:*

1. *If there exists any sequence (over n) of side information outcomes $[\ell_1, \ldots, \ell_M]$ such that $f_1(n)$, $f_2(n)$, $f_3(n)$ are all $o(\log(n))$, then $\eta_1(\rho, a, b) > 1$ must hold.*

2. *If there exists any sequence (over n) of side information outcomes $[\ell_1, \ldots, \ell_M]$ such that $f_1(n) = o(\log(n))$ and $f_2(n), f_3(n)$ evolve according to $-\beta \log(n) + o(\log(n))$ with $\beta > 0$, then $\eta_1(\rho, a, b) + \beta > 1$ must hold.*

3. *If there exists any sequence (over n) of side information outcomes $[\ell_1, \ldots, \ell_M]$ such that $f_1(n) = \beta_1 \log(n) + o(\log(n))$ with $0 < \beta_1 < \rho(a - b - bT)$ and furthermore $f_2(n) = o(\log(n))$, then $\eta_2(\rho, a, b, \beta_1) > 1$ must hold.*

4. *If there exists any sequence (over n) of side information outcomes $[\ell_1, \ldots, \ell_M]$ such that $f_1(n) = \beta_2 \log(n) + o(\log(n))$ with $0 < \beta_2 < \rho(a - b - bT)$ and furthermore $f_3(n) = o(\log(n))$, then $\eta_3(\rho, a, b, \beta_2) > 1$ must hold.*

5. *If there exists any sequence (over n) of side information outcomes $[\ell_1, \ldots, \ell_M]$ such that $f_1(n) = \beta_3 \log(n) + o(\log(n))$ with $0 < \beta_3 < \rho(a - b - bT)$ and furthermore $f_2(n) = -\beta'_3 \log(n) + o(\log(n))$, then $\eta_2(\rho, a, b, \beta_3) + \beta'_3 > 1$ must hold.*

6. *If there exists any sequence (over n) of side information outcomes $[\ell_1, \ldots, \ell_M]$ such that $f_1(n) = \beta_4 \log(n) + o(\log(n))$ with $0 < \beta_4 < \rho(a - b - bT)$ and furthermore $f_3(n) = -\beta'_4 \log(n) + o(\log(n))$, then $\eta_3(\rho, a, b, \beta_4) + \beta'_4 > 1$ must hold.*

Proof. For necessity, see Appendix 4.4.8. For sufficiency, see Appendix 4.4.9. □

Remark 6. *The six items in Theorem 4 are concurrent. For example, if some side information outcome sequences fall under Item 2 and some fall under Item 3, 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 7. *Theorem 4 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 4 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 4 combined with Lemma 36 state that exact recovery is achieved if and only if:*

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

Figure 4.2 shows the error exponent for the noisy label side information as a function of β.

Corollary 2. *For side information consisting of a fraction* $1 - \epsilon$ *of the labels revealed, Theorem 4 states that exact recovery is achieved if and only if:*

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

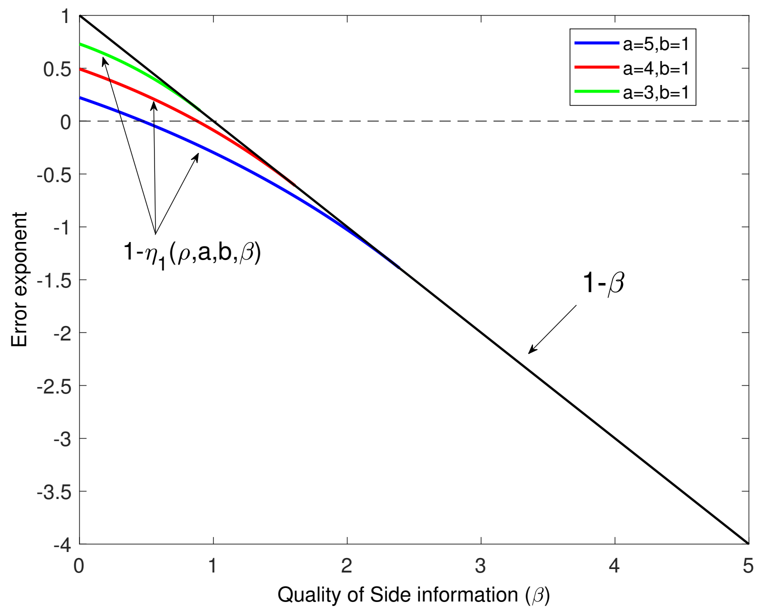

Figure 4.2. Error exponent for noisy side information.

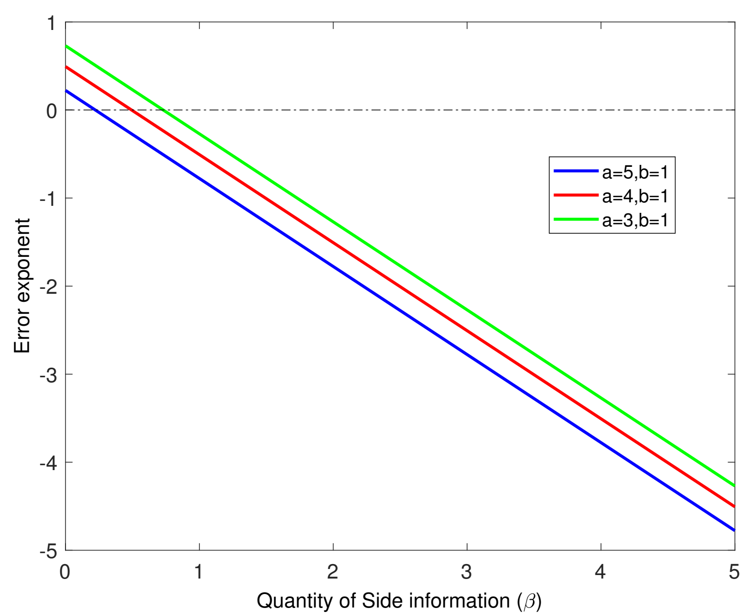

Figure 4.3. Error exponent for partially revealed side information.

Figure 4.3 shows the error exponent for partially revealed labels, as a function of β.

We now comment on the coverage of the regime (4.18). 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 $\Omega(\log n)$. This restricts Kp and Kq in a manner that is reflected in (4.18). Beyond that, in the system model of this chapter $K = o(n)$, so $\frac{\log(\frac{n}{K})}{\log(n)}$ is either $o(1)$ or approaching a constant $C \in (0, 1]$. The regime (4.18) 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.

4.3 Belief Propagation

Belief propagation for recovering a single community was studied *without* side information in (Hajek et al., 2018; Montanari, 2015) 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 straight forward 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:

$$\Lambda \triangleq \sum_{\ell=1}^{L} \frac{\alpha_{+,\ell}^2}{\alpha_{-,\ell}} \tag{4.29}$$

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

4.3.1 Bounded LLR

We begin by demonstrating the performance of belief propagation algorithm on a random tree with side information. Then, we show that the same performance is possible on a random *graph* drawn from $\mathcal{G}(n, K, p, q)$, using a coupling lemma (Hajek et al., 2018) expressing local approximation of random graphs by trees.

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_i^t. 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:

$$p_e^t \triangleq \frac{K}{n}\mathbb{P}(\hat{\tau}_0 = 0|\tau_0 = 1) + \frac{n-K}{n}\mathbb{P}(\hat{\tau}_0 = 1|\tau_0 = 0) \tag{4.30}$$

The maximum a posteriori (MAP) detector minimizes p_e^t and can be written in terms of the log-likelihood ratio as $\hat{\tau}_{MAP} = \mathbb{1}_{\{\Gamma_0^t \geq \nu\}}$, where $\nu = \log(\frac{n-K}{K})$ and:

$$\Gamma_0^t = \log\left(\frac{\mathbb{P}(T^t, \tilde{\boldsymbol{\tau}}^t | \tau_0 = 1)}{\mathbb{P}(T^t, \tilde{\boldsymbol{\tau}}^t | \tau_0 = 0)}\right) \tag{4.31}$$

The probability of error of the MAP estimator can be bounded as follows (Kobayashi and Thomas, 1967):

$$\frac{K(n-K)}{n^2}\rho^2 \leq p_e^t \leq \frac{\sqrt{K(n-K)}}{n}\rho \tag{4.32}$$

where $\rho = \mathbb{E}\left[e^{\frac{\Gamma_0^t}{2}} \middle| \tau_0 = 0\right]$.

Lemma 22. *Let $\mathcal{N}_i$ denote the children of node i, $N_i \triangleq |\mathcal{N}_i|$ and $h_i \triangleq \log\left(\frac{\mathbb{P}(\tilde{\tau}_i|\tau_i=1)}{\mathbb{P}(\tilde{\tau}_i|\tau_i=0)}\right)$. Then,*

$$\Gamma_i^{t+1} = -K(p-q) + h_i + \sum_{k \in \mathcal{N}_i} \log\left(\frac{\frac{p}{q}e^{\Gamma_k^t - \nu} + 1}{e^{\Gamma_k^t - \nu} + 1}\right) \tag{4.33}$$

Proof. See Appendix 4.4.12 □

Lower and Upper Bounds on ρ

Define for $t \geq 1$ and any node i:

$$\psi_i^t = -K(p-q) + \sum_{j \in \mathcal{N}_i} M(h_j + \psi_j^{t-1}) \tag{4.34}$$

where

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

Then, $\Gamma_i^{t+1} = h_i + \psi_i^{t+1}$ and $\psi_i^0 = 0\ \forall i \in T^t$. Let Z_0^t and Z_1^t denote random variables drawn according to the distribution of ψ_i^t 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, $\rho = \mathbb{E}\left[e^{\frac{1}{2}(Z_0^t + U_0)}\right] = \mathbb{E}\left[e^{\frac{U_0}{2}}\right]\mathbb{E}\left[e^{\frac{Z_0^t}{2}}\right]$. Define:

$$b_t \triangleq \mathbb{E}\left[\frac{e^{Z_1^t + U_1}}{1 + e^{Z_1^t + U_1 - \nu}}\right] \tag{4.35}$$

$$a_t \triangleq \mathbb{E}\left[e^{Z_1^t + U_1}\right] \tag{4.36}$$

Lemma 23. *Let $B = (\frac{p}{q})^{1.5}$. Then:*

$$\mathbb{E}[e^{\frac{U_0}{2}}]e^{\frac{-\lambda}{8}b_t} \leq \rho \leq \mathbb{E}[e^{\frac{U_0}{2}}]e^{\frac{-\lambda}{8B}b_t} \tag{4.37}$$

Proof. See Appendix 4.4.13. □

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

Lemma 24. *For all $t \geq 0$, if $\lambda \leq \frac{1}{\Lambda e}$, then $b_t \leq \Lambda e$.*

Proof. See Appendix 4.4.14. □

Lemma 25. *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,*

$$b_{t+1} \geq \Lambda e^{\lambda b_t}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu}{2}}) \tag{4.38}$$

Proof. See Appendix 4.4.15. □

Lemma 26. *The sequences a_t and b_t are non-decreasing in t.*

Proof. The proof follows directly from (Hajek et al., 2018, Lemma 5). □

Lemma 27. *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:*

$$b_{\bar{t}_o+\log^*(\nu)+2} \geq \Lambda e^{\frac{\lambda\nu}{2(C-\lambda)}}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu}{2}}) \tag{4.39}$$

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

Proof. See Appendix 4.4.16. □

Achievability and Converse for the MAP Detector

Lemma 28. *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:*

$$p_e^t \geq \frac{K(n-K)}{n^2}\mathbb{E}^2[e^{\frac{U_0}{2}}]e^{\frac{-\lambda\Lambda e}{4}} \tag{4.40}$$

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

$$p_e^t \leq \sqrt{\frac{K(n-K)}{n^2}}\mathbb{E}[e^{\frac{U_0}{2}}]e^{\frac{-\lambda\Lambda}{8B}e^{\frac{\lambda\nu}{2(C-\lambda)}}(1-\frac{\Lambda'}{\Lambda}e^{\frac{-\nu}{2}})} \tag{4.41}$$

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

$$p_e^t \leq \sqrt{\frac{K(n-K)}{n^2}}\mathbb{E}[e^{\frac{U_0}{2}}]e^{-\nu(r+\frac{1}{2})} = \frac{K}{n}e^{-\nu(r+o(1))} \tag{4.42}$$

for some $r > 0$.

Proof. The proof follows directly from (4.32) and Lemmas 24 and 27. □

Belief Propagation Algorithm 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 (Hajek et al., 2018) 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 4.3.1 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 29 (Coupling Lemma (Hajek et al., 2018)). *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:*

$$\mathbb{P}((\boldsymbol{G}_u^{\hat{t}}, \boldsymbol{x}_u^{\hat{t}}, \boldsymbol{Y}_u^{\hat{t}}) = (T^{\hat{t}}, \boldsymbol{\tau}^{\hat{t}}, \tilde{\boldsymbol{\tau}}^{\hat{t}})) \geq 1 - n^{-1+o(1)} \tag{4.43}$$

  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:*

$$\mathbb{P}((\boldsymbol{G}_u^{\hat{t}}, \boldsymbol{x}_u^{\hat{t}}, \boldsymbol{Y}_u^{\hat{t}}) = (T^{\hat{t}}, \boldsymbol{\tau}^{\hat{t}}, \tilde{\boldsymbol{\tau}}^{\hat{t}})) \geq 1 - n^{\frac{-1}{2}+o(1)} \tag{4.44}$$

Now, we are ready to present 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:

$$R_{i\to j}^{t+1} = h_i - K(p-q) + \sum_{k\in\mathcal{N}_i\setminus j} M(R_{k\to i}^t) \tag{4.45}$$

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 and $M(x) = \log(\frac{\frac{p}{q}e^{x-\nu}+1}{e^{x-\nu}+1})$. The messages are initialized to zero for all nodes i, i.e., $R_{i\to j}^0 = 0$ for all $i \in \{1, \cdots, n\}$ and $j \in \mathcal{N}_i$. Define the belief of node i at iteration $t+1$ as:

$$R_i^{t+1} = h_i - K(p-q) + \sum_{k\in\mathcal{N}_i} M(R_{k\to i}^t) \tag{4.46}$$

Algorithm 4.2 presents the proposed belief propagation algorithm for community recovery with side information.

Table 4.2. 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 (4.45). 4. For all nodes i, compute its belief R^t_i based on (4.46). 5. Output $\tilde{C} = \{$Nodes corresponding to K largest $R^t_i\}$.

If in Algorithm 4.2 we have $t = \hat{t}(n)$, according to Lemma 29 with probability converging to one $R^t_i = \Gamma^t_i$, where Γ^t_i was the log-likelihood defined for the random tree. Hence, the performance of Algorithm 4.2 is expected to be the same as the MAP estimator defined as $\hat{\tau}_{MAP} = \mathbb{1}_{\{\Gamma^t_i \geq \nu\}}$, where $\nu = \log(\frac{n-K}{K})$. The only difference is that the MAP estimator decides based on $\Gamma^t_i \geq \nu$ while Algorithm 4.2 selects the K largest R^t_i. To manage this difference, let $\hat{C}$ define the community recovered by the MAP estimator, i.e. $\hat{C} = \{i : R^t_i \geq \nu\}$. Since $\tilde{C}$ is the set of nodes with the K largest R^t_i. Then,

$$\begin{aligned} |C^* \triangle \tilde{C}| &\leq |C^* \triangle \hat{C}| + |\hat{C} \triangle \tilde{C}| \\ &= |C^* \triangle \hat{C}| + ||\hat{C}| - K| \end{aligned} \tag{4.47}$$

Moreover,

$$||\hat{C}| - K| \leq ||\hat{C}| - |C^*|| + ||C^*| - K| \leq |C^* \triangle \hat{C}| + ||C^*| - K| \tag{4.48}$$

Using (4.48) and substituting in (4.47):

$$|C^* \triangle \tilde{C}| \leq 2|C^* \triangle \hat{C}| + ||C^*| - K| \tag{4.49}$$

We will use (4.49) to prove weak recovery.

Weak Recovery

Theorem 5 (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 4.2 with* $t = \hat{t}(n)$ *resulting in estimated community* $\tilde{C}$*. Then:*

$$\frac{\mathbb{E}[|C^* \triangle \tilde{C}|]}{K} \to 0 \tag{4.50}$$

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

Proof. See Appendix 4.4.17. □

Theorem 6 (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*,*

$$\frac{\mathbb{E}[|C^* \triangle \hat{C}|]}{K} \geq (1 - \frac{K}{n})\mathbb{E}^2[e^{\frac{U_0}{2}}]e^{\frac{-\lambda\Lambda e}{4}} - o(1) \tag{4.51}$$

Proof. See Appendix 4.4.18. □

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 4.2.2, weak recovery is possible if and only if* $\lambda > \frac{1}{\Lambda e}$ *where* $\Lambda = \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_M=1}^{L_M} (\prod_{m=1}^{M} \frac{(\alpha_{+,\ell_m}^m)^2}{\alpha_{-,\ell_m}^m})$*.*

Exact Recovery

In Section 4.2.1, 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 4.1, where in Step 3 for weak recovery we use the belief propagation algorithm presented in Table 4.2.

Theorem 7. *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 27. Assume that (4.11) holds. Let* $\tilde{C}$ *be the estimated community produced by the modified version of Algorithm 4.1 with* $t = \hat{t}(n)$*. Then* $\mathbb{P}(\tilde{C} = C^*) \to 1$ *as* $n \to \infty$*.*

Proof. See Appendix 4.5. □

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 (Hajek et al., 2018):

$$\lambda > C\frac{K}{n}\log(\frac{n}{K}) \tag{4.52}$$

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:

$$K = \frac{cn}{\log(n)},\ q = \frac{b\log^2(n)}{n},\ p = 2q \tag{4.53}$$

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 4.4 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

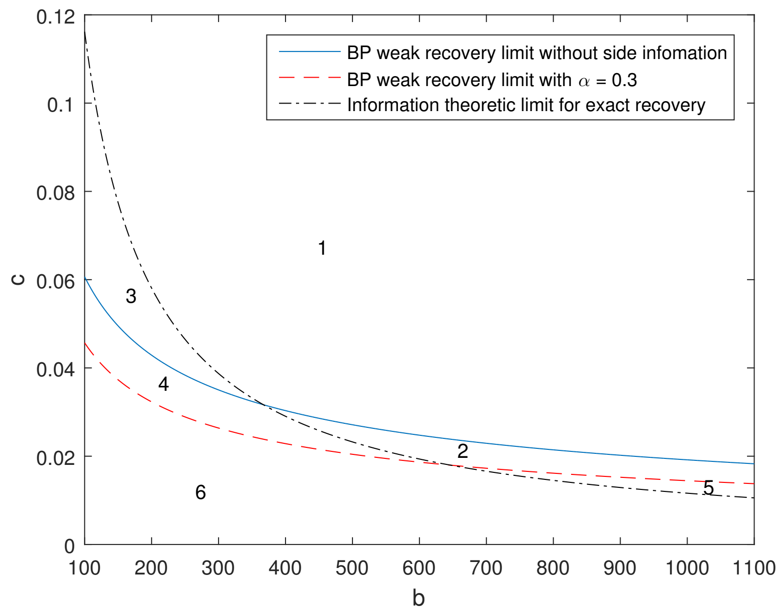

Figure 4.4. Phase diagram with $K = c\frac{n}{\log(n)}$, $q = \frac{b\log^2(n)}{n}$, $p = 2q$ and $\alpha = 0.3$ for b, c fixed as $n \to \infty$.

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.

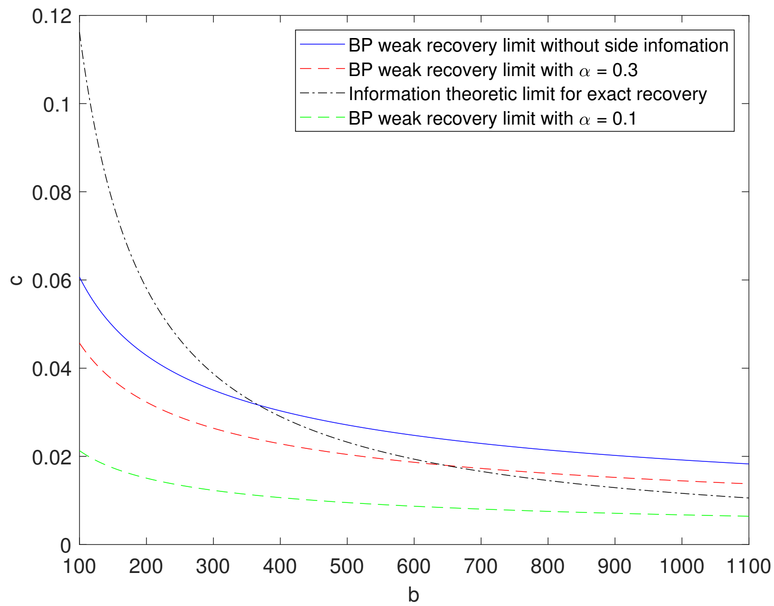

Figure 4.5. Phase diagram with $K = c\frac{n}{\log(n)}$, $q = \frac{b\log^2(n)}{n}$, $p = 2q$ and $\alpha = 0.3, 0.1$ for b, c fixed as $n \to \infty$.

Figure 4.5 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.

Application to Finite Data

This section explores the relevance of asymptotic results, obtained in this chapter, to finite data. The setup consists of a graph with $n = 10^4, K = 100, t = 10$ and side information consisting of noisy labels with error probability α. We study the performance of Algorithm 4.2 on this data set. The following performance metric is used $\zeta = \frac{1}{2K}\sum_{i=1}^{n}|x_i^* - \hat{x}_i|$. The normalization by $2K$, and the fact that the algorithm is guaranteed to return a community of known size K, defines the range of the error metric $\zeta \in [0, 1]$. Two scenarios are considered: First, $q = 5 \times 10^{-4}$ and $p = 10q$, which results in $\lambda \approx 0.041 < \frac{1}{e}$. The results are reported for different values of α in Table 4.3, which show that when $\lambda < \frac{1}{\Lambda e}$, significant residual error exists. On the other hand, when $\lambda >> \frac{1}{\Lambda e}$, error occurrences are rare. In the second scenario,

Table 4.3. Performance of belief propagation for $\lambda < \frac{1}{e}$.

α	ζ w/o side	$\lambda \times \Lambda e \approx$	ζ with side
0.1	0.95	0.903	0.75
0.01	0.95	10	0.4
0.001	0.95	100	0.05

Table 4.4. Performance of belief propagation for $\lambda > \frac{1}{e}$.

α	ζ w/o side	$\lambda \times \Lambda e \approx$	ζ with side
0.1	0.125	70	0.1
0.01	0.125	840	0.03
0.001	0.125	8551	0.02

$q = 5 \times 10^{-4}$ and $p = 80q$, resulting in $\lambda \approx 3.152 > \frac{1}{e}$. The results are reported for different values of α in Table 4.4. In this scenario, the performance of belief propagation without side information is much better compared with the first scenario because $\lambda > \frac{1}{e}$. The results also show that the performance is improved as α decreases.

4.3.2 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 (4.34) and Γ_i^t from (4.31) as well as the definitions of Z_0^t, Z_1^t, U_0 and U_1 defined directly afterward.

Lemma 30. *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 $t \geq 0$:*

$$\mathbb{E}[Z_0^{t+1}] = \frac{-\lambda}{2} b_t + o(1) \tag{4.54}$$

$$\mathbb{E}[Z_1^{t+1}] = \frac{\lambda}{2} b_t + o(1) \tag{4.55}$$

$$var(Z_0^{t+1}) = var(Z_1^{t+1}) = \lambda b_t + o(1) \tag{4.56}$$

Proof. See Appendix 4.5.1. □

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

Lemma 31. *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 $t \geq 0$:*

$$\sup_x \left|\mathbb{P}\left(\frac{Z_0^{t+1} + \frac{v_{t+1}}{2}}{\sqrt{v_{t+1}}} \leq x\right) - \phi(x)\right| \to 0 \tag{4.57}$$

$$\sup_x \left|\mathbb{P}\left(\frac{Z_1^{t+1} - \frac{v_{t+1}}{2}}{\sqrt{v_{t+1}}} \leq x\right) - \phi(x)\right| \to 0 \tag{4.58}$$

Proof. See Appendix 4.5.2. □

Lemma 32. *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,*

$$\lim_{nq,Kq\to\infty} \lim_{n\to\infty} \frac{\mathbb{E}[\hat{C}\triangle C^*]}{K} = \frac{n-K}{K}\mathbb{E}_{U_0}[Q(\frac{\nu + \frac{v_t}{2} - U_0}{\sqrt{v_t}})] + \mathbb{E}_{U_1}[Q(\frac{-\nu + \frac{v_t}{2} + U_1}{\sqrt{v_t}})] \tag{4.59}$$

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 30 and 31, 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 32 applies for side information with cardinality $L \geq 1$, and hence, generalizes (Kadavankandy et al., 2018) 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 33. *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 $\lambda > 0$:*

$$\lim_{\frac{n}{K} \to \infty} \lim_{nq, Kq \to \infty} \lim_{n \to \infty} \frac{\mathbb{E}[\hat{C} \triangle C^*]}{K} = 0 \tag{4.60}$$

Proof. See Appendix 4.5.3. □

Although Lemma 33 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 η, $\beta \in (0, 1)$ and $\gamma > 0$, define:

$$\begin{aligned} \alpha_{+,1} &= \mathbb{P}(y = 1 | x^* = 1) = \eta\beta \\ \alpha_{-,1} &= \mathbb{P}(y = 1 | x^* = 0) = \frac{\eta(1-\beta)}{(\frac{n-K}{K})^\gamma} \end{aligned} \tag{4.61}$$

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 33 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 (Kadavankandy et al., 2018) which was only for $\gamma = 1$.

Belief Propagation Algorithm for Community Recovery with Unbounded Side Information

Lemma 32 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 4.3.1, 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 29, which together with Equation (4.49) allow us to use Algorithm 4.2 (as long as $(np)^t = n^{o(1)}$). Let $\tilde{C}$ be the output of Algorithm 4.2, i.e., the set of nodes with the K largest R_i^t. Then, using Equation (4.49) 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 33 and the special case (4.61) hold. This also suggests that belief propagation (Algorithm 4.2) achieves weak recovery for any $\lambda > 0$ when Λ grows with $\frac{n}{K}$ arbitrarily slowly.

4.4 Appendix

4.4.1 Auxiliary Lemmas For Information Limits

Lemma 34. *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:

$$\hat{E}_{QU}(\theta, m_1, m_2) = E_{QU}(\theta, m_1, m_2) \tag{4.62}$$

$$\hat{E}_{PV}(\theta, m_1, m_2) = E_{PV}(\theta, m_1, m_2) \tag{4.63}$$

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

$$E_{QU}(\theta, m_1, m_2) \leq E_{QU}(\theta + \delta, m_1, m_2) \leq E_{QU}(\theta, m_1, m_2) + \delta \tag{4.64}$$

$$E_{PV}(\theta, m_1, m_2) \geq E_{PV}(\theta + \delta, m_1, m_2) \geq E_{PV}(\theta, m_1, m_2) - \delta \tag{4.65}$$

Proof. Equations (4.62) and (4.63) follow since $E_{PV}(\theta, m_1, m_2) = E_{QU}(\theta, m_1, m_2) - \theta$ and because:

$$\begin{aligned}
&E_{QU}(-m_1 D(Q||P) - m_2 D(U||V), m_1, m_2) = 0 \\
&E_{PV}(m_1 D(P||Q) + m_2 D(V||U), m_1, m_2) = 0 \\
&\psi'_{QU}(m_1, m_2, 0) = \psi'_{PV}(m_1, m_2, -1) = -m_1 D(Q||P) - m_2 D(U||V) \\
&\psi'_{QU}(m_1, m_2, 1) = \psi'_{PV}(m_1, m_2, 0) = m_1 D(P||Q) + m_2 D(V||U)
\end{aligned} \tag{4.66}$$

Equations (4.64) and (4.65) 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 35. *Assume $|L_G| \leq B$ and $|L_S| \leq B'$ for some positive constants B and B'. Define $B'' = \max\{B, B'\}$. Then, for $t \in [-1, 1]$ and $\eta \in [0, 1]$,*

$$\psi''_{QU}(m_1, m_2, t) \leq 2e^{5B''}\Big(\min\big\{m_1 D(Q||P) + m_2 D(U||V), m_1 D(P||Q) + m_2 D(V||U)\big\}\Big) \tag{4.67}$$

$$\psi_{QU}(m_1, m_2, t) \leq (m_1 D(Q||P) + m_2 D(U||V))(-t + e^{5B''} t^2) \tag{4.68}$$

$$E_{QU}\Big(m_1, m_2, -(1-\eta)(m_1 D(Q||P) + m_2 D(U||V))\Big) \geq \frac{\eta^2}{4e^{5B''}}(m_1 D(Q||P) + m_2 D(U||V)) \tag{4.69}$$

$$\psi''_{PV}(m_1, m_2, t) \leq 2e^{5B''}\Big(\min\big\{m_1 D(Q||P) + m_2 D(U||V), m_1 D(P||Q) + m_2 D(V||U)\big\}\Big) \tag{4.70}$$

$$\psi_{PV}(m_1, m_2, t) \leq (m_1 D(P||Q) + m_2 D(V||U))(t + e^{5B''} t^2) \tag{4.71}$$

$$E_{PV}(m_1, m_2, (1-\eta)(m_1 D(P||Q) + m_2 D(V||U))) \geq \frac{\eta^2}{4e^{5B''}}(m_1 D(P||Q) + m_2 D(V||U)) \tag{4.72}$$

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,

$$\psi''_{QU}(m_1,m_2,t) \le m_1\frac{\mathbb{E}_Q[L_G^2 e^{tL_G}]}{\mathbb{E}_Q[e^{tL_G}]} + m_2\frac{\mathbb{E}_U[L_S^2 e^{tL_S}]}{\mathbb{E}_U[e^{tL_S}]} \overset{(a)}{\le} m_1 e^{2B}\mathbb{E}_Q[L_G^2] + m_2 e^{2B'}\mathbb{E}_U[L_S^2] \quad (4.73)$$

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 (Apostol, 1962), 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 $x = L_S$:

$$\mathbb{E}_Q[L_G^2] \le 2e^B\mathbb{E}_Q[\phi(L_G)] = 2e^B D(Q||P) \quad (4.74)$$

$$\mathbb{E}_U[L_S^2] \le 2e^{B'}\mathbb{E}_U[\phi(L_S)] = 2e^{B'} D(U||V) \quad (4.75)$$

Combining (4.73), (4.74), (4.75) lead to $\psi''_{QU}(m_1,m_2,t) \le 2m_1e^{3B}D(Q||P) + 2m_2e^{3B'}D(U||V)$ for $t \in [-1,1]$. Similarly, it can shown for $t \in [0,2]$: $\psi''_{QU}(m_1,m_2,t) \le 2m_1e^{5B}D(Q||P) + 2m_2e^{5B'}D(U||V)$.

On the other hand, using $\phi(x) = e^{-x} - 1 + x$ with $|x| \le B$, it can be shown that $\psi''_{PV}(m_1,m_2,t) \le 2m_1e^{5B}D(P||Q) + 2m_2e^{5B'}D(V||U)$, for $t \in [0,2]$. By definition, $\psi_{QU}(m_1,m_2,t) = \psi_{PV}(m_1,m_2,t-1)$, and hence, $\psi''_{QU}(m_1,m_2,t) \le 2m_1e^{5B}D(P||Q) + 2m_2e^{5B'}D(V||U)$, for $t \in [-1,1]$, which concludes the proof of (4.67). The proof of (4.70) follows similarly.

Now since $\psi_{QU}(m_1,m_2,0) = 0$ and $\psi'_{QU}(m_1,m_2,0) = -m_1D(Q||P) - m_2D(U||V)$, then using Taylor's theorem with integral remainder, we have for $t \in [-1,1]$:

$$\begin{aligned}
&\psi_{QU}(m_1,m_2,t)\\
&= \psi_{QU}(m_1,m_2,0) + t\psi'_{QU}(m_1,m_2,0) + \int_t^0 (\lambda - t)\psi''_{QU}(m_1,m_2,t)d\lambda\\
&\overset{(a)}{\le} -t(m_1D(Q||P) + m_2D(U||V)) + e^{5B''}(m_1D(Q||P) + m_2D(U||V))t^2 \quad (4.76)
\end{aligned}$$

where (a) follows using (4.67). Similarly, it can be shown that:

$$\psi_{PV}(m_1, m_2, t) \leq t(m_1 D(P||Q) + m_2 D(V||U)) + e^{5B''}(m_1 D(P||Q) + m_2 D(V||U))t^2 \quad (4.77)$$

Combining (4.76) and (4.77) concludes the proof of (4.68), (4.71). Using (4.68) and (4.71), we get:

$$\begin{aligned}
&E_{QU}\Big(m_1, m_2, -(1-\eta)(m_1 D(Q||P) + m_2 D(U||V))\Big) \\
&\geq \sup_{t\in[0,1]} t(-(1-\eta)(m_1 D(Q||P) + m_2 D(U||V))) + t(m_1 D(Q||P) + m_2 D(U||V)) \\
&- e^{5B''}(m_1 D(Q||P) + m_2 D(U||V))t^2 \\
&= \frac{\eta^2}{4e^{5B''}}(m_1 D(Q||P) + m_2 D(U||V)) && (4.78)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&E_{PV}\Big(m_1, m_2, (1-\eta)(m_1 D(P||Q) + m_2 D(V||U))\Big) \\
&\geq \frac{\eta^2}{4e^{5B''}}(m_1 D(P||Q) + m_2 D(V||U)) && (4.79)
\end{aligned}$$

Combining (4.78) and (4.79) concludes the proof of (4.69), (4.72). □

Lemma 36. *$\eta_3(\rho, a, b, \beta) \geq \eta_2(\rho, a, b, \beta)$, 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, $\eta_3(\rho, a, b, \beta) \geq \beta$ for all $a \geq b > 0$.

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 $\beta = \rho(a - b - bT)$:

$$\eta_3(\rho, a, b, \beta) - \eta_2(\rho, a, b, \beta) = \eta_3(\rho, a, b, \beta) - \beta \geq 0 \quad (4.80)$$

Using (4.80) together with the fact that $\eta_3(\rho, a, b, \beta) - \eta_2(\rho, a, b, \beta)$ is concave in $\beta > 0$, leads to the conclusion that $\eta_3(\rho, a, b, \beta) \geq \eta_2(\rho, a, b, \beta)$ for $0 < \beta < \rho(a - b - bT)$. □

Lemma 37. *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 $a \in \mathbb{R}$:*

$$\mathbb{P}\big(S \geq a - \epsilon\big) \geq e^{-\left(t^* a - n\Gamma(t^*) + |t^*|\epsilon\right)}\left(1 - \frac{n\sigma_{\hat{X}}^2}{\epsilon^2}\right) \tag{4.81}$$

$$\mathbb{P}\big(S \leq a + \epsilon\big) \geq e^{-\left(t^* a - n\Gamma(t^*) + |t^*|\epsilon\right)}\left(1 - \frac{n\sigma_{\hat{X}}^2}{\epsilon^2}\right) \tag{4.82}$$

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 $\mu_{\hat{X}}, \sigma_{\hat{X}}^2$ are the mean and variance of $\hat{X}$, respectively.

Proof.

$$\begin{aligned}
&\mathbb{P}\big(S \geq a - \epsilon\big) \geq \mathbb{P}\big(a - \epsilon \leq S \leq a + \epsilon\big) \\
&= \int_{a-\epsilon \leq S \leq a+\epsilon} \mathbb{P}(x_1) \cdots \mathbb{P}(x_n) dx_1 \cdots dx_n \\
&\overset{(a)}{\geq} e^{-(ta - n\Gamma(t)) - |t|\epsilon} \int_{a-\epsilon \leq S \leq a+\epsilon} \prod_{i=1}^{n} \left(\frac{e^{tx_i}\mathbb{P}(x_i)}{\mathbb{E}_X[e^{tx}]} dx_i \right) \\
&\overset{(b)}{=} e^{-(ta - n\Gamma(t)) - |t|\epsilon} \mathbb{P}_{\hat{X}_n}\Big(a - \epsilon \leq S \leq a + \epsilon\Big) \\
&\overset{(c)}{\geq} e^{-(ta - n\Gamma(t)) - |t|\epsilon}\left(1 - \frac{n\sigma_{\hat{X}}^2 + (n\mu_{\hat{X}} - a)^2}{\epsilon^2}\right)
\end{aligned} \tag{4.83}$$

where, for all finite $\mathbb{E}[e^{tX}]$, (a) is true because $e^{t\sum x_i} \leq e^{n(ta + |t|\epsilon)}$ over the range of integration, (b) holds because $\frac{e^{tx}\mathbb{P}_X(x)}{\mathbb{E}_X[e^{tX}]}$ is a valid distribution (Dembo and Zeitouni, 2010), and (c) holds by Chebyshev inequality and by defining $\mu_{\hat{X}}, \sigma_{\hat{X}}^2$ 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 (4.83) leads to:

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

This concludes the proof of (4.81). The proof of (4.82) follows similarly.

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

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

which concludes the proof. $\square$

4.4.2 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., $\{j \,:\, x_j^* \neq x_i^*\}$. Then:

$$\begin{aligned}
&\frac{\mathbb{P}(\boldsymbol{G}, \boldsymbol{Y}, J, \boldsymbol{x}^*_{\backslash i,J} | x_i^* = 0)}{\mathbb{P}(\boldsymbol{G}, \boldsymbol{Y}, J, \boldsymbol{x}^*_{\backslash i,J} | x_i^* = 1)} \\
=&\frac{\mathbb{P}(\boldsymbol{G}|\boldsymbol{Y}, J, \boldsymbol{x}^*_{\backslash i,J}, x_i^* = 0)}{\mathbb{P}(\boldsymbol{G}|\boldsymbol{Y}, J, \boldsymbol{x}^*_{\backslash i,J}, x_i^* = 1)} \frac{\mathbb{P}(\boldsymbol{x}^*_{\backslash i,J}|J, x_i^* = 0, \boldsymbol{Y})}{\mathbb{P}(\boldsymbol{x}^*_{\backslash i,J}|J, x_i^* = 1, \boldsymbol{Y})} \frac{\mathbb{P}(\boldsymbol{Y}, J | x_i^* = 0)}{\mathbb{P}(\boldsymbol{Y}, J | x_i^* = 1)} \\
\overset{(a)}{=}&\frac{\mathbb{P}(\boldsymbol{G}|J, \boldsymbol{x}^*_{\backslash i,J}, x_i^* = 0)}{\mathbb{P}(\boldsymbol{G}|J, \boldsymbol{x}^*_{\backslash i,J}, x_i^* = 1)} \frac{\mathbb{P}(y_{i,1} \cdots, y_{i,M} | x_i^* = 0)}{\mathbb{P}(y_{i,1} \cdots, y_{i,M} | x_i^* = 1)} \frac{\mathbb{P}(y_{J,1}, \cdots, y_{J,M} | J, x_i^* = 0)}{\mathbb{P}(y_{J,1}, \cdots, y_{J,M} | J, x_i^* = 1)} \\
=&\Bigg(\prod_{\substack{k \neq i,J \\ x_k^* = 1}} \frac{Q(G_{ik}) P(G_{Jk})}{P(G_{ik}) Q(G_{Jk})} \Bigg) \Bigg(\prod_{m=1}^{M} \frac{U(y_{i,m}) V(y_{J,m})}{V(y_{i,m}) U(y_{J,m})} \Bigg)
\end{aligned} \tag{4.84}$$

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 $\mathbb{P}(\boldsymbol{x}^*_{\backslash i,J}|J, x_i^* = 0, \boldsymbol{Y}) = \mathbb{P}(\boldsymbol{x}^*_{\backslash i,J}|J, x_i^* = 1, \boldsymbol{Y})$.

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 (4.84), T is a sufficient statistic of $(G, \boldsymbol{Y}, J, \boldsymbol{x}^*_{\backslash i,J})$ 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:

$$\begin{aligned}
\mathbb{E}[d(\hat{\boldsymbol{x}}, \boldsymbol{x}^*)] &= \sum_{i=1}^{n} \mathbb{P}(x_i^* \neq \hat{x}_i) \\
&\geq \sum_{i=1}^{n} \min_{\tilde{x}_i(\boldsymbol{G}, \boldsymbol{Y})} \mathbb{P}(x_i^* \neq \tilde{x}_i)
\end{aligned}$$

$$\geq \sum_{i=1}^{n} \min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})} \mathbb{P}(x_i^* \neq \tilde{x}_i)$$

$$= n \min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})} \mathbb{P}(x_i^* \neq \tilde{x}_i)$$

$$= n \min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})} \left(\frac{K}{n}\mathbb{P}(x_i^* \neq \tilde{x}_i | x_i^* = 1) + \frac{n-K}{n}\mathbb{P}(x_i^* \neq \tilde{x}_i | x_i^* = 0) \right)$$

$$\geq n \min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})} \left(\frac{K}{n}\mathbb{P}(x_i^* \neq \tilde{x}_i | x_i^* = 1) + \frac{K}{n}\mathbb{P}(x_i^* \neq \tilde{x}_i | x_i^* = 0) \right)$$

$$= K \min_{\tilde{x}_i(\boldsymbol{G},\boldsymbol{Y},J,\boldsymbol{x}^*_{\setminus i,J})} \big(\mathbb{P}(x_i^* \neq \tilde{x}_i | x_i^* = 1) + \mathbb{P}(x_i^* \neq \tilde{x}_i | x_i^* = 0) \big) \tag{4.85}$$

Since by assumption, $\mathbb{E}[d(\hat{\boldsymbol{x}}, \boldsymbol{x}^*)] = o(K)$, then by (4.85), the sum of Type-I and II probabilities of error is $o(1)$, which implies that as $n \to \infty$ (Polyanskiy and Wu, 2017):

$$TV\Big(U^{\otimes M}V^{\otimes M}Q^{\otimes(K-1)}P^{\otimes(K-1)}, V^{\otimes M}U^{\otimes M}P^{\otimes(K-1)}Q^{\otimes(K-1)}\Big) \to 1 \tag{4.86}$$

where $TV(\cdot,\cdot)$ is the total variational distance between probability distributions. By properties of the total variational distance and KL divergence (Polyanskiy and Wu, 2017), 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 (4.86):

$$\begin{aligned} &D\Big(U^{\otimes M}V^{\otimes M}Q^{\otimes(K-1)}P^{\otimes(K-1)}\Big|\Big|V^{\otimes M}U^{\otimes M}P^{\otimes(K-1)}Q^{\otimes(K-1)}\Big) \\ &= M\Big(D(U||V) + D(V||U)\Big) + (K-1)\Big(D(P||Q) + D(Q||P)\Big) \to \infty \end{aligned} \tag{4.87}$$

Since the LLRs are bounded by assumption, using Lemma 35 in Appendix 4.4.1,

$$\begin{aligned} (K-1)D(P||Q) + MD(V||U) &= E_{QU}\Big((K-1)D(P||Q) + MD(V||U), K-1, M\Big) \\ &\geq E_{QU}\Big(-\frac{(K-1)D(Q||P) + MD(U||V)}{2}, K-1, M\Big) \\ &\geq C\Big((K-1)D(Q||P) + MD(U||V)\Big) \end{aligned} \tag{4.88}$$

for some positive constant C. Substituting in (4.87) leads to:

$$MD(V||U) + (K-1)D(P||Q) \to \infty \tag{4.89}$$

which proves the first condition in (4.6).

$\boldsymbol{x}^*$ 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.

$$I(\boldsymbol{G}, \boldsymbol{Y}; \boldsymbol{x}^*) \overset{(a)}{\geq} I(\hat{\boldsymbol{x}}(\boldsymbol{G}, \boldsymbol{Y}); \boldsymbol{x}^*) \overset{(b)}{\geq} \min_{\mathbb{E}[d(\tilde{\boldsymbol{x}}, \boldsymbol{x}^*)] \leq \epsilon_n K} I(\tilde{\boldsymbol{x}}(\boldsymbol{G}, \boldsymbol{Y}); \boldsymbol{x}^*) \tag{4.90}$$

$$\geq H(\boldsymbol{x}^*) - \max_{\mathbb{E}[d(\tilde{\boldsymbol{x}}, \boldsymbol{x}^*)] \leq \epsilon_n K} H(d(\tilde{\boldsymbol{x}}, \boldsymbol{x}^*))$$

$$\overset{(c)}{=} \log\left(\binom{n}{K}\right) - nh(\frac{\epsilon_n K}{n}) \overset{(d)}{\geq} K \log(\frac{n}{k})(1 + o(1)) \tag{4.91}$$

where (a) is due to the data processing inequality (Polyanskiy and Wu, 2017), 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 $p \leq \frac{1}{2}$ (Hajek et al., 2017), 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 (Polyanskiy and Wu, 2017):

$$I(\boldsymbol{G}, \boldsymbol{Y}; \boldsymbol{x}^*) = \min_{\tilde{Q}} D\Big(\mathrm{P}(\boldsymbol{G}, \boldsymbol{Y}|\boldsymbol{x}^*) \,||\, \tilde{\mathrm{Q}} \mid \mathrm{P}(\boldsymbol{x}^*)\Big)$$

$$\leq D\Big(\mathrm{P}(\boldsymbol{G}|\boldsymbol{x}^*) \prod_{m=1}^{M} (\mathrm{P}(\boldsymbol{y}_m|\boldsymbol{x}^*)) \Big|\Big| Q^{\otimes\binom{n}{2}} \prod_{m=1}^{M} (U^{\otimes n}) \,\big|\mathrm{P}(\boldsymbol{x}^*)\Big)$$

$$= \binom{K}{2} D(P||Q) + KMD(V||U) \tag{4.92}$$

Combining (4.91) and (4.92):

$$\liminf_{n\to\infty} \frac{(K-1)D(P||Q) + 2MD(V||U)}{\log(\frac{n}{K})} \geq 2 \tag{4.93}$$

which proves the second condition in (4.6).

4.4.3 Sufficiency of Theorem 1

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

$$e_1(S,T) \triangleq \sum_{i\in S}\sum_{j\in T} L_G(i,j) \tag{4.94}$$

$$e_2(S) \triangleq \sum_{i\in S}\sum_{m=1}^{M} L_S(i,m) \tag{4.95}$$

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

$$\hat{C} = \hat{C}_{ML} = \underset{\substack{C\subset\{1,\cdots,n\}\\ |C|=K}}{\arg\max}\big(e_1(C,C)+e_2(C)\big) \tag{4.96}$$

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^*.

$$\begin{aligned}
&e_1(\hat{C},\hat{C}) + e_2(\hat{C}) - \big(e_1(C^*,C^*)+e_2(C^*)\big)\\
&= e_1(\hat{C}\backslash C^*, \hat{C}\backslash C^*) + e_1(\hat{C}\backslash C^*, \hat{C}\cap C^*) - e_1(C^*\backslash\hat{C}, C^*) + e_2(\hat{C}\backslash C^*) - e_2(C^*\backslash\hat{C})
\end{aligned} \tag{4.97}$$

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

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

$$\cup \left\{S \subset C^*, T \subset (C^*)^c : |S| = |T| = K-r, e_1(T,T) + e_1(T, C^*\backslash S) + e_2(T) \geq \theta\right\}\Big) \tag{4.98}$$

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 (4.98) and a union bound:

$$\begin{aligned}
\mathbb{P}(R=r) \leq& \binom{K}{K-r} \mathbb{P}\Big(\sum_{i=1}^{a} L_{G,i} + \sum_{j=1}^{K-r}\sum_{m=1}^{M} L_{S,m,j} \leq \theta\Big) \\
&+ \binom{K}{K-r}\binom{n-K}{K-r} \mathbb{P}\Big(\sum_{i=1}^{a} L_{G,i} + \sum_{j=1}^{K-r}\sum_{m=1}^{M} L_{S,m,j} \geq \theta\Big) \\
\stackrel{(a)}{\leq}& e^{(K-r)\log(\frac{Ke}{K-r})} e^{-\sup_{t\geq 0} -t\theta - a\log_P(\mathbb{E}[e^{-tL_G}]) - (K-r)M\log_V(\mathbb{E}[e^{-tL_S}])} \\
&+ e^{(K-r)\log(\frac{(n-K)Ke^2}{(K-r)^2})} e^{-\sup_{t\geq 0} t\theta - a\log_Q(\mathbb{E}[e^{tL_G}]) - (K-r)M\log_U(\mathbb{E}[e^{tL_S}])} \\
\stackrel{(b)}{\leq}& e^{(K-r)\log(\frac{Ke}{K-r}) - E_{PV}(\theta,a,M(K-r))} + e^{(K-r)\log(\frac{(n-K)Ke^2}{(K-r)^2}) - E_{QU}(\theta,a,M(K-r))} \\
\stackrel{(c)}{=}& e^{(K-r)\log(\frac{Ke}{K-r}) - E_{PV}(\theta,a,M(K-r))} + e^{(K-r)\log(\frac{(n-K)Ke^2}{(K-r)^2}) - E_{PV}(\theta,a,M(K-r)) - \theta} \\
\stackrel{(d)}{\leq}& e^{(K-r)\log(\frac{Ke}{K-r}) - E_{PV}(\theta,a,M(K-r))} \\
&+ e^{-(K-r)\left((1-\eta)((\frac{K-1}{2})D(P||Q) + MD(V||U)) - \log(\frac{n-K}{K})\right)} e^{2(K-r)\log(\frac{e}{\epsilon}) - E_{PV}(\theta,a,M(K-r))} \\
\stackrel{(e)}{\leq}& 2e^{2(K-r)\log(\frac{e}{\epsilon}) - E_{PV}(\theta,a,M(K-r))}
\end{aligned} \tag{4.99}$$

where (a) holds by Chernoff bound and because $\binom{a}{b} \leq (\frac{ea}{b})^b$, (b) holds from Lemma 34 in Appendix 4.4.1, (c) holds because $E_{PV}(\theta, a, M(K-r)) = E_{QU}(\theta, a, M(K-r)) - \theta$, (d) holds because $a \geq \frac{(K-r)(K-1)}{2}$, $r \leq (1-\epsilon)K$ and (e) holds by assuming that $\liminf_{n\to\infty} \frac{(K-1)D(P||Q) + 2MD(V||U)}{\log(\frac{n}{K})} > 2$, which implies that

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

Lemma 34 in Appendix 4.4.1 shows that

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

Using $a \geq \frac{(K-r)(K-1)}{2}$ and substituting in (4.99),

$$\begin{aligned}\mathbb{P}(R=r) \leq& 2e^{-(K-r)\left(C(\frac{K-1}{2}D(P||Q)+MD(V||U))-2\log(\frac{e}{\epsilon})\right)} \\ \leq& 2e^{-(K-r)\left(\frac{C}{2}((K-1)D(P||Q)+MD(V||U))-2\log(\frac{e}{\epsilon})\right)} \end{aligned} \tag{4.100}$$

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,

$$\begin{aligned}\mathbb{P}(R \leq (1-\epsilon)K) =& \sum_{r=0}^{(1-\epsilon)K} \mathbb{P}(R=r) \leq \sum_{r=0}^{(1-\epsilon)K} 2e^{-(K-r)E} \\ \overset{(a)}{\leq}& 2\sum_{r'=\epsilon K}^{\infty} e^{-r'E} \leq 2\frac{e^{-\epsilon KE}}{1-e^{-E}} \overset{(b)}{\leq} o(1)\end{aligned} \tag{4.101}$$

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 3.

4.4.4 Proof of Lemma 20

Recall the definition of $\hat{C}$ from (4.96). 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, $|K' - K| \leq \frac{K}{\log(K)}$. 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 $n \to \infty$:

$$\frac{K}{K'} \to 1 \ , \ \frac{K-r}{K'-r} \to 1 \ , \ \frac{a}{a'} \to 1 \tag{4.102}$$

Following similar ideas as the proof of Theorem 3:

$$\begin{aligned}&\mathbb{P}(R=r) \\ &\leq \mathbb{P}\Big(\big\{C \in \{1, \cdots, n\} : |C| = K, |C \cap C^*| = r,\end{aligned}$$

$$e_1(\hat{C},\hat{C}) + e_2(\hat{C}) - e_1(C^*,C^*) - e_2(C^*) \geq 0\Big\}\Big)$$

$$= \mathbb{P}\Big(\Big\{S \subset C^*, T \subset (C^*)^c : |S| = K' - r, = |T| = K - r,$$

$$e_1(S,C^*) + e_2(S) \leq e_1(T,T) + e_1(T,C^*\backslash S) + e_2(T)\Big\}\Big)$$

$$\leq \mathbb{P}\Big(\Big\{S \subset C^* : |S| = K' - r, e_1(S,C^*) + e_2(S) \leq \theta\Big\}$$

$$\cup \Big\{\exists S \subset C^*, T \subset (C^*)^c : |S| = K' - r, |T| = K - r, e_1(T,T) + e_1(T,C^*\backslash S) + e_2(T) \geq \theta\Big\}\Big) \tag{4.103}$$

where $\theta = (1-\eta)(aD(P||Q) + (K-r)MD(V||U))$, for some $\eta \in (0,1)$. Using (4.103) and a union bound,

$$\begin{aligned}
&\mathbb{P}(R = r) \\
&\overset{(a)}{\leq} \binom{K'}{K'-r}\mathbb{P}\Big(\sum_{i=1}^{a'} L_{G,i} + \sum_{j=1}^{K'-r}\sum_{m=1}^{M} L_{S,m,j} \leq \theta\Big) \\
&+ \binom{K'}{K'-r}\binom{n-K'}{K-r}\mathbb{P}\Big(\sum_{i=1}^{a} L_{G,i} + \sum_{j=1}^{K-r}\sum_{m=1}^{M} L_{S,m,j} \geq \theta\Big) \\
&\overset{(b)}{\leq} e^{(K'-r)\log(\frac{K'e}{K'-r})} e^{-\sup_{t\geq 0} -t\theta - a'\log_P(\mathbb{E}[e^{-tL_G}]) - M(K'-r)\log_V(\mathbb{E}[e^{-tL_S}])} \\
&+ e^{(K'-r)\log(\frac{K'}{(K'-r)}) + (K-r)\log(\frac{(n-K)e}{(K-r)})} e^{-\sup_{t\geq 0} t\theta - a\log_Q(\mathbb{E}[e^{tL_G}]) - M(K-r)\log_U(\mathbb{E}[e^{tL_S}])} \\
&\overset{(c)}{\leq} e^{(K'-r)\log(\frac{K'e}{K'-r}) - (1-o(1))E_{PV}(\theta,a,M(K-r))} + e^{(K'-r)\log(\frac{K'}{(K'-r)}) + (K-r)\log(\frac{(n-K)e}{(K-r)}) - E_{QU}(\theta,a,M(K-r))} \\
&\overset{(d)}{=} e^{(K-r)\log(\frac{Ke}{K-r})(1+o(1)) - E_{PV}(\theta,a,M(K-r))(1+o(1))} + e^{(K-r)\log(\frac{(n-K)Ke^2}{(K-r)^2})(1+o(1)) - E_{PV}(\theta,a,M(K-r)) - \theta} \\
&\overset{(e)}{\leq} e^{(K-r)(1+o(1))\log(\frac{Ke}{K-r}) - (1+o(1))E_{PV}(\theta,a,M(K-r))} \\
&+ e^{-(K-r)(1+o(1))\big((1-\eta)((\frac{K-1}{2})D(P||Q) + MD(V||U)) - \log(\frac{n-K}{K})\big)} e^{2(1+o(1))(K-r)\log(\frac{e}{\epsilon}) - E_{PV}(\theta,a,M(K-r))} \\
&\overset{(f)}{\leq} 2e^{2(K-r)(1+o(1))\log(\frac{e}{\epsilon}) - (1+o(1))E_{PV}(\theta,a,M(K-r))}
\end{aligned} \tag{4.104}$$

where (a) holds for $L_{G,i}(L_{S,m,j})$ be i.i.d copies of $L_G(L_S)$, respectively, (b) holds by Chernoff bound and because $\binom{a}{b} \leq (\frac{ea}{b})^b$, (c) holds by using (4.102) and by Lemma 34 in Appendix 4.4.1, (d) holds by using (4.102) and because $E_{PV}(\theta, a, M(K-r)) = E_{QU}(\theta, a, M(K-$

$r)) - \theta$, (e) holds because $a \geq \frac{(K-r)(K-1)}{2}$, $r \leq (1-\epsilon)K$ and (f) holds by assuming that $\liminf_{n\to\infty} \frac{(K-1)D(P||Q)+2MD(V||U)}{\log(\frac{n}{K})} \geq 2$, which implies that $(1-\eta)((\frac{K-1}{2})D(P||Q) + MD(V||U)) - \log(\frac{n-K}{K}) \geq 0$.

The remainder of the proof follows similarly to Appendix 4.4.3 following (4.99).

4.4.5 Proof of Lemma 21

Lemma 38. *Suppose that* (4.11) *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* $\gamma = \frac{\log(\frac{n}{K})}{K}$*:*

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-\delta)} \tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big) = o(\frac{1}{n}) \tag{4.105}$$

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \leq K(1-\delta)\gamma\Big) = o(\frac{1}{K}) \tag{4.106}$$

Proof. By Chernoff bound:

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-\delta)} \tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big) \leq e^{-(1-\delta)\sup_{t\geq 0} tK\gamma - K\log(\mathbb{E}_Q[e^{tL_G}]) - \frac{M}{1-\delta}\log(\mathbb{E}_U[e^{tL_S}])} \tag{4.107}$$

From (4.5) it follows that for some positive ϵ_o:

$$\begin{aligned} K\gamma &\leq \frac{KD(P||Q)}{2+\epsilon_o} + \frac{MD(V||U)}{1+\frac{\epsilon_o}{2}} \\ &\leq KD(P||Q) + MD(V||U) \\ &\leq KD(P||Q) + \frac{M}{1-\delta}D(V||U) \end{aligned} \tag{4.108}$$

Hence, using Lemma 34 in Appendix 4.4.1, $\sup_{t\geq 0}$ 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 (Polyanskiy and Wu, 2017) 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 (4.107):

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-\delta)} \tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big) \leq e^{-(1-\delta)E_{QU}(K\gamma,K,M)}$$

$$\leq e^{-(1-\delta)(1+\epsilon)\log(n)} \tag{4.109}$$

where (4.109) follows since (4.11) holds by assumption, i.e., there exists $\epsilon \in (0,1)$: $E_{QU}(K\gamma, K, M) \geq (1+\epsilon)\log(n)$. Equation (4.109) implies that (4.105) holds for sufficiently small δ.

To show (4.106), Chernoff bound is used:

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \leq K(1-\delta)\gamma\Big) \\
&\stackrel{(a)}{\leq} e^{tK\gamma(1-\delta)+K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])+K\delta\log(\mathbb{E}_Q[e^{-tL_G}])} e^{M(1-\delta)\log(\mathbb{E}_V[e^{-tL_S}])+M\delta\log(\mathbb{E}_U[e^{-tL_S}])} \\
&= e^{(1-2\delta)(tK\gamma+K\log(\mathbb{E}_P[e^{-tL_G}])+M\frac{1-\delta}{1-2\delta}\log(\mathbb{E}_V[e^{-tL_S}]))} e^{\delta(tK\gamma+K\log(\mathbb{E}_Q[e^{-tL_G}])+M\log(\mathbb{E}_U[e^{-tL_S}]))} \\
&\stackrel{(b)}{\leq} e^{(1-2\delta)(tK\gamma+K\log(\mathbb{E}_P[e^{-tL_G}])+M\log(\mathbb{E}_V[e^{-tL_S}]))} e^{\delta(tK\gamma+K\log(\mathbb{E}_Q[e^{-tL_G}])+M\log(\mathbb{E}_U[e^{-tL_S}]))}
\end{aligned} \tag{4.110}$$

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 $t \geq 0$ (Polyanskiy and Wu, 2017).

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 (4.110),

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \leq K(1-\delta)\gamma\Big) \\
&\leq e^{-(1-2\delta)E_{PV}(K\gamma,K,M)} e^{\delta(K\gamma+K\log(\mathbb{E}_Q[e^{-tL_G}])+M\log(\mathbb{E}_U[e^{-tL_S}]))}
\end{aligned} \tag{4.111}$$

By Lemma 35 and convexity of $\psi_{QU}(t, m_1, m_2)$:

$$\psi_{QU}(-t, K, M) \leq \psi_{QU}(-1, K, M) \leq A(KD(Q||P) + MD(U||V)) \tag{4.112}$$

for some positive constant A. Moreover, by Lemma 35, $E_{QU}(K\gamma, K, M) \geq E_{QU}(0, K, M) \geq A_1(KD(Q||P)+MD(U||V))$, for some positive constant A_1. Hence, by substituting in (4.111), for some positive constant A_2:

$$\mathbb{P}\big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \le K(1-\delta)\gamma\big) \le e^{-(1-2\delta)E_{PV}(K\gamma,K,M)+\delta(K\gamma+A_2E_{QU}(K\gamma,K,M))}$$
$$\overset{(a)}{\le} e^{-E_{QU}(K\gamma,K,M)(1-2\delta-\delta A_2)+(1-\delta)K\gamma}$$
$$\overset{(b)}{=} e^{-\log(n)((1+\epsilon)(1-2\delta-\delta A_2)+\delta-1)-\log(K)(1-\delta)}$$
$$\overset{(c)}{=} o(\frac{1}{K}) \tag{4.113}$$

where (a) holds because $E_{PV}(K\gamma, K, M) = E_{QU}(K\gamma, K, M) - K\gamma$ from Lemma 35, (b) holds by the assumption that (4.11) holds, which implies that there exists $\epsilon \in (0,1)$: $E_{QU}(K\gamma, K, M) \ge (1+\epsilon)\log(n)$ and (c) holds for sufficiently small δ.

Equations (4.109) and (4.113) concludes the proof of Lemma 38. □

Define the event $E \triangleq \{(\hat{C}_k, C_k^*) : |\hat{C}_k \triangle C_k^*| \le \delta K \quad \forall k\}$; then conditioned on E we have:

$$|\hat{C}_k \cap C_k^*| \ge |\hat{C}_k| - |\hat{C}_k \triangle C_k^*| = \lceil K(1-\delta)\rceil - |\hat{C}_k \triangle C_k^*| \ge K(1-2\delta)$$

Thus, in Algorithm 4.1, 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 38 and (4.110). 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 38, with probability converging to 1,

$$r_i > K(1-\delta)\gamma, \qquad i \in C^*$$
$$r_i < K(1-\delta)\gamma, \qquad i \notin C^*$$

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

4.4.6 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:

$$\mathbb{P}\Big(\Big|Z - (1-\delta)K\Big| \geq \frac{K}{\log(K)}\Big) \leq e^{-\Omega(\frac{K}{\log^2(K)})} \tag{4.114}$$

A result of Hoeffding (Hoeffding, 1963, 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 (4.114) on the negative and positive side, individually. Putting them back together, we get a bound on the tails of $|C_k^*|$:

$$\begin{aligned}\mathbb{P}\Big(\Big||C_k^*| - (1-\delta)K\Big| \geq \frac{K}{\log(K)}\Big) &\leq e^{-\Omega(\frac{K}{\log^2(K)})} \\ &\leq o(1) \end{aligned}\tag{4.115}$$

Since (4.5) holds, for sufficiently small δ,

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

which together with (4.115) indicates, via Lemma 20, that ML achieves weak recovery. Thus, for any $1 \leq k \leq \frac{1}{\delta}$:

$$\mathbb{P}\Big(\frac{|\hat{C}_k \triangle C_k^*|}{K} \leq 2\epsilon + \frac{1}{\log(K)}\Big) \geq 1 - o(1) \tag{4.116}$$

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

$$\mathbb{P}\Big(\frac{|\hat{C}_k \triangle C_k^*|}{K} \leq 2\epsilon + \frac{1}{\log(K)}, \quad \forall k\Big) \geq 1 - o(1) \tag{4.117}$$

Since $\epsilon = o(1)$, the desired (4.9) holds.

4.4.7 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 39. *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:*

$$\mathbb{P}\Big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\theta_n - \tilde{\theta}_n\Big) \leq \frac{2}{K_o} \tag{4.118}$$

$$\mathbb{P}\Big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\theta_n\Big) \leq \frac{1}{n-K} \tag{4.119}$$

where

$$\tilde{\theta}_n \triangleq (K_o-1)D(P||Q) + 6\sqrt{K_o}\sigma \tag{4.120}$$

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

$$\begin{aligned} i_o &\triangleq \arg\min_{i \in C^*} e_1(i, C^*) + \sum_{m=1}^{M} L_S(i,m) \\ \tilde{C} &\triangleq C^* \backslash \{i_o\} \cup \{j\} \text{ for } j \notin C^* \end{aligned} \tag{4.121}$$

Also, define the following event:

$$F_M \triangleq \Big\{(\boldsymbol{G}, \boldsymbol{Y}) : \min_{i \in C^*} e_1(i, C^*) + \sum_{m=1}^{M} L_S(i,m) \leq \max_{j \notin C^*} e(j, C^* \backslash \{i_o\}) + \sum_{m=1}^{M} L_S(j,m)\Big\} \tag{4.122}$$

Since $\mathbb{P}(\text{ML fails}) = o(1)$, using (4.96):

$$\begin{aligned} &e_1(\tilde{C}, \tilde{C}) + e_2(\tilde{C}) - e_1(C^*, C^*) - e_2(C^*) \\ &= \Big(e(j, C^* \backslash \{i_o\}) + \sum_{m=1}^{M} L_S(j,m)\Big) - \Big(e_1(i, C^*) + \sum_{m=1}^{M} L_S(i,m)\Big) \end{aligned} \tag{4.123}$$

For observations belonging to F_M, the expression (4.123) is non-negative, implying ML fails with non-zero probability. Then,

$$\mathbb{P}(F_M) \leq \mathbb{P}(\text{ML fails}) = o(1) \tag{4.124}$$

since ML achieves exact recovery.

Define θ'_n, θ''_n and the events E_1 and E_2 as follows:

$$\theta'_n \triangleq \inf \left\{ x \in \mathbb{R} : \mathbb{P}\Big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)x - \tilde{\theta}_n \Big) \geq \frac{2}{K_o} \right\} \tag{4.125}$$

$$\theta''_n \triangleq \sup \left\{ x \in \mathbb{R} : \mathbb{P}\Big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)x \Big) \geq \frac{1}{n-K} \right\} \tag{4.126}$$

$$E_1 \triangleq \Big\{ (\boldsymbol{G}, \boldsymbol{Y}) : \max_{j \notin C^*} \Big(e(j, C^* \backslash \{i_o\}) + \sum_{m=1}^{M} L_S(j,m) \Big) \geq (K-1)\theta''_n \Big\} \tag{4.127}$$

$$E_2 \triangleq \Big\{ (\boldsymbol{G}, \boldsymbol{Y}) : \min_{i \in C^*} \Big(e_1(i, C^*) + \sum_{m=1}^{M} L_S(i,m) \Big) \leq (K-1)\theta'_n \Big\} \tag{4.128}$$

where $\tilde{\theta}_n$ is defined in (4.120).

Lemma 40. *$\mathbb{P}(E_1) = \Omega(1)$ and $\mathbb{P}(E_2) = \Omega(1)$.*

Proof.

$$\begin{aligned}
\mathbb{P}(E_1) &\overset{(a)}{=} 1 - \prod_{j \notin C^*} \mathbb{P}\Big(e(j, C^* \backslash \{i_o\}) + \sum_{m=1}^{M} L_S(j,m) < (K-1)\theta''_n \Big) \\
&= 1 - \Big(1 - \mathbb{P}\Big(e(j, C^* \backslash \{i_o\}) + \sum_{m=1}^{M} L_S(j,m) \geq (K-1)\theta''_n \Big) \Big)^{n-K} \\
S &\overset{(b)}{\geq} 1 - e^{\left(-(n-K)\mathbb{P}\left(e(j, C^* \backslash \{i_o\}) + \sum_{m=1}^{M} L_S(j,m) \geq (K-1)\theta''_n \right) \right)} \\
&\overset{(c)}{\geq} 1 - e^{-1}
\end{aligned} \tag{4.129}$$

where (a) holds because $e(j, C^* \backslash \{i_o\}) + \sum_{m=1}^{M} L_S(j,m)$ are i.i.d. for all $j \notin C^*$, (b) holds because $1 - x \leq e^{-x}$ $\forall x \in \mathbb{R}$ and (c) holds by definition of θ''_n. Thus, $\mathbb{P}(E_1) = \Omega(1)$.

To show $\mathbb{P}(E_2) = \Omega(1)$, we are confronted with the difficulty that $e_1(i, C^*)$ are not independent. Let T be the set of the first K_o indices in C^*, where $K_o \to \infty$ such that $K_o = o(K)$. Also, let $T' = \{i \in T : e_1(i, T) \le \tilde{\theta}_n\}$. Then,

$$\begin{aligned}
&\min_{i \in C^*} e_1(i, C^*) + \sum_{m=1}^{M} L_S(i, m) \le \min_{i \in T'} e_1(i, C^*) + \sum_{m=1}^{M} L_S(i, m) \\
&\le \min_{i \in T'} e_1(i, C^* \backslash T) + \sum_{m=1}^{M} L_S(i, m) + \tilde{\theta}_n
\end{aligned} \tag{4.130}$$

It follows that:

$$\begin{aligned}
&\mathbb{P}(E_2) \\
&\ge \mathbb{P}\Big(\min_{i \in T'} e_1(i, C^* \backslash T) + \sum_{m=1}^{M} L_S(i, m) \le (K-1)\theta'_n - \tilde{\theta}_n \Big) \\
&\overset{(a)}{=} 1 - \mathbb{P}\Big(\bigcap_{i \in T'} \Big\{ e_1(i, C^* \backslash T) + \sum_{m=1}^{M} L_S(i, m) > (K-1)\theta'_n - \tilde{\theta}_n \Big\} \Big) \\
&= 1 - \mathbb{P}\Big(\bigcap_{i \in T'} \Big\{ e_1(i, C^* \backslash T) + \sum_{m=1}^{M} L_S(i, m) > (K-1)\theta'_n - \tilde{\theta}_n \Big\} \Big| \, |T'| \ge \frac{K_o}{2} \Big) \mathbb{P}\big(|T'| \ge \frac{K_o}{2}\big) \\
&- \mathbb{P}\Big(\bigcap_{i \in T'} \Big\{ e_1(i, C^* \backslash T) + \sum_{m=1}^{M} L_S(i, m) > (K-1)\theta'_n - \tilde{\theta}_n \Big\} \Big| \, |T'| < \frac{K_o}{2} \Big) \times \mathbb{P}\big(|T'| < \frac{K_o}{2}\big) \\
&\ge 1 - \mathbb{P}\Big(\bigcap_{i \in T'} \Big\{ e_1(i, C^* \backslash T) + \sum_{m=1}^{M} L_S(i, m) > (K-1)\theta'_n - \tilde{\theta}_n \Big\} \Big| \, |T'| \ge \frac{K_o}{2} \Big) - \mathbb{P}\big(|T'| < \frac{K_o}{2}\big) \\
&\ge 1 - \Big(1 - \mathbb{P}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \le (K-1)\theta'_n - \tilde{\theta}_n \big) \Big)^{\frac{K_o}{2}} - \mathbb{P}\big(|T'| < \frac{K_o}{2}\big) \\
&\overset{(b)}{\ge} 1 - e^{\Big(-(\frac{K_o}{2}) \mathbb{P}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \le (K-1)\theta'_n - \tilde{\theta}_n \big) \Big)} - \mathbb{P}\big(|T'| < \frac{K_o}{2}\big) \\
&\overset{(c)}{\ge} 1 - e^{-1} - \mathbb{P}\big(|T'| < \frac{K_o}{2}\big)
\end{aligned}$$

where (a) holds because $e_1(i, C^* \backslash T) + \sum_{m=1}^{M} L_S(i, m)$ are i.i.d. for all $i \in T'$, (b) holds because $1 - x \le e^{-x} \; \forall x \in \mathbb{R}$, (c) holds by definition of θ'_n.

To conclude the proof, it remains to show that $\mathbb{P}(|T'| < \frac{K_o}{2}) = o(1)$. Recall $T' = \{i \in T : e_1(i, T) \le \tilde{\theta}_n\}$. For $i \in T$, $e_1(i, T) = G_i + H_i$, where $G_i = e_1(i, \{1 \cdots, i-1\})$ and

$H_i = e_1(i, \{i+1 \cdots, K_o\})$. Thus, by Chebyshev inequality:

$$\mathbb{P}\Big(G_i \geq (i-1)D(P||Q) + 3\sqrt{K_o}\sigma\Big) \leq \frac{1}{9}$$

for all $i \in T$. Therefore, $|\{i : G_i \leq (i-1)D(P||Q) + 3\sqrt{K_o}\sigma\}|$ is stochastically at least as large as a $\text{Bin}(K_o, \frac{8}{9})$ random variable. Thus,

$$\mathbb{P}\Big(\big|\{i : G_i \leq (i-1)D(P||Q) + 3\sqrt{K_o}\sigma\}\big| \geq \frac{3K_o}{4}\Big) \to 1 \tag{4.131}$$

as $K_o \to \infty$. Similarly,

$$\mathbb{P}\Big(\big|\{i : H_i \leq (K_o - i)D(P||Q) + 3\sqrt{K_o}\sigma\}\big| \geq \frac{3K_o}{4}\Big) \to 1 \tag{4.132}$$

as $K_o \to \infty$. Combining (4.131) and (4.132) and using the definition of $e_1(i, T)$:

$$\mathbb{P}(|T'| \geq \frac{K_o}{2}) \xrightarrow{K_o \to \infty} 1$$

which concludes the proof of the lemma. □

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

$$\begin{aligned} \mathbb{P}(E_1 \cap E_2 \cap F_M^c) &\geq \mathbb{P}(E_1 \cap E_2) - \mathbb{P}(F_M) = \mathbb{P}(E_1)\mathbb{P}(E_2) - o(1) \\ &= \Omega(1) \end{aligned} \tag{4.133}$$

where (4.133) holds since $\mathbb{P}(E_1) = \Omega(1)$ 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 (4.118) and (4.119) hold simultaneously.

□

The necessity of Theorem 2 expresses the following: subject to conditions (4.5), exact recovery implies (4.12). Lemma 39 shows that exact recovery implies (4.118) and (4.119). It remains to be shown that (4.118) and (4.119) imply (4.12). We show that by contraposition.

Assume (4.12) does not hold, then for arbitrarily small $\epsilon > 0$ and sufficiently large n

$$E_{QU}\big(\log(\frac{n}{K}), K, M\big) \leq (1-\epsilon)\log(n) \tag{4.134}$$

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 (4.5) holds, for sufficiently large n and arbitrary small $\epsilon_o > 0$:

$$\begin{aligned} K\gamma &\leq \frac{KD(P||Q)}{2+\epsilon_o} + \frac{MD(V||U)}{(1+\frac{\epsilon_o}{2})} \\ &\leq \frac{1}{1+\frac{\epsilon_o}{2}}(KD(P||Q) + MD(V||U)) \\ &\leq KD(P||Q) + MD(V||U) \end{aligned} \tag{4.135}$$

At $\theta_n = \gamma$:

$$\begin{aligned} &\mathbb{P}\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma\big) = \int\limits_{S\geq(K-1)\gamma} \mathbb{P}\big(\tilde{w}_1, \cdots, \tilde{w}_{K-1}, \tilde{z}\big) \\ &\overset{(a)}{\geq} \int\limits_{a-\delta\leq S\leq a+\delta} \big(\prod_{\ell=1}^{K-1} \mathbb{P}(\tilde{w}_\ell)\big)\big(\mathbb{P}(\tilde{z})\big) \overset{(b)}{=} \int\limits_{a-\delta\leq S\leq a+\delta} \left(\frac{\mathbb{E}[e^{tS}]e^{tS}}{\mathbb{E}[e^{tS}]e^{tS}}\right)\big(\prod_{\ell=1}^{K-1} \mathbb{P}(\tilde{w}_\ell)\big)\big(\mathbb{P}(\tilde{z})\big) \\ &\overset{(c)}{\geq} e^{-ta-|t|\delta+\psi_{QU}(K-1,M,t)} \int\limits_{a-\delta\leq S\leq a+\delta} \left(\prod_{\ell=1}^{K-1} \frac{\mathbb{P}(\tilde{w}_\ell)e^{t\tilde{w}_\ell}}{\mathbb{E}[e^{t\tilde{w}_\ell}]}\right)\left(\frac{\mathbb{P}(\tilde{z})e^{t\tilde{z}}}{\mathbb{E}[e^{t\tilde{z}}]}\right) \\ &\overset{(d)}{=} e^{-ta-|t|\delta+\psi_{QU}(K-1,M,t)} \mathbb{P}_{\tilde{Q}\tilde{U}}\big(a-\delta \leq S \leq a+\delta\big) \\ &\overset{(e)}{\geq} e^{-\left(ta-\psi_{QU}(K-1,M,t)\right)-|t|\delta} \left(1 - \frac{\big((K-1)\tilde{\sigma}_{L_G}^2 + M\tilde{\sigma}_{L_S}^2\big) + \big((K-1)\tilde{\mu}_{L_G} + M\tilde{\mu}_{L_S} - a\big)^2}{\delta^2}\right) \end{aligned} \tag{4.136}$$

where (a) holds because $\tilde{W}_\ell$ are i.i.d. and independent of $\tilde{Z}$, (b) holds for any $t \in \mathbb{R}$ such that $\mathbb{E}[e^{tS}]$ is finite, (c) holds by the definition of ψ_{QU} and because $a - \delta \leq S \leq a + \delta$, (d) holds because $\frac{\mathbb{P}(\tilde{W}_\ell)e^{t\tilde{W}_\ell}}{\mathbb{E}[e^{t\tilde{W}_\ell}]}$ 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 (e) holds from Chebyshev's inequality and by defining $\tilde{\sigma}^2_{L_G}$, $\tilde{\mu}_{L_G}$, $\tilde{\sigma}^2_{L_S}$ and $\tilde{\mu}_{L_S}$ to be 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, to find $t^* = \arg\sup_{t\in\mathbb{R}} ta - \psi_{QU}(K-1, M, t)$ we set the derivative to zero, finding

$$a = \psi'_{QU} = \frac{(K-1)\mathbb{E}_Q[L_G e^{tL_G}]}{\mathbb{E}_Q[e^{tL_G}]} + M\frac{\mathbb{E}_U[L_S e^{tL_S}]}{\mathbb{E}_U[e^{tL_S}]}.$$

Also, by the definition of $\tilde{Q}$ and $\tilde{U}$,

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

Thus, by substituting in (4.136):

$$\mathbb{P}_{QU}\big(\sum_{\ell=1}^{K-1}\tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma\big) \geq e^{-\left(t^*a - \psi_{QU}(K-1,M,t^*)\right) - |t^*|\delta}\big(1 - \frac{(K-1)\tilde{\sigma}^2_{L_G} + M\tilde{\sigma}^2_{L_S}}{\delta^2}\big) \tag{4.137}$$

By direct computation, and Lemma 35,

$$\begin{aligned}(K-1)\tilde{\sigma}^2_{L_G} + M\tilde{\sigma}^2_{L_S} &= \psi''_{QU}(K-1, M, t) \\ &\leq B\big((K-1)D(P||Q) + MD(V||U)\big) \end{aligned} \tag{4.138}$$

for some positive constant B. This allows us to eliminate the Chebyshev term (asymptotically) by setting

$$\delta = \big((K-1)D(P||Q) + MD(V||U)\big)^{\frac{2}{3}}.$$

Moreover, for sufficiently large n:

$$\begin{aligned}
a &= (K-1)\gamma + \delta \\
&\leq K\gamma + \delta \\
&\overset{(a)}{\leq} \frac{KD(P||Q)}{2+\epsilon_o} + \frac{MD(V||U)}{1+\frac{\epsilon_o}{2}} + (KD(P||Q) + MD(V||U))^{\frac{2}{3}} \\
&\leq KD(P||Q) + MD(V||U)(\frac{1}{1+\frac{\epsilon_o}{2}} + o(1)) \qquad (4.139)
\end{aligned}$$

where (a) holds from (4.135). Thus, for sufficiently large n,

$$-(K-1)D(Q||P) - MD(U||V) \leq a \leq (K-1)D(P||Q) + MD(V||U)$$

Hence, by Lemma 34,

$$\begin{aligned}
t^* &= \arg\sup_{t\in\mathbb{R}} ta - \psi_{QU}(K-1, M, t) \\
&= \arg\sup_{t\in[0,1]} ta - \psi_{QU}(K-1, M, t).
\end{aligned}$$

Using this result and substituting in (4.137):

$$\begin{aligned}
\mathbb{P}_{QU}\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma\big) &\geq e^{-E_{QU}(a,K-1,M)-\delta} \\
&\geq e^{-E_{QU}(a,K,M)-\delta} \qquad (4.140)
\end{aligned}$$

where (4.140) holds because $t^* \in [0,1]$ and $\log(\mathbb{E}_Q[e^{tL_G}]) = (t-1)D_t(P||Q) \leq 0$, where $D_t(V||U) \geq 0$ is the Rényi-divergence between distributions P and Q (Polyanskiy and Wu, 2017). Moreover,

$$\begin{aligned}
E_{QU}(a,K,M) &= E_{QU}((K-1)\gamma + \delta, K, M) \\
&\leq E_{QU}(K\gamma + \delta, K, M) \\
&\leq E_{QU}(K\gamma, K, M) + \delta \qquad (4.141)
\end{aligned}$$

where (4.141) holds because $t \in [0,1]$ and, by (4.139),

$$a \in \big[-KD(Q||P) - MD(U||V), KD(P||Q) + MD(V||U) \big]$$

Also, by Lemma 35, for some positive constant B:

$$\begin{aligned} E_{QU}(0, K-1, M) &\geq B((K-1)D(Q||P) + MD(U||V)) \\ &\geq B'((K-1)D(P||Q) + MD(V||U)) \end{aligned} \tag{4.142}$$

where (4.142) holds for some positive constant B' because for bounded LLR $D(Q||P) \approx D(P||Q)$ and $D(U||V) \approx D(V||U)$. Thus, for sufficiently large n, and for some positive constant B'':

$$\begin{aligned} \delta &= ((K-1)D(P||Q) + MD(V||U))^{\frac{2}{3}} \leq (B'' E_{QU}(0, K-1, M))^{\frac{2}{3}} \\ &\leq (B'' E_{QU}(K\gamma, K, M))^{\frac{2}{3}} \end{aligned} \tag{4.143}$$

Combining Equations (4.141), (4.142), (4.143):

$$\begin{aligned} E_{QU}(a, K, M) + \delta &\leq E_{QU}(K\gamma, K, M) + 2\delta \\ &\leq E_{QU}(K\gamma, K, M) + 2(B'' E_{QU}(K\gamma, K, M))^{\frac{2}{3}} \end{aligned} \tag{4.144}$$

Substituting in (4.140):

$$\begin{aligned} \mathbb{P}_{QU}\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma \big) &\geq e^{-(E_{QU}(K\gamma,K,M) + 2(B'' E_{QU}(K\gamma,K,M))^{\frac{2}{3}})} \\ &\overset{(a)}{\geq} e^{-((1-\epsilon)\log(n) + 2(B''(1-\epsilon)\log(n))^{\frac{2}{3}})} \\ &\geq e^{-(1-\epsilon)\log(n)(1+o(1))} \end{aligned} \tag{4.145}$$

where (a) comes from the contraposition assumption that (4.12) does not hold, i.e., $E_{QU}(K\gamma, K, M) \leq (1-\epsilon)\log(n)$ for arbitrary small $\epsilon > 0$. Equation (4.145) 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 (4.119) does not hold for $\theta_n = \gamma$.

Similarly, we will show that (4.118) does not hold for $\theta_n = \gamma$. Define

$$\begin{aligned} K_o &= \frac{K}{\log(K)} = o(K) \\ \delta' &= \frac{(K_o-1)(D(P||Q)-\gamma)+6\sqrt{K_o}\sigma}{(K-K_o)D(P||Q)+MD(V||U)} \end{aligned} \tag{4.146}$$

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 35 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 (4.136):

$$\begin{aligned} &\mathbb{P}_{PV}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\gamma + \tilde{\theta}_n\big) \\ &= \mathbb{P}_{PV}\bigg(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-K_o)(\gamma - \delta' D(P||Q) - \frac{\delta'}{K-K_o}MD(V||U))\bigg) \\ &\overset{(a)}{\geq} e^{-\left(t^*a-\psi_{PV}(K-K_o,M,t^*)\right)-|t^*|\delta}\big(1-\frac{(K-K_o)\tilde{\sigma}^2_{L_G}+M\tilde{\sigma}^2_{L_S}}{\delta^2}\big) \\ &\overset{(b)}{\geq} e^{-(t^*a-\psi_{PV}(K-K_o,M,t^*))-|t^*|\delta}(1-o(1)) \end{aligned} \tag{4.147}$$

where (a) holds for $t^* = \arg\sup_{t\in\mathbb{R}} ta - \psi_{PV}(K-K_o, M, t)$ and by defining two new probability distributions $\tilde{P}$ and $\tilde{V}$ over the same support of P and V, respectively and $\tilde{\sigma}^2_{L_G}$, $\tilde{\mu}_{L_G}$, $\tilde{\sigma}^2_{L_S}$ and $\tilde{\mu}_{L_S}$ to be the variances and means of L_G and L_S under $\tilde{P}$ and $\tilde{V}$, respectively. (b) holds by choosing

$$\delta = ((K-K_o)D(P||Q) + MD(V||U))^{\frac{2}{3}}$$

and noticing that for bounded LLR,

$$\begin{aligned} (K-K_o)\tilde{\sigma}^2_{L_G} + M\tilde{\sigma}^2_{L_S} &= \psi''(K-K_o, M, t) \\ &\leq B((K-K_o)D(P||Q) + MD(V||U)), \end{aligned}$$

by Lemma 35 for some positive constant B.

Moreover, for sufficiently large n:

$$\begin{aligned}
a &= (K-K_o)(\gamma - \delta' D(P||Q) - \frac{\delta'}{K-K_o} MD(V||U)) - \delta \\
&= (1-o(1))(K\gamma - K\delta' D(P||Q) - \delta' MD(V||U)) - \delta \\
&\overset{(a)}{\leq} (KD(P||Q) + MD(V||U))(\frac{1}{1+\frac{\epsilon_o}{2}} - \delta' - o(1)) \\
&\overset{(b)}{\leq} KD(P||Q) + MD(V||U) \qquad (4.148)
\end{aligned}$$

where (a) holds from (4.135) and (b) holds because $(K-1)D(P||Q) + MD(V||U) \to \infty$ and $\delta' = o(1)$. Thus

$$a \in [-KD(Q||P) - MD(U||V), KD(P||Q) + MD(V||U)].$$

By Lemma 34,

$$\begin{aligned}
t^* &= \arg\sup_{t\in\mathbb{R}} ta - \psi_{PV}(K-K_o, M, t) \\
&= \arg\sup_{t\in[-1,0]} ta - \psi_{PV}(K, M, t).
\end{aligned}$$

Substituting in (4.147):

$$\mathbb{P}_{PV}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\gamma + \tilde{\theta}_n\big) \geq e^{-(E_{PV}(a,K,M)+\delta)}(1-o(1)) \qquad (4.149)$$

Moreover,

$$E_{PV}(a, K, M) \leq E_{PV}(K\gamma, K, M) + \delta'(KD(P||Q) + MD(V||U)) + \delta \qquad (4.150)$$

which holds because $t \in [-1, 0]$ and $a \in [-KD(Q||P) - MD(U||V), KD(P||Q) + MD(V||U)]$ by (4.148). Also, by Lemma 35, for some positive constant B

$$\begin{aligned}
E_{PV}(K\gamma, K, M) &\geq E_{PV}(\frac{KD(P||Q) + MD(V||U)}{1+\frac{\epsilon_o}{2}}, K, M) \\
&\geq B(KD(P||Q) + MD(V||U)) \qquad (4.151)
\end{aligned}$$

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

$$\begin{aligned}\delta &= (KD(P||Q) + MD(V||U))^{\frac{2}{3}} \\ &\leq (B'E_{PV}(K\gamma, K, M))^{\frac{2}{3}}\end{aligned} \tag{4.152}$$

Combining equations (4.150), (4.151), (4.152):

$$\begin{aligned}&E_{PV}(a, K, M) + \delta \leq E_{PV}(K\gamma, K, M) + \delta'(KD(P||Q) + MD(V||U)) + 2\delta \\ &\leq E_{PV}(K\gamma, K, M) + \delta' B'' E_{PV}(K\gamma, K, M) + 2(B'E_{PV}(K\gamma, K, M))^{\frac{2}{3}}\end{aligned} \tag{4.153}$$

for some positive constants B' and B''. Substituting in (4.149):

$$\begin{aligned}\mathbb{P}_{PV}\Big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\gamma + \tilde{\theta}_n\Big) &\geq e^{-E_{PV}(K\gamma,K,M)(1+\delta' B'' + \frac{2(B')^{\frac{2}{3}}}{(E_{PV}(K\gamma,K,M))^{\frac{1}{3}}})}(1-o(1)) \\ &\overset{(a)}{=} e^{-E_{PV}(K\gamma,K,M)(1+o(1))} \\ &\overset{(b)}{=} e^{(K\gamma - E_{QU}(K\gamma,K,M))(1+o(1))} \\ &\overset{(c)}{\geq} e^{(\epsilon\log(n) - \log(K))(1+o(1))} \\ &\geq e^{(\epsilon\log(K) - \log(K))(1+o(1))} \\ &= e^{-\log(K)(1-\epsilon+o(1))}\end{aligned} \tag{4.154}$$

where (a) holds because $\delta' = o(1)$, (b) holds because from Lemma 34 $E_{PV}(K\gamma, K, M) = E_{QU}(K\gamma, K, M) - K\gamma$ and (c) is due to the contraposition assumption that (4.12) does not hold, i.e., $E_{QU}(K\gamma, K, M) \leq (1-\epsilon)\log(n)$ for arbitrary small $\epsilon > 0$.

Equation (4.154) 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 (4.118) does not hold for $\theta_n = \gamma$.

Thus, if (4.12) does not hold, both (4.145) and (4.154) show that (4.118) and (4.119) does not hold simultaneously at $\theta_n = \gamma$. Thus, for any $\theta_n > \gamma$, (4.118) will not hold and for any $\theta_n < \gamma$, (4.119) will not hold, and hence, if (4.12) does not hold, then there does not exist θ_n such that (4.118) and (4.119) hold simultaneously. This concludes the proof.

4.4.8 Necessity of Theorem 4

Recall that Definition 3 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 4.

$$\psi_Q(t,m_1) \triangleq m_1 \log(\mathbb{E}_Q[e^{tL_G}]) \tag{4.155}$$

$$\psi_P(t,m_1) \triangleq m_1 \log(\mathbb{E}_P[e^{tL_G}]) \tag{4.156}$$

$$E_Q(\theta,m_1) \triangleq \sup_{t\in[0,1]} t\theta - \psi_Q(t,m_1) \tag{4.157}$$

$$E_P(\theta,m_1) \triangleq \sup_{t\in[-1,0]} t\theta - \psi_P(t,m_1) \tag{4.158}$$

The quantities introduced in Definition 3 reduce to Definition 4 by setting $m_2 = 0$, therefore Lemmas 34 and 35 continue to hold.

In view of Lemma 39, it suffices to test whether there exists θ_n such that both (4.118) and (4.119) hold. We will show that if one of the conditions (1)-(6) of Theorem 4 is not satisfied, then there does not exist θ_n such that (4.118) and (4.119) hold simultaneously.

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

$$\begin{aligned}
&\mathbb{P}\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell + \tilde{Z} \geq (K-1)\gamma\big) \\
&= \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_M=1}^{L_M} \Big[(\prod_{m=1}^{M} \alpha^m_{-,\ell_m}) \mathbb{P}_Q\big(\sum_{\ell=1}^{K-1} \tilde{W}_\ell \geq (K-1)\gamma - \sum_{m=1}^{M} h^m_{\ell_m} \big) \Big] \\
&\overset{(a)}{\geq} \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_M=1}^{L_M} \Big[(\prod_{m=1}^{M} \alpha^m_{-,\ell_m}) e^{-(t^*a-(K-1)\log(\mathbb{E}_Q[e^{t^*L_G}]))-|t^*|\delta} (1-o(1)) \Big]
\end{aligned} \tag{4.159}$$

where (a) holds by Lemma 37, where $t^* = \arg\sup_{t\in\mathbb{R}} (ta - (K-1)\log(\mathbb{E}_Q[e^{tL_G}]))$[4].

[4]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.

Under (4.18):

$$KD(Q||P) = \rho(a - b - bT)(1 + o(1))\log(n)$$
$$KD(P||Q) = \rho(aT + b - a)(1 + o(1))\log(n)$$

Thus, according to conditions of Theorem 4,

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

So, by Lemma 34,

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

Without loss of generality, we focus on one term of the nested sum in (4.159). Then,

- If $\sum_{m=1}^{M} h_{\ell_m}^m = o(\log(n))$ and both $\sum_{m=1}^{M}\log(\alpha_{+,\ell_m}^m)$ and $\sum_{m=1}^{M}\log(\alpha_{-,\ell_m}^m)$ are $o(\log(n))$, then by evaluating the supremum and by substituting in (4.159),

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

  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 (4.119) does not hold for $\theta_n = \gamma$.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = o(\log(n))$ and $\sum_{m=1}^{M}\log(\alpha_{+,\ell_m}^m) = \sum_{m=1}^{M}\log(\alpha_{-,\ell_m}^m) = -\beta\log(n) + o(\log(n)), \beta > 0$, then by evaluating the supremum and by substituting in (4.159),

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

  Thus, if $\eta_1(\rho, a, b) + \beta \leq 1 - \varepsilon$ 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 (4.119) does not hold for $\theta_n = \gamma$.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = \beta\log(n)+o(\log(n)), 0<\beta<\rho(a-b-bT), \sum_{m=1}^{M}\log(\alpha_{+,\ell_m}^m)=o(\log(n))$, then by evaluating the supremum and by substituting in (4.159),

$$\mathbb{P}\big(\sum_{\ell=1}^{K-1}\tilde{W}_\ell+\tilde{Z}\geq (K-1)\gamma\big)\geq n^{-\eta_2(\rho,a,b,\beta)+o(1)}$$

  Thus, if $\eta_2(\rho,a,b,\beta)\leq 1-\varepsilon$ 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 (4.119) does not hold for $\theta_n=\gamma$.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = -\beta\log(n)+o(\log(n)), 0<\beta<\rho(a-b-bT), \sum_{m=1}^{M}\log(\alpha_{-,\ell_m}^m)=o(\log(n))$, then by evaluating the supremum and by substituting in (4.159),

$$\mathbb{P}\big(\sum_{\ell=1}^{K-1}\tilde{W}_\ell+\tilde{Z}\geq (K-1)\gamma\big)\geq n^{-\eta_3(\rho,a,b,\beta)+o(1)}$$

  Thus, if $\eta_3(\rho,a,b,\beta)\leq 1-\varepsilon$ 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 (4.119) does not hold for $\theta_n=\gamma$.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = \beta\log(n)+o(\log(n)), 0<\beta<\rho(a-b-bT), \sum_{m=1}^{M}\log(\alpha_{+,\ell_m}^m)=-\beta'\log(n)+o(\log(n))$, then by evaluating the supremum and by substituting in (4.159),:

$$\mathbb{P}\big(\sum_{\ell=1}^{K-1}\tilde{W}_\ell+\tilde{Z}\geq (K-1)\gamma\big)\geq n^{-\eta_2(\rho,a,b,\beta)-\beta'+o(1)}$$

  Thus, if $\eta_2(\rho,a,b,\beta)+\beta'\leq 1-\varepsilon$ 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 (4.119) does not hold for $\theta_n=\gamma$.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = -\beta\log(n)+o(\log(n)), 0<\beta<\rho(a-b-bT), \sum_{m=1}^{M}\log(\alpha_{-,\ell_m}^m)=-\beta'\log(n)+o(\log(n))$, then by evaluating the supremum and by substituting in (4.159),

$$\mathbb{P}\big(\sum_{\ell=1}^{K-1}\tilde{W}_\ell+\tilde{Z}\geq (K-1)\gamma\big)\geq n^{-\eta_3(\rho,a,b,\beta)-\beta'+o(1)}$$

  Thus, if $\eta_3(\rho,a,b,\beta)+\beta'\leq 1-\varepsilon$ 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 (4.119) does not hold for $\theta_n=\gamma$.

Now we show that (4.118) 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,

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \leq (K-1)\gamma + \tilde{\theta}_n\Big) \\
&\overset{(a)}{\geq} \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_L=1}^{L_M} \Big(\prod_{m=1}^{M} \alpha^m_{+,\ell_m}\Big) e^{-(t^* a - (K-K_o)\log(\mathbb{E}_P[e^{t^* L_G}])) - |t^*|\delta} (1-o(1)) \\
&\overset{(b)}{=} \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_L=1}^{L_M} \Big(\prod_{m=1}^{M} \alpha^m_{+,\ell_m}\Big) e^{-(\lambda^* a - (K-K_o)\log(\mathbb{E}_Q[e^{\lambda^* L_G}])) + a - |\lambda^* - 1|\delta} (1-o(1))
\end{aligned} \tag{4.160}$$

where (a) holds by Lemma 37, 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 34.

Thus, according to conditions of Theorem 4,

$$a \in \big[-KD(Q||P), KD(P||Q)\big]. \tag{4.161}$$

Thus, by Lemma 34, $\arg\sup_{t\in\mathbb{R}}$ is replaced by $\arg\sup_{t\in[-1,0]}$.

- 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 (4.160),

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

  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 (4.118) does not hold for $\theta_n = \gamma$.

- If $\sum_{m=1}^{M} h^m_{\ell_m} = o(\log(n))$ and $\sum_{m=1}^{M} \log(\alpha^m_{+,\ell_m}) = \sum_{m=1}^{M} \log(\alpha^m_{-,\ell_m}) = -\beta\log(n) + o(\log(n)), \beta > 0$, then by evaluating the supremum and by substituting in (4.160),

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

  Thus, if $\eta_1(\rho,a,b) + \beta \leq 1-\varepsilon$ 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 (4.118) does not hold for $\theta_n = \gamma$.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = \beta\log(n)+o(\log(n)), 0<\beta<\rho(a-b-bT), \sum_{m=1}^{M}\log(\alpha_{+,\ell_m}^m)=o(\log(n))$, then by evaluating the supremum and by substituting in (4.160),

$$\mathbb{P}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \le (K-1)\gamma + \tilde{\theta}_n\big) \ge n^{-\eta_2(\rho,a,b,\beta)+o(1)}$$

  Thus, if $\eta_2(\rho,a,b,\beta)\le 1-\varepsilon$ for some $0<\varepsilon<1$, then $K\mathbb{P}(\sum_{\ell=1}^{K-K_o} W_\ell + Z \le (K-1)\gamma+\tilde{\theta}_n)\ge n^{\varepsilon+o(1)}$ which shows that (4.118) does not hold for $\theta_n=\gamma$.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = -\beta\log(n)+o(\log(n)), 0<\beta<\rho(a-b-bT), \sum_{m=1}^{M}\log(\alpha_{-,\ell_m}^m)=o(\log(n))$, then by evaluating the supremum and by substituting in (4.160),

$$\mathbb{P}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \le (K-1)\gamma + \tilde{\theta}_n\big) \ge n^{-\eta_3(\rho,a,b,\beta)+o(1)}$$

  Thus, if $\eta_3(\rho,a,b,\beta)\le 1-\varepsilon$ for some $0<\varepsilon<1$, then $K\mathbb{P}(\sum_{\ell=1}^{K-K_o} W_\ell + Z \le (K-1)\gamma+\tilde{\theta}_n)\ge n^{\varepsilon+o(1)}$ which shows that (4.118) does not hold for $\theta_n=\gamma$.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = \beta\log(n)+o(\log(n)), 0<\beta<\rho(a-b-bT), \sum_{m=1}^{M}\log(\alpha_{+,\ell_m}^m)=-\beta'\log(n)+o(\log(n))$, then by evaluating the supremum and by substituting in (4.160),

$$\mathbb{P}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \le (K-1)\gamma + \tilde{\theta}_n\big) \ge n^{-\eta_2(\rho,a,b,\beta)-\beta'+o(1)}$$

  Thus, if $\eta_2(\rho,a,b,\beta)+\beta'\le 1-\varepsilon$ for some $0<\varepsilon<1$, then $K\mathbb{P}(\sum_{\ell=1}^{K-K_o} W_\ell + Z \le (K-1)\gamma+\tilde{\theta}_n)\ge n^{\varepsilon+o(1)}$ which shows that (4.118) does not hold for $\theta_n=\gamma$.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = -\beta\log(n)+o(\log(n)), 0<\beta<\rho(a-b-bT), \sum_{m=1}^{M}\log(\alpha_{-,\ell_m}^m)=-\beta'\log(n)+o(\log(n))$, then by evaluating the supremum and by substituting in (4.160),

$$\mathbb{P}\big(\sum_{\ell=1}^{K-K_o} W_\ell + Z \le (K-1)\gamma + \tilde{\theta}_n\big) \ge n^{-\eta_3(\rho,a,b,\beta)-\beta'+o(1)}$$

  Thus, if $\eta_3(\rho,a,b,\beta)+\beta'\le 1-\varepsilon$ for some $0<\varepsilon<1$, then $K\mathbb{P}(\sum_{\ell=1}^{K-K_o} W_\ell + Z \le (K-1)\gamma+\tilde{\theta}_n)\ge n^{\varepsilon+o(1)}$ which shows that (4.118) does not hold for $\theta_n=\gamma$.

To summarize, when $\theta_n = \gamma$, if one of the conditions (1)-(6) of Theorem 4 does not hold, then (4.118) and (4.119) cannot hold simultaneously. Thus, for any $\theta_n > \gamma$, (4.118) will not hold and for any $\theta_n < \gamma$, (4.119) will not hold, and hence, if one of the conditions (1)-(6) of Theorem 4 does not hold, then there does not exist θ_n such that (4.118) and (4.119) hold simultaneously. This concludes the proof of the necessary conditions.

Finally, we comment on how the proof would change if instead of the regime (4.18), K was chosen such that for all large n, $\log(\frac{n}{K}) = (C - o(1))\log(n)$ for some constant $C \in (0, 1]$. A key step in the proof was to ensure that θ in definition 4 is between $[-KD(Q||P),\ KD(P||Q)]$, e.g, see (4.161). Hence, the only modification needed is to take C into account. For example, when $\sum_{m=1}^{M} h_{\ell_m}^m = \beta \log(n) + o(\log(n))$ for some positive β, then a condition on β would be $-\rho(a - b - bT) < C \pm \beta < \rho(a - b - bT)$. The proofs for the modified regime would then follow a similar strategy as the proofs in this section. Similar modifications are needed for the sufficiency proofs as well.

4.4.9 Sufficiency of Theorem 4

The sufficient conditions are derived via Algorithm 4.1 provided in Section 4.2.1 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 20 cannot be used for the weak recovery step. Instead the maximum likelihood detector without side information provided in (Hajek et al., 2017) will be used.

The following lemma gives sufficient conditions for Algorithm 4.1 to achieve exact recovery.

Lemma 41. *Define $C_k^* = C^* \cap {S_k}^c$ and assume $\hat{C}_k$ achieves weak recovery, i.e.*

$$\mathbb{P}\big(|\hat{C}_k \triangle C_k^*| \leq \delta K \ \mathit{for}\ 1 \leq k \leq \frac{1}{\delta}\big) \rightarrow 1 \tag{4.162}$$

Under conditions (4.18), *if conditions* (1)-(6) *of Theorem 4 hold, then* $\mathbb{P}(\tilde{C} = C^*) \rightarrow 1$.

Proof. Please see Appendix 4.4.10 $\square$

In view of Lemma 41, it suffices to show that there exists an estimator that achieves weak recovery for a random cluster size and satisfies (4.162). We use the estimator presented in (Hajek et al., 2017, 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:

$$KD(P||Q) \to \infty \tag{4.163}$$

$$\liminf_{n\to\infty} \frac{(K-1)D(P||Q)}{\log(\frac{n}{K})} \geq 2 \tag{4.164}$$

$$\mathbb{P}\left(\left||C_k^*| - (1-\delta)K\right| \geq \frac{K}{\log(K)}\right) \leq o(1) \tag{4.165}$$

It is obvious that in the regime (4.18), both (4.163) and (4.164) are satisfied. Thus, it remains to show that (4.165) holds too. Let $\hat{C}_k$ be the ML estimator for C_k^* based on observing $\boldsymbol{G}_k$ defined in Algorithm 4.1. 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 $\mathrm{Bin}(n(1-\delta), \frac{K}{n})$. Therefore, the Chernoff bound for Z also holds for $|C_k^*|$. Thus,

$$\mathbb{P}\left(\left||C_k^*| - (1-\delta)K\right| \geq \frac{K}{\log(K)}\right) \leq o(1) \tag{4.166}$$

Thus, (4.165) holds, which implies that ML achieves weak recovery with K replaced with $\lceil(1-\delta)K\rceil$ in (Hajek et al., 2017, Lemma 4). Thus, from (Hajek et al., 2017, Lemma 4), for any $1 \leq k \leq \frac{1}{\delta}$:

$$\mathbb{P}\left(\frac{|\hat{C}_k \triangle C_k^*|}{K} \leq 2\epsilon + \frac{1}{\log(K)}\right) \geq 1 - o(1) \tag{4.167}$$

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

$$\mathbb{P}\left(\frac{|\hat{C}_k \triangle C_k^*|}{K} \leq 2\epsilon + \frac{1}{\log(K)} \quad \forall 1 \leq k \leq \frac{1}{\delta}\right) \geq 1 - o(1) \tag{4.168}$$

Since $\epsilon = o(1)$, the desired (4.162) holds.

4.4.10 Proof of Lemma 41

To prove Lemma 41, we follow essentially the same strategy used for Lemma 21 in Appendix 4.4.5. 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 42. *In the regime* (4.18)*, suppose conditions* (1)-(6) *of Theorem 4 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* $\gamma = \frac{\log(\frac{n}{K})}{K}$*:*

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-\delta)} \tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big) = o(\frac{1}{n}) \tag{4.169}$$

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \leq K(1-\delta)\gamma\Big) = o(\frac{1}{K}) \tag{4.170}$$

Proof. Using the Chernoff bound:

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{\ell=1}^{K(1-\delta)} \tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big) \\
&\leq \mathbb{P}\Big(\sum_{\ell=1}^{K} \tilde{W}_\ell + \tilde{Z} \geq K(1-\delta)\gamma\Big) \\
&\leq \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_M=1}^{L_M} \Big(\prod_{m=1}^{M} \alpha^m_{-,\ell_m}\Big) e^{-\sup_{t\geq 0} t(K(1-\delta)\gamma - \sum_{m=1}^{M} h^m_{\ell_m}) - K\log(\mathbb{E}_Q[e^{tL_G}])}
\end{aligned} \tag{4.171}$$

The terms inside the nested sum in (4.171) 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 $\sum_{m=1}^M h_{\ell_m}^m = \beta\log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT)$, $\sum_{m=1}^M \log(\alpha_{+,\ell_m}^m) = o(\log(n))$.

- $n^{-\eta_3(\rho,a,b,\beta)+o(1)}$, if $\sum_{m=1}^M h_{\ell_m}^m = -\beta\log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT)$, $\sum_{m=1}^M \log(\alpha_{-,\ell_m}^m) = o(\log(n))$.

- $n^{-\eta_2(\rho,a,b,\beta)-\beta'+o(1)}$, if $\sum_{m=1}^M h_{\ell_m}^m = \beta\log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT)$, $\sum_{m=1}^M \log(\alpha_{+,\ell_m}^m) = -\beta'\log(n) + o(\log(n))$.

- $n^{-\eta_3(\rho,a,b,\beta)-\beta'+o(1)}$, if $\sum_{m=1}^M h_{\ell_m}^m = -\beta\log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT)$, $\sum_{m=1}^M \log(\alpha_{-,\ell_m}^m) = -\beta'\log(n) + o(\log(n))$.

Since M and L_m are independent of n and finite, it follows that if items (1)-(6) of Theorem 4 are satisfied, then Equation (4.169) holds.

To show (4.170), Chernoff bound is used.

$$\mathbb{P}\Big(\sum_{\ell=1}^{K(1-2\delta)} W_\ell + \sum_{\ell=1}^{\delta K} \tilde{W}_\ell + Z \leq K(1-\delta)\gamma\Big)$$
$$\leq \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_M=1}^{L_M} (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta\log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \tag{4.172}$$

Without loss of generality, we focus on one term inside the nested sum in(4.172):

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

$$(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta\log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta}$$
$$\leq (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{(1-2\delta)\left(t(k\gamma - \frac{\sum_{m=1}^M h_{\ell_m}^m}{1-2\delta}) + K\log(\mathbb{E}_P[e^{-tL_G}])\right)} e^{\delta\left(tK\gamma + K\log(\mathbb{E}_Q[e^{-tL_G}])\right)} \tag{4.173}$$

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

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

Define $\theta \triangleq K\gamma - \frac{\sum_{m=1}^{M} h_{\ell_m}^m}{1-2\delta}$ and choose $t^* \in [0,1]$, such that $t^*\theta + K\log(\mathbb{E}[e^{-t^* L_G}]) = -E_P(\theta, K)$. Substituting in (4.173):

$$
\begin{aligned}
&(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^{M} h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{-(1-2\delta)E_P(\theta,K) + \delta\left(t^* K\gamma + K\log(\mathbb{E}_Q[e^{-t^* L_G}])\right)} \\
&\leq (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{-(1-2\delta)E_P(\theta,K) + \delta\left(K\gamma + K\log(\mathbb{E}_Q[e^{-t^* L_G}])\right)} \qquad (4.174)
\end{aligned}
$$

where the last inequality holds because $t^* \in [0,1]$. Also, by Lemma 35 and convexity of $\log(\mathbb{E}_Q[e^{-tL_G}])$, the following holds for some positive constant A:

$$
K\log(\mathbb{E}_Q[e^{-t^* L_G}]) \leq K\log(\mathbb{E}_Q[e^{-L_G}]) \leq AKD(Q||P) \qquad (4.175)
$$

Moreover, by Lemma 35, $E_P[\theta, K] = E_Q[\theta, K] - \theta$ and $E_Q[\theta, K] \geq E_Q[0, K] \geq A_1 KD(Q||P)$. Combining the last observation with (4.175), for some positive constant A_2,

$$
\begin{aligned}
&(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^{M} h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{-(1-2\delta)(E_Q(\theta,K)-\theta) + \delta K\gamma + \delta A_2 E_Q(\theta,K)} \\
&= (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{-E_Q(\theta,K)(1-2\delta-\delta A_2) + (1-2\delta)\theta + \delta K\gamma} \qquad (4.176)
\end{aligned}
$$

Since $\sum_{m=1}^{M} \log(\alpha_{+,\ell_m}^m) = o(\log(n))$, evaluating the supremum in $E_Q[\theta, K]$ and substituting in (4.176) leads to:

$$
\begin{aligned}
&(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^{M} h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq e^{-\log(n)(1-2\delta-\delta A_2)(\eta_1 + o(1))} \\
&\leq n^{-(1+\varepsilon)(1-2\delta-\delta A_2) + o(1))} \qquad (4.177)
\end{aligned}
$$

where (4.177) holds by assuming $\eta_1 \geq 1+\varepsilon$ for some $\varepsilon > 0$. Multiplying (4.177) by K:

$$K(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta}$$
$$\leq n^{1-(1+\varepsilon)(1-2\delta-\delta A_2)+o(1))} \tag{4.178}$$

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 Lemma 42.

- If $\sum_{m=1}^M h_{\ell_m}^m = o(\log(n))$ and $\sum_{m=1}^M \log(\alpha_{+,\ell_m}^m) = \sum_{m=1}^M \log(\alpha_{-,\ell_m}^m) = -\beta \log(n) + o(\log(n)), \beta > 0$, then:

$$(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta}$$
$$\leq \prod_{m=1}^{M} (\alpha_{+,\ell_m}^m) e^{-\log(n)(1-2\delta-\delta A_2)(\eta_1+o(1))} \tag{4.179}$$

Since $\sum_{m=1}^M \log(\alpha_{+,\ell_m}^m) = -\beta \log(n) + o(\log(n)), \beta > 0$:

$$(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta}$$
$$\leq e^{-\log(n)(1-2\delta-\delta A_2)(\eta_1 + \frac{\beta}{1-2\delta-\delta A_2} + o(1))}$$
$$\leq e^{-\log(n)(1-2\delta-\delta A_2)(\eta_1+\beta+o(1))} \tag{4.180}$$

where the last inequality holds because $0 < 1-2\delta-\delta A_2 < 1$ for sufficiently small δ. Thus:

$$K(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta}$$
$$\leq n^{1-(\eta_1+\beta)(1-2\delta-\delta A_2)+o(1)}$$
$$\leq n^{1-(1+\varepsilon)(1-2\delta-\delta A_2)+o(1))} \tag{4.181}$$

where the last inequality holds by assuming $\eta_1 + \beta \geq 1+\varepsilon$ for some $\varepsilon > 0$. 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 second case of Lemma 42.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = \beta \log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT), \sum_{m=1}^{M} \log(\alpha_{+,\ell_m}^m) = o(\log(n))$, then:

$$\begin{aligned}
&(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^{M} h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{(1-2\delta)\left(t(k\gamma - \sum_{m=1}^{M} h_{\ell_m}^m) + K\log(\mathbb{E}_P[e^{-tL_G}])\right)} e^{\delta\left(t(k\gamma - \sum_{m=1}^{M} h_{\ell_m}^m) + K\log(\mathbb{E}_Q[e^{-tL_G}])\right)}
\end{aligned} \tag{4.182}$$

Since $\sum_{m=1}^{M} h_{\ell_m}^m = \beta \log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT)$, it is easy to show that

$$K\gamma - \sum_{m=1}^{M} h_{\ell_m}^m \in [-KD(Q||P)\ ,\ KD(P||Q)]$$

Define $\theta \triangleq K\gamma - \sum_{m=1}^{M} h_{\ell_m}^m$ and choose $t^* \in [0,1]$, such that $t^*\theta + K\log(\mathbb{E}[e^{-t^* L_G}]) = -E_P(\theta, K)$. Substituting in (4.182):

$$\begin{aligned}
&(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^{M} h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{-(1-2\delta)E_P[\theta,K] + \delta\left(t^*(K\gamma - \sum_{m=1}^{M} h_{\ell_m}^m) + K\log(\mathbb{E}_Q[e^{-t^* L_G}])\right)}
\end{aligned} \tag{4.183}$$

By Lemma 35 and convexity of $\log(\mathbb{E}_Q[e^{-t^* L_G}])$, the following holds for some positive constant A:

$$K\log(\mathbb{E}_Q[e^{-t^* L_G}]) \leq K\log(\mathbb{E}_Q[e^{-L_G}]) \leq AKD(Q||P) \tag{4.184}$$

Moreover, since

$$-KD(Q||P) < K\gamma - \sum_{m=1}^{M} h_{\ell_m}^m < 0 \quad ,$$

it follows that $\theta = -(1-\tilde{\eta})KD(Q||P)$ for some $\tilde{\eta} \in (0,1)$. Thus, by Lemma 35, for some positive constant A_1:

$$\begin{aligned}
E_Q[\theta, K] &= E_Q[-(1-\tilde{\eta})KD(Q||P), K] \\
&\geq A_1 KD(Q||P)
\end{aligned}$$

$$\geq \frac{A_1}{A} K \log(\mathbb{E}_Q[e^{-t^* L_G}])$$

where the last inequality holds because of (4.184). Substituting in (4.183), for some positive constant A_2,

$$\begin{aligned}
&(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{-(1-2\delta)(E_Q[\theta,K]-\theta) + \delta A_2 E_Q[\theta,K]} \\
&\leq (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{-E_Q[\theta,K](1-2\delta-\delta A_2) + (1-2\delta)\theta} \qquad (4.185)
\end{aligned}$$

Since $\sum_{m=1}^M \log(\alpha_{+,\ell_m}^m) = o(\log(n))$, by evaluating the supremum in $E_Q[\theta, K]$, multiplying by K and substituting in (4.185):

$$\begin{aligned}
&K(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq K e^{-\log(n)(1-2\delta-\delta A_2)(\eta_2 - \beta + \frac{(1-2\delta)\beta}{1-2\delta-\delta A_2} + o(1))} \\
&\overset{(a)}{\leq} K e^{-\log(n)(1-2\delta-\delta A_2)(\eta_2 + o(1))} \\
&\leq n^{1-(1+\varepsilon)(1-2\delta-\delta A_2)+o(1))} \qquad (4.186)
\end{aligned}$$

where (a) holds for sufficiently small δ. 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 third case of Lemma 42.

- If $\sum_{m=1}^M h_{\ell_m}^m = -\beta \log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT), \sum_{m=1}^M \log(\alpha_{-,\ell_m}^m) = o(\log(n))$, then:

$$\begin{aligned}
&(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^M h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq (\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{(1-2\delta)\left(t(k\gamma - \sum_{m=1}^M h_{\ell_m}^m) + K\log(\mathbb{E}_P[e^{-tL_G}])\right)} e^{\delta\left(t(k\gamma - 2\sum_{m=1}^M h_{\ell_m}^m) + K\log(\mathbb{E}_Q[e^{-tL_G}])\right)}
\end{aligned}$$

$$(4.187)$$

Following similar analysis as in (4.185):

$$
\begin{aligned}
&(\prod_{m=1}^{M} \alpha^{m}_{+,\ell_m}) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^{M} h^{m}_{\ell_m}) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq (\prod_{m=1}^{M} \alpha^{m}_{+,\ell_m}) e^{-\log(n)(1-2\delta-\delta A_2)(\eta_3+o(1))} e^{-\sum_{m=1}^{M} h^{m}_{\ell_m}} \qquad (4.188)
\end{aligned}
$$

Since $\sum_{m=1}^{M} \log(\alpha^{m}_{-,\ell_m}) = o(\log(n))$, by multiplying by K:

$$
\begin{aligned}
&K(\prod_{m=1}^{M} \alpha^{m}_{+,\ell_m}) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^{M} h^{m}_{\ell_m}) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq K e^{-\log(n)(1-2\delta-\delta A_2)(\eta_3+o(1))} \\
&\leq K n^{-(\eta_3+o(1))(1-2\delta-\delta A_2)} \\
&\leq n^{1-(1+\varepsilon)(1-2\delta-\delta A_2)+o(1))} \qquad (4.189)
\end{aligned}
$$

where the last inequality holds by assuming $\eta_3 \geq 1+\varepsilon$ for some $\varepsilon > 0$. 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 fourth case of Lemma 42.

- If $\sum_{m=1}^{M} h^{m}_{\ell_m} = \beta\log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT), \sum_{m=1}^{M} \log(\alpha^{m}_{+,\ell_m}) = -\beta'\log(n) + o(\log(n))$, then following similar analysis as in (4.185):

$$
\begin{aligned}
&K(\prod_{m=1}^{M} \alpha^{m}_{+,\ell_m}) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^{M} h^{m}_{\ell_m}) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
&\leq K n^{-(1-2\delta-\delta A_2)(\eta_2+\beta'+o(1))} \\
&\leq n^{1-(1+\varepsilon)(1-2\delta-\delta A_2)+o(1))} \qquad (4.190)
\end{aligned}
$$

where the last inequality holds by assuming $\eta_2 + \beta' \geq 1+\varepsilon$ for some $\varepsilon > 0$. 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 fifth case of Lemma 42.

- If $\sum_{m=1}^{M} h_{\ell_m}^m = -\beta \log(n) + o(\log(n)), 0 < \beta < \rho(a-b-bT), \sum_{m=1}^{M} \log(\alpha_{-,\ell_m}^m) = -\beta' \log(n) + o(\log(n))$, then following similar analysis as in (4.185):

$$
\begin{aligned}
& K(\prod_{m=1}^{M} \alpha_{+,\ell_m}^m) e^{t(K(1-2\delta)\gamma - \sum_{m=1}^{M} h_{\ell_m}^m) + K(1-2\delta)\log(\mathbb{E}_P[e^{-tL_G}])} e^{K\delta \log(\mathbb{E}_Q[e^{-tL_G}]) + tK\gamma\delta} \\
& \leq K n^{-(1-2\delta-\delta A_2)(\eta_3+\beta'+o(1))} \\
& \leq n^{1-(1+\varepsilon)(1-2\delta-\delta A_2)+o(1))} \qquad (4.191)
\end{aligned}
$$

where the last inequality holds by assuming $\eta_3 + \beta' \geq 1 + \varepsilon$ for some $\varepsilon > 0$. 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 last case of Lemma 42.

□

The proof of Lemma 41 then follows similarly as the proof of Lemma 21.

4.4.11 Auxiliary Lemmas For Belief Propagation

Lemma 43. *Recall the definition of Γ_0^t from (4.31). For any measurable function $g(.)$:*

$$\mathbb{E}[g(\Gamma_0^t)|\tau_0 = 0] = \mathbb{E}[g(\Gamma_0^t)e^{-\Gamma_0^t}|\tau_0 = 1] \qquad (4.192)$$

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

$$
\begin{aligned}
\mathbb{E}[g(\Gamma_0^t)|\tau_0 = 0] &= \mathbb{E}_{Y|\tau_0=0}[g(\Gamma_0^t)] \\
&= \int_Y g(\Gamma_0^t) \frac{\mathbb{P}(Y|\tau_0 = 0)}{\mathbb{P}(Y|\tau_0 = 1)} \mathbb{P}(Y|\tau_0 = 1) \\
&= \int_Y g(\Gamma_0^t) e^{-\Gamma_0^t} \mathbb{P}(Y|\tau_0 = 1) \\
&= \mathbb{E}_{Y|\tau_0=1}[g(\Gamma_0^t)e^{-\Gamma_0^t}] \\
&= \mathbb{E}[g(\Gamma_0^t)e^{-\Gamma_0^t}|\tau_0 = 1] \qquad (4.193)
\end{aligned}
$$

□

Lemma 44. *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* $t \geq 0$

$$a_{t+1} = \mathbb{E}[e^{Z_1^t+U_1}] = \Lambda e^{\lambda b_t} \tag{4.194}$$

$$\mathbb{E}[e^{3(Z_0^t+U_0)}] = \mathbb{E}[e^{2(Z_1^t+U_1)}] = \mathbb{E}[e^{3U_0}]e^{3\lambda b_t + \frac{\lambda^2}{K(p-q)}\mathbb{E}[(\frac{e^{Z_1^t+U_1}}{1+e^{Z_1^t+U_1-\nu}})^2]} \tag{4.195}$$

Proof. The first equality in (4.194) holds by Lemma 43 for $g(x) = e^{2x}$. Similarly, the first equality in (4.195) holds by Lemma 43 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:

$$\begin{aligned} a_{t+1} &= \mathbb{E}[e^{2(Z_0^t+U_0)}] \\ &\overset{(a)}{=} e^{-2K(p-q)}\mathbb{E}[e^{2U_0}]\mathbb{E}[(\mathbb{E}[f^2(e^{Z_1^t+U_1-\nu})])^{H_u}]\mathbb{E}[(\mathbb{E}[f^2(e^{Z_0^t+U_0-\nu})])^{F_u}] \\ &\overset{(b)}{=} \Lambda e^{-2K(p-q)} e^{Kq(\mathbb{E}[f^2(e^{Z_1^t+U_1-\nu})]-1)} e^{(n-K)q(\mathbb{E}[f^2(e^{Z_0^t+U_0-\nu})]-1)} \end{aligned} \tag{4.196}$$

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 $f(x)$:

$$\begin{aligned} &Kq\big(\mathbb{E}\big[f^2(e^{Z_1^t+U_1-\nu})\big] - 1\big) + (n-K)q\big(\mathbb{E}\big[f^2(e^{Z_0^t+U_0-\nu})\big] - 1\big) \\ &= Kq\mathbb{E}\left[\frac{2(\frac{p}{q}-1)}{1+e^{-(Z_1^t+U_1-\nu)}} + \frac{(\frac{p}{q}-1)^2}{(1+e^{-(Z_1^t+U_1-\nu)})^2}\right] \\ &\quad + (n-K)q\mathbb{E}\left[\frac{2(\frac{p}{q}-1)}{1+e^{-(Z_0^t+U_0-\nu)}} + \frac{(\frac{p}{q}-1)^2}{(1+e^{-(Z_0^t+U_0-\nu)})^2}\right] \\ &\overset{(a)}{=} 2K(p-q) + Kq(\frac{p}{q}-1)^2\mathbb{E}\left[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\right] \\ &\overset{(b)}{=} 2K(p-q) + \lambda b_t \end{aligned} \tag{4.197}$$

where (a) holds by Lemma 43 and (b) holds by the definition of λ and b_t.

Using (4.197) and substituting in (4.196) concludes the proof of (4.194). The proof of (4.195) follows similarly using $f^3(x)$ instead of $f^2(x)$. $\square$

4.4.12 Proof of Lemma 22

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:

$$
\begin{aligned}
\Gamma_0^{t+1} &= \log\left(\frac{\mathbb{P}(T^{t+1}, \tilde{\tau}^{t+1}|\tau_0 = 1)}{\mathbb{P}(T^{t+1}, \tilde{\tau}^{t+1}|\tau_0 = 0)}\right) \\
&= \log\left(\frac{\mathbb{P}(N_0, \tilde{\tau}_0, \{T_k^t\}_{k\in\mathcal{N}_0}, \{\tilde{\tau}_k^t\}_{k\in\mathcal{N}_0}|\tau_0 = 1)}{\mathbb{P}(N_0, \tilde{\tau}_0, \{T_k^t\}_{k\in\mathcal{N}_0}, \{\tilde{\tau}_k^t\}_{k\in\mathcal{N}_0}|\tau_0 = 0)}\right) \\
&\overset{(a)}{=} \log\left(\frac{\mathbb{P}\big(N_0, \tilde{\tau}_0|\tau_0 = 1\big)}{\mathbb{P}\big(N_0, \tilde{\tau}_0|\tau_0 = 0\big)}\right) + \log\left(\frac{\prod_{k\in\mathcal{N}_0}\mathbb{P}\big(T_k^t, \tilde{\tau}_k^t|\tau_0 = 1\big)}{\prod_{k\in\mathcal{N}_0}\mathbb{P}\big(T_k^t, \tilde{\tau}_k^t|\tau_0 = 0\big)}\right) \\
&\overset{(b)}{=} \log\left(\frac{\mathbb{P}\big(N_0|\tau_0 = 1\big)}{\mathbb{P}\big(N_0|\tau_0 = 0\big)}\right) + \log\left(\frac{\mathbb{P}\big(\tilde{\tau}_0|\tau_0 = 1\big)}{\mathbb{P}\big(\tilde{\tau}_0|\tau_0 = 0\big)}\right) \\
&+ \sum_{k\in\mathcal{N}_0}\log\left(\frac{\sum_{\tau_k\in\{0,1\}}\mathbb{P}\big(T_k^t, \tilde{\tau}_k^t|\tau_k\big)\mathbb{P}\big(\tau_k|\tau_0 = 1\big)}{\sum_{\tau_k\in\{0,1\}}\mathbb{P}\big(T_k^t, \tilde{\tau}_k^t|\tau_k\big)\mathbb{P}\big(\tau_k|\tau_0 = 0\big)}\right) \\
&\overset{(c)}{=} -K(p-q) + h_0 + \sum_{k\in\mathcal{N}_0}\log\Big(\frac{\frac{p}{q}e^{\Gamma_k^t - \nu} + 1}{e^{\Gamma_k^t - \nu} + 1}\Big)
\end{aligned} \tag{4.198}
$$

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 $\tau_0 = 0$.

4.4.13 Proof of Lemma 23

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

$$\begin{aligned}\mathbb{E}\big[e^{\frac{Z_0^t}{2}}\big] &= e^{\frac{-K(p-q)}{2}}\mathbb{E}_{H_0}\left[(\mathbb{E}_{Z_1U_1}[f^{\frac{1}{2}}(e^{Z_1^t+U_1-\nu})])^{H_0}\right]\mathbb{E}_{F_0}\left[(\mathbb{E}_{Z_0U_0}[f^{\frac{1}{2}}(e^{Z_0^t+U_0-\nu})])^{F_0}\right] \\ &\overset{(a)}{=} e^{\frac{-K(p-q)}{2}}e^{Kq(\mathbb{E}[f^{\frac{1}{2}}(e^{Z_1^t+U_1-\nu})]-1)}e^{(n-K)q(\mathbb{E}[f^{\frac{1}{2}}(e^{Z_0^t+U_0-\nu})]-1)} \end{aligned} \tag{4.199}$$

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,

$$\sqrt{1+x} \leq 1 + \frac{x}{2} - \frac{x^2}{8(1+A)^{1.5}}, \qquad 0 \leq x \leq A \tag{4.200}$$

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

$$\begin{aligned}&\left(\frac{1+\frac{p}{q}e^{Z_0^t+U_0-\nu}}{1+e^{Z_0^t+U_0-\nu}}\right)^{\frac{1}{2}} \\ &= \left(1 + \frac{\frac{p}{q}-1}{1+e^{-(Z_0^t+U_0-\nu)}}\right)^{\frac{1}{2}} \\ &\leq 1 + \frac{1}{2}\frac{\frac{p}{q}-1}{1+e^{-(Z_0^t+U_0-\nu)}} - \frac{1}{8B}\frac{(\frac{p}{q}-1)^2}{(1+e^{-(Z_0^t+U_0-\nu)})^2}\end{aligned} \tag{4.201}$$

It follows that:

$$\begin{aligned}&Kq\big(\mathbb{E}[f^{\frac{1}{2}}(e^{Z_1^t+U_1-\nu})]-1\big) + (n-K)q\big(\mathbb{E}[f^{\frac{1}{2}}(e^{Z_0^t+U_0-\nu})]-1\big) \\ &\leq \frac{Kq(\frac{p}{q}-1)}{2}\left(\mathbb{E}\left[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\right] + e^{\nu}\mathbb{E}\left[\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\right]\right) \\ &- \frac{Kq(\frac{p}{q}-1)^2}{8B}\left(\mathbb{E}\left[\frac{1}{(1+e^{-(Z_1^t+U_1-\nu)})^2}\right] + e^{\nu}\mathbb{E}\left[\frac{1}{(1+e^{-(Z_0^t+U_0-\nu)})^2}\right]\right) \\ &\overset{(a)}{\leq} \frac{K(p-q)}{2} - \frac{K(p-q)^2}{8Bq}\mathbb{E}\left[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\right] \qquad (4.202) \\ &= \frac{K(p-q)}{2} - \frac{\lambda}{8B}b_t \qquad (4.203)\end{aligned}$$

where (a) holds by the following consequence of Lemma 43 (from Appendix 4.4.11):

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

$$\mathbb{E}\left[\frac{1}{(1+e^{-(Z_1^t+U_1-\nu)})^2}\right]+e^{\nu}\mathbb{E}\left[\frac{1}{(1+e^{-(Z_0^t+U_0-\nu)})^2}\right]=\mathbb{E}\left[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\right] \tag{4.204}$$

Using (4.199) and (4.203):

$$\mathbb{E}\left[e^{\frac{Z_0^t+U_0}{2}}\right]\leq \mathbb{E}\left[e^{\frac{U_0}{2}}\right]\, e^{\frac{-\lambda}{8B}b_t} \tag{4.205}$$

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

$$\mathbb{E}\left[e^{\frac{Z_0^t+U_0}{2}}\right]\geq \mathbb{E}\left[e^{\frac{U_0}{2}}\right]\, e^{\frac{-\lambda}{8}b_t} \tag{4.206}$$

4.4.14 Proof of Lemma 24

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 44 in Appendix 4.4.11, $a_{t+1}=\Lambda e^{\lambda b_t}$.

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 $\lambda\leq\frac{1}{\Lambda e}$, $b_t\leq\frac{v_{t+1}}{\lambda}\leq\Lambda e$.

4.4.15 Proof of Lemma 25

By definition of a_t, we have:

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

$$\leq \mathbb{E}\Big[\frac{e^{Z_1^t+U_1}}{1+e^{Z_1^t+U_1-\nu}}\Big]$$
$$= b_{t+1}$$

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

$$\begin{aligned}
b_{t+1} &\geq a_{t+1} - \mathbb{E}[e^{-\nu+2(Z_1^{t+1}+U_1)}] \\
&\overset{(a)}{=} \Lambda e^{\lambda b_t} - e^{-\nu}\Lambda' e^{3\lambda b_t + \frac{\lambda^2}{K(p-q)}\mathbb{E}\big[(\frac{e^{Z_1^t+U_1}}{1+e^{Z_1^t+U_1-\nu}})^2\big]} \\
&\overset{(b)}{\geq} \Lambda e^{\lambda b_t} - \Lambda' e^{C b_t - \nu} \\
&= \Lambda e^{\lambda b_t}\big(1 - \frac{\Lambda'}{\Lambda} e^{-\nu+(C-\lambda)b_t}\big) \\
&\overset{(c)}{\geq} \Lambda e^{\lambda b_t}\big(1 - \frac{\Lambda'}{\Lambda} e^{\frac{-\nu}{2}}\big)
\end{aligned} \tag{4.207}$$

where (a) holds from Lemma 44, (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 $b_t \leq \frac{\nu}{2(C-\lambda)}$.

4.4.16 Proof of Lemma 27

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
$$w_o\lambda\Lambda e^{\lambda D}(1-\frac{\Lambda'}{\Lambda}e^{-\nu_o}) - \lambda D \geq w_o. \tag{4.208}$$
- 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,

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

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 25:

$$\begin{aligned} b_{t+1} &\geq \Lambda e^{\lambda b_t}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu}{2}}) \\ &\geq \Lambda e^{\lambda b_t}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu_o}{2}}) \end{aligned} \tag{4.209}$$

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

$$\begin{aligned} b_1 &\geq \Lambda e^{\lambda b_0}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu_o}{2}}) \\ &\geq \Lambda(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu_o}{2}}) \\ &\geq \sqrt{\frac{\Lambda}{\lambda e}} \end{aligned} \tag{4.210}$$

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

$$\begin{aligned} b_{t+1} &\geq \Lambda e^{\lambda b_t}(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu_o}{2}}) \\ &\overset{(a)}{\geq} \Lambda e\lambda b_t(1 - \frac{\Lambda'}{\Lambda}e^{\frac{-\nu_o}{2}}) \\ &\overset{(b)}{\geq} \sqrt{\Lambda\lambda e}\, b_t \end{aligned} \tag{4.211}$$

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:

$$b_t \geq \Lambda(\lambda\Lambda e)^{\frac{t}{2}-1} \text{ for all } 1 \leq t \leq t^* + 1 \tag{4.212}$$

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, $b_{\bar{t}_o} \geq \Lambda(\lambda L e)^{\frac{\bar{t}_o}{2}-1}$. Let $t_o = \min\{t : b_t \geq \Lambda(\lambda\Lambda e)^{\frac{\bar{t}_o}{2}-1}\}$. Thus, by Lemma 26, we get $t_o \leq \bar{t}_o$. Moreover, by the choice of t_o and w_o:

$$w_o \leq \lambda(\Lambda(\lambda\Lambda e)^{\frac{\bar{t}_o}{2}-1} - D) \leq \lambda(b_{t_o} - D) \tag{4.213}$$

Now define sequence $(w_t : t \geq 0)$: $w_{t+1} = e^{w_t}$, where w_o was chosen according to (4.208). 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 25, (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

$$w_t \leq \lambda(b_{t_o+t} - D) \text{ for } 0 \leq t \leq t^* - t_o + 1. \tag{4.214}$$

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 (4.214), for $t = t_o + \bar{t}_1 + 1$:

$$b_{t_o+\bar{t}_1+1} \geq \frac{w_{\bar{t}_1+1}}{\lambda} + D \geq \frac{\nu}{2(C-\lambda)} \tag{4.215}$$

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 25:

$$b_{t^*+2} \geq \Lambda e^{\lambda b_{t^*+1}} (1 - \frac{\Lambda'}{\Lambda} e^{\frac{-\nu}{2}}) \tag{4.216}$$

By Lemma 26, the sequence b_t is non-decreasing in t. We also known $t^* + 2 \leq \bar{t}_o + \bar{t}_1 + 2$. Using (4.216):

$$b_{\bar{t}_o + \log^*(\nu) + 2} \geq \Lambda e^{\frac{\lambda \nu}{2(C-\lambda)}} (1 - \frac{\Lambda'}{\Lambda} e^{\frac{-\nu}{2}}) \tag{4.217}$$

which concludes one case of the proof.

When $b_{t^*+1} > \frac{\nu}{2(C-\lambda)}$, we use the truncation process (Hajek et al., 2018, 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 (Hajek et al., 2018, Lemma 6) depends only on the structure of the tree. In this chapter, 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 (4.216) and (4.217).

4.4.17 Proof of Theorem 5

The assumption $(np)^{\log^*(\nu)} = n^{o(1)}$ ensures that $(np)^{\hat{t}} = n^{o(1)}$. Since $\frac{K^2(p-q)^2}{q(n-K)} \to \lambda$, $p \geq q$ 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 29 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 4.2 whose output is $\hat{C} = \{i : R_i^{\hat{t}} \geq \nu\}$. Then for deterministic $|C^*| = K$, the following holds:

$$\begin{aligned} p_e &= \mathbb{P}(\text{No coupling}) p_{e|\text{no coupling}} + \mathbb{P}(\text{coupling}) p_{e|\text{coupling}} \\ &\leq n^{-1+o(1)} + \frac{K}{n} e^{-\nu(r+o(1))} \end{aligned} \tag{4.218}$$

where the last inequality holds by Lemmas 29 and 28 for some positive constant r. Multiplying (4.218) by $\frac{n}{K}$:

$$\frac{\mathbb{E}[|C^* \triangle \hat{C}|]}{K} \leq \frac{n^{o(1)}}{K} + e^{-\nu(r+o(1))} \to 0 \tag{4.219}$$

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

Now going back to Algorithm 4.2 and its output $\tilde{C}$, using Equation (4.49):

$$\frac{\mathbb{E}[|C^* \triangle \tilde{C}|]}{K} \leq 2\frac{\mathbb{E}[|C^* \triangle \hat{C}|]}{K} \to 0 \tag{4.220}$$

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

When $|C^*|$ is random such that $K \geq 3\log(n)$ and $\mathbb{P}(||C^*| - K| \geq \sqrt{3K\log(n)}) \leq n^{\frac{-1}{2}+o(1)}$, we have $\mathbb{E}[||C^*| - K|] \leq n^{\frac{1}{2}+o(1)}$. Thus, for $\tilde{C}$, using Equation (4.49):

$$\frac{\mathbb{E}[|C^* \triangle \tilde{C}|]}{K} \leq 2\frac{\mathbb{E}[|C^* \triangle \hat{C}|]}{K} + \frac{\mathbb{E}[||C^*| - K|]}{K} \to 0 \tag{4.221}$$

which concludes the proof.

4.4.18 Proof of Theorem 6

Since $(np+2)^{\hat{t}} = n^{o(1)}$, the coupling Lemma 29 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 $\hat{C}$:

$$\begin{aligned} p_e &= \mathbb{P}(\text{No coupling}) p_{e|\text{no coupling}} + \mathbb{P}(\text{coupling}) p_{e|\text{coupling}} \\ &\geq \frac{K(n-K)}{n^2} \mathbb{E}^2[e^{\frac{U_0}{2}}] e^{\frac{-\lambda\Lambda e}{4}} - n^{-1+o(1)} \end{aligned} \tag{4.222}$$

where the last inequality holds by Lemmas 29 and 28. Multiplying (4.222) by $\frac{n}{K}$:

$$\frac{\mathbb{E}[|C^*\triangle\hat{C}|]}{K} \geq \Big(1-\frac{K}{n}\Big)\mathbb{E}^2[e^{\frac{U_0}{2}}]e^{\frac{-\lambda\Lambda e}{4}} - o(1) \tag{4.223}$$

where the last inequality holds because $K = n^{1-o(1)}$. Thus, for $\lambda \leq \frac{1}{\Lambda e}$, $\frac{\mathbb{E}[|C^*\triangle\hat{C}|]}{K}$ 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.

4.5 Proof of Theorem 7

Let Z be a binomial random variable $\text{Bin}(n(1-\delta), \frac{K}{n})$. In view of Lemma 21, it suffices to verify (4.9) 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: $\mathbb{E}[\phi(|C_k^*|)] \leq \mathbb{E}[\phi(Z)]$. Therefore, Chernoff bound for Z also holds for $|C_k^*|$. This leads to:

$$\begin{aligned}\mathbb{P}\Big(\big||C_k^*| - (1-\delta)K\big| \geq \sqrt{3K(1-\delta)\log(n)}\Big) &\leq n^{-1.5+o(1)} \\ &\leq n^{\frac{-1}{2}+o(1)}\end{aligned} \tag{4.224}$$

Thus, by Theorem 5, belief propagation achieves weak recovery for recovering C_k^* for each k. Thus:

$$\mathbb{P}\big(|\hat{C}_k\triangle C_k^*| \leq \delta K \quad \text{for } 1 \leq k \leq \frac{1}{\delta}\big) \to 1 \tag{4.225}$$

which together with Lemma 21 conclude the proof.

4.5.1 Proof of Lemma 30

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

$$M(x) = \frac{\frac{p}{q}-1}{1+e^{-(x-\nu)}} - \frac{1}{2}\big(\frac{\frac{p}{q}-1}{1+e^{-(x-\nu)}}\big)^2 + O\Big(\big(\frac{\frac{p}{q}-1}{1+e^{-(x-\nu)}}\big)^3\Big) \tag{4.226}$$

Thus:

$$\begin{aligned}
\mathbb{E}[Z_0^{t+1}] =& - K(p-q) + Kq\mathbb{E}[M(Z_1^t + U_1)] + (n-K)q\mathbb{E}[M(Z_0^t + U_0)] \\
=& - K(p-q) + K(p-q)\mathbb{E}\Big[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\Big] + (n-K)(p-q)\mathbb{E}\Big[\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\Big] \\
& - \frac{K(p-q)^2}{2q}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big] - \frac{(n-K)(p-q)^2}{2q}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^0+U_0-\nu)}}\big)^2\Big] \\
& + O\Bigg(\frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^3\Big] + \frac{(n-K)(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\big)^3\Big]\Bigg)
\end{aligned} \tag{4.227}$$

Using Lemma 43 for $g(x) = \frac{1}{1+e^{-(x-\nu)}}$,

$$\begin{aligned}
K(p-q)\mathbb{E}\Big[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\Big] + (n-K)(p-q)\mathbb{E}\Big[\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\Big] \\
= K(p-q)
\end{aligned} \tag{4.228}$$

Similarly:

$$\begin{aligned}
\frac{K(p-q)^2}{2q}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big] + \frac{(n-K)(p-q)^2}{2q}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^0+U_0-\nu)}}\big)^2\Big] \\
= \frac{K(p-q)^2}{2q}\mathbb{E}\Big[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\Big]
\end{aligned} \tag{4.229}$$

and,

$$\begin{aligned}
\frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^3\Big] + \frac{(n-K)(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\big)^3\Big] \\
= \frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big]
\end{aligned} \tag{4.230}$$

Using (4.228), (4.229) and (4.230) and substituting in (4.227):

$$\begin{aligned}
\mathbb{E}[Z_0^{t+1}] =& - \frac{\lambda}{2}b_t + O\Bigg(\frac{K(p-q)^3}{q^2}\mathbb{E}[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2]\Bigg) \\
=& - \frac{\lambda}{2}b_t + o(1)
\end{aligned} \tag{4.231}$$

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

$$\begin{aligned}
\mathbb{E}[Z_1^{t+1}] &= \mathbb{E}[Z_0^{t+1}] + K(p-q)\mathbb{E}[M(Z_1^t+U_1)] \\
&= \mathbb{E}[Z_0^{t+1}] + \frac{K(p-q)^2}{q}\mathbb{E}\Big[\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\Big] + O\bigg(\frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big]\bigg) \\
&= \mathbb{E}[Z_0^{t+1}] + \lambda b_t + o(1) = \frac{\lambda}{2}b_t + o(1) \qquad (4.232)
\end{aligned}$$

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 $\mathrm{var}(Y) = \mathbb{E}[L]\mathbb{E}[X_1^2]$. Thus,

$$\begin{aligned}
&\mathrm{var}(Z_0^{t+1}) \\
&= Kq\,\mathbb{E}[M^2(Z_1^t+U_1)] + (n-K)q\,\mathbb{E}[M^2(Z_0^t+U_0)] \\
&\overset{(a)}{=} \frac{K(p-q)^2}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big] + \frac{(n-K)(p-q)^2}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\big)^2\Big] \\
&+ O\bigg(\frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^3\Big] + \frac{(n-K)(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_0^t+U_0-\nu)}}\big)^3\Big]\bigg) \\
&\overset{(b)}{=} \lambda b_t + o(1) \qquad (4.233)
\end{aligned}$$

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 (4.231).

Similarly,

$$\begin{aligned}
\mathrm{var}(Z_1^{t+1}) &= \mathrm{var}(Z_0^{t+1}) + O\bigg(\frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\big(\frac{1}{1+e^{-(Z_1^t+U_1-\nu)}}\big)^2\Big]\bigg) \\
&= \lambda b_t + o(1) \qquad (4.234)
\end{aligned}$$

4.5.2 Proof of Lemma 31

Before we prove the lemma, we need the following lemma from (Korolev and Shevtsova, 2012, Theorem 3).

Lemma 45. *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,*

$$\sup_x \left|\mathbb{P}\Big(\frac{S_\gamma - \gamma\mu}{\sqrt{\gamma(\mu^2+\sigma^2)}} \leq x\Big) - \phi(x)\right| \leq \frac{0.3041\rho^3}{\sqrt{\gamma(\mu^2+\sigma^2)^3}} \tag{4.235}$$

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

$$Z_0^{t+1} = -K(p-q) + \sum_{i=1}^{N_{nq}} X_i \tag{4.236}$$

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 (4.236), using the properties of compound Poisson distribution, and then applying Lemma 30:

$$nq\mathbb{E}[X_i^2] = \mathrm{var}(Z_0^{t+1}) = \lambda b_t + o(1) \tag{4.237}$$

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

$$\begin{aligned} nq\mathbb{E}[|X_i^3|] &\leq \frac{K(p-q)^3}{q^2}\mathbb{E}\Big[\Big(\frac{1}{1+e^{-(Z_1^t+U_1)+\nu}}\Big)^3\Big] + \frac{(n-K)(p-q)^3}{q^2}\mathbb{E}\Big[\Big(\frac{1}{1+e^{-(Z_0^t+U_0)+\nu}}\Big)^3\Big] \\ &\overset{(a)}{\leq} \frac{K(p-q)^3}{q^2} \\ &\overset{(b)}{=} o(1) \end{aligned} \tag{4.238}$$

where (a) holds by Lemma 43 for $g(x) = \frac{1}{1+e^{-(x-\nu)}}$ and (b) holds since $\frac{p}{q} \to 1$.

Combining (4.237) and (4.238) 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 45 yields:

$$\sup_x \left|\mathbb{P}\Big(\frac{Z_0^{t+1} + \frac{\lambda b_t}{2}}{\sqrt{\lambda b_t}} \leq x\Big) - \phi(x)\right| \to 0 \tag{4.239}$$

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

$$Z_1^{t+1} = -K(p-q) + \frac{1}{\sqrt{(n-K)q}} \sum_{i=1}^{N_{(n-K)q+Kp}} Y_i \tag{4.240}$$

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

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

Starting with (4.240), using the properties of compound Poisson distribution, and then applying Lemma 30:

$$((n-K)q+Kp)\mathbb{E}[Y_i^2] = \text{var}(Z_1^{t+1}) = \lambda b_t + o(1) \tag{4.241}$$

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

$$\begin{aligned}((n-K)q+Kp)\mathbb{E}[|Y_i^3|] &= nq\mathbb{E}[|X_i|^3] + K(p-q)\mathbb{E}\Big[\Big(\frac{\frac{p}{q}-1}{1+e^{-(Z_1^t+U_1)+\nu}}\Big)^3\Big] \\ &\leq o(1) \end{aligned} \tag{4.242}$$

where (4.242) holds since $\frac{p}{q} \to 1$.

Combining (4.241) and (4.242) yields $\frac{\mathbb{E}[|Y_i^3|]}{\sqrt{(n-K)q+Kp)\mathbb{E}^3[Y_i^2]}} \to 0$, which together with Lemma 45 yields:

$$\sup_x \big|\mathbb{P}\big(\frac{Z_1^{t+1} - \frac{\lambda b_t}{2}}{\sqrt{\lambda b_t}} \leq x\big) - \phi(x)\big| \to 0 \tag{4.243}$$

Hence, using (4.239) and (4.243), it suffices to show that $\lambda b_t \to v_{t+1}$, which implies that (4.57) and (4.58) 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,

$$b_{t+1} = \mathbb{E}[\frac{1}{e^{-\nu}+e^{-(Z_1^{t+1}+U_1)}}] = \mathbb{E}_{U_1}[\mathbb{E}_{Z_1}[\frac{1}{e^{-\nu}+e^{-(Z_1^t+u)}}]]$$

$$= \mathbb{E}_{U_1}[\mathbb{E}_{Z_1}[f(Z_1^{t+1}; u, \nu)]] = \mathbb{E}_{U_1}[\mathbb{E}_{Z_1}[\mathcal{E}_n]] \tag{4.244}$$

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 (4.243), 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:

$$\sup_{0<c<e^{\nu}} \left| F_{\mathcal{E}_n}(c) - F_{G(\lambda b_t)}\Big(\log\big(\frac{ce^{-u}}{1-ce^{-\nu}}\big)\Big) \right| \to 0 \tag{4.245}$$

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

$$\sup_{0<c<e^{\nu}} \left| F_{\mathcal{E}_n}(c) - F_{G(v_{t+1})}\Big(\log\big(\frac{ce^{-u}}{1-ce^{-\nu}}\big)\Big) \right| \to 0 \tag{4.246}$$

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:

$$\mathcal{E}_n \overset{i.d.}{\to} \frac{1}{e^{-\nu} + e^{-(G(v_{t+1})+u)}} \tag{4.247}$$

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

$$\mathbb{E}[\mathcal{E}_n^2] \overset{(a)}{\leq} e^{2\nu} \overset{(b)}{\leq} A \tag{4.248}$$

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 (4.246), (4.247) and (4.248), 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:

$$\begin{aligned} b_{t+1} &= \mathbb{E}_{U_1}\left[\mathbb{E}[\mathcal{E}_n]\right] \\ &\overset{(a)}{\to} \mathbb{E}_{U_1}\left[\mathbb{E}_Z\left[\frac{1}{e^{-\nu}+e^{-(\frac{v_{t+1}}{2}+\sqrt{v_{t+1}}Z)-u}}\right]\right] \\ &= \frac{v_{t+2}}{\lambda} \end{aligned} \tag{4.249}$$

where in (a) we define $Z \sim \mathcal{N}(0,1)$. Equation (4.249) implies that $\lambda b_{t+1} \to v_{t+2}$, which concludes the proof of the lemma.

4.5.3 Proof of Lemma 33

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

$$\begin{aligned} v_{t+1} &= \lambda\, \mathbb{E}_{Z,U_1}\left[\frac{1}{e^{-\nu}+e^{-(\frac{v_t}{2}+\sqrt{v_t}Z)-U_1}}\right] \\ &= \lambda\sum_{\ell=1}^{L}\frac{\alpha_{+,\ell}^2}{\alpha_{-,\ell}}\,\mathbb{E}_Z\left[\frac{1}{e^{-\nu(1-\frac{h_\ell}{\nu})}+e^{-(\frac{v_t}{2}+\sqrt{v_t}Z)}}\right] \\ &\overset{(a)}{=} \lambda\sum_{\ell=1}^{L}\frac{\alpha_{+,\ell}^2}{\alpha_{-,\ell}}\,\mathbb{E}_Z\left[\frac{1}{e^{-C_l\nu}+e^{-(\frac{v_t}{2}+\sqrt{v_t}Z)}}\right] \\ &\overset{(b)}{=} \lambda\Lambda e^{v_t}(1+o(1)) \end{aligned} \tag{4.250}$$

where (a) holds for positive constants C_ℓ, $\ell \in \{1,\cdots,L\}$ 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 $\lambda > 0$:

$$v_t \geq w_t \tag{4.251}$$

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

$$v_{t^*+2} \geq w_{t^*+2} \tag{4.252}$$

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 (4.250) and (4.252), we get for all sufficiently large κ and after $t^* + 2$ iterations of belief propagation (or for a tree of depth $t^* + 2$):

$$\mathbb{E}_{U_0}\Big[Q(\frac{\nu + \frac{v_{t^*+2}}{2} - U_0}{\sqrt{v_{t^*+2}}})\Big] = Q\Big(\frac{1}{2}\sqrt{v_{t^*+2}}(1 + o(1))\Big) \tag{4.253}$$

$$\mathbb{E}_{U_1}\Big[Q(\frac{-\nu + \frac{v_{t^*+2}}{2} + U_1}{\sqrt{v_{t^*+2}}})\Big] = Q\Big(\frac{1}{2}\sqrt{v_{t^*+2}}(1 + o(1))\Big) \tag{4.254}$$

Since $Q(x) \leq e^{-\frac{1}{2}x^2}$ for $x \geq 0$, then using (4.252), (4.253) and (4.254):

$$\frac{n-K}{K}Q\Big(\frac{1}{2}\sqrt{v_{t^*+2}}(1 + o(1))\Big) \to 0 \tag{4.255}$$

$$Q\Big(\frac{1}{2}\sqrt{v_{t^*+2}}(1 + o(1))\Big) \to 0 \tag{4.256}$$

Using (4.255) and (4.256) and Lemma 32, we get:

$$\lim_{\frac{n}{K}\to\infty} \lim_{nq,Kq\to\infty} \lim_{n\to\infty} \frac{\mathbb{E}[\hat{C}\triangle C^*]}{K} = 0 \tag{4.257}$$

CHAPTER 5

EXIT ANALYSIS FOR COMMUNITY DETECTION

The technical distinction and novelty of this chapter can be explained as follows: EXIT analysis was originally developed in the context of communication systems for bipartite graphs in which some nodes carry information while some other nodes represent the constraints on the data nodes (e.g. via parity check equations or the structure or memory of a communication channel). The present work aims to employ EXIT analysis in a scenario where the above conditions do not apply, and therefore the EXIT analysis must be developed anew for the scenario where each node in a general tree has both an individual label (information) as well as information that is applicable to other nodes. This gives rise to new EXIT equations. In other words, in the original EXIT analysis, all mutual information was calculated with respect to a subset of node labels, i.e., bit-node variables, whereas now all nodes have information. Since we are now interested in a graph that has a stochastic symmetry, the input/output belief propagation equations must be reinterpreted once again in terms of extrinsic information. This statement will be further clarified in the sequel while developing the details of EXIT equations.

5.1 System Models

Throughout this chapter, the community label of node i is denoted by x_i, the side information of node i by y_i, the vector of the nodes true labels by $\boldsymbol{x}^*$, the vector of the nodes side information by $\boldsymbol{y}$, and the observed graph by G. We assume that conditioned on $\boldsymbol{x}$, G and $\boldsymbol{y}$ are independent. The goal is to recover $\boldsymbol{x}^*$ from the observation of $\boldsymbol{G}$ and $\boldsymbol{y}$. The alphabet for y_i is denoted with $\{u_1, u_2, \cdots, u_M\}$, where M is the cardinality of side information which is assumed to be bounded and constant across n.

Two system models are considered. The first, the binary symmetric stochastic block model which consists of n nodes with $x_i \in \{\pm 1\}$. The node labels are independent and identically distributed across n, with 1 and -1 labels having equal probability. Each two nodes are connected with an edge with probability $\frac{a}{n}$ if the two nodes belong to the same community and with probability $\frac{b}{n}$, otherwise, for $a > b > 0$. In addition to the graph, each node independently observes side information, y_i, according to:

$$\alpha_{+,m} \triangleq \mathbb{P}(y_i = u_m | x_i = 1) \tag{5.1}$$

$$\alpha_{-,m} \triangleq \mathbb{P}(y_i = u_m | x_i = -1) \tag{5.2}$$

It is further assumed that as $n \to \infty$: $a, b \to \infty$ such that $\frac{a-b}{\sqrt{b}} = \mu$, for a fixed positive constant μ and that the average degree $\frac{(a+b)}{2} = n^{o(1)}$. The latter condition is crucial in our analysis, by enabling the approximation of the neighborhood of a given node in the graph by a tree (Mossel and Xu, 2016a,b).

The second model studied is the one-community stochastic block model, consisting of n nodes and containing a hidden community C^* with size $|C^*| = K$. Let $x_i = 1$ if $i \in C^*$ and $x_i = 0$ if $i \notin C^*$. 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, with $p \geq q$. For each node i, side information y_i is observed according to the distribution:

$$\alpha_{+,m} \triangleq \mathbb{P}(y_i = u_m | x_i = 1) \tag{5.3}$$

$$\alpha_{-,m} \triangleq \mathbb{P}(y_i = u_m | x_i = 0) \tag{5.4}$$

Define

$$\lambda \triangleq \frac{K^2(p-q)^2}{(n-K)q}. \tag{5.5}$$

We assume $\frac{K}{n}$, the LLR of side information and λ are constants independent of n, while $nq, Kq \xrightarrow{n\to\infty} \infty$, which implies that $\frac{p}{q} \xrightarrow{n\to\infty} 1$. Furthermore, $np = n^{o(1)}$.

5.2 Binary Symmetric Stochastic Block Model

Studying the performance of belief propagation with noisy-label side information was introduced in (Mossel and Xu, 2016b). This section generalizes the results to M-ary side information and introduces EXIT analysis as a new tool to study the performance of belief propagation for community detection. A key idea in our analysis is the relation between inference on graphs and inference on the corresponding Galton-Watson trees (Mossel and Xu, 2016b).

Definition 1. *For a node i, let $(T_i, \tau, \tilde{\tau})$ be a Poisson two-type branching process tree rooted at i, where τ is a ± 1 labeling of nodes in T_i. Let τ_i be chosen uniformly at random from $\{\pm 1\}$. Each node j in T_i will have $L_j \sim Pois(\frac{a}{2})$ children with label τ_j and $M_j \sim Pois(\frac{b}{2})$ children with label $-\tau_j$. Finally, for each node j, an M-ary side information $\tilde{\tau}_j$ is observed according to the conditional distributions $\alpha_{+,m}$ and $\alpha_{-,m}$.*

Let T_j^t be the sub-tree of T_i rooted at node j with depth t. The problem of inference on trees with side information is to estimate the label of the root τ_i given observation of $(T_i^t, \tilde{\tau}_{T_i^t})$, where $\tilde{\tau}_{T_i^t}$ is the side information of all the nodes in the tree rooted at i with depth t. It then follows that the error probability for an estimator $\hat{\tau}_i(T_i^t, \tilde{\tau}_{T_i^t})$ is:

$$\begin{aligned} q_{T^t} &= \frac{1}{2}\mathbb{P}(\hat{\tau}_i \\ &= 1|\tau_i = -1) + \frac{1}{2}\mathbb{P}(\hat{\tau}_i = -1|\tau_i = 1). \end{aligned}$$

Let $q^*_{T^t}$ be the error probability achieved by the optimal estimator, i.e. maximum a posteriori (MAP). Note that the MAP estimator for any node i can be written as: $\hat{\tau}_{MAP} = 2 \times 1_{\{\Gamma_i^t \geq 0\}} - 1$, where Γ_i^t is the log likelihood ratio and can be defined as:

$$\Gamma_j^t = \frac{1}{2}\log\left(\frac{\mathbb{P}(T_j^t, \tilde{\tau}_{T_j^t}|\tau_j = 1)}{\mathbb{P}(T_j^t, \tilde{\tau}_{T_j^t}|\tau_j = -1)}\right) \tag{5.6}$$

$\forall j \in T_i$. The log likelihood ratio Γ_j^t can be further computed via a recursive formula which is the basis for the belief propagation algorithm.

Lemma 1. *Let $\mathcal{N}_j$ denote the children of node j, $N_j \triangleq |\mathcal{N}_j|$, $\beta = \frac{1}{2}\log(\frac{a}{b})$ and $h_j \triangleq \frac{1}{2}\log\left(\frac{\mathbb{P}(\tilde{\tau}_j|\tau_j=1)}{\mathbb{P}(\tilde{\tau}_j|\tau_j=-1)}\right)$. Then, for all $t \geq 1$,*

$$\Gamma_j^t = h_j + \frac{1}{2}\sum_{k \in \mathcal{N}_j}\log\left(\frac{1 + e^{2\beta + 2\Gamma_k^{t-1}}}{e^{2\beta} + e^{2\Gamma_k^{t-1}}}\right) \tag{5.7}$$

Proof.

$$
\begin{aligned}
\Gamma_j^t =& \frac{1}{2}\log\left(\frac{\mathbb{P}(T_j^t, \tilde{\tau}_{T_j^t}|\tau_j = 1)}{\mathbb{P}(T_j^t, \tilde{\tau}_{T_j^t}|\tau_j = -1)}\right) \\
\overset{(a)}{=}& \frac{1}{2}\log\left(\frac{\mathbb{P}\big(N_j, \tilde{\tau}_j|\tau_j = 1\big)}{\mathbb{P}\big(N_j, \tilde{\tau}_j|\tau_j = -1\big)}\right) + \log\left(\frac{\prod_{k\in\mathcal{N}_j}\mathbb{P}\big(T_k^{t-1}, \tilde{\tau}_{T_k^{t-1}}|\tau_j = 1\big)}{\prod_{k\in\mathcal{N}_j}\mathbb{P}\big(T_k^{t-1}, \tilde{\tau}_{T_k^{t-1}}|\tau_j = -1\big)}\right) \\
\overset{(b)}{=}& \frac{1}{2}\log\left(\frac{\mathbb{P}\big(N_j|\tau_j = 1\big)}{\mathbb{P}\big(N_j|\tau_j = -1\big)}\right) + \frac{1}{2}\log\left(\frac{\mathbb{P}\big(\tilde{\tau}_j|\tau_j = 1\big)}{\mathbb{P}\big(\tilde{\tau}_j|\tau_j = -1\big)}\right) + \\
& \frac{1}{2}\sum_{k\in\mathcal{N}_j}\log\left(\frac{\sum_{\tau_k\in\{\pm1\}}\mathbb{P}\big(T_k^{t-1}, \tilde{\tau}_{T_k^{t-1}}|\tau_k\big)\mathbb{P}\big(\tau_k|\tau_j = 1\big)}{\sum_{\tau_k\in\{\pm1\}}\mathbb{P}\big(T_k^{t-1}, \tilde{\tau}_{T_k^{t-1}}|\tau_k\big)\mathbb{P}\big(\tau_k|\tau_j = -1\big)}\right) \\
\overset{(c)}{=}& h_j + \frac{1}{2}\sum_{k\in\mathcal{N}_j}\log\left(\frac{a\mathbb{P}\big(T_k^{t-1}, \tilde{\tau}_{T_k^{t-1}}|\tau_k = 1\big) + b\mathbb{P}\big(T_k^{t-1}, \tilde{\tau}_{T_k^{t-1}}|\tau_k = -1\big)}{b\mathbb{P}\big(T_k^{t-1}, \tilde{\tau}_{T_k^{t-1}}|\tau_k = 1\big) + a\mathbb{P}\big(T_k^{t-1}, \tilde{\tau}_{T_k^{t-1}}|\tau_k = -1\big)}\right) \\
\overset{(d)}{=}& h_j + \frac{1}{2}\sum_{k\in\mathcal{N}_j}\log\left(\frac{1 + e^{2\beta + 2\Gamma_k^{t-1}}}{e^{2\beta} + e^{2\Gamma_k^{t-1}}}\right)
\end{aligned}
$$

- (a) holds because conditioned on τ_j, $(N_j, \tilde{\tau}_j)$ are independent of the rest of the tree, and $(T_k^{t-1}, \tilde{\tau}_{T_k^{t-1}})$ are independent and identically distributed random variables $\forall k \in N_j$,
- (b) holds also because conditioned on τ_j, N_j and $\tilde{\tau}_j$ are independent,
- (c) holds because $N_j \sim \text{Pois}(\frac{a+b}{2})$ $\forall j \in T^t$, and for a node j, N_j children are generated $\sim \text{Pois}(\frac{a+b}{2})$, then for each node $k \in N_j$, $\tau_k = \tau_j$ with probability $\frac{a}{a+b}$ and $\tau_k = -\tau_j$ with probability $\frac{b}{a+b}$,
- (d) holds from the definition of β.

□

The above result clarifies the connection between inference on trees and the community detection problem addressed in this chapter. Let G_i^t be the sub-graph of G induced by the nodes whose distance to i is at most t, and $\boldsymbol{x}_A$ be a vector consisting of labels of nodes in a set of nodes A. Then, the following Lemma, proved in (Mossel and Xu, 2016b), shows the

Table 5.1. Belief propagation algorithm with side information.

Belief Propagation Algorithm
1: Input: $n, t \in \mathbb{N}$, G, $\boldsymbol{y}$.
2: Initialize: Set $R^0_{i\to j} = 0$, $\forall i \in G$ and $j \in \mathcal{N}_i$.
3: For all $i \in G$ and $j \in \mathcal{N}_i$, run for $t-1$ iterations: $R^{t-1}_{i\to j} = h_i + \sum_{k\in\mathcal{N}_i\setminus\{j\}} \log\left(\frac{1+e^{2\beta+2R^{t-2}_{k\to i}}}{e^{2\beta}+e^{2R^{t-2}_{k\to i}}}\right)$
4: For all $i \in G$, compute: $R^t_i = h_i + \sum_{k\in\mathcal{N}_i} \log\left(\frac{1+e^{2\beta+2R^{t-1}_{k\to i}}}{e^{2\beta}+e^{2R^{t-1}_{k\to i}}}\right)$
5: Return $\hat{\boldsymbol{x}}_{BP^t}$ with $\hat{\boldsymbol{x}}_{BP^t}(i) = 2 \times 1_{\{R^t_i \geq 0\}} - 1$.

feasibility of approximating $(G^t_i, \boldsymbol{x}_{G^t_i}, \boldsymbol{y}_{G^t_i})$ by $(T^t_i, \tau_{T^t_i}, \tilde{\tau}_{T^t_i})$ with probability approaching one under certain conditions on the depth t.

Lemma 2 ((Mossel and Xu, 2016b))**.** *For $t = t(n)$ such that $(\frac{a+b}{2})^t = n^{o(1)}$, there exists a coupling between $(G, \boldsymbol{x}, \boldsymbol{y})$ and $(T, \tau, \tilde{\tau})$ such that $(G^t_i, \boldsymbol{x}_{G^t_i}, \boldsymbol{y}_{G^t_i}) = (T^t_i, \tau_{T^t_i}, \tilde{\tau}_{T^t_i})$ with probability converging to* 1*.*

Lemma 2 suggests that the tree-based log likelihood ratio Γ^t_i, calculated in Lemma 1, is an asymptotically accurate representation for belief propagation in our problem. Let $\hat{\boldsymbol{x}}_{BP^t}$ be the output of the belief propagation algorithm after t iterations. The details of the belief propagation algorithm is presented in Table 5.1.

Define

$$p_{G,\boldsymbol{y}}(\hat{\boldsymbol{x}}) \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}\{x_i \neq \hat{x}_i\}$$

to be the expected fraction of misclassified nodes by an estimator $\hat{\boldsymbol{x}}$. The following lemma characterizes the asymptotic average behavior of grah-wide error as characterized by $p_{G,\boldsymbol{y}}(\hat{\boldsymbol{x}}_{BP^t})$.

Lemma 3. *For $t = t(n)$ such that $(\frac{a+b}{2})^t = n^{o(1)}$, $\lim_{n\to\infty} |p_{G,\boldsymbol{y}}(\hat{\boldsymbol{x}}_{BP^t}) - q^*_{T^t}| = 0$.*

Proof. By Lemma 2, $(G_i^t, \boldsymbol{x}_{G_i^t}, \boldsymbol{y}_{G_i^t}) = (T_i^t, \tau_{T_i^t}, \tilde{\tau}_{T_i^t})$ with probability converging to 1. This implies that $R_i^t = \Gamma_i^t$, and hence, $p_{G,\boldsymbol{y}}(\hat{\boldsymbol{x}}_{BP^t}) = q^*_{T^t} + o(1)$, where the $o(1)$ term comes from the coupling error of Lemma 2. $\square$

So far the results hold for all a and b as long as $\frac{(a+b)}{2} = n^{o(1)}$ and $\frac{a}{b} = \Theta(1)$. Now let $a = b + \mu\sqrt{b}$, for a fixed positive constant μ. Let U_+ and U_- be two random variables drawn according to the distribution of h_i conditioned respectively on $\tau_i = 1$ and $\tau_i = -1$. Then the following theorem describes a density evolution that evaluates $q^*_{T^t}$.

Theorem 8. *Assume as $n \to \infty$, $b \to \infty$ and $\frac{a-b}{\sqrt{b}} \to \mu$, for a fixed positive constant μ. Also, let $h(\nu) = \mathbb{E}[\tanh(\nu + \sqrt{\nu}Z + U_+)]$, where $Z \sim \mathcal{N}(0,1)$. Define $\bar{\nu}$ to be the smallest fixed point of $\nu = \frac{\mu^2}{4}h(\nu)$. Then:*

$$\lim_{t\to\infty}\lim_{n\to\infty} p_{G,\boldsymbol{y}}(\hat{\boldsymbol{x}}_{BP^t}) = \frac{1}{2}\Big(\mathbb{E}_{U_+}\big[Q\big(\frac{\bar{\nu}+U_+}{\sqrt{\bar{\nu}}}\big)\big] + \mathbb{E}_{U_-}\big[Q\big(\frac{\bar{\nu}-U_-}{\sqrt{\bar{\nu}}}\big)\big]\Big) \tag{5.8}$$

Proof. The proof has similarities with (Mossel and Xu, 2016b). For brevity, we only describe the new developments compared with (Mossel and Xu, 2016b) and the corresponding arguments.

Define

$$F(x) \triangleq \frac{1}{2}\log\big(\frac{e^{2x+2\beta}+1}{e^{2x}+e^{2\beta}}\big) \tag{5.9}$$

and for all $t \geq 1$, $\Phi_j^t = \sum_{k\in N_j} F(\Phi_k^{t-1} + h_k)$. Thus, for all $t \geq 0$,

$$\Gamma_j^t = h_j + \Phi_j^t. \tag{5.10}$$

We are interested in the moments of Φ_j^t conditioned on node label $\tau_j = -1$ and $\tau_j = 1$. For convenience of notation, we define new random variables W_+^t and W_-^t whose distribution is identical to Φ_j^t when τ_j is equal to 1 and -1, respectively.

Lemma 4. *For all $t \geq 0$,*

$$\mathbb{E}[W_{\pm}^{t+1}] = \pm\frac{\mu^2}{4}\mathbb{E}[\tanh(W_+^t + U_+)] + O(a^{-\frac{1}{2}})$$
$$var(W_{\pm}^{t+1}) = \frac{\mu^2}{4}\mathbb{E}[\tanh(W_+^t + U_+)] + O(a^{-\frac{1}{2}})$$

Proof. The proof for $\mathbb{E}[W_-^{t+1}]$ and $\text{var}(W_-^{t+1})$ departs from (Mossel and Xu, 2016b, Lemma 7.1) in the distribution of $U_\pm$.

Define $\psi(x) = \log(1+x) - x$. It then follows from Taylor expansion that $|\psi(x)| \leq x^2$. Then, $F(x)$, defined in (5.9), can be written as:

$$\begin{aligned} F(x) &= -\beta + \frac{1}{2}\log\big(1 + \frac{e^{4\beta}-1}{e^{-2(x-\beta)}+1}\big) \\ &= -\beta + \frac{e^{4\beta}-1}{2}f(x) + \frac{1}{2}\psi\big((e^{4\beta}-1)f(x)\big) \end{aligned}$$

where $f(x) = \frac{1}{1+e^{-2(x-\beta)}}$. It then follows that:

$$\begin{aligned} \Phi_j^{t+1} &= \sum_{k\in N_j} F(\Phi_k^t + h_k) \\ &= \sum_{k\in N_j}\Big[-\beta + \frac{e^{4\beta}-1}{2}f(\Phi_k^t + h_k) + \frac{1}{2}\psi\big((e^{4\beta}-1)f(\Phi_k^t + h_k)\big)\Big] \end{aligned} \tag{5.11}$$

Calculating the mean of the two sides of equation above conditioned on $\tau_j = \pm 1$,

$$\begin{aligned} \mathbb{E}[W_+^{t+1}] - \mathbb{E}[W_-^{t+1}] &= (\frac{e^{4\beta}-1}{4})(a-b)\mathbb{E}\Big[f(W_+^t + U_+) - f(W_-^t + U_-)\Big] \\ &+ \frac{a-b}{4}\mathbb{E}\Big[\psi\big((e^{4\beta}-1)f(W_+^t + U_+)\big) - \psi\big((e^{4\beta}-1)f(W_-^t + U_-)\big)\Big] \end{aligned} \tag{5.12}$$

By the definition of Γ_j^t and a change of measure, it follows that $\mathbb{E}[g(\Gamma_j^t)|\tau_j = -1] = \mathbb{E}[g(\Gamma_j^t)e^{-2\Gamma_j^t}|\tau_j = 1]$ for any measurable function g such that the expectations are well defined. Also, notice that:

$$(\frac{e^{4\beta}-1}{4})(a-b) = \frac{\mu^2}{2}\frac{a+b}{2b} = \frac{\mu^2}{2}(1 + \frac{a-b}{2b}) \tag{5.13}$$
$$= \frac{\mu^2}{2} + O(a^{-\frac{1}{2}}) \tag{5.14}$$

Moreover, since $|\psi(x)| \le x^2$ and $|f(x)| \le 1$, it follows that $\psi\big((e^{4\beta}-1)f(W_+^t+U_+)\big) - \psi\big((e^{4\beta}-1)f(W_-^t+U_-)\big) \le 2(e^{4\beta}-1)^2$. Therefore,

$$
\begin{aligned}
\frac{a-b}{4}\mathbb{E}\Big[\psi\big((e^{4\beta}-1)f(W_+^t+U_+)\big) - \psi\big((e^{4\beta}-1)f(W_-^t+U_-)\big)\Big] &\le \frac{a-b}{2}(e^{4\beta}-1)^2 \\
&= O(a^{\frac{-1}{2}}) \qquad (5.15)
\end{aligned}
$$

Combining (5.12), (5.14), and (5.15),

$$
\begin{aligned}
\mathbb{E}[W_+^{t+1}] &= \mathbb{E}[W_-^{t+1}] + \big(\frac{\mu^2}{2} + O(a^{\frac{-1}{2}})\big)\mathbb{E}\big[f(W_+^t+U_+)(1-e^{-2(W_+^t+U_+)})\big] + O(a^{\frac{-1}{2}}) \\
&\overset{(a)}{=} \frac{\mu^2}{4}\mathbb{E}[\tanh(W_+^t+U_+)] - O(a^{\frac{-1}{2}})\mathbb{E}[e^{-2(W_+^t+U_+)}] + O(a^{\frac{-1}{2}}) \\
&\overset{(b)}{=} \frac{\mu^2}{4}\mathbb{E}[\tanh(W_+^t+U_+)] + O(a^{\frac{-1}{2}}) \qquad (5.16)
\end{aligned}
$$

where (a) holds from the definition of $f(x)$, the definition of $\tanh(x)$ and the fact that $f(x) = \frac{1}{1+e^{-2x}} + O(a^{\frac{-1}{2}})$ and (b) holds because by change of measure $\mathbb{E}[e^{-2(W_+^t+U_+)}] = \mathbb{E}[e^{-2(W_-^t+U_-)}e^{2(W_-^t+U_-)}] = 1$. This concludes the proof for $\mathbb{E}[W_+^{t+1}]$. The proof for $\mathrm{var}(W_+^{t+1})$ follows similarly. $\square$

Lemma 5. *Assume $\alpha_{-,m}, \alpha_{+,m}$ are constants as $n \to \infty$. Let $h(\nu) = \mathbb{E}[\tanh(\nu+\sqrt{\nu}Z+U_+)]$, where $Z \sim \mathcal{N}(0,1)$. Define $(\nu_t : t \ge 0)$ recursively by $\nu_0 = 0$ and $\nu_{t+1} = \frac{\mu^2}{4}h(\nu_t)$. Then, for any fixed $t \ge 0$, as $n \to \infty$:*

$$
\sup_x \left| \mathbb{P}\left\{ \frac{W_\pm^t \mp \nu_t}{\sqrt{\nu^t}} \le x \right\} - \mathbb{P}\{Z \le x\} \right| = O(a^{-\frac{1}{2}}) \qquad (5.17)
$$

The proof of Lemma 5 departs from (Mossel and Xu, 2016b, Lemma 7.3) only in the distribution of $U_\pm$, and is therefore omitted for brevity.

In view of Lemmas 4, 5, for all j, $(\Phi_j^t|\tau_j = \pm 1) \sim \mathcal{N}(\pm\nu_t, \nu_t)$. Hence,

$$
\lim_{n\to\infty} \mathbb{P}(\Gamma_j^t > 0|\tau_j = -1) = \mathbb{E}_{U_-}\big[Q\big(\frac{\bar{\nu} - U_-}{\sqrt{\bar{\nu}}}\big)\big]
$$

$$
\lim_{n\to\infty} \mathbb{P}(\Gamma_j^t < 0|\tau_j = 1) = \mathbb{E}_{U_+}\big[Q\big(\frac{\nu_t + U_+}{\sqrt{\nu_t}}\big)\big]
$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy$. Hence, from Lemma 3,

$$\lim_{n\to\infty} p_{G,\boldsymbol{y}}(\hat{\boldsymbol{x}}_{BP^t}) = \lim_{n\to\infty} q^*_{T^t} = \frac{1}{2}\Big(\mathbb{E}_{U_+}\big[Q(\frac{\nu_t + U_+}{\sqrt{\nu_t}})\big] + \mathbb{E}_{U_-}\big[Q(\frac{\nu_t - U_-}{\sqrt{\nu_t}})\big]\Big)$$

It remains to show that $\lim_{t\to\infty} \nu_t = \bar{\nu}$.

Lemma 6. *Let* $h(\nu) = \mathbb{E}[\tanh(\nu + \sqrt{\nu}Z + U_+)]$*, where* $Z \sim \mathcal{N}(0,1)$*. Then,* $h(\nu)$ *is continuous on* $[0,\infty]$ *and* $h'(\nu) \geq 0$ *for* $\nu \in (0,\infty)$*.*

The proof of Lemma 6 departs from (Mossel and Xu, 2016b, Lemma 7.4) only in the distribution of $U_\pm$, and is therefore omitted for brevity.

Recall that $\nu_0 = 0$. By direct substitution $\nu_0 \leq \nu_1$. Now, let $\nu_{t+1} \geq \nu_t$. By Lemma 6,

$$\nu_{t+2} - \nu_{t+1} = \frac{\mu^2}{4}(h(\nu_{t+1}) - h(\nu_t)) = \frac{\mu^2}{4} h'(x) \tag{5.18}$$

for some $x \in (\nu_t, \nu_{t+1})$. By Lemma 6, $h'(x) \geq 0$ for $x \in (0,\infty)$. Thus, $\nu_{t+2} \geq \nu_{t+1}$, and hence, it has been shown by induction on t that ν_t is non-decreasing in t. Also, note that $\nu_0 = 0 \leq \bar{\nu}$. If we assume that $\nu_t \leq \bar{\nu}$, then by monotonicity of h, we have: $\nu_{t+1} = \frac{\mu^2}{4} h(\nu_t) \leq \frac{\mu^2}{4} h(\bar{\nu}) = \bar{\nu}$. Thus, $\lim_{t\to\infty} \nu_t = \bar{\nu}$.

□

5.2.1 Exit Analysis

Equation (5.8) characterizes the asymptotic residual error of belief propagation for recovering binary symmetric communities with side information. However, we seek answers to some natural and interesting questions that are not directly apparent by inspection from (5.8), such as: What is the effect of quality and quantity of side information on the residual error? How is this related to the amount of information provided by the graph about node labels? Can side information dominate the performance of belief propagation for community detection, and if so, under what conditions does that happen? In this section, we show that

EXIT charts can provide answers to these questions, via existence and location of crossing points of EXIT curves.

We begin by calculating the mutual information between the label of node i, x_i, and its belief at time t, namely R_i^t.

$$\begin{aligned}
&I(x_i, R_i^t) \\
&\qquad = 1 - H(x_i|R_i^t) \\
&\qquad = 1 - \frac{1}{2}\int_{-\infty}^{\infty}\left(\sum_{m=1}^{M}\alpha_{+,m}\frac{e^{\frac{-(y-(v_t+h_m))^2}{2v_t}}}{\sqrt{2\pi v_t}}\right)\log_2\left(1+\frac{\sum_{m=1}^{M}\alpha_{-,m}e^{\frac{-(y-(-v_t+h_m))^2}{2v_t}}}{\sum_{m=1}^{M}\alpha_{+,m}e^{\frac{-(y-(v_t+h_m))^2}{2v_t}}}\right)dy \\
&\qquad\quad - \frac{1}{2}\int_{-\infty}^{\infty}\left(\sum_{m=1}^{M}\alpha_{-,m}\frac{e^{\frac{-(y-(-v_t+h_m))^2}{2v_t}}}{\sqrt{2\pi v_t}}\right)\log_2\left(1+\frac{\sum_{m=1}^{M}\alpha_{+,m}e^{\frac{-(y-(v_t+h_m))^2}{2v_t}}}{\sum_{m=1}^{M}\alpha_{-,m}e^{\frac{-(y-(-v_t+h_m))^2}{2v_t}}}\right)dy
\end{aligned} \tag{5.19}$$

For simplicity and to show the power of EXIT analysis in drawing insights that cannot be easily deduced from belief propagation equations, we consider a concrete example with $M = 3$. More precisely, for each node i, we observe $y_i = x_i$ with probability $\epsilon(1-\alpha)$ or $y_i = -x_i$ with probability $\epsilon\alpha$ or $y_i = 0$ with probability $1-\epsilon$, independently at random, for $\alpha \in (0, 0.5)$ and $\epsilon \in [0, 1]$. Thus, $U_+ = -U_-$, where $U_+ \in \{\gamma, -\gamma, 0\}$ with probabilities $\epsilon(1-\alpha)$, $\epsilon\alpha$ and $1-\epsilon$, respectively, where $\gamma \triangleq \frac{1}{2}\log\left(\frac{1-\alpha}{\alpha}\right)$. Note that for fixed α and ϵ, $I(x_i, R_i^t)$ is function of ν_t only. Hence, we will denote it by $J(\nu_t)$.

Based on the belief propagation algorithm described in Table 5.1, at iteration t, node i receives the beliefs of all nodes $j \in N(i)$ calculated at iteration $(t-1)$. We denote the information node i receives from node j as I_{in}. Then, node i computes the new information it has at iteration t. We denote this information as I_{out}. Both I_{in} and I_{out} can be calculated using (5.19) as $J(\nu_{t-1})$ and $J(\nu_t)$, respectively. Since $J(\nu_t)$ is monotonically increasing in ν_t (Ten Brink, 2001), $J(\nu_t)$ is reversible. Thus, $\nu_t = J^{-1}(I(x_i, R_i^t))$. Moreover, ν_{t-1} and ν_t are related by

$$\nu_{t+1} = \frac{\mu^2}{4}h(\nu_t)$$

therefore I_{in} and I_{out} for node i are related as follows

$$I_{out} =$$
$$J\Bigg(\frac{\mu^2}{4}\Bigg[\epsilon(1-\alpha)\mathbb{E}_Z[\tanh(J^{-1}(I_{in}) + \sqrt{J^{-1}(I_{in})}Z + \gamma)] +$$
$$\epsilon\alpha\mathbb{E}_Z[\tanh(J^{-1}(I_{in}) + \sqrt{J^{-1}(I_{in})}Z - \gamma)] + (1-\epsilon)\mathbb{E}_Z[\tanh(J^{-1}(I_{in}) + \sqrt{J^{-1}(I_{in})}Z)]\Bigg]\Bigg) \tag{5.20}$$

There is a fundamental difference between using EXIT charts in the context of community detection in stochastic block models and EXIT charts in the standard context of coding theory. Taking Low Density Parity Check (LDPC) Codes as an example, each variable node i receives from a check node j the information or belief of that check node about whether the variable node is one or zero. Thus, the input log-likelihood ratio received by variable node i is actually calculated conditioned on the value of the variable node i. Community detection presents a different scenario: Each node i receives the belief of node j. However, the belief of node j is calculated conditioned on the value of node j, not the value of node i. This reflects the fundamental differences between the bipartite graph representing FEC codewords and a random graph representing relationships of randomly distributed node labels. The former is fundamentally asymmetric, where parity nodes carry no new information conditioned on bit nodes. On the contrary, in community detection, nodes are (stochastically) symmetric and all of them carry information.

In community detection, for a node i at iteration t, we define $I_{in} = I(x_j, R_j^{t-1})$, and $I_{out} = I(x_i, R_i^t)$. In other words, the amount of information transferred *from* each node outward represents how confident (in terms of mutual information) is the belief of that node about the value of its own label.

To compute J and J^{-1}, we apply curve fitting using the Levenberg Marquardt algorithm (Ten Brink, 2001). Figures 5.1, 5.2, and 5.3 show the EXIT curves for different values of μ, α and ϵ. From these figures, we can deduce the following:

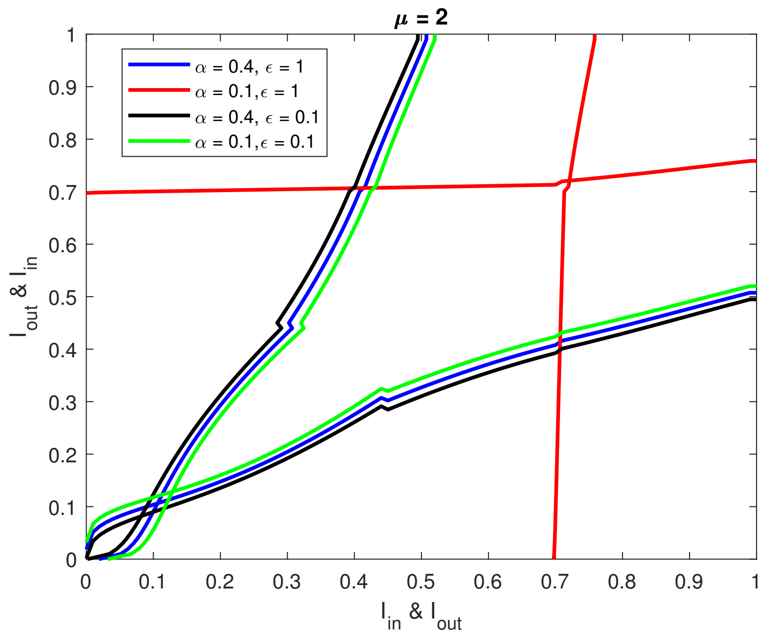

Figure 5.1. EXIT Chart for $\mu = 2$.

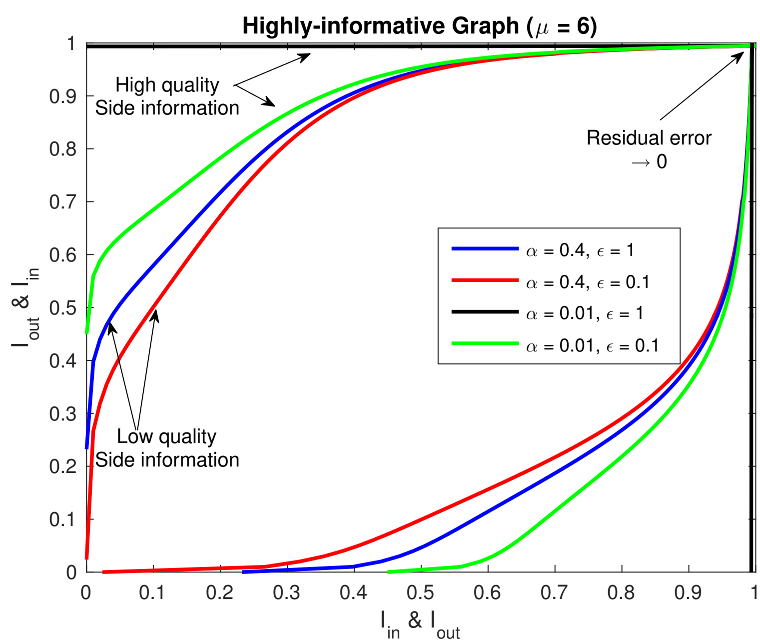

Figure 5.2. EXIT Chart for $\mu = 6$.

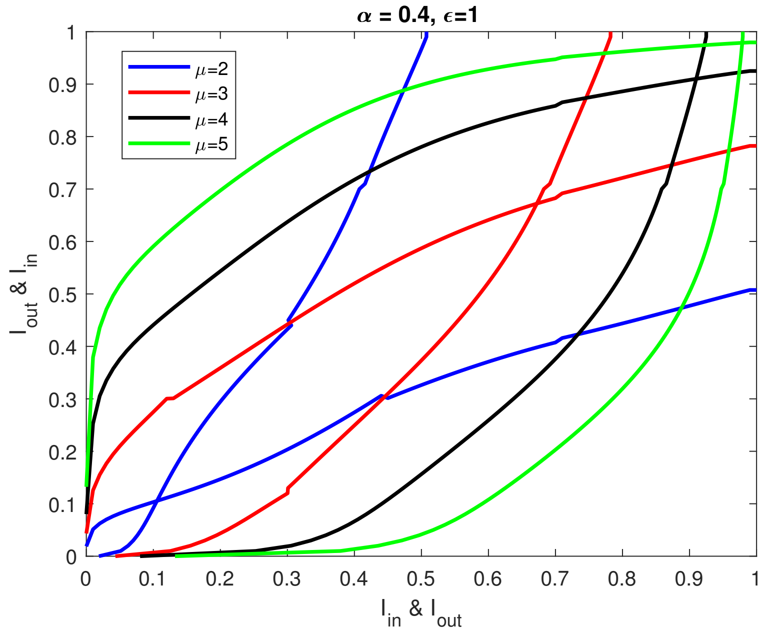

Figure 5.3. EXIT Chart for $\alpha = 0.4$ and $\epsilon = 1$.

- Side information, with any quantity (any $\epsilon \neq 0$), regardless of the quality (e.g. $\alpha = 0.4$), breaks the symmetry. Note that without side information the curves get stuck at the trivial $(0, 0)$ point, implying that the belief propagation algorithm is a trivial random guessing estimator (Saad et al., 2016). This is true for all values of μ.

- The starting point of the curves, which indicates the quality of the initial estimate, depends crucially on the values of μ, α, ϵ. For small values of μ, e.g. $\mu = 2$, EXIT charts reveal that the quantity of side information is not very important unless its quality is excellent. This can be seen in Figure 5.1: when $\alpha = 0.4$, the starting point for all values of $\epsilon \neq 0$ is almost the same. On the other hand, when $\alpha = 0.1$, the effect of ϵ on the starting point of the curve can be very significant, and the gap is around 0.7 between $\epsilon = 1$ and $\epsilon = 0.1$. For large values of μ, e.g. $\mu = 6$, the behavior changes. EXIT charts show that the effect of ϵ becomes more significant even when $\alpha = 0.4$. This is because larger values of μ imply larger difference between a and b, which means easier detection (quick convergence). Therefore when μ is large, the quality of the initial guess can make a bigger difference, proportionally.

- The intersection points on the curve exhibit almost the same behavior as the starting point. Note that the intersection point determines the value of $\bar{\nu}$, which determines the probability of error. In Figure 5.1, when $\alpha = 0.4$, the intersection points are very close in value for $\epsilon \neq 0$. This shows that the quantity of side information does not enhance the performance of belief propagation for small values of μ. On the other hand, when $\alpha = 0.1$, the effect of ϵ on the intersection point of the curve (i.e., probability of error) is significant, even when $\mu = 2$.

- EXIT charts also show that when the graph is not very informative, e.g., $\mu = 2$, even when side information provides significant information, e.g., when $\alpha = 0.1, \epsilon = 1$, the residual error does not improve markedly over the course of iterations. On the other hand, for highly-informative graphs, e.g., $\mu = 6$, even when side information provides a small amount of information, e.g., when $\alpha = 0.4, \epsilon = 0.1$, the eventual residual error improves significantly compared with the starting point.

- Although side information can break symmetry, even with high quality, e.g., $\alpha = 0.1$, unless $\epsilon \to 1$, one cannot hope to reach a vanishing fraction of misclassified nodes for a graph with small μ. This stems from the fact that the two communities are symmetric and for nodes with erased side information, the only source of information is the messages coming from its neighbors.

- When $\mu = 6$, for all values of $\alpha \in (0, 0.5)$ and $\epsilon \in (0, 1]$, one may achieve a vanishing fraction of misclassified nodes. This is because the only intersection point on the curve is approaching $(1, 1)$, which is the maximum mutual information available for binary variables.

- Figure 5.3 shows that as μ increases, there is always an intersection point. This suggests that one could not hope for vanishing residual error, i.e., weak recovery, except when $\mu \to \infty$ or $\alpha \to 0$. This suggests that belief propagation for recovering binary symmetric communities with side information does not have a phase transition for a finite μ.

5.3 One Community Stochastic Block Model

We begin by studying the performance of belief propagation on a random tree with side information. Then, we show that the same performance is possible on a random *graph* drawn according to the one community stochastic block model with side information, using a coupling lemma (Hajek et al., 2018).

Let T be an infinite tree with nodes indexed by variable i, each of them possessing a label $\tau_i \in \{0, 1\}$. The root is node $i = 0$. The sub-tree of depth t rooted at node i is denoted T_i^t. The sub-tree rooted at $i = 0$ with depth t is referenced often and is denoted simply 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 $\{u_1, \cdots, u_M\}$. 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 $\alpha_{+,m}, \alpha_{-,m}$, as earlier.

The goal is to infer the label τ_0 given observations T^t and $\tilde{\boldsymbol{\tau}}^t$. The error probability of the estimator $\hat{\tau}_0(T^t, \tilde{\boldsymbol{\tau}}^t)$ is:

$$p_e^t \triangleq \frac{K}{n}\mathbb{P}(\hat{\tau}_0 = 0|\tau_0 = 1) + \frac{n-K}{n}\mathbb{P}(\hat{\tau}_0 = 1|\tau_0 = 0) \tag{5.21}$$

The maximum a posteriori (MAP) detector minimizes p_e^t is given by $\hat{\tau}_{MAP} = 1_{\{\Gamma_0^t \geq \nu\}}$, where Γ_0^t is the log likelihood ratio,

$$\Gamma_0^t \triangleq \log\left(\frac{\mathbb{P}(T^t, \tilde{\boldsymbol{\tau}}^t|\tau_0 = 1)}{\mathbb{P}(T^t, \tilde{\boldsymbol{\tau}}^t|\tau_0 = 0)}\right) \tag{5.22}$$

and $\nu = \log(\frac{n-K}{K})$.

Lemma 7. *Let $\mathcal{N}_i$ denote the children of node i, $N_i \triangleq |\mathcal{N}_i|$ and $h_i \triangleq \log\left(\frac{\mathbb{P}(\tilde{\tau}_i|\tau_i=1)}{\mathbb{P}(\tilde{\tau}_i|\tau_i=0)}\right)$. Then,*

$$\Gamma_i^{t+1} = -K(p-q) + h_i + \sum_{k\in\mathcal{N}_i} \log\left(\frac{\frac{p}{q}e^{\Gamma_k^t - \nu} + 1}{e^{\Gamma_k^t - \nu} + 1}\right) \tag{5.23}$$

Proof. 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, as follows. The set of children of node i is denoted $\mathcal{N}_i$ with cardinality $N_i = |\mathcal{N}_i|$. If $\tau_i = 1$, the number of its children $N_i \sim \text{Poi}(Kp + (n-K)q)$ and each of these children j, independently of everything else has label $\tau_j = 1$ with probability $\frac{Kp}{Kp+(n-K)q}$ and $\tau_j = 0$ with probability $\frac{(n-K)q}{Kp+(n-K)q}$. If $\tau_i = 0$ the number of its children $N_i \sim \text{Poi}(nq)$ and each of these children j, independent from everything else, has label $\tau_j = 1$ with probability $\frac{K}{n}$ and $\tau_j = 0$ with probability $\frac{(n-K)}{n}$. Finally, for each node i in the tree, side information $\tilde{\tau}_i$ is observed according to $\alpha_{+,m}, \alpha_{-,m}$. Then:

$$\begin{aligned}
\Gamma_0^{t+1} &= \log\left(\frac{\mathbb{P}(T^{t+1}, \tilde{\tau}^{t+1}|\tau_0 = 1)}{\mathbb{P}(T^{t+1}, \tilde{\tau}^{t+1}|\tau_0 = 0)}\right)\\
&= \log\left(\frac{\mathbb{P}(N_0, \tilde{\tau}_0, \{T_k^t\}_{k\in\mathcal{N}_0}, \{\tilde{\tau}_k^t\}_{k\in\mathcal{N}_0}|\tau_0 = 1)}{\mathbb{P}(N_0, \tilde{\tau}_0, \{T_k^t\}_{k\in\mathcal{N}_0}, \{\tilde{\tau}_k^t\}_{k\in\mathcal{N}_0}|\tau_0 = 0)}\right)\\
&\overset{(a)}{=} \log\left(\frac{\mathbb{P}\big(N_0, \tilde{\tau}_0|\tau_0 = 1\big)}{\mathbb{P}\big(N_0, \tilde{\tau}_0|\tau_0 = 0\big)}\right) + \log\left(\frac{\prod_{k\in\mathcal{N}_0}\mathbb{P}\big(T_k^t, \tilde{\tau}_k^t|\tau_0 = 1\big)}{\prod_{k\in\mathcal{N}_0}\mathbb{P}\big(T_k^t, \tilde{\tau}_k^t|\tau_0 = 0\big)}\right)\\
&\overset{(b)}{=} \log\left(\frac{\mathbb{P}\big(N_0|\tau_0 = 1\big)}{\mathbb{P}\big(N_0|\tau_0 = 0\big)}\right) + \log\left(\frac{\mathbb{P}\big(\tilde{\tau}_0|\tau_0 = 1\big)}{\mathbb{P}\big(\tilde{\tau}_0|\tau_0 = 0\big)}\right)\\
&+ \sum_{k\in\mathcal{N}_0} \log\left(\frac{\sum_{\tau_k\in\{0,1\}}\mathbb{P}\big(T_k^t, \tilde{\tau}_k^t|\tau_k\big)\mathbb{P}\big(\tau_k|\tau_0 = 1\big)}{\sum_{\tau_k\in\{0,1\}}\mathbb{P}\big(T_k^t, \tilde{\tau}_k^t|\tau_k\big)\mathbb{P}\big(\tau_k|\tau_0 = 0\big)}\right)
\end{aligned}$$

$$\overset{(c)}{=} -K(p-q) + h_0 + \sum_{k\in\mathcal{N}_0} \log(\frac{\frac{p}{q}e^{\Gamma^t_k-\nu}+1}{e^{\Gamma^t_k-\nu}+1}) \tag{5.24}$$

where

- (a) holds because conditioned on τ_0 $(N_0, \tilde{\tau}_0)$ are independent of the rest of the tree and also $(T^t_k, \tilde{\tau}^t_k)$ 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 $\tau_0 = 0$.

□

The inference problem defined on the random tree is coupled to the recovering of a hidden community with side information through a coupling lemma (Hajek et al., 2018), which shows that under certain conditions, the neighborhood of a fixed node i in the graph is locally a tree with probability converging to one. Thus, the belief propagation algorithm defined for random trees 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 $\boldsymbol{G}^{\hat{t}}_u$ to be the subgraph containing all nodes that are at a distance at most $\hat{t}$ from node u and define $\boldsymbol{x}^{\hat{t}}_u$ and $\boldsymbol{Y}^{\hat{t}}_u$ to be the set of labels and side information of all nodes in $\boldsymbol{G}^{\hat{t}}_u$, respectively.

Lemma 8 (Coupling Lemma (Hajek et al., 2018))**.** *Suppose that $\hat{t}(n)$ are positive integers such that $(2+np)^{\hat{t}(n)} = n^{o(1)}$. 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:*

$$\mathbb{P}((\boldsymbol{G}^{\hat{t}}_u, \boldsymbol{x}^{\hat{t}}_u, \boldsymbol{Y}^{\hat{t}}_u) = (T^{\hat{t}}, \boldsymbol{\tau}^{\hat{t}}, \tilde{\boldsymbol{\tau}}^{\hat{t}})) \geq 1 - n^{-1+o(1)} \tag{5.25}$$

where for convenience of notation, the dependence of $\hat{t}$ on n is made implicit.

Table 5.2. 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 (5.26). 4. For all nodes i, compute its belief R^t_i based on (5.27). 5. Output $\tilde{C} = \{$Nodes corresponding to K largest $R^t_i\}$.

Now, we are ready to present 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:

$$R^{t+1}_{i\to j} = h_i - K(p-q) + \sum_{k\in\mathcal{N}_i\setminus j} M(R^t_{k\to i}) \tag{5.26}$$

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 and $M(x) = \log(\frac{\frac{p}{q}e^{x-\nu}+1}{e^{x-\nu}+1})$. The messages are initialized to zero for all nodes i, i.e., $R^0_{i\to j} = 0$ for all $i \in \{1,\cdots,n\}$ and $j \in \mathcal{N}_i$. Define the belief of node i at iteration $t+1$ as:

$$R^{t+1}_{i} = h_i - K(p-q) + \sum_{k\in\mathcal{N}_i} M(R^t_{k\to i}) \tag{5.27}$$

Algorithm 5.2 presents the proposed belief propagation algorithm for community recovery with side information.

If in Algorithm 5.2 we have $t = \hat{t}(n)$, according to Lemma 8 with probability converging to one $R^t_i = \Gamma^t_i$, where Γ^t_i was the log-likelihood defined for the random tree. Hence, the performance of Algorithm 5.2 is expected to be the same as the MAP estimator defined as $\hat{\tau}_{MAP} = 1_{\{\Gamma^t_i \geq \nu\}}$, where $\nu = \log(\frac{n-K}{K})$.

We now study the asymptotic behavior of Γ_i^t. Define for $t \geq 1$ and any node i:

$$\psi_i^t \triangleq -K(p-q) + \sum_{j \in \mathcal{N}_i} M(h_j + \psi_j^{t-1}) \tag{5.28}$$

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, $\Gamma_i^{t+1} = h_i + \psi_i^{t+1}$ and $\psi_i^0 = 0\ \forall i \in T^t$. Let Z_0^t and Z_1^t denote random variables drawn according to the distribution of ψ_i^t conditioned on $x_i = 0$ and $x_i = 1$, 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.

Lemma 9. *((Saad and Nosratinia, 2018a, Lemma 11)) Assume λ, $\frac{\alpha_{+,m}}{\alpha_{-,m}}$ and ν are constants independent of n while $nq, Kq \xrightarrow{n\to\infty} \infty$. Then, for all $t \geq 0$:*

$$\mathbb{E}[Z_0^{t+1}] = \frac{-\lambda}{2}b_t + o(1) \tag{5.29}$$

$$\mathbb{E}[Z_1^{t+1}] = \frac{\lambda}{2}b_t + o(1) \tag{5.30}$$

$$var(Z_0^{t+1}) = var(Z_1^{t+1}) = \lambda b_t + o(1) \tag{5.31}$$

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

Lemma 10. *((Saad and Nosratinia, 2018a, Lemma 12)) Assume λ, $\frac{\alpha_{+,m}}{\alpha_{-,m}}$ 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 $t \geq 0$:*

$$\sup_x \Big|\mathbb{P}\big(\frac{Z_0^{t+1}+\frac{v_{t+1}}{2}}{\sqrt{v_{t+1}}} \leq x\big) - \phi(x)\Big| \to 0 \tag{5.32}$$

$$\sup_x \Big|\mathbb{P}\big(\frac{Z_1^{t+1}-\frac{v_{t+1}}{2}}{\sqrt{v_{t+1}}} \leq x\big) - \phi(x)\Big| \to 0 \tag{5.33}$$

The following lemma characterizes the asymptotic residual error of belief propagation with side information for recovering a single community.

Lemma 11. *Assume* λ, $\frac{\alpha_{+,m}}{\alpha_{-,m}}$ *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,*

$$\lim_{nq,Kq\to\infty} \lim_{n\to\infty} \frac{\mathbb{E}[\hat{C}\triangle C^*]}{K} = \frac{n-K}{K}\mathbb{E}_{U_0}[Q(\frac{\nu + \frac{v_t}{2} - U_0}{\sqrt{v_t}})] + \mathbb{E}_{U_1}[Q(\frac{-\nu + \frac{v_t}{2} + U_1}{\sqrt{v_t}})] \tag{5.34}$$

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 9 and 10, using

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

□

5.3.1 Exit Analysis

An interesting and natural question is: does belief propagation with side information have a phase transition? If yes, what is the threshold? Equation (5.34) shows the residual asymptotic error of belief propagation for detecting one community with side information. However, it does not provide a direct answer regarding phase transition. This section demonstrates the utility of EXIT charts in the understanding of phase transition.

We begin by calculating the mutual information between the label of node i, x_i, and its belief at time t, R_i^t as follows:

$$\begin{aligned}
&I(x_i, R_i^t)\\
&= -\frac{K}{n}\log(\frac{K}{n}) - (1-\frac{K}{n})\log(1-\frac{K}{n}) - H(x_i|R_i^t)\\
&= -\frac{K}{n}\log(\frac{K}{n}) - (1-\frac{K}{n})\log(1-\frac{K}{n})
\end{aligned}$$

$$- \frac{K}{n} \int_{-\infty}^{\infty} \left(\sum_{m=1}^{M} \alpha_{+,m} \frac{e^{\frac{-(y-(v_t+h_m))^2}{2v_t}}}{\sqrt{2\pi v_t}} \right) \log_2 \left(1 + \frac{(n-K)\sum_{m=1}^{M} \alpha_{-,m} e^{\frac{-(y-(-v_t+h_m))^2}{2v_t}}}{K \sum_{m=1}^{M} \alpha_{+,m} e^{\frac{-(y-(v_t+h_m))^2}{2v_t}}} \right) dy$$
$$- \frac{n-K}{n} \int_{-\infty}^{\infty} \left(\sum_{m=1}^{M} \alpha_{-,m} \frac{e^{\frac{-(y-(-v_t+h_m))^2}{2v_t}}}{\sqrt{2\pi v_t}} \right) \log_2 \left(1 + \frac{K \sum_{m=1}^{M} \alpha_{+,m} e^{\frac{-(y-(v_t+h_m))^2}{2v_t}}}{(n-K)\sum_{m=1}^{M} \alpha_{-,m} e^{\frac{-(y-(-v_t+h_m))^2}{2v_t}}} \right) dy \tag{5.35}$$

where $h_m = \log(\frac{u_{+,m}}{u_{-,m}})$.

For a concrete demonstration of the capabilities of EXIT analysis, we use the following model for side information. Let $M = 2$, where for each node i, $y_i = x_i$ with probability $1-\alpha$, and $y_i = 1 - x_i$ with probability α, where $\alpha \in [0, 0.5]$. Note that for a fixed α, $I(x_i, R_i^t)$ is function of v_t only. Hence, we will denote it by $J(v_t)$.

Based on the belief propagation algorithm described in Table 5.2, at iteration t, node i receives the beliefs of all nodes $j \in \mathcal{N}_i$ calculated at iteration $(t-1)$. We denote the input information to node i from node j as I_{in}. Then, node i computes the new information it has at iteration t, which we call I_{out}. Note that I_{in} and I_{out} can be calculated using (5.35) as $J(v_{t-1})$ and $J(v_t)$, respectively. Since $J(v_t)$ is monotonically increasing in v_t (Ten Brink, 2001), $J(v_t)$ is reversible. Thus, $v_t = J^{-1}(I(x_i, R_i^t))$. Moreover, since v_{t-1} and v_t are related by:

$$v_t = \lambda \mathbb{E}_{Z,U_1} \left[\frac{1}{e^{-\nu} + e^{-(\frac{v_{t-1}}{2} + \sqrt{v_{t-1}} Z) - U_1}} \right],$$

we can define the relation between I_{in} and I_{out} for node i as follows:

$$\begin{aligned} I_{out} =& J\Bigg(\lambda \Bigg[\alpha \mathbb{E}_Z \Bigg[\Big(e^{-\nu} + e^{-(\frac{J^{-1}(I_{in})}{2} + \sqrt{J^{-1}(I_{in})} Z) - \log(\frac{\alpha}{1-\alpha})} \Big)^{-1} \Bigg] + \\ & (1-\alpha) \mathbb{E}_Z \Bigg[\Big(e^{-\nu} + e^{-(\frac{J^{-1}(I_{in})}{2} + \sqrt{J^{-1}(I_{in})} Z) - \log(\frac{1-\alpha}{\alpha})} \Big)^{-1} \Bigg] \Bigg] \Bigg) \end{aligned} \tag{5.36}$$

To compute J and J^{-1}, we apply curve fitting using the Levenberg Marquardt algorithm (Ten Brink, 2001). Figures 5.4, 5.5 and 5.6 show the EXIT curves for different values of λ, and α.

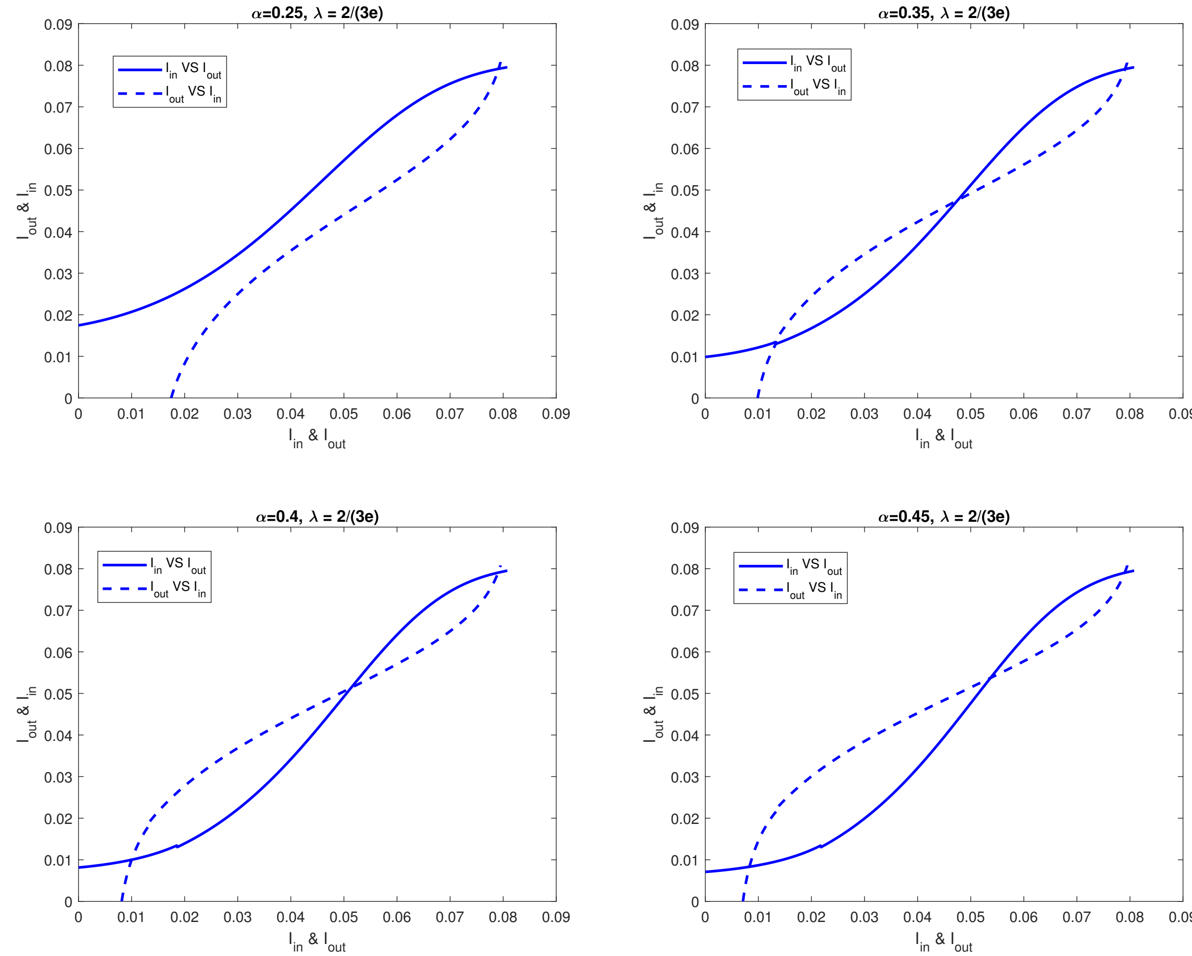

Figure 5.4. EXIT charts for one community detection with $\lambda = \frac{2}{3e}$ for different values of α.

- Figure 5.4 shows a threshold for λ, such that the EXIT curves do not intersect above this threshold and they do below the threshold. Hence, belief propagation with side information experiences a phase transition. Moreover, above the threshold, the maximum mutual information is attained, and hence, a vanishing residual error is possible (weak recovery). This particular example is constructed for a graph whose probability distribution does not provide sufficient information *alone* for weak recovery. This example demonstrates clearly the role of side information in weak recovery especially in conditions where, without it, weak recovery is not attainable. EXIT analysis thus confirms the threshold effect that was first reported in (Saad and Nosratinia, 2018a), but more importantly, EXIT demonstrates the phase transition behavior in a visually compelling manner that is easy to grasp, with relatively straight forward calculations.

- To elaborate, EXIT charts bring further clarity to the nature of the belief propagation threshold, by showing how the iterations of the belief propagation, at threshold, just barely manage to escape through a bottleneck and approach the maximum likelihood solution. EXIT also clearly demonstrates the residual error of belief propagation on the two sides of the phase transition (the jump in error probability at phase transition) which is not as easy to see via other analytical methods.

- Thus, the EXIT method demonstrates that while the thresholding phenomenon for belief propagation is indeed sharp in terms of transition across parameters of the model for the graph and side information, however, close to the threshold the belief propagation might pay a heavy price in terms of the number of iterations needed to converge. Thus, in the sense of the cost of the algorithm, the behavior of belief propagation near the threshold is something that is especially well understood via the EXIT analysis. The curvature (second derivative) of the EXIT curves at the point of bottleneck is an indication of the iterations needed close to the threshold. This effect is not visible to the other analytical methods that, typically, first let the number of iterations go to infinity, and then observe the (asymptotic, in iterations) performance of the belief propagation algorithm across the landscape of the parameters of the system model.

- As mentioned earlier, Fig. 5.4 shows the thresholding effect for the side information where the graphical information is fixed. In order to complete the picture, we also performed experiments where we hold the quality of the side information to be fixed (via a fixed α), while we allow the graph to become progressively more informative (characterized by improving λ). This result is shown in Figures 5.5 and 5.6. In these

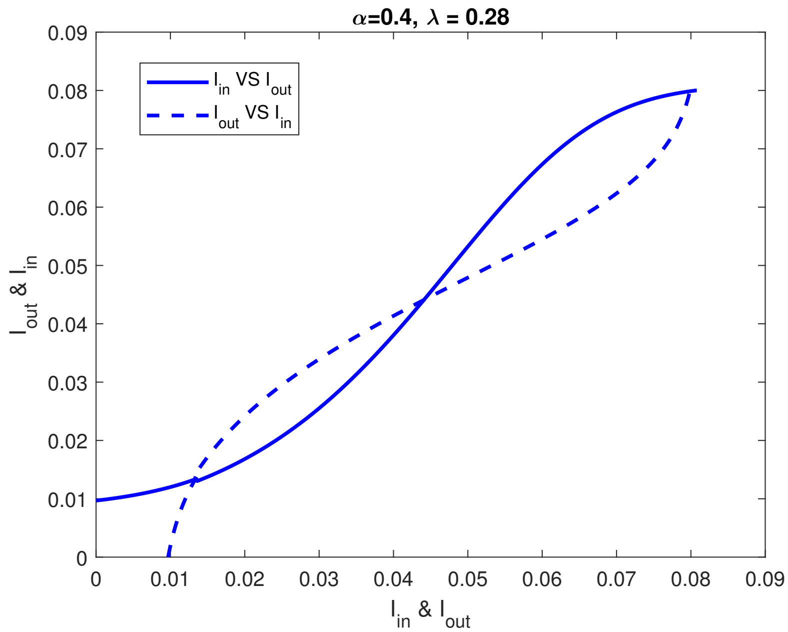

Figure 5.5. EXIT Chart for one community detection with $\alpha = 0.4$.

figures, the thresholding effect for graphical information is shown in the presence of side information.

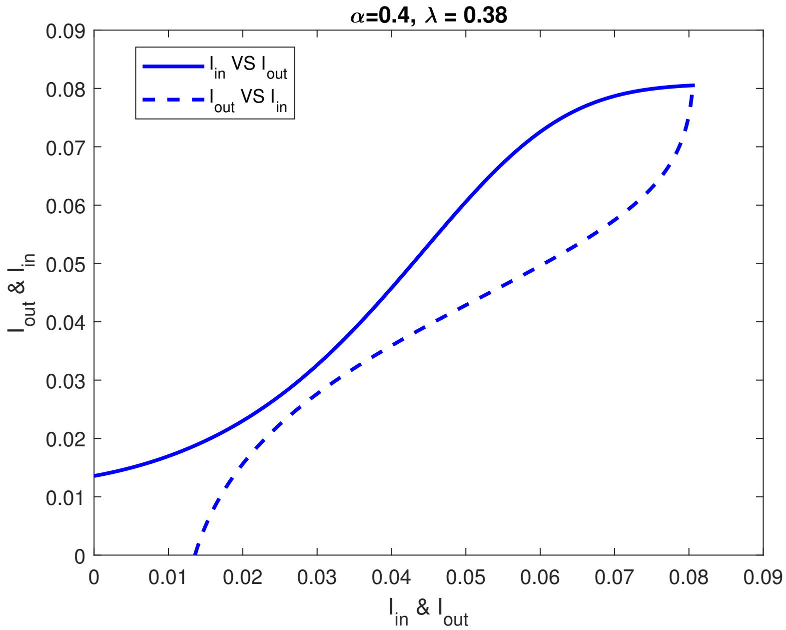

Figure 5.6. EXIT Chart for one community detection with $\alpha = 0.4$.

CHAPTER 6

CONCLUSION

This dissertation studies the community detection problem when non-graphical observations (side information) is available. Specifically, the following questions are considered: 1) when can side information change the fundamental limits of the community detection problem? 2) Can we devise efficient algorithms that incorporate side information? 3) what is the asymptotic performance of these algorithms?

Under the binary symmetric stochastic block model, we study the effect of quality and quantity of side information on the phase transition of exact recovery. To model quality, we propose three different discrete-valued side information models, where in all of them the LLR of side information is allowed to vary with n, while the number of observations per node is fixed. To model quantity, we assume each node observes a vector of i.i.d. observations, where the LLR is fixed and the dimension of the vector is allowed to vary with n. In all models, tight sufficient and necessary conditions for exact recovery are characterized. We show that for side information to change the phase transition of exact recovery, either the quality or quantity of side information has to grow at least as fast as $\Omega(\log(n))$. For the sufficient conditions, we propose a two-step efficient algorithm and show that it is asymptotically optimal for all the proposed models. Moreover, we characterize a more general phase transition when side information is infinite-valued or continuous.

Under the single community stochastic block model, we study the effect of quality and quantity of side information on the phase transition (information limits) of both weak and exact recovery, as well as the phase transition of an efficient algorithm, namely, belief propagation. 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. First, for the information limits, when the dimension of side information for each node varies but its LLR is fixed across n, tight necessary and sufficient conditions are calculated for both weak and exact recovery. The results show which classes of side information can change the information limits. Also, it is shown that under the same sufficient conditions, weak recovery is achievable even when the size of the community is random and unknown. The results for weak recovery are shown under maximum likelihood detection, while the results for exact recovery are shown under a two-stage algorithm. 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 the side information for each node has fixed dimension but varying LLR, we characterize tight necessary and sufficient conditions for exact recovery, and necessary conditions for weak recovery. Under varying LLR, our results apply to side information with finite alphabet.

Second, the phase transition of belief propagation in the presence of side information is characterized, where we assumed the side information per node has a fixed dimension. When the LLRs are fixed across n, tight necessary and sufficient conditions are calculated for weak recovery. We show that side information provides a gain that is proportional to the chi-squared distance between the conditional distributions of side information. Furthermore, it is shown that when belief propagation fails, no local algorithm can achieve weak recovery. We compare belief propagation phase transition for weak recovery with maximum likelihood, and it is shown than belief propagation is strictly inferior to the maximum likelihood detector. Numerical results on finite synthetic data-sets are presented that validated our asymptotic analysis and showed its relevance to even graphs of moderate size. The results show that when weak recovery is not asymptotically feasible, the fraction of misclassified nodes is significant, and when weak recovery is asymptotically feasible, the fraction of misclassified nodes is small. We also calculate conditions under which belief propagation followed by a local voting procedure achieves exact recovery. When the side information has variable LLR across n, the belief propagation misclassification rate is calculated using density evolution, and we show sufficient conditions under which weak recovery is always feasible.

Finally, we propose a new tool, namely EXIT, for the analysis of the performance of local message passing algorithms, e.g., belief propagation, for community detection with side information. EXIT analysis has been used to understand the behavior of iterative algorithms in the context of error control and communication systems. We apply EXIT analysis to single-community detection as well as to binary symmetric community detection, each with side information, and leveraged this technique to provide insights on: 1) The effect of the quality and quantity of side information on the performance of belief propagation, e.g. probability of error, 2) The asymptotic threshold for weak recovery, achieving a vanishing residual error, 3) The performance of belief propagation near the optimal threshold, 4) The performance of belief propagation through the first few iterations, and 5) Approximating the number of iterations needed for convergence.

REFERENCES

Abbe, E., A. Bandeira, and G. Hall (2016, January). Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory 62*(1), 471–487.

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

Abbe, E. and C. Sandon (2018). Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics 71*(7), 1334–1406.

Anandkumar, A., R. Ge, D. Hsu, and S. M. Kakade (2014, January). A tensor approach to learning mixed membership community models. *J. Mach. Learn. Res. 15*(1), 2239–2312.

Apostol, T. (1962). *Calculus*, Volume 2. Blaisdell Pub. Co.

Asadi, A. R., E. Abbe, and S. Verd (2017, June). Compressing data on graphs with clusters. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 1583–1587.

Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and newman-girvan and other modularities. *Proceedings of the National Academy of Sciences 106*(50), 21068–21073.

Bollobs, B. (1998). *Modern Graph Theory* (corrected ed.). Graduate texts in mathematics. Springer.

Cai, T. T. and X. Li (2015, June). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *The Annals of Statistics 43*(3), 1027–1059.

Cai, T. T., T. Liang, and A. Rakhlin (2016, Mar). Inference via Message Passing on Partially Labeled Stochastic Block Models. *arXiv e-prints*, arXiv:1603.06923.

Caltagirone, F., M. Lelarge, and L. Miolane (2018, July). Recovering asymmetric communities in the stochastic block model. *IEEE Transactions on Network Science and Engineering 5*(3), 237–246.

Chen, J. and B. Yuan (2006, September). Detecting functional modules in the yeast proteinprotein interaction network. *Bioinformatics 22*(18), 2283–2290.

Chen, Y., S. Sanghavi, and H. Xu (2014, Oct). Improved graph clustering. *IEEE Trans. Inform. Theory 60*(10), 6440–6455.

Chen, Y. and J. Xu (2016, January). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *J. Mach. Learn. Res. 17*(1), 882–938.

Coja-Oghlan, A. (2005). A spectral heuristic for bisecting random graphs. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, Philadelphia, PA, USA, pp. 850–859. Society for Industrial and Applied Mathematics.

Coja-oghlan, A. (2010, March). Graph partitioning via adaptive spectral techniques. *Comb. Probab. Comput. 19*(2), 227–284.

Csermely, P. (2008). Creative elements: network-based predictions of active centres in proteins and cellular and social networks. *Trends in Biochemical Sciences 33*(12), 569 – 576.

Decelle, A., F. Krzakala, C. Moore, and L. Zdeborová (2011a, December). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E 84*, 066106.

Decelle, A., F. Krzakala, C. Moore, and L. Zdeborová (2011b, Aug). Inference and phase transitions in the detection of modules in sparse networks. *Phys. Rev. Lett. 107*, 065701.

Dembo, A. and O. Zeitouni (2010). *Large deviations techniques and applications.* Berlin; New York: Springer-Verlag Inc.

Elchanan, M., J. Neeman, and S. Allan (2015). Consistency thresholds for the planted bisection model. In *ACM Symposium on Theory of Computing*, pp. 69–75.

Euler, L. (1736). Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae 8*, 128–140.

Fortunato, S. (2010a). Community detection in graphs. *Physics Reports 486*(3-5), 75 – 174.

Fortunato, S. (2010b, January). Community detection in graphs. *Physics Reports 486*(3), 75 – 174.

Girvan, M. and M. E. J. Newman (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences 99*(12), 7821–7826.

Hajek, B., Y. Wu, and J. Xu (2017, Aug). Information limits for recovering a hidden community. *IEEE Transactions on Information Theory 63*(8), 4729–4745.

Hajek, B., Y. Wu, and J. Xu (2018). Recovering a hidden community beyond the kestenstigum threshold in $o(|E|\log^*|V|)$ time. *Journal of Applied Probability 55*(2), 325352.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American statistical association 58*(301), 13–30.

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

Kadavankandy, A., K. Avrachenkov, L. Cottatellucci, and R. Sundaresan (2018, April). The power of side-information in subgraph detection. *IEEE Transactions on Signal Processing 66*(7), 1905–1919.

Kanade, V., E. Mossel, and T. Schramm (2016, October). Global and local information in clustering labeled block models. *IEEE Trans. Inform. Theory 62*(10), 5906–5917.

Kobayashi, H. and J. Thomas (1967). distance measures and related criteria. In *Allerton Conference Circuits and System Theory*.

Korolev, V. and I. Shevtsova (2012). An improvement of the berryesseen inequality with applications to poisson and mixed poisson random sums. *Scandinavian Actuarial Journal 2012*(2), 81–105.

Krishnamurthy, B. and J. Wang (2000). On network-aware clustering of web clients. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, SIGCOMM '00, pp. 97–110. ACM.

Lancichinetti, A. and S. Fortunato (2009, November). Community detection algorithms: A comparative analysis. *Phys. Rev. E 80*, 056117.

Massoulié, L. (2014). Community detection thresholds and the weak ramanujan property. In *ACM Symposium on Theory of Computing*, STOC '14, pp. 694–703.

Montanari, A. (2015, October). Finding one community in a sparse graph. *Journal of Statistical Physics 161*(2), 273–299.

Mossel, E., J. Neeman, and A. Sly (2014). Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields 162*(3), 431–461.

Mossel, E., J. Neeman, and A. Sly (2018, June). A proof of the block model threshold conjecture. *Combinatorica 38*(3), 665–708.

Mossel, E. and J. Xu (2016a, 23–26 Jun). Density evolution in the degree-correlated stochastic block model. In V. Feldman, A. Rakhlin, and O. Shamir (Eds.), *29th Annual Conference on Learning Theory*, Volume 49 of *Proceedings of Machine Learning Research*, Columbia University, New York, New York, USA, pp. 1319–1356. PMLR.

Mossel, E. and J. Xu (2016b). Local algorithms for block models with side information. In *ACM Conference on Innovations in Theoretical Computer Science*, ITCS '16, New York, NY, USA, pp. 71–80. ACM.

Newman, M. and A. Clauset (2016). Structure and inference in annotated networks. In *Nature communications*.

Paley, R. E. A. C. and A. Zygmund (1932). A note on analytic functions in the unit circle. *Mathematical Proceedings of the Cambridge Philosophical Society 28*(3), 266–272.

Polyanskiy, Y. and Y. Wu (2017, Jan.). Lecture notes on information theory.

Reddy, P. K., M. Kitsuregawa, P. Sreekanth, and S. S. Rao (2002). A graph based approach to extract a neighborhood customer community for collaborative filtering. In *Proceedings of the Second International Workshop on Databases in Networked Information Systems*, DNIS '02, pp. 188–200. Springer-Verlag.

Saad, H., A. Abotabl, and A. Nosratinia (2016, July). Exit analysis for belief propagation in degree-correlated stochastic block models. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 775–779.

Saad, H., A. Abotabl, and A. Nosratinia (2017, October). Exact recovery in the binary stochastic block model with binary side information. In *Allerton Conference on Communication, Control, and Computing*, pp. 822–829.

Saad, H. and A. Nosratinia (2018, June). Belief propagation with side information for recovering a single community. In *IEEE International Symposium on Information Theory*.

Saad, H. and A. Nosratinia (2018, Oct). Community detection with side information: Exact recovery under the stochastic block model. *IEEE Journal of Selected Topics in Signal Processing 12*(5), 944–958.

Saad, H. and A. Nosratinia (2018a). Recovering a single community with side information. *CoRR abs/1809.01738*.

Saad, H. and A. Nosratinia (2018b, June). Side information in recovering a single community: Information theoretic limits. In *IEEE International Symposium on Information Theory*.

Saad, H. and A. Nosratinia (2019, Feb). Exact recovery in community detection with continuous-valued side information. *IEEE Signal Processing Letters 26*(2), 332–336.

Saad, H. and A. Nosratinia (2019). EXIT analysis for community detection. *CoRR abs/1901.09656*.

Snijders, T. A. B. and K. Nowicki (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification 14*, 75–100.

Ten Brink, S. (2001). Convergence behavior of iteratively decoded parallel concatenated codes. *IEEE Transactions on Communications 49*(10), 1727–1737.

Wu, A. Y., M. Garland, and J. Han (2004). Mining scale-free networks using geodesic clustering. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, New York, NY, USA, pp. 719–724. ACM.

Xu, J., R. Wu, K. Zhu, B. Hajek, R. Srikant, and L. Ying (2014, June). Jointly clustering rows and columns of binary matrices: Algorithms and trade-offs. *SIGMETRICS Perform. Eval. Rev. 42*(1), 29–41.

Yang, J., J. McAuley, and J. Leskovec (2013, December). Community detection in networks with node attributes. In *IEEE International Conference on Data Mining*, pp. 1151–1156.

Yun, S. and A. Proutiere (2014). Community detection via random and adaptive sampling. In *Conference on Learning Theory*, pp. 138–175.

Zhang, A. and H. Zhou (2016, October). Minimax rates of community detection in stochastic block models. *The Annals of Statistics 44*(5), 2252–2280.

Zhang, P., F. Krzakala, J. Reichardt, and L. Zdeborová (2012). Comparative study for inference of hidden classes in stochastic block models. *Journal of Statistical Mechanics: Theory and Experiment 2012*(12), P12021.

Zhang, Y., E. Levina, and J. Zhu (2016). Community detection in networks with node features. *Electron. J. Statist. 10*(2), 3153–3178.

BIOGRAPHICAL SKETCH

Hussein Metwaly Saad received his BS degree in electrical engineering from the German University, Cairo, Egypt, and his MS degree in electrical engineering from Nile University, Giza, Egypt. He is currently working toward his PhD in electrical engineering at The University of Texas at Dallas, Richardson, TX, USA. His research interests include information theory and its applications in data science and machine learning. He received the Erik Jonsson Graduate Fellowship in 2014 from The University of Texas at Dallas.

CURRICULUM VITAE

Hussein Metwaly Saad

Summary

An electrical engineer with 7+ years of combined research and industrial experience in information theory, machine learning and wireless communications. I have a strong mathematical background and an in depth understanding of the theory behind machine learning algorithms, e.g. statistical learning theory. Moreover, I have a decent experience with Deep Learning architectures such as CNN, DNN, RNN, GCNN and others.

Key Skills

- In depth knowledge and solid background in mathematics: Information theory, Statistics, Statistical Learning Theory, and Random Processes.
- Good knowledge and experience in Graphical Models and its applications in Machine Learning and Deep Learning.
- Good knowledge and experience in deep learning architectures such as Convolutional Neural Networks, Recurrent Neural Networks, Graphical Convolutional Neural networks.
- Certified from Coursera: Deep Learning Specialization, Advanced Machine Learning Specialization, Data Structures and Algorithms Specialization.
- Good knowledge and experience in using tool kits such as: Keras and TensorFlow.
- Experience in Reinforcement learning and its applications to wireless communications.
- Theoretical performance analysis of communication systems.
- Programming Skills: Python, MATLAB, Java.

Professional Experience

The University of Texas at Dallas (UTD), Dallas, Texas

Research Assistant/ Teaching Assistant **September 2014 -August 2019**

- Conducting research on Information theory and its applications in Machine learning and Data science.
- Understanding and developing algorithms for Community Detection on graphs when non-graphical data is available.

- Understanding adversarial learning and developing robust learning algorithms with theoretical guarantees.

Varkon Semiconductors, Cairo, Egypt
System Design Engineer ***May 2013-June 2014***

- Design of a Digital Phase Locked Loop that is used for clock synchronization and recovery (floating and fixed point).
- Design of a Digital Farrow Filter (floating and fixed point and C code).

Nile University with Cooperation from Qatar University
Research Assistant ***October 2010-April 2013***

- Utilizing learning theory such as Reinforcement learning to perform resource allocation in femtocell networks.

Publications

1. H. Saad and A. Nosratinia, "EXIT Analysis for Community Detection," Submitted to IEEE Transactions on Signal Processing
2. H. Saad and A. Nosratinia, "Recovering a Single Community with Side Information," Submitted to IEEE Transaction of Information Theory
3. H. Saad and A. Nosratinia, "Exact Recovery in Community Detection with Continuous-Valued Side Information," IEEE Signal Processing Letters, vol. 26, pp. 332-336, 2019.
4. 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, June 2018.
5. H. Saad, A. Mohamed, and T. ElBatt, Cooperative Q-learning Techniques for Distributed Online Power Allocation in Femtocell Networks, Wiley Wireless Communications and Mobile Computing Journal, April 2013.
6. Hussein Saad and Aria Nosratinia, Belief Propagation with Side Information for Recovering a Single Community, IEEE International Symposium on Information Theory, July 2018.
7. Hussein Saad and Aria Nosratinia, Side Information in Recovering a Single Community: Information Theoretic Limits, at IEEE International Symposium on Information Theory, July 2018 .
8. H. Saad, A. Abotabl and A. Nosratinia, Exact recovery in the binary stochastic block model with binary side information, Allerton Conference on Communications, Control, and Computing, October 2017.
9. H. Saad, A. Abotabl and A. Nosratinia, EXIT analysis for belief propagation in degree-correlated stochastic block models, IEEE International Symposium on Information Theory, July 2016.

10. A. El Shafie, T. Khattab, H. Saad, and A. Mohamed, Optimal Cooperative Cognitive Relaying and Spectrum Access for an Energy Harvesting Cognitive Radio: Reinforcement Learning Approach, IEEE International Conference on Computing, Networking and Communications, Feb. 2015.

11. H. Saad, A. Mohamed, and T. ElBatt, A Cooperative Q-learning Approach for Distributed Resource Allocation in Multi-user Femtocell Networks, IEEE Wireless Communications and Networking Conference, April 2014.

12. H. Saad, A. Mohamed, and T. ElBatt, A Cooperative Q-Learning Approach for Online Power Allocation in Femtocell Networks, Proceedings of the IEEE 77th Vehicular Technology Conference Sep. 2013.

13. H. Saad, A. Mohamed, and T. ElBatt, Distributed Cooperative Q-learning for Power Allocation in Cognitive Femtocell Networks, Proceedings of the IEEE 76th Vehicular Technology Conference, Sep. 2012.

Education

The University of Texas at Dallas, Dallas, Tx
P.hD. in Electrical Engineering, ***August 2014 - Current***
Advisor: Prof. Aria Nosratinia.
Research: A Generalization of the Community Detection Problem via Side Information

Nile University, Egypt
M.Sc. in Electrical Engineering, ***October 2010 - April 2013*** Advisors: Tamer El-Batt and Amr Mohamed (Qatar University).
Research: Q-Learning Techniques for Distributed Online Resource Allocation in Femtocell Networks

The German University in Cairo, Egypt
B.Sc. in Electrical Engineering, ***Sept. 2005 - June 2010***