Side Information in the Binary Stochastic Block Model: Exact Recovery

Hussein Saad, Student Member, IEEE, Ahmed Abotabl, Student Member, IEEE and Aria Nosratinia, Fellow, IEEE

Abstract

In the community detection problem, one may have access to additional observations (side information) about the label of each node. This paper studies the effect of the quality and quantity of side information on the phase transition of exact recovery in the binary symmetric stochastic block model (SBM) with \( n \) nodes. When the side information consists of the label observed through a binary symmetric channel with crossover probability \( \alpha \), and when \( \log \left( \frac{1 - \alpha}{\alpha} \right) = O(\log(n)) \), it is shown that side information has a positive effect on phase transition; the new phase transition under this condition is characterized. When \( \alpha \) is constant or approaches zero sufficiently slowly, i.e., \( \log \left( \frac{1 - \alpha}{\alpha} \right) = o(\log(n)) \), it is shown that side information does not help exact recovery. When the side information consists of the label observed through a binary erasure channel with parameter \( \epsilon \), and when \( \log(\epsilon) = O(\log(n)) \), it is shown that side information improves exact recovery and the new phase transition is characterized. If \( \log(\epsilon) = o(\log(n)) \), then it is shown that side information is not helpful. The results are then generalized to an arbitrary side information of finite cardinality. Necessary and sufficient conditions are derived for exact recovery that are tight, except for one special case under \( M \)-ary side information. An efficient algorithm that incorporates the effect of side information is proposed that uses a partial recovery algorithm combined with a local improvement procedure. Sufficient conditions are derived for exact recovery under this efficient algorithm.

Index Terms

Community detection, Stochastic block model, Side information, Exact recovery.

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

The problem of learning or detecting community structures in random graphs has been studied in statistics [1]–[5], computer science [6]–[10] and theoretical statistical physics [11], [12], among others. The problem of detecting communities has many applications: finding like-minded people in social networks [13], improving recommendation systems [14], detecting protein complexes [15]. Among the different random graph models [16], [17], the stochastic block model (SBM) is widely used in the context of community detection [18]. This extension of the Erdos-Renyi model consists of $n$ nodes that belong to two communities, each pair connected with an edge with probability $p$ if the pair belongs to the same community and with probability $q$ if they do not. The prior distribution of the node labels is identical and independent, and often uniform (labels are equi-probable). The goal of community detection is to recover/detect the labels upon observing the graph edges.

The graphical structure of inference problems can lead to well-characterized asymptotic results (e.g. phase transitions) that can give insights on the performance of inference algorithms on large data sets. But a purely graphical model for observations also unfortunately limits the scope of the applicability of the model. For example, social networks such as Facebook and Twitter have access to much information other than the graph edges. 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.

In statistics, the problem of community detection with additional information such as “annotation” [19], “attributes” [20], or “features” [21] 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. This is unlike the present paper that is concerned with phase transitions and thresholds, but they nevertheless show the interest of the broader community in the issue of side-information in graph-based inference.

For reference, a few basic definitions are highlighted before continuing with the literature survey. Correlated recovery refers to community detection that performs better than random

We consider the binary stochastic block model.
guessing [22]–[26]. Weak recovery refers to a vanishing fraction of misclassified labels [27]–[29]. Exact recovery refers to recovering all communities with probability converging to one as $n \to \infty$ [18], [30], [31]. 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 [32] considered side information consisting of the label observed through a binary symmetric channel, showing that subject to side information the belief propagation under certain condition has the same residual error as the MAP estimator. They also showed weak recovery if the average degree grows with $n$. Cai et. al [33] considered side information consisting of the label observed through a binary erasure channel (BEC) with erasure probability $\epsilon \to 1$ as $n \to \infty$, demonstrating regimes for correlated recovery and weak recovery. Both [32], [33] present sufficient (but not necessary) conditions. Kadavankandy et al. [34] studies the single-community problem under side information consisting of the labels observed through a binary asymmetric channel, where they showed weak recovery in the sparse regime. Kanade et. al [35] showed that for symmetric communities, the phase transition of correlated recovery is not affected if labels are observed through a BEC whose erasure probability $\epsilon \to 1$ as $n$ grows. The same side information was shown to be helpful to correlated recovery of local algorithms. Caltagirone et. al [36] considered binary asymmetric communities, showing that in the presence of side information mentioned above, local algorithms achieve correlated recovery up to the phase transition threshold.

This paper studies the effect of side information on binary community detection, 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. Also, the literature has concentrated on the effect of side information only on belief propagation, which is not enough to determine phase transition. Second, only binary side information about binary labels has been studied (or binary side information with erasures). The more general case where side information has an arbitrary alphabet is motivated by many practical applications, but an M-ary side information has not been thoroughly studied either in the context of belief propagation or maximum likelihood. Finally, in many cases either necessary or sufficient conditions for recovery
is known, but not both.

In this paper, we study the effect of side information on exact recovery in the dense regime, i.e., when \( p = a \frac{\log n}{n} \) and \( q = b \frac{\log n}{n} \) with constants \( a > b > 0 \). We investigate the question: when and how much can side information affect the phase transition threshold of exact recovery? We study this question for the popular models of side information: (a) the labels are observed through a binary symmetric channel with crossover probability \( \alpha \in (0, 0.5) \), and (b) labels are observed through a binary erasure channel with erasure probability \( \epsilon \in (0, 1) \). We also generalize the existing models to include arbitrary side information distributions on finite alphabets. For exact recovery, necessary and sufficient conditions are derived that are tight. More specifically:

- For side information observed through a binary erasure channel with erasure probability \( \epsilon \in (0, 1) \), we show that when \( \log(\epsilon) \) is constant independent of \( n \), or, surprisingly, \( o(\log(n)) \), then side information does not help exact recovery and the phase transition is still \((\sqrt{a} - \sqrt{b})^2 > 2\). When \( \log(\epsilon) \) is \( O(\log(n)) \), then side information helps and the phase transition for exact recovery changes. More precisely, we show that when \( \log(\epsilon) = -\beta \log(n) \) for some \( \beta > 0 \), then: \((\sqrt{a} - \sqrt{b})^2 + 2\beta > 2 \) is necessary and sufficient for exact recovery. We provide our sufficient conditions for two detection algorithms, namely, Maximum Likelihood and an efficient algorithm which uses a partial recovery algorithm from the literature combined with a local improvement procedure that combines both the graph and side information. Also, during the proofs, we needed one lemma from [30], for which we provide an alternative and more compact proof.

- For side information observed through a binary symmetric channel with crossover probability \( \alpha \in (0, 0.5) \), we show that when \( c = \log\left(\frac{1 - \alpha}{\alpha}\right) \) is \( o(\log(n)) \), then side information does not help exact recovery and the phase transition is still \((\sqrt{a} - \sqrt{b})^2 > 2\). When \( \alpha \) is decreasing with \( n \) such that \( c = \beta \log(n) \) for some \( \beta > 0 \), then: for graphs with \( \frac{T(a-b)}{2} < \beta, \beta > 1 \) is necessary and sufficient for exact recovery and for graphs with \( \frac{T(a-b)}{2} > \beta, \eta(a, b, \beta) > 2 \) is necessary and sufficient for exact recovery, where \( \eta(a, b, \beta) = a + b + \beta - \frac{2\epsilon}{T} + \frac{\beta}{T} \log\left(\frac{\gamma + \beta}{\gamma - \beta}\right) \), where \( T = \log\left(\frac{a}{b}\right), \gamma = \sqrt{\beta^2 + abT^2} \). Unlike the partial noiseless side information, we provide the sufficient conditions only for the efficient algorithm.

---

2The exact recovery phase transition without side information is \((\sqrt{a} - \sqrt{b})^2 > 2\) [30].

3These results appear in the Allerton Conference on Communications, Control, and Computing 2017 [37].
We then generalize our results to M-ary side information with finite M. We provide necessary and sufficient conditions and show that they are tight, expect for one special case, by extending our efficient algorithm to M-ary side information. Surprisingly, we show that if for at least one \( m \in \{1, \ldots, M\} \) we have the log-likelihood ratio of this \( m^{th} \) side information to be \( o(\log(n)) \), then side information does not help exact recovery and the phase transition is still \( (\sqrt{a} - \sqrt{b})^2 > 2 \). If for all \( m \in \{1, \ldots, M\} \) we have the log-likelihood ratio of side information to be \( \Omega(\log(n)) \), then we need several conditions to be satisfied which will be specified later in Section [VI].

To illustrate our results, we show in Figures 1, 2 the error exponent, for some values of \( a, b \), for the side information observed through a binary erasure and binary symmetric channels as a function of \( \beta \). From the figures, we can see that the value of \( \beta \) needed for recovery depends on \( a, b \). For the binary erasure channel, for any given \( a > b > 0 \) with \( (\sqrt{a} - \sqrt{b})^2 < 2 \), the needed value of \( \beta \) for recovery, i.e., critical \( \beta \), is \( 1 - \frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + \delta \) for any arbitrary small \( \delta > 0 \). For the binary symmetric channel, for any given \( a > b > 0 \) with \( (\sqrt{a} - \sqrt{b})^2 < 2 \), the value of \( \beta \) needed for recovery can be determined as follows: if \( \eta(a, b, \frac{T(a-b)}{2}) > 2 \), then the critical \( \beta \) is the value of \( \beta \) that makes \( \eta > 2 \), in that case it is less than one. On the other hand, if \( \eta(a, b, \frac{T(a-b)}{2}) < 2 \), then it is \( 1 + \delta \) for arbitrary small \( \delta > 0 \).

Remark 1: For weak recovery, it is easy to see that for \( \alpha \) or \( \epsilon \to 0 \) as \( n \to \infty \) arbitrary slow, side information could help to achieve recovery regardless of the observed graph. However, for exact recovery, we still need a certain rate at which \( \alpha \) or \( \epsilon \to 0 \) to make the side information useful to the maximum likelihood detector. We will provide some intuition about why this is the case in our analysis in the next few sections.

II. System Model

In this paper, we consider the binary symmetric stochastic block model with side information. Let the labels of the two communities be 1 or \(-1\). Our observations are generated as follows: each node, out of \( n \) nodes, is assigned independently and uniformly at random to one of the two communities. After the assignment of nodes, each two nodes are connected with an edge independently with probability \( a \frac{\log(n)}{n} \) if the two nodes belong to the same community and with probability \( b \frac{\log(n)}{n} \), otherwise. Finally, for each node we observe a scalar side information. We consider three models of side information.
Fig. 1. Error exponent of the binary symmetric channel side information as a function of $\beta$.

Fig. 2. Error exponent of binary erasure channel side information as a function of $\beta$. 

$$1 - 0.5\eta(a,b,\beta) \text{ VS } 1 - \beta$$
First, for each node, we independently observe its true label with probability $(1 - \alpha)$ or the negative of its true label with probability $\alpha$, for $\alpha \in (0, 0.5)$. For the second model, for each node, we independently observe its true label with probability $1 - \epsilon$ or 0 with probability $\epsilon$, for $\epsilon \in (0, 1)$. For the third model, we consider $M$-ary side information, where for each node $i$ we independently observe $y_i \in \{u_1, u_2, \cdots, u_m\}$. We denote $\mathbb{P}(y_i = u_m | x_i = \pm 1) = \alpha_{\pm,m}$, and $\sum_{m=1}^{M} \alpha_{+,m} = \sum_{m=1}^{M} \alpha_{-,m} = 1$.

We denote the observed graph by $G = (V, E)$, the vector of nodes’ true assignment by $x^*$, and the vector of nodes’ side information by $y$. The goal is to recover the node assignment $x^*$ from the observation of $(G, y)$.

**Remark 2:** In [30] where the graph is the only observation, exact recovery is defined up to a global flip. This is due to the symmetry of the two communities and hence, recovering the true vector $x^*$ or $-x^*$ would be considered a success. This is not generally true with side information. For example, with side information observed through binary symmetric or binary erasure channels, symmetry is broken and success is only considered when the true vector $x^*$ is recovered.

### III. Binary Symmetric Channel

In this section, we consider side information consisting of the true labels observed through a binary symmetric channel with crossover probability $\alpha \in (0, 0.5)$. We provide tight necessary and sufficient conditions for different cases of $\alpha$. But before we proceed with the proofs we need the following lemmas.

First, we present the ML rule for detecting the communities. It is known that without side information, in the binary symmetric SBM, the ML will find two equally sized communities (of size $\frac{n}{2}$ each) that have the minimum number of edges between them, i.e., minimum cut [30]. So we need to determine the ML rule when it has side information. Denote the first and second communities by A and B, respectively. We will use communities A (1) and B (−1) interchangeably. Let the number of edges inside community $A$ and $B$ be $E(A)$ and $E(B)$, respectively. Also, let the total number of edges in the observed graph be $E_t$. Finally, let the number of $\{i \in A : y_i = 1\}$ and $\{i \in B : y_i = -1\}$ be $J_+(A)$ and $J_-(B)$, respectively. Then, the log-likelihood function can be written as:
\[
\log \left( \mathbb{P}(G, y|x) \right) \stackrel{(a)}{=} \log \left( \mathbb{P}(G|x) \right) + \log \left( \mathbb{P}(y|x) \right) \\
= \log \left( p^{E(A)+E(B)} q^{E_1-E(A)-E(B)} (1-p)^{2(\frac{1}{2})} - E(A) - E(B) (1-q)^{\frac{n^2}{4} - E_1 + E(A) + E(B)} \right) \\
+ \log \left( (1-\alpha)^{J_+(A)+J_-(B)} \alpha^{n-J_+(A)-J_-(B)} \right) \\
\stackrel{(b)}{=} R + T(E(A) + E(B))(1 + o(1)) + c(J_+(A) + J_-(B))
\]  

where \( T = \log \left( \frac{a}{b} \right) \), \( c = \log \left( \frac{1-\alpha}{\alpha} \right) \), (a) holds because \( G, y \) are independent given \( x \) and (b) holds by defining \( R \) to be a constant that contains all the terms that are independent of \( x \) and approximating \( \log \left( \frac{p(1-q)}{q(1-p)} \right) \) by \( (1 + o(1))T \) because \( (1-p), (1-q) \) both approach 1 as \( n \to \infty \).

Therefore, ML rule finds two equally sized communities that have the maximum the number of edges inside the communities weighted by \( T \) plus the number of \( y_i \) that matches the communities’ labels weighted by \( c \).

Using this result, we will now get a sufficient and necessary conditions for the event that ML fails to detect the communities. But we need the following definitions. Recall, we defined the true communities as \( A \) (1) and \( B \) (−1). Moreover, we define the following events.

\[
F = \{ \text{ML fails} \} \\
F_A = \{ \exists i \in A : T(E[i, B] - E[i, A]) - cy_i \geq T \} \\
F_B = \{ \exists j \in B : T(E[j, A] - E[j, B]) + cy_j \geq T \}
\]

where \( E[\cdot, \cdot] \) denotes the number of edges between two sets of nodes. The following two lemmas define lower and upper bounds of the probability of failure of ML.

**Lemma 1:** \( F \) will happen if both \( F_A \) and \( F_B \) happened.

**Proof:**

Define two new communities \( \hat{A} = A \setminus \{i\} \cup \{j\} \) and \( \hat{B} = B \setminus \{j\} \cup \{i\} \). Hence, we need to show that \( \log \left( \mathbb{P}(G, y|\hat{A}, \hat{B}) \right) \geq \log \left( \mathbb{P}(G, y|A, B) \right) \), which implies the failure of ML.

Let \( A_{ij} \sim Bern(q) \) be a random variable representing the existence of the edge between nodes \( i \) and \( j \). Then, using (1):

\[
\log \left( \mathbb{P}(G, y|\hat{A}, \hat{B}) \right) = R + T(E(\hat{A}) + E(\hat{B})) + c(J_+(\hat{A}) + J_-(\hat{B}))
\]
\[ R + T \left( E(A) + E(B) - E[i, A] + E[j, A] - E[j, B] + E[i, B] - 2A_{ij} \right) \]
\[ + c \left( J_+(A) + J_-(B) + J+(j) - J+(i) + J-(i) - J-(j) \right) \]
\[ = R + T \left( E(A) + E(B) \right) + c \left( J_+(A) + J_-(B) \right) - 2TA_{ij} \]
\[ + T \left( E[j, A] - E[j, B] \right) + cy_j + T \left( E[i, B] - E[i, A] \right) - cy_i \]
\[ \geq \log \left( \mathbb{P}(G, y | A, B) \right) + 2T \left( 1 - A_{ij} \right) \]
\[ \geq \log \left( \mathbb{P}(G, y | A, B) \right) \]

where \( a \) holds by the assumption that \( F_A \cap F_B \) happened and \( b \) holds because \( 1 - A_{ij} \geq 0 \) and \( T \geq 0 \). Hence, from the last inequality, we conclude that the ML will not coincide with the true assignment \( A, B \).

\textit{Lemma 2:} Suppose that \( F \) happens. Then, there exists \( k \) and sets \( A_w \subset A \) and \( B_w \subset B \) with \( |A_w| = |B_w| = k \), for \( 1 \leq k \leq \frac{n}{2} \) such that:

\[ T \left( E[B_w, A \setminus A_w] + E[A_w, B \setminus B_w] - E[B_w, B \setminus B_w] - E[A_w, A \setminus A_w] \right) + 2c \left( J_+(B_w) - J_+(A_w) \right) \geq 0 \]

\textit{Proof:} Let \( \hat{A} = A \setminus \{ A_w \} \cup \{ B_w \} \) and \( \hat{B} = B \setminus \{ B_w \} \cup \{ A_w \} \). Since \( F \) happened, then:

\[ \log \left( \mathbb{P}(G, y | \hat{A}, \hat{B}) \right) \geq \log \left( \mathbb{P}(G, y | A, B) \right) \]

and hence:

\[ T \left( E[A \setminus \{ A_w \}, A \setminus \{ A_w \}] + E[B_w, A \setminus \{ A_w \}] + E[B \setminus \{ B_w \}, B \setminus \{ B_w \}] + E[A_w, B \setminus \{ B_w \}] \right) + \]
\[ c \left( J_+(A \setminus \{ A_w \}) + J_+(B_w) + J_-(B \setminus \{ B_w \}) + J_-(A_w) \right) \geq \]
\[ T \left( E[A \setminus \{ A_w \}, A \setminus \{ A_w \}] + E[A_w, A \setminus \{ A_w \}] + E[B \setminus \{ B_w \}, B \setminus \{ B_w \}] + E[B_w, B \setminus \{ B_w \}] \right) + \]
\[ c \left( J_+(A \setminus \{ A_w \}) + J_+(A_w) + J_-(B \setminus \{ B_w \}) + J_-(B_w) \right) \]

By canceling out similar terms, this concludes our proof.

\textit{Remark 3:} In [30], due to symmetry, recovering the true assignment vector \( \mathbf{x}^* \) or \(-\mathbf{x}^* \) is considered a success. Thus, the error event was defined for \( 1 \leq k \leq \frac{n}{2} \). Here, since \( \alpha < 0.5 \),
then side information will break symmetry, and hence, exchanging a subset of nodes of size $k$ is not the same as exchanging a subset of nodes of size $\frac{n}{2} - k$. Thus, we defined our error event for $1 \leq k \leq \frac{n}{2}$. That is true also for side information observed through a binary erasure channel.

A. Necessary Conditions

Let $a > b > 0$ and define $c = \log(\frac{1-a}{a})$ and $\eta(a, b, \beta) = a + b + \beta - \frac{2\gamma}{T} + \frac{\beta}{T} \log(\frac{\gamma + \beta}{\gamma - \beta})$, where $T = \log(\frac{a}{b})$, $\gamma = \sqrt{\beta^2 + abT^2}$.

Theorem 1: For $c = o(\log(n))$, if $(\sqrt{a} - \sqrt{b})^2 < 2$, then ML fails in recovering the communities with probability bounded away from zero. On the other hand, for $c = \beta \log(n)$, for some $\beta > 0$, for graphs with $\frac{T(a-b)}{2} < \beta$, if $\beta < 1$ then ML fails in recovering the communities with probability bounded away from zero, and for graphs with $\frac{T(a-b)}{2} > \beta$, if $\eta(a, b, \beta) < 2$, then ML fails in recovering the communities with probability bounded away from zero.

Remark 4: Note that under the assumption that the number of edges connected to nodes $i, j \in$ community $A$ are independent, the proof would be much simpler. However, they are dependent. To overcome this dependency, we break the event $F_A$ into several events. Such idea was introduced in [30].

Proof:

Note that since $x^*$ is generated uniformly, then ML 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. Let $H$ be a subset of $A$ with $|H| = \frac{n}{\log(n)}$. We define the following events:

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

\[ F_i^H = \{ i \in H : T * E[i, A \setminus H] + cy_i + T + T \frac{\log(n)}{\log \log(n)} \leq T * E[i, B] \} \]

\[ \Delta = \{ \forall i \in H : \Delta_i \text{ is true} \} \]

\[ F^H = \{ \cup_{i \in H} F_i^H \} \]

We will prove this part of the theorem via several lemmas.
Lemma 3: If $\mathbb{P}(F_A) \geq \frac{2}{3}$, then $\mathbb{P}(F) \geq \frac{1}{3}$.

Proof:

If $\mathbb{P}(F_A) \geq \frac{2}{3}$, then by the symmetry of the graph and the side information, $\mathbb{P}(F_B) \geq \frac{2}{3}$ too. Also, by Lemma 1 $F_A \cap F_B \Rightarrow F$. Then, we have:

$$
\mathbb{P}(F) \geq \mathbb{P}(F_A) + \mathbb{P}(F_B) - 1 \geq \frac{2}{3} + \frac{2}{3} - 1 = \frac{1}{3}
$$

Lemma 4: If $\mathbb{P}(F^H) \geq \frac{9}{10}$ and $\mathbb{P}(\triangle) \geq \frac{9}{10}$, then $\mathbb{P}(F) \geq \frac{1}{3}$.

Proof: It is easy to see that $\triangle \cap F^H \Rightarrow F_A$. Hence,

$$
\mathbb{P}(F_A) \geq \mathbb{P}(F^H) + \mathbb{P}(\triangle) - 1 \geq \frac{9}{10} + \frac{9}{10} - 1 = \frac{2}{3}
$$

which together with Lemma 3 concludes the proof.

Based on the above Lemmas, our proof boils down to proving when $\mathbb{P}(F^H) \geq \frac{9}{10}$ and $\mathbb{P}(\triangle) \geq \frac{9}{10}$ are true.

Lemma 5: $\mathbb{P}(\triangle) \geq \frac{9}{10}$ for sufficiently large $n$.

Proof:

Let $W_i \sim Bern(p)$. Then, we have:

$$
\mathbb{P}(\triangle^c) = \mathbb{P}\left(\sum_{j=1}^{i-1} W_j + \sum_{j=i+1}^{n} W_j \geq \frac{\log(n)}{\log(\log(n))}\right)
$$

$$
\leq \mathbb{P}\left(\sum_{j=1}^{n} W_j \geq \frac{\log(n)}{\log(\log(n))}\right)
$$

$$
\leq \left(\frac{1}{e a \log(\log(n))}\right)^{-\frac{\log(n)}{\log(\log(n))}}
$$

where (a) holds from a multiplicative form of Chernoff bound, which states that for a sequence of $n$ i.i.d random variables $X_i$, $\mathbb{P}(\sum_{i=1}^{n} X_i \geq t\mu) \leq (\frac{e}{t})^{-t\mu}$, where $\mu = n\mathbb{E}[X]$. Thus, we get by union bound:
\[
\Pr(\triangle_i) \geq 1 - \frac{n}{\log^3(n)} \left( \frac{1}{e} a \log(\log(n)) \right)^{\log(\log(n))} \\
= 1 - e^{\log(n) - 3 \log(\log(n))} \left( \frac{\log(n) \log(\log(n))}{\log(\log(n)) - \log(\log(\log(n)))} \right)^{3 \log(\log(n)) - \log(\log(\log(n)))} \\
= 1 - e^{-2 \log(n) + o(\log(n))}
\]

Thus, for sufficiently large \( n \), \( \Pr(\triangle) \geq \frac{9}{10} \).

We will show when \( \Pr(F^H) \geq \frac{9}{10} \) is true in two steps. First we will show that if \( \Pr(F^H_i) > \frac{\log^3(n)}{n} \log(10) \), then \( \Pr(F^H) \geq \frac{9}{10} \). Then, we will show when \( \Pr(F^H_i) > \frac{\log^3(n)}{n} \log(10) \) is true.

**Lemma 6:** If \( \Pr(F^H_i) > \frac{\log^3(n)}{n} \log(10) \), then \( \Pr(F^H) \geq \frac{9}{10} \).

**Proof:**

\[
\Pr(F^H) = \Pr(\bigcup_{i \in H} F^H_i) \overset{(a)}{=} 1 - \Pr(\bigcap_{i \in H} (F^H_i)^c) = 1 - \left( 1 - \Pr(F^H_i) \right)^{\log^3(n)}
\]

where (a) holds because \( F^H_i \) are i.i.d random variables. So, for \( \Pr(F^H) \geq \frac{9}{10} \) to be true, we need \( (1 - \Pr(F^H_i))^{\frac{n}{\log^3(n)}} \leq \frac{1}{10} \). If \( \Pr(F^H_i) \) is not \( o(1) \), then clearly the inequality is true. On the other hand, if \( \Pr(F^H_i) \) is \( o(1) \), then:

\[
\lim_{n \to \infty} (1 - \Pr(F^H_i))^{\frac{n}{\log^3(n)}} = \lim_{n \to \infty} \left( 1 - \Pr(F^H_i) \right)^{(\Pr(F^H_i)(\frac{1}{\Pr(F^H_i)})^{\frac{n}{\log^3(n)}})} = \lim_{n \to \infty} e^{\log^3(n) \Pr(F^H)}
\]

So clearly, if \( \Pr(F^H_i) > \frac{\log^3(n)}{n} \log(10) \), then \( (1 - \Pr(F^H_i))^{\frac{n}{\log^3(n)}} \leq \frac{1}{10} \), which implies that \( \Pr(F^H) \geq \frac{9}{10} \).
The final step is to show when $\mathbb{P}(F^H_i) > \frac{\log^3(n)}{n} \log(10)$ is true. Recall that $c = \log(\frac{1-\alpha}{\alpha})$. Lemma 12 in the Appendix shows that $\mathbb{P}(F^H_i) > \frac{\log^3(n)}{n} \log(10)$ for sufficiently large $n$ if one of the following is satisfied:

- When $c = o(\log(n))$: if $(\sqrt{a} - \sqrt{b})^2 < 2$.
- When $c = \beta \log(n)$, $\beta > 0$: for any $a, b, \beta$, if $\eta(a, b, \beta) < 2$.
- When $c = \beta \log(n)$, $\beta > 0$: for $a, b, \beta$, if $\eta(a, b, \beta) > 2$ and $\beta < 1$.

\[ \text{B. Sufficient Conditions} \]

In this section we provide an efficient algorithm that achieves exact recovery down to the necessary conditions obtained in Section III-A. The algorithm is decomposed into two stages. In the first stage we use the algorithm proposed in [24] on the graph alone. Such algorithm is known to achieve weak recovery. Then, we modify its outcome using the graph and side information. This idea was used before in [30], [31], but it is the first time to be used with side information, to the best of our knowledge.

Note that based on Lemma 2 we can try to provide the sufficient conditions by bounding the failure event of maximum likelihood as $\mathbb{P}(F) \leq \sum_{k=1}^{n} (\frac{n}{k})^2 P_n^{(k)}$, where $1 \leq k \leq \frac{n}{2}$ and $P_n^{(k)} := \mathbb{P}\left( T \sum_{i=1}^{m} (Z_i - W_i) + c \sum_{i=1}^{k} (S_i - R_i) \geq 0 \right)$, $m = 2k(\frac{n}{2} - k)$, $W_i \sim \text{Bern}(p)$, $Z_i \sim \text{Bern}(q)$, $S_i \sim \text{Bern}(\alpha)$ and $R_i \sim \text{Bern}(1-\alpha)$. However, it is difficult to deal with a weighted sum of four independent Binomial random variables with different number of trials and different probability of success. The algorithm we provide overcome this difficulty because the second stage that exploits side information works on a node by node basis, and hence, deal with the side information as a Bernoulli random variable instead of Binomial.

The algorithm starts by splitting the information in the observed graph into two parts. The first part to be used by the partial recovery algorithm and the second part to be used for local modification with the observed side information. In order to make the two steps independent, we follow a similar idea as the one presented in [30]. First assume we have a complete graph on $n$ nodes. We split its edges into two random graphs each with $n$ nodes. We call these graphs $H_1$ and $H_2$. More precisely, we generate $H_1$ by keeping edges in the complete graph with probability $\frac{D}{\log(n)}$. $H_2$ will be the complement of $H_1$. Then, we define $G_1$ and $G_2$ as sub-graphs of the observed graph $G$ as: $G_1 = G \cap H_1$ and $G_2 = G \cap H_2$. 

August 17, 2017 DRAFT
Next, we apply the partial recovery algorithm presented in [24] on $G_1$. Notice that $G_1$ is a graph obtained from a binary symmetric SBM with connectivity parameters $(\frac{D a_n}{n}, \frac{D b_n}{n})$. Thus, the partial recovery algorithm is guaranteed to return two communities $A'$ and $B'$ such that $A'$, $B'$ coincide with the true community assignment $A$, $B$ on at least $(1 - \delta(D))n$ nodes, with $\delta(D) \to 0$ as $D \to \infty$ [24].

The last step is the local modifications step. Now that we have $G_2$, the side information $y$, $A'$ and $B'$, we locally modify the community assignment as follows: for a node $i \in A'$, we 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 $\frac{c}{T} y_i$ and for node $j \in B'$, we flip its membership if the number of edges between $j$ and $A'$ is greater than or equal the number of edges between $j$ and $B'$ minus $\frac{c}{T} y_i$. If the the number of flips in each cluster is not the same, keep the clusters unchanged.

**Theorem 2:** For $c = o(\log(n))$, if $(\sqrt{a} - \sqrt{b})^2 > 2$, then, there exists large enough $D$ such that, with high probability, the algorithm described above will successfully recover the communities from the observed graph and side information. On the other hand, for $c = \beta \log(n)$, for some $\beta > 0$, then the algorithm described above will successfully recover the communities from the observed graph and side information, if:

\[
\begin{align*}
\eta(a, b, \beta) > 2, \quad & \text{for } a, b, \beta : \beta < \frac{T(a-b)}{2} \\
\beta > 1, \quad & \text{for } a, b, \beta : \beta \geq \frac{T(a-b)}{2}
\end{align*}
\]

**Proof:**

Our goal here is to upper bound the error of misclassifying one node, then use a simple union bound over all nodes. If we assume for now (to be changed later) that $H_2$ is a complete graph, then we have four different cases for which a node could be misclassified according to our algorithm. Two cases could happen when the node $i \in A'$ and another two when the node $i \in B'$. The first two cases are displayed in Figures 3, 4. The remaining cases are similar.

Now under the assumption that $H_2$ is complete, then the four cases for error can be written as follows. Let $W_i \sim Bern(p)$, $Z_i \sim Bern(q)$ and $y_i \in \{1, -1\}$ with probabilities $(1 - \alpha), \alpha$, respectively. For simplicity, we will write $\delta$ instead of $\delta(D)$. Then, we have the following:

\[P_e = \mathbb{P}(\text{node } i \text{ is mislabeled})\]
Recall, we assumed that $H_2$ is a complete graph. However, using [Lemma 14 in [30]], it can be shown that the degree of any node in $H_2$ is at least $n(1 - 2D\log(n))$ with high probability. Hence, we will loosely upper bound (3) be removing $2D\log(n)$ from the first two terms on the right hand side. Thus, we get:

$$P_e \leq \mathbb{P}\left( \sum_{i=1}^{(1-\delta)n} Z_i + \sum_{i=1}^{(1-\delta)n} W_i \geq \sum_{i=1}^{(1-\delta)n} W_i + \sum_{i=1}^{(1-\delta)n} Z_i + \frac{c}{T}y_i \right)$$  \hspace{1cm} (3)

Now Lemma 17 shows that if $c = o(\log(n))$, then (4) can be upper bounded by (5) and if $c = \beta \log(n)$, for some $\beta > 0$, then for $a, b, \beta : \beta < \frac{T(a-b)}{2}$, (4) can be upper bounded by (6) and for $a, b, \beta : \beta > \frac{T(a-b)}{2}$, (4) can be upper bounded by (7).

$$P_e \leq n^{-\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + o(1)} + n^{-(1+\Omega(1))}$$  \hspace{1cm} (5)

$$P_e \leq (2 - \alpha)n^{-\frac{1}{2}\eta(a,b,\beta) + o(1)} + n^{-(1+\Omega(1))}$$  \hspace{1cm} (6)

$$P_e \leq (1 - \alpha)n^{-\frac{1}{2}\eta(a,b,\beta) + o(1)} + n^{-\beta} + n^{-(1+\Omega(1))}$$  \hspace{1cm} (7)
Hence, using a union bound (we will do the analysis for (6) only, (5) and (7) follow similarly), we get:

$$\mathbb{P}(\exists \text{ a misclassified node}) \leq (2 - \alpha)n^{1-\frac{1}{2}d(a,b,\beta)+o(1)} + n^{-\Omega(1)}$$

(8)

which shows that if $\eta(a,b,\beta) > 2$, then the algorithm described above will successfully recover the communities from the observed graph and side information with high probability. Note that for (7), Lemma 15 shows that for all $a$ and $b$, $\eta > 2$ if $\beta > 1$, and hence, for (7), $\beta > 1$ is sufficient.

IV. Binary Erasure Channel

In this section, we consider side information consisting of the true labels observed through a binary erasure channel with parameter $\epsilon \in (0, 1)$. We provide necessary and sufficient conditions that are tight for different cases for $\epsilon$. Unlike Section III, we provide the sufficient conditions for the Maximum Likelihood detector (known as information theoretic upper bound [30]), then we propose an efficient algorithm that succeeds all the way down to the threshold. But as we did in Section III before we proceed with the proofs we need the following lemmas.

First, we present the ML rule for detecting the communities. Note that since here the side information is either the true label or zero, thus, the maximum likelihood rule is the same as the rule without side information. In other words, the maximum likelihood rule is to maximize the number of edges inside the communities. However, the only difference is the feasible set. Instead of having a feasible set of all possible vectors of length $n \in \{\pm 1\}^n$ such that $\sum_{i=1}^{n} x_i = 0$, the new feasible set would be all vectors of length $n \in \{\pm 1\}^n$ such that $\sum_{i=1}^{n} x_i = 0$ and all these vectors match the non erased bits in the observed side information $y$. The reason why this is true, is because for any vector $x$ that does not match the observed $y$, then $\mathbb{P}(y|x) = 0$, and hence, the maximum likelihood will never choose such vector.

Now based on the above observation, we will now get a sufficient and necessary conditions for the event that ML fails to detect the communities. Recall that $F$ denotes the event of failure of maximum likelihood and $E[\cdot,\cdot]$ denotes the number of edges between two sets of nodes. We define the following events
\[ F_A = \{ \exists i \in A : (E[i, B] - E[i, A]) \geq 1 \text{ and } y_i = 0 \} \]
\[ F_B = \{ \exists j \in B : (E[j, A] - E[j, B]) \geq 1 \text{ and } y_j = 0 \} \]

**Lemma 7:** \( F \) will happen if both \( F_A \) and \( F_B \) happened.

**Proof:**

Define two new communities \( \hat{A} = A \{i\} \cup \{j\} \) and \( \hat{B} = B \{j\} \cup \{i\} \). Hence, we need to show that \( \log(\mathbb{P}(G, y | \hat{A}, \hat{B})) \geq \log(\mathbb{P}(G, y | A, B)) \), which implies the failure of ML.

Let \( A_{ij} \sim \text{Bern}(q) \) be a random variable representing the existence of the edge between nodes \( i \) and \( j \). Then, using the maximum likelihood rule, we have:

\[
\log(\mathbb{P}(G, y | \hat{A}, \hat{B})) = R + T(E(\hat{A}) + E(\hat{B})) + \log(\mathbb{P}(y | A, B))
\]
\[= R + T(E(A) + E(B)) + T(E[j, A] - E[i, A] - E[j, B] + E[i, B] - 2A_{ij})
\]
\[+ \log(\mathbb{P}(y | A, B))
\]
\[\geq \log(\mathbb{P}(G, y | A, B)) + 2T(1 - A_{ij}) \geq \log(\mathbb{P}(G, y | A, B))
\]

where \((a)\) holds by the assumption that \( F_A \cap F_B \) happened and \((b)\) holds because \((1 - A_{ij}) \geq 0\). Hence, from the last inequality, we conclude that the ML will not coincide with the true assignment \( A, B \).

**Lemma 8:** Suppose that maximum likelihood fails. Then, there exists \( k \) and sets \( A_w \subset A \) and \( B_w \subset B \) with \(|A_w| = |B_w| = k\), for \( 1 \leq k \leq \frac{n}{2} \) such that \((9)\) is true and the side information \( y_i = 0 \forall i \in A_w, B_w \).

\[ E[B_w, A \{A_w\}] + E[A_w, B \{B_w\}] - E[B_w, B \{B_w\}] - E[A_w, A \{A_w\}] \geq 0 \tag{9} \]

**Proof:**

Let \( \hat{A} = A \{A_w\} \cup \{B_w\} \) and \( \hat{B} = B \{B_w\} \cup \{A_w\} \). Since maximum likelihood is assumed to have failed, then: \( \log(\mathbb{P}(G, y | \hat{A}, \hat{B})) \geq \log(\mathbb{P}(G, y | A, B)) \). First, note that any node to be exchanged, has got to have erased side information, otherwise, it will never be exchanged. Also, note that conditioned on the community assignments, \( G, y \) are independent, and since all the
feasible community assignment vectors, \( x \), have the same \( \mathbb{P}(y|x) \), therefore, we are only left with \( \mathbb{P}(G|x) \) which gives (9).

\[ \mathbb{P}(G|x) \]

A. Necessary Conditions

**Theorem 3:** For \( \log(\varepsilon) = o(\log(n)) \), if \( (\sqrt{a} - \sqrt{b})^2 < 2 \), then ML fails in recovering the communities with probability bounded away from zero. On the other hand, for \( \log(\varepsilon) = -\beta \log(n) \), for some \( \beta > 0 \), if \( \frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + \beta < 1 \), then ML fails in recovering the communities with probability bounded away from zero.

**Proof:**

We will need some definitions as in Section [III-A]. Some of these definitions are the same but we include them here for completeness. Let \( H \) be a subset of \( A \) with \( |H| = \frac{n}{\log^3(n)} \). We define the following events:

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

\[
F_i^H = \{i \in H : y_i = 0 \text{ and } E[i, A\setminus H] + 1 + \frac{\log(n)}{\log \log(n)} \leq E[i, B]\}
\]

\[
\Delta = \{\forall i \in H : \Delta_i \text{ is true}\}
\]

\[
F^H = \{\cup_{i \in H} F_i^H\}
\]

We will prove this part of the theorem via several lemmas. Note that Lemmas 3, 4, 5, 6 extend directly to our case here using the above definitions. To complete the proof, we need to show when \( \mathbb{P}(F_i^H) > \frac{\log^3(n)}{n} \log(10) \) is true. This is proven in Lemma 13 in the Appendix, which shows that \( \mathbb{P}(F_i^H) > \frac{\log^3(n)}{n} \log(10) \) for sufficiently large \( n \) if one of the following is satisfied:

- When \( \log(\varepsilon) = o(\log(n)) \): if \( (\sqrt{a} - \sqrt{b})^2 < 2 \).
- When \( \log(\varepsilon) = -\beta \log(n) \), \( \beta > 0 \): for any \( a, b, \beta \), if \( \frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + \beta < 1 \).
B. Sufficient Conditions

Unlike Section [III], we provide our sufficient conditions for two detectors, namely, Maximum Likelihood and an efficient algorithm which uses a partial recovery algorithm from the literature combined with a local improvement procedure.

1) Maximum Likelihood:

**Theorem 4:** For $\log(\epsilon) = o(\log(n))$, if $(\sqrt{a} - \sqrt{b})^2 > 2$, then the ML detector exactly recovers the communities, with high probability. On the other hand, for $\log(\epsilon) = -\beta \log(n)$, for some $\beta > 0$, if $(\sqrt{a} - \sqrt{b})^2 + 2\beta > 2$, then the ML detector exactly recovers the communities, with high probability.

**Proof:**

We will use Lemma 8, from which we can define the following event:

$$P_n^{(k)} := \mathbb{P}\left(\sum_{i=1}^{m} (Z_i - W_i) \geq 0\right)$$

where $m = 2k(\frac{n}{2} - k)$, $W_i \sim Bern(p)$ and $Z_i \sim Bern(q)$.

Then, by a simple union bound we have:

$$\mathbb{P}(F) \leq \sum_{k=1}^{\frac{n}{2}} e^{2k \left(\frac{n}{2k}\right)^2 P_n^{(k)}}$$  \hspace{1cm} (10)

To bound $P_n^{(k)}$, we can use [Lemma 8 in [30]]. Lemma 16 in the Appendix shows an alternative proof which is easier and more compact. Using Lemma 16, we have:

$$P_n^{(k)} \leq e^{-2n \log(n) \frac{1}{2} (\sqrt{a} - \sqrt{b})^2}$$  \hspace{1cm} (11)

Using the fact that $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$, we can bound (10) as follows:

$$\sum_{k=1}^{\frac{n}{2}} e^{2k \left(\frac{n}{2k}\right)^2 P_n^{(k)}} \leq e^n + \sum_{k=1}^{\frac{n}{2}-1} e^{2k \left(\log(n) - 2 \log(2k) + 2\right) - 4k \left(\frac{1}{2} - \frac{1}{n}\right) \left(\frac{1}{2} (\sqrt{a} - \sqrt{b})^2 - \log(\epsilon) \right) \log(n)}$$

$$\leq o(1) + \sum_{k=1}^{\frac{n}{2}-1} e^{k \left(\log(n) - 2 \log(2k) + 2\right) - 4k \left(\frac{1}{2} - \frac{1}{n}\right) \left(\frac{1}{2} (\sqrt{a} - \sqrt{b})^2 - \log(\epsilon) \right) \log(n)}$$
where \((a)\) holds because \(\log(\epsilon) < 0\) and \(\sup_k 2\left(\frac{1}{2} - \frac{k}{n}\right) = 1\)

Now, we divide our analysis into two cases.

- when \(\log(\epsilon) = -\beta \log(n)\), for some \(\beta > 0\). In this case, we have:

\[
\begin{align*}
\text{O}(1) + \sum_{k=1}^{\frac{n}{2}} e^{2k\left(\frac{n}{2k}\right)^2} P_n^{(k)} & \leq \text{O}(1) + \sum_{k=1}^{\frac{n}{2}} e^{k\left(2\log(n) - 2\log(2k) + 2\right)} e^{-4k\left(\frac{1}{2} - \frac{k}{n}\right)\left(\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + \beta\right)} \log(n) \\
& \overset{(a)}{\leq} \text{O}(1) + 2 \sum_{k=1}^{\frac{n}{2}} e^{k\left(2\log(n) - 2\log(2k) + 2\right)} e^{-4k\left(\frac{1}{2} - \frac{k}{n}\right)\left(\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + \beta\right)} \log(n) \\
& \overset{(b)}{\leq} \text{O}(1) + 2 \sum_{k=1}^{\frac{n}{2}} e^{k\left(2\log(n) - 2\log(2k) + 2 - 4\left(\frac{1}{2} - \frac{k}{n}\right)(1+\delta) \log(n)\right)} \\
& \overset{(c)}{\leq} \text{O}(1) + 2 \sum_{k=1}^{\frac{n}{2}} e^{k\left(2 - 2\log(2k) + 4\frac{k}{n}\log(n) - \delta \log(n)\right)} \\
& \overset{(d)}{\leq} \text{O}(1) + 2n^{-\delta} \sum_{k=1}^{\frac{n}{2}} e^{-2k\left(-1 + \log(2k) - 2\frac{k}{n}\log(n)\right)} 
\end{align*}
\]

where \((a)\) holds because \(\left(\frac{n}{k}\right) = \left(\frac{\frac{n}{2}}{\frac{k}{2}}\right)\) and the range of \(m\) is the same (only decreasing not increasing) for \(1 \leq k \leq \frac{n}{4}\) and \(\frac{n}{4} \leq k \leq \frac{n}{2} - 1\), \((b)\) holds by assuming that \(\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + \beta > 1 + \delta\) for \(\delta > 0\) and \((c)\), \((d)\) hold because \(1 \leq k \leq \frac{n}{4}\).

Now notice that for sufficiently large \(n\), \(1 \leq k \leq \frac{n}{4}\) we have

\[
\log(2k) - \frac{2k}{n}\log(n) \geq \frac{1}{3}\log(2k)
\]

Hence for sufficiently large \(n\),

\[
\mathbb{P}(F) \leq 2n^{-\delta} \sum_{k=1}^{\frac{n}{4}} e^{-\frac{2}{3}k(\log(2k) - 3)} + o(1)
\]

which, together with the observation that \(\sum_{k=1}^{\frac{n}{4}} e^{-\frac{2}{3}k(\log(2k) - 3)}\) is \(O(1)\) concludes the proof of the second case of the theorem.
• when \( \log(\epsilon) = o(\log(n)) \). In this case, we have:

\[
o(1) + \sum_{k=1}^{\frac{k}{2}-1} \epsilon^{2k} \left( \frac{n}{2} \right)^2 P_n^{(k)} \leq o(1) + \sum_{k=1}^{\frac{k}{2}-1} \epsilon^{-k \left( 2 \log(n) - 2 \log(2k) + 2 \right)} e^{-4k \left( \frac{1}{2} - \frac{1}{\theta} \right) \left( \frac{1}{2} (\sqrt{a} - \sqrt{b})^2 + o(1) \right)} \log(n)
\]

\[
\leq Cn^{-\delta} + o(1)
\]

where \((a)\) holds as the same analysis done in the first case, just assume that \( \frac{1}{2} (\sqrt{a} - \sqrt{b})^2 > 1 + \delta \) for \( \delta > 0 \) instead of \( \frac{1}{2} (\sqrt{a} - \sqrt{b})^2 + \beta > 1 + \delta \). This concludes the proof of the first case of the theorem.

2) Efficient Algorithm: In this section we provide an efficient algorithm that achieves exact recovery down to the necessary conditions obtained in Section IV-A. The algorithm is decomposed into two stages. In the first stage we use the algorithm proposed in [24] on the graph alone. Such algorithm is known to achieve partial recovery. Then, we modify its outcome using the graph and side information. The first stage of the algorithm is the same as Section III-B. The second stage, the local modification is however different.

After the first stage, we have \( G_2 \), the side information \( y, A' \) and \( B' \). We locally modify the community assignment as follows: for any node \( i \), we flip its membership if \( y_i \neq 0 \) and does not match the assignment node \( i \) got from \( A' \) and \( B' \), or if \( y_i = 0 \) and the number of edges between \( i \) and the opposite community is greater than or equal the number of of edges between \( i \) and its own community. If the the number of flips in each cluster is not the same, keep the clusters unchanged.

Theorem 5: If \( \log(\epsilon) = o(\log(n)) \), then if \( (\sqrt{a} - \sqrt{b})^2 > 2 \), then, there exists large enough \( D \) such that, with high probability, the algorithm described above will successfully recover the communities from the observed graph and side information. On the other hand, if \( \log(\epsilon) = -\beta \log(n) \), for some \( \beta > 0 \), then the algorithm described above will successfully recover the communities from the observed graph and side information, if \( (\sqrt{a} - \sqrt{b})^2 + 2\beta > 2 \).

Proof:

Recall we defined \( P_c = \mathbb{P}(\text{node } i \text{ to be misclassified}) \). Following the same analysis as in the proof of Lemma 2, we get:
\[
P_e \leq \epsilon \mathbb{P}
\left( \sum_{i=1}^{(1-\delta)\frac{n}{2}} Z_i + \sum_{i=1}^{\delta \frac{n}{2}} W_i \geq \sum_{i=1}^{(1-\delta)\frac{n}{2}} W_i + \sum_{i=1}^{\delta \frac{n}{2}} Z_i \right)
\]

(15)

Now using Lemma 17 with \( c = 0 \) instead of \( c = o(\log(n)) \), then (4) can be upper bounded as:

\[
P_e \leq \epsilon n^{-\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + o(1)} + n^{-1+\Omega(1)}
\]

(16)

Following the analysis in Section IV-A, we divide the analysis into two cases:

- When \( \log(\epsilon) = -\beta \log(n) \), for some \( \beta > 0 \). In this case, we have:

\[
P_e \leq \epsilon n^{-\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 - \beta} + n^{-1+\Omega(1)}
\]

(17)

- When \( \log(\epsilon) = o(\log(n)) \). In this case, we have:

\[
P_e \leq \epsilon n^{-\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + o(1)} + n^{-1+\Omega(1)}
\]

(18)

Finally, using a union bound (we will do the analysis only for (17), (18) follows similarly), we get:

\[
\mathbb{P}(\exists \text{ a misclassified node}) \leq n^{1-(\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + \beta)} + n^{-\Omega(1)}
\]

(19)

which shows that if \( \frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + \beta > 1 \), then the algorithm described above will successfully recover the communities from the observed graph and side information.

V. M-ARY SIDE INFORMATION

In this section, we generalize our results to M-ary side information with finite M. More precisely, for each node \( i \) we independently observe \( y_i \in \{u_1, u_2, \cdots, u_m\} \). We denote \( \mathbb{P}(y_i = u_m| x_i = 1) = \alpha_{+,m} \) and \( \mathbb{P}(y_i = u_m| x_i = -1) = \alpha_{-,m} \), for \( \alpha_{+,m} > 0 \), \( \alpha_{-,m} > 0 \) and \( \sum_{m=1}^{M} \alpha_{+,m} = \sum_{m=1}^{M} \alpha_{-,m} = 1 \). Moreover, we define \( h_m = \log(\frac{\alpha_{+,m}}{\alpha_{-,m}}) \). We provide necessary conditions for exact recovery for different cases for \( h_m \) and propose an efficient algorithm that succeeds all the way down to the threshold. We emphasize that we studied the binary symmetric
and binary erasure models before the M-ary side information because their proofs are easier to follow and also for one case (that was proved to be sufficient and necessary for the binary symmetric model) we only prove that it is sufficient for the M-ary case but we could not prove its necessity. The proofs in this section are similar to the proof of the Theorems in Sections III, IV. Thus, we only provide proof sketches below.

Let $T = \log(\frac{a}{b})$ and recall that $F$ denotes the event of failure of maximum likelihood and $E[\cdot, \cdot]$ denotes the number of edges between two sets of nodes. Moreover, for a node $i$, we define $h_i = h_m$, if $y_i = u_m$. Then, we define the following events

$$F_A = \{\exists i \in A : T(E[i, B] - E[i, A]) - h_i \geq T\}$$

$$F_B = \{\exists j \in B : T(E[j, A] - E[j, B]) + h_j \geq T\}$$

**Lemma 9**: $F$ will happen if both $F_A$ and $F_B$ happened.

**Proof**:

Define two new communities $\hat{A} = A\{i\} \cup \{j\}$ and $\hat{B} = B\{j\} \cup \{i\}$. Hence, we need to show that $\log (\mathbb{P}(G, y|\hat{A}, \hat{B})) \geq \log (\mathbb{P}(G, y|A, B))$, which implies the failure of ML. Let the number of $i \in A : y_i = u_m$ and $i \in B : y_i = u_m$ be $J_{um}(A)$ and $J_{um}(B)$, respectively.

Using the maximum likelihood rule, we have:

$$\log (\mathbb{P}(G, y|\hat{A}, \hat{B})) = \log (\mathbb{P}(G|\hat{A}, \hat{B})) + \log (\mathbb{P}(y|\hat{A}, \hat{B}))$$

$$= R + T(E(A) + E(B) - E[i, A] + E[j, A] - E[j, B] + E[i, B] - 2A_{ij}) +$$

$$\sum_{m=1}^{M} J_{um}(A) \log(\alpha_{+,m}) + J_{um}(B) \log(\alpha_{-,m}) + \sum_{m=1}^{M} (J_{um}(j) - J_{um}(i))h_m$$

$$= \log (\mathbb{P}(G, y|A, B)) + T(-E[i, A] + E[j, A] - E[j, B] + E[i, B] - 2A_{ij})$$

$$+ h_j - h_i$$

$$\geq \log (\mathbb{P}(G, y|A, B)) + 2T(1 - A_{ij}) \geq \log (\mathbb{P}(G, y|A, B))$$

where $R$ is a constant representing all terms that are independent of $x$, $(a)$ holds by the assumption that $F_A \cap F_B$ happened and $(b)$ holds because $(1 - A_{ij}) \geq 0$. Hence, from the
last inequality, we conclude that the ML will not coincide with the true assignment $A, B$. ■

**Theorem 6:** If for at least one $m$, $\alpha_{+,m} \neq \alpha_{-,m}$ and $h_m = o(\log(n))$, then $(\sqrt{a} - \sqrt{b})^2 > 2$ is necessary and sufficient for exact recovery. On the other hand, if for at least one $m$, we have one of the following conditions:

$$\begin{cases} 
\alpha_{+,m} = \alpha_{-,m} = -\beta \log(n), \beta > 0 \\
h_m = \beta_1 \log(n), |\beta_1| < T\frac{(a-b)}{2}, \log(\alpha_{\text{sgn}(\beta_1),m}) = o(\log(n)) \\
h_m = \beta_2 \log(n), |\beta_2| < T\frac{(a-b)}{2}, \log(\alpha_{\text{sgn}(\beta_2),m}) = -\beta'_2 \log(n), \beta'_2 > 0
\end{cases}$$

then $\min_m \left( (\sqrt{a} - \sqrt{b})^2 + 2\beta, \eta(a, b, |\beta_1|), \eta(a, b, |\beta_2|) + 2\beta'_2 \right) > 2$ is necessary and sufficient for exact recovery, where $\eta$ is defined in section [III-A]

### A. Necessary Conditions

In this section we provide necessary conditions for exact recovery in the binary symmetric SBM with M-ary side information.

**Proof:** [Proof of converse part of Theorem 6]

Note that unlike Sections [III] [IV] the side information might not be symmetric. Hence, we need to define the events of Section [III-A] for both communities $A$ and $B$. Let $H_1$ and $H_2$ be subsets of $A$ and $B$, respectively, with $|H_1| = |H_2| = \frac{n}{\log^\eta(n)}$. We define the following events:

$$\triangle^1_i = \{ i \in H_1 : E[i, H_1] \leq \frac{\log(n)}{\log \log(n)} \}$$
$$F^H_1 = \{ i \in H_1 : TE[i, A \setminus H_1] + h_i + T + T\frac{\log(n)}{\log \log(n)} \leq TE[i, B] \}$$
$$\triangle_1 = \{ \forall i \in H_1 : \triangle^1_i \text{ is true} \}$$
$$F^H_1 = \{ \cup_{i \in H_1} F^H_1 \}$$

$$\triangle^2_i = \{ i \in H_2 : E[i, H_2] \leq \frac{\log(n)}{\log \log(n)} \}$$
$$F^H_2 = \{ i \in H_2 : TE[i, B \setminus H_2] - h_i + T + T\frac{\log(n)}{\log \log(n)} \leq TE[i, A] \}$$
$$\triangle_2 = \{ \forall i \in H_2 : \triangle^2_i \text{ is true} \}$$
\[ F_{H^2} = \left\{ \bigcup_{i \in H_2} F_i^{H_2} \right\} \]

Note that \( h_i \) is distributed according to \( \alpha_{+, m} \) and \( \alpha_{-, m} \) if node \( i \in A, B \), respectively. We will prove this part of the theorem via several lemmas. Note that Lemmas 3, 4, 5, 6 extend directly to our case here using the above definitions for both communities \( A \) and \( B \). To complete the proof, we need to show show when \( P(F_i^{H_1}) > \frac{\log^3(n)}{n} \log(10) \) and \( P(F_i^{H_2}) > \frac{\log^3(n)}{n} \log(10) \) are true.

**Lemma 10:** Both \( P(F_i^{H_1}) \) and \( P(F_i^{H_2}) \) are greater than \( \frac{\log^3(n)}{n} \log(10) \) for sufficiently large \( n \) if one of the following is true:

\[
\begin{align*}
&\text{For } m : \alpha_{+, m} \neq \alpha_{-, m}, h_m = o(\log(n)), \text{ if } (\sqrt{a} - \sqrt{b})^2 < 2 \\
&\text{For } m : \log(\alpha_{+, m}) = \log(\alpha_{-, m}) = -\beta \log(n), \beta > 0, \text{ if } (\sqrt{a} - \sqrt{b})^2 + 2\beta < 2 \\
&\text{For } m : h_m = \beta \log(n), |\beta| < T \frac{(a-b)}{2}, \log(\alpha_{sgn,m}) = o(\log(n)), \text{ if } \eta(a,b,|\beta|) < 2 \\
&\text{For } m : h_m = \beta \log(n), |\beta| < T \frac{(a-b)}{2}, \log(\alpha_{sgn,m}) = -\beta' \log(n), \beta' > 0, \text{ if } \eta(a,b,|\beta|) + 2\beta' < 2 \\
\end{align*}
\]

**Proof:** Let \( W_i \sim Bern(p) \), \( Z_i \sim Bern(q) \) and define \( l = \frac{n}{2} \) and \( \Gamma(t) = \log(E_X[e^{\epsilon x}]) \) for a random variable \( X \). Then, we have the following:

\[
P(F_i^{H_1}) = \mathbb{P}\left( \sum_{i=1}^{n} (Z_i) - \sum_{i=1}^{n} \left( W_i \right) \geq \frac{h_i}{T} + 1 + \frac{\log(n)}{\log(\log(n))} \right)
\]
\[
\geq \sum_{m=1}^{M} \alpha_{+, m} \mathbb{P}\left( \sum_{i=1}^{n} [Z_i - W_i] \geq \frac{h_m}{T} + 1 + \frac{\log(n)}{\log(\log(n))} \right)
\]
\[
\geq \sum_{m=1}^{M} \alpha_{+, m} e^{-l(t_m^* a_m - \Gamma(t_m^*) + |t_m^*| \delta)} (1 - o(1))
\]

(20)

where \( (a) \) holds by defining \( \delta = \frac{\log^2(n)}{n} \), \( a_m = \frac{1}{l} (\frac{h_m}{T} + 1 + \frac{\log(n)}{\log(\log(n))}) + \delta, t_m^* = \arg \sup_{t \in \mathbb{R}} a_m t - \Gamma(t) \)

and by using Lemma [14] in the Appendix. Similarly, we have:

\[
P(F_i^{H_2}) \geq \sum_{m=1}^{M} \alpha_{-, m} \mathbb{P}\left( \sum_{i=1}^{n} [Z_i - W_i] \geq -\frac{h_m}{T} + 1 + \frac{\log(n)}{\log(\log(n))} \right)
\]

August 17, 2017  DRAFT
have the following cases:

- If \( h_1 = o(\log(n)) \) and \( \alpha_{+1} \neq \alpha_{-1} \), then \( t_1^* = \frac{1}{2} T \) for both (20), (21). Hence, by substituting in (20), (21), we have:

\[
\mathbb{P}(F_i^{H_1}) \geq n^{-0.5(\sqrt{a} - \sqrt{b})^2 + o(1)} \sum_{m=2}^{M} \alpha_{+,m} e^{-l(t_m^a a_m - \Gamma(t_m^a) + |t_m^a| \delta)} (1 - o(1))
\]

(22)

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 \( n^{-1 + \frac{\varepsilon}{2}} \). If \( h_1 = 0 \) and \( \log(\alpha_{+1}) = \log(\alpha_{-1}) = -\beta \log(n) \), \( \beta > 0 \), then \( t_1^* = \frac{1}{2} T \) for both (20), (21). Hence, by substituting in (20), (21), we have:

\[
\mathbb{P}(F_i^{H_1}) \geq n^{-0.5(\sqrt{a} - \sqrt{b})^2 - \beta + o(1)} \sum_{m=2}^{M} \alpha_{+,m} e^{-l(t_m^a a_m - \Gamma(t_m^a) + |t_m^a| \delta)} (1 - o(1))
\]

(24)

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 \( n^{-1 + \frac{\varepsilon}{2}} > \frac{\log^3(n)}{n} \log(10) \) for sufficiently large \( n \).

- If \( h_1 = \beta \log(n) \), for \( 0 < \beta < T^a b^{-2} \), then \( t_1^* = \log(\frac{\gamma + \beta}{\delta T}) \) for (20) and \( t_1^* = \frac{1}{T} \log(\frac{\gamma - \beta}{\delta T}) \) for (21), where \( \gamma = \sqrt{\beta^2 + abT^2} \). Hence, by substituting in (20), (21), we have:

\[
\mathbb{P}(F_i^{H_1}) \geq e^{-\log(n)(0.5\eta(a,b,\beta) - \log(\alpha_{+1})/\log(n)) + o(1)} \sum_{m=2}^{M} \alpha_{+,m} e^{-l(t_m^a a_m - \Gamma(t_m^a) + |t_m^a| \delta)} (1 - o(1))
\]

(26)
\[
\mathbb{P}(F_i^{H_1}) \geq e^{-\log(n)\left(0.5\eta(a,b,\beta) - \frac{\log(\alpha_{-1})}{\log(n)} + o(1)\right)} + \sum_{m=2}^{M} \alpha_{-m} e^{-t \left(t_m^a a_m - \Gamma(t^*_m) + |t^*_m| \delta\right)} (1 - o(1)) \\
(27)
\]

Then, if \(\log(\alpha_{+1}) = o(\log(n))\), this implies that \(\frac{\log(\alpha_{-1})}{\log(n)} = -\beta + o(1)\). Hence we have:

\[
\mathbb{P}(F_i^{H_1}) \geq n^{-0.5\eta(a,b,\beta) + o(1)} + \sum_{m=2}^{M} \alpha_{+m} e^{-t \left(t_m^a a_m - \Gamma(t^*_m) + |t^*_m| \delta\right)} (1 - o(1)) \\
(28)
\]

\[
\mathbb{P}(F_i^{H_2}) \geq n^{-0.5\eta(a,b,\beta) + o(1)} + \sum_{m=2}^{M} \alpha_{-m} e^{-t \left(t_m^a a_m - \Gamma(t^*_m) + |t^*_m| \delta\right)} (1 - o(1)) \\
(29)
\]

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 \(n^{-1+\frac{\varepsilon}{2}} > \frac{\log^3(n)}{n} \log(10)\) for sufficiently large \(n\).

On the other hand, if \(\log(\alpha_{+1}) = -\beta' \log(n)\), this implies that \(\frac{\log(\alpha_{-1})}{\log(n)} = -\beta''\), for some \(\beta'' > 0\) and \(\beta = \beta'' - \beta'\). Hence we have:

\[
\mathbb{P}(F_i^{H_1}) \geq n^{-0.5\eta(a,b,\beta) + \beta' + o(1)} + \sum_{m=2}^{M} \alpha_{+m} e^{-t \left(t_m^a a_m - \Gamma(t^*_m) + |t^*_m| \delta\right)} (1 - o(1)) \\
(30)
\]

\[
\mathbb{P}(F_i^{H_2}) \geq n^{-0.5\eta(a,b,\beta) + \beta'' + o(1)} + \sum_{m=2}^{M} \alpha_{-m} e^{-t \left(t_m^a a_m - \Gamma(t^*_m) + |t^*_m| \delta\right)} (1 - o(1)) \\
= n^{-0.5\eta(a,b,\beta) + \beta' + o(1)} + \sum_{m=2}^{M} \alpha_{-m} e^{-t \left(t_m^a a_m - \Gamma(t^*_m) + |t^*_m| \delta\right)} (1 - o(1)) \\
(31)
\]

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 \(n^{-1+\frac{\varepsilon}{2}} > \frac{\log^3(n)}{n} \log(10)\) for sufficiently large \(n\).

- If \(h_1 = \beta \log(n)\), for \(T^\frac{b-a}{2} < \beta < 0\), then following the same analysis as the last point, we get the following:

  if \(\log(\alpha_{-1}) = o(\log(n))\), this implies that \(\frac{\log(\alpha_{+1})}{\log(n)} = -\beta + o(1)\). Hence we have:

\[
\mathbb{P}(F_i^{H_1}) \geq n^{-0.5\eta(a,b,|\beta|) + o(1)} + \sum_{m=2}^{M} \alpha_{+m} e^{-t \left(t_m^a a_m - \Gamma(t^*_m) + |t^*_m| \delta\right)} (1 - o(1)) \\
(31)
\]

\[
\mathbb{P}(F_i^{H_2}) \geq n^{-0.5\eta(a,b,|\beta|) + o(1)} + \sum_{m=2}^{M} \alpha_{-m} e^{-t \left(t_m^a a_m - \Gamma(t^*_m) + |t^*_m| \delta\right)} (1 - o(1)) \\
(32)
\]
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 $n^{-1+\frac{\varepsilon}{2}} > \frac{\log^3(n)}{n} \log(10)$ for sufficiently large $n$.

On the other hand, if $\log(\alpha_{-1}) = -\beta'' \log(n)$, this implies that $\log(\alpha_{+1}) = -\beta'$, for some $\beta' > 0$ and $\beta = \beta'' - \beta'$. Hence we have:

$$\mathbb{P}(F_i^{H_1}) \geq n^{-0.5\eta(a,b,|\beta|)-\beta''+o(1)} + \sum_{m=2}^{M} \alpha_{+,m} e^{-t_n^* \Gamma(t_n^* + |t_m^*| \delta)} (1 - o(1))$$  \hspace{1cm} (33)

$$\mathbb{P}(F_i^{H_2}) \geq n^{-0.5\eta(a,b,|\beta|)-\beta''+o(1)} + \sum_{m=2}^{M} \alpha_{-,m} e^{-t_n^* \Gamma(t_n^* + |t_m^*| \delta)} (1 - o(1))$$  \hspace{1cm} (34)

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 $n^{-1+\frac{\varepsilon}{2}} > \frac{\log^3(n)}{n} \log(10)$ for sufficiently large $n$.

\[\Box\]

\[\Box\]

**B. Sufficient Conditions**

In this section we prove the achievable part of Theorem 6 by providing an efficient algorithm that achieves exact recovery down to the optimal information theoretic threshold. The algorithm is decomposed into two stages. The first stage of the algorithm is the same as Section III-B. The second stage, the local modification is however different.

After the first stage, we have $G_2$, the side information $y$, $A'$ and $B'$. We locally modify the community assignment as follows: if a node $i \in A'$, we 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 $\frac{h}{n^2}$ and for node $j \in B'$, we flip its membership if the number of edges between $j$ and $A'$ is greater than or equal the number of edges between $j$ and $B'$ minus $\frac{h}{n^2}$. If the number of flips in each cluster is not the same, keep the clusters unchanged.

**Lemma 11**: There exists large enough $D$ such that, with high probability, the algorithm described above will successfully recover the communities from the observed graph and side information if the following are satisfied simultaneously:
\[
\begin{cases}
(\sqrt{a} - \sqrt{b})^2 > 2, \text{ for any } m : \alpha_{+m} \neq \alpha_{-m} \text{ and } h_m = o(\log(n)) \\
(\sqrt{a} - \sqrt{b})^2 + 2\beta > 2, \text{ for any } m : \log(\alpha_{+m}) = \log(\alpha_{-m}) = -\beta \log(n), \beta > 0 \\
\eta(a, b, |\beta|) > 2, \text{ for any } m : h_m = \beta \log(n), |\beta| < T\frac{(a-b)}{2}, \log(\alpha_{sgn(\beta), m}) = o(\log(n)) \\
\eta(a, b, |\beta|) + 2\beta' > 2 \text{ for any } m : h_m = \beta \log(n), |\beta| < T\frac{(a-b)}{2}, \log(\alpha_{sgn(\beta), m}) = -\beta' \log(n), \beta' > 0
\end{cases}
\]

Proof:

Recall we defined \( P_e = \mathbb{P}(\text{node } i \text{ to be misclassified}) \). Following the same analysis as in the proof of Lemma \([2]\) we get:

\[
P_e \leq \frac{1}{2} \sum_{m=1}^{M} \alpha_{+m} \mathbb{P} \left( \sum_{i=1}^{\delta} Z_i + \sum_{i=1}^{\delta} W_i \geq \sum_{i=1}^{\delta} W_i + \sum_{i=1}^{\delta} Z_i + \frac{h_m}{T} \right) + \left( \frac{1}{2} \sum_{m=1}^{M} \alpha_{-m} \mathbb{P} \left( \sum_{i=1}^{\delta} Z_i + \sum_{i=1}^{\delta} W_i \geq \sum_{i=1}^{\delta} W_i + \sum_{i=1}^{\delta} Z_i - \frac{h_m}{T} \right) \right)
\]

Using a similar lemma as Lemma \([17]\) we can show that for any term of the \( M \) terms above:

- if \( h_m = o(\log(n)) \) and \( \alpha_{+m} \neq \alpha_{-m} \), then this term can be upper bounded by \( n^{-\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + o(1)} + n^{-(1+\Omega(1))} \).
- if \( h_m = 0 \) and \( \log(\alpha_{+m}) = \log(\alpha_{-m}) = -\beta \log(n), \beta > 0 \), then this term can be upper bounded by \( n^{-\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 - \beta + o(1)} + n^{-(1+\Omega(1))} \).
- if \( h_m = \beta \log(n), |\beta| < T\frac{(a-b)}{2} \) and \( \log(\alpha_{sgn(\beta), m}) = o(\log(n)) \), then this term can be upper bounded by \( n^{-\frac{1}{2}\eta(a,b,|\beta|) + o(1)} + n^{-(1+\Omega(1))} \).
- if \( h_m = \beta \log(n), |\beta| < T\frac{(a-b)}{2} \) and \( \log(\alpha_{sgn(\beta), m}) = -\beta' \log(n), \beta' > 0 \), then this term can be upper bounded by \( n^{-\frac{1}{2}\eta(a,b,|\beta|) - \beta' + o(1)} + n^{-(1+\Omega(1))} \).
- if \( h_m = \beta \log(n), T\frac{(a-b)}{2} < |\beta| \) and \( \log(\alpha_{sgn(\beta), m}) = -\beta' \log(n), \beta' > 0 \), then this term can be upper bounded by \( n^{-\beta'} + n^{-(1+\Omega(1))} \).

Combining all the above with a simple union bound over all nodes concludes the proof. Note that the last case, which implies that \( \beta' > 1 \) is also a needed condition for recovery, was not included in Theorem \([6]\) as we only show that it is sufficient but we could not show that it is necessary for the \( M \)-ary case. However, we believe it is also necessary as shown, as a special case, in Lemma \([12]\).
VI. CONCLUSION AND OPEN PROBLEMS

In this paper, the binary symmetric stochastic block model is studied with side information. We addressed the problem of the effect of side information on the phase transition threshold for exact recovery. We first considered two popular models of side information, namely: side information observed through binary symmetric or binary erasure channels. We showed that, for both cases, side information does not always help. In fact even when the quality or quantity of side information improves as \( n \to \infty \), we showed that for side information to help exact recovery, we need a certain rate by which the quality or quantity improves. We also, provided an efficient algorithm which succeeds all the way down to the thresholds, for both cases, using a partial recovery algorithm combined with a local improvement procedure. Finally, we generalized our results to M-ary side information with finite M.

APPENDIX I

**Lemma 12:** Recall that \( c = \log\left(\frac{1-\alpha}{\alpha}\right) \) and \( \eta(a, b, \beta) = a + b + \beta - 2\gamma + \frac{\beta}{T} \log\left(\frac{\gamma + \beta}{\gamma - \beta}\right) \), where \( T = \log\left(\frac{a}{b}\right), \gamma = \sqrt{\beta^2 + abT^2} \). Also, recall from Section III-A that

\[
P(F_i^H) = P\left(\sum_{i=1}^{n} Z_i - \sum_{i=1}^{n} W_i - cy_i \geq T + T \frac{\log(n)}{\log\log(n)}\right)
\]

where \( W_i \sim T \ast Bern(p) \), \( Z_i \sim T \ast Bern(q) \) and \( y_i \in \{\pm 1\} \sim \{(1 - \alpha), \alpha\} \).

For sufficiently large \( n \), \( P(F_i^H) > \frac{\log^2(\alpha)}{n} \log(10) \), if one of the following is satisfied:

\[
\begin{align*}
\text{For } c = \beta \log(n), \beta > 0 &\quad \text{if } \eta(a, b, \beta) < 2 \\
\text{For } c = \beta \log(n), \beta > 0 &\quad \text{for any } a, b, \beta \\
\text{For } c = \beta \log(n), \beta > 0 &\quad \text{for } a, b, \beta : \beta > \frac{T(a-b)}{2} \quad \text{if } \eta(a, b, \beta) > 2 \text{ and } \beta < 1
\end{align*}
\]

**Proof:**

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

\[
P(F_i^H) = P\left(\sum_{i=1}^{n} Z_i - \sum_{i=1}^{n} W_i - cy_i \geq T + T \frac{\log(n)}{\log\log(n)}\right) \\
\geq P\left(\sum_{i=1}^{n} (Z_i - W_i) \geq cy_i + T + T \frac{\log(n)}{\log\log(n)}\right)
\]
\[(1 - \alpha)\mathbb{P}\left(\frac{1}{l} \sum_{i=1}^{n} [Z_i - W_i] \geq \frac{1}{l}(c + T + T \frac{\log(n)}{\log \log(n)})\right) + \alpha\mathbb{P}\left(\frac{1}{l} \sum_{i=1}^{n} [Z_i - W_i] \geq \frac{1}{l}(-c + T + T \frac{\log(n)}{\log \log(n)})\right) \geq (1 - \alpha)e^{-t\left(t_1^*a_1 - \Gamma(t_1^*) + |t_1^*|\delta\right)}(1 - o(1)) + e^{-t\left(t_2^*a_2 - \Gamma(t_2^*) + |t_2^*|\delta\right)}(1 - o(1)) \quad (36)\]

where \(a\) holds by defining \(\delta = \frac{\log \frac{2}{\delta(n)}}{l}\), \(a_1 = \frac{1}{T}(c + T + T \frac{\log(n)}{\log \log(n)}) + \delta\), \(a_2 = \frac{1}{T}(-c + T + T \frac{\log(n)}{\log \log(n)}) + \delta\), \(t_1^* = \arg \sup_{t \in \mathbb{R}} ta_1 - \Gamma(t)\), \(t_2^* = \arg \sup_{t \in \mathbb{R}} ta_2 - \Gamma(t)\) and by using Lemma \([14]\).

Now, we calculate \(t_1^*\) and \(t_2^*\). First, we simplify the function inside the supremum. Note that both supremums are very similar, so we will do the analysis for one of them, and the second should follow similarity.

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

Now, it is easy to check that the right hand side is concave in \(t \in \mathbb{R}\). Hence, taking the derivative with respect to \(t\), we get:

\[
a_1 - \frac{q\left(\frac{a}{b}\right)^t}{1 - q(1 - \left(\frac{a}{b}\right)^t)} + \frac{p\left(\frac{a}{b}\right)^{-t}}{1 - p(1 - \left(\frac{a}{b}\right)^{-t})} = \frac{\log(n)}{n} \left(\frac{2c}{\log(n)} + \frac{2T}{\log(n)} + \frac{2T}{\log(\log(n))} - \frac{2}{\log^2(n)} - \frac{b\left(\frac{a}{b}\right)^t}{1 - q(1 - \left(\frac{a}{b}\right)^t)} + \frac{a\left(\frac{a}{b}\right)^{-t}}{1 - p(1 - \left(\frac{a}{b}\right)^{-t})}\right) = 0 \quad (37)\]

Now, we divide our analysis into two cases:

- **Case one: \(c = o(\log(n))\)**. In that case, the first four terms in (37) is \(o(1)\). This suggests that \(t^* = \frac{1}{2}\). Hence, by substituting back in (37), we get:

\[
ta_1 - \Gamma(t) = \frac{1}{2}a_1 - \log \left(1 - q(1 - \left(\sqrt{\frac{a}{b}}\right))^t\right) - \log \left(1 - p(1 - \left(\sqrt{\frac{b}{a}}\right)^{-t})\right) \leq \frac{1}{2}a_1 + \frac{q(1 - \left(\sqrt{\frac{a}{b}}\right))}{1 - q(1 - \left(\sqrt{\frac{a}{b}}\right))} + \frac{p(1 - \left(\sqrt{\frac{b}{a}}\right))}{1 - p(1 - \left(\sqrt{\frac{b}{a}}\right))} \leq \frac{\log(n)}{n} \left(\frac{b(1 - \left(\sqrt{\frac{a}{b}}\right))}{1 - q(1 - \left(\sqrt{\frac{a}{b}}\right))} + \frac{a(1 - \left(\sqrt{\frac{b}{a}}\right))}{1 - p(1 - \left(\sqrt{\frac{b}{a}}\right))} + o(1)\right)\]
\[
\frac{\log(n)}{n} \left( (\sqrt{a} - \sqrt{b})^2 + o(1) \right)
\]  

(38)

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 \( (1 - q(1 - (\sqrt{\frac{b}{a}}))) \) → 1 as \( n \to \infty \). Hence, substituting in one of the supremums of (36), we have:

\[
e^{-l(t^*_a - \Gamma(t^*_a) + |t^*_a|^d)}(1 - o(1)) \geq e^{-\log(n)} \left( \frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + o(1) \right)(1 - o(1))
\]

Finally, following the same steps for the second supremum and substituting in (36), we have:

\[
\mathbb{P}(F^H_i) \geq (1 - \alpha)n^{-0.5(\sqrt{a} - \sqrt{b})^2 + o(1)} + \alpha n^{-0.5(\sqrt{a} - \sqrt{b})^2 + o(1)}
\]

\[
= n^{-0.5(\sqrt{a} - \sqrt{b})^2 + o(1)}
\]

Thus, if \((\sqrt{a} - \sqrt{b})^2 \leq 2 - \varepsilon\) for some \( 0 < \varepsilon < 2 \), then \( \mathbb{P}(F^H_i) \geq n^{-1 + \varepsilon} > \frac{\log^2(n)}{n} \log(10) \) for sufficiently large \( n \). This proves the first case of Lemma 12.

- **Case two:** \( \alpha = \frac{1}{n^2} \), \( \beta > 0 \). In this case \( c = \beta \log(n) + o(1) \). Hence, substituting in (37), 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 (37) and following the same ideas as in (38), we get:

\[
ta_1 - \Gamma(t) \leq \frac{\log(n)}{n} \left( 2\beta t^* + b(1 - \left( a \over b \right)^t) + a(1 - (a \over b)^{-t}) + o(1) \right)
\]

\[
= \frac{\log(n)}{n} \left( \frac{2\beta}{T} \log(\frac{\gamma + \beta}{bT}) + a + b - \frac{\gamma + \beta}{T} - \frac{abT}{\gamma + \beta} + o(1) \right)
\]

\[
= \frac{\log(n)}{n} \left( a + b + \beta - \frac{2\gamma}{T} + \frac{\beta}{T} \log(\frac{\gamma + \beta}{\gamma - \beta}) + o(1) \right)
\]

\[
= \frac{\log(n)}{n} (\eta(a, b, \beta) + o(1))
\]

(39)

Hence, substituting in one of the supremums of (36), we have:

\[
e^{-l(t^*_a - \Gamma(t^*_a) + |t^*_a|^d)}(1 - o(1)) \geq e^{-\frac{\log(n)}{n} (\eta(a, b, \beta) + o(1))}(1 - o(1))
\]

Finally, following the same steps for the second supremum and substituting in (36), we have:
\[\mathbb{P}(F_i^H) \geq (1 - \alpha) n^{-0.5n(a,b,\beta) + \alpha(n)} + \alpha n^{-0.5n(a,b,\beta) + \beta + \alpha(n)}\]
\[= n^{-0.5n(a,b,\beta) + \alpha(n)} (2 - \alpha)\]

Thus, if \(\eta(a, b, \beta) \leq 2 - \varepsilon\) for some \(0 < \varepsilon < 2\), then \(\mathbb{P}(F_i^H) \geq n^{-1+\frac{\alpha}{2}} > \frac{\log^3(n)}{n} \log(10)\) for sufficiently large \(n\). This proves the second case of Lemma 12.

For the last case of Lemma 12, we begin as in (36) but take a different approach:

\[\mathbb{P}(F_i^H) \geq \mathbb{P}\left(\sum_{i=1}^{\frac{n}{2}} |Z_i - W_i| \geq c y_i + T + T \frac{\log(n)}{\log \log(n)}\right)\]
\[= (1 - \alpha)(1 - \mathbb{P}\left(\sum_{i=1}^{\frac{n}{2}} |Z_i - W_i| \leq c + T + T \frac{\log(n)}{\log \log(n)}\right)) + \alpha \mathbb{P}\left(\sum_{i=1}^{\frac{n}{2}} |Z_i - W_i| \leq (-c + T + T \frac{\log(n)}{\log \log(n)})\right)\]
\[\geq (1 - \alpha)e^{-n \sup_{t > 0} t (c + T + T \frac{\log(n)}{\log \log(n)})} - \frac{1}{n} \log \left(E(e^{-t \sum_{i=1}^{\frac{n}{2}} (Z_i - W_i)})\right)\]
\[- \alpha e^{-n \sup_{t > 0} t (-c + T + T \frac{\log(n)}{\log \log(n)})} - \frac{1}{n} \log \left(E(e^{-t \sum_{i=1}^{\frac{n}{2}} (Z_i - W_i)})\right)\]

where (a) holds by Chernoff bound. Note that unlike the previous cases, here the supremum is only on \(t > 0\). Again, we will now focus on calculating one of the supremums in (40).

By direct computation of the logarithmic term we get:

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

where (a) follows from the fact that \(W_i, Z_i\) are independent random variables \(\forall i\), and (b) holds because \(\log(1 - x) \leq -x\). Substituting in one of the supremums:
\[
\sup_{t>0} \frac{-t}{n} \left( c + T + T \log(n) \right) - \frac{1}{n} \log \left( \mathbb{E}(e^{-t(\sum_{i=1}^{n} Z_i - W_i)}) \right) \\
\geq \frac{\log(n)}{n} \sup_{t>0} -t(\beta + o(1)) + \frac{1}{2} \left( a + b - b\left(\frac{a}{b}\right)^{-t} - a\left(\frac{a}{b}\right)^{t} \right)
\]

Again, by concavity of the last equation in \(t\), we calculate the first derivative to get:

\[
-\beta - \frac{aT}{2} \left(\frac{a}{b}\right)^{t} + \frac{bT}{2} \left(\frac{a}{b}\right)^{-t} = 0
\]

Hence, following the same analysis as before, we can show 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 we need \(t\) to be greater than zero. Thus, we need \(\beta < \frac{T(b-a)}{2}\) for the first supremum, which can not be true, since \(\beta\) 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, we need \(\beta > \frac{T(a-b)}{2}\) for \(t\) to be positive.

Thus, assuming \(\beta > \frac{T(a-b)}{2}\), substituting in (40), we get:

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

where \((b)\) 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 \(\mathbb{P}(F_i^H) \geq n^{-1}(n^{\varepsilon_1} - n^{\varepsilon_2}) > \frac{\log^3(n)}{n} \log(10)\) for sufficiently large \(n\). This proves the third and last case of Lemma 12.

\textit{Lemma 13}: Recall from Section IV-A that

\[
\mathbb{P}(F_i^H) = e^{\mathbb{P}\left( \sum_{i=1}^{n} Z_i - \sum_{i=1}^{n} W_i \geq 1 + \frac{\log(n)}{\log(\log(n))} \right)}
\]

where \(W_i \sim \text{Bern}(p)\), \(Z_i \sim \text{Bern}(q)\) and \(\epsilon \in (0,1)\).

For sufficiently large \(n\), \(\mathbb{P}(F_i^H) > \frac{\log^3(n)}{n} \log(10)\), if one of the following is satisfied:
\[
\begin{align*}
\text{For } \log(\epsilon) &= o(\log(n)), \quad \text{if } (\sqrt{a} - \sqrt{b})^2 < 2 \\
\text{For } \log(\epsilon) &= -\beta \log(n), \quad \beta > 0, \quad \text{if } (\sqrt{a} - \sqrt{b})^2 + 2\beta < 2
\end{align*}
\]

Proof:

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

\[
\mathbb{P}(F_i^H) = e^{\mathbb{P}\left( \sum_{i=1}^{\frac{n}{2}} (Z_i) - \sum_{i=1}^{\frac{n}{2}} (W_i) \geq 1 + \frac{\log(n)}{\log \log(n)} \right)}
\]

\[
\geq e^{\mathbb{P}\left( \sum_{i=1}^{\frac{n}{2}} (Z_i - W_i) \geq 1 + \frac{\log(n)}{\log \log(n)} \right)}
\]

\[
\overset{(a)}{=} e^{-l\left( \log_2(n) \right)} \left( 1 - o(1) \right)
\]

where \((a)\) holds by defining \( \delta = \frac{\log 2}{l}, a = \frac{1}{l} (1 + \frac{\log(n)}{\log \log(n)}) + \delta, \ t^* = \arg \sup_{t \in \mathbb{R}} at - \Gamma(t) \) and by using Lemma [14] in the Appendix.

Now, we calculate \( t^* \). First, we simplify the function inside the supremum.

\[
\begin{align*}
ta - \Gamma(t) &= ta - \log \left( 1 - q(1 - e^t) \right) - \log \left( 1 - p(1 - e^{-t}) \right) \\
&= t^* a = t^* \left( \log_2(n) \right)
\end{align*}
\]

By concavity of the above function in \( t \) and by taking the derivative with respect to \( t \) and equating to zero, we get:

\[
\begin{align*}
a - \frac{q e^t}{1 - q(1 - e^t)} + \frac{p e^{-t}}{1 - p(1 - e^{-t})} = \\
\frac{\log(n)}{n} \left( \frac{2}{\log(n)} + \frac{2}{\log \log(n)} + \frac{2}{\log \log(n)} - \frac{b e^t}{1 - q(1 - e^t)} + \frac{a e^{-t}}{1 - p(1 - e^{-t})} \right) = 0
\end{align*}
\]

Now, since the first three terms in \((45)\) is \( o(1) \). Thus, this suggests that \( t^* = \frac{1}{2} T \) and \( T = \log \left( \frac{a}{b} \right) \). Hence, by substituting back in \((44)\), we get:

\[
ta - \Gamma(t) = \frac{1}{2} a T - \log \left( 1 - q(1 - (\sqrt{a} / b)) \right) - \log \left( 1 - p(1 - (\sqrt{b} / a)) \right)
\]
\[ \begin{align*}
&\leq \frac{1}{2} aT + \frac{q(1 - (\sqrt{\frac{a}{b}}))}{1 - q(1 - (\sqrt{\frac{a}{b}}))} + \frac{p(1 - (\sqrt{\frac{b}{a}}))}{1 - p(1 - (\sqrt{\frac{b}{a}}))} \\
&= \frac{\log(n)}{n} ((\sqrt{a} - \sqrt{b})^2 + o(1))
\end{align*} \] (46)

where \((a)\) holds because \(\log(1 - x) \geq -\frac{x}{1-x}\). Hence, substituting in the supremum of (43), we have:

\[ e^{-t^*(a - \Gamma(t^*) + |t^*|\delta)} (1 - o(1)) \geq e^{-\log(n)} \left( \frac{1}{2} (\sqrt{a} - \sqrt{b})^2 + o(1) \right) (1 - o(1)) \] (47)

Finally, substituting in (43), we have:

\[ \mathbb{P}(F_i^H) \geq \epsilon n^{-0.5(\sqrt{a} - \sqrt{b})^2 + o(1)} \] (48)

Thus, if \(\log(\epsilon) = o(\log(n))\), then, it is clear that if \((\sqrt{a} - \sqrt{b})^2 \leq 2 - \epsilon\) for some \(0 < \epsilon < 2\), then \(\mathbb{P}(F_i^H) \geq n^{-1+\epsilon} > \frac{\log(n)}{n} \log(10)\) for sufficiently large \(n\). This proves the first case of Lemma 13.

On the other hand, if \(\log(\epsilon) = -\beta \log(n)\), for some \(\beta > 0\), then, it is clear that if \((\sqrt{a} - \sqrt{b})^2 + 2\beta \leq 2 - \epsilon\) for some \(0 < \epsilon < 2\), then \(\mathbb{P}(F_i^H) \geq n^{-1+\frac{\epsilon}{2}} > \frac{\log(n)}{n} \log(10)\) for sufficiently large \(n\). This proves the second and last case of Lemma 13.

**Lemma 14:** 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 \(a, \epsilon \in \mathbb{R}\):

\[ \mathbb{P}\left( \frac{1}{n} \sum_{i=1}^{n} X_i \geq a - \epsilon \right) \geq e^{-n \left( t^*(a - \Gamma(t^*) + |t^*|\delta) \right)} \left( 1 - \frac{\sigma_X^2}{n\epsilon^2} \right) \]

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

**Proof:**

\[ \begin{align*}
\mathbb{P}\left( \frac{1}{n} \sum_{i=1}^{n} X_i \geq a - \epsilon \right) &\geq \mathbb{P}(a - \epsilon \leq \frac{1}{n} \sum_{i=1}^{n} X_i \leq a + \epsilon) \\
&= \int_{a - \epsilon \leq \frac{1}{n} \sum_{i=1}^{n} X_i \leq a + \epsilon} \mathbb{P}(x_1) \cdots \mathbb{P}(x_n) dx_1 \cdots dx_n
\end{align*} \]
which concludes our proof.

We have:

Now, choose \( t = t^* = \arg \sup_{t \in \mathbb{R}} ta - \Gamma(t) \). Since this function is concave in \( t \in \mathbb{R} \), then by setting the first derivative to zero, we have \( a = \frac{E_X[xe^{tx}]}{E[e^{tx}]} \). Also, by direct computation of \( \mu_{\hat{X}} \), we can show that \( \mu_{\hat{X}} = \frac{E_X[xe^{tx}]}{E[e^{tx}]} \). This means that at \( t = t^* \), we have \( \mu_{\hat{X}} = a \). Thus, substituting back in (49), we get:

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

Now, in our model \( \epsilon = \frac{\log^2(n)}{n} \) and \( X = Z - W \), where \( Z \sim T^*\text{Bern}(q) \) and \( W \sim T^*\text{Bern}(p) \), where \( T = \log\left(\frac{p}{q} \right) \). Hence, we can easily show that \( \sigma^2_{\hat{X}} \) is in the order of \( \frac{\log(n)}{n} \), and hence, we have:

\[
P\left( \frac{1}{n} \sum_{i=1}^{n} X_i \geq a - \epsilon \right) \geq e^{-n(ta - \Gamma(t) + |t|\epsilon)} \left( 1 - o(1) \right)
\]

which concludes our proof.

\[\text{Lemma 15:} \text{ Define } \eta(a,b,\beta) = a + b + \beta - 2\frac{\gamma}{T} + \frac{\beta}{T} \log\left(\frac{\gamma + \beta}{\gamma - \beta}\right), \text{ where } T = \log\left(\frac{p}{q}\right), \gamma = \sqrt{\beta^2 + 4bT^2} \text{ and } a, b, \beta > 0 \text{ with } a > b. \text{ Then, } \beta > 1 \implies \eta > 2.\]

\[\text{Proof:}\]

Let \( a + b - \beta - 2\frac{\gamma}{T} + \frac{\beta}{T} \log\left(\frac{\gamma + \beta}{\gamma - \beta}\right) = \psi(a,b,\beta) \). Then, from the definition of \( \eta \), we have:
\[ \eta(a, b, \beta) - 2\beta = \psi(a, b, \beta) \quad (50) \]

Note that \( \psi(a, b, \beta) \) is convex in \( \beta \). Hence, at the optimal \( \beta^* \), we can show that \( \log(\frac{\gamma^* + \beta^*}{\gamma^* - \beta^*}) = T \).

Using this fact and by substituting in (50), we have:

\[ \eta(a, b, \beta) - 2\beta \geq a + b - 2\frac{\gamma^*}{T} \quad (51) \]

Now by the definition of \( \gamma \), we have \( \frac{\gamma^* + \beta^*}{\gamma^* - \beta^*} = \frac{a}{b} \), we have \( \frac{a}{b} = \frac{abT^2}{(\gamma^* - \beta^*)^2} \), which implies that \( \gamma^* = bT + \beta^* \). Hence, by substituting in (51), we get:

\[ \eta(a, b, \beta) - 2\beta \geq a - b - 2\frac{\beta^*}{T} \quad (52) \]

Also, we can show that at \( \beta^* \), \( \gamma^* = \beta^*(\frac{a + b}{a - b}) \). This implies that \( \beta^* = \frac{T(a - b)}{2} \). Substituting in (52), we get: \( \eta(a, b, \beta) - 2\beta \geq 0 \), which implies that \( \eta > 2 \) when \( \beta > 1 \).

\[ \text{Lemma 16: Define } P_n^{(k)} := \mathbb{P}( \sum_{i=1}^m (Z_i - W_i) \geq 0 ) \text{, where } m = 2k(\frac{a}{2} - k), W_i \sim \text{Bern}(p), Z_i \sim \text{Bern}(q). \text{ Then,} \]

\[ P_n^{(k)} \leq e^{-2m \frac{\log(a)}{n}(\frac{1}{2}(\sqrt{a} - \sqrt{b})^2)} \]

\[ \text{Proof:} \]

\[ P_n^{(k)} = \mathbb{P}\left( \sum_{i=1}^m (Z_i - W_i) \geq 0 \right) \]

\[ \leq e^{-m \sup_{t > 0} - \frac{1}{m} \log(\mathbb{E}[e^{t \sum_{i=1}^m (Z_i - W_i)}])} \]

\[ \leq e^{-m \frac{\log(a)}{n} \sup_{t > 0} a + b - be^t - ae^{-t}} \]

\[ \leq e^{-2m \frac{\log(a)}{n}(\frac{1}{2}(\sqrt{a} - \sqrt{b})^2)} \quad (53) \]

where (a) holds by Chernoff bound, (b) holds because \( Z_i \) and \( W_i \) are both i.i.d \( \forall i \) and (c) holds by evaluating the supremum to get \( t^* = \frac{1}{2}T \) and using the fact that \( \log(1 - x) \leq -x \).
Lemma 17: Let $W_i \sim Bern(p)$, $Z_i \sim Bern(q)$, $y_i \in \{1, -1\}$ with probabilities $(1 - \alpha), \alpha$, respectively. Define $P_e = \mathbb{P}(\sum_{i=1}^{(1-\delta)\frac{n}{2}} Z_i + \sum_{i=1}^{\delta \frac{n}{2}} W_i \geq \sum_{i=1}^{(1-\delta)\frac{n}{2}} W_i + \sum_{i=1}^{\delta \frac{n}{2} - \frac{2D}{\log(n)} n} Z_i + \epsilon y_i)$. Then,

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

Proof:

By upper bounding $P_e$, we get:

$$
P_e \leq \mathbb{P} \left( \sum_{i=1}^{(1-\delta)\frac{n}{2}} Z_i + \sum_{i=1}^{\delta \frac{n}{2}} W_i \geq \sum_{i=1}^{(1-\delta)\frac{n}{2}} W_i + \frac{c}{T} y_i \right)$$

$$
\leq \mathbb{P} \left( \sum_{i=1}^{\frac{n}{2}} Z_i + \sum_{i=1}^{\delta \frac{n}{2}} W_i \geq \sum_{i=1}^{(1-\delta)\frac{n}{2}} W_i + \frac{c}{T} y_i \right)$$

$$
\leq \mathbb{P} \left( \sum_{i=1}^{\frac{n}{2}} Z_i - \sum_{i=1}^{\delta \frac{n}{2}} W_i + \sum_{i=1}^{\delta \frac{n}{2}} W_i \geq \frac{c}{T} y_i \right)$$

$$
\leq \mathbb{P} \left( \sum_{i=1}^{\frac{n}{2}} (Z_i - W_i) \geq \frac{c}{T} y_i - \psi \delta \log(n) \right) + \mathbb{P} \left( \sum_{i=1}^{\delta \frac{n}{2}} W_i \geq \psi \delta \log(n) \right)$$

$$
= (1 - \alpha) \mathbb{P} \left( \sum_{i=1}^{\frac{n}{2}} (Z_i - W_i) \geq \frac{c}{T} - \psi \delta \log(n) \right) + \alpha \mathbb{P} \left( \sum_{i=1}^{\frac{n}{2}} (Z_i - W_i) \geq - \frac{c}{T} - \psi \delta \log(n) \right) + \mathbb{P} \left( \sum_{i=1}^{\delta \frac{n}{2}} W_i \geq \psi \delta \log(n) \right)
$$

where (a) holds by defining $\psi = \frac{1}{\delta \sqrt{\log(\frac{n}{2})}}$.

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

\[
\Pr \left( \sum_{i=1}^{\delta n + \frac{2D}{\log(n)}} W_i \geq \psi \delta \log(n) \right) \leq \left( \frac{\psi \delta \log(n)}{ae(\delta \log(n) + 2D)} \right)^{-\psi \delta \log(n)}
\]

\[
= \left( \frac{\psi}{ae(1 + \frac{2D}{\delta \log(n)})} \right)^{-\log(1+\frac{2D}{\delta \log(n)})}
\]

\[
= e^{\log(n) \left( 1 + \frac{\log(1+\frac{2D}{\delta \log(n)})}{\log(\frac{1}{\delta})} \right)}
\]

\[
= n^{-\sqrt{\frac{\log(1+\frac{2D}{\delta \log(n)})}{\log(\frac{1}{\delta})}} + o(1) (55)}
\]

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, we get:

\[
\Pr \left( \sum_{i=1}^{\delta n + \frac{2D}{\log(n)}} W_i \geq \psi \delta \log(n) \right) \leq n^{-(1+\Omega(1))} (56)
\]

Now for the first term in (54), we use Chernoff bound as follows:

\[
(1 - \alpha)\Pr \left( \sum_{i=1}^{n} (Z_i - W_i) \geq \frac{c}{\sqrt{T}} - \psi \delta \log(n) \right) + \alpha \Pr \left( \sum_{i=1}^{n} (Z_i - W_i) \geq \frac{c}{\sqrt{T}} - \psi \delta \log(n) \right)
\]

\[
\leq (1 - \alpha)\Pr \left( \sum_{i=1}^{\frac{n}{2}} (Z_i - W_i) \geq \frac{c}{\sqrt{T}} - \psi \delta \log(n) \right) + \alpha \Pr \left( \sum_{i=1}^{\frac{n}{2}} (Z_i - W_i) \geq \frac{c}{\sqrt{T}} - \psi \delta \log(n) \right)
\]

\[
\leq (1 - \alpha)e^{-\frac{\log(n)}{2} \sup_{t_1 > 0} \left( \frac{c}{\sqrt{T \log(n)}} - \psi \delta \log(n) \right) - \frac{2t_1}{\sqrt{n}} - \psi \delta \log(n) - \log(\mathbb{E}[e^{t_1^2 - t_1^2 W}])} + \alpha e^{-\frac{\log(n)}{2} \sup_{t_2 > 0} \left( \frac{c}{\sqrt{T \log(n)}} - \psi \delta \log(n) \right) - \frac{2t_2}{\sqrt{n}} - \psi \delta \log(n) - \log(\mathbb{E}[e^{t_2^2 - t_2^2 W}])}
\]

\[
\leq (1 - \alpha)e^{-\frac{\log(n)}{2} \sup_{t_1 > 0} \left( \frac{c}{\sqrt{T \log(n)}} - \psi \delta \log(n) \right) - \frac{2t_1}{\sqrt{n}} - \psi \delta \log(n) - \log(\mathbb{E}[e^{t_1^2 - t_1^2 W}])} + \alpha e^{-\frac{\log(n)}{2} \sup_{t_2 > 0} \left( \frac{c}{\sqrt{T \log(n)}} - \psi \delta \log(n) \right) - \frac{2t_2}{\sqrt{n}} - \psi \delta \log(n) - \log(\mathbb{E}[e^{t_2^2 - t_2^2 W}])}
\]

\[
= (a) e^{-\frac{\log(n)}{2} \sup_{t_1 > 0} \left( \frac{c}{\sqrt{T \log(n)}} - \psi \delta \log(n) \right) + \alpha e^{-\frac{\log(n)}{2} \sup_{t_2 > 0} \left( \frac{c}{\sqrt{T \log(n)}} - \psi \delta \log(n) \right) + \frac{2t_2}{\sqrt{n}} - \psi \delta \log(n) - \log(\mathbb{E}[e^{t_2^2 - t_2^2 W}])}
\]

\[
(57)
\]

where \((a)\) holds because \( \log(1 - x) \leq -x \). Now, we calculate \( t_1^* \) and \( t_2^* \).

Note that \( \psi \delta \to 0 \) as \( D \to \infty \). Hence, we will replace \( \psi \delta \) by \( o(1) \) for sufficiently large \( D \). Now we divide our analysis into three cases:

- If \( c = o(\log(n)) \), this suggests that \( t_1^* = t_2^* = \frac{1}{2} T \). Hence, substituting in (57), we get:

\[
(1 - \alpha)\Pr \left( \sum_{i=1}^{\frac{n}{2}} (Z_i - W_i) \geq \frac{c}{\sqrt{T}} - \psi \delta \log(n) \right) + \alpha \Pr \left( \sum_{i=1}^{\frac{n}{2}} (Z_i - W_i) \geq \frac{c}{\sqrt{T}} - \psi \delta \log(n) \right)
\]

August 17, 2017 DRAFT
\[ \leq n^{-\frac{1}{2}(\sqrt{a} - \sqrt{b})^2 + o(1)} \]  

\[ (1 - \alpha)\mathbb{P}\left( \sum_{i=1}^{n} (Z_i - W_i) \geq \frac{c}{T} - \psi \delta \log(n) \right) + \alpha \mathbb{P}\left( \sum_{i=1}^{n} (Z_i - W_i) \geq -\frac{c}{T} - \psi \delta \log(n) \right) \leq (2 - \alpha) n^{-\frac{1}{2}(a,b)} + o(1) \]  

\[ (1 - \alpha)\mathbb{P}\left( \sum_{i=1}^{n} (Z_i - W_i) \geq \frac{c}{T} - \psi \delta \log(n) \right) + \alpha \mathbb{P}\left( \sum_{i=1}^{n} (Z_i - W_i) \geq -\frac{c}{T} - \psi \delta \log(n) \right) \leq (1 - \alpha) n^{-\frac{1}{2}n(a,b) + o(1)} + n^{-\beta} \]

Using the last three displayed equations and \[ (56) \] and substituting in \[ (54) \] concludes the proof of the lemma.

**References**

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

[2] A. Zhang and H. Zhou, “Minimax rates of community detection in stochastic block models,” *arXiv:1507.05313*, Nov. 2015.

[3] P. J. Bickel and A. Chen, “A nonparametric view of network models and newmangirvan and other modularities,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 50, pp. 21 068–21 073, 2009.

[4] T. T. Cai and X. Li, “Robust and computationally feasible community detection in the presence of arbitrary outlier nodes,” *arXiv:1404.6000v4*, June 2015.

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

[6] S. Jafar, “Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices,” *ICML, In proceedings of*, Feb. 2014.

[7] A. Coja-oghlan, “Graph partitioning via adaptive spectral techniques,” *Comb. Probab. Comput.*, vol. 19, no. 2, pp. 227–284, Mar. 2010.
[8] A. Coja-Oghlan, “A spectral heuristic for bisecting random graphs,” in *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA ’05. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2005, pp. 850–859. [Online]. Available: http://dl.acm.org/citation.cfm?id=1070432.1070552
[9] X. L. T. Tony Cai, “A tensor approach to learning mixed membership community models,” *arXiv:1302.2684v4*, Oct. 2013.
[10] Y. Chen, S. Sanghavi, and H. Xu, “Improved graph clustering,” *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6440–6455, Oct 2014.
[11] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, “Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications,” *Phys. Rev. E*, vol. 84, p. 066106, Dec. 2011.
[12] P. Zhang, K. F., ReichardtJ., and Z. L., “Comparative study for inference of hidden classes in stochastic block models,” *arXiv:1207.2328v2*, July 2012.
[13] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
[14] J. Xu, R. Wu, K. Zhu, B. Hajek, R. Srikant, and L. Ying, “Jointly clustering rows and columns of binary matrices: Algorithms and trade-offs,” *arXiv:1310.0512*, Feb. 2014.
[15] J. Chen and B. Yuan, “Detecting functional modules in the yeast proteinprotein interaction network,” *Bioinformatics*, vol. 22, no. 18, pp. 2283–2290, Sept. 2006.
[16] A. Lancichinetti and S. Fortunato, “Community detection algorithms: A comparative analysis,” *Phys. Rev. E*, vol. 80, p. 056117, Nov. 2009.
[17] S. Fortunato, “Community detection in graphs,” *arXiv:0906.0612v2*, Jan. 2010.
[18] E. Abbe, A. Bandeira, and G. Hall, “Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms,” *arXiv:1503.00609*, Mar. 2015.
[19] M. Newman and A. Clauset, “Structure and inference in annotated networks,” *arXiv:1507.04001v1*, 2015.
[20] J. Yang, J. McAuley, and J. Leskovec, “Community detection in networks with node attributes,” in *IEEE International Conference on Data Mining*, Dec. 2013, pp. 1151–1156.
[21] Y. Zhang, E. Levina, and J. Zhu, “Community detection in networks with node features,” *arXiv:1509.01173v1*, 2015.
[22] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, “Inference and phase transitions in the detection of modules in sparse networks,” *Phys. Rev. Lett.*, vol. 107, p. 065701, Aug 2011. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.107.065701
[23] E. Mossel, J. Neeman, and A. Sly, “Reconstruction and estimation in the planted partition model,” *Probability Theory and Related Fields*, vol. 162, no. 3, pp. 431–461, 2014.
[24] L. Massoulie, “Community detection thresholds and the weak ramanujan property,” *arXiv:1311.3085v1*, Nov. 2013.
[25] E. Mossel, J. Neeman, and A. Sly, “A proof of the block model threshold conjecture,” *arXiv:1311.4115v4*, Aug 2015.
[26] E. Abbe and C. Sandon, “Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic bp, and the information-computation gap,” *arXiv:1512.09080v4*, Sep 2016.
[27] S. Yun and A. Proutiere, “Community detection via random and adaptive sampling,” *arXiv:1402.3072v2*, Feb 2014.
[28] E. Mossel and J. Xu, “Density evolution in the degree-correlated stochastic block model,” in *29th Annual Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, V. Feldman, A. Rakhlin, and O. Shamir, Eds., vol. 49. Columbia University, New York, New York, USA: PMLR, 23–26 Jun 2016, pp. 1319–1356. [Online]. Available: http://proceedings.mlr.press/v49/mossel16.html
[29] H. Saad, A. Abotabl, and A. Nosratinia, “Exit analysis for belief propagation in degree-correlated stochastic block models,” in IEEE International Symposium on Information Theory (ISIT), July 2016, pp. 775–779.

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

[31] M. Elchanan, J. Neeman, and S. Allan, “Consistency thresholds for the planted bisection model,” arXiv:1407.1591v4, Mar. 2016.

[32] E. Mossel and J. Xu, “Local algorithms for block models with side information,” in ACM Conference on Innovations in Theoretical Computer Science, ser. ITCS ’16. New York, NY, USA: ACM, 2016, pp. 71–80. [Online]. Available: http://doi.acm.org/10.1145/2840728.2840749

[33] T. Tony Cai, T. Liang, and A. Rakhlin, “Inference via message passing on partially labeled stochastic block models,” arXiv:1603.06923v1, Mar. 2016.

[34] A. Kadavankandy, K. Avrachenkov, L. Cottatellucci, and R. Sundaresan, “The power of side-information in subgraph detection,” arXiv:1611.04847v3, Mar. 2017.

[35] V. Kanade, E. Mossel, and T. Schramm, “Global and local information in clustering labeled block models,” arXiv:1404.6325v4, July 2014.

[36] F. Caltagirone, M. Lelarge, and L. Miolane, “Recovering asymmetric communities in the stochastic block model,” arXiv:1610.03680v2, Mar. 2017.

[37] H. Saad, A. Abotabl, and A. Nosratinia, “Exact recovery in the binary stochastic block model with binary side information,” in Allerton Conference on Communications, Control, and Computing, Oct 2017.

[38] A. Dembo and O. Zeitouni, Large deviations techniques and applications. Berlin; New York: Springer-Verlag Inc, 2010.