Belief Propagation with Side Information for Recovering a Single Community

Hussein Saad and Aria Nosratinia
Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX 75083-0688, USA,
E-mail: hussein.saad@utdallas.edu; aria@utdallas.edu.

***Abstract*—In this paper, we study the effect of side information on the recovery of a hidden community of size K inside a graph consisting of n nodes with $K = o(n)$. We focus on side information with finite cardinality and bounded (as $n \to \infty$) log-likelihood ratios (LLRs). We calculate tight necessary and sufficient conditions for weak recovery of the labels subject to observation of the graph and side information under belief propagation (BP). Also, we show that BP with side information is strictly inferior to the maximum likelihood detector without side information. Finally, we validate our results through simulations on finite synthetic data-sets that shows the power of our asymptotic results in characterizing the performance even at finite n.**

***Index Terms*—Community detection, Stochastic block model, Side information, Belief Propagation.**

I. Introduction

The problem of learning or detecting community structures in random graphs has been studied in statistics [1], [2], computer science [3], [4] and theoretical statistical physics [5]. In this paper, we consider the problem of finding a single sub-graph (community) hidden in a large graph, where the community size is much smaller than the graph size. This problem arises in many applications such as fraud detection in auction networks and web-graphs [6], [7].

Among the different random graph models, the stochastic block model (SBM) is widely used in the context of community detection [8], [9]. We use the stochastic block model for one community [10], [11], [12], [13], which is characterized by the following parameters: n is the number of nodes in the graph, K is the size of the community, p is the probability of having an edge between any two nodes inside the community, and q is the probability of having an edge otherwise. The goal is to recover/detect the hidden community upon observing the graph edges.

The problem of finding a hidden community upon observing *only* the graph has been studied in [10], [11], [12]. The information limit of *weak recovery* (expected number of misclassified nodes is $o(K)$) and *exact recovery* (probability of correctly recovering all the labels converges to one) of a hidden community have been established in [11]. The limit of the belief propagation (BP) algorithm for weak recovery and exact recovery has been also established in [12], [10] in terms of a signal to noise ratio parameter $\lambda = \frac{K^2(p-q)^2}{(n-k)q}$, where it was shown that BP achieves weak recovery if and only if $\lambda > \frac{1}{e}$. Moreover, BP followed by a local voting procedure was shown to achieve exact recovery if exact recovery is information theoretically possible and if $\lambda > \frac{1}{e}$.

Graphical structures have been the main focus of the literature of community detection. However, in many practical applications, non-graphical relevant information is available that can aid the inference. For example, social networks such as Facebook and Twitter have access to much information other than the graph edges. A citation network has the authors names, keywords, and abstracts of papers, and therefore may provide significant additional information beyond the co-authoring relationships. This paper presents new results on the utility of side information in community detection, in particular shedding light on the conditions under which side information can improve the limit of weak recovery of a local algorithm based on belief propagation for detecting a hidden community.

A few results have recently appeared in the literature regarding the community detection problem in the presence of additional (non-graphical) information. In the context of detecting two symmetric communities: (1) [14] showed that, under certain condition, belief propagation with noisy label information has the same residual error as the maximum a-posteriori estimator. (2) Cai *et. al* [15] demonstrated regimes for BP to achieve weak recovery upon observing a vanishing fraction of labels. In [14], [15], a converse result was not established. Also, the graph parameters are chosen such that node degrees alone are not informative. Our work is different from the above settings, in that we deal with a single community, and the degrees can be informative in revealing node identities. In the context of detecting a single community, Kadavankandy *et al.* [13] used density evolution to characterize the performance of BP with binary side information consisting of noisy labels with vanishing noise, i.e. unbounded likelihood ratios, with a specific rate of growth. They showed that BP achieves weak recovery for any non-vanishing $\lambda > 0$. Our work is different in that we focus on side information with bounded log-likelihood ratios which is more practical. Also, we consider more general side information whose alphabet does not match the number/identity of communities which is motivated by practical scenarios. Moreover, our results suggests that BP can achieve weak recovery for any $\lambda > 0$ with side information whose log-likelihood ratios are unbounded with an arbitrary rate of growth which generalizes the results obtained in [13]. In the interest of completeness, [16], [17],

This work was supported in part by the NSF grant 1718551.

978-1-5386-4780-6/18/$31.00©2018 IEEE

[18] also considered side information but in a different context other than belief propagation.

In this paper, a BP algorithm that leverage side information consisting of M outcomes with finite $M$is proposed for detecting a hidden community of size $K = o(n)$. We consider side information with bounded log-likelihood ratios (LLR) and study the effect of such side information on the weak recovery limits of BP. More specifically:

- We show that the proposed BP algorithm achieves weak recovery if and only if $\lambda > \frac{1}{Le}$, where $L \geq 1$ is a function of the side information LLR. Moreover, we show that for any local algorithm (to be defined later), weak recovery is not possible if $\lambda \leq \frac{1}{Le}$.
- We also compare the new limit with the information limit without side information for weak and exact recovery established in [11]. We show that BP, which is near-linear in computational complexity with respect to n, is strictly inferior to the maximum likelihood detector, which has exponential complexity.
- Moreover, we provide numerical results on finite synthetic data-sets that validate our asymptotic analysis showing the power of our asymptotic results in characterizing the performance even at finite n

II. System Model and Definitions

We consider the stochastic block model for a hidden community with side information. Let $\mathcal{G}(n, K, p, q)$ denote the ensemble of graphs with n nodes, a hidden community C^* with size $|C^*| = K$ and an edge between a pair of nodes is drawn with probability p if both nodes are in C^* and probability q otherwise. Denote by $G(V, E)$ a graph realization of $\mathcal{G}(n, K, p, q)$. Let x_i denotes the label of node $i \in \{1, \cdots, n\}$, where $x_i = 1$ if $i \in C^*$ and $x_i = 0$ if $i \notin C^*$. Also, let $\boldsymbol{x}^* \in \{0, 1\}^n$ denotes the vector of the true labels. Finally, for each node i a scalar side information $y_i \in \{u_1, u_2, \cdots, u_M\}$, $M < \infty$ is observed. Let $\mathbb{P}(y_i = u_m | x_i = 1) = \alpha_{+,m}$ and $\mathbb{P}(y_i = u_m | x_i = 0) = \alpha_{-,m}$, for $\alpha_{+,m} \geq 0$, $\alpha_{-,m} \geq 0$ and $\sum_{m=1}^M \alpha_{+,m} = \sum_{m=1}^M \alpha_{-,m} = 1$.

This paper studies the problem of recovery the hidden community upon observing G and the vector of nodes' side information by $\boldsymbol{y}$. Let $\hat{\boldsymbol{x}}(G, \boldsymbol{y})$ be an estimator of $\boldsymbol{x}^*$ given G and $\boldsymbol{y}$. The following assumptions and definitions are used throughout the paper:

As $n \to \infty$: $K \to \infty$ such that $K = o(n)$, $p \geq q$, $\frac{p}{q} = \theta(1)$, $\limsup_{n \to \infty} p < 1$, $np = n^{o(1)}$ and $\frac{K^2(p-q)^2}{(n-k)q} \to \lambda$ and λ is a positive constant. Also, as $n \to \infty$, both $\frac{\alpha_{+,m}}{\alpha_{-,m}}$ and $\frac{\alpha_{-,m}}{\alpha_{+,m}}$ are bounded for all $m \in \{1, \cdots, M\}$.

An estimator $\hat{\boldsymbol{x}}(G, \boldsymbol{y})$ is said to achieve weak recovery if, as $n \to \infty$, $\frac{d(\hat{\boldsymbol{x}}, \boldsymbol{x}^*)}{K} \to 0$ in probability, where $d(.,.)$ denotes the hamming distance. It was shown in [11] that such a definition for weak recovery is equivalent to the existence of an estimator $\hat{\boldsymbol{x}}$ such that $\mathbb{E}[d(\hat{\boldsymbol{x}}, \boldsymbol{x}^*)] = o(K)$. We will use this equivalence throughout this paper. Also, an estimator $\hat{\boldsymbol{x}}(G, \boldsymbol{y})$ is said to achieve exact recovery if, as $n \to \infty$, $\mathbb{P}(\hat{\boldsymbol{x}} = \boldsymbol{x}^*) \to 1$.

Finally, we denote the expectation of the likelihood ratio of the side information conditioned on $x = 1$ by:

$$L \triangleq \sum_{m=1}^{M} \frac{\alpha_{+,m}^2}{\alpha_{-,m}} \tag{1}$$

III. Belief Propagation Algorithm

A. Belief Propagation on a Random Tree with Side Information

In this section, we study an inference problem on a random tree. Fix a node u and let T_u be an infinite tree rooted at u. For a node $i \in T_u$, let T_i^t be a sub-tree of T_u rooted at i with depth t. Let $\tau_i \in \{0, 1\}$ denote the label of node i in T_u. The tree is generated as follows. Assume $\tau_u = Bernoulli(\frac{K}{n})$. For any $i \in T_u$, let H_i denote the number of its children j such that $\tau_j = 1$ and F_i denote the number of its children j such that $\tau_j = 0$. Assume $H_i \sim Poisson(Kp)$ if $\tau_i = 1$, $H_i \sim Poisson(Kq)$ if $\tau_i = 0$ and $F_i \sim Poisson((n-K)q)$ for all i. Finally, for any node $i \in T_u$, let $\tilde{\tau}_i \in \{u_1, \cdots, u_M\}, M < \infty$ denote node i side information. Assume, $\mathbb{P}(\tilde{\tau}_i = u_m | \tau_i = 1) = \alpha_{+,m}$ and $\mathbb{P}(\tilde{\tau}_i = u_m | \tau_i = 0) = \alpha_{-,m}$.

We wish to infer the label of root u given observations the tree T_u^t and side information $\tilde{\tau}_{T_u^t}$, where $\tilde{\tau}_{T_u^t}$ is the set of side information of all nodes i in T_u^t. Based on these definitions, the probability of error of an estimator $\hat{\tau}_u(T_u^t, \tilde{\tau}_{T_u^t})$ can be written as:

$$p_e^t \triangleq \frac{K}{n} \mathbb{P}(\hat{\tau}_u = 0 | \tau_u = 1) + \frac{n-K}{n} \mathbb{P}(\hat{\tau}_u = 1 | \tau_u = 0) \tag{2}$$

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

$$\Gamma_u^t = \log \left(\frac{\mathbb{P}(T_u^t, \tilde{\tau}_{T_u^t} | \tau_u = 1)}{\mathbb{P}(T_u^t, \tilde{\tau}_{T_u^t} | \tau_u = 0)} \right) \tag{3}$$

The probability of error of the MAP estimator can be bounded as follows [19]:

$$\frac{K(n-K)}{n^2} \rho^2 \leq p_e^t \leq \frac{\sqrt{K(n-K)}}{n} \rho \tag{4}$$

where $\rho = \mathbb{E}[e^{\frac{\Gamma_u^t}{2}} | \tau_u = 0]$. In the remainder of this section, we will drive upper and lower bounds on ρ.

Lemma 1. *Let N_u denote the children of node u and $h_u = \log\left(\frac{\mathbb{P}(\tilde{\tau}_u | \tau_u = 1)}{\mathbb{P}(\tilde{\tau}_u | \tau_u = 0)}\right)$. Then,*

$$\Gamma_u^{t+1} = -K(p-q) + h_u + \sum_{k \in N_u} \log \left(\frac{\frac{p}{q} e^{\Gamma_k^t - \nu} + 1}{e^{\Gamma_k^t - \nu} + 1} \right) \tag{5}$$

Proof. Omitted for brevity. □

1) Lower and Upper Bounds on ρ:
Define for $t \geq 1$, $\psi_u^t = -K(p-q) + \sum_{j \in N_u} F(h_j + \psi_j^{t-1})$, where $F(x) = \log(\frac{\frac{p}{q} e^{x-\nu} + 1}{e^{x-\nu} + 1}) = \log(1 + \frac{\frac{p}{q} - 1}{1 + e^{-(x-\nu)}})$. Then, $\Gamma_u^{t+1} = h_u + \psi_u^{t+1}$ and $\psi_i^0 = 0 \; \forall i \in T_u^t$. Also, let Z_0^t and Z_1^t denote the distribution of ψ_u^t conditioned on $\tau_u = 0$ and $\tau_u = 1$, respectively. Similarly, let U_0 and U_1 denote the distribution

of h_u conditioned on $\tau_u = 0$ and $\tau_u = 1$, respectively. Thus, $\rho = \mathbb{E}[e^{\frac{1}{2}(Z_0^t+U_0)}] = \mathbb{E}[e^{\frac{U_0}{2}}]\mathbb{E}[e^{\frac{Z_0^t}{2}}]$. Let $f(x) = \frac{1+\frac{p}{q}x}{1+x}$. Also, define:

$$b_t \triangleq \mathbb{E}[\frac{e^{Z_1^t+U_1}}{1+e^{Z_1^t+U_1-\nu}}] \tag{6}$$

$$a_t \triangleq \mathbb{E}[e^{Z_1^t+U_1}] \tag{7}$$

Lemma 2. *Let $B = (\frac{p}{q})^{1.5}$. Then:*

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

Proof. Omitted for brevity. □

Thus, to bound ρ, we need to bound b_t.

Lemma 3. *For all $t \geq 0$, $b_t \leq Le$, if $\lambda \leq \frac{1}{Le}$.*

Proof. Omitted for brevity. □

It remains to lower bound b_t for $\lambda > \frac{1}{Le}$.

Lemma 4. *Let $L^{'} = \mathbb{E}[e^{3U_0}]$. Assume that $b_t \leq \frac{\nu}{2(C-\lambda)}$. Then,*

$$b_{t+1} \geq a_{t+1} \geq Le^{\lambda b_t}(1 - \frac{L^{'}}{L}e^{\frac{-\nu}{2}}) \tag{9}$$

Proof. Omitted for brevity. □

Lemma 5. *The sequences a_t and b_t are non-decreasing in t.*

Proof. The proof is very similar to that of [12, Lemma 5], and is omitted for brevity. □

Lemma 6. *Define $\log^*(\nu)$ to be the number of times the logarithm function must be iteratively applied to ν to get a result less than or equal to one. Let $C = \lambda(2 + \frac{p}{q})$ and $L^{'} = \mathbb{E}[e^{3U_0}]$. Suppose $\lambda > \frac{1}{Le}$. Then there are constants $\bar{t}_o$ and ν_o depending only on λ and L such that:*

$$b_{\bar{t}_o+\log^*(\nu)+2} \geq Le^{\frac{\lambda\nu}{2(C-\lambda)}}(1 - \frac{L^{'}}{L}e^{\frac{-\nu}{2}}) \tag{10}$$

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

Proof. Omitted for brevity. □

2) Achievability and Converse for the MAP Detector:

Lemma 7. *Recall the definition of the MAP estimator $\hat{\tau}_{MAP} = 1_{\Gamma_u^t \geq \nu}$, where $\nu = \log(\frac{n-K}{K})$ and $\Gamma_u^t = \log(\frac{\mathbb{P}(T_u^t, \tilde{\tau}_{T_u^t}|\tau_u=1)}{\mathbb{P}(T_u^t, \tilde{\tau}_{T_u^t}|\tau_u=0)})$. Let $L^{'} = \mathbb{E}[e^{3U_0}]$. Define $C = \lambda(2 + \frac{p}{q})$ and $B = (\frac{p}{q})^{1.5}$, which is bounded because $\frac{p}{q} = \theta(1)$.*

If $0 < \lambda \leq \frac{1}{Le}$, then:

$$p_e^t \geq \frac{K(n-K)}{n^2}\mathbb{E}^2[e^{\frac{U_0}{2}}]e^{\frac{-\lambda Le}{4}} \tag{11}$$

If $\lambda > \frac{1}{Le}$, then:

$$p_e^t \leq \sqrt{\frac{K(n-K)}{n^2}}\mathbb{E}[e^{\frac{U_0}{2}}]e^{\frac{-\lambda L}{8B}e^{\frac{\lambda\nu}{2(C-\lambda)}}(1-\frac{L^{'}}{L}e^{\frac{-\nu}{2}})} \tag{12}$$

Moreover, by assumption, we have $\nu \to \infty$. Then,

$$p_e^t \leq \sqrt{\frac{K(n-K)}{n^2}}\mathbb{E}[e^{\frac{U_0}{2}}]e^{-\nu(r+\frac{1}{2})} = \frac{K}{n}e^{-\nu(r+o(1))} \tag{13}$$

for some $r > 0$.

Proof. The proof follows directly from (4) and Lemmas 3 and 6. □

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

```
Start with graph $G$ and side information $\boldsymbol{y}$.
Set $R_{i\to j}^0 = 0, \forall i \in \{1, \cdots, n\}$ and $j \in N_i$.
For all $i \in \{1, \cdots, n\}$ and $j \in N_i$, run for $t_f - 1$ iterations of
    belief propagation as in (15).
For all $i \in \{1, \cdots, n\}$, compute its belief $R_i^{t_f}$ based on (16):
Return $\tilde{C}$, the set of $K$ indices in $\{1, \cdots, n\}$ with the largest $R_i^{t_f}$.
```

B. Belief Propagation Algorithm for Community Recovery with Side Information

In this section, we relate the inference problem defined on the random tree to the problem of recovering a hidden community with side information. This can be done via a coupling lemma [12] that shows that under certain conditions, the neighborhood of a fixed node i in the graph is locally like a tree with probability converging to one, and hence, the BP algorithm defined for random trees in Section III-A can be used on the graph as well. The proof of these coupling lemmas depends only on the tree structure. Since in this paper, the side information is independent of the tree structure given the labels. This implies that the coupling lemmas hold for our case as well. We state the coupling lemma for completion.

Lemma 8. *Suppose that $K, p, q, \alpha_{+,m}, \alpha_{-,m}$ for all $m \in \{1, \cdots, M\}$ and t_f depend on n such that t_f is a positive integer and $(2 + np)^{t_f} = n^{o(1)}$. Then:*

If $|C^*| = K$, then for any fixed node u, there exists a coupling between $(G, \boldsymbol{x}, \boldsymbol{y})$ and $(T_u, \tau_{T_u}, \tilde{\tau}_{T_u})$ such that:

$$\mathbb{P}((G_u^{t_f}, \boldsymbol{x}_u^{t_f}, \boldsymbol{y}_u^{t_f}) = (T_u^{t_f}, \tau_{T_u^{t_f}}, \tilde{\tau}_{T_u^{t_f}})) \geq 1 - n^{-1+o(1)} \tag{14}$$

Proof. The proof follows directly from [12, Lemma 15]. □

We are now ready to present the BP algorithm for community recovery with side information. Define the message transmitted from node i to its neighbor node j at iteration $t+1$ as:

$$R_{i\to j}^{t+1} = h_i - K(p-q) + \sum_{k\in N_i\backslash j} F(R_{k\to i}^t) \tag{15}$$

where $h_i = \log(\frac{\mathbb{P}(y_i|x_i=1)}{\mathbb{P}(y_i|x_i=0)})$, N_i is the set of neighbors of node i and $F(x) = \log(\frac{\frac{p}{q}e^{x-\nu}+1}{e^{x-\nu}+1})$. The messages are initialized to zero for all nodes i, i.e., $R_{i\to j}^0 = 0$ for all $i \in \{1, \cdots, n\}$ and $j \in N_i$. Define the belief of node i at iteration $t+1$ as:

$$R_i^{t+1} = h_i - K(p-q) + \sum_{k\in N_i} F(R_{k\to i}^t) \tag{16}$$

Algorithm I presents the proposed BP algorithm for community recovery with side information.

By Lemma 8, with probability converging to one, we have $R_i^{t_f} = \Gamma_i^{t_f}$, where $\Gamma_i^{t_f}$ was the log-likelihood defined for the random tree. Hence, we expect the performance of Algorithm I to be the same as the MAP estimator defined as $\hat{\tau}_{MAP} = 1_{\Gamma_i^{t_f} \geq \nu}$, where $\nu = \log(\frac{n-K}{K})$. The only difference is that the MAP estimator decides based on $\Gamma_i^{t_f} \geq \nu$ while in Algorithm I the selection is based on the largest $R_i^{t_f}$. Let $\hat{C}$ define the community recovered by the MAP estimator, i.e. $\hat{C} = \{i : R_i^{t_f} \geq \nu\}$. Since $\tilde{C}$ is the set of nodes with the K largest $R_i^{t_f}$. Then, we have either $\tilde{C} \subset \hat{C}$ or $\hat{C} \subset \tilde{C}$. Thus,

$$\begin{aligned} |C^* \triangle \tilde{C}| &\leq |C^* \triangle \hat{C}| + |\hat{C} \triangle \tilde{C}| = |C^* \triangle \hat{C}| + |K - |\hat{C}|| \\ &= |C^* \triangle \hat{C}| + ||C^*| - |\hat{C}|| \leq 2|C^* \triangle \hat{C}| \end{aligned} \tag{17}$$

1) Weak Recovery:

Theorem 1. *[Achievability] Suppose that $(np)^{\log^*(n)} = n^{o(1)}$. Assume $\lambda > \frac{1}{Le}$. Let $t_f = \bar{t}_o + \log^*(\nu) + 2$, where $\bar{t}_o$ is a constant depending only on λ and L. Let $\tilde{C}$ be the output of Algorithm I. Then,*

$$\frac{\mathbb{E}[|C^* \triangle \tilde{C}|]}{K} \to 0 \tag{18}$$

for $|C^*| = K$.

Proof. Omitted for brevity. □

Remark 1. *In [13] the authors used a different approach, namely, density evolution to study the effect of **binary noisy** side information with unbounded LLR on the performance of BP. The authors considered the case where the LLR grows as $\log(\frac{n}{K})$, i.e., L grows as $\frac{n}{K}$, and showed that BP achieves weak recovery for any $\lambda > 0$. Our result here suggests that for unbounded LLR, BP can achieve weak recovery for $L \to \infty$ arbitrary slow. A comprehensive study of the unbounded LLR case is still an open problem.*

Theorem 2. *[Converse] Assume $\lambda \leq \frac{1}{Le}$. Let $t_f \in \mathbb{N}$ depend on n such that $(2+np)^{t_f} = n^{o(1)}$. Then, for any estimator $\hat{C}$ such that for each node u in the graph, x_u^* is estimated based on G and $\boldsymbol{y}$ in a neighborhood of radius t_f from u,*

$$\frac{\mathbb{E}[|C^* \triangle \hat{C}|]}{K} \geq (1 - \frac{K}{n})\mathbb{E}^2[e^{\frac{U_0}{2}}]e^{\frac{-\lambda Le}{4}} - o(1) \tag{19}$$

where $\mathbb{E}^2[e^{\frac{U_0}{2}}] = (\sum_{m=1}^{M} \alpha_{-,m}\sqrt{\frac{\alpha_{+,m}}{\alpha_{-,m}}})^2$ which is bounded by assumption.

Proof. Omitted for brevity. □

2) Comparison with Information Theoretic Limits:

The information limits for weak recovery without side information was established in [11]. Since $K \to \infty$, the information limits for weak recovery without side information reduces to $\liminf_{n\to\infty} \frac{Kd(p||q)}{2\log(\frac{n}{K})} \geq 1$ [11], where $d(p||q)$ is the binary Kullback-Leibler divergence. In terms of λ, it was shown in [12] that the former condition can be written as:

$$\lambda > C\frac{K}{n}\log(\frac{n}{K}) \tag{20}$$

for some positive constant C. Thus, weak recovery only demands a vanishing λ. On the other hand, we showed that BP achieves weak recovery for $\lambda > \frac{1}{Le}$, where L is bounded and greater than one. This implies a gap between the information limits and BP limits for weak recovery. Note that since $L \geq 1$, then the gap is smaller compared to the BP limit without side information, i.e., $\lambda > \frac{1}{e}$.

To illustrate our results, consider the following regime:

$$K = \frac{cn}{\log(n)},\ q = \frac{b\log^2(n)}{n},\ p = 2q \tag{21}$$

for fixed positive b, c as $n \to \infty$. In the above regime, $\lambda = c^2 b$ and $Kd(p||q) \approx \log(n)$, and hence, weak recovery is always asymptotically possible without side information, and by extension, with side information. Moreover, it was shown in [12] that exact recovery is asymptotically possible if $cb(1 - \frac{1+\log\log(2)}{\log(2)}) > 1$. We focus on side information with $M = 2$, where each node observes the true label passed though a binary symmetric channel with cross-over probability α.

Figure 1 shows the curve $\{(b,c) : c^2 b = \frac{1}{e}\}$, i.e., the weak recovery limit of BP without side information, the curve $\{(b,c) : c^2 b = \frac{1}{Le}\}$, i.e., the weak recovery limit of BP with side information and the curve $\{(b,c) : cb(1 - \frac{1+\log\log(2)}{\log(2)}) = 1\}$, i.e., the information theoretic limit for exact recovery. It was shown in [12], that BP without side information achieves exact recovery if $cb(1 - \frac{1+\log\log(2)}{\log(2)}) > 1$ and $\lambda > \frac{1}{e}$. From the figure, it can be shown that side information helps BP to achieve recovery in regimes where it was known to fail without side information. In region 1, exact recovery is provided by the BP algorithm plus voting procedure with or without the help of side information. In region 2, we conjecture that exact recovery is provided by the proposed BP algorithm with side information plus the same voting procedure provided in [12]. In region 3, weak recovery is provided by the BP algorithm with or without the help of side information, but exact recovery is not asymptotically possible. In region 4, weak recovery is provided by the BP algorithm only with the help of side information, but exact recovery is not asymptotically possible. In region 5, exact recovery is asymptotically possible, but BP without side information or with side information whose $\alpha = 0.3$ can not achieve even weak recovery (needs less α, i.e., better quality of side information). In region 6, weak recovery, but not exact recovery, is asymptotically possible and BP without side information or with side information whose $\alpha = 0.3$ can not achieve weak recovery.

3) Numerical Results:

This section aims to validate our theoretical findings on the synthetic model. We show that even at finite n, the performance of BP is illuminated by the theoretic limits found in this paper. We run Algorithm I on a graph generated with $n = 10^4, K = 100, t_f = 10$. We assume side information consisting of the true labels passed through a binary symmetric channel with cross-over probability α. We use the following performance metric $\zeta = \frac{\sum_{i=1}^{n}|x_i^* - \hat{x}_i|}{K} \in [0, 2]$. Two scenarios are considered: (1) $\lambda < \frac{1}{e}$, where we used $q = 5 \times 10^{-4}$ and $p = 10q$, which results in $\lambda \approx 0.041$. The results are reported for different values of α in Table II, which show that when $\lambda < \frac{1}{Le}$, the fraction of error is close to its maximum which

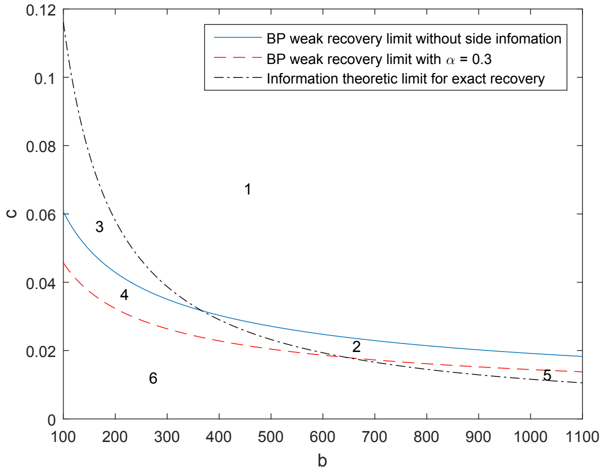

Fig. 1. Phase diagram with $K = c\frac{n}{\log(n)}$, $q = \frac{b\log^2(n)}{n}$, $p = 2q$ and $\alpha = 0.3$ for b, c fixed as $n \to \infty$.

implies that weak recovery is not attainable. On the other hand, when $\lambda >> \frac{1}{Le}$, the fraction of error is closer to zero. (2) $\lambda > \frac{1}{e}$, where we used $q = 5 \times 10^{-4}$ and $p = 80q$, which results in $\lambda \approx 3.152$. The results are reported for different values of α in Table III. In this scenario, the results show that the performance of BP without side information is much better compared to the first scenario, which occur because $\lambda > \frac{1}{e}$. The results also show that the performance is improved as α decreases.

TABLE II
PERFORMANCE OF BP FOR $\lambda < \frac{1}{e}$.

α	ζ w/o side	$\lambda \times Le \approx$	ζ with side
0.1	1.92	0.903	1.5
0.01	1.92	10	0.8
0.001	1.92	100	0.1

TABLE III
PERFORMANCE OF BP FOR $\lambda > \frac{1}{e}$.

α	ζ w/o side	$\lambda \times Le \approx$	ζ with side
0.1	0.25	70	0.2
0.01	0.25	840	0.06
0.001	0.25	8551	0.04

REFERENCES

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

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

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

[4] A. Coja-oghlan, “Graph partitioning via adaptive spectral techniques,” *Comb. Probab. Comput.*, vol. 19, no. 2, pp. 227–284, Mar. 2010.

[5] 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.

[6] A. Beutel, W. Xu, V. Guruswami, C. Palow, and C. Faloutsos, “Copycatch: Stopping group attacks by spotting lockstep behavior in social networks,” in *International Conference on World Wide Web*, 05 2013, pp. 119–130.

[7] D. H. Chau, S. Pandit, and C. Faloutsos, “Detecting fraudulent personalities in networks of online auctioneers,” in *European Conference on Principle and Practice of Knowledge Discovery in Databases*, 2006, pp. 103–114.

[8] S. Fortunato, “Community detection in graphs,” *arXiv:0906.0612v2*, Jan. 2010.

[9] 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*, July 2016, pp. 775–779.

[10] A. Montanari, “Finding one community in a sparse graph,” *arXiv:1502.05680v2*, Jul. 2015.

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

[12] ——, “Recovering a hidden community beyond the spectral limit in $o(|e|\log^*|v|)$ time,” *arXiv:1510.02786v2*, Jun. 2017.

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

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

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

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

[17] 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.

[18] H. Saad and A. Nosratinia, “Side information in the binary stochastic block model: Exact recovery,” *accepted in IEEE Journal of Selected Topics in Signal Processing*, 2018.

[19] H. Kobayashi and J. Thomas, “distance measures and related criteria,” in *Allerton Conference Circuits and System Theory*, 1967.