The Power of Side-Information in Subgraph Detection

Arun Kadavankandy, Konstantin Avrachenkov, Laura Cottatellucci, Rajesh Sundaresan

The Power of Side-information in Subgraph Detection

Arun Kadavankandy, Konstantin Avrachenkov, Laura Cottatellucci and Rajesh Sundaresan

***Abstract*—In this work, we tackle the problem of hidden community detection. We consider Belief Propagation (BP) applied to the problem of detecting a hidden Erdős-Rényi (ER) graph embedded in a larger and sparser ER graph, in the presence of side-information. We derive two related algorithms based on BP to perform subgraph detection in the presence of two kinds of side-information. The first variant of side-information consists of a set of nodes, called cues, known to be from the subgraph. The second variant of side-information consists of a set of nodes that are cues with a given probability. It was shown in past works that BP without side-information fails to detect the subgraph correctly when a so-called effective signal-to-noise ratio (SNR) parameter falls below a threshold. In contrast, in the presence of non-trivial side-information, we show that the BP algorithm achieves asymptotically zero error for any value of a suitably defined phase-transition parameter. We validate our results on synthetic datasets and a few real world networks.**

I. Introduction

1) Problem Motivation: We consider the problem of hidden community detection in graphs in the presence of side-information. In various disciplines graphs have been used to model, in a parsimonious fashion, relationships between heterogenous data. The presence of a dense hidden community in such graphs is usually indicative of interesting phenomena in the associated real-world network.

An application of dense subgraph detection in Signal Processing is the problem of Correlation Mining [15]. Consider a network with nodes representing correlated signals and weighted links representing pairwise correlations. The problem of detecting a group of closely correlated signals is then a dense subgraph detection problem on the constructed graph [15]. Dense subgraph detection also finds application in real-world computer and social networks; for e.g., in detecting fraudulent activity [6], [10], [33]. It can also be viewed as a signal recovery problem on graphs [13], [34].

A majority of subgraph detection algorithms try to find a subset of nodes that maximizes some objective such as the average link density within the subset [24]. A good way to benchmark the performance of various community detection algorithms is to validate them on generative graph models with inherent community structure. In this work, we model the hidden community as a small but well-connected Erdős-Rényi graph embedded within a larger but sparser Erdős-Renyi graph. This model was used in [26] to capture terrorist transactions in a computer network. It is a special case of the Stochastic Block Model (SBM), widely used to assess the performance of different community detection algorithms [32].

The study of subgraph detection on generative models is interesting in itself from an algorithmic perspective. Recent works on hidden community detection and related problems demonstrate the presence of sharp phase transitions in the range of parameter values between three regimes: easy (detection achievable with relatively small computational costs), hard (computationally taxing, but detectable), and impossible to detect [9], [17], [29]. We provide more details on these phenomena while reviewing prior works in the next subsection. The novel aspect of this paper is a theoretical study of the impact of side-information on this computational barrier. The form of side-information we consider is the identity of special nodes called cues that are known to belong to the subgraph, either deterministically or with some level of certainty. One often has access to such prior knowledge in real-world applications [5], [37], [38].

In this paper we focus on local distributed algorithms that are essential for graph clustering application when the graph is distributed over a computer cluster or a cloud. Furthermore, in graph applications involving extremely large sizes as is the typical case in Big Data problems, the full graph is not available to the algorithm, and one is often interested in a local solution or in a local *low-conductance cut* in graph clustering terminology. Such considerations have been pursued in the works [4], [16] and references therein.

By developing and analyzing the asymptotic performance of a local algorithm based on Belief Propagation (BP), we show that even a small amount of side-information can lead to the disappearance of the computational barrier. BP is an efficient way to perform approximate Maximum Likelihood (ML) detection on certain types of graphs using distributed and local message passing [25]. It belongs to the class of guilt-by-association schemes [23] and has been successfully applied to many practical problems in graphs such as fraud detection [10] and data mining [22].

2) Previous works: Consider a graph with n nodes that contains a hidden community of size K. The edge probability between any two nodes within the community is p and it is q otherwise, such that $p > q$. The parameters p, q and K can in general be functions of n. This model, denoted by $G(K, n, p, q)$, was already considered in [20], [26], [27] in

A. Kadavankandy is with CentraleSupélec, Paris(email:arun.kadavankandy@supelec.fr)

K. Avrachenkov is with Inria, Sophia Antipolis, France(email:konstantin.avratchenkov@inria.fr).

L. Cottatellucci was with Department of Communication Systems, EURECOM, Sophia-Antipolis, France. She is with Institute of Digital Communications, Friedrich-Alexander Universität, Erlangen-Nürnberg, Germany, (e-mail:laura.cottatellucci@fau.de).

R. Sundaresan is with Indian Institute of Science, Bangalore, India(email:rajeshs@iisc.ac.in).

the context of anomaly detection.

A special case of the above model is the hidden clique model with $p = 1$ and $q = 1/2$. The study of clique detection algorithms demonstrate the presence of phase transitions in the subgraph size K between impossible, hard and easy regimes. If $K \leq 2(1-\epsilon)\log_2(n)$, the clique is impossible to detect; however, an exhaustive search detects the clique nodes when $K \geq 2(1+\epsilon)\log_2(n)$. In contrast, the smallest clique size that can be detected in polynomial time is believed to be $c\sqrt{n}$ [3] for some $c > 0$, and the minimum clique-size that can be detected in nearly-linear time is believed to be $\sqrt{n/e}$ [14].

The computational barriers for subgraph detection in a sparse graph without cues were studied in [17], [18], [29]. In [29] the author investigated the performance of ML detection and BP, and analyzed the phase transition with respect to an effective signal-to-noise ratio (SNR) parameter λ defined as

$$\lambda = \frac{K^2(p-q)^2}{(n-K)q}. \tag{1}$$

The larger the λ, the easier it is to detect the subgraph. Subgraph recovery was considered under a parameter setting where $K = \kappa n, p = a/n$ and $q = b/n$, where κ, a and b are constants independent of n. It was shown under this setting that, for any $\lambda > 0$, an exhaustive search can detect the subgraph with success probability approaching one as $\kappa \to 0$. However BP, which has quasi-linear time complexity, achieves non-trivial success probability only when $\lambda > 1/e$ in the same regime. Further, for $\lambda < 1/e$, the success probability of the algorithm is bounded away from one. This demonstrates the existence of a computational barrier for local algorithms.

In [18] the authors show that when $K = o(n)$, i.e., when $\kappa \to 0$, and p, q are such that $a = np = n^{o(1)}$ and $p/q = O(1)$, ML detection succeeds when $\lambda = \Omega(\frac{K}{n}\log(\frac{n}{K}))$, i.e., detection is possible even when the SNR parameter goes to zero so long as it does not go to zero too fast. Under the same parameter setting, it was shown that BP succeeds in detecting the subgraph with the fraction of misdetected nodes going to zero, only when $\lambda > 1/e$ [17]. Therefore, $\lambda = 1/e$ represents a computational barrier for BP in the subgraph detection problem without side-information.

In the present work, we examine the impact of side-information on the above computational barrier. To the best of our knowlege, ours is the first theoretical study of the performance of local algorithms for subgraph detection in the presence of side-information in $G(K, n, p, q)$. In [28], the authors compared, but only empirically, several guilt-by-association schemes for subgraph detection with cues.

There exist many works on the effect of side-information in the context of identifying multiple communities [2], [8], [9], [30]. These works considered a different variant of the SBM where nodes are partitioned into two or more communities, with dense links inside communities and sparse links across communities. The authors of [8] and [30] consider a BP algorithm to detect two equal-sized communities. In [30], the side-information is such that all nodes indicate their community information after passing it through a binary symmetric channel with error rate α. They show that when $\alpha < 1/2$, i.e., when there is non-trivial side-information, there is no computational barrier and BP works all the way down to the detectability threshold called the Kesten-Stigum threshold [1]. In [8], a vanishing fraction $n^{-o(1)}$ of nodes reveal their true communities. Again, there is no computational barrier and BP works all the way down to the detectability threshold. A fuller picture is available in [9], which considers asymmetric communities and asymmetric connection probabilities within communities. In this setting, the authors of [9] demonstrate the presence of all three regimes (easy to detect, hard to detect but possible via exhaustive search, and impossible to detect) as a function of the size of the smallest community. In contrast, [30] and [8] consider equal-sized communities with the same edge probability within each community.

In [8], [9], [30], the 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, i.e., the average degree of a node within the subgraph $Kp + (n-K)q$ is greater than nq, the average degree of a node outside. In this setting we show that the computational barrier disappears when side-information is available. Additionally, our results cannot be obtained as a special case of the results in [2], [8], [9], [30].

3) Summary of Results: We consider subgraph detection in $G(K, n, p, q)$ with two types of side-information:

1) A fraction α of subgraph nodes are revealed to the detector, which we call reliable cues. This represents the case of perfect side-information.
2) A similar number of nodes are marked as cues, but they are unreliable, i.e., imperfect side-information.

These two types of side-information are typical in semi-supervised clustering applications [5], [37], [38].

We use BP for subgraph detection to handle these two kinds of side-information. Our computations are local and distributed and require only neighbourhood information for each node in addition to the graph parameters p, q and K.

We analyze the detection performance of our algorithm when $p = a/n, q = b/n$ with a, b fixed and $K = \kappa n$ with κ fixed, as in the regime of [29]. We derive recursive equations for the distributions of BP messages in the limit as the graph size n tends to infinity. These recursions allow for numerical computation of the error rates for finite values of a, b and κ.

Based on these recursions, we obtain closed form expressions for the distributions when $a, b \to \infty$. We then show that when there is non-trivial side-information, the expected fraction of misclassified nodes goes to zero as $\kappa \to 0$, for any positive value of the phase transition parameter λ_α or λ, for perfect or imperfect side-information, made explicit later. Thus the computational barrier of $\lambda = 1/e$ for BP without side-information disappears when there is side-information.

We validate our theoretical findings by simulations. To demonstrate the practical usefulness of our algorithm we also apply it to subgraph detection on real-world datasets. The algorithm for imperfect side-information with its numerical validation on synthetic datasets was submitted for presented at ISIT 2017 [19]. The rest of the material, such as the algorithm for perfect side-information, all the proofs and numerical results on real-world datasets, is new in this journal version.

4) Organization: The rest of the paper is organized as follows. In Subsection I-5 we delineate useful notation. In Section II we describe the model and define the problem in detail. In Section III, we present our algorithm with perfect cues and explain the steps in its derivation. In Section IV we derive the asymptotic distribution of BP messages. In particular, in section IV-A, we prove our main result on the asymptotic error rate of our algorithm. In Section V we present our algorithm with imperfect side-information and provide a result on its asymptotic error rate. In Section VI we present results on our experiments on the synthetic graph as well as a few real-world graphs. In Section VII, we conclude with some suggestions for future work. Some proofs are relegated to supplementary material for lack of space.

5) Notation and Nomenclature: A graph node is denoted by a lower case letter such as i. The graph distance between two nodes i and j is the length of the shortest sequence of edges to go from i to j. The neighbourhood of a node i, denoted by δi is the set of one-hop neighbours of i, i.e., nodes that are at a graph distance of one. Similarly, we also work with t-hop neighbours of i, denoted as G_i^t, the set of nodes within a distance of t from i. Note that $G_i^1 = \delta i$. We use the following symbols to denote set operations: $C = A\backslash B$ is the set of elements that belong to A and not B and Δ denotes the symmetric difference, i.e., $A\Delta B = (A \cup B)\backslash(A \cap B)$. Also $|C|$ denotes the cardinality of the set C. The indicator function for an event A is denoted by $\mathbf{1}(A)$, i.e., $\mathbf{1}(A) = 1$ if A is true and 0 otherwise. The symbol $\sim$ denotes the distribution of a random variable (rv), for example $X \sim \text{Poi}(\gamma)$ means that X is a Poisson distributed rv with mean γ. Also, $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2. The symbol $\xrightarrow{D}$ denotes convergence in distribution.

II. Model and Problem Definition

Let $G(K, n, p, q)$ be a random undirected graph with n nodes and a hidden community S such that $|S| = K$. Let $\mathcal{G} = (V, E)$ be a realization of $G(K, n, p, q)$. An edge between two nodes appears independently of other edges such that $\mathbb{P}((i,j) \in E | i, j \in S) = p$ and $\mathbb{P}((i,j) \in E | i \in S, j \notin S) = \mathbb{P}((i,j) \in E | i, j \notin S) = q$. We assume that S is chosen uniformly from V among all sets of size K. Additionally let $p = a/n$ and $q = b/n$, where a and b are constants independent of n. Such graphs, with average degree $O(1)$, are called diluted graphs. We use a function $\sigma : V \to \{0,1\}^n$ to denote community membership such that $\sigma_i = 1$ if $i \in S$ and 0 otherwise. Next we describe the model for selecting C, the set of cues. To indicate which nodes are cues, we introduce a function $c : V \to \{0,1\}^n$ such that $c_i = 1$ if i is a cued vertex and $c_i = 0$ otherwise. The model for cues depends on the type of side-information: perfect or imperfect.

The side-information models are as follows:

1) **Perfect side-information**: In this case the cues are reliable, i.e., they all belong to the subgraph. To construct C we sample nodes as follows

$$\mathbb{P}(c_i = 1 | \sigma_i = x) = \begin{cases} \alpha & \text{if } x = 1 \\ 0 & \text{if } x = 0, \end{cases}$$

for some $\alpha \in (0, 1)$. Under this model we have

$$n\mathbb{P}(c_i = 1) = \sum_{i \in V} \mathbb{P}(c_i = 1 | \sigma_i = 1)\mathbb{P}(\sigma_i = 1) = \alpha K. \tag{2}$$

2) **Imperfect side-information**: Under imperfect side-information, the cues are unreliable. We generate C by sampling nodes from V as follows using a fixed $\beta \in (0, 1]$. For any $i \in V$:

$$\mathbb{P}\left(c_i = 1 | \sigma_i = x\right) = \begin{cases} \alpha\beta & \text{if } x = 1, \\ \frac{\alpha K(1-\beta)}{(n-K)} & \text{if } x = 0. \end{cases} \tag{3}$$

Under this model we have for any $i \in V$,

$$\begin{aligned} \mathbb{P}(c_i = 1) &= \mathbb{P}(\sigma_i = 1)\mathbb{P}(c_i = 1 | \sigma_i = 1) \\ &\quad + \mathbb{P}(\sigma_i = 0)\mathbb{P}(c_i = 1 | \sigma_i = 0) \\ &= \frac{K}{n}\alpha\beta + \frac{(n-K)}{n}\frac{\alpha K(1-\beta)}{(n-K)} \\ &= \alpha K/n; \end{aligned}$$

hence it matches with (2) of the perfect side-information case. It is easy to verify that under the above sampling

$$\mathbb{P}\left(\sigma_i = 1 | c_i = 1\right) = \beta, \tag{4}$$

which provides us with the interpretation of $|\log(\beta/(1-\beta))|$ as a reliability parameter for cue information.

Given G, C our objective is to infer the labels $\{\sigma_i, i \in V\backslash C\}$. The optimal detector minimizing the expected number of misclassified nodes is the per-node MAP detector given as [18]:

$$\hat{\sigma}_i = \mathbf{1}\left(R_i > \log\frac{\mathbb{P}(\sigma_i = 0)}{\mathbb{P}(\sigma_i = 1)}\right),$$

where

$$R_i = \log\left(\frac{\mathbb{P}(G, C | \sigma_i = 1)}{\mathbb{P}(G, C | \sigma_i = 0)}\right)$$

is a log-likelihood ratio of the detection problem. Observe that this detector requires the observation of the whole graph. Our objective then is to compute R_i for each i using a local BP algorithm and identify some parameter ranges for which it is useful. Specifically, we want to show that a certain barrier that exists for BP when $\alpha = 0$ disappears when $\alpha\beta > 0$.

III. Belief Propagation Algorithm for Detection with Perfect Side-information

In this section we present the BP algorithm, Algorithm 1, which performs detection in the presence of perfect side-information. In order to avoid erroneous likelihood computations, the number of steps t_f of the BP algorithm must be small enough not to encounter loops. By Lemma 3, this can be ensured with high probability if $t_f < \frac{\log(n)}{\log(np)} + 1$.

We provide here a brief overview of the algorithm. At step t of Algorithm 1, each node $u \in V\backslash C$ updates its own log-likelihood ratio based on its t-hop neighbourhood:

$$R_u^t := \log\left(\frac{\mathbb{P}(G_u^t, C_u^t | \sigma_u = 1)}{\mathbb{P}(G_u^t, C_u^t | \sigma_u = 0)}\right), \tag{5}$$

where G_u^t is the set of t-hop neighbours of u and C_u^t is the set of cues in G_u^t, i.e., $C_u^t = G_u^t \cap C$. The beliefs are updated according to (8). The messages transmitted to u by the nodes $i \in \delta u$, the immediate neighbourhood of u, are given by

$$R_{i\to u}^t := \log\left(\frac{\mathbb{P}(G_i^t\backslash u, C_i^t\backslash u|\sigma_i = 1)}{\mathbb{P}(G_i^t\backslash u, C_i^t\backslash u|\sigma_i = 0)}\right), \tag{6}$$

where $G_i^t\backslash u$ and $C_i^t\backslash u$ are defined as above, but excluding the contribution from node u. Node i updates $R_{i\to u}^t$ by acquiring messages from its neighbours, except u, and aggregating them according to (7). If node u is isolated, i.e., $\delta u = \emptyset$, there are no updates for this node. It can be checked that the total computation time for t_f steps of BP is $O(t_f|E|)$.

Algorithm 1 BP with perfect side-information

1: Initialize: Set $R_{i\to j}^0$ to 0, for all $(i,j) \in E$ with $i, j \notin C$. Let $t_f < \frac{\log(n)}{\log(np)} + 1$. Set $t = 0$.
2: For all directed pairs $(i,u) \in E$, such that $i, u \notin C$:

$$R_{i\to u}^{t+1} = -K(p-q) + \sum_{l\in C_i^1, l\neq u} \log\left(\frac{p}{q}\right) + \sum_{l\in\delta i\backslash C_i^1, l\neq u} \log\left(\frac{\exp(R_{l\to i}^t - \upsilon)(p/q)+1}{\exp(R_{l\to i}^t - \upsilon)+1}\right), \tag{7}$$

where $\upsilon = \log(\frac{n-K}{K(1-\alpha)})$.
3: Increment t, if $t < t_f - 1$ go back to 2, else go to 4
4: Compute $R_u^{t_f}$ for every $u \in V\backslash C$ as follows:

$$R_u^{t+1} = -K(p-q) + \sum_{l\in C_u^1} \log\left(\frac{p}{q}\right) + \sum_{l\in\delta u\backslash C_u^1} \log\left(\frac{\exp(R_{l\to u}^t - \upsilon)(p/q)+1}{\exp(R_{l\to u}^t - \upsilon)+1}\right) \tag{8}$$

5: The output set is the union of C and the $K - |C|$ set of nodes in $V\backslash C$ with the largest values of $R_u^{t_f}$.

The detailed derivation of the algorithm can be found in Appendix A. The derivation consists of two steps. First we establish a coupling between G_u^t, the t-hop neighbourhood of a node u of the graph and a specially constructed Galton-Watson (G-W) tree[1] T_u^t of depth t rooted on u. This coupling ensures that for a carefully chosen $t = t_f$ the neighbourhood $G_u^{t_f}$ of the node is a tree with probability tending to one as $n \to \infty$ (i.e., with high probability (w.h.p)). The second step of the derivation involves deriving the recursions (7) and (8) to compute (6) and (5) respectively, using the tree coupling.

The output of the algorithm is C along with the set of $K - |C|$ nodes with the largest value of log-likelihoods $R_i^{t_f}$. In the following section we derive the asymptotic distributions of the BP messages as the graph size tends to infinity, so as to quantify the error performance of the algorithm.

[1]Detailed in Appendix A

IV. Asymptotic Error Analysis

In this section we analyze the distributions of BP messages $R_{i\to u}^t$ given $\{\sigma_i = 1\}$ and given $\{\sigma_i = 0\}$ for $i \in V\backslash C$. First, we derive a pair of recursive equations for the asymptotic distributions of the messages $R_{i\to u}^t$ given $\{\sigma_i = 0, c_i = 0\}$ and given $\{\sigma_i = 1, c_i = 0\}$ in the limit as $n \to \infty$ in Lemma 1. In Proposition 1 we present the asymptotic distributions of the messages in the large degree regime where $a, b \to \infty$. This result will enable us to derive the error rates for detecting the subgraph in the large degree regime (Theorem 1). Finally, we contrast this result with Proposition 2 from [29], which details the limitation of local algorithms.

Instead of studying $R_{i\to u}^t$ directly, we look at the log-likelihood ratios of the posterior probabilities of σ_i given as

$$\widetilde{R}_i^t = \log\left(\frac{\mathbb{P}(\sigma_i = 1|G_i^t, C_i^t, c_i = 0)}{\mathbb{P}(\sigma_i = 0|G_i^t, C_i^t, c_i = 0)}\right)$$

and the associated messages $\widetilde{R}_{i\to u}^t$. By Bayes rule, $\widetilde{R}_{i\to u}^t = R_{i\to u}^t - \upsilon$, where

$$\upsilon = \log\left(\frac{\mathbb{P}(\sigma_i = 0|c_i = 0)}{\mathbb{P}(\sigma_i = 1|c_i = 0)}\right) = \log\left(\frac{n-K}{K(1-\alpha)}\right).$$

It is worthwhile to note that our analysis of BP recursions proceeds in two steps. In the first step, we keep the parameters b, a constant and we take the limit where the graph size n goes to infinity. This helps us to apply the tree coupling result in Lemma 3 and establish a set of recursive distributional equations, the exact solutions to which will provide us with the limiting distribution of BP messages in the limit $n \to \infty$. This result is given in Lemma 1. However, for finite degrees, this set of equations is difficult to solve and does not give any insight into the performance of our algorithm, although it can be solved by iterative methods. Therefore, in Proposition 1, we derive the limiting form of the distributions in the limit $b \to \infty$. Letting $b \to \infty$ facilitates the analysis by letting us use the CLT in Lemma 2 to arrive at a closed form expression of the distributions, which can in turn be used to show an error bound for BP given in Theorem 1.

Let ξ_0^t, ξ_1^t be rvs with the same distribution as the messages $\widetilde{R}_{i\to u}^t$ given $\{\sigma_i = 0, c_i = 0\}$ and given $\{\sigma_i = 1, c_i = 0\}$, respectively in the limit as $n \to \infty$. Based on the tree coupling in Lemma 3 of Appendix A, it can be shown that these rvs satisfy the recursive distributional evolutionary equations given in the following lemma.

Lemma 1 *The random variables ξ_0^t and ξ_1^t satisfy the following recursive distributional equations with initial conditions $\xi_0^0 = \xi_1^0 = \log\left(\kappa(1-\alpha)/(1-\kappa)\right)$.*

$$\xi_0^{(t+1)} \overset{\mathrm{D}}{=} h + \sum_{i=1}^{L_{0c}} \log(\rho) + \sum_{i=1}^{L_{00}} f(\xi_{0,i}^{(t)}) + \sum_{i=1}^{L_{01}} f(\xi_{1,i}^{(t)}) \tag{9}$$

$$\xi_1^{(t+1)} \overset{\mathrm{D}}{=} h + \sum_{i=1}^{L_{1c}} \log(\rho) + \sum_{i=1}^{L_{10}} f(\xi_{0,i}^{(t)}) + \sum_{i=1}^{L_{11}} f(\xi_{1,i}^{(t)}), \tag{10}$$

where $\overset{\mathrm{D}}{=}$ denotes equality in distribution, $h = -\kappa(a-b) - \upsilon$, $\rho := p/q = a/b$, and the function f is defined as

$$f(x) := \log\left(\frac{\exp(x)\rho + 1}{\exp(x) + 1}\right). \tag{11}$$

The rvs $\xi_{0,i}^t, i = 1, 2, \ldots$ are independent and identically distributed (iid) with the same distribution as ξ_0^t. Similarly $\xi_{1,i}^t, i = 1, 2, \ldots$ are iid with the same distribution as ξ_1^t. Furthermore, $L_{00} \sim Poi((1-\kappa)b), L_{01} \sim Poi(\kappa b(1-\alpha)), L_{10} \sim Poi((1-\kappa)b), L_{11} \sim Poi(\kappa a(1-\alpha)), L_{0c} \sim Poi(\kappa b\alpha)$ and $L_{1c} \sim Poi(\kappa p\alpha)$.

Proof: This follows from (7) and the tree coupling in Lemma 3 of Appendix A. ■

We define a *phase transition parameter* for the detection problem in the presence of perfect side-information as:

$$\lambda_\alpha = \frac{K^2(p-q)^2(1-\alpha)^2}{(n-K)q} = \frac{\kappa^2(a-b)^2(1-\alpha)^2}{(1-\kappa)b}, \quad (12)$$

where the factor $(1-\alpha)^2$ arises from the fact that we are now trying to detect a smaller subgraph of size $K(1-\alpha)$.

We now present one of our main results, on the distribution of BP messages in the limit of large degrees as $a, b \to \infty$ such that λ_α is kept fixed.

Proposition 1 *In the regime where λ_α and κ are held fixed and $a, b \to \infty$, we have*

$$\xi_0^{t+1} \xrightarrow{D} \mathcal{N}\left(-\log\frac{1-\kappa}{\kappa(1-\alpha)} - \frac{1}{2}\mu^{(t+1)}, \mu^{(t+1)}\right)$$

$$\xi_1^{t+1} \xrightarrow{D} \mathcal{N}\left(-\log\frac{1-\kappa}{\kappa(1-\alpha)} + \frac{1}{2}\mu^{(t+1)}, \mu^{(t+1)}\right).$$

The variance $\mu^{(t)}$ satisfies the following recursion with initial condition $\mu^{(0)} = 0$:

$$\mu^{(t+1)} = \lambda_\alpha\alpha\frac{1-\kappa}{(1-\alpha)^2\kappa} + \lambda_\alpha\mathbb{E}\left(\frac{(1-\kappa)}{\kappa(1-\alpha) + (1-\kappa)\exp(-\mu^{(t)}/2 - \sqrt{\mu^{(t)}}Z)}\right), \quad (13)$$

where the expectation is taken w.r.t. $Z \sim \mathcal{N}(0,1)$.

Before providing a short sketch of the proof of the above proposition, we state a CLT for Poisson sums from [17].

Lemma 2 *[17, Lemma 11] Let $S_\gamma = X_1 + X_2 + \ldots + X_{N_\gamma}$, where X_i, for $i = 1, 2, \ldots N_\gamma$, are independent, identically distributed rv with mean μ, variance σ^2 and $\mathbb{E}(|X_i^3|) \le g^3$, and for some $\gamma > 0$, N_γ is a Poi(γ) rv independent of X_i : $i = 1, 2, \ldots, N_\gamma$. Then*

$$\sup_x \left|\mathbb{P}\left(\frac{S_\gamma - \gamma\mu}{\sqrt{\gamma(\mu^2+\sigma^2)}}\right) - \Phi(x)\right| \le \frac{C_{BE}g^3}{\sqrt{\gamma(\mu^2+\sigma^2)^3}},$$

where $\Phi(x) = \mathbb{P}(Z \le x)$, with $Z \sim \mathcal{N}(0,1)$ and $C_{BE} = 0.3041$.

We now provide a sketch of the proof of Proposition 1; the details can be found in Appendix B.

Sketch of Proof of Proposition 1: The proof proceeds primarily by applying the expectation and variance operators to both sides of (9) and (10) and applying various reductions. First notice that when $a, b \to \infty$ and λ and κ are held constant, we have $\rho \to 1$ as follows:

$$\rho = a/b = 1 + \sqrt{\frac{\lambda_\alpha(1-\kappa)}{(1-\alpha)^2\kappa^2 b}}. \quad (14)$$

Then using Taylor's expansion of $\log(1+x)$ we can expand the function $f(x)$ in (11) up to second order as follows:

$$f(x) = (\rho-1)\frac{e^x}{1+e^x} - \frac{1}{2}(\rho-1)^2\left(\frac{e^x}{1+e^x}\right)^2 + O(b^{-3/2}). \quad (15)$$

We use these expansions to simplify the expressions for the means and variances of (9) and (10). Then, by a change of measure, we express them in terms of functionals of a single rv, ξ_1^t. We then use induction to show that the variance $\mu^{(t+1)}$ satisfies the recursion (13) and use Lemma 2 to prove Gaussianity. ■

In the following subsection, we use Proposition 1 to derive the asymptotic error rates of the detector in Algorithm 1.

A. Detection Performance

Let us use the symbol $\overline{S}$ to denote the subgraph nodes with the cued nodes removed, i.e., $\overline{S} = S \backslash C$. This is the set that we aim to detect. The output of Algorithm 1, $\widehat{S}$ is the set of nodes with the top $K - |C|$ beliefs. We are interested in bounding the expected number of misclassified nodes $\mathbb{E}(|\overline{S}\Delta\widehat{S}|)$. Let $\widehat{S}$ be the output set of the algorithm excluding cues since the cues are always correctly detected. Note that $|\overline{S}| = |\widehat{S}| = K - |C|$. To characterize the performance of the detector, we need to choose a performance measure. In [29], a rescaled probability of success was used to study the performance of a subgraph detector without cues, defined as

$$P_{\text{succ}}(\widehat{\sigma}) = \mathbb{P}(i \in \widehat{S} | i \in S) + \mathbb{P}(i \notin \widehat{S} | i \notin S) - 1, \quad (16)$$

where $\widehat{\sigma}_i = \mathbf{1}(i \in \widehat{S})$, and the dependence of $P_{\text{succ}}(\widehat{\sigma})$ on n is implicit. In our work, we study the following error measure, which is the average fraction of misclassified nodes, also considered in [17], which for the uncued case is defined as

$$\mathcal{E} := \frac{\mathbb{E}(|S\Delta\widehat{S}|)}{K}.$$

Observe that $0 \le \mathcal{E} \le 2$. In particular $\mathcal{E} = 2$ if the algorithm misclassifies all the subgraph nodes. We now show that these two measures are roughly equivalent. For simplicity we consider the case where there are no cues, but the extension to the cued case is straightforward. Since our algorithm always outputs K nodes as the subgraph, i.e., $|\widehat{S}| = K$, the following is true for any estimate $\widehat{\sigma}$ of σ :

$$r_n := \sum_{i=1}^n \mathbf{1}(\widehat{\sigma}_i = 0, i \in S) = \sum_{i=1}^n \mathbf{1}(\widehat{\sigma}_i = 1, i \notin S), \quad (17)$$

i.e., the number of misclassified subgraph nodes is equal to the number of misclassified nodes outside the subgraph. We can rewrite the error measure $\mathcal{E}$ in terms of r_n, since

$$\frac{|S\Delta\widehat{S}|}{K} = \frac{2r_n}{K}. \quad (18)$$

Next notice that we can rewrite $P_{\text{succ}}(\widehat{\sigma})$ as follows.

$$\begin{aligned} P_{\text{succ}}(\widehat{\sigma}) &= 1 - \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{P}(\widehat{\sigma}_i = 0 | i \in S) + \mathbb{P}(\widehat{\sigma}_i = 1 | i \notin S)\right) \\ &\overset{\text{(a)}}{=} 1 - \sum_{i=1}^{n}\left(\frac{\mathbb{P}(\widehat{\sigma}_i = 0, i \in S)}{K} + \frac{\mathbb{P}(\widehat{\sigma}_i = 1, i \notin S)}{n-K}\right) \\ &\overset{\text{(b)}}{=} 1 - \left(\frac{\mathbb{E}(r_n)}{K} + \frac{\mathbb{E}(r_n)}{n-K}\right) = 1 - \frac{n\mathbb{E}(r_n)}{K(n-K)}, \end{aligned} \tag{19}$$

where in step (a) we used Bayes rule with $\mathbb{P}(i \in S) = \frac{K}{n}$. Since $1 \leq \frac{n}{n-K} \leq 2$, we get

$$1 - 2\mathbb{E}(r_n)/K \leq P_{\text{succ}}(\widehat{\sigma}) \leq 1 - \mathbb{E}(r_n)/(K). \tag{20}$$

Hence from (18) and (20), $P_{\text{succ}}(\widehat{\sigma}) \to 1$ if and only if $\frac{\mathbb{E}(|S\Delta\widehat{S}|)}{K} \to 0$.

In the following proposition, we state and prove the main result concerning the asymptotic error performance of Algorithm 1.

Theorem 1 *For any $\lambda_\alpha > 0, \alpha > 0$,*

$$\lim_{b\to\infty}\lim_{n\to\infty}\frac{\mathbb{E}(|\overline{S}\Delta\widehat{S}|)}{K(1-\alpha)} \leq 2\sqrt{\frac{1-\kappa}{\kappa(1-\alpha)}}e^{-\frac{1}{8}\frac{\alpha\lambda_\alpha(1-\kappa)}{\kappa(1-\alpha)^2}}. \tag{21}$$

Consequently,

$$\lim_{\kappa\to 0}\lim_{b\to\infty}\lim_{n\to\infty}\frac{\mathbb{E}(|\overline{S}\Delta\widehat{S}|)}{K(1-\alpha)} = 0.$$

Proof: Let $\widehat{S}_0$ be the MAP estimator given by

$$\widehat{S}_0 = \left\{i : R_i^t > \log\frac{1-\kappa}{\kappa(1-\alpha)}\right\}. \tag{22}$$

Since $\widehat{S}$ is the set of nodes with the top $K - |C|$ beliefs, we have either $\widehat{S} \subset \widehat{S}_0$ or $\widehat{S}_0 \subset \widehat{S}$. Therefore,

$$\begin{aligned} |\overline{S}\Delta\widehat{S}| &\leq |\overline{S}\Delta\widehat{S}_0| + |\widehat{S}\Delta\widehat{S}_0| \\ &= |\overline{S}\Delta\widehat{S}_0| + |K - |C| - |\widehat{S}_0|| \\ &= |\overline{S}\Delta\widehat{S}_0| + ||\overline{S}| - |\widehat{S}_0|| \\ &\leq 2|\overline{S}\Delta\widehat{S}_0|, \end{aligned} \tag{23}$$

where the last step follows because the symmetric difference between two sets is lower bounded by the difference of their sizes. If we can bound $\frac{\mathbb{E}(|\overline{S}\Delta\widehat{S}_0|)}{K(1-\alpha)}$ by one-half the expression in (21) the result of the Proposition follows. The proof of this upper bound uses Proposition 1 and is given in Appendix C. ∎

Theorem 1 states that the detectability threshold does not exist for BP with cues.

Note that Algorithm 1 is local until the last step of picking $K - |C|$ nodes with the largest values of $R_i^{t_f}$, which requires a fusion node to implement. However, as can be observed from (22) and (23), the statement about the error performance remains true even for a local version of this algorithm where each node independently employs a suitable thresholding of the final belief $R_i^{t_f}$. In conclusion, a local algorithm consisting of BP followed by thresholding of the beliefs is sufficient to get close to the maximum-likelihood algorithm's performance.

This is in stark contrast to the performance of BP when there is no side-information. In that case, as stated in the following theorem from [29], the performance of any local algorithm suffers when the parameter $\lambda < 1/e$. Thus, there is a discontinuity in the performance of BP in the sense that when $\alpha > 0$, a sharp threshold that exists in the absence of side-information disappears.

In the following LOC denotes the class of all local algorithms, i.e., algorithms that take as input the local neighbourhood of a node.

Proposition 2 *[29, Theorem 1] If $\lambda < 1/e$, then all local algorithms have success probability uniformly bounded away from one; in particular,*

$$\sup_{T\in\text{LOC}}\lim_{n\to\infty} P_{\text{succ}}(T) \leq \frac{e-1}{4},$$

and therefore

$$\inf_{T\in\text{LOC}}\lim_{n\to\infty}\mathcal{E}(T) \geq \frac{5-e}{4} > 1/2.$$

V. Imperfect Side Information

In this section, we develop a BP algorithm under the more realistic assumption of imperfect side information, where the available cue information is not completely reliable. This is true of humanly classfied data available for many semi-supervised learning problems.

Our BP algorithm can easily take into account imperfection in side information. Suppose we know the parameters α and β defined in (2) and (4) respectively, or their estimates thereof. We remark that unlike Algorithm 1, which only has to detect the uncued subgraph nodes, our algorithm needs to explore the whole graph, since we do not know a priori which cues are correct. As before, for a node u, we wish to compute the following log-likelihood ratio in a distributed manner:

$$R_u^t = \log\left(\frac{\mathbb{P}(G_u^t, c_u, C_u^t | \sigma_u = 1)}{\mathbb{P}(G_u^t, c_u, C_u^t | \sigma_u = 0)}\right),$$

where c_u is the indicator variable of whether u is a cued node, and C_u^t is the cued information of the t-hop neighbourhood of u, excluding u. Note that we can expand R_u^t as follows

$$\begin{aligned} R_u^t &= \log\left(\frac{\mathbb{P}(G_u^t, C_u^t | \sigma_u = 1, c_u)}{\mathbb{P}(G_u^t, C_u^t | \sigma_u = 0, c_u)}\right) + \log\left(\frac{\mathbb{P}(c_u | \sigma_u = 1)}{\mathbb{P}(c_u | \sigma_u = 0)}\right) \\ &= \log\left(\frac{\mathbb{P}(G_u^t, C_u^t | \sigma_u = 1)}{\mathbb{P}(G_u^t, C_u^t | \sigma_u = 0)}\right) + \log\left(\frac{\mathbb{P}(c_u | \sigma_u = 1)}{\mathbb{P}(c_u | \sigma_u = 0)}\right), \end{aligned} \tag{24}$$

where in the second step we dropped the conditioning w.r.t. c_u because (G_u^t, C_u^t) is independent of the cue information of node u given σ_u. Let $h_u = \log\left(\frac{\mathbb{P}(c_u|\sigma_u=1)}{\mathbb{P}(c_u|\sigma_u=0)}\right)$. Then it is easy to see from (3) that

$$h_u = \begin{cases} \log\left(\frac{\beta(1-\kappa)}{(1-\beta)\kappa}\right), & \text{if } u \in C, \\ \log\left(\frac{(1-\alpha\beta)(1-\kappa)}{(1-\kappa-\alpha\kappa+\alpha\kappa\beta)}\right), & \text{otherwise.} \end{cases} \tag{25}$$

Algorithm 2 BP with imperfect cues

1: Initialize: Set $R^0_{i\to j}$ to 0, for all $(i,j)\in E$. Let $t_f < \frac{\log(n)}{\log(np)}+1$. Set $t=0$.
2: For all directed pairs $(i,u)\in E$:

$$R^{t+1}_{i\to u} = -K(p-q)+h_i + \sum_{l\in\delta i, l\neq u}\log\left(\frac{\exp(R^t_{l\to i}-\nu)(p/q)+1}{\exp(R^t_{l\to i}-\nu)+1}\right), \tag{26}$$

where $\nu=\log(\frac{n-K}{K})$.
3: Increment t; if $t<t_f-1$ go back to 2, else go to 4
4: Compute $R^{t_f}_u$ for every $u\in V$ as follows:

$$R^{t+1}_u = -K(p-q)+h_u + \sum_{l\in\delta u}\log\left(\frac{\exp(R^t_{l\to u}-\nu)(p/q)+1}{\exp(R^t_{l\to u}-\nu)+1}\right) \tag{27}$$

5: Output $\widehat{S}$ as K set of nodes in V with the largest values of $R^{t_f}_u$.

The recursion for the first term in (24) can be derived along the same lines as the derivation of Algorithm 1 and is skipped. The final BP recursions are given in Algorithm 2.

In order to analyze the error performance of this algorithm we derive the asymptotic distributions of the messages $R^t_{u\to i}$, for $\{\sigma_u=0\}$ and $\{\sigma_u=1\}$. Note that, since we now assume that we do not know the exact classification of any of the subgraph nodes, we need to detect K nodes, and hence the phase transition parameter is defined as

$$\lambda = \frac{K^2(p-q)^2}{(n-K)q}. \tag{28}$$

The following proposition presents the asymptotic distribution of the messages $R^t_{u\to i}$ in the limit of $n\to\infty$ and in the large degree regime where $a,b\to\infty$.

Proposition 3 *Let $n\to\infty$. In the regime where λ and κ are held fixed and $a,b\to\infty$, the message $R^t_{u\to i}$ given $\{\sigma_u=j\}$, where $j=\{0,1\}$ converges in distribution to $\Gamma^t_j+h_u$ where h_u is defined in (25). The rvs Γ^t_j have the following distribution:*

$$\Gamma^t_0\sim\mathcal{N}(-\mu^{(t)}/2,\mu^{(t)}),\ \text{and}\ \Gamma^t_1\sim\mathcal{N}(\mu^{(t)}/2,\mu^{(t)}),$$

where $\mu^{(t)}$ satisfies the following recursion with $\mu^{(0)}=0$,

$$\mu^{(t+1)} = \alpha\beta^2\lambda\mathbb{E}\left(\frac{(1-\kappa)/\kappa}{\beta+(1-\beta)e^{(-\mu^{(t)}/2-\sqrt{\mu^{(t)}}Z)}}\right)+(1-\alpha\beta)^2\lambda \mathbb{E}\left(\frac{(1-\kappa)}{\kappa(1-\alpha\beta)+(1-\kappa-\alpha\kappa+\alpha\kappa\beta)e^{(-\mu^{(t)}/2-\sqrt{\mu^{(t)}}Z)}}\right), \tag{29}$$

and the expectation is w.r.t. (with respect to) $Z\sim\mathcal{N}(0,1)$.

Proof: The proof proceeds by deriving the recursive distributional equations that the message distributions satisfy in the limit $n\to\infty$, and then applying the large degree limit of $a,b\to\infty$ to these recursions.The details are in the supplementary material. ∎

The above proposition immediately leads to the following result on the asymptotic error rate of Algorithm 2.

Theorem 2 *For any $\lambda>0,\alpha>0,\beta>0$,*

$$\lim_{b\to\infty}\lim_{n\to\infty}\frac{\mathbb{E}(|\hat{S}\Delta S|)}{K} \leq 2\left(\alpha\sqrt{\beta(1-\beta)}+\sqrt{(1-\alpha\beta)(\frac{1-\kappa}{\kappa}-\alpha(1-\beta))}\right)e^{-\frac{\lambda\alpha\beta^2(1-\kappa)}{8\kappa}}.$$

Consequently,

$$\lim_{\kappa\to0}\lim_{b\to\infty}\lim_{n\to\infty}\frac{\mathbb{E}(|S\Delta\widehat{S}|)}{K}=0.$$

Proof: The proof essentially analyzes the properties of the recursion (29) and is similar to the proof of Theorem 1. See supplementary material for details. ∎

Note on number of steps for asymptotic weak recovery: It is worthwhile to note that, while running BP for t_f number of steps, where t_f is defined in Algorithms 1 and 2 leads to improved error performance, in order to get asymptotic weak recovery, it is sufficient to run BP for one step. This follows because the error bounds we derived in Theorems 1 and 2 rely on a lower bound on the asymptotic mean of the messages $\mu^{(t)}$, which is valid for any $t>0$, specifically for $t=1$.

VI. Numerical Experiments

In this section we provide numerical results to validate our theoretical findings on the synthetic model as well as on two real-world datasets. We compare the performance of BP to another seed-based community detection algorithm, the personalized PageRank, which is widely used for local community detection [4].

A. Synthetic dataset

First we show that the limitation of local algorithms described in Proposition 2 is overcome by BP when there is non-trivial side-information. Proposition 2 says that when $\lambda<1/e$, $\mathcal{E}(T)>1/2$ for any local algorithm T. We run our Algorithm 1, on a graph generated with $\alpha=0.1,\kappa=5\times10^{-4},b=100$ and $n=10^6$. For $\lambda=1/4<1/e$, we get an average value of $\mathcal{E}=0.228<1/2$. Thus it is clear that our algorithm overcomes the computational threshold of $\lambda=1/e$.

Next, we study the performance of Algorithm 2 when there is noisy side-information with $\beta=0.8$. For $\lambda=1/3<1/e$, we get an average error rate of $0.3916<1/2$ clearly beating the threshold of $\lambda=1/e$. Thus we have demonstrated that both with perfect and imperfect side-information, our algorithm overcomes the $\lambda=1/e$ barrier of local algorithms.

Next, we verify that increasing α improves the performance of our algorithm as expected. In Figure 1, we plot the variation of $\mathcal{E}$ of Algorithm 1 as a function of α. Our parameter setting

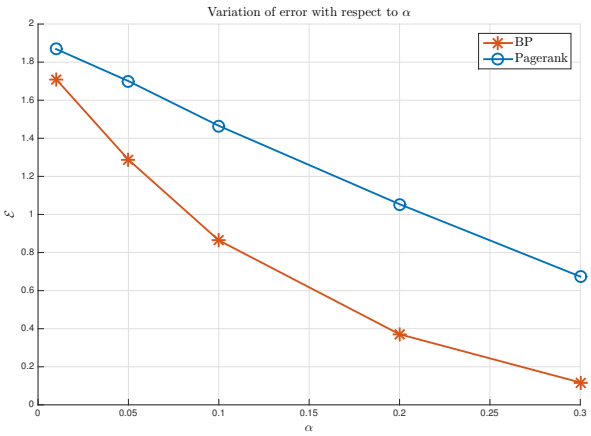

Fig. 1: Performance of BP Algo 1 as a function of α

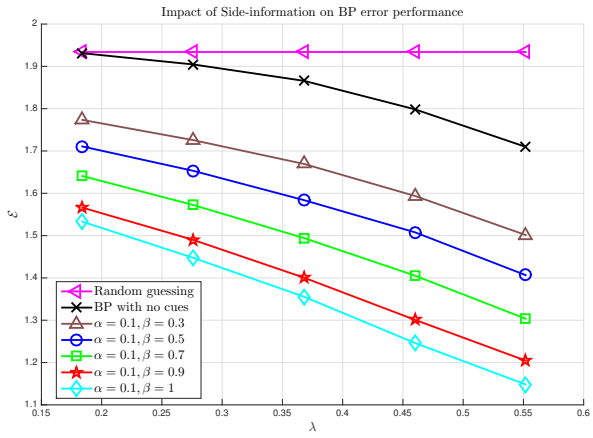

Fig. 2: Comparison of BP for subgraph detection for different amounts of side-information

is $\kappa = 0.01, b = 100$, and $\lambda = 1/2$ with $n = 10^4$. In the figure, we also plot the error rate $\mathcal{E}$ obtained by personalized PageRank under the same setting, with damping factor $\alpha_{pr} = 0.9$ [4]. The figure demonstrates that BP benefits more as the amount of side-information is increased than PageRank does.

Next, we compare the performance of BP algorithm without side-information given in [29] to our algorithm with varying amounts of side-information. We choose the setting where $n = 10^4, b = 140$ and $\kappa = 0.033$ for different values of λ by varying p. In Figure 2 we plot the metric $\mathcal{E}$ against λ for different values of β, with $\alpha = 0.1$. For $\beta = 1$ we use Algorithm 1. We can see that even BP with noisy side-information performs better than standard BP with no side-information. In addition, as expected increasing β improves the error performance.

Finally, we note that as observed in the previous section, in terms of asymptotic error performance in the presence of side-information, even BP run for a single step achieves asymptotically zero error. In order to assess the performance of BP when run for more than one step, we simulated a case where $n = 10^4$,$\kappa = 0.01, b = 100$ and $\lambda = 1/2$ with $\alpha = .05$ and $\beta = 1$. In Figure 3, we plot the error $\mathcal{E}$ for up to 3 steps of BP. Clearly, the performance of BP for only one step is quite unsatisfactory. This finding makes a case for running BP for more than one step. In the future, we hope to correctly characterize the error improvement when BP is run for more than one step.

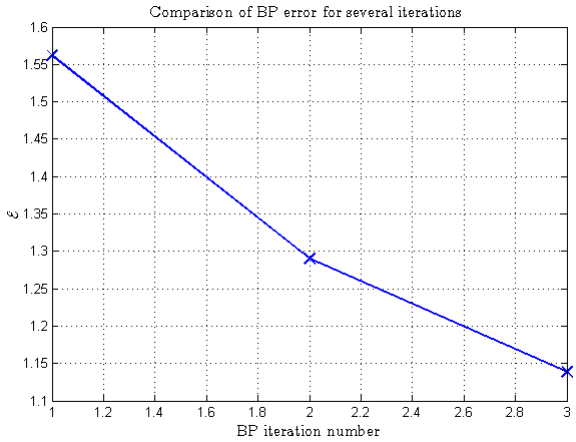

Fig. 3: BP error performance vs number of steps of BP

B. Real-world datasets

We consider two real-world networks: The USPS dataset and the Reuters-911 dataset. For these two datasets we compare the performance of BP with personalized PageRank in terms of recall rate $\mathcal{R}$ defined as

$$\mathcal{R} = \frac{|S \cap \hat{S}|}{|S|},$$

where S is the true community and $\hat{S}$ is its estimate. This is a commonly used metric for community detection applications [35]. We use $\alpha_{\text{pr}} = 0.9$ as the damping factor of PageRank. We describe the datasets and the results obtained by our algorithms below.

1) USPS dataset: The USPS dataset contains 9296 scanned images of size 16 $\times$ 16, which can represented by a feature vector of size 256 $\times$ 1 with values from -1 to +1 [37]. First, we construct a graph from this dataset, where nodes represent scanned images, by adding a link between a node and its three nearest neighbours, where the distance is defined as the euclidean distance between the images represented as feature vectors. The resulting graph is undirected with a minimum degree of at least 3. This is an instance of the k nearest neighbour graph, with $k = 3$.

On this graph we run BP and PageRank separately for each of the 10 communities for $\alpha = 0.01$ and $\alpha = 0.05$ (Figure 4). It can be seen from Figure 4, that the performance of BP is strictly worse than that of PageRank. This result points to the importance of having the correct initialization for the BP parameters. Indeed, in our underlying model for BP, we assumed that there is only one dense community in a sparse network, in which case, as demonstrated in Figure1, BP outperforms PageRank by a big margin. However in the USPS graph, there are ten dense communities, and therefore

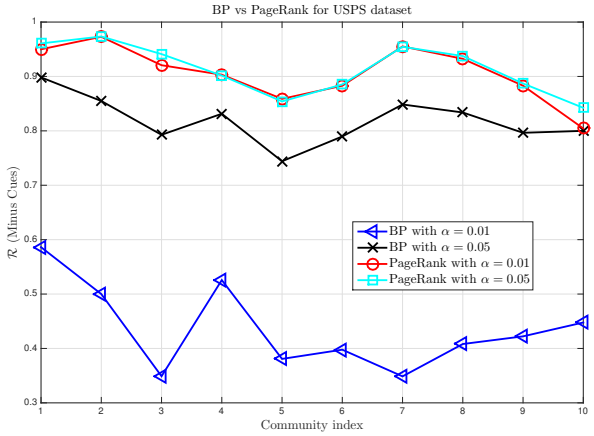

Fig. 4: Comparison of BP for subgraph detection for different amounts of side-information

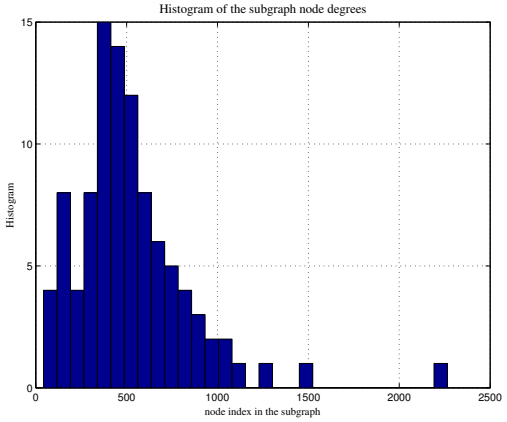

Fig. 5: Histogram of the degrees of the nodes in the dense subgraph in the Reuters graph

it deviates significantly from our underlying model.

2) Reuters911 Dataset: In this subsection we consider a graph that is closer to our assumed model. We consider the Reuters911 dataset also used in [11]. It is made up of words from all news released by Reuters for 66 days since September 11, 2001. Table 5 in [11] shows a group of 99 collocated words in this dataset.This subset represents the largest dense community to be detected in this dataset with an average degree of 520 while the average degree of nodes outside this subgraph is 18.48. A graph of size $n = 13332$ is generated from this dataset by adding a link between two words if they appear together in a sentence. The resulting graph is undirected and unweighted. We compare BP and Pagerank on this dataset for one and two cues. The cues we use are the words *pentagon* and *11*, with node degrees 432 and 43 respectively. In Figures 5, 6, we provide for illustrative purposes, the histogram of the degrees inside the subgraph and outside the subgraph respectively. In Table I we show the recall values $\mathcal{R}$ of PageRank and BP, excluding cues. Clearly, BP performs better.

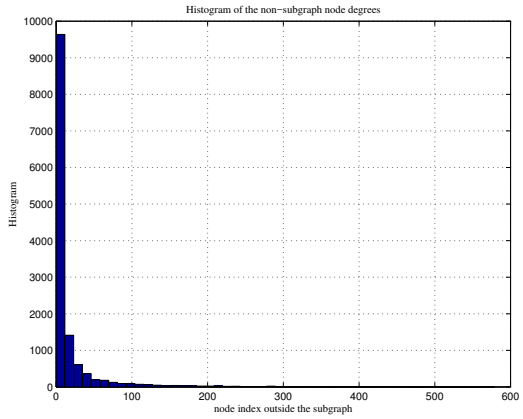

Fig. 6: Histogram of the degrees of the nodes outside the target subgraph in the Reuters graph

Class 0	#of cues = 1	#of cues = 2
BP	**0.7143**	**0.7216**
PageRank	0.6327	0.6392

TABLE I: Reuters911 recall results

VII. Conclusions and Future Extensions

In this work we developed a local distributed BP algorithm that takes advantage of side-information to detect a dense subgraph embedded in a sparse graph. We obtained theoretical results based on density evolution on trees to show that it achieves zero asymptotic error regardless of the parameter λ, unlike BP without cues, where there is a non-zero detectability threshold. We then validated our theoretical results by simulating our algorithm on a synthetic dataset and showing that, in the presence of both noise-less and noisy side-information, our BP algorithm overcomes the error bound of local algorithms when $\lambda < 1/e$.

An intuition why the side information we consider is so effective is that it provides global knowledge of the subgraph location, in addition to the local information available to the BP algorithm. Our results are also in line with the similar drastic improvements brought about by side-information in local community detection in other models as reported in [21], [36].

We then applied our algorithm to two real-world datasets: USPS and Reuters911 and compared its performance with personalized PageRank. Our results indicate that the relative improvement in BP depends on the closeness of the dataset to the underlying graph model used to derive BP. In the future, we would like to do non-asymptotic analysis when a, b and κ are functions of n. Extension to dense graphs would also be interesting, where traditional BP and tree coupling-based analysis will not work owing to the presence of loops.

It is also important to note that the BP algorithm overcomes the $1/e$ barrier because the fraction of cued nodes α is strictly positive. This naturally leads one to a question for the future on whether there is a critical rate for $\alpha \to 0$ above which local algorithms with cued nodes continue to perform as well as maximum-likelihood detection.

Finally, we would like to note that the simple SBM graph used in this paper is not often a good fit for graphs encountered in practice. There exist several modifications of the SBM in the literature such as the degree-corrected SBM [31] and the Random Intersection Model [12]. Generative models like SBM provide a good analytical platform for studying the impact of side-information on community detection algorithms, and can lead to the study of more realistic models in the future.

ACKNOWLEDGEMENTS

This work was partly funded by the French Government (National Research Agency, ANR) through the "Investments for the Future" Program reference #ANR-11-LABX-0031-01 and Indo-French CEFIPRA Collaboration Grant No.5100-IT1 "Monte Carlo and Learning Schemes for Network Analytics." We would like to thank Bruce Hajek for very useful comments that helped to improve this paper.

APPENDIX A
DESCRIPTION OF G-W TREE AND DERIVATION OF ALGORITHM 1

We derive Algorithm 1 by establishing a coupling formulation between a t-hop neighbourhood G_u^t of node u and a Galton-Watson (G-W) tree rooted at u constructed as follows. Let T_u^t be a labelled Galton-Watson (G-W) tree of depth t rooted at node u constructed as follows (as in [17]): The label τ_u at node u is chosen at random in the following way:

$$\mathbb{P}(\tau_u = 1) = \frac{K}{n}, \qquad \mathbb{P}(\tau_u = 0) = \frac{n-K}{n}.$$

The number of children N_u of the root u is Poisson-distributed with mean $d_1 = Kp + (n-K)q$ if $\tau_u = 1$ and mean $d_0 = nq$ if $\tau_u = 0$. Each child is also assigned a label. The number of children i with label $\tau_i = 1$ is Poisson distributed with mean Kp if $\tau_u = 1$ and mean Kq if $\tau_i = 0$. The number of children with label $\tau_i = 0$ is Poisson distributed with mean $(n-K)q$ for both $\tau_u = 0$ and $\tau_u = 1$. By the independent splitting property of Poisson random variables, this is equivalent to assigning the label $\tau_i = 1$ to each child i by sampling a Bernoulli random variable with probability (w.p.) Kp/d_1 if $\tau_u = 1$ and Kq/d_0 if $\tau_u = 0$. Similarly $\tau_i = 0$ w.p. $(n-K)q/d_1$ and $(n-K)q/d_0$ for $\tau_u = 0$ and 1 respectively. Namely, if i is a child of u,

$$\mathbb{P}(\tau_i = 1|\tau_u = 1) = \frac{Kp}{d_1}, \quad \mathbb{P}(\tau_i = 1|\tau_u = 0) = \frac{Kq}{d_0}. \quad (30)$$

We then assign the cue indicator function $\widetilde{c}$ such that $\widetilde{c}_i = 1$ w.p. α if $\tau_i = 1$ and $\widetilde{c}_i = 0$ if $\tau_i = 0$. The process is repeated up to depth t giving us $\widetilde{C}_u^t$, the set of cued neighbours. Now we have the following coupling result between (G_u^t, σ^t, C_u^t), the neighbourhood of u and the node labels of that neighbourhood and $(T_u^t, \tau^t, \widetilde{C}_u^t)$, the depth-$t$ tree T_u^t and its labels due to [17].

Lemma 3 *[17, Lemma 15] For t such that $(np)^t = n^{o(1)}$, there exists a coupling such that $(G_u^t, \sigma^t, C_u^t) = (T_u^t, \tau^t, \widetilde{C}_u^t)$ with probability $1 - n^{-1+o(1)}$, where equality of graphs means that the GW tree T_u^t is identical in topology to G_u^t with node labels τ^t and cues C_u^t identical to σ^t and $\widetilde{C}_u^t$, respectively.*

We now derive the recursions for the likelihood ratios on the tree T_u^t. For large n with high probability, by the coupling formulation, R_u^t also satisfy the same recursions. For notational simplicity, from here onwards we represent the cue labels on the tree by c and the set of cued neighbours by C_u^t, just as for the original graph. We use Λ_u^t to denote the likelihood ratio of node u computed on a tree defined as below:

$$\Lambda_u^{t+1} = \log\left(\frac{\mathbb{P}(T_u^{t+1}, C_u^{t+1}|\tau_u = 1)}{\mathbb{P}(T_u^{t+1}, C_u^{t+1}|\tau_u = 0)}\right).$$

By virtue of tree construction, if the node u has N_u children, the N_u subtrees rooted on these children are jointly independent given τ_u thanks to the coupling property in Lemma 3 [2]. We use this fact to split Λ_u^{t+1} in two parts.

$$\begin{aligned}\Lambda_u^{t+1} &= \log\left(\frac{\mathbb{P}(T_u^{t+1}, C_u^{t+1}|\tau_u = 1)}{\mathbb{P}(T_u^{t+1}, C_u^{t+1}|\tau_u = 0)}\right) \\ &= \log\left(\frac{\mathbb{P}(N_u|\tau_u = 1)}{\mathbb{P}(N_u|\tau_u = 0)}\right) + \quad (31) \\ &\sum_{i\in\delta u} \log\left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 0)}\right), \quad (32)\end{aligned}$$

by the independence property of subtress T_i^t rooted on $i \in \delta u$. Since by Lemma 3, the degrees are Poisson,

$$\mathbb{P}(N_u|\tau_u = 1) = d_1^{N_u} e^{-d_1}/N_u!,$$

and similarly for $\mathbb{P}(N_u|\tau_u = 0)$. Therefore we have

$$\begin{aligned}\log\left(\frac{\mathbb{P}(N_u|\tau_u = 1)}{\mathbb{P}(N_u|\tau_u = 0)}\right) &= N_u \log\left(\frac{d_1}{d_0}\right) - (d_1 - d_0) \\ &= N_u \log\left(\frac{d_1}{d_0}\right) - K(p-q). \quad (33)\end{aligned}$$

Next we look at the second term in (32). We analyze separately the case of $c_i = 1$ and $c_i = 0$ for $i \in \delta_u$, i.e, the cued and uncued children are handled separately.

Case 1 ($c_i = 1$): We have

$$\begin{aligned}&\log\left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 0)}\right) \\ &\overset{(a)}{=} \log\left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t, \tau_i = 1|\tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t, \tau_i = 1|\tau_u = 0)}\right) \\ &= \log\left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_i = 1)\mathbb{P}(\tau_i = 1|\tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_i = 1)\mathbb{P}(\tau_i = 1|\tau_u = 0)}\right) \\ &\overset{(b)}{=} \log\left(\frac{Kp/d_1}{Kq/d_0}\right), \quad (34)\end{aligned}$$

where in step (a) we applied the fact that $c_i = 1$ implies $\tau_i = 1$, and in (b) we used (30).

Case 2 ($c_i = 0$): Observe that $\mathbb{P}(c_i = 0|\tau_i = 1) = 1 - \alpha$ and $\mathbb{P}(c_i = 0|\tau_i = 0) = 1$. Note that

[2] Under the assumptions of Lemma 3 we can construct a Poisson tree identical to the t-hop neighbourhood of a given node. In addition, it can be concluded from the Lemma that all the subtrees of this tree are non-overlapping and hence independent.

$$\begin{aligned}
&\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 1)\\
&\quad = \quad \mathbb{P}(T_i^t, C_i^t|\tau_i = 1)\mathbb{P}(c_i|\tau_i = 1)\mathbb{P}(\tau_i = 1|\tau_u = 1)\\
&\qquad + \quad \mathbb{P}(T_i^t, C_i^t|\tau_i = 0)\mathbb{P}(c_i|\tau_i = 0)\mathbb{P}(\tau_i = 0|\tau_u = 1)\\
&\quad = \quad \mathbb{P}(T_i^t, C_i^t|\tau_i = 1)(1-\alpha)\frac{Kp}{d_1}\\
&\qquad + \quad \mathbb{P}(T_i^t, C_i^t|\tau_i = 0)\frac{(n-K)q}{d_1}. \qquad (35)
\end{aligned}$$

Similarly, we can show

$$\begin{aligned}
\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 0) = \mathbb{P}(T_i^t, C_i^t|\tau_i = 1)\frac{Kq}{d_0}(1-\alpha)\\
+ \quad \mathbb{P}(T_i^t, C_i^t|\tau_i = 0)\frac{(n-K)q}{d_0}. \qquad (36)
\end{aligned}$$

Let us define

$$\Lambda_{i\to u}^t := \log\left(\frac{\mathbb{P}(T_i^t, C_i^t|\tau_i = 1)}{\mathbb{P}(T_i^t, C_i^t|\tau_i = 0)}\right),$$

the message that i sends to u at step t. Using the above definition, (35), and (36) we get

$$\begin{aligned}
&\log\left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 0)}\right)\\
&\quad = \quad \log\left(\frac{e^{\Lambda_{i\to u}^t}\frac{Kp}{d_1}(1-\alpha) + \frac{(n-K)q}{d_1}}{e^{\Lambda_{i\to u}^t}\frac{Kq}{d_0}(1-\alpha) + \frac{(n-K)q}{d_0}}\right)\\
&\quad = \quad \log\left(\frac{d_0}{d_1}\right) + \log\left(\frac{e^{\Lambda_{i\to u}^t}\frac{Kp}{(n-K)q}(1-\alpha)+1}{e^{\Lambda_{i\to u}^t}\frac{K}{(n-K)}(1-\alpha)+1}\right) \qquad (37)
\end{aligned}$$

We then use the substitution $\nu := \log((n-K)/K)$ in the above equation. Finally combining (33), (34) and (37) and replacing Λ_u^t with R_u^t and $\Lambda_{i\to u}^t$ with $R_{i\to u}^t$, we arrive at (8). The recursive equation (7) can be derived in exactly the same way by looking at the children of $i \in \delta u$.

Appendix B
Proof of Proposition 1

Since the statistical properties of R_u^t and Λ_u^t are the same in the $n \to \infty$ limit, we analyze the distribution of Λ_u^t. Let us define the posterior likelihood for τ_u given by

$$\widetilde{\Lambda}_i^t = \log\left(\frac{\mathbb{P}(\tau_i = 1|T_i^t, C_i^t, c_i = 0)}{\mathbb{P}(\tau_i = 0|T_i^t, C_i^t, c_i = 0)}\right).$$

Note that $\mathbb{P}(\tau_i = 1|c_i = 0) = \kappa(1-\alpha)/(1-\kappa\alpha)$ and $\mathbb{P}(\tau_i = 0|c_i = 0) = (1-\kappa)/(1-\kappa\alpha)$ are the prior probabilities of the uncued vertices. For convenience we use an overline for the symbols of expectation $\overline{\mathbb{E}}$ and probability $\overline{\mathbb{P}}$ to denote conditioning w.r.t $\{c_i = 0\}$.

By a slight abuse of notation, let ξ_0^t and ξ_1^t denote the rvs whose distributions are the same as the distributions of $\widetilde{\Lambda}_i^t$ given $\{c_i = 0, \tau_i = 0\}$ and $\{c_i = 0, \tau_i = 1\}$ respectively in the limit $n \to \infty$. We need a relationship between P_0 and P_1, the probability measures of ξ_0^t and ξ_1^t respectively, stated in the following lemma.

Lemma 4

$$\frac{dP_0}{dP_1}(\xi) = \frac{\kappa(1-\alpha)}{1-\kappa}\exp(-\xi).$$

In other words for any integrable function $g(\cdot)$

$$\overline{\mathbb{E}}[g(\widetilde{\Lambda}_u^t)|\tau_u = 0] = \frac{\kappa(1-\alpha)}{1-\kappa}\overline{\mathbb{E}}[g(\widetilde{\Lambda}_u^t)e^{-\widetilde{\Lambda}_u^t}|\tau_u = 1].$$

Proof: Following the logic in [29], we show this result for $g(\widetilde{\Lambda}_u^t) = \mathbf{1}(\widetilde{\Lambda}_u \in A)$, A being some measurable set. The result for general g then follows because any integrable function can be obtained as the limit of a sequence of such rvs [7]. Let $Y = (T_u^t, C_u^t)$, the observed rv. Therefore

$$\begin{aligned}
\overline{\mathbb{E}}\left(\mathbf{1}\left(\widetilde{\Lambda}_u^t \in A\right)|\tau_u = 0\right) &= \overline{\mathbb{P}}\left(\widetilde{\Lambda}_u^t \in A|\tau_u = 0\right)\\
&= \frac{\overline{\mathbb{P}}(\widetilde{\Lambda}_u^t \in A, \tau_u = 0)}{\overline{\mathbb{P}}(\tau_u = 0)}\\
&= \frac{\overline{\mathbb{E}}_Y\left(\overline{\mathbb{P}}(\widetilde{\Lambda}_u^t \in A, \tau_u = 0|Y)\right)}{\overline{\mathbb{P}}(\tau_u = 0)}\\
&= \overline{\mathbb{E}}_Y\left[\frac{\mathbf{1}(\widetilde{\Lambda}_u^t \in A)\overline{\mathbb{P}}(\tau_u = 0|Y)}{\overline{\mathbb{P}}(\tau_u = 0)}\right]\\
&\overset{(a)}{=} \overline{\mathbb{E}}_Y\left(\frac{\mathbf{1}(\widetilde{\Lambda}_u^t \in A)\overline{\mathbb{P}}(\tau_u = 1|Y)}{e^{\widetilde{\Lambda}_u^t}\overline{\mathbb{P}}(\tau_u = 0)}\right)\\
&= \frac{\overline{\mathbb{P}}(\tau_u = 1)}{\overline{\mathbb{P}}(\tau_u = 0)}\overline{\mathbb{E}}_1(\mathbf{1}(\widetilde{\Lambda}_u^t \in A)e^{-\widetilde{\Lambda}_u^t})\\
&= \frac{\kappa(1-\alpha)}{1-\kappa}\overline{\mathbb{E}}_1(\mathbf{1}(\widetilde{\Lambda}_u^t \in A)e^{-\widetilde{\Lambda}_u^t}),
\end{aligned}$$

where in (a) we used the fact that $\frac{\overline{\mathbb{P}}(\tau_u=0|Y)}{\overline{\mathbb{P}}(\tau_u=1|Y)} = \exp(-\widetilde{\Lambda}_u^t)$, and $\mathbb{E}_1$ denotes expectation conditioned on the event $\{\tau_u = 1\}$. ■

Proof:

Since λ_α and κ are fixed and $b \to \infty$, from (12) we have

$$\rho := a/b = 1 + \sqrt{\frac{\lambda_\alpha(1-\kappa)}{(1-\alpha)^2\kappa^2 b}} = 1 + O(b^{-1/2}). \qquad (38)$$

Following [29], we prove the result by induction on t. First let us verify the result holds when $t = 0$, for the initial condition that $\xi_0^0 = \xi_1^0 = -\upsilon$. We only do this for ξ_0^t since the steps are similar for ξ_1^t. Observe that

$$\begin{aligned}
f(-\upsilon) &= \log\left(\frac{\frac{\kappa(1-\alpha)\rho}{(1-\kappa)} + 1}{\frac{\kappa(1-\alpha)}{(1-\kappa)} + 1}\right)\\
&= \log\left(1 + (\rho-1)\frac{\kappa(1-\alpha)}{1-\kappa\alpha}\right)\\
&\overset{(a)}{=} (\rho-1)\frac{\kappa(1-\alpha)}{1-\kappa\alpha} - \frac{(\rho-1)^2}{2}\frac{\kappa^2(1-\alpha)^2}{(1-\kappa\alpha)^2}\\
&\quad + O(b^{-3/2}), \qquad (39)
\end{aligned}$$

where (a) follows from (38), and Taylor's expansion around $\rho = 1$. Similarly,

$$f^2(-\upsilon) = (\rho-1)^2\frac{\kappa^2(1-\alpha)^2}{(1-\kappa\alpha)^2} + O(b^{-3/2}), \qquad (40)$$

$$\log(\rho) = \log(1+(\rho-1)) = \sqrt{\frac{\lambda_\alpha(1-\kappa)}{(1-\alpha)^2\kappa^2 b}} - \frac{\lambda_\alpha(1-\kappa)}{2(1-\alpha)^2\kappa^2 b} + O(b^{-3/2}), \tag{41}$$

and

$$\log^2(\rho) = \frac{\lambda_\alpha(1-\kappa)}{(1-\alpha)^2\kappa^2 b} + O(b^{-3/2}). \tag{42}$$

Let us verify the induction result for $t=0$. Using the recursion (9) with $\xi_0^0 = \log\frac{\kappa(1-\alpha)}{1-\kappa} = -\upsilon$, we can express $\mathbb{E}\xi_0^1$ as

$$\mathbb{E}\xi_0^1 = -\kappa b(\rho-1) - \upsilon + \kappa b\alpha\log(\rho) + b(1-\kappa\alpha)f(-\upsilon).$$

Now using (39) and (41) we obtain

$$\begin{aligned}\mathbb{E}\xi_0^1 &= -\kappa\sqrt{\frac{\lambda_\alpha b(1-\kappa)}{(1-\alpha)^2\kappa^2}} - \upsilon + \kappa\alpha\sqrt{\frac{\lambda_\alpha(1-\kappa)b}{(1-\alpha)^2\kappa^2}} \qquad (43)\\ &\quad - \frac{\lambda_\alpha(1-\kappa)\alpha}{2(1-\alpha)^2\kappa} \\ &\quad + \sqrt{\frac{\lambda_\alpha(1-\kappa)b}{(1-\alpha)^2\kappa^2}}\kappa(1-\alpha) - \frac{\lambda_\alpha(1-\kappa)}{2(1-\kappa\alpha)} + O(b^{-1/2}) \\ &= -\upsilon - \frac{\lambda_\alpha(1-\kappa)}{2(1-\alpha)^2\kappa}\alpha - \frac{\lambda_\alpha(1-\kappa)}{2(1-\kappa\alpha)} + O(b^{-1/2}). \qquad (44)\end{aligned}$$

We also obtain, using the formula for the variance of a Poisson random variable

$$\begin{aligned}\mathrm{Var}\xi_0^1 &= \log^2(\rho)\kappa b\alpha + f^2(-\upsilon)(1-\kappa)b + f^2(-\upsilon)\kappa b(1-\alpha) \\ &\overset{(a)}{=} \frac{\lambda_\alpha\alpha(1-\kappa)}{(1-\alpha)^2\kappa} + \frac{(1-\kappa)\lambda_\alpha}{1-\kappa\alpha} + O(b^{-1/2}), \qquad (45)\end{aligned}$$

where in (a) we used (42) and (40). Comparing (44) and (45), after letting $b\to\infty$ with $\mu^{(1)}$ in (13) using $\mu^{(0)}=0$, we can verify the mean and variance recursions. Next we use Lemma 2 to prove gaussianity. Note that we can express $\xi_0^1 - h$ as the Poisson sum of iid mixture random variables as follows

$$\xi_0^1 - h = \sum_{i=1}^{L_0} X_i,$$

where $L_0 \sim \mathrm{Poi}(b)$, and $\mathcal{L}(X_i) = \kappa\alpha\mathcal{L}(\log(\rho)) + (1-\kappa)\mathcal{L}(f(-\upsilon)) + (\kappa(1-\alpha))\mathcal{L}(f(-\upsilon))$, keeping in mind the independent splitting property of Poissons, where $\mathcal{L}$ denotes the law of a rv[3]. Next we calculate $\mathbb{E}(|X_i|^3)$. It is easy to show using (39) and (41) that

$$\mathbb{E}(|X_i|^3) = \kappa\alpha\log^3(b) + (1-\kappa\alpha)|f^3(-\upsilon)| = O(b^{-3/2}). \tag{46}$$

Therefore the upper bound of Lemma 2 with $\lambda = b$ becomes

$$\frac{C_{BE}\mathbb{E}(|X_i|^3)}{\sqrt{\gamma(\mu^2+\sigma^2)^3}} = \frac{O(b^{-3/2})}{\sqrt{b\Omega(b^{-3})}} = O(b^{-1/2}).$$

By Lemma 2, taking $b\to\infty$ we obtain the convergence to Gaussian.

Having shown the induction hypothesis for $t=0$, we now assume it holds for some $t>0$. By using (11), (15) and Lebesgue's dominated convergence theorem [7, Theorem 16.4] we obtain

$$\begin{aligned}\mathbb{E}f(\xi_1^t) &= (\rho-1)\mathbb{E}\left(\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}}\right) \\ &\quad - \frac{(\rho-1)^2}{2}\mathbb{E}\left(\frac{e^{2\xi_1^t}}{(1+e^{\xi_1^t})^2}\right) + O(b^{-3/2}),\end{aligned} \tag{47}$$

and by using Lemma 4 in addition we obtain

$$\begin{aligned}\mathbb{E}f(\xi_0^t) &= (\rho-1)\frac{\kappa(1-\alpha)}{1-\kappa}\mathbb{E}\left(\frac{1}{1+e^{\xi_1^t}}\right) \\ &\quad - \frac{(\rho-1)^2\kappa(1-\alpha)}{2(1-\kappa)}\mathbb{E}\left(\frac{e^{\xi_1^t}}{(1+e^{\xi_1^t})^2}\right) + O(b^{-3/2}).\end{aligned} \tag{48}$$

Now we take the expectation of both sides of (9) and (10). Using the fact that $\mathbb{E}\sum_{i=1}^L X_i = \mathbb{E}X_i\mathbb{E}L$ if $L\sim$ Poi and X_i are independent and identically distributed (iid) rv, we obtain

$$\begin{aligned}\mathbb{E}(\xi_0^{t+1}) &= h + \log\left(\frac{p}{q}\right)\kappa b\alpha + \mathbb{E}\left(f(\xi_0^t)\right)(1-\kappa)b \\ &\quad + \mathbb{E}\left(f(\xi_1^t)\right)\kappa b(1-\alpha)\end{aligned} \tag{49}$$

and

$$\begin{aligned}\mathbb{E}(\xi_1^{t+1}) &= h + \log\left(\frac{p}{q}\right)\kappa a\alpha + \mathbb{E}\left(f(\xi_0^t)\right)(1-\kappa)b \\ &\quad + \mathbb{E}\left(f(\xi_1^t)\right)\kappa a(1-\alpha).\end{aligned} \tag{50}$$

We now substitute (48) and (47) in (49) to get:

$$\begin{aligned}&\mathbb{E}(\xi_0^{t+1}) \\ &= h + \kappa b\alpha\log(\rho) \\ &+ (1-\kappa)b\Bigg[(\rho-1)\frac{\kappa(1-\alpha)}{1-\kappa}\mathbb{E}\left(\frac{1}{1+e^{\xi_1^t}}\right) \\ &- \frac{(\rho-1)^2\kappa(1-\alpha)}{2(1-\kappa)}\mathbb{E}\left(\frac{e^{\xi_1^t}}{(1+e^{\xi_1^t})^2}\right) + O(b^{-3/2})\Bigg] \\ &+ \kappa b(1-\alpha)\Bigg[(\rho-1)\mathbb{E}\left(\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}}\right) \\ &- \frac{(\rho-1)^2}{2}\mathbb{E}\left(\frac{e^{2\xi_1^t}}{(1+e^{\xi_1^t})^2}\right) + O(b^{-3/2})\Bigg],\end{aligned}$$

which on simplifying and grouping like terms gives

$$\begin{aligned}\mathbb{E}(\xi_0^{t+1}) &= h + \kappa b\alpha\log(\rho) + \kappa(a-b)(1-\alpha) \\ &\quad - \frac{\lambda_\alpha(1-\kappa)}{2(1-\alpha)\kappa}\mathbb{E}\left(\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}}\right) + O(b^{-1/2}).\end{aligned}$$

Substituting $h = -\kappa(a-b) - \log\left(\frac{1-\kappa}{\kappa(1-\alpha)}\right)$, we get

$$\begin{aligned}\mathbb{E}(\xi_0^{t+1}) &= -\log\left(\frac{1-\kappa}{\kappa(1-\alpha)}\right) - \alpha\kappa(a-b) + \kappa b\alpha\log(\rho) \\ &\quad - \frac{\lambda_\alpha(1-\kappa)}{2\kappa(1-\alpha)}\mathbb{E}\left(\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}}\right) + O(b^{-1/2}).\end{aligned}$$

Using (41) we get

[3]Clearly X_i are iid with mean $\mu = \kappa\alpha\log(\rho) + (1-\kappa\alpha)f(-\nu) = \Omega(1/\sqrt{b})$ and $\sigma^2 = \Omega(1/b)$, both of which are bounded (fixed b and as $n\to\infty$). Also $\mu^2+\sigma^2 = \Omega(1/b)$.

$$
\begin{aligned}
&-\alpha\kappa(a-b)+\kappa b\alpha\log(\rho)\\
&= \kappa b\alpha(\log(\rho)-(\rho-1))\\
&= \kappa b\alpha\left(-\frac{\lambda_\alpha(1-\kappa)}{2\kappa^2 b(1-\alpha)^2}+O(b^{-3/2})\right)\\
&= -\frac{\lambda_\alpha\alpha(1-\kappa)}{2(1-\alpha)^2\kappa}+O(b^{-1/2}).
\end{aligned}
$$

Finally we obtain

$$
\begin{aligned}
\mathbb{E}(\xi_0^{t+1}) = &-\log\left(\frac{1-\kappa}{\kappa(1-\alpha)}\right)-\frac{\lambda_\alpha\alpha(1-\kappa)}{2(1-\alpha)^2\kappa}\\
&-\lambda_\alpha\frac{(1-\kappa)}{2(1-\alpha)\kappa}\mathbb{E}\left(\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}}\right)+O(b^{-1/2}).
\end{aligned}
\tag{51}
$$

Using exactly the same simplifications we can get

$$
\begin{aligned}
\mathbb{E}(\xi_1^{t+1}) = &-\log\left(\frac{1-\kappa}{\kappa(1-\alpha)}\right)+\frac{\alpha\lambda_\alpha(1-\kappa)}{2\kappa(1-\alpha)^2}\\
&+\frac{\lambda_\alpha(1-\kappa)}{2\kappa(1-\alpha)}\mathbb{E}\left(\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}}\right)+O(b^{-1/2}).
\end{aligned}
\tag{52}
$$

Our next goals are to compute $\text{var}(\xi_0^{t+1})$ and $\text{var}(\xi_1^{t+1})$. Towards this, observe that $f^2(x)=(\rho-1)^2\left(\frac{e^x}{1+e^x}\right)^2+O(b^{-3/2})$. Therefore

$$
\mathbb{E}(f^2(\xi_0^t))=(\rho-1)^2\mathbb{E}\left(\frac{e^{2\xi_0^t}}{(1+e^{\xi_0^t})^2}\right)+O(b^{-3/2}),
$$

and using Lemma 4 the above becomes

$$
\mathbb{E}(f^2(\xi_0^t))=(\rho-1)^2\frac{\kappa(1-\alpha)}{1-\kappa}\mathbb{E}\left(\frac{e^{\xi_1^t}}{(1+e^{\xi_1^t})^2}\right)+O(b^{-3/2}).
\tag{53}
$$

Similarly,

$$
\mathbb{E}\left(f^2(\xi_1^t)\right)=(\rho-1)^2\mathbb{E}\left(\frac{e^{2\xi_1^t}}{(1+e^{\xi_1^t})^2}\right)+O(b^{-3/2}).
\tag{54}
$$

Now we use the formula for the variance of Poisson sums $\text{Var}\sum_{i=1}^L X_i=\mathbb{E}(X_i^2)\mathbb{E}(L)$ to get

$$
\begin{aligned}
\text{Var}(\xi_0^{t+1}) = &\log^2(\rho)\kappa b\alpha+(1-\kappa)b\mathbb{E}(f^2(\xi_0^t))\\
&+\kappa b(1-\alpha)\mathbb{E}(f^2(\xi_1^t))
\end{aligned}
$$

$$
\begin{aligned}
\text{Var}(\xi_1^{t+1}) = &\log^2(\rho)\kappa a\alpha+(1-\kappa)b\mathbb{E}(f^2(\xi_0^t))\\
&+\kappa a(1-\alpha)\mathbb{E}(f^2(\xi_1^t)).
\end{aligned}
$$

Substituting (53) and (54) into the above equations and letting $b\to\infty$, we get

$$
\lim_{b\to\infty}\text{Var}(\xi_1^{t+1})=\lim_{b\to\infty}\text{Var}(\xi_0^{t+1})=\mu^{(t+1)},
$$

where

$$
\mu^{(t+1)}=\frac{\lambda_\alpha\alpha(1-\kappa)}{\kappa(1-\alpha)^2}+\frac{\lambda_\alpha(1-\kappa)}{\kappa(1-\alpha)}\mathbb{E}\left(\frac{\exp\xi_1^t}{1+\exp(\xi_1^t)}\right).
\tag{55}
$$

Using $\mu^{(t+1)}$ of (55) in (51) and (52) we get

$$
\begin{aligned}
\mathbb{E}(\xi_0^{t+1}) &= -\log\left(\frac{(1-\kappa)}{\kappa(1-\alpha)}\right)-\frac{1}{2}\mu^{(t+1)}+O(b^{-1/2})\\
\mathbb{E}(\xi_1^{t+1}) &= -\log\left(\frac{(1-\kappa)}{\kappa(1-\alpha)}\right)+\frac{1}{2}\mu^{(t+1)}+O(b^{-1/2}).
\end{aligned}
\tag{56}
$$

Now we use the fact the induction assumption that $\xi_1^t\to\mathcal{N}(\mathbb{E}(\xi_1^t),\mu^{(t)})$. Since the function $e^{\xi_1^t}/(1+e^{\xi_1^t})$ is bounded, by Lebesgue's dominated convergence theorem [7, Theorem 16.4] this means $\mathbb{E}(1/(1+e^{-\xi_1^t}))\to\mathbb{E}(1/(1+e^{-\mathcal{N}(\mathbb{E}(\xi_1^t),\mu^{(t)})}))$ as $b\to\infty$.We can write $\mathcal{N}(\mathbb{E}(\xi_1^t),\mu^{(t)})=\sqrt{\mu^{(t)}}Z+\mathbb{E}(\xi_1^t)$, where $Z\sim\mathcal{N}(0,1)$. Therefore we obtain

$$
\begin{aligned}
\mathbb{E}\left(\frac{1}{1+e^{-\xi_1^t}}\right) &= \mathbb{E}\left(\frac{1}{1+e^{-\sqrt{\mu^{(t)}}Z}\frac{(1-\kappa)}{\kappa(1-\alpha)}e^{-\frac{\mu^{(t)}}{2}}}\right)\\
&= \mathbb{E}\left(\frac{\kappa(1-\alpha)}{\kappa(1-\alpha)+(1-\kappa)e^{(-\sqrt{\mu^t}Z-\frac{\mu^{(t)}}{2})}}\right).
\end{aligned}
$$

Substituting the above into (55) gives us the recursion for $\mu^{(t+1)}$ given in (13).

Next we prove Gaussianity. Consider

$$
\begin{aligned}
&\xi_0^{t+1}-\mathbb{E}(\xi_0^{t+1})\\
&= \log\left(\frac{p}{q}\right)(L_{0c}-\mathbb{E}(L_{0c}))+\sum_{i=1}^{L_{00}}(f(\xi_{0,i}^t)-\mathbb{E}(f(\xi_0^t)))+\\
&\quad\sum_{i=1}^{L_{01}}(f(\xi_{1,i}^t)-\mathbb{E}(f(\xi_1^t)))+(L_{00}-\mathbb{E}(L_{00}))\mathbb{E}(f(\xi_0^t))+\\
&\quad(L_{01}-\mathbb{E}(L_{01}))\mathbb{E}(f(\xi_1^t)).
\end{aligned}
\tag{57}
$$

Let us look at the second term. Let $X_i=f(\xi_{0,i}^t)-\mathbb{E}f(\xi_{0,i}^t)$. Then it can be shown that $\mathbb{E}X_i^2=O(1/b)$. Let $D:=\sum_{i=1}^{L_{00}}X_i-\sum_{i=1}^{\mathbb{E}L_{00}}X_i$. In the second term the summation is taken up to $i\le\mathbb{E}L_{00}$. Then $\mathbb{E}(D^2)=|\sum_{i=1}^{\delta}X_i|^2$, where $\delta\le|L_{00}-\mathbb{E}L_{00}|+1$, where the extra 1 is because $\mathbb{E}L_{00}$ may not be an integer. Therefore $\mathbb{E}D^2=\mathbb{E}\delta\mathbb{E}|X_1|^2\le(C/b)((1-\kappa)b+1)^{1/2}=O(1/\sqrt{b})$. Thus, we can replace the Poisson upper limits of the summations in the second and third terms of (57) by their means, leading to

$$
\begin{aligned}
\xi_0^{t+1}-\mathbb{E}(\xi_0^{t+1}) = &\log\left(\frac{p}{q}\right)(L_{0c}-\mathbb{E}(L_{0c}))\\
&+\sum_{i=1}^{\mathbb{E}(L_{00})}(f(\xi_{0,i}^t)-\mathbb{E}(f(\xi_0^t)))\\
&+\sum_{i=1}^{\mathbb{E}(L_{01})}(f(\xi_{1,i}^t)-\mathbb{E}(f(\xi_1^t)))\\
&+(L_{00}-\mathbb{E}(L_{00}))\mathbb{E}f(\xi_0^t)\\
&+(L_{01}-\mathbb{E}(L_{01}))\mathbb{E}(f(\xi_1^t))+o_p(1),
\end{aligned}
\tag{58}
$$

where $o_p(1)$ indicates a rv that goes to zero in probability in the limit. The combined variance of all other terms approaches $\mu^{(t+1)}$, defined in (13), as $b\to\infty$ and it is finite for a fixed t. Now since we have an infinite sum of independent rvs

as $a, b \to \infty$, with zero mean and finite variance, from the standard CLT, we can conclude that the distribution tends to $\mathcal{N}(0, \mu^{t+1})$. The argument for ξ_1^{t+1} is identical. ∎

APPENDIX C
FINISHING THE PROOF OF THEOREM 1

Proof: We bound $\mathbb{E}(|\overline{S}\Delta\widehat{S}_0|)/(K(1-\alpha))$ as follows:

$$\lim_{b\to\infty}\lim_{n\to\infty}\frac{\mathbb{E}(|\overline{S}\Delta\widehat{S}_0|)}{K(1-\alpha)} = \lim_{b\to\infty}\lim_{n\to\infty}\left(\frac{\mathbb{E}\left(\sum_{i=1}^n \mathbf{1}_{\sigma_i\neq\widehat{\sigma}_i}\right)}{K-K\alpha}\right)$$
$$\leq \lim_{b\to\infty}\left(\frac{(1-\kappa)}{\kappa(1-\alpha)}\mathbb{P}(\xi_0^t\geq 0)+ \mathbb{P}(\xi_1^t\leq 0)\right), \quad (59)$$

since

$$\mathbb{E}\left(\sum_{i=1}^n \mathbf{1}_{\sigma_i\neq\widehat{\sigma}_i}\right) = n(\mathbb{P}(c_i=0,\sigma_i=0)\mathbb{P}(R_i^t>\upsilon|c_i=0,\sigma_i=0)+ \mathbb{P}(c_i=0,\sigma_i=1)\mathbb{P}(R_i^t<\upsilon|c_i=0,\sigma_i=1)), \quad (60)$$

and since $R_i^t - R_{i\to u}^t = O(b^{-1/2})$. Indeed, given the $b\to\infty$ limit in (59), the bound $O(b^{-1/2})$ allows us to replace R_i^t in (60) by the distribution limit when $n\to\infty$, which is ξ_0^t or ξ_1^t when conditioned on $\{\sigma_i=0\}$ or $\{\sigma_i=1\}$ respectively, for an arbitrary i. We now analyze each term in (59) separately. By Proposition 1 we have

$$\lim_{b\to\infty}\mathbb{P}(\xi_1^t\leq 0) = Q\left(\frac{1}{\sqrt{\mu^{(t)}}}\left(\frac{\mu^{(t)}}{2}-\log\frac{(1-\kappa)}{\kappa(1-\alpha)}\right)\right)$$

where $Q(\cdot)$ denotes the standard Q function. Notice that by (13) we have that $\mu^{(t)} \geq \lambda_\alpha\alpha(1-\kappa)/(\kappa(1-\alpha)^2)$, since $\mathbb{E}\left(\frac{1-\kappa}{\kappa(1-\alpha)+(1-\kappa)\exp(-\mu/2-\sqrt{\mu}Z)}\right)\geq 0$. In addition, by (55), $\mu^{(t)} \leq \frac{\lambda_\alpha(1-\kappa)}{\kappa(1-\alpha)^2}$. Note that the lower bound on $\mu^{(t)}$ is not useful when $\alpha=0$. Therefore by using the Chernoff bound for the Q function, $Q(x)\leq\frac{1}{2}e^{-x^2/2}$, we get

$$\lim_{b\to\infty}\mathbb{P}(\xi_1^t\leq 0) \leq \frac{1}{2}e^{-\frac{1}{2\mu^{(t)}}(\frac{\mu^{(t)}}{2}-\log\left(\frac{1-\kappa}{\kappa(1-\alpha)}\right))^2}$$
$$= \frac{1}{2}e^{-\frac{\mu^{(t)}}{8}(1-\frac{2}{\mu^{(t)}}\log\left(\frac{1-\kappa}{\kappa(1-\alpha)}\right))^2}$$
$$\leq \frac{1}{2}e^{-\frac{\mu^{(t)}}{8}}e^{\frac{1}{2}\log(\frac{1-\kappa}{\kappa(1-\alpha)})}$$
$$= \frac{1}{2}\sqrt{\frac{1-\kappa}{\kappa(1-\alpha)}}e^{-\frac{\mu^{(t)}}{8}}, \quad (61)$$

where we used the fact that $(1-x)^2\geq 1-2x$ for any $x>0$. By employing similar reductions, we can show

$$\lim_{b\to\infty}\left(\frac{(1-\kappa)}{\kappa(1-\alpha)}\right)\mathbb{P}(\xi_0^t\geq 0)\leq\frac{1}{2}\sqrt{\frac{1-\kappa}{\kappa(1-\alpha)}}e^{-\frac{\mu^{(t)}}{8}}. \quad (62)$$

Substituting (61) and (62) back in (59) and using the fact that $\mu^{(t)}\geq\lambda_\alpha\alpha(1-\kappa)/(\kappa(1-\alpha)^2)$, we get

$$\lim_{b\to\infty}\lim_{n\to\infty}\frac{\mathbb{E}(|\overline{S}\Delta\widehat{S}_0|)}{K(1-\alpha)}\leq\sqrt{\frac{1-\kappa}{\kappa(1-\alpha)}}e^{-\frac{\lambda_\alpha\alpha(1-\kappa)}{8\kappa(1-\alpha)^2}}.$$

Then using (23) we get the desired result in (21) . ∎

REFERENCES

[1] 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 preprint arXiv:1512.09080*, 2015.

[2] A. E. Allahverdyan, G. Ver Steeg, and A. Galstyan, "Community detection with and without prior information," *Europhysics Letters*, vol. 90, no. 1, p. 18002, 2010.

[3] N. Alon, M. Krivelevich, and B. Sudakov, "Finding a large hidden clique in a random graph," *Random Structures and Algorithms*, vol. 13, no. 3-4, pp. 457–466, 1998.

[4] R. Andersen and F. Chung, "Detecting sharp drops in PageRank and a simplified local partitioning algorithm," *Theory Appl. Model. Comput.*, vol. 4484/2007, no. 3, pp. 1–12, 2007.

[5] K. Avrachenkov, P. Gonçalves, A. Mishenin, and M. Sokol, "Generalized optimization framework for graph-based semi-supervised learning," in *Proceedings of the Twelfth SIAM International Conference on Data Mining, Anaheim, California, USA, April 26-28, 2012.*, 2012, pp. 966–974.

[6] A. Beutel, W. Xu, V. Guruswami, C. Palow, and C. Faloutsos, "Copycatch: stopping group attacks by spotting lockstep behavior in social networks," in *Proceedings of the 22nd WWW*. ACM, 2013, pp. 119–130.

[7] P. Billingsley, *Probability and measure*. John Wiley & Sons, 2008.

[8] T. T. Cai, T. Liang, and A. Rakhlin, "Inference via message passing on partially labeled stochastic block models," *arXiv preprint arXiv:1603.06923*, 2016.

[9] F. Caltagirone, M. Lelarge, and L. Miolane, "Recovering asymmetric communities in the stochastic block model," in *Allerton 2016 54th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, United States, Sep. 2016.

[10] D. H. Chau, S. Pandit, and C. Faloutsos, "Detecting fraudulent personalities in networks of online auctioneers," in *PKDD*. Springer, 2006, pp. 103–114.

[11] J. Chen and Y. Saad, "Dense subgraph extraction with application to community detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 7, pp. 1216–1230, 2012.

[12] P.-Y. Chen and A. O. Hero, "Phase transitions in spectral community detection," *IEEE transactions on signal processing*, vol. 63, no. 16, pp. 4339–4347, 2015.

[13] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovačević, "Signal recovery on graphs: Variation minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4609–4624, 2015.

[14] Y. Deshpande and A. Montanari, "Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time," *Foundations of Computational Mathematics*, vol. 15, no. 4, pp. 1069–1128, 2015.

[15] H. Firouzi, B. Rajaratnam, and A. O. Hero III, "Predictive correlation screening: Application to two-stage predictor design in high dimension." in *AISTATS*, 2013, pp. 274–288.

[16] D. Gleich and K. Kloster, "Seeded pagerank solution paths," *European Journal of Applied Mathematics*, pp. 1–34, 2016.

[17] B. Hajek, Y. Wu, and J. Xu, "Recovering a Hidden Community Beyond the Spectral Limit in $O(|E|\log^*|V|)$ Time," *arXiv Prepr. arXiv1510.02786*, 2015.

[18] ——, "Information limits for recovering a hidden community," in *Information Theory (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 1894–1898.

[19] A. Kadavankandy, K. Avrachenkov, L. Cottatellucci, and R. Sundaresan, "Belief propagation for subgraph detection with imperfect side-information," in *Submitted to ISIT 2017*. IEEE, Awaiting review.

[20] A. Kadavankandy, L. Cottatellucci, and K. Avrachenkov, "Characterization of L^1-norm statistic for Anomaly Detection in Erdös Rényi Graphs," in *CDC*. IEEE, 2016.

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

[22] U. Kang, D. H. Chau, and C. Faloutsos, "Mining large graphs: Algorithms, inference, and discoveries," in *2011 IEEE 27th International Conference on Data Engineering*. IEEE, 2011, pp. 243–254.

[23] D. Koutra, T.-Y. Ke, U. Kang, D. H. P. Chau, H.-K. K. Pao, and C. Faloutsos, "Unifying guilt-by-association approaches: Theorems and fast algorithms," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2011, pp. 245–260.

[24] V. E. Lee, N. Ruan, R. Jin, and C. Aggarwal, "A survey of algorithms for dense subgraph discovery," in *Managing and Mining Graph Data*. Springer, 2010, pp. 303–336.

[25] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.

[26] T. Mifflin, C. Boner, G. Godfrey, and J. Skokan, "A random graph model for terrorist transactions," in *2004 IEEE Aerosp. Conf. Proc.*, vol. 5. IEEE, 2004, pp. 3258–3264.

[27] B. Miller, N. Bliss, and P. J. Wolfe, "Subgraph detection using eigenvector l1 norms," in *Advances in Neural Information Processing Systems*, 2010, pp. 1633–1641.

[28] B. A. Miller, S. Kelley, R. S. Caceres, and S. T. Smith, "Residuals-based subgraph detection with cue vertices," in *2015 49th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2015, pp. 1530–1534.

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

[30] E. Mossel and J. Xu, "Local Algorithms for Block Models with Side Information," in *ITCS '16*. New York, New York, USA: ACM Press, jan 2016, pp. 71–80.

[31] T. Qin and K. Rohe, "Regularized spectral clustering under the degree-corrected stochastic blockmodel," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2013. [Online]. Available: http://papers.nips.cc/paper/5099-regularized-spectral-clustering-under-the-degree-corrected-stochastic-blockmodel

[32] K. Rohe, S. Chatterjee, and B. Yu, "Spectral clustering and the high-dimensional stochastic blockmodel," *Ann. Stat.*, pp. 1878–1915, 2011.

[33] S. T. Smith, E. K. Kao, K. D. Senne, G. Bernstein, and S. Philips, "Bayesian discovery of threat networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5324–5338, 2014.

[34] X. Wang, P. Liu, and Y. Gu, "Local-set-based graph signal reconstruction," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2432–2444, 2015.

[35] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowledge and Information Systems*, vol. 42, no. 1, pp. 181–213, 2015.

[36] P. Zhang, C. Moore, and L. Zdeborová, "Phase transitions in semisupervised clustering of sparse networks," *Physical Review E*, vol. 90, no. 5, p. 052802, 2014.

[37] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," *Advances in neural information processing systems*, vol. 16, no. 16, pp. 321–328, 2004.

[38] X. Zhu, Z. Ghahramani, J. Lafferty *et al.*, "Semi-supervised learning using gaussian fields and harmonic functions," in *ICML*, vol. 3, 2003, pp. 912–919.

Arun Kadavankandy is currently a postdoctoral researcher at CentraleSupélec, Paris, working with Romain Couillet at LANEAS research group. He obtained his Ph.D. at Inria Sophia Antipolis in July 2017, on the topic of 'Spectral analysis of random graphs with application to clustering and sampling' under the supervision of Konstantin Avrachenkov, Inria Sophia Antipolis and Laura Cottatellucci, Eurecom. He obtained a master's degee in Signal Processing from the Indian Institute of Science Bangalore in 2011. His research interests include graph-based machine learning, random graph analysis, random matrix theory and distributed algorithms.

Konstantin Avrachenkov received Master degree in Control Theory from St. Petersburg State Polytechnic University (1996), Ph.D. degree in Mathematics from University of South Australia (2000) and Habilitation (Doctor of Science) from University of Nice Sophia Antipolis (2010). Currently, he is a Director of Research at Inria Sophia Antipolis, France. He is an associate editor of International Journal of Performance Evaluation and ACM TOMPECS. His main research interests are Markov processes, singular perturbation theory, optimization, game theory and analysis of complex networks.

Laura Cottatellucci (S'01-M'07) obtained the Master degree from La Sapienza University, the Ph.D. from Technical University of Vienna (2006) and the Habilitation from University of Nice-Sophia Antipolis. She worked in Telecom Italia (1995–2000) as responsible of industrial projects and as senior research in ftw Austria (Apr. 2000-Sept. 2005). She was research fellow in INRIA Sophia Antipolis (Oct.-Dec. 2005) and at University of South Australia (2006). In the period Dec. 2006-Nov. 2017, she was assistant professor in EURECOM. Currently, she is professor at Friedrich-Alexander Universität. Cottatellucci is associate editor for IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON SIGNAL PROCESSING. Her research interests lie in the areas of large system analysis and algorithm design for wireless communications and complex networks, random matrix theory, and game theory.

Rajesh Sundaresan (S'96-M'00-SM'06) received the B.Tech. degree in electronics and communication from the Indian Institute of Technology Madras, the M.A. and Ph.D. degrees in electrical engineering from Princeton University in 1996 and 1999, respectively. From 1999 to 2005, he worked at Qualcomm Inc. on the design of communication algorithms for wireless modems. Since 2005, he has been with the Indian Institute of Science where he is currently a Professor in the Department of Electrical Communication Engineering and an associate faculty in the Robert Bosch Centre for Cyber-Physical Systems. His interests are in the areas of communication, computation, and control over networks. He was an associate editor of the IEEE TRANSACTIONS ON INFORMATION THEORY for the period 2012-2015.