Subgraph Detection with cues using Belief Propagation

A.Kadavankandy, K. Avrachenkov,
L. Cottatellucci, and Rajesh Sundaresan.

RESEARCH REPORT

N° XXX

Oct 2016

Project-Team Maestro

ISSN 0249-6399   ISRN INRIA/RR--XXX--FR+ENG

Subgraph Detection with cues using Belief Propagation

A.Kadavankandy*, K. Avrachenkov,
L. Cottatellucci, and Rajesh Sundaresan.

Project-Team Maestro

Research Report n° XXX — Oct 2016 — 17 pages

Abstract: We consider an Erdős-Rényi graph with n nodes and edge probability q that is embedded with a random subgraph of size K with edge probabilities p such that $p > q$. We address the problem of detecting the subgraph nodes when only the graph edges are observed, along with some extra knowledge of a small fraction of subgraph nodes, called cued vertices or cues. We employ a local and distributed algorithm called belief propagation (BP). Recent works on subgraph detection without cues have shown that global maximum likelihood (ML) detection strictly outperforms BP in terms of asymptotic error rate, namely, there is a threshold condition that the subgraph parameters should satisfy below which BP fails in achieving asymptotically zero error, but ML succeeds. In contrast, we show that when the fraction of cues is strictly bounded away from zero, i.e., when there exists non-trivial side-information, BP achieves zero asymptotic error even below this threshold, thus approaching the performance of ML detection.

Key-words: Belief Propagation, Subgraph Detection, Semisupervised Learning, Random Graphs

* Corresponding author, arun.kadavankandy@inria.fr

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

La Detection de Sousgraphes en presence des indices grâce au Belief Propagation

Résumé : Nous considérons un graphe Erdős-Rényi qui a n sommets dont q est la probabilité d'arrêtes. La dessus il y un sousgraphe placé sur leurs m sommets selectionnés aléatoirement et leur probabilité d'arrêtes est p, en sorte que $p > q$. Nous proposons un algorithme distribué aux calculs locales à chaque sommet, tiré du "Belief Propagation" (BP), qui détecte les sommets du sousgraphe, quand on connait une fraction de sommets du sousgraphe en tant qu'indices. Des recherches récentes ont prouvé que la prestation du BP dans l'absence des indices est strictement inférior par rapport à la detection globale du maximum de vraisemblance (DMV). A l'opposé, ici on prouve qu'en presence des indices, la prestation du BP est à l'hauteur de celle de DMV, dans la sens où le premier reussie à detecter la sousgraphe avec une erreur qui tend a zéro, à chaque fois le dernier peut le faire, dans la limite où le nombre de sommets du graph tend l'infinité.

Mots-clés : Belief Propagation, Detection de Sousgraphes, Semisupervised Learning, Graphes Aléatoires

1 Introduction

Detecting a small community of highly connected nodes in a sparse network is an important problem in data mining, machine learning, and theoretical computer science. This problem is linked to threat detection, anomaly detection, fault detection etc. in a network. Please see [1] for a survey. The hidden subgraph model with both the subgraph and the background modelled as ER graphs with different edge densities was proposed in [10] to study anomalous transactions in a computer network.

In this model the background graph is ER with n nodes and edge probability q. A random subset of $K < n/2$ vertices has the edge probabilities within it changed to $p > q$, without affecting any other edge. This graph, denoted $G(K, n, p, q)$, can model a network with a hidden community [10]. See [8,9] for works on detecting the presence of such a subgraph in a given graph and detecting the subgraph nodes. A stream of recent works in this area suggests that there is a subgraph size below which detection is impossible by means of nearly linear-time algorithms, and above which efficient algorithms have been identified ([2,7,9,11]). When $q = 1/2$ and $p = 1$, the subgraph detection problem reduces to the well-known clique detection problem ([2], [9], [5]).

We use the framework of analysis developed in [7] and [11]. In [11] the author considers the problem of detecting the hidden community in $G(K, n, p, q)$ under the assumption that $p = a/n, q = b/n$ and $K = \kappa/n$, where a, b, κ are constants. A parameter λ (defined later) is introduced to characterise the "strength" of the subgraph. They develop a local BP algorithm and show it achieves zero asymptotic error when $\lambda > 1/\exp(1)$, whereas if $\lambda \leq 1/\exp(1)$, BP does not do better than random guessing. In contrast, the ML detector achieves zero error asymptotically for any $\lambda > 0$. In [7], the authors consider a more general setting of sparse graphs and prove that BP succeeds when $\lambda > 1/e$. In [6] global ML detection is shown to achieve zero error rate asymptotically if $\lambda = \Theta((K/n)\log(K/n))$.

The optimality of BP with side information is shown in [12] and [4] for community detection on SBM with symmetric communities, but to the best of our knowledge no theoretical studies of BP have been made when cued vertices are available for detecting a small subgraph in $G(K, n, p, q)$.

Our contributions: In our work we consider the subgraph detection problem where some side information in the form of cued nodes is available. This fits within the framework of semisupervised learning. We develop a BP algorithm that detects the nodes of the subgraph in the presence of cues and prove that when the graph is dilute, with $p, q = \Theta(1/n)$, the fraction of miss-classified nodes approaches zero for any $\lambda > 0$ when there is a strictly positive fraction of cues. In other words we show that BP with cues succeeds in the entire regime where ML succeeds [11, Proposition 4.1].

The paper is organised as follows: In section 2 we describe our graph model and the problem. In section 3, we present our algorithm and its derivation. In section 4 we derive the asymptotic distribution of BP messages. In section 5 we prove our main result on the asymptotic error rate of our algorithm. In section 6 we present some simulation results to back up the theory.

2 Model and Problem Definition

Let $\mathcal{G} = (V, E)$ be a realisation of $G(K, n, p, q)$. Let S be the set of subgraph nodes and C be the set of cued nodes. The latter is chosen from S by independent Bernoulli sampling with probability (w.p.) $\alpha < 1$. 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 dilute graphs. The results in this paper presuppose that $\kappa = K/n$ is a constant independent of n. Our aim is to propose a candidate set $\widehat{S}$ given $\mathcal{G}$ and C, assuming p, q, K and α are known, using local and recursive updates provided by BP. Note that this problem is identical to detecting the hidden labels σ_i of the graph nodes assigned such that $\sigma_i = 1$ if $i \in S$ and $\sigma_i = 0$ otherwise.

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, which are the set of nodes at a distance of t from 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 when $B \subset A$, and Δ denotes the set difference, i.e., $A \Delta B = (A \cup B) \backslash (A \cap B)$. The symbol $\sim$ denotes the distribution of a random variable (rv), for example $X \sim \text{Poi}(\lambda)$ means that X is a Poisson distributed rv with rate λ. 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 [3].

3 Belief Propagation Algorithm for Detection in the Presence of Cues

In this section we describe the local and distributed BP algorithm (1), which performs detection in the presence of side-information available in the form of cued nodes. The algorithm has two stages: message passing (2), and belief updation (1). At step t of Algorithm 1, each node $u \in V \backslash C$ updates its own log-likelihood ratio:

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

where G_u^t denotes the subgraph induced by the t-neighbourhood of u and C_u^t is the set of cues in G_u^t. This computation is local, because it uses only messages transmitted to u by its neighbours $i \in \delta u$, given by

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

where G_i^t and C_i^t are defined as done for u. It can be checked that the total computation time for t_f steps of BP is $O(t_f |E|)$.

Recall that the optimum detector that minimises the expected number of misclassified nodes is the ML detector [6] given as:

$$\hat{\sigma}_i = \mathbf{1}_{\{R_i > \log(n - K / K(1-\alpha))\}},$$

where

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

The output set size here may not be exactly equal to K, but this can be mitigated by some post-processing (for example, if the size is larger than K, we can simply pick the top K values of the likelihood function). This detector however requires the observation of the whole graph, and cannot be implemented in a distributed fashion. In addition, it is not computationally feasible, since it requires marginalising over 2^n pair-wise dependent random variables over a large graph. We would like algorithms that, for a decision at a node u of the graph, rely only on the observation of the t-neighbourhood G_u^t of u, BP being one of them.

A small neighbourhood of a large sparse graph can be approximated by a tree. This is formalised in Lemma 1. In the following we present a Poisson random tree, which can be coupled with G_u^t.

Algorithm 1 BP with 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 $(i,u)\in E$, such that $u\notin C$:

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

3: If $t < t_f - 1$ go back to 2, else go to 4

4: Compute $R^{t_f}_u$ for every $u\in V\backslash C$ as follows:

$$R^{t+1}_{u} = -K(p-q) + \sum_{l\in C_u} \log\frac{p}{q} + \sum_{l\in \delta u\backslash C_u} \log\frac{\exp(R^t_{l\to u}-\nu)(p/q)(1-\alpha)+1}{\exp(R^t_{l\to u}-\nu)(1-\alpha)+1} \tag{2}$$

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

Let T^t_u be a labelled Galton-Watson (G-W) tree of depth t rooted at node u constructed as follows (as in [7]): The label τ_u at node u is chosen at random in the following way:

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

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 rvs, this is equivalent to assigning the label $\tau_i = 1$ to each child i by sampling a Bernoulli rv 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 $\tau_u = 0, 1$ and 1 respectively. Namely, if i is a child of u,

$$\mathbb{P}(\tau_i = 1|\tau_u = 1) = \frac{Kp}{d_1}, \qquad\qquad \mathbb{P}(\tau_i = 1|\tau_u = 0) = \frac{Kq}{d_0}. \tag{3}$$

We then assign the cue indicator function c such that $c_i = 1$ w.p. α if $\tau_i = 1$ and $c_i = 0$ if $\tau_i = 0$. The process is repeated up to depth t giving us C^t_u, the set of cued neighbours.

Consider the problem of estimating the label τ_u of node $u\notin C$ based on an observation of T^t_u and C^t_u. The optimal ML detector is given as

$$\hat{\tau}_u = \mathbf{1}_{\{\Lambda^t_u > \log(\frac{(n-K)}{K(1-\alpha)})\}},$$

where $\Lambda^t_u = \log(\mathbb{P}(T^t_u, C^t_u|\tau_u = 1)/\mathbb{P}(T^t_u, C^t_u|\tau_u = 0))$. By the following coupling lemma established in [7], the detection of label σ_u based on G^t_u is statistically identical to the detection of τ_u based on T^t_u :

LEMMA 1 *[7] For t such that $(np)^t = n^{o(1)}$, there exists a coupling such that $(G_u^t, \sigma^t) = (T_u^t, \tau^t)$ with probability $1 - n^{-1+o(1)}$.*

In our case since $p = a/n$, any $t = o(\log(n))$ satisfies the condition of the above lemma.

Consequently, the likelihood ratios in a small neighbourhood G_u^t of u are statistically identical to the likelihoods derived on the corresponding G-W tree, which are the BP messages. Hence we proceed by deriving BP recursions in Algorithm 1 for node u assuming G_u^t is a tree. Consider a node $u \in V \backslash C$. We can express the likelihood ratio at u based on an observation of T_u^{t+1}, C_u^{t+1} as

$$\begin{aligned}\Lambda_u^{t+1} &= \log \frac{\mathbb{P}(T_u^{t+1}, C_u^{t+1} | \tau_u = 1)}{\mathbb{P}(T_u^{t+1}, C_u^{t+1} | \tau_u = 0)} \\ &= \log \frac{\mathbb{P}\{N_u | \tau_u = 1\}}{\mathbb{P}\{N_u | \tau_u = 0\}} + \sum_{i \in \delta u} \log \frac{\mathbb{P}(T_i^t, c_i, C_i^t | \tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t | \tau_u = 0)}, \qquad (4)\end{aligned}$$

by independence of the children of u given τ_u. Moreover,

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

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

$$\begin{aligned}\log \frac{\mathbb{P}\{N_u | \tau_u = 1\}}{\mathbb{P}\{N_u | \tau_u = 0\}} &= N_u \log \frac{d_1}{d_0} - (d_1 - d_0) \\ &= N_u \log \frac{d_1}{d_0} - K(p - q). \qquad (5)\end{aligned}$$

Next we look at the second term in (4). We analyse separately the cued neighbours of u and the non-cue neighbours.

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

$$\begin{aligned}&\log \frac{\mathbb{P}(T_i^t, c_i, C_i^t | \tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t | \tau_u = 0)} \qquad (6) \\ &= \log \left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t, \tau_i = 1 | \tau_u = 1) + \overset{0}{\mathbb{P}(T_i^t, c_i, C_i^t, \tau_i = 0 | \tau_u = 1)}}{\mathbb{P}(T_i^t, c_i, C_i^t, \tau_i = 1 | \tau_u = 0) + \overset{0}{\mathbb{P}(T_i^t, c_i, C_i^t, \tau_i = 0 | \tau_u = 0)}} \right) \\ &\overset{(a)}{=} \log \left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t, \tau_i = 1 | \tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t, \tau_i = 1 | \tau_u = 0)} \right) \\ &= \log \left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t | \tau_i = 1) \mathbb{P}(\tau_i = 1 | \tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t | \tau_i = 1) \mathbb{P}(\tau_i = 1 | \tau_u = 0)} \right) \\ &\overset{(b)}{=} \log \frac{Kp/d_1}{Kq/d_0}, \qquad (7)\end{aligned}$$

where in step (a) we applied the fact that $\mathbb{P}(c_i = 1, \tau_i = 0) = 0$ and in (b) we used (3).

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

$$\begin{aligned}&\mathbb{P}(T_i^t, c_i, C_i^t | \tau_u = 1) \qquad (8) \\ &= \mathbb{P}(T_i^t, C_i^t | \tau_i = 1) \mathbb{P}(c_i | \tau_i = 1) \mathbb{P}(\tau_i = 1 | \tau_u = 1) \\ &\quad + \mathbb{P}(T_i^t, C_i^t | \tau_i = 0) \mathbb{P}(c_i | \tau_i = 0) \mathbb{P}(\tau_i = 0 | \tau_u = 1) \\ &= \mathbb{P}(T_i^t, C_i^t | \tau_i = 1)(1 - \alpha) \frac{Kp}{d_1} + \mathbb{P}(T_i^t, C_i^t | \tau_i = 0) \frac{(n - K)q}{d_1}.\end{aligned}$$

Similarly, we can show

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

Let us define

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

the message that i sends to u at step t. We plug this into (8). Finally combining (5), (7) and (8) and replacing Λ_u^t with R_u^t and $\Lambda_{i\to u}^t$ with $R_{i\to u}^t$, we arrive at (2). The recursive equation (1) can be derived in exactly the same way by looking at the children of $i \in \delta u$. □

4 Asymptotic Error Analysis

We analyse the distributions of BP messages Λ_i^t given $\tau_i = 1$ and $\tau_i = 0$. This will help us to bound the error rate on a tree. This equals the error rate on $\mathcal{G}$ asymptotically since by the coupling Lemma 1 the two are the same with a probability that tends to 1. Notice that since we only focus on non-cued vertices the prior distribution after the observation of cues changes. Therefore $\mathbb{P}\{\tau_i = 1|c_i = 0\} = K(1-\alpha)/(n-K\alpha)$ and $\mathbb{P}\{\tau_i = 0|c_i = 0\} = (n-K)/(n-K\alpha)$ are the prior probabilities of the uncued vertices. For convenience we put a line over the symbols for expectation and probability to denote conditioning w.r.t $\{c_i = 0\}$ when considering the posterior distributions (eg: $\overline{\mathbb{E}}, \overline{\mathbb{P}}$). Define $\upsilon = \log\left(\frac{n-K}{K(1-\alpha)}\right)$.

Instead of studying the distribution of Λ_i^t, $i \in V\backslash C$, we look at the log of the ratio of the a-posteriori probabilities of τ_i given as

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

This is just a matter of choice since by Bayes rule it holds that $\widetilde{\Lambda}_i^t = \Lambda_i^t - \upsilon$. Let ξ_0^{t+1}, ξ_1^{t+1} be the random variables with the same distribution as the messages $\widetilde{\Lambda}_i^{t+1}$ given $\tau_i = 0$ and $\tau_i = 1$ respectively, conditioned on $\{c_i = 0\}$, in the limit as $n \to \infty$. In view of the coupling formulation, it is then straightforward to show that they satisfy the following two recursive distributional evolutionary equations with initial conditions $\xi_0^0 = \xi_1^0 = \log \kappa(1-\alpha)/(1-\kappa)$:

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

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

where, $\stackrel{\mathrm{D}}{=}$ means that the L.H.S has the same distribution as the R.H.S. and $h = -K(p-q) - \log(\frac{n-K}{K(1-\alpha)}) = -\kappa(a-b) - \log(\frac{1-\kappa}{\kappa(1-\alpha)})$ and the function f is defined as

$$f(\cdot) \equiv \log\left(\frac{\exp(\cdot)(p/q)+1}{\exp(\cdot)+1}\right).$$

The rvs $\xi_{0,i}^t$ are independent and identically distributed (iid) and identically distributed to ξ_0^t, and $\xi_{1,i}^t$ are iid with the same distribution as ξ_1^t. Furthermore, $L_{00} \sim \text{Poi}((n-K)q)$ is the rv that equals the number of children of u with label 0 if $\tau_u = 0$, and $L_{01} \sim \text{Poi}(Kq(1-\alpha))$, the number of children with label 1 when $\tau_u = 0$. Similarly $L_{10} \sim \text{Poi}((n-K)q)$ and $L_{11} \sim \text{Poi}(Kp(1-\alpha))$ denote the number of children of u with label 0 and 1 respectively when $\tau_u = 1$. Lastly, L_{0c} and L_{1c} are the number of cued children of u when $\tau_u = 0$ and $\tau_u = 1$ respectively with $L_{0c} \sim \text{Poi}(Kq\alpha)$ and $L_{1c} \sim Poi(Kp\alpha)$. We define the parameter λ, interpreted as an effective SNR [11] of the detection problem, as

$$\lambda = \frac{K^2(p-q)^2}{(n-K)q} \tag{11}$$

$$= \frac{\kappa^2(a-b)^2}{(1-\kappa)b} = \frac{\kappa^2 b(a/b-1)^2}{(1-\kappa)}. \tag{12}$$

If P_0 and P_1 are the probability measures of ξ_0^t and ξ_1^t respectively, then they are related as follows.

Lemma 2

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

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 [11], 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 [3]. Let $Y = (T_u^t, C_u^t)$, the observed rv. Therefore

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

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 $\{\tau_u = 1\}$. $\square$

Note that the distributional equations (9) and (10) give the asymptotic distributions of the messages on the graph $\mathcal{G}$ as $n \to \infty$. These equations do not depend on n because of the choice of p, q and K. For ease of analysis we will presently study the distributions in the limit where $a, b \to \infty$. This limit is taken after $n \to \infty$. Ideally, one would like to analyse the distributions for finite a and b, but this is left for future work. We have the following result on the Gaussianity of the asymptotic messages in the limit where $a, b \to \infty$, after $n \to \infty$.

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

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

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

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

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

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

Remark: When $\alpha = 0$ (13) reduces to the recursion given in [11] as expected.
Proof: Since λ is fixed and $b \to \infty$, we have

$$\rho \equiv a/b = 1 + \sqrt{\frac{\lambda(1-\kappa)}{\kappa^2 b}} = 1 + O(b^{-1/2}), \tag{14}$$

by (12) since λ and κ are fixed. In the proof we use Berry-Essen inequality for Poisson sums [7, Lemma 11]

Lemma 3 *Let $S_\lambda = X_1 + X_2 + \ldots X_{N_\lambda}$, where $X_i : i \geq 1$ are independent, identically distributed random variables with mean μ, variance σ^2 and $\mathbb{E}[|X_i^3|] \leq g^3$, and for some $\lambda > 0$, N_λ is a Poi(λ) random variable independent of $(X_i : i \geq 1)$. Then*

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

where $C_{BE} = 0.3041$.

Following [11], 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 for ξ_1^t the steps are similar. Observe that

$$\begin{aligned} & f(-\upsilon) \\ &= \log\left(\frac{\frac{pK(1-\alpha)}{q(n-K)}+1}{\frac{K(1-\alpha)}{(n-K)}+1}\right) \\ &= \log\left(1 + (\rho-1)\frac{\kappa(1-\alpha)}{1-\kappa\alpha}\right) \\ &\overset{(a)}{=} (\rho-1)\frac{\kappa(1-\alpha)}{1-\kappa\alpha} - \frac{(\rho-1)^2}{2}\frac{\kappa^2(1-\alpha)^2}{(1-\kappa\alpha)^2} + \\ & \quad O(b^{-3/2}), \end{aligned} \tag{15}$$

where (a) follows from (14), and Taylor's expansion around $\rho = 1$. Similarly,

$$f^2(-\upsilon) = (\rho - 1)^2 \frac{\kappa^2(1-\alpha)^2}{(1-\kappa\alpha)^2} + O(b^{-3/2}), \tag{16}$$

$$\log(\rho) = \log(1 + (\rho - 1)) = \sqrt{\frac{\lambda(1-\kappa)}{\kappa^2 b}} - \frac{\lambda(1-\kappa)}{2\kappa^2 b} + O(b^{-3/2}), \tag{17}$$

and

$$\log^2(\rho) = \frac{\lambda(1-\kappa)}{\kappa^2 b} + O(b^{-3/2}) \tag{18}$$

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

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

Now using (15) and (17) we obtain

$$\begin{aligned}\mathbb{E}\xi_0^1 &= -\kappa\sqrt{\frac{\lambda b(1-\kappa)}{\kappa^2}} - \upsilon + \kappa\alpha\sqrt{\frac{\lambda(1-\kappa)b}{\kappa^2}} - \frac{\lambda(1-\kappa)\alpha}{2\kappa} \\ &+ \sqrt{\frac{\lambda(1-\kappa)b}{\kappa^2}}\kappa(1-\alpha) - \frac{(1-\alpha)^2}{2(1-\kappa\alpha)}\lambda(1-\kappa) \qquad (19) \\ &+ O(b^{-1/2}) \\ &= -\upsilon - \frac{\lambda(1-\kappa)}{2\kappa}\alpha - \frac{(1-\alpha)^2}{2(1-\kappa\alpha)}\lambda(1-\kappa) + O(b^{-1/2}), \qquad (20)\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}\xi_0^1 &= \log^2(\rho)\kappa b\alpha + f^2(-\upsilon)(1-\kappa)b + f^2(-\upsilon)\kappa b(1-\alpha) \\ &\overset{(a)}{=} \frac{\lambda\alpha(1-\kappa)}{\kappa} + \frac{(1-\alpha)^2(1-\kappa)\lambda}{1-\kappa\alpha},\end{aligned} \tag{21}$$

where in (a) we used (18) and (16). Comparing (20) and (21) with $\mu^{(1)}$ in (13) using $\mu^{(0)} = 0$, we can verify the mean and variance recursions. Next we use Lemma3 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 $L_0 \sim \mathrm{Poi}(nq) = \mathrm{Poi}(b)$, and $\mathcal{L}(X_i) = \kappa\alpha\mathcal{L}(p/q) + (1-\kappa)b\mathcal{L}(f(-\upsilon)) + (\kappa b(1-\alpha))\mathcal{L}(f(-\upsilon))$, keeping in mind the independent splitting property of Poissons, where $\mathcal{L}$ denotes the law of a random variable. Then by comparing with the form in Lemma 3, $\lambda = b$, and the term $\lambda(\mu^2+\sigma^2) = \mathrm{Var}\xi_0^1 = b(\mu^2+\sigma^2)$, which is finite. Next we calculate $\mathbb{E}|X_i|^3$. It is easy to see that

$$\mathbb{E}|X_i|^3 = \kappa\alpha\log^3(b) + ((1-\kappa) + \kappa(1-\alpha))f^3(-\upsilon) \tag{22}$$

$$= O(b^{-3/2}). \tag{23}$$

Hence $b\mathbb{E}|X_i|^3 = O(b^{-1/2})$. Therefore the RHS of Lemma (3) becomes

$$\begin{aligned}\frac{C_{BE}\mathbb{E}|X_i|^3}{\sqrt{\lambda(\mu^2+\sigma^2)^3}} &= \frac{C_{BE}\mathbb{E}|X_i|^3}{\sqrt{b^3/b^2(\mu^2+\sigma^2)^3}} \\ &= \frac{C_{BE}b\mathbb{E}|X_i|^3}{\sqrt{(b(\mu^2+\sigma^2))^3}} \\ &= O(b^{-1/2}).\end{aligned}$$

Having shown the induction hypothesis for $t = 0$, we now assume it holds for some t. Notice that $f(x) = (\rho-1)\frac{e^x}{1+e^x} - \frac{1}{2}(\rho-1)^2(\frac{e^x}{1+e^x})^2 + O(b^{-3/2})$, by Taylor's expansion, and using (14). Then by using dominated convergence theorem [3] and Lemma 2 we obtain

$$\begin{aligned}\mathbb{E}f(\xi_0^t) = (\rho-1)\frac{\kappa(1-\alpha)}{1-\kappa}\mathbb{E}\frac{1}{1+e^{\xi_1^t}} - \\ \frac{(\rho-1)^2\kappa(1-\alpha)}{2(1-\kappa)}\mathbb{E}\frac{e^{\xi_1^t}}{(1+e^{\xi_1^t})^2} + O(b^{-3/2})\end{aligned} \tag{24}$$

and

$$\mathbb{E}f(\xi_1^t) = (\rho-1)\mathbb{E}\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}} - \frac{(\rho-1)^2}{2}\mathbb{E}\frac{e^{2\xi_1^t}}{(1+e^{\xi_1^t})^2} + O(b^{-3/2}). \tag{25}$$

Now we take the expectation of both sides of (9) and (10). Here we use 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) random variables, hence obtaining

$$\mathbb{E}\xi_0^{t+1} = h + \log(\frac{p}{q})\kappa b\alpha + \mathbb{E}f(\xi_0^t)(1-\kappa)b + \mathbb{E}f(\xi_1^t)\kappa b(1-\alpha) \tag{26}$$

and

$$\mathbb{E}\xi_1^{t+1} = h + \log(\frac{p}{q})\kappa a\alpha + \mathbb{E}f(\xi_0^t)(1-\kappa)b + \mathbb{E}f(\xi_1^t)\kappa a(1-\alpha). \tag{27}$$

We now substitute (24) and (25) in (26) to get:

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

which on simplifying and grouping like terms becomes

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

Since $h = -\kappa(a-b) - \log\left(\frac{1-\kappa}{\kappa(1-\alpha)}\right)$

$$\mathbb{E}\xi_0^{t+1} = -\log\left(\frac{1-\kappa}{\kappa(1-\alpha)}\right) - \alpha\kappa(a-b) + \kappa b\alpha\log(\rho) - \frac{\lambda(1-\kappa)(1-\alpha)}{2\kappa}\mathbb{E}\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}}.$$

Using (17) we get

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

Finally we obtain

$$\mathbb{E}\xi_0^{t+1} = -\log(\frac{1-\kappa}{\kappa(1-\alpha)}) - \frac{\alpha\lambda(1-\kappa)}{2\kappa} - \lambda\frac{(1-\kappa)(1-\alpha)}{2\kappa}\mathbb{E}(\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}}) + O(b^{-1/2}). \tag{28}$$

Using exactly the same simplifications we can get

$$\mathbb{E}\xi_1^{t+1} = -\log(\frac{1-\kappa}{\kappa(1-\alpha)}) + \frac{\alpha\lambda(1-\kappa)}{2\kappa} + \lambda\frac{(1-\kappa)(1-\alpha)}{2\kappa}\mathbb{E}(\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}}) + O(b^{-1/2}). \tag{29}$$

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

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

and using Lemma 2 the above becomes

$$\mathbb{E}f^2(\xi_0^t) = (\rho-1)^2\frac{\kappa(1-\alpha)}{1-\kappa}\mathbb{E}\frac{e^{\xi_1^t}}{(1+e^{\xi_1^t})^2} + O(b^{-3/2}). \tag{30}$$

Similarly,

$$\mathbb{E}f^2(\xi_1^t) = (\rho-1)^2\mathbb{E}\frac{e^{2\xi_1^t}}{(1+e^{\xi_1^t})^2} + O(b^{-3/2}). \tag{31}$$

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

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

Substituting (30) and (31) into the above equations we get

$$\mathrm{Var}\xi_1^{t+1} = \mathrm{Var}\xi_0^{t+1} = \frac{\lambda\alpha(1-\kappa)}{\kappa} + \frac{\lambda(1-\kappa)(1-\alpha)}{\kappa} \mathbb{E}\frac{\exp\xi_1^t}{1+\exp(\xi_1^t)}. \tag{32}$$

Let us use $\mu^{(t+1)}$ to denote $\mathrm{Var}\xi_1^{(t+1)} = \mathrm{Var}\xi_0^{(t+1)}$. Then

$$\begin{aligned} \mathbb{E}\xi_0^{t+1} &= -\log\left(\frac{(1-\kappa)}{\kappa(1-\alpha)}\right) - \frac{1}{2}\mu^{(t+1)} + O(b^{-1/2}) \\ \mathbb{E}\xi_1^{t+1} &= -\log\left(\frac{(1-\kappa)}{\kappa(1-\alpha)}\right) + \frac{1}{2}\mu^{(t+1)} + O(b^{-1/2}). \end{aligned} \tag{33}$$

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 $1/(1+e^{-\xi_1^t})$ is bounded, by Bounded Convergence Theorem this means $\mathbb{E}[1/(1+e^{-\xi_1^t})] \to \mathbb{E}[1/(1+e^{-\mathcal{N}(\mathbb{E}\xi_1^t,\mu^{(t)})})]$. 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 can write and using (33) we obtain

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

Substituting the above into (32) gives us the recursion for $\mu^{(t+1)}$ given in (13).

Next we prove Gaussianity. Consider

$$\begin{aligned} &\xi_0^{t+1} - \mathbb{E}\xi_0^{t+1} \\ &= \log\left(\frac{p}{q}\right)(L_{0c} - \mathbb{E}L_{0c}) + \sum_{i=1}^{L_{00}}(f(\xi_{0,i}^t) - \mathbb{E}f(\xi_0^t)) + \\ &\quad \sum_{i=1}^{L_{01}}(f(\xi_{1,i}^t) - \mathbb{E}f(\xi_1^t)) + (L_{00} - \mathbb{E}L_{00})\mathbb{E}f(\xi_0^t) + \\ &\quad (L_{01} - \mathbb{E}L_{01})\mathbb{E}f(\xi_1^t). \end{aligned} \tag{34}$$

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 $D \equiv \sum_{i=1}^{L_{00}} X_i - \sum_{i=1}^{\mathbb{E}L_{00}} X_i$. Here the summation is taken up to $i \leq \mathbb{E}L_{00}$. Then $\mathbb{E}D^2 = |\sum_{i=1}^{\delta} X_i|^2$, where $\delta \leq |L_{00} - \mathbb{E}L_{00}|+1$, where the extra 1 is because $\mathbb{E}L_{00}$ may not be an integer. Therefore $\mathbb{E}D^2 = \mathbb{E}\delta\mathbb{E}|X_1|^2 \leq (C/b)((1-\kappa)b+1)^{1/2} = O(1/\sqrt{b})$. Thus, we can replace the Poisson upper limits of the summations in the second and third terms of (34) by their means, leading to

$$\begin{aligned} \xi_0^{t+1} - \mathbb{E}\xi_0^{t+1} = \log\left(\frac{p}{q}\right)(L_{0c} - \mathbb{E}L_{0c}) + \sum_{i=1}^{\mathbb{E}L_{00}}(f(\xi_{0,i}^t) - \mathbb{E}f(\xi_0^t)) \\ + \sum_{i=1}^{\mathbb{E}L_{01}}(f(\xi_{1,i}^t) - \mathbb{E}f(\xi_1^t)) + (L_{00} - \mathbb{E}L_{00})\mathbb{E}f(\xi_0^t) + \\ (L_{01} - \mathbb{E}L_{01})\mathbb{E}f(\xi_1^t) + o_p(1), \end{aligned} \tag{35}$$

where $o_p(1)$ indicates a random variable that goes to zero in probability in the limit.

The variance of the above term is μ^{t+1}, defined in (13), and it is finite for a fixed t. Now since we have an infinite sum of independent random variables as $a, b \to \infty$, with zero mean and finite variance, from standard CLT we can conclude that the distribution tends $\mathcal{N}(0, \mu^{t+1})$. $\square$

5 Detection Method

It is shown [7] that asymptotically the tests

$$\widehat{S}_0 = \{i : R_i^t > \log \frac{1-\kappa}{\kappa(1-\alpha)}\},$$

and $\widehat{S}$, the output of Algorithm 1, have the same fraction of miss-classified nodes. So we now go on to show that $\widehat{S}_0$ weakly recovers S, i.e., the expected fraction of missclassified nodes approaches 0, and the result for $\widehat{S}$ follows. By Lemma 1 we work with Λ_i instead of R_i. Consider the estimator on the tree:

$$\widehat{\tau}_i = \begin{cases} 1 & \text{if } \Lambda_i^t \geq \log \frac{1-\kappa}{\kappa(1-\alpha)}, \\ 0 & \text{otherwise.} \end{cases}$$

Alternatively, $\widehat{\tau}_i = \mathbf{1}_{\{\widetilde{\Lambda}_i^t \geq 0\}}$. The above estimator minimises the following error probability:

$$p_e = \overline{\mathbb{P}}\{\tau_i = 1\}\overline{\mathbb{P}}(\widehat{\tau}_i = 0 | i \in S) + \overline{\mathbb{P}}(\widehat{\tau}_i = 1 | i \notin S)\overline{\mathbb{P}}\{\tau_i = 0\}.$$

In the following proposition, we state and prove the main result of our paper. We show that the expected fraction of miss-classified nodes goes to zero for an infinitesimally small subgraph size, for any $\lambda > 0$. This implies that BP with cue beats BP without cues, which requires $\lambda > 1/e$ for zero asymptotic error rate ([7, 11]).

Proposition 2 *In the regime where $a, b \to \infty$ we have*

$$\lim_{\kappa \to 0} \frac{\mathbb{E} S \Delta \widehat{S}}{K(1-\alpha)} \to 0,$$

for any $\lambda > 0$, i.e., the expected fraction of miss-classified nodes tends to zero, as long as α is strictly positive.

Proof: We upperbound the error rate of $\widehat{S}_0$ and the result for $\widehat{S}$ follows based on the explanation in Section 5. By Lemma 1, the Λ_u^t and R_u^t have the same distributions on an event whose probability goes to 1. Therefore it is sufficient to bound the error for the tree, as follows:

$$\begin{aligned} \frac{\mathbb{E} S \Delta \widehat{S}_0}{K(1-\alpha)} &= \left(\frac{n - K\alpha}{K - K\alpha}\right) p_e \\ &\overset{(a)}{=} \left(\frac{(1-\kappa)}{\kappa(1-\alpha)}\right) P_0^t(\xi > 0) + P_1^t(\xi < 0) \end{aligned} \tag{36}$$

where in (a) P_0^t and P_1^t denote probabilities w.r.t. the distributions of ξ_0^t, ξ_1^t respectively. We now analyse the asymptotic value of each term in (36) in the limit as $\kappa \to 0$ with α fixed. By Proposition 1 we have that in the limit where $a \to \infty$ and $b \to \infty$,

$$P_1^t(\xi < 0) = Q\left(\frac{1}{\sqrt{\mu^{(t)}}}\left(\frac{\mu^{(t)}}{2} - \log(\frac{(1-\kappa)}{\kappa(1-\alpha)})\right)\right)$$

where $Q(\cdot)$ denotes the standard Q function. Notice that by (13) we have that $\mu^{(t)} \geq \lambda\alpha(1-\kappa)/\kappa$, since $F(\mu) \equiv \mathbb{E}\frac{1-\kappa}{\kappa(1-\alpha)+(1-\kappa)\exp(-\mu/2-\sqrt{\mu}Z)} \geq 0$. In addition, by (32), $\mu^{(t)} \leq \frac{\lambda(1-\kappa)}{\kappa}$. Therefore $\mu^{(t)} = \Theta(\frac{1}{\kappa})$. Note that the lower bound on $\mu^{(t)}$ is not useful when $\alpha = 0$. Consequently $\lim_{\kappa\to 0} \frac{1}{\mu^{(t)}} \log(\frac{(1-\kappa)}{\kappa(1-\alpha)}) = 0$. Therefore:

$$\begin{aligned} P_1^t(\xi < 0) &= \frac{1}{\kappa} Q(\sqrt{\mu^{(t)}}(1 + O(\kappa))) \\ &\leq \exp(-\Theta(1/\kappa)) \to 0. \end{aligned}$$

Similarly we have

$$\frac{1}{\kappa} P_0^t(\xi > 0) \tag{37}$$

$$= \frac{1}{\kappa} Q(\frac{\log(\frac{(1-\kappa)}{\kappa(1-\alpha)}) + \frac{\mu}{2}}{\sqrt{\mu^{(t)}}}) \tag{38}$$

$$\leq \frac{1}{\kappa} \exp(-\Theta(\frac{1}{\kappa})) \tag{39}$$

$$\to 0. \tag{40}$$

Substituting these back in (36) the result then follows. □

6 Numerical Experiments

In this section we provide simulation results to corroborate our theoretical findings and also to demonstrate the performance improvement of Algorithm 1 in the presence of side-information. We fix $n = 10^4, b = 100$, and $\kappa = 0.005$, giving $K = 50$. Next we sweep over different values of λ in the range $[0.1, 0.8]$ and average over 1000 graph realisations to find the fraction of miss-classified subgraph nodes for each value of λ. In Figure 1 we have the ratio between the number of subgraph nodes wrongly classified by the algorithm and the number of unlabelled subgraph nodes on the y-axis and λ on the x-axis. This demonstrates that there is a marked improvement in the performance of BP with the introduction of cues.

In Figure 2 we plot the theoretical error of Algorithm 1 given in (36) against κ for the two cases of $\alpha = 0$ (no cues) and $\alpha = 0.1$ (10% cues) for $\lambda = \frac{1}{2e}$. We have chosen this value of λ in order to be below the detectability threshold of $\lambda = \frac{1}{e}$ of BP without cues. We can observe that contrary to when $\alpha = 0$, with $\alpha = 0.1$ the error decreases as κ decreases, as proved in our analysis. We also observed this in our simulations where we obtained an error rate of 73.86% for $\kappa = 4 \times 10^{-4}$ ($n = 5 \times 10^4$) with $\alpha = 0.1$, whereas it was 0.995 when $\alpha = 0$.

7 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 SNR parameter λ, unlike BP without cues, where there is a non-zero detectability threshold. We also obtained some simulation results on synthetic graphs to demonstrate the improvement in error rates in the presence of cues. In the future, we would like to investigate non-asymptotic properties of the algorithm for finite a and b and when $K = o(n)$.

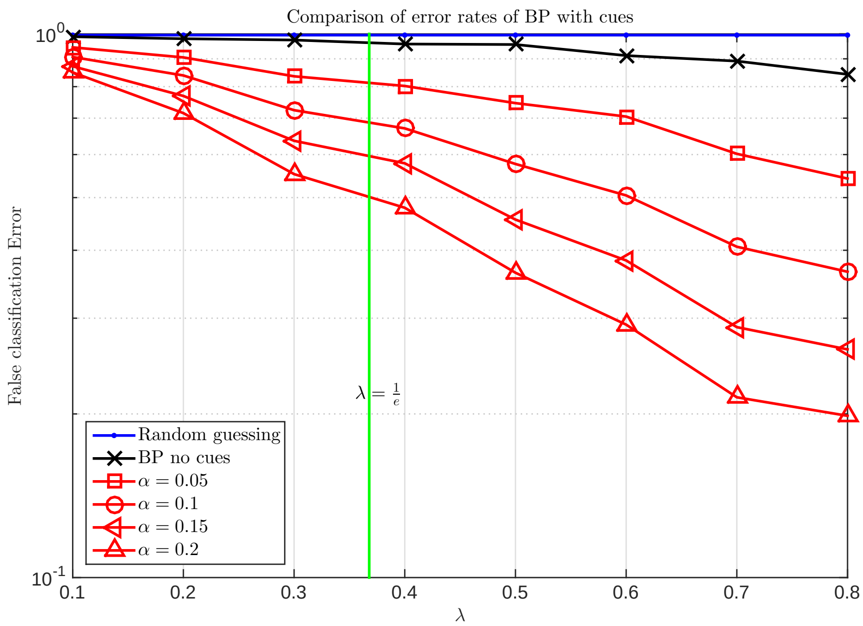

Figure 1: Comparison of error rates for different α

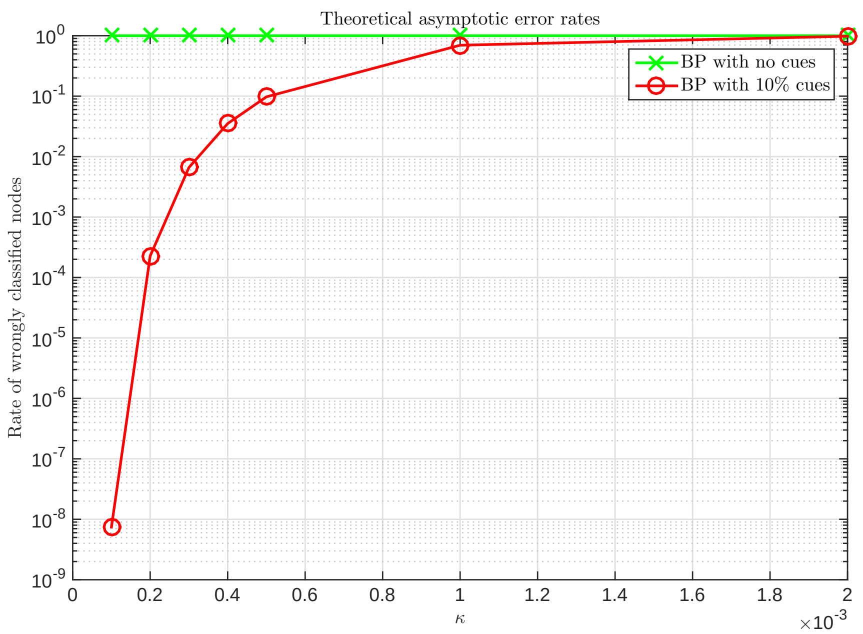

Figure 2: Theoretical error rates for small κ

References

[1] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," *Data Mining and Knowledge Discovery*, vol. 29, no. 3, pp. 626–688, 2015.

[2] 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.

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

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

[5] 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.

[6] B. Hajek, Y. Wu, and J. Xu, "Information limits for recovering a hidden community," *arXiv preprint arXiv:1509.07859*, 2015.

[7] ——, "Recovering a Hidden Community Beyond the Spectral Limit in $O(|E|\log^*|V|)$ Time," *arXiv Prepr. arXiv1510.02786*, 2015.

[8] A. Kadavankandy, L. Cottatellucci, and K. Avrachenkov, "Characterization of L^1-norm statistic for Anomaly Detection in Erdös Rényi Graphs," in *CDC*. IEEE, 2016. [Online]. Available: http://www.eurecom.fr/en/publication/list/author/cottatellucci-laura

[9] A. Martinsson, "Lovasz θ function , SVMs and Finding Dense Subgraphs," *J. Mach. Learn. Res.*, vol. 14, pp. 3495–3536, 2013.

[10] T. Mifflin, C. Boner, G. Godfrey, and J. Skokan, "A random graph model for terrorist transactions," in *2004 IEEE Aerosp. Conf. Proc. (IEEE Cat. No.04TH8720)*, vol. 5. IEEE, 2004, pp. 3258–3264.

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

[12] E. Mossel and J. Xu, "Local Algorithms for Block Models with Side Information," in *Proc. 2016 ACM Conf. Innov. Theor. Comput. Sci. - ITCS '16*. New York, New York, USA: ACM Press, jan 2016, pp. 71–80.

Contents

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399