

# The Power of Side-information in Subgraph Detection

Arun Kadavankandy, Konstantin Avrachenkov, Laura Cottatellucci, Rajesh Sundaresan

► **To cite this version:**

Arun Kadavankandy, Konstantin Avrachenkov, Laura Cottatellucci, Rajesh Sundaresan. The Power of Side-information in Subgraph Detection. [Research Report] RR-8974, Inria Sophia Antipolis. 2017, pp.37. <hal-01394889v4>

**HAL Id: hal-01394889**

**<https://hal.inria.fr/hal-01394889v4>**

Submitted on 6 Mar 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# The Power of Side-information in Subgraph Detection

Arun Kadavankandy, Konstantin Avrachenkov, Laura Cottatellucci  
and Rajesh Sundaresan

**RESEARCH  
REPORT**

**N° 8974**

Feb 2017

Project-Team Neo





## The Power of Side-information in Subgraph Detection

Arun Kadavankandy, Konstantin Avrachenkov,  
Laura Cottatellucci and Rajesh Sundaresan

Project-Team Neo

Research Report n° 8974 — Feb 2017 — 37 pages

**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 an 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 the SNR parameter. We validate our results through simulations on synthetic datasets as well as on a few real world networks.

**Key-words:** Belief Propagation, Dense community detection, Cavity Method, Detectability, Stochastic Block Model

**RESEARCH CENTRE  
SOPHIA ANTIPOLIS – MÉDITERRANÉE**

2004 route des Lucioles - BP 93  
06902 Sophia Antipolis Cedex

## Le Pouvoir d'Information Supplémentaire en Detection des Sousgraphes

**Résumé :** Dans ce travail, nous nous attaquons au problème de la detection d'une communauté cachée dans un grand graphe. Nous considérons Belief Propagation (BP) pour cela, et le modele que nous utilisons est celui d'un petite graph Erdős-Rényi qui est place dans un plus grande graphe, aussi Erdős-Rényi. Nous assumons qu'il y a des informations supplémentaires de deux categories : la premiere où certains sommets qui appartiennent à la communauté concernée sont connus, qui s'appellent des sommets indices (parfaites informations supplémentaires), et la deuxième où certains sommets sont donnés, mais qu'une partie d'eux ne sont correctes (informations supplémentaires imparfaites). L'algorithme basé sur Belief propagation que nous présentons peut fonctionner en presence de ces deux categories d'information supplémentaires, et en l'analysant nous prouvons que meme une très peu de leur quantité enlève un seuil de « detectabilité » qui existe dans Belief Propagation sans aucunes informations supplémentaires. Nous fournissons aussi des résultats d'experiences conduites sur des graphes synthétiques et aussi deux graphes reels.

**Mots-clés :** Belief Propagation, Graphes Aléatoires non-orientés, Detection de Commaunautes, Stochastic Block Model, Methode de Cavite, Seuil de Detectabilite

# 1 Introduction

## 1.1 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 example application of dense subgraph detection in Signal Processing is the problem of Correlation Mining [1]. Given a network of correlated signals, a graph is formed with nodes representing 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 [1]. Dense subgraph detection also finds application in real-world computer and social networks; for e.g., in detecting fraudulent activity [2–4]. It can, in addition, be viewed as a signal recovery problem on graphs [5, 6].

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 [7]. 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-Rényi graph. This model was used in [8] to capture terrorist transactions in a computer network. It is a special case of the Stochastic Block Model (SBM), which has been widely used to assess the performance of different community detection algorithms [9].

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 [10–12]. 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 [13–15].

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 ML detection on certain types of graphs using distributed and local message passing [16]. It belongs to the class of guilt-by-association schemes [17] and has been successfully applied to many practical problems in graphs such as fraud detection [2] and data mining [18].

## 1.2 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 [19–21] 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  $c\sqrt{n}$  [22] for some  $c > 0$ , and the minimum clique-size that can be detected in nearly-linear time is believed to be  $\sqrt{n/e}$  [23].

The computational barriers for subgraph detection in a sparse graph without cues were studied in [10, 11, 24]. In [11] the author investigated the performance of Maximum Likelihood (ML) detection and BP, and analyzed the phase transition with respect to an effective signal-to-noise ratio (SNR) parameter  $\lambda$  defined as

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

The larger the  $\lambda$ , 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  $\kappa$ ,  $a$  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  $\kappa \rightarrow 0$ . 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 [24] the authors show that when  $K = o(n)$ , i.e., when  $\kappa \rightarrow 0$ , 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  $\lambda > 1/e$  [10]. 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 knowledge, 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 [25], 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 [12, 26–28]. 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 [27] and [28] consider a BP algorithm to detect two equal-sized communities. In [28], the side-information is such that all nodes indicate their community information after passing it through a binary symmetric channel with error rate  $\alpha$ . 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 [29]. In [27], 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 [12], which considers asymmetric communities and asymmetric connection probabilities within communities. In this setting, the authors of [12] 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, [28] and [27] consider equal-sized communities with the same edge probability within each community. In [12, 27, 28], 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 the subgraph. In this setting we show that the computational barrier disappears when side-

information is available. We emphasize that our results cannot be obtained as a special case of the results in [12, 26–28].

### 1.3 Summary of Results

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

1. A fraction  $\alpha$  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 [13–15].

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  $\kappa$  fixed, as in the regime of [11]. Under this setting, 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  $\kappa$ .

Based on these recursions, we obtain closed form expressions for the distributions when  $a, b \rightarrow \infty$ . We then show that when there is non-trivial side-information, the expected fraction of misclassified nodes goes to zero as  $\kappa \rightarrow 0$ , for any positive value of the respective SNR parameter  $\lambda_\alpha$  or  $\lambda$ , 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 review to ISIT 2017 [30]. 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.

### 1.4 Organization

The rest of the paper is organized as follows. In Subsection 1.5 we delineate useful notation. In Section 2 we describe the model and define the problem in detail. In Section 3, we present our algorithm with perfect cues and explain the steps in its derivation. In Section 4 we derive the asymptotic distribution of BP messages. In particular, in section 4.1, we prove our main result on the asymptotic error rate of our algorithm. In Section 5 we present our algorithm with imperfect side-information and provide a result on its asymptotic error rate. In Section 6 we present results on our experiments on the synthetic graph as well as a few real-world graphs. In Section 7, we conclude with some suggestions for future work. Some proofs are relegated to supplementary material for lack of space.

### 1.5 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  $\delta i$  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:  $C = A \setminus B$  is the set of elements that belong to  $A$  and not  $B$  and  $\Delta$  denotes the set difference, i.e.,  $A \Delta B = (A \cup B) \setminus (A \cap B)$ . 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  $\gamma$ . Also,  $\mathcal{N}(\mu, \sigma^2)$  denotes the Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . The symbol  $\xrightarrow{D}$  denotes convergence in distribution.

## 2 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  $\mathbb{P}((i, j) \in E | i \in S, j \notin S) = \mathbb{P}((i, j) \in E | i \notin S, j \in S) = q$ . 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  $\sigma : V \rightarrow \{0, 1\}^n$  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  $c : V \rightarrow \{0, 1\}^n$  such that (s.t.)  $c_i = 1$  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

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

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  $i \in V$ :

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

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

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

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  $\{\sigma_i, i \in V \setminus C\}$ . The optimal detector that minimizes the expected number of misclassified nodes is the per-node MAP detector given as [24]:

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

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 Belief Propagation (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  $\alpha\beta > 0$ .

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

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

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

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

where  $G_i^t \setminus u$  and  $C_i^t \setminus u$  are defined as above, but excluding the contribution from node  $u$ . Node  $i$  updates  $R_{i \rightarrow u}^t$  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  $O(t_f |E|)$ .

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<sup>1</sup>  $T_u^t$  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  $n \rightarrow \infty$  (i.e., with high probability (w.h.p)). The

<sup>1</sup>Detailed in Appendix A

---

**Algorithm 1** BP with perfect side-information
 

---

- 1: Initialize: Set  $R_{i \rightarrow j}^0$  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  $i, u \notin C$ :

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

where  $v = \log\left(\frac{n-K}{K(1-\alpha)}\right)$ .

- 3: Increment  $t$ , if  $t < t_f - 1$  go back to 2, else go to 4
- 4: Compute  $R_u^{t_f}$  for every  $u \in V \setminus C$  as follows:

$$R_u^{t+1} = -K(p-q) + \sum_{l \in C_u^1} \log\left(\frac{p}{q}\right) + \sum_{l \in \delta u \setminus C_u^1} \log\left(\frac{\exp(R_{l \rightarrow u}^t - v)(p/q) + 1}{\exp(R_{l \rightarrow u}^t - v) + 1}\right) \quad (8)$$

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

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.

## 4 Asymptotic Error Analysis

In this section we analyze the distributions of BP messages  $R_{i \rightarrow u}^t$  given  $\{\sigma_i = 1\}$  and given  $\{\sigma_i = 0\}$  for  $i \in V \setminus C$ . First, we derive a pair of recursive equations for the asymptotic distributions of the messages  $R_{i \rightarrow u}^t$  given  $\{\sigma_i = 0, c_i = 0\}$  and given  $\{\sigma_i = 1, c_i = 0\}$  in the limit as  $n \rightarrow \infty$  in Lemma 1. In Proposition 1 we present the asymptotic distributions of the messages in the large degree regime where  $a, b \rightarrow \infty$ . 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 [11], which details the limitation of local algorithms.

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

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

and the associated messages  $\tilde{R}_{i \rightarrow u}^t$ . By Bayes rule,  $\tilde{R}_{i \rightarrow u}^t = R_{i \rightarrow u}^t - v$ , where

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

Let  $\xi_0^t, \xi_1^t$  be rvs with the same distribution as the messages  $\tilde{R}_{i \rightarrow u}^t$  given  $\{\sigma_i = 0, c_i = 0\}$  and given  $\{\sigma_i = 1, c_i = 0\}$ , respectively in the limit as  $n \rightarrow \infty$ . 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  $\xi_0^t$  and  $\xi_1^t$  satisfy the following recursive distributional equations with initial conditions  $\xi_0^0 = \xi_1^0 = \log(\kappa(1-\alpha)/(1-\kappa))$ .*

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

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

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

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

The rvs  $\xi_{0,i}^t, i = 1, 2, \dots$  are independent and identically distributed (iid) with the same distribution as  $\xi_0^t$ . Similarly  $\xi_{1,i}^t, i = 1, 2, \dots$  are iid with the same distribution as  $\xi_1^t$ . Furthermore,  $L_{00} \sim \text{Poi}((1-\kappa)b)$ ,  $L_{01} \sim \text{Poi}(\kappa b(1-\alpha))$ ,  $L_{10} \sim \text{Poi}((1-\kappa)b)$ ,  $L_{11} \sim \text{Poi}(\kappa a(1-\alpha))$ ,  $L_{0c} \sim \text{Poi}(\kappa b \alpha)$  and  $L_{1c} \sim \text{Poi}(\kappa p \alpha)$ .

*Proof.* This follows from (7) and the tree coupling in Lemma 3 of Appendix A.  $\square$

We define the effective SNR 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  $a, b \rightarrow \infty$  such that  $\lambda_\alpha$  is kept fixed.

**Proposition 1.** *In the regime where  $\lambda_\alpha$  and  $\kappa$  are held fixed and  $a, b \rightarrow \infty$ , we have*

$$\begin{aligned} \xi_0^{t+1} &\xrightarrow{D} \mathcal{N}\left(-\log\frac{1-\kappa}{\kappa(1-\alpha)} - \frac{1}{2}\mu^{(t+1)}, \mu^{(t+1)}\right) \\ \xi_1^{t+1} &\xrightarrow{D} \mathcal{N}\left(-\log\frac{1-\kappa}{\kappa(1-\alpha)} + \frac{1}{2}\mu^{(t+1)}, \mu^{(t+1)}\right). \end{aligned}$$

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 Lemma from [10], which we need for our derivations.

**Lemma 2.** [10, Lemma 11] *Let  $S_\gamma = X_1 + X_2 + \dots + X_{N_\gamma}$ , where  $X_i$ , for  $i = 1, 2, \dots, N_\gamma$ , are independent, identically distributed rv with mean  $\mu$ , variance  $\sigma^2$  and  $\mathbb{E}(|X_i^3|) \leq g^3$ , and for some  $\gamma > 0$ ,  $N_\gamma$  is a  $\text{Poi}(\gamma)$  rv independent of  $X_i : i = 1, 2, \dots, N_\gamma$ . Then*

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

where  $C_{BE} = 0.3041$ .

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  $a, b \rightarrow \infty$  and  $\lambda$  and  $\kappa$  are held constant, we have  $\rho \rightarrow 1$  as follows:

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

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,  $\xi_1^t$ . We then use induction to show that the variance  $\mu^{(t+1)}$  satisfies the recursion (13) and use Lemma 2 to prove Gaussianity.  $\square$

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

#### 4.1 Detection Performance

Let us use the symbol  $\bar{S}$  to denote the subgraph nodes with the cued nodes removed, i.e.,  $\bar{S} = S \setminus C$ . This is the set that we aim to detect. The output of Algorithm 1,  $\hat{S}$  is the set of nodes with the top  $K - |C|$  beliefs. We are interested in bounding the expected number of misclassified nodes  $\mathbb{E}(|\bar{S} \Delta \hat{S}|)$ . Let  $\hat{S}$  be the output set of the algorithm excluding cues since the cues are always correctly detected. Note that  $|\bar{S}| = |\hat{S}| = K - |C|$ . To characterize the performance of the detector, we need to choose a performance measure. In [11], a rescaled probability of success was used to study the performance of a subgraph detector without cues, defined as

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

where  $\hat{\sigma}_i = \mathbf{1}(i \in \hat{S})$ , and the dependence of  $P_{\text{succ}}(\hat{\sigma})$  on  $n$  is implicit. In our work, we study the following error measure, which is the average fraction of misclassified nodes, also considered in [10], which for the uncued case is defined as

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

Observe that  $0 \leq \mathcal{E} \leq 2$ . 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.,  $|\hat{S}| = K$ , the following is true for any estimate  $\hat{\sigma}$  of  $\sigma$ :

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

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.

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

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

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

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

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

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

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

Consequently,

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

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

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

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,

$$\begin{aligned} |\overline{S}\Delta\widehat{S}| &\leq |\overline{S}\Delta\widehat{S}_0| + |\widehat{S}\Delta\widehat{S}_0| \\ &= |\overline{S}\Delta\widehat{S}_0| + |K - |C| - |\widehat{S}_0|| \\ &= |\overline{S}\Delta\widehat{S}_0| + |\overline{S}| - |\widehat{S}_0| \\ &\leq 2|\overline{S}\Delta\widehat{S}_0|, \end{aligned} \quad (22)$$

where the last step follows because the set 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.  $\square$

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

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 [11], the performance of any local algorithm suffers when the SNR parameter  $\lambda < 1/e$ . 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.** [11, Theorem 1] *If  $\lambda < 1/e$ , then all local algorithms have success probability uniformly bounded away from one; in particular,*

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

and therefore

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

## 5 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 classified 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  $\alpha$  and  $\beta$  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

$$\begin{aligned} R_u^t &= \log \left( \frac{\mathbb{P}(G_u^t, C_u^t | \sigma_u = 1, c_u)}{\mathbb{P}(G_u^t, C_u^t | \sigma_u = 0, c_u)} \right) + \log \left( \frac{\mathbb{P}(c_u | \sigma_u = 1)}{\mathbb{P}(c_u | \sigma_u = 0)} \right) \\ &= \log \left( \frac{\mathbb{P}(G_u^t, C_u^t | \sigma_u = 1)}{\mathbb{P}(G_u^t, C_u^t | \sigma_u = 0)} \right) + \log \left( \frac{\mathbb{P}(c_u | \sigma_u = 1)}{\mathbb{P}(c_u | \sigma_u = 0)} \right), \end{aligned} \quad (23)$$

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  $\sigma_u$ . 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

$$h_u = \begin{cases} \log \left( \frac{\beta(1-\kappa)}{(1-\beta)\kappa} \right), & \text{if } u \in C, \\ \log \left( \frac{(1-\alpha\beta)(1-\kappa)}{(1-\kappa-\alpha\kappa+\alpha\kappa\beta)} \right), & \text{otherwise.} \end{cases} \quad (24)$$

The recursion for the first term in (23) 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  $R_{u \rightarrow i}^t$ , for  $\{\sigma_u = 0\}$  and  $\{\sigma_u = 1\}$ . Note that, since we now assume

**Algorithm 2** BP with imperfect cues

- 1: Initialize: Set  $R_{i \rightarrow j}^0$  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  $(i, u) \in E$ :

$$R_{i \rightarrow u}^{t+1} = -K(p - q) + h_i + \sum_{l \in \delta i, l \neq u} \log \left( \frac{\exp(R_{l \rightarrow i}^t - \nu)(p/q) + 1}{\exp(R_{l \rightarrow i}^t - \nu) + 1} \right), \quad (25)$$

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  $R_u^{t_f}$  for every  $u \in V$  as follows:

$$R_u^{t+1} = -K(p - q) + h_u + \sum_{l \in \delta u} \log \left( \frac{\exp(R_{l \rightarrow u}^t - \nu)(p/q) + 1}{\exp(R_{l \rightarrow u}^t - \nu) + 1} \right) \quad (26)$$

- 5: Output  $\hat{S}$  as  $K$  set of nodes in  $V$  with the largest values of  $R_u^{t_f}$ .

that we do not know the exact classification of any of the subgraph nodes, we need to detect  $K$  nodes, and hence the effective SNR parameter is defined as

$$\lambda = \frac{K^2(p - q)^2}{(n - K)q}. \quad (27)$$

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

**Proposition 3.** *Let  $n \rightarrow \infty$ . In the regime where  $\lambda$  and  $\kappa$  are held fixed and  $a, b \rightarrow \infty$ , the message  $R_{u \rightarrow i}^t$  given  $\{\sigma_u = j\}$ , where  $j = \{0, 1\}$  converges in distribution to  $\Gamma_j^t + h_u$  where  $h_u$  is defined in (24). The rvs  $\Gamma_j^t$  have the following distribution:*

$$\begin{aligned} \Gamma_0^t &\sim \mathcal{N}(-\mu^{(t)}/2, \mu^{(t)}), \text{ and} \\ \Gamma_1^t &\sim \mathcal{N}(\mu^{(t)}/2, \mu^{(t)}), \end{aligned}$$

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

$$\begin{aligned} \mu^{(t+1)} &= \alpha\beta^2\lambda\mathbb{E} \left( \frac{(1 - \kappa)/\kappa}{\beta + (1 - \beta)e^{(-\mu^{(t)}/2 - \sqrt{\mu^{(t)}}Z)}} \right) + (1 - \alpha\beta)^2\lambda \\ &\mathbb{E} \left( \frac{(1 - \kappa)}{\kappa(1 - \alpha\beta) + (1 - \kappa - \alpha\kappa + \alpha\kappa\beta)e^{(-\mu^{(t)}/2 - \sqrt{\mu^{(t)}}Z)}} \right), \end{aligned} \quad (28)$$

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

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

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



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

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

Consequently,

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

*Proof.* The proof essentially analyzes the properties of the recursion (28) and is similar to the proof of Theorem 1. See supplementary material for details.  $\square$

## 6 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 [31].

### 6.1 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  $\alpha = 0.1, \kappa = 5 \times 10^{-4}, b = 100$  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  $\alpha$  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  $\alpha$ . Our parameter setting 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  $\alpha_{pr} = 0.9$  [31]. 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 [11] 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  $\lambda$  by varying  $p$ . In Figure 2 we plot the metric  $\mathcal{E}$  against  $\lambda$  for different values of  $\beta$ , 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  $\beta$  improves the error performance.

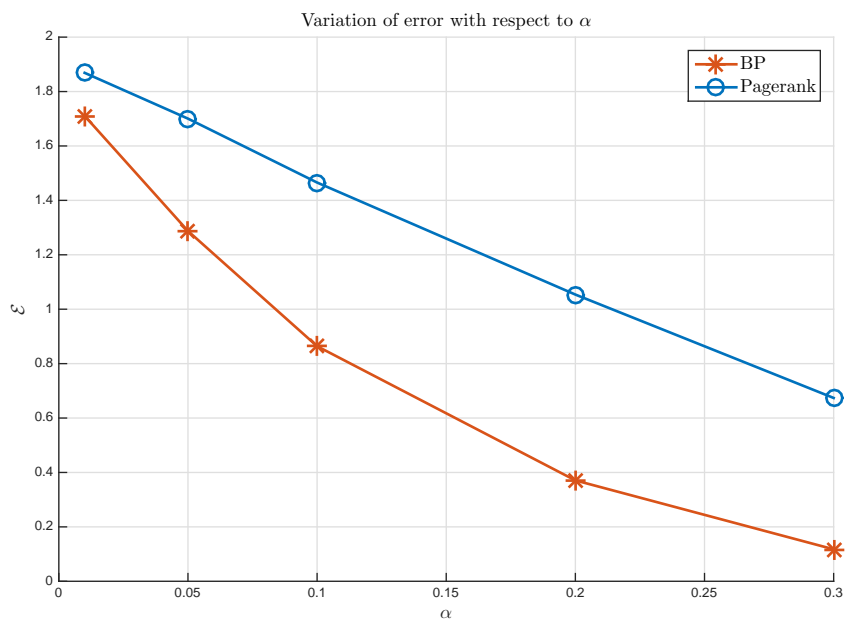


Figure 1: Performance of BP Algo 1 as a function of  $\alpha$

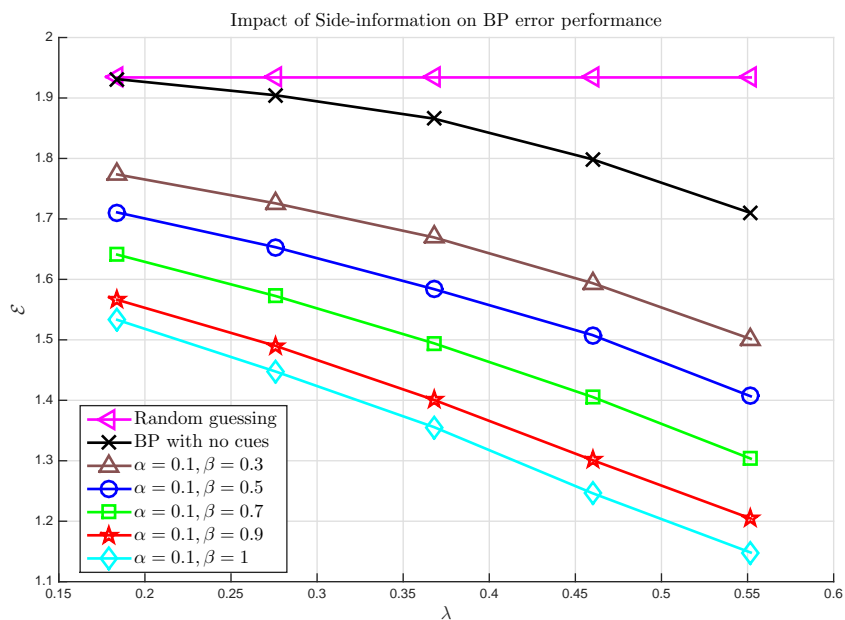


Figure 2: Comparison of BP for subgraph detection for different amounts of side-information

Class 0	#of cues = 1	#of cues = 2
BP	<b>0.7143</b>	<b>0.7216</b>
PageRank	0.6327	0.6392

Table 1: Reuters911 recall results

## 6.2 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

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

where  $S$  is the true community and  $\hat{S}$  is its estimate. This is a commonly used metric for community detection applications [32]. 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.

### 6.2.1 USPS dataset

The USPS dataset contains 9296 scanned images of size  $16 \times 16$ , which can be represented by a feature vector of size  $256 \times 1$  with values from  $-1$  to  $+1$  [33]. 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 3). It can be seen from Figure 3, 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 Figure 1, BP outperforms PageRank by a big margin. However in the USPS graph, there are ten dense communities, and therefore it deviates significantly from our underlying model.

### 6.2.2 Reuters911 Dataset

In this subsection we consider a graph that is closer to our assumed model. We consider the Reuters911 dataset also used in [34]. It is made up of words from all news released by Reuters for 66 days since September 11, 2001. Table 5 in [34] shows a group of 99 collocated words in this dataset. This subset represents the largest dense community to be detected in this dataset. 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*. In Table 1 we show the recall values  $\mathcal{R}$  of PageRank and BP, excluding cues. Clearly, BP performs better.

## 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

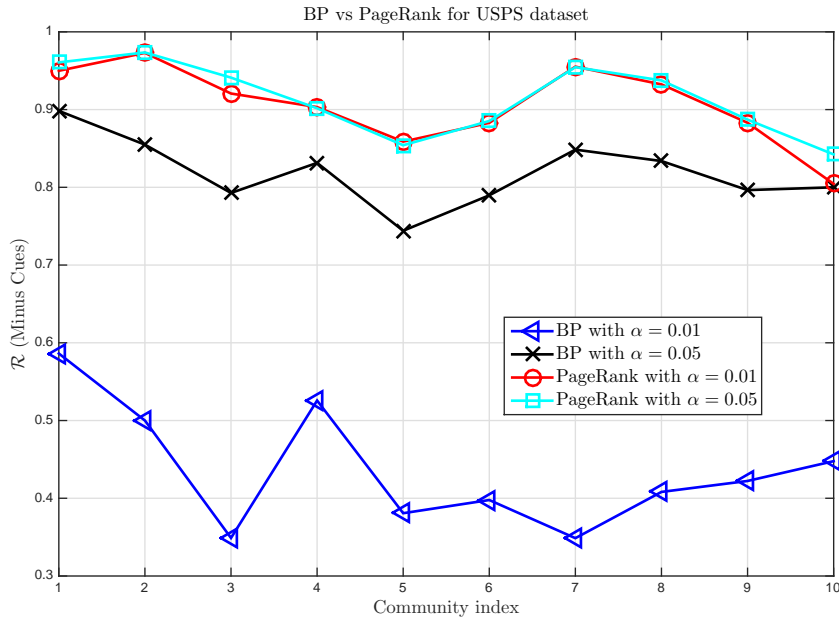


Figure 3: Comparison of BP for subgraph detection for different amounts of side-information

results based on density evolution on trees to show that it achieves zero asymptotic error regardless of the SNR parameter  $\lambda$ , 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  $\lambda < 1/e$ . 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  $\kappa$  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.

## 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.”

## 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 [10]): The label  $\tau_u$  at node  $u$  is chosen at random in the following way:

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

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  $u$ ,

$$\mathbb{P}(\tau_i = 1 | \tau_u = 1) = \frac{Kp}{d_1}, \quad \mathbb{P}(\tau_i = 1 | \tau_u = 0) = \frac{Kq}{d_0}. \quad (29)$$

We then assign the cue indicator function  $\tilde{c}$  such that  $\tilde{c}_i = 1$  w.p.  $\alpha$  if  $\tau_i = 1$  and  $\tilde{c}_i = 0$  if  $\tau_i = 0$ . The process is repeated up to depth  $t$  giving us  $\tilde{C}_u^t$ , the set of cued neighbours. Now we have the following coupling result between  $(G_u^t, \sigma^t, C_u^t)$ , the neighbourhood of  $u$  and the node labels of that neighbourhood and  $(T_u^t, \tau^t, \tilde{C}_u^t)$ , the depth- $t$  tree  $T_u^t$  and its labels due to [10].

**Lemma 3.** [10, Lemma 15] *For  $t$  such that  $(np)^t = n^{o(1)}$ , there exists a coupling such that  $(G_u^t, \sigma^t, C_u^t) = (T_u^t, \tau^t, \tilde{C}_u^t)$  with probability  $1 - n^{-1+o(1)}$ .*

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  $\Lambda_u^t$  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  $\tau_u$ . We use this fact to split  $\Lambda_u^{t+1}$  in two parts.

$$\begin{aligned} \Lambda_u^{t+1} &= \log \left( \frac{\mathbb{P}(T_u^{t+1}, C_u^{t+1} | \tau_u = 1)}{\mathbb{P}(T_u^{t+1}, C_u^{t+1} | \tau_u = 0)} \right) \\ &= \log \left( \frac{\mathbb{P}(N_u | \tau_u = 1)}{\mathbb{P}(N_u | \tau_u = 0)} \right) + \end{aligned} \quad (30)$$

$$\sum_{i \in \delta u} \log \left( \frac{\mathbb{P}(T_i^t, c_i, C_i^t | \tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t | \tau_u = 0)} \right), \quad (31)$$

by the independence property of subtree  $T_i^t$  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

$$\begin{aligned} \log\left(\frac{\mathbb{P}(N_u|\tau_u = 1)}{\mathbb{P}(N_u|\tau_u = 0)}\right) &= N_u \log\left(\frac{d_1}{d_0}\right) - (d_1 - d_0) \\ &= N_u \log\left(\frac{d_1}{d_0}\right) - K(p - q). \end{aligned} \quad (32)$$

Next we look at the second term in (31). 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

$$\begin{aligned} &\log\left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 0)}\right) \\ &\stackrel{(a)}{=} \log\left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t, \tau_i = 1|\tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t, \tau_i = 1|\tau_u = 0)}\right) \\ &= \log\left(\frac{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_i = 1)\mathbb{P}(\tau_i = 1|\tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t|\tau_i = 1)\mathbb{P}(\tau_i = 1|\tau_u = 0)}\right) \\ &\stackrel{(b)}{=} \log\left(\frac{Kp/d_1}{Kq/d_0}\right), \end{aligned} \quad (33)$$

where in step (a) we applied the fact that  $c_i = 1$  implies  $\tau_i = 1$ , and in (b) we used (29).

*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

$$\begin{aligned} &\mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 1) \\ &= \mathbb{P}(T_i^t, C_i^t|\tau_i = 1)\mathbb{P}(c_i|\tau_i = 1)\mathbb{P}(\tau_i = 1|\tau_u = 1) \\ &\quad + \mathbb{P}(T_i^t, C_i^t|\tau_i = 0)\mathbb{P}(c_i|\tau_i = 0)\mathbb{P}(\tau_i = 0|\tau_u = 1) \\ &= \mathbb{P}(T_i^t, C_i^t|\tau_i = 1)(1 - \alpha)\frac{Kp}{d_1} \\ &\quad + \mathbb{P}(T_i^t, C_i^t|\tau_i = 0)\frac{(n - K)q}{d_1}. \end{aligned} \quad (34)$$

Similarly, we can show

$$\begin{aligned} \mathbb{P}(T_i^t, c_i, C_i^t|\tau_u = 0) &= \mathbb{P}(T_i^t, C_i^t|\tau_i = 1)\frac{Kq}{d_0}(1 - \alpha) \\ &\quad + \mathbb{P}(T_i^t, C_i^t|\tau_i = 0)\frac{(n - K)q}{d_0}. \end{aligned} \quad (35)$$

Let us define

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

the message that  $i$  sends to  $u$  at step  $t$ . Using the above definition, (34), and (35) we get

$$\begin{aligned}
& \log \left( \frac{\mathbb{P}(T_i^t, c_i, C_i^t | \tau_u = 1)}{\mathbb{P}(T_i^t, c_i, C_i^t | \tau_u = 0)} \right) \\
&= \log \left( \frac{e^{\Lambda_{i \rightarrow u}^t} \frac{Kp}{d_1} (1 - \alpha) + \frac{(n-K)q}{d_1}}{e^{\Lambda_{i \rightarrow u}^t} \frac{Kq}{d_0} (1 - \alpha) + \frac{(n-K)q}{d_0}} \right) \\
&= \log \left( \frac{d_0}{d_1} \right) + \log \left( \frac{e^{\Lambda_{i \rightarrow u}^t} \frac{Kp}{(n-K)q} (1 - \alpha) + 1}{e^{\Lambda_{i \rightarrow u}^t} \frac{K}{(n-K)} (1 - \alpha) + 1} \right). \tag{36}
\end{aligned}$$

We then use the substitution  $\nu := \log((n-K)/K)$  in the above equation. Finally combining (32), (33) and (36) and replacing  $\Lambda_u^t$  with  $R_u^t$  and  $\Lambda_{i \rightarrow u}^t$  with  $R_{i \rightarrow u}^t$ , we arrive at (8). The recursive equation (7) can be derived in exactly the same way by looking at the children of  $i \in \delta u$ .

## B Proof of Proposition 1

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

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

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  $\xi_0^t$  and  $\xi_1^t$  denote the rvs whose distributions are the same as the distributions of  $\tilde{\Lambda}_i^t$  given  $\{c_i = 0, \tau_i = 0\}$  and  $\{c_i = 0, \tau_i = 1\}$  respectively in the limit  $n \rightarrow \infty$ . We need a relationship between  $P_0$  and  $P_1$ , the probability measures of  $\xi_0^t$  and  $\xi_1^t$  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  $g(\cdot)$*

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

*Proof.* Following the logic in [11], we show this result for  $g(\tilde{\Lambda}_u^t) = \mathbf{1}(\tilde{\Lambda}_u^t \in A)$ ,  $A$  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 [35]. Let  $Y = (T_u^t, C_u^t)$ , the observed rv. Therefore

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

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

*Proof.* Since  $\lambda_\alpha$  and  $\kappa$  are fixed and  $b \rightarrow \infty$ , from (12) we have

$$\rho := a/b = 1 + \sqrt{\frac{\lambda_\alpha(1 - \kappa)}{(1 - \alpha)^2 \kappa^2 b}} = 1 + O(b^{-1/2}). \quad (37)$$

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  $\xi_0^0 = \xi_1^0 = -v$ . We only do this for  $\xi_0^t$  since the steps are similar for  $\xi_1^t$ . Observe that

$$\begin{aligned}
f(-v) &= \log \left( \frac{\frac{\kappa(1-\alpha)\rho}{(1-\kappa)} + 1}{\frac{\kappa(1-\alpha)}{(1-\kappa)} + 1} \right) \\
&= \log \left( 1 + (\rho - 1) \frac{\kappa(1 - \alpha)}{1 - \kappa\alpha} \right) \\
&\stackrel{(a)}{=} (\rho - 1) \frac{\kappa(1 - \alpha)}{1 - \kappa\alpha} - \frac{(\rho - 1)^2}{2} \frac{\kappa^2(1 - \alpha)^2}{(1 - \kappa\alpha)^2} \\
&\quad + O(b^{-3/2}),
\end{aligned} \quad (38)$$

where (a) follows from (37), and Taylor's expansion around  $\rho = 1$ . Similarly,

$$f^2(-v) = (\rho - 1)^2 \frac{\kappa^2(1 - \alpha)^2}{(1 - \kappa\alpha)^2} + O(b^{-3/2}), \quad (39)$$



$$\begin{aligned}\log(\rho) &= \log(1 + (\rho - 1)) \\ &= \sqrt{\frac{\lambda_\alpha(1-\kappa)}{(1-\alpha)^2\kappa^2b}} - \frac{\lambda_\alpha(1-\kappa)}{2(1-\alpha)^2\kappa^2b} + O(b^{-3/2}),\end{aligned}\quad (40)$$

and

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

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

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

Now using (38) and (40) we obtain

$$\begin{aligned}\mathbb{E}\xi_0^1 &= -\kappa\sqrt{\frac{\lambda_\alpha b(1-\kappa)}{(1-\alpha)^2\kappa^2}} - v + \kappa\alpha\sqrt{\frac{\lambda_\alpha(1-\kappa)b}{(1-\alpha)^2\kappa^2}} \\ &\quad - \frac{\lambda_\alpha(1-\kappa)\alpha}{2(1-\alpha)^2\kappa} \\ &\quad + \sqrt{\frac{\lambda_\alpha(1-\kappa)b}{(1-\alpha)^2\kappa^2}}\kappa(1-\alpha) - \frac{\lambda_\alpha(1-\kappa)}{2(1-\kappa\alpha)} + O(b^{-1/2}) \\ &= -v - \frac{\lambda_\alpha(1-\kappa)}{2(1-\alpha)^2\kappa}\alpha - \frac{\lambda_\alpha(1-\kappa)}{2(1-\kappa\alpha)} + O(b^{-1/2}).\end{aligned}\quad (42)$$

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

$$\begin{aligned}\text{Var}\xi_0^1 &= \log^2(\rho)\kappa b\alpha + f^2(-v)(1-\kappa)b + f^2(-v)\kappa b(1-\alpha) \\ &\stackrel{(a)}{=} \frac{\lambda_\alpha\alpha(1-\kappa)}{(1-\alpha)^2\kappa} + \frac{(1-\kappa)\lambda_\alpha}{1-\kappa\alpha} + O(b^{-1/2}),\end{aligned}\quad (43)$$

where in (a) we used (41) and (39). Comparing (43) and (44), after letting  $b \rightarrow \infty$  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  $L_0 \sim \text{Poi}(b)$ , and  $\mathcal{L}(X_i) = \kappa\alpha\mathcal{L}(\log(\rho)) + (1-\kappa)\mathcal{L}(f(-v)) + (\kappa(1-\alpha))\mathcal{L}(f(-v))$ , keeping in mind the independent splitting property of Poissons, where  $\mathcal{L}$  denotes the law of a rv<sup>2</sup>. Next we calculate  $\mathbb{E}(|X_i|^3)$ . It is easy to show using (38) and (40) that

$$\mathbb{E}(|X_i|^3) = \kappa\alpha \log^3(b) + (1-\kappa\alpha)|f^3(-v)| = O(b^{-3/2}).\quad (44)$$

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

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

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

By Lemma 2, taking  $b \rightarrow \infty$  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 [35, Theorem 16.4] we obtain

$$\mathbb{E}f(\xi_1^t) = (\rho - 1)\mathbb{E}\left(\frac{e^{\xi_1^t}}{1 + e^{\xi_1^t}}\right) - \frac{(\rho - 1)^2}{2}\mathbb{E}\left(\frac{e^{2\xi_1^t}}{(1 + e^{\xi_1^t})^2}\right) + O(b^{-3/2}), \quad (46)$$

and by using Lemma 4 in addition we obtain

$$\mathbb{E}f(\xi_0^t) = (\rho - 1)\frac{\kappa(1 - \alpha)}{1 - \kappa}\mathbb{E}\left(\frac{1}{1 + e^{\xi_1^t}}\right) - \frac{(\rho - 1)^2\kappa(1 - \alpha)}{2(1 - \kappa)}\mathbb{E}\left(\frac{e^{\xi_1^t}}{(1 + e^{\xi_1^t})^2}\right) + O(b^{-3/2}). \quad (47)$$

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  $L \sim \text{Poi}$  and  $X_i$  are independent and identically distributed (iid) rv, we obtain

$$\mathbb{E}(\xi_0^{t+1}) = h + \log\left(\frac{p}{q}\right)\kappa b\alpha + \mathbb{E}(f(\xi_0^t))(1 - \kappa)b + \mathbb{E}(f(\xi_1^t))\kappa b(1 - \alpha) \quad (48)$$

and

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

We now substitute (47) and (46) in (48) to get:

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

which on simplifying and grouping like terms gives

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

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

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

Using (40) we get

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

Finally we obtain

$$\mathbb{E}(\xi_0^{t+1}) = -\log\left(\frac{1-\kappa}{\kappa(1-\alpha)}\right) - \frac{\lambda_\alpha\alpha(1-\kappa)}{2(1-\alpha)^2\kappa} - \lambda_\alpha \frac{(1-\kappa)}{2(1-\alpha)\kappa} \mathbb{E}\left(\frac{e^{\xi_1^t}}{1+e^{\xi_1^t}}\right) + O(b^{-1/2}). \quad (50)$$

Using exactly the same simplifications we can get

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

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

$$\mathbb{E}(f^2(\xi_0^t)) = (\rho-1)^2 \frac{\kappa(1-\alpha)}{1-\kappa} \mathbb{E}\left(\frac{e^{\xi_1^t}}{(1+e^{\xi_1^t})^2}\right) + O(b^{-3/2}). \quad (52)$$

Similarly,

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

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

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

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

Substituting (52) and (53) into the above equations and letting  $b \rightarrow \infty$ , we get

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

where

$$\mu^{(t+1)} = \frac{\lambda_\alpha\alpha(1-\kappa)}{\kappa(1-\alpha)^2} + \frac{\lambda_\alpha(1-\kappa)}{\kappa(1-\alpha)} \mathbb{E}\left(\frac{\exp \xi_1^t}{1+\exp(\xi_1^t)}\right). \quad (54)$$

Using  $\mu^{(t+1)}$  of (54) in (50) and (51) we get

$$\begin{aligned}\mathbb{E}(\xi_0^{t+1}) &= -\log\left(\frac{(1-\kappa)}{\kappa(1-\alpha)}\right) - \frac{1}{2}\mu^{(t+1)} + O(b^{-1/2}) \\ \mathbb{E}(\xi_1^{t+1}) &= -\log\left(\frac{(1-\kappa)}{\kappa(1-\alpha)}\right) + \frac{1}{2}\mu^{(t+1)} + O(b^{-1/2}).\end{aligned}\quad (55)$$

Now we use the fact the induction assumption that  $\xi_1^t \rightarrow \mathcal{N}(\mathbb{E}(\xi_1^t), \mu^{(t)})$ . Since the function  $e^{\xi_1^t}/(1+e^{\xi_1^t})$  is bounded, by Lebesgue's dominated convergence theorem [35, Theorem 16.4] this means  $\mathbb{E}(1/(1+e^{-\xi_1^t})) \rightarrow \mathbb{E}(1/(1+e^{-\mathcal{N}(\mathbb{E}(\xi_1^t), \mu^{(t)})}))$  as  $b \rightarrow \infty$ . 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

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

Substituting the above into (54) gives us the recursion for  $\mu^{(t+1)}$  given in (13).

Next we prove Gaussianity. Consider

$$\begin{aligned}\xi_0^{t+1} - \mathbb{E}(\xi_0^{t+1}) &= \log\left(\frac{p}{q}\right)(L_{0c} - \mathbb{E}(L_{0c})) + \sum_{i=1}^{L_{00}}(f(\xi_{0,i}^t) - \mathbb{E}(f(\xi_0^t))) + \\ &\quad \sum_{i=1}^{L_{01}}(f(\xi_{1,i}^t) - \mathbb{E}(f(\xi_1^t))) + (L_{00} - \mathbb{E}(L_{00}))\mathbb{E}(f(\xi_0^t)) + \\ &\quad (L_{01} - \mathbb{E}(L_{01}))\mathbb{E}(f(\xi_1^t)).\end{aligned}\quad (56)$$

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  $i \leq \mathbb{E}L_{00}$ . 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 (56) by their means, leading to

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

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  $b \rightarrow \infty$  and it is finite for a fixed  $t$ . Now since we have an infinite sum of independent rvs as  $a, b \rightarrow \infty$ , 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  $\xi_1^{t+1}$  is identical.  $\square$

## C Finishing the proof of Theorem 1

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

$$\begin{aligned} \lim_{b \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\mathbb{E}(|\overline{S}\Delta\widehat{S}_0|)}{K(1-\alpha)} &= \lim_{b \rightarrow \infty} \lim_{n \rightarrow \infty} \left( \frac{\mathbb{E}(\sum_{i=1}^n \mathbf{1}_{\sigma_i \neq \widehat{\sigma}_i})}{K - K\alpha} \right) \\ &\leq \lim_{b \rightarrow \infty} \left( \frac{(1-\kappa)}{\kappa(1-\alpha)} \mathbb{P}(\xi_0^t \geq 0) + \right. \\ &\quad \left. \mathbb{P}(\xi_1^t \leq 0) \right), \end{aligned} \quad (58)$$

since

$$\begin{aligned} \mathbb{E} \left( \sum_{i=1}^n \mathbf{1}_{\sigma_i \neq \widehat{\sigma}_i} \right) &= \\ &n(\mathbb{P}(c_i = 0, \sigma_i = 0) \mathbb{P}(R_i^t > v | c_i = 0, \sigma_i = 0) + \\ &\quad \mathbb{P}(c_i = 0, \sigma_i = 1) \mathbb{P}(R_i^t < v | c_i = 0, \sigma_i = 1)), \end{aligned} \quad (59)$$

and since  $R_i^t - R_{i \rightarrow u}^t = O(b^{-1/2})$ . Indeed, given the  $b \rightarrow \infty$  limit in (58), the bound  $O(b^{-1/2})$  allows us to replace  $R_i^t$  in (59) by the distribution limit when  $n \rightarrow \infty$ , which is  $\xi_0^t$  or  $\xi_1^t$  when conditioned on  $\{\sigma_i = 0\}$  or  $\{\sigma_i = 1\}$  respectively, for an arbitrary  $i$ . We now analyze each term in (58) separately. By Proposition 1 we have

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

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 (54),  $\mu^{(t)} \leq \frac{\lambda_\alpha (1-\kappa)}{\kappa(1-\alpha)^2}$ . 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

$$\begin{aligned} \lim_{b \rightarrow \infty} \mathbb{P}(\xi_1^t \leq 0) &\leq \frac{1}{2} e^{-\frac{1}{2\mu^{(t)}} \left( \frac{\mu^{(t)}}{2} - \log \left( \frac{1-\kappa}{\kappa(1-\alpha)} \right) \right)^2} \\ &= \frac{1}{2} e^{-\frac{\mu^{(t)}}{8} \left( 1 - \frac{2}{\mu^{(t)}} \log \left( \frac{1-\kappa}{\kappa(1-\alpha)} \right) \right)^2} \\ &\leq \frac{1}{2} e^{-\frac{\mu^{(t)}}{8}} e^{\frac{1}{2} \log \left( \frac{1-\kappa}{\kappa(1-\alpha)} \right)} \\ &= \frac{1}{2} \sqrt{\frac{1-\kappa}{\kappa(1-\alpha)}} e^{-\frac{\mu^{(t)}}{8}}, \end{aligned} \quad (60)$$

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

$$\lim_{b \rightarrow \infty} \left( \frac{(1-\kappa)}{\kappa(1-\alpha)} \right) \mathbb{P}(\xi_0^t \geq 0) \leq \frac{1}{2} \sqrt{\frac{1-\kappa}{\kappa(1-\alpha)}} e^{-\frac{\mu^{(t)}}{8}}. \quad (61)$$

Substituting (69) and (70) back in (58) and using the fact that  $\mu^{(t)} \geq \lambda_\alpha \alpha (1 - \kappa) / (\kappa (1 - \alpha)^2)$ , we get

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

Then using (22) we get the desired result in (21).  $\square$

## References

- [1] 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.
- [2] D. H. Chau, S. Pandit, and C. Faloutsos, "Detecting fraudulent personalities in networks of online auctioneers," in *PKDD*. Springer, 2006, pp. 103–114.
- [3] 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.
- [4] S. T. Smith, E. K. Kao, K. D. Senne, G. Bernstein, and S. Phillips, "Bayesian discovery of threat networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5324–5338, 2014.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] T. Miffin, 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.
- [9] K. Rohe, S. Chatterjee, and B. Yu, "Spectral clustering and the high-dimensional stochastic blockmodel," *Ann. Stat.*, pp. 1878–1915, 2011.
- [10] 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.
- [11] A. Montanari, "Finding one community in a sparse graph," *Journal of Statistical Physics*, vol. 161, no. 2, pp. 273–299, 2015.
- [12] 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.
- [13] 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.

- [14] 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.
- [15] 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.
- [16] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.
- [17] 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.
- [18] 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.
- [19] T. Miffin, 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.
- [20] B. A. Miller, N. T. Bliss, and P. P. J. Wolfe, “Subgraph detection using eigenvector L1 norms,” in *Adv. Neural Inf. Process. Syst. 23 24th Annu. Conf. Neural Inf. Process. Syst. 2010, NIPS 2010*, 2010, pp. 1–9.
- [21] A. Kadavankandy, L. Cottatellucci, and K. Avrachenkov, “Characterization of  $L^1$ -norm statistic for Anomaly Detection in Erdős Rényi Graphs,” in *CDC*. IEEE, 2016.
- [22] 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.
- [23] 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.
- [24] B. Hajek, Y. Wu, and J. Xu, “Information limits for recovering a hidden community,” in *Information Theory (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 1894–1898.
- [25] 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.
- [26] A. E. Allahverdyan, G. Ver Steeg, and A. Galstyan, “Community detection with and without prior information,” *EPL (Europhysics Letters)*, vol. 90, no. 1, p. 18002, 2010.
- [27] T. T. Cai, T. Liang, and A. Rakhlin, “Inference via message passing on partially labeled stochastic block models,” *arXiv preprint arXiv:1603.06923*, 2016.
- [28] 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.

- 
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf, “Learning with Local and Global Consistency,” *Adv. Neural Inform. Process. Syst.*, vol. 16, pp. 321–328, 2004.
- [34] 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.
- [35] P. Billingsley, *Probability and measure*. John Wiley & Sons, 2008.
- [36] N. Alon and J. H. Spencer, *The probabilistic method*. John Wiley & Sons, 2004.



## Supplementary Material

*Proof of Proposition 3.* We derive the conditional distributions of the messages  $R_{u \rightarrow i}^t$  for a finite  $t$  given  $\{\sigma_u = 0\}$  and given  $\{\sigma_u = 1\}$ . In this limit the tree coupling of Lemma 3 holds with a slightly modified construction of the tree to accommodate the difference in the generation of cued nodes. It is similar to the tree coupling in Lemma 3, with the only difference being the generation of cues. At any level of the tree, a node  $u$  is labelled a cue such that  $\mathbb{P}(c_u = 1 | \tau_u = 1) = \alpha\beta$  and  $\mathbb{P}(c_u = 1 | \tau_u = 0) = \kappa\alpha(1 - \beta)/(1 - \kappa)$ , so that the equalities in (2) and (4) hold, where  $c_u$  denotes the cue membership of node  $u$  on the tree. Let  $F_{u \rightarrow i}^t$  be such that  $R_{u \rightarrow i}^t = F_{u \rightarrow i}^t + h_u$ , for any two neighbouring nodes  $i$  and  $u$ . Then, it can be seen from (25) that  $F_{u \rightarrow i}^t$  satisfies the following recursion

$$F_{u \rightarrow i}^{t+1} = -\kappa(a - b) + \sum_{l \in \delta u, l \neq i} f_{\text{isi}}(F_l^t + h_l), \quad (62)$$

where  $f_{\text{isi}}(x) := \log\left(\frac{e^{(x-\nu)}\rho+1}{e^{(x-\nu)}+1}\right)$ . Let  $\Psi_0^t, \Psi_1^t$  be the rvs that have the conditional asymptotic distribution of  $F_{u \rightarrow i}^t$  given  $\{\sigma_u = 0\}$  and  $\{\sigma_u = 1\}$  respectively in the limit  $n \rightarrow \infty$ . Then, by studying the recursion (62) on the tree we can conclude that  $\Psi_0^t, \Psi_1^t$  satisfy the following recursive distributional equations

$$\begin{aligned} \Psi_0^{t+1} &\stackrel{\text{D}}{=} -\kappa(a - b) + \sum_{i=0}^{L_{01c}} f_{\text{isi}}(\Psi_{1i}^t + B_c) + \\ &\sum_{i=0}^{L_{01n}} f_{\text{isi}}(\Psi_{1i}^t + B_n) + \sum_{i=0}^{L_{00c}} f_{\text{isi}}(\Psi_{0i}^t + B_c) + \\ &\sum_{i=0}^{L_{00n}} f_{\text{isi}}(\Psi_{0i}^t + B_n), \end{aligned} \quad (63)$$

$$\begin{aligned} \Psi_1^{t+1} &\stackrel{\text{D}}{=} -\kappa(a - b) + \sum_{i=0}^{L_{11c}} f_{\text{isi}}(\Psi_{1i}^t + B_c) + \\ &\sum_{i=0}^{L_{11n}} f_{\text{isi}}(\Psi_{1i}^t + B_n) + \sum_{i=0}^{L_{10c}} f_{\text{isi}}(\Psi_{0i}^t + B_c) + \\ &\sum_{i=0}^{L_{10n}} f_{\text{isi}}(\Psi_{0i}^t + B_n), \end{aligned} \quad (64)$$

where  $\stackrel{\text{D}}{=}$  represents equality in distribution, and the random sums are such that  $L_{01c} \sim \text{Poi}(\kappa b \alpha \beta)$ ,  $L_{01n} \sim \text{Poi}(\kappa b (1 - \alpha \beta))$ ,  $L_{00c} \sim \text{Poi}(\kappa b \alpha (1 - \beta))$ ,  $L_{00n} \sim \text{Poi}(b(1 - \kappa - \kappa \alpha (1 - \beta)))$ ,  $L_{11c} \sim \text{Poi}(\kappa a \alpha \beta)$ ,  $L_{11n} \sim \text{Poi}(\kappa a (1 - \alpha \beta))$ ,  $L_{10c} \sim \text{Poi}(\kappa b \alpha (1 - \beta))$ , and  $L_{10n} \sim \text{Poi}(b(1 - \kappa - \kappa \alpha (1 - \beta)))$ ,  $B_c = \log\left(\frac{\beta(1 - \kappa)}{(1 - \beta)\kappa}\right)$ ;  $B_n = \log\left(\frac{(1 - \alpha \beta)(1 - \kappa)}{(1 - \kappa - \alpha \kappa + \alpha \kappa \beta)}\right)$ ; and  $\Psi_{0,i}^t$  and  $\Psi_{1,i}^t$  are iid rvs with the same distribution as  $\Psi_0^t$  and  $\Psi_1^t$  respectively.

We now derive the asymptotic distributions  $\Psi_0^{t+1}$  and  $\Psi_1^{t+1}$  when  $a, b \rightarrow \infty$  such that  $\lambda = \frac{\kappa^2(a-b)^2}{(1-\kappa)b}$  and  $\kappa$  are fixed. Observe that  $\rho = 1 + \sqrt{\frac{\lambda(1-\kappa)}{\kappa^2 b}} = 1 + \sqrt{\frac{r}{b}}$ , where  $r := \frac{\lambda(1-\kappa)}{\kappa^2}$ . Notice

that if  $P_0 \sim \mathcal{L}(\Psi_0^t)$  and  $P_1 = \mathcal{L}(\Psi_1^t)$ , we have, since  $\Psi = \log\left(\frac{dP_1}{dP_0}\right)$ , that  $\frac{dP_0}{dP_1}(\Psi) = e^{-\Psi}$ . Also

$$f_{\text{isi}}(x) = \log\left(1 + (\rho - 1)\frac{e^{x-\nu}}{1 + e^{x-\nu}}\right) \quad (65)$$

$$= \sqrt{\frac{r}{b}}\left(\frac{e^{x-\nu}}{1 + e^{x-\nu}}\right) - \quad (66)$$

$$\frac{1}{2}\frac{r}{b}\left(\frac{e^{x-\nu}}{1 + e^{x-\nu}}\right)^2 + O(b^{-3/2}), \quad (67)$$

and

$$f_{\text{isi}}^2(x) = \frac{r}{b}\left(\frac{e^{x-\nu}}{1 + e^{x-\nu}}\right)^2 + O(b^{-3/2}). \quad (68)$$

Now we can reformulate the recursions in (63) and (64) as a Poisson sum as follows:

$$\Psi_0^{t+1} \stackrel{D}{=} -\kappa(a - b) + \sum_{l=1}^{L_0} X_l \quad (69)$$

$$\Psi_1^{t+1} \stackrel{D}{=} -\kappa(a - b) + \sum_{l=1}^{L_1} Y_l, \quad (70)$$

where  $L_0 = \text{Poi}(b)$ ,  $L_1 = \text{Poi}(\kappa a + (1 - \kappa)b)$  and  $X_l$  and  $Y_l$  are mixture rvs with laws defined as follows:

$$\begin{aligned} \mathcal{L}(X_l) &= \alpha\kappa(1 - \beta)\mathcal{L}(f_{\text{isi}}(\Psi_0^t + B_c)) + (1 - \kappa)(1 - \alpha(1 - \beta)e^{-\nu})\mathcal{L}(f_{\text{isi}}(\Psi_0^t + B_n)) \\ &\quad + \alpha\kappa\beta\mathcal{L}(f_{\text{isi}}(\Psi_1^t + B_c)) + \kappa(1 - \alpha\beta)\mathcal{L}(f_{\text{isi}}(\Psi_1^t + B_n)), \end{aligned}$$

$$\begin{aligned} \mathcal{L}(Y_l) &= \frac{\alpha\kappa b(1 - \beta)}{\kappa a + (1 - \kappa)b}\mathcal{L}(f_{\text{isi}}(\Psi_0^t + B_c)) + \frac{(1 - \kappa)b(1 - \alpha(1 - \beta)e^{-\nu})}{\kappa a + (1 - \kappa)b}\mathcal{L}(f_{\text{isi}}(\Psi_0^t + B_n)) \\ &\quad + \frac{\kappa a \alpha \beta}{\kappa a + (1 - \kappa)b}\mathcal{L}(f_{\text{isi}}(\Psi_1^t + B_c)) + \frac{\kappa a(1 - \alpha\beta)}{\kappa a + (1 - \kappa)b}\mathcal{L}(f_{\text{isi}}(\Psi_1^t + B_n)). \end{aligned}$$

Observe that we have  $B_c - \nu = \log\left(\frac{\beta}{1 - \beta}\right)$  and  $B_n - \nu = \log\left(\frac{\kappa(1 - \alpha\beta)}{(1 - \kappa - \alpha\kappa(1 - \beta))}\right)$ . We can calculate  $\mathbb{E}(X_l)$  as

$$\begin{aligned} \mathbb{E}(X_l) &= \alpha\kappa\beta\sqrt{\frac{r}{b}} + \kappa(1 - \alpha\beta)\sqrt{\frac{r}{b}} - \alpha\kappa\beta\frac{r}{2b}\mathbb{E}\left(\frac{e^{\Psi_1^t + B_c - \nu}}{1 + e^{\Psi_1^t + B_c - \nu}}\right) \\ &\quad - \kappa(1 - \alpha\beta)\frac{r}{2b}\mathbb{E}\left(\frac{e^{\Psi_1^t + B_n - \nu}}{1 + e^{\Psi_1^t + B_n - \nu}}\right) + O(b^{-3/2}), \end{aligned}$$

which gives,

$$\mathbb{E}(X_l) = \kappa\sqrt{\frac{r}{b}} - \alpha\kappa\beta\frac{r}{2b}\mathbb{E}\left(\frac{e^{\Psi_1^t + B_c - \nu}}{1 + e^{\Psi_1^t + B_c - \nu}}\right) - \kappa(1 - \alpha\beta)\frac{r}{2b}\mathbb{E}\left(\frac{e^{\Psi_1^t + B_n - \nu}}{1 + e^{\Psi_1^t + B_n - \nu}}\right) + O(b^{-3/2}).$$

Similarly

$$\mathbb{E}(X_l^2) = \alpha\kappa\beta\frac{r}{b}\mathbb{E}\left(\frac{e^{\Psi_1^t + B_c - \nu}}{1 + e^{\Psi_1^t + B_c - \nu}}\right) + \frac{r\kappa(1 - \alpha\beta)}{b}\mathbb{E}\left(\frac{e^{\Psi_1^t + B_n - \nu}}{1 + e^{\Psi_1^t + B_n - \nu}}\right) + O(b^{-3/2}),$$

and

$$\begin{aligned}
& \mathbb{E}(|X_l^3|) \\
&= \alpha\kappa\beta\left(\frac{r}{b}\right)^{3/2}\mathbb{E}\left(\frac{e^{2(\Psi_1^t+B_c-\nu)}}{(1+e^{\Psi_1^t+B_c-\nu})^2}\right) + \\
& \quad \frac{\kappa(1-\alpha\beta)r^{3/2}}{b^{3/2}}\mathbb{E}\left(\frac{e^{2(\Psi_1^t+B_n-\nu)}}{(1+e^{\Psi_1^t+B_n-\nu})^2}\right) \\
& \quad + O(b^{-2}).
\end{aligned} \tag{71}$$

Similarly we can calculate the moments of  $Y_l$  as follows:

$$\begin{aligned}
\mathbb{E}(Y_l) &= \frac{\alpha\kappa b\beta}{\kappa a + (1-\kappa)b}\sqrt{\frac{r}{b}}\mathbb{E}\left(\frac{1+\rho e^{\Psi_1^t+B_c-\nu}}{1+e^{\Psi_1^t+B_c-\nu}}\right) \\
& \quad - \frac{r\alpha\kappa\beta}{2(\kappa a + (1-\kappa)b)}\mathbb{E}\left(\frac{e^{\Psi_1^t+B_c-\nu}(1+\rho e^{\Psi_1^t+B_c-\nu})}{(1+e^{\Psi_1^t+B_c-\nu})^2}\right) \\
& \quad + \frac{\kappa b(1-\alpha\beta)}{\kappa a + (1-\kappa)b}\sqrt{\frac{r}{b}}\mathbb{E}\left(\frac{1+\rho e^{\Psi_1^t+B_n-\nu}}{1+e^{\Psi_1^t+B_n-\nu}}\right) \\
& \quad - \frac{r\kappa(1-\alpha\beta)}{2(\kappa a + (1-\kappa)b)}\mathbb{E}\left(\frac{e^{\Psi_1^t+B_n-\nu}(1+\rho e^{\Psi_1^t+B_n-\nu})}{(1+e^{\Psi_1^t+B_n-\nu})^2}\right) + O(b^{-3/2}),
\end{aligned}$$

giving

$$\begin{aligned}
\mathbb{E}(Y_l) &= \kappa\sqrt{rb}\frac{1}{\kappa a + (1-\kappa)b} + \frac{r\alpha\beta\kappa}{2(\kappa a + (1-\kappa)b)}\mathbb{E}\left(\frac{e^{\Psi_1^t+B_c-\nu}}{1+e^{\Psi_1^t+B_c-\nu}}\right) \\
& \quad + \frac{r\kappa(1-\alpha\beta)}{2(\kappa a + (1-\kappa)b)}\mathbb{E}\left(\frac{e^{\Psi_1^t+B_n-\nu}}{1+e^{\Psi_1^t+B_n-\nu}}\right) + O(b^{-3/2}).
\end{aligned}$$

In addition,

$$\begin{aligned}
\mathbb{E}(Y_l^2) &= \frac{\alpha\kappa\beta r}{\kappa a + (1-\kappa)b}\mathbb{E}\left(\frac{e^{\Psi_1^t+B_c-\nu}}{1+e^{\Psi_1^t+B_c-\nu}}\right) + \frac{\kappa r(1-\alpha\beta)}{\kappa a + (1-\kappa)b}\mathbb{E}\left(\frac{e^{\Psi_1^t+B_n-\nu}}{1+e^{\Psi_1^t+B_n-\nu}}\right) \\
& \quad + O(b^{-2})
\end{aligned}$$

and

$$\begin{aligned}
& \mathbb{E}(|Y_l^3|) \\
&= \frac{\alpha\kappa\beta r^{3/2}}{(\kappa a + (1-\kappa)b)b^{1/2}}\mathbb{E}\left(\frac{e^{2\Psi_1^t+2B_c-2\nu}}{(1+e^{\Psi_1^t+B_c-\nu})^2}\right) \\
& \quad + \frac{\kappa(1-\alpha\beta)r^{3/2}}{(\kappa a + (1-\kappa)b)b^{1/2}}\mathbb{E}\left(\frac{e^{2\Psi_1^t+2B_n-2\nu}}{(1+e^{\Psi_1^t+B_n-\nu})^2}\right) \\
& \quad + O(b^{-2}).
\end{aligned} \tag{72}$$

Let us define  $\mu^{(t)}$  as

$$\mu^{(t+1)} = \alpha\beta\kappa r\mathbb{E}\left(\frac{1}{1+e^{-\Psi_1^t-B_c+\nu}}\right) + \kappa r(1-\alpha\beta)\mathbb{E}\left(\frac{1}{1+e^{-\Psi_1^t-B_n+\nu}}\right). \tag{73}$$

Finally we have

$$\begin{aligned}
\mathbb{E}(\Psi_0^{t+1}) &= -\kappa(a-b) + b\mathbb{E}(X_t) \\
&= -\frac{\alpha\kappa\beta r}{2}\mathbb{E}\left(\frac{e^{\Psi_1^t+B_c-\nu}}{1+e^{\Psi_1^t+B_c-\nu}}\right) \\
&\quad - \frac{\kappa(1-\alpha\beta)r}{2}\mathbb{E}\left(\frac{e^{\Psi_1^t+B_n-\nu}}{1+e^{\Psi_1^t+B_n-\nu}}\right) + O(b^{-1/2}) \\
&= -\frac{\mu^{t+1}}{2} + O(b^{-1/2}),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}(\Psi_1^{t+1}) &= -\kappa(a-b) + (\kappa a + (1-\kappa)b)\mathbb{E}(Y_t) \\
&= \frac{\alpha\kappa\beta r}{2}\mathbb{E}\left(\frac{e^{\Psi_1^t+B_c-\nu}}{1+e^{\Psi_1^t+B_c-\nu}}\right) \\
&\quad + \frac{\kappa(1-\alpha\beta)r}{2}\mathbb{E}\left(\frac{e^{\Psi_1^t+B_n-\nu}}{1+e^{\Psi_1^t+B_n-\nu}}\right) + O(b^{-1/2}) \\
&= \frac{\mu^{(t+1)}}{2} + O(b^{-1/2}).
\end{aligned}$$

In addition, for the variances of  $\Psi_0^{t+1}$  and  $\Psi_1^{t+1}$  we have

$$\begin{aligned}
\text{Var}(\Psi_0^{t+1}) &= b\mathbb{E}(X_t^2) \\
&= \alpha\beta\kappa r\mathbb{E}\left(\frac{e^{\Psi_1^t+B_c-\nu}}{1+e^{\Psi_1^t+B_c-\nu}}\right) \\
&\quad + \kappa(1-\alpha\beta)r\mathbb{E}\left(\frac{e^{\Psi_1^t+B_n-\nu}}{1+e^{\Psi_1^t+B_n-\nu}}\right) + O(b^{-1/2}) \\
&= \mu^{(t+1)} + O(b^{-1/2}), \tag{74}
\end{aligned}$$

and similarly

$$\text{Var}(\Psi_1^{t+1}) = (\kappa a + (1-\kappa)b)\mathbb{E}(Y_t^2) \tag{75}$$

$$= \mu^{(t+1)} + O(b^{-1/2}). \tag{76}$$

Now we need to show the Gaussianity of the messages  $\Psi_0^t$  and  $\Psi_1^t$ , which we show using Lemma 2. For (69) the upperbound in Lemma 2 becomes

$$\begin{aligned}
\frac{C_{BE}\mathbb{E}(|X_i|^3)}{\sqrt{\gamma(\mu^2 + \sigma^2)^3}} &= \frac{C_{BE}b\mathbb{E}(|X_i|^3)}{\sqrt{(b(\mu^2 + \sigma^2))^3}} \\
&= \frac{C_{BE}b\mathbb{E}(|X_i|^3)}{\text{Var}(\Psi_0^{t+1})^{3/2}} \tag{77}
\end{aligned}$$

Similarly for (70) we get

$$\frac{C_{BE}\mathbb{E}(|Y_i|^3)}{\sqrt{\gamma(\mu^2 + \sigma^2)^3}} = \frac{C_{BE}(\kappa a + (1-\kappa)b)\mathbb{E}(|Y_i|^3)}{\text{Var}(\Psi_1^{t+1})^{3/2}}. \tag{78}$$

In Lemma 6 stated and proved below, we show that  $\mu^{(t+1)} \geq \alpha\beta^2\lambda\frac{1-\kappa}{\kappa}$ . Therefore for any  $\kappa < 1/2$ , we have

$$\text{Var}(\Psi_0^{t+1}) = \text{Var}(\Psi_0^t) \geq \frac{\alpha\beta^2\lambda}{2} + O(b^{-1/2}) = \Theta(1),$$

under the assumptions of the proposition. In addition we have  $b\mathbb{E}(|X_i|^3) = O(b^{-1/2})$  and  $(\kappa a + (1-\kappa)b)\mathbb{E}(|Y_i|^3) = O(b^{-1/2})$  from (71) and (72). Thus the bounds given in (77) and (78) both tend to zero as  $b \rightarrow \infty$ .

Hence by Lemma 2, we obtain that  $\Psi_1^t \rightarrow \mathcal{N}(\frac{\mu^{(t)}}{2}, \mu^{(t)})$  and  $\Psi_0^t \rightarrow \mathcal{N}(-\frac{\mu^{(t)}}{2}, \mu^{(t)})$  as  $b \rightarrow \infty$ , where from (73),  $\mu^{(t)}$  satisfies the following recursion with initial condition  $\mu^{(0)} = 0$ :

$$\mu^{(t+1)} = \alpha\beta\lambda\mathbb{E}\left(\frac{(1-\kappa)}{\kappa + (1-\kappa)e^{-\sqrt{\mu^{(t)}}Z - \frac{\mu^{(t)}}{2}} - B_c}\right) + (1-\alpha\beta)\lambda\mathbb{E}\left(\frac{(1-\kappa)}{\kappa + (1-\kappa)e^{-\sqrt{\mu^{(t)}}Z - \frac{\mu^{(t)}}{2}} - B_n}\right). \quad (79)$$

Consequently, the distributions of the messages  $R_{u \rightarrow i}^t$  in the limit of  $n \rightarrow \infty$  converge to  $\Gamma_j^t + h_u$ , given  $\{\sigma_u = j\}$ , where  $\Gamma_1^t \sim \mathcal{N}(\frac{\mu^{(t)}}{2}, \mu^{(t)})$  and  $\Gamma_0^t \sim \mathcal{N}(-\frac{\mu^{(t)}}{2}, \mu^{(t)})$ , in the large degree limit where  $b \rightarrow \infty$ .  $\square$

### Proving the bound on $\mu^{(t)}$

Let  $F(\mu)$  be defined as

$$\begin{aligned} F(\mu) &= \alpha\beta^2\lambda\mathbb{E}\left(\frac{(1-\kappa)/\kappa}{\beta + (1-\beta)\exp(-\mu/2 - \sqrt{\mu}Z)}\right) + \\ &\quad (1-\alpha\beta)^2\lambda \\ &\quad \mathbb{E}\left(\frac{(1-\kappa)}{\kappa(1-\alpha\beta) + (1-\kappa - \alpha\kappa + \alpha\kappa\beta)e^{(-\mu/2 - \sqrt{\mu}Z)}}\right). \end{aligned}$$

Then  $\mu^{(t)}$  satisfies the recursion  $\mu^{(t+1)} = F(\mu^{(t)})$ , by substituting for  $B_c$  and  $B_n$  in (79). Below we show a lower bound on  $F(\mu)$ . For its proof we need the following Lemma from [36].

**Lemma 5.** [36, Theorem 6.2.1] *If  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  are two non-decreasing functions, then  $\mathbb{E}(fg) \geq \mathbb{E}(f)\mathbb{E}(g)$ .*

Now we state our result on  $F(\mu)$ .

**Lemma 6.** *For  $0 < \beta < 1$ ,*

$$F(\mu) \geq \alpha\beta^2\lambda\frac{1-\kappa}{\kappa}.$$

*Proof.* We show that

$$g_\beta = \mathbb{E}\left(\frac{1}{\beta + (1-\beta)\exp(-\mu/2 - \sqrt{\mu}Z)}\right)$$

is nonincreasing for  $0 \leq \beta \leq 1$  as shown below. Let  $X = \exp(-\sqrt{\mu}Z)$ . Then  $\frac{d}{d\beta}(g_\beta) = \mathbb{E}\left(\frac{\exp(-\mu/2)X-1}{(\beta+(1-\beta)e^{-\mu/2}X)^2}\right)$ . Now we show  $\frac{d}{d\beta}(g_\beta) < 0$  using Lemma 5. In Lemma 5, let  $f =$

$\exp(-\mu/2)X$  and  $g = \frac{-1}{(\beta+(1-\beta)e^{-\mu/2}X)^2}$ . Clearly these are non-decreasing in  $X$ . Therefore  $\mathbb{E}(fg) \geq \mathbb{E}(f)\mathbb{E}(g) = \mathbb{E}(g)$ , since  $\mathbb{E}(f) = \mathbb{E}e^{-\mu/2}e^{\sqrt{\mu}Z} = 1$ . Therefore we have

$$\mathbb{E}\left(\frac{-e^{-\mu/2}X}{(\beta+(1-\beta)e^{-\mu/2}X)^2}\right) \geq \mathbb{E}\left(\frac{-1}{(\beta+(1-\beta)e^{-\mu/2}X)^2}\right),$$

hence  $\frac{dg_\beta}{d\beta} < 0$  for all  $\beta$ . Therefore  $1 = g_\beta(1) \leq g_\beta(\beta)$  for  $\beta < 1$ . The result then follows by substituting this lower bound in the definition of  $F(\mu)$  and observing that the second term is strictly non-negative.  $\square$

## Proof of Theorem 2

*Proof.* Notice that when we set  $\beta = 1$  the recursion (28) becomes the same as (13). Also, when  $\beta = 0$  we can retrieve the recursion for standard BP without side-information, i.e., and from this it can be gleaned that the asymptotic error rate is zero only if  $\lambda > 1/e$ .

Let us now consider  $0 < \beta < 1$ . By Lemma 6, we have

$$\alpha\beta^2\lambda\frac{1-\kappa}{\kappa} \leq \mu^{(t)} \leq \lambda\frac{(1-\kappa)}{\kappa}.$$

Hence  $\mu^{(t)} = \Theta\left(\frac{1-\kappa}{\kappa}\right)$ . The asymptotic distributions of the messages are as follows:

$$\begin{aligned} \Gamma_{0,0}^t &\sim \mathcal{N}(-\mu^{(t)}/2, \mu^{(t)}) + \log\left(\frac{(1-\alpha\beta)(1-\kappa)}{(1-\kappa-\alpha\kappa+\alpha\kappa\beta)}\right) \\ \Gamma_{0,1}^t &\sim \mathcal{N}(-\mu^{(t)}/2, \mu^{(t)}) + \log\left(\frac{\beta(1-\kappa)}{\kappa(1-\beta)}\right) \\ \Gamma_{1,0}^t &\sim \mathcal{N}(\mu^{(t)}/2, \mu^{(t)}) + \log\left(\frac{(1-\alpha\beta)(1-\kappa)}{(1-\kappa-\alpha\kappa+\alpha\kappa\beta)}\right) \\ \Gamma_{1,1}^t &\sim \mathcal{N}(\mu^{(t)}/2, \mu^{(t)}) + \log\left(\frac{\beta(1-\kappa)}{\kappa(1-\beta)}\right), \end{aligned}$$

where  $\Gamma_{j,k}^t$  is the rv with the asymptotic distribution of the messages  $R_{u \rightarrow i}^t$  in the limit of  $n \rightarrow \infty$  and  $b \rightarrow \infty$ , given  $\{\sigma_u = j, c_u = k\}$ . We can now write the probability of error  $p_e^\beta$  of the per-node MAP detector  $\hat{S}_0$  as

$$\begin{aligned} p_e^\beta &= p_e^\beta(i|\sigma_i = 0, c_i = 0)P(\sigma_i = 0, c_i = 0) + \\ & p_e^\beta(i|\sigma_i = 0, c_i = 1)P(\sigma_i = 0, c_i = 1) \\ & + p_e^\beta(i|\sigma_i = 1, c_i = 0)P(\sigma_i = 1, c_i = 0) \\ & + p_e^\beta(i|\sigma_i = 1, c_i = 1)P(\sigma_i = 1, c_i = 1) \\ & = P_{0,0}(R_i^t > \nu)\pi_{0,0} + P_{0,1}(R_i^t > \nu)\pi_{0,1} \\ & + P_{1,0}(R_i^t < \nu)\pi_{1,0} + P_{1,1}(R_i^t < \nu)\pi_{1,1}, \end{aligned}$$

is the error rate of Algorithm 2, where  $p_e^\beta(i|\sigma_i = 0, c_i = 0)$  denotes the probability that node  $i$  is misclassified, given  $\{\sigma_i = 0, c_i = 0\}$  and  $\pi_{0,1} = \mathbb{P}(\sigma_i = 0, c_i = 1)$  etc. Then the expected fraction

of mislabelled nodes  $\frac{\mathbb{E}(|\hat{S}_0 \Delta S|)}{K}$  in the limit  $n \rightarrow \infty, b \rightarrow \infty$  is

$$\begin{aligned}
& \lim_{b \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{np_e^\beta}{K} \\
&= Q \left( \frac{\frac{\mu^{(t)}}{2} + \log\left(\frac{\beta}{1-\beta}\right)}{\sqrt{\mu^{(t)}}} \right) \alpha\beta + \\
& \quad (1 - \alpha\beta) \\
& \quad Q \left( \frac{\frac{\mu^{(t)}}{2} - \log\left(\frac{1-\kappa}{\kappa} \left(\frac{1-\frac{\alpha\kappa(1-\beta)}{1-\kappa}}{1-\alpha\beta}\right)\right)}{\sqrt{\mu^{(t)}}} \right) + \\
& \quad \alpha(1 - \beta)Q \left( \frac{\frac{\mu^{(t)}}{2} - \log\left(\frac{\beta}{1-\beta}\right)}{\sqrt{\mu^{(t)}}} \right) + \\
& \quad \left(\frac{1-\kappa}{\kappa} - \alpha(1 - \beta)\right)Q \left( \frac{\frac{\mu^{(t)}}{2} - \log\left(\frac{(1-\alpha\beta)\kappa}{(1-\kappa-\alpha\kappa+\alpha\kappa\beta)}\right)}{\sqrt{\mu^{(t)}}} \right).
\end{aligned}$$

We can show, by a calculation similar to the one followed in the proof of Theorem 1, that

$$\begin{aligned}
& \lim_{b \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{np_e^\beta}{K} \\
& \leq \left( \alpha\sqrt{\beta(1-\beta)} + \right. \\
& \quad \left. \sqrt{(1-\alpha\beta)\left(\frac{1-\kappa}{\kappa} - \alpha(1-\beta)\right)} \right) e^{-\frac{\lambda\alpha\beta^2(1-\kappa)}{8\kappa}}.
\end{aligned}$$

Finally by a similar calculation to (22),

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

□

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Problem Motivation . . . . .	3
1.2	Previous works . . . . .	3
1.3	Summary of Results . . . . .	5
1.4	Organization . . . . .	5
1.5	Notation and Nomenclature . . . . .	5

---

<b>2</b>	<b>Model and Problem Definition</b>	<b>6</b>
<b>3</b>	<b>Belief Propagation Algorithm for Detection with Perfect Side-information</b>	<b>7</b>
<b>4</b>	<b>Asymptotic Error Analysis</b>	<b>8</b>
4.1	Detection Performance . . . . .	10
<b>5</b>	<b>Imperfect Side Information</b>	<b>12</b>
<b>6</b>	<b>Numerical Experiments</b>	<b>14</b>
6.1	Synthetic dataset . . . . .	14
6.2	Real-world datasets . . . . .	16
6.2.1	USPS dataset . . . . .	16
6.2.2	Reuters911 Dataset . . . . .	16
<b>7</b>	<b>Conclusions and Future Extensions</b>	<b>16</b>
<b>A</b>	<b>Description of G-W tree and derivation of Algorithm 1</b>	<b>17</b>
<b>B</b>	<b>Proof of Proposition 1</b>	<b>20</b>
<b>C</b>	<b>Finishing the proof of Theorem 1</b>	<b>26</b>





**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](http://inria.fr)

ISSN 0249-6399