# Density Evolution in the Degree-correlated Stochastic Block Model

**Elchanan Mossel**                                                    MOSSEL@WHARTON.UPENN.EDU
*Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA*

**Jiaming Xu**                                                          JIAMINGXU@BERKELEY.EDU
*Simons Institute for the Theory of Computing, University of California, Berkeley, Berkeley, CA*

## Abstract

There is a recent surge of interest in identifying the sharp recovery thresholds for cluster recovery under the stochastic block model. In this paper, we address the more refined question of how many vertices that will be misclassified on average. We consider the binary form of the stochastic block model, where $n$ vertices are partitioned into two clusters with edge probability $a/n$ within the first cluster, $c/n$ within the second cluster, and $b/n$ across clusters. Suppose that as $n \to \infty$, $a = b + \mu\sqrt{b}$, $c = b + \nu\sqrt{b}$ for two fixed constants $\mu, \nu$, and $b \to \infty$ with $b = n^{o(1)}$. When the cluster sizes are balanced and $\mu \neq \nu$, we show that the minimum fraction of misclassified vertices on average is given by $Q(\sqrt{v^*})$, where $Q(x)$ is the Q-function for standard normal, $v^*$ is the unique fixed point of $v = \frac{(\mu-\nu)^2}{16} + \frac{(\mu+\nu)^2}{16}\mathbb{E}[\tanh(v + \sqrt{v}Z)]$, and $Z$ is standard normal. Moreover, the minimum misclassified fraction on average is attained by a local algorithm, namely belief propagation, in time linear in the number of edges. Our proof techniques are based on connecting the cluster recovery problem to tree reconstruction problems, and analyzing the density evolution of belief propagation on trees with Gaussian approximations.

**Keywords:** Belief propagation, Density evolution, Community detection

## 1. Introduction

The problem of cluster recovery under the stochastic block model has been intensely studied in statistics Holland et al. (1983); Snijders and Nowicki (1997); Bickel and Chen (2009); Cai and Li (2014); Zhang and Zhou (2015); Gao et al. (2015), computer science (where it is known as the planted partition problem) Dyer and Frieze (1989); Jerrum and Sorkin (1998); Condon and Karp (2001); McSherry (2001); Coja-Oghlan (2005, 2010); Chen et al. (2014); Anandkumar et al. (2014); Chen and Xu (2014), and theoretical statistical physics Decelle et al. (2011a); Zhang et al. (2012); Decelle et al. (2011b). In the simplest binary form, the stochastic block model assumes that $n$ vertices are partitioned into two clusters with edge probability $a/n$ within the first cluster, $c/n$ within the second cluster, and $b/n$ across the two clusters. The goal is to reconstruct the underlying clusters from the observation of the graph. Different reconstruction goals can be considered depending on how the model parameters $a, b, c$ scale with $n$ (See Abbe and Sandon (2015) for more discussions):

- *Exact recovery (strong consistency).* If the average degree is $\Omega(\log n)$, it is possible to exactly recover the clusters (up to a permutation of cluster indices) with high probability. In the case with two equal-sized clusters, and $a = c = \alpha \log n/n$ and $b = \beta \log n/n$ for two fixed $\alpha, \beta > 0$, a sharp exact recovery threshold $(\sqrt{\alpha} - \sqrt{\beta})^2 \geq 2$ has been found in Mossel et al. (2015b); Abbe et al. (2014) and it is further shown that semi-definite programming can achieve the sharp threshold in Hajek et al. (2014); Bandeira (2015). The threshold for

two unequal-sized clusters is proved in Hajek et al. (2015a). Exact recovery threshold with a fixed number of clusters has been identified in Hajek et al. (2015a); Yun and Proutiere (2014a); Agarwal et al. (2015), and more generally in Abbe and Sandon (2015); Perry and Wein (2015) with heterogeneous cluster sizes and edge probabilities.

- *Weak recovery (weak consistency)*. If the average degree is $\Omega(1)$, one can hope for misclassifying only $o(n)$ vertices with high probability, which is known as weak recovery or weak consistency. In the setting with two approximately equal-sized clusters and $a = c$, it is shown in Yun and Proutiere (2014b); Mossel et al. (2015b) that weak recovery is possible if and only $(a - b)^2/(a + b) \to \infty$.

- *Correlated recovery (non-trivial detection)*. If the average degree is $\Theta(1)$, exact recovery or weak recovery becomes hopeless as the resulting graph under the stochastic block model will have at least a constant fraction of isolated vertices. Moreover, it is easy to see that even vertices with constant degree cannot be labeled accurately given all the other vertices' labels are revealed. Thus one goal in the sparse graph regime is to find a partition positively correlated with the true one (up to a permutation of cluster indices), which is also called non-trivial detection. In the setting with two approximately equal-sized clusters and $a = c$, it was first conjectured in Decelle et al. (2011a) and later proven in Mossel et al. (2015a, 2013b); Massoulié (2014) that correlated recovery is feasible if and only if $(a - b)^2 > 2(a + b)$. A spectral method based on the non-backtracking matrix is recently shown to achieve the sharp threshold in Bordenave et al. (2015).

In practice, one may be interested in the finer question of how many vertices that will be misclassified on expectation or with high probability. In the two equal-sized clusters setting, previous results on exact recovery, weak recovery, and correlated recovery provide conditions under which the minimum fraction of misclassified vertices on average is $o(1/n)$, $o(1)$, and strictly smaller than $1/2$, respectively. By assuming $(a - b)^2/(a + b) \to \infty$, recent work Zhang and Zhou (2015); Gao et al. (2015) showd that the expected misclassified fraction decays to zero exponentially fast and gives a sharp characterization of the decay exponent under a minimax framework. However, all these previous results do not shed light on the important question of when it is possible to misclassify only $\epsilon$ fraction of vertices on expectation, for any finite $\epsilon \in (0, 1/2)$. To the best of our knowledge, it is an open problem to find a closed-form expression of the expected misclassified fraction in terms of the model parameters. In this paper, we give such a simple formula in the special case of two approximately equal-sized clusters. Specifically, suppose that

$$a = b + \sqrt{b}\mu, \qquad c = b + \sqrt{b}\nu, \qquad b \to \infty, \qquad b = n^{o(1)}, \tag{1}$$

for two fixed constants $\mu, \nu$. We further assume that $\mu \neq \nu$ so that the vertex degrees are statistically correlated with the cluster structure, and hence the name of the *degree-correlated stochastic block model*. We show that the minimum fraction of misclassified vertices on average is given by $Q(\sqrt{v^*})$, where $Q(x)$ is the Q-function for standard normal, $v^*$ is the unique fixed point of $v = \frac{(\mu-\nu)^2}{16} + \frac{(\mu+\nu)^2}{16}\mathbb{E}\left[\tanh(v + \sqrt{v}Z)\right]$, and $Z$ is standard normal. Moreover, the minimum expected misclassified fraction can be attained by a local algorithm, namely belief propagation (BP) algorithm (See Algorithm 1), in time $O(nb^2)$. The local belief propagation algorithm can be viewed as an iterative algorithm which improves on the misclassified fraction on average step by step; running belief propagation for one iteration reduces to the simple thresholding algorithm

2

based on vertex degrees. It is crucial to assume $\mu \neq \nu$ for the above results to hold, otherwise it is well-known (see e.g. Kanade et al. (2014)) that no local algorithm can even achieve the non-trivial detection. Nevertheless, under a slightly stronger assumption that $b \to \infty$ and $b = o(\log n)$, we show that if $\mu = \nu$ with $|\mu| > 2$, local belief propagation combined with a global algorithm capable of non-trivial detection when $|\mu| > 2$, attains the minimum expected misclassified fraction $Q(\sqrt{\overline{v}})$ in polynomial-time, where $\overline{v}$ is the largest fixed point of $v = \frac{\mu^2}{4}\mathbb{E}\left[\tanh(v + \sqrt{v}Z)\right].$

When the clusters sizes are unbalanced, i.e., one cluster is of size approximately $\rho n$ for $\rho \in (0, 1/2)$, we give a lower bound on the minimum expected misclassified fraction, and an upper bound attained by the local belief propagation algorithm. However, we are unable to prove that the upper bound matches the lower bound. In fact, numerical experiments suggest that there exists a gap between the upper and lower bound when the cluster sizes are very unbalanced, i.e., $\rho$ is close to $0$.

Our proofs are mainly based on two useful techniques introduced in previous work. First, in the regime (1), the observed graph is locally tree-like, so we connect the cluster recovery problem to reconstruction problems on trees, and for the tree problems, the optimal estimator can be computed by belief propagation algorithm. Such connection has been investigated before in Mossel et al. (2015a, 2013a); Mossel and Xu (2015). Second, we characterize the density evolution of belief propagation on trees with Gaussian approximations, and as a result, we get a recursion with the largest fixed point corresponding to a lower bound on the minimum expected misclassified fraction, and the smallest fixed point correspond to the expected misclassified fraction attained by the local belief propagation algorithm. Density evolution has been widely used for the analysis of multiuser detection Montanari and Tse (2006) and sparse graph codes Richardson and Urbanke (2008); Mezard and Montanari (2009), and more recently has been introduced for the analysis of finding a single community in a sparse graph Montanari (2015). As a final piece, we prove that in the balanced cluster case, the recursion has a unique fixed point using the ideas of symmetric random variables Richardson and Urbanke (2008); Montanari (2005) and the first-order stochastic dominance, thus establishing the tightness of the lower bound and the optimality of the local BP simultaneously.

We point out that local algorithm by itself is a thriving research area (see Lyons and Nazarov (2011); Hatami et al. (2012); Gamarnik and Sudan (2014) and the references therein). Intuitively, local algorithms are one type of algorithms that make decision for each vertex just based on the neighborhood of small radius around the vertex; these algorithms are by design easy to run in a distributed fashion. Under the context of community detection, local algorithms determine which community each vertex lies in just based on the local neighborhood around each vertex (see Montanari (2015) for a formal definition). Recent work Mossel and Xu (2015) shows that with the aid of extra noisy label information on cluster structure, the local algorithms can be optimal in minimizing the expected misclassified fraction in the stochastic block model. In comparison, we show that when the vertex degrees are correlated with the cluster structure, the local algorithms can be optimal even without the extra noisy label information.

In closing, we compare our results with the recent results in Montanari (2015); Hajek et al. (2015b), which studied the problem of finding a single community of size $\rho n$ in a sparse graph. When $\nu = 0$, i.e. $b = c$, the stochastic block model considered in this paper, specializes to the single community model studied in Montanari (2015), and the recursion of density evolution derived in this paper reduces to the recursion derived in (Montanari, 2015, Eq. (36)). It is shown in Montanari (2015); Hajek et al. (2015b) that the local algorithm is strictly suboptimal comparing to the global exhaustive search when $\rho \to 0$. In contrast, we show that if $\rho = 1/2$, the local algorithm is optimal

in minimizing the expected fraction of misclassified vertices as long as $\mu \neq \nu$, and give a sharp characterization of the minimum expected misclassified fraction.

**Parallel Independent Work** Zhang et al. (2015) independently studied the problem of cluster recovery under the degree-correlated stochastic block model with multiple clusters. Based on the cavity method and numerical simulations, it is shown that with at most four clusters of unequal sizes but same in and out degrees, the non-trivial detection threshold phenomenon disappears, making the minimum fraction of misclassified vertices on average a continuous function of model parameters. In comparison, in the regime (1) with two equal-sized clusters and $\mu \neq \nu$, we give a more precise answer, showing that the fraction of misclassified vertices on average is $Q(\sqrt{v^*})$, where $v^*$ is the unique fixed point of $v = \frac{(\mu-\nu)^2}{16} + \frac{(\mu+\nu)^2}{16} \mathbb{E}\left[\tanh(v + \sqrt{v}Z)\right]$. Moreover, it is shown in Zhang et al. (2015) that with more than four clusters of unequal sizes, there exists a regime where two stable fixed points coexist, with the smaller one corresponding to the performance of local belief propagation, and the larger one corresponding to the performance of belief propagation initialized based on the true cluster structure. We find that the same phenomenon also happens in the case of two clusters with very unbalanced sizes and different in and out degrees (See Section 2.4 for details).

We recently became aware of the work Deshpande et al. (2015), who studied the problem of cluster recovery under the stochastic block model in the symmetric setting with two equal-sized clusters and $a = c$. By assuming that $\frac{(a-b)^2}{2(a+b)(1-(a+b)/2n)} \to \mu$ for a fixed constant $\mu$ and $(a + b)(1 - (a + b)/2n) \to \infty$, a sharp characterization of the per-vertex mutual information between the vertex labels and the graph is given in terms of $\mu$ and $\overline{v}$, where $\overline{v}$ is the largest fixed point of $v = \frac{\mu^2}{4}\mathbb{E}\left[\tanh(v + \sqrt{v}Z)\right]$. In comparison, we show that the minimum fraction of vertices misclassified on expectation is given by $Q(\sqrt{\overline{v}})$ and it is attainable in polynomial time with an additional technical assumption that $b = o(\log n)$. Interestingly, the point (a) of Lemma 6.1 in Deshpande et al. (2015) is a special case of Lemma 14 with $\rho = 1/2$ in our paper. The proof of Lemma 6.1 given in Deshpande et al. (2015) and the proof of Lemma 14 given in this paper are similar: both used the ideas of symmetric random variables Richardson and Urbanke (2008); Montanari (2005). One slight difference is that to prove the concavity of the mapping in the recursion when $\rho = 1/2$, we used the first-order stochastic dominance, while Deshpande et al. (2015) computes the second-order derivative.

## 2. Model and Main Results

We consider the binary stochastic block model with $n$ vertices partitioned into two clusters, where each vertex is independently assigned into the first cluster with probability $\rho \in (0, 1)$ and the second cluster with probability $\bar{\rho} \triangleq 1 - \rho$.[1] Each pair of vertices is connected independently with probability $a/n$ if two vertices are in the first cluster, with probability $c/n$ if they are in the second cluster, and with probability $b/n$ if they are in two different clusters. Let $G = (V, E)$ denote the observed graph and $A$ denote the adjacency matrix of the graph $G$. Let $\sigma$ denote the underlying vertex labeling such that $\sigma_i = +$ if vertex $i$ is in the first cluster and $\sigma_i = -$ otherwise. The model parameters $\{\rho, a, b, c\}$

---

1. Notice that the cluster sizes are random and concentrate on $\rho n$ and $(1 - \rho)n$. A slightly different model assumes that the vertices are partitioned into two clusters of deterministic sizes, exactly given by $\rho n$ and $(1 - \rho)n$. The two models behave similarly, but for ease of analysis, we focus on the random cluster size model in this paper.

are assumed to be known, and the goal is to estimate the vertex labeling $\sigma$ from the observation of $G$. More precisely, we have the following definition.

**Definition 1** *The* reconstruction problem on the graph *is the problem of inferring $\sigma$ from the observation of $G$. The expected fraction of vertices misclassified by an estimator $\widehat{\sigma}$ is given by*

$$p_G(\widehat{\sigma}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}\left\{\sigma_i \neq \widehat{\sigma}_i\right\}. \tag{2}$$

*Let $p_G^*$ denote the minimum expected misclassified fraction among all possible estimators based on $G$.*

The optimal estimator in minimizing the error probability $\mathbb{P}\left\{\sigma_i \neq \widehat{\sigma}_i\right\}$ is the maximum a posterior (MAP) estimator, which is given by $2 \times \mathbf{1}_{\{\mathbb{P}\{\sigma_i=+|G\} \geq \mathbb{P}\{\sigma_i=-|G\}\}} - 1$, and the minimum error probability is given by $\frac{1}{2} - \frac{1}{2}\mathbb{E}\left[|\mathbb{P}\left\{\sigma_i = +|G\right\} - \mathbb{P}\left\{\sigma_i = -|G\right\}|\right]$. Hence, the minimum expected misclassified fraction $p_G^*$ is given by

$$p_G^* = \frac{1}{2} - \frac{1}{2n} \sum_{i=1}^{n} \mathbb{E}\left[|\mathbb{P}\left\{\sigma_i = +|G\right\} - \mathbb{P}\left\{\sigma_i = -|G\right\}|\right]$$
$$= \frac{1}{2} - \frac{1}{2}\mathbb{E}\left[|\mathbb{P}\left\{\sigma_i = +|G\right\} - \mathbb{P}\left\{\sigma_i = -|G\right\}|\right], \tag{3}$$

where the second equality holds due to the symmetry among vertices. In the special case with $\rho = 1/2$ and $a = c$, the two clusters are symmetric; thus $p_G^* = 1/2$ and one can only hope to estimate $\sigma$ up to a global flip of sign. In general, computing the MAP estimator is computationally intractable and it is unclear whether the minimum expected misclassified fraction $p_G^*$ can be achieved by some estimator computable in polynomial-time.

Throughout this paper, we assume that $\rho$ is fixed and focus on the regime (1). As the average degree is $n^{o(1)}$, it is well-known that a local neighborhood of a vertex is a tree with high probability. Thus, it is natural to study the local algorithms. More precisely, we consider a local belief propagation algorithm to approximate the MAP estimator in the next subsection.

### 2.1. Local Belief Propagation Algorithm

Our local belief propagation algorithm is given in Algorithm 1. Specifically, let $\partial i$ denote the set of neighbors of $i$, and $F(x) = \frac{1}{2}\log\left(\frac{e^{2x}\rho a + \bar{\rho}b}{e^{2x}\rho b + \bar{\rho}c}\right)$. Let $d_+ = \rho a + \bar{\rho}b$ and $d_- = \rho b + \bar{\rho}c$ denote the expected vertex degree in the first and second cluster, respectively. Define the message transmitted from vertex $i$ to vertex $j$ at $t$-th iteration as

$$R_{i \to j}^t = \frac{-d_+ + d_-}{2} + \sum_{\ell \in \partial i \setminus \{j\}} F(R_{\ell \to i}^{t-1}), \tag{4}$$

with initial conditions $R_{i \to j}^0 = 0$ for all $i \in [n]$ and $j \in \partial i$. Then we define the belief of vertex $u$ at $t$-th iteration $R_u^t$ to be

$$R_u^t = \frac{-d_+ + d_-}{2} + \sum_{\ell \in \partial u} F(R_{\ell \to u}^{t-1}). \tag{5}$$

---

**Algorithm 1** Belief propagation for cluster recovery

---

1: Input: $n \in \mathbb{N}$, $\rho \in (0, 1)$, $a/b, c/b$, adjacency matrix $A \in \{0, 1\}^{n \times n}$, and $t \in \mathbb{N}$.
2: Initialize: Set $R_{i \to j}^0 = 0$ for all $i \in [n]$ and $j \in \partial i$.
3: Run $t - 1$ iterations of message passing as in (4) to compute $R_{i \to j}^{t-1}$ for all $i \in [n]$ and $j \in \partial i$.
4: Compute $R_i^t$ for all $i \in [n]$ as per (5).
5: Return $\widehat{\sigma}_{\mathrm{BP}}^t$ with $\widehat{\sigma}_{\mathrm{BP}}^t(i) = 2 \times \mathbf{1}_{\left\{R_i^t \geq -\varphi\right\}} - 1$, where $\varphi = \frac{1}{2} \log \frac{\rho}{1 - \rho}$.

---

As we will show in Section 3.1, the message passing as in (4) and (5) exactly computes the log likelihood ratio for a problem of inferring $\sigma_u$ on a suitably defined tree model with root $u$. Moreover, in the regime (1), there exists a coupling such that the local neighborhood of a fixed vertex $u$ is the same as the tree model rooted at $u$ with high probability. These two observations together suggest that $R_u^t$ is a good approximation of $\frac{1}{2} \log \frac{\mathbb{P}\{G|\sigma_u=+\}}{\mathbb{P}\{G|\sigma_u=-\}}$, and thus we can estimate $\sigma_u$ by truncating $R_u^t$ at the optimal threshold $-\varphi = \frac{1}{2} \log \frac{1-\rho}{\rho}$, according to the MAP rule.

We can see from (4) that in each BP iterations, each vertex $i$ needs to compute $|\partial i|$ outgoing messages. To this end, $i$ can first compute $R_i^t$ according to (5), and then subtract $F(R_{j \to i}^{t-1})$ from $R_i^t$ to get $R_{i \to j}^t$ for every neighbor $j$ of $i$. In this way, each BP iteration runs in time $O(m)$, where $m$ is the total number of edges. Hence $\widehat{\sigma}_{\mathrm{BP}}^t$ can be computed in time $O(tm)$.

Finally, notice that Algorithm 1 needs to know the parameters $\{\rho, a/b, c/b\}$. For the main results of this paper continue to hold, the values of the parameters are only needed to know up to $o(1)$ additive errors. In fact, there exists a fully data-driven procedure to consistently estimate those parameters, see e.g., Hajek et al. (2015a)[Appendix B].

## 2.2. Main Results

The following theorem characterizes the expected fraction of vertices misclassified by $\widehat{\sigma}_{\mathrm{BP}}^t$ as $n \to \infty$; it also gives a lower bound on the minimum expected misclassified fraction as $n \to \infty$. Furthermore, in the case $\rho = 1/2$ and $\mu \neq \nu$, $\widehat{\sigma}_{\mathrm{BP}}^t$ achieves the lower bound as $t \to \infty$ after $n \to \infty$.

**Theorem 2** *Assume $\rho \in (0, 1)$ is fixed and consider the regime (1). Let*

$$h(v) = \mathbb{E}\left[\tanh(v + \sqrt{v}Z + \varphi)\right],$$

*where $Z \sim \mathcal{N}(0, 1)$ and $\varphi = \frac{1}{2} \log \frac{\rho}{1-\rho}$. Let $\lambda = \frac{\rho(\mu+\nu)^2}{8}$ and $\theta = \frac{\rho(\mu-\nu)^2}{8} + \frac{(1-2\rho)\nu^2}{4}$. Define $\underline{v}$ and $\overline{v}$ to be the smallest and largest fixed point of*

$$v = \theta + \lambda h(v),$$

*respectively[2]. Define $(v_t : t \geq 0)$ recursively by $v_0 = 0$ and $v_{t+1} = \theta + \lambda h(v_t)$. Let $\widehat{\sigma}_{\mathrm{BP}}^t$ denote the estimator given by Belief Propagation applied for $t$ iterations, as defined in Algorithm 1. Then*

---

2. The existence of fixed points of $v \mapsto \theta + \lambda h(v)$ follows from Brouwer's fixed-point theorem and the fact that $h(v) \leq 1$.

$\lim_{t\to\infty} v_t = \underline{v}$, $(\rho\mu - \bar{\rho}\nu)^2/4 \le \underline{v} \le \overline{v} \le (\rho\mu^2 + \bar{\rho}\nu^2)/4$, *and*

$$\lim_{n\to\infty} p_G(\widehat{\sigma}_{\mathrm{BP}}^t) = \rho Q\left(\frac{v_t + \varphi}{\sqrt{v_t}}\right) + (1-\rho)Q\left(\frac{v_t - \varphi}{\sqrt{v_t}}\right),$$

$$\liminf_{n\to\infty} p_G^* \ge \rho Q\left(\frac{\overline{v} + \varphi}{\sqrt{\overline{v}}}\right) + (1-\rho)Q\left(\frac{\overline{v} - \varphi}{\sqrt{\overline{v}}}\right),$$

*where* $Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y^2/2}\mathrm{d}y$. *Moreover, if* $\rho = 1/2$ *and* $\mu \ne \nu$, *then* $\underline{v} = \overline{v} = v^*$, *and thus*

$$\lim_{t\to\infty}\lim_{n\to\infty} p_G(\widehat{\sigma}_{\mathrm{BP}}^t) = \lim_{n\to\infty} p_G^* = Q(\sqrt{v^*}),$$

*where* $v^*$ *is the unique fixed point of* $v = \frac{(\mu-\nu)^2}{16} + \frac{(\mu+\nu)^2}{16}\mathbb{E}\left[\tanh(v + \sqrt{v}Z)\right]$.

If $\rho\mu \ne \bar{\rho}\nu$ so that the vertex degrees are statistically correlated with the cluster structure, we have $\underline{v} > 0$ and thus $\lim_{t\to\infty}\lim_{n\to\infty} p_G(\widehat{\sigma}_{\mathrm{BP}}^t) \ge \min\{\rho, 1-\rho\}$. Hence, the local application of BP strictly outperforms the trivial estimator, which always guesses the label of all vertices to be $+1$ if $\rho \ge 1/2$ and $-1$ if $\rho < 1/2$. In the balanced case $\rho = 1/2$, the local BP achieves the minimum expected misclassified fraction. Numerical experiments further indicate that the local BP is still optimal in the unbalanced case provided that $\rho$ is not close to 0 or 1; however, we do not have a proof (See Section 2.4 for more discussions).

If $\rho\mu = \bar{\rho}\nu$, then $\underline{v} = 0$ and thus

$$\rho Q\left(\frac{\underline{v} + \varphi}{\sqrt{\underline{v}}}\right) + (1-\rho)Q\left(\frac{\underline{v} - \varphi}{\sqrt{\underline{v}}}\right) = \min\{\rho, 1-\rho\}.$$

In this case, our local application of BP cannot do better than the trivial estimator. In fact, the local neighborhoods are statistically uncorrelated with the cluster structure, and one can further argue that no local algorithm can achieve non-trivial detection (see e.g. Kanade et al. (2014)). Although local algorithms are bound to fail, there might still exist some efficient global algorithms which achieve the minimum expected misclassified fraction. The following theorem shows that this is indeed the case when $\rho = 1/2$, $\mu = \nu$ and $b = o(\log n)$.

**Theorem 3** *Assume* $\rho = 1/2$, $a = c = b + \sqrt{b}\mu$ *for some fixed constant* $\mu$, *and* $b \to \infty$ *such that* $b = o(\log n)$. *For an estimator* $\widehat{\sigma}$ *based on graph* $G$, *define the fraction of vertices misclassified by* $\widehat{\sigma}$ *as*

$$\mathcal{O}(\widehat{\sigma}, \sigma) = \frac{1}{n}\min\left\{\sum_{i=1}^n \mathbf{1}_{\{\sigma_i \ne \widehat{\sigma}_i\}}, \sum_{i=1}^n \mathbf{1}_{\{\sigma_i \ne -\widehat{\sigma}_i\}}\right\}. \tag{6}$$

*If* $|\mu| > 2$, *then*

$$\lim_{n\to\infty}\inf_{\widehat{\sigma}} \mathbb{E}\left[\mathcal{O}(\widehat{\sigma}, \sigma)\right] = Q\left(\sqrt{\overline{v}}\right), \tag{7}$$

*where the infimum ranges over all possible estimators* $\widehat{\sigma}$ *based on graph* $G$; $\overline{v} > 0$ *is the largest fixed point of* $v = \frac{\mu^2}{4}\mathbb{E}\left[\tanh(v + \sqrt{v}Z)\right]$. *Moreover, there is a polynomial-time estimator such that for every* $\epsilon > 0$, *it misclassifies at most* $Q\left(\sqrt{\overline{v}}\right) - \epsilon$ *fraction of vertices on expectation.*

In contrast to (2), the fraction of vertices misclassified by $\widehat{\sigma}$ is defined up to a global flip of signs of $\widehat{\sigma}$ in (6). This is because in the case $\rho = 1/2$ and $a = c$, due to the symmetry between $+$ and $-$, $\sigma$ and $-\sigma$ have the same distribution conditional on graph $G$. Thus, it is impossible to reliably estimate the sign of $\sigma$ based on graph $G$.

Note that $|\mu| = 2$ corresponds to the Kesten-Stigum bound Kesten and Stigum (1966). It is shown in Mossel et al. (2015a) that if $|\mu| < 2$, correlated recovery is impossible and thus the minimum expected misclassified fraction is 0; Remarkably, Massoulié (2014); Mossel et al. (2013b); Bordenave et al. (2015) prove that correlated recovery is efficiently achievable if $|\mu| > 2$. Our results further show that in this case with $b \to \infty$ and $b = o(\log n)$, the minimum expected misclassified fraction is $Q\left(\sqrt{\overline{v}}\right)$ and it can also be attained in polynomial-time. The proof of Theorem 3 is mainly based on two observations. First, it is shown in Mossel et al. (2013a) that the local BP is able to improve a clustering that is slightly better than a random guess to achieve the minimum expected misclassified fraction if $|\mu| > C$ for a universal constant $C > 0$. Second, we find that if $|\mu| > 2$, the recursion $v = \frac{\mu^2}{4}\mathbb{E}\left[\tanh(v + \sqrt{v}Z)\right]$ derived in the density evolution analysis of local BP has only two fixed points: 0 and $\overline{v} > 0$, where 0 is unstable and $\overline{v}$ is stable. These two results together establish that if $|\mu| > 2$, then running the local BP for $t$ iterations with a correlated initialization provided by a non-trivial detection algorithm is able to attain the minimum expected misclassified fraction $Q\left(\sqrt{\overline{v}}\right)$ as $t \to \infty$.

## 2.3. Proof Ideas

The proof of Theorem 2 is based on two useful tools. First, we connect the cluster recovery problem to the reconstruction problem on trees. Second, we use the density evolution with Gaussian approximations to give a sharp characterization of error probabilities of tree reconstruction problems, in terms of fixed points of a recursion.

- To bound from below the minimum expected misclassified fraction, we bound the error probability of inferring $\sigma_u$ for a specific vertex $u$. Following Mossel et al. (2015a), we consider an oracle estimator, which in addition to the graph structure, the exact labels of all vertices at distance exactly $t$ from $u$ are also revealed. As in Mossel et al. (2015a), it is possible to show that the best estimator in this case is given by BP for $t$ levels initialized using the exact labels at distance $t$. In contrast, the expected fraction of vertices misclassified by the local BP algorithm approximately equals to the error probability of inferring $\sigma_u$ solely based on the neighborhood of vertex $u$ of radius $t$, without having access to the exact labels of vertices at distance $t$.

- We characterize the density evolution of the local BP and the BP lower bound using Gaussian approximations, and get a recursion with the largest fixed point corresponding to the BP lower bound, and the smallest fixed point corresponding to the expected fraction of vertices misclassified by the local BP as $t \to \infty$. In the balanced cluster cases, we further show that there is a unique fixed point for the recursion, and thus the BP lower bound matches the expected fraction of vertices misclassified by the local BP.

## 2.4. Numerical Experiments and Open Problems

In the case with $\rho = 1/2$ and $\mu \neq \nu$, we show that $v = \theta + \lambda h(v)$ has a unique fixed point and thus the local BP is optimal; the key idea is to prove that $h(v)$ is concave in this case. Numerical
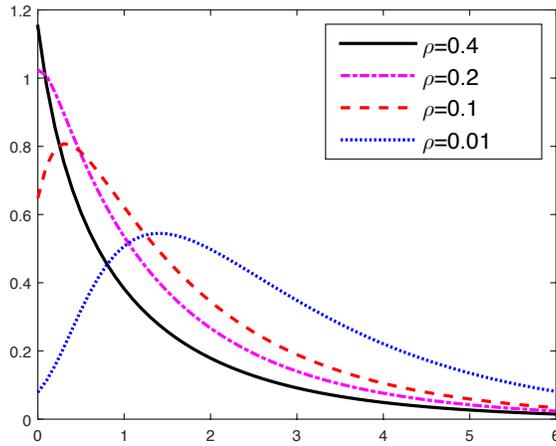
Figure 1: Numerical calculations of $h'(v)$ (y axis) versus $v \in [0,6]$ (x axis) with different $\rho$. It shows that $h(v)$ is concave when $\rho \geq 0.2$ and $h(v)$ becomes convex for $v$ small when $\rho \leq 0.1$.
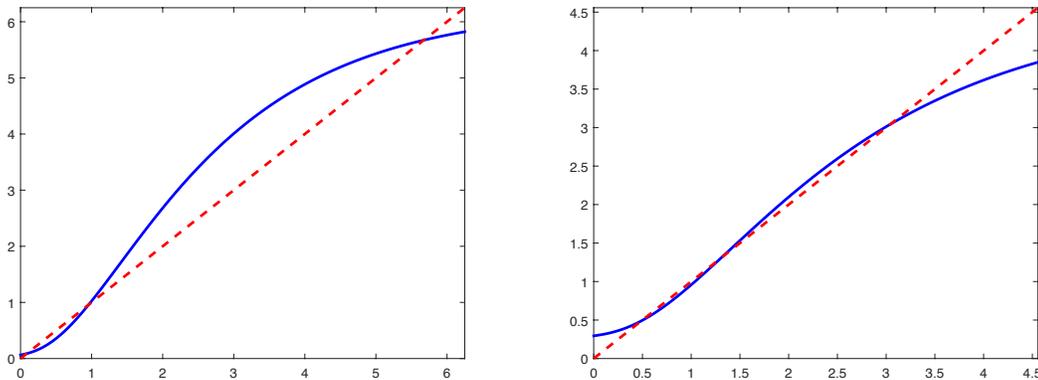


Figure 2: The plot of $\theta + \lambda h(v)$ (y axis) versus $v$ (x axis) in the case $\rho = 0.01$. Left frame: $\mu = 50$ and $\nu = 0$; right frame: $\mu = 40$ and $\nu = 1.5$. It shows that $v = \theta + \lambda h(v)$ has three fixed points: The smallest one is $\underline{v}$ corresponding to the performance of local BP; the largest one is $\overline{v}$ corresponding the lower bound on expected misclassified fraction; the intermediate one is unstable.

calculations, depicted in Fig. 1, show that $h(v)$ is still concave if $\rho \geq 0.2$, suggesting that the local BP is still optimal in roughly balanced cluster size cases. However, $h(v)$ becomes convex for $v$ small when $\rho \leq 0.1$.

It is intriguing to investigate when $v = \theta + \lambda h(v)$ has a unique fixed point. If $\rho = 0.01$, numerical experiments, depicted in Fig. 2, shows that $v = \theta + \lambda h(v)$ may have multiple fixed points, suggesting that the local BP may be suboptimal. However, in the case with $\mu = \nu$ and $\rho \neq 1/2$, numerical experiments indicate that there is always a unique fixed point.

**Conjecture 4** *If $\mu = \nu$, then $v = \theta + \lambda h(v)$ has a unique fixed point for all $\rho \in (0, 1/2) \cup (1/2, 1)$.*

Notice that in the case with $\mu = \nu$ and $\rho = 1/2$, $\theta = 0$, $\lambda = \mu^2/4$, and $h(v) = \mathbb{E}\left[\tanh(v + \sqrt{v}Z)\right]$. We have shown in Lemma 14 that $h$ is non-decreasing, concave, and $\lim_{v \to 0} h'(v) = 1$. Thus if $|\mu| < 2$, there is a unique fixed point at zero, which is stable; if $|\mu| > 2$, there are two fixed points: one is zero which is unstable and the other is $\overline{v} > 0$ which is stable.

### 2.5. Notation and Organization of the Paper

For any positive integer $n$, let $[n] = \{1, \ldots, n\}$. For any set $T \subset [n]$, let $|T|$ denote its cardinality and $T^c$ denote its complement. We use standard big $O$ notations, e.g., for any sequences $\{a_n\}$ and $\{b_n\}$, $a_n = \Theta(b_n)$ if there is an absolute constant $c > 0$ such that $1/c \leq a_n/b_n \leq c$. Let $\mathrm{Bern}(p)$ denote the Bernoulli distribution with mean $p$ and $\mathrm{Binom}(n, p)$ denote the binomial distribution with $n$ trials and success probability $p$. All logarithms are natural and we use the convention $0 \log 0 = 0$. We say a sequence of events $\mathcal{E}_n$ holds with high probability if $\mathbb{P}\{\mathcal{E}_n\} \to 1$.

The rest of this paper is organized as follows. Section 3 focuses on the inference problems on the tree model. The analysis of the belief propagation algorithm on trees and the proofs of our main theorems are given in Section 4. The proofs of auxiliary lemmas can be found in Appendix A.

## 3. Inference Problems on Galton-Watson Tree Model

In this section, we first introduce the inference problems on Galton-Watson trees, and then relate it to the cluster recovery problem under the stochastic block model.

**Definition 5** *For a vertex $u$, we denote by $(T_u, \tau)$ the following Poisson two-type branching process tree rooted at $u$, where $\tau$ is a $\{\pm\}$ labeling of the vertices of $T$. Let $\tau_u = +$ with probability $\rho$ and $\tau_u = -1$ with probability $\bar{\rho}$, where $\bar{\rho} = 1 - \rho$. Now recursively for each vertex $i$ in $T_u$, given its label $\tau_i = +$, $i$ will have $L_i \sim \mathrm{Pois}(\rho a)$ children $j$ with $\tau_j = +$ and $M_i \sim \mathrm{Pois}(\bar{\rho}b)$ children $j$ with $\tau_j = -$; given its label $\tau_i = -1$, $i$ will have $L_i \sim \mathrm{Pois}(\rho b)$ children $j$ with $\tau_j = +$ and $M_i \sim \mathrm{Pois}(\bar{\rho}c)$ children $j$ with $\tau_j = -$.*

For any vertex $i$ in $T_u$, let $T_i^t$ denote the subtree of $T_u$ of depth $t$ rooted at vertex $i$, and $\partial T_i^t$ denote the set of vertices at the boundary of $T_i^t$. With a bit abuse of notation, let $\tau_A$ denote the vector consisting of labels of vertices in $A$, where $A$ could be either a set of vertices or a subgraph in $T_u$. We first consider the problem of estimating the label of root $u$ given the observation of $T_u^t$ and $\tau_{\partial T_u^t}$. Notice that the labels of vertices in $T_u^{t-1}$ are not observed.

**Definition 6** *The* detection problem on the tree with exact information at the boundary *is the problem of inferring $\tau_u$ from the observation of $T_u^t$ and $\tau_{\partial T_u^t}$. The error probability for an estimator $\widehat{\tau}_u(T_u^t, \tau_{\partial T_u^t})$ is defined by*

$$p_{T^t}(\widehat{\tau}_u) = \rho \mathbb{P}\left\{\widehat{\tau}_u = - | \tau_u = +\right\} + \bar{\rho}\mathbb{P}\left\{\widehat{\tau}_u = + | \tau_u = -\right\}.$$

*Let $p_{T^t}^*$ denote the minimum error probability among all estimators based on $T_u^t$ and $\tau_{\partial T_u^t}$.*

The optimal estimator in minimizing $p_{T^t}$, is the maximum a posterior (MAP) estimator, which can be expressed in terms of log likelihood ratio:

$$\widehat{\tau}_{\mathrm{ML}} = 2 \times \mathbf{1}_{\{\Lambda_u^t \geq -\varphi\}} - 1,$$

where

$$\Lambda_i^t \triangleq \frac{1}{2} \log \frac{\mathbb{P}\left\{T_i^t, \tau_{\partial T_i^t} | \tau_i = +\right\}}{\mathbb{P}\left\{T_i^t, \tau_{\partial T_i^t} | \tau_i = -\right\}},$$

for all $i$ in $T_u$, and $\varphi = \frac{1}{2} \log \frac{\rho}{1-\rho}$. Thus, the minimum error probability $p_{T^t}^*$ is given by

$$p_{T^t}^* = \frac{1}{2} - \frac{1}{2}\mathbb{E}\left[\left|\mathbb{P}\left\{\tau_u = +|T_u^t, \tau_{\partial T_u^t}\right\} - \mathbb{P}\left\{\tau_u = -|T_u^t, \tau_{\partial T_u^t}\right\}\right|\right]. \tag{8}$$

We then consider the problem of estimating $\tau_u$ given observation of $T_u^t$. Notice that in this case the true labels of vertices in $T_u^t$ are not observed.

**Definition 7** *The* detection problem on the tree *is the problem of inferring $\tau_u$ from the observation of $T_u^t$. The error probability for an estimator $\widehat{\tau}_u(T_u^t)$ is defined by*

$$q_{T^t}(\widehat{\tau}_u) = \rho\mathbb{P}\left\{\widehat{\tau}_u = -|\tau_u = +\right\} + \bar{\rho}\mathbb{P}\left\{\widehat{\tau}_u = +|\tau_u = -\right\}.$$

*Let $q_{T^t}^*$ denote the minimum error probability among all estimators based on $T_u^t$.*

In passing, we remark that the only difference between Definition 6 and Definition 7 is that the exact labels at the boundary of the tree is revealed to estimators in the former and hidden in the latter. The optimal estimator in minimizing $q_T^t$, is the maximum a posterior (MAP) estimator, which can be expressed in terms of the log likelihood ratio:

$$\widehat{\tau}_{\mathrm{MAP}} = 2 \times \mathbf{1}_{\{\Gamma_u^t \geq -\varphi\}} - 1,$$

where

$$\Gamma_i^t \triangleq \frac{1}{2} \log \frac{\mathbb{P}\left\{\mathcal{T}_i^t | \tau_u = +\right\}}{\mathbb{P}\left\{\mathcal{T}_i^t | \tau_u = -\right\}}.$$

for all $i$ in $T_u$, and $\varphi = \frac{1}{2} \log \frac{\rho}{1-\rho}$. The minimum error probability $q_{T^t}^*$ is given by

$$q_{T^t}^* = \frac{1}{2} - \frac{1}{2}\mathbb{E}\left[\left|\mathbb{P}\left\{\tau_u = +|T_u^t\right\} - \mathbb{P}\left\{\tau_u = -|T_u^t\right\}\right|\right], \tag{9}$$

If $d_+ = d_-$, then the distribution of $T_u^t$ conditional on $\tau_u = +$ is the same as that conditional on $\tau_u = -$. Thus, $\Gamma_u^t = 0$ and the MAP estimator reduces to the trivial estimator, which always guesses the label to be $+$ if $\rho \geq 1/2$ and $-$ if $\rho < 1/2$, and $q_{T^t}^* = \min\{\rho, \bar{\rho}\}$. If $d_+ \neq d_-$, then $T_u$ becomes statistically correlated with $\tau_u$, and it is possible to do better than the trivial estimator based on $T_u^t$.

For the tree model, the likelihoods can be computed exactly via a belief propagation algorithm. The following lemma gives a recursive formula to compute $\Lambda_i^t$ and $\Gamma_i^t$; no approximations are needed. Let $\partial i$ denote the set of children of vertex $i$.

**Lemma 8** *Recall $F(x) = \frac{1}{2} \log\left(\frac{e^{2x}\rho a + \bar{\rho}b}{e^{2x}\rho b + \bar{\rho}c}\right)$. For $t \geq 0$,*

$$\Lambda_i^{t+1} = \frac{-d_+ + d_-}{2} + \sum_{j \in \partial i} F(\Lambda_j^t), \tag{10}$$

$$\Gamma_i^{t+1} = \frac{-d_+ + d_-}{2} + \sum_{j \in \partial i} F(\Gamma_j^t), \tag{11}$$

*with $\Lambda_i^0 = \infty$ if $\tau_i = +$ and $\Lambda_i^0 = -\infty$ if $\tau_i = -$; $\Gamma_i^0 = 0$ for all $i$.*

### 3.1. Connection between the Graph Problem and Tree Problems

For the reconstruction problem on graph, recall that $p_G(\widehat{\sigma}_{\mathrm{BP}}^t)$ denote the expected fraction of vertices misclassified by $\widehat{\sigma}_{\mathrm{BP}}^t$ as per (2); $p_G^*$ is the minimum expected misclassified fraction. For the reconstruction problems on tree, recall that $p_{T^t}^*$ is the minimum error probability of estimating $\tau_u$ based on $T_u^t$ and $\tau_{\partial T_u^t}$ as per (8); $q_{T^t}^*$ is minimum error probability of estimating $\tau_u$ based on $T_u^t$ as per (9). In this section, we show that in the limit $n \to \infty$, $p_G(\widehat{\sigma}_{\mathrm{BP}}^t)$ equals to $q_{T^t}^*$, and $p_G^*$ is bounded by $p_{T^t}^*$ from the below for any $t \geq 1$. Notice that $q_{T^t}^*$ and $p_{T^t}^*$ depend on $n$ only through the parameters $a$, $b$, and $c$.

A key ingredient is to show that $G$ is locally tree-like with high probability in the regime $b = n^{o(1)}$. Let $G_u^t$ denote the subgraph of $G$ induced by vertices whose distance to $u$ is at most $t$ and let $\partial G_u^t$ denote the set of vertices whose distance from $u$ is precisely $t$. In the following, for ease of notation, we write $T_u^t$ as $T^t$ and $G_u^t$ as $G^t$ when there is no ambiguity. With a bit abuse of notation, let $\sigma_A$ denote the vector consisting of labels of vertices in $A$, where $A$ could be either a set of vertices or a subgraph in $G$. The following lemma proved in Mossel et al. (2015a) shows that we can construct a coupling such that $(G^t, \sigma_{G^t}) = (T^t, \tau_{T^t})$ with probability converging to 1 when $b^t = n^{o(1)}$.

**Lemma 9** *For $t = t(n)$ such that $b^t = n^{o(1)}$, there exists a coupling between $(G, \sigma)$ and $(T, \tau)$ such that $(G^t, \sigma_{G^t}) = (T^t, \tau_{T^t})$ with probability converging to 1.*

Suppose that $(G^t, \sigma_{G^t}) = (T^t, \tau_{T^t})$ holds, then by comparing BP iterations (4) and (5) with the recursions of log likelihood ratio $\Gamma^t$ given in (11), we find that $R_u^t$ exactly equals to $\Gamma_u^t$, i.e., the BP algorithm defined in Algorithm 1 exactly computes the log likelihood ratio $\Gamma_u^t$ for the tree model. Building upon this intuition, the following lemma shows that $p_G(\widehat{\sigma}_{\mathrm{BP}}^t)$ equals to $q_{T^t}^*$ as $n \to \infty$.

**Lemma 10** *For $t = t(n)$ such that $b^t = n^{o(1)}$,*

$$\lim_{n\to\infty} |p_G(\widehat{\sigma}_{\mathrm{BP}}^t) - q_{T^t}^*| = 0.$$

**Proof** In view of Lemma 9, we can construct a coupling such that $(G^t, \sigma_{G^t}) = (T^t, \tau_{T^t})$ with probability converging to 1. On the event $(G^t, \sigma_{G^t}) = (T^t, \tau_{T^t})$, we have that $R_u^t = \Gamma_u^t$. Hence,

$$p_G(\widehat{\sigma}_{\mathrm{BP}}^t) = q_{T^t}^* + o(1), \tag{12}$$

where $o(1)$ term comes from the coupling error. $\blacksquare$

The following lemma shows that $p_G^*$ is lower bounded by $p_{T^t}^*$ as $n \to \infty$.

**Lemma 11** *For $t = t(n)$ such that $b^t = n^{o(1)}$,*

$$\limsup_{n\to\infty} (p_G^* - p_{T^t}^*) \geq 0.$$

We pause a while to give some intuition behind the lemma. To lower bound $p_G^*$, it suffices to lower bound the error probability of estimating $\sigma_u$ for a given vertex $u$ based on graph $G$. To this end, we consider an oracle estimator, which in addition to the graph structure, the exact labels of all vertices at distance exactly $t$ from $u$ are also revealed. We further show that once the exact

labels at distance $t$ are conditioned, $\sigma_u$ becomes asymptotically independent of the labels of all vertices at distance larger than $t$ from $u$. Hence, effectively the oracle estimator is equivalent to the MAP estimator solely based on the graph structure in $G_u^t$ and the exact labels at distance $t$. By the coupling lemma, $G_u^t$ is a tree with high probability, and thus the error probability of the oracle estimator asymptotically equals to $p_{T^t}^*$.

## 4. Gaussian Density Evolution

In the previous subsection, we have argued that in the limit $n \to \infty$, $p_G(\widehat{\sigma}_{\mathrm{BP}}^t)$ equals to $q_{T^t}^*$, and $p_G^*$ is bounded by $p_{T^t}^*$ from the below. In this section, we analyze recursions (10) and (11) using density evolution analysis with Gaussian approximations, and derive simple formulas for $p_{T^t}^*$ and $q_{T^t}^*$ in the limit $n \to \infty$. Afterwards, we give the proof of Theorem 2.

Notice that $\Gamma_i^t$ is a function of $T_i^t$ alone. Since the subtrees $\{T_i^t\}_{i \in \partial u}$ conditional on $\tau_u$ are independent and identically distributed, $\{\Gamma_i^t\}_{i \in \partial u}$ conditional on $\tau_u$ are also independent and identically distributed. Thus, in view of the recursion (11), $\Gamma_u^t$ can be viewed as a sum of i.i.d. random variables. When the expected degree of $u$ tends to infinity, due to the central limit theorem, we expect that the distribution of $\Gamma_u^t$ conditional on $\tau_u$ is approximately Gaussian. Moreover, the construction of the subtree $T_i^t$ conditional on $\tau_i$ is the same as the construction of $T_u^t$ conditional on $\tau_u$. Therefore, for any $i \in \partial u$, the distribution of $\Gamma_i^t$ conditional on $\tau_i$ is the same as the distribution of $\Gamma_u^t$ conditional on $\tau_u$. Similar conclusions hold for $\Lambda_i^t$ as well.

Let $Z_{\pm}^t$ ($W_{\pm}^t$) denote a random variable that has the same distribution as $\Gamma_u^t$ ($\Lambda_u^t$) conditional on $\tau_u = \pm$. The following lemma provides expressions of the mean and variance of $Z_+^t$ and $Z_-^t$. Recall that $\lambda = \frac{\rho(\mu+\nu)^2}{8}$ and $\theta = \frac{\rho(\mu-\nu)^2}{8} + \frac{(1-2\rho)\nu^2}{4}$.

**Lemma 12** *For all $t \geq 0$,*

$$\mathbb{E}\left[Z_{\pm}^{t+1}\right] = \pm\theta \pm \lambda\mathbb{E}\left[\tanh(Z_+^t + \varphi)\right] + O(b^{-1/2}), \tag{13}$$

$$\mathrm{var}\left(Z_{\pm}^{t+1}\right) = \theta + \lambda\mathbb{E}\left[\tanh(Z_+^t + \varphi)\right] + O(b^{-1/2}). \tag{14}$$

Recall that $(v_t : t \geq 0)$ satisfies $v_0 = 0$ and

$$v_{t+1} = \theta + \lambda h(v_t) = \theta + \lambda\mathbb{E}\left[\tanh(v_t + \sqrt{v_t}Z + \varphi)\right],$$

where $Z \sim \mathcal{N}(0, 1)$. Similarly, define $(w_t : t \geq 1)$ by $w_1 = \theta + \lambda = \frac{\rho\mu^2 + \bar{\rho}\nu^2}{4}$ and

$$w_{t+1} = \theta + \lambda h(w_t) = \theta + \lambda\mathbb{E}\left[\tanh(w_t + \sqrt{w_t}Z + \varphi)\right].$$

The following lemma shows that for any fixed $t \geq 0$, $Z_{\pm}^t$ and $W_{\pm}^t$ are approximately Gaussian.

**Lemma 13** *For any $t \geq 0$, as $n \to \infty$,*

$$\sup_x \left|\mathbb{P}\left\{\frac{Z_{\pm}^t \mp v_t}{\sqrt{v_t}} \leq x\right\} - \mathbb{P}\left\{Z \leq x\right\}\right| = O(b^{-1/2}). \tag{15}$$

*Similarly, for any $t \geq 1$, as $n \to \infty$,*

$$\sup_x \left|\mathbb{P}\left\{\frac{W_{\pm}^t \mp w_t}{\sqrt{w_t}} \leq x\right\} - \mathbb{P}\left\{Z \leq x\right\}\right| = O(b^{-1/2}). \tag{16}$$

Before proving Theorem 2, we also need a key lemma, which shows that $h$ is continuous and non-decreasing, and $h$ is concave if $\varphi = 0$.

**Lemma 14** *$h(v)$ is continuous on $[0, \infty)$ and for $v \in (0, +\infty)$,*

$$h'(v) = \mathbb{E}\left[\left(1 - \tanh(v + \sqrt{v}Z + \varphi)\right)\left(1 - \tanh^2(v + \sqrt{v}Z + \varphi)\right)\right] \geq 0. \qquad (17)$$

*Furthermore, if $\varphi = 0$, then $h'(v) \geq h'(w)$ for $0 < v < w < \infty$.*

Finally, we are ready to prove Theorem 2 based on Lemma 13 and Lemma 14.
**Proof** [Proof of Theorem 2] In view of Lemma 13,

$$\lim_{n\to\infty} \mathbb{P}\left\{\Gamma_u^t \geq -\varphi|\tau_u = -\right\} = Q\left(\frac{v_t - \varphi}{\sqrt{v_t}}\right),$$

$$\lim_{n\to\infty} \mathbb{P}\left\{\Gamma_u^t \leq -\varphi|\tau_u = +\right\} = Q\left(\frac{v_t + \varphi}{\sqrt{v_t}}\right).$$

Hence, it follows from Lemma 10 that

$$\lim_{n\to\infty} p_G(\widehat{\sigma}_{\mathrm{BP}}^t) = \lim_{n\to\infty} q_{T^t}^* = \mathbb{E}\left[Q\left(\frac{v_t + U}{\sqrt{v_t}}\right)\right],$$

where $U = -\varphi$ with probability $1 - \rho$ and $U = \varphi$ with probability $\rho$.

We prove that $v_{t+1} \geq v_t$ for $t \geq 0$ by induction. Recall that

$$v_0 = 0 \leq (\rho\mu - \bar\rho\nu)^2/4 = \theta + \lambda h(v_0) = v_1.$$

Suppose $v_{t+1} \geq v_t$ holds; we shall show the claim also holds for $t + 1$. In particular, since $h$ is continuous on $[0, \infty)$ and differential on $(0, \infty)$, it follows from the mean value theorem that

$$v_{t+2} - v_{t+1} = \lambda\left(h(v_{t+1}) - h(v_t)\right) = \lambda h'(x)(v_{t+1} - v_t),$$

for some $x \in (v_t, v_{t+1})$. Lemma 14 implies that $h'(x) \geq 0$ for $x \in (0, \infty)$, it follows that $v_{t+2} \geq v_{t+1}$. Hence, $v_t$ is non-decreasing in $t$. Next we argue that $v_t \leq \underline{v}$ for all $t \geq 0$ by induction, where $\underline{v}$ is the smallest fixed point of $v = \theta + \lambda h(v)$. For the base case, $v_0 = 0 \leq \underline{v}$. If $v_t \leq \underline{v}$, then by the monotonicity of $h$, $v_{t+1} = \theta + \lambda h(v_t) \leq \theta + \lambda h(\underline{v}) = \underline{v}$. Hence, $v_t \leq \underline{v}$ and thus $\lim_{t\to\infty} v_t \leq \underline{v}$. By the continuity of $h$, $\lim_{t\to\infty} v_t$ is also a fixed point of $v = \theta + \lambda h(v)$, and consequently $\lim_{t\to\infty} v_t = \underline{v}$. Therefore,

$$\lim_{t\to\infty}\lim_{n\to\infty} p_G(\widehat{\sigma}_{\mathrm{BP}}^t) = \lim_{t\to\infty}\lim_{n\to\infty} q_{T^t}^* = \mathbb{E}\left[Q\left(\frac{\underline{v} + U}{\sqrt{\underline{v}}}\right)\right].$$

Next, we prove the claim for $p_G^*$. In view of Lemma 13,

$$\lim_{n\to\infty} \mathbb{P}\left\{\Lambda_u^t \geq -\varphi|\tau_u = -\right\} = Q\left(\frac{w_t - \varphi}{\sqrt{w_t}}\right),$$

$$\lim_{n\to\infty} \mathbb{P}\left\{\Lambda_u^t \leq -\varphi|\tau_u = +\right\} = Q\left(\frac{w_t + \varphi}{\sqrt{w_t}}\right).$$

Hence, it follows from Lemma 11 that

$$\liminf_{n\to\infty} p_G^* \geq \lim_{n\to\infty} p_{T^t}^* = \mathbb{E}\left[Q\left(\frac{w_t + U}{\sqrt{w_t}}\right)\right].$$

Recall that $w_1 = \theta + \lambda \geq w_t$. By the same argument of proving $v_t$ is non-decreasing, one can show that $w_t$ is non-increasing in $t$. Also, by the same argument of proving $v_t$ is upper bounded by $\underline{v}$, one can show that $w_t$ is lower bounded by $\overline{v}$, where $\overline{v}$ is the largest fixed point of $v = \theta + \lambda h(v)$. Thus, $\lim_{t\to\infty} w_t = \overline{v}$ and $\overline{v} \leq w_1 = \theta + \lambda = (\rho\mu^2 + \bar{\rho}\nu^2)/4$. Therefore,

$$\liminf_{n\to\infty} p_G^* = \lim_{t\to\infty}\liminf_{n\to\infty} p_G^* \geq \lim_{t\to\infty}\lim_{n\to\infty} p_{T^t}^* = \mathbb{E}\left[Q\left(\frac{\overline{v} + U}{\sqrt{\overline{v}}}\right)\right].$$

If $\varphi = 0$ and $\mu \neq \nu$, then $v_1 > 0$ and Lemma 14 shows that $h'(v) \geq h'(w)$ for all $0 < v < w < \infty$. Since $v_1 = \theta + \lambda h(0) > 0$ and $\underline{v} = \theta + \lambda h(\underline{v})$, it must hold that $\lambda h'(\underline{v}) < 1$. Thus $\lambda h'(v) < 1$ for all $v \geq \underline{v}$ and consequently $\theta + \lambda h(v) < v$ for all $v > \underline{v}$. Hence, $\overline{v} = \underline{v} = v^*$, where $v^*$ is the unique fixed point of $v = \frac{(\mu-\nu)^2}{16} + \frac{(\mu+\nu)^2}{16}\mathbb{E}\left[\tanh(v + \sqrt{v}Z)\right]$. Therefore,

$$\liminf_{n\to\infty} p_G^* \geq \lim_{t\to\infty}\lim_{n\to\infty} p_{T^t}^* = \lim_{t\to\infty}\lim_{n\to\infty} q_{T^t}^* = \lim_{t\to\infty}\lim_{n\to\infty} p_G(\widehat{\sigma}_{\mathrm{BP}}^t) = Q(\sqrt{v^*}).$$

Since $p_G^*$ is the minimum expected misclassified fraction, it also holds that $\limsup_{n\to\infty} p_G^* \leq \lim_{n\to\infty} p_G(\widehat{\sigma}_{\mathrm{BP}}^t)$ for all $t \geq 1$ and consequently

$$\limsup_{n\to\infty} p_G^* \leq \lim_{t\to\infty}\lim_{n\to\infty} p_G(\widehat{\sigma}_{\mathrm{BP}}^t).$$

Combing the last two displayed equations gives that

$$\lim_{n\to\infty} p_G^* = \lim_{t\to\infty}\lim_{n\to\infty} p_G(\widehat{\sigma}_{\mathrm{BP}}^t) = Q(\sqrt{v^*}).$$

∎

### 4.1. Degree-Uncorrelated Case

As remarked in Section 2.2, in the case $\rho\mu = \bar{\rho}\nu$, the vertex degrees are statistically uncorrelated with the cluster structure, and no local algorithms is capable of non-trivial detection. However, it is still possible that local algorithms combined with some efficient global algorithms achieve the minimum expected misclassified fraction. In this subsection, we show that it is indeed the case, if $\rho = 1/2$, $\mu = \nu$ with $|\mu| > 2$, and $b = o(\log n)$. The algorithm as described in Algorithm 2 is introduced in Mossel et al. (2013a) and we give the full description for completeness.

Notice that Algorithm 2 runs in time polynomial in $n$. The algorithm consists of two main steps. First, we apply some global algorithm to get a correlated clustering when $|\mu| > 2$, for example, the algorithm studied in Mossel et al. (2013c). Then, we apply the local BP algorithm to boost the correlated clustering in the first step to achieve the minimum expected misclassified fraction. To ensure the first and second step are independent of each other, for each vertex $u$, we first withhold the $(t-1)$-local neighborhood of $u$, and then apply the global detection algorithms on the reduced set of vertices. The clustering on the reduced set of vertices is used as the initialization to the local

---

**Algorithm 2** Local Belief propagation Plus Correlated Recovery

---

1: Input: $n \in \mathbb{N}$, $a = c, b > 0$, $\rho = 1/2$, adjacency matrix $A \in \{0, 1\}^{n \times n}$, $t \in \mathbb{N}$.

2: Take $U \subset V$ to be a random subset of size $\lfloor \sqrt{n} \rfloor$. Let $u_* \in U$ be a random vertex in $U$ with at least $\frac{\log n}{2 \log(\log n / b)}$ neighbors in $V \setminus U$.

3: For $u \in V \setminus U$ do

1. Run a polynomial-time estimator capable of correlated recovery on the subgraph induced by vertices not in $G_u^{t-1}$ and $U$, and let $W_u^+$ and $W_u^-$ denote the output of the partition.

2. Relabel $W_u^+$ and $W_u^-$ such that if $a > b$, then $u_*$ has more neighbors in $W_u^+$ than $W_u^-$; otherwise, $u_*$ has more neighbors in $W_u^-$ than $W_u^+$. Let $\alpha_u$ denote the fraction of vertices misclassified by the partition $(W_u^+, W_u^-)$.

3. For all $i \in \partial G_u^t$ and $j \in \partial G_u^{t-1}$, define $R_{i \to j}^0 = \frac{1}{2} \log \frac{1 - \alpha_u}{\alpha_u}$ if $i \in W_u^+$, and $R_{i \to j}^0 = -\frac{1}{2} \log \frac{1 - \alpha_u}{\alpha_u}$ if $i \in W_u^-$.

4. Run $t - 1$ iterations of message passing as in (4) to compute $R_{i \to u}^{t-1}$ for all $u$'s neighbors $i$.

5. Compute $R_u^t$ as per (5), and let $\hat{\sigma}_{\mathrm{BP}}^t(u) = +$ if $R_u^t \geq -\varphi$; otherwise let $\hat{\sigma}_{\mathrm{BP}}^t(u) = -$.

4: For $u \in U$, let $\hat{\sigma}_{\mathrm{BP}}^t(u)$ equal to $+$ or $-$ uniformly at random. Output $\hat{\sigma}_{\mathrm{BP}}^t$.

---

belief propagation algorithm running on the withheld $(t - 1)$-local neighborhood of $u$. In this way, the outcome of the global detection algorithm based on the reduced set of vertices is independent of the edges between the withheld $t$-local neighborhood of $u$ and the reduced set of vertices, as well as the edges within the withheld set.

There is also a subtle issue to overcome. We run the global detection algorithm once for each vertex, and the global detection algorithm cannot reliably estimate the sign of the true $\sigma$ due to the symmetry between $+$ and $-$. Therefore, different runs of the global detection algorithm may have different estimates of the sign of $\sigma$. We need a way to coordinate different runs of the global detection algorithms to have the same estimate of the sign of $\sigma$. To this end, a small random subset $U$ is reserved and a vertex of high degree $u_*$ in $U$ is served as an anchor. In every runs of the global detection algorithms, we relabel the partition if necessary, to ensure that $u_*$ will always have more neighbors with estimated $+$ labels than neighbors with estimated $-$ labels if $a > b$, and the other way around if $a < b$.

Finally, we caution the reader that in addition to the model parameters $a, b$, after each run of the global detection algorithm, the algorithm requires knowing $\alpha_u$, which is the fraction of vertices misclassified by the partition $(W_u^+, W_u^-)$. In the main analysis, we assume the exact value of $\alpha_u$ is known for simplicity. One can check that only an estimator $\hat{\alpha}_u = \alpha_u + o(1)$ with high probability is needed for Theorem 3 to hold. In Appendix B, we give an efficient and data-driven procedure to construct such a consistent estimate of $\alpha_u$.

Next, in the limit $n \to \infty$, we give a lower bound on the minimum expected misclassified fraction, and an upper bound attainable by $\hat{\sigma}_{\mathrm{BP}}^t$. Then we show that the lower and upper bound match with each other in the double limit, where first $n \to \infty$ and then $t \to \infty$.

16

Recall that the fraction of vertices misclassified by $\widehat{\sigma}$ is defined up to a global flip of signs of $\widehat{\sigma}$ as in (6). The following lemma shows that the minimum expected misclassified fraction is still lower bounded by $p_{T^t}^*$. Its proof is very similar to the proof of Lemma 11. The key new challenge is that $\mathbb{E}\left[\mathcal{O}(\sigma, \widehat{\sigma})\right]$ does not reduce to the error probability of estimating $\sigma_u$ for a given vertex $u$ directly.

**Lemma 15** *For $t = t(n)$ such that $b^t = n^{o(1)}$,*

$$\limsup_{n \to \infty} \left( \inf_{\widehat{\sigma}} \mathbb{E}\left[\mathcal{O}(\sigma, \widehat{\sigma})\right] - p_{T^t}^* \right) \geq 0,$$

*where $p_{T^t}^*$ is defined in (8) under the tree model with $\rho = 1/2$ and $a = c$ defined in Definition 5.*

In the following, we relate the expected fraction of vertices misclassified by $\widehat{\sigma}_{\mathrm{BP}}^t$ as defined in Algorithm 2 to an estimation problem on the tree model. In particular, consider the tree model with $\rho = 1/2$ and $a = c$ as defined in Definition 5. Fix an $\alpha \in [0, 1/2]$. Let $\widetilde{\tau}_i = \tau_i$ with probability $1 - \alpha$ and $\widetilde{\tau}_i = -\tau_i$ with probability for $\alpha$, independently for all $i \in T_u$. Then $\widetilde{\tau}_{\partial T_u^t}$ is a $\alpha$-noisy version of $\tau_{\partial T_u^t}$. Let $\widetilde{q}_{T^t, \alpha}$ denote the minimum error probability of inferring $\tau_u$ based on $T_u^t$ and $\widetilde{\tau}_{\partial T_u^t}$. The optimal estimator achieving $\widetilde{q}_{T^t, \alpha}$ is the MAP estimator given by

$$\widehat{\tau}_{\mathrm{MAP}} = 2 \times \mathbf{1}_{\left\{\widetilde{\Gamma}_u^t \geq -\varphi\right\}} - 1,$$

where

$$\widetilde{\Gamma}_i^t \triangleq \frac{1}{2} \log \frac{\mathbb{P}\left\{T_i^t, \widetilde{\tau}_{\partial T_i^t} | \tau_u = +1\right\}}{\mathbb{P}\left\{T_i^t, \widetilde{\tau}_{\partial T_i^t} | \tau_u = -1\right\}}$$

for all $i$ in $T_u$. The minimum error probability $\widetilde{q}_{T^t, \alpha}$ is given by

$$\widetilde{q}_{T^t, \alpha} = \frac{1}{2} \mathbb{P}\left\{\widetilde{\Gamma}_u^t < -\varphi | \tau_u = +\right\} + \frac{1}{2} \mathbb{P}\left\{\widetilde{\Gamma}_u^t \geq -\varphi | \tau_u = -\right\}$$

$$= \frac{1}{2} - \frac{1}{2} \mathbb{E}\left[\left|\mathbb{P}\left\{\tau_u = +|T_u^t, \widetilde{\tau}_{\partial T_u^t}|\right\} - \mathbb{P}\left\{\tau_u = -|T_u^t, \widetilde{\tau}_{\partial T_u^t}\right\}\right|\right],$$

It follows from the definition that $\widetilde{q}_{T^t, \alpha}$ is non-decreasing in $\alpha$. Also, $\widetilde{q}_{T^t, \alpha} = p_{T^t}^*$ if $\alpha = 0$ and $\widetilde{q}_{T^t, \alpha} = q_{T^t}^*$ if $\alpha = 1/2$. The following lemma shows that the fraction of vertices misclassified by $\widehat{\sigma}_{\mathrm{BP}}^t$ as defined in Algorithm 2 is asymptotically no larger than $\widetilde{q}_{T^t, \alpha}$ for some $\alpha \in [0, 1/2)$.

**Lemma 16** *There exists an $\alpha \in [0, 1/2)$ such that for $t = t(n)$ with $b^t = n^{o(1)}$,*

$$\limsup_{n \to \infty} \left( \mathbb{E}\left[\mathcal{O}(\sigma, \widehat{\sigma}_{\mathrm{BP}}^t)\right] - \widetilde{q}_{T^t, \alpha} \right) \leq 0.$$

The following lemma gives a characterization of the distribution of $\widetilde{\Gamma}_u^t$ based on the density evolution with Gaussian approximations.

**Lemma 17** *Let $\widetilde{Z}_+^t$ and $\widetilde{Z}_{-1}$ denote a random variable that has the same distribution as $\widetilde{\Gamma}_u^t$ conditioning on $\tau_u = +$ and $\tau_u = -$, respectively. For any $t \geq 1$, as $n \to \infty$,*

$$\sup_x \left| \mathbb{P}\left\{\frac{\widetilde{Z}_\pm^t \mp u_t}{\sqrt{u_t}} \leq x\right\} - \mathbb{P}\left\{Z \leq x\right\} \right| = O(b^{-1/2}), \tag{18}$$

*where $u_1 = \frac{(1-2\alpha)^2 \mu^2}{4}$ and $u_{t+1} = \frac{\mu^2}{4} \mathbb{E}\left[\tanh(u_t + \sqrt{u_t} Z)\right]$.*

We are ready to prove Theorem 3 by combing Lemma 15, Lemma 16, and Lemma 17.
**Proof** [Proof of Theorem 3] In view of Lemma 17, for $t \geq 0$,

$$\lim_{n \to \infty} \mathbb{P}\left\{ \widetilde{\Gamma}_u^t \geq 0 | \tau_u = - \right\} = \lim_{n \to \infty} \mathbb{P}\left\{ \widetilde{\Gamma}_u^t \leq 0 | \tau_u = + \right\} = Q(\sqrt{u_t}),$$

It follows from Lemma 16 that there exists an $\alpha \in [0, 1/2)$ such that

$$\limsup_{n \to \infty} \mathbb{E}\left[\mathcal{O}(\sigma, \widehat{\sigma}_{\mathrm{BP}}^t)\right] \leq \lim_{n \to \infty} \widetilde{q}_{T^t, \alpha} = Q(\sqrt{u_t}).$$

Let $\widetilde{h}(v) = \mathbb{E}\left[\tanh(v + \sqrt{v}Z)\right]$. In view of Lemma 14, $\widetilde{h}$ is non-decreasing and concave in $[0, \infty)$, and $\lim_{v \to 0} \widetilde{h}'(v) = 1$. Notice that $h(0) = 0$, and thus 0 is a fixed point of $v = \frac{\mu^2}{4}\widetilde{h}(v)$. Moreover, by the mean value theorem, for $v > 0$, $\widetilde{h}(v) = h(0) + \widetilde{h}'(\xi)v$ for some $\xi \in (0, v)$. Thus $\frac{\mu^2}{4}\widetilde{h}(v) = \frac{\mu^2}{4}\widetilde{h}'(\xi)v$. By the assumption that $\mu > 2$, and $\lim_{v \to 0} \widetilde{h}'(v) = 1$, it follows that there exists a $v_* > 0$ such that $\frac{\mu^2}{4}\widetilde{h}(v) > v$ for all $v \in (0, v_*)$. Furthermore, $\widetilde{h}(v) \leq 1$ and hence $\frac{\mu^2}{4}\widetilde{h}(v) < v$ for all $v$ sufficiently large. Since $\widetilde{h}$ is continuous, $v = \frac{\mu^2}{4}\widetilde{h}(v)$ must have nonzero fixed points. Let $\overline{v}$ denote the smallest nonzero fixed point. Then $\overline{v} > 0$, $\frac{\mu^2}{4}\widetilde{h}(v) > v$ for all $v \in (0, \overline{v})$, and $\frac{\mu^2}{4}\widetilde{h}'(\overline{v}) < 1$. Because $\widetilde{h}$ is concave, $\widetilde{h}'(v) \leq \widetilde{h}'(\overline{v})$ for all $v \geq \overline{v}$. Thus $\frac{\mu^2}{4}\widetilde{h}(v) > v$ for all $v > \overline{v}$. Therefore, $\overline{v}$ is the unique nonzero fixed point and also the largest fixed point. It follows that if $u_1 < \overline{v}$, then $\{u_t\}$ is a non-decreasing sequence upper bounded by $\overline{v}$. If $u_1 > \overline{v}$, then $\{u_t\}$ is a non-increasing sequence lower bounded by $\overline{v}$. Since $u_1 > 0$, it follows that $\lim_{t \to \infty} u_t = \overline{v}$. Hence,

$$\limsup_{n \to \infty} \inf_{\widehat{\sigma}} \mathbb{E}\left[\mathcal{O}(\sigma, \widehat{\sigma})\right] \leq \lim_{t \to \infty} \limsup_{n \to \infty} \mathbb{E}\left[\mathcal{O}(\sigma, \widehat{\sigma}_{\mathrm{BP}}^t)\right] \leq \lim_{t \to \infty} \lim_{n \to \infty} \widetilde{q}_{T^t, \alpha} = Q(\sqrt{\overline{v}}). \tag{19}$$

It follows from Theorem 2 and Lemma 15 that

$$\liminf_{n \to \infty} \inf_{\widehat{\sigma}} \mathbb{E}\left[\mathcal{O}(\sigma, \widehat{\sigma})\right] \geq \lim_{t \to \infty} \lim_{n \to \infty} p_{T^t}^* = Q(\sqrt{\overline{v}}). \tag{20}$$

The theorem follows by combining the last two displayed equations. ∎

## 5. Acknowledgement

## References

E. Abbe and C. Sandon. Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms. *arXiv 1503.00609*, March 2015.

E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. arXiv 1405.3267, October 2014.

N. Agarwal, A. S. Bandeira, K. Koiliaris, and A. Kolla. Multisection in the stochastic block model using semidefinite programming. arXiv 1507.02323, July 2015.

A. Anandkumar, R. Ge, D. Hsu, and S. M. Kakade. A tensor spectral approach to learning mixed membership community models. *Journal of Machine Learning Research*, 15:2239–2312, June 2014.

A.S. Bandeira. Random Laplacian matrices and convex relaxations. arXiv 1504.03987, April 2015.

P. J. Bickel and A. Chen. A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Science*, 106(50):21068–21073, 2009.

C. Bordenave, M. Lelarge, and L. Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. ArXiv 1501.06087, January 2015. URL http://arxiv.org/abs/1501.06087.

T. Cai and X. Li. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *arXiv preprint arXiv:1404.6000*, 2014.

Y. Chen and J. Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. In *Proceedings of ICML 2014 (Also arXiv:1402.1267)*, Feb 2014.

Y. Chen, S. Sanghavi, and H. Xu. Improved graph clustering. *IEEE Transactions on Information Theory. An earlier version of this work appeared under the title "Clustering Sparse Graphs" at NIPS 2012*, 60(10):6440–6455, Oct 2014. ISSN 0018-9448.

A. Coja-Oghlan. A spectral heuristic for bisecting random graphs. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '05, pages 850–859, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.

A. Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Comb. Probab. Comput.*, 19 (2):227–284, March 2010.

A. Condon and Richard M. Karp. Algorithms for graph partitioning on the planted partition model. *Random Structures and Algorithms*, 18(2):116–140, 2001.

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106, Dec 2011a.

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Phys. Rev. Lett.*, 107:065701, 2011b.

Y. Deshpande, E. Abbe, and A. Montanari. Asymptotic mutual information for the two-groups stochastic block model. *arXiv:1507.08685*, July 2015.

M. E. Dyer and A. M. Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, 1989.

D. Gamarnik and M. Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 369–376. ACM, 2014.

C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving optimal misclassification proportion in stochastic block model. *arXiv:1505.03772*, 2015.

B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. arXiv1 412.6156, Nov. 2014.

B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *arXiv:1502.07738*, 2015a.

B. Hajek, Y. Wu, and J. Xu. Recovering a hidden community beyond the spectral limit in $O(|E| \log^* |V|)$ time. arXiv 1510.02786, October 2015b.

H. Hatami, L. Lovász, and B. Szegedy. Limits of local-global convergent graph sequences. *arXiv preprint arXiv:1205.4356*, 2012.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

M. Jerrum and G. B. Sorkin. The Metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(1–3):155–175, 1998.

V. Kanade, E. Mossel, and T. Schramm. Global and local information in clustering labeled block models. *In Proceedings of Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 779–810, 2014.

H. Kesten and B. P. Stigum. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *Ann. Math. Statist.*, 37:1463–1481, 1966.

V. Korolev and I. Shevtsova. An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums. *Scandinavian Actuarial Journal*, 2012(2):81–105, 2012.

R. Lyons and F. Nazarov. Perfect matchings as iid factors on non-amenable groups. *European Journal of Combinatorics*, 32(7):1115–1125, 2011.

L. Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the Symposium on the Theory of Computation (STOC)*, 2014.

F. McSherry. Spectral partitioning of random graphs. In *Proceedings of IEEE Conference on the Foundations of Computer Science (FOCS)*, pages 529–537, 2001.

M. Mezard and A. Montanari. *Information, Physics, and Computation*. Oxford University Press, Inc., New York, NY, USA, 2009. ISBN 019857083X, 9780198570837.

A. Montanari. Tight bounds for LDPC and LDGM codes under MAP decoding. *IEEE Transactions on Information Theory*, 51(9):3221–3246, Sept 2005.

A. Montanari. Finding one community in a sparse random graph. arXiv 1502.05680, Feb 2015.

A. Montanari and D. Tse. Analysis of belief propagation for non-linear problems: The example of cdma (or: How to prove tanaka's formula). In *Information Theory Workshop, 2006. ITW '06 Punta del Este. IEEE*, pages 160–164, March 2006.

E. Mossel and J. Xu. Local algorithms for block models with side information. In *Accepted to The 7th Innovations in Theoretical Computer Science (ITCS) conference. arXiv:1508.02344*, 2015.

E. Mossel, J. Neeman, and A. Sly. Belief propogation, robust reconstruction, and optimal recovery of block models. arXiv 1309.1380, 2013a.

E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. arXiv 1311.4115, 2013b.

E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *arxiv:1311.4115*, 2013c.

E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. To appear in Probability Theory and Related Fields. The Arxiv version of this paper is titled Stochastic Block Models and Reconstruction, 2015a.

E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for the planted bisection model. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 69–75, New York, NY, USA, 2015b. ACM.

W. Perry and A. S. Wein. A semidefinite program for unbalanced multisection in the stochastic block model. *arXiv:1507.05605v1*, 2015.

T. Richardson and R. Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008. ISBN 9780511791338. Cambridge Books Online.

T. A. B. Snijders and K. Nowicki. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14(1):75–100, 1997.

S. Yun and A. Proutiere. Accurate community detection in the stochastic block model via spectral algorithms. arXiv 1412.7335, 2014a.

S. Yun and A. Proutiere. Community detection via random and adaptive sampling. In *Proceedings of The 27th Conference on Learning Theory*, 2014b.

A. Y. Zhang and H. H. Zhou. Minimax rates of community detection in stochastic block models. *arXiv:1507.05313*, 2015.

P. Zhang, F. Krzakala, J. Reichardt, and L. Zdeborová. Comparitive study for inference of hidden classes in stochastic block models. *Journal of Statistical Mechanics : Theory and Experiment*, 2012.

P. Zhang, C. Moore, and M. E. J. Newman. Community detection in networks with unequal groups. *arXiv:1509.00107*, Sep. 2015.

## Appendix A. Additional Proofs

### A.1. Proof of Lemma 8

By definition, $\Lambda_i^0 = +\infty$ if $\tau_i = +$ and $\Lambda_i^0 = -\infty$ if $\tau_i = -$, and $\Gamma_i^0 = 0$ for all $i$. We prove the claim for $\Gamma_i^t$ with $t \geq 1$; the claim for $\Lambda_i^t$ with $t \geq 1$ follows similarly.

A key point is to use the independent splitting property of the Poisson distribution to give an equivalent description of the numbers of children of each type for any vertex in the tree. Instead of separately generating the number of children of each type, we can first generate the total number of children and then independently and randomly select the type of each child. For every vertex $i$ in $T_u$, let $N_i$ denote the total number of its children. If $\tau_i = +$ then $N_i \sim \mathrm{Pois}(d_+)$, and for each child $j \in \partial i$, independently of everything else, $\tau_j = +$ with probability $\rho a/d_+$ and $\tau_j = -$ with probability $\bar\rho b/d_+$, where $d_+ = \rho a + \bar\rho b$. If $\tau_i = -$ then $N_i \sim \mathrm{Pois}(d_-)$, and for each child $j \in \partial i$, independently of everything else, $\tau_j = +$ with probability $\rho b/d_-$ and $\tau_j = +$ with probability $\bar\rho c/d_-$, where $d_- = \rho b + \bar\rho c$. With this view, the observation of the total number of children $N_i$ of vertex $i$ gives some information, and then the conditionally independent messages from those children give additional information on $\tau_i$. Specifically,

$$
\begin{aligned}
\Gamma_i^{t+1} &= \frac{1}{2}\log\frac{\mathbb{P}\left\{T_i^{t+1}|\tau_i=+\right\}}{\mathbb{P}\left\{T_i^{t+1}|\tau_i=-\right\}} \overset{(a)}{=} \frac{1}{2}\log\frac{\mathbb{P}\left\{N_i|\tau_i=+\right\}}{\mathbb{P}\left\{N_i|\tau_i=-\right\}} + \frac{1}{2}\sum_{j\in\partial i}\log\frac{\mathbb{P}\left\{T_j^t|\tau_i=+\right\}}{\mathbb{P}\left\{T_j^t|\tau_i=-\right\}} \\[2mm]
&\overset{(b)}{=} \frac{-d_+ + d_-}{2} + \frac{1}{2}N_i\log\frac{d_+}{d_-} + \frac{1}{2}\sum_{j\in\partial i}\log\frac{\sum_{x\in\{\pm\}}\mathbb{P}\left\{\tau_j=x|\tau_i=+\right\}\mathbb{P}\left\{T_j^t|\tau_j=x\right\}}{\sum_{x\in\{\pm\}}\mathbb{P}\left\{\tau_j=x|\tau_i=-\right\}\mathbb{P}\left\{T_j^t|\tau_j=x\right\}} \\[2mm]
&\overset{(c)}{=} \frac{-d_+ + d_-}{2} + \frac{1}{2}\sum_{j\in\partial i}\log\frac{\rho a\mathbb{P}\left\{T_j^t|\tau_j=+\right\} + \bar\rho b\mathbb{P}\left\{T_j^t|\tau_j=-\right\}}{\rho b\mathbb{P}\left\{T_i^t|\tau_j=+\right\} + \frac{1}{2}\bar\rho c\mathbb{P}\left\{T_i^t|\tau_j=-\right\}} \\[2mm]
&\overset{(d)}{=} \frac{-d_+ + d_-}{2} + \frac{1}{2}\sum_{j\in\partial i}\log\frac{e^{2\Gamma_j^t}\rho a + \bar\rho b}{e^{2\Gamma_j^t}\rho b + \bar\rho c},
\end{aligned}
$$

where $(a)$ holds because $N_i$ and $T_j^t$ for $j \in \partial u$ are independent conditional on $\tau_i$; $(b)$ follows because $N_i \sim \mathrm{Pois}(d_+)$ if $\tau_i = +$ and $N_i \sim \mathrm{Pois}(d_-)$ if $\tau_i = -$, and $T_j^t$ is independent of $\tau_i$ conditional on $\tau_j$; $(c)$ follows from the definition of $T_u$ as $\tau_j \sim 2 * \mathrm{Bern}(\rho a/d_+) - 1$ (resp. $2 * \mathrm{Bern}(\rho b/d_-) - 1$) conditional on $\tau_i = +$ (resp. $-$); $(d)$ follows from the definition of $\Gamma_j^t$.

### A.2. Proof of Lemma 11

We will show that as $n \to \infty$, $p_G^*$ is bounded by $p_{T^t}^*$ from the below for any $t \geq 1$. Before that, we need a key lemma which shows that conditional on $(G^t, \sigma_{\partial G^t})$, $\sigma_u$ is almost independent of the graph structure outside of $G^t$. The proof is similar to that of (Mossel et al., 2015a, Proposition 4.2) which deals with the special case $\rho = 1/2$ and $a = c$. The key challenge here is that when $\rho \neq 1/2$ or $a \neq c$, the overall effect of the non-edges depends on $\sigma$ and some extra care has to be taken (see (23) for details).

**Lemma 18** *For $t = t(n)$ such that $b^t = n^{o(1)}$, there exists a sequence of events $\mathcal{E}_n$ such that $\mathbb{P}\{\mathcal{E}_n\} \to 1$ as $n \to \infty$, and on event $\mathcal{E}_n$,*

$$\mathbb{P}\left\{\sigma_u = x | G^t, \sigma_{\partial G^t}\right\} = (1 + o(1))\mathbb{P}\left\{\sigma_u = x | G, \sigma_{\partial G^t}\right\}, \quad \forall x \in \{\pm\}. \tag{21}$$

*Moreover, on event $\mathcal{E}_n$, $(G^t, \sigma_{G^t}) = (T^t, \tau_{T^t})$ holds.*

**Proof** Recall that $G^t$ is the subgraph of $G$ induced by vertices whose distance from $u$ is at most $t$. Let $A^t$ denote the set of vertices in $G^{t-1}$, $B^t$ denote the set of vertices in $G^t$, and $C^t$ denote the set of vertices in $G$ but not in $G^t$. Then $A^t \cup \partial G^t = B^t$ and $A^t \cup \partial G^t \cup C^t = V$. Define $s_A = \sum_{i \in A^t} \sigma_i$ and $s_C = \sum_{i \in C^t} \sigma_i$. Let

$$\mathcal{E}_n = \{(\sigma_C, G^t) : |s_C| \leq n^{0.6}, |B| \leq n^{0.1}, (G^t, \sigma_{G^t}) = (T^t, \tau_{T^t})\}.$$

By the assumption $b^t = n^{o(1)}$, it follows that $(G^t, \sigma_{G^t}) = (T^t, \tau_{T^t})$ and $|B| = n^{o(1)}$ with probability converging to 1 (see (Mossel et al., 2015a, Proposition 4.2) for a proof). Note that $s_C = 2X - |C|$ for some $X \sim \text{Binom}(|C|, \rho)$. Letting $\alpha_n = n^{0.6}$, in view of the Bernstein inequality,

$$\mathbb{P}\left\{\left|s_C - (2\rho - 1)|C|\right| > \alpha_n\right\} = \mathbb{P}\left\{\left|X - \rho|C|\right| > \alpha_n/2\right\} \leq 2e^{-\frac{\alpha_n^2/4}{|C|/2 + \alpha_n/3}} = o(1),$$

where the last equality holds because $|C| \leq n$ and $\alpha_n/\sqrt{n} \to \infty$. In conclusion, we have that $\mathbb{P}\{\mathcal{E}_n\} \to 1$ as $n \to \infty$.

To prove that (21) holds, it suffices to show that on event $\mathcal{E}_n$,

$$\mathbb{P}\left\{\sigma_u = x | G^t, \sigma_{\partial G^t}\right\} = (1 + o(1))\mathbb{P}\left\{\sigma_u = x | G^t, \sigma_{\partial G^t}, \sigma_C\right\}, \quad \forall x \in \{\pm\}. \tag{22}$$

In particular, on event $\mathcal{E}_n$,

$$\mathbb{P}\left\{\sigma_u = x | G, \sigma_{\partial G^t}\right\} = \sum_{\sigma_C} \mathbb{P}\left\{\sigma_u = x, \sigma_C | G, \sigma_{\partial G^t}\right\}$$

$$= \sum_{\sigma_C} \mathbb{P}\left\{\sigma_C | G, \sigma_{\partial G^t}\right\} \mathbb{P}\left\{\sigma_u = x | G, \sigma_{\partial G^t}, \sigma_C\right\}$$

$$= \sum_{\sigma_C} \mathbb{P}\left\{\sigma_C | G, \sigma_{\partial G^t}\right\} \mathbb{P}\left\{\sigma_u = x | G^t, \sigma_{\partial G^t}, \sigma_C\right\}$$

$$= (1 + o(1))\mathbb{P}\left\{\sigma_u = x | G^t, \sigma_{\partial G^t}\right\},$$

where the third equality holds, because conditional on $(G^t, \sigma_{\partial G^t}, \sigma_C)$, $\sigma_u$ is independent of the graph structure outside of $G^t$; the last equality follows due to (22). Hence, we are left to show the desired (22) holds.

Recall that $G = (V, E)$. For any two sets $U_1, U_2 \subset V$, define

$$\Phi_{U_1, U_2}(G, \sigma) = \prod_{(u,v) \in U_1 \times U_2} \phi_{uv}(G, \sigma),$$

where $(u, v)$ denotes an unordered pair of vertices and

$$\phi_{uv}(G, L, \sigma) = \begin{cases} a/n & \text{if } \sigma_u = \sigma_v = +, (u, v) \in E \\ c/n & \text{if } \sigma_u = \sigma_v = -, (u, v) \in E \\ b/n & \text{if } \sigma_u \neq \sigma_v, (u, v) \in E \\ 1 - a/n & \text{if } \sigma_u = \sigma_v = +, (u, v) \notin E \\ 1 - c/n & \text{if } \sigma_u = \sigma_v = -, (u, v) \notin E \\ 1 - b/n & \text{if } \sigma_u \neq \sigma_v, (u, v) \notin E \end{cases}$$

Then the joint distribution of $\sigma$ and $G$ is given by

$$\mathbb{P}\left\{\sigma, G, \widetilde{\sigma}\right\} = 2^{-n}\Phi_{B,B}\ \Phi_{C,C}\ \Phi_{\partial G^t,C}\ \Phi_{A,C}.$$

Notice that $A$ and $C$ are disconnected. We claim that on event $\mathcal{E}_n$, $\Phi_{A,C}$ only depend on $\sigma_C$ through the $o(1)$ term. In particular, on event $\mathcal{E}_n$,

$$
\begin{aligned}
\Phi_{A,C}(G,\sigma) &= \prod_{(u,v)\in A\times C} \phi_{uv}(G,\sigma) \\
&= \left(1-\frac{a}{n}\right)^{(|A|+s_A)(|C|+s_C)/4}\left(1-\frac{c}{n}\right)^{(|A|-s_A)(|C|-s_C)/4}\left(1-\frac{b}{n}\right)^{(|A||C|-s_As_C)/2} \\
&= (1+o(1))\left(1-\frac{a}{n}\right)^{\rho(|A|+s_A)|C|/2}\left(1-\frac{c}{n}\right)^{\bar{\rho}(|A|-s_A)|C|/2}\left(1-\frac{b}{n}\right)^{(|A||C|-s_A(2\rho-1)|C|)/2} \\
&\triangleq (1+o(1))K(\sigma_A,|C|), \tag{23}
\end{aligned}
$$

where the second equality holds because $u\in A$ and $v\in C$ implies that $(u,v)\notin E$ and thus $\phi_{uv}$ is either $1-a/n$, $1-c/n$, or $1-b/n$, depending on $\sigma_u$ and $\sigma_v$; the third equality holds because $(|A|+|s_A|)|s_C - (2\rho-1)|C|| \le 2\alpha_n|B| = o(n)$; the last equality holds for some $K(\sigma_A,|C|)$ which only depends on $\sigma_A$ and $|C|$. As a consequence,

$$\mathbb{P}\left\{\sigma, G, \mathcal{E}_n\right\} = (1+o(1))2^{-n}\ K(\sigma_A,,|C|)\ \Phi_{B,B}\ \Phi_{C,C}\ \Phi_{\partial G^t,C}\ .$$

It follows that

$$\mathbb{P}\left\{\sigma_u = x, G^t, \sigma_{\partial G^t}, \sigma_C, \mathcal{E}_n\right\} = (1+o(1))2^{-n}\sum_{\sigma_{A\setminus\{u\}}} K(\sigma_A,|C|)\ \Phi_{B,B}$$

and thus

$$
\begin{aligned}
\mathbb{P}\left\{\sigma_u = x, G^t, \sigma_{\partial G^t}, \mathcal{E}_n\right\} &= \sum_{\sigma_C}\mathbb{P}\left\{\sigma_u = x, G^t, \sigma_{\partial G^t}, \sigma_C, \mathcal{E}_n\right\} \\
&= (1+o(1))2^{-n}\sum_{\sigma_{A\setminus\{u\}}} K(\sigma_A,|C|)\ \Phi_{B,B}\sum_{\sigma_C}\mathbf{1}_{\{|s_C|\le n^{0.6}\}}.
\end{aligned}
$$

By Bayes' rule,

$$
\begin{aligned}
\mathbb{P}\left\{\sigma_u = x|G^t, \sigma_{\partial G^t}, \mathcal{E}_n\right\} &= \frac{\mathbb{P}\left\{\sigma_u = x, G^t, \sigma_{\partial G^t}, \mathcal{E}_n\right\}}{\mathbb{P}\left\{G^t, \sigma_{\partial G^t}, \mathcal{E}_n\right\}} \\
&= (1+o(1))\frac{\sum_{\sigma_{A\setminus\{u\}}} K(\sigma_A,|C|)\ \Phi_{B,B}}{\sum_{\sigma_A} K(\sigma_A,|C|)\ \Phi_{B,B}} \\
&= (1+o(1))\frac{\mathbb{P}\left\{\sigma_u = x, G^t, \sigma_{\partial G^t}, \sigma_C, \mathcal{E}_n\right\}}{\mathbb{P}\left\{G^t, \sigma_{\partial G^t}, \sigma_C, \mathcal{E}_n\right\}} \\
&= (1+o(1))\mathbb{P}\left\{\sigma_u = x|G^t, \sigma_{\partial G^t}, \sigma_C, \mathcal{E}_n\right\}.
\end{aligned}
$$

Hence, the desired (22) follows on event $\mathcal{E}_n$. ∎

**Proof** [Proof of Lemma 11] In view of the definition of $p_G^*$ given in (3),

$$p_G^* = \frac{1}{2} - \frac{1}{2} \left| \mathbb{P}\left\{\sigma_u = +|G\right\} - \mathbb{P}\left\{\sigma_u = -|G\right\} \right|.$$

Consider estimating $\sigma_u$ based on $G$. For any $t \in \mathbb{N}$, suppose a genie reveals the labels of all vertices whose distance from $u$ is precisely $t$, and let $\widehat{\sigma}_{\mathrm{Oracle},t}$ denote the optimal oracle estimator given by

$$\widehat{\sigma}_{\mathrm{Oracle},t}(u) = 2 \times \mathbf{1}_{\left\{\mathbb{P}\left\{\sigma_u = +|G, \sigma_{\partial G^t}\right\} \geq \mathbb{P}\left\{\sigma_u = -|G, \sigma_{\partial G^t}\right\}\right\}} - 1.$$

Let $p_G(\widehat{\sigma}_{\mathrm{Oracle},t})$ denote the error probability of the oracle estimator, which is given by

$$p_G(\widehat{\sigma}_{\mathrm{Oracle},t}) = \frac{1}{2} - \frac{1}{2} \left| \mathbb{P}\left\{\sigma_u = +|G, \sigma_{\partial G^t}\right\} - \mathbb{P}\left\{\sigma_u = -|G, \sigma_{\partial G^t}\right\} \right|$$

Since $\widehat{\sigma}_{\mathrm{Oracle},t}(u)$ is optimal with the extra information $\sigma_{\partial G^t}$, it follows that $p_G(\widehat{\sigma}_{\mathrm{Oracle},t}) \leq p_G^*$ for all $t$ and $n$. Lemma 18 implies that there exists a sequence of events $\mathcal{E}_n$ such that $\mathbb{P}\left\{\mathcal{E}_n \to 1\right\}$ and on event $\mathcal{E}_n$,

$$\mathbb{P}\left\{\sigma_u = x|G, \sigma_{\partial G^t}\right\} = (1 + o(1))\mathbb{P}\left\{\sigma_u = x|G^t, \sigma_{\partial G^t}\right\}, \quad \forall x \in \{\pm\},$$

and $(G^t, \sigma_{G^t}) = (T^t, \tau_{T^t})$. It follows that

$$p_G(\widehat{\sigma}_{\mathrm{Oracle},t}) = \frac{1}{2} - \frac{1}{2}\mathbb{E}\left[\left|\mathbb{P}\left\{\sigma_u = +|G, \sigma_{\partial G^t}\right\} - \mathbb{P}\left\{\sigma_u = -|G, \sigma_{\partial G^t}\right\}\right| \mathbf{1}_{\{\mathcal{E}_n\}}\right] + o(1)$$

$$= \frac{1}{2} - \frac{1}{2}\mathbb{E}\left[\left|\mathbb{P}\left\{\sigma_u = +|G^t, \sigma_{\partial G^t}\right\} - \mathbb{P}\left\{\sigma_u = -|G^t, \sigma_{\partial G^t}\right\}\right| \mathbf{1}_{\{\mathcal{E}_n\}}\right] + o(1)$$

$$= \frac{1}{2} - \frac{1}{2}\mathbb{E}\left[\left|\mathbb{P}\left\{\tau_u = +|T^t, \tau_{\partial T^t}\right\} - \mathbb{P}\left\{\tau_u = -|T^t, \tau_{\partial T^t}\right\}\right| \mathbf{1}_{\{\mathcal{E}_n\}}\right] + o(1)$$

$$= \frac{1}{2} - \frac{1}{2}\mathbb{E}\left[\left|\mathbb{P}\left\{\tau_u = +|T^t, \tau_{\partial T^t}\right\} - \mathbb{P}\left\{\tau_u = -|T^t, \tau_{\partial T^t}\right\}\right|\right] + o(1)$$

$$= p_{T^t}^* + o(1).$$

Hence,

$$\limsup_{n\to\infty}\left(p_G^* - p_{T^t}^*\right) \geq \limsup_{n\to\infty}\left(p_G(\widehat{\sigma}_{\mathrm{Oracle},t}) - p_{T^t}^*\right) = 0.$$

$\blacksquare$

## A.3. Proof of Lemma 12

We first prove the claims for $Z_-^{t+1}$. By the definition of $\Gamma_u^t$ and the change of measure, we have

$$\mathbb{E}\left[g(\Gamma_u^t)|\tau_u = -\right] = \mathbb{E}\left[g(\Gamma_u^t)\mathrm{e}^{-2\Gamma_u^t}|\tau_u = +\right],$$

where $g$ is any measurable function such that the expectations above are well-defined. It follows that

$$\mathbb{E}\left[g(Z_-^t)\right] = \mathbb{E}\left[g(Z_+^t)\mathrm{e}^{-2Z_+^t}\right]. \tag{24}$$

Define $\psi(x) \triangleq \log(1 + x) - x + x^2/2$. It follows from the Taylor expansion that $|\psi(x)| \leq |x|^3$. Then

$$
\begin{aligned}
F(x) &= \frac{1}{2} \log \left( \frac{\mathrm{e}^{2x} \rho a + \bar{\rho} b}{\mathrm{e}^{2x} \rho b + \bar{\rho} c} \right) = \frac{1}{2} \log \frac{b}{c} + \frac{1}{2} \log \left( \frac{\mathrm{e}^{2x} (\rho a)/(\bar{\rho} b) + 1}{\mathrm{e}^{2x} (\rho b)/(\bar{\rho} c) + 1} \right) \\
&= \frac{1}{2} \log \frac{b}{c} + \frac{1}{2} \log \left( 1 + \frac{\mathrm{e}^{4\beta} - 1}{1 + \mathrm{e}^{-2x} (\bar{\rho} c)/(\rho b)} \right) \\
&= \frac{1}{2} \log \frac{b}{c} + \frac{\left( \mathrm{e}^{4\beta} - 1 \right)}{2} f(x) - \frac{\left( \mathrm{e}^{4\beta} - 1 \right)^2}{4} f^2(x) + \frac{1}{2} \psi \left( (\mathrm{e}^{4\beta} - 1) f(x) \right),
\end{aligned}
$$

where $\beta = \frac{1}{2} \log \frac{\sqrt{ac}}{b}$ and $f(x) = \frac{1}{1 + \mathrm{e}^{-2x} (\bar{\rho} c)/(\rho b)}$. Since $|\psi(x)| \leq |x|^3$ and $|f(x)| \leq 1$, it follows that

$$
F(x) = \frac{1}{2} \log \frac{b}{c} + \frac{\left( \mathrm{e}^{4\beta} - 1 \right)}{2} f(x) - \frac{\left( \mathrm{e}^{4\beta} - 1 \right)^2}{4} f^2(x) + O \left( |\mathrm{e}^{4\beta} - 1|^3 \right), \tag{25}
$$

Therefore, in view of (11),

$$
\begin{aligned}
\Gamma_u^{t+1} &= \frac{-d_+ + d_-}{2} + \sum_{\ell \in \partial u} F(\Gamma_\ell^t) \\
&= \frac{-d_+ + d_-}{2} + \frac{1}{2} \sum_{\ell \in \partial u} \left[ \log \frac{b}{c} + \left( \mathrm{e}^{4\beta} - 1 \right) f(\Gamma_\ell^t) - \frac{\left( \mathrm{e}^{4\beta} - 1 \right)^2}{2} f^2(\Gamma_\ell^t) + O \left( |\mathrm{e}^{4\beta} - 1|^3 \right) \right].
\end{aligned}
$$

By conditioning the label of vertex $u$ is $-$, it follows that

$$
\begin{aligned}
\mathbb{E} \left[ Z_-^{t+1} \right] &= \frac{-d_+ + d_-}{2} + \frac{1}{2} \log(b/c) d_- + \frac{\left( \mathrm{e}^{4\beta} - 1 \right)}{2} \left( \rho b \mathbb{E} \left[ f(Z_+^t) \right] + \bar{\rho} c \mathbb{E} \left[ f(Z_-^t) \right] \right) \\
&\quad - \frac{\left( \mathrm{e}^{4\beta} - 1 \right)^2}{4} \left( \rho b \mathbb{E} \left[ f^2(Z_+^t) \right] + \bar{\rho} c \mathbb{E} \left[ f^2(Z_-^t) \right] \right) + O \left( b |\mathrm{e}^{4\beta} - 1|^3 \right).
\end{aligned}
$$

In view of (24), we have that

$$
\rho b \mathbb{E} \left[ f(Z_+^t) \right] + \bar{\rho} c \mathbb{E} \left[ f(Z_-^t) \right] = \rho b \mathbb{E} \left[ f(Z_+^t)(1 + \mathrm{e}^{-2Z_+^t} (\bar{\rho} c)/(\rho b)) \right] = \rho b, \tag{26}
$$

$$
\rho b \mathbb{E} \left[ f^2(Z_+^t) \right] + \bar{\rho} c \mathbb{E} \left[ f^2(Z_-^t)) \right] = \rho b \mathbb{E} \left[ f^2(Z_+^t)(1 + \mathrm{e}^{-2Z_+^t} (\bar{\rho} c)/(\rho b)) \right] = \rho b \mathbb{E} \left[ f(Z_+^t) \right]. \tag{27}
$$

Hence,

$$
\mathbb{E} \left[ Z_-^{t+1} \right] = \frac{-d_+ + d_-}{2} + \frac{1}{2} \log(b/c) d_- + \frac{\left( \mathrm{e}^{4\beta} - 1 \right)}{2} \rho b - \frac{\left( \mathrm{e}^{4\beta} - 1 \right)^2 \rho b}{4} \mathbb{E} \left[ f(Z_+^t) \right] + O \left( b |\mathrm{e}^{4\beta} - 1|^3 \right).
$$

Notice that

$$
\log \frac{b}{c} = - \log \left( 1 + \frac{c - b}{b} \right) = \frac{b - c}{b} + \frac{(b - c)^2}{2b^2} + O \left( \frac{|b - c|^3}{b^3} \right). \tag{28}
$$

As a consequence,

$$- d_+ + d_- + \log(b/c)d_- + \left(e^{4\beta} - 1\right)\rho b$$

$$= -\rho a - \bar{\rho}b + b + \bar{\rho}(c - b) + \log(b/c)(b + \bar{\rho}(c - b)) + \rho\frac{ac - b^2}{b}$$

$$= -\rho a - \bar{\rho}b + b + \bar{\rho}(c - b) + (b - c) - \frac{\bar{\rho}(b - c)^2}{b} + \frac{(b - c)^2}{2b} + \rho\frac{ac - b^2}{b} + O\left(\frac{|b - c|^3}{b^2}\right)$$

$$= \rho(-a + b - c + ac/b) + (1/2 - \bar{\rho})(b - c)^2/b + O\left(\frac{|b - c|^3}{b^2}\right)$$

$$= \rho(a - b)(c - b)/b + (\rho - 1/2)(b - c)^2/b + O\left(\frac{|b - c|^3}{b^2}\right)$$

$$= \rho\mu\nu + (\rho - 1/2)\nu^2 + O(b^{-1/2}),$$

where the last equality holds due to $(a - b)/\sqrt{b} = \mu$ and $(c - b)/\sqrt{b} = \nu$ for fixed constants $\mu$ and $\nu$. Moreover,

$$\left(e^{4\beta} - 1\right)^2 \rho b = \frac{\rho(ac - b^2)^2}{b^3} = \rho\left(\frac{a - b}{\sqrt{b}} + \frac{c - b}{\sqrt{b}} + \frac{(a - b)(c - b)}{b\sqrt{b}}\right)^2 = \rho(\mu + \nu)^2) + O(b^{-1/2}),$$

and $b|e^{4\beta} - 1|^3 = O(b^{-1/2})$. Assembling the last four displayed equations gives that

$$\mathbb{E}\left[Z_-^{t+1}\right] = \frac{1}{2}\rho\mu\nu + \frac{(2\rho - 1)\nu^2}{4} - \frac{\rho(\mu + \nu)^2}{4}\mathbb{E}\left[f(Z_+^t)\right] + O(b^{-1/2}).$$

Finally, recall that $\varphi = \frac{1}{2}\log\frac{\rho}{1-\rho}$ and thus

$$\left|f(x) - \frac{1}{1 + e^{-2(x+\varphi)}}\right| = \frac{e^{-x}(\bar{\rho}/\rho)|1 - c/b|}{(1 + e^{-x}(\bar{\rho}c)/(\rho b))(1 + e^{-x}(\bar{\rho}/\rho))} \le |1 - c/b| = O(b^{-1/2}). \quad (29)$$

It follows that

$$\mathbb{E}\left[Z_-^{t+1}\right] = \frac{1}{2}\rho\mu\nu + \frac{(2\rho - 1)\nu^2}{4} - \frac{\rho(\mu + \nu)^2}{4}\mathbb{E}\left[\frac{1}{1 + e^{-2(Z_+^t + \varphi)}}\right] + O(b^{-1/2})$$

$$= -\frac{\rho(\mu - \nu)^2}{8} + \frac{(2\rho - 1)\nu^2}{4} - \frac{\rho(\mu + \nu)^2}{8}\mathbb{E}\left[\tanh(Z_+^t + \varphi)\right] + O(b^{-1/2}),$$

where in the last equality we used the fact that $\frac{1}{1+e^{-x}} = \frac{1}{2}(\tanh(x) + 1)$. Recall that $\lambda = \frac{\rho(\mu+\nu)^2}{8}$ and $\theta = \frac{\rho(\mu-\nu)^2}{8} + \frac{(1-2\rho)\nu^2}{4}$. Therefore, we get the desired equality:

$$\mathbb{E}\left[Z_-^{t+1}\right] = -\theta - \lambda\mathbb{E}\left[\tanh(Z_+^t + \varphi)\right] + O(b^{-1/2}).$$

Next we calculate $\text{var}(Z_-^{t+1})$. For $Y = \sum_{i=1}^L X_i$, where $L$ is Poisson distributed, and $\{X_i\}$ are i.i.d. with finite second moments, one can check that $\text{var}(Y) = \mathbb{E}[L]\mathbb{E}\left[X_1^2\right]$. In view of (11),

$$\text{var}(Z_-^{t+1}) = \rho b\mathbb{E}\left[F^2(Z_+^t)\right] + \bar{\rho}c\mathbb{E}\left[F^2(Z_-^t)\right],$$

In view of (25) and the fact that $e^{4\beta} - 1 = o(1)$, we have that

$$F^2(x) = \frac{1}{4}\log^2\left(\frac{b}{c}\right) + \frac{1}{2}\log\frac{b}{c}\left(e^{4\beta} - 1\right)f(x) + \frac{1}{4}\left(e^{4\beta} - 1\right)^2(1 - \log(b/c))f^2(x) + O\left(|e^{4\beta} - 1|^3\right),$$

Thus,

$$\text{var}(Z_-^{t+1}) = \frac{1}{4}\log^2\left(\frac{b}{c}\right)d_- + \frac{1}{2}\log\frac{b}{c}\left(e^{4\beta} - 1\right)\left[\rho b\mathbb{E}\left[f(Z_+^t)\right] + \bar{\rho}c\mathbb{E}\left[f(Z_-^t)\right]\right]$$
$$+ \frac{1}{4}\left(e^{4\beta} - 1\right)^2(1 - \log(b/c))\left[\rho b\mathbb{E}\left[f^2(Z_+^t)\right] + \bar{\rho}c\mathbb{E}\left[f^2(Z_-^t)\right]\right] + O\left(b|e^{4\beta} - 1|^3\right)$$

Applying (26) and (27), we get that

$$\text{var}(Z_-^{t+1}) = \frac{1}{4}\log^2\left(\frac{b}{c}\right)d_- + \frac{\rho b\left(e^{4\beta} - 1\right)}{2}\log\frac{b}{c} + \frac{\rho b\left(e^{4\beta} - 1\right)^2}{4}\left(1 - \log\frac{b}{c}\right)\mathbb{E}\left[f(Z_+^t)\right]$$
$$+ O\left(b|e^{4\beta} - 1|^3\right).$$

In view of (28), we have that

$$\log^2(b/c)d_- + 2\rho b\left(e^{4\beta} - 1\right)\log(b/c) = \log^2(b/c)(b + \bar{\rho}(c - b)) + 2\rho\frac{ac - b^2}{b}\log(b/c)$$
$$= \frac{(b - c)^2}{b} + 2\rho(b - c)\frac{ac - b^2}{b^2} + O\left(\frac{|b - c|^3}{b^2}\right)$$
$$= \nu^2 - 2\rho\nu(\mu + \nu) + O(b^{-1/2}),$$

and that

$$\rho b\left(e^{4\beta} - 1\right)^2(1 - \log(b/c)) = \frac{\rho(ac - b^2)^2}{b^3}(1 - \log(b/c)) = \rho(\mu + \nu)^2 + o(1).$$

Moreover, we have shown that $b|e^{4\beta} - 1|^3 = O(b^{-1/2})$. Assembling the last three displayed equations give that

$$\text{var}(Z_-^{t+1}) = \frac{\nu^2}{4} - \frac{\rho\nu(\mu + \nu)}{2} + \frac{\rho(\mu + \nu)^2}{4}\mathbb{E}\left[f(Z_+^t)\right] + O(b^{-1/2}).$$

Finally, in view of (29), we get that

$$\text{var}(Z_-^{t+1}) = \frac{\nu^2}{4} - \frac{\rho\nu(\mu + \nu)}{2} + \frac{\rho(\mu + \nu)^2}{4}\mathbb{E}\left[\frac{1}{1 + e^{-2(Z_+^t + \varphi)}}\right] + O(b^{-1/2})$$
$$= \frac{\rho(\mu - \nu)^2}{8} + \frac{(1 - 2\rho)\nu^2}{4} + \frac{\rho(\mu + \nu)^2}{8}\mathbb{E}\left[\tanh(Z_+^t + \varphi)\right] + O(b^{-1/2})$$
$$= \theta + \lambda\mathbb{E}\left[\tanh(Z_+^t + \varphi)\right] + O(b^{-1/2}).$$

The claims for $Z_+^{t+1}$ can be proved similarly as above. We provide another proof by exploiting the symmetry. In particular, note that our tree model is parameterized by $(\rho, a, b, c)$ with labels $+$ and $-$. Consider another parametrization $(\rho', a', b', c')$ with labels $+'$ and $-'$, where $\rho' = \bar{\rho}$, $a' = c$,

$b' = b$, $c' = a$, $+' = -$, $-' = +$. Let $\widetilde{Z}^t_{+'}$ and $\widetilde{Z}^t_{-'}$ denote the random variables corresponding to $Z^t_+$ and $Z^t_-$, respectively. Then, one can check that $\widetilde{Z}^t_{+'}$ has the same distribution as $-Z^t_-$ and $\widetilde{Z}^t_{-'}$ has the same distribution as $-Z^t_+$. We have shown that

$$\mathbb{E}\left[\widetilde{Z}^{t+1}_{-'}\right] = -\frac{\rho'(\mu'-\nu')^2}{8} + \frac{(2\rho'-1)(\nu')^2}{4} - \frac{\rho'(\mu'+\nu')^2}{8}\mathbb{E}\left[\tanh(\widetilde{Z}^t_{+'}+\varphi')\right] + O(b^{-1/2}),$$

where $\mu' = \frac{a'-b'}{b'} = \frac{c-b}{b} = \nu$ and similarly $\nu' = \mu$, and $\varphi' = -\varphi = \frac{1}{2}\log\frac{1-\rho}{\rho}$. It follows that

$$\mathbb{E}\left[Z^{t+1}_+\right] = \frac{\bar{\rho}(\mu-\nu)^2}{8} - \frac{(2\bar{\rho}-1)\mu^2}{4} - \frac{\bar{\rho}(\mu+\nu)^2}{8}\mathbb{E}\left[\tanh(Z^t_- + \varphi)\right] + O(b^{-1/2}), \qquad (30)$$

Applying $g(x) = \tanh(x-\varphi)$ into (24), we get that

$$\mathbb{E}\left[\tanh(Z^t_+ + \varphi)\right] + \mathrm{e}^{-2\varphi}\mathbb{E}\left[\tanh(Z^t_- + \varphi)\right] = \mathbb{E}\left[\tanh(Z^t_+ + \varphi)\left(1 + \mathrm{e}^{-2(Z^t_+ + \varphi)}\right)\right]$$

$$= \mathbb{E}\left[1 - \mathrm{e}^{-2(Z^t_+ + \varphi)}\right] = 1 - \mathrm{e}^{-2\varphi},$$

where the last equality by the change of measure: $\mathbb{E}\left[\mathrm{e}^{-Z^t_+}\right] = 1$. Hence,

$$\mathbb{E}\left[\tanh(Z^t_- + \varphi)\right] = \frac{1 - \mathrm{e}^{-2\varphi} - \mathbb{E}\left[\tanh(Z^t_+ + \varphi)\right]}{\mathrm{e}^{-2\varphi}} = \frac{\rho - \bar{\rho} - \rho\mathbb{E}\left[\tanh(Z^t_+ + \varphi)\right]}{1-\rho}.$$

It follows from (30) that

$$\mathbb{E}\left[Z^{t+1}_+\right] = \frac{\bar{\rho}(\mu-\nu)^2}{8} - \frac{(2\bar{\rho}-1)\mu^2}{4} - \frac{(\mu+\nu)^2(\rho-\bar{\rho})}{8} + \frac{\rho(\mu+\nu)^2}{8}\mathbb{E}\left[\tanh(Z^t_+ + \varphi)\right] + O(b^{-1/2})$$

$$= \frac{\rho(\mu-\nu)^2}{8} + \frac{(1-2\rho)\nu^2}{4} + \frac{\rho(\mu+\nu)^2}{8}\mathbb{E}\left[\tanh(Z^t_+ + \varphi)\right] + O(b^{-1/2})$$

$$= \theta + \lambda\mathbb{E}\left[\tanh(Z^t_+ + \varphi)\right] + O(b^{-1/2}).$$

Finally, note that

$$\mathsf{var}(Z^{t+1}_+) = \mathsf{var}(\widetilde{Z}^{t+1}_{-'}) = -\mathbb{E}\left[\widetilde{Z}^{t+1}_{-'}\right] + O(b^{-1/2}) = \mathbb{E}\left[Z^{t+1}_+\right] + O(b^{-1/2}).$$

Combing the last two displayed equations completes the proof.

### A.4. Proof of Lemma 13

The following lemma is useful for proving the distributions of $Z^t_+$ and $Z^t_-$ are approximately Gaussian.

**Lemma 19** (*Analog of Berry-Esseen inequality for Poisson sums (Korolev and Shevtsova, 2012, Theorem 3).*) *Let $S_d = X_1 + \cdots + X_{N_d}$, where $X_i : i \geq 1$ are independent, identically distributed random variables with finite second moment, and for some $d > 0$, $N_d$ is a $\mathrm{Pois}(d)$ random variable independent of $(X_i : i \geq 1)$. Then*

$$\sup_x \left|\mathbb{P}\left\{\frac{S_d - \mathbb{E}[S_d]}{\sqrt{\mathsf{var}(S_d)}} \leq x\right\} - \mathbb{P}\{Z \leq x\}\right| \leq \frac{C_{BE}\mathbb{E}\left[|X_1|^3\right]}{\sqrt{d(\mathbb{E}[X_1^2])^3}},$$

*where $\mathbb{E}[S_d] = d\mathbb{E}[X_1]$, $\mathsf{var}(S_d) = d\mathbb{E}[X_1^2]$, and $C_{BE} = 0.3041$.*

**Proof** [Proof of Lemma 13] We prove the lemma by induction over $t$. We first consider the base case. For $Z^t$, the base case $t = 0$ trivially holds, because $\Gamma_u^0 \equiv 0$ and $v_0 = 0$. For $W^t$, we need to check the base case $t = 1$. Recall that $\Lambda_\ell^0 = \infty$ if $\tau_\ell = +$ and $\Lambda_\ell^0 = -\infty$ if $\tau_\ell = -$. Notice that $F(\infty) = \frac{1}{2}\log(a/b)$ and $F(-\infty) = \frac{1}{2}\log(b/c)$. Hence,

$$\Lambda_u^1 = \frac{-d_+ + d_-}{2} + \sum_{i=1}^{N_d} X_i, \tag{31}$$

where conditional on $\tau_u = \pm$, $N_d \sim \text{Pois}(d_\pm)$ is independent of $\{X_i\}$; $\{X_i\}$ are i.i.d. such that conditional on $\tau_u = +$, $X_i = \frac{1}{2}\log(a/b)$ with probability $(\rho a)/d_+$ and $X_i = \frac{1}{2}\log(b/c)$ with probability $(\bar{\rho} b)/d_+$; conditional on $\tau_u = -$, $X_i = \frac{1}{2}\log(a/b)$ with probability $(\rho b)/d_-$ and $X_i = \frac{1}{2}\log(b/c)$ with probability $(\bar{\rho} c)/d_-$. Taylor expansion yields that

$$\log(a/b) = \log\left(1 + \frac{a-b}{b}\right) = \frac{a-b}{b} - \frac{\mu^2}{2b} + O\left(b^{-3/2}\right)$$

$$\log(b/c) = -\log\left(1 + \frac{c-b}{b}\right) = \frac{b-c}{b} + \frac{\nu^2}{2b} + O\left(b^{-3/2}\right),$$

Since $F$ is monotone,

$$\mathbb{E}\left[X_1^2\right] \geq \min\{|\log(a/b)|^2, |\log(b/c)|^2\} = \Omega\left(\min\left\{\frac{(a-b)^2}{b^2}, \frac{(c-b)^2}{b^2}\right\}\right) = \Omega(b^{-1})$$

$$\mathbb{E}\left[|X_1|^3\right] \leq \max\{|\log(a/b)|^3, |\log(b/c)|^3\} = O\left(\frac{|a-b|^3 + |b-c|^3}{b^3}\right) = O(b^{-3/2}).$$

Thus, in view of Lemma 19, we get that

$$\sup_x \left|\mathbb{P}\left\{\frac{W_\pm^1 - \mathbb{E}\left[W_\pm^1\right]}{\sqrt{\text{var}\left(W_\pm^1\right)}} \leq x\right\} - \mathbb{P}\{Z \leq x\}\right| \leq O(b^{-1/2}).$$

By conditioning the label of $u$ is $-$, it follows from (31) that

$$\mathbb{E}\left[W_-^1\right] = \frac{1}{2}\left[-d_+ + d_- + \log(a/b)\rho b + \log(b/c)\bar{\rho} c\right] = -\frac{\rho\mu^2 + \bar{\rho}\nu^2}{4} + O(b^{-1/2}) = -w_1 + O(b^{-1/2})$$

$$\text{var}\left(W_-^1\right) = \frac{1}{4}\log^2(a/b)\rho b + \frac{1}{4}\log^2(b/c)\bar{\rho} b = \frac{\rho\mu^2 + \bar{\rho}\nu^2}{4} + O(b^{-1/2}) = w_1 + O(b^{-1/2}),$$

where we used the fact that $w_1 = \theta + \lambda = \frac{\rho\mu^2 + \bar{\rho}\nu^2}{4}$ by definition. Similarly, by conditioning the label of $u$ is $+$, it follows that

$$\mathbb{E}\left[W_+^1\right] = \frac{1}{2}\left[-d_+ + d_- + \log(a/b)\rho a + \log(b/c)\bar{\rho} b\right] = \frac{\rho\mu^2 + \bar{\rho}\nu^2}{4} + O(b^{-1/2}) = w_1 + O(b^{-1/2})$$

$$\text{var}\left(W_+^1\right) = \frac{1}{4}\log^2(a/b)\rho a + \frac{1}{4}\log^2(b/c)\bar{\rho} b = \frac{\rho\mu^2 + \bar{\rho}\nu^2}{4} + O(b^{-1/2}) = w_1 + O(b^{-1/2}).$$

Hence, we get the desired equality:

$$\sup_x \left|\mathbb{P}\left\{\frac{W_\pm^1 \mp w_1}{\sqrt{w_1}} \leq x\right\} - \mathbb{P}\{Z \leq x\}\right| \leq O(b^{-1/2}).$$

In view of (10) and (11), $\Lambda^t$ and $\Gamma^t$ satisfy the same recursion. Moreover, by definition, $v_t$ and $w_t$ also satisfy the same recursion. Thus, to finish the proof of the lemma, it suffices to show that: suppose (15) holds for $t$, then it also holds for $t+1$. We prove the claim for $Z_-^{t+1}$; the claim for $Z_+^{t+1}$ follows similarly. In view of the recursion given in (11),

$$Z_-^{t+1} = \frac{-d_+ + d_-}{2} + \sum_{i=1}^{N_d} Y_i,$$

where $N_d \sim \mathrm{Pois}(d)$ is independent of $\{Y_i\}$; $\{Y_i\}$ are i.i.d. such that $Y_i = F(Z_+^t)$ with probability $\rho b/d_-$ and $Y_i = F(Z_-^t)$ with probability $\bar{\rho}c/d_-$. Since $F$ is monotone, $F(\infty) = \log(a/b)$, and $F(-\infty) = \log(b/c)$, it follows that

$$\mathbb{E}\left[Y_1^2\right] \geq \min\{|\log(a/b)|^2, |\log(b/c)|^2\} = \Omega(b^{-1})$$
$$\mathbb{E}\left[|Y_1|^3\right] \leq \max\{|\log(a/b)|^3, |\log(b/c)|^3\} = O(b^{-3/2}).$$

In view of Lemma 19, we get that

$$\sup_x \left| \mathbb{P}\left\{ \frac{Z_-^{t+1} - \mathbb{E}\left[Z_-^{t+1}\right]}{\sqrt{\mathrm{var}\left(Z_-^{t+1}\right)}} \leq x \right\} - \mathbb{P}\left\{Z \leq x\right\} \right| = O(b^{-1/2}). \tag{32}$$

It follows from Lemma 12 that

$$\mathbb{E}\left[Z_-^{t+1}\right] = -\theta - \lambda\mathbb{E}\left[\tanh(Z_+^t + \varphi)\right] + O(b^{-1/2}).$$
$$\mathrm{var}\left(Z_-^{t+1}\right) = \theta + \lambda\mathbb{E}\left[\tanh(Z_+^t + \varphi)\right] + O(b^{-1/2}).$$

Using the area rule of expectation, we have that

$$\mathbb{E}\left[\tanh(Z_+^t + \varphi)\right]$$
$$= \int_0^1 \tanh'(t)\mathbb{P}\left\{Z_+^t + \phi \geq t\right\}\mathrm{d}t - \int_{-1}^0 \tanh'(t)\mathbb{P}\left\{Z_+^t + \varphi \leq t\right\}$$
$$= \int_0^1 \tanh'(t)\mathbb{P}\left\{v_t + \sqrt{v_t}Z + \varphi \geq t\right\}\mathrm{d}t - \int_{-1}^0 \tanh'(t)\mathbb{P}\left\{v_t + \sqrt{v_t}Z + \varphi \leq t\right\} + O(b^{-1/2})$$
$$= \mathbb{E}\left[\tanh(v_t + \sqrt{v_t}Z + \varphi)\right] + O(b^{-1/2}).$$

where the second equality follows from the induction hypothesis and the fact that $|\tanh'(t)| \leq 1$. Recall that $v_{t+1} = \theta + \lambda\mathbb{E}\left[\tanh(v_t + \sqrt{v_t}Z + \varphi)\right]$. Hence, $\mathbb{E}\left[Z_-^{t+1}\right] = -v_{t+1} + O(b^{-1/2})$ and $\mathrm{var}\left(Z_-^{t+1}\right) = v_{t+1} + O(b^{-1/2})$. As a consequence, in view of (32), the desired (15) holds for $Z_-^{t+1}$. ∎

### A.5. Proof of Lemma 14

By definition,

$$h(v) = \mathbb{E}\left[\tanh\left(v + \sqrt{v}Z + \varphi\right)\right].$$

Since $|\tanh(x)| \le 1$, the continuity of $h$ follows from the dominated convergence theorem. We next show $h'(v)$ exists for $v \in (0, \infty)$. Notice that $\tanh'(x + \sqrt{x}Z + \varphi) = (1 - \tanh^2(x + \sqrt{x}Z + \varphi))(1 + x^{-1/2}Z/2)$ for $x \in (0, \infty)$, and

$$\left| \left(1 - \tanh^2(x + \sqrt{x}Z + \varphi)\right)(1 + x^{-1/2}Z/2) \right| \le 1 + x^{-1/2}|Z|/2.$$

Since $|Z|$ is integrable, by the dominated convergence theorem, $\mathbb{E}\left[\tanh'(x + \sqrt{x}Z + \varphi)\right]$ exists and is continuous in $x$ over $(0, \infty)$. Therefore, $x \to \mathbb{E}\left[\tanh'(x + \sqrt{x}Z + \varphi)\right]$ is integrable over $x \in (0, \infty)$. It follows that

$$h(v) = \mathbb{E}\left[\tanh(\varphi) + \int_0^v \tanh'(x + \sqrt{x}Z + \varphi)\mathrm{d}x\right] = \tanh(\varphi) + \int_0^v \mathbb{E}\left[\tanh'(x + \sqrt{x}Z + \varphi)\right]\mathrm{d}x,$$

where the second equality holds due to Fubini's theorem. Hence,

$$h'(v) = \mathbb{E}\left[\left(1 - \tanh^2(v + \sqrt{v}Z + \varphi)\right)(1 + v^{-1/2}Z/2)\right].$$

Using the integration by parts, we can get that

$$\mathbb{E}\left[\left(1 - \tanh^2(v + \sqrt{v}Z + \varphi)\right)\sqrt{v}Z\right]$$
$$= \int_{-\infty}^{\infty} (1 - \tanh^2(v + x + \varphi))\frac{x}{\sqrt{2\pi v}}\mathrm{e}^{-x^2/2v}\mathrm{d}x$$
$$= -v\int_{-\infty}^{\infty} (1 - \tanh^2(v + x + \varphi))\left(\frac{1}{\sqrt{2\pi v}}\mathrm{e}^{-x^2/2v}\right)'\mathrm{d}x$$
$$= -v(1 - \tanh^2(v + x + \varphi))\frac{1}{\sqrt{2\pi v}}\mathrm{e}^{-x^2/2v}\Big|_{-\infty}^{+\infty} + v\int_{-\infty}^{\infty} (1 - \tanh^2(v + x + \varphi))'\frac{1}{\sqrt{2\pi v}}\mathrm{e}^{-x^2/2v}\mathrm{d}x$$
$$= -2v\mathbb{E}\left[\tanh(v + \sqrt{v}Z + \varphi)(1 - \tanh^2(v + \sqrt{v}Z + \varphi))\right].$$

By combing the last two displayed equations, we get (17).

Next, we prove the concavity of $h$ in the special case with $\varphi = 0$. We will use the following equality coming from the change of measure: For $k \in \mathbb{N}$,

$$\mathbb{E}\left[\tanh^{2k}(\sqrt{v}Z + v)\right] = \mathbb{E}\left[\tanh^{2k-1}(\sqrt{v}Z + v)\right].$$

It follows from Lemma 14 that

$$h'(v) = \mathbb{E}\left[\left(1 - \tanh(v + \sqrt{v}Z)\right)\left(1 - \tanh^2(v + \sqrt{v}Z)\right)\right]$$
$$= \mathbb{E}\left[\left(1 - \tanh^2(\sqrt{v}Z + v)\right)^2\right]$$
$$= \mathbb{E}\left[\left(1 - \tanh^2\left(\sqrt{v}\,|Z + \sqrt{v}|\right)\right)^2\right],$$

where the last equality holds because $\tanh^2(x)$ is even in $x$. For $0 < v < w < \infty$ and all $z \in \mathbb{R}$,

$$\tanh^2\left(\sqrt{w}\,|z + \sqrt{v}|\right) \ge \tanh^2\left(\sqrt{v}\,|z + \sqrt{v}|\right).$$

Thus,

$$h'(v) = \mathbb{E}\left[\left(1 - \tanh^2\left(\sqrt{v}\,|Z + \sqrt{v}|\right)\right)^2\right] \ge \mathbb{E}\left[\left(1 - \tanh^2\left(\sqrt{w}\,|Z + \sqrt{v}|\right)\right)^2\right]. \tag{33}$$

Let $X = |Z + \sqrt{v}|$ and $Y = |Z + \sqrt{w}|$. Then for any $x \geq 0$,

$$\mathbb{P}\{X \leq x\} = \mathbb{P}\{-x - \sqrt{v} \leq Z \leq x - \sqrt{v}\} \geq \mathbb{P}\{-x - \sqrt{w} \leq Z \leq x - \sqrt{w}\} = \mathbb{P}\{Y \leq x\}.$$

Hence, $X$ is first-order stochastically dominated by $Y$. Since $\left(1 - \tanh^2\left(\sqrt{w}x\right)\right)^2$ is non-increasing in $x$ for $x \geq 0$, it follows that

$$\mathbb{E}\left[\left(1 - \tanh^2\left(\sqrt{w}X\right)\right)^2\right] \geq \mathbb{E}\left[\left(1 - \tanh^2\left(\sqrt{w}Y\right)\right)^2\right].$$

Thus by (33),

$$h'(v) \geq \mathbb{E}\left[\left(1 - \tanh^2\left(\sqrt{w}\,|Z + \sqrt{w}|\right)\right)^2\right] = h'(w).$$

### A.6. Proof of Lemma 15

The proof is very similar to the proof of Lemma 11; the key new challenge is that $\mathbb{E}\left[\mathcal{O}(\sigma, \widehat{\sigma})\right]$ does not directly reduce to the error probability of estimating $\sigma_u$ based on graph $G$. We need a key lemma.

**Lemma 20** *Fix any $t \geq 1$ and any two different vertices $i$ and $j$. For estimator $\widehat{\sigma}(G)$ of $\sigma$ based on $G$,*

$$\mathbb{E}\left[\left(\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_i\}} - \frac{1}{2}\right)\left(\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_j\}} - \frac{1}{2}\right)\right] \leq \left(\frac{1}{2} - p^*_{T^t}\right)^2 + o(1). \tag{34}$$

**Proof** Fix any $t \geq 1$. Recall that $G_u^t$ denotes the subgraph induced by vertices whose distance from $u$ is at most $t$ and $\partial G_u^t$ denotes the set of vertices whose distance from $u$ is precisely $t$. Let $(T_i, \tau_{T_i})$ and $(T_j, \tau_{T_j})$ denote two independent copies of the tree model with $\rho = 1/2$ and $a = c$ defined in Definition 5. The coupling lemma given in Lemma 9 and Lemma 18 can be strengthened so that there exists a sequence of events $\mathcal{E}_n$ such that $\mathbb{P}\{\mathcal{E}_n\} \to 1$ and on event $\mathcal{E}_n$, $(G_i^t, \sigma_{G_i^t}) = (T_i^t, \tau_{T_i^t})$, $(G_j^t, \sigma_{G_j^t}) = (T_j^t, \tau_{T_j^t})$, and

$$\mathbb{P}\left\{\sigma_i, \sigma_j \middle| G, \sigma_{\partial G_i^t}, \sigma_{\partial G_j^t}\right\} = (1 + o(1))\mathbb{P}\left\{\sigma_i \middle| G, \sigma_{\partial G_i^t}\right\}\mathbb{P}\left\{\sigma_j \middle| G, \sigma_{\partial G_j^t}\right\} \tag{35}$$

$$\mathbb{P}\left\{\sigma_i \middle| G, \sigma_{\partial G_i^t}\right\} = (1 + o(1))\mathbb{P}\left\{\sigma_i \middle| G_i^t, \sigma_{\partial G_i^t}\right\} \tag{36}$$

$$\mathbb{P}\left\{\sigma_j \middle| G, \sigma_{\partial G_j^t}\right\} = (1 + o(1))\mathbb{P}\left\{\sigma_j \middle| G_j^t, \sigma_{\partial G_j^t}\right\}. \tag{37}$$

For $u = i, j$, define

$$X_u = \mathbb{P}\left\{\sigma_u = + \middle| G, \sigma_{\partial G_u^t}\right\} - \mathbb{P}\left\{\sigma_u = - \middle| G, \sigma_{\partial G_u^t}\right\}$$
$$Y_u = \mathbb{P}\left\{\tau_u = + \middle| T_u^t, \tau_{\partial T_u^t}\right\} - \mathbb{P}\left\{\tau_u = - \middle| T_u^t, \tau_{\partial T_u^t}\right\}.$$

Then for any estimator $\widehat{\sigma}(G)$, we have that

$$
\begin{aligned}
&\mathbb{E}\left[\left(\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_i\}} - 1/2\right)\left(\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_j\}} - 1/2\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_i\}} - 1/2\right)\left(\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_j\}} - 1/2\right)\Big| G, \sigma_{\partial G_i^t}, \sigma_{\partial G_j^t}\right]\mathbf{1}_{\{\mathcal{E}_n\}}\right] + o(1) \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_i\}} - 1/2\right)|G, \sigma_{\partial G_i^t}\right]\mathbb{E}\left[\left(\mathbf{1}_{\{\widehat{\sigma}_j \neq \sigma_j\}} - 1/2\right)|G, \sigma_{\partial G_j^t}\right]\mathbf{1}_{\{\mathcal{E}_n\}}\right] + o(1) \\
&\leq \frac{1}{4}\mathbb{E}\left[|X_i||X_j|\mathbf{1}_{\{\mathcal{E}_n\}}\right] + o(1) \\
&= \frac{1}{4}\mathbb{E}\left[|Y_i||Y_j|\mathbf{1}_{\{\mathcal{E}_n\}}\right] + o(1) = \frac{1}{4}\mathbb{E}\left[|Y_i||Y_j|\right] + o(1) \\
&= \frac{1}{4}\mathbb{E}\left[|Y_i|\right]\mathbb{E}\left[|Y_j|\right] + o(1) = (1/2 - p_{T^t}^*)^2 + o(1),
\end{aligned}
$$

where the first and fourth equality follows due to $\mathbb{P}\{\mathcal{E}_n\} \to 1$; the second equality holds due to (35); the first inequality holds due to the fact that $\mathbb{P}\left\{\sigma_i \neq x | G, \sigma_{\partial G_i^t}\right\}$ is maximized at $x = -$ if $X_i \geq 0$ and at $x = +$ if $X_i < 0$; the third inequality holds due to (36), (37), $(G_i^t, \sigma_{G_i^t}) = (T_i^t, \tau_{T_i^t})$, and $(G_j^t, \sigma_{G_j^t}) = (T_j^t, \tau_{T_j^t})$ ; the firth equality follows because $(T_i, \tau_{T_i})$ and $(T_j, \tau_{T_j})$ are independent; the last equality follows because $p_{T^t}^* = 1/2 - \mathbb{E}\left[|Y_u|\right]/2$ by definition. Hence we get the desired (34). ∎

**Proof** [Proof of Lemma 15] Fix any estimator $\widehat{\sigma}(G)$. Notice that by definition of $\mathcal{O}(\sigma, \widehat{\sigma})$,

$$
\mathcal{O}(\sigma, \widehat{\sigma}) = \frac{1}{2} - \left|\frac{1}{n}\sum_{i \in [n]}\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_i\}} - \frac{1}{2}\right|.
$$

Therefore,

$$
\mathbb{E}[\mathcal{O}(\sigma, \widehat{\sigma})] = \frac{1}{2} - \mathbb{E}\left[\left|\frac{1}{n}\sum_{i \in [n]}\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_i\}} - \frac{1}{2}\right|\right] \geq \frac{1}{2} - \mathbb{E}\left[\left(\frac{1}{n}\sum_{i \in [n]}\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_i\}} - \frac{1}{2}\right)^2\right]^{1/2}, \quad (38)
$$

where we used the Cauchy-Schwartz in the last inequality. Furthermore,

$$
\mathbb{E}\left[\left(\frac{1}{n}\sum_{i \in [n]}\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_i\}} - \frac{1}{2}\right)^2\right] = \frac{1}{4n} + \frac{n-1}{n}\mathbb{E}\left[\left(\mathbf{1}_{\{\widehat{\sigma}_1 \neq \sigma_1\}} - \frac{1}{2}\right)\left(\mathbf{1}_{\{\widehat{\sigma}_2 \neq \sigma_2\}} - \frac{1}{2}\right)\right],
$$

where we used the symmetry among vertices. Applying Lemma 20, we get that

$$
\mathbb{E}\left[\left(\frac{1}{n}\sum_{i \in [n]}\mathbf{1}_{\{\widehat{\sigma}_i \neq \sigma_i\}} - \frac{1}{2}\right)^2\right] \leq \left(\frac{1}{2} - p_{T^t}^*\right)^2 + o(1).
$$

Combining the last displayed equation with (38) and noticing that $p_{T^t}^* \leq 1/2$, we get the desired equality $\mathbb{E}[\mathcal{O}(\sigma, \widehat{\sigma})] \geq p_{T^t}^* + o(1)$. Since $\widehat{\sigma}$ is arbitrary, it follows that $\inf_{\widehat{\sigma}} \mathbb{E}[\mathcal{O}(\sigma, \widehat{\sigma})] \geq p_{T^t}^* + o(1)$ and the proof is complete. ∎

### A.7. Proof of Lemma 16

Before the main proof, we need a key lemma, which gives a recursive formula of $\widetilde{\Gamma}_i^t$ on the tree model. Its proof is almost identical to the proof of Lemma 8 and thus omitted.

**Lemma 21** *For $t \geq 0$,*

$$\widetilde{\Gamma}_i^{t+1} = \frac{1}{2} \sum_{j \in \partial i} \log \frac{\exp\left(2\widetilde{\Gamma}_j^t\right) a + b}{\exp\left(2\widetilde{\Gamma}_j^t\right) b + a}. \tag{39}$$

*with $\widetilde{\Gamma}_i^0 = \frac{1}{2} \log \frac{1-\alpha}{\alpha}$ if $\widetilde{\tau}_i = +$ and $\widetilde{\Gamma}_i^0 = -\frac{1}{2} \log \frac{1-\alpha}{\alpha}$ if $\widetilde{\tau}_i = -$.*

Let $V^+ = \{i \in V : \sigma_i = +\}$ and $V^- = \{i \in V : \sigma_i = -\}$. For an Erdős-Rényi random graph with edge probability $b/n$, it is well-known that if $b \to \infty$ and $b = o(\log n)$, the maximum degree is at least $\frac{\log n}{\log(\log n/b)}$ with high probability (see (Hajek et al., 2014, Appendix A) for a proof). Thus, with high probability, at least one vertex in $U$ has more than $\frac{\log n}{2\log(\log n/b)}$ neighbors in $V \backslash U$, so that $u_*$ is well-defined. Due to the symmetry between $+$ and $-$, without loss of generality, assume that $\sigma_{u_*} = +$. By the assumption that $|\mu| > 2$ and $b = n^{o(1/\log\log n)}$, it is proved in (Mossel et al., 2013a, Lemma 5.7) that there exists an $\alpha \in (0, 1/2)$ and a polynomial-time estimator such that for any $u \in V \backslash U$, when we apply the estimator in Step 3.1 of Algorithm 2, its output satisfies $|W_u^+ \Delta V^+| \leq \alpha n$ and $|W_u^- \Delta V^-| \leq \alpha n$ after relabeling defined in Step 3.2 of Algorithm 2. Recall that $\alpha_u$ is the fraction of vertices misclassified by the partition $(W_u^+, W_u^-)$. Thus, $\alpha_u \leq \alpha$.

Fix a vertex $u \in V \backslash U$. For all $i \in \partial G_u^t$, let $\widetilde{\sigma}_i = +$ if $i \in W_u^+$ and $\widetilde{\sigma}_i = -$ if $i \in W_u^-$ after the labeling defined in Step 3.2 of Algorithm 2. It is argued in (Mossel et al., 2013a, Section 5.2) that for each $i \in \partial G_u^t$, independently at random, $\widetilde{\sigma}_i = \sigma_i$ with probability $1 - \alpha_u$, and $\widetilde{\sigma}_i = -\sigma_i$ with probability $\alpha_u$. Consider the tree model $(T_u, \tau)$ with $\rho = 1/2$ and $a = c$, where for each vertex $i \in T_u$, independently at random, $\widetilde{\tau}_i = \tau_i$ with probability $1 - \alpha_u$ and $\widetilde{\tau}_i = -\tau_i$ with probability $\alpha_u$. By the coupling lemma given in Lemma 9, we can construct a coupling such that $(G_u^t, \sigma_{G_u^t}, \widetilde{\sigma}_{\partial G_u^t}) = (T_u^t, \tau_{T_u^t}, \widetilde{\tau}_{\partial T_u^t})$ holds with probability converging to 1. Moreover, on the event $(G_u^t, \sigma_{G_u^t}, \widetilde{\sigma}_{\partial G_u^t}) = (T_u^t, \tau_{T^t}, \widetilde{\tau}_{\partial T_u^t})$, we have that $R_u^t = \widetilde{\Gamma}_u^t$ in view of the definition of $R_u^t$ given in Algorithm 2, and the recursive formula of $\widetilde{\Gamma}_u^t$ given in Lemma 21. Hence,

$$p_G(\widehat{\sigma}_{\mathrm{BP}}^t) = \widetilde{q}_{T^t, \alpha_u} + o(1),$$

where the $o(1)$ term comes from the coupling error. Since $\widetilde{q}_{T^t, \alpha}$ is non-decreasing in $\alpha$, it follows that

$$p_G(\widehat{\sigma}_{\mathrm{BP}}^t) \leq \widetilde{q}_{T^t, \alpha} + o(1).$$

By the definition of $\mathcal{O}(\widehat{\sigma}, \sigma)$ given in (6),

$$\mathbb{E}\left[\mathcal{O}(\widehat{\sigma}_{\mathrm{BP}}^t, \sigma)\right] \leq p_G(\widehat{\sigma}_{\mathrm{BP}}^t),$$

and the lemma follows by combining the last two displayed equations.

### A.8. Proof of Lemma 17

Recall that $u_{t+1} = \frac{\mu^2}{4}\mathbb{E}\left[\tanh(u_t + \sqrt{u_t}Z)\right]$ with $u_1 = \frac{(1-2\alpha)^2\mu^2}{4}$. In the case $\rho = 1/2$ and $\mu = \nu$, $\theta = 0$ and $\lambda = \frac{\mu^2}{4}$, and hence $u_t$ and $v_t$ satisfy the same recursion. Also, comparing (39) to (11), $\widetilde{\Gamma}$ and $\Gamma$ satisfy the same recursion with $\rho = 1/2$ and $\mu = \nu$. Therefore, in view of the proof of Lemma 13, to prove the lemma, it suffices to show that in the base case with $t = 1$,

$$\sup_x \left| \mathbb{P}\left\{ \frac{\widetilde{Z}_\pm^1 \mp u_1}{\sqrt{u_1}} \leq x \right\} - \mathbb{P}\left\{Z \leq x\right\} \right| = O(b^{-1/2}), \tag{40}$$

Recall that $\widetilde{\Gamma}_i^0 = \frac{1}{2}\log\frac{1-\alpha}{\alpha}$ if $\widetilde{\tau}_i = +$ and $\widetilde{\Gamma}_i^0 = -\frac{1}{2}\log\frac{1-\alpha}{\alpha}$ if $\widetilde{\tau}_i = -$. Also, for all $i \in \partial u$, independently at random $\widetilde{\tau}_i = \tau_i$ with probability $1 - \alpha$, and $\widetilde{\tau}_i = \tau_i$ with probability $\alpha$. Let $x^* = \frac{1}{2}\log\frac{a-\alpha(a-b)}{b+\alpha(a-b)}$, $d = \frac{a+b}{2}$ and $\eta = \frac{b+(a-b)\alpha}{a+b}$. Thus, in view of the recursion given in (39), $\widetilde{\Gamma}_u^1 = \sum_{i=1}^{N_d} X_i$, where conditional on $\tau_u = \pm$, $N_d \sim \text{Pois}(d)$ is independent of $\{X_i\}$; $\{X_i\}$ are i.i.d. such that conditional on $\tau_u = +$, $X_i = x^*$ with probability $1 - \eta$ and $X_i = -x^*$ with probability $\eta$; conditional on $\tau_u = -$, $X_i = x^*$ with probability $\eta$ and $X_i = -x^*$ with probability $1 - \eta$. Taylor expansion yields that

$$x^* = \frac{(1-2\alpha)(a-b)}{2b} + O(b^{-1}).$$

By conditioning the label of $u$ is $-$, it follows that

$$\mathbb{E}\left[\widetilde{Z}_-^1\right] = -d(1-2\eta)x = -\frac{(1-2\alpha)(a-b)}{2}x = -\frac{(1-2\alpha)^2(a-b)^2}{4b} + O(b^{-1/2})$$

$$= -u_1 + O(b^{-1/2}),$$

$$\text{var}\left(\widetilde{Z}_-^1\right) = dx^2 = \frac{(1-2\alpha)^2(a-b)^2}{4b} + O(b^{-1/2}) = u_1 + O(b^{-1/2}).$$

In view of Lemma 19, we get that

$$\sup_x \left| \mathbb{P}\left\{ \frac{\widetilde{Z}_\pm^1 - \mathbb{E}\left[\widetilde{Z}_\pm^1\right]}{\sqrt{\text{var}\left(\widetilde{Z}_\pm^1\right)}} \leq x \right\} - \mathbb{P}\left\{Z \leq x\right\} \right| \leq O(b^{-1/2}).$$

Hence, we proved (40) for $\widetilde{Z}_-^1$. By symmetry between $-$ and $+$, the desired (40) also holds for $\widetilde{Z}_+^1$.

## Appendix B. A Data-driven Choice of the Parameter $\alpha$ in Algorithm 2

Algorithm 2 requires the knowledge of $\alpha_u$, which is given by $\alpha_u = |W_u^+ \Delta V^+|/n = |W_u^- \Delta V^-|/n$. In this section, we show that there exists an efficient estimator $\widehat{\alpha}_u$ such that $\widehat{\alpha}_u = \alpha_u + o(1)$ with high probability. Our estimation procedure is given in Algorithm 3.

**Lemma 22** *Let $\widehat{\alpha}_u$ be the output of Algorithm 3. Then with probability converging to 1, $\widehat{\alpha}_u = \alpha_u + o(1)$.*

---

**Algorithm 3** Estimation of $\alpha_u$

---

1: Take $U \subset V$ to be a random subset of size $\lfloor \sqrt{n} \rfloor$ and $S \subset V$ to be a random subset of size $\lfloor n/\log b \rfloor$. Let $u_* \in U$ be a random vertex in $U$ with at least $\frac{\log n}{2\log(\log n/b)}$ neighbors in $V \backslash U \backslash S$.

2: For $u \in V \backslash U \backslash S$ do

    1. Run a polynomial-time estimator capable of correlated recovery on the subgraph induced by vertices not in $G_u^{t-1}$ and $U \cup S$, and let $W_u^+$ and $W_u^-$ denote the output of the partition.

    2. Relabel $W_u^+$ and $W_u^-$ such that if $a > b$, then $u_*$ has more neighbors in $W_u^+$ than $W_u^-$; otherwise, $u_*$ has more neighbors in $W_u^-$ than $W_u^+$.

    3. Take $T \subset W_u^+ \cup W_u^-$ to be a random subset of size $\lfloor \sqrt{n} \rfloor$. Let $T_0 \subset T$ denote the set of vertices with at least $\frac{\log n}{2\log(\log n/b)}$ neighbors in $S$. Let $T_1$ denote a random subset of $T_0$ with size $\lfloor \frac{\log n}{b} \rfloor$.

    4. Run a polynomial-time estimator capable of correlated recovery on the subgraph induced by vertices not in $G_u^{t-1}$ and $U \cup T$. Let $W^+$ and $W^-$ denote the output of the partition. Relabel $(W^+, W^-)$ in the same way as $(W_u^+, W_u^-)$.

    5. Let $T_1^+$ consists of vertices $i \in T_1$ with more neighbors in $W^+$ than $W^-$; let $T_1^- = T_1 \backslash T_1^+$. Define $\widehat{\alpha}_u = \frac{|T_1^+ \cap W_u^-| + |T_1^- \cap W_u^+|}{|T_1|}$.

---

**Proof** We assume $a > b$ in the proof; the case $a < b$ can be proved similarly. Let $k^* = \frac{\log n}{2\log(\log n/b)}$. For any vertex $i$ in $T$, let $d_i$ denote its number of neighbors in $S$. Then $d_i$ is stochastically lower bounded by $\mathrm{Binom}(|S|, b/n)$. Since $|S| = \lfloor n/\log b \rfloor$, it follows that (see (Hajek et al., 2014, Appendix A) for a proof)

$$-\log \mathbb{P}\{d_i \geq k^*\} \leq \frac{1}{2}\log n - \log\log n.$$

Because $\{d_i\}_{i \in T}$ are independent, the cardinality of set $T_0$ is stochastically lower bounded by $\mathrm{Binom}(\lfloor \sqrt{n} \rfloor, \log n/\sqrt{n})$. Therefore, with high probability $|T_0| \geq \frac{1}{2}\log n$ and thus $T_1$ is well-defined.

Define the event $E_1$ to be that $T_1^+ = T_1 \cap V^+$ and $T_1^- = T_1 \cap V^-$. We claim that $\mathbb{P}\{E_1\} \to 1$. In fact, fix any vertex $i \in T_1$, suppose $\sigma_i = +$ without loss of generality and let $\mathcal{N}(i)$ denote the set of its neighbors in $V \backslash T$. Let

$$|W^+ \Delta V^+|/n \leq \delta n, \quad |W^- \Delta V^-| \leq \delta n.$$

Then $\delta \in [0, 1/2)$. Notice that $d_i$ is independent of the partition $W^+$ and $W^-$. Thus, conditional on $d_i$, $|\mathcal{N}(i) \cap W^+|$ is stochastically lower bounded by $\mathrm{Binom}(d_i, \frac{a}{a+b} - \frac{(a-b)\delta}{a+b})$ and $|\mathcal{N}(i) \cap W^-|$ is stochastically upper bounded by $\mathrm{Binom}(d_i, \frac{b}{a+b} + \frac{(a-b)\delta}{a+b})$. It follows from the Chernoff bound

that conditional on $d_i$,

$$\mathbb{P}\left\{|\mathcal{N}(i) \cap W^+| < \frac{(a-b)d_i}{2(a+b)}\right\} \leq \mathrm{e}^{-\Omega(d_i(a-b)^2/a^2)},$$

$$\mathbb{P}\left\{|\mathcal{N}(i) \cap W^-| > \frac{(a-b)d_i}{2(a+b)}\right\} \leq \mathrm{e}^{-\Omega(d_i(a-b)^2/a^2)}.$$

Due to $d_i \geq k^*$ for all $i \in T_0$, it yields that

$$\mathbb{P}\left\{|\mathcal{N}(i) \cap W^+| < |\mathcal{N}(i) \cap W^-|\right\} \leq 2\mathrm{e}^{-\Omega(k^*(a-b)^2/a^2)}.$$

Applying the union bound, we get that

$$\mathbb{P}\left\{\exists i \in T_1 \cap V^+, |\mathcal{N}(i) \cap W^+| < |\mathcal{N}(i) \cap W^-|\right\} \leq \frac{2\log n}{b}\mathrm{e}^{-\Omega(k^*(a-b)^2/a^2)}.$$

By the assumption that $a - b/\sqrt{a} = O(1)$ and $b = o(\log n)$, it follows that

$$\frac{\log n}{b}\mathrm{e}^{-\Omega(k^*(a-b)^2/a^2)} \leq \frac{\log n}{b}\exp\left(-\Omega\left(\frac{\log n}{b\log(\log n/b)}\right)\right) = o(1),$$

Combing the last two displayed equations gives that with high probability, for all $i \in T_1 \cap V^+$, $|\mathcal{N}(i) \cap W^+| > |\mathcal{N}(i) \cap W^-|$ and thus $i \in T_1^+$. Similarly, one can show that with high probability, for all $i \in T_1 \cap V^-$, $|\mathcal{N}(i) \cap W^+| < |\mathcal{N}(i) \cap W^-|$ and thus $i \in T_1^-$. Hence, $\mathbb{P}\{E_1\} \to 1$.

Finally, we show $\widehat{\alpha}_u = \alpha_u + o(1)$ with high probability. Let $W_u = W_u^+ \cup W_u^-$. Notice that $T_1$ is randomly chosen and independent of the partition $(W_u^+, W_u^-)$. Hence for $i \in T_1$, it lies in $(W_u^- \cap V^+) \cup (W_u^+ \cap V^-)$ with probability $\alpha_u$. Therefore,

$$|T_1 \cap V^+ \cap W_u^-| + |T_1 \cap V^- \cap W_u^+| \sim \mathrm{Binom}(|T_1|, \alpha_u).$$

Define the event $E_2$ to be

$$\frac{|T_1 \cap V^+ \cap W_u^-| + |T_1 \cap V^- \cap W_u^+|}{|T_1|} = \alpha_u\left(1 + (b/\log n)^{1/4}\right).$$

Then it follows from the Chernoff bounds that $\mathbb{P}\{E_2\} \to 1$. By the union bound, $\mathbb{P}\{E_1 \cap E_2\} \to 1$. Notice that on the event $E_1 \cap E_2$,

$$\widehat{\alpha}_u = \frac{|T_1 \cap V^+ \cap W_u^-| + |T_1 \cap V^- \cap W_u^+|}{|T_1|} = \alpha_u\left(1 + (b/\log n)^{1/4}\right).$$

Therefore, we conclude that $\widehat{\alpha}_u = \alpha_u + o(1)$ with high probability. ∎