Adversarially-Robust Inference on Trees via Belief Propagation

Samuel B. Hopkins* and Anqi Li†

April 2, 2024

Abstract

We introduce and study the problem of posterior inference on tree-structured graphical models in the presence of a malicious adversary who can corrupt some observed nodes. In the well-studied *broadcasting on trees* model, corresponding to the ferromagnetic Ising model on a d-regular tree with zero external field, when a natural signal-to-noise ratio exceeds one (the celebrated *Kesten-Stigum threshold*), the posterior distribution of the root given the leaves is bounded away from $\mathrm{Ber}(1/2)$, and carries nontrivial information about the sign of the root. This posterior distribution can be computed exactly via dynamic programming, also known as belief propagation.

We first confirm a folklore belief that a malicious adversary who can corrupt an inverse-polynomial fraction of the leaves of their choosing makes this inference impossible. Our main result is that accurate posterior inference about the root vertex given the leaves *is* possible when the adversary is constrained to make corruptions at a ρ-fraction of randomly-chosen leaf vertices, so long as the signal-to-noise ratio exceeds $O(\log d)$ and $\rho \le c\varepsilon$ for some universal $c > 0$. Since inference becomes information-theoretically impossible when $\rho \gg \varepsilon$, this amounts to an information-theoretically optimal fraction of corruptions, up to a constant multiplicative factor. Furthermore, we show that the canonical belief propagation algorithm performs this inference.

Contents

*Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139. SBH is supported by NSF CAREER award No. 2238080 and MLA@CSAIL. samhop@mit.edu.

†Department of Pure Mathematics and Mathematical Statistics (Centre for Mathematical Sciences), University of Cambridge, Cambridge CB30WB, England, UK. AL is supported by the Trinity Studentship in Mathematics. Part of this work was completed while AL was an intern at Microsoft Research (New England). anqili@mit.edu.

1 Introduction

Posterior inference is the central problem in Bayesian statistics: given a multivariate probability distribution, often specified by a *graphical model*, and observed values for a subset of the random variables in this distribution, the goal is to infer resulting conditional, or *posterior*, distribution on the unobserved variables. Bayesian methods allow rich domain knowledge to be incorporated into inference, via prior distributions, and posterior distributions offer an expressive language to describe the uncertainty remaining in unobserved variables given observations. Observations are not necessarily distributed iid conditioned on unobserved variables – a graphical model may specify a complex joint distribution among observed variables.

How robust are posterior inferences to corruptions or errors in observed data? Or, to misspecifiation of the underlying probabilistic model? The rapidly developing field of robust estimation addresses similar questions in the *frequentist* setting where the goal is to construct point estimators which retain provable accuracy guarantees when a small fraction of otherwise iid samples have been maliciously corrupted [DK23]. But, to our knowledge, relatively little attention has been paid to questions of robustness to adversarial (i.e. non-probabilistic) contamination in posterior inference.

We study robustness of posterior inference to adversarial corruption and model misspecification for a protoypical family of graphical models: broadcast processes on trees (see [EKPS00] and references therein).

Definition 1.1 (Broadcasting and inference in regular trees)**.** *Given a depth-t tree with vertices V where each non-leaf node has d children, we consider the following joint probability distribution on $\{\pm 1\}^V$. First, assign the root vertex a uniform draw from $\{-1, +1\}$. Then, recursively assign each child vertex the same sign as its parent with probability $\frac{1+\varepsilon}{2}$ and otherwise the opposite sign. Observing the signs of the leaf nodes σ_L in the broadcast tree, our goal is to infer the posterior distribution of the sign σ_R of the root.*

The broadcast process on trees is a useful "model organism" for our purposes:

1. Like many graphical models used in practice, it is tree-structured. This means that posterior inference is algorithmically tractable via dynamic programming, in this setting called "belief propagation".

2. Some graphical models exhibit *correlation decay*, meaning that the covariance/mutual information between subsets of random variables decays (exponentially) in the graph distance between the corresponding subsets. Strong forms of correlation decay (e.g., strong spatial mixing) imply that the distribution on any unobserved node far enough in graph distance to every observed node is insensitive to the values of the observed nodes. This means that any corruptions in observations won't affect the posterior on the unobserved node, but also means that even without corruptions, no interesting inference can be made about the unobserved node.

   As d and ε increase, the broadcast process exhibits long-range correlations, because the leaf signs σ_L collectively carry information about the sign σ_R of the root vertex. In particular, when $d\varepsilon^2 > 1$ (the celebrated *Kesten-Stigum threshold*), this holds even in the limit of infinite tree depth, that is, $\lim_{t\to\infty} I(\sigma_L; \sigma_R) > 0$. Thus, there is nontrival inference to be performed, but by the same token, incorrect observations of the leaf signs could adversely affect the accuracy of this inference.

We introduce and study several models for adversarially-robust inference in a broadcast tree. In each of these models, a malicious adversary observes the leaf signs σ_L resulting from a broadcast process and may flip a subset of them – the specifics of how this subset is chosen are important, and vary across our models. The corrupted leaf signs are then passed to an inference algorithm which aims to output the conditional distribution of the root sign *conditioned on the signs of the non-corrupted leaves*. Crucially, we do not know which subset of leaf signs has been flipped at inference time.

Organization In the remainder of this section, we give a high level overview of our results (Section 1.1), discuss how our results can be interpreted in terms of model misspecification (Section 1.2), discuss some related work (Section 1.3) and provide some proof ideas for our main results (Section 1.4).

1.1 Results

The first adversarial model we consider is the simplest and most powerful: for some $\rho > 0$, the *ρ-fraction adversary* can choose any ρd^t leaves (out of d^t leaves in total) and flip their signs. In this setting we confirm what we believe to be folklore [Pol23]:[1] for any $\rho > d^{-\Omega(t)}$, this adversary can make the posterior distribution at the root unidentifiable from given the (corrupted) leaf signs. In what follows, for a random variable X, we write $\{X\}$ for the distribution of X, and for an event E, we write $\{X \mid E\}$ for the distribution of X conditioned on E. We also write $d_{TV}(\cdot, \cdot)$ for the total variation distance between two distributions.

Theorem 1.2 (Proof in Section 5)**.** *There exists $\varepsilon_0 > 0$ such that for every ρ, d and $\varepsilon < \varepsilon_0$, there exists a $\varepsilon^{O(t)}$-fraction adversary A such that if (σ_R, σ_L) is distributed according to the broadcast process with parameters d, ε,*

$$d_{TV}(\{\sigma_L \mid \sigma_R = 1\}, \{A(\sigma_L) \mid \sigma_R = -1\}) \leq e^{-\Omega(t)}.$$

An alternative interpretation of Theorem 1.2 is that the Wasserstein distance, with respect to the Hamming metric, between the distributions $\{\sigma_L \mid \sigma_R = 1\}$ and $\{\sigma_L \mid \sigma_R = -1\}$, decays exponentially with t.

Theorem 1.2 implies that no algorithm can reliably distinguish whether $\sigma_R = 1$ or $\sigma_R = -1$, with advantage better than $e^{-\Omega(t)}$ over random guessing, in the presence of an $\varepsilon^{O(t)}$-fraction adversary. Accurately computing the posterior distribution $\{\sigma_R \mid \sigma_L\}$ and then sampling from it would distinguish these cases with nonvanishing advantage as $t \to \infty$ (if $\varepsilon^2 d > 1$). Hence, computing the posterior in the presence of this adversary is impossible.

While this may make it appear that posterior inference in the broadcast tree is inherently non-robust, recent works [MNS16, YP22] tell a contrasting story about a weaker adversary. The *ρ-random* adversary simply flips each leaf sign independently with probability $0 \leq \rho < 1/2$. It turns out that the posterior at the root vertex can still be accurately recovered in the presence of the ρ-random adversary as long as $\rho(\varepsilon, d)$ is held fixed as $t \to \infty$ [MNS16, YP22] – this is sometimes call "robust reconstruction." This suggests the question:

Is posterior inference in the broadcast process possible in the presence of an adversary more malicious than the ρ-random adversary?

We introduce a *semirandom* adversary, whose power lies in between the worst-case adversary of Theorem 1.2 and the ρ-random one. For the semirandom adversary, the locations of allowed sign flips are random, but the decision whether to make a flip is made adversarially, in full knowledge of all of the leaf signs.

Definition 1.3 (ρ-semirandom adversary)**.** *Fix $\rho > 0$. A ρ-semirandom adversary receives leaf signs σ_L and flips an independent coin x_u for each leaf u which is heads with probability ρ. For each leaf u, if u's coin is heads, the adversary may choose to flip the sign σ_u.*

Our main result shows that when the signal-to-noise ratio $d\varepsilon^2$ exceeds the Kesten-Stigum threshold by a logarithmic factor, there is $\rho(\varepsilon, d) > 0$ such that the distribution of the root vertex can be successfully inferred even in the presence of a ρ-semirandom adversary, for large-enough depth t. In what follows, we write $(\sigma_R, \sigma_L) \sim \mathcal{D}_{d,\varepsilon,t}$ to denote (σ_R, σ_L) distributed according to the broadcast on tree process with parameters d, ε run up to depth t.

[1] However, we are not aware of anywhere it is written in the literature.

Theorem 1.4 (Main theorem, follows from Lemma 3.2 and Lemma 3.11)**.** *For any $\delta > 0$ there exists C such that for any d, ε satisfying $d\varepsilon^2 > C \log \frac{d}{1-\varepsilon}$ there exists $\rho_0(\varepsilon) = \Omega(\varepsilon)$ and $t_0(\delta, d)$ such that if $\rho < \rho_0$ and $t > t_0$, for any ρ-semirandom adversary A, the belief-propagation algorithm* BP *satisfies*

$$\mathop{\mathbb{E}}_{\sigma_L \sim \mathcal{D}_{d,\varepsilon,t}} d_{TV}(\{\sigma_R \mid \sigma_L\}, \mathit{BP}(A(\sigma_L))) \leq \delta \,,$$

where the expectation is taken over the broadcast process σ_R, σ_L as well as the random choice of which vertices the adversary A may choose to corrupt.

The algorithm BP in Theorem 1.4 simply computes the posterior distribution of the root spin σ_R as if the leaf spins had been $A(\sigma_L)$ rather than σ_L, via dynamic programming – this is the canonical method to compute posterior distributions in tree-structured graphical models [EKPS00, MNS16]. Our analysis of belief propagation borrows techniques from the arguments of [MNS16] for the random-adversary case, but since the semirandom adversary can introduce nasty dependencies among leaf vertices, these arguments are far from transferring immediately.

We can also replace the assumption that the allowed corruptions are in random locations with a natural deterministic assumption on the pattern of allowed corruption locations. Concretely, if $d\varepsilon^2 > C \log \frac{d}{1-\varepsilon}$, then for every $c > 0$ there is k such that if the adversary makes at most c corruptions in every height-k subtree of the broadcast tree, we show that our algorithm successfully infers the distribution at the root vertex. We capture this in Theorem 4.2.

Open question: robustness down to the KS threshold Theorem 1.4 leaves an important open question: is robustness against a semirandom adversary possible for all $d\varepsilon^2 > 1$, as in the random-adversary case? Or, does the semirandom adversary shift the information-theoretic phase transition from non-recoverability to recoverability of the root spin away from 1?[2]

1.2 Adversarial Corruption versus Model Misspecification

Adversarial robustness is of course desirable when inference is performed with potentially-corrupted data. But is it useful beyond protection against malicious data poisoning?

Background: corruptions and misspecification in frequentist statistics In (frequentist) robust statistics, there is an appealing relationship between adversarial corruption of a subset of otherwise-iid samples and learning/estimation under model misspecification. Suppose we design a learning algorithm which takes samples from some distribution D in a class of distributions $\mathcal{D}$, of which 1% have been corrupted by an adversary, and successfully learns some $\hat{D}$ which is close to D in total variation distance.

Now, suppose that $\mathcal{D}$ is misspecified, in the sense that it does not contain the ground truth distribution D, but only contains some $D' \in \mathcal{D}$ such that $TV(D, D') \leq 0.001$. Since the adversary could have coupled D and D' to make corrupted samples from D' look as though they are from D, then even given samples from D the algorithm must still learn some $\hat{D}$ with small $TV(\hat{D}, D)$.

Misspecified Bayesian models Adversarially-robust algorithms in our setting are also robust against an appropriate notion of model misspecification. Concretely, fix a joint probability distribution $\mu(x_0, x_1, \ldots, x_n)$ and random variables $X_0, \ldots, X_n$ jointly distributed according to μ, and consider the posterior inference problem where we observe a joint sample $X_1, \ldots, X_n$ and aim to output the distribution $\{X_0 \mid X_1, \ldots, X_n\}$.

[2] In the related *stochastic blockmodel* setting, a so-called "monotone" adversary is known to shift the analogous non-recoverability threshold by a constant factor [MPW16]. In our setting, a monotone adversary would correspond to one who observes the root sign σ_R and may flip any leaf sign σ_L to agree with σ_R.

Now, let $\mathcal{S} \subseteq 2^{[n]}$ be a set of possible subsets of observed variables which are correctly specified by μ, the remaining ones being misspecified – formally, we introduce the following definition:

Definition 1.5. *An algorithm* `ALG` *solves the $\mathcal{S}$-misspecified inference problem for μ with error δ if for every $S \in \mathcal{S}$ and every μ' such that $\mu|_S = \mu'|_S$,*

$$\mathop{\mathbb{E}}_{x=(x_S, x_{\overline{S}}) \sim \mu'} d_{TV}(\texttt{ALG}(x), \{X_0 \mid X_S = x_S\}) \leq \delta \,.$$

Importantly, the algorithm `ALG` only knows $\mathcal{S}$, not the particular S or μ'.

In Definition 1.5, μ' is our "ground truth" description of the world, and μ is our misspecified model. Given a sample $x_1, \ldots, x_n$ from $\mu'|_{\{1,\ldots,n\}}$, if we knew μ' then the best inference we could make about x_0 would be the conditional distribution of x_0 given $x_1, \ldots, x_n$. But there is a problem – we do not know μ'. If we knew S, we could discard the observations whose distributions we know nothing about, and infer $\{X_0 \mid X_S\}$. We do not know S, but Definition 1.5 requires `ALG` to compete with a hypothetical algorithm that does.

Of course, many other possible notions of model misspecification for the posterior inference problem are possible – we make no claim that this definition is universally applicable. Exploring alternative notions of graphical model misspecification and their consequences for posterior inference is a fascinating open direction.

Interpreting our results as robustness to model misspecification In the context of the broadcast process on trees, we take X_0 above to be the sign at the root vertex σ_R, and $X_1, \ldots, X_n$ to be the leaf vertex signs σ_L. Theorem 1.2 shows if we take $\mathcal{S}$ to be all subsets of size $(1-\rho)d^t$ then the $\mathcal{S}$-misspecified problem is impossible. Theorem 1.4 has the following corollary, showing that there is a large set $\mathcal{S}$ of possible misspecifications against which robust inference is possible (indeed, a random $\mathcal{S}$ suffices).

Corollary 1.6. *Under the same hypotheses on $\delta, d, \varepsilon, \rho$ as Theorem 1.4, there is some $t_0(d, \delta)$ such that if $t \geq t_0$ then there exists $\mathcal{S} \subseteq 2^{[d^t]}$ with $|\mathcal{S}| \geq 2^{\Omega_{\varepsilon,d}(d^t)}$ and an algorithm which solves the $\mathcal{S}$-misspecified inference problem with error δ, where $\Omega_{\varepsilon,d}(\cdot)$ indicates asymptotic behavior as $t \to \infty$.*

The proof of this corollary constructs a ρ-semi-random adversary which can sample from a misspecification distribution μ'. We defer the proof to Section A.

1.3 Related Work

Algorithmic Robust Statistics Algorithmic robust statistics in high dimensions has developed rapidly in both statistics and computer science in the last decade. This field has focused on parameter estimation, distribution learning, and prediction using (otherwise-)iid samples of which a small fraction have been maliciously corrupted. The recent book [DK23] provides a comprehensive overview.

We highlight one parallel between our results and robust supervised learning. In PAC learning, the *Massart noise* model [MN06] is a parallel to our semirandom adversary. In the Massart noise model, for each labeled example $(x, f(x))$, the learner gets to see (x, y), where y is selected by flipping a coin which comes up heads with probability $\rho < 1/2$ and then allowing an adversary who sees x the chance to make a (randomized) decision whether to let $y = f(x)$ or $y = 1 - f(x)$. The Massart noise model is an important middle ground between randomized noise models (e.g. *random classification noise* [Kea90]) and nastier noise models (e.g. agnostic learning), and has recently led to several algorithmic advances [DGT19, DKTZ20].

A couple of recent works in algorithmic robust statistics design "robustified" versions of belief propagation or its dense-graph analogue, approximate message passing [IS23, LM22], often in the context

of adversarially-robust algorithms analogues of the "linearized-BP" algorithm for community detection (often called the "nonbacktracking spectral algorithm" [Abb18]) in the stochastic blockmodel [BMR21, DdNS22].

Finally, our work has a parallel in [CKMY22], which studies Bayesian inference with corruption in a Gaussian time-series setting (Kalman filtering). This work also finds a strong information-theoretic impossibility result for corruptions at adversarially-chosen times (corresponding for us to adversarially-chosen leaf locations), and an efficient inference algorithm when corruptions are made at random times but with knowledge of the rest of the time series.

Broadcasting on Trees Broadcasting on trees is an important special case of the ferromagnetic Ising model, and the problem of reconstructing the root vertex label from observations at the leaves has been extensively studied in probability, statistical physics, and computer science [Mos04]. It has significant connections to Markov-chain mixing times [BKMP05, MSW04], phylogenetic reconstruction [DMR06], and community detection [MNS15]. Of particular interest because of an important application to inference in the stochastic block model [MNS16, YP22] is the "robust reconstruction" problem, originally introduced in [JM04] to study sharpness of the Kesten-Stigum phase transition. From the perspective of our work, this is the root-inference problem in the presence of a random adversary.

Robust Bayesian Inference Ensuring that Bayesian inferences are robust to inaccuracies in the choices of priors and models is a longstanding concern dating at least to the 1950s [Goo50, BB86], with too vast a literature to survey here. Well-studied approaches involve choosing flat-tailed or noninformative priors and placing a hyperprior on a family of models to capture uncertainty about the true model.

1.4 Overview of Proofs

Optimal recovery from ρ-semirandom adversary In Section 3.1, we prove Theorem 1.4 in the regime of small ε, that is $\varepsilon \le \varepsilon^*$ for some small ε^*. In Section 3.2, we finish the proof of Theorem 1.4 by addressing the case that $\varepsilon > \varepsilon^*$. Both proofs use the following template. The belief propagation algorithm produces, for each vertex u, a "belief" $Z_u \in [-1, 1]$, which implicitly specifies an inferred posterior distribution $\{\sigma_u \,|\, \sigma_{\text{decendents}(u)}\}$ by specifying the bias of that distribution. Z_u is a function of $Z_{u1}, \dots, Z_{ud}$, where $u1, \dots, ud$ are children of u and $Z_u = \sigma_u$ for a leaf-vertex σ.

Let X_u denote the belief which would have been produced by BP if it were run on a non-corrupted leaf spins σ_L, and Z_u the belief produced when BP is run with corrupted leaf spins. Our goal will be to show that $|X_u - Z_u| < \delta$ for δ as small as we like, so long as u is at large-enough height in the tree.

We begin by showing that when $d\varepsilon^2$ exceeds Kesten-Stigum threshold by a $O(\log d)$ multiplicative factor, the difference $|Z_u - X_u|$ in our estimated belief Z_u for any vertex u at height 1 is moderately close to the actual "ground truth" belief X_u. That is, we show that

$$\mathbb{E} \max_{\text{adversary}} |Z_u - X_u| = O(\varepsilon)\,,$$

where the expectation is taken over the randomness in the broadcast process and also the allowed corruption locations x_u in the ρ-semirandom adversary (Lemma 3.4), and $\max_{\text{adversary}}$ denotes maximizing over choices of leaf-spin flips made by the ρ-semirandom adversary at the allowed locations x.

Next, we employ a contraction argument to show that this "worst case" perturbation of beliefs contracts as we move up the broadcast tree. More precisely, in Lemma 3.6 we basically show that if u is the parent of ui then

$$\mathbb{E} \max_{\text{adversary}} |Z_u - X_u| \le \frac{1}{2} \mathbb{E} \max_{\text{adversary}} |Z_{ui} - X_{ui}|.$$

This would then show that by taking a sufficiently large tree, we are able to recover the belief at the root to arbitrarily high precision.

We note that [MNS16] employs a similar contraction argument. However, the random adversary of [MNS16] flips each leaf spin independently. In comparison, our ρ-semirandom adversary can introduce new long-range correlations between the leaves in our broadcast tree.

Our contraction argument uses a first order Taylor expansion of the belief propagation function. Fix a vertex u with children $u1, \dots, ud$. In Lemma 3.6, we effectively show that there is some $f : [-1, 1] \to \mathbb{R}$ which captures the effect of each Z_{ui} on $BP(Z_{u1}, \dots, Z_{ud})$ in the sense that

$$\begin{aligned}
\mathbb{E} \max_{\text{adversary}} |Z_u - X_u| &= \mathbb{E} \max_{\text{adversary}} |BP(Z_{u1}, \dots, Z_{ud}) - BP(X_{u1}, \dots, X_{ud})| \\
&\le \mathbb{E} \max_{\text{adversary}} \left| \prod_{i=1}^{d} f(Z_{ui}) - \prod_{i=1}^{d} f(X_{ui}) \right| \\
&\le \mathbb{E} \left[\max_{\text{adversary}} \sum_{i=1}^{d} \max\{|f'(Z_{ui})|, |f'(X_{ui})|\} \cdot |X_{ui} - Z_{ui}| \cdot \max \left\{ \prod_{j \neq i} f(Z_{uj}), \prod_{j \neq i} f(X_{uj}) \right\} \right]
\end{aligned}$$

where the second inequality above follows from an application of the mean value theorem (plus some additional facts about monotonicity of f).

Now, if $|X_{ui} - Z_{ui}|$ were independent of Z_{uj}, X_{uj} (conditioned on σ_u), as it would be in the random adversary setting, we could modify the above chain of inequalities to condition each one on σ_u and then use this independence to separate the $|X_{ui} - Z_{ui}|$ term from the $\max\{\prod_{j \neq i} f(Z_{uj}), \prod_{j \neq i} f(X_{uj})\}$ term. As long as we could show that the latter term was small, we would be able to obtain a contraction in this way.

But because of long range correlations introduced by the adversary, even though $|X_{ui} - Z_{ui}|$ is expected to be small by the induction hypothesis, it is conceivable that the adversary can coordinate its flips in a way that blows up $\max\{|f'(Z_{ui})|, |f'(X_{ui})|\} \max\left\{\prod_{j \neq i} f(Z_{uj}), \prod_{j \neq i} f(X_{uj})\right\}$. To circumvent this, we carefully reintroduce independence by splitting the ρ-semi-random adversary into several *independent* "local" adversaries who can only see small subtrees. Formally, this corresponds to continuing the above with the bound

$$\le \mathbb{E} \left[\sum_{i=1}^{d} \left(\max_{\text{adversary}} \max\{|f'(Z_{ui})|, |f'(X_{ui})|\} \cdot |X_{ui} - Z_{ui}| \right) \cdot \left(\max_{\text{adversary}} \max \left\{ \prod_{j \neq i} f(Z_{uj}), \prod_{j \neq i} f(X_{uj}) \right) \right\} \right]$$

Strengthening the induction hypothesis to include $\mathbb{E} \max_{\text{adversary}} |Z_{uj} - X_{uj}| = O(\varepsilon)$ will allow us to bound remaining terms above involving f and f', leading to our contraction.

Information-theoretic lower bound: Proof of Theorem 1.2 To prove that a ρ-fraction adversary makes posterior inference impossible, we construct a coupling between the distributions $\{\sigma_L \mid \sigma_R = 1\}$ and $\{\sigma'_L \mid \sigma'_R = -1\}$ such that with all but exponentially-small probability, σ_L and σ'_L differ on a $d^{-\Omega(t)}$-fraction of coordinates. The key observation is that, for a spin σ_u and children $\sigma_{u1}, \dots, \sigma_{ud}$, the distributions $\{\sigma_{u1}, \dots, \sigma_{ud} \mid \sigma_u = 1\}$ and $\{\sigma_{u1}, \dots, \sigma_{ud} \mid \sigma_u = -1\}$ have Wasserstein distance roughly εd. If we are allowed to flip an ε-fraction of coordinates, we can couple the distributions successfully with high probability. This ε-fraction propagates down a height-t tree to become an ε^t fraction of necessary flips.

This argument can even be adapted to the ρ-semirandom adversary when ρ is $\Omega(\varepsilon)$ (see Theorem 5.3), by showing that the coupling needs only to flip signs which the ρ-semirandom adversary is allowed allowed to flip.

Spread adversary In Theorem 4.2 we show that for every $c > 0$ there exists k such that if an adversary makes at most c corruptions in every height-k subtree of the broadcast tree, then we can optimally infer the root.

The algorithm for Theorem 4.2 proceeds in two stages: first we run a "noise injection" phase – flipping each leaf node independently with small probability – and then we run belief propagation. We appeal to the bounded-sensitivity property of the posterior inference function in Theorem 4.4 (which is where the noise injection phase arises) in order to establish the base case of the contraction. The inductive step of the contraction is identical to that of Theorem 1.4.

2 Preliminaries

For a set Ω, we write $\Delta(\Omega)$ to denote the space of probability distributions over Ω. If $\Omega = \{\pm 1\}$, a distribution $\nu \in \Delta(\{\pm 1\})$ is also associated to a *belief* $B = \mathbb{E}_{b\sim\nu}\, b$. For $\nu, \nu' \in \Delta(\{\pm 1\})$, clearly $d_{TV}(\nu, \nu') \leq O(|B - B'|)$ where B, B' are the respective beliefs.

For a random variable X and event E, we write $\{X \mid E\}$ for the distribution of X conditioned on E. For distributions μ, ν, we write $d_{TV}(\mu, \nu)$ for the total variation distance between μ and ν. We write $\mathcal{D}_{d,\varepsilon,t}$ to denote the random process of generating spins on depth t d-regular trees according to the broadcast on tree process with parameters d, ε. In an abuse of notation, we will often write $(\sigma_R, \sigma_L) \sim \mathcal{D}_{d,\varepsilon,t}$ to mean a sample from the marginal distribution of $\mathcal{D}_{d,\varepsilon,t}$ on the spins of the root R and the d^t spins of the leaves L.

3 ρ-semi-random Adversary

In this section we prove our main result, Theorem 1.4. We make no attempt at optimizing the relevant constants and prove Theorem 1.4 with

$$t_0 = \log(d) + \log(\delta^{-1}), \text{ and } \quad \rho = \frac{\varepsilon}{4}. \tag{1}$$

Given the output of a broadcast on tree process that was run up till depth t, we recursively apply belief propagation until we reach the root and output the belief thus obtained. Specifically, suppose the spins of the leaves are given by $Z_{i,0}$ $(1 \leq i \leq d^t)$, then recursively define

$$Z_{j,i} = BP(Z_{dj+1,i-1}, Z_{dj+2,i-1}, \ldots, Z_{d(j+1)-1,i-1})$$

for $1 \leq i \leq t$ and $1 \leq j \leq d^i$, where $BP : \mathbb{R}^d \to \mathbb{R}$ is the belief propagation function given by

$$BP(X_1, \ldots, X_d) = \frac{\prod_{i=1}^d (1 + \varepsilon X_i) - \prod_{i=1}^d (1 - \varepsilon X_i)}{\prod_{i=1}^d (1 + \varepsilon X_i) + \prod_{i=1}^d (1 - \varepsilon X_i)}.$$

We argue separately in the cases of small and large ε.

3.1 Small-ε case

In this section we will prove Theorem 1.4 when $\varepsilon < \varepsilon^*$ for some sufficiently small constant ε^*; we capture this in Lemma 3.2. This small-ε case already captures the main ideas in the proof of Theorem 1.4; we defer the large-ε case to Section 3.2. Before we state Lemma 3.2 we introduce some convenient notation.

Definition 3.1. *Given d, ε, ρ, t, for a function $f : \{\pm 1\}^{d^t} \to \mathbb{R}$, let $\mathbb{E} \max_{\mathsf{adversary}} f(x)$ be the expected value of:*

- *Sampling $\sigma_L \sim \mathcal{D}_{d,\varepsilon,t}$ and a $0/1$-valued vector x of length d^t with entries independently equal to 1 with probability ρ and 0 otherwise*
- *Given σ_L and x, return $\max f(\sigma'_L)$ where σ_L and σ'_L differ only on the coordinates indicated by x.*

The main result in this section is:

Lemma 3.2. *There exist absolute constants C and $\varepsilon^* > 0$ such that if $d\varepsilon^2 \geq C\log(d)$ and $\varepsilon \leq \varepsilon^*$ then for any $\delta > 0$, with $\rho(\varepsilon, d), t_0(\delta, d)$ as in (1), for any $t \geq t_0$,*

$$\mathbb{E} \max_{\text{adversary}} |X_{\text{root,t}} - Z_{\text{root,t}}| \leq \delta.$$

Note that Lemma 3.2 establishes Theorem 1.4 in the case $\varepsilon \leq \varepsilon^*$ since $X_{\text{root,t}}$ is the bias corresponding to the posterior distribution of the root. Before we describe the main ideas behind the proof, we need one more piece of notation. If $x \in \{0,1\}^{d^t}$ is the random bitstring produced by the ρ-semi-random adversary, we define:

Definition 3.3. *Given a vector $x \in \{0,1\}^{d^t}$ indexed by leaves of a $(d+1)$-ary tree of depth t and an (internal) node u in that tree, we write x_u to denote the restriction of x to the leaves of the subtree rooted at u.*

The key steps in our contraction argument are given by the following two lemmas. The first, Lemma 3.4 captures that even at height-1 internal nodes, the adversary cannot (in expectation) corrupt the beliefs by more than $O(1/d)$. We will prove Lemma 3.4 later in this section.

Lemma 3.4. *There exists an absolute constant C such that for any $k \geq 1$, if $d\varepsilon^2 \geq C\log\left(\frac{d}{1-\varepsilon}\right)$ then for any vertex u at level $t-1$, we have*

$$\mathbb{E} \max_{\text{adversary}} \left[|X_{u,1} - Z_{u,1}| \,\middle|\, \sigma_u = +1 \right] \leq \frac{(1-\varepsilon)^{1/4}}{100d}.$$

We note that because we are above the Kesten-Stigum regime, it follows that $\frac{(1-\varepsilon)^{1/4}}{100d} \leq \frac{\varepsilon}{100}$, and so we can recover the beliefs at level $t-1$ up to error $O(\varepsilon)$. The $O(\varepsilon)$ bound on the RHS of Lemma 3.4 is crucial as the base case for an induction up the remaining $t-1$ levels of the tree. Our induction hypothesis, captured below in Lemma 3.6, relies on the computed beliefs $Z_{u,r}$ at a higher level $r \geq 1$ being distance at most $\lesssim \varepsilon$ to $X_{u,r}$ (in expectation).

For the both the base case and inductive step of the contraction argument, we will start by rearranging the belief propagation function definition of $X_{u,r}$ to $X_{u,r} = 2/(1 + \prod_{i \leq d} \frac{1-\varepsilon X_{ui,r-1}}{1+\varepsilon X_{ui,r-1}}) - 1$. We obtain a simpler bound on $X_{ur} - Z_{ur}$ by then applying an elementary-calculus bound on $\frac{1}{1+x} - \frac{1}{1+y}$ to $|X_{u,r} - Z_{u,r}|$:

Claim 3.5 (Proof in Section A)**.** *For any $0 < p < 1$ and any $x, y \geq 0$, we have $\left|\frac{1}{1+x} - \frac{1}{1+y}\right| \leq \frac{1}{p}|x^p - y^p|$.*

So, it will be enough to bound $\left| \left(\prod_{i=1}^d \frac{1-\varepsilon X_{ui,r}}{1+\varepsilon X_{ui,r}}\right)^{1/2} - \left(\prod_{i=1}^d \frac{1-\varepsilon Z_{ui,r}}{1+\varepsilon Z_{ui,r}}\right)^{1/2} \right|$. Conditioned on the values of $|X_{ui,r} - Z_{ui,r}|$, the extremal values for $\left(\prod_{i=1}^d \frac{1-\varepsilon Z_{ui,r}}{1+\varepsilon Z_{ui,r}}\right)^{1/2}$ are achieved when $\text{sign}(X_{ui,r} - Z_{ui,r})$ is the same for all i; we introduce random variables $Y_{v,r} = X_{v,r} - |X_{v,r} - Z_{v,r}|$ and $Y'_{v,r} = X_{v,r} + |X_{v,r} - Z_{v,r}|$ to capture the hypothetical situation that all these signs have lined up.

The following Lemma 3.6 says that both the resulting extremal values are close to $\left(\prod_{i=1}^d \frac{1-\varepsilon X_{ui,r}}{1+\varepsilon X_{ui,r}}\right)^{1/2}$. The technical key underlying our contraction argument is that we get a bound relying only on the assumption that

$$\max_{i \leq d} \mathbb{E}\left[\max_{\text{adversary}} |Y_{ui,r} - X_{ui,r}| \,\middle|\, \sigma_{ui} = +1 \right]$$

is in turn bounded. It would be much easier to argue under the stronger assumption of a bound on

$$\mathbb{E}\left[\max_{i\le d}\max_{\text{adversary}}|Y_{ui,r}-X_{ui,r}|\;\middle|\;\sigma_u=+1\right]. \tag{2}$$

The latter would amount to the assumption that the adversary has been unable to have much effect on the beliefs computed at *any* of the children of u. But this would be too strong an assumption to use inductively – the risk is that at level $r+1$ of the induction we would need an assumption on

$$\mathbb{E}\left[\max_{i,j\le d}\max_{\text{adversary}}|Y_{uij,r}-X_{uij,r}|\;\middle|\;\sigma_u=+1\right], \tag{3}$$

and so on. This eventually would amount to a union bound aiming to show that $|Y_{u,k}-X_{u,k}|\le\varepsilon$ simultaneously for every height-k vertex u, but this simply isn't true – since the adversary gets to corrupt a constant fraction of leaves, there are subtrees of the broadcast tree where they could achieve, say, $|X_{u,k}-Y_{u,k}|>0.1$.

In the non-adversarial setting, [MNS15] can make a similar argument, avoiding an assumption like (2) by leveraging the independence of the beliefs $Z_{ui,k}$ conditioned on the sign σ_u. Crucially for them, the signs of $Z_{ui,k}-X_{ui,k}$ conditioned on σ_u are independent across $i=1,\dots,d$, which leads important cancellations. This independence fails in the adversarial setting, because the adversary's choices whether to corrupt a leaf vertex may depend on signs of other far-away leaves! We carefully re-introduce independence by replacing the adversary who sees the whole tree with "local" adversaries who see only subtrees – we must argue that these local adversaries are not too much weaker than the original ρ-semirandom adversary. We show this argument in Section 3.1.2.

Lemma 3.6. *Define $Y_{v,r}=X_{v,r}-|X_{v,r}-Z_{v,r}|$, and $Y'_{v,r}=X_{v,r}+|X_{v,r}-Z_{v,r}|$. There exist constants $\varepsilon^*, C', R>0$ such that if $\varepsilon\le\varepsilon^*$, $\varepsilon^2 d>C'$, the following holds: let $r\ge 1$, for any vertex u at level $t-r-1$ with children $u1,\dots,ud$ and suppose $\xi:=\mathbb{E}\left[\max_{\text{adversary}}|Y_{ui,r}-X_{ui,r}|\;\middle|\;\sigma_{ui}=+1\right]\le\frac{\varepsilon}{10}$, then*

$$\mathbb{E}\left[\max_{\text{adversary}}\left|\left(\prod_{i=1}^{d}\frac{1-\varepsilon X_{ui,r}}{1+\varepsilon X_{ui,r}}\right)^{1/2}-\left(\prod_{i=1}^{d}\frac{1-\varepsilon Y_{ui,r}}{1+\varepsilon Y_{ui,r}}\right)^{1/2}\right|\;\middle|\;\sigma_u=+1\right]\le de^{-R(d-1)\varepsilon^2}\xi.$$

Furthermore, the above holds with $Y'_{ui,k}, Y'_{ui,r}$ replacing $Y_{ui,k}, Y_{ui,r}$ respectively.

Given Lemmas 3.4 and 3.6, we show how to deduce Lemma 3.2.

Proof of Lemma 3.2. Let $\alpha=\varepsilon/10$. We prove by induction on $r\ge 1$ that for any vertex u at level r, we have

$$\mathbb{E}\left[\max_{\text{adversary}}|X_{u,t-r-1}-Z_{u,t-r-1}|\;\middle|\;\sigma_u=+1\right]\le 2^{-(r-1)}\alpha. \tag{4}$$

Having proved this, it will suffice to take $t_0\asymp\log(\delta^{-1})$ and iteratively apply (4).

The base case when $r=1$ follows from Lemma 3.4. Assume the claim is true for r and we aim to prove the claim for $r+1$. Fix a vertex u at level $t-(r+1)$ and let its children be $u1,\dots,ud$.

By taking $p = \frac{1}{2}$ in Claim 3.5, it follows for any k that

$$|X_{u,k+1} - Z_{u,k+1}| = |BP(X_{u1,k}, \dots, X_{ud,k}) - BP(Z_{u1,k}, \dots, Z_{ud,k})|$$
$$\leq 4 \left| \left(\prod_i \frac{1 - \varepsilon X_{ui,k}}{1 + \varepsilon X_{ui,k}} \right)^{1/2} - \left(\prod_i \frac{1 - \varepsilon Z_{ui,k}}{1 + \varepsilon Z_{ui,k}} \right)^{1/2} \right|. \tag{5}$$

By the monotonicity of $x \mapsto \frac{1-\varepsilon x}{1+\varepsilon x}$, we have the pointwise inequalities

$$\frac{1 - \varepsilon Y'_{ui,k}}{1 + \varepsilon Y'_{ui,k}} \leq \frac{1 - \varepsilon Z_{ui,k}}{1 + \varepsilon Z_{ui,k}} \leq \frac{1 - \varepsilon Y_{ui,k}}{1 + \varepsilon Y_{ui,k}}$$

and therefore

$$\left| \left(\prod_i \frac{1 - \varepsilon X_{ui,k}}{1 + \varepsilon X_{ui,k}} \right)^{1/2} - \left(\prod_i \frac{1 - \varepsilon Z_{ui,k}}{1 + \varepsilon Z_{ui,k}} \right)^{1/2} \right|$$
$$\leq \max \left\{ \left| \left(\prod_i \frac{1 - \varepsilon X_{ui,k}}{1 + \varepsilon X_{ui,k}} \right)^{1/2} - \left(\prod_i \frac{1 - \varepsilon Y_{ui,k}}{1 + \varepsilon Y_{ui,k}} \right)^{1/2} \right|, \left| \left(\prod_i \frac{1 - \varepsilon X_{ui,k}}{1 + \varepsilon X_{ui,k}} \right)^{1/2} - \left(\prod_i \frac{1 - \varepsilon Y'_{ui,k}}{1 + \varepsilon Y'_{ui,k}} \right)^{1/2} \right| \right\}. \tag{6}$$

That is, the "worst case adversary" is the one that is able to perturb the spins at level t such that running belief propagation on these spins produces beliefs $\{B_{ui,k}\}$ satisfying $\operatorname{sign}(B_{ui,k} - Z_{ui,k}) = \operatorname{sign}(B_{uj,k} - Z_{uj,k})$ for all $1 \leq i, j \leq k$. It follows that

$$\mathbb{E}\left[\max_{\text{adversary}} |X_{u,t-r-1} - Z_{u,t-r-1}| \,\middle|\, \sigma_u = +1 \right]$$
$$\leq 4\,\mathbb{E}\left[\max_{\text{adversary}} \left| \left(\prod_i \frac{1 - \varepsilon X_{ui,t-r}}{1 + \varepsilon X_{ui,t-r}} \right)^{1/2} - \prod_i \left(\frac{1 - \varepsilon Z_{ui,t-r}}{1 + \varepsilon Z_{ui,t-r}} \right)^{1/2} \right| \,\middle|\, \sigma_u = +1 \right]$$
$$\leq 4\,\mathbb{E}\left[\max_{\text{adversary}} \left| \left(\prod_i \frac{1 - \varepsilon X_{ui,t-r}}{1 + \varepsilon X_{ui,t-r}} \right)^{1/2} - \prod_i \left(\frac{1 - \varepsilon Y_{ui,t-r}}{1 + \varepsilon Y_{ui,t-r}} \right)^{1/2} \right| \,\middle|\, \sigma_u = +1 \right]$$
$$+ 4\,\mathbb{E}\left[\max_{\text{adversary}} \left| \left(\prod_i \frac{1 - \varepsilon X_{ui,t-r}}{1 + \varepsilon X_{ui,t-r}} \right)^{1/2} - \prod_i \left(\frac{1 - \varepsilon Y'_{ui,t-r}}{1 + \varepsilon Y'_{ui,t-r}} \right)^{1/2} \right| \,\middle|\, \sigma_u = +1 \right]$$

Applying Lemma 3.6 to the above gives

$$\mathbb{E}\left[\max_{\text{adversary}} |X_{u,t-r-1} - Z_{u,t-r-1}| \,\middle|\, \sigma_u = +1 \right] \leq 16d(2^{-r-1}\alpha)e^{-R(d-1)\varepsilon^2}$$

and by choosing C sufficiently large such that $\varepsilon^2 d > C \log(d)$, we can make the above smaller than $2^{-(r+1)-1}\alpha$ as desired. □

3.1.1 Base case: level $t - 1$

In this subsection, we prove Lemma 3.4. Fix a vertex u at level $t - 1$, and suppose the sum of the uncorrupted spins of its d children is S while the sum of the corrupted spins of its d children is S'.

Proof idea: The main idea is that a Chernoff bound gives $S, S' = \Theta(\varepsilon d)$. Vaguely speaking, if most of the input to the belief function BP function is $+1$, then we would expect the output of BP to also be quite close to 1. By writing $|Z_{u,1} - X_{u,1}| \le \left| \left(\prod_{i=1}^{d} \frac{1-\varepsilon X_{ui,r}}{1+\varepsilon X_{ui,r}} \right)^{1/2} - \left(\prod_{i=1}^{d} \frac{1-\varepsilon Z_{ui,r}}{1+\varepsilon Z_{ui,r}} \right)^{1/2} \right|$ à la Claim 3.5, we can convert $S, S' = \Omega(\varepsilon d)$ and $\varepsilon^2 d \gtrsim \log(d)$ to a $O(1/d)$ bound on the magnitude $|Z_{u,1} - X_{u,1}|$.

Proof of Lemma 3.4. Let S denote the random variable given by the sum of the d uncorrupted spins of the children of u and let S' be the sum of the d corrupted spins of the children of u. Then we have that $\mathbb{E}[S \mid \sigma_u = +1] = \varepsilon d$, and in particular the Chernoff bound implies that for sufficiently large d (which we can arrange by taking a sufficiently large C), we have

$$\mathbb{P}\left[S \le \frac{\varepsilon d}{2} \;\middle|\; \sigma_u = +1 \right] \le \exp(-\varepsilon d/8) \le (1-\varepsilon/2)^d \le \frac{(1-\varepsilon)^{1/4}}{800d}.$$

By a Chernoff bound, since $\rho = \varepsilon/4$ in (1), we also have for sufficiently large d,

$$\mathbb{P}\left[|x_u| \ge \frac{\varepsilon d}{3} \right] \le \exp\left(-\Omega\left(\varepsilon d\right)\right) \le (1-\varepsilon/2)^d \le \frac{(1-\varepsilon)^{1/4}}{800d}. \tag{7}$$

In particular, it follows that

$$\mathbb{P}\left(S \ge \frac{\varepsilon d}{2} \text{ and } \min_{\text{adversary}} S' \ge \frac{\varepsilon d}{6} \right) \ge \left(1 - \frac{(1-\varepsilon)^{1/4}}{800d} \right)^2 \ge 1 - \frac{(1-\varepsilon)^{1/4}}{400d}.$$

Next, by Claim 3.5, note that we can bound

$$\begin{aligned}
\mathbb{E} \max_{\text{adversary}} |Z_{u,1} - X_{u,1}| &\le 2\,\mathbb{E}\left[\max_{\text{adversary}} \left| \left(\prod_{i=1}^{d} \frac{1-\varepsilon X_{ui,r}}{1+\varepsilon X_{ui,r}} \right)^{1/2} - \left(\prod_{i=1}^{d} \frac{1-\varepsilon Z_{ui,r}}{1+\varepsilon Z_{ui,r}} \right)^{1/2} \right| \;\middle|\; S \ge \frac{\varepsilon d}{2},\, S' \ge \frac{\varepsilon d}{6} \right] \\
&\quad + 2 \cdot \frac{(1-\varepsilon)^{1/4}}{400d} \\
&= 2\,\mathbb{E}\left[\max_{\text{adversary}} \left| \left(\frac{1-\varepsilon}{1+\varepsilon} \right)^{S'/2} - \left(\frac{1-\varepsilon}{1+\varepsilon} \right)^{S/2} \right| \;\middle|\; S \ge \frac{\varepsilon d}{2},\, S' \ge \frac{\varepsilon d}{6} \right] + \frac{(1-\varepsilon)^{1/4}}{200d} \\
&\le 2\left(\frac{1-\varepsilon}{1+\varepsilon} \right)^{\frac{\varepsilon d}{4}} + 2\left(\frac{1-\varepsilon}{1+\varepsilon} \right)^{\frac{\varepsilon d}{12}} + \frac{(1-\varepsilon)^{1/4}}{200d} \\
&\le 4e^{-\Omega_C(\varepsilon^2 d)} + \frac{(1-\varepsilon)^{1/4}}{200d} \\
&\le \frac{(1-\varepsilon)^{1/4}}{100d}
\end{aligned}$$

by taking a sufficiently large C. □

3.1.2 Contraction in the presence of an adversary

We turn to the proof of Lemma 3.6. In the proof, we relate $\left| \left(\prod_{i=1}^{d} \frac{1-\varepsilon X_{ui,r}}{1+\varepsilon X_{ui,r}} \right)^{1/2} - \left(\prod_{i=1}^{d} \frac{1-\varepsilon Y_{ui,r}}{1+\varepsilon Y_{ui,r}} \right)^{1/2} \right|$ with $|X_{ui,r} - Y_{ui,r}|$ (whose expected value is bounded by assumption) by applying the mean value theorem.

Specifically, we write

$$\left|\left(\prod_{i=1}^{d}\frac{1-\varepsilon X_{ui,r}}{1+\varepsilon X_{ui,r}}\right)^{1/2}-\left(\prod_{i=1}^{d}\frac{1-\varepsilon Y_{ui,r}}{1+\varepsilon Y_{ui,r}}\right)^{1/2}\right|\le\sum_{i=1}^{d}|X_{ui,r}-Y_{ui,r}|\max_{y_1\in[Y_{u1,r},X_{u1,r}]}\frac{\partial f}{\partial x}(y_i)\prod_{j\neq i}f(y_j)\tag{8}$$

where $f(x)=\left(\frac{1-\varepsilon x}{1+\varepsilon x}\right)^{1/2}$. The following two claims allow us to bound the size of each of the terms in the derivative. We comment that the proof of Lemma 3.6 strongly utilizes the independence of the locations in x where a flip is allowed. This is because a priori we are only able to work with the expected value version of (8), and it is this independence that allows us to split the products and bound them term by term.

Claim 3.7. *Let $f(x)=\left(\frac{1-x}{1+x}\right)^{1/2}$. There exists $c>0$ such that if $|x|\le c$, then we have $|f(x)|\le 1-x+\frac{3x^2}{5}$.*

Claim 3.8. *There are $\varepsilon^*,\kappa>0$ such that for all $\varepsilon<\varepsilon^*$, all $x\in[-2,2]$ and all $a\in[-2,2]$, we have*

$$\left|\frac{\partial}{\partial x}\left(\frac{1-\varepsilon(x+a)}{1+\varepsilon(x+a)}\right)^{1/2}\right|\le\kappa.$$

We defer the proofs of Claims 3.7 and 3.8 to the Appendix. We also quote without proof the following lemma from [MNS16].

Lemma 3.9 ([MNS16, Lemma 3.5])**.** *For a vertex u at level s, let $u_1,\dots,u_{d^r}$ denote its d^r children at level $s+r$. Let S_r denote the sum of the spins $\sum_{i=1}^{d^r}\sigma_{u_i}$. Then*

$$\mathrm{Var}[S_r\,|\,\sigma_u=+1]=(1-\varepsilon)^2d^r\frac{(\varepsilon^2d)^r-1}{\varepsilon^2d-1}.$$

Using this second moment bound on the majority vote, we can deduce that the ground truth beliefs $X_{ui,k}$ are all close to 1 in the regime where we exceed the Kesten-Stigum threshold by a constant.

Lemma 3.10. *There exists $C>0$ such that for all d,ε with $\varepsilon^2d>C$, for all $s\ge 1$, we have*

$$\mathbb{E}[X_{u,s}\,|\,\sigma_u=+1]\ge\frac{7}{8}.$$

Proof. Let S be the sum of the spins of the children s levels down from u. By the optimality of $\mathrm{sign}(X_{u,s})$ as an estimate for σ_u given the spins of children s levels down, we have

$$\begin{aligned}\frac{\mathbb{E}\,|X_{u,s}|+1}{2}&\ge\mathbb{P}(\mathrm{sign}(S)>0)\\&\ge 1-\frac{\mathrm{Var}[s\,|\,\sigma_u=+1]}{(\mathbb{E}[S\,|\,\sigma_u=+1])^2}\\&=1-\frac{\varepsilon^{2s}d^{2s}}{(\varepsilon^sd^s)^2(\varepsilon^2d-1)}\\&\ge 1-\frac{1}{C-1}\end{aligned}$$

where in the second inequality we applied the Chebyshev inequality. Observe that

$$\mathbb{P}[X_{u,s}<0\,|\,\sigma_u=+1]\ge\frac{1-\mathbb{E}[|X_{u,s}|\,|\,\sigma_u=+1]}{2}.$$

Consequently,

$$\begin{aligned}\mathbb{E}[X_{u,s} < 0 \,|\, \sigma_u = +1] &\geq \mathbb{E}[|X_{u,s}| \,|\, \sigma_u = +1] - 2\mathbb{P}[X_{u,s} < 0 \,|\, \sigma_u = +1] \\ &= 2\,\mathbb{E}[|X_{u,s}| \,|\, \sigma_u = +1] - 1 \\ &\geq 1 - \frac{4}{C-1} \\ &\geq \frac{7}{8}\end{aligned}$$

for sufficiently large C. □

Proof of Lemma 3.6. We will prove the case of $Y_{ui,r}$; the $Y'_{ui,r}$ case follows by nearly identical reasoning.

Now, let $f(x) = \left(\frac{1-\varepsilon x}{1+\varepsilon x}\right)^{1/2}$, let $g(y_1, \ldots, y_d) = \prod_{j \leq d} f(y_j)$, and let $g_i(y_1, \ldots, y_d) = \frac{\partial g}{\partial y_i}$. By applying the mean value theorem, we can write

$$\begin{aligned}&\mathbb{E}\left[\max_{\text{adversary}} |g(X_{u1,r}, \ldots, X_{ud,r}) - g(Y_{u1,r}, \ldots, Y_{ud,r})| \;\middle|\; \sigma_u = +1\right] \\ &\leq \mathbb{E}\left[\max_{\text{adversary}} \sum_{i=1}^{d} |X_{ui,r} - Y_{ui,r}| \cdot \max_{\substack{y_1 \in [Y_{u1,r}, X_{u1,r}] \\ \vdots \\ y_d \in [Y_{ud,r}, X_{ud,r}]}} |g_i(y_1, \ldots, y_d)| \;\middle|\; \sigma_u = +1\right] \qquad (9) \\ &\leq \kappa\,\mathbb{E}\left[\sum_{i=1}^{d} \max_{\text{adversary}} |X_{ui,r} - Y_{ui,r}| \cdot \max_{\text{adversary}} \max_{\substack{y_1 \in [Y_{u1,r}, X_{u1,r}] \\ \vdots \\ y_d \in [Y_{ud,r}, X_{ud,r}]}} \prod_{\substack{j=1 \\ j \neq i}}^{d} f(y_j) \;\middle|\; \sigma_u = +1\right].\end{aligned}$$

In the above, κ is the constant from Claim 3.8. In the second inequality, we introduced the *"local" adversaries* who only looks at the subtree beneath ui. At this point, we note that $\max_{\text{adversary}} |X_{ui,r} - Y_{ui,r}|$ only depends on the randomness which we revealed in the subtree below ui, while $\max_{\text{adversary}} \max_{\substack{y_1 \in [Y_{u1,r}, X_{u1,r}] \\ \vdots \\ y_d \in [Y_{ud,r}, X_{ud,r}]}} \prod_{\substack{j=1 \\ j \neq i}}^{d} f(y_j)$ only depends on the randomness that we reveal in the subtrees below uj for $j \neq i$. Because of the tree structure, it follows that if we condition on $\sigma_u = +1$, then $\max_{\text{adversary}} |X_{ui,r} - Y_{ui,r}|$ is independent of $\max_{\text{adversary}} \max_{\substack{y_1 \in [Y_{u1,r}, X_{u1,r}] \\ \vdots \\ y_d \in [Y_{ud,r}, X_{ud,r}]}} \prod_{\substack{j=1 \\ j \neq i}}^{d} f(y_j)$. It is crucial that the derivative of the belief propagation function has this separation property, allowing us to leverage the independence to write:

$$\mathbb{E}\left[\max_{\text{adversary}}\left|\prod_{i=1}^{d} f(X_{ui,r}) - \prod_{i=1}^{d} f(Y_{ui,r})\right| \,\middle|\, \sigma_u = +1\right]$$

$$\leq \kappa \sum_{i=1}^{d} \mathbb{E}\left[\max_{\text{adversary}} |X_{ui,r} - Y_{ui,r}| \,\middle|\, \sigma_{ui} = +1\right] \cdot \mathbb{E}\left[\max_{\text{adversary}} \max_{\substack{y_1 \in [Y_{u1,r}, X_{u1,r}] \\ \vdots \\ y_d \in [Y_{ud,r}, X_{ud,r}]}} \prod_{\substack{j=1 \\ j \neq i}}^{d} f(y_j) \,\middle|\, \sigma_u = +1\right]$$

$$\leq \kappa d \xi \, \mathbb{E}\left[\max_{\text{adversary}} \max_{\substack{y_1 \in [Y_{u1,r}, X_{u1,r}] \\ \vdots \\ y_d \in [Y_{ud,r}, X_{ud,r}]}} \prod_{j=1}^{d-1} f(y_j) \,\middle|\, \sigma_u = +1\right].$$

where in the last inequality we used symmetry of the subtrees and also the fact that

$$\mathbb{E}\left[\max_{\text{adversary}} |X_{ui,k} - Y_{ui,k}| \,\middle|\, \sigma_{ui} = +1\right] = \mathbb{E}\left[\max_{\text{adversary}} |X_{ui,k} - Y_{ui,k}| \,\middle|\, \sigma_{ui} = -1\right] = \mathbb{E}\left[\max_{\text{adversary}} |X_{ui,k} - Y_{ui,k}|\right].$$

This equality holds because $\max_{\text{adversary}} |X_{ui,k} - Y_{ui,k}|$ measures the magnitude of the change in beliefs of the worst adversary, and there by flipping $+1$ to -1 and vice versa there is a coupling between the broadcast processes when $ui = +1$ and $ui = -1$ which leaves this magnitude invariant. In particular, this implies that

$$\mathbb{E}\left[\max_{\text{adversary}} |X_{ui,r} - Y_{ui,r}| \,\middle|\, \sigma_{ui} = +1\right] \leq \xi$$

by the hypothesis of the claim. Next, we write

$$\mathbb{E}\left[\max_{\text{adversary}} \max_{\substack{y_1 \in [Y_{u1,r}, X_{u1,r}] \\ \vdots \\ y_d \in [Y_{ud,r}, X_{ud,r}]}} \prod_{j=1}^{d-1} f(y_j) \,\middle|\, \sigma_u = +1\right] \leq \mathbb{E}\left[\max_{\substack{y_1 \in [Y_{u1,r}, X_{u1,r}] \\ \vdots \\ y_d \in [Y_{ud,r}, X_{ud,r}]}} \prod_{j=1}^{d-1} \max_{\text{adversary}} f(y_j) \,\middle|\, \sigma_u = +1\right]$$

$$= \prod_{j=1}^{d-1} \mathbb{E}\left[\max_{\text{adversary}} \max_{\substack{y_1 \in [Y_{u1,r}, X_{u1,r}] \\ \vdots \\ y_d \in [Y_{ud,r}, X_{ud,r}]}} f(y_j) \,\middle|\, \sigma_u = +1\right]$$

$$\leq \prod_{j=1}^{d-1} \mathbb{E}\left[\max_{\text{adversary}} f(X_{uj,r} - |Y_{uj,r} - X_{uj,r}|) \,\middle|\, \sigma_u = +1\right]$$

where the third equality follows because of independence of x_{ui} and x_{uj} for $i \neq j$ and the final inequality

follows from the monotonicity of f so that $\max_{x\in[a,b]} f(x) = f(a)$. Now we apply Claim 3.7 to the each term in the above inequality to obtain

$$\begin{aligned}
&\mathbb{E}\left[\max_{\text{adversary}}\left|\prod_{i=1}^{d} f(X_{ui,r}) - \prod_{i=1}^{d} f(Y_{ui,r})\right| \,\middle|\, \sigma_u = +1\right] \\
&\le \kappa d\xi \prod_{j=1}^{d-1} \mathbb{E}\left[\max_{\text{adversary}}\left(1 - \varepsilon(X_{uj,r} - |Y_{uj,r} - X_{uj,r}|) + \frac{3\varepsilon^2(X_{uj,r} - |Y_{uj,r} - X_{uj,r}|)^2}{5}\right) \,\middle|\, \sigma_u = +1\right] \\
&\le \kappa d\xi \prod_{j=1}^{d-1} \mathbb{E}\left[1 - \varepsilon(X_{uj,r} - \max_{\text{adversary}} |Y_{uj,r} - X_{uj,r}|) + \frac{3\varepsilon^2 \max_{\text{adversary}}(X_{uj,r} - |Y_{uj,r} - X_{uj,r}|)^2}{5} \,\middle|\, \sigma_u = +1\right] \\
&\le \kappa d\xi \prod_{j=1}^{d-1} \left(1 - \varepsilon\,\mathbb{E}[X_{uj,r}|\sigma_u = +1] + \varepsilon\,\mathbb{E}\max_{\text{adversary}}[|Y_{uj,r} - X_{uj,r}||\sigma_u = +1]\right. \\
&\qquad\qquad \left. + \frac{3\varepsilon^2\,\mathbb{E}\max_{\text{adversary}}[(X_{uj,r} - |Y_{uj,r} - X_{uj,r}|)^2|\sigma_u = +1]}{5}\right) \\
&\le \kappa d\xi \prod_{j=1}^{d-1} e^{-\varepsilon\,\mathbb{E}[X_{uj,r}|\sigma_u=+1] + \varepsilon\,\mathbb{E}\max_{\text{adversary}}[|Y_{uj,r} - X_{uj,r}||\sigma_u=+1] + \frac{3\varepsilon^2\,\mathbb{E}\max_{\text{adversary}}[(X_{uj,r} - |Y_{uj,r} - X_{uj,r}|)^2|\sigma_u=+1]}{5}} \\
&\le \kappa d\xi \prod_{j=1}^{d-1} e^{-\varepsilon\,\mathbb{E}[X_{uj,r}|\sigma_u=+1] + \varepsilon\xi + \frac{3\varepsilon^2\,\mathbb{E}\max_{\text{adversary}}[(X_{uj,r} - |Y_{uj,r} - X_{uj,r}|)^2|\sigma_u=+1]}{5}},
\end{aligned}$$

where for the penultimate inequality we used the fact that $1 - x \le e^{-x}$ for all $x \in \mathbb{R}$.

To simplify this further, note that by Lemma 3.10 we have

$$\begin{aligned}
\mathbb{E}[X_{ui,r}|\sigma_u = +1] &= \frac{1+\varepsilon}{2}\,\mathbb{E}[X_{ui,r}|\sigma_{ui} = +1] + \frac{1-\varepsilon}{2}\,\mathbb{E}[X_{ui,r}|\sigma_{ui} = -1] \\
&= \mathbb{E}\left[\frac{1+\varepsilon}{2}\cdot X_{ui,r} - \frac{1-\varepsilon}{2}\cdot X_{ui,r} \mid \sigma_{ui} = +1\right] \\
&= \varepsilon\,\mathbb{E}[X_{ui,r} \mid \sigma_{ui} = +1] \ge \frac{7\varepsilon}{8}. \qquad (10)
\end{aligned}$$

And furthermore $\mathbb{E}\max_{\text{adversary}}[(X_{uj,r} - |Y_{uj,r} - X_{uj,r}|)^2 \,\big|\, \sigma_u = +1] \le 1 + 4\xi$, since we have both $|X_{uj,r}|^2 \le 1$, $\mathbb{E}\max_{\text{adversary}}[2X_{uj,r}|Y_{uj,r} - X_{uj,k}| \,\big|\, \sigma_u = +1] \le 2\,\mathbb{E}\max_{\text{adversary}}[|Y_{uj,r} - X_{uj,r}| \,\big|\, \sigma_u = +1] \le 2\xi$ and $\mathbb{E}\max_{\text{adversary}}[|Y_{uj,r} - X_{uj,r}|^2 \,\big|\, \sigma_u = +1] \le 2\xi$. This allows us to write

$$\mathbb{E}\left[\max_{\text{adversary}}\left|\prod_{i=1}^{d} f(X_{ui,r}) - \prod_{i=1}^{d} f(Y_{ui,r})\right| \,\middle|\, \sigma_u = +1\right] \le \kappa d\xi e^{(d-1)\varepsilon^2\left(-\frac{7}{8} + \xi\varepsilon^{-1} + \frac{3+12\xi}{5}\right)}$$

as desired. □

3.2 Large-ε case

In this subsection, we aim to prove the following result.

Lemma 3.11. *For any $0 < \varepsilon^* < 1$, there is some $C(\varepsilon^*) > 0$ such that for all $\varepsilon > \varepsilon^*$ and d satisfying $\varepsilon^2 d > -C\log(1-\varepsilon)$, then with $\rho(\varepsilon), t_0(\delta, d)$ as in (1), for any $t \geq t_0$,*

$$\mathbb{E} \max_{\text{adversary}} |X_{\text{root,t}} - Z_{\text{root,t}}| \leq \delta.$$

Here, the expectation $\mathbb{E}$ is taken with respect to the underlying broadcast process which produces leaf spins σ_L, and the Ber(ρ) variables x indicating which leaf spins can be corrupted by the adversary.

The gist of the proof of Lemma 3.11 is similar to that of Lemma 3.4, with the main caveat being that approximations such as Claim 3.7 no longer work, and so we need to use other estimates to bound the belief propagation function. To that end, we first state a technical lemma that we need to bound the "worst case" derivative of the belief propagation function.

Lemma 3.12 (Effectively [MNS16, Lemma 3.16])**.** *For any $0 < \varepsilon^* < 1$, there is some $d^*(\varepsilon^*)$, some $\lambda = \lambda(\varepsilon^*) < 1$ such that for all $1 > \varepsilon \geq \varepsilon^*$, $d \geq d^*$ there exists $\nu(\varepsilon, \varepsilon^*)$ such that if $|\xi| \leq \nu$ and $Y_i = \min\{\max\{X_{ui,k} - \xi, -1\}, 1\}$ for $k \geq 1$,*

$$\mathbb{E}\left[\sqrt{\frac{1-\varepsilon Y_i}{1+\varepsilon Y_i}} \,\middle|\, \sigma_u = +1\right] \leq \lambda.$$

In fact, we can take $\nu(\varepsilon, \varepsilon^*) = \frac{(\varepsilon^*)^2(1-\varepsilon)^{1/4}}{8\sqrt[4]{2}}$.

Because the proof of this lemma is near verbatim that of [MNS16, Lemma 3.16], we defer its proof to Subsection A.

First, we prove an analogue of Lemma 3.6 in the setting of large ε. For technical reasons, in the following we define the truncated random variables

$$Y_{ui,k} = \max\{-1, X_{ui,k} - |X_{ui,k} - Z_{ui,k}|\},$$

and

$$Y'_{ui,k} = \min\{X_{ui,k} + |X_{ui,k} - Z_{ui,k}|, 1\}.$$

These random variables effectively behave like their counterparts from before, in terms of recording the "worst case" adversary. In fact, we observe that this definition of $Y_{ui,k}$ and $Y'_{ui,k}$ is quite natural; -1 and 1 are the extreme points that the adversary can perturb the beliefs. The reason why such a truncation is necessary is basically because when we apply equation 5, we need to ensure that $\frac{1-\varepsilon Y_{ui,k}}{1+\varepsilon Y_{ui,k}}$ avoids the singularity $-\varepsilon^{-1}$ of $\frac{1-\varepsilon x}{1+\varepsilon x}$.

Lemma 3.13. *For any $0 < \varepsilon^* < 1$, there is some $C(\varepsilon^*) > 0$ such that for every $1 > \varepsilon > \varepsilon^*$ and d such that $\varepsilon^2 d > C$, the following holds: let $r \geq 0$, for any vertex u at level $t - r - 1$ with children $u1, \ldots, ud$ and suppose $\xi := \mathbb{E}\left[\max_{\text{adversary}} |Y_{ui,r} - X_{ui,r}| \,\middle|\, \sigma_{ui} = +1\right] \leq \frac{(1-\varepsilon)\nu(\varepsilon,\varepsilon^*)}{d}$, where ν is as in Lemma 3.12, then*

$$\mathbb{E}\left[\max_{\text{adversary}} \left| \left(\prod_{i=1}^{d} \frac{1-\varepsilon X_{ui,r}}{1+\varepsilon X_{ui,r}}\right)^{1/2} - \left(\prod_{i=1}^{d} \frac{1-\varepsilon Y_{ui,r}}{1+\varepsilon Y_{ui,r}}\right)^{1/2} \right| \,\middle|\, \sigma_u = +1\right] \leq \frac{\xi}{100}.$$

Furthermore, the above holds with $Y'_{ui,r}$ replacing $Y_{ui,r}$.

Proof of Lemma 3.13. First, by applying Markov's inequality on the hypothesis of the claim, note that for any $1 \leq i \leq d$ we have

$$\mathbb{P}\left[\max_{\text{adversary}} |Y_{ui,r} - X_{ui,r}| \geq \nu \,\middle|\, \sigma_u = +1\right] \leq \frac{1-\varepsilon}{d}. \tag{11}$$

Because of the independence of x_{ui} from x_{uj} for $i \neq j$, we have for any $I \subset [d]$ such that $I = \{i_1, \ldots, i_m\}$,

$$\mathbb{P}\left[\max_{\text{adversary}} |Y_{ui,r} - X_{ui,r}| \geq \nu \,\forall i \in I \;\middle|\; \sigma_u = +1\right] = \prod_{i=1}^{m} \mathbb{P}\left[\max_{\text{adversary}} |Y_{ui,r} - X_{ui,r}| \geq \nu \;\middle|\; \sigma_u = +1\right] \leq \left(\frac{1-\varepsilon}{d}\right)^{|I|}.$$

Let $f(x) = \sqrt{\frac{1-\varepsilon x}{1+\varepsilon x}}$. By the independence of x_{ui} from x_{uj} for $i \neq j$, we apply the mean value theorem to obtain

$$\begin{aligned}
&\mathbb{E}\left[\max_{\text{adversary}} \left|\prod_{i=1}^{d} f(X_{ui,r}) - \prod_{i=1}^{d} f(Y_{ui,r})\right| \;\middle|\; \sigma_u = +1\right] \\
&\leq \sum_{i=1}^{d} \mathbb{E}\left[\max_{\text{adversary}} |X_{ui,r} - Y_{ui,r}| \;\middle|\; \sigma_{ui} = +1\right] \cdot \frac{1}{(1-\varepsilon)^{3/2}} \cdot \prod_{\substack{j=1 \\ j\neq i}}^{d} \mathbb{E}\left[\max_{\text{adversary}} \max_{\substack{y_1 \in [Y_{u1,r}, X_{u1,r}] \\ \vdots \\ y_d \in [Y_{ud,k}, X_{ud,k}]}} f(y_j) \;\middle|\; \sigma_u = +1\right] \\
&\leq \frac{d \cdot \xi}{(1-\varepsilon)^{3/2}} \mathbb{E}\left[\prod_{j=1}^{d-1} \max_{\text{adversary}} f(Y_{uj,r}) \;\middle|\; \sigma_u = +1\right]
\end{aligned}$$

where the second inequality holds by monotonocity of f and the symmetry of subtrees rooted at $u1, \ldots, ud$. For the first inequality, note that since $\left|\frac{df}{dx}\right|$ is convex on $[-1, 1]$ with a minimum at $x = \frac{1}{2}$ and we truncated the random variables $Y_{ui,k}$ so that we always have $X_{ui,k}, Y_{ui,k} \in [-1, 1]$, we can bound $\left|\frac{df}{dx}\right|$ by its value at the endpoints

$$\max\left\{\left|\frac{df}{dx}(-1)\right|, \left|\frac{df}{dx}(1)\right|\right\} = \max\left\{\frac{\varepsilon}{\sqrt{1-\varepsilon}(\varepsilon+1)^{3/2}}, \frac{\varepsilon}{\sqrt{1+\varepsilon}(1-\varepsilon)^{3/2}}\right\} \leq \frac{1}{(1-\varepsilon)^{3/2}}.$$

For the first inequality, we also used the fact that

$$\mathbb{E}\left[\max_{\text{adversary}} |X_{ui,r} - Y_{ui,r}| \;\middle|\; \sigma_{ui} = +1\right] = \mathbb{E}\left[\max_{\text{adversary}} |X_{ui,r} - Y_{ui,r}| \;\middle|\; \sigma_u = +1\right] \leq \xi$$

which is fundamentally because $|X_{ui,r} - Y_{ui,r}|$ measures the worst-case *magnitude* by which an adversary can perturb the beliefs.

Let $\mathcal{B}_I$ be the event that $\max_{\text{adversary}} |Y_{ui,r} - X_{ui,r}| \geq \nu$ for all $i \in I$ and $\max_{\text{adversary}} |Y_{ui,r} - X_{ui,r}| \leq \nu$ for all $i \notin I$.

By the law of total probability,

$$\begin{aligned}
&\mathbb{E}\left[\prod_{j=1}^{d-1} \max_{\text{adversary}} f(Y_{uj,r}) \;\middle|\; \sigma_u = +1\right] \\
&\leq \sum_{I\subset[d]} \mathbb{P}[\mathcal{B}_I \mid \sigma_u = +1]\, \mathbb{E}\left[\prod_{j=1}^{d-1} \max_{\text{adversary}} f(Y_{uj,r}) \;\middle|\; \mathcal{B}_I\,,\ \sigma_u = +1\right] \\
&\leq \sum_{I\subset[d]} \mathbb{P}[\mathcal{B}_I \mid \sigma_u = +1] \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^{\frac{|I|}{2}} \mathbb{E}\left[\prod_{i\notin I} \sqrt{\frac{1-\varepsilon(\max\{-1, X_{ui,r}-\nu\})}{1+\varepsilon(\max\{-1, X_{ui,r}-\nu\})}} \;\middle|\; \mathcal{B}_I\,,\ \sigma_u = +1\right] \\
&\leq \sum_{I\subset[d]} \mathbb{P}[\mathcal{B}_I \mid \sigma_u = +1] \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^{\frac{|I|}{2}} \prod_{i\notin I} \mathbb{E}\left[\sqrt{\frac{1-\varepsilon(\max\{-1, X_{ui,r}-\nu\})}{1+\varepsilon(\max\{-1, X_{ui,r}-\nu\})}} \;\middle|\; \mathcal{B}_I\,,\ \sigma_u = +1\right].
\end{aligned}$$

Next, we use Lemma 3.12 to bound $\mathbb{E}\left[\sqrt{\frac{1-\varepsilon(\max\{-1,X_{ui,r}-\nu\})}{1+\varepsilon(\max\{-1,X_{ui,r}-\nu\})}} \;\middle|\; \mathcal{B}_I\,,\ \sigma_u = +1\right]$. Now, we note that

$$\begin{aligned}
&\mathbb{E}\left[\sqrt{\frac{1-\varepsilon(\max\{-1, X_{ui,r}-\nu\})}{1+\varepsilon(\max\{-1, X_{ui,r}-\nu\})}} \;\middle|\; \mathcal{B}_I\,,\ \sigma_u = +1\right] \\
&= \mathbb{E}\left[\sqrt{\frac{1-\varepsilon(\max\{-1, X_{ui,r}-\nu\})}{1+\varepsilon(\max\{-1, X_{ui,r}-\nu\})}} \;\middle|\; |\max_{\text{adversary}} |Y_{ui,k} - X_{ui,k}| < \nu\,,\ \sigma_u = +1\right] \\
&\leq \frac{\mathbb{E}\left[\sqrt{\frac{1-\varepsilon(\max\{-1,X_{ui,r}-\nu\})}{1+\varepsilon(\max\{-1,X_{ui,r}-\nu\})}} \;\middle|\; \sigma_u = +1\right]}{\mathbb{P}[\max_{\text{adversary}} |Y_{ui,r} - X_{ui,r}| < \nu \mid \sigma_u = +1]} \\
&\leq \frac{\lambda}{1-\frac{1}{d}} \\
&\leq \widetilde{\lambda},
\end{aligned}$$

where the second equality follows since x_{ui} is independent from x_{uj} for $i \neq j$, the penultimate inequality follows by applying Lemma 3.12 to bound the numerator and applying (11) to bound the denominator and the last equality follows for some $\widetilde{\lambda} < 1$ by taking sufficiently large $C(\varepsilon^*)$ so that $d > C(\varepsilon^*)$ is sufficiently large. Putting all this together, it follows that

$$\begin{aligned}
&\mathbb{E}\left[\prod_{j=1}^{d-1} \max_{\text{adversary}} f(Y_{uj,r}) \;\middle|\; \sigma_u = +1\right] \\
&\leq \sum_{m=0}^{d-1} \binom{d-1}{m} \left(\frac{1-\varepsilon}{d}\right)^m \left(\frac{2}{1-\varepsilon}\right)^{\frac{m}{2}} \widetilde{\lambda}^{d-m-1} \\
&\leq \left(\widetilde{\lambda} + O\left(\frac{1}{d}\right)\right)^{d-1}
\end{aligned}$$

for sufficiently large $d > d^*$. Combining these bounds, we obtain

$$\mathbb{E}\left[\max_{\text{adversary}}\left|\prod_{i=1}^{d} f(X_{ui,r}) - \prod_{i=1}^{d} f(Y_{ui,r})\right| \,\middle|\, \sigma_u = +1\right] \le \frac{d\left(\widetilde{\lambda} + O\left(\frac{1}{d}\right)\right)^{d-1}}{(1-\varepsilon)^{3/2}} \cdot \xi \le \frac{\xi}{100}$$

by taking sufficiently large $C(\varepsilon^*)$ to ensure that

$$\left(\frac{\lambda}{1-\frac{1}{d}} + O\left(\frac{1}{d}\right)\right)^{d-1} \ge d^{-1}\left(\frac{\lambda}{1-\frac{1}{d}} + O\left(\frac{1}{d}\right)\right)^{\frac{d}{2}} \ge d^{-1}\left(\frac{\lambda}{1-\frac{1}{d}} + O\left(\frac{1}{d}\right)\right)^{-\frac{C}{4}\log(1-\varepsilon)} \ge \frac{(1-\varepsilon)^{3/2}}{100d},$$

as desired. □

With this inductive contraction lemma in place, the proof of Lemma 3.11 is immediate.

Proof of Lemma 3.11. Let $\nu = \frac{(1-\varepsilon)\nu(\varepsilon,\varepsilon^*)}{d}$. We induct on $r \ge 1$ to prove that

$$\mathbb{E}\left[\max_{\text{adversary}} |X_{u,t-r} - Z_{u,t-r}| \,\middle|\, \sigma_u = +1\right] \le 2^{-(r-1)}\nu. \tag{12}$$

The base case of $r = 1$ is Lemma 3.4 by taking C sufficiently large, and the inductive step follows by applying Lemma 3.13. □

4 (c,k)-spread adversary

Finally, we introduce a deterministic adversary - the (c,k)-spread adversary - for which we can also accurately reconstruct the spin of the root.

Definition 4.1. *Let $c > 0$ and $k > 0$ be given. The* (c,k)-spread adversary *is an adversary that receives leaf signs σ_L and is allowed to flip c of the leaf spins of his choice among every height k subtree.*

Theorem 4.2. *There exists a universal constant $C > 0$ with the following property. For any $c > 0$ and $d, \varepsilon > 0$ such that $\varepsilon^2 d > C\log(d)$, there exists some $k(c)$ and t_0 such that if $t > t_0$ then there exists an algorithm $\texttt{ALG}$ which takes as input an element of $\{\pm 1\}^{d^t}$ and outputs an element of $\Delta(\pm 1)$ with the property that for any (c,k)-spread adversary $\texttt{A}$ acting on the broadcast, then*

$$\mathop{\mathbb{E}}_{(\sigma_R,\sigma_L)\sim\mathcal{D}_{d,\varepsilon,t},\, x} d_{TV}(\{\sigma_R \mid \sigma_L\}, \texttt{ALG}(A_\rho(\sigma_L))) \le \delta\,.$$

As before, we implement a contraction argument. Fix d, ε. Note, however, that for instance when $c \ge d$, we cannot expect the base case of our contraction argument as in Lemma 3.4 to work because the adversary could choose to concentrate all of his flips in one subtree of height 1. For the base case of the induction, we instead look at a subtree of height k; this is captured by the following Theorem 4.4.

Definition 4.3. *The* c-flip adversary *is an adversary that is allowed to flip c of the leaves of his choice after witnessing the entire broadcast tree process.*

Theorem 4.4. *There exists $C > 0$ such that the following holds: for every $c, d, \varepsilon, \delta > 0$ there exists an algorithm $\texttt{A}$ and a constant $K(c) > 0$ such that if $d\varepsilon^2 > C$ and if $t \ge \log(K\delta^{-1})$, there exists an algorithm $\texttt{ALG}$: $\{\pm 1\}^{d^t} \to \Delta(\pm 1)$ such that for any c-flip adversary $\texttt{A}$, we have*

$$\mathop{\mathbb{E}}_{(\sigma_R,\sigma_L)\sim\mathcal{D}_{d,\varepsilon,t}} d_{TV}(\{\sigma_R \mid \sigma_L\}, \texttt{ALG}(\texttt{A}(\sigma_L))) \le \delta.$$

Proof of Theorem 4.4. Let $\psi = \min\{\delta/(8c), \log(1+\delta/4)/(4c)\}$. To define the algorithm $\mathtt{A}$, consider first the following random process $\mathcal{N}$. A draw $\sigma_L \sim \mathcal{N}$ corresponds to the output $\{\pm 1\}^{(d-1)^t}$ from running the broadcast on tree process with flip probability $\frac{1-\varepsilon}{2}$ down a d-regular tree (i.e., each non-leaf node has $d-1$ children) and then passing each bit of the leaves through independent binary symmetric channels with flip probability $\frac{1-\psi}{2}$.

We also define the "clean" process $\mathcal{T}$: a draw $\sigma_L \sim \mathcal{T}$ corresponds to the $\{\pm 1\}^{(d-1)^t}$ spins of the leaves at level t of a broadcast on tree process with flip probability $\frac{1-\varepsilon}{2}$.

We let

$$\mathtt{A}_{\mathcal{N}}(x) = \mathbb{P}_{\mathcal{N}}(\sigma_R = 1 \,\Big|\, \sigma_L = x) - \mathbb{P}_{\mathcal{N}}(\sigma_R = -1 \,\Big|\, \sigma_L = x)\,,$$

and

$$\mathtt{A}_{\mathcal{T}}(x) = \mathbb{P}_{\mathcal{T}}(\sigma_R = 1 \,\Big|\, \sigma_L = x) - \mathbb{P}_{\mathcal{T}}(\sigma_R = -1 \,\Big|\, \sigma_L = x)\,,$$

both of which can be computed by belief propagation. Ultimately, we will take $\mathtt{A} = \mathtt{A}_{\mathcal{N}}$.

The goal is to find some K such that for $t \geq \log(K\delta^{-1})$, we have the bound

$$\mathop{\mathbb{E}}_{\sigma_L \sim \mathcal{T}} \max_{v_1,\ldots,v_c \in L} \mathop{\mathbb{E}}_{x \sim \mathrm{Ber}(\frac{1-\psi}{2})^{(d-1)^t}} \left| \mathtt{A}_{\mathcal{N}}\left(\sigma_L \oplus x \oplus \bigoplus_{i=1}^{c} e_{v_i}\right) - \mathtt{A}_{\mathcal{T}}(\sigma_L) \right| \leq \delta. \tag{13}$$

We claim that [MNS16, Lemmas 3.4, 3.5] gives the quantitative bound that there is some constant $K(c) > 0$ such that if $t \geq \log(K\delta^{-1})$, then

$$\mathop{\mathbb{E}}_{\sigma_L \sim \mathcal{N}} |\mathtt{A}_{\mathcal{N}}(\sigma_L)| \geq 1 - \frac{4}{\varepsilon^2 d}. \tag{14}$$

Define S to be the sum of all the spins drawn from the process $\mathcal{N}$. Then [MNS16, Lemmas 3.4, 3.5] gives

$$\begin{cases} \mathbb{E}[S|\sigma_{\mathsf{root}} = +1] = \psi\varepsilon^t d^t, \\ \mathrm{Var}[S|\sigma_{\mathsf{root}} = +1] = (1-\psi^2)d^t + (1-\varepsilon^2)\psi^2 \frac{((\varepsilon^2 d)^t - 1)d^t}{\varepsilon^2 d - 1}. \end{cases}$$

Since $\psi \asymp_c \delta$, there exists $K(c)$ such that if $t \geq \log(K\delta^{-1})$, then $(\varepsilon^2 d)^t > \frac{\psi^{-2}}{2}$, meaning that $(1-\psi^2)d^t$ is dominated by $(1-\varepsilon^2)\psi^2 \frac{((\varepsilon^2 d)^t - 1)d^t}{\varepsilon^2 d - 1}$. In particular, in this case we have the (crude) bound of

$$\mathrm{Var}[S|\sigma_{\mathsf{root}} = +1] \leq \frac{3}{2}(1-\varepsilon^2)\psi^2 \frac{((\varepsilon^2 d)^t - 1)d^t}{\varepsilon^2 d - 1}.$$

By Chebyshev's inequality, it follows that if $t \geq \log(K\delta^{-1})$ then

$$\mathbb{P}[S > 0|\sigma_{\mathsf{root}} = +1] \geq 1 - \frac{2}{\varepsilon^2 d}.$$

Because $\mathrm{sign}(\mathtt{A}_{\mathcal{N}}(x))$ is the optimal estimator of σ_{root} given σ_L, we have

$$\begin{aligned} \frac{1 + \mathbb{E}\,|\mathtt{A}_{\mathcal{N}}(x)|}{2} &= \mathbb{P}[\mathrm{sign}(\mathtt{A}_{\mathcal{N}}(x)) = +1|\sigma_{\mathsf{root}} = +1] \\ &\geq \mathbb{P}[\mathrm{sign}(S) = +1|\sigma_{\mathsf{root}} = +1] \\ &\geq 1 - \frac{2}{\varepsilon^2 d} \end{aligned}$$

which upon rearranging gives (14).

Combining (14) with the proof of [MNS16, Theorem 3.3], we have that if $t \geq \log(K\delta^{-1})$, then

$$\mathop{\mathbb{E}}_{\sigma_L\sim\mathcal{T}} \mathop{\mathbb{E}}_{x\sim\mathrm{Ber}(\frac{1-\psi}{2})^{(d-1)^t}} |\mathtt{A}_{\mathcal{N}}(x\oplus\sigma_L) - \mathtt{A}_{\mathcal{T}}(\sigma_L)| \leq \frac{1}{2} \mathop{\mathbb{E}}_{\sigma_L\sim\mathcal{T}_{t-1}} \mathop{\mathbb{E}}_{x\sim\mathrm{Ber}(\frac{1-\psi}{2})^{(d-1)^{t-1}}} |\mathtt{A}_{\mathcal{N}}(x\oplus\sigma_L) - \mathtt{A}_{\mathcal{T}_{t-1}}(\sigma_L)|.$$

By iterating this contraction result, it follows that if $t \geq \log(\max\{2,K\}\delta^{-1})$, then we have

$$\mathop{\mathbb{E}}_{\sigma_L\sim\mathcal{T}} \mathop{\mathbb{E}}_{x\sim\mathrm{Ber}(\frac{1-\psi}{2})^{(d-1)^t}} |\mathtt{A}_{\mathcal{N}}(x\oplus\sigma_L) - \mathtt{A}_{\mathcal{T}}(\sigma_L)| \leq \delta/2.$$

By the triangle inequality, in order to obtain the desired conclusion it suffices therefore to prove

$$\mathop{\mathbb{E}}_{\sigma_L\sim\mathcal{T}} \max_{v_1,\ldots,v_c\in L} \mathop{\mathbb{E}}_{x\sim\mathrm{Ber}(\frac{1-\psi}{2})^{(d-1)^t}} \left|\mathtt{A}_{\mathcal{N}}\left(\sigma_L\oplus x\oplus\bigoplus_{i=1}^{c} e_{v_i}\right) - \mathtt{A}_{\mathcal{N}}(x\oplus\sigma_L)\right| \leq \delta/2.$$

By symmetry and the triangle inequality, it suffices to prove that

$$\mathop{\mathbb{E}}_{\sigma'_L\sim\mathcal{T}} \max_{v_1,\ldots,v_c\in L} \mathop{\mathbb{E}}_{x\sim\mathrm{Ber}(\frac{1-\psi}{2})^{(d-1)^t}} \left|\mathbb{P}_{\mathcal{N}}\left(\sigma_R = 1|\sigma_L = \sigma'_L\oplus x\oplus\bigoplus_{i=1}^{c} e_{v_i}\right) - \mathbb{P}_{\mathcal{N}}(\sigma_R = 1|\sigma_L = \sigma'_L\oplus x)\right| \leq \delta/4. \quad (15)$$

In fact, we will show that the difference above is at most $\delta/4$ for *any* choice of $\sigma'_L, v_1, \ldots, v_c, x$. For any $x_L \in \{\pm1\}^{(d-1)^t}$, using Bayes' rule, we can write

$$\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}[\sigma_R = 1 \,\big|\, \sigma_L = x_L] = \frac{\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}[\sigma_L = x_L|\sigma_R = 1]\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}[\sigma_R = 1]}{\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}[\sigma_L = x_L]}$$

and for any choice of $v_1, \ldots, v_c$,

$$\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}\left[\sigma_R = 1 \,\Big|\, \sigma_L = x_L\oplus\bigoplus_{i=1}^{c} e_{v_i}\right] = \frac{\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}\left[\sigma_L = x_L\oplus\bigoplus_{i=1}^{c} e_{v_i}|\sigma_R = 1\right]\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}[\sigma_R = 1]}{\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}[\sigma_L = x_L\oplus e_v]}.$$

Note that $(\frac{1-\psi}{2})^c \leq \frac{\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}\left[\sigma_L=x_L\oplus\bigoplus_{i=1}^{c} e_{v_i}\Big|\sigma_R=1\right]}{\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}[\sigma_L=x_L|\sigma_R=1]} \leq (\frac{1+\psi}{2})^c$ and $(\frac{1-\psi}{2})^c \leq \frac{\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}\left[\sigma_L=x_L\oplus\bigoplus_{i=1}^{c} e_{v_i}\right]}{\mathbb{P}[\sigma_L=x_L]} \leq (\frac{1+\psi}{2})^c$. That is,

$$1-2\psi c \leq \left(\frac{\frac{1-\psi}{2}}{\frac{1+\psi}{2}}\right)^c \leq \frac{\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}\left[\sigma_R = 1 \,\Big|\, \sigma_L = x_L\oplus\bigoplus_{i=1}^{c} e_{v_i}\right]}{\mathbb{P}_{(\sigma_R,\sigma_L)\sim\mathcal{N}}[\sigma_R = 1 \,\big|\, \sigma_L = x_L]} \leq \left(\frac{\frac{1+\psi}{2}}{\frac{1-\psi}{2}}\right)^c \leq e^{4\psi c}. \quad (16)$$

Our choice of ψ gives the desired inequality 15. □

As mentioned earlier, we use Theorem 4.4 as a primitive to kickstart our contraction argument. To be more precise, our algorithm and the relevant parameters in our proof of Theorem 4.2 are as follows. Fix d, ε, δ and let $K(c)$ be the constant from Theorem 4.4 and set

$$k = K(c)\,,\ t_0 = 10^5(\log(d) - \log(1-\varepsilon) + \log(\delta^{-1})). \quad (17)$$

Given the output of a broadcast on tree process that was run up till depth t, we partition the bottom k levels of the tree into subtrees of size d^k (for an illustration, refer to Figure 1). On the leaves of each of these subtrees, run the algorithm from Theorem 4.4 and suppose the outputs are beliefs $Z_{i,k}$ $(1 \leq i \leq d^{t-k})$ for the d^{t-k} nodes at level $t-k$ of the broadcast tree. Now, recursively apply belief propagation until we

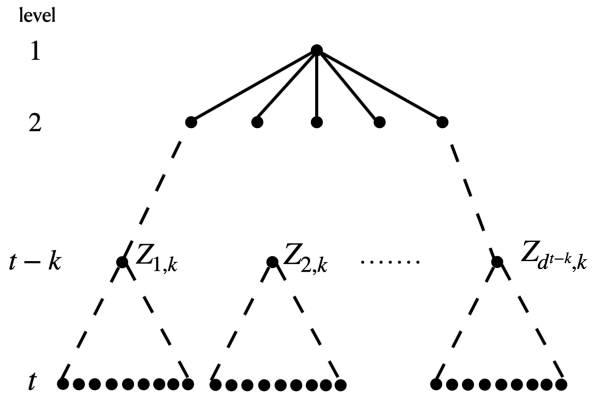

Figure 1: An illustration of the partition of the levels of the tree in the algorithm.

reach the root and output the belief thus obtained. Specifically, recursively define

$$Z_{j,i} = BP(Z_{dj+1,i-1}, Z_{dj+2,i-1}, \ldots, Z_{d(j+1)-1,i-1})$$

for $k \le i \le t$ and $1 \le j \le d^i$, where $BP : \mathbb{R}^d \to \mathbb{R}$ is the belief propagation function given by

$$BP(X_1, \ldots, X_d) = \frac{\prod_{i=1}^d (1 + \varepsilon X_i) - \prod_{i=1}^d (1 - \varepsilon X_i)}{\prod_{i=1}^d (1 + \varepsilon X_i) + \prod_{i=1}^d (1 - \varepsilon X_i)}.$$

We can restate Theorem 4.4 in terms of an upper bound on the maximum perturbation from the ground truth $X_{u,k}$.

Lemma 4.5 (Consequence of Theorem 4.4 for the parameters in (17))**.** *There exists an absolute constant C such that if $d\varepsilon^2 \ge C \log(d)$ then for any vertex u at level $t-k$ where $t \ge t_0$ where k and t_0 are as in* (17)*, we have*

$$\mathbb{E} \max_{\text{adversary}} \left[|X_{u,k} - Z_{u,k}| \,\middle|\, \sigma_u = +1 \right] \le \frac{\nu(\varepsilon, \varepsilon^*)}{10d}.$$

Proof of Theorem 4.2. We induct on $r \ge 1$ to prove that

$$\mathbb{E} \left[\max_{\text{adversary}} |X_{u,t-r} - Z_{u,t-r}| \,\middle|\, \sigma_u = +1 \right] \le 2^{-(r-k)} \nu. \tag{18}$$

The base case of $r = k$ is Lemma 4.5, and the inductive step follows by applying Lemma 3.6 when $\varepsilon \le \varepsilon^*$ and Lemma 3.13 when $\varepsilon > \varepsilon^*$. To finish up, it suffices to take $t_0 \asymp \log(\delta^{-1})$ and iteratively apply (18). □

5 Information-theoretic lower bounds

First, we establish the folklore result that the ρ-fraction adversary is very powerful and can in fact completely erode any information in the leaves about the posterior distribution of the root.

Definition 5.1. *Fix $\rho > 0$. The* ρ-fraction adversary *is allowed to look at all the d^t spins of the leaves and then flip the signs of ρd^t leaves of his choosing.*

Theorem 5.2. *For every ρ, d and ε, there exists some t_0 such that if the ρ-fraction adversary is allowed to corrupt the leaves of the broadcast on tree process with parameters (d, ε) that was run up to level $t \ge t_0$, it is information*

theoretically impossible to recover the root vertex. In fact, we prove that for $(\sigma_R, \sigma_L) \sim \mathcal{D}_{d,\varepsilon t}$ there exists a ρ-fraction adversary $\mathtt{A}$ for which we have

$$d_{TV}(\{\sigma_L \mid \sigma_R = 1\}, \{A(\sigma_L) \mid \sigma_R = -1\}) \leq e^{-\Omega(t)}.$$

The quantitative bound in the theorem implies that it is impossible to have an algorithm $\mathtt{A}$ that recovers the root vertex with an advantage larger than $e^{-\Omega(t)}$.

Proof. Set $t_0 = \log \rho^{-1}/(\log \frac{1+2\varepsilon}{4\varepsilon}) + 1$ and throughout this proof we consider $t \geq t_0$. Let $\mathcal{D}_t^+$ be the distribution of the leaves at level t of a broadcast process with parameters (ε, d) with root being $+1$ and $\mathcal{D}_t^-$ be the distribution with root being -1. And, let $D_{\leq t}^+, D_{\leq t}^-$ be the analogous joint distributions on the first t levels of the broadcast tree.

We claim that it suffices to exhibit a coupling $(\mathbf{x}, \pi_t(\mathbf{x})) \in \{\pm 1\}^{(d-1)^t} \times \{\pm 1\}^{(d-1)^t}$ between $\mathbf{x} \sim \mathcal{D}_t^+$ and $\pi_t(\mathbf{x}) \sim \mathcal{D}_t^-$ such that with probability $1 - e^{-\Omega(t)}$, we have $\mathrm{dist}(\mathbf{x}, \pi_t(\mathbf{x})) \leq (4e\varepsilon)^t (d-1)^t$ where dist refers to Hamming distance between the two strings.

Indeed, consider the following ρ-fraction adversary $\mathbf{A}$: for $\mathbf{x} \sim \mathcal{D}_t^+$, if $\mathrm{dist}(\mathbf{x}, \pi_t(\mathbf{x})) \leq (4e\varepsilon)^t (d-1)^t$ then let $\mathbf{A}(\mathbf{x}) = \pi_t(\mathbf{x})$, where by our choice of parameters we have that $(4e\varepsilon)^t \leq \rho$. Otherwise, let $\mathbf{A}(\mathbf{x}) = \mathbf{x}$. Let $\mathcal{A}_t^+$ denote the distribution corresponding to $\mathbf{A}(\mathbf{x})$ where $\mathbf{x} \sim \mathcal{D}_t^+$. In particular, this means that

$$d_{TV}(\mathcal{D}_t^-, \mathcal{A}_t^+) \leq e^{-\Omega(t)}.$$

It suffices therefore to exhibit the coupling π_t as claimed. Define $\pi_t(\mathbf{x})$ for $\mathbf{x} \sim \mathcal{D}_t^+$ to act as follows: (in the following, let $\xi = \frac{4\varepsilon}{1+2\varepsilon}$) here one should think of marked vertices as the vertices where the adversary did not make any changes, and the goal of the adversary is to flip unmarked vertices that are labelled '+' to '-'

- Sample $\mathbf{y} \sim \mathcal{D}_{\leq t}^+ \,\Big|\, \mathbf{x}$ (that is, $\mathcal{D}_{\leq t}^+$ conditioned on the leaves of the tree taking labels $\mathbf{x}$).
- Let $\mathbf{y}'(\mathsf{root}) = -1$.
- Traverse down the tree starting from the the nodes at level 1, that is, from the children of the root. If the current node v is marked, let $\mathbf{y}'(v) = \mathbf{y}(v)$ and mark all its children. Otherwise, if v is unmarked and $\mathbf{y}(v) = -1$, then mark v and all its children and set $\mathbf{y}'(v) = \mathbf{y}(v)$. Else, if v is unmarked and $\mathbf{y}(v) = +1$, with probability ξ leave it unmarked and set $\mathbf{y}'(v) = -1$. Otherwise with probability $1 - \xi$, mark v and its children and set $\mathbf{y}'(v) = +1$.
- Let $\pi_t(\mathbf{x})$ be the restriction of $\mathbf{y}'$ to the vertices at level t.

In order to prove that this coupling has the desired properties, we first show that with high probability, the Hamming distance between $\mathbf{x}$ and $\pi_t(\mathbf{x})$ is small.

Note that if for a vertex v at the t-th level, we have $\pi_t(\mathbf{x}(v)) \neq \mathbf{x}(v)$ then v has to be unmarked, and furthermore there is a path of t unmarked vertices connecting v to the root. The latter occurs with probability at most ξ^t. Applying Markov's inequality immediately gives

$$\mathbb{P}[\mathrm{dist}(\mathbf{x}, \pi_t(\mathbf{x})) \geq \rho^t (d-1)^t] \leq e^{-\Omega(t)}.$$

Next, we need to show that $\pi_t(\mathbf{x}) \sim \mathcal{D}_t^-$. We will actually show that $\mathbf{y}' \sim \mathcal{D}_{\leq t}^-$ by inducting on t. For the base case, note that for any vertex w on level 1, we have that

$$\mathbb{P}[\mathbf{y}'(w) = -] = \left(\frac{1}{2} - \varepsilon\right) + \left(\frac{1}{2} + \varepsilon\right)\xi = \frac{1}{2} + \varepsilon$$

as desired.

For $s \le t$, we write $\mathbf{y}'_{\le s}$ to denote the restriction of $\mathbf{y}'$ to labels of vertices in levels $r \le s$. Note that it suffices to prove that for any vertex v at level $s+1$ with parent w,

$$\mathbb{P}[\mathbf{y}'(v) = +|\mathbf{y}'_{\le s}] = \frac{1}{2} + \mathbf{y}'(w)\varepsilon.$$

Indeed, note that

$$\begin{aligned}
&\mathbb{P}[\mathbf{y}'(v) = + \mid \mathbf{y}'_{\le s}] \\
&\quad = \mathbb{P}[\mathbf{y}'(v) = + \mid \mathbf{y}'_{\le s}, w \text{ marked}]\mathbb{P}[w \text{ marked} \mid \mathbf{y}'_{\le s}] + \mathbb{P}[\mathbf{y}'(v) = + \mid \mathbf{y}'_{\le s}, w \text{ unmarked}]\mathbb{P}[w \text{ unmarked} \mid \mathbf{y}'_{\le s}] \\
&\quad = \left(\frac{1}{2} + \mathbf{y}'(w)\varepsilon\right)\mathbb{P}[w \text{ marked} \mid \mathbf{y}'_{\le s}] + \mathbb{P}[\mathbf{y}'(v) = + \mid \mathbf{y}'(w) = -, w \text{ unmarked}]\mathbb{P}[w \text{ unmarked} \mid \mathbf{y}'_{\le s}].
\end{aligned}$$

Since

$$\begin{aligned}
\mathbb{P}[\mathbf{y}'(v) = + \mid \mathbf{y}'(w) = -, w \text{ unmarked}] &= 1 - \mathbb{P}[\mathbf{y}'(v) = - \mid \mathbf{y}'(w) = -, \mathbf{y}(w) = +, w \text{ unmarked}] \\
&= 1 - \left(\frac{1}{2} - \varepsilon + \left(\frac{1}{2} + \varepsilon\right)\xi\right) = \frac{1}{2} - \varepsilon \\
&= \frac{1}{2} + \mathbf{y}'(w)\varepsilon
\end{aligned}$$

and $\mathbb{P}[w \text{ marked} \mid \mathbf{y}'_{\le s}] + \mathbb{P}[w \text{ unmarked} \mid \mathbf{y}'_{\le s}] = 1$, the above indeed implies that

$$\mathbb{P}[\mathbf{y}'(v) = +|\mathbf{y}'_{\le s}] = \frac{1}{2} + \mathbf{y}'(w)\varepsilon$$

as desired. □

Another natural question is whether the bound on ρ in Theorem 1.4 is tight; that is, can we prove any information theoretic lower bounds on ρ-semi-random adversaries? To that end, we have the following upper bound of $\rho \lesssim \varepsilon$. This means that at least in the range of $d\varepsilon^2 \gtrsim \log \frac{d}{1-\varepsilon}$ being a log-factor above the Kesten-Stigum threshold, belief propagation achieves the optimal ρ (up to constant factors) in terms of being robust against ρ-semi-random adversaries.

Theorem 5.3. *For every d and ε, there exists t_0 such that if a $\left(\frac{4\varepsilon}{1+2\varepsilon}\right)$-semi-random adversary is allowed to corrupt the leaves of the broadcast on tree process with parameters (d, ε) that was run up to level $t \ge t_0$, then it is information theoretically impossible to recover the root vertex. That is, for $(\sigma_R, \sigma_L) \sim \mathcal{D}_{d,\varepsilon,t}$ there exists a $\left(\frac{4\varepsilon}{1+2\varepsilon}\right)$-semi-random adversary $\boldsymbol{A}$ such that*

$$d_{TV}(\{\sigma_L \mid \sigma_R = 1\}, \{\boldsymbol{A}(\sigma_L) \mid \sigma_R = -1\}) \le e^{-\Omega(t)}.$$

Proof. Set $t_0 = 1/(\log(2e\varepsilon)^{-1})$. We adapt the adversary that we used to prove the information theoretic bound in Theorem 5.2. Retain the notations as in the proof of Theorem 5.2, where π_t defines a coupling on $(\mathcal{D}_t^+, \mathcal{D}_t^-)$. Consider the following adversary $\mathbf{B}$: for $\mathbf{x} \sim \mathcal{D}_t^+$, if all the sites of $\pi_t(\mathbf{x}) \triangle \mathbf{x}$ (i.e. the sites where the two vectors differ) are selected as possible sites for corruption, and we call this event *good*, then output $\pi_t(\mathbf{x})$. Otherwise, return $\mathbf{x}$. Let $\mathcal{B}_t^+$ denote the distribution corresponding to $\mathbf{B}(\mathbf{x})$ where $\mathbf{x} \sim \mathcal{D}_t^+$.

We claim that the good event happens with probability 1. To that end, we traverse down the tree and mark vertices as in the process described in Theorem 5.2 until we reach level $t-1$. For each of these marked vertices on level $t-1$, we simulate the process of choosing whether to mark its children with the coin flips in the $\left(\frac{4\varepsilon}{1+2\varepsilon}\right)$-semi-random adversary. In other words we couple random coin flips to mark the children and change their spins with the coin tosses in the adversary. The bound on the semi-random robust gain follows from Theorem 5.2 as well. □

Acknowledgements

We thank Guy Bresler and Po-Ling Loh for helpful conversations as this manuscript was being prepared.

A Deferred proofs

Proof of Corollary 1.6. We claim that an equivalent formulation of Theorem 1.4 is that for any $\delta', d, \varepsilon, \rho$ satisfying suitable hypothesis, there exists $t_0'(d, \delta')$ such that for any adversary $\mathtt{B}$ that takes as input σ_L and $S \subset [n]$ and is only allowed to flip the signs of the restriction $\sigma_L|_S$, there exists an algorithm $\mathtt{ALG}$ with the following guarantee:

$$\mathop{\mathbb{E}}_{(\sigma_R,\sigma_L)\sim\mathcal{D}_{d,\varepsilon,t},S\sim\Delta_{\rho,t}} d_{TV}\left(\{\sigma_R \mid \sigma_L\}, \mathtt{ALG}(\mathtt{B}(\sigma_L, S))\right) \le \delta'$$

for all $t \ge t_0'$ where $\Delta_{\rho,t}$ is the uniform distribution on subsets $S \subset [d^t]$ of cardinality $\frac{\rho d^t}{2} \le |S| \le 2\rho d^t$.

To this end, note that if $k \le m$ then

$$\mathop{\mathbb{E}}_{(\sigma_R,\sigma_L)\sim\mathcal{D}_{d,\varepsilon,t},S\sim\Delta_{\rho,t}} \max_{|S|=k} d_{TV}\left(\{\sigma_R \mid \sigma_L\}, \mathtt{ALG}(\mathtt{B}(\sigma_L, S))\right) \le \mathop{\mathbb{E}}_{(\sigma_R,\sigma_L)\sim\mathcal{D}_{d,\varepsilon,t},S\sim\Delta_{\rho,t}} \max_{|S|=m} d_{TV}\left(\{\sigma_R \mid \sigma_L\}, \mathtt{ALG}(\mathtt{B}(\sigma_L, S))\right).$$

Let $\Delta'_{\rho,t}$ be the distribution on subsets $S \subset [d^t]$ of cardinality $\rho d^t \le |S| \le 2\rho d^t$ where $\mathbb{P}[S] = \rho^{|S|}(1-\rho)^{n-|S|}$. Note that $\rho^m(1-\rho)^{n-m}$ is decreasing on the range $\rho d^t \le m \le 2\rho d^t$ since $\rho < \frac{1}{2}$. Consequently,

$$\begin{aligned}
\mathop{\mathbb{E}}_{(\sigma_R,\sigma_L),S\sim\Delta_{\rho,n}} d_{TV}\left(\{\sigma_R \mid \sigma_L\}, \mathtt{ALG}(\mathtt{B}(\sigma_L, S))\right) &\le \mathop{\mathbb{E}}_{(\sigma_R,\sigma_L),S\sim\Delta'_{\rho,n}} d_{TV}\left(\{\sigma_R \mid \sigma_L\}, \mathtt{ALG}(\mathtt{B}(\sigma_L, S))\right) \\
&\le \mathop{\mathbb{E}}_{(\sigma_R,\sigma_L)} d_{TV}(\{\sigma_R \mid \sigma_L\}, \mathtt{ALG}(A(\sigma_L))) + \mathbb{P}[|\mathbf{x}| \ge 2\rho n] \\
&\le \mathop{\mathbb{E}}_{(\sigma_R,\sigma_L)} d_{TV}(\{\sigma_R \mid \sigma_L\}, \mathtt{ALG}(A(\sigma_L))) + e^{-\rho d^t/6},
\end{aligned}$$

where the final inequality follows from the Chernoff bound. This implies that we can take $t_0'(d, \delta') = t(d, \delta + e^{-\rho d^t/6})$.

By an application of Markov's inequality, it follows that 0.99 of subset $S \subset [n]$ have the property that

$$\mathop{\mathbb{E}}_{(\sigma_R,\sigma_L)\sim\mathcal{D}_{d,\varepsilon,t}} d_{TV}\left(\{\sigma_R \mid \sigma_L\}, \mathtt{ALG}(\mathtt{B}(\sigma_L, S))\right) \le \delta. \tag{19}$$

We claim that we can take $\mathcal{S}$ to consist of the subsets S which satisfy (19). First we check that these subsets have the desired property. Indeed, it suffices to note that for any μ with $\mu|_S = \mathcal{D}_{d,\varepsilon,t}|_S$, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\sigma_L\sim\mu|_L} d_{TV}(\mathtt{ALG}(\sigma_L, S), \{\sigma_R|\sigma_L\}) &= \mathop{\mathbb{E}}_{\sigma_L|_S} \mathop{\mathbb{E}}_{\sigma'_L|_{\overline{S}}} d_{TV}\left(\mathtt{ALG}(\sigma_L|_S, \sigma'_L|_{\overline{S}}, S), \mathop{\mathbb{E}}_{\sigma_L|_{\overline{S}}} \{\sigma_R|\sigma_L\}\right) \\
&\le \mathop{\mathbb{E}}_{\sigma_L,\sigma'_L|_{\overline{S}}} d_{TV}\left(\mathtt{ALG}(\sigma_L|_S, \sigma'_L|_{\overline{S}}, S), \{\sigma_R \mid \sigma_L\}\right)
\end{aligned}$$

where the first inequality follows because of the convexity of the total variation distance and the second inequality follows by (19).

Finally, we show that $\mathcal{S}$ is large in size. This follows because

$$|\mathcal{S}| \geq 0.99\binom{d^t}{\rho d^t} \geq \exp(\rho d^t) = 2^{\Omega_{\varepsilon,d}(d^t)}$$

where the second inequality is true for sufficiently large t. □

Proof of Claim 3.5. Let $f(x) = \frac{1}{1+x}$ and $g(x) = x^p$. We claim that

$$|f'(x)| \leq g'(x)/p = x^{p-1}.$$

The desired inequality then follows from the fundamental theorem of calculus. Now, $|f'(x)| = (1+x)^{-2}$ and so when $x \geq 1$, we have $|f'(x)| \leq x^{-2} \leq x^{p-1}$ and when $x \leq 1$ we have $|f'(x)| \leq 1 \leq x^{p-1}$. □

Proof of Claim 3.7. It suffices to note that for sufficiently small x, we have

$$\left(\frac{1-x}{1+x}\right)^{1/2} \leq 1 - x + \frac{3x^2}{5}.$$

Indeed, the above holds as for sufficiently small x, we have

$$(1+x)\left(1 - x + \frac{3x^2}{5}\right)^2 = 1 - x + \frac{x^2}{5} + O(x^3) \geq 1 - x,$$

as desired. □

Proof of Claim 3.8. It suffices to note that for

$$\frac{d}{dx}\left(\frac{\alpha - x}{\beta + x}\right)^{1/2} = \frac{-(\alpha+\beta)}{4(\alpha-x)^{1/2}(\beta+x)^{3/2}}.$$

By setting $\alpha = 1 - \varepsilon a$ and $\beta = 1 + \varepsilon a$, it follows that for sufficiently small ε, it follows that the desired quantity is bounded above by a constant. □

Proof of Lemma 3.12. In what follows, denote $\eta := \frac{1-\varepsilon}{2}$. We adapt the proof of [MNS16, Lemma 3.16]. For simplicity of notation denote $f(x) = \sqrt{\frac{1-\varepsilon x}{1+\varepsilon x}}$. By applying Markov's inequality to Lemma 3.10, it follows that for some $\tau = \tau(\varepsilon^*)$ to be chosen and sufficiently large $\varepsilon^2 d$ relative to τ, we have

$$\mathbb{P}[X_{ui,k} \geq 1 - \tau\eta^{1/4} \mid \sigma_u = +1] \geq 1 - \tau\eta^{3/4} - \eta.$$

Since $Y_i \geq X_{ui,k} - |\xi|$, it follows that

$$\mathbb{P}\left[Y_i \geq 1 - \tau\eta^{1/4} - |\xi| \,|\, \sigma_u = +1\right] \geq \mathbb{P}\left[X_{ui,k} - |\xi| \geq 1 - \tau\eta^{1/4} - |\xi| \,|\, \sigma_u = +1\right] \geq 1 - \tau\eta^{3/4} - \eta =: \alpha.$$

The remainder of the proof is near verbatim that of [MNS16, Lemma 3.16] and we repeat it here for completeness. Since $f(x)$ is decreasing in x, it follows that

$$\mathbb{E}[f(X) \mid \sigma_u = +1] \leq f(s)\mathbb{P}[X \geq s \mid \sigma_u = +1] + f(-1)\mathbb{P}[X \leq s \mid \sigma_u = +1]$$

for any random variable X supported on $[-1, 1]$ and $s \in [-1, 1]$. Applying this for $s = 1 - \tau\eta^{1/4} - \xi$ and

$X = Y_i$, the above probability estimate implies that

$$\mathbb{E}[f(Y_i) \mid \sigma_u = +1] \leq f(1 - \tau\eta^{1/4} - |\xi|)\alpha + f(-1)(1-\alpha).$$

We claim that we can make each of the terms above bounded away from $\frac{1}{2}$ by taking $\tau(\varepsilon^*) = \frac{(\varepsilon^*)^2}{8}$ and $\nu(\varepsilon, \varepsilon^*) = \frac{(\varepsilon^*)^2\eta^{1/4}}{8}$. Note that this choice of $\nu(\varepsilon, \varepsilon^*)$ satisfies $\nu = \tau\eta^{1/4}$.

We can compute

$$\begin{aligned}
f(1 - \tau\eta^{1/4} - |\xi|)(1 - \tau\eta^{3/4} - \eta) &\leq \sqrt{\frac{2\eta + \varepsilon\tau\eta^{1/4} + \varepsilon\xi}{2 - (2\eta + \varepsilon\tau\eta^{1/4} + \varepsilon|\xi|)}} \cdot (1 - \tau\eta^{3/4} - \eta) \\
&\leq \left(\sqrt{\frac{\eta}{1-\eta}} + \frac{1}{\sqrt{2\eta}}(\varepsilon\tau\eta^{1/4} + \varepsilon|\xi|)\right) \cdot (1 - \tau\eta^{3/4} - \eta) \\
&\leq \sqrt{\eta(1-\eta)} + \frac{1}{\sqrt{2\eta}}\varepsilon(\tau\eta^{1/4} + |\xi|)(1-\eta) \\
&\leq \sqrt{\eta(1-\eta)} + \sqrt{2}\tau\eta^{1/4}\sqrt{1-\eta}
\end{aligned}$$

where the second inequality follows by Taylor expanding $\sqrt{x}1 - x$ around $x = \eta$, and our choice of parameters ensure that $2(\eta + \varepsilon\nu) \leq 2(\eta + \nu) < 1$ so that the derivative of $\sqrt{x}1 - x$ is bounded. Finally, it can be checked that the above is bounded away from $\frac{1}{2}$ for our choice of parameters for our choice of τ and ν since $\eta \in \left(0, \frac{1-\varepsilon^*}{2}\right]$; in fact our choice of parameters ensures that $\sqrt{\eta(1-\eta)} + 2\tau\eta^{1/4}\sqrt{1-\eta} \leq \frac{1}{2} - c(\varepsilon^*)$ for some $c(\varepsilon^*) > 0$. This follows because for $0 < \varepsilon < 1$, we have that $\eta^{1/4}\sqrt{1-\eta} \leq \frac{1}{2^{3/4}} + \frac{\varepsilon}{2^{3/4}}$ and we also have $\sqrt{\eta(1-\eta)} \leq \frac{1}{2} - \frac{\varepsilon^2}{4}$ and so we can write

$$\sqrt{\eta(1-\eta)} + 2\tau\eta^{1/4}\sqrt{1-\eta} \leq \frac{1}{2} - (\varepsilon^*)^2\left(\frac{1}{4} - \frac{1}{4 \cdot 2^{3/4}}\right).$$

We can also compute

$$f(-1)(2\tau\eta^{3/4} + \eta) = 2\tau\eta^{1/4}\sqrt{1-\eta} + \sqrt{\eta(1-\eta)}.$$

Note that $\eta(1-\eta)$ is bounded away from $\frac{1}{2}$ on the interval $\eta \in \left(0, \frac{1-\varepsilon^*}{2}\right]$ by our earlier calculation.

It follows that there is some $d^*(\varepsilon)$ such that for all $d \geq d^*(\varepsilon^*)$, there is some $\lambda(\varepsilon^*) < 1$ such that

$$\mathbb{E}\left[\sqrt{f(Y_i)} \,\middle|\, \sigma_u = +1\right] \leq \lambda$$

as desired. □

References

[Abb18] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

[BB86] James Berger and L Mark Berliner. Robust bayes and empirical bayes analysis with ε-contaminated priors. *The Annals of Statistics*, pages 461–486, 1986.

[BKMP05] Noam Berger, Claire Kenyon, Elchanan Mossel, and Yuval Peres. Glauber dynamics on trees and hyperbolic graphs. *Probability Theory and Related Fields*, 131(3):311–340, 2005.

[BMR21] Jess Banks, Sidhanth Mohanty, and Prasad Raghavendra. Local statistics, semidefinite programming, and community detection. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1298–1316. SIAM, 2021.

[CKMY22] Sitan Chen, Frederic Koehler, Ankur Moitra, and Morris Yau. Kalman filtering with adversarial corruptions. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 832–845, 2022.

[DdNS22] Jingqiu Ding, Tommaso d'Orsi, Rajai Nasser, and David Steurer. Robust recovery for stochastic block models. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 387–394. IEEE, 2022.

[DGT19] Ilias Diakonikolas, Themis Gouleakis, and Christos Tzamos. Distribution-independent pac learning of halfspaces with massart noise. *Advances in Neural Information Processing Systems*, 32, 2019.

[DK23] Ilias Diakonikolas and Daniel M Kane. *Algorithmic High-dimensional Robust Statistics*. Cambridge University Press, 2023.

[DKTZ20] Ilias Diakonikolas, Vasilis Kontonis, Christos Tzamos, and Nikos Zarifis. Learning halfspaces with massart noise under structured distributions. In *Conference on Learning Theory*, pages 1486–1513. PMLR, 2020.

[DMR06] Constantinos Daskalakis, Elchanan Mossel, and Sébastien Roch. Optimal phylogenetic reconstruction. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 159–168, 2006.

[EKPS00] William Evans, Claire Kenyon, Yuval Peres, and Leonard J Schulman. Broadcasting on trees and the ising model. *Annals of Applied Probability*, pages 410–433, 2000.

[Goo50] Isidore Jacob Good. *Probability and the Weighing of Evidence*. Charles Griffin & Co., Ltd., London; Hafner Publishing Co., New York, 1950.

[IS23] Misha Ivkov and Tselil Schramm. Semidefinite programs simulate approximate message passing robustly. *arXiv preprint arXiv:2311.09017*, 2023.

[JM04] Svante Janson and Elchanan Mossel. Robust reconstruction on trees is determined by the second eigenvalue. *Ann. Probab.*, 32(3B):2630–2649, 2004.

[Kea90] Michael J Kearns. *The computational complexity of machine learning*. MIT press, 1990.

[LM22] Allen Liu and Ankur Moitra. Minimax rates for robust community detection. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 823–831. IEEE, 2022.

[MN06] Pascal Massart and Élodie Nédélec. Risk bounds for statistical learning. *Ann. Statist.*, 34(5):2326–2366, 2006.

[MNS15] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162:431–461, 2015.

[MNS16] Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. *Ann. Appl. Probab.*, 26(4):2211–2256, 2016.

[Mos04] Elchanan Mossel. Survey: Information flow on trees. *arXiv preprint math/0406446*, 2004.

[MPW16] Ankur Moitra, William Perry, and Alexander S. Wein. How robust are reconstruction thresholds for community detection? In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 828–841. ACM, New York, 2016.

[MSW04] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Glauber dynamics on trees: boundary conditions and mixing time. *Communications in Mathematical Physics*, 250:301–334, 2004.

[Pol23] Yury Polyanskiy. Personal communication, December 2023. Personal communication with Yury Polyanskiy.

[YP22] Qian Yu and Yury Polyanskiy. Ising model on locally tree-like graphs: Uniqueness of solutions to cavity equations. *arXiv preprint arXiv:2211.15242*, 2022.