Broadcasting on Random Networks

Anuran Makur[*], Elchanan Mossel[†], and Yury Polyanskiy[*]
[*]EECS and [†]Mathematics Departments, Massachusetts Institute of Technology
Email: {a_makur, elmos, yp}@mit.edu

***Abstract*—We study a generalization of the problem of broadcasting on trees to the setting of directed acyclic graphs (DAGs). At time 0, a source vertex X transmits a uniform bit along binary symmetric channels (BSCs) to a set of vertices called layer 1. Each vertex except X has indegree d. At time $k \geq 1$, vertices at layer k apply d-input Boolean processing functions to their received bits and send out the results to vertices at layer $k+1$. We say that broadcasting is possible if we can reconstruct X with probability of error bounded away from $\frac{1}{2}$ using the values of all vertices at an arbitrarily deep layer k. This question is closely related to models of reliable computation and storage, probabilistic cellular automata, and information flow in biological networks.**

In this work, we analyze randomly constructed DAGs and demonstrate that broadcasting is only possible if the BSC noise level is below a certain (degree and function dependent) critical threshold. Specifically, for every $d \geq 3$, we identify the critical threshold for random DAGs with layers of size $\Omega(\log(k))$ and majority processing functions. For $d = 2$, we establish a similar result for the NAND processing function. Furthermore, for odd $d \geq 3$, we prove that the identified thresholds cannot be improved by other processing functions if reconstruction is required from a single vertex. Finally, for any BSC noise level, in quasi-polynomial or randomized polylogarithmic time in the depth, we construct deterministic bounded degree DAGs with layers of size $\Theta(\log(k))$ that admit reconstruction using lossless expander graphs.

I. Introduction

We study a generalization of the problem of *broadcasting* on trees to the setting of directed acyclic graphs (DAGs). In the broadcasting on trees problem, we are given a noisy tree T whose vertices are Bernoulli random variables and edges are independent binary symmetric channels (BSCs) with common crossover probability $\delta \in \left(0, \frac{1}{2}\right)$. Given that the root is an unbiased random bit, the goal is to decode the root from the bits at the kth layer of the tree as $k \to \infty$. It is well-known that $(1-2\delta)^2 \operatorname{br}(T) > 1$ if and only if the minimum probability of error in decoding is bounded away from $\frac{1}{2}$ for all k, where $\operatorname{br}(T)$ denotes the branching number of the tree [1]–[3]. A consequence of this result is that reconstruction is impossible for trees with sub-exponentially many vertices at each layer.

In our problem of broadcasting on *bounded indegree DAGs*, where all vertices are Bernoulli random variables and all edges are BSCs as before, there are two principal differences: (a) unlike trees, layer sizes are sub-exponential in the depth for DAGs; (b) a DAG vertex has several incoming signals, and its value is obtained by applying a *Boolean processing function*. The latter aspect enables the possibility of information fusion at the vertices, and our main goal is to understand whether the benefits of (b) overpower the shortcoming of (a) and permit reconstruction of the root bit with sub-exponential layer size.

This work has two main contributions. Firstly, via a probabilistic argument using random DAGs (defined in subsection I-B), we demonstrate the existence of bounded degree DAGs with layer size $\Omega(\log(k))$ in the depth k which permit recovery of the root bit for sufficiently low δ's in section II. Secondly, we provide explicit deterministic constructions of such DAGs using regular bipartite lossless expander graphs in section III. In particular, the constituent expander graphs for the first k layers can be constructed in either deterministic quasi-polynomial time or randomized polylogarithmic time in k. Together, these results imply that in terms of economy of storing information, *DAGs are doubly-exponentially more efficient than trees*.

A. Motivation

Broadcasting on DAGs has several natural interpretations. Perhaps most pertinently, it captures the feasibility of reliably communicating through Bayesian networks in the field of *communication networks*. Indeed, suppose a sender communicates a sequence of bits to a receiver through a large network. If broadcasting is impossible on this network, then the "wavefront of information" for each bit decays irrecoverably through the network, and the receiver cannot reconstruct the sender's message regardless of the coding scheme employed.

Broadcasting on DAGs is also a close variant of the model of *reliable computation using noisy circuits*, cf. [4], [5]. Suppose we want to remember a bit using a noisy circuit. The "von Neumann approach" is to take multiple perfect clones of the bit and recursively apply noisy gates in order to reduce the overall noise [6], [7]. In contrast, the broadcasting perspective is to start from a single bit and repeatedly create noisy clones and apply perfect gates to these clones so that one can recover the bit reasonably well. Thus, the broadcasting model can be construed as a noisy circuit with perfect gates at the vertices and edges or wires that independently make errors.

Furthermore, the broadcasting model plays a role in various discrete probability questions. For example, the special case of trees corresponds to *ferromagnetic Ising models in statistical physics*. Specifically, reconstruction is impossible on a tree if and only if the free boundary Gibbs state of the corresponding

[*]This work was supported in part by the Center for Science of Information, a National Science Foundation (NSF) Science and Technology Center, under Grant CCF-09-39370, in part by the NSF CAREER Award under Grant CCF-12-53205, and in part by Schneider Electric, Lenovo Group (China) Limited and the Hong Kong Innovation and Technology Fund (ITS/066/17FP) under the HKUST-MIT Research Alliance Consortium.

[†]This work was supported in part by the NSF Grants CCF-1665252 and DMS-1737944, and in part by the DOD ONR Grant N00014-17-1-2598.

Ising model is extremal [2], [3, Section 2.2]. On a different front, broadcasting on 2D regular grids can be perceived as 1D *probabilistic cellular automata* (see e.g. [8, Section 1] for a definition) with boundary conditions that limit the layer sizes, and the feasibility of broadcasting on 2D grids is closely related to questions of ergodicity of probabilistic cellular automata.

Finally, special cases of the model have also found applications in theoretical computer science. Indeed, broadcasting on trees plays a crucial role in understanding *ancestral data and phylogenetic tree reconstruction*, cf. [9], [10], and phase transitions for *random constraint satisfaction* problems, see e.g. [11], [12]. Moreover, the existence results obtained here on DAGs suggest it might be possible to reconstruct other biological networks, such as phylogenetic networks or pedigrees, even if the growth of the network is very mild. It is also interesting to explore if there are connections between broadcasting on general DAGs and random constraint satisfaction.

B. Random DAG Model

A *random DAG model* consists of an infinite DAG with fixed vertices that are Bernoulli random variables and randomly generated edges that are independent BSCs. Let the root or source vertex be $X_{0,0} \sim \mathsf{Ber}\big(\frac{1}{2}\big)$, and let $X_k = (X_{k,0}, \dots, X_{k,L_k-1})$ be the collection of vertices at distance $k \in \mathbb{N} \triangleq \{0,1,2,\dots\}$ from the root, where $L_k \in \mathbb{N}\backslash\{0\}$ is the number of vertices at distance k. In particular, we have $X_0 = (X_{0,0})$ and $L_0 = 1$.

For any $k \in \mathbb{N}\backslash\{0\}$ and any $j \in [L_k] \triangleq \{0,\dots,L_k-1\}$, we first independently and uniformly select $d \in \mathbb{N}\backslash\{0\}$ vertices $X_{k-1,i_1},\dots,X_{k-1,i_d}$ from X_{k-1}, and then construct d directed edges $(X_{k-1,i_1},X_{k,j}),\dots,(X_{k-1,i_d},X_{k,j})$. This process generates the underlying DAG (or directed multigraph) structure, and we let G denote this random DAG variable.

To define a *Bayesian network* on G, we fix a sequence of Boolean functions at the vertices and a crossover probability $\delta \in \big(0,\frac{1}{2}\big)$. Then, for every $k \in \mathbb{N}\backslash\{0\}$ and $j \in [L_k]$, given $i_1,\dots,i_d$ and $X_{k-1,i_1},\dots,X_{k-1,i_d}$, we define:

$$X_{k,j} = f_k(X_{k-1,i_1} \oplus Z_{k,j,1},\dots,X_{k-1,i_d} \oplus Z_{k,j,d}) \qquad (1)$$

where $f_k : \{0,1\}^d \to \{0,1\}$ is the processing function at level k, $\oplus$ denotes XOR, and $Z_{k,j,i}$ are i.i.d $\mathsf{Ber}(\delta)$ random variables that are independent of everything else. The propagation process of $X_{0,0}$ through G is completely characterized by (1).

Under this setup, our objective is to determine whether $X_{0,0}$ can be decoded from the observations X_k as $k \to \infty$. Let us define the proportion of 1's at level $k \in \mathbb{N}$ as:

$$\sigma_k \triangleq \frac{1}{L_k}\sum_{m=0}^{L_k-1} X_{k,m} \qquad (2)$$

where $\sigma_0 = X_{0,0}$. Given $\sigma_{k-1} = \sigma$, $X_{k-1,i_1},\dots,X_{k-1,i_d}$ are i.i.d. $\mathsf{Ber}(\sigma)$, and as a result, each $X_{k,j}$ is the output of f_k upon inputting a d-length i.i.d. $\mathsf{Ber}(\sigma * \delta)$ string, where $\sigma * \delta \triangleq \sigma(1-\delta)+\delta(1-\sigma)$. It is straightforward to verify that σ_k is a sufficient statistic of X_k for performing inference about σ_0. Therefore, we consider the Markov chain $\{\sigma_k : k \in \mathbb{N}\}$ in our achievability results. In particular, given σ_k, inferring the value of σ_0 is a binary hypothesis testing problem with minimum probability of error $\mathbb{P}\big(f^k_{\mathsf{ML}}(\sigma_k) \neq \sigma_0\big)$, where $f^k_{\mathsf{ML}}(\sigma_k) \in \{0,1\}$ is the maximum likelihood (ML) decision rule without knowledge of the random DAG realization G. We say that reconstruction of the root bit $X_{0,0}$ is possible when:

$$\lim_{k\to\infty} \mathbb{P}\big(f^k_{\mathsf{ML}}(\sigma_k) \neq \sigma_0\big) < \frac{1}{2}\,. \qquad (3)$$

On the other hand, we will consider the Markov chain $\{X_k : k \in \mathbb{N}\}$ conditioned on G in our converse results. We say that reconstruction of the root bit $X_{0,0}$ is impossible when:

$$\lim_{k\to\infty} \left\| P^+_{X_k|G} - P^-_{X_k|G} \right\|_{\mathsf{TV}} = 0 \quad G\text{-}a.s. \qquad (4)$$

where $\|\cdot\|_{\mathsf{TV}}$ is the *total variation (TV) distance*, $P^+_{X_k|G}$ and $P^-_{X_k|G}$ are conditional distributions of X_k given $\{X_0 = 1, G\}$ and $\{X_0 = 0, G\}$, respectively, and G-$a.s.$ stands for "almost surely with respect to the law of G." Moreover, the condition (4) is equivalent to $\lim_{k\to\infty} \mathbb{P}\big(h^k_{\mathsf{ML}}(X_k,G) \neq X_0 \big| G\big) = \frac{1}{2}$ G-$a.s.$, because for every realization of the random DAG G:

$$\mathbb{P}\big(h^k_{\mathsf{ML}}(X_k,G) \neq X_0 \big| G\big) = \frac{1}{2}\Big(1 - \left\| P^+_{X_k|G} - P^-_{X_k|G} \right\|_{\mathsf{TV}}\Big) \qquad (5)$$

where $h^k_{\mathsf{ML}}(X_k,G) \in \{0,1\}$ is the ML decision rule with knowledge of the random DAG realization G.

II. Reconstruction on Random DAGs

In this section, we present two main results for random DAG models with general indegree $d \geq 3$ and $d = 2$, respectively. Note that $d = 1$ corresponds to the well-studied tree setting.

A. Phase Transition for Indegree $d \geq 3$

When $d \geq 3$, taking a majority vote of the d inputs at each vertex intuitively has good "local error correction" properties. So, we fix all Boolean functions in the random DAG model to be the *majority* rule, and prove that this model exhibits a phase transition phenomenon around a critical threshold of:

$$\delta_{\mathsf{maj}} \triangleq \frac{1}{2} - \frac{2^{d-2}}{\left\lceil \frac{d}{2} \right\rceil \binom{d}{\lceil \frac{d}{2} \rceil}} \qquad (6)$$

where $\lceil\cdot\rceil$ denotes the ceiling function.

Theorem 1 ($d \geq 3$ and Majority Processing)**.** *Let $C(\delta,d)$ and $D(\delta,d)$ be the constants defined in (12) and (10). Consider a random DAG model with $d \geq 3$ and all majority processing functions, where ties are broken by outputting random bits.*

1) *If $\delta \in (0,\delta_{\mathsf{maj}})$ and the layer size $L_k \geq C(\delta,d)\log(k)$ for all sufficiently large k (depending on δ and d),[1] then reconstruction is possible in the sense that:*

$$\limsup_{k\to\infty} \mathbb{P}\Big(\hat{S}_k \neq \sigma_0\Big) < \frac{1}{2}$$

where $\hat{S}_k \triangleq \mathbb{1}\big\{\sigma_k \geq \frac{1}{2}\big\}$ is the majority decoder. Hence, reconstruction is also possible in the sense of (3).

2) *If $\delta \in \big(\delta_{\mathsf{maj}},\frac{1}{2}\big)$ and the layer size $L_k = o\big(D(\delta,d)^{-k}\big)$, then reconstruction is impossible in the sense of (4).*

[1] Note that $\log(\cdot)$ and $\exp(\cdot)$ have base e in this paper.

Proof Outline. We outline the proof here, and refer readers to the full version of this paper [13, Section III] for details.

Suppose we are given that $\sigma_{k-1} = \sigma$ for any $k \in \mathbb{N}\backslash\{0\}$. Then, for all $j \in [L_k]$, $X_{k,j}$ are i.i.d. $\mathsf{Ber}(g(\sigma))$, where e.g. when d is odd, the function $g : [0,1] \rightarrow [0,1]$ is defined as:

$$g(\sigma) \triangleq \mathbb{P}(X_{k,j} = 1 | \sigma_{k-1} = \sigma) = \mathbb{E}[\sigma_k | \sigma_{k-1} = \sigma] \quad (7)$$

$$= \sum_{i=\frac{d+1}{2}}^{d} \binom{d}{i} (\sigma * \delta)^i (1 - \sigma * \delta)^{d-i}. \quad (8)$$

Moreover, the *Margulis-Russo formula* yields [14, Section 2]:

$$g'(\sigma) = (1-2\delta)\, \frac{d+1}{2} \binom{d}{\frac{d+1}{2}} \left((\sigma * \delta)(1 - \sigma * \delta)\right)^{\frac{d-1}{2}} \quad (9)$$

which means g' is positive on $[0,1]$, increasing on $\left[0, \frac{1}{2}\right]$, and decreasing on $\left[\frac{1}{2}, 1\right]$. Hence, g is increasing on $[0,1]$, convex on $\left[0, \frac{1}{2}\right]$, concave on $\left[\frac{1}{2}, 1\right]$, and has Lipschitz constant:

$$D(\delta, d) \triangleq g'\left(\frac{1}{2}\right) = (1-2\delta) \left(\frac{1}{2}\right)^{d-1} \frac{d+1}{2} \binom{d}{\frac{d+1}{2}}. \quad (10)$$

These properties also hold for even d. Let δ_{maj} be the critical value in (6) such that $D(\delta_{\mathsf{maj}}, d) = 1$. Then, there are two regimes of δ of interest.

1) Achievability: Suppose $\delta \in (0, \delta_{\mathsf{maj}})$ so that $D(\delta, d) > 1$. In this case, g has three fixed points at $\sigma = 1 - \hat{\sigma}, \frac{1}{2}, \hat{\sigma}$, where $\hat{\sigma} \in \left(\frac{1}{2}, 1\right)$, as $g\left(\frac{1}{2}\right) = \frac{1}{2}$ and $g(1-\sigma) = 1 - g(\sigma)$ using (8).

We first construct a useful *monotone Markovian coupling* $\{(X_k^-, X_k^+) : k \in \mathbb{N}\}$ between the Markov chains $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$, which are versions of the Markov chain $\{X_k : k \in \mathbb{N}\}$ initialized at $X_0^+ = 1$ and $X_0^- = 0$, respectively. For every realization of G, we couple these chains so that along any edge BSC, e.g. $(X_{k,j}, X_{k+1,i})$, either $X_{k,j}^+$ and $X_{k,j}^-$ are both copied with probability $1-2\delta$, or $X_{k+1,i}^+ = X_{k+1,i}^- = \mathsf{Ber}\left(\frac{1}{2}\right)$ *a.s.* for a shared independent $\mathsf{Ber}\left(\frac{1}{2}\right)$ bit with probability 2δ. In other words, $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$ "run" on the same underlying DAG G with common BSCs. Since the majority rule is monotone non-decreasing, this coupling is also monotone, i.e. $X_{k,j}^+ \geq X_{k,j}^-$ *a.s.* for every $k \in \mathbb{N}$ and $j \in [L_k]$.

Next, observe that $\gamma(\epsilon) \triangleq g(\hat{\sigma} - \epsilon) - (\hat{\sigma} - \epsilon) > 0$ for all sufficiently small $\epsilon > 0$, because $g'(\hat{\sigma}) < 1$ and $g(\hat{\sigma}) = \hat{\sigma}$. Fix any such $\epsilon = \epsilon(\delta, d) > 0$. Then, Hoeffding's inequality yields:

$$\mathbb{P}\left(\sigma_k^+ < g\left(\sigma_{k-1}^+\right) - \gamma(\epsilon) \,\middle|\, \sigma_{k-1}^+\right) \leq \exp\left(-2L_k \gamma(\epsilon)^2\right)$$

for every $k \in \mathbb{N}\backslash\{0\}$, where σ_k^+ and σ_k^- are defined using X_k^+ and X_k^-, respectively, according to (2). Hence, we have:

$$\mathbb{P}\left(\sigma_k^+ < \hat{\sigma} - \epsilon \,\middle|\, \sigma_{k-1}^+ \geq \hat{\sigma} - \epsilon\right) \leq \exp\left(-2L_k \gamma(\epsilon)^2\right) \quad (11)$$

because $\sigma_k^+ < \hat{\sigma} - \epsilon = g(\hat{\sigma} - \epsilon) - \gamma(\epsilon)$ implies that $\sigma_k^+ < g(\sigma_{k-1}^+) - \gamma(\epsilon)$ when $\sigma_{k-1}^+ \geq \hat{\sigma} - \epsilon$ (since g is increasing).

Fix any $\tau > 0$ and any sufficiently large $K = K(\epsilon, \tau) \in \mathbb{N}$ such that $\sum_{m>K} \exp\left(-2L_m \gamma(\epsilon)^2\right) \leq \tau$. Note that such K exists because $L_m \geq C(\delta, d) \log(m)$ for all sufficiently large m (depending on δ and d), where we define:

$$C(\delta, d) \triangleq \frac{1}{\gamma(\epsilon(\delta, d))^2} > 0\,. \quad (12)$$

Now let $E \triangleq \left\{\sigma_K^+ \geq \hat{\sigma} - \epsilon, \sigma_K^- \leq 1 - \hat{\sigma} + \epsilon\right\}$, and observe using the Hoeffding based bound in (11) that:

$$\mathbb{P}\left(\bigcap_{k>K} \left\{\sigma_k^+ \geq \hat{\sigma} - \epsilon\right\} \,\middle|\, E\right) \geq \prod_{k>K} 1 - \exp\left(-2L_k \gamma(\epsilon)^2\right)$$
$$\geq 1 - \sum_{k>K} \exp\left(-2L_k \gamma(\epsilon)^2\right)$$

where (11) can be shown to hold with the additional conditioning required. Therefore, we have for any $k > K$:

$$\mathbb{P}\left(\sigma_k^+ \geq \hat{\sigma} - \epsilon \,\middle|\, E\right), \mathbb{P}\left(\sigma_k^- \leq 1 - \hat{\sigma} + \epsilon \,\middle|\, E\right) \geq 1 - \tau \quad (13)$$

where the $\mathbb{P}\left(\sigma_k^- \leq 1 - \hat{\sigma} + \epsilon \middle| E\right)$ case holds mutatis mutandis.

Finally, notice that for all $k > K$:

$$\mathbb{P}\left(\sigma_k^+ \geq \frac{1}{2}\right) - \mathbb{P}\left(\sigma_k^- \geq \frac{1}{2}\right)$$
$$\geq \mathbb{E}\left[\mathbb{1}\left\{\sigma_k^+ \geq \frac{1}{2}\right\} - \mathbb{1}\left\{\sigma_k^- \geq \frac{1}{2}\right\} \,\middle|\, E\right] \mathbb{P}(E)$$
$$\geq \left(\mathbb{P}\left(\sigma_k^+ \geq \hat{\sigma} - \epsilon \,\middle|\, E\right) - \mathbb{P}\left(\sigma_k^- > 1 - \hat{\sigma} + \epsilon \,\middle|\, E\right)\right) \mathbb{P}(E)$$
$$\geq (1 - 2\tau)\mathbb{P}(E) > 0$$

where the first inequality uses the monotonicity of our Markovian coupling, $\mathbb{1}\{\sigma_k^+ \geq \frac{1}{2}\} \geq \mathbb{1}\{\sigma_k^- \geq \frac{1}{2}\}$ *a.s.*, the second inequality holds because $1 - \hat{\sigma} + \epsilon < \frac{1}{2} < \hat{\sigma} - \epsilon$, and the final inequality follows from (13). It is straightforward to verify that this implies that $\limsup_{k\rightarrow\infty} \mathbb{P}(\hat{S}_k \neq \sigma_0) < \frac{1}{2}$.

2) Converse: Suppose $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2}\right)$ so that $D(\delta, d) < 1$. In this case, the only fixed point of g is $\sigma = \frac{1}{2}$.

First, using our monotone coupling and the maximal coupling representation of TV distance, it can be shown that:

$$\mathbb{E}\left[\left\|P_{X_k|G}^+ - P_{X_k|G}^-\right\|_{\mathsf{TV}}\right] \leq \mathbb{P}\left(X_k^+ \neq X_k^-\right)$$
$$\leq L_k \, \mathbb{E}\left[\sigma_k^+ - \sigma_k^-\right] \quad (14)$$

where the second inequality follows from the union bound, the relation $\mathbb{P}(X_{k,j}^+ \neq X_{k,j}^-) = \mathbb{E}[X_{k,j}^+ - X_{k,j}^-]$, and (2). Then, we can bound $\mathbb{E}\left[\sigma_k^+ - \sigma_k^-\right]$ using the Lipschitz continuity of g and the monotonicity of our coupling. Indeed, observe that:

$$\mathbb{E}\left[\sigma_k^+ - \sigma_k^-\right] = \mathbb{E}\left[g\left(\sigma_{k-1}^+\right) - g\left(\sigma_{k-1}^-\right)\right]$$
$$\leq D(\delta, d)\, \mathbb{E}\left[\sigma_{k-1}^+ - \sigma_{k-1}^-\right]$$

where the equality follows from the tower property and (7). Therefore, we recursively have:

$$\mathbb{E}\left[\left\|P_{X_k|G}^+ - P_{X_k|G}^-\right\|_{\mathsf{TV}}\right] \leq L_k D(\delta, d)^k \quad (15)$$

where we use (14) and $\mathbb{E}\left[\sigma_0^+ - \sigma_0^-\right] = 1$. Letting $k \rightarrow \infty$ produces $\lim_{k\rightarrow\infty} \mathbb{E}[\|P_{X_k|G}^+ - P_{X_k|G}^-\|_{\mathsf{TV}}] = 0$ because $L_k = o(D(\delta, d)^{-k})$ (with $D(\delta, d) < 1$). Finally, a monotonicity and bounded convergence theorem argument yields (4). ■

Part 1 of Theorem 1 illustrates that reconstruction is possible on random DAGs with majority rule processing using the majority decoder $\hat{S}_k$ when $\delta \in (0, \delta_{\mathsf{maj}})$, while part 2 establishes that even if the ML decoder knows G and has access to X_k, it cannot beat the δ_{maj} critical threshold in all but a zero measure set of DAGs. We remark that the δ_{maj} critical threshold in (6) has appeared in past literature. For example, reliable computation using formulae with d-input δ-noisy gates, where $d \geq 3$ is odd, is impossible if and only if $\delta \geq \delta_{\mathsf{maj}}$, cf. [6], [7]. In fact, the analysis of the fixed point structure of g when $d = 3$ and $\delta_{\mathsf{maj}} = \frac{1}{6}$ can be traced back to von Neumann's seminal work [4]. Furthermore, the recursive structure of g was also analyzed in [14] in the context of recursive reconstruction on periodic trees. However, our analysis also requires significant applications of concentration of measure and coupling arguments not used in these works.

Part 2 of Theorem 1 is only a partial converse result. We conjecture that: In the random DAG model with $L_k = O(\log(k))$ and fixed $d \geq 3$, reconstruction is impossible for all choices of Boolean processing functions (which may vary between vertices and be graph dependent) when $\delta \geq \delta_{\mathsf{maj}}$. In fact, it is known that this conjecture is true when $\delta > \frac{1}{2} - \frac{1}{2\sqrt{d}}$, cf. [5], [13, Section II-C]. The ensuing proposition establishes another special case of our conjecture. It portrays that the ML decoder based on a single vertex, e.g. $X_{k,0}$, cannot reconstruct $X_{0,0}$ in all but a vanishing fraction of DAGs when $\delta \geq \delta_{\mathsf{maj}}$, although reconstruction is possible using $X_{k,0}$ when $\delta < \delta_{\mathsf{maj}}$.

Proposition 1 (Single Vertex Reconstruction)**.** *As in Theorem 1, consider a random DAG model with $d \geq 3$.*

1) *If $\delta \in (0, \delta_{\mathsf{maj}})$, $L_k \geq C(\delta, d)\log(k)$ for all sufficiently large k, and all processing functions are the majority rule, then reconstruction is possible in the sense that:*

$$\limsup_{k\to\infty} \mathbb{P}(X_{k,0} \neq X_{0,0}) < \frac{1}{2}$$

where $X_{k,0}$ is the single vertex decoder.

2) *If $\delta \in \left[\delta_{\mathsf{maj}}, \frac{1}{2}\right)$, d is odd, $\lim_{k\to\infty} L_k = \infty$, and $R_k \triangleq \inf_{n\geq k} L_n = O\left(d^{2k}\right)$, then for all choices of processing functions, reconstruction is impossible in the sense that:*

$$\lim_{k\to\infty} \mathbb{E}\left[\left\|P^{+}_{X_{k,0}|G} - P^{-}_{X_{k,0}|G}\right\|_{\mathsf{TV}}\right] = 0\,.$$

This is proved in [13, Appendix A], and part 2 exploits the aforementioned impossibility results on reliable computation.

We next present an immediate corollary of Theorem 1 which demonstrates that *deterministic DAGs* (i.e. Bayesian networks on specific realizations of G) admitting reconstruction of the root bit with logarithmic layer sizes in the depth do exist.

Corollary 1 (Existence of Deterministic DAGs)**.** *For every $d \geq 3$, $\delta \in (0, \delta_{\mathsf{maj}})$, and layer sizes $L_k \geq C(\delta, d)\log(k)$ for all sufficiently large k, there exists a deterministic DAG $\mathcal{G}$ with all majority processing functions such that:*

$$\lim_{k\to\infty} \mathbb{P}\left(h^k_{\mathsf{ML}}(X_k, \mathcal{G}) \neq X_0\right) < \frac{1}{2}\,.$$

This follows from a probabilistic method argument; see [13, Appendix B]. Since $\delta_{\mathsf{maj}} \to \frac{1}{2}$ as $d \to \infty$, a consequence of Corollary 1 is that for any $\delta \in \left(0, \frac{1}{2}\right)$, there exists a deterministic DAG with sufficiently large d and $L_k = \Omega(\log(k))$ which admits reconstruction of the root bit.

B. Phase Transition for Indegree $d = 2$

Our second main result considers the $d = 2$ setting, where it is not immediately obvious that deterministic DAGs admitting reconstruction exist. Indeed, it is not clear which processing functions are good for "local error correction" in this scenario. We fix all Boolean functions in the random DAG model to be the NAND rule. It is straightforward to verify that for the purposes of broadcasting, this is equivalent to a random DAG model with all AND functions at even levels and all OR functions at odd levels. For simplicity, we analyze this model at even levels, and establish a phase transition phenomenon around a critical threshold of $\delta_{\mathsf{nand}} \triangleq \frac{3-\sqrt{7}}{4}$.

Theorem 2 ($d = 2$ and NAND Processing)**.** *Consider a random DAG model with $d = 2$, all AND processing functions at even levels, and all OR processing functions at odd levels.*

1) *Suppose $\delta \in (0, \delta_{\mathsf{nand}})$. Then, there exist $C(\delta) > 0$ and $t = t(\delta) \in (0, 1)$ such that if $L_k \geq C(\delta)\log(k)$ for all sufficiently large k (depending on δ), then reconstruction is possible in the sense that:*

$$\limsup_{k\to\infty} \mathbb{P}\left(\hat{T}_{2k} \neq \sigma_0\right) < \frac{1}{2}$$

where $\hat{T}_k \triangleq \mathbb{1}\{\sigma_k \geq t\}$ is a thresholding decoder. Hence, reconstruction is also possible in the sense of (3).

2) *Suppose $\delta \in \left(\delta_{\mathsf{nand}}, \frac{1}{2}\right)$. Then, there exists $D(\delta) \in (0, 1)$ such that if $L_k = o\left(E(\delta)^{-k/2}\right)$ and $\liminf_{k\to\infty} L_k > \frac{2}{E(\delta)-D(\delta)}$ for any $E(\delta) \in (D(\delta), 1)$, then reconstruction is impossible in the sense of (4).*

Theorem 2 is an analogue of Theorem 1 for $d = 2$, and is proved in [13, Section IV] using the same proof technique. Moreover, analogues of part 1 of Proposition 1 and Corollary 1 also hold for Theorem 2. As before, the δ_{nand} threshold has appeared in the reliable computation literature. In particular, it is well-known that reliable computation using formulae consisting of δ-noisy NAND gates is possible when $\delta < \delta_{\mathsf{nand}}$ [15], and reliable computation using formulae with general 2-input δ-noisy gates is impossible when $\delta \geq \delta_{\mathsf{nand}}$ [16].

C. Optimality of Logarithmic Layer Size Growth

The next result shows that if L_k grows sub-logarithmically with the depth, then reconstruction is impossible for deterministic and random DAGs regardless of the decision rule used.

Proposition 2 (Layer Size Impossibility Result)**.** *For any deterministic DAG with parameters $\delta \in \left(0, \frac{1}{2}\right)$ and $d \in \mathbb{N}\backslash\{0\}$, if $L_k \leq \log(k)/(d\log(1/(2\delta)))$ for all sufficiently large k, then for all choices of Boolean processing functions, reconstruction is impossible in the sense that $\lim_{k\to\infty} \left\|P^{+}_{X_k} - P^{-}_{X_k}\right\|_{\mathsf{TV}} = 0$, where $P^{+}_{X_k}$ and $P^{-}_{X_k}$ denote the conditional distributions of X_k given $X_0 = 1$ and $X_0 = 0$, respectively.*

Proposition 2 is proved in [13, Appendix C]. Moreover, under the conditions of Proposition 2, reconstruction is also impossible for random DAG models in the sense of (4). Thus, our assumption that $L_k \geq C\log(k)$ for reconstruction to be possible in our previous results is in fact necessary.

III. Explicit Construction of DAGs where Broadcasting is Possible

Finally, we present an explicit construction of deterministic bounded degree DAGs such that $L_k = \Theta(\log(k))$ and reconstruction is possible using the majority decision rule. Our construction is based on a variant of regular bipartite lossless expander graphs. Using results like [17, Lemma 1] and [18, Proposition 1, Appendix II] which establish the existence of expander graphs via the probabilistic method, we show in [13, Corollary 2] that for any $d \in \mathbb{N}\backslash\{0\}$ and every sufficiently large $n \in \mathbb{N}\backslash\{0\}$ (depending on d), there exists a d-regular bipartite graph $B_n = (U_n, V_n, E_n)$ with two disjoint sets of degree d vertices U_n and V_n such that $|U_n| = |V_n| = n$, undirected edge set E_n (where multiple edges are allowed between two vertices), and the *lossless expansion property*:

$$\forall S \subseteq U_n,\ \ |S| = \frac{n}{d^{6/5}} \Rightarrow |\Gamma(S)| \geq \left(1 - \frac{2}{d^{1/5}}\right) d|S| \quad (16)$$

where $\Gamma(S) \triangleq \{v \in V_n : \exists u \in S, \{u,v\} \in E_n\}$ denotes the *neighborhood* of S. Note that we only require subsets of U_n to expand (not V_n). Moreover, strictly speaking, $nd^{-6/5}$ must be an integer, but we neglect this detail for simplicity. In the sequel, we refer to graphs B_n that satisfy (16) as *d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graphs*. The next theorem constructs deterministic DAGs with majority processing where reconstruction is possible by concatenating d-regular bipartite lossless $(d^{-6/5}, d-2d^{4/5})$-expander graphs.

Theorem 3 (Expander Based DAG Construction)**.** *Fix any $\delta \in \left(0, \frac{1}{2}\right)$, any sufficiently large odd $d = d(\delta) \geq 5$ satisfying:*

$$\frac{8}{d^{1/5}} + d^{6/5} \exp\left(-\frac{(1-2\delta)^2(d-4)^2}{8d}\right) \leq \frac{1}{2}, \quad (17)$$

and any sufficiently large $N = N(\delta) \in \mathbb{N}$ such that $M \triangleq \exp(N/(4d^{12/5})) \geq 2$ and for every $n \geq N$, there exists a d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graph $B_n = (U_n, V_n, E_n)$. Define $L_0 = 1$, $L_1 = N$, and $L_k = 2^m N$ for all $m, k \in \mathbb{N}$ such that $M^{\lfloor 2^{m-1} \rfloor} < k \leq M^{2^m}$, where $\lfloor \cdot \rfloor$ is the floor function, and $L_k = \Theta(\log(k))$. Then, either in deterministic quasi-polynomial time $O(\exp(\Theta(\log(r)\log\log(r))))$, *or if N additionally satisfies $N \geq 11/(5d^{-6/5}(1 - d^{-6/5}))$, in* randomized polylogarithmic time $O(\log(r)\log\log(r))$, *we can construct levels $0, \ldots, r$ of a deterministic DAG with layer sizes L_k, indegrees bounded by d, outdegrees bounded by $2d$, and the following edge configuration:*

1) *Every vertex in X_1 has one directed edge from $X_{0,0}$.*
2) *For every pair of consecutive levels k and $k+1$ such that $L_{k+1} = L_k$, the directed edges from X_k to X_{k+1} are given by the edges of B_{L_k}, where we identify the vertices in U_{L_k} with X_k and V_{L_k} with X_{k+1}, respectively.*
3) *For every pair of consecutive levels k and $k+1$ such that $L_{k+1} = 2L_k$, we partition the vertices in X_{k+1} into two sets, $X^1_{k+1} = (X_{k+1,0}, \ldots, X_{k+1,L_k - 1})$ and $X^2_{k+1} = (X_{k+1,L_k}, \ldots, X_{k+1,L_{k+1}-1})$, so that the directed edges from X_k to X^i_{k+1} are given by B_{L_k} for $i = 1, 2$, where we identify U_{L_k} with X_k and V_{L_k} with X^i_{k+1}, as before.*

Furthermore, if this infinite deterministic DAG has all identity processing functions in level $k = 1$, and all majority processing functions in levels $k \geq 2$, then reconstruction is possible:

$$\limsup_{k \to \infty} \mathbb{P}\left(\hat{S}_k \neq X_0\right) < \frac{1}{2}$$

where $\hat{S}_k = \mathbb{1}\{\sigma_k \geq \frac{1}{2}\}$ is the majority decoder.

We refer readers to [13, Section V] for a proof and a detailed discussion. Our quasi-polynomial time algorithm constructs the desired expander graphs by exhaustively enumerating over all d-regular bipartite graphs and testing property (16) by brute force. We expound our randomized polylogarithmic time *Monte Carlo algorithm* in [13, Section V]. In closing, we note that the question of finding a deterministic polynomial time construction of DAGs that admit reconstruction remains open.

References

[1] H. Kesten and B. P. Stigum, "A limit theorem for multidimensional Galton-Watson processes," *The Annals of Mathematical Statistics*, vol. 37, no. 5, pp. 1211–1223, October 1966.

[2] P. M. Bleher, J. Ruiz, and V. A. Zagrebnov, "On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice," *Journal of Statistical Physics*, vol. 79, no. 1-2, pp. 473–482, April 1995.

[3] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman, "Broadcasting on trees and the Ising model," *The Annals of Applied Probability*, vol. 10, no. 2, pp. 410–433, May 2000.

[4] J. von Neumann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components," in *Automata Studies*, ser. Annals of Mathematics Studies, C. E. Shannon and J. McCarthy, Eds., vol. 34. Princeton, NJ, USA: Princeton University Press, 1956, pp. 43–98.

[5] W. S. Evans and L. J. Schulman, "Signal propagation and noisy circuits," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2367–2373, November 1999.

[6] B. Hajek and T. Weller, "On the maximum tolerable noise for reliable computation by formulas," *IEEE Transactions on Infomation Theory*, vol. 37, no. 2, pp. 388–391, March 1991.

[7] W. S. Evans and L. J. Schulman, "On the maximum tolerable noise of k-input gates for reliable computation by formulas," *IEEE Transactions on Information Theory*, vol. 49, no. 11, pp. 3094–3098, November 2003.

[8] L. F. Gray, "A reader's guide to Gacs's "positive rates" paper," *Journal of Statistical Physics*, vol. 103, no. 1-2, pp. 1–44, April 2001.

[9] E. Mossel, "On the impossibility of reconstructing ancestral data and phylogenies," *Journal of Computational Biology*, vol. 10, no. 5, pp. 669–676, July 2003.

[10] ——, "Phase transitions in phylogeny," *Transactions of the American Mathematical Society*, vol. 356, no. 6, pp. 2379–2404, June 2004.

[11] F. Krząkała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, "Gibbs states and the set of solutions of random constraint satisfaction problems," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 104, no. 25, pp. 10 318–10 323, June 2007.

[12] A. Montanari, R. Restrepo, and P. Tetali, "Reconstruction and clustering in random constraint satisfaction problems," *SIAM Journal on Discrete Mathematics*, vol. 25, no. 2, pp. 771–808, July 2011.

[13] A. Makur, E. Mossel, and Y. Polyanskiy, "Broadcasting on random directed acyclic graphs," November 2018, arXiv:1811.03946 [cs.IT].

[14] E. Mossel, "Recursive reconstruction on periodic trees," *Random Structures and Algorithms*, vol. 13, no. 1, pp. 81–97, August 1998.

[15] W. Evans and N. Pippenger, "On the maximum tolerable noise for reliable computation by formulas," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1299–1305, May 1998.

[16] F. Unger, "Noise threshold for universality of two-input gates," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3693–3698, August 2008.

[17] M. S. Pinsker, "On the complexity of a concentrator," in *Proceedings of the 7th International Teletraffic Congress (ITC)*, Stockholm, Sweden, June 13-20 1973, pp. 318/1–318/4.

[18] M. Sipser and D. A. Spielman, "Expander codes," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1710–1722, November 1996.