

Subexponential convergence for information aggregation on regular trees

Yashodhan Kanoria and Andrea Montanari

Abstract—We consider the decentralized binary hypothesis testing problem on trees of bounded degree and increasing depth. For a regular tree of depth t and branching factor $k \geq 2$, we assume that the leaves have access to independent and identically distributed noisy observations of the ‘state of the world’ s . Starting with the leaves, each node makes a decision in a finite alphabet \mathcal{M} , that it sends to its parent in the tree. Finally, the root decides between the two possible states of the world based on the information it receives.

We prove that the error probability vanishes only subexponentially in the number of available observations, under quite general hypotheses. More precisely, in the case of binary messages, decay is subexponential for any decision rule. For general (finite) message alphabet \mathcal{M} , decay is subexponential for ‘node-oblivious’ decision rules, that satisfy a mild irreducibility condition. In the latter case, we propose a family of decision rules with close-to-optimal asymptotic behavior.

I. INTRODUCTION

Let $G = (V, E)$ be a (possibly infinite) network rooted at node \emptyset . Assume that independent and identically distributed noisy observations of an hidden random variable $s \in \{0, 1\}$ are available at a subset $U \subseteq V$ of the vertices. Explicitly, each $i \in U$ has access to a private signal $x_i \in \mathcal{X}$ where $\{x_i\}_{i \in U}$ are independent and identically distributed, conditional on s . The ‘state of the world’ s is drawn from a prior probability distribution $\pi = (\pi_0, \pi_1)$. The objective is to aggregate information about s at the root node under communication constraints encoded by the network structure, while minimizing the error probability at \emptyset .

We ask the following question:

How much does the error probability at the root node \emptyset increase due to these communication constraints?

In order to address this question, consider a sequence of information aggregation problems indexed by t . Information is revealed in a subset of the vertices $U_t \subseteq V$. There are t rounds in which information aggregation occurs. In each round, a subset of the nodes in V make ‘decisions’ that are broadcasted to their neighbors. In round 0, nodes $i \in U_t$ with distance $d(\emptyset, i) = t$ (with $d(\cdot, \cdot)$ being the graph distance) broadcast a decision $\sigma_i \in \mathcal{M}$ to their neighbors, with \mathcal{M} a finite alphabet. In the next round, nodes $i \in V$ with distance $d(\emptyset, i) = t - 1$ broadcast a decision $\sigma_i \in \mathcal{M}$ to their neighbors. And so on, until the neighbors of \emptyset announce their decisions in round t . Finally, the root makes its decision. The decision of any node i is a function of decisions of i ’s

neighbors in earlier rounds, and, if $i \in U_t$, on the private signal x_i received by i .

Clearly, the root can possibly access only the private information available at nodes $i \in V$ with $d(\emptyset, i) \leq t$ (with $d(\cdot, \cdot)$ the graph distance). We can therefore assume, without loss of generality, that $U_t \subseteq \{i \in V : d(\emptyset, i) \leq t\}$. It is convenient to think of U_t as the *information horizon* at time t .

Consider first the case in which communication is unconstrained. This can be modeled by considering the graph with vertices $V = \{\emptyset, 1, 2, 3, \dots\}$ and edges $E = \{(\emptyset, 1), (\emptyset, 2), (\emptyset, 3), \dots\}$. In other words, this is a star network, with the root at the center. Without loss of generality, we take $U_t = \{1, \dots, |U_t|\}$, with $|U_t| \uparrow \infty$ as $t \rightarrow \infty$.

A simple procedure for information aggregation would work as follows. Each node i computes the log-likelihood ratio (LLR) $\ell(x_i)$ corresponding to the observed signal, and quantizes it to a value σ_i . The root adds up the quantized LLRs and decides on the basis of this sum. It follows from basic large deviation theory [2] that, under mild regularity assumptions, the error probability decreases exponentially in the number of observations, i.e.,

$$\mathbb{P}\{\sigma_\emptyset \neq s\} = \exp\{-\Theta(|U_t|)\}. \quad (1)$$

This result is extremely robust:

- (1) It holds for any non-trivial alphabet $|\mathcal{M}| \geq 2$;
- (2) Using concentration-of-measure arguments [3], [4] it is easy to generalize it to families of weakly dependent observations [5];
- (3) It can be generalized to network structures G with weak communications constrains. For instance, [6] proved that the error probability decays exponentially in the number of observations for trees of bounded depth. The crucial observation here is that such networks have large degree diverging with the number of vertices. In particular, for a tree of depth t , the maximum degree is at least $n^{1/t}$.

At the other extreme, Hellmann and Cover [7] considered the case of a line network. In our notations, we have $V = \{\emptyset, 1, 2, 3, \dots\}$, $E = \{(\emptyset, 1), (1, 2), (2, 3), \dots\}$, and $U_t = \{1, 2, \dots, t\}$. In [7] they proved that, as long as the LLRs are bounded (namely $|\ell(x_i)| \leq C$ almost surely for some constant C), and the decision rule is independent of the node, the error probability remains bounded away from 0 as $t \rightarrow \infty$.

If the decision rule is allowed to depend on the node, the error probability can vanish as $t \rightarrow \infty$ provided $|\mathcal{M}| \geq 3$ [8], [9]. Despite this, even if the probability of error decays to 0, it does so much more slowly than for highly connected

Department of Electrical Engineering and Department of Statistics, Stanford University. Supported by 3Com Corporation Stanford Graduate Fellowship and NSF grants CCF-0743978 and CCF-0915145.

networks. Namely, Tay, Tsitsiklis and Win [10] proved that

$$\mathbb{P}\{\sigma_\emptyset \neq s\} = \exp\{-O(|U_t|^\rho)\} \quad (2)$$

for some $\rho < 1$. In other words, the communication constraint is so severe that, after t steps, the amount of information effectively used by the root is equivalent to a vanishingly small fraction of the one within the ‘information horizon’.

These limit cases naturally lead to the general question: Given a rooted network (G, \emptyset) , a sequence of information horizons $\{U_t\}_{t \geq 1}$ and a finite alphabet \mathcal{M} , can information be aggregated at the root in such a way that the error probability decays exponentially in $|U_t|$? The question is wide open, in particular for networks of with average degree bounded or increasing slowly (e.g. logarithmically) with the system size.

Networks with moderate degree arise in a number of practical situations. Within decentralized detection applications, moderate degree is a natural assumption for interference-limited wireless networks. In particular, systems in which a single root node communicates with a significant fraction of the sensors are likely to scale poorly because of interference at the root. Standard models for wireless ad hoc networks [11] are indeed based on random geometric graphs where each node is connected to a logarithmic number of neighbors.

A different domain of applications for models of decentralized decision making is social learning [12]. In this case, each node corresponds to an agent, and the underlying graph is the social network across which information is exchanged. Also in this case, it is reasonable to assume that each agent has a number of neighbors which is bounded, or diverges slowly as the total number of agents grows. In many graph-theoretic models of social networks [13], although a small number of nodes can have large degree, the average degree is bounded or grows logarithmically with the network size.

Given the slow progress with extreme network structures (line networks and highly-connected networks), the study of general moderate degree networks appears extremely challenging. In this paper we focus on regular trees. More precisely, we let G be the (infinite) regular tree with branching factor $k \geq 2$, rooted at \emptyset (each node has k descendants and, with the exception of the root, one parent). The information horizon U_t is formed by all the nodes at distance t from the root, hence $|U_t| = k^t$. Under a broad set of assumptions, we prove that the probability of error decays subexponentially in the size of the information set, cf. Eq. (2), where $\rho = \rho_{\mathcal{M}} < 1$ depends on the size of the alphabet $|\mathcal{M}| = m$.

More precisely, we establish subexponential convergence in the following cases:

- 1) For binary messages $|\mathcal{M}| = 2$ and any choice of the decision rule. In fact, we obtain a precise characterization of the smallest possible error probability in this case.
- 2) For general message alphabet $3 \leq |\mathcal{M}| < \infty$ provided the decision rule does not depend on the node, and satisfies a mild ‘irreducibility’ condition (see Section IV-B for a definition).

In the latter case, one expects that exponential convergence is recovered as the message set gets large. Indeed we prove that, for a given $k \geq 2$, the optimal exponent in Eq. (2) obeys

$$1 - \frac{C_1}{|\mathcal{M}|} \leq \rho_{\mathcal{M}} \leq 1 - \exp\{-C_2|\mathcal{M}|\}. \quad (3)$$

The upper bound follows from our general proof for irreducible decision rules, while the lower bound is obtained by constructing an explicit decision rule that achieves it.

Our investigation leaves several interesting open problems. First, it would be interesting to compute the optimal exponent $\rho = \rho(k, \mathcal{M})$ for given degree of the tree and size of the alphabet. Even the behavior of the exponent for large alphabet sizes is unknown at the moment (cf. Eq. (3)). Second, the question of characterizing the performance limits of general, node-dependent decision rules remains open for $|\mathcal{M}| \geq 3$. Third, it would be interesting to understand the case where non-leaf nodes also get private signals, e.g., $U_t = \{i : i \in V, d(\emptyset, i) \leq t\}$. Finally, this paper focuses on tree of bounded degree. It would be important to explore generalization to other graph structures, namely trees with slowly diverging degrees (which could be natural models for the local structure of preferential attachment graphs [14]), and loopy graphs. Our current results can be extended to trees of diverging degree only in the case of binary signals. In this case we obtain that the probability of error is subexponential

$$\mathbb{P}\{\sigma_\emptyset \neq s\} = \exp\{-o(|U_t|)\} \quad (4)$$

as soon as the degree is sub-polynomial, i.e. $k = o(n^a)$ for all $a > 0$.

The rest of the paper is organized as follows: Section II defines formally the model for information aggregation. Section III presents our results for binary messages $|\mathcal{M}| = 2$. Section IV treats the case of decision rules that do not depend on the node, with general \mathcal{M} .

II. MODEL DEFINITION

As mentioned in the introduction, we assume the network $G = (V, E)$ to be an (infinite) rooted k -ary tree, i.e. a tree where each node has k descendants and one parent (with the exception of the root, that has no parent). Independent noisy observations (‘private signals’) of the state of the world s are provided to the nodes at t -th generation $U_t = \{i \in V : d(\emptyset, i) = t\}$. These will be also referred to as the ‘leaves’. Define $n \equiv |U_t| = k^t$. Formally, the state of the world $s \in \{0, 1\}$ is drawn according to the prior π and for each $i \in U_t$ an independent observation $x_i \in \mathcal{X}$ is drawn with probability distribution $p_0(\cdot)$ (if $s = 0$) or $p_1(\cdot)$ (if $s = 1$). For notational simplicity in this short version, we assume that $\mathcal{X} = \{0, 1\}$, and that $p_s(x) = (1 - \delta)\mathbb{I}(x = s) + \delta\mathbb{I}(x \neq s)$ for $s = 0, 1$; where $\delta \in (0, 1/2)$. Also, we exclude degenerate cases by taking $\pi_0, \pi_1 > 0$. We refer to the two events $\{s = 0\}$ and $\{s = 1\}$ as the hypotheses H_0 and H_1 .

In round 0, each leaf i sends a message $\sigma_i \in \mathcal{M}$ to its parent at level 1. In round 1, the each node j at level 1 sends a message $\sigma_j \in \mathcal{M}$ to its parent at level 2. Similarly

up to round t . Finally, the root node ϕ makes a decision $\sigma_\phi \in \{0, 1\}$ based on the k messages it receives. The objective is to minimize $\mathbb{P}_{\text{err}} \equiv \mathbb{P}(\sigma_\phi \neq s)$. We call a set of decision rules *optimal* if it minimizes \mathbb{P}_{err} .

We will denote by ∂i the set of children of node i . We denote the probability of events under H_0 by $\mathbb{P}_0(\cdot)$, and the probability of events under H_1 by $\mathbb{P}_1(\cdot)$. Finally, we denote by f_i the decision rule at node i in the tree. If i is not a leaf node and $i \neq \phi$, then $f_i : \mathcal{M}^k \rightarrow \mathcal{M}$. The root makes a binary decision $f_\phi : \mathcal{M}^k \rightarrow \{0, 1\}$. If i is a leaf node, it maps its private signal to a message, $f_i : \{0, 1\} \rightarrow \mathcal{M}$. In general, f_i 's can be randomized.

III. BINARY MESSAGES

In this section, we consider the case $\mathcal{M} = \{0, 1\}$, i.e., the case of binary messages.

Consider the case $\pi_0 = \pi_1 = 1/2$. Define the majority decision rule at non-leaf node i as follows: σ_i takes the value of the majority of $\sigma_{\partial i}$ (ties are broken uniformly at random).

It is not hard to see that if we implement majority updates at all non-leaf nodes, we achieve

$$\mathbb{P}_{\text{maj}}(\sigma_\phi \neq s) = \exp \left\{ -\Omega \left(\lfloor (k+1)/2 \rfloor^t \right) \right\} \quad (5)$$

Note that this is an upper bound on error probability under majority updates.

Our main result shows that, in fact, this is essentially the best that can be achieved.

Theorem 3.1: Fix π . There exists $C < \infty$ such that for all $k \in \mathbb{N}$ and $t \in \mathbb{N}$, for any combination of decision rules at the nodes, we have

$$\mathbb{P}(\sigma_\phi \neq s) \geq \exp \left\{ -C \left(\frac{k+1}{2} \right)^t \right\} \quad (6)$$

In particular, the error probability decays subexponentially in the number of private signals $n = k^t$, even with the optimal protocol.

A. Proof of Theorem 3.1

We prove the theorem for the case $\pi_0 = \pi_1 = 1/2$. The proof easily generalizes to arbitrary π .

Also, without loss of generality we can assume that, for every node i ,

$$\frac{\mathbb{P}(s = 1 | \sigma_i = 1)}{\mathbb{P}(s = 0 | \sigma_i = 1)} \geq \frac{\mathbb{P}(s = 1 | \sigma_i = 0)}{\mathbb{P}(s = 0 | \sigma_i = 0)} \quad (7)$$

(otherwise simply exchange the symbols and modify the decision rules accordingly).

Denote by η_i^I the (negative) logarithm of the ‘type I error’ in σ_i , i.e. $\eta_i^I \equiv -\log(\mathbb{P}(s = 0, \sigma_i = 1))$. Denote by η_i^{II} the (negative) logarithm of the ‘type II error’ in σ_i , i.e. $\eta_i^{II} \equiv -\log(\mathbb{P}(s = 1, \sigma_i = 0))$.

The following is the key lemma in our proof of Theorem 3.1.

Lemma 3.2: Given $\delta > 0$, there exists $C \equiv C(\delta) > 0$ such that for any k we have the following: There exists an

optimal set of decision rules such that for any node i at level $\tau \in \mathbb{N}$,

$$\eta_i^I \eta_i^{II} \leq C^2 ((k+1)/2)^{2\tau}. \quad (8)$$

Proof: [Proof of Theorem 3.1] Applying Lemma 3.2 to the root ϕ , we see that $\min(\eta_\phi^I, \eta_\phi^{II}) \leq C((k+1)/2)^t$. The result follows immediately. ■

Lemma 3.2 is proved using the fact that there is an optimal set of decision rules that correspond to deterministic likelihood ratio tests (LRTs) at the non-leaf nodes.

Definition 3.3: Choose a node i . Fix the decision functions of all descendants of i . Define $L_i(\sigma_{\partial i}) = \mathbb{P}(H_1 | \sigma_{\partial i}) / \mathbb{P}(H_0 | \sigma_{\partial i})$.

a) The decision function f_i is a *monotone deterministic likelihood ratio test* if:

- (i) It is deterministic.
- (ii) There is a threshold θ such that

$$\begin{aligned} \mathbb{P}(f_i = 1, L_i < \theta) &= 0 \\ \mathbb{P}(f_i = 0, L_i > \theta) &= 0 \end{aligned}$$

b) The decision function f_i is a *deterministic likelihood ratio test* if either f_i or f_i^c is a monotone deterministic likelihood ratio test. Here f_i^c is the Boolean complement of f_i .

The next lemma is an easy consequence of a beautiful result of Tsitsiklis [15]. Though we state it here only for binary message alphabet, it easily generalizes to arbitrary finite \mathcal{M} .

Lemma 3.4: There is a set of monotone deterministic likelihood ratio tests at the nodes that achieve the minimum possible $\mathbb{P}(\sigma_\phi \neq s)$.

The proof is available in the full version [1].

Clearly, if f_i is a monotone LRT, Eq. (7) holds. In fact, we can argue that there is a set of deterministic monotone LRTs with strict inequality in Eq. (7), i.e., such that

$$\frac{\mathbb{P}(s = 1 | \sigma_i = 1)}{\mathbb{P}(s = 0 | \sigma_i = 1)} > \frac{\mathbb{P}(s = 1 | \sigma_i = 0)}{\mathbb{P}(s = 0 | \sigma_i = 0)} \quad (9)$$

holds for all i , that are optimal. See the full version [1] for details.

Definition 3.5: Let $\bar{\alpha}$ and $\bar{\beta}$ be binary vectors of the same length τ . We say $\bar{\alpha} \succeq \bar{\beta}$ if $\alpha_i \geq \beta_i$ for all $i \in \{1, 2, \dots, \tau\}$. We now prove Lemma 3.2.

Proof: [Proof of Lemma 3.2]

From Lemma 3.4 and Eq. (9), we can restrict attention to monotone deterministic LRTs satisfying Eq. (9).

We proceed via induction on level τ . For any leaf node i , we know that $\eta_i^I = \eta_i^{II} = -\log(\delta/2)$. Choosing $C = -\log(\delta/2)$, Eq. (8) clearly holds for all nodes at level 0. Suppose Eq. (8) holds for all nodes at level τ . Let i be a node at level $\tau+1$. Let its children be $\partial i = \{c_1, c_2, \dots, c_k\}$. Without loss of generality, assume

$$\eta_{c_1}^I \geq \eta_{c_2}^I \geq \dots \geq \eta_{c_k}^I \quad (10)$$

Claim A: We can also assume

$$\eta_{c_1}^{II} \leq \eta_{c_2}^{II} \leq \dots \leq \eta_{c_k}^{II} \quad (11)$$

Recall that $f_i : \{0, 1\}^k \rightarrow \{0, 1\}$ is the decision rule at node i . Assume the first bit in the input corresponds to σ_{c_1} , the second corresponds to σ_{c_2} , and so on. Using Lemma 3.4, we can assume that f_i implements a deterministic likelihood ratio test. Define the k -bit binary vectors $\underline{\omega}^0 = (111\dots 1)$, $\underline{\omega}^1 = (011\dots 1)$, \dots , $\underline{\omega}^k = (00\dots 0)$. From Lemma 3.4 and Eq. (9), it follows that $f_i(\underline{\omega}^j) = \mathbb{I}(j < j_0)$ for some $j_0 \in \{0, 1, \dots, k, k+1\}$.

Claim B: Without loss of generality, we can assume that $j_0 \neq 0$ and $j_0 \neq k+1$.

See the full version [1] for proofs of Claims A and B.

Using Claim B, we can assume $j_0 \in \{1, \dots, k\}$ without loss of generality. Now $\underline{\omega} \succeq \underline{\omega}^{j_0-1}$ contribute to type I error and $\underline{\omega} \preceq \underline{\omega}^{j_0}$ contribute to type II error. It follows that

$$\eta_i^l \leq \sum_{j=j_0}^k \eta_{c_j}^l \leq (k - j_0 + 1) \eta_{c_{j_0}}^l, \quad (12)$$

$$\eta_i^h \leq \sum_{j=1}^{j_0} \eta_{c_j}^h \leq j_0 \eta_{c_{j_0}}^h, \quad (13)$$

where we have used the ordering on the error exponents (Eqs. (10) and (11)). Eqs. (12) and (13) lead immediately to

$$\eta_i^l / \eta_{c_{j_0}}^l + \eta_i^h / \eta_{c_{j_0}}^h \leq (k+1). \quad (14)$$

Now, for any $x, y \geq 0$, we have $x + y \geq 2\sqrt{xy}$. Plugging $x = \eta_i^l / \eta_{c_{j_0}}^l$ and $y = \eta_i^h / \eta_{c_{j_0}}^h$, we obtain from Eq. (14)

$$\eta_i^l \eta_i^h \leq \left(\frac{k+1}{2}\right)^2 \eta_{c_{j_0}}^l \eta_{c_{j_0}}^h. \quad (15)$$

By our induction hypothesis $\eta_{c_{j_0}}^l \eta_{c_{j_0}}^h \leq C^2((k+1)/2)^{2\tau}$. Thus, $\eta_i^l \eta_i^h \leq C^2((k+1)/2)^{2(\tau+1)}$ as required. Induction completes the proof. ■

IV. ‘NODE-OBLIVIOUS’ RULES WITH NON-BINARY MESSAGES

In this section we allow a general finite message alphabet \mathcal{M} that need not be binary. However, we restrict attention to the case of *node-oblivious* rules: The decision rules f_i at all nodes in the tree, except the leaves and the root, must be the same. We denote this ‘internal node’ decision rule by $f : \mathcal{M}^k \rightarrow \mathcal{M}$. Also, the decision rules used at each of the leaf nodes should be same. We denote the leaf decision rule by $g : \{0, 1\} \rightarrow \mathcal{M}$. The decision rule at the root is denoted by $h = f_\emptyset : \mathcal{M}^k \rightarrow \{0, 1\}$. We call such (f, g, h) a node-oblivious decision rule vector.

Define $m \equiv |\mathcal{M}|$. In Section IV-A, we present a scheme that achieves

$$\mathbb{P}(\sigma_\emptyset \neq s) = \exp\left\{-\Omega\left(\left\{k(1-1/m)\right\}^t\right)\right\}, \quad (16)$$

when the error probability in the private signals is sufficiently small. Next, under appropriate assumptions, we show that the decay of error probability must be sub-exponential in the number of private signals k^t .

A. An efficient scheme

For convenience, we label the messages as

$$\mathcal{M} = \left\{ \frac{-m+1}{2}, \frac{-m+3}{2}, \dots, \frac{m-1}{2} \right\} \quad (17)$$

The labels have been chosen so as to be suggestive (in a quantitative sense, see below) of the inferred log-likelihood ratio. Further, we allow the messages to be treated as real numbers (corresponding to their respective labels) that can be operated on. In particular, the quantity $S_i \equiv \sum_{c \in \partial i} \sigma_c$ is well defined for a non-leaf node i .

The node-oblivious decision rule we employ at a non-leaf node $i \neq \emptyset$ is

$$f(\sigma_{\partial i}) = \begin{cases} \left\lfloor \frac{S_i/k+(m-1)/2}{1-1/m} \right\rfloor - \frac{m-1}{2}, & \text{if } S_i \leq 0 \\ \left\lceil \frac{S_i/k-(m-1)/2}{1-1/m} \right\rceil + \frac{m-1}{2}, & \text{if } S_i > 0 \end{cases} \quad (18)$$

Note that the rule is symmetric with respect to an inversion of sign, except that $S_i = 0$ is mapped to the message $1/2$ when m is even.

The rule $g(x_i)$ used at the leafs is simply $g(1) = (m-1)/2$ and $g(0) = -(m-1)/2$. The decision rule at the root is

$$h(\sigma_{\partial \emptyset}) = \begin{cases} 1, & \text{if } S_\emptyset \geq 0 \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

If we associate H_0 with negative quantities, and H_1 with positive quantities, then again, the rule at the leafs is symmetric, and the rule at the root is essentially symmetric (except for the case $S_\emptyset = 0$).

Lemma 4.1: Consider the node-oblivious decision rule vector (f, g, h) defined above. For $k \geq 2$ and $m \geq 3$, there exists $\delta_0 \equiv \delta_0(m, k) > 0$ such that the following is true for all $\delta < \delta_0$:

(i) Under H_0 , for node i at level $\tau \geq 0$, we have

$$-\log \mathbb{P}[\sigma_i = -(m-1)/2 + l] \geq (l/m) \{k(1-1/m)\}^\tau \quad (20)$$

for $l = 1, 2, \dots, m-1$.

(ii) Under H_1 , for node i at level $\tau \geq 0$, we have

$$-\log \mathbb{P}[\sigma_i = (m-1)/2 - l] \geq (l/m) \{k(1-1/m)\}^\tau \quad (21)$$

for $l = 1, 2, \dots, m-1$.

Proof: We prove (i) here. The proof of (ii) is analogous.

Assume H_0 . Define $\gamma \equiv k(1-1/m)$ and $C \equiv k \log m / (k-1)$. We show that, in fact, for suitable choice of δ_0 the following holds: If $\delta < \delta_0$, then for any node i at any level $\tau \geq 0$,

$$-\log \mathbb{P}[\sigma_i = -(m-1)/2 + l] \geq (l/m) \gamma^\tau + C \quad (22)$$

We proceed by induction on τ . Consider i at level $\tau = 0$. We have $\mathbb{P}_0[\sigma_i = -(m-1)/2 + l] = 0$ for $l = 1, 2, \dots, m-2$ and $\mathbb{P}_0[\sigma_i = (m-1)/2] = \delta$. Choosing $\delta_0 \equiv \exp(-1-C)$,

we can ensure that Eq. (22) holds at level 0. Note that for $k \gg 1$, we have $\delta_0 \approx 1/(em)$.

Now suppose Eq. (22) holds at level τ . Consider node i at level $\tau + 1$. From Eq. (18), for $\sigma_i = -(m-1)/2 + l$ we need

$$S_i \geq k[-(m-1)/2 + l(1-1/m)] \quad (23)$$

For every $\sigma_{\partial i} = (-(m-1)/2 + l_1, -(m-1)/2 + l_2, \dots, -(m-1)/2 + l_k)$ such that Eq. (23) holds, we have $\sum_{j=1}^k l_j \geq kl(1-1/m)$. Thus,

$$\begin{aligned} \mathbb{P}_0(\sigma_{\partial i}) &\leq \exp\left(-kC - (1/m)\gamma^\tau \sum_{j=1}^k l_j\right) \\ &\leq \exp(-kC - (1/m)l\gamma^{\tau+1}) \end{aligned} \quad (24)$$

Obviously, there are at most m^k such $\sigma_{\partial i}$. Thus,

$$\begin{aligned} \mathbb{P}_0[\sigma_i = -(m-1)/2 + l] \\ &\leq m^k \exp(-kC - (1/m)l\gamma^{\tau+1}) \\ &= \exp(-C - (1/m)l\gamma^{\tau+1}) \end{aligned}$$

Thus, Eq.(22) holds at level $\tau + 1$. Induction completes the proof. \blacksquare

Theorem 4.2: For $k \geq 2$ and $m \geq 3$, there exists $\delta_0 \equiv \delta_0(m, k) > 0$, and a node-oblivious decision rule vector, such that the following is true: For any $\delta < \delta_0$, we have

$$\begin{aligned} \mathbb{P}[\sigma_\phi \neq s] &\leq \exp\left\{-\frac{m-1}{2m} \{k(1-1/m)\}^t\right\} \\ &= \exp\left\{-\frac{m-1}{2m} n^\rho\right\} \end{aligned} \quad (25)$$

with $\rho \equiv 1 + \log(1-1/m)/\log k$.

The theorem follows easily from Lemma 4.1 and the root decision rule Eq. (19). See the full version [1] for details.

B. Subexponential decay of error probability

Define $n \equiv k^t$, i.e., n is the number of private signals received, one at each leaf. The scheme presented in the previous section allows us to achieve error probability that decays like $\exp(-\Omega(\{k(1-1/m)\}^t)) = \exp(-\Omega(n^\rho))$, where $\rho = 1 + \log(1-1/m)/\log k \approx 1 - 1/(m \log k)$ for $m \gg 1$. In this section we show that under appropriate assumptions, error probability that decays exponentially in n , i.e., $\exp(-\Theta(n))$, is not achievable with node-oblivious rules.

In this section we call the letters of the message alphabet $\mathcal{M} = \{1, 2, \dots, m\}$. For simplicity, we consider only deterministic node-oblivious rules, though our results and proofs extend easily to randomized rules.

We define here a directed graph \mathcal{G} with vertex set \mathcal{M} and edge set \mathcal{E} that we define below. We emphasize that \mathcal{G} is distinct from the tree on which information aggregation is occurring. There is a directed edge from node $\mu_i \in \mathcal{M}$ to node $\mu_j \in \mathcal{M}$ in \mathcal{G} if there exists $\bar{\alpha} \in \mathcal{M}^k$ such that μ_j appears at least once in $\bar{\alpha}$ and $f(\bar{\alpha}) = \mu_i$. Informally, $(\mu_i, \mu_j) \in \mathcal{E}$ if μ_i can be ‘caused’ by a message vector

received from children that includes μ_j . We call \mathcal{G} the *dependence graph*.

We make the following irreducibility assumptions on the node-oblivious decision rule vectors (f, g, h) under consideration (along with leaf and root decision rules).

Assumption 1: The dependence graph \mathcal{G} is strongly connected. In other words, for any $\mu_i \in \mathcal{M}$ and $\mu_j \in \mathcal{M}$, there is a directed path from μ_i to μ_j in \mathcal{G} .

Assumption 2: There exists a level $\tau_p \in \mathbb{N}$ such that for node i at level τ_p , we have $\mathbb{P}_0(\sigma_i = \mu) > 0$ for all $\mu \in \mathcal{M}$.

Note that $\mathbb{P}_0(\sigma_i = \mu) > 0$ implies $\mathbb{P}_1(\sigma_i = \mu) > 0$ by absolute continuity of $\mathbb{P}_0(x_i)$ w.r.t. $\mathbb{P}_1(x_i)$.

Assumption 3: There exists $\mu_- \in \mathcal{M}$, $\mu_+ \in \mathcal{M}$, $\eta > 0$ and $\tau_d \in \mathbb{N}$ such that, for all $\tau > \tau_d$ the following holds: For node i at level τ , we have $\mathbb{P}_0(\sigma_i = \mu_-) > \eta$ and $\mathbb{P}_1(\sigma_i = \mu_+) > \eta$.

In other words, we assume there is one ‘dominant’ message under each of the two possible hypothesis.

It is not hard to verify that for $k \geq 2$, $m \geq 3$ and $\delta < \delta_0(m, k)$ (where δ_0 is same as in Lemma 4.1 and Theorem 4.2), the scheme presented in the previous section satisfies all four of our assumptions. In other words, the assumptions are all satisfied in the regime where our scheme has provably good performance.

Definition 4.3: Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that is strongly connected. For $u, v \in \mathcal{V}$, let d_{uv} be the length of the shortest path from u to v . Then the *diameter* of \mathcal{G} is defined as

$$\text{diameter}(\mathcal{G}) \equiv \max_{u \in \mathcal{V}} \max_{v \in \mathcal{V}, v \neq u} d_{uv}.$$

Theorem 4.4: Fix m and $k \geq 2$. Consider any node-oblivious decision rule vector (f, g, h) such that Assumptions 1, 2 and 3 are satisfied. Let d be the diameter of the dependence graph \mathcal{G} . Then, there exists $C \equiv C(f, m, k) < \infty$ such that we have

$$\mathbb{P}[\sigma_\phi \neq s] \geq \exp\{-Cn^{\bar{\rho}}\}, \quad (26)$$

where $\bar{\rho} \equiv 1 + \frac{\log(1-k^{-d})}{d \log k} < 1$.

Now \mathcal{G} has m vertices, so clearly $d \leq m-1$. The following corollary is immediate.

Corollary 4.5: Fix m and $k \geq 2$. Consider any node-oblivious decision rule vector (f, g, h) such that Assumptions 1, 2 and 3 are satisfied. Then, there exists $C \equiv C(f, m, k) < \infty$ such that we have

$$\mathbb{P}[\sigma_\phi \neq s] \geq \exp\{-Cn^\rho\}, \quad (27)$$

where $\rho \equiv 1 + \frac{\log(1-k^{-(m-1)})}{(m-1) \log k} < 1$.

Thus, we prove that under the above irreducibility assumptions, the error must decay subexponentially in the number of private signals available at the leaves.

Remark 4.6: We have $\mathbb{P}_0(\sigma_{\partial \phi} = (\mu_-, \mu_-, \dots, \mu_-)) > \eta^k$. It follows that we must have $f_\phi(\mu_-, \mu_-, \dots, \mu_-) = 0$ (else the probability of error is bounded below by $\pi_0 \eta^k$ for any t). Similarly, we must have $f_\phi(\mu_+, \mu_+, \dots, \mu_+) = 1$. In particular, $\mu_- \neq \mu_+$.

The following lemma is straightforward to obtain (see [1]).

Lemma 4.7: If Assumption 2 holds, then for a node i at any level $\tau > \tau_p$, we have $\mathbb{P}_0(\sigma_i = \mu) > 0$ for all $\mu \in \mathcal{M}$.

Lemma 4.8 can be thought of as a quantitative version of Lemma 4.7, showing that the probability of the least frequent message decays subexponentially.

Lemma 4.8: Suppose Assumptions 1, 2 and 3 are satisfied. Fix $s \in \{0, 1\}$. Consider a node i at level τ . Define $\zeta_\tau \equiv \min_{\mu \in \mathcal{M}} \mathbb{P}(\sigma_i = \mu | H_s)$. Let $\tau_* = \max(\tau_p, \tau_d)$ (cf. Assumptions 2, 3). Let $d = \text{diameter}(\mathcal{G})$. There exists $C' \equiv C'(f, m, k) < \infty$ such that for any $a \in \mathbb{N} \cup \{0\}$ and $b \in \{0, 1, \dots, d-1\}$, we have,

$$\zeta_{\tau_*+ad+b} \geq \exp\{-C'(k^d - 1)^a\} \quad (28)$$

Proof: Assume H_0 holds, i.e. $s = 0$. The proof for $s = 1$ is analogous.

We prove that, in fact, the following stronger bound holds:

$$-\log(\zeta_{\tau_*+ad+b}) \leq C'(k^d - 1)^a - \log(1/\eta)/(k^d - 2). \quad (29)$$

We proceed via induction on a . First consider $a = 0$. Consider a node i at level $\tau_* + b$ for $b \in \{0, 1, \dots, d-1\}$. Consider the descendants of node i at level τ_* . For any $\mu \in \mathcal{M}$, we know from Lemma 4.7 that there must be *some* assignment of messages to the descendants, such that $\sigma_i = \mu$. It follows that

$$\zeta_{\tau_*+b} \geq \zeta_{\tau_*}^{k^b} \quad (30)$$

Thus, choosing $C' = k^{d-1}(-\log \zeta_{\tau_*}) + \log(1/\eta)/(k^d - 2)$, we can ensure that Eq. (29) holds for $a = 0$ and all $b \in \{0, 1, \dots, d-1\}$.

Now suppose Eq. (29) holds for some $a \in \mathbb{N} \cup \{0\}$. Consider a node i at level $\tau_* + (a+1)d + b$. Let \mathcal{D} be the set of descendants of node i at level $\tau_* + ad + b$. Note that $|\mathcal{D}| = k^d$. Consider any $\mu \in \mathcal{M}$. By Assumption 1, there is a directed path in \mathcal{G} of length at most d going from μ to μ_- . By Remark 4.6, we know that $(\mu_-, \mu_-) \in \mathcal{E}$. It follows that there is a directed path in \mathcal{G} of length *exactly* d going from μ to μ_- . Thus, there must be an assignment of messages $\sigma_{\mathcal{D}}$ to nodes in \mathcal{D} , including at least one occurrence of μ_- , such that $\sigma_i = \mu$. Using Assumption 3, we deduce that

$$\zeta_{\tau_*+(a+1)d+b} \geq \eta \zeta_{\tau_*+ad+b}^{k^d-1}$$

Rewriting as

$$-\log \zeta_{\tau_*+(a+1)d+b} \leq (k^d - 1)(-\log \zeta_{\tau_*+ad+b}) + \log(1/\eta),$$

and combining with Eq. (29), we obtain

$$-\log(\zeta_{\tau_*+(a+1)d+b}) \leq C'(k^d - 1)^{a+1} - \log(1/\eta)/(k^d - 2).$$

Induction completes the proof. \blacksquare

Theorem 4.4 follows easily from Lemma 4.8 (see [1]).

Remark 4.9: For the scheme presented in Section IV-A, we have $d \approx \log_k m$, where $d = \text{diameter}(\mathcal{G})$. For any $\epsilon > 0$,

Theorem 4.4 provides a lower bound on error probability with $\bar{\rho} \leq 1 - C_1/m^{1+\epsilon}$ for some $C_1 \equiv C_1(k, \epsilon) > 0$. This closely matches the m dependence of the upper bound on error probability we proved in Theorem 4.2.

C. Discussion of the irreducibility assumptions

We already mentioned that the efficient node-oblivious rule presented in Section IV-A satisfies all of Assumptions 1, 2 and 3. Moreover, it is natural to expect that similar schemes based on propagation of quantized likelihood ratio estimates should also satisfy our assumptions.

In the full version, we discuss our assumptions taking the cases of binary and ternary messages as examples. We find that for the case of binary messages, each of the irreducibility assumptions *must* be satisfied by any node-oblivious rule for which error probability decays to 0. For the ternary case as well, we argue that schemes violating our irreducibility assumptions should be suboptimal.

We argued above that our irreducibility assumptions are quite reasonable in various circumstances. In fact, we expect the assumptions to be a proof artifact, and conjecture that a subexponential convergence bound holds for general node-oblivious rules. A possible approach to eliminate our assumptions would be to prune the message alphabet \mathcal{M} , discarding letters that never appear, or appear with probability bounded by $\exp(-\Omega(k^t))$ (because they require descendants from a strict subset of \mathcal{M}).

REFERENCES

- [1] Y. Kanoria and A. Montanari, Subexponential convergence for information aggregation on regular trees, arXiv:1104.2939 (2011).
- [2] A. Dembo and O. Zeitouni, Large Deviations Theory and Applications Techniques and Applications, Springer 1998
- [3] M. Ledoux, The Concentration of Measure Phenomenon, American Mathematical Society, 2005
- [4] D. P. Dubhashi and A. Panconesi, Concentration of Measure for the Analysis of Randomized Algorithms, Cambridge University Press, 2009
- [5] Y. Kanoria and A. Montanari, unpublished (2011)
- [6] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, Bayesian Detection in Bounded Height Tree Networks, *IEEE Transactions on Signal Processing*, vol. 57, no. 10, Oct. 2009.
- [7] M. E. Hellman and T. M. Cover, Learning with Finite Memory, *The Annals of Mathematical Statistics*, vol. 41, no. 3, Jun. 1970, 765-782
- [8] T. M. Cover, Hypothesis testing with finite statistics, *The Annals of Mathematical Statistics*, vol. 40, no. 3, Jun. 1969, 828-835
- [9] J. Kopolowitz, Necessary and sufficient memory size for m-hypothesis testing, *IEEE Trans. on Inform. Theory*, vol. 21, no. 1, Jan. 1975, 44-46
- [10] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, On the sub-exponential decay of detection error probabilities in long tandems, *IEEE Trans. on Inform. Theory*, vol. 54, no. 10, pp. 4767-4771, Oct. 2008.
- [11] P. Gupta and P. R. Kumar, The capacity of wireless networks, *IEEE Trans. on Inform. Theory*, vol. 46, no. 2, pp. 388-404, Mar. 2000.
- [12] D. Acemoglu, M.A. Dahleh, I. Lobel, and A. Ozdaglar, Bayesian Learning in Social Networks, LIDS report 2780, to appear in Review of Economic Studies, 2010.
- [13] M. E. J. Newman, D. J. Watts and S. H. S. Strogatz, Random graph models of social networks, *Proc. Natl. Acad. Sciences*, vol. 99, 2566-2572, 2002
- [14] N. Berger, C. Borgs, J. T. Chayes, A. Saberi, A Weak Local Limit for Preferential Attachment Graphs, 2009
- [15] J.N. Tsitsiklis, Extremal Properties of Likelihood-Ratio Quantizers, *IEEE Transactions on Communications*, Vol. 41, No. 4, 1993, pp. 550-558.