Learning in Hierarchical Social Networks

Zhenliang Zhang, *Student Member, IEEE*, Edwin K. P. Chong, *Fellow, IEEE*, Ali Pezeshki, *Member, IEEE*, William Moran, *Member, IEEE*, and Stephen D. Howard, *Member, IEEE*

Abstract—We study a social network consisting of agents organized as a hierarchical M-ary rooted tree, common in enterprise and military organizational structures. The goal is to aggregate information to solve a binary hypothesis testing problem. Each agent at a leaf of the tree, and only such an agent, makes a direct measurement of the underlying true hypothesis. The leaf agent then generates a message and sends it to its supervising agent, at the next level of the tree. Each supervising agent aggregates the messages from the M members of its group, produces a summary message, and sends it to its supervisor at the next level, and so on. Ultimately, the agent at the root of the tree makes an overall decision. We derive upper and lower bounds for the Type I and Type II error probabilities associated with this decision with respect to the number of leaf agents, which in turn characterize the converge rates of the Type I, Type II, and total error probabilities. We also provide a message-passing scheme involving non-binary message alphabets and characterize the exponent of the error probability with respect to the message alphabet size.

Index Terms—Bayesian learning, convergence rate, decentralized detection, tree structure, hypothesis testing, social learning.

I. Introduction

WE consider a binary hypothesis testing problem and an associated social network that attempts (jointly) to solve the problem. The network consists of a set of agents with interconnections among them. Each of the agents makes a measurement of the underlying true hypothesis, observes the past actions of his neighboring agents, and makes a decision to optimize an objective function (e.g., probability of error). In this paper, we are interested in the following questions: Will the agents *asymptotically learn* the underlying true hypothesis? More specifically, will the overall network decision converges in probability to the correct decision as the network size

(number of agents) increases? If so, how fast is the convergence with respect to the network size? In general, the answers to these questions depend on the social network structure. There are two structures primarily studied in the previous literature.

- Feedforward structure: Each of a set of agents makes a decision in sequence based on its private measurement and the decisions of some or all previous agents. For example, we usually decide on which restaurant to dine in or which movie to go to based on our own taste and how popular they appear to be with previous patrons. Investors often behave similarly in asset markets.
- Hierarchical tree structure: Each agent makes a decision based on its private measurement and the decisions of its descendent agents in the tree. This structure is common in enterprises, military hierarchies, political structures, on-line social networks, and even engineering systems (e.g., sensor networks).

The problem of social learning as described above is closely related to the decentralized detection problem. The latter concerns decision making in a sensor network, where each of the sensors is allowed to transmit a summarized message of its measurement (using a compression function) to an overall decision maker (usually called the fusion center). The goal typically is to characterize the optimal compression functions such that the error probability associated with the detection decision at the fusion center is minimized. However, this problem becomes intractable as the network structure gets complicated. Much of the recent work studies the decentralized detection problems in the asymptotic regime, focusing on the problems of the convergence and convergence rate of the error probability.

A. Related Work

The literature on social learning is vast spanning various disciplines including signal processing, game theory, information theory, economics, biology, physics, computer science, and statistics. Here we only review the relevant asymptotic learning results in the two aforementioned network structures.

1) Feedforward Structure: Suppose that a set of agents make decisions sequentially about the underlying truth θ, which equals one of two hypotheses. The first agent makes a measurement of θ and generates a binary decision d_1, which is observed by all the other agents. The second agent makes its decision d_2 based on its own measurement and d_1. Recursively, the decision d_N of the Nth agent is based on its own measurement and the decisions observed from agents 1 to $N-1$. Banerjee [3] and Bikchandani *et al.* [4] show that in the case where the agent signals only allow bounded private belief; i.e., the likelihood-ratio of each signal is bounded, if the first two agents make the same decision, then the rest of the agents

would simply copy this decision ignoring their own measurements, even if their own measurements indicate the opposite hypothesis. This interesting phenomenon is also known as *herding*. Moreover, we have $\lim_{N \to \infty} \mathbb{P}(d_N = \theta) < 1$, which means that the agent decisions do not converge in probability to the underlying true hypothesis as the number of agents goes to infinity; i.e., the agents cannot learn asymptotically. Smith and Sorensen [5] show that if the agent signals allow unbounded private beliefs; i.e., the likelihood-ratio of each signal can be greater than any constant, then these agents learn asymptotically. In other words, the agent decisions converge in probability to the underlying true hypothesis: $\lim_{N \to \infty} \mathbb{P}(d_N = \theta) = 1$. Krishnamurthy [6], [7] studies this problem from the perspective of quickest time change detection. A similar scenario where agents make decisions sequentially but each agent only observes the decision from its immediate previous agent (also known as *tandem network*) is considered in [8]–[12]. Veeravalli [11] shows that the error probability converges sub-exponentially with respect to the number N of agents in the case where the private measurements are independent and identically Gaussian distributed. Tay *et al.* [12] show that the error probability in general converges sub-exponentially and derive an upper bound for the convergence rate of the error probability. Djuric and Wang [13] investigate the evolution of social belief in these structures. Lobel *et al.* [14] derive a lower bound for the convergence rate of the error probability in the feedforward structure where each agent observes a decision randomly from all the previous agents.

2) Hierarchical Tree Structure: In many relevant situations, the social network structure is very complicated, wherein each individual makes its decision not by learning from all the past agent decisions, but from only a subset of agents that are directly connected to this individual. For complex network structures, Jadbabaie *et al.* [15] study the social learning problem from a non-Bayesian perspective. Acemoglu *et al.* [16] provide some sufficient conditions for agents to learn asymptotically from a Bayesian perspective. Cattivelli and Sayed [17] study this problem using a diffusion approach. However, analyzing the convergence rate on learning for complex structures remains largely open.

Recent studies suggest that social networks often exhibit hierarchical structures [18]–[28]. These structures naturally arise from the concept of social hierarchy, which has been observed and extensively studied in fish, birds, and mammals [18]. Hierarchical structures can also be observed in networks of human societies [19]; for example, in enterprise organizations, military hierarchies, political structures [22], and even online social networks [26].

In the special case where the tree height is 1, this structure is usually referred as the *star configuration* [29]–[46]. With the assumption of (conditional) independence of the agent measurements, the error probability in the star configuration converges exponentially with respect to the number N of agents. Tree networks with bounded height (greater than 1) are considered in [47]–[55]. In a tree network, measurements are summarized by leaf agents into smaller messages and sent to their parent agents, each of which fuses all the messages it receives with its own measurement (if any) and then forwards the new message to its parent agent at the next level. This process takes place throughout the tree, culminating at the root where an overall decision is made. In this way, information from each agent is aggregated at the root via a multihop path. Note that the information is 'degraded' along the path. Therefore, the convergence rate for tree networks cannot be better than that of the star configuration. More specifically, under the Bayesian criterion, the error probability converges exponentially fast to 0 with an error exponent that is worse than the one associated with the star configuration [51].

The error probability convergence rate in trees with unbounded height was considered in [56] and [57]. We study in [56] the error probability convergence rate in balanced binary relay trees, where each nonleaf agent in this tree has two child agents and all the leaf agents are at the same distance from the root. Hence, this situation represents the worst-case scenario in the sense that the minimum distance from the root to the leaves is the largest. We show that if each agent in the tree aggregates the messages from its child agents using the unit-threshold likelihood-ratio test, then we can derive tight upper and lower bounds for the total error probability at the root, which characterize the convergence rate of the total error probability. Kanoria and Montanari [57] provide an upper bound for the convergence rate of the error probability in M-ary relay trees (directed trees where each nonleaf node has indegree M and outdegree 1), with any combination of fusion rules for all nonleaf agents. Their result gives an upper bound on the rate at which an agent can learn from others in a social network. To elaborate further, the authors of [57] provide the following upper bound for the convergence rate of the error probability P_N with any combination of fusion rules:

$$\log_2 P_N^{-1} = O\left(N^{\log_M \frac{M+1}{2}}\right). \tag{1}$$

They also provide the following asymptotic lower bound for the convergence rate in the case of majority dominance rule with random tie-breaking: $\log_2 P_N^{-1} = \Omega(N^{\log_M \lfloor (M+1)/2 \rfloor})$. In the case where M is odd, the majority dominance rule achieves the upper bound in (1), which shows that the bound is the optimal convergence rate. However, in the case where M is even, there exists a gap between these two bounds because of the floor function in the second bound. In this case, [57] leaves two questions open:

Q1. Does the majority dominance rule achieve the upper bound in (1)?

Q2. Do there exist other strategies that achieve the upper bound in (1)?

In our paper, for the case where M is even, we answer the first question definitively by showing that the majority dominance rule does *not* achieve the upper bound in (1). For the second question, we provide a strategy that is closer to achieving the upper bound in (1) than the majority dominance rule.

Our paper also differs from (and complements) [57] in a number of other ways. For example, our analysis also includes non-asymptotic results. Moreover, we also consider the

Bayesian likelihood-ratio test[1] (the fusion rule for *Bayesian learning*) as an alternative fusion rule, not considered in [57]. These differences should become clear as we clarify the contributions of this paper in the next section.

B. Contributions

In this paper, we consider the learning problem in social networks configured as M-ary relay trees. Each agent at the leaf level, and only such an agent, takes a direct measurement of the underlying truth and generates a message, which is sent to its parent agent. Each intermediate agent in the tree receives messages from its child agents and aggregates them into a new message, which is again sent to its parent agent at the next level. This process takes place at each nonleaf agent culminating at the root, where a final decision is made. In this way, the information from the leaf agents is aggregated into a summarized form at the decision maker at the root. This hierarchical structure is of interest because it represents the worst-case scenario in the sense that the leaf agents are maximally far away from the decision maker at the root.

In the study of social networks, M-ary relay trees arise naturally. First, as pointed out before, many organizational structures are well described in this way. Also, it is well-known that many real-world social networks, including email networks [58] and the Internet [59], are *scale-free networks*; i.e., the probability $P(\ell)$ that ℓ links are connected to a agent is $P(\ell) \sim c\ell^{-\gamma}$, where c is a normalization constant and the parameter $\gamma \in (2, 3)$. In other words, the number of links does not depend on the network size and is bounded with high probability. Moreover, Newman *et al.* [60] show that the average degree in a social network is bounded or grows very slowly as the network size increases. Therefore, to study the learning problem in social networks, it is reasonable to assume that each nonleaf agent in the tree has a finite number of child agents, in which case the tree height grows unboundedly as the number of agents goes to infinity.

In this paper, we study two ways of aggregating information: the majority dominance rule (a typical non-Bayesian rule) and the Bayesian likelihood-ratio test. Our contributions are as follows:

1) In both cases, we have derived non-asymptotic bounds for the error probabilities with respect to the number of leaf nodes N. These bounds in turn characterize the asymptotic decay rates of the error probabilities.

2) Suppose that the majority dominance rule with random tie-breaking is applied throughout the tree. In the case where M is even, we derive the exact convergence rate of the error probability: $\log_2 P_N^{-1} = \Theta(N^{\log_M \lfloor (M+1)/2 \rfloor})$. Therefore, we show that the majority dominance rule with random tie-breaking does *not* achieve the upper bound in (1). (In the case where M is odd, our asymptotic decay rate is consistent with the result in [57].)

3) Suppose that the Bayesian likelihood-ratio tests is applied. We show that the convergence rate of the error probability

[1]By the Bayesian likelihood-ratio test, we mean a likelihood-ratio test in which the threshold is given by the ratio of the prior probabilities.

is $\log_2 P_N^{-1} = \Omega(N^{\log_M \lfloor (M+1)/2 \rfloor})$. Therefore, the convergence rate in this case is not worse than that in the majority dominance case. Hence in the case where M is odd, the Bayesian likelihood-ratio test also achieves the upper bound in (1).

4) In the case where M is even, we study an alternative majority dominance strategy, which achieves a strictly faster convergence rate than the majority dominance rule with random tie-breaking. The convergence rate of the total error probability using this strategy is $\log_2 P_N^{-1} = \Theta(N^{\log_M \sqrt{M(M+2)/2}})$. The upper bound in (1) involves an *arithmetic* mean of $M + 2$ and M. In contrast, the above rate involves the *geometric* mean of $M + 2$ and M. Therefore, the gap between this rate and the upper bound in (1) is small and almost negligible when M is large. We also show that the Bayesian likelihood-ratio test achieves this convergence rate under certain conditions.

5) We propose a message-passing scheme involving non-binary message alphabets. We derive explicit convergence rates of the total error probabilities in the following cases: any combination of fusion rules, majority dominance rule with random tie-breaking, Bayesian likelihood-ratio test, and alternative majority dominance strategy. We also derive tight upper and lower bounds for the average message size as explicit functions of the spanning factor M.

II. PROBLEM FORMULATION

We consider the problem of binary hypothesis testing between H_0 and H_1, with $\mathbb{P}_0$ and $\mathbb{P}_1$ as the probability measures associated with the two hypotheses. The social network is organized as an M-ary relay tree shown in Fig. 1, in which leaf agents (circles) are agents making independent measurements of the underlying true hypothesis. Only these leaves have direct access to the measurements in the tree structure. These leaf agents then make binary decisions based on their measurements and forward their decisions (messages) to their parent agents at the next level. Each nonleaf agent, with the exception of the root, is a *relay* agent (diamond), which aggregates M binary messages received from its child agents into one new binary message and forwards it to its parent agent again. This process takes place at each agent, culminating at the root (rectangle) where the final decision is made between the two hypotheses based on the messages received. We denote the number of leaf agents by N, which also represents the number of measurements. The height of the tree is $\log_M N$, which grows unboundedly as the number of leaf agents goes to infinity.

We assume that the decisions at all the leaf agents are independent given each hypothesis, and that they have identical Type I error probability (also known as false alarm probability, denoted by α_0) and identical Type II error probability (also known as missed detection probability, denoted by β_0). In this paper, we answer the following questions about the Type I and Type II error probabilities:

- How do they change as we move upward in the tree?
- What are their explicit forms as functions of N?
- Do they converge to 0 at the root?
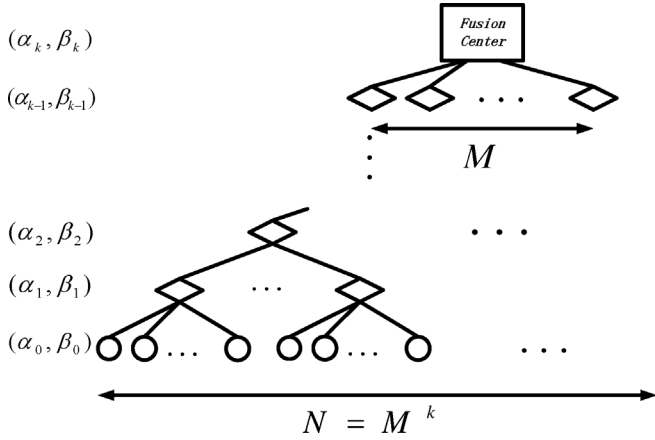- If yes, how fast will they converge with respect to N?

Fig. 1. An M-ary relay tree with height k. Circles represent leaf agents making direct measurements. Diamonds represent relay agents which fuse M binary messages. The rectangle at the root makes an overall decision.

For each nonleaf agent, we consider two ways of aggregating M binary messages:

- In the first case, each nonleaf agent simply aggregates M binary messages into a new binary decision using the majority dominance rule (with random tie-breaking), which is a typical non-Bayesian fusion rule. This way of aggregating information is common in daily life (e.g., voting). For this fusion rule, we provide explicit recursions for the Type I and Type II error probabilities as we move towards the root. We derive bounds for the Type I, Type II, and total error probabilities at the root as explicit functions of N, which in turn characterize the convergence rates.
- In the second case, each nonleaf agent knows the error probabilities associated with the binary messages received and it aggregates M binary messages into a new binary decision using the Bayesian likelihood-ratio test, which is locally optimal in the sense that the total error probability after fusion is minimized. We derive an upper bound for the total error probability, which shows that the convergence speed of the total error probability using this fusion rule is at least as fast as that using the majority dominance rule.

III. ERROR PROBABILITY BOUNDS AND ASYMPTOTIC CONVERGENCE RATES: MAJORITY DOMINANCE

In this section, we consider the case where each nonleaf agent uses the majority dominance rule. We derive explicit upper and lower bounds for the Type I, Type II, and total error probabilities with respect to N. Then, we use these bounds to characterize the asymptotic convergence rates.

A. Error Probability Bounds

We divide our analysis into two cases: oddary tree (M odd) and evenary tree (M even). In each case, we first derive the recursions for the Type I and Type II error probabilities and show that all agents at level k have the same error probability pair (α_k, β_k). Then, we study the step-wise reduction of each kind of error probability. From these we derive upper and lower bounds for the Type I, Type II, and the total error probability at the root.

1) Oddary Tree: We first study the case where the degree of branching M is an odd integer. Consider an

agent at level k, which aggregates M binary messages $\mathbf{u}_i^{k-1} = \{u_1^{k-1}, u_2^{k-1}, \ldots, u_M^{k-1}\}$ from its child agents at level $k-1$, where $u_t^{k-1} \in \{0, 1\}$ for all t. Suppose that u_o^k is the output binary message after fusion, which is again sent to the parent agent at the next level. The majority dominance rule, when M is odd, is simply

$$u_o^k := \begin{cases} 1, & \text{if } \sum_{t=1}^{M} u_t^{k-1} \geq M/2, \\ 0, & \text{if } \sum_{t=1}^{M} u_t^{k-1} \leq M/2. \end{cases}$$

Suppose that the binary messages $\{u_t^{k-1}\}_{t=1}^M$ have identical Type I error probability α and identical Type II error probability β. Then, the Type I and Type II error probability pair (α', β') associated with the output binary message u_o^k is given by:

$$\alpha' = \mathbb{P}_0\left(u_o^k = 1\right) = \prod_{t=1}^{M} \mathbb{P}_0\left(u_t^{k-1} = 1\right)$$

$$+ \binom{M}{1} \mathbb{P}_0\left(u_s^{k-1} = 0\right) \prod_{t=1}^{M-1} \mathbb{P}_0\left(u_t^{k-1} = 1\right) + \cdots$$

$$+ \binom{M}{(M-1)/2} \prod_{s=1}^{(M-1)/2} \mathbb{P}_0\left(u_s^{k-1} = 0\right)$$

$$\times \prod_{t=1}^{(M+1)/2} \mathbb{P}_0\left(u_t^{k-1} = 1\right)$$

$$= f(\alpha),$$

where $f(\alpha) := \alpha^M + \cdots + \binom{M}{(M-1)/2}\alpha^{(M+1)/2}(1-\alpha)^{(M-1)/2}$ and

$$\beta' = \mathbb{P}_1\left(u_o^k = 0\right) = \prod_{t=1}^{M} \mathbb{P}_1\left(u_t^{k-1} = 0\right)$$

$$+ \binom{M}{1} \mathbb{P}_1\left(u_s^{k-1} = 1\right) \prod_{t=1}^{M-1} \mathbb{P}_1\left(u_t^{k-1} = 0\right) + \cdots$$

$$+ \binom{M}{(M-1)/2} \prod_{s=1}^{(M+1)/2} \mathbb{P}_1\left(u_s^{k-1} = 1\right)$$

$$\times \prod_{t=1}^{(M-1)/2} \mathbb{P}_1\left(u_t^{k-1} = 0\right)$$

$$= f(\beta).$$

We assume that all the binary messages from leaf agents have the same error probability pair (α_0, β_0). Hence, all agent decisions at level 1 will have the same error probability pair after fusion: $(\alpha_1, \beta_1) = (f(\alpha_0), f(\beta_0))$. By induction, we have

$$(\alpha_{k+1}, \beta_{k+1}) = (f(\alpha_k), f(\beta_k)), k = 0, 1, \ldots, \log_M N - 1,$$

where (α_k, β_k) represents the error probability pair for agents at the kth level of the tree. Note that the recursions for α_k and β_k are identical. Hence, it suffices to consider only the Type I error probability α_k in deriving the error probability bounds. Before proceeding, we provide the following lemma.

Lemma 1: Let $h_k^M(x) = x^k + \binom{M}{1}x^{k-1}(1-x) + \cdots + \binom{M}{k}(1-x)^k$, where k and M are integers. Suppose that $0 < k < M$. Then, h_k^M is a monotone decreasing function of $x \in (0, 1)$.

Proof: We use induction in M to prove the claim. First we note that $h_0^M(x) = 1$ for all M. Suppose that $M = 2$. Then, we have $h_1^2(x) = 2 - x$. Suppose that $M = 3$. Then, we have $h_1^3(x) = 3 - 2x$ and $h_2^3(x) = x^2 - 3x + 3$. Clearly, in these cases h_k^M are monotone decreasing functions of $x \in (0, 1)$.

Now suppose that h_k^j are monotone decreasing functions of $x \in (0, 1)$ for all $j = 2, \ldots, m - 1$ and $k = 1, \ldots, j - 1$. We wish to show that h_k^m are monotone decreasing functions of $x \in (0, 1)$ for all $k = 1, \ldots, m - 1$. We know that the binomial coefficients satisfy

$$\binom{m}{i} = \binom{m-1}{i-1} + \binom{m-1}{i}$$
$$= \binom{m-1}{i-1} + \binom{m-2}{i-1} + \binom{m-2}{i} = \cdots$$
$$= \binom{m-1}{i-1} + \binom{m-2}{i-1} + \cdots + \binom{k}{i-1} + \binom{k}{i}.$$

We apply the above expansion for all the coefficients in $h_k^m(x)$:

$$h_k^m(x) = x^k + \binom{m}{1}x^{k-1}(1-x) + \cdots + \binom{m}{k}(1-x)^k$$
$$= x^k + \binom{k}{1}x^{k-1}(1-x) + \cdots + \binom{k}{k}(1-x)^k$$
$$+ \binom{k}{0}x^{k-1}(1-x) + \cdots$$
$$+ \binom{k}{k-1}(1-x)^k + \cdots$$
$$+ \binom{m-1}{0}x^{k-1}(1-x) + \cdots$$
$$+ \binom{m-1}{k-1}(1-x)^k$$
$$= 1 + (1-x)h_{k-1}^k(x) + \cdots + (1-x)h_{k-1}^{m-1}(x)$$
$$= 1 + (1-x)\sum_{j=k}^{m-1} h_{k-1}^j(x).$$

By the induction hypothesis, h_{k-1}^j are monotone decreasing for all $j = k, \ldots, m - 1$. Moreover, it is easy to see that h_{k-1}^j are positive for all $j = k, \ldots, m - 1$. Therefore, because the product of two positive monotone decreasing functions is also monotone decreasing, h_k^m is a monotone decreasing function of $x \in (0, 1)$. This completes the proof. ∎

Next we will analyze the step-wise shrinkage of the Type I error probability after each fusion step. This analysis will in turn provide upper and lower bounds for the Type I error probability at the root.

Proposition 1: Consider an M-ary relay tree, where M is an odd integer. Suppose that we apply the majority dominance rule as the fusion rule. Then, for all k we have $1 \leq \alpha_{k+1}/\alpha_k^{(M+1)/2} \leq \binom{M}{(M-1)/2}$.

Proof: Consider the ratio of α_{k+1} and $\alpha_k^{(M+1)/2}$:
$\alpha_{k+1}/\alpha_k^{(M+1)/2} = \alpha_k^{(M-1)/2} + \binom{M}{1}\alpha_k^{(M-3)/2}(1 - \alpha_k) + \cdots + \binom{M}{(M-1)/2}(1 - \alpha_k)^{(M-1)/2}$. First, we derive the lower bound of the ratio. We know that $1 = (\alpha_k + 1 - \alpha_k)^{(M-1)/2} = \alpha_k^{(M-1)/2} + \binom{(M-1)/2}{1}\alpha_k^{(M-3)/2}(1-\alpha_k) + \cdots + \binom{(M-1)/2}{(M-1)/2}(1-$

$\alpha_k)^{(M-1)/2}$. Moreover, it is easy to see that $\binom{M}{k} \geq \binom{(M-1)/2}{k}$ for all $k = 1, 2, \ldots, (M - 1)/2$. Consequently, we have $\alpha_{k+1}/\alpha_k^{(M+1)/2} \geq 1$. Next, we derive the upper bound of the ratio. By Lemma 1, we know that the ratio $\alpha_{k+1}/\alpha_k^{(M+1)/2}$ is monotone increasing as $\alpha_k \to 0$. Hence, we have $\alpha_{k+1}/\alpha_k^{(M+1)/2} \leq \binom{M}{(M-1)/2}$. ∎

The bounds in Proposition 1 hold for all $\alpha_k \in (0, 1)$. Furthermore, the upper bound is achieved at the limit as $\alpha_k \to 0$: $\lim_{\alpha_k \to 0} \alpha_{k+1}/\alpha_k^{(M+1)/2} = \binom{M}{(M-1)/2}$. Using the above proposition, we now derive upper and lower bounds for $\log_2 \alpha_k^{-1}$.

Theorem 1: Consider an M-ary relay tree, where M is an odd integer. Let $\lambda_M = (M + 1)/2$. Suppose that we apply the majority dominance rule as the fusion rule. Then, for all k we have $\lambda_M^k(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M}) \leq \log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}$.

Proof: From the inequalities in Proposition 1, we have $\alpha_{k+1} = c_k \alpha_k^{(M+1)/2} = c_k \alpha_k^{\lambda_M}$, where $c_k \in [1, \binom{M}{(M-1)/2}]$. From these we obtain $\alpha_k = c_{k-1}c_{k-2}^{\lambda_M} \cdots c_0^{\lambda_M^{k-1}} \alpha_0^{\lambda_M^k}$, where $c_i \in [1, \binom{M}{(M-1)/2}]$ for all i, and $\log_2 \alpha_k^{-1} = -\log_2 c_{k-1} - \lambda_M \log_2 c_{k-2} - \cdots - \lambda_M^{k-1} \log_2 c_0 + \lambda_M^k \log_2 \alpha_0^{-1}$. Since $\log_2 c_i \in [0, \log_2 \binom{M}{(M-1)/2}]$, we have $\log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}$. Moreover, we obtain $\log_2 \alpha_k^{-1} \geq -\log_2 \binom{M}{(M-1)/2} - \lambda_M \log_2 \binom{M}{(M-1)/2} - \cdots - \lambda_M^{k-1} \log_2 \binom{M}{(M-1)/2} + \lambda_M^k \log_2 \alpha_0^{-1} = -(\lambda_M^k - 1)/(\lambda_M - 1) \log_2 \binom{M}{(M-1)/2} + \lambda_M^k \log_2 \alpha_0^{-1} \geq \lambda_M^k(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{(M-1)/2}) = \lambda_M^k(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M}))$. ∎

The bounds for $\log_2 \beta_k^{-1}$ are similar and they are omitted for brevity. Note that our result holds for all finite integer k. In addition, our approach provides explicit bounds for both Type I and Type II error probabilities respectively. From the above results, we immediately obtain bounds at the root simply by substituting $k = \log_M N$ into the bounds in Theorem 1.

Corollary 1: Let $P_{F,N}$ be the Type I error probability at the root of an M-ary relay tree, where M is an odd integer. Suppose that we apply the majority dominance rule as the fusion rule. Then, we have $N^{\log_M \lambda_M}(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M}) \leq \log_2 P_{F,N}^{-1} \leq N^{\log_M \lambda_M} \log_2 \alpha_0^{-1}$.

2) Evenary Tree: We now study the case where M is even and derive upper and lower bounds for the Type I error probabilities. The majority dominance rule in this case is

$$u_o^k := \begin{cases} 1, & \text{if } \sum_{t=1}^M u_t^{k-1} > M/2, \\ 1 \text{ w.p. } P_b, & \text{if } \sum_{t=1}^M u_t^{k-1} = M/2, \\ 0 \text{ w.p. } 1 - P_b, & \text{if } \sum_{t=1}^M u_t^{k-1} = M/2, \\ 0, & \text{if } \sum_{t=1}^M u_t^{k-1} < M/2, \end{cases}$$

where $P_b \in (0, 1)$ denotes the Bernoulli parameter for tie-breaking. We first assume that the tie-breaking is fifty-fifty; i.e., $P_b = 1/2$. We will show later that this assumption can be relaxed. The recursions for the Type I and Type II error probabilities are as follows:

$$\alpha_k = \mathbb{P}_0\left(u_o^k = 1\right) = \prod_{t=1}^M \mathbb{P}_0\left(u_t^{k-1} = 1\right)$$

$$+ \binom{M}{1} \mathbb{P}_0 \left(u_s^{k-1} = 0 \right) \prod_{t=1}^{M-1} \mathbb{P}_0 \left(u_t^{k-1} = 1 \right) + \cdots$$

$$+ \frac{1}{2} \binom{M}{M/2} \prod_{s=1}^{M/2} \mathbb{P}_0 \left(u_s^{k-1} = 0 \right) \prod_{t=1}^{M/2} \mathbb{P}_0 \left(u_t^{k-1} = 1 \right)$$

$$= g(\alpha_{k-1}),$$

where $g(\alpha_{k-1}) := \alpha_{k-1}^M + \cdots + (1/2)\binom{M}{M/2}\alpha_{k-1}^{M/2}(1 - \alpha_{k-1})^{M/2}$ and

$$\beta_k = \mathbb{P}_1 \left(u_o^k = 0 \right) = \prod_{t=1}^{M} \mathbb{P}_1 \left(u_t^{k-1} = 0 \right)$$

$$+ \binom{M}{1} \mathbb{P}_1 \left(u_s^{k-1} = 1 \right) \prod_{t=1}^{M-1} \mathbb{P}_1 \left(u_t^{k-1} = 0 \right) + \cdots$$

$$+ \frac{1}{2} \binom{M}{M/2} \prod_{s=1}^{M/2} \mathbb{P}_1 \left(u_s^{k-1} = 1 \right) \prod_{t=1}^{M/2} \mathbb{P}_1 \left(u_t^{k-1} = 0 \right)$$

$$= g(\beta_{k-1}).$$

Next we study the step-wise reduction of each type of error probability when each nonleaf agent uses the majority dominance rule. Again it suffices to consider α_k since the recursions are the same.

Proposition 2: Consider an M-ary relay tree, where M is an even integer. Suppose that we apply the majority dominance rule as the fusion rule. Then, for all k we have $1 \leq \alpha_{k+1}/\alpha_k^{M/2} \leq \binom{M}{M/2}/2$.

The proof is given in Appendix A. The upper bound is achieved at the limit as $\alpha_k \to 0$; i.e., $\lim_{\alpha_k \to 0} \alpha_{k+1}/\alpha_k^{M/2} = \binom{M}{M/2}/2$.

In deriving the above results, we assumed that the tie-breaking rule uses $P_b = 1/2$. Suppose now that the tie is broken with Bernoulli distribution with some arbitrary probability $P_b \in (0, 1)$. Then, it is easy to show that $P_b \leq \alpha_{k+1}/\alpha_k^{M/2} \leq 2^M$. The bounds above are not as tight as those in Proposition 2. However, the asymptotic convergence rates remain the same as we shall see later. Next we derive upper and lower bounds for the Type I error probability at each level k.

Theorem 2: Consider an M-ary relay tree, where M is an even integer. Let $\lambda_M = M/2$. Suppose that we apply the majority dominance rule as the fusion rule. Then, for all k we have $\lambda_M^k (\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M})) \leq \log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}$.

The proof is given in Appendix B. Similar to the oddary tree case, we can provide upper and lower bounds for the Type I error probability at the root.

Corollary 2: Let $P_{F,N}$ be the Type I error probability at the root of an M-ary relay tree, where M is an even integer. Suppose that we apply the majority dominance rule as the fusion rule. Then, we have $N^{\log_M \lambda_M} (\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M})) \leq \log_2 P_{F,N}^{-1} \leq N^{\log_M \lambda_M} \log_2 \alpha_0^{-1}$.

Remark 1: Notice that the above result is only useful when $M \geq 4$. For the case where $M = 2$ (balanced binary relay trees), we have $\alpha_{k+1} = \alpha_k^2 + \alpha_k(1 - \alpha_k) = \alpha_k$ and $\beta_{k+1} = \beta_k^2 + \beta_k(1 - \beta_k) = \beta_k$; that is, the Type I and Type II error probabilities remain the same after fusing with the majority dominance rule.

Remark 2: We have provided a detail analysis in [56] of the convergence rate of the total error probability in balanced binary relay trees ($M = 2$) using the unit-threshold likelihood-ratio test at every nonleaf agent. We show explicit upper and lower bounds for the total error probability at the root as function of the number N of leaf agents, which in turn characterizes the convergence rate $\sqrt{N}$. Moreover, we show that the unit-threshold likelihood-ratio test, which is locally optimal, is close-to-globally optimal in terms of the reduction in the total error probability (see [61] for details).

Remark 3: Notice that the bounds in Corollaries 1 and 2 have the same form. Therefore, the odd and even cases can be unified if we simply let $\lambda_M = \lfloor (M+1)/2 \rfloor$.

In the next section, we use the bounds above to derive upper and lower bounds for the total error probability at the root in the majority dominance rule case.

3) Total Error Probability Bounds: In this section, we provide upper and lower bounds for the total error probability P_N at the root. Let π_0 and π_1 be the prior probabilities for the two underlying hypotheses. It is easy to see that $P_N = \pi_0 P_{F,N} + \pi_1 P_{M,N}$, where $P_{F,N}$ and $P_{M,N}$ correspond to the Type I and Type II error probabilities at the root. With the bounds for each type of error probability in the case where the majority dominance rule is used, we provide bounds for the total error probability as follows.

Theorem 3: Consider an M-ary relay tree, let $\lambda_M = \lfloor (M+1)/2 \rfloor$. Suppose that we apply the majority dominance rule as the fusion rule. Then, we have $N^{\log_M \lambda_M} (\log_2 \max\{\alpha_0, \beta_0\}^{-1} - \log_2 \binom{M}{\lambda_M})) \leq \log_2 P_N^{-1} \leq N^{\log_M \lambda_M} (\pi_0 \log_2 \alpha_0^{-1} + \pi_1 \log_2 \beta_0^{-1})$.

Proof: From the definition of P_N; that is, $P_N = \pi_0 P_{F,N} + \pi_1 P_{M,N}$, we have $P_N \leq \max\{P_{F,N}, P_{M,N}\}$. In addition, we know that α_k and β_k have the same recursion. Therefore, the maximum between the Type I and Type II error probabilities at the root corresponds to the maximum at the leaf agents. Hence, we have $N^{\log_M \lambda_M} (\log_2 \max\{\alpha_0, \beta_0\}^{-1} - \log_2 \binom{M}{\lambda_M})) \leq \log_2 P_N^{-1}$.

By the fact that $\log_2 x^{-1}$ is a convex function, we have $\log_2 P_N^{-1} \leq (\pi_0 \log_2 P_{F,N}^{-1} + \pi_1 \log_2 P_{M,N}^{-1})$. Therefore, we have $\log_2 P_N^{-1} \leq N^{\log_M \lambda_M} (\pi_0 \log_2 \alpha_0^{-1} + \pi_1 \log_2 \beta_0^{-1})$. ∎

These non-asymptotic results are useful. For example, if we want to know how many measurements are required such that $P_N \leq \epsilon$, the answer is simply to find the smallest N that satisfies the inequality in Theorem 3; i.e.,

$$N^{\log_M \lambda_M} \left(\log_2 \max\{\alpha_0, \beta_0\}^{-1} - \log_2 \binom{M}{\lambda_M} \right) \geq \log_2 \epsilon^{-1}.$$

Hence we have

$$N \geq \left(\frac{\log_2 \epsilon^{-1}}{\log_2 \max\{\alpha_0, \beta_0\}^{-1} - \log_2 \binom{M}{\lambda_M}} \right)^{\log_{\lambda_M} M}.$$

The growth rate for the number of measurements is $\Theta((\log_2 \epsilon^{-1})^{\log_{\lambda_M} M})$.

B. Asymptotic Convergence Rates

In this section, we study the convergence rates of error probabilities in the asymptotic regime as $N \to \infty$. We use the fol-

lowing notation to characterize the scaling law of the asymptotic decay rate. Let j and h be positive functions defined on positive integers. We write $j(N) = O(h(N))$ if there exists a positive constant c_1 such that $j(N) \leq c_1 h(N)$ for sufficiently large N. We write $j(N) = \Omega(h(N))$ if there exists a positive constant c_2 such that $j(N) \geq c_2 h(N)$ for sufficiently large N. We write $j(N) = \Theta(h(N))$ if $j(N) = O(h(N))$ and $j(N) = \Omega(h(N))$.

From Corollaries 1 and 2, we can easily derive the decay rates of the Type I and Type II error probabilities. For example, for the Type I error probability, we have the following.

Proposition 3: Consider an M-ary relay tree, let $\lambda_M = \lfloor (M + 1)/2 \rfloor$. Suppose that we apply the majority dominance rule as the fusion rule. Then, we have $\log_2 P_{F,N}^{-1} = \Theta(N^{\log_M \lambda_M})$.

Proof: To analyze the asymptotic rate, we may assume that α_0 is sufficiently small. More specifically, we assume that $\alpha_0 < 1/\binom{M}{\lambda_M}$. In this case, the bounds in Corollaries 1 and 2 show that $\log_2 P_{F,N}^{-1} = \Theta(N^{\log_M \lambda_M})$. ∎

Remark 4: Note that $\log_M \lambda_M$ is monotone increasing with respect to M. Moreover, as M goes to infinity, the limit of $\log_M \lambda_M$ is 1. That is to say, when M is very large, the decay is close to exponential, which is the rate for star configuration and bounded-height trees. In terms of tree structures, when M is very large, the tree becomes short, and therefore achieves similar performance to that of bounded-height trees.

Remark 5: From the fact that the Type I and Type II error probabilities follow the same recursion, it is easy to see that the Type II error probability at the root also decays to 0 with exponent $N^{\log_M \lambda_M}$.

Next, we compute the decay rate of the total error probability.

Corollary 3: Consider an M-ary relay tree, let $\lambda_M = \lfloor (M + 1)/2 \rfloor$. Suppose that we apply the majority dominance rule as the fusion rule. Then, we have $\log_2 P_N^{-1} = \Theta(N^{\log_M \lambda_M})$.

For the total error probability at the root, we have similar arguments with that for individual error probabilities. For large M, the decay of the total error probability is close to exponential.

IV. ERROR PROBABILITY BOUNDS AND ASYMPTOTIC CONVERGENCE RATES: BAYESIAN LIKELIHOOD-RATIO TEST

In this section, we consider the case where the Bayesian likelihood-ratio test is used as the fusion rule. We derive an upper bound for the total error probability, which in turn characterizes the convergence rate. We show that the convergence rate in this case is at least as fast or faster than that with the majority dominance rule.

Theorem 4: Let $\mathbb{P}_N$ be the total error probability at the root in the case where the Bayesian likelihood-ratio test is used as the fusion rule in M-ary relay trees. We have

$$
\log_2 \mathbb{P}_N^{-1}
$$
$$
\geq N^{\log_M \lambda_M} \left(\log_2 L_0^{-1} - \log_2 \left(\frac{2\binom{M}{\lambda_M} \max(\pi_0, \pi_1)}{\min(\pi_0, \pi_1)^{\lambda_M}} \right) \right).
$$

Proof: In the case where the majority dominance rule is used, from Propositions 1 and 2, it is easy to show that $1/2 \leq (\alpha_{k+1} + \beta_{k+1})/(\alpha_k^{\lambda_M} + \beta_k^{\lambda_M}) \leq 2\binom{M}{\lambda_M}$.

Since x^{λ_M} is a convex function for all $M \geq 2$, we have $(\alpha_k^{\lambda_M} + \beta_k^{\lambda_M})/2 \geq ((\alpha_k + \beta_k)/2)^{\lambda_M}$, which implies the following: $2^{-\lambda_M + 1} \leq (\alpha_k^{\lambda_M} + \beta_k^{\lambda_M})/((\alpha_k + \beta_k)^{\lambda_M}) \leq 1$. Hence, we obtain $2^{-\lambda_M} \leq (\alpha_{k+1} + \beta_{k+1})/((\alpha_k + \beta_k)^{\lambda_M}) \leq 2\binom{M}{\lambda_M}$. From these bounds and the fact that $\min(\pi_0, \pi_1)(\alpha_k + \beta_k) \leq \pi_0 \alpha_k + \pi_1 \beta_k \leq \max(\pi_0, \pi_1)(\alpha_k + \beta_k)$, we have $\frac{2^{-\lambda_M} \min(\pi_0, \pi_1)}{\max(\pi_0, \pi_1)^{\lambda_M}} \leq \frac{\pi_0 \alpha_{k+1} + \pi_1 \beta_{k+1}}{(\pi_0 \alpha_k + \pi_1 \beta_k)^{\lambda_M}} \leq \frac{2\binom{M}{\lambda_M} \max(\pi_0, \pi_1)}{\min(\pi_0, \pi_1)^{\lambda_M}}$. Note that $\pi_0 \alpha_k + \pi_1 \beta_k$ is the total error probability for agents at level k and we denote it by L_k.

The Bayesian likelihood-ratio test is the optimal rule in the sense that the total error probability is minimized after fusion. Let L_k^{LRT} be the total error probability after fusing with the Bayesian likelihood-ratio test. We have $L_{k+1}^{LRT}/L_k^{\lambda_M} \leq L_{k+1}/L_k^{\lambda_M} \leq (2\binom{M}{\lambda_M} \max(\pi_0, \pi_1))/(\min(\pi_0, \pi_1)^{\lambda_M})$. Using a similar approach as that used in proving Theorem 1, we can derive the following lower bound for $\log_2 \mathbb{P}_N^{-1}$:

$$
\log_2 \mathbb{P}_N^{-1}
$$
$$
\geq N^{\log_M \lambda_M} \left(\log_2 L_0^{-1} - \log_2 \left(\frac{2\binom{M}{\lambda_M} \max(\pi_0, \pi_1)}{\min(\pi_0, \pi_1)^{\lambda_M}} \right) \right).
$$
∎

From the above bound, we immediately obtain the following.

Corollary 4: Consider an M-ary relay tree, and let $\lambda_M = \lfloor (M + 1)/2 \rfloor$. Suppose that we apply the Bayesian likelihood-ratio test as the fusion rule. Then, we have $\log_2 \mathbb{P}_N^{-1} = \Omega(N^{\log_M \lambda_M})$.

Note that in the case where the majority dominance rule is used, the convergence rate is exactly $\Theta(N^{\log_M \lambda_M})$. Therefore, the convergence rate for the Bayesian likelihood-ratio test is at least as good as that for the majority dominance rule.

V. ASYMPTOTIC OPTIMALITY OF FUSION RULES

In this section, we discuss the asymptotic optimality of the two fusion rules considered in our paper by comparing our asymptotic convergence rates with those in [57], in which it is shown that with any combination of fusion rules, the convergence rate is upper bounded as

$$
\log_2 P_N^{-1} = O\left(N^{\log_M \frac{M+1}{2}} \right). \tag{2}
$$

A. Oddary Case

In the oddary tree case, if each nonleaf agent uses the majority dominance rule, then the upper bound in (2) is achieved; i.e., $\log_2 P_N^{-1} = \Theta(N^{\log_M \lfloor (M+1)/2 \rfloor}) = \Theta(N^{\log_M (M+1)/2})$. This result is also mentioned in [57]. Tay *et al.* [51] find a similar result in bounded-height trees; that is, if the degree of branching for all the agents except those at level 1 is an odd constant, then the majority dominance rule achieves the optimal exponent.

Now we consider the case where each nonleaf agent uses the Bayesian likelihood-ratio test. Since the convergence rate for this fusion rule is at least as good as that for the majority dominance rule, it is evident that the Bayesian likelihood-ratio test, which is only locally optimal (the total error probability after each fusion is minimized), achieves the globally optimal convergence rate. This result is also of interest in decentralized detection problems, in which the objective is usually to find the

globally optimal strategy. In oddary trees, the myopically optimal Bayesian likelihood-ratio test, which is relevant to social learning problems because of the selfishness of agents, is essentially globally optimal in terms of achieving the optimal exponent.

Remark 6: Suppose that each nonleaf agent uses the Bayesian likelihood-ratio test and we assume that the two hypotheses are equally likely. In this case, the output message is given by the unit-threshold likelihood-ratio test: $(\prod_{t=1}^{M} \mathbb{P}_1(u_t^{k-1}))/(\prod_{t=1}^{M} \mathbb{P}_0(u_t^{k-1})) \underset{H_0}{\overset{H_1}{\gtrless}} 1$. If the Type I and Type II error probabilities at level 0 are equal; i.e., $\alpha_0 = \beta_0$, then the unit-threshold likelihood-ratio test reduces to the majority dominance rule. The bounds for the error probabilities in this case and those in the majority dominance rule case are identical.

B. Evenary Case

In the evenary tree case, our results show that with the majority dominance rule, we have

$$\log_2 P_N^{-1} = \Theta\left(N^{\log_M \lfloor \frac{M+1}{2} \rfloor}\right) = \Theta\left(N^{\log_M \frac{M}{2}}\right). \quad (3)$$

This characterizes the explicit convergence rate of the total error probability (cf. [57], in which there is a gap between the upper and lower bounds for $\log_2 P_N^{-1}$). It is evident that the majority dominance rule in this evenary tree case does not achieve the upper bound in (2). However, the gap between the rates described in (2) and (3) becomes smaller and more negligible as the degree M of branching grows.

In the case of binary relay trees ($M = 2$), the gap is most significant because the total error probability does not change after fusion with the majority dominance rule. In contrast, we have shown in [56] that the likelihood-rate test achieves convergence rate $\sqrt{N}$. For $M \geq 4$, we have shown that the convergence rate using the Bayesian likelihood-ratio test is at least as good as that using the majority dominance rule.

Now we consider the case where the *alternative majority dominance strategy* (tie is broken alternatively for agents at consecutive levels) is used throughout the tree. In this case we have $\alpha_k = \alpha_{k-1}^M + \binom{M}{1}\alpha_{k-1}^{M-1}(1 - \alpha_{k-1}) + \cdots + \binom{M}{M/2}\alpha_{k-1}^{M/2}(1 - \alpha_{k-1})^{M/2}$ and $\alpha_{k+1} = \alpha_k^M + \binom{M}{1}\alpha_k^{M-1}(1 - \alpha_k) + \cdots + \binom{M}{M/2-1}\alpha_k^{M/2+1}(1 - \alpha_k)^{M/2-1}$. Using Lemma 1, it is easy to show that

$$1 \leq \frac{\alpha_k}{\alpha_{k-1}^{M/2}} \leq \binom{M}{M/2} \quad \text{and}$$

$$1 \leq \frac{\alpha_{k+1}}{\alpha_k^{M/2+1}} \leq \binom{M}{M/2 - 1}. \quad (4)$$

Theorem 5: Consider an M-ary relay tree, where M is an even integer, and let $\lambda_M = M/2$. Suppose that we apply the alternative majority dominance strategy. Then, for even k we have $\lambda_M^{k/2}(\lambda_M + 1)^{k/2}(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M})) \leq \log_2 \alpha_k^{-1} \leq \lambda_M^{k/2}(\lambda_M + 1)^{k/2}\log_2 \alpha_0^{-1}$.

Proof: The case where $M = 2$ is easy to show using the recursion for α_k and the proof is omitted. Now let us consider the case where $M \geq 4$. From the inequalities in (4), we have $\alpha_{k+1} = c_k\alpha_k^{\lambda_M + 1} = c_k c_{k-1}^{\lambda_M}\alpha_k^{\lambda_M(\lambda_M + 1)}$, where

c_{k-1} and $c_k \in [1, \binom{M}{M/2}]$. From these we obtain $\alpha_k = c_{k-1}c_{k-2}^{\lambda_M}c_{k-3}^{\lambda_M(\lambda_M+1)} \cdots c_0^{\lambda_M^{k/2}(\lambda_M+1)^{k/2-1}} \alpha_0^{\lambda_M^{k/2}(\lambda_M+1)^{k/2}}$, where $c_i \in [1, \binom{M}{M/2}]$ for all i, and $\log_2 \alpha_k^{-1} = -\log_2 c_{k-1} - \cdots - \lambda_M^{k/2}(\lambda_M + 1)^{k/2-1}\log_2 c_0 + \lambda_M^{k/2}(\lambda_M + 1)^{k/2}\log_2 \alpha_0^{-1}$. Since $\log_2 c_i \in [0, \log_2 \binom{M}{M/2}]$, we have $\log_2 \alpha_k^{-1} \leq \lambda_M^{k/2}(\lambda_M + 1)^{k/2}\log_2 \alpha_0^{-1}$. Moreover, we have $\log_2 c_i \leq \log_2 \binom{M}{M/2}$. Hence,

$$\log_2 \alpha_k^{-1} \geq -\log_2 \binom{M}{\lambda_M}(1 + \lambda_M + \lambda_M(\lambda_M + 1)$$
$$+ \cdots + \lambda_M^{k/2}(\lambda_M + 1)^{k/2-1})$$
$$+ \lambda_M^{k/2}(\lambda_M + 1)^{k/2}\log_2 \alpha_0^{-1}. \quad (5)$$

Next we use induction to show that

$$1 + \lambda_M + \lambda_M(\lambda_M + 1) + \cdots + \lambda_M^{k/2}(\lambda_M + 1)^{k/2-1}$$
$$\leq \lambda_M^{k/2}(\lambda_M + 1)^{k/2}. \quad (6)$$

Suppose that $k = 2$. Then, we have $1 + \lambda_M \leq \lambda_M(\lambda_M + 1)$, which holds because $\lambda_M \geq 2$. Suppose that (6) holds when $k = k_0$. We wish to show that it also holds when $k = k_0 + 1$, in which case we have

$$1 + \lambda_M + \cdots + \lambda_M^{k_0/2}(\lambda_M + 1)^{k_0/2-1}$$
$$+ \lambda_M^{k_0/2}(\lambda_M + 1)^{k_0/2} + \lambda_M^{k_0/2+1}(\lambda_M + 1)^{k_0/2}$$
$$\leq 2\lambda_M^{k_0/2}(\lambda_M + 1)^{k_0/2} + \lambda_M^{k_0/2+1}(\lambda_M + 1)^{k_0/2}$$
$$\leq 2\lambda_M^{k_0/2+1}(\lambda_M + 1)^{k_0/2} \leq \lambda_M^{k_0/2+1}(\lambda_M + 1)^{k_0/2+1}.$$

Therefore, we have proved (6). Substituting this result in (5), we obtain the desired lower bound.   ∎

The bounds for $\log_2 \beta_k^{-1}$ are similar and they are omitted for brevity.

Corollary 5: Let $P_{F,N}$ be the Type I error probability at the root of an M-ary relay tree, where M is an even integer. Suppose that we apply the alternative majority dominance strategy. Then, we have $N^{\log_M \sqrt{M(M+2)}/2}(\log_2 \alpha_0^{-1} - \log_2 \binom{M}{\lambda_M})) \leq \log_2 P_{F,N}^{-1} \leq N^{\log_M \sqrt{M(M+2)}/2}\log_2 \alpha_0^{-1}$.

Corollary 6: Let P_N be the total error probability at the root of an M-ary relay tree, where M is an even integer. Suppose that we apply the alternative majority dominance strategy. Then, we have $\log_2 P_{F,N}^{-1} = \Theta(N^{\log_M \sqrt{M(M+2)/2}})$ and $\log_2 P_N^{-1} = \Theta(N^{\log_M \sqrt{M(M+2)/2}})$.

Note that when $M = 2$, $\log_2 P_N^{-1} = \Theta(\sqrt{N})$. Therefore, the decay rate with this strategy is identical with that using the Bayesian likelihood-ratio test. This is not surprising because we show in [56] that the Bayesian likelihood-ratio test is essentially either 'AND' rule or 'OR' rule depending on the values of the Type I and II error probabilities. We also show that the same rule will repeat no more than two consecutive times. Therefore, the decay rate in this case is the same as that using the alternative majority dominance strategy.

For the case where $M \geq 4$, suppose that α_0 and β_0 are sufficiently small and sufficiently close to each other. Then, it is easy to show that the Bayesian likelihood-ratio test is majority dominance rule with tie-breaking given by the values of the Type I and II error probabilities. Moreover, we can show that the same tie-breaking will repeat no more than two
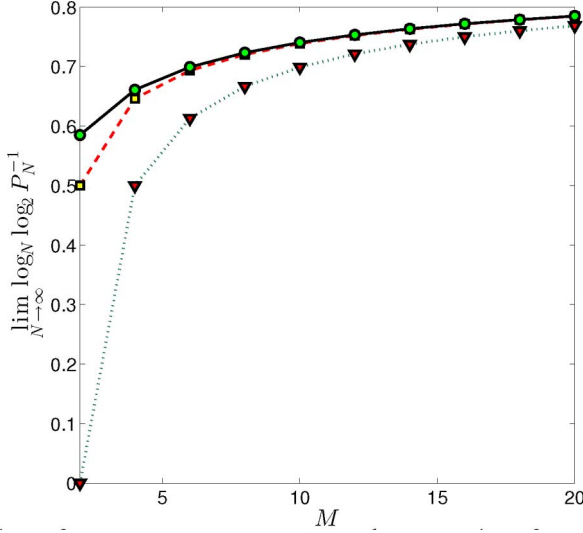
Fig. 2. Plot of error exponents versus the spanning factor M. Dashed (red) line represents the alternative majority dominance strategy. Dotted (blue) line represents the majority dominance rule with random tie-breaking. Solid (black) line represents the upper bound on exponent in (2).

consecutive times. In this case, the error probability decays as $\Theta(N^{\log_M \sqrt{M(M+2)/2}})$.

Recall that the upper bound for the decay rate of the total error probability with all combinations of fusion rules is $O(N^{\log_M (M+1/2)})$, which involves an arithmetic mean of $M + 2$ and M. In contrast, the decay rate using the alternative majority dominance strategy and Bayesian likelihood-ratio test involves the geometric mean of $M + 2$ and M, which means that these two strategies are almost asymptotic optimal, especially when M is large.

The convergence rate of the total error probability using the alternative majority dominance strategy is better that that the random tie-breaking case. For illustration purposes, in Fig. 2 we plot the exponent for the decay rate of the total error probability versus the spanning factor M in these two cases. For comparison purposes, we also plot the exponent in the upper bound (2). We can see from Fig. 2 that alternative majority dominance strategy achieves a larger exponent than that of the majority dominance rule with random tie-breaking. Moreover, the gap between the exponents in the alternative majority dominance strategy case and the upper bound (2) is small and almost negligible.

VI. NON-BINARY MESSAGE ALPHABETS

In the previous sections, each agent in the tree is only allowed to pass a binary message to its supervising agent at the next level. A natural question is, what if each agent can transmit a 'richer' message? In this section, we provide a message-passing scheme that allows general message alphabet of size $\mathcal{D}$ (non-binary). We call this M-ary relay tree with message alphabet size $\mathcal{D}$ an $(M, \mathcal{D})$-tree. We have studied the convergence rates of $(M, 2)$-trees by investigating how fast the total error probability decays to 0. What about the convergence rate when $\mathcal{D}$ is an arbitrary finite integer?

We denote by u_o^k the output message for each agent at the kth level after fusing M input messages $\mathbf{u}_i^{k-1} =$
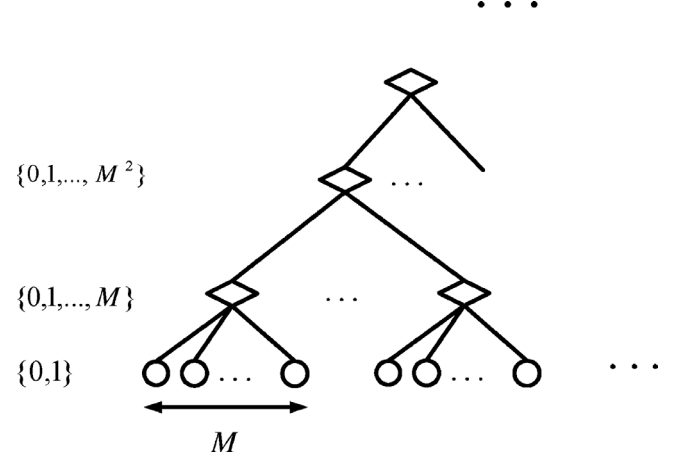

Fig. 3. A message-passing scheme for non-binary message alphabets in an M-ary relay tree.

$\{u_1^{k-1}, u_2^{k-1}, \ldots, u_M^{k-1}\}$ from its child agents at the $(k - 1)$th level, where $u_t^{k-1} \in \{0, 1, \ldots, \mathcal{D}\}$ for all $t \in \{1, 2, \ldots, M\}$.

Case I: First, we consider an $(M, \mathcal{D})$-tree with height k_0, in which there are M^{k_0} leaf agents, and the message alphabet size is sufficiently large; more precisely,

$$\mathcal{D} \geq M^{k_0 - 1} + 1. \tag{7}$$

For our analysis, we need the following terminology:

Definition: Given a nonleaf agent in the tree, a *subtree leaf* of this agent is any leaf agent of the subtree rooted at the agent. An *affirmative subtree leaf* is any subtree leaf that sends a message of '1' upward.

Suppose that each leaf agent still generates a binary message $u_o^0 \in \{0, 1\}$ and sends it upward to its parent agent. Moreover, each intermediate agent simply sums up the messages it receives from its immediate child agents and sends the summation to its parent agent; that is, $u_o^k = \sum_{t=1}^M u_t^{k-1}$. Then we can show that the output message for each agent at the kth level is an integer from $\{0, 1, \ldots, M^k\}$ for all $k \in \{0, 1, \ldots, k_0 - 1\}$. Moreover, this message essentially represents the number of its affirmative subtree leaf.

Because of inequality (7), at each level k in the tree, the message alphabet size $\mathcal{D}$ is large enough to represent all possible values of u_o^k ($k \in \{0, \ldots, k_0 - 1\}$). In particular, the root (at level k_0) knows the number of its affirmative subtree leaves. In this case, the convergence rate is the same as that of the star configuration, where each leaf agent sends a binary message to the root directly. Recall that in the star configurations, the total error probability decays exponentially fast to 0.

Case II: We now consider the case where the tree height is very large; i.e., (7) does not hold. As shown in Fig. 3, we apply the scheme described in Case I; that is, the leaf agents send binary compressions of their measurements upward to their parent agents. Moreover, each intermediate agent simply sends the sum of the messages received to its parent agent:

$$u_o^k = \sum_{t=1}^M u_t^{k-1}. \tag{8}$$

From the assumption of large tree height, it is easy to see that the message alphabet size is not large enough for all the relay

agents to use the fusion rule described in (8). With some abuse of notation, we let k_0 to be the integer $k_0 = \lfloor \log_M(\mathcal{D}-1) \rfloor + 1$ (here, k_0 is not the height of the tree; it is strictly less than the height). Note that $M^{k_0-1} + 1 \le \mathcal{D} < M^{k_0} + 1$.

From the previous analysis, we can see that with this scheme, each agent at the k_0th level knows the number of its affirmative subtree leaves. Therefore, it is equivalent to consider the case where each agent at level k_0 connects to its M^{k_0} subtree leaves directly (all the intermediate agents in the subtree can be ignored). However, we cannot use the fusion rule described in (8) for the agents at k_0th level to generate the output messages because the message alphabet size is not large enough. Hence, we let each agent at level k_0 aggregate the M^{k_0} binary messages from its subtree leaves into a new binary message (using some fusion rule). By doing so, the output message from each agent at the k_0th level is binary again. Henceforth, we can simply apply the fusion rule (8) and repeat this process throughout the tree, culminating at the root. We now provide an upper bound for the asymptotic decay rate in this case.

Theorem 6: The convergence rate of the total error probability for an $(M, \mathcal{D})$-tree is equal to that for an $(M^{k_0}, 2)$-tree, where $k_0 = \lfloor \log_M(\mathcal{D}-1) \rfloor + 1$. In particular, let P_N be the total error probability at the root for an $(M, \mathcal{D})$-tree. With any combination of fusion rules at level ℓk_0, $\ell = 1, 2, \ldots$, we have $\log_2 P_N^{-1} = O(N^\rho)$, where $\rho := \frac{\ln(M^{k_0}+1)}{\ln M^{k_0}} - \frac{\log_M 2}{k_0}$.

Proof: Consider an $(M, \mathcal{D})$-tree with the scheme described above. It is easy to see that equivalently we can consider a tree where the leaf agents connect to the agents at the k_0th level directly. In addition, because of the recursive strategy applied throughout the tree, it suffices to consider the tree where the agents at the ℓk_0th level connect to the agents at the $(\ell+1)k_0$th level directly for all non-negative integers ℓ. Therefore, the convergence rate of an $(M, \mathcal{D})$-tree is equal to that of the corresponding $(M^{k_0}, 2)$-tree.

In the asymptotic regime, the decay rate in $(M, 2)$-trees is bounded above as follows [57]: $\log_2 P_N^{-1} = O(N^{\log_M (M+1)/2})$. Therefore, the decay rate for $(M^{k_0}, 2)$-trees is also bounded above as $\log_2 P_N^{-1} = O(N^{\log_{M^{k_0}} (M^{k_0}+1)/2})$, which upon simplification gives the desired result. ∎

Suppose that each agent at level ℓk_0 for all ℓ uses the majority dominance rule. Then, we can derive the convergence rate for the total error probability as follows.

Theorem 7: Consider $(M, \mathcal{D})$-trees where the majority dominance rule is used. Let $k_0 = \lfloor \log_M(\mathcal{D}-1) \rfloor + 1$. We have $\log_2 P_N^{-1} = \Theta(N^\varrho)$, where

$$\varrho := \begin{cases} \frac{\ln(M^{k_0}+1)}{\ln M^{k_0}} - \frac{\log_M 2}{k_0}, & \text{if } M \text{ is odd,} \\ 1 - \frac{\log_M 2}{k_0}, & \text{if } M \text{ is even.} \end{cases}$$

Proof: By Theorem 6, the performance of $(M, \mathcal{D})$-trees is equal to that of $(M^{k_0}, 2)$-trees, where $k_0 = \lfloor \log_M(\mathcal{D}-1) \rfloor + 1$. For the asymptotic rate, we have $\log_2 P_N^{-1} = \Theta(N^{\log_{M^{k_0}} \lfloor (M^{k_0}+1)/2 \rfloor})$, which upon simplification gives the desired result. ∎

Remark 7: Notice that $\lim_{M \to \infty} \ln(M^{k_0}+1)/\ln M^{k_0} = 1$, which means that the even and odd cases in the expression for ϱ are similar when M is large.

Remark 8: From Theorem 7, we can see that with larger message alphabet size, the total error probability decays more quickly. However, the change in the decay exponent is not significant because k_0 depends on $\mathcal{D}$ logarithmically. Furthermore, if M is large, then the change in the performance is less sensitive to the increase in $\mathcal{D}$.

Remark 9: Comparing the results in Theorems 6 and 7, we can see that the majority dominance rule achieves the optimal exponent in the oddary case and it almost achieves the optimal exponent in the evenary case.

For the Bayesian likelihood-ratio test, we have the following result.

Theorem 8: The convergence rate using the likelihood-ratio test is at least as good as that using the majority dominance rule; i.e., $\log_2 \mathbb{P}_N^{-1} = \Omega(N^\varrho)$.

In the case where M is even, we can derive the decay rate using the alternative majority dominance strategy.

Theorem 9: The convergence rate using the alternative majority dominance strategy is $\log_2 P_N^{-1} = \Omega(N^\sigma)$, where $\sigma = \frac{1}{2}\left(\frac{1+\ln(M^{k_0}+2)}{\ln M^{k_0}}\right) - \frac{\log_M 2}{k_0}$.

Theorem 8 and 9 follow by applying the same arguments as those made in proofs of Corollary 4 and Theorem 6 and the proofs are omitted for brevity.

The message-passing scheme provided here requires message alphabets with maximum size $\mathcal{D}$. However, most of the agents use much 'smaller' messages. For example, the leaf agents generate binary messages. It is interesting to characterize the *average* message size used in our scheme. Because of the recursive strategy, it suffices to calculate the average message size in a subtree with height $k_0 - 1$ since the message sizes in our scheme repeat every k_0 levels. The message size (in bits) for agents at level $t \in \{0, 1, \ldots, k_0 - 1\}$ is $\log_2(M^t + 1)$ and the number of agents at level t is M^{k_0-t}. Hence, the average size $\bar{b}(k_0)$ in bits used in our scheme is

$$\bar{b}(k_0) = \frac{M^{k_0} + \ldots + M \log_2(M^{k_0-1}+1)}{M^{k_0} + M^{k_0-1} + \ldots + M}$$
$$= \frac{\sum_{t=0}^{k_0-1} M^{k_0-t} \log_2(M^t+1)}{\sum_{t=0}^{k_0-1} M^{t+1}}.$$

We have $\log_2(M^t+1) > \log_2 M^t = t \log_2 M$ and $\log_2(M^t+1) < \log_2(2M^t) = 1 + t \log_2 M$ for all $t \ge 1$. Hence, the average size in bits is lower bounded as

$$\bar{b}(k_0) > \frac{M^{k_0} + M^{k_0-1} \log_2 M + \cdots + M(k_0 - 1) \log_2 M}{M^{k_0} + M^{k_0-1} + \cdots + M}$$
$$= \frac{M^{k_0}}{M^{k_0} + M^{k_0-1} + \cdots + M}$$
$$\quad + \frac{\log_2 M(M^2(M^{k_0-1} - 1) - M(M-1)(k_0-1))}{(M^{k_0} + M^{k_0-1} + \cdots + M)(M-1)^2}$$
$$= \frac{M^{k_0} - M^{k_0-1}}{M^{k_0} - 1} + \frac{M \log_2 M}{M-1}$$
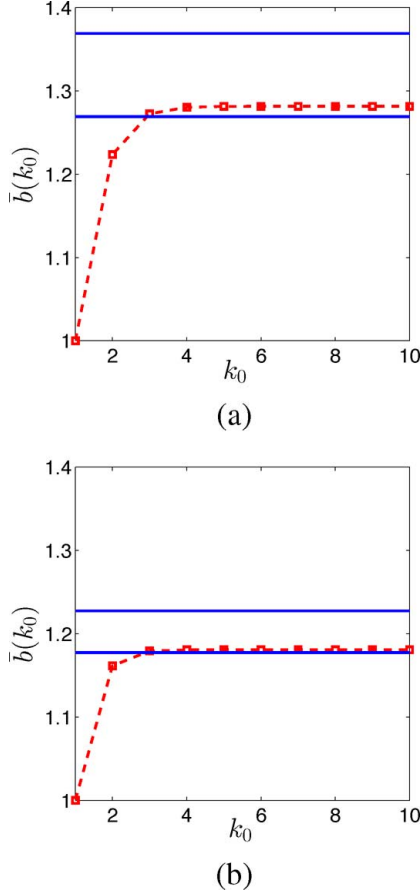$$\quad \times \frac{M^{k_0-1} - 1 - M(M-1)(k_0-1)}{M^{k_0} - 1}.$$

Fig. 4. (a) Average message size (dashed red line) in $M = 10$ case. (b) Average message size (dashed red line) in $M = 20$ case. The blue lines represent the bounds in (9).

In addition, it is upper bounded as

$$\bar{b}(k_0) < 1 + \frac{M \log_2 M}{M - 1} \frac{M^{k_0 - 1} - 1 - M(M - 1)(k_0 - 1)}{M^{k_0} - 1}$$
$$\leq 1 + \frac{\log_2 M}{M - 1}.$$

Recall that, with sufficiently large k_0, the error probability convergence rates are close to exponential. However, from the above bounds the average message size in terms of bits in our scheme is still very small, specifically for sufficiently large k_0 we have

$$1 + \frac{\log_2 M}{M - 1} - \frac{1}{M} \leq \bar{b}(k_0) \leq 1 + \frac{\log_2 M}{M - 1}. \tag{9}$$

Fig. 4 shows plots of the average message sizes $\bar{b}(k_0)$ versus k_0 in the $M = 10$ and 20 cases. Note that as M increases, the average message size becomes smaller and the bounds in (9) become tighter.

VII. CONCLUDING REMARKS

We have studied the social learning problem in the context of M-ary relay trees. We have analyzed the step-wise reductions of the Type I and Type II error probabilities and derived upper and lower bounds for each error probability at the root as explicit functions of N, which characterize the convergence rates for Type I, Type II, and the total error probabilities. We have shown that the majority dominance rule is not better than the Bayesian likelihood-ratio test in terms of convergence rate. We have studied the convergence rate using the alternative majority dominance strategy. Last, we have provided a message-passing scheme which increases the convergence rate of the total error probability. We have shown quantitatively how the convergence rate varies with respect to the message alphabet sizes. This scheme is very efficient in terms of the average message size used for communication.

Many interesting questions remain. Social networks usually involve very complex topologies. For example, the degree of branching may vary among different agents in the network. The convergence rate analysis for general complex structures is still wide open. Another question involves the assumption that the agent measurements are conditionally independent. It is of interest to study the scenario where these agent measurements are correlated. This scenario has been studied in the star configuration [62]–[64] but not in any other structures yet. Yet another question is related with the assumption that the communications and agents are perfectly reliable. We would like to study the rate of learning in cases where communications and agents are non-ideal [65].

APPENDIX A
PROOF OF PROPOSITION 2

Proof: We consider the ratio of α_{k+1} and $\alpha_k^{M/2}$: $\alpha_{k+1}/\alpha_k^{M/2} = \alpha_k^{M/2} + \binom{M}{1}\alpha_k^{(M-2)/2}(1 - \alpha_k) + \cdots + (1/2)\binom{M}{M/2}(1 - \alpha_k)^{M/2}$. First, we show the lower bound of the ratio. We know that $(\alpha_k + 1 - \alpha_k)^{M/2} = \alpha_k^{M/2} + \binom{M/2}{1}\alpha_k^{(M-2)/2}(1-\alpha_k) + \cdots + \binom{M/2}{M/2}(1-\alpha_k)^{M/2} = 1$ and $\binom{M}{k} \geq \binom{M/2}{k}$ for all $k = 1, 2, \ldots, M/2$. Moreover, we have $\binom{M}{M/2}/2 \geq \binom{M/2}{M/2} = 1$. In consequence, we have $\alpha_{k+1}/\alpha_k^{M/2} \geq 1$. Notice that $\alpha_{k+1}/\alpha_k^{M/2} = h_{M/2}^M(\alpha_k)/2 + h_{M/2-1}^M(\alpha_k)/2$. By Lemma 1, the ratio is monotone increasing as $\alpha_k \to 0$. Hence, we have $\alpha_{k+1}/\alpha_k^{M/2} \leq (1/2)\binom{M}{M/2}$. ∎

APPENDIX B
PROOF OF THEOREM 2

Proof: From the inequalities in Proposition 2, we have $\alpha_{k+1} = c_k \alpha_k^{M/2} = c_k \alpha_k^{\lambda_M}$, where $c_k \in [1, \binom{M}{M/2}/2]$. From these we obtain $\alpha_k = c_{k-1}c_{k-2}^{\lambda_M} \cdots c_0^{\lambda_M^{k-1}} \alpha_0^{\lambda_M^k}$, where $c_i \in [1, \binom{M}{M/2}/2]$ for all i, and $\log_2 \alpha_k^{-1} = -\log_2 c_{k-1} - \lambda_M \log_2 c_{k-2} - \cdots - \lambda_M^{k-1} \log_2 c_0 + \lambda_M^k \log_2 \alpha_0^{-1}$. Since $\log_2 c_i \in [0, \log_2 \binom{M}{M/2} - 1]$, we have $\log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}$. Moreover, we obtain $\log_2 \alpha_k^{-1} \geq \lambda_M^k (\log_2 \alpha_0^{-1} - \log_2 \binom{M}{M/2})$. ∎

REFERENCES

[1] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Detection performance of M-ary relay trees with non-binary message alphabets," in *Proc. Statist. Signal Process. Workshop*, Ann Arbor, MI, Aug. 5–8, 2012, pp. 796–799.

[2] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Rate of learning in hierarchical social networks," in *Proc. 50th Allerton Conf.*, Urbana, IL, USA, Oct. 2012, to be published.

[3] A. V. Banarjee, "A simple model for herd behavior," *Quart. J. Econ.*, vol. 107, no. 3, pp. 797–817, Aug. 1992.

[4] S. Bikchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as information cascades," *J. Political Econ.*, vol. 100, no. 5, pp. 992–1026, Oct. 1992.

[5] L. Smith and P. Sorensen, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, no. 2, pp. 371–398, Mar. 2000.

[6] V. Krishnamurthy, "Bayesian sequential detection with phase-distributed change time and nonlinear penalty—A POMDP lattice programming approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 7096–7124, Oct. 2011.

[7] V. Krishnamurthy, "Quickest detection POMDPs with social learning: Interaction of local and global decision makers," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5563–5587, Aug. 2012.

[8] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Part I—Tandem structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 5, pp. 1044–1059, Sep./Oct. 1991.

[9] R. Viswanathan, S. C. A. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 366–376, Jul. 1988.

[10] J. D. Papastravrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 639–653, Jul. 1992.

[11] V. V. Veeravalli, "Topics in Decentralized Detection," Ph.D. Dissertation, Univ. Illinois, Urbana-Champaign, IL, USA, 1992.

[12] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the sub-exponential decay of detection error probabilities in long tandems," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4767–4771, Oct. 2008.

[13] P. M. Djuric and Y. Wang, "Disturbed Bayesian learning in multiagent systems: Improving our understanding of its capabilities and limitations," *IEEE Signal Process. Mag.*, vol. 29, no. 2, pp. 65–76, Mar. 2012.

[14] I. Lobel, D. Acemoglu, M. Dahleh, and A. Ozdaglar, "Lower bounds on the rate of learning in social networks," in *Proc. IEEE American Control Conf.*, St. Louis, MO, USA, Jun. 10–12, 2009, pp. 2825–2830.

[15] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *PIER Working Paper 11-025*, Aug. 2011.

[16] D. Acemoglu, M. Dahleh, I. Lobel, and A. E. Ozdaglar, "Bayesian learning in social networks," *Rev. Econ. Stud.*, vol. 78, no. 4, 2011.

[17] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Tran. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, May 2011.

[18] K. A. Jameson, M. C. Appleby, and L. C. Freeman, "Finding an appropriate order for a hierarchy based on probabilistic dominance," *Animal Behaviour*, vol. 57, no. 5, pp. 991–998, May 1999.

[19] S. Wasserman, K. Faust, and D. Iacobucci, *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge, U.K.: Cambridge Univ. Press, Nov. 1994.

[20] D. J. Watts, P. S. Dodds, and M. E. J. Newman, "Identity and search in social networks," *Science*, vol. 296, no. 5571, pp. 1302–1305, May 2002.

[21] A. Clauset, C. Moore, and M. E. J. Newman, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, pp. 98–101, May 2008.

[22] A. S. Maiya and T. Y. Berger-Wolf, "Inferring the maximum likelihood hierarchy in social networks," in *Proc. 12th IEEE Intl. Conf. Comput. Sci., Eng. (CSE'09)*, Vancouver, BC, Canada, Aug. 2009.

[23] Y. Kanoria and O. Tamuzy, "Tractable Bayesian social learning," in *Proc. IEEE Int. Symp. Inf. Theory*, Boston, MA, USA, Jul. 2012, pp. 2721–2725.

[24] L. Lopez, J. F. F. Mendes, and M. A. F. Sanjuan, "Hierarchical social networks and information flow," *Phys. A*, vol. 316, no. 1–4, pp. 695–708, Dec. 2002.

[25] C. C. Eckel and R. K. Wilson, "Social learning in a social hierarchy: An experimental study," in *Assoc. Adv. Sci.*, Washington, D.C., Feb. 16–22, 2000.

[26] M. Gupte, P. Shankar, J. Li, S. Muthukrishnan, and L. Iftode, "Finding hierarchy in directed online social networks," in *Proc. 20th Int. World Wide Web Conf. (WWW)*, Hyderabad, India, Mar. 28–Apr. 1 2011.

[27] A. E. Motter, T. Nashikawa, and Y.-C. Lai, "Large-scale structural organization of social networks," *Phys. Rev. E*, vol. 68, no. 3, p. 036105, Sep. 2003.

[28] J. D. Noh, "Exact scaling properties of a hierarchical network model," *Phys. Rev. E*, vol. 67, no. 4, p. 045103, Apr. 2003.

[29] S. Choi, B. Celen, and S. Kariv, "Learning in networks: An experimental study," Dec. 2005, UCLA Tech. Rep..

[30] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, no. 4, pp. 501–510, Jul. 1981.

[31] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.

[32] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.

[33] J. N. Tsitsiklis, "Decentralized detection," *Adv. Statist. Signal Process.*, vol. 2, pp. 297–344, 1993.

[34] G. Polychronopoulos and J. N. Tsitsiklis, "Explicit solutions for some simple decentralized detection problems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 2, pp. 282–292, Mar. 1990.

[35] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Asymptotic performance of a censoring sensor network," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4191–4209, Nov. 2007.

[36] P. K. Willett and D. Warren, "The suboptimality of randomized tests in distributed and quantized detection systems," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 355–361, Mar. 1992.

[37] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I—Fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[38] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II—Advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.

[39] T. M. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 2, pp. 469–476, Apr. 1998.

[40] B. Chen and P. K. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 693–699, Feb. 2005.

[41] B. Liu and B. Chen, "Channel-optimized quantizers for decentralized detection in sensor networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3349–3358, Jul. 2006.

[42] B. Chen and P. K. Varshney, "A Bayesian sampling approach to decision fusion using hierarchical models," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1809–1818, Aug. 2002.

[43] A. Kashyap, "Comments on on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1274–1275, Mar. 2006.

[44] H. Chen, B. Chen, and P. K. Varshney, "Further results on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 828–832, Feb. 2009.

[45] G. Fellouris and G. V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 534–548, Jan. 2011.

[46] J. A. Gubner, L. L. Scharf, and E. K. P. Chong, "Exponential error bounds for binary detection using arbitrary binary sensors and an all-purpose fusion rule in wireless sensor networks," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Process.*, Taipei, Taiwan, Apr. 19–24, 2009, pp. 2781–2784.

[47] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Part II—Tree structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 1, pp. 211–221, Jan./Feb. 1993.

[48] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Data fusion trees for detecion: Does architecture matter?," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4155–4168, Sep. 2008.

[49] W. P. Tay and J. N. Tsitsiklis, "Error exponents for decentralized detection in tree networks," in *Networked Sensing Information and Control*, V. Saligrama, Ed. New York, NY, USA: Springer-Verlag, 2008, pp. 73–92.

[50] A. R. Reibman and L. W. Nolte, "Design and performance comparison of distributed detection networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 6, pp. 789–797, Nov. 1987.

[51] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Bayesian detection in bounded height tree networks," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4042–4051, Oct. 2009.

[52] A. Pete, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks with multiple event structures," *IEEE Trans. Autom. Control*, vol. 39, no. 8, pp. 1702–1707, Aug. 1994.

[53] O. P. Kreidl and A. S. Willsky, "An efficient message-passing algorithm for optimizing decentralized detection networks," *IEEE Trans. Autom. Control*, vol. 55, no. 3, pp. 563–578, Mar. 2010.

[54] S. Alhakeem and P. K. Varshney, "A unified approach to the design of decentralized detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 9–20, Jan. 1995.

[55] Y. Lin, B. Chen, and P. K. Varshney, "Decision fusion rules in multi-hop wireless sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 2, pp. 475–488, Apr. 2005.

[56] Z. Zhang, A. Pezeshki, W. Moran, S. D. Howard, and E. K. P. Chong, "Error probability bounds for balanced binary relay trees," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3548–3563, Jun. 2012.

[57] Y. Kanoria and A. Montanari, "Subexponential convergence for information aggregation on regular trees," in *Proc. Joint 50th IEEE Conf. Decision Control Eur. Control Conf.*, Orlando, FL, USA, Dec. 12–15, 2011, pp. 5317–5322.

[58] H. Ebel, L.-I. Mielsch, and S. Bornholdt, "Scale-free topology of e-mail networks," *Phys. Rev. E*, vol. 66, no. 3, p. 035103, Sep. 2002.

[59] S.-H. Yook, H. Jeong, and A.-L. Barabasi, "Modeling the Internet's large-scale topology," *Proc. Nat. Acad. Sci.*, vol. 99, no. 21, pp. 13382–13386, Oct. 2002.

[60] M. E. J. Newman, D. J. Watts, and S. H. S. Strogatz, "Random graph models of social networks," *Proc. Natl. Acad. Sci.*, vol. 99, no. Suppl 1, pp. 2566–2572, Feb. 2002.

[61] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Submodularity and optimality of fusion rules in balanced binary relay trees," in *Proc. 51th IEEE Conf. on Decision and Control*, Maui, HI, USA, Dec. 10–13, 2012, pp. 3802–3807.

[62] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5099–5106, Nov. 2006.

[63] W. Li and H. Dai, "Distributed detection in large-scale sensor networks with correlated sensor observations," in *Proc. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep. 2005.

[64] H. Chen, B. Chen, and P. K. Varshney, "A new framework for distributed detection with conditionally dependent observations," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1409–1419, Mar. 2012.

[65] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Detection performance of balanced binary relay trees with node and link failures," *IEEE Trans. Signal Process.*, 2013, to be published.
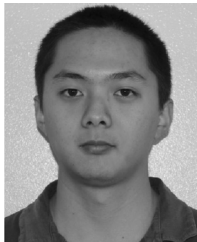
Zhenliang Zhang (S'09) received the B.S. (physics) degree from Special Class for Gifted Young, University of Science and Technology of China, Hefei, China, in 2008.

He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, Colorado State University, CO. His research interests include distributed signal processing, complex network, and optimization.

Edwin K. P. Chong (F'04) received the B.E. degree with first class honors from the University of Adelaide, South Australia, in 1987; and the M.A. and Ph.D. degrees in 1989 and 1991, respectively, both from Princeton University, where he held an IBM Fellowship. He joined the School of Electrical and Computer Engineering at Purdue University in 1991, where he was named a University Faculty Scholar in 1999. Since August 2001, he has been a Professor of Electrical and Computer Engineering and Professor of Mathematics at Colorado State University. His current research interests span the areas of stochastic modeling and control, optimization methods, and communication and sensor networks. He coauthored the best-selling book, *An Introduction to Optimization* (4th Edition, Wiley-Interscience, 2013). He received the NSF CAREER Award in 1995 and the ASEE Frederick Emmons Terman Award in 1998. He was a co-recipient of the 2004 Best Paper Award for a paper in the journal *Computer Networks*. In

2010, he received the IEEE Control Systems Society Distinguished Member Award.

Prof. Chong was the founding chairman of the IEEE Control Systems Society Technical Committee on Discrete Event Systems, and served as an IEEE Control Systems Society Distinguished Lecturer. He is currently a Senior Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and also serves on the editorial boards of *Computer Networks* and the *Journal of Control Science and Engineering*. He has served as a member of the IEEE Control Systems Society Board of Governors and is currently Vice President for Financial Activities. He has also served on the organizing committees of several international conferences. He has been on the program committees for the IEEE Conference on Decision and Control, the American Control Conference, the IEEE International Symposium on Intelligent Control, IEEE Symposium on Computers and Communications, and the IEEE Global Telecommunications Conference. He has also served in the executive committees for the IEEE Conference on Decision and Control, the American Control Conference, the IEEE Annual Computer Communications Workshop, the International Conference on Industrial Electronics, Technology & Automation, and the IEEE International Conference on Communications. He was the Conference (General) Chair for the Conference on Modeling and Design of Wireless Networks, part of SPIE ITCom 2001. He was the General Chair for the 2011 Joint 50th IEEE Conference on Decision and Control and European Control Conference.

Ali Pezeshki (S'95–M'05) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Tehran, Tehran, Iran, in 1999 and 2001, respectively. He received his Ph.D. degree in electrical engineering at Colorado State University in 2004. In 2005, he was a postdoctoral research associate with the Electrical and Computer Engineering Department at Colorado State University. From January 2006 to August 2008, he was a postdoctoral research associate with The Program in Applied and Computational Mathematics at Princeton University.

Since August 2008, he has been an assistant professor with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO. His research interests are in statistical signal processing and coding theory and their applications to active/passive sensing.

William Moran (M'95) received the B.S. (Hons.) degree in mathematics from the University of Birmingham in 1965 and the Ph.D. degree in mathematics from the University of Sheffield, London, U.K., in 1968. He is currently a Professor of electrical engineering with the University of Melbourne, Australia, where he is the Research Director of Defence Science Institute and Technical Director of Melbourne Systems Laboratory. Previously, he was a Professor of mathematics with the University of Adelaide and Flinders University. He also serves as a Consultant to the Australian Department of Defence through the Defence Science and Technology Organization. His research interests are in signal processing, particularly with radar applications, waveform design and radar theory, and sensor management. He also works in various areas of mathematics, including harmonic analysis and number theory and has published widely in these areas.

Stephen D. Howard (M'10) graduated in 1982 from La Trobe University, Melbourne, Australia. He received his M.S. and Ph.D. degrees in mathematics from La Trobe University in 1984 and 1990, respectively. He joined the Australian Defence Science and Technology Organisation (DSTO) in 1991, where he has been involved in research in the area of electronic surveillance and radar systems. He has led the DSTO research effort into the development of algorithms in all areas of electronic surveillance, including radar pulse train deinterleaving, precision radar parameter estimation and tracking, estimation of radar intrapulse modulation, and advanced geolocation techniques. Since 2003, he has led the DSTO long-range research program in radar resource management and waveform design. He is currently the recipient of a three year DSTO research fellowship.