Multivariate Belief Formation for Distributed Multitask Learning

Malek Khammassi
EPFL
Lausanne, Switzerland
malek.khammassi@epfl.ch

Vincenzo Matta
University of Salerno
Fisciano, Italy
vmatta@unisa.it

Ali H. Sayed
EPFL
Lausanne, Switzerland
ali.sayed@epfl.ch

***Abstract*—In traditional social learning, a network of agents wish to learn a common truth or hypothesis. However, in many real-world applications, different agents may observe data related to different hypotheses, resulting in clusters or communities within the network. This setting gives rise to a multitask decision-making (MDM) problem. Existing MDM approaches primarily modify the network topology to reflect the community structures, limiting each agent's interactions to its local cluster. While effective, this philosophy imposes strict neighborhood-based identifiability conditions on the agents. This work shows that an alternative approach to solve the decision-making problem with almost-sure guarantees goes in the opposite direction. Instead of becoming self-interested, each agent should be curious about all the truths co-existing in the network. We prove that when the agents know the number of clusters, they almost surely learn the true states of all agents.**

***Index Terms*—multitask learning, opinion formation, social learning, multitask decision-making.**

I. Introduction and Motivation

Social learning is a distributed inference process in which the agents in a network form beliefs about an unknown state by iteratively combining their private observations with the information received from their neighbors. Inspired by human and animal learning behavior, this framework leverages local interactions to achieve global knowledge without requiring a centralized coordinator, even when local observations are noisy or limited.

Several works have examined the convergence properties of social learning and established conditions under which the agents successfully learn the true state [1]–[10]. While these approaches have been successful in numerous applications, they predominantly focus on scenarios where all agents have the same underlying true state. In many real-world settings, however, different agents may observe data related to different hypotheses. This heterogeneity might correspond to different decision-making tasks or different data, driven by the underlying structure of the network. Applications include distributed sensor networks, where different regions have distinct environmental states, and personalized recommendation systems, where user communities may have heterogeneous preferences that must be inferred simultaneously. In such cases, the assumption of a single true state across the network is no longer valid. Traditional social learning methods, designed to achieve consensus, lead the entire network toward a global agreement rather than enabling agents to identify their own underlying true states [11].

This challenge gives rise to the multitask decision-making (MDM) problem, in which the agents are organized into clusters, each associated with a different true state. This problem has been tackled before in the social learning literature. In [12], social learning is studied in community-structured graphs, where stochastic block models define topological communities based on connectivity. It is assumed that agents within a community interact more frequently and share the same true hypothesis, while cross-community interactions remain limited. The authors show that adaptive social learning [13] enables such networks to learn the truth on average. However, this result holds only if topological communities—characterized by high connectivity—align with communities of shared true states. When the inter-community connectivity is high, this assumption fails, leading to incorrect learning outcomes. In real-world networks, agents' true states do not necessarily adhere to predefined topological clusters. For example, in social opinion networks, individuals may be grouped by geography or profession but hold diverse political beliefs, making it difficult to structure their truths into topological communities. In contrast to [12], our work considers strongly connected networks [14] and defines communities solely based on the agents' true states. Thus, agents belong to the same community if they share the same underlying truth, regardless of their position or connectivity within the network. Furthermore, the agents are not assumed to have prior knowledge of these communities.

In the context of strongly connected networks, the work [15] addresses heterogeneous truths in social learning, without imposing topological communities. The algorithm therein encourages the agents to become self-interested by modifying the network topology so that agents only interact with neighbors that share the same true state. The authors characterize the asymptotic behavior of the agents' beliefs and identify conditions under which all agents can correctly learn their true hypotheses. While this approach prevents the agents from being misled by incorrect information sources, it imposes strict local identifiability requirements, limiting its applicability in real-world settings. In particular, the work in [15] assumes that whenever two neighboring agents have different true states, their indistinguishable hypothesis sets must be disjoint,

ensuring that no agent confuses its true hypothesis with that of a neighbor. However, in many practical scenarios, local observations are too noisy for neighboring agents to reliably distinguish their true states. For instance, in industrial IoT fault detection, machines can classify their equipment's health states using sensor data such as vibration, pressure, and temperature. Due to proximity, load, or configuration, neighboring machines with different underlying faults may record nearly identical readings, making their states indistinguishable.

Our approach to solving the MDM problem takes the opposite direction of [15]. Rather than promote self-interested behavior, we allow the agents to become curious and interested in learning not only their own truths but also the truths of all other agents in the network. Reformulating the problem in this manner provides the agents with a common objective: to learn the true states across the entire network. This approach transforms the MDM problem into a collaborative process that facilitates consensus while preserving the distinct truths of the different agents. In this way, by facilitating information flow across the entire network, our approach bypasses strict neighborhood-based identifiability conditions, allowing neighboring agents to have common indistinguishable hypotheses. We establish the optimality of our approach by proving that it enables each agent to almost surely learn the true states of all agents. Consequently, agents benefit not only from cooperation with those in the same cluster but also from insights gained through interactions with agents from other clusters.

II. Problem Formulation

In a traditional social learning setting, a set of K cooperating agents, each collecting streaming observations about a common phenomenon, learn the hypothesis that best describes the phenomenon. This hypothesis, referred to as the *true state* or the *true hypothesis*, is chosen from a set of plausible hypotheses $\Theta = \{\theta_1, \ldots, \theta_H\}$. In the heterogeneous setting of this work, each agent can have a different true hypothesis. Specifically, we assume that there are C distinct true hypotheses distributed among the agents, where $1 \leq C < H$. Thus, we rule out the case where all hypotheses are simultaneously true across the network. We define the *global true state/hypothesis* $s^\star$ as the collection of all true states across the agents, namely $s^\star = \big(s^\star_{(1)}, \ldots, s^\star_{(K)}\big)$, where each component $s^\star_{(k)} \in \Theta$ represents the true hypothesis of the agent k. Obviously, some states can appear repeated within $s^\star$. For this reason, and since we have exactly C distinct true hypotheses distributed among the agents, we define the *global set of plausible hypotheses* as

$$\mathcal{S}_C(\Theta) = \Big\{ s = \big(s_{(1)}, \ldots, s_{(K)}\big) \in \Theta^K \text{ such that } s \text{ has } C \text{ distinct entries} \Big\}. \tag{1}$$

Therefore, instead of each agent k focusing solely on identifying its true hypothesis from the set Θ, it will focus on learning the global true state $s^\star$ from within the set $\mathcal{S}_C(\Theta)$.

At each time step $i \geq 1$, each agent $k \in \{1, \ldots, K\}$ receives an observation $\boldsymbol{x}_{k,i}$ from its observation space $\mathcal{X}_k$. Each agent k is equipped with a set of likelihood models, $\{L_k(\cdot|\theta)\}_{\theta\in\Theta}$, which are known only to that agent. The observations of agent k are governed by the k-th component of the global true state $s^\star$. In other words, for each agent k, the observation $\boldsymbol{x}_{k,i}$ is drawn from the distribution $L_k(\cdot|s^\star_{(k)})$.

In traditional social learning, each agent k starts with an initial belief vector representing a probability mass function over the set of plausible hypotheses Θ and subsequently updates it based on its local observations and the belief vectors of its neighbors. To adapt this approach to the heterogeneous setting, we propose that each agent k begins with an initial belief vector $\boldsymbol{\mu}_{k,0}$, which represents a probability distribution over the global set of hypotheses $\mathcal{S}_C(\Theta)$ so that the dimension of $\boldsymbol{\mu}_{k,0}$ is the cardinality of $\mathcal{S}_C(\Theta)$. At each time step i, upon receiving a new observation $\boldsymbol{x}_{k,i}$, the agent uses its likelihood models $\{L_k(\cdot|\theta)\}_{\theta\in\Theta}$ to perform a *local Bayesian update*. This step integrates the new observation into the previous belief vector $\boldsymbol{\mu}_{k,i-1}$ to produce an *intermediate* belief vector $\boldsymbol{\psi}_{k,i}$. Following the local update, each agent k performs a *combination step* by aggregating the intermediate beliefs received from its neighbors. This aggregation is carried out using a weighted geometric or arithmetic averaging rule, resulting in the *private* belief vector $\boldsymbol{\mu}_{k,i}$.

The social learning algorithm in this heterogeneous setting thus evolves iteratively over time according to the following update rule for $s \in \mathcal{S}_C(\Theta)$, $k \in \{1, \ldots, K\}$, and $i \geq 1$:

$$\boldsymbol{\psi}_{k,i}(s) \propto L_k\left(\boldsymbol{x}_{k,i} \mid s_{(k)}\right) \boldsymbol{\mu}_{k,i-1}(s), \tag{2}$$

$$\boldsymbol{\mu}_{k,i}(s) \propto \prod_{\ell\in\mathcal{N}_k} \left[\boldsymbol{\psi}_{\ell,i}(s)\right]^{a_{\ell k}}, \tag{3}$$

where the proportionality symbol $\propto$ indicates that the entries of $\boldsymbol{\mu}_{k,i}$ and $\boldsymbol{\psi}_{k,i}$ are normalized to add up to 1. The quantity $a_{\ell k}$ is a nonnegative weight assigned by agent k to the information received from neighbor ℓ satisfying the following conditions:

$$0 \leq a_{\ell k} \leq 1, \quad \sum_{\ell=1}^{K} a_{\ell k} = 1, \quad a_{\ell k} = 0 \text{ for } \ell \notin \mathcal{N}_k, \tag{4}$$

where $\mathcal{N}_k$ denotes the neighborhood of agent k (which includes agent k as well). The collection of the weights $a_{\ell k}$ forms the combination matrix A of the graph connecting the agents.

Although we have reformulated the MDM problem to align with the traditional social learning framework, the algorithm described in (2)-(3) operates over belief vectors of significantly higher dimensionality, specifically, $|\mathcal{S}_C(\Theta)|$ instead of $|\Theta| = H$. As a result, each agent's belief has size

$$\begin{aligned} |\mathcal{S}_C(\Theta)| &= \binom{H}{C} \times \#\text{surjective mappings from } \{1, \ldots, K\} \\ &\qquad \text{to a set of } C \text{ distinct elements.} \\ &= \binom{H}{C}\, S(K, C)\, C! \end{aligned} \tag{5}$$

where $S(K, C)$ is the Stirling number of the second kind [16], which counts the number of ways to partition a set of K

objects into C nonempty subsets. Then, the multiplication by $C!$ accounts for the number of ways we can label the C nonempty subsets, which is the number of permutations among C elements.

We see that the size of the global set of plausible hypotheses grows exponentially with K, H, and C. However, for relatively small networks with a constrained set of plausible hypotheses, the computational requirements remain manageable with modern computing capabilities. Consider the example of a network with $K = 10$ agents, $H = 3$ possible hypothesis states, and $C = 2$ observation classes. Then, $|\mathcal{S}_C(\Theta)| = 3066$. Each entry of the belief vector is stored as a 64-bit double-precision floating-point number, requiring 8 bytes per entry. Consequently, storing or transmitting a complete belief vector requires approximately 24.52 KB. This storage requirement is negligible for contemporary devices, which can efficiently manage gigabytes even with consumer-grade hardware. From a communication perspective, the communication overhead for an agent k is $\mathcal{N}_k \times 24.52$ KB. For a network of 10 agents, the communication overhead per agent can be upperbounded by $10 \times 24.52 = 245.2$ KB. Given that modern network infrastructure can handle data transfers on the order of megabytes per second, this communication overhead is manageable.

III. Main Results

In this section, we characterize the learning behavior of the social learning algorithm described in (2)-(3).

Assumption 1 (Primitive matrix) *The combination matrix A is assumed to be primitive [17].* ■

A sufficient condition for a primitive combination matrix is the existence of bidirectional paths with non-zero weights between any pair of distinct nodes, along with at least one self-loop, indicating that there exists at least one agent k for which $a_{kk} > 0$.

Under Assumption 1, the combination matrix A is irreducible [17]. By the Perron-Frobenius theorem [17], A has a spectral radius equal to 1 and a single eigenvalue at 1, associated with the Perron vector π, which is scaled to have all positive entries summing to 1, namely,

$$A\pi = \pi, \quad \sum_{k=1}^{K} \pi_k = 1, \quad \pi_k > 0 \text{ for } k = 1, 2, \ldots, K. \tag{6}$$

We introduce the following assumption on the statistical model for the observations.

Assumption 2 (Statistical model) *Let $\boldsymbol{x}_i \triangleq \{\boldsymbol{x}_{k,i}\}_{k=1}^{K}$ collect all observations from the agents at time i, and let $s = \big(s_{(1)}, \ldots, s_{(K)}\big) \in \mathcal{S}_C(\Theta)$. The joint likelihood at time i satisfies*

$$L(\boldsymbol{x}_i|s) = \prod_{k=1}^{K} L_k(\boldsymbol{x}_{k,i}|s_{(k)}). \tag{7}$$

■

Furthermore, we introduce the following assumption on the intial beliefs of the agents.

Assumption 3 (Positive initial beliefs) *The initial belief vectors of all agents are positive, i.e., $\mu_{k,0}(s) > 0$ for each agent $k \in \{1, \ldots, K\}$ and all $s \in \mathcal{S}_C(\Theta)$.* ■

We also introduce, for $k \in \{1, \ldots, K\}$ and $\theta \in \Theta$ with $\theta \neq s^\star_{(k)}$, the Kullback-Leibler (KL) divergence between $L_k(.|\theta)$ and $L_k(.|s^\star_{(k)})$,

$$D_k(s^\star_{(k)}, \theta) \triangleq \mathbb{E}_{s^\star_{(k)}} \left[\log \frac{L_k(\boldsymbol{x}|s^\star_{(k)})}{L_k(\boldsymbol{x}|\theta)} \right], \tag{8}$$

where the subscript on the expectation operator means that the expectation is computed under $L_k(\cdot|s^\star_{(k)})$.

Theorem 1 (Truth learning) *We construct a subset of agents $\mathcal{R} \subset \{1, \ldots, K\}$ such that for every distinct θ that appears in $s^\star$, there exists a unique agent in $\mathcal{R}$ that has θ as its true hypothesis:*

$$\forall \theta \in \mathrm{Set}(s^\star), \exists!\, \ell \in \mathcal{R} : s^\star_{(\ell)} = \theta.$$

Under Assumptions 1-3, and if $\mathcal{R}$ satisfies the following two conditions:

1) *Every agent in $\mathcal{R}$ can distinguish its true hypothesis from the rest of the plausible hypotheses:*

$$\forall k \in \mathcal{R} : D_k(s^\star_{(k)}, \theta) > 0, \, \forall \theta \neq s^\star_{(k)}. \tag{9}$$

2) *Every agent outside of $\mathcal{R}$ can distinguish its true hypothesis from every other hypothesis that appears in $s^\star$:*

$$\forall k \notin \mathcal{R} : D_k(s^\star_{(k)}, \theta) > 0, \, \forall \theta \in \mathrm{Set}(s^\star), \, \theta \neq s^\star_{(k)},$$

where $\mathrm{Set}\big(.\big)$ denotes the set of distinct elements in its vector argument, then each agent $k \in \{1, \ldots, K\}$ learns the truth almost surely:

$$\lim_{i \to \infty} \boldsymbol{\mu}_{k,i}(s^\star) = 1 \quad a.s. \tag{10}$$

■

Proof: Due to space limitations, we outline a sketch of the proof. Algorithm (2)–(3) represents traditional social learning over the global set of plausible hypotheses $S_C(\Theta)$ instead of the typical set of plausible hypotheses Θ. Thus, we follow similar steps to those for traditional social learning [11] to show that under Assumptions 1-3, for all $k \in \{1, \ldots, K\}$, for all $s \in \mathcal{S}_C(\Theta)$, $s \neq s^\star$

$$\lim_{i \to \infty} \frac{1}{i} \log \frac{\boldsymbol{\mu}_{k,i}(s^\star)}{\boldsymbol{\mu}_{k,i}(s)} = \sum_{k=1}^{K} \pi_k D_k\big(s^\star_{(k)}, s_{(k)}\big) \quad a.s. \tag{11}$$

In order to prove (10), we need to show that for all agents k and for all $s \neq s^\star$

$$\sum_{k=1}^{K} \pi_k D_k\big(s^\star_{(k)}, s_{(k)}\big) > 0. \tag{12}$$

In what follows, we show that conditions 1)-2) of Theorem 1 imply (12). Let $s \neq s^\star$:

If there exists an agent $\ell \in \mathcal{R}$ such that $s^\star_{(\ell)} \neq s_{(\ell)}$, then in view of condition 1), $D_\ell\big(s^\star_{(\ell)}, s_{(\ell)}\big) > 0$. Therefore, (12) holds.

If for all agents $k \in \mathcal{R}$, $s^\star_{(k)} = s_{(k)}$, we first observe that $|\mathcal{R}| = C$ and suppose without loss of generality that $\mathcal{R} = \{1, \ldots, C\}$. Thus,

$$\sum_{k=1}^{K} \pi_k D_k\big(s^\star_{(k)}, s_{(k)}\big) = \sum_{k=C+1}^{K} \pi_k D_k\big(s^\star_{(k)}, s_{(k)}\big). \tag{13}$$

There exists at least one agent $\ell' \in \{C+1, \ldots, K\}$ such that $s^\star_{(\ell')} \neq s_{(\ell')}$ because otherwise s would be equal to $s^\star$. Now, we will show that $s_{(\ell')} \in \mathrm{Set}(s^\star)$. Suppose, for the sake of contradiction, that $s_{(\ell')} \notin \mathrm{Set}(s^\star)$. Since the number of distinct hypothesis that appear in $s^\star$ is C and s and $s^\star$ have the same hypothesis for agents $1, \ldots, C$, then the number of distinct hypotheses in s must be greater than or equal to C. Now, since $s_{(\ell')}$ does not appear in $s^\star$, the number of distinct hypotheses in s becomes greater than or equal to $C+1$. This implies that $s \notin \mathcal{S}_C(\Theta)$, which is a contradiction. Therefore, we have $\ell' \notin \mathcal{R}$ and $s_{(\ell')} \in \mathrm{Set}(s^\star)$, which in view of condition 2) implies that $D_k\big(s^\star_{(k)}, s_{(k)}\big) > 0$. Thus, (12) follows from (13). ■

Theorem 1 presents conditions for truth learning. While Assumptions 1-3 are standard in social learning, conditions 1)-2) are specific to the heterogeneous setting of this work. The subset of agents $\mathcal{R}$ is constructed such that for each distinct hypothesis θ in $s^\star$, there is exactly one agent $\ell \in \mathcal{R}$ with $s^\star_{(\ell)} = \theta$. Thus, $\mathcal{R}$ can be viewed as a set of representative agents, where each cluster is uniquely represented by one agent. We refer to $\mathcal{R}$ as the representative set. Condition 1) requires that each agent in $\mathcal{R}$ has local identifiability, meaning they can distinguish their true hypothesis from all others in Θ using only their local observations. This assumption ensures that even with heterogeneous states, a small subset of well-defined agents (i.e., $\mathcal{R}$) can guide the network toward accurate inference. Condition 2) requires agents outside $\mathcal{R}$ to only distinguish the hypotheses appearing in $s^\star$ from their true hypothesis. In other words, they are not required to differentiate any hypothesis that does not appear in $s^\star$ from their true hypotheses.

In summary, the two conditions in Theorem 1 ensure the existence of a representative set $\mathcal{R}$, where each agent in $\mathcal{R}$ has local identifiability, while agents outside $\mathcal{R}$ only need to distinguish hypotheses present in $s^\star$. Consequently, any agent k outside $\mathcal{R}$ can confuse its true hypothesis with any $\theta \in \Theta \setminus \mathrm{Set}(s^\star)$ $\big($i.e., $D_k\big(s^\star_{(k)}, \theta\big) = 0\ \big)$ without affecting the algorithm's ability to guarantee truth learning. This property reflects a form of partial identifiability, where some agents cannot distinguish certain hypotheses from their true hypotheses using only local observations.

IV. Illustrative Examples

We consider a strongly connected network of $K = 8$ agents, as shown in Fig. 1, where agents 1, 2, and 8 have self-loops that are omitted for clarity. The combination matrix A is constructed using the uniform-averaging rule [14], [17], resulting in a left-stochastic matrix that satisfies Assumption 1. We consider $\Theta = \{0, 0.5, 1, 1.5\}$ and we set $C = 2$. The global true state is given by $s^\star = (1, 0.5, 0.5, 0.5, 1, 1, 1, 1)$, where we recall that the k-th entry corresponds to the true state of agent k. In Fig. 1, agents with true state 0.5 are marked in black, while those with true state 1 are marked in orange.

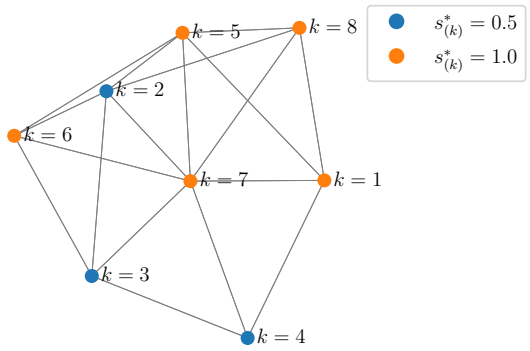


Fig. 1: Strongly connected network with $K = 8$ agents, where k denotes the agent index and $s^\star_{(k)}$ denotes its true hypothesis.

The likelihood models of the agents belong to a family of Gaussian distributions with mean $\theta \in \Theta$ and standard deviation 1. We also assume that certain agents are unable to locally distinguish some hypotheses from their true states, as presented in the identifiability setup in Table I. The configuration in Table I satisfies conditions 1)–2) of Theorem 1. Specifically, the set of representative agents in this example is $\mathcal{R} = \{1, 2\}$, since these agents uniquely represent each true hypothesis present in $s^\star$ (i.e., each cluster). Consequently, agents 1 and 2 have local identifiability. In contrast, the remaining agents are unable to distinguish their respective true states from the hypotheses 0 and 1.5.

In this setting, the truth learning conditions in [15] fail, leading to incorrect outcomes. Specifically, although agents 7 and 3 are neighbors with different true states, neither can distinguish 1.5 and 0 from their true hypotheses, violating condition (30) in [15]. Our proposed approach overcomes this limitation by eliminating neighborhood-based identifiability constraints, enabling truth learning even in such cases. This comes at the cost of requiring agents 1 and 2 to have local identifiability, which guides the whole network into truth learning.

To illustrate the results of Theorem 1, we initialize the beliefs of the multitask social learning algorithm in (2)–(3) uniformly and run it for 1000 iterations. For comparison, we also run the traditional social learning algorithm under the same setting [11]. In Fig. 2, we plot the belief evolution over time for agent 7 for both algorithms.

As shown in Fig. 2a, agent 7 in the traditional social learning algorithm converges to a belief vector that is maximized at $\theta = 1$, despite its true hypothesis being $s^\star_{(7)} = 0.5$. This outcome demonstrates how the consensus-driven nature of traditional

TABLE I: Identifiability setup for the agents.

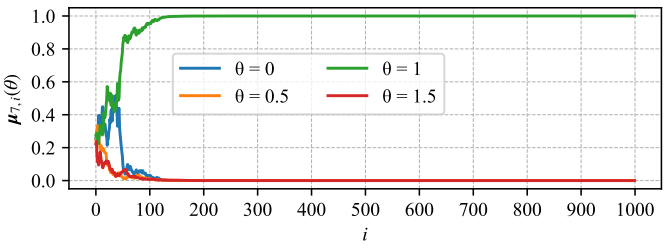
Agent k	**Likelihood model:** $L_k(\cdot\|\theta)$			
	$\theta = 0$	$\theta = 0.5$	$\theta = 1$	$\theta = 1.5$
1, 2	f_0	$f_{0.5}$	f_1	$f_{1.5}$
3, 4	$f_{0.5}$	$f_{0.5}$	f_1	$f_{0.5}$
5, 6, 7, 8	f_1	$f_{0.5}$	f_1	f_1

social learning causes all agents to incorrectly converge to a single hypothesis (i.e., $\theta = 1$) which corresponds to the true state of the majority of agents in this scenario.

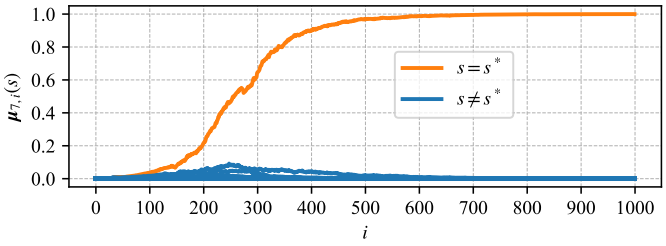
On the other hand, agent 7 in multitask social learning, as shown in Fig. 2b, converges to the hypothesis $s = (1, 0.5, 0.5, 0.5, 1, 1, 1, 1) = s^\star$. This not only gives agent 7 access to its true state 1 by simply reading the 7-th component of s but also gives it access to the true states of the remaining of agents by simply considering the hypothesis corresponding to their index in s.

For a clearer understanding of the network behavior, in Fig. 3 we show the network after convergence for both traditional and multitask social learning. On the left, we depict the network after convergence of the multitask social learning algorithm, where the color of each node k corresponds to the k-th component of the agent's belief at iteration 1000. On the right, we show the network after convergence for the traditional social learning algorithm, where each node's color represents the hypothesis that maximizes the agent's belief at iteration 1000. In both plots, the inferred hypothesis of agent k is denoted as $\hat{s}^\star_{(k)}$.

As shown in Fig. 2, the traditional social learning algorithm drives the entire network toward a global consensus, leading each agent to incorrectly believe that its true hypothesis is $\theta = 1$. In contrast, when compared with Fig.1, we observe that the multitask social learning algorithm enables each agent to correctly identify its true hypothesis.



(a) Belief evolution over time of agent 7 for traditional social learning.



(b) Belief evolution over time of agent 7 for multitask social learning.

Fig. 2: Comparison between traditional and multitask social learning.

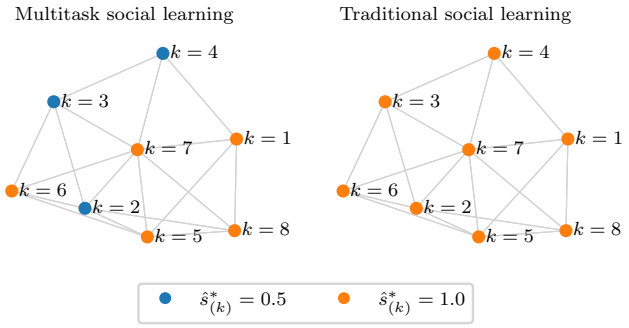


Fig. 3: Comparison between multitask and traditional social learning for the whole network.

References

[1] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dyn. Games Appl.*, vol. 1, no. 1, pp. 3–49, 2011.

[2] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games Econ. Behav.*, vol. 76, no. 1, pp. 210–225, Sep 2012.

[3] X. Zhao and A. H. Sayed, "Learning over social networks via diffusion adaptation," in Proc. *Asilomar Conf. Signals, Syst. and Comput.*, 2012, pp. 709–713.

[4] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi, "Information heterogeneity and the speed of learning in social networks," *J. Econ. Theory*, vol. 148, no. 3, pp. 579–608, 2013.

[5] H. Salami, B. Ying, and A. H. Sayed, "Social learning over weakly connected graphs," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 2, pp. 222–238, Feb. 2017.

[6] A. Nedić, A. Olshevsky, and C. A. Uribe, "Fast convergence rates for distributed non-Bayesian learning," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 5538–5553, Mar. 2017.

[7] M. Pooya, T.-S. Alireza, and J. Ali, "A theory of non-Bayesian social learning," *Econometrica*, vol. 86, no. 2, pp. 445–490, Mar. 2018.

[8] A. Lalitha, T. Javidi, and A. D. Sarwate, "Social learning and distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6161–6179, May 2018.

[9] P. Molavi, K. R. Rad, A. Tahbaz-Salehi, and A. Jadbabaie, "On consensus and exponentially fast social learning," in *Amer. Control Conf. (ACC)*, 2012, pp. 2165–2170.

[10] V. Krishnamurthy and H. V. Poor, "Social learning and Bayesian games in multiagent signal processing: How do local and global decision makers interact?" *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 43–57, May 2013.

[11] V. Matta, V. Bordignon, and A. H. Sayed, *Social Learning: Opinion Formation and Decision-Making over Graphs*. NOW Book Series on Information and Learning Sciences, 2025.

[12] V. Shumovskaia, M. Kayaalp, and A. H. Sayed, "Social learning in community structured graphs," *IEEE Trans. Signal Process.*, vol. 72, pp. 2812–2826, 2024.

[13] V. Bordignon, V. Matta, and A. H. Sayed, "Adaptive social learning," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 6053–6081, Jul. 2021.

[14] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Found. Trends Mach. Learn.*, vol. 7, no. 4–5, p. 311–801, jul 2014.

[15] K. Ntemos, V. Bordignon, S. Vlaski, and A. H. Sayed, "Self-aware social learning over graphs," *IEEE Trans. Inf. Theory*, vol. 69, no. 8, pp. 5299–5317, 2023.

[16] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*. Addison–Wesley, 1988.

[17] A. H. Sayed, *Inference and Learning from Data*. Cambridge University Press, 2022.