DISTRIBUTED DECISION-MAKING FOR COMMUNITY STRUCTURED NETWORKS

Valentina Shumovskaia, Mert Kayaalp, and Ali H. Sayed

École Polytechnique Fédérale de Lausanne (EPFL)

ABSTRACT

Traditional social learning frameworks consider environments with a homogeneous state where each agent receives observations conditioned on the same hypothesis. In this work, we study the distributed hypothesis testing problem for graphs with a community structure, assuming that each cluster receives data conditioned on some different true state. This situation arises in many scenarios, such as when sensors are spatially distributed, or when individuals in a social network have differing views or opinions. We show that the adaptive social learning strategy is not only superior in nonstationary environments, but also allows each cluster to discover its own truth.

Index Terms— Social learning, hypothesis testing, diffusion strategies, adaptive learning, multitask learning, personalized learning, decision-making.

1. INTRODUCTION AND RELATED WORK

The social learning framework [1–16] is a popular non-Bayesian approach for solving distributed hypothesis testing (or decision-making) problems over graphs. In these problems, the objective is to learn and track an unknown true state of nature (or hypothesis) from streaming data. Such formulations can be used, for example, to model opinion formation over social networks where agents exchange beliefs on a topic of interest to arrive at a collaborative conclusion. For instance, the work [17] considers a special application of social learning to identify influential users over Twitter.

In traditional social learning algorithms, at each iteration, agents update their confidence level about each possible hypothesis. They do so by updating a probability mass function called the *belief* vector. One of the remarkable variations of social learning is the adaptive social learning strategy (ASL) [7]. ASL infuses these methodologies with adaptation and tracking abilities and allows agents to track drifts in the underlying hypothesis. This method resolves the stubbornness issue that plagues traditional solutions where agents resist changing their opinions.

The non-Bayesian social learning approach is an effective alternative to finding the fully Bayesian estimate of the unknown hypothesis. This is because the Bayesian solution over graphs is generally NP-hard and requires full knowledge of the graph topology and of the likelihood models [10, 11, 18]. The non-Bayesian approach considered herein, and in the aforementioned works, relies on a fully decentralized implementation, allowing each agent to keep their observations private and to exchange beliefs solely with their immediate neighbors.

All these previous works assume that the agents receive observations arising from the same state of nature (or the same hypothesis). In this work, we relax this assumption and allow agents to receive observations from different models (or hypotheses). We therefore focus on studying the limiting behavior of the decision-making process and examine how the presence of agents with different states affects collaborative learning.

The main challenge is the following. If we assume a multiple hypotheses scenario and apply any of the traditional social learning strategies [1, 2, 4, 9], then it is known that these methods will force all agents to converge to the same common state that best describes the observations. This state will not be necessarily optimal or the true one for the individual agents. In other words, the traditional techniques force consensus while the multiple hypotheses scenario requires the use of a strategy that would allow diversity of models across the agents. In this regard, we show in this work that the adaptation parameter δ, which is used by the adaptive social learning strategy (ASL) to enable tracking, actually plays an additional role. It allows the final belief vector for each agent to become dependent on the *local* subnetwork. As a result, different parts of the graph can now arrive at different conclusions. This means that adaptation helps infuse diversity into the convergence behavior of the network and allows agents to approach different limiting states.

For graphs involving a multiplicity of models, one useful topology is that of a community-structured graph such as the Stochastic Block Model (SBM) [19–24]. It defines each community as a collection of agents that have a large probability of connection with each other, while the probability of connection between communities is small. The SBM has been used in the literature for the analysis of exchanges over social networks such as Twitter [23–26], and for studying opinion polarization [27,28]. This motivates us to study social learning under polar true hypotheses. Assuming that each cluster receives data conditioned on a different true state, we are interested in determining conditions that enable each of the communities to discover their own truth. We will show in this work that, for such block models and under reasonable constraints on the adaptation parameter δ, it is possible for each cluster in the network to discover its true hypothesis (or model) by relying on the adaptive social learning strategy [7].

2. SOCIAL LEARNING MODEL

Consider a collection of agents denoted by $\mathcal{N}$, interacting to form belief vectors that reflect their confidence in each possible hypothesis $\theta \in \Theta$ based on their private streaming observations and on interactions with their immediate neighbors. These interactions are governed by a combination graph $A \in [0,1]^{|\mathcal{N}|\times|\mathcal{N}|}$, where nonzero elements indicate an edge between two nodes. For any two connected agents, the value $a_{\ell k} = [A]_{\ell,k} > 0$ determines the level of trust that agent $k \in \mathcal{N}$ assigns to information arriving from agent $\ell \in \mathcal{N}$. The combination matrix is assumed to be left-stochastic, i.e., the weights on every column add up to one:

$$\sum_{\ell\in\mathcal{N}} a_{\ell k} = 1,\ \forall k \in \mathcal{N}. \tag{1}$$

Emails: {valentina.shumovskaia, mert.kayaalp, ali.sayed}@epfl.ch.
This work was supported in part by grant 205121-184999 from the Swiss National Science Foundation.

Additionally, the combination graph is assumed to be strongly connected, which means that there exists at least one self-loop with a positive weight, i.e., $a_{kk} > 0$ for some k and, moreover, there exists a path with positive weights between any two nodes [29]. It follows from these conditions and the Perron-Frobenius theorem [30, Chapter 8], [29] that the power matrix A^s converges to $u\mathbb{1}^\mathsf{T}$ as $s \to \infty$ at an exponential rate, where u refers to the Perron eigenvector of A, namely,

$$Au = u, \qquad u_\ell > 0, \qquad \sum_{\ell \in \mathcal{N}} u_\ell = 1. \tag{2}$$

At each moment i, agent k receives some random observation $\boldsymbol{\zeta}_{k,i}$ from the environment. In traditional social learning algorithms it is assumed that there exists a single global true state of nature, denoted by $\theta^\star \in \Theta$. The aim of the social learning algorithm then becomes to enable all agents to discover the value of $\theta^\star$ from the streaming observations. In this work, however, we allow the truth to be agent-dependent. That is, the observations of agent k are conditioned on the local model $\theta_k^\star$ and we write $L_k(\boldsymbol{\zeta}_{k,i}|\theta_k^\star)$.

The traditional social learning strategy is described as follows. Each agent $k \in \mathcal{N}$ assigns an initial belief $\boldsymbol{\mu}_{k,0}(\theta) \in [0,1]$ for each state $\theta \in \Theta$ such that the total confidence sums up to one, i.e., $\sum_\theta \boldsymbol{\mu}_{k,0}(\theta) = 1$. In order not to exclude any hypothesis beforehand, we assume that each component of the belief vector $\boldsymbol{\mu}_{k,0}$ is strictly positive. Then, at each iteration i, each agent k receives an observation $\boldsymbol{\zeta}_{k,i}$ and performs a local Bayesian update [2, 4, 9]:

$$\boldsymbol{\psi}_{k,i}(\theta) = \frac{L_k(\boldsymbol{\zeta}_{k,i}|\theta)\boldsymbol{\mu}_{k,i-1}(\theta)}{\sum_{\theta' \in \Theta} L_k(\boldsymbol{\zeta}_{k,i}|\theta')\boldsymbol{\mu}_{k,i-1}(\theta')}, \quad \forall k \in \mathcal{N}. \tag{3}$$

The vector $\boldsymbol{\psi}_{k,i}$ is a probability mass function and we refer to it as the *intermediate (public)* belief. The qualification "public" refers to the fact that this vector is shared among neighbors. For the adaptive social learning (ASL) strategy from [7], the update step (3) is replaced by:

$$\boldsymbol{\psi}_{k,i}(\theta) = \frac{L_k^\delta(\boldsymbol{\zeta}_{k,i}|\theta)\boldsymbol{\mu}_{k,i-1}^{1-\delta}(\theta)}{\sum_{\theta' \in \Theta} L_k^\delta(\boldsymbol{\zeta}_{k,i}|\theta')\boldsymbol{\mu}_{k,i-1}^{1-\delta}(\theta')}, \quad \forall k \in \mathcal{N}. \tag{4}$$

The step-size $\delta \in (0,1)$ infuses into the algorithm the ability to track drifts in the underlying models (hypotheses).

Following (3) or (4), the agents perform geometric averaging of public beliefs of their neighbors [4, 7, 9]:

$$\boldsymbol{\mu}_{k,i}(\theta) = \frac{\prod_{\ell \in \mathcal{N}_k} \boldsymbol{\psi}_{\ell,i}^{a_{\ell k}}(\theta)}{\sum_{\theta' \in \Theta} \prod_{\ell \in \mathcal{N}_k} \boldsymbol{\psi}_{\ell,i}^{a_{\ell k}}(\theta')}, \quad \forall k \in \mathcal{N}. \tag{5}$$

The resulting vector $\boldsymbol{\mu}_{k,i}$ is referred to as the *private* belief.

At every iteration, each agent estimates the true state based on their private belief $\boldsymbol{\mu}_{k,i}$ (it can also use $\boldsymbol{\psi}_{k,i}$):

$$\widehat{\boldsymbol{\theta}}_{k,i} \triangleq \arg\max_{\theta \in \Theta} \boldsymbol{\mu}_{k,i}(\theta). \tag{6}$$

3. TRUTH LEARNING

It is known that under traditional social learning (3) and (5), even if the agents have different local truths, the network will reach consensus on some subset $\Theta^\star$ that minimizes [4]:

$$\min_\theta \sum_{k \in \mathcal{N}} u_k D_{\mathrm{KL}}\big(L_k(\theta_k^\star)||L_k(\theta)\big) \tag{7}$$

Here, the notation D_{KL} denotes the Kullback-Leibler divergence between two distributions:

$$D_{\mathrm{KL}}\big(L_k(\theta)||L_k(\theta')\big) \triangleq \mathbb{E}_{\boldsymbol{\zeta} \sim L_k(\boldsymbol{\zeta}|\theta)} \log \frac{L_k(\boldsymbol{\zeta}|\theta)}{L_k(\boldsymbol{\zeta}|\theta')} \tag{8}$$

The algorithm behaves conservatively and forces the agents to agree on one optimal subset $\Theta^\star$ independently of the graph structure. There is no guarantee that for each individual agent k, its true hypothesis $\theta_k^\star$ is present in the subset $\Theta^\star$.

The main advantage of the adaptive strategy (4)–(5) is its capacity to adapt to changes in the models over time. The hyperparameter $\delta \in (0,1)$ plays an important role: the higher δ is, the more importance is attached to newly received samples. According to [7, Theorem 1], the log-belief ratio of private beliefs converges in distribution to the following random variable:

$$\log \frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta')} \xrightarrow[i \to \infty]{d} \boldsymbol{\rho}_k(\theta, \theta') \tag{9}$$

defined as

$$\boldsymbol{\rho}_k(\theta, \theta') \triangleq \delta \sum_{\ell \in \mathcal{N}} \sum_{i=0}^{\infty} (1-\delta)^t [A^{t+1}]_{\ell k} \log \frac{L_\ell(\boldsymbol{\zeta}_{\ell,i}|\theta)}{L_\ell(\boldsymbol{\zeta}_{\ell,i}|\theta')} \tag{10}$$

Its expected value is given by:

$$\begin{aligned}\mathbb{E}\boldsymbol{\rho}_k(\theta, \theta') = \delta \sum_{\ell \in \mathcal{N}} \sum_{t=0}^{\infty} (1-\delta)^t [A^{t+1}]_{\ell k} \\ \times \Big(D_{\mathrm{KL}}\big(L_\ell(\theta_\ell^\star)||L_\ell(\theta')\big) - D_{\mathrm{KL}}\big(L_\ell(\theta_\ell^\star)||L_\ell(\theta)\big)\Big)\end{aligned} \tag{11}$$

Assuming finiteness of second-order moments for the log-likelihoods, the variance of $\boldsymbol{\rho}_k$ is on the order of δ:

$$\mathrm{Var}(\boldsymbol{\rho}_k(\theta, \theta')) = O(\delta). \tag{12}$$

The result implies that the expectation $\mathbb{E}\boldsymbol{\rho}_k(\theta, \theta')$ determines which hypothesis agent k will prioritize in the steady-state *on average*. As opposed to the case when $\delta \to 0$ studied in [7], there is no almost sure convergence guarantee toward a final solution, and the log beliefs will fluctuate around their mean $\mathbb{E}\boldsymbol{\rho}_k(\theta, \theta')$ with the variance on the order of $O(\delta)$.

These results reveal that each agent k can arrive at its own locally optimal solution because the expression on the right-hand side of (10) depends on k. This is in contrast to traditional social learning where the beliefs of all agents converge to the same zero value with a rate determined by the network divergence. In the ASL, the final inference for each agent depends on the local network, and thus on the observations and the true states of these agents: observe from (11) that each agent gives higher importance to its close neighbors. In particular, the weight $(1-\delta)$ scales the immediate one-hop neighbors (namely, those agents ℓ for which $a_{\ell k}$ is non-zero), while the weight $(1-\delta)^2$ scales the agents from the 2-hop neighborhood, and so on. This way, as the value of δ increases, the influence of further connected nodes diminishes.

This observation suggests that under certain network conditions, such as community structured graphs, and for large enough δ, each agent k should be able to arrive on average to their own truth $\theta_k^\star$. This is because over these graphs there is a higher probability for each agent to be connected to the nodes that share the same underlying truth. We will verify that this is indeed the case.

(a) True states, colors stand for hypotheses (b) Predicted states, $\delta = 0.01$ (c) Predicted states, $\delta = 0.1$.

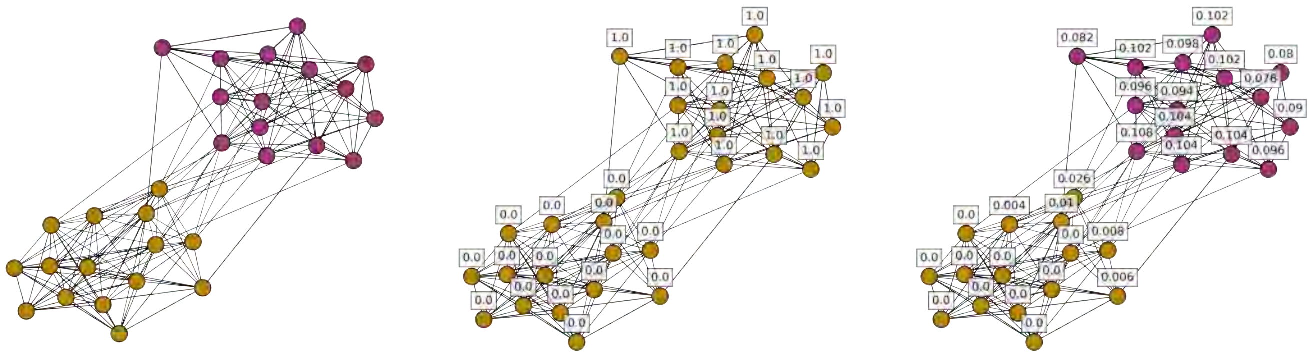

Fig. 1: The algorithm's performance in identifying the true state of each node, using the adaptive social learning strategy. The probabilities of error (shown inside the boxes) $\mathbb{P}(\widehat{\boldsymbol{\theta}}_{k,i} \neq \theta_k^\star)$ are approximated based on 500 iterations.

4. STOCHASTIC BLOCK MODEL

4.1. Combination Matrix

The Stochastic Block Model (SBM) is a generative model to produce graphs with community structure, where each community is a collection of agents that have a large probability of connection with each other, and the probabilities of connection between communities are smaller [19–21]. It is a popular framework for modeling social networks and graphs with polar opinions [31]. Polar opinions reflect the fact that there might be no single truth. For example, over a network of spatially distributed sensors, different sensors might experience different weather conditions (different temperature or precipitation). This motivates us to study social learning under polar true hypotheses. We remark that the majority of works on community-structured graphs focus their analysis on the case of graphs with two communities [19, 20]. While extending the experimental part to more general cases with multiple communities is feasible, theoretical bounds often become intractable. In this work, we will similarly focus on the common scenario with two communities.

We describe next the SBM model. We denote by $\boldsymbol{E}$ the adjacency matrix of the network. The entries of the adjacency matrix are assumed to be drawn from a Bernoulli distribution, $\boldsymbol{E} \sim Bernoulli(P)$, conditioned on the probability matrix $P \in [0,1]^{|\mathcal{N}|\times|\mathcal{N}|}$ (i.e., each entry $\boldsymbol{E}_{\ell,k} \sim Bernoulli(P_{\ell,k})$ is generated independently). We introduce the main idea by considering a model with two communities of sizes n_0 and n_1 such that $n_0 + n_1 = |\mathcal{N}|$. Under this model, the probability matrix P takes the following block form:

$$P \triangleq \left[\begin{array}{c|c} p_0 \mathbb{1}_{n_0}\mathbb{1}_{n_0}^\mathsf{T} & q_0 \mathbb{1}_{n_0}\mathbb{1}_{n_1}^\mathsf{T} \\ \hline q_1 \mathbb{1}_{n_1}\mathbb{1}_{n_0}^\mathsf{T} & p_1 \mathbb{1}_{n_1}\mathbb{1}_{n_1}^\mathsf{T} \end{array}\right] \tag{13}$$

where $\mathbb{1}_n$ is a column vector of ones of size n and we let $q_0,\ q_1 < \min\{p_0, p_1\}$. This form of P allows us to generate graphs with clearly defined communities, as illustrated in Fig. 2a.

Agents in the network will communicate with each other according to a combination protocol that is defined by some matrix A. We assume the combination weights are set using the averaging rule [29]. This way, each column is normalized and agents give equal confidence to their neighbors with each entry equal to $[\boldsymbol{A}]_{\ell,k} = \boldsymbol{E}_{\ell,k}/\sum_\ell \boldsymbol{E}_{\ell,k}$. We can show that for any integer power $t < |\mathcal{N}|$, the moments of $\boldsymbol{A}$ can be approximated[1] by:

$$\mathbb{E}\boldsymbol{A}^t = \overline{A}^t + O\left(\min\{n_0, n_1\}^{-4/3}\right) \tag{14}$$

where $\overline{A}$ is the left-stochastic matrix given by:

$$\overline{A} \triangleq \left[\begin{array}{c|c} \frac{p_0}{p_0 n_0 + q_1 n_1}\mathbb{1}_{n_0}\mathbb{1}_{n_0}^\mathsf{T} & \frac{q_0}{q_0 n_0 + p_1 n_1}\mathbb{1}_{n_0}\mathbb{1}_{n_1}^\mathsf{T} \\ \hline \frac{q_1}{p_0 n_0 + q_1 n_1}\mathbb{1}_{n_1}\mathbb{1}_{n_0}^\mathsf{T} & \frac{p_1}{q_0 n_0 + p_1 n_1}\mathbb{1}_{n_1}\mathbb{1}_{n_1}^\mathsf{T} \end{array}\right] \tag{15}$$

We show one example of such a combination matrix in Fig. 2b, based on the adjacency network from Fig. 2a.

(a) An SBM graph model with two communities. (b) Combination matrix for the SBM graph in part (a).

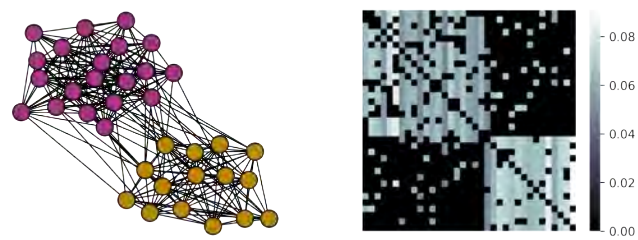

Fig. 2: Network illustration with $n_0 = 20$, $n_1 = 15$, $p_0 = 0.8$, $p_1 = 0.9$, $q_0 = q_1 = 0.1$.

We additionally assume that all agents within the same community receive data arising from the same underlying model (or hypothesis). In our particular two-communities case, we assume the binary hypotheses set $\Theta = \{\theta_0, \theta_1\}$. Thus, we let the first n_0 agents $k \in \mathcal{C}_0 \triangleq \{1, \dots, n_0\}$ to follow hypothesis θ_0, and the remaining agents $k \in \mathcal{C}_1 \triangleq \{n_0 + 1, \dots, |\mathcal{N}|\}$ to follow hypothesis θ_1. For simplicity, we assume that agents in each community (or cluster) have the same level of informativeness measured in terms of the KL divergence between the two models, as defined by the following statement.

[1]The proof is omitted for brevity; proofs can be found in the preprint [32].

Assumption 1 (**Homogeneous likelihoods).** *Within each cluster $\mathcal{C}_i$, agents have the same level of informativeness. For any $k \in \mathcal{C}_0$:*

$$d_0 \triangleq D_{\text{KL}}\big(L_k(\theta_0)\,||L_k(\theta_1)\big) \tag{16}$$

and for any $k \in \mathcal{C}_1$:

$$d_1 \triangleq D_{\text{KL}}\big(L_k(\theta_1)\,||L_k(\theta_0)\big) \tag{17}$$

While this assumption is not strictly necessary, it is introduced for the sake of clarity and for analytical tractability of the results. It holds, for example, when all likelihoods within each community are equal, i.e., agents receive samples from the same or similar sources.

4.2. Truth Learning

By using properties (2) and (14) of the combination matrix $\boldsymbol{A}$ and Assumption 1, we can show that in steady state, the log-ratio of beliefs can be bounded as follows[1]. For agent $k \in \mathcal{C}_0$, we get that:

$$\begin{aligned}\mathbb{E}\boldsymbol{\rho}_k(\theta_0,\theta_1) \geq (1-\delta)\cdot \frac{q_0 n_0 r_0 d_0 - q_1 n_1 r_1 d_1}{q_0 n_0 r_0 + q_1 n_1 r_1} \\ + \delta \cdot \frac{p_0 n_0 d_0 - q_1 n_1 d_1}{r_0} + O(\min\{n_0, n_1\}^{-4/3})\end{aligned} \tag{18}$$

and for agent $k \in \mathcal{C}_1$, the log-ratio is bounded from above:

$$\begin{aligned}\mathbb{E}\boldsymbol{\rho}_k(\theta_0,\theta_1) \leq (1-\delta)\cdot \frac{q_0 n_0 r_0 d_0 - q_1 n_1 r_1 d_1}{q_0 n_0 r_0 + q_1 n_1 r_1} \\ - \delta \cdot \frac{p_0 n_0 d_0 - q_1 n_1 d_1}{r_0} + O(\min\{n_0, n_1\}^{-4/3}).\end{aligned} \tag{19}$$

where expectations are taken w.r.t. $\boldsymbol{\zeta}_{\ell,i}$ and $\boldsymbol{A}$. The desired outcome is a positive sign for (18) and a negative sign for (19), ensuring that each community will arrive at its own truth. Remarkably, the first term in both (18) and (19) is the same. If we let $\delta \to 0$, the whole network will converge to the more prevalent hypothesis, and will therefore behave similarly to traditional social learning. In contrast, a reasonably large δ would allow the less dominant cluster (i.e., the cluster with smaller informativeness level d_i) to drive itself to their own truth. It can be shown that we can find such δ that allows the desired behavior[1].

Theorem 1 (**Log-belief ratios for the SBM).** *If the probabilities between clusters are sufficiently smaller than the probabilities inside the clusters, more specifically:*

$$p_0 n_0 d_0 - q_1 n_1 d_1 > 0, \quad p_1 n_1 d_1 - q_0 n_0 d_0 > 0 \tag{20}$$

Then, there exist a $\delta_0 \in (0,1)$, that for any $\delta > \delta_0$, on average, each cluster converges to its own hypothesis, i.e. both $\lim_{i\to\infty} \mathbb{E}\log\frac{\boldsymbol{\mu}_{k,i}(\theta_0)}{\boldsymbol{\mu}_{k,i}(\theta_1)}$ and $\lim_{i\to\infty} \mathbb{E}\log\frac{\boldsymbol{\psi}_{k,i}(\theta_0)}{\boldsymbol{\psi}_{k,i}(\theta_1)}$ are strictly positive or strictly negative depending on the cluster.

5. COMPUTER SIMULATIONS

We consider the SBM graph shown in Figure 1a with connection probabilities $p_0 = p_1 = 0.8$ and $q_0 = q_1 = 0.1$. Each cluster has 15 agents and all agents have equal likelihood models with Bernoulli distributions: $L_k(\theta_0) = Bernoulli(0.1)$ and $L_\ell(\theta_1) = Bernoulli(0.5)$. Agents of the first block follow hypothesis θ_0 and agents from the second block follow θ_1. The Kullback-Leibler divergence for the first group is smaller since $0.37 = d_0 < d_1 = 0.51$.

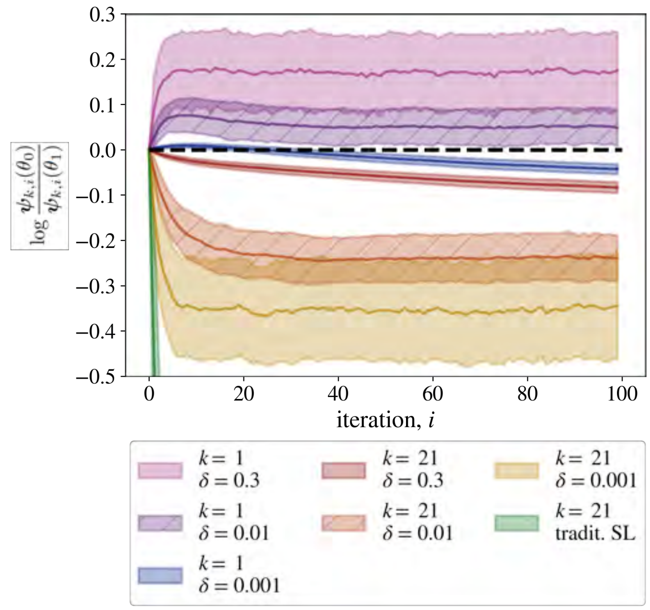

Fig. 3: Evolution of log-belief ratios $\log\frac{\boldsymbol{\psi}_{i,k}(\theta_0)}{\boldsymbol{\psi}_{i,k}(\theta_1)}$ over time with different step-size δ (mean and standard deviations over 500 algorithm runs). Agent 1 belongs to the first cluster and follows θ_0, while agent 21 belongs to the second cluster and follows θ_1.

Therefore, for small $\delta \to 0$, the network is expected to converge to θ_1.

Figure 3 illustrates Theorem 1. We let the network follow the ASL strategy with different step-size parameters δ. We see that when $\delta = 0.01$, the log-ratios of both group are below zero. When $\delta = 0.1$, the log-belief ratio of the first group is able (on average) to stay above the zero threshold, therefore its expectation converges to hypothesis θ_0. And, for larger $\delta = 0.3$, we observe further increase in the gap. However, it is evident that the gap grows together with the variance, and therefore leads to an increased probability of error $\mathbb{P}(\arg\max_\theta \boldsymbol{\psi}_{k,i}(\theta) \neq \theta_k^\star)$ for the second cluster in the steady state. Figure 1 shows probability of errors for different values of δ, that support the same considerations.

6. CONCLUSIONS

This work examines the behavior of social learning algorithms under multiple hypotheses. We show that traditional social learning techniques behave conservatively and the whole network converges to a consensus solution, which is not necessarily optimal at the individual or cluster level. In other words, if the majority of high-informative agents operate with observations following the true hypothesis, malicious or malfunctioning agents (the agents that receive observations from another state) will not be able to drive the network to a wrong conclusion. Adaptive social learning strategies behave similarly when the adaptation hyperparameter $\delta \to 0$. However, for sufficiently large $\delta > 0$, the ASL strategy is the preferred choice for graphs with community structure (such as SBM). In this case, the learning strategy is able to discover the underlying hypotheses of the individuals. The results were illustrated by computer simulations.

7. REFERENCES

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

[2] X. Zhao and A. H. Sayed, "Learning over social networks via diffusion adaptation," in *Proc. Asilomar Conference on Signals, Systems and Computers*, 2012, pp. 709–713.

[3] H. Salami, B. Ying, and A. H. Sayed, "Social learning over weakly connected graphs," *IEEE Trans. Signal and Information Processing over Networks*, vol. 3, no. 2, pp. 222–238, 2017.

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

[5] P. Molavi, A. Tahbaz-Salehi, and A. Jadbabaie, "Foundations of non-Bayesian social learning," *Columbia Business School Research Paper*, no. 15-95, 2017.

[6] ——, "A theory of non-Bayesian social learning," *Econometrica*, vol. 86, no. 2, pp. 445–490, 2018.

[7] V. Bordignon, V. Matta, and A. H. Sayed, "Adaptive social learning," *IEEE Transactions on Information Theory*, vol. 67, no. 9, pp. 6053–6081, 2021.

[8] ——, "Partial information sharing over social learning networks," *IEEE Transactions on Information Theory*, vol. 69, no. 3, pp. 2033–2058, 2023.

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

[10] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *The Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.

[11] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games and Economic Behavior*, vol. 45, no. 2, pp. 329–346, 2003.

[12] Y. İnan, M. Kayaalp, E. Telatar, and A. H. Sayed, "Social learning under randomized collaborations," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 115–120.

[13] J. Z. Hare, C. A. Uribe, L. Kaplan, and A. Jadbabaie, "Non-Bayesian social learning with uncertain models," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4178–4193, 2020.

[14] C. A. Uribe, A. Olshevsky, and A. Nedić, "Nonasymptotic concentration rates in cooperative learning–part i: Variational non-Bayesian social learning," *IEEE Transactions on Control of Network Systems*, vol. 9, no. 3, pp. 1128–1140, 2022.

[15] ——, "Nonasymptotic concentration rates in cooperative learning—part ii: Inference on compact hypothesis sets," *IEEE Transactions on Control of Network Systems*, vol. 9, no. 3, pp. 1141–1153, 2022.

[16] M. Cemri, V. Bordignon, M. Kayaalp, V. Shumovskaia, and A. H. Sayed, "Asynchronous social learning," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[17] V. Shumovskaia, M. Kayaalp, M. Cemri, and A. H. Sayed, "Discovering influencers in opinion formation over social graphs," *IEEE Open Journal of Signal Processing*, vol. 4, pp. 188–207, 2023.

[18] J. Hkazla, A. Jadbabaie, E. Mossel, and M. A. Rahimian, "Bayesian decision making in groups is hard," *Operations Research*, vol. 69, no. 2, pp. 632–654, 2021.

[19] E. Abbe, "Community detection and stochastic block models: recent developments," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6446–6531, 2017.

[20] E. Abbe, A. S. Bandeira, and G. Hall, "Exact recovery in the stochastic block model," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 471–487, 2015.

[21] K. Rohe, S. Chatterjee, and B. Yu, "Spectral clustering and the high-dimensional stochastic blockmodel," *Annals of Statistics*, vol. 39, 07 2010.

[22] C. Lee and D. J. Wilkinson, "A review of stochastic block models and extensions for graph clustering," *Applied Network Science*, vol. 4, no. 1, pp. 1–50, 2019.

[23] A. Goldenberg, A. X. Zheng, S. E. Fienberg, E. M. Airoldi *et al.*, "A survey of statistical network models," *Foundations and Trends® in Machine Learning*, vol. 2, no. 2, pp. 129–233, 2010.

[24] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[25] N. Zalmout and M. Ghanem, "Multidimensional community detection in Twitter," in *Proc. International Conference for Internet Technology and Secured Transactions (ICITST-2013)*, 2013, pp. 83–88.

[26] D. Cherepnalkoski and I. Mozetič, "Retweet networks of the European parliament: Evaluation of the community structure," *Applied Network Science*, vol. 1, pp. 1–20, 2016.

[27] U. Chitra and C. Musco, "Analyzing the impact of filter bubbles on social network polarization," in *Proc. International Conference on Web Search and Data Mining*, 2020, pp. 115–123.

[28] N. Bhalla, A. Lechowicz, and C. Musco, "Local edge dynamics and opinion polarization," in *Proc. Sixteenth ACM International Conference on Web Search and Data Mining*, 2023, pp. 6–14.

[29] A. H. Sayed, *Inference and Learning from Data*. Cambridge University Press, 2022, vols. 1–3.

[30] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, NY, 2013.

[31] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Physical Review E*, vol. 84, no. 6, p. 066106, 2011.

[32] V. Shumovskaia, M. Kayaalp, and A. H. Sayed, "Social learning in community structured graphs," *arXiv:2312.12186*, 2023.