

Received October 26, 2020, accepted November 4, 2020, date of publication November 9, 2020, date of current version November 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3036669

Learning the Truth by Weakly Connected Agents in Social Networks Using Multi-Armed Bandit

OLUSOLA TOLULOPE ODEYOMI¹, (Graduate Student Member, IEEE)

Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS 67260-0083, USA

e-mail: otodeyomi@shockers.wichita.edu

This work was supported by the Petroleum Technology Development Fund in Nigeria.

ABSTRACT This article provides a study into the social network where influential personalities collaborate positively among themselves to learn an underlying truth over time, but may have misled their followers to believe a false information. Most existing work that study leader-follower relationships in a social network model the social network as a graph network, and apply non-Bayesian learning to train the weakly connected agents to learn the truth. Although this approach is popular, it has the limitation of assuming that the truth - otherwise called the true state - is time-invariant. This is not practical in social network, where streams of information are released and updated every second, making the true state arbitrarily time-varying. Thus, this article improves on existing work by introducing online reinforcement learning into the graph theoretic framework. Specifically, multi-armed bandit technique is applied. A multi-armed bandit algorithm is proposed and used to train the weakly connected agents to converge to the most stable state over time. The speed of convergence for these weakly connected agents trained with the proposed algorithm is slower by 66% on average, when compared to the speed of convergence for strongly connected agents trained with the state-of-the-art algorithm. This is because weakly connected agents are difficult to train. However, the speed of convergence of these weakly connected agents can be improved by approximately 50% on average, by fine-tuning the learning rate of the proposed algorithm. The sublinearity of the regret bound for the proposed algorithm is compared to the sublinearity of the regret bound for the state-of-the-art algorithm for strongly connected networks.

INDEX TERMS Weakly connected agents, non-Bayesian learning, diffusion learning, online learning, multi-armed bandit, regret.

I. INTRODUCTION

The social network has grown over the years to become a platform for influential personalities to sell their beliefs to followers within their sphere of influence. Even if the beliefs of these influential personalities are wrong, it is easy for them to access first-hand information and cooperate among themselves to learn the truth over time. However, it is difficult for the misinformed followers to learn the truth by virtue of their inherent weakness to independently find the truth. They are subject to the dominating influence of these influential personalities. Take for instance the prevalent attitude among adolescents, where they form strong loyalty to celebrities of their interest, and inherit the enemies of these celebrities. In reality, most of the celebrity fights on social media are merely for attention seeking. The rival celebrities may resolve

their differences among themselves, but these adolescents endlessly engage in social media battle against themselves. This is a social learning problem. As a result of the exponential growth of social networks, it is imperative to study the behavior of humans in social networks, and to proffer solutions to the menace of negative social influence. There have been lots of interest among researchers to study the spread of negative information in social networks [1]–[4]. A study in [5] discussed the connection between social relationships among humans and health.

Graph theory is commonly used for theoretical research on social networks. Influential personalities or agents form a strongly connected subnetwork, while their followers form a weakly connected subnetwork. Agents in the strongly connected subnetwork can easily communicate among themselves to learn the truth. On the other hand, agents in the weakly connected subnetwork are dominated by agents in the strongly connected subnetwork, whether or not they

The associate editor coordinating the review of this manuscript and approving it for publication was Zeshui Xu¹.

communicate among themselves. At least two common graph theoretic approaches are used to study the interaction process [6], [7] among agents in a graph network. The Bayesian method [6], [8]–[10] is the first approach, where agents rely on some prior beliefs and on Bayes' rule to update their beliefs. The non-Bayesian method [11], is the second approach, where each agent initially obtain an intermediate belief based on an observed private signal, using the Bayes' rule. Then, the agent cooperates with its neighboring agents to update its belief. Agents asymptotically learn the truth - also known as the *underlying true state* - following this approach. The underlying true state is the unknown parameter of interest that the agents seek to uncover. Although, each agent observes a private signal at each time instance, this private signal is only fuzzy about the true state. No agent can directly observe the true state; hence, the agents rely on interactions among themselves to learn this true state over time. Thus, the true state is said to be underlying. The state-of-the-art non-Bayesian learning approaches use *diffusion learning* [12]. The performance of diffusion learning is good in situations where learning is continuous [13], [14].

There are lots of interesting work on training weakly connected agents to learn the true state, when the true state is time-invariant. The authors in [15]–[18] applied a graph theoretic approach and linear algebraic manipulations to a partitioned adjacency matrix resulting from the graph network. The weakly connected agents were able to converge to the true state, although with a belief probability less than one. This means that the weakly connected agents could still be controlled by the strongly connected agents, and this is a limitation of the work. The authors in [19] overcame this limitation by using a log-intermediate belief and updated the belief with an exponential function. The speed of convergence improved, and the weakly connected agents' belief probabilities reached one for the true state over time. In [20], the authors derived some closed-form expressions to determine how close both strongly connected agents and weakly connected agents get to their limiting points in terms of mean-square-error. The mean-square-error performance of weakly connected agents is determined by the mean-square-error of the strongly connected agents. In [21], the authors did not only proposed a model that trained weakly connected agents to converge to the true state, but also studied the reverse problem of learning the network topology given that the weakly connected agents received some measurements from the strongly connected agents. One of their main results is that for topology learning to occur, the number of hypotheses or states, must not be less than the number of strongly connected subnetworks. This means that if there are two strongly connected subnetworks in the graph network, then, the number of states - which include the true state - must be at least two.

Most existing literature that applied graph theory to study truth-learning in the social network assumed that the true state is *time-invariant* [11], [13], [14], [22]. An analogy of the approach in this existing literature is described as follows: At the initial time, all agents are doubtful of the true state.

However, after cooperation among one another, the beliefs of the influential agents in the strongly connected subnetwork will converge to the true state over time. On the other hand, the agents in the weakly connected subnetwork will lose their own beliefs and will be subtly manipulated to accept the false beliefs of agents in the strongly connected subnetwork. A limitation of this social learning model is that it does not consider the dynamic nature of social network. This is not practical, since streams of information are released and updated every second in social networks [23], [24]. Thus, the true state is *arbitrarily time-varying*. Hence, the analyses and results in this existing literature are not applicable to a dynamic setting. It has been shown in [25], [26], that convergence of agents to the true state is difficult to achieve when the true state is dynamic. For the setting discussed in this article, where the true state is arbitrarily time-varying, common existing approaches are therefore inadequate; hence online (reinforcement) learning is introduced into the graph theoretic framework.

Online learning is an aspect of machine learning where agents receive information sequentially. Online learning has shown to perform well in predicting the time-varying true state for strongly connected agents unlike conventional social learning methods that fail. For instance, [27] proposed an online learning approach that can help strongly connected agents predict the time-varying true state. Although, the performance of the agents often come with some *regret*. Minimizing this regret is the main goal of online learning. Online learning strategies have the flexibility to work with different forms of observed private signals. For instance, in most existing literature, the private signals observed by the agents are strictly linear; however online learning can work for arbitrary signals [28]. There are many online learning strategies. The multi-armed bandit technique, which is an online learning strategy, has proven to be one of the most successful for social learning even in difficult situations, where cooperation among the agents is difficult [27], [29]. In such a setting, an agent may be denied feedback information from its neighboring agents making consensus difficult.

Although, the multi-armed bandit technique has proven to be quite effective for training strongly connected agents to learn the truth when the true state is arbitrarily time-varying [27], [30], it is yet to be applied for weakly connected agents. Thus, this article applies multi-armed bandit technique to help weakly connected agents predict the time-varying true state. The regret incurred by these weakly connected agents is expected to be higher than the regret incurred by strongly connected agents. A preliminary study of this work can be found in [31].

A. RESEARCH CONTRIBUTIONS

The contributions of this article are as follows:

- 1) It studies the problem of negative influence from influential personalities to their followers in social networks. To address this problem, the social network is

modeled as a graph network, where the influential personalities form strongly connected subnetworks, and their followers form weakly connected subnetworks.

- 2) It applies non-Bayesian learning and multi-armed bandit technique, which is an online learning strategy, to help the agents in a weakly connected subnetwork learn a time-varying true state. Hence, a non-stochastic multi-armed bandit algorithm is proposed.
- 3) It provides simulation results that show that the weakly connected agents converge to the most stable state, despite the arbitrarily time-varying nature of the true state. The most stable state is the true state that appears to be the most stable among the sequence of arbitrarily time-varying true states over the time horizon. Also, the sublinearity of the proposed algorithm for the weakly connected subnetwork is compared with the sublinearity of the state-of-the-art algorithm for strongly connected subnetwork already established in literature.

The rest of the paper is organized as follows: Section II explains the system model in detail, Section III explains the proposed algorithm, Section IV gives some theoretical results, Section V discusses simulation results, Section VI concludes the findings, and Appendix A shows the proof of Theorem 1.

II. SYSTEM MODEL

A. NETWORK MODEL

Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{1, \dots, N\}$ represents the set of agents in the network with $|\mathcal{V}| = N$. Let a pair of non-negative scalar weights $\{a_{jk}, a_{kj}\} \in \mathcal{E}$ be assigned to the edges connecting agents $k \in \mathcal{V}$ and $j \in \mathcal{V}$. The edge weight a_{jk} represents the weight assigned to the directed path from j to k . Similarly, a_{kj} represents the weight assigned to the directed path from k to j . The network is said to be strongly connected if there exists a directed path in both ways connecting any two agents of a neighborhood, and there is at least one self-loop, i.e., $a_{kk} > 0$ for an agent k . The presence of at least a self-loop means that in the strongly connected subnetwork, there is at least one agent who uses its own opinions in its decision-making process. Such an agent is said to be self-conscious. Agents not connected by an edge have a weight of 0 for that direction. This implies that it is possible to have $a_{jk} > 0$, but $a_{kj} = 0$. Adopting the definition for a weakly connected network in [16], a weakly connected network is defined as a network that acts as a receiver only and can be dominated. The neighborhood of any agent, say agent k , is the set of agents connected to k . Denote the neighborhood of agent k as \mathcal{N}_k . Note that agent k is a member of its own neighborhood. The adjacency matrix of the graph is a square matrix whose elements are the weights of the edges linking any two agents. Denote the adjacency matrix as A . When each column vector elements in the adjacency matrix sum up to one, then the matrix is said to be left-stochastic, i.e.,

$$a_{jk} \geq 0, \quad \sum_j a_{jk} = 1. \quad (1)$$

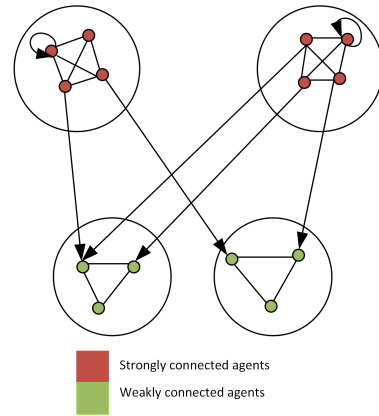


FIGURE 1. A graph network consisting of two strongly connected subnetworks controlling two weakly connected subnetworks.

A graph network may be a combination of subnetworks, as shown in Fig. 1. The top two subnetworks are strongly connected, while the bottom two subnetworks are weakly connected. All subnetworks in Fig. 1 have directed arrows, but the directed arrows in each subnetwork are purposely omitted to avoid confusion with the directed arrows indicating domination. Also, while the strongly connected subnetworks have at least one self-loop, the weakly connected subnetworks have no self-loop. As common in graph theory [12], [13], [16], a strongly connected subnetwork is left-stochastic and has a spectral radius of one, i.e., the eigenvalues are bounded by one. A strongly connected subnetwork also follows the Perron-Frobenius theorem, and has a single eigenvalue at one, while other eigenvalues are strictly inside a unit disc. On the other hand, a weakly connected subnetwork is not left-stochastic and does not include the strongly connected subnetwork in its neighborhood.

B. DIFFUSION LEARNING

In diffusion learning, all agents start with a uniform prior belief over each state in the network. Let $\Theta = \{\theta_1, \dots, \theta_M\}$ be the set of all possible states that can be detected by a graph network, and let $\theta_t^* \in \Theta$ denote the unknown time-varying true state of the network at time t . Intuitively, Θ represents a bounded set of discrete information containing a time-varying truth. Take for the purpose of illustration only, θ_t^* could mean the price of stock that fluctuates arbitrarily over time, and Θ could mean a set of all known stock prices, any of which could be θ_t^* . Here, the strongly connected agents could mean a collection of some wealthy entrepreneurs and stockbrokers. The weakly connected agents could mean some social media followers of these entrepreneurs who are novice to the stock market, but wish to invest in stock. Each agent in the entire graph network has some prior belief. The prior belief for the agent $k \in \mathcal{V}$ for instance, is denoted as $\mu_{k,0}(\theta) = \frac{1}{M}$ at time $t = 0$. It is to be noted that the agent $k \in \mathcal{V}$ in this context refers to any agent in the entire graph network. Each agent will update its belief at each time $t \geq 1$ first by observing a random private signal. Intuitively, this private

signal represents a side observation not fully informative but accessible by each agent. From the above illustration, the private signal could mean some rumors about the stock prices. The observed private signal of the agent k is denoted as $S_{k,t}$, and it is drawn from some known likelihood function $L_k(\cdot|\theta_t^*)$ that is dependent on the time-varying true state θ_t^* . $S_{k,t}$ is a member of a finite state space $\mathcal{Z}_{k,t}$. The private signals are independent over time and over all agents. These signals are not fully informative which necessitates cooperation among the agents i.e.,

$$S_{k,t} = \theta_t^* + n, \quad \forall k \in \mathcal{V}, t \leq T \quad (2)$$

where $n \sim \mathcal{N}(0, 1)$. The observed private signal is a noisy version of the underlying time-varying true state. The agent k uses this observed private signal to compute the likelihood $L_k(S_{k,t}|\theta)$ over each state $\theta \in \Theta$ as follows:

$$L_k(S_{k,t}|\theta) = \frac{1}{\sqrt{2\pi\sigma_{k,t}^2}} \exp\{-\frac{(S_{k,t} - \theta)^2}{2\sigma_{k,t}^2}\} \quad (3)$$

where $\sigma_{k,t}^2$ is the variance of agent k at time t . Then, the agent uses the Bayesian rule to generate an intermediate belief as follows:

$$\psi_{k,t}(\theta) = \frac{\mu_{k,t-1}(\theta)L_k(S_{k,t}|\theta)}{\sum_{\theta' \in \Theta} \mu_{k,t-1}(\theta')L_k(S_{k,t}|\theta')} \quad (4)$$

where $\psi_{k,t}(\theta)$ is the intermediate belief of the agent k at round t . To account for the exploration-exploitation trade-off in online learning, an intermediate probability $p_{k,t}(\theta)$ is introduced. This intermediate probability is computed from the intermediate belief. This is given as:

$$p_{k,t}(\theta) = (1 - \gamma)\mu_{k,t-1}(\theta) + \gamma\psi_{k,t}(\theta), \quad (5)$$

where γ is the exploration parameter. The agent k cooperates with other agents in its neighborhood to compute a consensus probability $P_{k,t}(\theta)$, using the weight connections in the adjacency matrix. This is illustrated as follows:

$$P_{k,t}(\theta) = \sum_{j \in \mathcal{N}_k} a_{jk} p_{j,t}(\theta). \quad (6)$$

If k is a strongly connected agent, then, $a_{kk} \geq 0$ in (6), and if k is a weakly connected agent, then, $a_{kk} = 0$. Because there is at least an agent in the strongly connected subnetwork with self-loop, there is at least an agent that will use a weighted portion of its own intermediate probability to compute the consensus probability according to (6). However, in the weakly connected subnetwork, since there is no self-loop, there is no agent that uses its intermediate probability to compute its consensus probability. This implies that none of the weakly connected agents uses its opinions in its decision-making process.

C. ONLINE LEARNING

The network is set up as an online learning problem that is best described as a game between the agents and an oblivious adversary. The goal of the adversary is to maximize the regret. Thus, the agents must be able to make smart decisions to outwit the adversary and minimize regret. Before the game begins, the adversary fixes the loss $l_{k,t}(\theta) \in [0, 1]$ for each agent k at each time t and over all the states. The time-varying true state incurs no loss, i.e., $l_{k,t}(\theta_t^*) = 0$. The agents can minimize regret by ensuring it accurately predict the time-varying true state on most occasion, and thus incur minimum number of losses throughout the entire duration of the game. The agents' performance can be benchmark against an oracle that has some knowledge of the game setting, and would prefer to stick to a state that incurs the lowest possible losses over the entire duration of the game. Therefore, each agent's expected regret is the difference between the total expected loss incurred by the agent that follows a randomized approach in predicting the time-varying true state, and the total expected loss of the oracle who chooses to stick to the best fixed state θ^* for the entire duration of the game. This is given as:

$$R(T) = \mathbb{E}_{\mathcal{F}_T} \left[\sum_{t=1}^T \sum_{\theta \in \Theta} p_{k,t}(\theta) l_{k,t}(\theta) - \sum_{t=1}^T l_{k,t}(\theta^*) \right], \quad (7)$$

where $R(T)$ is the expected regret over the time horizon T , and the filtration $\mathcal{F}_T = \sigma(S_{k,1}, \dots, S_{k,T}, l_{k,1}, \dots, l_{k,T}, \theta_1, \dots, \theta_T)$ represents the history of all observed private signals, states chosen and incurred losses.

III. PROPOSED ALGORITHM

The proposed algorithm is an adversarial multi-armed bandit algorithm designed to help weakly connected agents to predict the time-varying true state. The parameters for the algorithm are the feedback graph, the learning rate $\eta \geq 0$ and the exploration parameter $\gamma \in (0, \frac{1}{2}]$. The input to the algorithm is the adjacency matrix A , and the number of states M . The output of the algorithm is the belief $\mu_{k,t}(\theta) \forall \theta \in \Theta$. The operation of the algorithm is discussed as follows: In step 0, each agent's belief $\mu_{k,0}(\theta)$ is initialized over the state $\theta \in \Theta$ as a uniform distribution. For each round of the algorithm, the following steps are executed: In step 1, the intermediate probability $p_{k,t}(\theta)$ is computed. This involves a trade-off between exploitation and exploration with the parameter γ . In exploitation, the algorithm sticks to the past belief of each agent about the states, while in exploration, the algorithm combines the likelihood of each agent over the states with the effect of domination from the strongly connected subnetwork. The trade-off is necessary to minimize regret. Thus, $p_{k,t}(\theta)$ is computed with the introduction of the *domination number* δ (see Definition 1). The exploration parameter is shared evenly between $\psi_{k,t}(\theta)$ and $\frac{1}{\delta}$. In step 2, the consensus probability is computed from (6). In step 3, a state is drawn at random according to the consensus probability distribution $P_{k,t}$. Loss is incurred for the chosen state. In step 4, the estimated loss

Proposed Algorithm: Online Diffusion Learning for Weakly Connected Network

Parameters: Feedback graph, learning rate $\eta > 0$.

V is the set of weakly connected agents and E is the set of edges.

Exploration parameter $\gamma \in (0, \frac{1}{2}]$.

Input: The adjacency matrix A , $|\Theta| = M$

Output: The belief $\mu_{k,t}(\theta) \forall \theta \in \Theta$

Step 0: Initialize $\mu_{k,0}(\theta) = \frac{1}{M}$

For each round $t \in \{1, \dots, T\}$

Step 1: Compute $p_{k,t}(\theta) = (1 - \gamma)\mu_{k,t-1}(\theta) + [\frac{\gamma}{2}\psi_{k,t}(\theta) + \frac{\gamma}{2\delta}] \forall \theta \in \Theta$

Step 2: Compute $P_{k,t}(\theta) = \sum_{j \in \mathcal{N}_k} a_{jk} p_{j,t}(\theta)$, $P_{k,t} = (P_{k,t}(\theta_1), \dots, P_{k,t}(\theta_M))$

Step 3: Draw state $\theta_t \sim P_{k,t}$ and incur loss $l_{k,t}(\theta_t) \in [0, 1]$

Step 4: Compute

$$\hat{l}_{k,t}(\theta) = \frac{l_{k,t}(\theta)}{P_{k,t}(\theta)} \mathbb{I}\{\theta = \theta_t\} \quad \forall \theta \in \Theta$$

Step 5: Update

$$\mu_{k,t}(\theta) = \frac{\mu_{k,t-1}(\theta) \exp(-\eta \hat{l}_{k,t}(\theta))}{\sum_{\theta' \in \Theta} \mu_{k,t-1}(\theta') \exp(-\eta \hat{l}_{k,t}(\theta'))} \quad \forall \theta \in \Theta$$

Step 6: end

over the whole state is computed. This is important because the algorithm needs to update the belief over all the states but does not know the value of the losses for unchosen states at each time. This is typical of multi-armed bandit settings. This estimated loss is an unbiased estimate of the true loss in expectation, as shown in Lemma 1. In step 5, the algorithm updates the belief of each agents over all the states using an exponential function. Notice that the belief update equation is normalized to ensure that the sum of the beliefs over all the states is one. Step 6 ends an iteration of the algorithm. The algorithm repeats from step 1 until the time horizon is reached.

The goal of the proposed algorithm is to converge to the most stable state from the arbitrary sequence of time-varying true states over the time horizon. The most stable state appears to be the true state that is the most stable from the arbitrary sequence $\theta_1^*, \dots, \theta_T^*$. The belief probability of all agents over the most stable state is expected to reach a value of 1 over time. Intuitively, the agents in a weakly connected subnetwork will converge to the state that appears to be the truth in most occasion.

IV. THEORETICAL RESULTS

In this section, theoretical results are presented.

Definition 1: The weak domination number of a graph G is denoted by $\delta(G)$, and is the smallest size of any subset $D \subseteq \mathcal{V}$, which belongs to the strongly connected subnetwork and dominates the weakly connected subnetwork.

Remark: In standard graph theory, a weakly dominating set $\mathcal{D} \subseteq \mathcal{V}$, is the set of nodes that dominates the weakly connected subnetwork. Computing a maximal dominating set is NP-hard but it can be efficiently approximated within a logarithmic factor using the greedy algorithm [32].

Assumption 1: The exploration over the subset D and the intermediate belief $\psi_{k,t}(\theta)$ is assumed to be uniform.

This assumption is useful for the proof of Theorem 1.

Lemma 1: The estimated loss $\hat{l}_{k,t}(\theta)$, is an unbiased estimate of the true loss $l_{k,t}(\theta)$ in expectation, and it is given as

$$\begin{aligned} \mathbb{E}_{\mathcal{F}_t/\mathcal{F}_{t-1}} [\hat{l}_{k,t}(\theta) | \mathcal{F}_{t-1}] &= l_{k,t}(\theta) \text{ and} \\ \mathbb{E}_{\mathcal{F}_t/\mathcal{F}_{t-1}} [\hat{l}_{k,t}(\theta)^2 | \mathcal{F}_{t-1}] &= \frac{l_{k,t}(\theta)^2}{P_{k,t}(\theta)}. \end{aligned} \quad (8)$$

Proof:

$$\begin{aligned} &\mathbb{E}_{\mathcal{F}_t/\mathcal{F}_{t-1}} [\hat{l}_{k,t}(\theta) | \mathcal{F}_{t-1}] \\ &= \sum_{\Theta} \frac{l_{k,t}(\theta)}{P_{k,t}(\theta)} P_{k,t}(\theta_t) \mathbb{I}\{\theta_t = \theta\} = l_{k,t}(\theta) \\ &\mathbb{E}_{\mathcal{F}_t/\mathcal{F}_{t-1}} [\hat{l}_{k,t}(\theta)^2 | \mathcal{F}_{t-1}] \\ &= \sum_{\Theta} \left(\frac{l_{k,t}(\theta)}{P_{k,t}(\theta)} \right)^2 P_{k,t}(\theta_t) \mathbb{I}\{\theta_t = \theta\} \\ &= \frac{l_{k,t}(\theta)^2}{P_{k,t}(\theta)}. \end{aligned}$$

Theorem 1: The upper bound on the expectation of the regret in the proposed algorithm for the weakly connected network is $\mathcal{O}((\delta \ln M)^2 T^{2/3})$ when $\gamma = \min \left\{ \frac{(\delta \ln M)^2}{T^{1/3}}, \frac{1}{2} \right\}$, $\eta = \frac{\gamma^2}{\delta}$, and $T \geq M^3 \ln M / \delta^2$.

Proof: See Appendix A.

Remark: The regret bound for the proposed algorithm is worse than the regret bound for strongly connected network obtained in [30], which is $\mathcal{O}(\sqrt{\alpha T \ln M})$, where α represents the independence number of the strongly connected graph. However, the regret bound of the proposed algorithm is comparable to the regret bound of EXP3.G for weakly observable graphs of order $\tilde{\mathcal{O}}(T^{2/3})$ [33], where the tilde symbol represents the removal of some constant parameters. In EXP3.G, the loss feedback is not strictly bandit like the proposed algorithm, and each weakly connected agent is allowed to observe the losses of its neighbors. Also, the proposed regret bound is comparable to the regret bound of

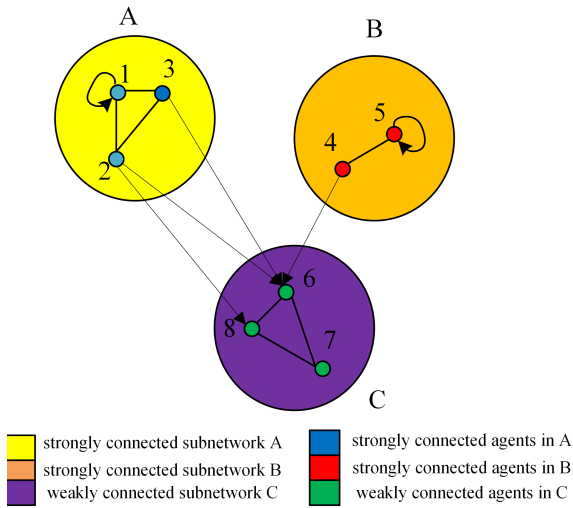


FIGURE 2. Two strongly connected subnetworks A and B controlling weakly connected subnetwork C.

Lazy Revealing Action algorithm with order $\tilde{O}(T^{2/3})$ for the full information setting [34], where each weakly connected agents can observe the losses of all agents in the graph network. Despite the restrictions of the bandit setting, the proposed algorithm has the same regret bound as the EXP3.G and Lazy Revealing Action algorithms with less restrictive settings. This is an advantage of the proposed algorithm.

V. SIMULATION RESULTS

The simulation uses the network in Fig. 2 consisting of two strongly connected subnetworks A and B and a weakly connected subnetwork C. The adjacency matrix for the weakly connected subnetwork is given as:

$$A = \begin{bmatrix} 0 & 0.2 & 0.8 \\ 0.5 & 0 & 0.1 \\ 0.3 & 0.4 & 0 \end{bmatrix}$$

It can be observed from Fig. 2 that none of the weakly connected agents 6, 7 and 8 in the weakly connected subnetwork C has a self-loop. This means that for a weakly connected agent k (which may be any of 6,7 or 8), its self-loop weight $a_{kk} = 0$. Since there is not a single self-loop in subnetwork C, none of the weakly connected agents uses its own opinion for decision making. A weakly connected subnetwork is known not to be left-stochastic; hence, the adjacency matrix A can be formed such that $a_{kk} = 0$, and $\sum_j a_{jk} < 1$ with $a_{jk} \geq 0$. The domination number in Fig. 2 is 2 since two nodes from the strongly connected subnetworks are sufficient to influence the weakly connected subnetwork. For instance, node 2 and node 4 from the two strongly connected subnetworks in Fig. 2 are sufficient to send one-way information from both strongly connected subnetworks to the weakly connected subnetwork. It is not important to show the convergence of the strongly connected agents, as this is shown already in [30]. However, it is of importance to show that the weakly connected agents

can converge to the most stable state, albeit, at a slower rate compared to the strongly connected agents.

Assuming there are five states, i.e., $\Theta = \{\theta_1, \dots, \theta_5\}$, any of which can be the true state at each time t , since the true state varies arbitrarily over time. This time-varying true state θ_t^* is randomly chosen using the randi function in MATLAB. If algorithm 1 is implemented, each weakly connected agent will converge to the most stable state over time, despite the time-varying nature of the true state. The parameters used for the simulation are: $t = 1, \dots, 400$; $\eta = 0.2$; and $\gamma = 0.1$. The private signals observed by each agent k is drawn from a time-varying Gaussian distribution $\mathcal{N}(\theta_t^*, 1)$, where θ_t^* represents the arbitrarily time-varying true state at time t , and the distribution is centered around this time-varying true state. The variance of the distribution is 1. It is to be noted that at each time t , each agent draws its observation from this distribution independently. At the start of the algorithm, the belief $\mu_{k,0}(\theta)$ is uniformly distributed over the state. This means that $\mu_{k,0}(\theta) = \frac{1}{5}$. The algorithm is iterated 50 times. The settings for the simulation parameters are very similar to what was used in [30]. However, the algorithms in [30] are different from the proposed algorithm due to the presence of domination in the weakly connected subnetwork.

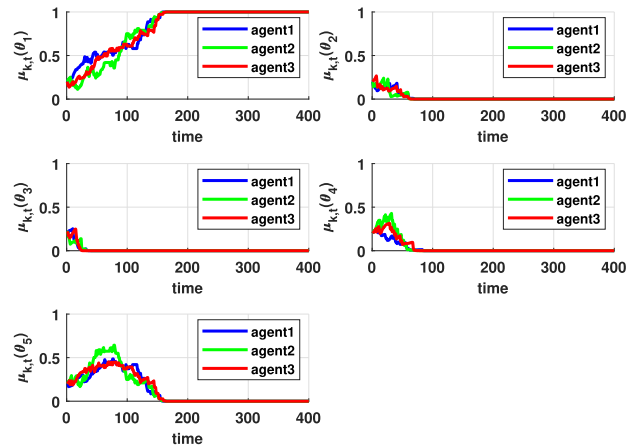


FIGURE 3. Convergence of agents' belief in weakly connected subnetwork C to θ_1 at $\eta = 0.1$ at the 1st iteration.

Figs. 3-5 show how the weakly connected agents converge to the most stable state using algorithm 1. Each of the figures have five plots. Four of those plots show how the beliefs of agents 6, 7 and 8 converge to zero for some states. However, the beliefs of the agents 6, 7 and 8 will converge to a value of 1 only for the most stable state. The most stable state appears to vary from iteration to iteration because the sequence of the true states $\theta_1^*, \dots, \theta_T^*$ varies arbitrarily over the iteration.

Fig. 3 shows the convergence of the beliefs of agents 6, 7 and 8 at the 1st iteration when $\eta = 0.1$. It can be seen that the beliefs of these agents for states $\theta_2, \theta_3, \theta_4$ and θ_5 go to zero over time. However, the beliefs of these weak agents converge to θ_1 with a value of 1 at $t = 161$. Thus, θ_1 is the most stable state at the 1st iteration.

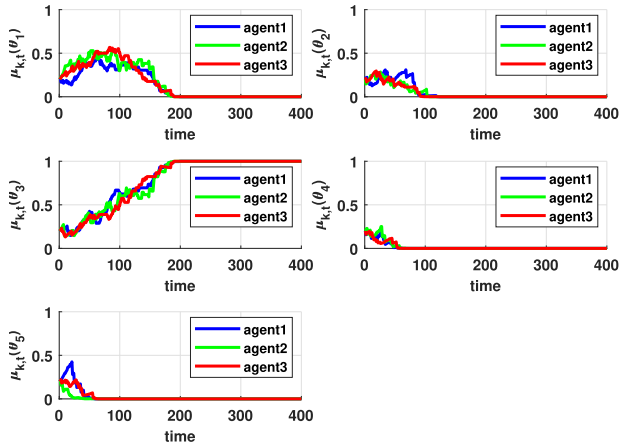


FIGURE 4. Convergence of agents' belief in weakly connected subnetwork C to θ_1 at $\eta = 0.1$ at the 50th iteration.

Fig. 4 shows the convergence of the beliefs of agents 6, 7 and 8 at the 50th iteration with $\eta = 0.1$. It can be seen that the beliefs of these weakly connected agents for states $\theta_1, \theta_2, \theta_4$ and θ_5 go to zero over time. However, the beliefs of these agents converge to θ_3 with a value of 1 at time $t = 189$. Thus, θ_3 is the most stable state at the 50th iteration. The most stable state at the 50th iteration differs from the most stable state at the 1st iteration. Hence, the most stable state varies arbitrarily over the number of iterations. The convergence of these weakly connected agents is slow compared to the convergence of strongly connected agents shown in [30]. We can compare the average speed of convergence for weakly connected agents, using the proposed algorithm, to the average speed of convergence for strongly connected agents in [30], using the state-of-the-art algorithm, over 50 iterations. On average, the weakly connected agents converge at time $t = 171$, while on average, the strongly connected agents converge at time $t = 103$. This means that the weakly connected agents converge 66% times slower than the strongly connected agents.

To show how important is the learning rate for fine-tuning algorithm 1, a higher learning rate $\eta = 0.2$ is used in Fig. 5. At the 50th iteration, the beliefs of the weakly connected agents 6, 7 and 8 go to zero for states $\theta_1, \theta_2, \theta_4$ and θ_5 . However, the beliefs of these weakly connected agents converge to θ_3 with a value of 1 at time $t = 124$. Here, the convergence is improved. Thus, fine-tuning the learning rate can improve the speed of convergence of the weakly connected agents. On average, the weakly connected agents converge at time $t = 120$. This is approximately 50% improvement.

Fig. 6 shows the comparison of the sublinearity of the regret bound of algorithm 1, which is given as $\mathcal{O}((\delta \ln M)^2 T^{2/3})$, to the regret bound obtained in [30], which is $\mathcal{O}(\sqrt{\alpha T \ln M})$ for strongly connected agents. The sublinearity is defined as $\lim_{t \rightarrow \infty} \frac{R(t)}{t}$. It can be seen that the regret bound of algorithm 1 decays slowly compared to that of the strongly connected agents in [30], when other parameters of the regret bound are kept constant except time.

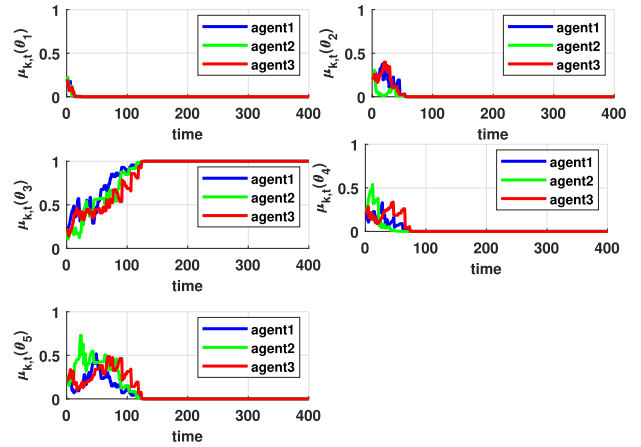


FIGURE 5. Convergence of agents' belief in weakly connected subnetwork C to θ_3 at $\eta = 0.2$ at the 50th iteration.

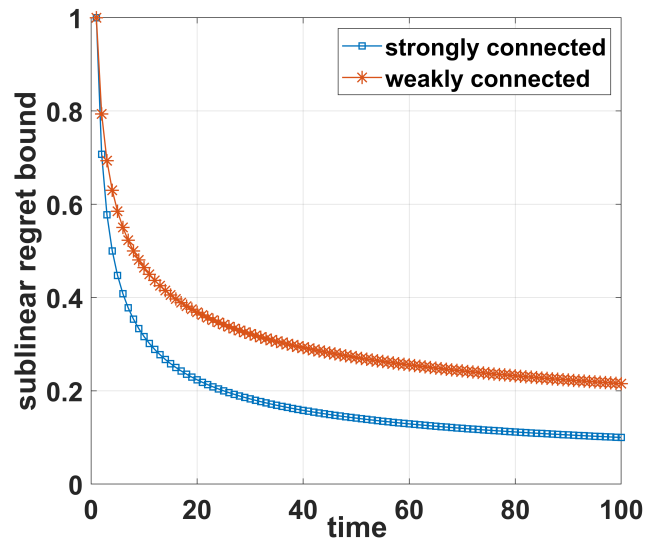


FIGURE 6. Sublinear regret bound comparison between the algorithm designed for strongly connected agents in [30] and the proposed algorithm designed for weakly connected agents.

We can compare the results in this article with the results obtained in [16] and [21], for the case where the true state is time-invariant. From Fig. 9 in [16], it can be seen that the weakly connected agents converge to the true state with a belief probability less than 1. This means that the weakly connected agents are still manipulated by the strongly connected agents. This model is not very good in practice, as the goal should be to train the weakly connected agents to learn the true state with a belief probability of 1. On the other hand, the authors in [21] showed that the weakly connected agents can learn the time-invariant true state with a belief probability of 1, as shown in Fig. 8 in [21]. The convergence in this article is comparable to what is obtained in [21], even though the true state is arbitrarily time-varying here.

This work can be extended to situations, where the strongly connected agents have limited control over the weakly connected agents. More so, it will be interesting to improve the

regret bound for the weakly connected agents in this article. A much challenging problem will be situations where the transfer of information from the strongly connected agents to the weakly connected agents is both ways. Also, the algorithm may be trained on massive social media data that can be modeled as a massive graph network, instead of using a graph network with few number of nodes.

VI. CONCLUSION

In conclusion, this article studied leader-follower relationships in social networks, and proposed how followers can overcome manipulations from leaders (or influential personalities), in order to learn an arbitrarily time-varying truth. These influential personalities form strongly connected sub-networks, while their followers form weakly connected sub-networks. It has been shown in existing work that strongly connected agents can cooperate among themselves to learn this time-varying truth - otherwise called the time-varying true state. However, training weakly connected agents to learn the arbitrarily time-varying true state is yet to be investigated. Thus, this article focused on training weakly connected agents to converge to the most stable state over time from the arbitrary sequence of time-varying true states over the time horizon. To achieve this, a non-stochastic multi-armed bandit algorithm is proposed, and it is shown by simulation that the beliefs of weakly connected agents can converge to this most stable state. Also, it is shown that the most stable state varies randomly over the number of iterations. The speed of convergence for these weakly connected agents trained with the proposed algorithm is slower by 66% on average, when compared to the speed of convergence for strongly connected agents using the state-of-the-art algorithm already established in the literature. This is because weakly connected agents are harder to train. Fine-tuning the learning rate of the proposed algorithm can improve the speed of convergence of the weakly connected agents by approximately 50% on average. Finally, The sublinearity of the regret bound for the proposed algorithm is compared with the sublinearity of the regret bound for the state-of-the-art algorithm for strongly connected agents.

APPENDIX A

By equating the upperbound and lowerbound on the fraction $\ln \frac{W_{k,T}}{W_{k,0}}$, we can subtly find the upperbound of the regret. This approach is well-known in multi-armed bandit (Lemma 4 and Theorem 2 in [33], [35]).

Let $\mu_{k,t}(\theta) := \frac{w_{k,t}(\theta)}{W_{k,t}}$ and $W_{k,t} := \sum_{\theta \in \Theta} w_{k,t-1}(\theta) \exp(-\eta \hat{l}_{k,t}(\theta))$. Verify that the update equation for $\mu_{k,t}(\theta)$ in Step 5 of the proposed algorithm is same as the definition of $\mu_{k,t}(\theta)$ here.

$$\frac{W_{k,t}}{W_{k,t-1}} = \frac{\sum_{\theta \in \Theta} w_{k,t-1}(\theta) \exp(-\eta \hat{l}_{k,t}(\theta))}{W_{k,t-1}}$$

$$\begin{aligned} &= \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \exp(-\eta \hat{l}_{k,t}(\theta)) \\ &\leq \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \left(1 - \eta \hat{l}_{k,t}(\theta) + \eta^2 \hat{l}_{k,t}^2(\theta)\right) \end{aligned}$$

where we use

$$\begin{aligned} e^x &\leq 1 + x + x^2 \text{ for all } x \leq 1 \\ &\leq 1 - \eta \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \hat{l}_{k,t}(\theta) + \eta^2 \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \hat{l}_{k,t}^2(\theta) \end{aligned} \tag{A-1}$$

where we use $\sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \leq 1$ in (A-1) using $\ln(1-x) \leq -x$

$$\begin{aligned} \ln \frac{W_{k,t}}{W_{k,t-1}} &= \ln \left(1 - \eta \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \hat{l}_{k,t}(\theta) + \eta^2 \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \hat{l}_{k,t}^2(\theta)\right) \\ &\leq -\eta \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \hat{l}_{k,t}(\theta) + \eta^2 \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \hat{l}_{k,t}^2(\theta). \end{aligned}$$

Summing over $t = 1, \dots, T$, we get

$$\ln \frac{W_{k,T}}{W_0} \leq \sum_{t=1}^T \sum_{\theta \in \Theta} \left(-\eta \mu_{k,t-1}(\theta) \hat{l}_{k,t}(\theta) + \eta^2 \mu_{k,t-1}(\theta) \hat{l}_{k,t}^2(\theta)\right) \tag{A-2}$$

We apply telescoping series in (A-2) to arrive at $\ln \frac{W_{k,T}}{W_0}$. Thus, we have found the upperbound on $\ln \frac{W_{k,T}}{W_0}$.

Now, we proceed to find the lowerbound on $\ln \frac{W_{k,T}}{W_0}$. For any fixed $\theta_f \in \Theta$,

$$\ln \frac{W_{k,T}}{W_{k,0}} \geq \ln \frac{w_{k,T}(\theta_f)}{W_{k,0}} = -\eta \sum_{t=1}^T \hat{l}_{k,t}(\theta_f) - \ln W_{k,0} \tag{A-3}$$

where $W_{k,0} = M$. In (A-3), we obtain $\ln w_{k,T}(\theta_f) = -\eta \sum_{t=1}^T \hat{l}_{k,t}(\theta_f)$ as follows:

$$\begin{aligned} w_{k,T}(\theta_f) &= w_{k,T-1}(\theta_f) \exp(-\eta \hat{l}_{k,T}(\theta_f)) \\ &= w_{k,T-2}(\theta_f) \exp\left(-\eta \left(\hat{l}_{k,T}(\theta_f) + \hat{l}_{k,T-1}(\theta_f)\right)\right) \\ &\vdots \\ &= w_{k,0}(\theta_f) \exp\left(-\eta \sum_{t=1}^T \hat{l}_{k,t}(\theta_f)\right) \end{aligned}$$

But $w_{k,0}(\theta_f) = 1$ because $W_{k,0} = M$ and $\mu_{k,0}(\theta_f) = \frac{1}{M}$. Applying the natural logarithm, we arrive at $-\eta \sum_{t=1}^T \hat{l}_{k,t}(\theta_f)$.

Now, to find the upperbound on the regret, we equate (A-2) and (A-3)

$$\sum_{t=1}^T \sum_{\theta \in \Theta} \left(-\eta \mu_{k,t-1}(\theta) \hat{l}_{k,t}(\theta) + \eta^2 \mu_{k,t-1}(\theta) \hat{l}_{k,t}^2(\theta)\right)$$

$$\begin{aligned} &\geq -\eta \sum_{t=1}^T \hat{l}_{k,t}(\theta_f) - \ln M \\ &\sum_{t=1}^T \sum_{\theta \in \Theta} \eta \mu_{k,t-1}(\theta) \hat{l}_{k,t}(\theta) - \eta \sum_{t=1}^T \hat{l}_{k,t}(\theta_f) \\ &\leq \sum_{t=1}^T \sum_{\theta \in \Theta} \eta^2 \mu_{k,t-1}(\theta) \hat{l}_{k,t}^2(\theta) + \ln M. \end{aligned}$$

Dividing by η , taking conditional expectation and introducing min function to show the fixed choice of the oracle with the lowest possible accumulated loss over the time horizon, we get

$$\begin{aligned} &\mathbb{E}_{\mathcal{F}_t/\mathcal{F}_{t-1}} \left[\sum_{t=1}^T \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \hat{l}_{k,t}(\theta) \middle| \mathcal{F}_{t-1} \right] \\ &- \mathbb{E}_{\mathcal{F}_t/\mathcal{F}_{t-1}} \left[\min_{\theta_f \in \Theta} \sum_{t=1}^T \hat{l}_{k,t}(\theta_f) \middle| \mathcal{F}_{t-1} \right] \\ &\leq \mathbb{E}_{\mathcal{F}_t/\mathcal{F}_{t-1}} \left[\sum_{t=1}^T \sum_{\theta \in \Theta} \eta \mu_{k,t-1}(\theta) \hat{l}_{k,t}^2(\theta) \middle| \mathcal{F}_{t-1} \right] + \frac{\ln M}{\eta} \end{aligned}$$

Opening the bracket yields

$$\begin{aligned} &\sum_{t=1}^T \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \mathbb{E}_{\mathcal{F}_t/\mathcal{F}_{t-1}} \left[\hat{l}_{k,t}(\theta) \middle| \mathcal{F}_{t-1} \right] \\ &- \min_{\theta_f \in \Theta} \sum_{t=1}^T \mathbb{E}_{\mathcal{F}_t/\mathcal{F}_{t-1}} \left[\hat{l}_{k,t}(\theta_f) \middle| \mathcal{F}_{t-1} \right] \\ &\leq \sum_{t=1}^T \sum_{\theta \in \Theta} \eta \mu_{k,t-1}(\theta) \mathbb{E}_{\mathcal{F}_t/\mathcal{F}_{t-1}} \left[\hat{l}_{k,t}^2(\theta) \middle| \mathcal{F}_{t-1} \right] + \frac{\ln M}{\eta} \end{aligned} \tag{A-4}$$

Applying Lemma 1 in (A-4), we get

$$\begin{aligned} &\sum_{t=1}^T \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) l_{k,t}(\theta) - \sum_{t=1}^T l_{k,t}(\theta^*) \\ &\leq \eta \sum_{t=1}^T \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \frac{l_{k,t}^2(\theta)}{P_{k,t}(\theta)} + \frac{\ln M}{\eta} \\ &\sum_{t=1}^T \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) l_{k,t}(\theta) - \sum_{t=1}^T l_{k,t}(\theta^*) \\ &\leq \eta \sum_{t=1}^T \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) \frac{1}{P_{k,t}(\theta)} + \frac{\ln M}{\eta} \end{aligned} \tag{A-5}$$

where we use $l_{k,t}^2(\theta) \leq 1$ in (A-5)

Using assumption 1, because of uniform exploration over \mathcal{D} and ψ_{kt}

$$p_{k,t}(\theta) \geq \frac{\gamma}{2\delta}; \text{ hence } P_{k,t}(\theta) \geq \frac{\gamma}{2\delta} \tag{A-6}$$

(A-6) is gotten from Step 1 and Step 2 of the proposed algorithm. We can find the bound on $\sum_{\theta \in \Theta} \frac{\mu_{k,t-1}(\theta)}{P_{k,t}(\theta)}$ and

substitute into (A-5).

$$\sum_{\theta \in \Theta} \frac{\mu_{k,t-1}(\theta)}{P_{k,t}(\theta)} \leq \frac{2\delta}{\gamma} \tag{A-7}$$

we use $\sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) = 1$ in (A-7). Substituting (A-7) into (A-5), we get

$$\begin{aligned} &\sum_{t=1}^T \sum_{\theta \in \Theta} \mu_{k,t-1}(\theta) l_{k,t}(\theta) - \sum_{t=1}^T l_{k,t}(\theta^*) \\ &\leq \eta \sum_{t=1}^T \frac{2\delta}{\gamma} + \frac{\ln M}{\eta} \end{aligned} \tag{A-8}$$

$$p_{k,t}(\theta) \leq \mu_{k,t-1}(\theta) + \gamma \tag{A-9}$$

(A-9) is gotten from Step 1 of the proposed algorithm since $\psi_{k,t}(\theta) \leq 1$. Apply (A-9) in (A-8) and take expectation to obtain;

$$\begin{aligned} &\mathbb{E}_{\mathcal{F}_t} \left[\sum_{t=1}^T \sum_{\theta \in \Theta} p_{k,t}(\theta) l_{k,t}(\theta) - \sum_{t=1}^T l_{k,t}(\theta^*) \right] \\ &\leq \eta T \frac{2\delta}{\gamma} + \gamma T + \frac{\ln M}{\eta} \end{aligned}$$

choosing $\eta = \frac{\gamma^2}{\delta}$, $\gamma = \min \left\{ \frac{(\delta \ln M)^2}{T^{1/3}}, \frac{1}{2} \right\}$ and $T \geq M^3 \ln M / \delta^2$ gives the regret bound of

$$O((\delta \ln M)^2 T^{2/3})$$

ACKNOWLEDGMENT

The author thanks his kind Ph.D. supervisor Dr. Gergely Zaruba for his support of this manuscript. Also, the author thanks the anonymous reviewers of this manuscript.

REFERENCES

- [1] J. S. He, M. Han, S. Ji, T. Du, and Z. Li, "Spreading social influence with both positive and negative opinions in online networks," *Big Data Mining Anal.*, vol. 2, no. 2, pp. 100–117, Jun. 2019.
- [2] L. Muchnik, S. Aral, and S. J. Taylor, "Social influence bias: A randomized experiment," *Science*, vol. 341, no. 6146, pp. 647–651, Aug. 2013.
- [3] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan, "Influence maximization in social networks when negative opinions may emerge and propagate," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2011, pp. 379–390.
- [4] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 24, pp. 8788–8790, Jun. 2014.
- [5] S. Cohen, "Social relationships and health," *Amer. Psychologist*, vol. 59, no. 8, p. 676, 2004.
- [6] C. Chamley, *Rational Herds: Economic Models of Social Learning*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [7] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dyn. Games Appl.*, vol. 1, no. 1, pp. 3–49, Mar. 2011.
- [8] D. Acemoglu, M. Dahleh, A. Ozdaglar, and A. Tahbaz-Salehi, "Observational learning in an uncertain world," in *Proc. 49th IEEE Conf. Decis. Control (CDC)*, Dec. 2010, pp. 6645–6650.
- [9] 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.
- [10] V. Krishnamurthy, "Interactive sensing and decision making in social networks," *Found. Trends Signal Process.*, vol. 7, nos. 1–2, pp. 1–196, 2014.

- [11] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games Econ. Behavior*, vol. 76, no. 1, pp. 210–225, 2012.
- [12] X. Zhao and A. H. Sayed, "Learning over social networks via diffusion adaptation," in *Proc. Conf. Rec. Forty 6th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2012, pp. 709–713.
- [13] A. Sayed, "Adaptation, learning, and optimization over networks," *Found. Trends Mach. Learn.*, vol. 7, nos. 4–5, pp. 311–801, 2014.
- [14] A. H. Sayed, "Adaptive networks," *Proc. IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.
- [15] B. Ying and A. H. Sayed, "Learning by weakly-connected adaptive agents," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 5788–5792.
- [16] H. Salami, B. Ying, and A. H. Sayed, "Social learning over weakly connected graphs," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 3, no. 2, pp. 222–238, Jun. 2017.
- [17] H. Salami, B. Ying, and A. H. Sayed, "Diffusion social learning over weakly-connected graphs," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 4119–4123.
- [18] H. Salami, B. Ying, and A. H. Sayed, "Belief control strategies for interactions over weakly-connected graphs," 2018, *arXiv:1801.05479*. [Online]. Available: <http://arxiv.org/abs/1801.05479>
- [19] V. Matta, A. Santos, and A. H. Sayed, "Exponential collapse of social beliefs over weakly-connected heterogeneous networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 5267–5271.
- [20] B. Ying and A. H. Sayed, "Information exchange and learning dynamics over weakly connected adaptive networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1396–1414, Mar. 2016.
- [21] V. Matta, V. Bordignon, A. Santos, and A. H. Sayed, "Interplay between topology and social learning over weak graphs," *IEEE Open J. Signal Process.*, vol. 1, pp. 99–119, 2020.
- [22] M. H. DeGroot, "Reaching a consensus," *J. Amer. Statist. Assoc.*, vol. 69, no. 345, pp. 118–121, Mar. 1974.
- [23] S. Shahrampour, S. Rakhlin, and A. Jadbabaie, "Online learning of dynamic parameters in social networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2013–2021.
- [24] S. Vaswani. (2015). *Influence Maximization in Bandit and Adaptive Settings (T)*. University of British Columbia. [Online]. Available: <https://open.library.ubc.ca/collections/ubctheses/24/items/1.0166665>
- [25] G. Moscarini, M. Ottaviani, and L. Smith, "Social learning in a changing world," *Econ. Theory*, vol. 11, no. 3, pp. 657–665, 1998.
- [26] K. Dasaratha, B. Golub, and N. Hak, "Learning from neighbors about a changing state," 2018, *arXiv:1801.02042*. [Online]. Available: <http://arxiv.org/abs/1801.02042>
- [27] O. T. Odeyomi, H. M. Kwon, and D. A. Murrell, "Time-varying truth prediction in social networks using online learning," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2020, pp. 171–175.
- [28] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, "Distributed estimation of dynamic parameters: Regret analysis," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2016, pp. 1066–1071.
- [29] R. K. Kolla, K. Jagannathan, and A. Gopalan, "Collaborative learning of stochastic bandits over a social network," *IEEE/ACM Trans. Netw.*, vol. 26, no. 4, pp. 1782–1795, Aug. 2018.
- [30] O. T. Odeyomi, "Learning the truth in social networks using multi-armed bandit," *IEEE Access*, vol. 8, pp. 137692–137701, 2020.
- [31] O. T. Odeyomi, H. M. Kwon, D. A. Murrell, "Learning the truth by weakly connected agents in social networks using online learning," in *Proc. Int. Symp. Netw., Comput. Commun. (ISNCC)*, Feb. 2020, pp. 171–175.
- [32] V. V. Vazirani, *Approximation Algorithms*. New York, NY, USA: Springer, 2013.
- [33] N. Alon, N. Cesa-Bianchi, O. Dekel, and T. Koren, "Online learning with feedback graphs: Beyond bandits," in *Proc. Annu. Conf. Learn. Theory*, vol. 40. Microtome Publishing, 2015, pp. 1–13.
- [34] Z. Feng and P.-L. Loh, "Online learning with graph-structured feedback against adaptive adversaries," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 931–935.
- [35] N. Alon, N. Cesa-Bianchi, O. Dekel, and T. Koren, "Online learning with feedback graphs: Beyond bandits," 2015, *arXiv:1502.07617*. [Online]. Available: <http://arxiv.org/abs/1502.07617>



OLUSOLA TOLULOPE ODEYOMI (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electronic and electrical engineering from Obafemi Awolowo University, Nigeria, in 2011 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS, USA. He has been a Lecturer with Obafemi Awolowo University, since 2014. His research

interests include machine learning, deep learning, social networks, natural language processing, and bioinformatics. He was a recipient of the Petroleum Technology Development Fund (PTDF) Full Ph.D. Scholarship from Nigeria.

• • •