Cooperative Online Learning:
Keeping your Neighbors Updated

Nicolò Cesa-Bianchi   NICOLO.CESA-BIANCHI@UNIMI.IT
Dipartimento di Informatica & DSRC, Università degli Studi di Milano, Milano, Italy

Tommaso R. Cesari   TOMMASO.CESARI@UNIMI.IT
Dipartimento di Informatica & DSRC, Università degli Studi di Milano, Milano, Italy
ANITI, Artificial and Natural Intelligence Toulouse Institute, Toulouse, France
TSE, Toulouse School of Economics, Toulouse, France

Claire Monteleoni   CMONTEL@COLORADO.EDU
Department of Computer Science, University of Colorado, Boulder, Colorado

Editors: Aryeh Kontorovich and Gergely Neu

Abstract

We study an asynchronous online learning setting with a network of agents. At each time step, some of the agents are activated, requested to make a prediction, and pay the corresponding loss. The loss function is then revealed to these agents and also to their neighbors in the network. Our results characterize how much knowing the network structure affects the regret as a function of the model of agent activations. When activations are stochastic, the optimal regret (up to constant factors) is shown to be of order $\sqrt{\alpha T}$, where T is the horizon and α is the independence number of the network. We prove that the upper bound is achieved even when agents have no information about the network structure. When activations are adversarial the situation changes dramatically: if agents ignore the network structure, a $\Omega(T)$ lower bound on the regret can be proven, showing that learning is impossible. However, when agents can choose to ignore some of their neighbors based on the knowledge of the network structure, we prove a $\mathcal{O}(\sqrt{\overline{\chi}T})$ sublinear regret bound, where $\overline{\chi} \geq \alpha$ is the clique-covering number of the network.

Keywords: online mirror descent, regret minimization, multiagent learning

1. Introduction

Distributed asynchronous online learning settings with communication constraints arise naturally in several applications. For example, large-scale learning systems are often geographically distributed, and in domains such as finance or online advertising, each agent must serve high volumes of prediction requests. If agents keep updating their local models in an online fashion, then bandwidth and computational constraints may preclude a central processor from having access to all the observations from all sessions, and synchronizing all local models at the same time. An example in a different domain is mobile sensor networks cooperating towards a common goal, such as environmental monitoring. Sensor readings provide instantaneous, full-information feedback and energy-saving constraints favor short-range communication over long-range. Online learning algorithms distributed over spatial locations have already been proposed for problems in the field of climate informatics by McQuade and Monteleoni (2012, 2017), and have shown empirical performance advantages compared to their global (i.e., non-spatially distributed) online learning counterparts.

Motivated by these real-life applications, we introduce and analyze an online learning setting in which a network of agents solves a common online convex optimization problem by sharing feedback with their network neighbors. Agents do not have to be synchronized. At each time step, only some of the agents are requested to make a prediction and pay the corresponding loss: we call these agents "active". Because the feedback (i.e., the current loss function) received by the active agents is communicated to their neighbors, both active agents and their neighbors can use the feedback to update their local models. The lack of global synchronization implies that agents who are not requested to make a prediction get "free feedback" whenever someone is active in their neighborhood. Since in online convex optimization the sequence of loss functions is fully arbitrary, it is not clear whether this free feedback can improve the system's performance. In this paper, we characterize under which conditions and to what extent such improvements are possible.

Our goal is to control the network regret, which we define by summing the average instantaneous regret of the active agents at each time step. In order to build some intuition on this problem, consider the following two extreme cases where, for the sake of simplicity, we assume exactly one agent is active at each time step. If no communication is possible among the agents, then each agent v learns in isolation over the subset T_v of time steps when they are active. Assuming each agent runs a standard online learning algorithm with regret bounded by $\mathcal{O}(\sqrt{T})$ —such as Online Mirror Descent (OMD)— the network regret is at most of order $\sum_v \sqrt{T_v} \leq \sqrt{NT}$, where $T = \sum_v T_v$ and N is the number of agents. Next, consider a fully connected graph, where agents share their feedback with the rest of the network. Each local instance of OMD now sees the same loss sequence as the other instances, so the sequence of predictions is the same, no matter which agents are chosen to be active. The network regret is then bounded by $\mathcal{O}(\sqrt{T})$, as in the single-instance case. Our goal is to understand the regret when the communication network corresponds to an arbitrary graph G.

We consider two natural activation mechanisms for the agents: stochastic and adversarial. In the stochastic setting, at each time step t each agent v is independently active with probability q_v, where q_v is a fixed and unknown number in $[0, 1]$. Under this assumption, we show that when each agent runs OMD, the network regret is $\mathcal{O}(\sqrt{\alpha T})$, where $\alpha \leq N$ is the independence number of the communication graph. Note that this bound smoothly interpolates the two extreme cases of no communication ($\alpha = N$) and full communication ($\alpha = 1$). From this viewpoint, α can be viewed as the number of "effective instances" that are implicitly maintained by the system. It is not hard to prove that this upper bound cannot be improved upon: fix a network G and a maximal independent set in G of size α. Define $q_v = 1/\alpha$ if v belongs to the independent set and 0 otherwise. Then no two nodes that can ever become active are adjacent in G, and we reduced the problem to that of learning with α non-commmunicating agents over T/α time steps. Since there are instances of the standard online convex optimization problem on which any agent strategy has regret $\Omega(\sqrt{T})$, we obtain that the network regret must be at least of order $\alpha\sqrt{T/\alpha} = \sqrt{\alpha T}$. Note that this lower bound also applies to algorithms that have complete preliminary knowledge of the graph structure, and can choose to ignore or process any feedback coming from their neighbors. In contrast, the OMD instances used to prove the upper bound are fully oblivious both to the graph structure and to the source of their feedback (i.e., whether their agent is active as opposed to being the neighbor of an active agent).

In the adversarial activation setting, nodes are activated according to some unknown deterministic schedule. Surprisingly, under the same assumption of obliviousness about the feedback source which we used to prove the $\mathcal{O}(\sqrt{\alpha T})$ upper bound for stochastic activations, we show that on certain network topologies a deterministic schedule of activations can force a *linear* regret on *any* algorithm, thus making learning impossible. On the other hand, if agents are free to use feedback only from

a subset of their neighbors chosen with knowledge of the graph structure, then the network regret of OMD is $\mathcal{O}(\sqrt{\overline{\chi}T})$, where $\overline{\chi}$ is the clique-covering number of the communication graph. Hence, unlike the stochastic case, where the knowledge of the graph is not required to achieve optimality, in the adversarial case the ability of choosing the feedback source based on the graph structure is both a necessary and sufficient condition for sublinear regret.

The extension of the OMD analysis to a multiagent setting with communication (Theorem 4), and the lower bound for the adversarial activation setting (Theorem 7) are the main technical novelties of the paper.

2. Related Works

The study of cooperative nonstochastic online learning on networks was pioneered by Awerbuch and Kleinberg (2008), who investigated a bandit setting in which the communication graph is a clique, users are clustered so that the loss function at time t may differ across clusters, and some users may be non-cooperative. More recently, a similar line of work was pursued by Cesa-Bianchi et al. (2016), where they derive graph-dependent regret bounds for nonstochastic bandits on arbitrary networks when the loss function is the same for all nodes and the feedbacks are broadcast to the network with a delay corresponding to the shortest path distance on the graph. Although their regret bounds —like ours— are expressed in terms of the network independence number, this happens for reasons that are very different from ours, and by means of a different analysis. In their setting all agents are simultaneously active at each time step, and sharing the feedback serves the purpose of reducing the variance of the importance-weighted loss estimates. A node with many neighbors observes the current loss function evaluated at all the points corresponding to actions played by the neighbors. Hence, in that context cooperation serves to bring the bandit feedback closer to a full information setting.

In contrast, we study a full information setting in which agents get free and meaningful feedback only when they are not requested to predict.[1] Therefore, in our setting cooperation corresponds to faster learning (through the free feedback that is provided over time) *within the full information model*, as opposed to (Cesa-Bianchi et al., 2016) where cooperation *increases feedback within a single time-step*. An even more recent work considering bandit networks studies a stochastic bandit model with simultaneous activation and constraints on the amount of communication between neighbors (Martínez-Rubio et al., 2018). Their regret bounds scale with the spectral gap of the communication network. The work of Sahu and Kar (2017) investigates a different partial information model of prediction with expert advice where each agent is paired with an expert, and agents see only the loss of their own expert. The communication model includes delays, and the regret bound depends on a quantity related to the mixing time of a certain random walk on the network. Zhao et al. (2019) study a decentralized online learning setting in which losses are characterized by two components, one adversarial and another stochastic. They show upper bounds on the regret in terms of a constant representing the magnitude of the adversarial component and another constant measuring the randomness of the stochastic part.

The idea of varying the amount of feedback available to learning agents has also appeared in single-agent settings. In the *sleeping experts* model (Freund et al., 1997), different subsets of actions are available to the learner at different time steps. In our multi-agent setting, instead, actions are

---

1. Two adjacent agents that are simultaneously active exchange their feedback, but this does not bring any new information to either agent because we are in a full information setting and the loss function is the same for all nodes.

always available while the agents are occasionally sleeping. An algorithmic reduction between the two settings seems unlikely to exist because actions and agents play completely different roles in the learning process. In the *learning with feedback graphs* model (Alon et al., 2015; Mannor and Shamir, 2011), each selection of an action reveals to the learner the loss of the actions that are adjacent to it in a given graph. In our model, each time an active agent plays an action, the loss vector is revealed to the agents that are adjacent to the active learner. There is again a similarity between actions and agents in the two settings, but to the best of our knowledge there is no algorithmic reduction from multi-agent problems to single-agent problems. Yet, it should not come as a surprise that some general graph-theoretic tools —like Lemma 3— are used in the analysis of both single-agent and multi-agent models.

A very active area of research involves distributed extensions of online convex optimization, in which the global loss function is defined as a sum of local convex functions, each associated with an agent. Agents are run over the local optimization problem corresponding to their local functions and communicate with their neighborhood to find a point in the decision set approximating the loss of the best global action. This problem has been studied in various settings: distributed convex optimization —see, e.g., (Duchi et al., 2012; Scaman et al., 2018) and references therein, distributed online convex optimization (Hosseini et al., 2013), and a dynamic regret extension of distributed online convex optimization (Shahrampour and Jadbabaie, 2018). Unlike our work, these papers consider distributed extensions of OMD (and Nesterov dual averaging) based on generalizations of the consensus problems. The resulting performance bounds scale inversely in the spectral gap of the communication network.

3. Preliminaries and definitions

Let $G = (V, E)$ be a communication network, i.e., an undirected graph over a set V of N *agents*. Without loss of generality, assume $V = \{1, \ldots, N\}$. For any agent $v \in V$, we denote by $\mathcal{N}_v$ the set of nodes containing the agent v and the neighborhood $\{w \in V \mid (v, w) \in E\}$. The *independence number* α_G is the cardinality of the biggest *independent set* of G, i.e., the cardinality of the biggest subset of agents, no two of which are neighbors.

We study the following cooperative online convex optimization protocol: initially, hidden from the agents, the environment picks a sequence of subsets $S_1, S_2, \ldots \subseteq V$ of *active* agents and a sequence of differentiable convex real loss functions $\ell_1, \ell_2, \ldots$ defined on a convex decision set $\mathcal{X} \subset \mathbb{R}^d$. Then, for each time step $t \in \{1, 2, \ldots\}$,

1. each agent $v \in S_t$ predicts with $\boldsymbol{x}_t(v) \in \mathcal{X}$,
2. each agent $v \in \bigcup_{v \in S_t} \mathcal{N}_v$ receives ℓ_t as feedback,
3. the system incurs the loss $\frac{1}{|S_t|} \sum_{v \in S_t} \ell_t(\boldsymbol{x}_t(v))$ (defined as 0 when $S_t \equiv \varnothing$).

We assume each agent v runs an instance of the same online algorithm. Each instance learns a local model generating predictions $\boldsymbol{x}_t(v)$. This local model is updated whenever a feedback ℓ_t is received. We call *paid feedback* the feedback ℓ_t received by v when $v \in S_t$ (i.e., the agent is active) and *free feedback* the feedback ℓ_t received by v when $v \in \left(\bigcup_{v \in S_t} \mathcal{N}_v \right) \setminus \{S_t\}$ (i.e., the agent is not active but in the neighborhood of some active agent). The goal is to minimize the *network regret* as a function of the unknown number T of time steps,

$$R_T = \sum_{t=1}^{T} \frac{1}{|S_t|} \sum_{v \in S_t} \ell_t(\boldsymbol{x}_t(v)) - \inf_{\boldsymbol{x} \in \mathcal{X}} \sum_{t=1}^{T} \ell_t(\boldsymbol{x}) \tag{1}$$

Note that only the losses of active agents contribute to the network regret.

4. Online Mirror Descent

Algorithm 1: Online Mirror Descent

Input: σ_t-strongly convex regularizers $g_t \colon \mathcal{X} \to \mathbb{R}$ for $t \in \{1, 2, \ldots\}$
Initialization: $\boldsymbol{\theta}_1 = \mathbf{0} \in \mathbb{R}^d$
for $t = 1, 2, \ldots$ **do**
$\quad$ choose $\boldsymbol{w}_t = \nabla g_t^*(\boldsymbol{\theta}_t)$
$\quad$ observe $\nabla \ell_t(\boldsymbol{w}_t) \in \mathbb{R}^d$
$\quad$ update $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \nabla \ell_t(\boldsymbol{w}_t)$

We now review the standard Online Mirror Descent algorithm (OMD) —see Algorithm 1— and its analysis. Let $f \colon \mathcal{X} \to \mathbb{R}$ be a convex function. We say that $f^* \colon \mathbb{R}^d \to \mathbb{R}$ is the *convex conjugate* of f if $f^*(\boldsymbol{x}) = \sup_{\boldsymbol{w} \in \mathcal{X}} (\boldsymbol{x} \cdot \boldsymbol{w} - f(\boldsymbol{w}))$. We say that f is *σ-strongly convex* on $\mathcal{X}$ with respect to a norm $\|\cdot\|$ if there exists $\sigma \geq 0$ such that, for all $\boldsymbol{u}, \boldsymbol{w} \in \mathcal{X}$ we have $f(\boldsymbol{u}) \geq f(\boldsymbol{w}) + \nabla f(\boldsymbol{w}) \cdot (\boldsymbol{u} - \boldsymbol{w}) + \frac{\sigma}{2} \|\boldsymbol{u} - \boldsymbol{w}\|^2$. The following well-known result can be found in (Shalev-Shwartz, 2012, Lemma 2.19 and subsequent paragraph).

Lemma 1 *Let $f \colon \mathcal{X} \to \mathbb{R}$ be a strongly convex function on $\mathcal{X}$. Then the convex conjugate f^* is everywhere differentiable on $\mathbb{R}^d$.*

The following result —see, e.g., (Orabona et al., 2015, bound (6) in Corollary 1 with F set to zero)— shows an upper bound on the regret of OMD.

Theorem 2 *Let $g \colon \mathcal{X} \to \mathbb{R}$ be a differentiable function σ-strongly convex with respect to $\|\cdot\|$. Then the regret of OMD run with $g_t = \frac{\sqrt{t}}{\eta} g$, for $\eta > 0$, satisfies*

$$\sum_{t=1}^T \ell_t(\boldsymbol{x}_t) - \inf_{\boldsymbol{x} \in \mathcal{X}} \sum_{t=1}^T \ell_t(\boldsymbol{x}) \leq \frac{D}{\eta}\sqrt{T} + \frac{\eta}{2\sigma} \sum_{t=1}^T \frac{1}{\sqrt{t}} \|\nabla \ell_t\|_*^2$$

where $D = \sup g$ and $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. If $\sup \|\nabla \ell_t\|_* \leq L$, then choosing $\eta = \sqrt{2\sigma D}/L$ gives $R_T \leq L\sqrt{2DT/\sigma}$.

A popular instance of OMD is the standard online gradient descent algorithm, corresponding to choosing $\mathcal{X}$ equal to a closed Euclidean ball centered at the origin, and setting $g = \frac{1}{2} \|\cdot\|^2$ for all t, where $\|\cdot\|$ is the Euclidean norm. Another instance is the Hedge algorithm for prediction with expert advice, corresponding to choosing $\mathcal{X}$ equal to the probability simplex, and setting $g(\boldsymbol{p}) = \sum_i p_i \ln p_i$.

5. Stochastic Activations

In this section we analyze the performance of OMD when the sets S_t of active agents are chosen stochastically. As discussed in the introduction, in this setting we do not require any ad-hoc interface between each OMD instance and the rest of the network. In particular, we make the following assumption.

Assumption 1 (Oblivious network interface) *An online algorithm A is run with an oblivious network interface if for each agent v it holds that:*

1. *v runs an instance A_v of A*
2. *A_v uses the same initialization and learning rate as the other instances*
3. *A_v makes predictions and updates while being oblivious to whether $v \in S_t$ or $v \in \bigcup_{u \in S_t} \mathcal{N}_u \setminus S_t$*

This assumption implies that each instance is oblivious to both the network topology and the location of the agent in the network. Moreover, instances make an update whenever they have the opportunity to do so, (i.e., whenever they or some of their neighbors are active). The purpose of this assumption is to show that communication might help OMD even without any network-specific tuning. In concrete applications, one might use ad-hoc OMD variants that rely on the knowledge of the task at hand, and decrease the regret even further. However, the lower bound proven in Section 6 shows that the regret cannot be decreased significantly even when agents have full knowledge of the graph.

We start by considering a slightly simplified stochastic activation setting, where only a single agent is activated at each time step (i.e., $|S_t| = 1$ for all t). The more general stochastic case is analyzed at the end of this section.

We assume that the active agents $v_1, v_2, \ldots$ are drawn i.i.d. from an unknown fixed distribution q on V. The main result of this section is an upper bound on the regret of the network when all agents run the basic OMD (Algorithm 1) with an oblivious network interface. We show that in this case the network achieves the same regret guarantee as the single-agent OMD (Theorem 2) multiplied by the square root of independence number of the communication network.

Before proving the main result, we state a combinatorial lemma that allows to upper bound the sum of a ratio of probabilities over the vertices of an undirected graph with the independence number of the graph (Griggs, 1983; Mannor and Shamir, 2011). The proof is included for completeness.

Lemma 3 *Let $G = (V, E)$ be an undirected graph with independence number α_G and q any probability distribution on V such that $Q_v = \sum_{w \in \mathcal{N}_v} q_v > 0$ for all $v \in V$. Then*

$$\sum_{v \in V} \frac{q_v}{Q_v} \le \alpha_G$$

Proof Initialize $V_1 = V$, fix $w_1 \in \arg\min_{w \in V_1} Q_w$, and denote $V_2 = V \setminus \mathcal{N}_{w_1}$. For $k \ge 2$ fix $w_k \in \arg\min_{w \in V_k} Q_w$ and shrink $V_{k+1} = V_k \setminus \mathcal{N}_{w_k}$ until $V_{k+1} = \varnothing$. Being G undirected, we have $w_k \notin \bigcup_{s=1}^{k-1} \mathcal{N}_{w_s}$, therefore the number m of times that an action can be picked this way is upper bounded by α_G. Denoting $\mathcal{N}'_{w_k} = V_k \cap \mathcal{N}_{w_k}$, this implies

$$\sum_{v \in V} \frac{q_v}{Q_v} = \sum_{k=1}^{m} \sum_{v \in \mathcal{N}'_{w_k}} \frac{q_v}{Q_v} \le \sum_{k=1}^{m} \sum_{v \in \mathcal{N}'_{w_k}} \frac{q_v}{Q_{w_k}} \le \sum_{k=1}^{m} \frac{\sum_{v \in \mathcal{N}_{w_k}} q_v}{Q_{w_k}} = m \le \alpha_G$$

∎

The following holds for any differentiable function $g \colon \mathcal{X} \to \mathbb{R}$, σ-strongly convex with respect to some norm $\|\cdot\|$.

Theorem 4 *Consider a network $G = (V, E)$ of N agents and assume $S_t = \{v_t\}$ for each t, where v_t is drawn i.i.d. from some fixed and unknown distribution on V. If all agents run OMD with an*

oblivious network interface and using $g_t = \frac{\sqrt{t}}{\eta} g$, for $\eta > 0$, then the network regret satisfies

$$\mathbb{E}[R_T] \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{\alpha_G T}$$

where $D \geq \sup g$, $L \geq \sup \|\nabla \ell_t\|_*$, and $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. In particular, choosing $\eta = \sqrt{2\sigma D}/L$ gives $\mathbb{E}[R_T] \leq L\sqrt{2D\alpha_G T/\sigma}$.

Proof Fix $\boldsymbol{x} \in \mathcal{X}$, any sequence of realizations $v_1, \ldots, v_T$, and any v in the support $V' \subset V$ of the activation distribution q. Note that the OMD instance run by v, makes an update at time t only when $v \in \mathcal{N}_{v_t}$. Hence, letting $r_t(v) = \ell_t(\boldsymbol{x}_t(v)) - \ell_t(\boldsymbol{x})$ and applying Theorem 2,

$$\sum_{t=1}^{T} r_t(v) \mathbb{I}\{v \in \mathcal{N}_{v_t}\} \leq \frac{D}{\eta} \sqrt{T_v} + \frac{\eta L^2}{2\sigma} \sum_{t=1}^{T} \frac{\mathbb{I}\{v \in \mathcal{N}_{v_t}\}}{\sqrt{\sum_{s=1}^{t} \mathbb{I}\{v \in \mathcal{N}_{v_s}\}}} \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{T_v} \quad (2)$$

where $T_v = \sum_{t=1}^{T} \mathbb{I}\{v \in \mathcal{N}_{v_t}\}$, the addends after the first inequality are intended to be null when the denominator is zero, and we used $\sum_{s=1}^{T_v} s^{-1/2} \leq 2\sqrt{T_v}$. Note that $r_t(v)$ is independent of v_t, as it only depends on the subset of v_s, $s \in \{1, \ldots, t-1\}$, such that $v \in \mathcal{N}_{v_s}$. Denote by Q_v the probability $\mathbb{P}(v \in \mathcal{N}_{v_t}) = \sum_{w \in \mathcal{N}_v} q(w) > 0$. Let $\mathcal{F}_{t-1}$ be the σ-algebra generated by $\{v_1, \ldots, v_{t-1}\}$. Since Q_v is independent of t, $\mathbb{P}(v \in \mathcal{N}_{v_t} \mid \mathcal{F}_{t-1}) = Q_v$. Therefore, taking expectation with respect to $v_1, \ldots, v_T$ on both sides of (2), using $\mathbb{E}[T_v] = Q_v T$, and applying Jensen's inequality, yields

$$\mathbb{E}\left[\sum_{t=1}^{T} r_t(v) Q_v \right] \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{Q_v T} \quad (3)$$

Now, letting $R_T(\boldsymbol{x}) = \sum_{t=1}^{T} r_t(v_t)$, we have that $\mathbb{E}[R_T(\boldsymbol{x})]$ is equal to

$$\mathbb{E}\left[\sum_{v \in V'} \sum_{t=1}^{T} r_t(v) \mathbb{I}\{v_t = v\} \right] = \mathbb{E}\left[\sum_{v \in V'} \sum_{t=1}^{T} r_t(v) \mathbb{E}\big[\mathbb{I}\{v_t = v\} \mid \mathcal{F}_{t-1}\big] \right] = \sum_{v \in V'} q_v \mathbb{E}\left[\sum_{t=1}^{T} r_t(v) \right]$$

Dividing both sides of (3) by $Q_v > 0$, we can write

$$\mathbb{E}[R_T(\boldsymbol{x})] \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sum_{v \in V'} q_v \sqrt{\frac{T}{Q_v}} \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{T \sum_{v \in V'} \frac{q_v}{Q_v}} \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{\alpha T}$$

where in the last two inequalities we applied Jensen's inequality and Lemma 3. Observing that $\mathbb{E}[R_T] = \sup_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}[R_T(\boldsymbol{x})]$ and recalling that $\boldsymbol{x}$ was chosen arbitrarily in $\mathcal{X}$ concludes the proof. $\blacksquare$

Note that the proof of the previous result gives a tighter upper bound on the network regret in terms of the independence number $\alpha' \leq \alpha$ of the subgraph induced by the support V' of q.

Next, we consider the setting in which we allow the activation of more than one agent per time step. At the beginning of the process, the environment draws an i.i.d. sequence of Bernoulli random variables $X_1(v), X_2(v), \ldots$ with some unknown fixed parameter $q_v \in [0, 1]$ for each agent $v \in V$. The active set at time t is then defined as $S_t = \{v \in V \mid X_t(v) = 1\}$. Note that, unlike the previous setting, now $\sum_{v \in V} q_v \neq 1$ in general.

7

We state an upper bound on the regret that the network incurs if all agents run OMD with an oblivious network interface (for a proof, see Appendix A). Our upper bound is expressed in terms of a constant depending on the probabilities of activating each agent and such that $Q \leq 1.6(\alpha_G + 1)$. The result holds for any differentiable function $g \colon \mathcal{X} \to \mathbb{R}$, σ-strongly convex with respect to some norm $\|\cdot\|$.

Theorem 5 *Consider a network $G = (V, E)$ of N agents. Assume that, at each time step t each agent v is independently activated with probability $q_v \in [0, 1]$. If all agents run OMD with an oblivious network interface and using $g_t = \frac{\sqrt{t}}{\eta} g$, for $\eta > 0$, the network regret satisfies*

$$\mathbb{E}[R_T] \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{QT}$$

for some nonnegative $Q \leq 1.6(\alpha_G + 1)$, $D \geq \sup g$, and $L \geq \sup \|\nabla \ell_t\|_*$, where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. In particular, choosing $\eta = \sqrt{2\sigma D}/L$ gives $\mathbb{E}[R_T] \leq L\sqrt{2DQT/\sigma}$.

In order to compare the previous upper bound to Theorem 4, consider the case $q_v = q$ for all $v \in V$. Without loss of generality, assume $q > 0$ (the regret is zero when q vanishes). Then

$$Q = Q(q) = \frac{1}{N} \sum_{v \in V} \frac{1 - (1 - q)^N}{1 - (1 - q)^{|\mathcal{N}_v|}}$$

(for a proof, see Theorem 10 in Appendix A and proceed as in the proof of Lemma 9). A direct computation of the sign of the first derivative of the addends $q \mapsto \frac{1 - (1-q)^N}{1 - (1-q)^{|\mathcal{N}_v|}}$ shows that these functions are decreasing in q, hence $1 = \lim_{q \to 1^-} Q(q) \leq Q \leq \lim_{q \to 0^+} Q(q) = \sum_{v \in V} \frac{1}{|\mathcal{N}_v|} \leq \alpha_G$ where the last inequality follows by Lemma 3. Note that the lower bound $Q \geq 1$ is attained if the probabilities of picking agents at each time step are all 1. In this case all agents are activated at each time step, the graph structure over the set of agents becomes irrelevant and the model reduces to a single-agent problem. The inequality $Q(q) \leq \alpha_G$ is not a coincidence due to the constant q. Indeed, one can prove that this is always the case, up to a small constant factor (for a proof., see Lemma 11 in Appendix A).

The previous results shows that paying the average price of multiple activations is never worse (up to a small constant factor) than drawing a single agent per time step, and it can be significantly better. A similar argument shows a tighter bound $Q \leq \max\{3, \alpha_G\}$ when the activation probabilities satisfy $\sum_{v \in V} q_v = 1$, which allows to recover the upper bound on the network regret proven in Theorem 4. This is consistent with the intuition that —in expectation— picking a single agent at random according to a distribution $q = (q_1, \ldots, q_N)$ is the same as picking each v independently with probability q_v. Similarly to the case $|S_t| = 1$, the previous result gives a tighter upper bound on the network regret in terms of the independence number $\alpha' \leq \alpha$ of the subgraph induced by the subset V' of V containing all agents v with $q_v > 0$. Note that the setting discussed in this section smoothly interpolates between the single-agent setting ($q_v = 1$ for all v), cooperative learning with one agent stochastically activated at each time step ($\sum_v q_v = 1$), and beyond ($\sum_v q_v < 1$), where a non trivial fraction of the total number rounds is skipped.

6. Lower Bound for Stochastic Activations

In this section we show that, for any communication network G with stochastic agent activations, the best possible regret rate is of order $\Omega(\sqrt{\alpha_G T})$. This holds even when agents are not restricted to

use an oblivious network interface. The idea is that if the distribution from which active agents are drawn is supported on an independent set of cardinality α_G, then the problem reduces to that of an edgeless graph with α_G agents.

Theorem 6 *There exists a convex decision set in $\mathbb{R}^d$ such that, for each communication network G and for arbitrary (and possibly different) online learning algorithms run by the agents, $\mathbb{E}[R_T] = \Omega(\sqrt{\alpha T})$ for some sequence $(S_1, \ell_1), \ldots, (S_T, \ell_T)$, where $S_t = \{v_t\}$, v_t is drawn i.i.d. from some fixed distribution on V, and the expectation is taken with respect to the random draw of the $v_1, \ldots, v_T$.*

Proof We sketch the proof for the case $|S_t| = 1$. Let $\mathcal{X}$ be the probability simplex in $\mathbb{R}^d$. Let $G = (V, E)$ be any communication graph and α its independence number. We consider linear losses defined on $\mathcal{X}$. Let q be the uniform distribution over a maximal independent set $A = \{a_1, \ldots, a_\alpha\} \subset V$. Fix now any cooperative online linear optimization algorithm for this setting. Since each active agent v_t belongs to A for all $t \in \{1, \ldots, T\}$ with probability 1, it suffices to analyze the updates of the algorithm for these agents. Indeed, no other agent incurs any loss at any time-step. Since A is an independent set, each agent a_i makes an update at round t if and only if $v_t = a_i$. This happens with probability $q(a_i) = 1/\alpha$, independently of t. Each agent a_i is therefore running an independent single-agent online linear optimization problem for an average of T/α rounds. It is well-known (Hazan, 2016, Theorem 3.2) that any algorithm for online linear optimization on the simplex with losses bounded in $[0, 1]$ incurs $\Omega(\sqrt{T/\alpha})$ regret over T/α rounds in the worst case. Consequently, the regret of the network satisfies $R_T = \Omega(\alpha\sqrt{T/\alpha}) = \Omega(\sqrt{\alpha T})$. ∎

An analogous lower bound can be proven for the case of multiple agent activations per time step. Indeed, define $q_v = 1/\alpha$ for each agent v belonging to some fixed maximal independent set and $q_v = 0$ otherwise. This again leads to α independent single-agent online linear optimization problems for an average of T/α rounds each, and an argument similar to the one in the proof of Theorem 6 gives the result.

7. Adversarial Activations

In this section we drop the stochasticity assumption on the agents' activations and focus on the case where active agents are picked from V by an adversary. The goal is to control the regret (1) for *any* individual sequence of pairs $(\ell_1, S_1), (\ell_2, S_2), \ldots$ where ℓ_t is a convex loss and $S_t \subseteq V$, without any stochastic assumptions on the mechanism generating these pairs. For the rest of this section, we focus on the special case where $|S_t| = 1$ for all t and denote by v_t the active node at time t.

We start by proving that learning with adversarial activations is impossible if we use an oblivious network interface. We prove this result in the setting of prediction with expert advice with two actions and binary losses, a special case of online convex optimization. The idea of the lower bound is that if the communication network is a star graph, the environment is able to make both actions look equally good to all peripheral agents, even if one of the two actions is actually better than the other. This is done by drawing the good action at random, then activating an agent depending on the outcome of the draw. For a small fraction of the times the good action has loss one, the central agent is activated. Since the central agent shares feedback with all peripheral agents, we can amplify this loss by a factor of N, and thus make the good action look to all peripheral agents as bad as the bad action.

Theorem 7 *For each $N > 3$ there exists a convex decision set in $\mathbb{R}^2$ and a graph G with N vertices such that, whenever N agents are run on G using instances of any online learning algorithm with an oblivious network interface, then $R_T = \Omega(T)$ for some sequence $(\ell_1, v_1), \ldots, (\ell_T, v_T)$ of convex losses and active agents.*

Proof Fix $N > 3$ and let $\mathcal{X}$ be the probability simplex in $\mathbb{R}^2$. Let $G = (V, E)$ be the star graph with central agent a_0, and peripheral agents $a_1, \ldots, a_{N-1}$. Because our losses are linear on $\mathcal{X}$, the online convex optimization problem is equivalent to prediction with expert advice with two experts (or actions), and we may denote losses using loss vectors $\ell_t = (\ell_t(1), \ell_t(2))$ where 1 and 2 index the actions. A *good action* $J \in \{1, 2\}$ is drawn uniformly at random. Denote the other one (i.e., the *bad* one) by J_B. To keep notation tidy, we define loss vectors by $\ell_t = (\ell_t(J), \ell_t(J_B))$. Fix any $\varepsilon \in \left(0, \frac{N-1}{2(N-2)}\right)$. The loss vectors ℓ_t are drawn i.i.d. at random, according to the following joint distribution:

$$\mathbb{P}(\ell_t = (0, 1)) = \frac{1}{2} \qquad \mathbb{P}(\ell_t = (1, 0)) = \frac{1}{2} - \varepsilon + \frac{\varepsilon}{N-1} \qquad \mathbb{P}(\ell_t = (0, 0)) = \varepsilon - \frac{\varepsilon}{N-1}$$

Recall that only a single agent v_t is active at any time. At each time step t, the adversary decides whether to activate the central agent a_0 or a peripheral agent, depending on the realization of ℓ_t. If $\ell_t(J) = 0$, then a random peripheral agent is activated. Otherwise, we set

$$\mathbb{P}(\ell_t = (1, 0), v_t = a_0) = \frac{\varepsilon}{N-1} \quad \text{and} \quad \mathbb{P}(\ell_t = (1, 0), v_t = a_i) = \frac{1/2 - \varepsilon}{N-1} \quad a_1, \ldots, a_{N-1}$$

Note that when $v_t = a_0$, then all peripheral agents receive feedback ℓ_t. Similarly, when a peripheral agent is active at time t, then a_0 receives feedback ℓ_t. For $b_1, b_2 \in \{0, 1\}$, let $E(a_i, b_1, b_2)$ be the event: agent a_i receives the loss vector $\ell_t = (b_1, b_2)$ as feedback. The following statements then hold for each peripheral agent a_i,

$$\mathbb{P}(E(a_i, 0, 1)) = \frac{1/2}{N-1} \qquad \mathbb{P}(E(a_i, 0, 0)) = \frac{\varepsilon}{N-1} - \frac{\varepsilon}{(N-1)^2}$$

$$\mathbb{P}(E(a_i, 1, 0)) = \frac{1/2 - \varepsilon}{N-1} + \frac{\varepsilon}{N-1} = \frac{1/2}{N-1}$$

Hence, each instance managed by a peripheral agent observes loss vectors $(1, 0)$ and $(0, 1)$ with the same probability proportional to $1/2$, and loss vector $(0, 0)$ with probability proportional to $\varepsilon(N-1)/(N-2)$. Since the network interface is oblivious, the instance cannot distinguish between paid and free feedback (which would reveal the good action), and incurs an expected loss of $1/2$ each time $\ell_t \in \{(0, 1), (1, 0)\}$. Using the fact that a peripheral agent is active when $\ell_t \in \{(0, 1), (1, 0)\}$ with probability $1/2 + 1/2 - \varepsilon = 1 - \varepsilon$, the system's expected total loss is at least $\frac{1-\varepsilon}{2}T$ (we lower bound the loss of the central agent by zero). Since the expected loss of J is $\left(1/2 - \varepsilon + \frac{\varepsilon}{N-1}\right)T$, the expected regret of the system satisfies

$$\mathbb{E}[R_T] \geq \left(\frac{1 - \varepsilon}{2} - \frac{1}{2} + \varepsilon - \frac{\varepsilon}{N-1}\right)T \geq \frac{T}{8}$$

where we picked $\varepsilon = (N-1)/(N-2)$ and used $(N-3)/(N-2) \geq 1/2$ in the last inequality. Therefore, there exists some sequence $(\ell_1, S_1), \ldots, (\ell_T, S_T)$ such that $R_T \geq T/8$, concluding the proof. $\blacksquare$

We complement the above negative result by showing that when algorithms are run without the oblivious network interface, and agents are free to use feedback only from a subset of their neighbors chosen with knowledge of the graph structure, then the network regret of OMD is $\mathcal{O}(\sqrt{\overline{\chi}_G T})$. The quantity $\overline{\chi}_G$ is the clique-covering number of the communication graph G, which corresponds to the smallest cardinality of a clique cover of G (a clique cover is a partition of the vertices such that the nodes in every element of the partition form a clique in the graph). The intuition behind this result is simple: fix a clique cover and let the agents in the same clique of the cover know each other. Now, if each agent ignores all feedback coming from agents in other cliques, then the agents in the same clique make exactly the same sequence of prediction and updates. Therefore, the effective number of OMD instances that are being run is equal to $\overline{\chi}_G$.

The following result holds for any differentiable function $g \colon \mathcal{X} \to \mathbb{R}$, σ-strongly convex with respect to some norm $\|\cdot\|$.

Theorem 8 *Consider a network $G = (V, E)$ of N agents, a clique cover $\{K_1, \ldots, K_M\}$ where $M = \overline{\chi}_G$, and let $K(v)$ be the unique element of the cover which each $v \in V$ belongs to. For any sequence $v_1, v_2, \ldots \in V$ of active agents, assume each agent $v \in V$ runs OMD using $g_t = \frac{\sqrt{t}}{\eta} g$ (with $\eta > 0$) while making updates only at those time steps t such that $v_t \in K(v)$. Then the network regret satisfies*

$$\mathbb{E}[R_T] \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{\overline{\chi}_G T}$$

where $D \geq \sup g$, $L \geq \sup \|\nabla \ell_t\|_*$, and $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. In particular, choosing $\eta = \sqrt{2\sigma D}/L$ gives $\mathbb{E}[R_T] \leq L\sqrt{2D\overline{\chi}_G T/\sigma}$.

Proof Fix any clique K_c and any $v \in K_c$. Let T_c be the time steps such that $v_t \in K_c$. Since each agent $v \in K_c$ ignores the feedback coming from other cliques, the nodes in K_c perform exactly the same updates, and therefore make exactly the same predictions. This means that, for any $t \in T_c$, the predictions in the set $\{\boldsymbol{x}_t(v) \mid v \in K_c\}$ are all equal to the same common value denoted by $\boldsymbol{x}_t(K_c)$. Fix any $\boldsymbol{x} \in \mathcal{X}$ and, for any $t \in T_c$, let $r_t(K_c) = \ell_t(\boldsymbol{x}_t(K_c)) - \ell_t(\boldsymbol{x})$. By Theorem 2 we have that

$$\sum_{t \in T_c} r_t(K_c) \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{T_c} \,.$$

Therefore, recalling that $r_t(v_t) = \ell_t(\boldsymbol{x}_t(v_t)) - \ell_t(\boldsymbol{x})$ and using Jensen's inequality,

$$\sum_{t=1}^{T} r_t(v_t) = \sum_{c=1}^{\overline{\chi}_G} \sum_{t \in T_c} r_t(K_c) \leq \sum_{c=1}^{\overline{\chi}_G} \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{T_c} \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{\overline{\chi}_G T}$$

concluding the proof. $\blacksquare$

Theorems 7 and 8 show that with adversarial activations the knowledge of the graph is crucial for learning (e.g., for achieving sublinear regret). Since $\overline{\chi}_G \geq \alpha_G$, it is not clear whether the better rate $\sqrt{\alpha_G T}$ can be proven in the adversarial activation setting when agents do not use the oblivious network interface.

8. Conclusions

We introduced a cooperative learning setting in which agents, sitting on the nodes of a communication network, run instances of an online learning algorithm with the common goal of minimizing their regret. In order to investigate how the knowledge of the graph topology affects regret in cooperative online learning under different activation mechanisms, we introduced the notion of oblivious network interface. This prevents agents from doing any network-specific tuning or even accessing their neighborhood structure. When activations are stochastic, we showed that sharing losses among neighbors is enough to guarantee optimal regret rates even with the oblivious network interface. Surprisingly, when activations are adversarial the situation changes completely. There exist problem instances in which any algorithm that runs with the oblivious network interface suffers linear regret. In this case knowing graph structure is not only necessary to perform optimally, but even to have sublinear regret.

Other interesting variants of this settings could be studied in the future. For example, at the beginning of each round, active agents could be allowed to ask the predictions of some of their neighbors, and base their prediction upon it. In this case, we conjecture that the optimal regret rate would scale with the dominating number δ_G of the graph, which is always smaller or equal to the independence number.

Acknowledgments

References

Noga Alon, Nicolo Cesa-Bianchi, Ofer Dekel, and Tomer Koren. Online learning with feedback graphs: Beyond bandits. In *Annual Conference on Learning Theory*, volume 40. Microtome Publishing, 2015.

Baruch Awerbuch and Robert Kleinberg. Competitive collaborative learning. *Journal of Computer and System Sciences*, 74(8):1271–1288, 2008.

Nicolo Cesa-Bianchi, Claudio Gentile, Yishay Mansour, and Alberto Minora. Delay and cooperation in nonstochastic bandits. *JMLR Workshop and Conference Proceedings (COLT 2016)*, 49:605–622, 2016.

John C Duchi, Alekh Agarwal, and Martin J Wainwright. Dual averaging for distributed optimization: Convergence analysis and network scaling. *IEEE Transactions on Automatic control*, 57(3): 592–606, 2012.

Yoav Freund, Robert E Schapire, Yoram Singer, and Manfred K Warmuth. Using and combining predictors that specialize. In *In Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing*. Citeseer, 1997.

Jerrold R Griggs. Lower bounds on the independence number in terms of the degrees. *Journal of Combinatorial Theory, Series B*, 34(1):22–39, 1983.

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

Saghar Hosseini, Airlie Chapman, and Mehran Mesbahi. Online distributed optimization via dual averaging. In *52nd Annual IEEE Conference on Decision and Control (CDC)*, pages 1484–1489. IEEE, 2013.

Shie Mannor and Ohad Shamir. From bandits to experts: On the value of side-observations. In *Advances in Neural Information Processing Systems*, pages 684–692, 2011.

David Martínez-Rubio, Varun Kanade, and Patrick Rebeschini. Decentralized cooperative stochastic multi-armed bandits. *arXiv preprint arXiv:1810.04468*, 2018.

Scott McQuade and Claire Monteleoni. Global climate model tracking using geospatial neighborhoods. In *Proc. Twenty-Sixth AAAI Conference on Artificial Intelligence, Special Track on Computational Sustainability and AI*, pages 335–341, 2012.

Scott McQuade and Claire Monteleoni. Spatiotemporal global climate model tracking. *Large-Scale Machine Learning in the Earth Sciences; Data Mining and Knowledge Discovery Series. Srivastava, A., Nemani R., Steinhaeuser, K. (Eds.), CRC Press, Taylor & Francis Group*, 2017.

Francesco Orabona, Koby Crammer, and Nicolo Cesa-Bianchi. A generalized online mirror descent with applications to classification and regression. *Machine Learning*, 99(3):411–435, 2015.

Anit Kumar Sahu and Soummya Kar. Dist-Hedge: A partial information setting based distributed non-stochastic sequence prediction algorithm. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 528–532. IEEE, 2017.

Kevin Scaman, Francis Bach, Sébastien Bubeck, Laurent Massoulié, and Yin Tat Lee. Optimal algorithms for non-smooth distributed optimization in networks. In *Advances in Neural Information Processing Systems*, pages 2745–2754, 2018.

Shahin Shahrampour and Ali Jadbabaie. Distributed online optimization in dynamic environments using mirror descent. *IEEE Transactions on Automatic Control*, 63(3):714–725, 2018.

Shai Shalev-Shwartz. Introduction to online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

Yawei Zhao, Chen Yu, Peilin Zhao, and Ji Liu. Decentralized online learning: Take benefits from others' data without sharing your own to track global trend. *arXiv preprint arXiv:1901.10593*, 2019.

Appendix A. Stochastic Activations: Multiple Agents

In this section we present all missing results related to the stochastic activation model with multiple activations per time step. Recall that, at the beginning of the process, the environment draws an i.i.d. sequence of Bernoulli random variables $X_1(v), X_2(v), \ldots$ with some unknown fixed parameter $q_v \in [0, 1]$ for each agent $v \in V$. The active set at time t is then defined as $S_t = \{v \in V \mid X_t(v) = 1\}$. Note that, unlike when only one agent is active at each time step, now $\sum_{v \in V} q_v \neq 1$ in general. Before the main result, we give some definitions and prove a technical combinatorial lemma that is leveraged in the analysis.

Denote by V' the set of all agents $v \in V$ such that $q_v > 0$. For each $v \in V'$, let

$$c_v = \sum_{S \subset \{1, \ldots, N\} \setminus \{v\}} \frac{\lambda_{S,v}}{1 + |S|} \tag{4}$$

where the convex coefficients $\lambda_{S,v}$ are defined by

$$\left(\prod_{w=1}^{N} q_w \right) \left(\prod_{u \in \{1, \ldots, N\} \setminus (\{v\} \cup S)} (1 - q_u) \right)$$

Let also Q_v be the probability

$$\mathbb{P}\left(v \in \bigcup_{w \in S_t} \mathcal{N}_w \right) = 1 - \prod_{w \in \mathcal{N}_v} (1 - q_w) > 0 \tag{5}$$

that agent v is updated at time t —note that Q_v is independent of t.

Lemma 9 *Let $X(1), \ldots, X(m)$ be independent Bernoulli random variables with strictly positive parameters $q_1, \ldots, q_m$ respectively. Then, for all $v \in \{1, \ldots, m\}$,*

$$\mathbb{E}\left[\frac{X(v)}{\sum_{w=1}^{m} X(w)} \right] = q_v c_v$$

where we define $X(v)/\sum_{w=1}^{m} X(w) = 0$ when $X(v) = 0$.

Proof Fix any $v \in \{1, \ldots, m\}$. Let S_v be the set $\{1, \ldots, m\} \setminus \{v\}$ and let $\mathcal{F}_v$ be the σ-algebra generated by $\{X(w) \mid w \in S_v\}$. Then

$$\mathbb{E}\left[\frac{X(v)}{\sum_{w=1}^{m} X(w)} \right] = \mathbb{E}\left[\mathbb{E}\left[\frac{X(v)}{\sum_{w=1}^{m} X(w)} \Big| \mathcal{F}_v \right] \right] = q_v \mathbb{E}\left[\frac{1}{1 + \sum_{w \in S_v} X(w)} \right]$$

Denote the last expectation by c_v. Since for all $x \neq 0$, $\int_0^\infty e^{-tx} \mathrm{d}t = \frac{1}{x}$, Fubini's theorem yields

$$
\begin{aligned}
c_v &= \int_0^\infty \mathbb{E}\left[e^{-t\left(1 + \sum_{w \in S_v} X(w)\right)} \right] \mathrm{d}t \\
&= \int_0^\infty e^{-t} \prod_{w \in S_v} \mathbb{E}\left[e^{-tX_t(w)} \right] \mathrm{d}t \\
&= \int_0^\infty e^{-t} \prod_{w \in S_v} (q_w e^{-t} + 1 - q_w) \mathrm{d}t \\
&= \int_0^1 \prod_{w \in S_v} (q_w x + 1 - q_w) \mathrm{d}x \\
&= \int_0^1 \sum_{S \subset S_v} x^{|S|} \left(\prod_{w \in S} q_w \right) \left(\prod_{u \in S_v \setminus S} (1 - q_u) \right) \mathrm{d}x
\end{aligned}
$$

Now set $\lambda_{S,v} = \left(\prod_{w \in S} q_w \right)\left(\prod_{S_v \setminus S} (1 - q_u) \right)$ and note that $\sum_{S \subset S_v} \lambda_{S,v} = \prod_{w \in S_v} (q_w + 1 - q_w) = 1$. Substituting $\lambda_{S,v}$ in the last identity gives

$$
c_v = \sum_{S \subset S_v} \lambda_{S,v} \int_0^1 x^{|S|} \mathrm{d}x = \sum_{S \subset S_v} \frac{\lambda_{S,v}}{1 + |S|}
$$

∎

We now give an upper bound on the regret that the network incurs if all agents run OMD with an oblivious network interface. Our upper bound is expressed in terms of a constant depending on the probabilities of activating each agent and such that $Q \leq 1.6(\alpha_G + 1)$. The result holds for any differentiable function $g \colon \mathcal{X} \to \mathbb{R}$, σ-strongly convex with respect to some norm $\|\cdot\|$.

Theorem 10 *Consider a network $G = (V, E)$ of N agents. Assume that, at each time step t each agent v is independently activated with probability $q_v \in [0, 1]$. If all agents run OMD with an oblivious network interface and using $g_t = \frac{\sqrt{t}}{\eta} g$, for $\eta > 0$, the network regret satisfies*

$$
\mathbb{E}[R_T] \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{QT}
$$

where $Q = \sum_{v \in V'} (q_v c_v)/Q_v$, $D \geq \sup g$, $L \geq \sup \|\nabla \ell_t\|_*$, and $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. In particular, choosing $\eta = \sqrt{2\sigma D}/L$ gives $\mathbb{E}[R_T] \leq L\sqrt{2DQT/\sigma}$.

Proof Fixing an arbitrary $x \in \mathcal{X}$, setting $r_t(v) = \ell_t(x_t(v)) - \ell_t(x)$, and proceeding as in Theorem 2 yields, for each $v \in V'$,

$$
\mathbb{E}\left[\sum_{t=1}^T r_t(v) \right] \leq \left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma} \right) \sqrt{\frac{T}{Q_v}} \tag{6}
$$

Now we write $\mathbb{E}[R_T] = \sup_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}[R_T(\boldsymbol{x})]$, where

$$\mathbb{E}[R_T(\boldsymbol{x})] = \mathbb{E}\left[\sum_{t=1}^{T} \frac{1}{\sum_{w \in V} X_t(w)} \sum_{v \in V'} r_t(v) X_t(v)\right]$$

$$= \sum_{t=1}^{T} \sum_{v \in V'} \mathbb{E}\left[\frac{X_t(v)}{\sum_{w \in V} X_t(w)}\right] \mathbb{E}[r_t(v)]$$

$$= \sum_{v \in V'} q_v c_v \sum_{t=1}^{T} \mathbb{E}[r_t(v)] \qquad (7)$$

and the last identity follows by Lemma 9. Putting identity (7) and inequality (6) together gives

$$\mathbb{E}[R_T(\boldsymbol{x})] \leq \left(\sum_{v \in V'} q_v c_v \sqrt{\frac{1}{Q_v}}\right)\left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma}\right)\sqrt{T} \leq \sqrt{\sum_{v \in V'} \frac{q_v c_v}{Q_v}}\left(\frac{D}{\eta} + \frac{\eta L^2}{2\sigma}\right)\sqrt{T}$$

where in the last inequality we used Jensen inequality and $\sum_{v \in V'} q_v c_v \leq 1$. This concludes the proof. ∎

We now prove that the inequality $Q(q) \leq \alpha_G$ is always true up to a small constant factor.

Lemma 11 *Let $G = (V, E)$ be an undirected graph. For all $v \in V$, choose numbers $q_v \in (0, 1]$ and define c_v and Q_v as in (4) and (5) respectively. Then*

$$Q = \sum_{v \in V} \frac{q_v c_v}{Q_v} \leq \frac{\alpha_G + 1}{1 - e^{-1}}$$

Proof Let $P_v = \sum_{w \in \mathcal{N}_v} q_w$, $V_1 = \{v \in V \mid P_v \geq 1\}$, and $V_0 = \{v \in V \mid P_v < 1\}$. We begin by splitting the sum as follows

$$\sum_{v \in V} \frac{q_v c_v}{Q_v} = \sum_{v \in V_1} \frac{q_v c_v}{Q_v} + \sum_{v \in V_0} \frac{q_v c_v}{Q_v}$$

We upper bound the two terms separately. Since the minimum $\min_{v \in V_1} Q_v$ is attained when $q_v = 1/|\mathcal{N}_v|$ for all $v \in \mathcal{N}_v$, we can lower bound, for each $v \in V_1$,

$$Q_v \geq 1 - \left(1 - \frac{1}{|\mathcal{N}_v|}\right)^{|\mathcal{N}_v|} \geq 1 - e^{-1}$$

This together with $\sum_{v \in V} q_v c_v \leq 1$ yields

$$\sum_{v \in V_1} \frac{q_v c_v}{Q_v} \leq \frac{1}{1 - e^{-1}}$$

To upper bound the sum over V_0, we first use the inequality $1 - x \leq e^{-x}$ that holds for all $x \in [0, 1]$. Setting $x = q_w$ gives

$$Q_v \geq 1 - \exp\left(-\sum_{w \in \mathcal{N}_v} q_w\right) = 1 - e^{-P_v}$$

16

For all $v \in V_0$, we can then use the inequality $1 - e^{-x} \geq (1 - e^{-1})x$, holding for all $x \in [0, 1]$. Setting $x = P_v < 1$ we conclude that $Q_v \geq (1 - e^{-1})P_v$ for all $v \in V_0$. Finally, using $c_v \leq 1$ we can write

$$\sum_{v \in V_0} \frac{c_v q_v}{Q_v} \leq \frac{1}{1 - e^{-1}} \sum_{v \in V} \frac{q_v}{P_v} \leq \frac{\alpha_G}{1 - e^{-1}}$$

where the last inequality follows by Lemma 3. Putting everything together gives the result. ∎