Online Learning with Feedback Graphs: The True Shape of Regret

Tomáš Kocák [1] **Alexandra Carpentier** [1]

Abstract

Sequential learning with feedback graphs is a natural extension of the multi-armed bandit problem where the problem is equipped with an underlying graph structure that provides additional information - playing an action reveals the losses of all the neighbors of the action. This problem was introduced by Mannor & Shamir (2011) and received considerable attention in recent years. It is generally stated in the literature that the minimax regret rate for this problem is of order $\sqrt{\alpha T}$, where α is the independence number of the graph, and T is the time horizon. However, this is proven only when the number of rounds T is larger than α^3, which poses a significant restriction for the usability of this result in large graphs. In this paper, we define a new quantity R^*, called the *problem complexity*, and prove that the minimax regret is proportional to R^* for any graph and time horizon T. Introducing an intricate exploration strategy, we define the EXP3-EX algorithm that achieves the minimax optimal regret bound and becomes the first provably optimal algorithm for this setting, even if T is smaller than α^3.

1. Introduction

In this paper, we consider a sequential decision-making problem in an adversarial environment. This problem consists of N actions, T rounds, and sequence of losses $(\ell_{t,i})_{(t,i)\in[T]\times[N]}$ where $[K] \triangleq \{1, \dots, K\}$. Each loss $\ell_{t,i}$ is associated with round t and action i. We do not impose any statistical assumptions on the losses provided by the environment. Instead, we assume that the losses are set by an oblivious adversary before the learning process begins. The only assumption on the losses is that they are bounded in $[0, 1]$, otherwise, the losses can be completely arbitrary and change in every round.

[1]Institute of Mathematics, University of Potsdam, Germany. Correspondence to: Tomáš Kocák <kocak@math.uni-potsdam.de>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

The learning process, or the game, proceeds in rounds. In round t, the learner picks one of the actions denoted by i_t and incurs associated loss ℓ_{t,i_t}. The learner also observes loss ℓ_{t,i_t} itself and possibly, losses of some other actions. The set of observations depends on the feedback scheme, we discuss different feedback schemes later.

The goal of the learner is to minimize the total loss received at the end of the game, after T rounds. This is equivalent to minimizing the difference between the total loss of the learner and the loss of the strategy that plays the best-fixed action in hindsight, after T rounds. We refer to this difference as regret and define it as

$$R_T \triangleq \max_{i\in[N]} \mathbb{E}\Big[\sum_{t\in[T]} (\ell_{t,i_t} - \ell_{t,i})\Big],$$

where the expectation is taken over the potential randomization of the environment and the learner.

A quantity of interest to characterize the difficulty of such a sequential decision-making problem is what we will refer to here as the minimax regret, namely the regret incurred by the best possible strategy - the choice of actions $(i_t)_t$ - on the most difficult possible bandit problem - the choice of loss sequences $(\ell_{t,i})_{t,i}$. Note that the minimax regret depends on the feedback scheme considered.

Traditionally, this problem is studied under different feedback schemes. The most relevant schemes for our paper are the following:

Full-information feedback (Cesa-Bianchi et al., 1997; Littlestone & Warmuth, 1994; Vovk, 1990), sometimes called prediction with expert advice. This feedback is the simplest since the learner has access to all losses. At the end of round t the learner observes whole loss vector $(\ell_{t,1} \dots, \ell_{t,N})$. The minimax rate for this feedback scheme is $\sqrt{T\log(N)}$ and is attained by the EXP algorithm (Cesa-Bianchi & Lugosi, 2006). Note that having access to all the losses in every round causes the minimax rate scale only as $\sqrt{\log(N)}$ with the number of actions.

Bandit feedback (Thompson, 1933; Robbins, 1952; Auer et al., 1995). In every round, the learner observes only the loss of the selected action, namely ℓ_{t,i_t}, while the losses of other actions are not disclosed. The minimax rate for

this feedback scheme is $\sqrt{NT}$ (Audibert & Bubeck, 2010) and is attained by INF (Implicitly Normalized Forecaster) algorithm by Audibert & Bubeck (2010). Having only one observation per round results in a scaling of the regret with $\sqrt{N}$ which is significantly worse than in the full-information feedback.

Graph feedback (Mannor & Shamir, 2011; Alon et al., 2013; 2015; 2017; Kocák et al., 2014; 2016a;b; Esposito et al., 2022). In the graph feedback setting, the actions are vertices of a graph and in every round, the learner observes the loss of the selected action (so that the setting is strongly observable) as well as the losses of all its neighbors - see Section 2 for a precise definition. This is the feedback scheme that we consider in this paper, which is an intermediary between full-information and bandit feedback and contains both these settings. Similarly to what happens in the bandit setting, the algorithms for bandits with graph feedback need to balance *exploration* of actions with *exploitation* of already acquired knowledge. In the graph feedback setting, however, different actions might provide different amounts of exploration, as an action also provides information on the losses of its neighbors. So that balancing exploration and exploitation in this context is more delicate, and efficient algorithms will need to adapt to the graph structure - and the minimax regret will also be graph dependent. In this setting, a relevant graph-dependent quantity is the independence number α of the graph (see Definition 2.3). Several algorithms with different approaches have been proposed, ELP (Mannor & Shamir, 2011), EXP3-SET and EXP3-DOM (Alon et al., 2013), EXP3-IX and FPL-IX (Kocák et al., 2014), EXP3.G (Alon et al., 2015). While these algorithms differ in their approach to exploration, assumptions on the graph disclosure, or computational complexity, the common denominator is that all of these algorithms' upper bounds on the regret, in the case of strongly observable graphs, are of order $\sqrt{\alpha T}$ up to logarithmic terms, regardless of time horizon T. All of the aforementioned algorithms were inspired by the lower bound for the setting proposed by Mannor & Shamir (2011), which states that if $T \geq 374\alpha^3$, the minimax regret is lower bounded by a quantity of the order $\sqrt{\alpha T}$ - see Proposition 2.4 for a precise quotation of their result. This poses the question of what happens for large graphs - or equivalently when T is small - and whether current algorithms are also optimal in this case. This is a very important question since even in a moderately large problem and for some graphs, where the independence number is in the hundreds, we need to have millions of rounds for this assumption to hold.

Partial monitoring. A bit further from the setting that we consider in this paper, yet related to it, is the field of partial monitoring (Rustichini, 1999; Audibert & Bubeck, 2010; Lattimore & Szepesvári, 2019; 2020), where the action selection is decoupled from the feedback. An example of this which is very relevant for us is weakly observed graphs - see (Alon et al., 2015) - which is a generalization of the graph feedback setting where not all self-loops are included, which means that one does not necessarily observes the loss of the action that one selects. The algorithm EXP3.G therein takes advantage of small dominating sets of vertices - i.e. sets of vertices whose joint set of neighbors are all vertices, see Section 2 for a precise definition - to explore efficiently the vertices, and then focus on promising actions. While not developed for the setting considered in this paper, this approach opens however interesting perspectives in cases of large graphs with a few very connected vertices, and we will discuss this more in detail in Subsection 2.2.

1.1. Contribution

In this paper, we focus on the setting with graph feedback - see Section 2 - and our aim is to pinpoint the minimax regret in the missing case presented in the corresponding paragraph above, namely for large graphs where T is of smaller order than α^3.

The first important remark that we make in this paper is that there are some simple cases of large graphs where it is possible to achieve a minimax regret of much smaller order than $\sqrt{\alpha T}$, which is the current best known upper bound. This is e.g. the case when there is one action that is connected in the graph to all other actions, and that is therefore very informative. In this case, if T is of smaller order than α^3, a minimax optimal strategy would make heavy use of this action in order to explore the other actions, even if this action is sub-optimal. We detail such an example in 2.2. This is very different from what current algorithms in the (strongly observable) graph bandit literature do and is more related to some strategies in partial monitoring, see e.g. (Lattimore & Szepesvári, 2019) and also (Alon et al., 2015) that we will discuss in detail later. Starting from this remark, the main result of this paper is to pinpoint, for any time horizon T and any given graph, the minimax regret up to logarithmic terms. We first provide a more refined lower bound in Section 3, that holds for any graph and time horizon - therefore also in the case where $T < 374\alpha^3$ which is not covered by the state of the art lower bound in Mannor & Shamir (2011) - and that involves a more subtle graph dependent quantity than the independence number. Then, in Section 4, we provide EXP3-EX algorithm (EX stands for **E**xplicit e**X**ploration) that matches this lower bound up to logarithmic terms, and whose particularity is that it explores informative actions in a refined and explicit way.

2. Problem Setting

In this section, we formally define the setting introduced by Mannor & Shamir (2011) and provide all the notation used

throughout the paper.

We consider an online learning game with a directed observability graph $G = (V, E)$ over the set of actions $V = [N]$ with the set of edges $E \subseteq [N] \times [N]$. The graph contains all the self-loops, i.e. $(i, i) \in E$ for every $i \in V$. The indicator function of an edge from node i to node j is defined as $G_{i,j} \triangleq \mathbb{I}\{(i, j) \in E\}$. The game takes place over T rounds. Before the game starts the environment, potentially adversarial, assigns losses $\{\ell_{t,i}\}_{(t,i)\in[T]\times[N]}$ to every action i and round t. We only assume that $\ell_{t,i} \in [0, 1]$ for any $t \leq T, i \leq N$.

In every round t, the learner picks an action $i_t \in [N]$, incurs the loss ℓ_{t,i_t}, and observes the losses $\ell_{t,i}$ of all out neighbors of i_t, i.e. of all $i \in V$ such that $(i_t, i) \in E$[1]. Note that in our setting, we always observe the loss of the chosen action since the graph contains all the self-loops. The performance of the learner is then measured in terms of regret - sometimes also called pseudo-regret, or also expected regret - as explained in the introduction

$$R_T \triangleq \max_{i\in[N]} \mathbb{E}\left[\sum_{t\in[T]} (\ell_{t,i_t} - \ell_{t,i})\right],$$

where the expectation is taken over the potential randomization of the environment and the learner.

2.1. Auxiliary Definitions and Statements

This section sums up all the necessary graph-related definitions we use later throughout the paper.

In bandits with graph feedback, the learner's task is to select an action and observe the losses of its neighbors. Each loss observation can have different sources, either the learner selected the action itself or one of its neighbors. The following definition provides us with a tool to define side observations and their sources more easily.

Definition 2.1. Let $G = (V, E)$ be a graph with the set of vertices V and the set of edges E. We define the out-neighborhood of vertex $i \in V$ as

$$N_i^{out} \triangleq \{j \in V \,:\, (i, j) \in E\}$$

and the in-neighborhood of vertex $i \in V$ as

$$N_i^{in} \triangleq \{j \in V \,:\, (j, i) \in E\}$$

Playing only a few actions can provide the learner with information about many other actions. Dominating sets and numbers provide a convenient way to describe this phenomenon.

Definition 2.2. Let $G = (V, E)$ be a graph with the set of vertices V and the set of edges E. We say that $D \subseteq V$ is a dominating set of $B \subseteq V$ (or that D dominates B) from $A \subseteq V$ if $D \subseteq A$ and $B \subseteq \cup_{i\in D} N_i^{out}$. We define the dominating number of B from A as $\delta^A(B) \triangleq \min |D|$ where the minimum is taken over all dominating sets D of B from A. In case no such D exists, we define $\delta^A(B)$ as ∞. Further, we say that $\delta(B) \triangleq \delta^V(B)$ is the dominating number of B and $\delta \triangleq \delta^V(V)$ is the dominating number of the graph.

On the other hand, if the actions are not connected by an edge, playing one action does not provide any additional information about the other actions. This is captured in the following definition of independent sets.

Definition 2.3. Let $G = (V, E)$ be a graph with the set of vertices V and the set of edges E. We say that $I \subseteq V$ is an independent set of G if for every $i, j \in I$ s.t. $i \neq j$, vertices i and j are not connected by an edge, i.e. $(i, j) \notin E$. Independence number α of G is the size of the largest independent set of G, i.e.

$$\alpha \triangleq \max_{I\in\{J\subseteq V \,:\, J \text{ is independent}\}} |I|.$$

2.2. Lower Bound by Mannor & Shamir (2011) and Motivational Example

In this subsection, we quote formally an important and state-of-the-art result of the literature and discuss why some relevant graph feedback examples are not optimally resolved by existing algorithms.

The following proposition restates the lower bound result by Mannor & Shamir (2011).

Proposition 2.4. *Let G be an observability graph with independence number α. Then there exists a series of losses $\{\ell_{t,i}\}_{(t,i)\in[T]\times[N]}$ such that for every $T \geq 374\alpha^3$ and any learner, the expected regret is at least $0.06\sqrt{\alpha T}$*

It is important to note that the statement assumes that T needs to be large - $T \geq 374\alpha^3$ - for the lower bound to hold. As mentioned in the introduction, some existing algorithms match this lower bound up to logarithmic factors (Kocák et al., 2014, Corollary 1) (Alon et al., 2015, Theorem 1). These algorithms also function when $T < 374\alpha^3$ where the best known upper bounds on the regret are of order $\sqrt{\alpha T}$ up to logarithmic factors. However, since the existing lower bound stated above does not cover this case, it is therefore unclear whether those algorithms are optimal or not.

The following lemma demonstrates that $\sqrt{\alpha T}$ is indeed not the correct rate.

Lemma 2.5. *Let $G = (V, E)$ be a graph with $|V| = N$ and $E = \{(N, i) : i \in [N - 1]\}$ (see Figure 1). Then, there exists an algorithm such that the regret upper bound of this algorithm is of $\delta^{1/3}T^{2/3}$ where $\delta = 1$ is the dominating number of G.*

[1] We write $N_{i_t}^{out}$ for this set, see definition 2.1 later.

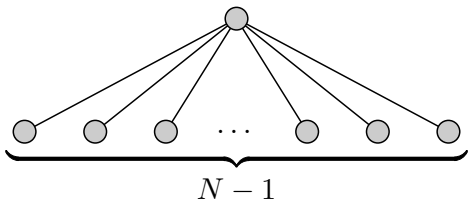

Figure 1. Bandit problem with one hub action that observes all other $N-1$ actions.

The independence number of the graph from the previous lemma is $N-1$. This also means that whenever $T \ll \alpha^3$ in the lemma above, the regret bound of $\delta^{1/3}T^{2/3}$ is an improvement over the regret bound of $\sqrt{\alpha T}$. See Appendix A for the proof and further discussion.

2.3. Problem Complexity

We have seen some indications (e.g. the example in Lemma 2.5) that for small T, the minimax regret might not scale with $\sqrt{\alpha T}$.

In what follows, we define the graph-dependent problem complexity that will later appear in our lower and upper bounds. This quantity is complex and depends on the graph in a refined way. It is however what one would expect for a worst-case stochastic problem - namely a problem where losses $\ell_{i,t}$ are independent and sampled according to a distribution depending on i. In order to introduce the problem complexity, we resort to intuitions from the stochastic setting, although we will analyze the problem in an adversarial setting, and provide precise results later, see Theorems 3.1 and 4.4.

Assume that we are given a set containing all promising actions that could be optimal, given available information - let us call it I, in a stochastic setting it would typically be a set of actions whose empirical mean confidence intervals intersect one of the actions with higher lower confidence bound. Let us oversimplify the problem and assume in this informal explication that all these actions but one have a small gap $\Delta > 0$ with respect to the optimal action. The optimal action is also in I and has a gap of 0. When playing an action in I, one incurs an average instantaneous regret of Δ except if one samples the optimal action. We are facing the following choice when we try to find the optimal action: we can either sample in I directly and have small instantaneous regret - namely Δ - or we can sample outside of this set and have an instantaneous regret that is in all generality bounded by 1. However, sampling outside of I might still be interesting if some of the actions there are connected to many actions in I, providing in this way very informative feedback on many actions therein - see e.g. Figure 1 where even if the hub action is clearly sub-optimal from the samples, it might still be interesting to take advantage of it. At the end of the budget and if one wants to have found the optimal action - and to therefore not pay an instantaneous regret of at least Δ at each round - one would need as is usual in stochastic bandits to have observed all actions - from inside or outside of I- at least $1/\Delta^2$ times. In the stochastic setting, we would therefore expect that the most difficult graph bandit problems would correspond to the worst choice of (I, Δ).

These considerations drive us to the first definition of the *problem complexity* Q^*.

Definition 2.6. Let $G = (V, E)$ be a graph and T be the number of rounds. Then the problem complexity Q_I^* for given set $I \subseteq V$ is defined as

$$Q_I^* \triangleq \max_{\Delta \in (0,1/2]} Q_{I,\Delta}^*$$

where

$$Q_{I,\Delta}^* \triangleq \min_{\boldsymbol{\pi} \in \Pi} \min \left[T \sum_{i \in I} \pi_i \Delta + T \sum_{i \notin I} \pi_i,\, T\Delta \right]$$

and

$$\Pi = \Bigg\{ \boldsymbol{\pi} \in \mathbb{R}_+^N : \sum_{i \in [N]} \pi_i \leq 1, \\ T \sum_{i \in [N]} \pi_i G_{i,j} \geq 1/\Delta^2, \forall j \in I \Bigg\}.$$

Moreover, we define the problem complexity Q^* as

$$Q^* \triangleq \max_{I \subseteq V} Q_I^*.$$

$Q_{I,\Delta}^*$ would correspond to the regret of the best stationary policy π over a problem as described above, for a fixed set I and gap Δ. The worst-case problem is then obtained by taking the worst case of set I and gap Δ.

Unfortunately, the quantity defined in Definition 2.6 is very unintuitive, in that it is unclear how it relates to quantities such as dominating numbers, and independence numbers, of (sub-)graphs. we, therefore, define another relevant notion of the problem complexity R^*.

Definition 2.7. Let $G = (V, E)$ be a graph and T be a number of rounds. Then the problem complexity R_I^* for given set $I \subseteq V$ is defined as

$$R_I^* \triangleq \min_{J \subseteq I} \max \left\{ \delta^I(J)^{\frac{1}{2}} T^{\frac{1}{2}},\, \delta^V(I \setminus J)^{\frac{1}{3}} T^{\frac{2}{3}} \right\}.$$

Moreover, we define the problem complexity R^* as

$$R^* \triangleq \max_{I \subseteq V} R_I^*.$$

This definition is much more tractable from a graph perspective, as it involves only two relevant graph-dependent quantities, namely the dominating set $\delta^I(J)$ of J from I, and the dominating number $\delta^V(I \setminus J)$ of $I \setminus J$ from V. Here the choice of the optimal policy is reduced to only choosing the set J that is best explored from inside of I, and we then select the worst case of I.

Interestingly, the following lemma shows that both definitions of the problem complexity are almost equivalent and differ only up to a logarithmic factor. From now on, whenever talking about the problem complexity, we specify which definition we use and the reasons why.

Lemma 2.8. *Let $G = (V, E)$ be a graph and $I \subseteq V$ be any set of actions. Then for the problem complexities Q_I^* and R_I^*, the following inequalities hold.*

$$R_I^*/(10 \log N) \leq Q_I^* \leq 2R_I^*$$

The proof of this lemma can be found in Appendix B.

3. Lower Bound

Ever since the introduction of the setting by Mannor & Shamir (2011), the lower bound in Proposition 2.4 was used to drive the ideas for the algorithms. However, in general, this lower bound holds only when $T \geq 374\alpha^3$.

Even though approaches of algorithms and their upper bound analyses differ from paper to paper, most of them are able to match the lower bound for $T \geq 374\alpha^3$. Without the lower bound for regimes where $T < 374\alpha^3$, there was no incentive for the algorithms to strive for a different rate than the one suggested by Proposition 2.4. In this section, we present a new lower bound that holds regardless of the value of T and thus, extend the result in Proposition 2.4.

The following theorem shows a regret lower bound that scales with the problem complexity R^* and is one of the main results of our paper.

Theorem 3.1. *Let $G = (V, E)$ be a directed graph with $N = |V|$ and T be a number of rounds. Then, for any learner, there exists a sequence of randomized losses such that regret R_T of the learner is lower bounded as*

$$R_T \geq \frac{Q^*}{2^7} \geq \frac{R^*}{2^7 10 \log N}.$$

where Q^* and R^* are problem complexities

Proof idea. The idea of the proof follows standard lower bound proof steps, see e.g. (Lattimore & Szepesvári, 2020, Chapter 15), with our problem-specific twist. The idea is to create a set of "difficult" stochastic bandit problems and show that no matter what the learner does, there always will be at least two different problems that the learner can not distinguish.

We create the problems by first choosing a set of near-optimal actions I and then setting the gap of every action outside of I to 1 and inside of I to some small constant Δ. The only exception is the optimal action. For different problems, we choose different optimal actions from I and set its gap to 0.

Using information-theoretic tools, we can show that every action needs to be explored enough, i.e. at least $1/\Delta^2$ times, in order to be able to distinguish the problems. The result of the theorem is then obtained by carefully choosing the gap parameter Δ and the set of difficult actions I.

The detailed proof of the theorem can be found in Appendix C. Note that the lower bound in Theorem 3.1 scales with either of the definitions of the problem complexity. Later, we show that the rate depending on the problem complexity is indeed minimax optimal and we comment on the connection to the rate in the previous papers in Section 5.

4. Algorithm

The algorithm for our setting, similarly to the previous papers, uses exponential weights to define a probability distribution over the set of actions and then samples according to this distribution. Similarly to EXP3.G algorithm by Alon et al. (2015), we add extra exploration to some actions. This extra exploration adapts to the estimated quality of each action, but also to its informativeness, i.e. to how much it is connected to other promising actions on the graph.

The construction of the exploration distribution is rather intricate and is the main algorithmic contribution of this paper, and we will discuss it in detail later. We first present the main algorithm, with part devoted to the construction of this exploration distribution. We then present associated theoretical guarantees.

4.1. Main Algorithm

As is usual in the literature, our algorithm (EXP3-EX presented in Algorithm 1) updates at each time t a probability distribution $\mathbf{p}_t$. It then plays at each round action i_t, which in the graph setting reveals the losses $\ell_{t,i}$ of all its neighbors $i \in N_{i_t}^{out}$. These losses enable us to update unbiased estimates of the (cumulative) loss estimates, which will be used in the algorithm.

(Cumulative) Loss estimates. In the graph setting, the probability $P_{t,i}$ of observing loss $\ell_{t,i}$ is simply the sum of probabilities of playing any of the in-neighbors of arm i and is defined as $P_{t,i} \triangleq \sum_{j \in N_i^{in}} p_{t,j}$. This allows us, at the end of each round t, to construct, conditionally unbiased loss

estimates $\hat{\ell}_{t,i}$ of the loss of each action

$$\hat{\ell}_{t,i} \triangleq \frac{\ell_{t,i}\mathbb{I}\{i \in N_{i_t}^{out}\}}{P_{t,i}} \qquad \text{for all} \qquad i \in [N]$$

and to also update the cumulative loss estimates as

$$\widehat{L}_{t,i} \triangleq \widehat{L}_{t-1,i} + \hat{\ell}_{t,i} \qquad \text{for all} \qquad i \in [N].$$

We now describe the construction of the distribution $\mathbf{p}_t$. As in (Alon et al., 2015) and in several other papers from the literature, we mix a distribution based on exponential weights - i.e. we update at each time t the exponential weights $\mathbf{w}_t$ and define a normalized distribution $\mathbf{q}_t$, as in (Auer et al., 2002) - with an exploration distribution $\mathbf{u}_t$. We postpone the construction of the learning distribution to Section 4.2 (summary in Definition 4.3), as it is intricate and our main algorithmic contribution. We describe below how we construct $\mathbf{w}_t, \mathbf{q}_t, \mathbf{p}_t$, based on $\mathbf{u}_t$.

Recallibration of the parameters. Our algorithm first recalibrates at every step the learning rate η_t of the EXP3 part of the algorithm - we describe later in how it is chosen. Our algorithm uses it to also callibrates the mixing probability $\gamma_t \triangleq \min\{(\eta_t T)^{-1}, 1/2\}$ of sampling the exploration distribution $\mathbf{u}_t$.

(Renormalized) Exponential weights. Based on the cumulative loss estimate $\widehat{L}_{t-1,i}$ of arm i at time t, we can define as in (Auer et al., 2002) the exponential weights $w_{t,i}$ as

$$w_{t,i} \triangleq \exp(-\eta_t \widehat{L}_{t-1,i}) \qquad \text{for all} \qquad i \in [N].$$

Using these weights, we can construct a distribution simply by re-normalizing them to define

$$q_{t,i} \triangleq \frac{w_{t,i}}{W_t} \triangleq \frac{w_{t,i}}{\sum_{j\in[N]} w_{t,j}} \qquad \text{for all} \qquad i \in [N]. \quad (1)$$

Mixed distribution. Based on $\mathbf{u}_t$, we define our sampling distribution as

$$p_{t,i} \triangleq (1-\gamma_t)q_{t,i} + \gamma_t u_{t,i} \qquad \text{for all} \qquad i \in [N]. \quad (2)$$

So far the algorithm is not different from the majority of algorithms designed for the graph setting and it is summarized in Algorithm 1. The key difference lies in the exploration distributions $(u_{t,i})_{i\in[N]}$ leveraging the structure of the graph, and in the learning rates η_t defined later in Theorem 4.4. Especially exploration distributions $(u_{t,i})_{i\in[N]}$ set our algorithm apart from the previous algorithms and enables us to improve the upper bound to match the newly proposed lower bound. The following section explains all the details necessary for the definition of the exploration distributions.

Algorithm 1 EXP3-EX

Input: $G = (V, E)$, $\widehat{L}_{0,i} = 0$ for all $i \in [N]$
for $t = 1$ **to** T **do**
  Set learning rate η_t (see Theorem 4.4)
  $\gamma_t = \min\{(\eta_t T)^{-1}, 1/2\}$
  $w_{t,i} = (1/N)\exp(-\eta \widehat{L}_{t-1,i})$
  $W_t = \sum_{i\in[N]} w_{t,i}$
  $q_{t,i} = w_{t,i}/W_t$
  $p_{t,i} = (1-\gamma_t)q_{t,i} + \gamma_t u_{t,i}$ (see Definition 4.3)
  Choose $i_t \sim \boldsymbol{p}_t = (p_{t,1}, \ldots, p_{t,N})$
  Observe losses $\ell_{t,i}$ for $i \in N^{out}(i_t)$
  $P_{t,i} = \sum_{j\in N^{in}(i)} p_{t,j}$
  $\hat{\ell}_{t,i} = \ell_{t,i}\mathbb{I}\{i \in N^{out}(i_t)\}/P_{t,i}$
  $\widehat{L}_{t,i} = \widehat{L}_{t-1,i} + \hat{\ell}_{t,i}$
end for

4.2. Mixing Distribution and Exploration

In the graph setting, the interest of an algorithm for sampling an arm is not only characterized by the quality of this arm - i.e. minus its cumulative loss at time t - but also by the informativeness of this algorithm on other relevant arms - namely, whether or not it is connected to many arms with small cumulative loss at time t. While a classical adversarial bandit algorithm would take into account the first of these two factors, we need to add extra exploration to take into account the second factor, namely the connections of the arms through the graph structure.

The idea of the algorithm is to homogenize the actions by grouping them up according to their cumulative loss as well as the amount of information they provide and then define the exploration for each partition separately. We create the partitioning in two steps.

Partitioning of the actions into sets $(I_{t,k,l})_{t,k,l}$. For every round t, we create partitions $\{I_{t,k}\}_{k\in[K+1]}$, for $K = \lceil 5\log_2(N)\rceil$, such that the normalized exponential weights of arms, defined in Equation 1, are similar within a partition. More precisely define

$$I_{t,k} \triangleq \left\{i \in [N] : q_{t,i} \in (2^{-k}, 2^{-k+1}]\right\}.$$

The last partition $I_{t,K+1}$ contains the rest of the arms, i.e.

$$I_{t,K+1} \triangleq \left\{i \in [N] : q_{t,i} \le 2^{-K}\right\}.$$

Note that for every action $i \in I_{t,K+1}$, $q_{t,i}$ can be upper bounded by $1/N^5$.

We further subdivide each set $I_{t,k}$ into subsets that are roughly homogeneous in terms of the numbers of neighbors in $I_{t,k}$. For every arm $i \in I_{t,k}$, we define $\deg_{t,k}(i)$ as the number of neighbors of i within partition $I_{t,k}$:

$$\deg_{t,k}(i) \triangleq |\{j \in I_{t,k} : (i,j) \in E\}|.$$

For $L = \lceil \log_2(N) \rceil$, we will further subdivide each set $I_{t,k}$ into subsets $\{I_{t,k,l}\}_{l\in[L]}$ that are roughly homogeneous in terms of numbers of neighbors in $I_{t,k}$ - namely, arm $i \in I_{t,k,l}$ if and only if $\deg_{t,k}(i) \in (N2^{-l}, N2^{-l+1}]$. The following definition summarizes the construction of the partitions.

Definition 4.1. Let $K \triangleq \lceil 5\log_2(N) \rceil$ and $L \triangleq \lceil \log_2(N) \rceil$, then for every $(t,k,l) \in [T] \times [K] \times [L]$ we define

$$I_{t,k,l} \triangleq \Big\{ i \in [N] : q_{t,i} \in \big(2^{-k}, 2^{-k+1}\big], \\ \deg_{t,k}(i) \in \big(N2^{-l}, N2^{-l+1}\big] \Big\}$$

and

$$I_{t,K+1} \triangleq \Big\{ i \in [N] : q_{t,i} \le 2^{-K} \Big\}.$$

Partition in exploration sets from inside and outside of $I_{t,k,l}$. Inspired by the definition of the problem complexity in Definition 2.7, we can define a splitting of every set $I_{t,k,l}$, for $k \in [K]$ and $l \in [L]$, into two parts, $J_{t,k,l}$ and $J'_{t,k,l} \triangleq I_{t,k,l} \setminus J_{t,k,l}$ that minimize expression

$$\max\Big\{ \delta^{I_{t,k,l}}(J_{t,k,l})^{\frac{1}{2}} T^{\frac{1}{2}},\, \delta^{V}(J'_{t,k,l})^{\frac{1}{3}} T^{\frac{2}{3}} \Big\}.$$

We write $R^*_{I_{t,k,l}}$ the value of this minimum for each given set $I_{t,k,l}$. As we discussed in Section 2.3, an optimal exploration of the set $J_{t,k,l}$ can be done using actions in $I_{t,k,l}$, while an optimal exploration of $J'_{t,k,l}$ can be performed using actions outside $I_{t,k,l}$.

In order to construct our exploration distribution, we would like to have access to the sets $J_{t,k,l}$ and $J'_{t,k,l}$, and more specifically to some (approximate) dominating sets, in order to be able to define the exploration distribution. While it is possible in theory to find these sets based on the $I_{t,k,l}$ and on the graph, solving the optimization problem that leads to them can be computationally very expensive.

For this reason, we do not work directly with the sets $J_{t,k,l}, J'_{t,k,l}$, but rather with some approximations that are computationally tractable. Such approximations exist, as stated in Corollary 4.2 below, and are described in Appendix D.

Corollary 4.2. *The algorithm described in Appendix D, which is polynomial time in N (as it consists in solving a linear optimization problem under linear constraints) outputs partitions $\bar{J}_{t,k,l}$, $\bar{J}'_{t,k,l}$ of $I_{t,k,l}$ together with their corresponding dominating sets $\bar{D}_{t,k,l}$, $\bar{D}'_{t,k,l}$, which satisfy*

$$|\bar{D}_{t,k,l}| \le \log(N)\delta^{I_{t,k,l}}(\bar{J}_{t,k,l}),$$
$$|\bar{D}'_{t,k,l}| \le \log(N)\delta^{V}(\bar{J}'_{t,k,l}),$$

and

$$\max\Big\{ |\bar{D}_{t,k,l}|^{\frac{1}{2}} T^{\frac{1}{2}},\, |\bar{D}'_{t,k,l}|^{\frac{1}{3}} T^{\frac{2}{3}} \Big\} \le \\ \le 24 \times 10^4 \log(N)^4 \sqrt{\log(N)} R^*_{I_{t,k,l}}.$$

As mentioned, $\bar{J}_{t,k,l}$ (resp. $\bar{J}'_{t,k,l}$) serves as a surrogate of $J_{t,k,l}$ (resp. $J'_{t,k,l}$) and dominating set $\bar{D}_{t,k,l}$ (resp. $\bar{D}'_{t,k,l}$) is an approximation of the smallest dominating set of $\bar{J}_{t,k,l}$ (resp. $\bar{J}'_{t,k,l}$) from $I_{t,k,l}$ (resp. V). While the full construction of these sets is deferred to Appendix D, we discuss and sketch briefly their construction in Subsection 2.3.

Having an efficient way of computing partitions $\bar{J}'_{t,k,l}$ and their dominating sets $\bar{D}'_{t,k,l}$ allows us to define the following exploration distribution

Definition 4.3. let $I_{t,k,l}$, for $(t,k,l) \in [T] \times [K] \times [L]$ be a partition of V from Definition 4.1 and $\bar{D}'_{t,k,l}$ be a dominating set, from Corollary 4.2. Then, we can define,

$$u_{t,i} \triangleq \frac{1}{KL+1}\left(\frac{1}{N} + \sum_{k\in[K]} \sum_{l\in[L]} u^{k,l}_{t,i} \right) \tag{3}$$

where

$$u^{k,l}_{t,i} = \frac{1}{|\bar{D}'_{t,k,l}|} \qquad \text{for all} \qquad i \in \bar{D}'_{t,k,l}$$
$$u^{k,l}_{t,i} = 0 \qquad \text{for all} \qquad i \notin \bar{D}'_{t,k,l}.$$

Distribution $(u_{t,i})_{i\in[N]}$ can be seen as a mixture of uniform distributions where the term $1/N$ in Equation 3 corresponds to the uniform distribution over all the actions and $(u^{k,l}_{t,i})_{i\in[N]}$ corresponds to the uniform distribution over set $\bar{D}'_{t,k,l}$ which, as a consequence, secures exploration of $\bar{J}'_{t,k,l}$.

4.3. Main Upper Bound Theorem

Utilization of the exploration distributions $(u_{t,i})_{i\in[N]}$ from the previous section and appropriately tuned learning rates η_t enable us to prove the optimal regret upper bound for Algorithm 1 stated in the following theorem.

Theorem 4.4. *Let learning rate η_t is defined as*

$$\min_{s\in[t]} \min_{k\in[K]} \min_{l\in[L]} \min\Big\{ |\bar{D}_{s,k,l}|^{-\frac{1}{2}} T^{-\frac{1}{2}}, |\bar{D}'_{s,k,l}|^{-\frac{1}{3}} T^{-\frac{2}{3}} \Big\},$$

where we remind that $\bar{D}_{t,k,l}$ and $\bar{D}'_{t,k,l}$ be the dominating sets outputted by the algorithm described in Appendix D. Then the regret of Algorithm 1 is upper bounded as

$$R_T \le 24 \times 10^4 \log(N)^5 D R^*$$

for

$$D = 4KL + 2 + \big((KL)^2 + KL + 1\big)\log(N),$$
$$K = \lceil 5\log_2(N) \rceil,$$
$$L = \lceil \log_2(N) \rceil.$$

Proof idea. The proof of the theorem relies heavily on the partitioning from the Definition 4.1 by decomposing the

regret along the partitions. Careful construction of partitions allows us to show that the actions corresponding to one individual partition contribute to regret with no more than R^*, up to logarithmic factors. The fact that the number of partitions $KL + 1$ is only polylogarithmic in the number of actions allows us to obtain the final regret bound as the sum of regret bounds for individual partitions.

The detailed proof of the theorem can be found in Appendix E.

5. Discussion

We have presented the regret lower bound for the setting (Section 3, Theorem 3.1) as well as the matching, up to logarithmic terms, regret upper bound for the proposed algorithm (Section 4, Theorem 4.4). Together, these two theorems prove that the minimax rate for online learning with feedback graphs is proportional to the problem complexity R^* (Definition 2.7). In this section, we compare the minimax rate presented in this paper to the previously known results and emphasize the improvements that we bring to the setting. We focus mainly on two regimes:

When T is large enough when compared to α^3, we recover results from the literature, namely a minimax rate of order $\sqrt{\alpha T}$ up to logarithmic terms. We do it by showing that the problem complexity R^* is equal to $\sqrt{\alpha T}$ when T is large enough.

When T is small, we demonstrate the existence of graphs for which rate $\sqrt{\alpha T}$ is far from optimal. An important consequence of this statement is that all the algorithms proposed in the previous papers prove only suboptimal regret upper bounds for some graphs and budgets T. This also means that Algorithm 1 is the first provably optimal algorithm for the setting in all possible problems and regimes.

5.1. Regime when T is Large

Previous papers proved that the minimax regret scales with $\sqrt{\alpha T}$ whenever $T \geq 374\alpha^3$ while the minimax regret presented in this paper scales with the problem complexity R^* instead. The following corollary shows that the two rates are up to log factors the same when T is large enough.

Corollary 5.1. *Let G be a graph with independence number α. Then for any $T \geq \alpha^3$, the problem complexity R^* simplifies to*

$$R^* = \sqrt{\alpha T}.$$

The proof for this corollary can be found in Appendix F.1. As our upper and lower bounds in Theorems 3.1 and 4.4 match R^* up to logarithmic terms, we recover the results from the literature.

5.2. Regime when T is Small

From the previous section, we know that the minimax rate for large enough T is $\sqrt{\alpha T}$. It is also true that most of the prior algorithms can achieve a regret upper bound that scales with $\sqrt{\alpha T}$. However, at first glance, it is not obvious how significant the improvement of the newly defined problem complexity is.

We return back to the example from Lemma 2.5 and introduce a couple of examples demonstrating that rate $\sqrt{\alpha T}$ can be significantly sub-optimal.

Example 1. Lemma 2.5 provides an example where the graph contains $N - 1$ independent vertices and one hub connected to all other vertices. The following Corollary states that the minimax rate for this graph indeed scales with $\delta^{1/3}T^{2/3}$ instead of $\sqrt{\alpha T}$.

Corollary 5.2. *Let $G = (V, E)$ be a graph on N vertices with one hub, i.e. the set of edges is $E = \{(N, i) : i \in [N - 1]\}$. Then for any $T < \alpha^3$, the problem complexity R^* simplifies to*

$$R^* = T^{\frac{2}{3}}.$$

The proof for this corollary can be found in Appendix F.2. This result also shows that with the increasing number of actions, the gap between $\sqrt{\alpha T}$ and the problem complexity can be arbitrarily large.

Example 2. Generalizing the previous example, we can create a graph consisting of two parts. A star graph with $1 + N_1$ vertices and N_2 independent vertices without any edges. Now the problem complexity is of order $(T^{2/3} + \sqrt{N_2 T}) \wedge \sqrt{(N_1 + N_2)T}$ while $\alpha = N_1 + N_2$ and $\delta = 1 + N_2$. If either $N_2 \geq N_1$, or $T \geq N_1^3$, the problem complexities R^* and Q^* are of order $\sqrt{(N_1 + N_2)T} = \sqrt{\alpha T}$ (up to logarithmic terms) as predicted by Alon et al. (2015). However, if $N_2 < N_1$ and $T < N_1^3$ (large star graph), then the problem complexity is of order $T^{2/3}$. This is an example where the minimax rate is much smaller than $\delta^{1/3}T^{2/3}$ or $\sqrt{\alpha T}$. This example also illustrates that the minimax rates from the previous papers are not valid when T is small enough. In contrast, the true minimax rate scales with the problem complexity which demonstrates that it is important to adapt locally to the graph and global quantities like the dominating number, or the independence number, are not complex enough to describe the problem complexity.

Example 3. Expanding the previous example, we consider a graph where we have $\sum_{k \leq K}(k + 1)m_k$ vertices. This graph consists, for each $k \in \{1, \ldots, K\}$, of m_k star graphs with $k + 1$ vertices each with no connection to each other. In this case $\alpha = \sum_{k \leq K} k m_k$ and $\delta = \sum_{k \leq K} m_k$. Now, write A for the set of indexes k such that $\sqrt{m_k k T} \geq m_k^{1/3} T^{2/3}$. The problem complexities Q^*, R^*

are of order $\sqrt{T\sum_{k\notin A} m_k k} + (\sum_{k\in A} m_k)^{1/3}T^{2/3})$. For a graph containing some large star graphs, e.g. whenever $\sup_k m_k k^3 \geq T$, the rate is of order, up to logarithmic terms, of $(\sum_{k\in A} m_k)^{1/3}T^{2/3}$. This can be significantly smaller than $\delta^{1/3}T^{2/3}$ if A is very different from $\{1,\dots,n\}$, e.g. when the graph contains a small number of very large star graphs and a moderate number of small star graphs - an extreme case being in the previous example.

These examples highlight that in the case of large graphs that are not homogeneous in the size of their hubs, the problem complexity is not driven by quantities like the dominating number or the independence number, but by some related quantities that are local in the graph. Our algorithm is able to adapt to such local structures.

5.3. Exploration Distribution

We believe that the exploration distribution in Definition 4.3 plays a crucial role in adapting EXP3 algorithm to the setting for small T and that the algorithm is suboptimal without it. In general, exponential weights encourage playing actions with small cumulative loss but neglect actions that are highly informative, i.e. connected to many other actions. To correct this behavior, we look at every partition $I_{t,k,l}$ and identify the set $\bar{J}'_{t,k,l}$ from Corollary 4.2 and its dominating set $\bar{D}'_{t,k,l}$. We already know that the optimal way of exploring $\bar{J}'_{t,k,l}$ is by playing more informative actions in $\bar{D}'_{t,k,l}$. To enforce this behavior in the EXP3 algorithm we simply add extra uniform exploration to actions in $\bar{D}'_{t,k,l}$.

Acknowledgements

The work of A. Carpentier is partially supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether grant MuSyAD (CA 1488/1-1), by the DFG CRC 1294 'Data Assimilation', Project A03, by the DFG Forschungsgruppe FOR 5381 "Mathematical Statistics in the Information Age - Statistical Efficiency and Computational Tractability", Project TP 02, by the Agence Nationale de la Recherche (ANR) and the DFG on the French-German PRCI ANR ASCAI CA 1488/4-1 "Aktive und Batch-Segmentierung, Clustering und Seriation: Grundlagen der KI".

References

Alon, N., Cesa-Bianchi, N., Gentile, C., and Mansour, Y. From bandits to experts: A tale of domination and independence. *Neural Information Processing Systems*, 2013.

Alon, N., Cesa-Bianchi, N., Dekel, O., and Koren, T. Online learning with feedback graphs: Beyond bandits. *Conference on Learning Theory*, 2015.

Alon, N., Cesa-Bianchi, N., Gentile, C., Mannor, S., Mansour, Y., and Shamir, O. Nonstochastic multi-armed bandits with graph-structured feedback. *SIAM Journal on Computing*, 2017.

Audibert, J.-Y. and Bubeck, S. Regret bounds and minimax policies under partial monitoring. *Journal of Machine Learning Research*, 2010.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. Gambling in a rigged casino: The adversarial multi-armed bandit problem. *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, 1995.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. The nonstochastic multi-armed bandit problem. *SIAM Journal on Computing*, 2002.

Cesa-Bianchi, N. and Lugosi, G. *Prediction, learning, and games.* Cambridge University Press, 2006.

Cesa-Bianchi, N., Freund, Y., Haussler, D., Helmbold, D. P., Schapire, R. E., and Warmuth, M. K. How to use expert advice. *Journal of the ACM*, 1997.

Chvatal, V. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, 1979.

Cohen, M. B., Lee, Y. T., and Song, Z. Solving linear programs in the current matrix multiplication time. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. Association for Computing Machinery, 2019.

Esposito, E., Fusco, F., van der Hoeven, D., and Cesa-Bianchi, N. Learning on the edge: Online learning with stochastic feedback graphs. In *Advances in Neural Information Processing Systems*, 2022.

Györfi, L. and Ottucsák, G. Sequential prediction of unbounded stationary time series. *IEEE Transactions on Information Theory*, 2007.

Kocák, T., Neu, G., Valko, M., and Munos, R. Efficient learning by implicit exploration in bandit problems with side observations. *Neural Information Processing Systems*, 2014.

Kocák, T., Neu, G., and Valko, M. Online learning with noisy side observations. *International Conference on Artificial Intelligence and Statistics*, 2016a.

Kocák, T., Neu, G., and Valko, M. Online learning with Erdős-Rényi side-observation graphs. *Conference on Uncertainty in Artificial Intelligence*, 2016b.

Lattimore, T. and Szepesvári, C. An information-theoretic approach to minimax regret in partial monitoring. *Conference on Learning Theory*, 2019.

Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, 2020.

Littlestone, N. and Warmuth, M. K. The weighted majority algorithm. *Information and Computation*, 1994.

Mannor, S. and Shamir, O. From bandits to experts: On the value of side-observations. *Neural Information Processing Systems*, 2011.

Robbins, H. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 1952.

Rustichini, A. Minimizing regret: The general case. *Games and Economic Behavior*, 1999.

Thompson, W. R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 1933.

Vovk, V. Aggregating strategies. *Proceedings of the third annual workshop on Computational learning theory*, 1990.

A. Suboptimality of $\sqrt{\alpha T}$ for some Graphs and Proof of Lemma 2.5

Consider a graph with $N-1$ independent vertices and one hub vertex connected to all other vertices - see Figure 1. The independence number α of this graph is $N-1$. This also means that the regret bounds of existing algorithms scale with $\sqrt{TN}$. This rate is also attained by classical minimax bandit algorithms that do not take the graph structure into account at all. This also means that there is no theoretical advantage to using existing graph-based algorithms.

However, there are several indications that playing the hub vertex might be a good idea. Especially, when T is small and the graph is large, i.e. N is large.

The first indication comes from the stochastic bandits where even simple *Explore then Commit* algorithm that samples the hub vertex $T^{2/3}$ times and then commits to the empirically best action achieves regret of $T^{2/3}$ (Lattimore & Szepesvári, 2020, Theorem 6.1.).

Another indication comes from the EXP3.G algorithm by Alon et al. (2015). This algorithm work in a slightly more general setting where the feedback graph can be weakly observable, i.e. the learner does not necessarily knows the loss of the selected actions. We can transform our example from Figure 1 to the weakly observable case by not revealing the loss of the selected action to the learner whenever the learner selects a non-hub action. This makes the problem more difficult. Applying the EXP3.G algorithm would result in the regret bound of $\delta^{1/3}T^{2/3}$ where δ is the dominating number of the graph, in our case $\delta = 1$.

B. Proof of Lemma 2.8

We start with the lower bound on Q_I^*. The proof is conducted by connecting set Π to the underlying graph and selecting a specific δ that gives us a desirable lower bound.

Every vector $\boldsymbol{\pi}$ can be associated with a non-adaptive algorithm that plays each arm i with probability π_i, regardless of past observations. Condition $T\sum_{i\in[N]}\pi_i G_{i,j} \geq \frac{1}{\Delta^2}$ means that, in expectation, every action in I needs to be observed at least $1/\Delta^2$ times. From now on, when we talk about the number of observations of the algorithm or the number of times the algorithm plays an action, we mean the quantity in expectation. Now, we can split set I into subsets

$$J_{I,\boldsymbol{\pi}} \triangleq \{j \in I : \sum_{i\in I}\pi_i G_{i,j} \geq \frac{1}{2}\sum_{i\in[N]}\pi_i G_{i,j}\} \tag{4}$$

and

$$J'_{I,\boldsymbol{\pi}} \triangleq \{j \in I : \sum_{i\in I}\pi_i G_{i,j} < \frac{1}{2}\sum_{i\in[N]}\pi_i G_{i,j}\} = I \setminus J_{I,\boldsymbol{\pi}}. \tag{5}$$

Set $J_{I,\boldsymbol{\pi}}$ contains all the actions of I with at least half of the observations coming from actions in I while $J'_{I,\boldsymbol{\pi}}$ contains all the actions of I with at least half of the observations coming from actions in $V \setminus I$

Before proceeding, we state a technical graph proposition (Alon et al., 2015, Lemma 8)

Proposition B.1. *Let $G = (V, E)$ be a graph over $|V| = N$ vertices, and let $I \subseteq V$ be a set of vertices whose smallest dominating set is of size $\delta(I)$. Then, I contains an independent set U of size at least $\delta(I)/(50\log N)$, with the property that any vertex of G dominates at most $\log N$ vertices of U*

Applying Proposition B.1 to set $J_{I,\boldsymbol{\pi}}$ and a subgraph of G, induced by I, implies existence of set $U_{I,\boldsymbol{\pi}} \subseteq J_{I,\boldsymbol{\pi}}$, such that

$$|U_{I,\boldsymbol{\pi}}| \geq \frac{\delta^I(J_{I,\boldsymbol{\pi}})}{50\log|I|} \geq \frac{\delta^I(J_{I,\boldsymbol{\pi}})}{50\log N}$$

with a property that every vertex of I dominates at most $\log|I| \leq \log N$ nodes from $U_{I,\boldsymbol{\pi}}$. Since $U_{I,\boldsymbol{\pi}} \subseteq J_{I,\boldsymbol{\pi}}$, using the definition of $J_{I,\boldsymbol{\pi}}$ together with the assumption that the number of observations of each arm in I is at least $1/\Delta^2$, we know that

$$T\sum_{i\in I}\pi_i G_{i,j} \geq \frac{T}{2}\sum_{i\in[N]}\pi_i G_{i,j} \geq \frac{1}{2\Delta^2}$$

for every $j \in U_{I,\boldsymbol{\pi}}$. Summing over $j \in U_{I,\boldsymbol{\pi}}$ and using the fact that each arm from I can provide at most $\log N$ observations of arms in $U_{I,\boldsymbol{\pi}}$, we need the algorithm associated with $\boldsymbol{\pi}$ to play actions from I at least

$$\frac{|U_{I,\boldsymbol{\pi}}|}{2\Delta^2 \log N} \geq \frac{\delta^I(J_{I,\boldsymbol{\pi}})}{100\Delta^2 \log^2 N}$$

times to ensure enough observations in $U_{I,\boldsymbol{\pi}}$. This gives us

$$T \sum_{i \in I} \pi_i \geq \frac{\delta^I(J_{I,\boldsymbol{\pi}})}{100\Delta^2 \log^2 N}, \tag{6}$$

Applying Proposition B.1 to $J'_{I,\boldsymbol{\pi}}$ and graph G implies existence of set $U'_{I,\boldsymbol{\pi}} \subseteq J'_{I,\boldsymbol{\pi}}$ such that

$$|U'_{I,\boldsymbol{\pi}}| \geq \frac{\delta^V(J'_{I,\boldsymbol{\pi}})}{50 \log |V|} = \frac{\delta^V(J'_{I,\boldsymbol{\pi}})}{50 \log N}$$

with a property that every vertex of V dominates at most $\log |V| = \log N$ nodes from $U'_{I,\boldsymbol{\pi}}$. Since $U'_{I,\boldsymbol{\pi}} \subseteq J'_{I,\boldsymbol{\pi}}$, using the definition of $J'_{I,\boldsymbol{\pi}}$ together with the assumption that the number of observations of each arm in I is at least $1/\Delta^2$, we know that

$$T \sum_{i \notin I} \pi_i G_{i,j} \geq \frac{T}{2} \sum_{i \in [N]} \pi_i G_{i,j} \geq \frac{1}{2\Delta^2}$$

for every $j \in U'_{I,\boldsymbol{\pi}}$. Summing over $j \in U'_{I,\boldsymbol{\pi}}$ and te fact that each arm from V can provide at most $\log N$ observations of arms in $U'_{I,\boldsymbol{\pi}}$, we need the algorithm associated with $\boldsymbol{\pi}$ to play an action from V at least

$$\frac{|U_{I,\boldsymbol{\pi}}|}{2\Delta^2 \log N} \geq \frac{\delta^I(J_{I,\boldsymbol{\pi}})}{100\Delta^2 \log^2 N}$$

times to ensure enough observations in $U'_{I,\boldsymbol{\pi}}$. This gives us

$$T \sum_{i \notin I} \pi_i \geq \frac{\delta^I(J_{I,\boldsymbol{\pi}})}{100\Delta^2 \log^2 N}, \tag{7}$$

Applying (6) and (7) to the definition of Q^*_I gives us

$$\begin{aligned} Q^*_I = \max_{\Delta} \min_{\boldsymbol{\pi} \in \Pi} \min \left[T \sum_{i \in I} \pi_i \Delta + T \sum_{i \notin I} \pi_i, T\Delta \right] &\geq \max_{\Delta} \min_{\boldsymbol{\pi} \in \Pi} \min \left[\frac{\delta^I(J_{I,\boldsymbol{\pi}})}{100\Delta \log^2 N} + \frac{\delta^V(J'_{I,\boldsymbol{\pi}})}{100\Delta^2 \log^2 N}, T\Delta \right] \\ &\geq \max_{\Delta} \min_{J \subseteq I} \min \left[\frac{\delta^I(J)}{100\Delta \log^2 N} + \frac{\delta^V(I \setminus J)}{100\Delta^2 \log^2 N}, T\Delta \right] \end{aligned}$$

Now, we are able to choose a specific value of Δ to lower bound Q^*_I further. In particular, we choose the following two values of Δ

$$\begin{aligned} \Delta = \left(\tfrac{\delta^I(J)}{100T \log^2 N} \right)^{1/2} \quad &\Longrightarrow \quad Q^*_I \geq \left(\tfrac{T\delta^I(J)}{100 \log^2 N} \right)^{1/2} \\ \Delta = \left(\tfrac{\delta^V(I \setminus J)}{100T \log^2 N} \right)^{1/3} \quad &\Longrightarrow \quad Q^*_I \geq \left(\tfrac{T^2\delta^V(I \setminus J)}{100 \log^2 N} \right)^{1/3} \end{aligned}$$

Taking Δ which results in a larger lower bound, we obtain

$$Q^*_I \geq \frac{1}{10 \log N} \min_{J \subseteq I} \max \left[\delta^I(J)^{\frac{1}{2}} T^{\frac{1}{2}},\, \delta^V(I \setminus J)^{\frac{1}{3}} T^{\frac{2}{3}} \right] = \frac{1}{10 \log N} R^*_I.$$

This concludes the lower bound proof. The idea of the upper bound proof is to select a specific distribution $\boldsymbol{\pi} \in \Pi$ over the set of actions and choose the optimal Δ. This would give us the desirable upper bound on Q_I^*.

Let $J^* \subseteq I$ be one of the minimizers in the definition of R_I^*,

$$J^* \in \arg\min_{J \subseteq I} \max \left[\delta^I(J)^{\frac{1}{2}} T^{\frac{1}{2}},\, \delta^V(I \setminus J)^{\frac{1}{3}} T^{\frac{2}{3}} \right].$$

Let D be a dominating set of J^* from I and D' be a dominating set of $I \setminus J^*$ from V such that $|D| = \delta^I(J^*)$ and $D' = \delta^V(I \setminus J^*)$. For a given bandit problem, with optimal arm i^*, we define $\boldsymbol{\pi}$ as

$$\pi_i = \begin{cases} T^{-1}\Delta^{-2} & \text{if } i \in D \cup D' \setminus \{i^*\} \\ 1 - |D \cup D' \setminus \{i^*\}| T^{-1}\Delta^{-2} & \text{if } i = i^* \\ 0 & \text{otherwise.} \end{cases}$$

Distribution $\boldsymbol{\pi}$ is now a mixture of uniform distribution over $D \cup D' \setminus \{i^*\}$ with rest of the mass put on arm i^*. Note that for every arm $j \in I$, we know that

$$T \sum_{i \in [N]} \pi_i G_{i,j} \geq T \sum_{i \in D \cup D'} \pi_i G_{i,j} = T \sum_{i \in D \cup D'} \frac{1}{T\Delta^2} G_{i,j} \geq \frac{T}{T\Delta} = \frac{1}{\Delta^2},$$

from the definition of $\boldsymbol{\pi}$ and the fact that nodes from D and D' dominate every node in I. Therefore, $\boldsymbol{\pi}$ is one of the distributions in Π. Using this $\boldsymbol{\pi}$, we can upper bound Q_I^* as

$$\begin{aligned} Q_I^* &\leq \max_{\Delta} \min \left[\sum_{i \in D} \frac{T\Delta}{T\Delta^2} + \sum_{i \in D'} \frac{T}{T\Delta^2}, T\Delta \right] = \max_{\Delta} \min \left[\frac{|D|}{\Delta} + \frac{|D'|}{\Delta^2}, T\Delta \right] \\ &= \max_{\Delta} \min \left[\frac{\delta^I(J^*)}{\Delta} + \frac{\delta^V(I \setminus J^*)}{\Delta^2}, T\Delta \right] \leq 2 \max_{\Delta} \min \left[\max \left(\frac{\delta^I(J^*)}{\Delta}, \frac{\delta^V(I \setminus J^*)}{\Delta^2} \right), T\Delta \right]. \end{aligned}$$

The last expression is maximized for $\Delta = \max \left[\delta^I(J^*)^{\frac{1}{2}} T^{\frac{1}{2}},\, \delta^V(I \setminus J^*)^{\frac{1}{3}} T^{\frac{2}{3}} \right]$ which implies $Q_I^* \leq 2R_I^*$. □

C. Lower Bound Proofs

Before we proceed with the proof of Theorem 3.1, we need the following information-theoretic lemma that provides a foundation for the lower-bound proof.

Lemma C.1. *Let $G = (V, E)$ be a graph and fix a set $I \subseteq V$ with $|I| \geq 2$. Let $1/2 \geq \Delta > 0$ be such that $1/\Delta^2 \leq T/2^6$. For any algorithm, there exists a problem, such that regret R_T of the algorithm can be bounded as*

$$R_T \geq \frac{1}{2^7} \min_{\boldsymbol{\pi} \in \Pi} \min \left[T \sum_{i \in I} \pi_i \Delta + T \sum_{i \notin I} \pi_i,\, T\Delta \right]$$

where

$$\Pi = \left\{ \boldsymbol{\pi} \in \mathbb{R}^{+N} : \sum_{i \in [N]} \pi_i = 1, T \sum_{i \in [N]} \pi_i G_{i,j} \geq 1/\Delta^2, \forall j \in I \right\}.$$

Proof of Lemma C.1

Throughout the proof, we fix a set of nodes $I \subseteq V$, some node $j_0 \in I$, and any bandit policy. We use them for the rest of the proof unless said otherwise. The proof proceeds in several steps.

Step 1: Construction of a set of difficult problems. First, we define a set of difficult stochastic problems, indexed by $j \in I$, for a given set I. Then, we analyze the performance of the fixed algorithm on this set of problems. For problem j, the samples of each arm i are independent and distributed according to $P_i^{(j)} \triangleq \mathcal{N}(\mu_i^{(j)}, 1)$, parametrized by expectations $\mu_i^{(j)}$.

For j_0, we define the problem as

$$\mu_i^{(j_0)} \triangleq \begin{cases} 1/2 + \Delta/2 & \text{for} \quad i = j_0 \\ 1/2 & \text{for} \quad i \in I \setminus \{j_0\} \\ 0 & \text{for} \quad i \notin I \end{cases}$$

For $j \in I \setminus \{j_0\}$, we define the problem as

$$\mu_i^{(j)} \triangleq \begin{cases} 1/2 + \Delta & \text{for} \quad i = j \\ 1/2 + \Delta/2 & \text{for} \quad i = j_0 \\ 1/2 & \text{for} \quad i \in I \setminus \{j_0, j\} \\ 0 & \text{for} \quad i \notin I \end{cases}$$

there are several important observations to make:

- Arm $j \in I$ is optimal for problem j.
- The arms for problems $j \neq j_0$ and j_0 differ only for arm j and difference is Δ.
- For any problem j, arms that are not in I have a gap (distance from the optimal arm) at least $1/2$.

In what follows and for the policy we fixed at the beginning of the proof, we denoted the expectation and the probability, under the environment of bandit problem j, by $\mathbb{E}^{(j)}$ and $\mathbb{P}^{(j)}$.

Let $T_{T,i}$ denotes the number of rounds, up to round T, in which our fixed algorithm plays action i. We can associate the algorithm with a probability vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)$, under environment j_0, defined as

$$\pi_i \triangleq \frac{\mathbb{E}^{(j_0)}[T_{T,i}]}{T} \qquad \text{for all} \qquad i \in [N].$$

Each $T\pi_i$ represents the expected number of rounds our algorithm spends playing arm i, under environment j_0.

We also write $R_T^{(j)}$ for the expected regret of the fixed policy in problem j. Note that

$$R_T^{(j_0)} \geq \frac{T}{2}\Bigg[\sum_{i \in I \setminus \{j_0\}} \pi_i \Delta + \sum_{i \notin I} \pi_i\Bigg]. \tag{8}$$

For the rest of the proof, we assume the existence of arm $\bar{j} \in I \setminus \{j_0\}$ such that

$$T \sum_{i \in [N]} \pi_i G_{i,\bar{j}} \leq \frac{2}{\Delta^2}. \tag{9}$$

Note that the left-hand side of the previous inequality represents the expected number of observations of arm $\bar{j}$ using the algorithm under environment j_0. Markov inequality for any j gives us

$$\mathbb{P}^{(j_0)}\Bigg[\sum_{i \in [N]} G_{i,j} T_{T,i} \geq 2^4 T \sum_i G_{i,j} \pi_i\Bigg] \leq 2^{-4},$$

which for $\bar{j}$, using assumption (9), translates to

$$\mathbb{P}^{(j_0)}\Bigg[\sum_{i \in [N]} G_{i,\bar{j}} T_{T,i} \geq \frac{2^5}{\Delta^2}\Bigg] \leq 2^{-4}. \tag{10}$$

Now, let us define the event

$$F \triangleq \left\{T_{T,\bar{j}} \leq \frac{2^5}{\Delta^2}\right\}.$$

Note that complementary event $\bar{F}$ lower-bounds the number of rounds in which the algorithm played action $\bar{j}$ which is a sub-event of the event in (10) that lower-bounds the number of rounds in which the algorithm observed action $\bar{j}$. Therefore, we get

$$\mathbb{P}^{(j_0)}[\bar{F}] = \mathbb{P}^{(j_0)}\big[T_{T,\bar{j}} > 2^5/\Delta^2\big] \leq 2^{-4}. \tag{11}$$

Step 2: Bound on the regret using KL divergence. Note that action j_0 is optimal for problem j_0 and that action $\bar{j}$ is optimal for problem $\bar{j}$. Since choosing an action that is suboptimal leads to an instantaneous regret at least $\Delta/2$, we also have

$$R_T^{(\bar{j})} \geq \left(T - \frac{2^5}{\Delta^2}\right) \frac{\Delta}{2} \mathbb{P}^{(\bar{j})}\left[F\right] \geq \frac{T\Delta}{4} \mathbb{P}^{(\bar{j})}\left[F\right]. \tag{12}$$

Then, Bretagnolle-Huber inequality (see, e.g., (Lattimore & Szepesvári, 2020, Theorem 14.2)) implies that

$$\mathbb{P}^{(j_0)}\left[\bar{F}\right] + \mathbb{P}^{(\bar{j})}\left[F\right] \geq \frac{1}{2} \exp\Big(- \mathrm{KL}\left(\mathbb{P}^{(j_0)}, \mathbb{P}^{(\bar{j})}\right)\Big).$$

Applying Equation (11) implies

$$\mathbb{P}^{(\bar{j})}\left[F\right] \geq \frac{1}{2} \exp\Big(- \mathrm{KL}\left(\mathbb{P}^{(j_0)}, \mathbb{P}^{(\bar{j})}\right)\Big) - 2^{-4}.$$

This allows us to further bound $R_T^{(\bar{j})}$ in (12) as

$$R_T^{(\bar{j})} \geq \frac{T\Delta}{4} \left[\frac{1}{2} \exp\Big(- \mathrm{KL}\left(\mathbb{P}^{(j_0)}, \mathbb{P}^{(\bar{j})}\right)\Big) - 2^{-4}\right]. \tag{13}$$

Step 3: Information-theoretic bound on the number of pulls. Because of the graph structure, the number of observations of arm i at time T is $N_{T,i} \triangleq \sum_{k\in[N]} G_{k,i} T_{T,k}$. So that for any problem j, the Kullback-Leibler divergence between $\mathbb{P}^{(j)}$ and $\mathbb{P}^{(j_0)}$ can be rewritten as follows (Lattimore & Szepesvári, 2020, immediate corollary of Lemma 15.1)

$$\mathrm{KL}\left(\mathbb{P}^{(j_0)}, \mathbb{P}^{(\bar{j})}\right) = \sum_{i\in[N]} \mathbb{E}^{(j_0)}\left[N_{T,i}\right] \mathrm{KL}\left(\mathbb{P}_i^{(j_0)}, \mathbb{P}_i^{(\bar{j})}\right).$$

Using definition of bandit problems j_0 and $\bar{j}$, we get

$$\mathrm{KL}\left(\mathbb{P}^{(j_0)}, \mathbb{P}^{(\bar{j})}\right) = \mathbb{E}^{(j_0)}\left[\sum_{i\in[N]} G_{i,\bar{j}} T_{T,i}\right] \mathrm{KL}\left(\mathbb{P}_{\bar{j}}^{(j_0)}, \mathbb{P}_{\bar{j}}^{(\bar{j})}\right) = \left[\sum_{i\in[N]} G_{i,\bar{j}} \pi_i T\right] \frac{\Delta^2}{2}$$

since, for any $j \in I \setminus \{j_0\}$, problems j and j_0 differ only on action j and since $\mathrm{KL}(\mathbb{P}_j^{(j_0)}, \mathbb{P}_j^{(\bar{j})}) = \Delta^2/2$. Re-ordering terms and using the assumption in Equation (9), we get

$$\mathrm{KL}\left(\mathbb{P}^{(j_0)}, \mathbb{P}^{(\bar{j})}\right) = \frac{\Delta^2}{2} T \sum_{i\in[N]} G_{i,\bar{j}} \pi_i \leq 1 \tag{14}$$

Step 4: Lower bound on the regret depending on $\boldsymbol{\pi}$. Combining Equations (14) and (13), we get

$$R_T^{(\bar{j})} \geq \frac{T\Delta}{4} \left[\frac{1}{2} \exp\left(-1\right) - 2^{-4}\right] \geq \frac{T\Delta}{2^6}.$$

In case an arm $\bar{j}$, satisfying condition (9), exists, regret $R_T^{(\bar{j})}$ is bounded by $(T\Delta)/2^6$. In case no such arm $\bar{j}$ exists, i.e. $T\sum_i \pi_i G_{i,j} \geq 2/\Delta^2$ for all $j \in I \setminus \{j_0\}$, we can bound $R_T^{(j_0)}$, using (8), as

$$R_T^{(j_0)} \geq \frac{T}{2}\Big[\sum_{i\notin I\setminus\{j_0\}} \pi_i \Delta + \sum_{i\notin I} \pi_i\Big].$$

Therefore, we have that for any policy, it holds that

$$R^* \geq \max_{j\in I} R_T^{(j)} \geq \min_{\pi\in\mathbb{R}^{+N}: \sum_i \pi_i = 1, T\sum_i \pi_i G_{i,j} \geq 2/\Delta^2, \forall j\in I\setminus\{j_0\}} \min\left[\frac{T}{2}\Big[\sum_{i\in I\setminus\{j_0\}} \pi_i \Delta + \sum_{i\notin I} \pi_i\Big], \frac{T\Delta}{2^6}\right].$$

We have proved this inequality for any $j_0 \in I$ so that if we take two $j_0', j_0'' \in I : j_0' \neq j_0''$, we would get:

$$R^* \geq \max_{j_0 \in \{j_0', j_0''\}} \min_{\pi \in \mathbb{R}^{+N} : \sum_i \pi_i = 1, T \sum_i \pi_i G_{i,j} \geq 2/\Delta^2, \forall j \in I \setminus \{j_0\}} \min \left[\frac{T}{2} \Big[\sum_{i \in I \setminus \{j_0\}} \pi_i \Delta + \sum_{i \notin I} \pi_i \Big], \frac{T\Delta}{2^6} \right].$$

Since for any function A of π we have $\max(\min_{\pi : \pi_i > a_i} A(\pi), \min_{\pi : \pi_i > b_i} A(\pi)) \geq \min_{\pi : \pi_i > (a_i + b_i)/2} A(\pi)/2$ this implies the result of the lemma, namely

$$R^* \geq \frac{1}{2^7} \min_{\pi \in \mathbb{R}^{+N} : \sum_i \pi_i = 1, T \sum_i \pi_i G_{i,j} \geq 1/\Delta^2, \forall j \in I} \min \left[T \sum_{i \in I} \pi_i \Delta + T \sum_{i \notin I} \pi_i, T\Delta \right].$$

C.1. Proof of Theorem 3.1

The starting point of the proof is the statement of Lemma C.1. From the definition of the problem complexity Q^*, we can directly show that for any $I \subset V$

$$R_T \geq \frac{Q_I^*}{2^7}.$$

Since this inequality holds for any $I \subset V$, it holds also for I that maximizes Q_I^*. This implies that

$$R_T \geq \max_{I \subset V} \frac{Q_I^*}{2^7} = \frac{Q^*}{2^7}. \tag{15}$$

Applying Lemma 2.8 to (15), we have

$$R_T \geq \frac{Q^*}{2^7} \geq \frac{R^*}{2^7 10 \log N}. \quad \square$$

D. Efficient Construction of the Proxies $\bar{J}_{t,k,l}$, $\bar{J}'_{t,k,l}$ and Dominating Sets $\bar{D}_{t,k,l}$, $\bar{D}'_{t,k,l}$ from Corollary 4.2

In this section, we construct the partition of the set $I_{t,k,l}$ through sets $\bar{J}_{t,k,l}$, $\bar{J}'_{t,k,l}$ and the associated dominating sets $\bar{D}_{t,k,l}$, $\bar{D}'_{t,k,l}$ from Corollary 4.2.

In order to do this, we provide a construction for an arbitrary set I, and can then apply this procedure to the $I_{t,k,l}$. Consider therefore any arbitrary set $I \subset V$.

For any $1 \geq \Delta > 0$, consider the optimisation problem $Q_{I,\Delta}^*$ from Definition 2.6, namely:

$$Q_{I,\Delta}^* \triangleq \min_{\boldsymbol{\pi} \in \Pi} Q_{I,\Delta}(\pi) \triangleq \min_{\boldsymbol{\pi} \in \Pi} \min \left[T \sum_{i \in I} \pi_i \Delta + T \sum_{i \notin I} \pi_i,\ T\Delta \right] \tag{16}$$

where

$$\Pi = \Pi^\Delta = \left\{ \boldsymbol{\pi} \in \mathbb{R}_+^N : \sum_{i \in [N]} \pi_i \leq 1,\ T \sum_{i \in [N]} \pi_i G_{i,j} \geq 1/\Delta^2, \forall j \in I \right\}.$$

This is a linear optimization problem under linear constraints. If the set of solutions is not empty, we know e.g. by using Algorithm *Main* from (Cohen et al., 2019) with precision $\Delta^2/(N^4 T^2)$ that there there is a solver for such problem doing less than $N^3 \log(T^2 N^5/\Delta^2)$ basic operations and that would output a solution π^Δ such that $\pi^\Delta \in \Pi$ and

$$Q_{I,\Delta}^* \leq Q_{I,\Delta}(\pi^\Delta) \leq Q_{I,\Delta}^* + 1. \tag{17}$$

See Theorem 1 from (Cohen et al., 2019) for more details. From there, we use π^Δ, which is the solution of algorithm outputted by Algorithm *Main* from (Cohen et al., 2019) with precision $\Delta^2/(N^4 T^2)$ on the optimization problem from above (characterized by Δ).

Based on this, the procedure that we use is described in Algorithm 2. We consider a logarithmic grid of Δ of size $\lfloor \log(NT) \rfloor + 2$, and associated to the policies $\bar{\pi}^\Delta$ constructed on this grid, we construct sets $\bar{J}^\Delta = J_{I,\pi^\Delta}$ as in (4), and the

Algorithm 2 ALGORITHM OUTPUTTING PROXIES OF J, J' AND OF THEIR DOMINATING SETS

Input: $G = (V, E)$, set $I \subset V$.
for $\Delta \in \{2^{-1}, 2^{-2}, ..., 2^{-(\lfloor \log(NT) \rfloor + 1)}\}$ **do**
  Compute $\bar{\pi}^{\Delta}$ as the output of Algorithm *Main* from (Cohen et al., 2019) applied on Problem 16 with precision $\Delta^2/(N^4T^2)$
  Set $\bar{J}^{\Delta} = J_{I,\pi^{\Delta}}$ (as defined in Equation (4))
  Set $\bar{J}^{',\Delta} = I \setminus J^{\Delta}$
  Set $\bar{D}^{\Delta}$ as the output of Algorithm 3 on $\bar{J}^{\Delta}$
  Set $\bar{D}^{',\Delta}$ as the output of Algorithm 3 on $\bar{J}^{',\Delta}$
end for
Set $\bar{\Delta} = \arg\min_{\Delta}\{\sqrt{\bar{D}^{\Delta}T} \wedge [(\bar{D}^{',\Delta})^{1/3}T^{2/3}]\}$
Set $\bar{\pi} = \pi^{\bar{\Delta}}, \bar{J} = \bar{J}^{\bar{\Delta}}, \bar{J}' = \bar{J}^{',\bar{\Delta}}, \bar{D} = \bar{D}^{\bar{\Delta}}, \bar{D}' = \bar{D}^{',\bar{\Delta}}$
Output: $\bar{J}, \bar{J}', \bar{D}, \bar{D}'$

associated complement $\bar{J}^{',\Delta}$, and then compute a greedy approximation of their dominating sets as described in Algorithm 3. The description and discussion of Algorithm 3 is postponed to Subsection D.2, as it is a very standard result. We finally take the stationary policy $\bar{\pi}$ corresponding to the best possible case π^{Δ} for a criterion that resembles the one of Definition 2.7, and output the associated sets $\bar{J}, \bar{J}', \bar{D}, \bar{D}'$.

The following lemma holds for Algorithm 2.

Lemma D.1. *Let $G = (V, E)$ be a graph and $I \subset V$ be any subset of vertices. Then Algorithm 2 applied to I runs in less than $30N^3 \log(TN)^2$ iterations, and produces sets $\bar{J} \subset I$, $\bar{J}' \triangleq I \setminus \bar{J}$, $\bar{D} \subset I$, and $\bar{D}' \subset V$ such that $\bar{D}$ dominates set $\bar{J}$, $\bar{D}'$ dominates $\bar{J}'$,*

$$\max\left\{\delta^I(\bar{J})^{\frac{1}{2}}T^{\frac{1}{2}},\ \delta^V(\bar{J}')^{\frac{1}{3}}T^{\frac{2}{3}}\right\} \leq 24 \times 10^4 \log(N)^4 R^*_I,$$

and

$$|\bar{D}| \leq \delta^I(\bar{J}) \log(|I|), \qquad |\bar{D}'| \leq \delta^V(\bar{J}') \log(|V|).$$

Corollary 4.2 is a direct application of this lemma.

D.1. Proof of Lemma D.1

Let $\Delta \in \{2^{-1}, 2^{-2}, ..., 2^{-(\lfloor \log(NT) \rfloor + 1)}\}$. Note fist that for any $1/2 \geq \Delta' > 0$ such that $\Delta'/2 \leq \Delta \leq \Delta'$, we have for any policy π that

$$Q_{I,\Delta'}(\pi)/2 \leq Q_{I,\Delta}(\pi) \leq Q_{I,\Delta'}(\pi),$$

by definition of $Q_{I,\Delta}(.)$. Moreover for $\Delta' \leq 2^{-(\lfloor \log(NT) \rfloor + 1)}$, we have that for any stationary policy π:

$$Q_{I,\Delta'}(\pi) \leq T\Delta' \leq 1/N.$$

So that if, for any Δ', we write $G(\Delta')$ for the projection of Δ' on the closest larger element of $\Delta \in \{2^{-1}, 2^{-2}, ..., 2^{-(\lfloor \log(NT) \rfloor + 1)}\}$, we have

$$Q_{I,\Delta'}(\pi)/2 \leq Q_{I,G(\Delta')}(\pi) + 1/N \leq Q_{I,\Delta'}(\pi) + 2/N.$$

This first implies taking $\pi = \pi^{G(\Delta')}$ and since $\Pi^{G(\Delta')} \in 2\Pi^{\Delta'}$ (where $2\Pi^{\Delta'}$ is the set $\{x : x/2 \in 2\Pi^{\Delta'}\}$)

$$Q^{\star}_{I,\Delta'}/2 \leq Q^{\star}_{I,G(\Delta')} + 1/N.$$

This also implies by taking the minimum over $\pi \in \Pi^{\Delta'}$

$$Q^{\star}_{I,\Delta'} \leq \min_{\pi \in \Pi^{\Delta'}} Q_{I,G(\Delta')}(\pi) + 1/N \leq 2Q^{\star}_{I,\Delta'} + 2/N.$$

And so since $\Pi^{\Delta'} \subset \Pi^{G(\Delta')}$ we have

$$Q^\star_{I,G(\Delta')} + 1/N \leq Q^\star_{I,\Delta'} + 2/N.$$

So that in the end for any Δ'

$$Q^\star_{I,\Delta'}/2 \leq Q^\star_{I,G(\Delta')} \leq 2Q^\star_{I,\Delta'} + 1/N.$$

So that by Equation (17)

$$Q^\star_{I,\Delta'}/2 \leq Q_{I,G(\Delta')}(\bar{\pi}^{G(\Delta')}) \leq 2Q^\star_{I,\Delta'} + 1/N + 1.$$

Set $\Delta = G(\Delta')$. Following exactly the same steps as in the proof of Proposition 2.8 in Section B, we know that for $J_{I,\bar{\pi}^\Delta}$ defined as in Equation (4) we have

$$Q_{I,\Delta}(\bar{\pi}^\Delta) \geq \frac{1}{100\log^2 N} \min\left[\frac{\delta^I(J_{I,\bar{\pi}^\Delta})}{\Delta} + \frac{\delta^V(I \setminus J_{I,\bar{\pi}^\Delta})}{\Delta^2}, T\Delta\right].$$

We remind that $J, J' = I \setminus J$ minimise the equation in Definition (2.7). Consider now $\Delta^\star = G(400\log^2 N\left[\sqrt{\frac{\delta^I(J)}{T}} \vee \left(\frac{\delta^V(J')}{T}\right)^{1/3}\right])$. For $\Delta^\star$ in particular, we have from the previous equation and from Equation (17)

$$2Q^\star_{I,\Delta^\star} + 1 + 1/N \geq Q_{I,\Delta^\star}(\bar{\pi}^{\Delta^\star}) \geq \frac{1}{100\log^2 N} \min\left[\frac{\delta^I(J_{I,\bar{\pi}^{\Delta^\star}})}{\Delta^\star} + \frac{\delta^V(I \setminus J_{I,\bar{\pi}^{\Delta^\star}})}{\Delta^{\star 2}}, T\Delta^\star\right].$$

So that from Proposition 2.8 and the last equation

$$\begin{aligned}4\max\left[\sqrt{\delta^I(J)T}, (\delta^V(J))^{1/3}T^{2/3}\right] &\geq 4R^*_I \geq 2Q^*_I \geq 2Q^\star_{I,\Delta^\star} \\ &\geq \frac{1}{100\log^2 N} \min\left[\frac{\delta^I(J_{I,\bar{\pi}^{\Delta^\star}})}{\Delta^\star} + \frac{\delta^V(I \setminus J_{I,\bar{\pi}^{\Delta^\star}})}{\Delta^{\star 2}}, T\Delta^\star\right] - 1 - 1/N.\end{aligned}$$

Since by definition of $\Delta^\star$ we have that $\frac{1}{100\log^2 N}T\Delta^\star - 1 - 1/N > 4\max\left[\sqrt{J^\star T}, (J'^\star)^{1/3}T^{2/3}\right]$ this implies

$$4\max\left[\sqrt{\delta^I(J)T}, (\delta^V(J'))^{1/3}T^{2/3}\right] \geq \frac{1}{100\log^2 N}\left[\frac{\delta^I(J_{I,\bar{\pi}^{\Delta^\star}})}{\Delta^\star} + \frac{\delta^V(I \setminus J_{I,\bar{\pi}^{\Delta^\star}})}{\Delta^{\star 2}}\right] - 1 - 1/N.$$

By definition of $\Delta^\star$, we have

$$4\max\left[\sqrt{\delta^I(J)T}, (\delta^V(J'))^{1/3}T^{2/3}\right] \geq \frac{1}{8\times 10^4\log^4 N}\left[\frac{\delta^I(J_{I,\bar{\pi}^{\Delta^\star}})}{\sqrt{\delta^I(J)}}\sqrt{T} + \frac{\delta^V(I \setminus J_{I,\bar{\pi}^{\Delta^\star}})}{\delta^V(J')^{2/3}}T^{2/3}\right] - 1 - 1/N,$$

i.e.

$$48\times 10^4\log^4 N\max\left[\sqrt{\delta^I(J)T}, (\delta^V(J'))^{1/3}T^{2/3}\right] \geq \max\left[\frac{\delta^I(J_{I,\bar{\pi}^{\Delta^\star}})}{\sqrt{\delta^I(J)}}\sqrt{T}, \frac{\delta^V(I \setminus J_{I,\bar{\pi}^{\Delta^\star}})}{\delta^V(J')^{2/3}}T^{2/3}\right].$$

So that in the end

$$48\times 10^4\log^4 N\max\left[\sqrt{\delta^I(J)T}, (\delta^V(J'))^{1/3}T^{2/3}\right] \geq \max\left[\sqrt{\delta^I(J_{I,\bar{\pi}^{\Delta^\star}})T}, \delta^V(I \setminus J_{I,\bar{\pi}^{\Delta^\star}})^{1/3}T^{2/3}\right].$$

By Subsection D.2, we therefore have

$$48\times 10^4\log^5 N\max\left[\sqrt{\delta^I(J)T}, (\delta^V(J'))^{1/3}T^{2/3}\right] \geq \max\left[\sqrt{\bar{D}_{I,\bar{\pi}^{\Delta^\star}}T}, (\bar{D}'_{I,\bar{\pi}^{\Delta^\star}})^{1/3}T^{2/3}\right].$$

Since by definition of the algorithm we have $\max\left[\sqrt{\bar{D}_{I,\bar{\pi}^{\Delta^\star}}T}, (\bar{D}'_{I,\bar{\pi}^{\Delta^\star}})^{1/3}T^{2/3}\right] \geq \max\left[\sqrt{\bar{D}T}, (\bar{D}')^{1/3}T^{2/3}\right]$, this concludes the proof.

Algorithm 3 GREEDY ALGORITHM FOR DOMINATING SET

Input:
$G = (V, E)$, sets $A, B \subset V$ such that A dominates B.
$D = \emptyset$
repeat
  $d = \arg\max_{v \in A} |N_v^{out} \cap B|$ (ties resolved arbitrarily)
  $D = D \cup d$
  $B = B \setminus N_d^{out}$
until —B— = 0
Output: D

D.2. Greedy Algorithm for Dominating Set

Finding the smallest dominating set is an NP-hard problem, however, a simple greedy algorithm can find an approximate solution, only a logarithmic factor away from the optimal solution. The algorithm is described in Algorithm 3 and the theoretical guarantees can be found in Theorem D.2.

Theorem D.2. *Let $G = (V, E)$ be a graph with two sets of vertices $A, B \subset V$ such that A dominates B. Then Algorithm 3 produces set $D \subset A$ that dominates B from A, such that*

$$|D| \leq \log(N)\delta^A(B).$$

Moreover, the computational complexity of Algorithm 3 is at most linear in the number of vertices.

This is a standard result that can be found for example in (Chvatal, 1979).

E. Proof of Theorem 4.4

We start the proof with a standard proposition that can be used as the first step in the analysis of most of the algorithms based on EXP3

Proposition E.1. *Let everything be defined as in Algorithm 1. Then we have*

$$\mathbb{E}\left[\sum_{t\in[T]}\sum_{i\in[N]} q_{t,i}\ell_{t,i} - \min_{k\in[N]}\sum_{t\in[T]}\ell_{t,k}\right]\mathbb{E}\left[\frac{\log N}{\eta_{T+1}}\right] + \mathbb{E}\left[\sum_{t\in[T]}\frac{\eta_t}{2}\sum_{i\in[N]}\frac{q_{t,i}}{P_{t,i}}\right].$$

Proof. The proof of this proposition is based on the proof by (Györfi & Ottucsák, 2007). First, lets define W'_{t+1}, similarly to W_{t+1}, using learning rate from round t instead of $t+1$.

$$W'_{t+1} = \frac{1}{N}\sum_{i\in[N]}\exp(-\eta_t \widehat{L}_{t+1,i}).$$

Following standard analysis of EXP3 algorithms with adaptive learning rate (e.g. (Kocák et al., 2014)), we can obtain

$$\begin{aligned}
\frac{1}{\eta_t}\log\frac{W'_{t+1}}{W_t} &= \frac{1}{\eta_t}\log\sum_{i\in[N]}\frac{(1/N)\exp(-\eta_t\widehat{L}_{t,i})}{W_t}\\
&= \frac{1}{\eta_t}\log\sum_{i\in[N]}\frac{w_{t,i}\exp(-\eta_t\hat{\ell}_{t,i})}{W_t}\\
&= \frac{1}{\eta_t}\log\sum_{i\in[N]}q_{t,i}\exp(-\eta_t\hat{\ell}_{t,i})\\
&\le \frac{1}{\eta_t}\log\sum_{i\in[N]}q_{t,i}\left(1-\eta_t\hat{\ell}_{t,i}+\frac{1}{2}(\eta_t\hat{\ell}_{t,i})^2\right)\\
&= \frac{1}{\eta_t}\log\left(1-\eta_t\sum_{i\in[N]}q_{t,i}\hat{\ell}_{t,i}+\frac{\eta_t^2}{2}\sum_{i\in[N]}q_{t,i}(\hat{\ell}_{t,i})^2\right)\\
&\le -\sum_{i\in[N]}q_{t,i}\hat{\ell}_{t,i}+\frac{\eta_t}{2}\sum_{i\in[N]}q_{t,i}(\hat{\ell}_{t,i})^2,
\end{aligned}$$

where we used inequality $\exp(-x)\le 1-x+x^2/2$ for $x\ge 0$ as well as inequality $\log(1-x)\le -x$ that holds for any x. Rearranging the terms in the previous inequality, we obtain

$$\sum_{i\in[N]}q_{t,i}\hat{\ell}_{t,i}\le\left[\left(\frac{\log W_t}{\eta_t}-\frac{\log W_{t+1}}{\eta_{t+1}}\right)+\left(\frac{\log W_{t+1}}{\eta_{t+1}}-\frac{\log W'_{t+1}}{\eta_t}\right)\right]+\frac{\eta_t}{2}\sum_{i\in[N]}q_{t,i}(\hat{\ell}_{t,i})^2 \tag{18}$$

The second term in the brackets can be further bounded using

$$W_{t+1}=\sum_{i\in[N]}\exp(-\eta_{t+1}\widehat{L}_{t,i})=\sum_{i\in[N]}\exp(-\eta_t\widehat{L}_{t,i})^{\frac{\eta_{t+1}}{\eta_t}}\le\left(\sum_{i\in[N]}\exp(-\eta_t\widehat{L}_{t,i})\right)^{\frac{\eta_{t+1}}{\eta_t}}=(W'_{t+1})^{\frac{\eta_{t+1}}{\eta_t}},$$

where we applied Jensen's inequality to the concave function $x^{\frac{\eta_{t+1}}{\eta_t}}$ thanks to the assumption that $\eta_{t+1}\le\eta_t$. Therefore, we obtain

$$\left(\frac{\log W_{t+1}}{\eta_{t+1}}-\frac{\log W'_{t+1}}{\eta_t}\right)\le 0,$$

which further simplifies (18) to obtain

$$\sum_{i\in[N]}q_{t,i}\hat{\ell}_{t,i}\le\left(\frac{\log W_t}{\eta_t}-\frac{\log W_{t+1}}{\eta_{t+1}}\right)+\frac{\eta_t}{2}\sum_{i\in[N]}q_{t,i}(\hat{\ell}_{t,i})^2$$

The next step is summing over time and taking expectation

$$\mathbb{E}\left[\sum_{t\in[T]}\sum_{i\in[N]}q_{t,i}\ell_{t,i}\right]\le\mathbb{E}\left[\frac{\log W_{T+1}}{\eta_{T+1}}\right]+\mathbb{E}\left[\sum_{t\in[T]}\frac{\eta_t}{2}\sum_{i\in[N]}\frac{q_{t,i}}{P_{t,i}}\ell_{t,i}^2\right].$$

Lower-bounding W_{T+1} by $\max_{k\in[N]}w_{T+1,k}$ and using the definition of exponential weights, we obtain

$$\mathbb{E}\left[\sum_{t\in[T]}\sum_{i\in[N]}q_{t,i}\ell_{t,i}-\min_{k\in[N]}\sum_{t\in[T]}\ell_{t,k}\right]\le\mathbb{E}\left[\frac{\log N}{\eta_{T+1}}\right]+\mathbb{E}\left[\sum_{t\in[T]}\frac{\eta_t}{2}\sum_{i\in[N]}\frac{q_{t,i}}{P_{t,i}}\right].$$

□

Using the definition of $p_{t,i}$ from Equation (2), we can lower bound the left-hand side of the inequality in Proposition E.1 as

$$\mathbb{E}\left[\sum_{t\in[T]}\sum_{i\in[N]} q_{t,i}\ell_{t,i} - \min_{k\in[N]}\sum_{t\in[T]}\ell_{t,k}\right] \geq R_T - 2\mathbb{E}\left[\sum_{t\in[T]}\gamma_t\right].$$

This changes the inequality from Proposition E.1 to

$$R_T \leq \mathbb{E}\left[2\sum_{t\in[T]}\gamma_t + \frac{\log N}{\eta_{T+1}} + \sum_{t\in[T]}\frac{\eta_t}{2}\sum_{i\in[N]}\frac{q_{t,i}}{P_{t,i}}\right]. \tag{19}$$

The next step is to upper bound the last term inside of the expectation, starting with the inner most sum. We can split this sum into several parts, depending on the partition in which the arm is situated and upper bound each par separately.

$$\sum_{i\in[N]}\frac{q_{t,i}}{P_{t,i}} = \sum_{k\in[K]}\sum_{l\in[L]}\sum_{i\in\bar{J}_{t,k,l}}\frac{q_{t,i}}{P_{t,i}} \tag{20}$$

$$+\sum_{k\in[K]}\sum_{l\in[L]}\sum_{i\in\bar{J}'_{t,k,l}}\frac{q_{t,i}}{P_{t,i}} \tag{21}$$

$$+\sum_{i\in I_{t,K+1}}\frac{q_{t,i}}{P_{t,i}} \tag{22}$$

Bounding expression (20). Partition $\bar{J}_{t,k,l}$ contains only arms i with $q_{t,i}\in(2^{-k},2^{-k+1}]$ and $\deg_{t,k}(i)\in(N2^{-l},N2^{-l+1}]$. Therefore, $q_{t,i}$ can be simply upper bounded by the largest possible value of $q_{t,i}$, which is 2^{-k+1}. We also know that $i\in\bar{J}_{t,k,l}$ has at least $N2^{-l}$ neighbors in $I_{t,k}$ each of which has the probability of being played $p_{t,j}$ which is at least $q_{t,j}/2$ from the definition of $p_{t,j}$ and the fact that $\gamma_t\leq 1/2$. Now, we lower bound $P_{t,i}$ by the smallest possible number of neighbors in $I_{t,k}$ and the smallest possible value of $q_{t,j}$ for neighbors j of i to obtain $P_{t,i}\geq N2^{-l}2^{-k}2^{-1}$. This gives us the following upper bound on (20)

$$(20)\leq\sum_{k\in[K]}\sum_{l\in[L]}\frac{|\bar{J}_{t,k,l}|2^{-k+1}}{N2^{-l}2^{-k}2^{-1}}\leq\sum_{k\in[K]}\sum_{l\in[L]}\frac{|\bar{J}_{t,k,l}|2^{3}}{N2^{-l+1}}\leq\sum_{k\in[K]}\sum_{l\in[L]}2^3|\bar{D}_{t,k,l}|.$$

The last inequality holds thanks to the fact that $|\bar{D}_{t,k,l}|$ is the size of the dominating set of $\bar{J}_{t,k,l}$ with every vertex dominating at most $N2^{-l+1}$ other nodes. This means that the number of possible edges $|\bar{D}_{t,k,l}|N2^{-l+1}$ from $\bar{D}_{t,k,l}$ to $\bar{J}_{t,k,l}$ needs to be larger than the number of vertices in $\bar{J}_{t,k,l}$.

Bounding expression (21). Since $\bar{D}'_{t,k,l}$ is a dominating set of $\bar{J}'_{t,k,l}$, we know that every vertex of $\bar{J}'_{t,k,l}$ is dominated by at least one vertex from the dominating set $\bar{D}'_{t,k,l}$. For each vertex of this dominating set, we used a mixing in Definition 4.3, and therefore we observe every vertex of $\bar{D}'_{t,k,l}$ with probability at least $\gamma_t/((KL+1)|\bar{D}'_{t,k,l}|)$. As a consequence, we have that for every $i\in J'_{t,k,l}$, probability of observing i is at least $\gamma_t/((KL+1)|D'_{t,k,l}|)$. Thisgives us the following upper bound on (21)

$$(21)\leq\sum_{k\in[K]}\sum_{l\in[L]}\frac{(KL+1)|\bar{D}'_{t,k,l}|}{\gamma_t}$$

Bounding expression (22). We know that every arm i from $I_{t,K+1}$ is such that $q_{t,i}$, by definition, is smaller than $1/N^5$. From the definition of exploration distribution in Definition 4.3, we know that the corresponding $P_{t,i}$ is lower bounded by $\gamma_t/((KL+1)N)$. This gives us the following upper bound on (22)

$$(22)\leq\sum_{i\in I_{t,K+1}}\frac{(KL+1)N}{\gamma_t N^5}\leq\frac{1}{\gamma_t}.$$

The last inequality holds thanks to the fact that $(KL+1)\leq N^2$, for any positive integer N, from the definition of K and L. Note that in particular

$$(22)\leq\sum_{k\in[K]}\sum_{l\in[L]}\frac{(KL+1)\log(N)|\bar{D}'_{t,k,l}|}{\gamma_t},$$

i.e. we can upper bound the last term using the same expression as in the upper bound of (21).

Using all three bounds together, we obtain the following upper bound for $\frac{\eta_t}{2}\sum_{i\in[N]}\frac{q_{t,i}}{P_{t,i}}$

$$\frac{\eta_t}{2}\sum_{i\in[N]}\frac{q_{t,i}}{P_{t,i}} \leq \frac{\eta_t}{2}\sum_{\substack{k\in[K]\\ l\in[L]}}\left(8|\bar{D}_{t,k,l}| + \frac{2(KL+1)\log(N)|\bar{D}'_{t,k,l}|}{\gamma_t}\right). \tag{23}$$

Now we can define an auxiliary learning rate $\eta_{t,k,l}$ for each individual summand as

$$\eta_{t,k,l} \triangleq \min\left(|\bar{D}_{t,k,l}|^{-\frac{1}{2}}T^{-\frac{1}{2}}, (|\bar{D}'_{t,k,l}|)^{-\frac{1}{3}}T^{-\frac{2}{3}}\right)$$

Note that η_t is defined as $\min_{s\in[t],k\in[K],l\in[L]}\eta_{s,k,l}$ and therefore we can upper bound it by $\eta_{t,k,l}$ in (23) to obtain

$$\begin{aligned}
\frac{\eta_t}{2}\sum_{i\in[N]}\frac{q_{t,i}}{P_{t,i}} &\leq \sum_{\substack{k\in[K]\\ l\in[L]}}\left(4\eta_t|\bar{D}_{t,k,l}| + \eta_t^2 T(C-4)|\bar{D}'_{t,k,l}|\right)\\
&\leq \sum_{\substack{k\in[K]\\ l\in[L]}}\left(4\eta_{t,k,l}|\bar{D}_{t,k,l}| + \eta_{t,k,l}^2 T(C-4)|\bar{D}'_{t,k,l}|\right)\\
&= C\sum_{\substack{k\in[K]\\ l\in[L]}}\max\left(|\bar{D}_{t,k,l}|^{\frac{1}{2}}T^{-\frac{1}{2}}, |\bar{D}'_{t,k,l}|^{\frac{1}{3}}T^{-\frac{1}{3}}\right)\\
&\leq C\sum_{\substack{k\in[K]\\ l\in[L]}} 24\times 10^4\log(N)^4\log N\frac{R^*}{T}
\end{aligned}$$

for $C = 4 + (KL+1)\log(N)$. Summing over time, we obtain the following upper bound

$$\sum_{t\in[T]}\frac{\eta_t}{2}\sum_{i\in[N]}\frac{q_{t,i}}{P_{t,i}} \leq C\sum_{\substack{k\in[K]\\ l\in[L]}}\max\left(\delta_*^{\frac{1}{2}}T^{\frac{1}{2}}, \delta'^{\frac{1}{3}}_*T^{\frac{2}{3}}\right) \tag{24}$$

Now we are able to bound the last term in (19). The next step is bounding the first two terms to obtain the regret upper bound. In order to do so, we can use the definitions of γ_t and η_t and the fact that $\{\eta_t\}_{t\in[T]}$ is a non-increasing sequence to obtain

$$\begin{aligned}
2\sum_{t\in[T]}\gamma_t + \frac{\log N}{\eta_{T+1}} &= 2\sum_{t\in[T]}\min\left\{\frac{1}{T\eta_t}, \frac{1}{2}\right\} + \frac{\log N}{\eta_{T+1}}\\
&\leq 2\sum_{t\in[T]}\frac{1}{T\eta_t} + \frac{\log N}{\eta_{T+1}}\\
&\leq 2\sum_{t\in[T]}\frac{1}{T\eta_{T+1}} + \frac{\log N}{\eta_{T+1}} = \frac{2+\log N}{\eta_{T+1}}\\
&\leq 24\times 10^4\log(N)^4\log N(2+\log N)R^*
\end{aligned} \tag{25}$$

The proof is concluded by applying bounds (24) and (25) to expression (19)

F. Discussion

This section covers the proofs of corollaries in Section 5.

F.1. Proof of Corollary 5.1

The starting point of this proof is the Definition 2.7. The idea is to show that for $T \geq \alpha^3$, it is optimal to take $J = I$, in the minimization part of the problem complexity definition, regardless of set I.

Let us fix I and assume that the optimization problem is maximized for some $J \neq I$. This also means that $I \setminus J$ is a non-empty set and therefore, needs to be dominated by at least one node, i.e. $\delta^V(I \setminus J) \geq 1$. Note that the independence number of a graph is always an upper bound on the dominating number since the largest independent set is connected to all the vertices in the graph. This means that the independence number α of G can be lower bounded by the independence number $\delta^I(I)$ of the graph induced by I which in turn can be lower bounded by $\delta^I(J)$, for any $J \subseteq I$. Using these observations, together with assumption $T \geq \alpha^3$, we get

$$\delta^V(I \setminus J)^{\frac{1}{3}} T^{\frac{2}{3}} \geq T^{\frac{2}{3}} \geq \alpha^{\frac{1}{2}} T^{\frac{1}{2}} \geq \delta^V(V)^{\frac{1}{2}} T^{\frac{1}{2}} \geq \delta^I(I)^{\frac{1}{2}} T^{\frac{1}{2}} \geq \delta^I(J)^{\frac{1}{2}} T^{\frac{1}{2}}.$$

However, setting $J = I$ would decrease $\delta^V(I \setminus J)^{\frac{1}{3}} T^{\frac{2}{3}}$ to 0 and therefore, improve the minimization problem from the problem complexity definition. Now that we know that the optimal J is equal to I, we are ready to find I that maximizes the problem complexity. Let I be the largest independent set. The size of I is now α and the only way to dominate $J = I$ using only nodes from I is by using all the nodes, therefore, $\delta^I(J) = |I| = \alpha$ and the problem complexity $R^* = \max(0, \sqrt{\alpha T})$.

F.2. Proof of Corollary 5.2

The starting point of this proof is the Definition 2.7. We know that the graph contains one vertex connected to all other nodes. Therefore, $\delta^V(I \setminus J)$ is either 0, if $I \setminus J$ is an empty set, or 1, if $I \setminus J$ is non-empty.

In case $\delta^V(I \setminus J) = 1$, the minimization part of the problem complexity suggests that J can be empty set since it does not change the value of $\delta^V(I \setminus J)$ while reducing $\delta^I(J)$ to 0, regardless of the choice of I.

In case $\delta^V(I \setminus J) = 0$, we know that $I = J$ and therefore, the value $\delta^I(J)$ can be upperbounded by α - case when I contains all $N - 1 = \alpha$ independent vertices.

Since only these two cases are possible, we have a freedom to choose J for which the value is smaller. Therefore, the problem complexity R^* can be computed as

$$R^* = \min\left\{\alpha^{\frac{1}{2}} T^{\frac{1}{2}},\, 1^{\frac{1}{3}} T^{\frac{2}{3}}\right\}.$$

Using assumption $T < \alpha^3$ gives us $\alpha^{\frac{1}{2}} T^{\frac{1}{2}} > 1^{\frac{1}{3}} T^{\frac{2}{3}}$ and therefore $R^* = T^{\frac{2}{3}}$, which concludes the proof.