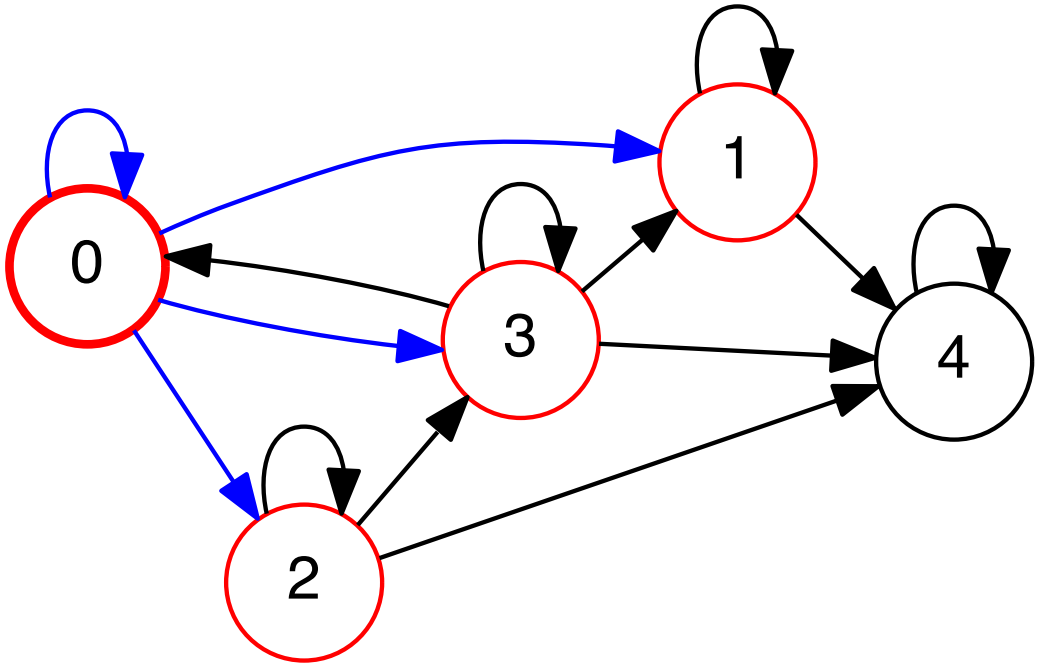
Online Learning with Feedback Graphs

MEHRYAR MOHRI MOHRI@
GOOGLE RESEARCH & COURANT INSTITUTE

Motivation

- Online learning with side observation (Mannor and Shamir, 2011):
  - side observation modeled as feedback graph.
  - full information and bandit: special cases.
  - intermediate regret guarantees expressed in terms of graph properties (mas-number, independence number, domination number).

Feedback Graph

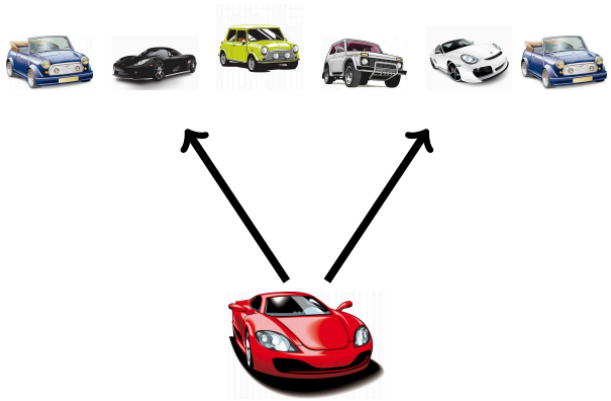


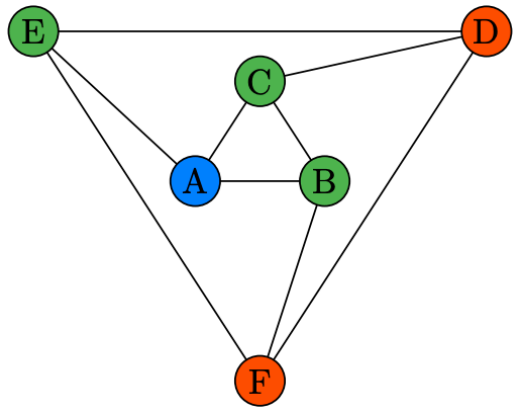
- If arm 0 is selected, then the losses of arms 0, 1, 2, and 3 are observed (but not the loss of arm 4).

Applications

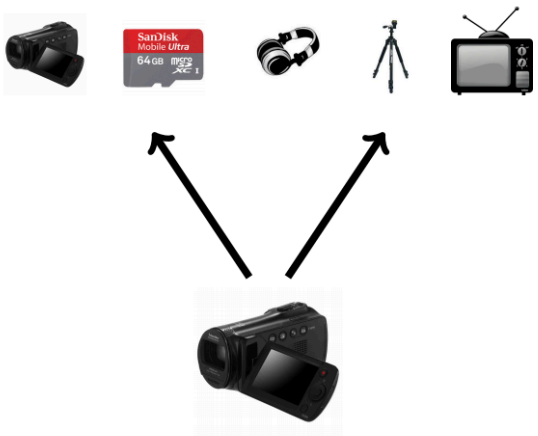
(Valko, 2016)

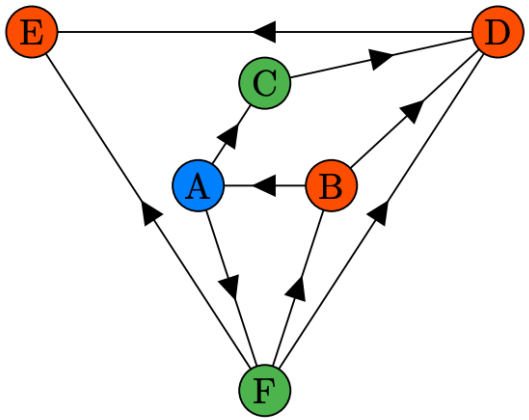
- Undirected graph:





- Directed graph:





Graph Theory Notions

(Goddard and Henning, 2013)

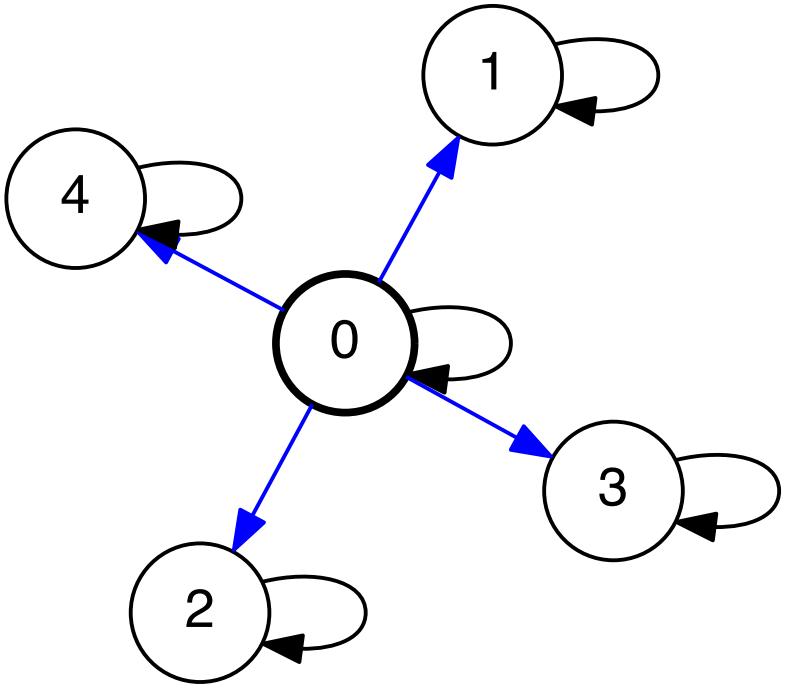
- Given a directed graph $G = (V, E)$ (self-loops ignored),
  - the mas-number of G, $\mu(G)$, is the size of the maximum acyclic subgraph of G.
  - a subset of the vertices is independent if no two vertices in it are adjacent; the independence number of G, $\alpha(G)$, is the size of the maximum independent set in G.
  - a dominating set of G is a subset $S \subseteq V$ such that every vertex not in S is adjacent to S; the domination number of G, $\gamma(G)$, is the minimum size of a dominating set.
  - it follows that for any graph: $\gamma(G) \leq \alpha(G) \leq \mu(G)$.

Graph Theory Notions

- Computing domination number is NP-hard since it is equivalent to the minimum vertex cover problem. But, it can be approximated modulo logarithmic factor via greedy set cover:
  - at each round select vertex with largest uncovered adjacent set.
- When G is undirected (symmetric edges), then,

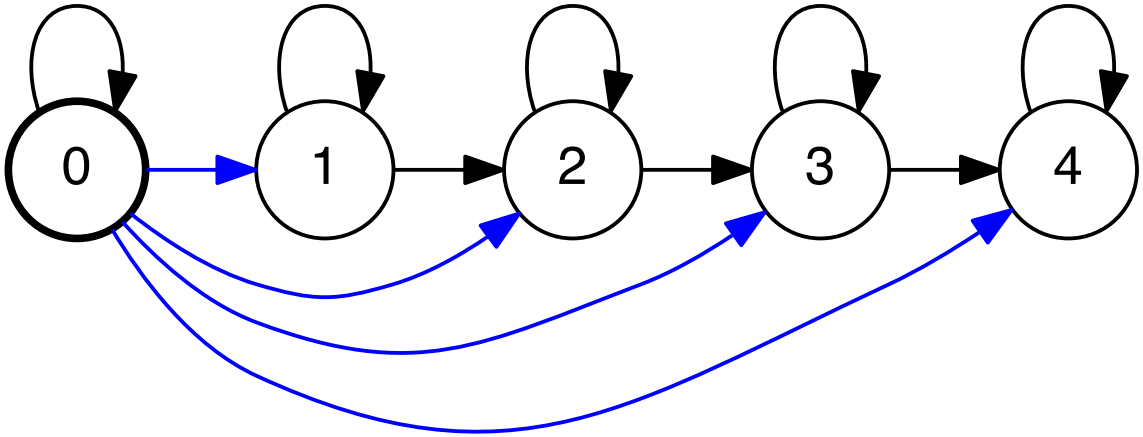
$$\alpha(G) = \mu(G).$$

Examples



- Star graph: $\gamma(G) = 1, \alpha(G) = n - 1, \mu(G) = n.$

Example



- Auction graph: $\gamma(G) = 1, \alpha(G) = (n-1)/2, \mu(G) = n.$

Adversarial Setting

Protocols

- Graph information:
  - pre-informed setting: feedback graph received before selecting arm.
  - uninformed setting: feedback graph received after selecting arm.
- Time-dependent or fixed feedback graph.

EXP3-SET Algorithm

- (Alon et al., 2013): variant of EXP3;
  - uninformed setting.
  - directed feedback graphs $G_t = (V, E_t)$.
  - surrogate loss:

$$\widetilde{\ell}_{i,t} = \frac{\ell_{i,t}}{q_{i,t}} \mathbb{I}\{(I_t, i) \in E_t\}$$

$$q_{i,t} = \underbrace{\sum_{j\,:\,(j,i)\in E_t} p_{j,t}}_{\text{Probability of observing } i.}$$

EXP3-SET

EXP3-SET(η)

$\forall i \in V, w_{i,1} \leftarrow 1$
for $t \leftarrow 1$ **to** T **do**
  $\forall i \in V, \mathsf{p}_{i,t} \leftarrow \frac{w_{i,t}}{\sum_{j \in V} w_{j,t}}$
  SAMPLE($I_t \sim \mathsf{p}_t$)
  RECEIVE($\{(j, \ell_{j,t}) \colon (I_t, j) \in E_t\}$)
  RECEIVE(G_t)
  for $i \leftarrow 1$ **to** $|V|$ **do**
    $q_{i,t} \leftarrow \sum_{j \colon (j,i) \in E_t} p_{j,t}$ ▷ probability of observing i.
    $\widetilde{\ell}_{i,t} \leftarrow \frac{\ell_{i,t}}{q_{i,t}} \mathbb{I}\{(I_t, i) \in E_t\}$
    $w_{i,t+1} \leftarrow e^{-\eta \widetilde{\ell}_{i,t}} w_{i,t}$

EXP3-SET Guarantee

- Theorem: the pseudo-regret of EXP3-SET can be bounded as follows:

$$\overline{\mathrm{Reg}}(\text{EXP3-SET}) \leq \frac{\log K}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \mathbb{E}[Q_t],$$

  - where $Q_t = \sum_{i \in V} \frac{\mathsf{p}_{i,t}}{\mathsf{q}_{i,t}}$.

Proof

- Potential: $\Phi_{t+1} = \log \sum_{i=1}^{K} e^{-\eta \widetilde{L}_{i,t}}$, with $\widetilde{L}_{i,t} = \sum_{s=1}^{t} \widetilde{\ell}_{i,s}$.
- Upper bound:

$$\begin{aligned}
\Phi_{t+1} - \Phi_t = \log \frac{\sum_{i=1}^{K} e^{-\eta \widetilde{L}_{i,t}}}{\sum_{i=1}^{N} e^{-\eta \widetilde{L}_{i,t-1}}} &= \log \frac{\sum_{i=1}^{K} e^{-\eta \widetilde{L}_{i,t-1}} e^{-\eta \widetilde{\ell}_{i,t}}}{\sum_{i=1}^{N} e^{-\eta \widetilde{L}_{i,t-1}}} \\
&= \log \left[\mathop{\mathrm{E}}_{i \sim \mathsf{p}_t} \left[e^{-\eta \widetilde{\ell}_{i,t}} \right] \right] \\
&\leq \mathop{\mathrm{E}}_{i \sim \mathsf{p}_t} \left[e^{-\eta \widetilde{\ell}_{i,t}} \right] - 1 \qquad (\log x \leq x - 1) \\
&\leq \mathop{\mathrm{E}}_{i \sim \mathsf{p}_t} \left[-\eta \widetilde{\ell}_{i,t} + \frac{\eta^2}{2} \widetilde{\ell}_{i,t}^2 \right] \quad (e^{-x} \leq 1 - x + \tfrac{x^2}{2}).
\end{aligned}$$

- Summing up:

$$\Phi_{T+1} - \Phi_1 \leq -\eta \sum_{t=1}^{T} \sum_{i \in V} p_{i,t} \widetilde{\ell}_{i,t} + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_{i \in V} p_{i,t} \widetilde{\ell}_{i,t}^2.$$

Proof

- Lower bound:

$$\Phi_{T+1} - \Phi_1 = \log\left[\sum_{i=1}^{K} e^{-\eta \widetilde{L}_{i,T}} - \log K\right] \geq -\eta \widetilde{L}_{j,T} - \log K = -\eta \sum_{t=1}^{T} \widetilde{\ell}_{j,t} - \log K.$$

- Comparison:

$$\sum_{t=1}^{T} \sum_{i \in V} p_{i,t} \widetilde{\ell}_{i,t} \leq \sum_{t=1}^{T} \widetilde{\ell}_{j,t} + \frac{\log K}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{i \in V} p_{i,t} \widetilde{\ell}_{i,t}^2.$$

- Using conditional expectation $\underset{t}{\mathrm{E}} = \underset{I_t \sim \mathsf{p}_t}{\mathrm{E}}[\cdot | I_1, \ldots, I_{t-1}]$:

$$\sum_{t=1}^{T} \sum_{i \in V} \mathbb{E}\left[p_{i,t} \underset{t}{\mathbb{E}}[\widetilde{\ell}_{i,t}]\right] \leq \sum_{t=1}^{T} \mathbb{E}\left[\underset{t}{\mathbb{E}}[\widetilde{\ell}_{j,t}]\right] + \frac{\log K}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{i \in V} \mathbb{E}\left[p_{i,t} \underset{t}{\mathbb{E}}[\widetilde{\ell}_{i,t}^2]\right].$$

Proof

- Observe that:

$$\mathop{\mathbb{E}}_{t}\left[\widetilde{\ell}_{i,t}\right] = \mathop{\mathbb{E}}_{I_t \sim \mathsf{p}_t}\left[\frac{\ell_{i,t}}{q_{i,t}}\,\mathbb{I}\{(I_t, i) \in E\}\right]$$
$$= \sum_{j \in V} p_{j,t}\frac{\ell_{i,t}}{q_{i,t}}\,\mathbb{I}\{(j,i) \in E\} = q_{i,t}\frac{\ell_{i,t}}{q_{i,t}} = \ell_{i,t}.$$

- Similarly,

$$\mathop{\mathbb{E}}_{t}\left[\widetilde{\ell}^2_{i,t}\right] = \mathop{\mathbb{E}}_{I_t \sim \mathsf{p}_t}\left[\frac{\ell^2_{i,t}}{q^2_{i,t}}\,\mathbb{I}\{(I_t, i) \in E\}\right]$$
$$= \sum_{j \in V} p_{j,t}\frac{\ell^2_{i,t}}{q^2_{i,t}}\,\mathbb{I}\{(j,i) \in E\} = q_{i,t}\frac{\ell^2_{i,t}}{q^2_{i,t}} = \frac{\ell^2_{i,t}}{q_{i,t}} \leq \frac{1}{q_{i,t}}.$$

Proof

- Thus,

$$\sum_{t=1}^{T}\sum_{i\in V}\mathbb{E}[p_{i,t}\ell_{i,t}] \leq \sum_{t=1}^{T}\mathbb{E}[\ell_{j,t}] + \frac{\log K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i\in V}\mathbb{E}\left[\frac{p_{i,t}}{q_{i,t}}\right],$$

  - and,

$$\overline{\mathrm{Reg}}(\text{EXP3-SET}) \leq \frac{\log K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\mathbb{E}[Q_{i,t}].$$

EXP3-SET Guarantee

- Theorem: the pseudo-regret of EXP3-SET can be bounded as follows for directed graphs:

$$\overline{\text{Reg}}(\text{EXP3-SET}) \leq \frac{\log K}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \mathbb{E}[\mu(G_t)].$$

  - for $\mathbb{E}[\mu(G_t)] \leq \mu_t$,

$$\overline{\text{Reg}}(\text{EXP3-SET}) \leq \sqrt{2(\log K) \sum_{t=1}^{T} \mu_t}.$$

Proof

- For any graph (dropping time indices),

$$\sum_{i=1}^{K} \frac{\mathsf{p}_i}{\sum_{j \in \mathrm{IN}(i)} \mathsf{p}_j} \leq \mu(G).$$

  - Construct subset of vertices V' inducing acyclic graph such that $\sum_{i=1}^{K} \frac{\mathsf{p}_i}{\sum_{j \in \mathrm{IN}(i)} \mathsf{p}_j} \leq |V'|$.
  - define $i_1 = \mathrm{argmin}_{i \in V} \sum_{j \in \mathrm{IN}(i)} \mathsf{p}_j$ and remove that vertex from the graph as well as all $j \in \mathrm{IN}(i_1)$ and all edges entering or leaving these vertices.
  - Observe that:

$$\sum_{k \in \mathrm{IN}(i_1)} \frac{p_k}{\sum_{j \in \mathrm{IN}(k)} \mathsf{p}_j} \leq \sum_{k \in \mathrm{IN}(i_1)} \frac{p_k}{\sum_{j \in \mathrm{IN}(i_1)} \mathsf{p}_j} = 1.$$

Proof

- Thus, $\sum_{i=1}^{K} \frac{\mathsf{p}_i}{\sum_{j \in \mathrm{IN}(i)} \mathsf{p}_j} \leq \sum_{i \notin \mathrm{IN}(i_1)} \frac{\mathsf{p}_i}{\sum_{j \in \mathrm{IN}(i)} \mathsf{p}_j} + 1.$

  - reiterating until no vertex is left, with $V' = \{i_1, \ldots, i_k\}$,

$$\sum_{k \in \mathrm{IN}(i_1)} \frac{p_k}{\sum_{j \in \mathrm{IN}(k)} \mathsf{p}_j} \leq |V'|.$$

  - the graph induced by V' cannot contain cycles since at each step all incoming edges of i_r and source vertices of those edges are removed.

EXP3-SET Guarantee

- Corollary: the pseudo-regret of EXP3-SET can be bounded as follows for undirected graphs:

$$\overline{\mathrm{Reg}}(\text{EXP3-SET}) \leq \frac{\log K}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \mathbb{E}[\alpha(G_t)].$$

  - for $\alpha(G_t) \leq \alpha_t$,

$$\overline{\mathrm{Reg}}(\text{EXP3-SET}) \leq \sqrt{2(\log K) \sum_{t=1}^{T} \alpha_t}.$$

Adversarial Setting

- (Mannor and Shamir, 2011): introduced online learning with side information modeled as feedback graph.
- (Alon et al., 2013): directed feedback graphs, variants of EXP3.
- (Alon et al., 2015): algorithm with $O(T^{\frac{2}{3}})$ regret for weakly observable graphs (vertex with no self-loop or no entering edge from all other vertices).
- (Alon et al., 2014): high probability bounds based on mas- number.
- (Neu 2015): high probability bounds based on independence number, *implicit exploration*.

Stochastic Setting

UCB-N

- (Caron et al., 2012): UCB-type algorithm;
  - number of observations of arm i up to time $(t-1)$, $O_{i,t-1}$.
  - average reward of arm i up to time $(t-1)$, $\overline{X}_{i,t-1}$.

$$\overline{X}_{i,t-1} = \frac{1}{O_{i,t-1}} \sum_{s=1}^{t-1} X_{i,s} 1_{i \in N(I_s)}.$$

  - arm selected at time t: $I_t = \operatorname{argmax}_{i \in [K]} \overline{X}_{i,t-1} + \sqrt{\frac{2 \log t}{O_{i,t-1}}}$.

UCB-N

UCB-N(G)

1 $\forall i, \overline{X}_i, O_i \leftarrow 0$
2 **for** $t \leftarrow 1$ **to** T **do**
3 $\quad I_t \leftarrow \operatorname{argmax}_{i \in [K]} \overline{X}_i + \sqrt{\frac{2 \log T}{O_i}}$
4 $\quad$ **for** $k \in N(I_t)$ **do**
5 $\quad\quad O_k \leftarrow O_k + 1$
6 $\quad\quad \overline{X}_k \leftarrow \frac{1}{O_k} X_k + (1 - \frac{1}{O_k}) \overline{X}_k$

Graph Theory Notions

- A clique in an undirected graph $G = (V, E)$: subset of V with any two vertices being adjacent.
- A clique covering $\mathcal{C}$ of G is a set of cliques such that

$$V = \bigcup_{C \in \mathcal{C}} C.$$

UCB-N Guarantee

- Theorem: the pseudo-regret of UCB-N can be bounded as follows for undirected graphs:

$$\overline{\text{Reg}}(\text{UCB-N}) \leq \inf_{\mathcal{C}} \left\{ 8 \sum_{C \in \mathcal{C}} \frac{\max_{i \in C} \Delta_i}{\min_{i \in C} \Delta_i^2} \log T \right\} + \left(1 + \frac{\pi^2}{3}\right) \sum_{i=1}^{K} \Delta_i.$$

Proof

- Lemma: for any $s \geq 0$, for $T_C(t-1) = \sum_{i \in C} T_i(t-1)$,

$$\sum_{t=1}^{T} \sum_{i \in C} 1_{I_t = i} \, \Delta_i \leq s \left(\max_{i \in C} \Delta_i \right) + \sum_{t=s+1}^{T} \sum_{i \in C} 1_{I_t = i} \, 1_{T_C(t-1) \geq s} \, \Delta_i.$$

- Proof: observe that

$$\sum_{t=1}^{T} \sum_{i \in C} 1_{I_t = i} \, \Delta_i = \sum_{t=1}^{T} \sum_{i \in C} 1_{I_t = i} \, 1_{T_C(t-1) < s} \, \Delta_i + \sum_{t=1}^{T} \sum_{i \in C} 1_{I_t = i} \, 1_{T_C(t-1) \geq s} \, \Delta_i.$$

  - Now, for $t^* = \max \{ t \leq T \colon 1_{T_C(t-1) < s} \neq 0 \}$,

$$\sum_{t=1}^{T} \sum_{i \in C} 1_{I_t = i} \, 1_{T_C(t-1) < s} \, \Delta_i \leq \left(\max_{i \in C} \Delta_i \right) \sum_{t=1}^{t^*} \sum_{i \in C} 1_{I_t = i} \, 1_{T_C(t-1) < s}.$$

  - By definition of t^*, the number of non-zero terms in the sum is at most s.

Proof

- For any i and t define $\eta_{i,t-1} = \sqrt{\frac{2 \log t}{O_i(t-1)}}$. At time t, if i is selected, then

$$(\widehat{\mu}_{i,t-1} + \eta_{i,t-1}) - (\widehat{\mu}_{i^*,t} + \eta_{i^*,t-1}) \geq 0$$
$$\Leftrightarrow [\widehat{\mu}_{i,t-1} - \mu_{i,t-1} - \eta_{i,t-1}] + [2\eta_{i,t-1} - \Delta_i] + [\mu^* - \widehat{\mu}_{i^*,t-1} - \eta_{i^*,t-1}] \geq 0.$$

  Thus, at least of one of these three terms is non-negative. Also, if one is non-positive, at least one of the other two is non-negative.

Proof

- To bound the pseudo-regret, we bound $\sum_{i \in C} \mathrm{E}[T_i(T)]$. Observe first that

$$O_i(t-1) \geq s_C = \max_{i \in C} \left\lceil \frac{8 \log T}{\Delta_i^2} \right\rceil \geq \frac{8 \log T}{\min_{i \in C} \Delta_i^2} \Rightarrow \forall i \in C, \Delta_i - 2\eta_{i,t-1} \geq 0.$$

- Thus,

$$\begin{aligned}
\sum_{i \in C} \mathbb{E}[T_i(T)] \Delta_i &= \mathbb{E}\left[\sum_{t=1}^{T} \sum_{i \in C} 1_{I_t = i}\right] \\
&\leq s_C \left(\max_{i \in C} \Delta_i\right) + \mathbb{E}\left[\sum_{t=s_C+1}^{T} \sum_{i \in C} 1_{I_t = i} \, 1_{T_C(t-1) \geq s_C} \Delta_i\right] \\
&\leq s_C \left(\max_{i \in C} \Delta_i\right) + \mathbb{E}\left[\sum_{t=s_C+1}^{T} \sum_{i \in C} 1_{I_t = i} \, 1_{O_i(t-1) \geq s_C} \Delta_i\right] \\
&\leq s_C \left(\max_{i \in C} \Delta_i\right) + \sum_{t=s_C+1}^{T} \sum_{i \in C} \Delta_i \, \mathbb{P}[\widehat{\mu}_{i,t-1} - \mu_{i,t-1} - \eta_{i,t-1} \geq 0] + \Delta_i \, \mathbb{P}[\mu^* - \widehat{\mu}_{i^*,t-1} - \eta_{i^*,t-1} \geq 0] \\
&\leq s_C \left(\max_{i \in C} \Delta_i\right) + \sum_{t=s_C+1}^{T} \sum_{i \in C} \Delta_i \sum_{t=1}^{T} \frac{2}{t^4}.
\end{aligned}$$

Proof

- The pseudo-regret of the algorithm can thus be upper-bounded as follows:

$$\begin{aligned}
\overline{\mathrm{Reg}}(\text{UCB-N}) &= \sum_{C \in \mathcal{C}} \sum_{i \in C} \mathbb{E}[T_i(T)] \Delta_i \\
&\leq \sum_{C \in \mathcal{C}} s_C \left(\max_{i \in C} \Delta_i \right) + \sum_{C \in \mathcal{C}} \sum_{t=s_C+1}^{T} \sum_{i \in C} \Delta_i \sum_{t=1}^{T} \frac{2}{t^4} \\
&\leq \sum_{C \in \mathcal{C}} \left(\max_{i \in C} \Delta_i \right) \frac{8 \log T}{\min_{i \in C} \Delta_i^2} + \sum_{C \in \mathcal{C}} \sum_{i \in C} \Delta_i \left(1 + \frac{\pi^2}{3} \right).
\end{aligned}$$

Stochastic Setting

- (Caron et al., 2012): UCB-type algorithm for undirected feedback graphs in stochastic setting; guarantees in terms of graph clique structure.
- (Cohen et al., 2016): full feedback graph never revealed, regret guarantee based on independence number, contrast with adversarial setting where bandit bound remains optimal.
- (Buccapatnam et al., 2014): LP-based solution, regret guarantee based on domination number, lower bound.

Stochastic Setting

- (Buccapatnam et al., 2017): more general setting covering (Cohen et al., 2016).
- (Lykouris et al., 2020): analysis of algorithms using *layering technique;* e.g. independence number guarantee for UCB-N.
- (Cortes et al., 2019): sleeping experts with dependent losses and awake sets.
- (Cortes et al., 2020): dependent losses and feedback graphs varying stochastically.
- (Marinov, MM, Zimmert, 2022): notion of optimal finite-time regret not uniquely defined in this context! Algorithm with quasi-optimal pseudo-regret for a *meaningful* notion.

Extensions

- (Valko, 2016): general survey of feedback graphs.
- (Kocak et al., 2016): online learning with noisy side information.
- (Arora et al., 2019): adversarial setting with feedback graphs and switching costs.
- (Dann et al., 2020): reinforcement learning with feedback graphs.