Online learning with feedback graphs and switching costs

A Proof of Theorem 1

Proof. Without loss of generality, let the independent sequence set $\mathcal{I}(G_{1:T})$ formed of actions (or "arms") from 1 to $\beta(G_{1:T})$. Given the sequence of feedback graphs $G_{1:T}$, let T_i be the number of times the action $i \in \mathcal{I}(G_{1:T}) = [\beta(G_{1:T})]$ is selected by the player in T rounds. Let T_Δ be the total number of times the actions are selected from the set $[K]\backslash\mathcal{I}(G_{1:T})$. Let $\mathbb{E}_i$ denote expectation conditioned on $X = i$, and $\mathbb{P}_i$ denote the probability conditioned on $X = i$. Additionally, we define $\mathbb{P}_0$ as the probability conditioned on event $\epsilon_1 = 0$. Therefore, under $\mathbb{P}_0$, all the actions in the independent sequence set, i.e. $i \in \mathcal{I}(G_{1:T})$, incur an expected regret of $1/2$, whereas, the expected regret of actions $i \in [K]\backslash\mathcal{I}(G_{1:T})$ is $1/2 + \epsilon_2$. Let $\mathbb{E}_0$ be the corresponding conditional expectation. For all $i \in [K]$ and $t \leq T$, $\ell_t(i)$ and $\ell_t^c(i)$ denote the unclipped and clipped loss of the action i respectively. Assuming the unclipped losses are observed by the player, then $\mathcal{F}$ is the sigma field generated by the unclipped losses, and $S_t(i_t)$ is the set of actions whose losses are observed at time t, following the selection of i_t, according to the feedback graph G_t. The observed sequence of unclipped losses will be referred as $\ell_{1:T}^o$. Additionally, $\mathcal{F}'$ is the sigma field generated by the clipped losses, for all $t \in [T]$, $\ell_t'(i)$ where $i \in S_t(i_t)$, and the observed sequence of clipped losses will be referred as $\ell_{1:T}'^o$. By definition, $\mathcal{F}' \subseteq \mathcal{F}$.

Let $i_1, \ldots, i_T$ be the sequence of actions selected by a player over the time horizon T. Then, the regret R^c of the player corresponding to clipped loses is

$$R^c = \sum_{t=1}^{T} \ell_t^c(i_t) + c \cdot M_s - \min_{i\in[K]} \sum_{t=1}^{T} \ell_t^c(i), \tag{1}$$

where M_s is the number of switches in the action selection sequence $i_1, \ldots, i_T$, and c is the cost of each switch in action. Now, we define the regret R which corresponds to the unclipped loss function in Algorithm 1 as following

$$R = \sum_{t=1}^{T} \ell_t(i_t) + c \cdot M_s - \min_{i\in[K]} \sum_{t=1}^{T} \ell_t(i). \tag{2}$$

Using (Dekel et al., 2014, Lemma 4), we have

$$\mathbb{P}\left(\text{For all } t \in [T], \frac{1}{2} + W_t \in \left[\frac{1}{6}, \frac{5}{6}\right]\right) \geq \frac{5}{6}. \tag{3}$$

Thus, for all $T > \max\{\beta(G_{1:T}), 6\}$, we have $\epsilon_1 = \epsilon_2 < 1/6$. If $B = \{\text{For all } t \in [T] : 1/2 + W_t \in [1/6, 5/6]\}$ occurs and $\epsilon_1 = \epsilon_2 < 1/6$, then for all $i \in [K]$, $\ell_t^c(i) = \ell_t(i)$ which implies $R^c = R$ (see (1) and (2)). Now, if the event B does not occur, then the losses at any time t satisfy

$$\ell_t(i) - \ell_t^c(i) \leq (\epsilon_1 + \epsilon_2).$$

Therefore, we have

$$c \cdot M_s \leq R^c \leq R \leq c \cdot M_s + (\epsilon_1 + \epsilon_2)T.$$

Now, for $T > \max\{\beta(G_{1:T}), 6\}$, we have

$$\mathbb{E}[R] - \mathbb{E}[R^c] = (1 - \mathbb{P}(B))\mathbb{E}[R - R^c | B \text{ does not occur}] \leq \frac{(\epsilon_1 + \epsilon_2)T}{6}. \tag{4}$$

Thus, (4) lower bounds the actual regret R^c in terms of regret R. Now, we derive the lower bound on regret R corresponding to the unclipped loses. Using the definition of R, we have

$$\begin{aligned}
\mathbb{E}[R] &= \max_{i\in[K]} \mathbb{E}[\sum_{t=1}^{T} \ell_t(i_t) - \sum_{t=1}^{T} \ell_t(i)] + \mathbb{E}[M_s] \\
&= \frac{1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i[\sum_{t=1}^{T} \ell_t(i_t) - \min_{i\in[K]} \sum_{t=1}^{T} \ell_t(i)] + \mathbb{E}[M_s] \\
&= \frac{1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i\Bigg[\sum_{j\in\mathcal{I}(G_{1:T})\setminus\{i\}} \frac{1}{2}T_j + \left(\frac{1}{2}-\epsilon_1\right)T_i + \left(\frac{1}{2}+\epsilon_2\right)T_\Delta - \left(\frac{1}{2}-\epsilon_1\right)T\Bigg] + \mathbb{E}[M_s] \\
&= \frac{1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i\Bigg[\sum_{j=1}^{\beta(G_{1:T})} \frac{1}{2}T_j + \left(\frac{1}{2}+\epsilon_2\right)T_\Delta - \epsilon_1 T_i - \left(\frac{1}{2}-\epsilon_1\right)T\Bigg] + \mathbb{E}[M_s] \\
&\overset{(a)}{=} \frac{1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i\Big[\epsilon_2 T_\Delta + \epsilon_1(T-T_i)\Big] + \mathbb{E}[M_s] \\
&\overset{(b)}{\geq} \epsilon_1\left(T - \frac{1}{\beta(G_{1:T})} \sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i\big[T_i\big]\right) + \mathbb{E}[M_s],
\end{aligned} \tag{5}$$

where (a) follows from $\sum_{j=1}^{\beta(G_{1:T})} T_j + T_\Delta = T$, and (b) follows from $\epsilon_2 T_\Delta \geq 0$.

Now, we upper bound the $\mathbb{E}_i\big[T_i\big]$ in (5) to obtain the lower bound on the expected regret $\mathbb{E}[R]$. Since the player is deterministic, the event $\{i_t = i\}$ is $\mathcal{F}'$ measurable. Therefore, we have

$$\mathbb{P}_i(i_t = i) - \mathbb{P}_0(i_t = i) \leq d_{TV}^{\mathcal{F}'}(P_0, P_i) \overset{(a)}{\leq} d_{TV}^{\mathcal{F}}(P_0, P_i),$$

where $d_{TV}^{\mathcal{F}}(P_0, P_i) = \sup_{A\in\mathcal{F}} |\mathbb{P}_0(A) - \mathbb{P}_i(A)|$ is the total variational distance between the two probability measures, and (a) follows from $\mathcal{F}' \subseteq \mathcal{F}$. Summing the above equation over $t \in [T]$ and $i \in \mathcal{I}(G_{1:T})$ yields

$$\sum_{i=1}^{\beta(G_{1:T})} \big(\mathbb{E}_i[T_i] - \mathbb{E}_0[T_i]\big) \leq T \cdot \sum_{i=1}^{\beta(G_{1:T})} d_{TV}^{\mathcal{F}}(P_0, P_i).$$

Rearranging the above equation and using $\sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_0[T_i] = \mathbb{E}_0[\sum_{i=1}^{\beta(G_{1:T})} T_i] = T$, we get

$$\sum_{i=1}^{\beta(G_{1:T})} \mathbb{E}_i[T_i] \leq T \cdot \sum_{i=1}^{\beta(G_{1:T})} d_{TV}^{\mathcal{F}}(P_0, P_i) + T.$$

Combining the above equation with (5), we get

$$\begin{aligned}
\mathbb{E}[R] &\geq \epsilon_1 T - \frac{\epsilon_1 T}{\beta(G_{1:T})} \cdot \sum_{i=1}^{\beta(G_{1:T})} d_{TV}^{\mathcal{F}}(P_0, P_i) - \frac{\epsilon_1 T}{\beta(G_{1:T})} + \mathbb{E}[M_s] \\
&\overset{(a)}{\geq} \frac{\epsilon_1 T}{2} - \frac{\epsilon_1 T}{\beta(G_{1:T})} \cdot \sum_{i=1}^{\beta(G_{1:T})} d_{TV}^{\mathcal{F}}(P_0, P_i) + \mathbb{E}[M_s],
\end{aligned} \tag{6}$$

where (a) uses the fact that $\beta(G_{1:T}) > 1$. Next, we upper bound the second term in the right hand side of (6). Using Pinsker's inequality, we have

$$d_{TV}^{\mathcal{F}}(P_0, P_i) \leq \sqrt{\frac{1}{2} D_{KL}(\mathbb{P}_0(\ell_{1:T}^o) || \mathbb{P}_i(\ell_{1:T}^o))}, \tag{7}$$

where $\ell^o_{1:T}$ are the losses observed by the player over the time horizon T. Using the chain rule of relative entropy to decompose $D_{KL}(\mathbb{P}_0(\ell^o_{1:T})||\mathbb{P}_0(\ell^o_{1:T}))$, we get

$$\begin{aligned} D_{KL}(\mathbb{P}_0(\ell^o_{1:T})||\mathbb{P}_i(\ell^o_{1:T})) &= \sum_{t=1}^{T} D_{KL}(\mathbb{P}_0(\ell^o_t|\ell^o_{1:t-1})||\mathbb{P}_i(\ell^o_t|\ell^o_{1:t-1})) \\ &= \sum_{t=1}^{T} D_{KL}(\mathbb{P}_0(\ell^o_t|\ell^o_{\rho^*(t)})||\mathbb{P}_i(\ell^o_t|\ell^o_{\rho^*(t)})), \end{aligned} \tag{8}$$

where $\rho^*(t)$ is the set of time instances $0 \leq k \leq t$ encountered when operation $\rho(.)$ in Algorithm 1 is applied recursively to t. Now, we deal with each term $D_{KL}(\mathbb{P}_0(\ell^o_t|\ell^o_{\rho^*(t)})||\mathbb{P}_i(\ell^o_t|\ell^o_{\rho^*(t)}))$ in the summation individually. For $i \in \mathcal{I}(G_{1:T})$, we separate this computation into four cases: i_t is such that loss of action i is observed at both time instances t and $\rho(t)$ i.e. $i \in S_t(i_t)$ and $i \in S_t(i_{\rho(t)})$; i_t is such that loss of action i is observed at time instance t but not at time instance $\rho(t)$ i.e. $i \in S_t(i_t)$ and $i \notin S_t(i_{\rho(t)})$; i_t is such that loss of action i is not observed at time instance t but is observed at time instance $\rho(t)$ i.e. $i \notin S_t(i_t)$ and $i \in S_t(i_{\rho(t)})$; i_t is such that loss of action i is not observed at both time instances t and $\rho(t)$ i.e. $i \notin S_t(i_t)$ and $i \notin S_t(i_{\rho(t)})$. Note that at a single time instance the loss of only one single action can be observed from $\mathcal{I}(G_{1:T})$ arms.

Case 1: Since the loss of action i is observed from the independent sequence set $\mathcal{I}(G_{1:T})$ at both the time instances, the loss distribution for the action i is $\ell^o_t(i)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(i), \sigma^2)$ for both $\mathbb{P}_0$ and $\mathbb{P}_i$. For all $j \in [K]\backslash\mathcal{I}(G_{1:T})$, the loss distribution is $\ell^o_t(j)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(i) + \epsilon_1 + \epsilon_2, \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$.

Case 2: Since the loss of action i is observed from the independent sequence set $\mathcal{I}(G_{1:T})$ at time instance t but not at $\rho(t)$, therefore, there exists an action $k' \in \mathcal{I}(G_{1:T})\backslash\{i\}$ from the independent sequence set which was observed at time instance $\rho(t)$. Then, the loss distribution for the action i is $\ell^o_t(i)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(k'), \sigma^2)$ under $\mathbb{P}_0$, and $\ell^o_t(i)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(k') - \epsilon_1, \sigma^2)$ under $\mathbb{P}_i$. For all $j \in [K]\backslash\mathcal{I}(G_{1:T})$, the loss distribution is $\ell^o_t(j)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(k') + \epsilon_2, \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$.

Case 3:Since the action i is observed from the independent sequence set $\mathcal{I}(G_{1:T})$ at time instance $\rho(t)$ but not at t, therefore, there exists an action $k' \in \mathcal{I}(G_{1:T})\backslash\{i\}$ from the independent sequence set which was observed at time instance t. Then, the loss distribution for the arm k' is $\ell^o_t(k')|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(i), \sigma^2)$ under $\mathbb{P}_0$, and $\ell^o_t(k')|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(i) + \epsilon_1, \sigma^2)$ under $\mathbb{P}_i$. For all $j \in [K]\backslash\mathcal{I}(G_{1:T})$, the loss distribution is $\ell^o_t(j)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(i) + \epsilon_1 + \epsilon_2, \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$.

Case 4: Let k^* be the arm from the independent sequence set observed at time instance $\rho(t)$. Since the arm i is not observed from the independent sequence set $\mathcal{I}(G_{1:T})$ at the time instances t and $\rho(t)$, therefore the loss distribution for all arms $k' \in \mathcal{I}(G_{1:T})\backslash\{i\}$ is $\ell^o_t(k')|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(k^*), \sigma^2)$ for both $\mathbb{P}_0$ and $\mathbb{P}_i$. For all $j \in [K]\backslash\mathcal{I}(G_{1:T})$, the loss distribution is $\ell^o_t(j)|\ell^o_{\rho^*(t)} \sim \mathcal{N}(\ell_{\rho(t)}(k^*)+\epsilon_2, \sigma^2)$ under both $\mathbb{P}_0$ and $\mathbb{P}_i$.

Therefore, we have

$$\begin{aligned} D_{KL}(\mathbb{P}_0(\ell^o_t|\ell^o_{\rho^*(t)})||\mathbb{P}_i(\ell^o_t|\ell^o_{\rho^*(t)})) &= \mathbb{P}_0(i \in S_t(i_t), i \notin S_{\rho(t)}(i_{\rho(t)})) \cdot D_{KL}(\mathcal{N}(0,\sigma^2)||\mathcal{N}(-\epsilon_1,\sigma^2)) \\ &\quad + \mathbb{P}_0(i \notin S_t(i_t), i \in S_{\rho(t)}(i_{\rho(t)})) \cdot D_{KL}(\mathcal{N}(0,\sigma^2)||\mathcal{N}(\epsilon_1,\sigma^2)) \\ &= \frac{\epsilon_1^2}{2\sigma^2}\mathbb{P}_0(B_t), \end{aligned} \tag{9}$$

where $B_t = \{i \in S_t(i_t), i \notin S_{\rho(t)}(i_{\rho(t)}) \cup i \notin S_t(i_t), i \in S_{\rho(t)}(i_{\rho(t)})\}$. The event B_t implies that the player has switched at least once between the feedback systems $S_t(k_1)$ and $S_{\rho(t)}(k_2)$ such that $i \in S_t(k_1)$ but $i \notin S_{\rho(t)}(k_2)$ or vice-versa. Let N_i be the number of times a player switches from the feedback system which includes i to the feedback system which does not include i and vice-versa. Then, using (8) and (9), we have

$$D_{KL}(\mathbb{P}_0(\ell^o_{1:T})||\mathbb{P}_i(\ell^o_{1:T})) \leq \frac{\epsilon_1^2\omega(\rho)}{2\sigma^2}\mathbb{E}_0[N_i], \tag{10}$$

where $\omega(\rho)$ is the width of process $\rho(.)$ (see Definition 2 in Dekel et al. (2014)) and is bounded above by $2\log_2(T)$. Combining (7) and (10), we have

$$\sup_{A\in\mathcal{F}}(\mathbb{P}_0(A)-\mathbb{P}_i(A)) \leq \frac{\epsilon_1}{\sigma}\sqrt{\log_2(T)\mathbb{E}_0[N_i]}. \tag{11}$$

If $M_s \geq \epsilon_1 T$, then $\mathbb{E}[R'] > \epsilon_1 T$. Thus, the claimed lower bound follows. Now, let us assume $M_s \leq \epsilon_1 T$. For all $i \in \mathcal{I}(G_{1:T})$, we have

$$\begin{aligned}\mathbb{E}_0[M_s]-\mathbb{E}_i[M_s] &= \sum_{m=1}^{\lfloor\epsilon_1 T\rfloor}\mathbb{P}_0(M_s\geq m)-\mathbb{P}_i(M_s\geq m)) \\ &\leq \epsilon_1 T\cdot d_{TV}^{\mathcal{F}}(\mathbb{P}_0,\mathbb{P}_i).\end{aligned} \tag{12}$$

Using the above equation, we have

$$\begin{aligned}\mathbb{E}_0[M_s]-\mathbb{E}[M_s] &= \frac{1}{\beta(G_{1:T})}\sum_{i=1}^{\beta(G_{1:T})}(\mathbb{E}_0[M_s]-\mathbb{E}_i[M_s]) \\ &\leq \frac{\epsilon_1 T}{\beta(G_{1:T})}\sum_{i=1}^{\beta(G_{1:T})} d_{TV}^{\mathcal{F}}(\mathbb{P}_0,\mathbb{P}_i).\end{aligned} \tag{13}$$

Now, combining (4), (6), (11)and (13), we obtain

$$\begin{aligned}\mathbb{E}[R'] &\geq \frac{\epsilon_1 T}{6}-\frac{\epsilon_1 T}{\beta(G_{1:T})}\sum_{i=1}^{\beta(G_{1:T})}\frac{\epsilon_1}{\sigma}\sqrt{\log_2(T)\mathbb{E}_0[N_i]}+c\cdot\mathbb{E}_0[M_s] \\ &\overset{(a)}{\geq} \frac{\epsilon_1 T}{6}-\frac{\epsilon_1^2 T}{\sigma\sqrt{\beta(G_{1:T})}}\sqrt{2\log_2(T)\mathbb{E}_0[M_s]}+c\cdot\mathbb{E}_0[M_s] \\ &\overset{(b)}{\geq} \frac{c^{1/3}\beta(G_{1:T})^{1/3}T^{2/3}}{54\log_2(T)}-\frac{c^{1/3}\beta(G_{1:T})^{1/3}T^{2/3}}{162\log_2(T)} \\ &= \frac{c^{1/3}\beta(G_{1:T})^{1/3}T^{2/3}}{81\log_2(T)},\end{aligned} \tag{14}$$

where (a) follows from the concavity of $\sqrt{x}$ and $\sum_i^{\beta(G_{1:T})} N_i \leq 2M_s$, (b) follows from the fact that the right hand side is minimized for $\sqrt{\mathbb{E}_0[M_s]} = \epsilon^2 T\sqrt{\log_2(T)}/2c\sigma\sqrt{\beta(G_{1:T})}$. The claim of the theorem now follows. □

B Proof of Lemma 2

$\beta(G_{1:T})$ is the cardinality of $\mathcal{I}(G_{1:T})$. Let $1, 2, \ldots \beta(G_{1:T})$ actions belong to the set $\mathcal{I}(G_{1:T})$. Then, the adversary selects an action uniformly at random from the set $\mathcal{I}(G_{1:T})$ say j, and assigns the loss sequence to action j using independent Bernoulli random variable with parameter $0.5-\epsilon$, where $\epsilon = \sqrt{\beta(G_{1:T})/T}$. For all $i \in \mathcal{I}(G_{1:T})/\{j\}$, losses are assigned using independent Bernoulli random variable with parameter 0.5. For all $i \notin \mathcal{I}(G_{1:T})$, the losses are assigned using independent Bernoulli random variable with parameter 1. The proof of the lemma follows along the same lines as Theorem 5 in (Alon et al. (2017)).

C Proof of Theorem 3

Proof of this theorem uses the result from Theorem 1. Since the loss sequence is assigned independently to each sub-sequence U_m where $m \in [M]$. Using Theorem 1, there exists a constant b_m such that

$$\begin{aligned}&\mathbf{E}\left[\sum_{t=1}^{T}(\ell_t(i_t)\mathbf{1}(G_t\in U_m)+cW_m\right]-\min_{i\in U_m}\sum_{t=1}^{T}(\ell_t(i)\mathbf{1}(G_t\in U_m) \\ &\geq b_m c^{1/3}\beta(U_m)^{1/3}N(U_m)^{2/3}/\log(T),\end{aligned} \tag{15}$$

where W_m is number of switches performed within the sequence U_m. Since

$$\sum_{m\in[M]} W_m \leq \sum_{t=1}^{T} \mathbf{1}(i_t \neq i_{t-1}),$$

there exist a constant b such that the expected regret of any algorithm $\mathcal{A}$ is at least

$$b\, c^{1/3} \sum_{m\in[M]} \beta(U_m)^{1/3} N(U_m)^{2/3} / \log T.$$

D Proof of Lemma 4

Proof. The proof follows from contradiction and is along the same lines as the proof of Theorem 4 in Dekel et al. (2014). Let $\mathcal{A}$ performs at most $\tilde{O}((\beta(G_{1:T})^{1/2}T)^{\alpha})$ switches for any sequence of loss function over T rounds with $\beta + \alpha/2 < 1$. Then, there exists a real number γ such that $\beta < \gamma < 1 - \alpha/2$. Then, assign $c = (\beta(G_{1:T})^{1/2}T)^{3\gamma-2}$. Thus, the expected regret, including the switching cost, of the algorithm is

$$\tilde{O}((\beta(G_{1:T})^{1/2}T)^{\beta} + (\beta(G_{1:T})^{1/2}T)^{3\gamma-2}(\beta(G_{1:T})T)^{\alpha}) = \tilde{o}(\beta(G_{1:T})^{1/2}T)^{\gamma},$$

over a sequence of losses assigned by the adversary because $\beta < \gamma$ and $\alpha < 2 - 2\gamma$. However, according to Theorem 1, the expected regret is at least $\tilde{\Omega}(\beta(G_{1:T})^{1/3}(\beta(G_{1:T})^{1/2}T)^{(3\gamma-2)/3}T^{2/3}) = \tilde{\Omega}((\beta(G_{1:T})T)^{\gamma})$. Hence, by contradiction, the proof of the lemma follows. □

E Proof of Theorem 5

Proof. Let $t_1, t_2 \ldots, t_{\sigma(T)}$ be the sequence of time instances at which the event E^t occurs during the duration T of the game. We define $\{r_j = t_{j+1} - t_j\}_{1\leq j\leq T}$ as the sequence of inter-event times between the events E^t. Let $\text{mas}(G_{(1)}), \ldots, \text{mas}(G_{(T)})$ denote the sequence in the decreasing order of size of maximal acyclic graphs, i.e. $\text{mas}(G_{(1)})$ (or $\text{mas}(G_{(T)})$) is the maximum (or minimum) size of maximal acyclic graph observed in sequence $G_{1:T} = \{G_1, \ldots G_T\}$. Using the definition of E^t, note that r_j is a random variable bounded by $T^{1/3}c^{2/3}/\text{mas}(G_{(T)})^{1/3}$. For all $1 \leq j \leq \sigma(T)$, the ratio of total weights of actions at round t_j and t_{j+1} is

$$\begin{aligned}
\frac{W_{t_{j+1}}}{W_{t_j}} &= \sum_{i\in[K]} \frac{w_{i,t_{j+1}}}{W_{t_j}} \\
&= \sum_{i\in[K]} \frac{w_{i,t_j} \exp(-\eta \ell'_{t_j+r_j-1}(i))}{W_{t_j}} \\
&= \sum_{i\in[K]} p_{i,t_j} \exp(-\eta \ell'_{t_j+r_j-1}(i)) \\
&\overset{(a)}{\leq} \sum_{i\in[K]} p_{i,t_j} \left(1 - \eta \ell'_{t_j+r_j-1}(i) + \frac{1}{2}\eta^2 \ell'^2_{t_j+r_j-1}(i)\right) \\
&= 1 - \eta \sum_{i\in[K]} p_{i,t_j} \cdot \ell'_{t_j+r_j-1}(i) + \frac{\eta^2}{2} \sum_{i\in[K]} p_{i,t_j} \cdot \ell'^2_{t_j+r_j-1}(i),
\end{aligned} \tag{16}$$

where (a) follows from the fact that, for all $x \geq 0$, $e^{-x} \leq 1 - x - x^2/2$. Now, taking logs on both sides of (16), summing over $t_1, t_2, \ldots t_{\sigma(T)}$, and using $\log(1 + x) \leq x$ for all $x > -1$, we get

$$\log \frac{W_{t_{\sigma(T)+1}}}{W_1} \leq -\eta \sum_{j=1}^{\sigma(T)} \sum_{i\in[K]} p_{i,t_j} \cdot \ell'_{t_j+r_j-1}(i) + \frac{\eta^2}{2} \sum_{j=1}^{\sigma(T)} \sum_{i\in[K]} p_{i,t_j} \cdot \ell'^2_{t_j+r_j-1}(i). \tag{17}$$

For all actions $k' \in [K]$, we also have

$$\log \frac{W_{t_{\sigma(T)+1}}}{W_1} \geq \log \frac{w_{k',t_{\sigma(T)+1}}}{W_1} \geq -\eta \sum_{j=1}^{\sigma(T)} \ell'_{t_j+r_j-1}(k') - \log(K). \tag{18}$$

Combining (17) and (18), for all $k' \in [K]$, we obtain

$$\sum_{j=1}^{\sigma(T)} \sum_{i\in[K]} p_{i,t_j} \cdot \ell'_{t_j+r_j-1}(i) - \sum_{j=1}^{\sigma(T)} \ell'_{t_j+r_j-1}(k') \leq \frac{\log(K)}{\eta} + \frac{\eta}{2} \sum_{j=1}^{\sigma(T)} \sum_{i\in[K]} p_{i,t_j} \cdot \ell'^2_{t_j+r_j-1}(i). \tag{19}$$

Now, for all $i \in [K]$, the conditional expectation of $\ell'_{t_j+r_j-1}(i)$ is

$$\begin{aligned}
\mathbb{E}\Big[\ell'_{t_j+r_j-1}(i)\Big|p_{t_j}, r_j\Big] &= \sum_{t=t_j}^{t_j+r_j-1} \sum_{k':i\in S_t(k')} p_{k',t_j} \cdot \frac{\ell_t(i)}{q_{i,t}}, \\
&= \sum_{t=t_j}^{t_j+r_j-1} \frac{\ell_t(i)}{q_{i,t}} \cdot \sum_{k':i\in S_t(k')} p_{k',t_j}, \\
&= \sum_{t=t_j}^{t_j+r_j-1} \ell_t(i).
\end{aligned} \tag{20}$$

Therefore, we have that for all $i \in [K]$, the conditional expectation

$$\mathbb{E}\Big[\sum_{j=1}^{\sigma(T)} \ell'_{t_j+r_j-1}(i)\Big|\{p_{t_j}, r_j\}_{1\leq j\leq\sigma(T)}]\Big] = \sum_{j=1}^{\sigma(T)} \sum_{t=t_j}^{t_j+r_j-1} \ell_t(i) = \sum_{t=1}^{T} \ell_t(i). \tag{21}$$

Now, the expectation of second term in right hand side of (19) is

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^{\sigma(T)} \sum_{i\in[K]} p_{i,t_j} \cdot \ell'^2_{t_j+r_j-1}(i)\right] &= \mathbb{E}\left[\sum_{j=1}^{\sigma(T)} \mathbb{E}\Big[\sum_{i\in[K]} p_{i,t_j} \ell'^2_{t_j+r_j-1}(i)|\{p_{t_j}, r_j\}_{1\leq j\leq\sigma(T)}\Big]\right] \\
&\overset{(a)}{\leq} \mathbb{E}\left[\sum_{j=1}^{\sigma(T)} \text{mas}(G_{t_j:t_j+r_j-1}) r_j^2\right],
\end{aligned} \tag{22}$$

where $\text{mas}(G_{t_j:t_j+r_j-1}) = \max_{n\in[t_j,t_j+r_j-1]} \text{mas}(G_n)$, and (a) follows from the fact that, for all $i \in [K]$ and $t \leq T$, $\ell_t(i) \leq 1$, and $\sum_{i\in[K]} p_{i,t}/q_{i,t} \leq \text{mas}(G_t)$(Alon et al., 2017, Lemma 10).

Now, we bound $\sum_{j=1}^{\sigma(T)} \text{mas}(G_{t_j:t_j+r_j-1}) r_j^2$. We write the following optimization problem:

$$\max_{\{r_j\}_{1\leq j\leq T}} \sum_{j=1}^{T} \text{mas}(G_{t_j:t_j+r_j-1}) r_j^2, \text{ subject to} \tag{23}$$

$$\sum_{j=1}^{T} r_j = T,$$

$$0 \leq r_j \leq \frac{T^{1/3} c^{2/3}}{\text{mas}^{1/3}(G_{(T)})}.$$

Since the objective function is submodular and the constraints are linear, the ratio of the solution of the greedy algorithm and the optimal solution is at most $(1 - 1/e)$ (Nemhauser and Wolsey (1978)). Therefore, the optimal solution o^* of the above optimization problem is

$$o^* \leq \sum_{t=1}^{t^*} \frac{T^{2/3} \text{mas}(G_{(t)}) c^{4/3}}{(1-1/e)\text{mas}^{2/3}(G_{(T)})}, \tag{24}$$

where $t^* = \lceil T^{2/3}c^{-2/3}\text{mas}^{1/3}(G_{(T)})\rceil$. Using (19), (20), (21), (22) and (24), we have

$$\mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\sum_{i\in[K]} p_{i,k_j}\sum_{t=k_j}^{k_j+r_j-1}\ell_t(i) - \sum_{j=1}^{T}\ell_t(k')\right] \leq \frac{\log(K)}{\eta} + \frac{\eta}{2}\sum_{t=1}^{t^*}\frac{T^{2/3}c^{4/3}\text{mas}(G_{(t)})}{(1-1/e)\text{mas}^{2/3}(G_{(T)})}. \quad (25)$$

Additionally, the player switches its action only if E^t is true. Thus, using (25) and $c(i,j) = c$, for all $i, j \in [K]$, we have

$$R^{\mathcal{A}}(l_{1:T}, \mathcal{C}) \leq \frac{\log(K)}{\eta} + \frac{\eta}{2}\sum_{t=1}^{t^*}\frac{T^{2/3}c^{4/3}\text{mas}(G_{(t)})}{(1-1/e)\text{mas}^{2/3}(G_{(T)})} + c\cdot\mathbb{E}[\sum_{t=2}^{T}\mathbf{1}(i_t \neq i_{t-1})]. \quad (26)$$

Now, we bound $\mathbb{E}[\sum_{t=2}^{T}\mathbf{1}(i_t \neq i_{t-1})]$. E_1^t occurs with probability 1, and does not contribute to any SC. E_2^t can lead to at most $\lceil T^{2/3}c^{-2/3}\text{mas}^{1/3}(G_{(T)})\rceil$ switches. Now, let E_3^t causes N_T switches. Then, we have

$$\begin{aligned}
\mathbb{E}[N_T] &= \mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\mathbf{1}(i_{t_{j+1}} \neq i_{t_j}, E_3^{t_j} \text{ is true})\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\mathbb{E}\Big[\mathbf{1}(i_{t_{j+1}} \neq i_{t_j}, E_3^{t_j} \text{ is true})\Big|\{p_{t_j}, r_j\}_{1\leq j\leq\sigma(T)}\Big]\right] \\
&\leq \mathbb{E}\left[\sum_{j=1}^{\sigma(T)}\mathbb{E}\Big[\sum_{i\in[K],k'\in[K]\setminus\{i\}}\mathbb{P}(i_{t_j}=i\big|E_3^{t_j}\text{ is true})\mathbb{P}(i_{t_{j+1}}=k'\big|i_{t_j}=i)\Big|\{p_{t_j},r_j\}_{1\leq j\leq\sigma(T)}\Big]\right] \\
&= \mathbb{E}\Big[\sum_{j=1}^{\sigma(T)}\sum_{i\in[K],k'\in[K]\setminus\{i\}} p_{i,t_j}p_{k',t_{j+1}}\Big] \\
&\overset{(a)}{\leq} \sum_{t=1}^{T} c^{-2/3}\text{mas}^{1/3}(G_{(T)})t^{-1/3} = c^{-2/3}\text{mas}^{1/3}(G_{(T)})T^{2/3},
\end{aligned} \quad (27)$$

where (a) follows from Lemma 1 in this section. Thus, the number of switches are $2c^{-2/3}\text{mas}^{1/3}(G_{(T)})T^{2/3}$, and the SC is $2c^{1/3}\text{mas}^{1/3}(G_{(T)})T^{2/3}$.

Part (iii) of the theorem follows by combining the results from (i) and (ii). Part (iv) follows from the fact that if G_t is undirected, $\text{mas}(G_t) = \alpha(G_t)$. □

Lemma 1. *Given $i \in [K]$ is chosen at time instance t_j, for all $k' \in [K]\setminus\{i\}$, we have*

$$p_{i,t_j}\cdot p_{k',t_{j+1}} \leq (t_{j+1})^{-1/3}.$$

Proof. Given i is chosen at time instance t_j, for all $k' \in [K]\setminus\{i\}$, we have

$$\begin{aligned}
\frac{p_{k',t_{j+1}}}{p_{i,t_{j+1}}} &= \frac{p_{k',1}\exp(-\eta\hat{\ell}_{t_{j+1}}(k'))}{p_{i,t_j}\exp(-\eta\ell'_{t_j+r_j-1}(i))} \\
&\overset{(a)}{=} \frac{p_{k',1}\exp(-\eta(\hat{\ell}_{t_j}(k') + \ell'_{t_j+r_j-1}(k')))}{p_{i,t_j}\exp(-\eta\ell'_{t_j+r_j-1}(i))} \\
&\overset{(b)}{\leq} \frac{\exp\big(-\eta(\hat{\ell}_{t_j}(k') + \ell'_{t_j+r_j-1}(k') - \ell'_{t_j+r_j-1}(i))\big)}{p_{i,t_j}} \\
&\overset{(c)}{\leq} \frac{\exp\big(-\eta(\epsilon_{t_{j+1}}/\eta)\big)}{Kp_{i,t_j}} \\
&= \frac{\exp\big(-\epsilon_{t_{j+1}}\big)}{p_{i,t_j}},
\end{aligned} \quad (28)$$

where (a) follows from the fact that $\hat{\ell}_{t_{j+1}}(k') = \hat{\ell}_{t_j}(k') + \ell'_{t_j+r_j-1}(k')$; (b) follows from $p_{k',1} = 1/K$; (c) follows from the fact that for all $k \in [K]\backslash\{i\}$, $\hat{\ell}_{k,t-1} - \ell'_{i,t-1} > \epsilon_t/\eta$ as the increment in $\ell'_{i,t-1}$ is bounded by $1/q_{i,t-1}$. Now, replacing $\epsilon_t \geq \log(tc^2/\text{mas}(G_{(T)}))/3$ in (28), we have

$$p_{i,t_j} \cdot p_{k',t_{j+1}} \leq c^{-2/3}\text{mas}^{1/3}(G_{(T)})t_{j+1}^{-1/3}. \tag{29}$$

□

F Proof of Theorem 6

Proof. We borrow the notations from the proof of Theorem 5. Using the fact that η_t is decreasing in t and (19), we have

$$\sum_{j=1}^{\sigma(T)} \sum_{i\in[K]} p_{i,t_j} \cdot \ell'_{t_j+r_j-1}(i) - \min_{k'\in[K]} \sum_{j=1}^{\sigma(T)} \ell'_{t_j+r_j-1}(k') \leq \frac{\log(K)}{\eta_T} + \sum_{j=1}^{\sigma(T)} \frac{\eta_{t_j}}{2} \sum_{i\in[K]} p_{i,t_j} \cdot \ell'^2_{t_j+r_j-1}(i). \tag{30}$$

Now, taking expectation on both the sides and using the fact that expectation of the $\min(.)$ is smaller than the $\min(.)$ of the expectation, we have

$$\begin{aligned}
&\mathbf{E}\left[\sum_{j=1}^{\sigma(T)} \sum_{i\in[K]} p_{i,t_j} \cdot \ell'_{t_j+r_j-1}(i)\right] - \min_{k'\in[K]} \mathbf{E}\left[\sum_{j=1}^{\sigma(T)} \ell'_{t_j+r_j-1}(k')\right] \\
&\leq \frac{\log(K)}{\eta_T} + \mathbf{E}\left[\sum_{j=1}^{\sigma(T)} \frac{\eta_{t_j}}{2}\epsilon_{t_j}\mathbf{E}\left[\sum_{i\in[K]} p_{i,t_j} \cdot \ell'^2_{t_j+r_j-1}(i)|p_{t_j}, r_j, \mathbf{1}(i_t \text{ is selected using } p_t)\right]\right], \\
&\overset{(a)}{\leq} \frac{\log(K)}{\eta_T} + \mathbf{E}\left[\sum_{j=1}^{\sigma(T)} \frac{\eta_{t_j}}{2}\epsilon_{t_j}\mathbf{E}[\text{mas}(G_{t_j:t_j+r_j-1})r_j^2|\mathbf{1}(i_t \text{ is selected using } p_t)]\right], \\
&\overset{(b)}{\leq} \frac{\log(K)}{\eta_T} + \mathbf{E}\left[\sum_{j=1}^{\sigma(T)} \frac{\eta_{t_j}}{2}\epsilon_{t_j}\frac{2\cdot\text{mas}(G_{t_j:t_j+r_j-1})}{\epsilon_{t_j}^2}\right], \\
&= \frac{\log(K)}{\eta_T} + \mathbf{E}\left[\sum_{j=1}^{\sigma(T)} \frac{\eta_{t_j}}{2}\frac{2\cdot\text{mas}(G_{t_j:t_j+r_j-1})}{\epsilon_{t_j}}\right], \\
&\overset{(c)}{\leq} \frac{\log(K)}{\eta_T} + \mathbf{E}\left[\sum_{j=1}^{\sigma(T)} \frac{2\log(K)}{\text{mas}^{2/3}(G_{(T)})}\text{mas}(G_{(j)})\right], \\
&\overset{(d)}{\leq} \frac{\log(K)}{\eta_T} + \sum_{j=1}^{\mathbf{E}[\sigma(T)]} \frac{2\log(K)}{\text{mas}^{2/3}(G_{(T)})}\text{mas}(G_{(j)})
\end{aligned} \tag{31}$$

where (a) follows from (22), (b) follows from the fact that since the probability of selecting a new action is at most ϵ_{t_j}, the mean and the variance of the geometric random variable r_j is bounded by $1/\epsilon_{t_j}^2$ and $(1-\epsilon_{t_j})/\epsilon_{t_j}^2$ respectively, (c) follows from the value of η_t and ϵ_t, and (d) follows from the fact that $\text{mas}(G_{(j)})/\text{mas}(G_{(T)})$ is a monotonic non increasing sequence in j, therefore the summation is a concave function and the inequality follows from the Jensen's inequality.

Now, we bound the $\mathbf{E}[\sigma(T)]$ in (31). This also gives a bound on the number of switches performed by the algorithm. We have

$$\begin{aligned}
\mathbf{E}[\sigma(T)] &= \sum_{t=1}^{T} \mathbf{E}[\mathbf{1}(i_t \neq i_{t-1})], \\
&\leq \sum_{t=1}^{T} \epsilon_t, \\
&\leq 0.5\text{mas}^{1/3}(G_{(T)})T^{2/3}c^{1/3}
\end{aligned} \tag{32}$$

□

References

Alon, N., Cesa-Bianchi, N., Gentile, C., Mannor, S., Mansour, Y., and Shamir, O. (2017). Nonstochastic multi-armed bandits with graph-structured feedback. *SIAM Journal on Computing*, 46(6):1785–1826.

Dekel, O., Ding, J., Koren, T., and Peres, Y. (2014). Bandits with switching costs: T 2/3 regret. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 459–467. ACM.

Nemhauser, G. L. and Wolsey, L. A. (1978). Best algorithms for approximating the maximum of a submodular set function. *Mathematics of operations research*, 3(3):177–188.