Pure Exploration with Feedback Graphs (Supplementary Material)

Contents

A Additional Numerical Results

In this section of the appendix we describe the graphs used in the numerical results; the details of the algorithms used and exhibits additional numerical results.

A.1 Graphs details

Here we briefly describe the graphs used in the numerical results. Also refer to the code for more details.

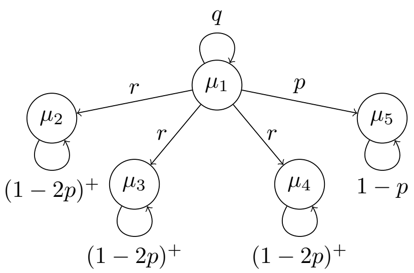

Figure 6: Loopy star graph.

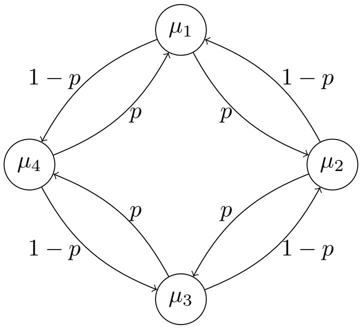

Figure 7: Ring graph.

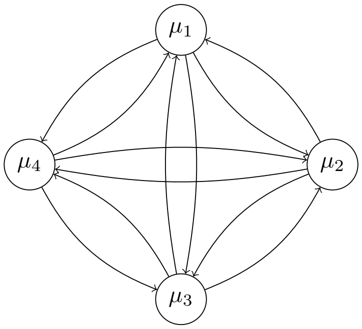

Figure 8: Loopless clique graph.

A.1.1 Loopy star graph

The loopy star graph (fig. 6) is the only graph with self-loops in our experiments. However, it depends on several parameter (p, q, r) that affect the underlying topology. The rewards are Gaussianly distributed, with variance 1. The best arm has average reward 1, while sub-optimal arms have average reward 0.5.

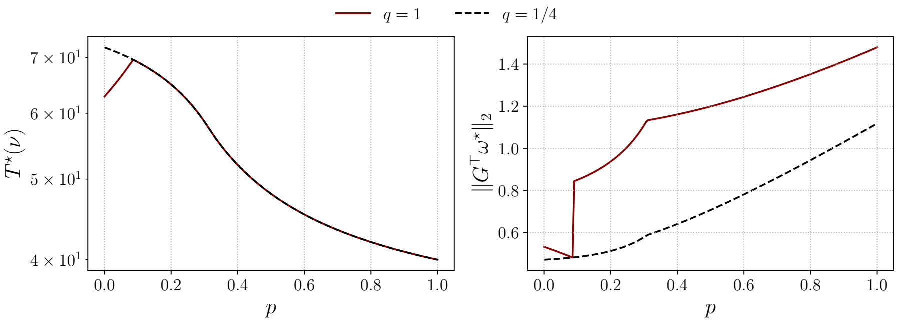

Figure 9: On the left: characteristic time of the loopy star graph with $K = 5, r = 0.5$ for different values of p, q (the best arm is $v = 5$). On the right: plot of $\|m^\star\|$ as a function of $\omega^\star$.

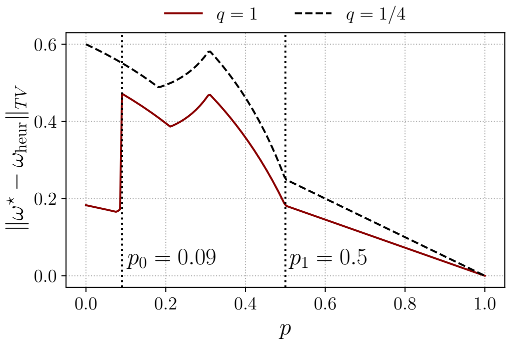

Figure 10: Difference between $\omega^\star$ and ω_{heur} in the loopy star graph (same setting as in fig. 9).

We simulated two settings:

1. Main setting: we set $p = 1/5, q = 1/4, r = 1/5$. Hence, self-loops may bring more information. The best arm is $v = 5$. In fig. 9 and fig. 10 we depict $T^\star(\nu), \|G^\top \omega^\star\|$ and the difference $\|\omega^\star - \omega_{\text{heur}}\|$ for this setting. Notably, for increasing values of p the approximate solution ω_{heur} converges to the optimal solution.
2. Alternative setting: we set $p = 0, q = 1/4, r = \frac{1-2q}{4(K-1)}$. Hence, it is not worth for the agent to choose $v = 1$, and the optimal arm is $v = 1$. In fig. 11 and fig. 12 we depict $T^\star(\nu), \|G^\top \omega^\star\|$ and the difference $\|\omega^\star - \omega_{\text{heur}}\|$ for this setting. Also in this case for increasing values of p the approximate solution ω_{heur} converges to the optimal solution.

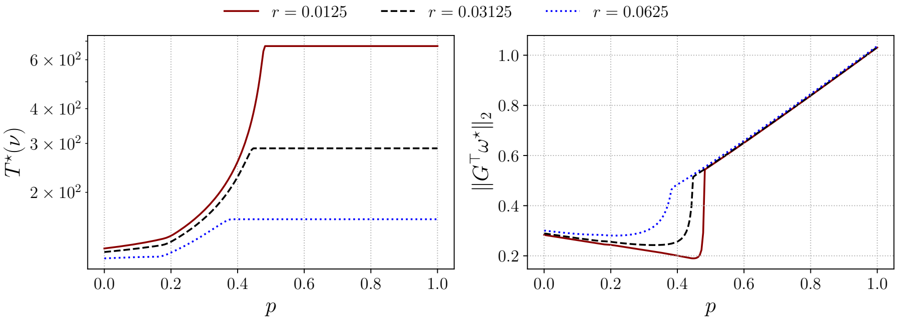

Figure 11: On the left: characteristic time of the loopy star graph in the alternative setting with $K = 5$ and $r = \alpha\frac{1-2q}{4(K-1)}, \alpha \in \{0.1, 0.25, 0.5\}$ (the best arm is $v = 1$). On the right: plot of $\|m^\star\|$ as a function of $\omega^\star$.

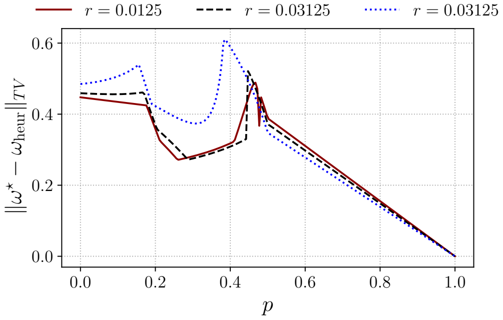

Figure 12: Difference between $\omega^\star$ and ω_{heur} in the loopy star graph (same setting as in fig. 11).

A.1.2 Ring graph

In this graph (fig. 7) each node is connected to two adjacent nodes. For each node u, the feedback from the node on the right v_r (clockwise direction) is "seen" with probability p, i.e., $G_{u,v_r} = p$; on the other hand, the feedback from the node on the "left" v_l (anti-clockwise direction) is "seen" with probability $1 - p$, i.e., $G_{u,v_l} = 1 - p$. All other edges have 0 probability. We used a value of $p = 0.3$ and rewards are Gaussian (with variance 1), with mean values linearly distributed in $[0, 1]$ across the K arms.

In fig. 13 and fig. 14 we depict $T^\star(\nu), \|G^\top \omega^\star\|$ and the difference $\|\omega^\star - \omega_{\text{heur}}\|$ for this setting. In this case the approximate solution ω_{heur} converges to the optimal solution for $p \to 0$ or $p \to 1$.

A.1.3 Loopless clique graph

This graph (fig. 8) is fully connected without any self-loops (the probabilities in the figure are omitted to avoid cluttering). Assuming the vertices $u \in V$ are numbered according to the natural numbers $V = \{1, 2, \ldots, K\}$, the

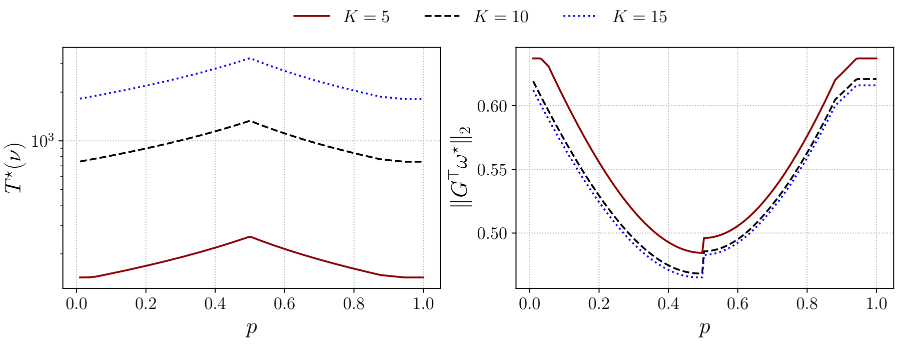

Figure 13: On the left: characteristic time of the ring graph for different values of K, p. On the right: plot of $\|m^\star\|$ as a function of $\omega^\star$.

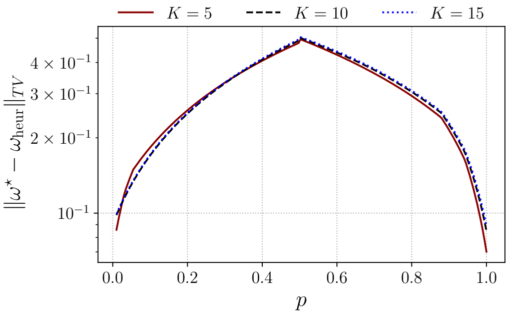

Figure 14: Difference between $\omega^\star$ and ω_{heur} in the ring graph (same setting as in fig. 13).

edge probabilities are

$$G_{u,v} = \begin{cases} 0 & v = u, \\ p/u & \forall v \neq u \wedge v \notin 2\mathbb{N}, \\ 1 - (p/u) & \text{otherwise.} \end{cases}$$

where $2\mathbb{N}$ is the set of even numbers. We see that in this type of graph it may be better to choose a vertex according to its index depending on what type of feedback the learner seeks. We used a value of $p = 0.5$ and rewards are Gaussian (with variance 1), with mean values linearly distributed in $[0, 1]$ across the K arms.

In fig. 15 and fig. 16 we depict $T^\star(\nu)$, $\|G^\top \omega^\star\|$ and the difference $\|\omega^\star - \omega_{\text{heur}}\|$ for this setting. In this case the approximate solution ω_{heur} does not seem to converge to $\omega^\star$ for any value of p.

A.2 Numerical results

In this section we present additional details on the numerical results.

A.2.1 Algorithms and results

In addition to TaS-FG and heuristic TaS-FG we also used EXP3.G (Alon et al., 2015), an algorithm for regret minimization in the adversarial case. We also compare with two variant of UCB: (1) UCB-FG-E, which acts greedily with respect to the upper confidence bound of $\left(\hat{G}(t)\hat{\mu}(t)\right)$; (2) UCB-FG-V, which selects $\arg\max_v \hat{G}^{\text{ucb}}_{v,\hat{a}^{\text{ucb}}(t)}(t)$, where $\hat{G}^{\text{ucb}}(t), \hat{a}^{\text{ucb}}(t)$ are, respectively, the UCB estimates of G and $a^\star$ at time t. For all algorithms, the graph estimator $\hat{G}(t)$ was initialized in an optimistic way, i.e., $\hat{G}_{u,v}(1) = 1$ for all $u, v \in V$.

EXP3.G algorithm. EXP3.G Alon et al. (2015) initializes two vectors $p, q \in \mathbb{R}^K$ uniformly, so that $p_i = q_i = 1/K$ for $i = 1, \dots, K$. At every time-step, an action V_t is drawn from $V_t \sim p_t$, where $p_t \leftarrow (1 - \eta)q_t + \eta\mathcal{U}$ with $\eta \in (0, 1)$ being an exploration facotr and $\mathcal{U}$ is the uniform distribution over $\{1, \dots, K\}$.

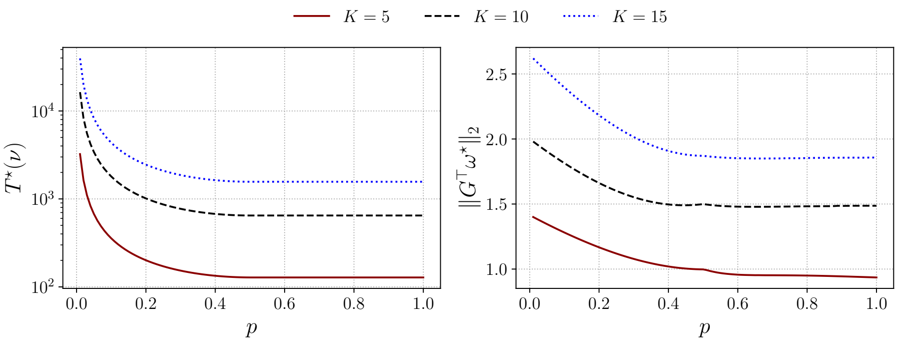

Figure 15: On the left: characteristic time of the loopless graph for different values of K, p. On the right: plot of $\|m^\star\|$ as a function of $\omega^\star$.

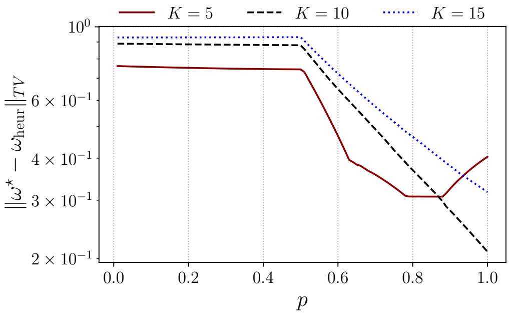

Figure 16: Difference between $\omega^\star$ and ω_{heur} in the loopless graph (same setting as in fig. 15).

After observing the feedback, the algorithm sets

$$\hat{q}_t \leftarrow q_t \exp(-\eta x_t) \text{ and } q_t \leftarrow \hat{q}_t / \sum_u \hat{q}_{t,u},$$

where $x_{t,u} = -Z_{t,u} / \sum_{v \in N_{in}(u)} p_{t,v}$. In the experiments, we let $\eta = 3/10$.

UCB-FG-E algorithm. This method is a variant of UCB that acts greedily with respect to the upper confidence bound of $\left(\hat{G}(t)\hat{\mu}(t)\right)$. In practice, we let $\hat{\mu}_u^{\text{ucb}}(t) = \hat{\mu}_u(t) + \sqrt{\frac{2\ln(1+t)}{M_u(t)}}$, $\hat{G}_{u,v}^{\text{ucb}}(t) = \hat{G}_{u,v}(t) + \sqrt{\frac{\ln(1+t)}{2N_u(t)}}$ and select the action to take according to $V_t = \arg\max_u \hat{G}^{\text{ucb}}(t)\hat{\mu}^{\text{ucb}}(t)$.

UCB-FG-V algorithm. This method is a variant of UCB that selects $V_t \arg\max_v \hat{G}^{\text{ucb}}_{v,\hat{a}^{\text{ucb}}(t)}(t)$, where $\hat{G}^{\text{ucb}}(t), \hat{a}^{\text{ucb}}(t)$ are, respectively, the UCB estimates of G and $a^\star$ at time t (note that $a^{\text{ucb}}(t) = \arg\max_u \hat{\mu}_u^{\text{ucb}}(t)$).

In fig. 17 we depict the sample complexity of the algorithms for different values of K, δ. Note that the sample complexity τ is not normalized, and results were computed over 100 seeds.

A.2.2 Libraries and computational resources

Libraries used in the experiments. We set up our experiments using Python 3.10.12 (Van Rossum and Drake Jr, 1995) (For more information, please refer to the following link `http://www.python.org`), and made use of the following libraries: NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), Seaborn (Waskom et al., 2017), Pandas (McKinney et al., 2010), Matplotlib (Hunter, 2007), CVXPY (Diamond and Boyd, 2016). As numerical optimizer we used Gurobi 10.0.1 (Gurobi Optimization, LLC, 2024).

New code is published under the MIT license. To run the code, please, read the attached `README.md` file for instructions.

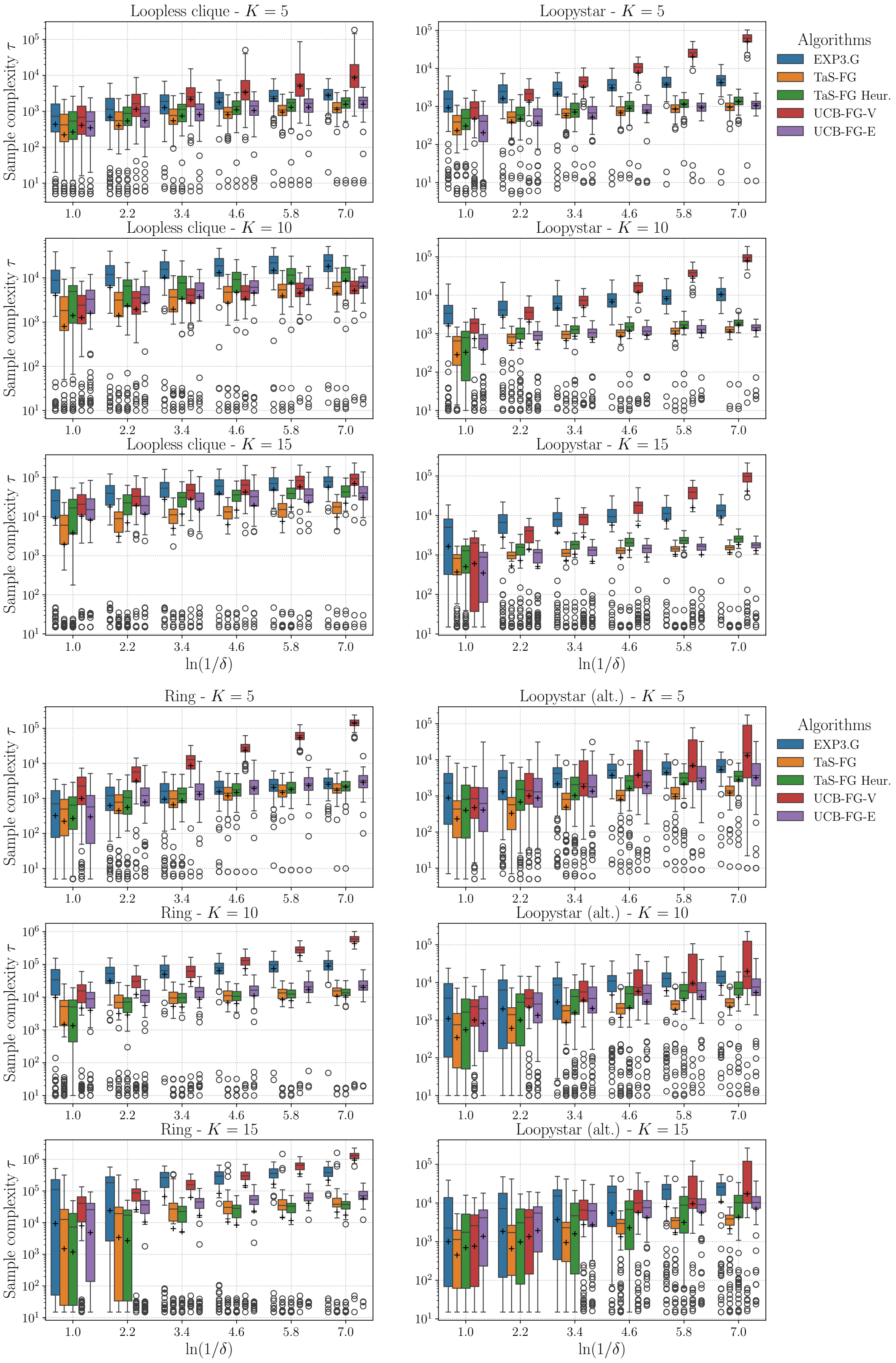

Figure 17: Sample complexity results, from top-left (clockwise): loopless clique, loopystar, ring and alternative loopystar graphs. Box plots are not normalized, and were computed over 100 seeds. Boxes indicate the interquartile range, while the median and mean values are, respectively, the solid line and the + sign in black.

Computational resources. Experiments were run on a shared cluster node featuring a Linux OS with 2 fourteen-core 2.6 GHz Intel Gold 6132 and 384GB of ram. The total computation time per core to design the experiments, debug the code and obtain the final results was roughly of 82 hours/core. To obtain the final results over 100 seeds we estimated that 10/hours/core are sufficient.

B Feedback Graph Properties

In the following we list some properties for feedback graphs.

Lemma 1. *For a strongly observable graph G We have that $|SO(G)| \geq \max(\sigma(G), \alpha(G))$.*

Proof. First, note that $\sigma(G) \leq |SO(G)|$, since all vertices with a self-loop are strongly observable.

To prove that $|SO(G)| \geq \alpha(G)$, by contradiction, assume that $\alpha(G) > |SO(G)|$. First, for any $I \in \mathcal{I}, u \in I$, u is strongly observable. Since $|I| = \alpha(G)$, then at-least $|SO(G)| \geq \alpha(G)$, which is a contradiction. □

Lemma 2. *Assume $|\mathcal{I}(G)| > 1$ and let $I_1, I_2 \in \mathcal{I}(G)$, with $I_1 \neq I_2$. Then, for any $u \in I_1$ we have that there exists $v \in I_2 \setminus I_1$ satisfying $G_{uv} > 0$ or $G_{vu} > 0$.*

Proof. Let $I_1, I_2 \in \mathcal{I}(G)$ satisfying $I_1 \neq I_2 \Rightarrow \exists v \in I_2 \setminus I_1$. Let $u \in I_1$. By contradiction, for all $v \in I_2 \setminus I_1$ we have that $G_{u,v} = 0$ and $G_{v,u} = 0$. In that case, we can construct a new a new set $\tilde{I} = I_1 \cup \{v\}$ such that $G_{u,v} = 0$ and $G_{v,u} = 0$ for any $u, v \in \tilde{I}$, implying that $\tilde{I} \in \mathcal{I}(G)$. But $|I_1| < |\tilde{I}|$, which contradicts the fact that $I_1 \in \mathcal{I}(G)$. □

Lemma 3. *For a strongly observable graph if $\alpha(G) > 1$ then $\forall I \in \mathcal{I}(G), \forall u \in I$ we have $\{u\} \in N_{in}(u)$, that is, all vertices in I have self-loops. As a corollary, we have that $\sigma(G) \geq \alpha(G)$ if $\alpha(G) > 1$.*

Proof. By contradiction, assume there exists $I_0 \in \mathcal{I}(G)$ with $u \in I_0$ such that $\{u\} \notin N_{in}(u)$. Since u is strongly observable, it means that $V \setminus \{u\} \in N_{in}(u)$.

Now, consider the case $\alpha(G) > 1$. If that is the case, then let $v \in I_0$. By strong observability of u, we have $\{v\} \in N_{in}(u)$, which contradicts the fact that I_0 is an independent set.

The latter statement is a consequence of the fact that all vertices in every I have self-loops. □

Corollary 2. *Consider a strongly observable graph with $\alpha(G) > 1$. Then, $\forall I \in \mathcal{I}$ there exists $I_0 \in \mathcal{I}$ such that $I \ll I_0$.*

Proof. By contradiction assume that $\exists I \in \mathcal{I}(G)$ such that $\forall I_0 \in \mathcal{I}(G)$ there exists $u(I_0) \in I$ such that $N_{in}(u(I_0)) \cap I_0 = \emptyset$, where $u : \mathcal{I}(G) \to V$. However, since the graph is strongly observable, we have that either (1) $\{u(I_0)\} \in N_{in}(u(I_0))$, or (2) $V \setminus \{u(I_0)\} \subset N_{in}(u(I_0))$ or (3) both.

Consider the case $\alpha(G) > 1$. By lemma 3 then each vertex in I has a self-loop. Hence taking $I_0 = I$ leads to $N_{in}(u(I)) \cap I \ni \{u(I)\}$, which is a contradiction. □

Lemma 4. *Consider a strongly observable graph. Then $|SO(G)| = \sigma(G) = \alpha(G)$ only for bandit feedback graphs. As a corollary, for non-bandit feedback graphs we have $|SO(G)| > \alpha(G)$.*

Proof. The first part of the lemma is easy to prove as $\alpha(G)$ is maximal when all the vertices have only self-loops, thus $\alpha(G) = |SO(G)| = K$.

To prove the second part, note that it always holds true that $K = |SO(G)| \geq \alpha(G)$ by lemma 1. However, equality is reached only for bandit feedback graphs. Therefore, for other feedback graphs it holds that $|SO(G)| > \alpha(G)$. □

Lemma 5. *Consider a set of vertices with self-loops in a graph $L(G)$ of size $n+1$. Assume the independence number of $L(G)$ is n. Then, at-most n vertices in $L(G)$ are needed to dominate $L(G)$.*

Proof. The proof is simple, and follows from the fact that since $\alpha(L(G)) = n$, then there must exists $v, u \in L(G)$ such that $u \in N_{out}(v)$. Then, $L(G) \setminus \{u\}$ dominates $L(G)$. □

Lemma 6. *For any set G satisfying $\alpha(G) = k$, we must have* $\max(1, k + |V| - |G|) \leq \alpha(V) \leq k$ *for any subset V of G.*

Proof. The right hand-side is trivial since any subset $V \subset G$ can have at-most k independent vertices. The left hand-side follows from the fact that removing an element from G can at-most reduce the number of independent vertices by 1. □

Lemma 7. *Let $E(G) = SO(G) \setminus L(G)$ be the set of strongly observable vertices without a self-loop. The following statements are true*

- *If $SO(G) \setminus E(G) \neq \emptyset$, then at-most $\sigma(G) - \left\lfloor \frac{\sigma(G)}{\alpha(G)+1} \right\rfloor$ vertices in $L(G)$ are required to dominate the set of strongly observable vertices $SO(G)$.*
- *If $SO(G) \setminus E(G) = \emptyset$, then 2 vertices are required to dominate the set of strongly observable vertices $SO(G)$.*

Proof. Let $L(G) = \{v \in V : \{v\} \in N_{in}(v)\}$ to be the set of vertices with a self-loop. It follows that $L(G) \cup E(G) = SO(G)$.

First, note that $|SO(G)| \geq |L(G)| = \sigma(G)$. For any $v \in E(G)$, we must have that $N_{in}(v) = V \setminus \{u\}$.

Case 1. For the first statement, note that any vertex in $L(G)$ dominates also $E(G)$. Therefore we only need to find the domination number of $L(G)$. Consider the following cases:

1. If $\alpha(G) = 1$ and $\sigma(G) = 1$, then it suffices to pick $v \in L(G)$ to dominate $E(G)$, since $v \in N_{in}(u)$ for any $u \in E(G)$, hence $E(G) \subset N_{out}(v)$. Since v has a self-loop, and $L(G) = \{v\}$, it satisfies $L(G) \subset N_{out}(v)$. Therefore $SO(G) \subset N_{out}(v)$.
2. If $\alpha(G) = 1$ and $\sigma(G) = 2$, then it must either be $G_{u,v} > 0$ or $G_{v,u} > 0$ for $L(G) = \{u, v\}$. Similarly as before, one can choose one vertex in $L(G)$ to satisfy $SO(G) \subseteq N_{out}(v)$.
3. If $\alpha(G) = 1$ and $\sigma(G) = 3$ then we can build two sets $L_1(G), L_2(G)$, each of size 2, such that $L_1(G) \cup L_2(G) = L(G)$ and $|L_1(G)| = |L_2(G)|$. By lemma 6 we have $\alpha(L_1(G)) = \alpha(L_2(G)) = 1$. Then, because of the previous case with $\sigma(G) = 2$, each set $L_i(G)$ is dominated by 1 vertex, thus 2 vertices are required to dominate $L(G)$. Since any vertex in $L(G)$ also dominates $E(G)$, then 2 vertices are needed to dominate $SO(G)$.
4. If $\alpha(G) = 1$ and $\sigma(G) > 1$, then we need at most $\sigma(G) - \lfloor \sigma(G)/2 \rfloor$ vertices to dominate $L(G)$ (and thus also $SO(G)$). By induction, assume it holds for $\sigma(G_n) = n$, where we indicate by G_n the set where the number of self-loops vertices is n, and by V_n we denote the corresponding set of vertices of the graph.
   - We add a new vertex $\tilde{v}$ to V_n to create $V_{n+1} = V_n \cup \{\tilde{v}\}$ and G_{n+1}, such that $\tilde{v}$ has a self-loop and $\alpha(G_{n+1}) = \alpha(G_n) = 1$. We also let $\alpha_n = \alpha(G_n), \sigma_n = \sigma(G_n)$, etc.
   - If σ_{n+1} is even, we can create $k = \sigma_{n+1}/2$ sets $L_i(G_{n+1})$ each of size 2, so that their union is equal to $L(G_{n+1})$, i.e., $\cup_{i=1}^k L_i(G_{n+1}) = L(G_{n+1})$. From the case $\alpha(G) = 1, \sigma(G) = 2$, only one vertex from in each $L_i(G_{n+1})$ is needed to dominate that subset. Hence the domination number is at most $k = \sigma_{n+1}/2 = \sigma_{n+1} - \sigma_{n+1}/2$.
   - If σ_{n+1} is not even, then we simply take the set of dominating vertices of $L(G_n)$ and add $\tilde{v}$: $L(G_n) \cup \{\tilde{v}\}$. Then, the number of dominating vertices is
   $$\sigma_n/2 + 1 = (\sigma_{n+1} - 1)/2 + 1 \pm \sigma_{n+1} = \sigma_{n+1} - \frac{\sigma_{n+1} - 1}{2} = \sigma_{n+1} - \lfloor \sigma_{n+1}/2 \rfloor.$$
5. Now, assume $1 < \alpha(G)$ and $\alpha(G) \geq \sigma(G)$.
   - Then $L(G)$ is an independent set, and we need $\sigma(G)$ elements to dominate $L(G)$. Since any element in $L(G)$ also dominates $E(G)$, we need at-most $\sigma(G)$ vertices to dominate $SO(G)$.
6. If $1 < \alpha(G) < \sigma(G)$ we prove by induction that the domination number of $L(G)$ is at most $\sigma(G) - \lfloor \frac{\sigma(G)}{\alpha+1} \rfloor$.
   - As before, let $\alpha_n = \alpha(G_n)$, $\sigma_n = \sigma(G_n)$. We also indicate by V_n the set of vertices.
   - Divide $L(G_n)$ into subsets $L_1(G_n), \dots, L_k(G_n)$ each of size $\alpha_n + 1$ with $k = \lfloor \sigma_n/(\alpha_n + 1) \rfloor$ such that $\cup_{i=1}^k L_i(G_n) = L(G_n)$. Hence, if $\alpha_n + 1$ is a divisor of σ_n, we can obtain k non-overlapping sets, otherwise we have two sets with an overlapping element. By lemma 6 we also have that $\alpha(L_i(G_n)) \leq \alpha_n$ for any $i = 1, \dots, k$.
   - Suppose the result holds for n. We add another vertex $\tilde{v}$ to the graph, such that $\tilde{v}$ has a self-loop. We denote by $V_{n+1} = V_n \cup \{\tilde{v}\}$ the new set of vertices, similarly G_{n+1}, etc.
     - If $\alpha_{n+1} = \alpha_n$, then adding $\tilde{v}$ has not increased the independence number. Assume $\alpha_{n+1} + 1$ is not a divisor of σ_{n+1}. We partition $L_{G_{n+1}}$ into $k = \lfloor \sigma_{n+1}/(1 + \alpha_{n+1}) \rfloor$ sets $L_i(G_{n+1})$ each of size α_{n+1},

plus an additional set, which obviously is of size $r = \sigma_{n+1} - (\alpha_{n+1}+1)k$. Each of these subsets is at most dominated by α_{n+1} vertices by lemma 5, except the last one, which is dominated at most by r vertices. Then, the domination number is at most

$$\alpha_{n+1}k + r = \alpha_{n+1}k + \sigma_{n+1} - (\alpha_{n+1}+1)k = \sigma_{n+1} - k = \sigma_{n+1} - \left\lfloor \frac{\sigma_{n+1}}{\alpha_{n+1}+1} \right\rfloor .$$

- On the other hand if $\alpha_{n+1} = \alpha_n$, and $\alpha_{n+1}+1$ is a divisor of σ_{n+1}, then we can apply the same argument as before, and split $L(G_{n+1})$ into $k = \sigma_{n+1}/(\alpha_{n+1}+1)$ subsets each of size $\alpha_{n+1}+1$. By lemma 5, each subset has α_{n+1} vertices that dominate that subset. Hence we have the following number of dominating vertices

$$\alpha_{n+1}\frac{\sigma_{n+1}}{\alpha_{n+1}+1} = \sigma_{n+1} + \alpha_{n+1}\frac{\sigma_{n+1}}{\alpha_{n+1}+1} - \sigma_{n+1} = \sigma_{n+1} - \frac{\sigma_{n+1}}{\alpha_{n+1}+1} = \sigma_{n+1} - \left\lfloor \frac{\sigma_{n+1}}{\alpha_{n+1}+1} \right\rfloor .$$

- If $\alpha_{n+1} = \alpha_n + 1$ then adding $\tilde{v}$ increased the independence number . Nonetheless, we can apply the same arguments as before. If $\alpha_{n+1}+1$ is a divisor of σ_{n+1} then we can repeat the same argument as above. Construct $k = \sigma_{n+1}/(1+\alpha_{n+1})$ sets $L_i(G_{n+1})$ such that $\cup_i L_i(G_{n+1}) = L(G_{n+1})$, each of size $\alpha_{n+1}+1$ and at most dominated by α_{n+1} vertices by lemma 5. Hence we obtain that the domination number is at most $\sigma_{n+1} - \sigma_{n+1}/(\alpha_{n+1}+1) = \sigma_{n+1} - \lfloor \sigma_{n+1}/(\alpha_{n+1}+1) \rfloor$.

  Otherwise, if $\alpha_{n+1}+1$ is not a divisor of σ_{n+1}, we partition $L_{G_{n+1}}$ into $k = \lfloor \sigma_{n+1}/(1+\alpha_{n+1}) \rfloor$ sets $L_i(G_{n+1})$ each of size α_{n+1}, plus an additional set, which obviously is of size $r = \sigma_{n+1} - (\alpha_{n+1}+1)k$. Each of these subsets is at most dominated by α_{n+1} vertices by lemma 5, except the last one, which is dominated at most by r vertices. Then, the domination number is at most

$$\alpha_{n+1}k + r = \alpha_{n+1}k + \sigma_{n+1} - (\alpha_{n+1}+1)k = \sigma_{n+1} - k = \sigma_{n+1} - \left\lfloor \frac{\sigma_{n+1}}{\alpha_{n+1}+1} \right\rfloor .$$

In conclusion, for $\sigma(G) > \alpha(G)$ one can always say that the number of dominating vertices is at-most $\sigma(G) - \lfloor \frac{\sigma(G)}{\alpha(G)+1} \rfloor$. On the other hand, if $\sigma(G) \leq \alpha(G)$ we obtain that $\sigma(G)$ vertices are needed at-most. Since in this case $\lfloor \frac{\sigma(G)}{\alpha(G)+1} \rfloor = 0$, we can conclude that in general the domination number is at-most $\sigma(G) - \left\lfloor \frac{\sigma_{n+1}}{\alpha_{n+1}+1} \right\rfloor$.

Case 2. For the second statement, we have that all vertices in the graph do not have a self-loops, but all vertices are strongly observable. So each vertex u has an in-degree number of $N_{in}(u) = K - 1$. Hence, two vertices are needed to dominate the entire graph.

□

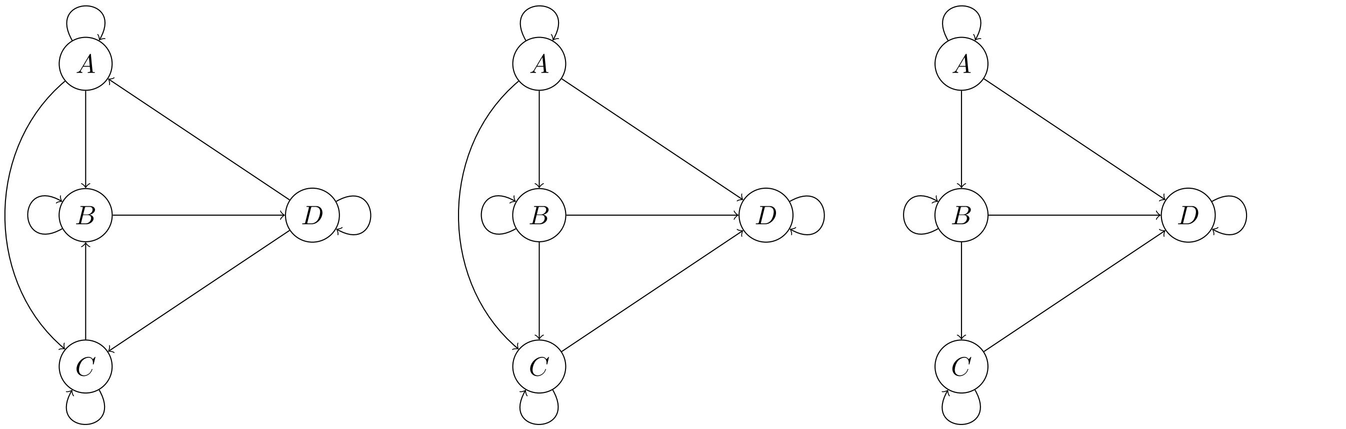

Figure 18: Example of strongly observable graphs and their domination number. **On the left**: a graph with $\sigma(G) = 4$ and $\alpha(G) = 1$. The smallest sets of vertices that dominate this graph are $\{A, B\}, \{B, D\}$. The maximally independent sets are $\mathcal{I} = \{\{A\}, \{B\}, \{C\}, \{D\}\}$. Note that $\sigma(G) - \lfloor \sigma(G)/(\alpha(G)+1) \rfloor = 4 - \lfloor 4/2 \rfloor = 2$. **In the middle**: a graph with $\alpha(G) = 1, \sigma(G) = 4$. The smallest set of vertices that dominate the graph is $\{A\}$. The maximally independent sets are $\mathcal{I} = \{\{A\}, \{B\}, \{C\}, \{D\}\}$. We have $\sigma(G) - \lfloor \sigma(G)/(\alpha(G)+1) \rfloor = 2$. **On the right**: a graph with $\alpha(G) = 2, \sigma(G) = 4$. The smallest sets of vertices that dominate the graph are $\{A, B\}, \{A, C\}$. The maximally independent sets are $\mathcal{I} = \{\{A, C\}\}$. We have $\sigma(G) - \lfloor \sigma(G)/(\alpha(G) + 1) \rfloor = 4 - \lfloor 4/3 \rceil = 3$.

C Sample Complexity Lower Bounds

The sample complexity analysis delves on the required minimum amount of *evidence* needed to discern between different hypotheses (e.g., vertex v is optimal vs vertex v is not optimal). The evidence is quantified by the log-likelihood ratio of the observations under the true model and a *confusing model.* This confusing model, is usually, the model that is *statistically* closer to the true model, while admitting a different optimal vertex.

To state the lower bounds, we first define the concept of *absolute continuity* between two models. For any two models $\nu = \{G, (\nu_u)_{u \in V}\}, \nu' = \{G', (\nu'_u)_{u \in V}\}$ with the same number of vertices we say that ν is absolutely continuous w.r.t ν', that is $\nu \ll \nu'$, if for all $(v, u) \in V^2$ we have $\nu_{v,u} \ll \nu'_{v,u}$

Given this definition of absolute continuity, we can define the set of confusing models as follows

$$\mathrm{Alt}(\nu) \coloneqq \{\nu' : a^\star(\nu) \neq a^\star(\nu'), \nu \ll \nu'\},$$

which is the set of models for which $a^\star(\nu)$ is not optimal. We also denote by $\mathrm{Alt}_u(\nu) = \{\nu' : \mu'_u > \mu'_{a^\star}\}$ the set of models where $u \neq a^\star(\nu)$ may be the optimal vertex in ν'.

C.1 The uninformed setting

In this section we prove theorem 1 and proposition 4. We start by stating a general expression of the lower bound[2].

C.1.1 General lower bound expression

In theorem 4 we state a general expression for $T^\star(\nu)$ and then show in the next sections the proofs of theorem 1 and proposition 4.

[2] Note that in the uploaded paper, due to an oversight, assumption 1 is incorrectly stated. In our results we use the assumption that the graph is observable. In this full version of the paper we correct this typo (see assumption 1) and explicitly mention that our results are valid for an observable graph. While extending our results to accommodate non-fully observable graphs with at-least 2 observable vertices (incl. the best one) is straightforward, it necessitates minor modifications. Since all results stated are valid for an observable graph, to maintain consistency with the uploaded paper, and the supplementary material, we decided to just correct the assumption in this full version of the paper, and highlight this change.

Theorem 4. *For any δ-PC algorithm and any model ν satisfying assumption 1, we have that*

$$\mathbb{E}_\nu[\tau] \geq T^\star(\nu)\mathrm{kl}(\delta, 1-\delta), \tag{11}$$

where $(T^\star(\nu))^{-1}$ is equivalent to the following two expressions

$$(T^\star(\nu))^{-1} = \begin{cases} \sup_{\omega\in\Delta(V)} \inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{v\in V} \omega_v \sum_{u\in N_{out}(v)} \mathrm{KL}(\nu_{v,u}, \nu'_{v,u}), \\ \sup_{\omega\in\Delta(V)} \inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{u\in V} \sum_{v\in N_{in}(u)} \omega_v \mathrm{KL}(\nu_{v,u}, \nu'_{v,u}). \end{cases} \tag{12}$$

Proof. Consider two bandit models $\nu = \{G, (\nu_u)_u\}, \nu' = \{G', (\nu'_u)_u\}$ with the same number of vertices and unique optimal vertex in both models, such that $\nu \ll \nu'$. For each v there exists a measure λ_v such that ν_v and ν'_v have, respectively, densities f_v and f'_v. Similarly, for $\nu_{v,u}$ and $\nu'_{v,u}$ there exists a measure $f_{v,u}$ and $f'_{v,u}$ respectively.

First expression. Hence, consider the log-likelihood ratio between ν and ν' of the data observed up to time t, and consider writing it in terms of the out-neighborhood of the vertices selected by the algorithm:

$$\begin{aligned} L_t &= \ln \frac{\mathrm{d}\mathbb{P}_\nu(V_1, Z_1, \ldots, V_t, Z_t)}{\mathrm{d}\mathbb{P}_{\nu'}(V_1, Z_1, \ldots, V_t, Z_t)}, \\ &= \sum_{n=1}^t \sum_{u\in N_{out}(V_n)} \ln\left(\frac{f_{v,u}(Z_{n,u})}{f'_{v,u}(Z_{n,u})}\right), \\ &= \sum_{n=1}^t \sum_{v\in V} \sum_{u\in N_{out}(v)} \mathbf{1}_{\{V_n=v\}} \ln\left(\frac{f_{v,u}(Z_{n,u})}{f'_{v,u}(Z_{n,u})}\right), \\ &= \sum_{v\in V} \sum_{u\in N_{out}(v)} \sum_{s=1}^{N_v(t)} \ln\left(\frac{f_{v,u}(W_{s,(v,u)})}{f'_{v,u}(W_{s,(v,u)})}\right), \end{aligned}$$

where $(W_{s,(v,u)})_s$ is an i.i.d. sequence of samples observed from $\nu_{v,u}$. Hence, if we take the expectation with respect to ν, by Wald's lemma, we have that

$$\begin{aligned} \mathbb{E}_\nu[L_t] &= \sum_{v\in V} \sum_{u\in N_{out}(v)} \mathbb{E}_\nu[N_v(t)]\mathrm{KL}(\nu_{v,u}, \nu'_{v,u}), \\ &= \sum_{v\in V} \mathbb{E}_\nu[N_v(t)] \sum_{u\in N_{out}(v)} \mathrm{KL}(\nu_{v,u}, \nu'_{v,u}). \end{aligned}$$

Therefore, applying (Kaufmann et al., 2016, Lemma 1) at $t = \tau$, we find that for any δ-PC algorithm we have that

$$\sum_{v\in V} \mathbb{E}_\nu[N_v(\tau)] \sum_{u\in N_{out}(v)} \mathrm{KL}(\nu_{v,u}, \nu'_{v,u}) \geq \mathrm{kl}(\delta, 1-\delta).$$

Consider the set of confusing models $\mathrm{Alt}(\nu) = \{\nu' = (\mu', G') : a^\star(\mu) \neq a^\star(\mu'), \nu \ll \nu'\}$, and define the selection rate of a vertex v as $\omega_v = \mathbb{E}_\nu[N_v(\tau)]/\mathbb{E}_\nu[\tau]$. Then, by minimizing over the set of confusing models, and then optimizing $\omega = (\omega_v)_{v\in V}$ over the simplex $\Delta(V)$, we obtain

$$\mathbb{E}_\nu[\tau] \underbrace{\sup_{\omega\in\Delta(V)} \inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{v\in V} \omega_v \sum_{u\in N_{out}(v)} \mathrm{KL}(\nu_{v,u}, \nu'_{v,u})}_{=:(T^\star(\nu))^{-1}} \geq \mathrm{kl}(\delta, 1-\delta).$$

and therefore $\mathbb{E}_\nu[\tau] \geq T^\star(\nu)\mathrm{kl}(\delta, 1-\delta)$.

Second expression. The second version of $T^\star(\nu)$ comes from considering the in-neighborhood of v for each vertex:

$$
\begin{aligned}
L_t &= \sum_{n=1}^{t} \sum_{u\in V} \sum_{v\in N_{in}(u)} \mathbf{1}_{\{V_n=v\}} \ln\left(\frac{f_{v,u}(Z_{n,u})}{f'_{v,u}(Z_{n,u})}\right),\\
&= \sum_{u\in V} \sum_{v\in N_{in}(u)} \sum_{s=1}^{N_v(t)} \ln\left(\frac{f_{v,u}(W_{s,(v,u)})}{f'_{v,u}(W_{s,(v,u)}))}\right)
\end{aligned}
$$

Hence

$$
\mathbb{E}_\nu[L_t] = \sum_{u\in V} \sum_{v\in N_{in}(u)} \mathbb{E}_\nu[N_v(t)]\mathrm{KL}(\nu_{v,u}, \nu'_{v,u}),
$$

from which we can immediately conclude the proof by following the same steps as for the previous expression. □

C.1.2 Continuous vs discrete rewards

Before proceeding further in our analysis, we rewrite the KL-divergence in terms of the associated Radon-Nykodim derivatives. Note that for a product random variable $Z = XY$ with $Z \sim \nu_{X,Y}$, we have that $\mathbb{P}_Z(A) = \int_A f_Z(z)\mathrm{d}\mu(z)$ with respect to some dominating measure $\mu(z)$. We consider some cases:

- **Continuous case:** for Y distributed as a Bernoulli of parameter p, and X as a continuous r.v. with density $f_X(x)$ we have that $\mathbb{P}_Z(A) = (1-p)\mathbf{1}_{\{0\in A\}} + p\int_A f_X(z)\mathrm{d}\lambda(z)$ where λ is the Lebesgue measure Let $\mu(A) = \delta_0(A) + \lambda(A)$ be the dominating measure. To find $f_Z(z)$ we can apply the Radon-Nykodim derivative in $z = 0$, which tells us that

  $$
  \mathbb{P}_Z(0) = 1 - p = \int_{\{0\}} f_Z(z)\mathrm{d}\mu(z) = f_Z(0).
  $$

  On the other hand for $z \neq 0$ we have

  $$
  \mathbb{P}_Z(A) = p\int_A f_X(z)\mathrm{d}\lambda(z) = \int_A f_Z(z)\mathrm{d}\mu(z) = \int_A f_Z(z)\mathrm{d}\lambda(z) \Rightarrow f_Z(z) = pf_X(z) \text{ a.e.}
  $$

  Therefore

  $$
  f_Z(z) = (1-p)\mathbf{1}_{\{z=0\}} + pf_X(z)\mathbf{1}_{\{z\neq 0\}}.
  $$

  In words, the "continuous" part has no contribution to the overall probability mass when $z = 0$, since the Lebesgue measure of $\{X = 0\}$ is 0, while for $z \neq 0$ the main contribution comes from the continuous part. In the setting studied in this paper, the intuition is that when we observe 0, then almost surely we know its due to the edge not being activated.

- **Discrete case:** for Y distributed as a Bernoulli of parameter p, and X as a categorical r.v. over $\{0, \dots, N\}$ with probabilities $\{q_0, \dots, q_N\}$ we have that $\mu(A) = \sum_{i=0}^N \delta_i(A)$, and

  $$
  f_Z(z) = (1-p)\mathbf{1}_{\{z=0\}} + p\left[\sum_{i=0}^N q_i \mathbf{1}_{\{z=i\}}\right],
  $$

  hence $\mathbb{P}_Z(Z = 0) = 1 - p + pq_0$ and $\mathbb{P}(Z = i) = pq_i$ for $i \in \{1, \dots, N\}$.

C.1.3 The continuous case: proof of theorem 1

We now consider the continuous case. From the second expression in the theorem theorem 4 we derive the result of theorem 1.

Proof of theorem 1. We continue from the result of theorem 4. Note that $\mathrm{Alt}(\nu) = \{\nu' = (G', \{\nu'_u\}_u) \mid \exists v_0 \neq a^\star : \mu'_{v_0} > \mu'_{a^\star}\}$, where $a^\star = a^\star(\mu)$. Hence, letting $\mathrm{Alt}_v(\nu) = \{\nu' \mid \mu'_v > \mu'_{a^\star}\}$, we have $\mathrm{Alt}(\nu) = \cup_{v\neq a^\star}\mathrm{Alt}_v(\nu)$.

Therefore, due to the properties of the KL divergence we obtain

$$\begin{aligned}
&\inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{u\in V}\sum_{v\in N_{in}(u)} \omega_v \mathrm{KL}(\nu_{v,u},\nu'_{v,u}) \\
&\quad = \min_{u\neq a^\star} \inf_{\nu'\in\mathrm{Alt}_u(\nu):\mu'_v>\mu'_{a^\star}} \sum_{u\in V}\sum_{v\in N_{in}(u)} \omega_v \mathrm{KL}(\nu_{v,u},\nu'_{v,u}).
\end{aligned}$$

Following the discussion in appendix C.1.2, we can write

$$\begin{aligned}
\mathrm{KL}(\nu_{v,u},\nu'_{v,u}) &= \mathbb{E}_{Z\sim\nu_{v,u}}\left[\ln\frac{\mathrm{d}\mathbb{P}_{\nu_{v,u}}(Z)/\mathrm{d}\mu(Z)}{\mathrm{d}\mathbb{P}'_{\nu_{v,u}}(Z)/\mathrm{d}\mu(Z)}\right], \\
&= \mathbb{E}_{Z\sim\nu_{v,u}}\left[\ln\frac{f_{v,u}(Z)}{f'_{v,u}(Z)}\right], \\
&= (1-G_{v,u})\ln\frac{1-G_{v,u}}{1-G'_{v,u}} + G_{v,u}\mathbb{E}_{Z\sim\nu_u}\left[\ln\frac{G_{v,u}f_u(Z)}{G'_{v,u}f'_u(Z)}\right], \\
&= \mathrm{kl}(G_{v,u},G'_{v,u}) + G_{v,u}\mathbb{E}_{Z\sim\nu_u}\left[\ln\frac{f_u(Z)}{f'_u(Z)}\right], \\
&= \mathrm{kl}(G_{v,u},G'_{v,u}) + G_{v,u}\mathrm{KL}(\nu_u,\nu'_u).
\end{aligned}$$

Therefore, noting that the constraint involves only the pair $(\nu'_u,\nu'_{a^\star})$ through their parameters $(\mu'_u,\mu'_{a^\star})$, we conclude that

$$\begin{aligned}
&\inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{u\in V}\sum_{v\in N_{in}(u)} \omega_v \mathrm{KL}(\nu_{v,u},\nu'_{v,u}) \\
&\quad = \min_{u\neq a^\star} \inf_{\nu'\in\mathrm{Alt}_u(\nu):\mu'_v>\mu'_{a^\star}} \sum_{u\in V}\sum_{v\in N_{in}(u)} \omega_v \left(\mathrm{kl}(G_{v,u},G'_{v,u}) + G_{v,u}\mathrm{KL}(\nu_u,\nu'_u).\right), \\
&\quad = \min_{u\neq a^\star} \inf_{\nu':\mu'_u\geq\mu'_{a^\star}} \sum_{v\in N_{in}(u)} \omega_v G_{v,u}\mathrm{KL}(\nu_u,\nu'_u) + \sum_{w\in N_{in}(a^\star)} \omega_w G_{w,a^\star}\mathrm{KL}(\nu_{a^\star},\nu'_{a^\star}), \\
&\quad = \min_{u\neq a^\star} \inf_{\nu':\mu'_u\geq\mu'_{a^\star}} m_u\mathrm{KL}(\nu_u,\nu'_u) + m_{a^\star}\mathrm{KL}(\nu_{a^\star},\nu'_{a^\star}),
\end{aligned}$$

where $m_u \coloneqq \sum_{v\in N_{in}(u)} \omega_v G_{v,u}$ and $m_{a^\star} = \sum_{\omega\in N_{in}(a^\star)} \omega_w G_{w,a^\star}$.

Therefore, by optimizing over ν' as in (Garivier and Kaufmann, 2016, Lemma 3) we obtain

$$(T^\star(\nu))^{-1} = \sup_{\omega\in\Delta(V)} \min_{u\neq a^\star} (m_u+m_{a^\star}) I_{\frac{m_{a^\star}}{m_u+m_{a^\star}}}(\nu_{a^\star},\nu_u) \text{ s.t. } m_u = \sum_{v\in N_{in}(u)} \omega_v G_{v,u}.$$

□

C.1.4 The discrete case: proof of proposition 4

We find that if $(\nu_u)_{u\in V}$ are Bernoulli distributed, or more generally, then we obtain that it is not possible, in general, to estimate the best vertex. The reason is simple: without knowing which edge was activated, the learner does not know how to discern between the randomness of the reward and the randomness of the edge.

Proof of proposition 4. Let $n\in\mathbb{R}_+^K$ and consider the optimization problem

$$\inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{u\in V}\sum_{v\in N_{in}(u)} n_v \mathrm{KL}(\nu_{v,u},\nu'_{v,u}).$$

If $\nu_{v,u}$ is a Bernoulli distribution of parameter $q_{v,u} \coloneqq G_{v,u}\mu_u$ (sim. for $\nu'_{v,u}$ with $q'_{v,u} \coloneqq G'_{v,u}\mu'_u$), then we can write

$$\inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{u\in V}\sum_{v\in N_{in}(u)} n_v \mathrm{KL}(\nu_{v,u},\nu'_{v,u}) = \inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{u\in V}\sum_{v\in N_{in}(u)} n_v \mathrm{kl}(q_{v,u},q'_{v,u}).$$

Using the fact that $\nu \in \mathrm{Alt}(\nu)$ is observable, it must imply that there exists $(v,u) \in V^2$ such that $\mu'_u > \mu'_{a^\star}$ and $G'_{v,u} > 0$. Similarly, $\exists w \in V$ s.t. $G_{w,a^\star} > 0$. Hence, by absolute continuity, we have $G'_{w,a^\star} > 0$, otherwise the event $\mathcal{E} = \{Z_{w,a^\star} = 1\}$ would satisfy $\mathbb{P}_\nu(\mathcal{E}) > 0$ and $\mathbb{P}_{\nu'}(\mathcal{E}) = 0$.

Hence, similarly as before, we have that

$$\begin{aligned}
&\inf_{\nu' \in \mathrm{Alt}(\nu)} \sum_{u \in V} \sum_{v \in N_{in}(u)} n_v \mathrm{kl}(q_{v,u}, q'_{v,u}), \\
&\quad = \min_{u \neq a^\star} \min_{v \in N_{in}(u), w \in N_{in}(a^\star)} \inf_{\mu', G' : \mu'_u \geq \mu'_{a^\star}, G'_{v,u}, G'_{w,a^\star} \geq 0} n_v \mathrm{kl}(q_{v,u}, q'_{v,u}) + n_w \mathrm{kl}(q_{w,a^\star}, q'_{w,a^\star}).
\end{aligned}$$

Therefore, for $u \neq a^\star$, $v \in N_{in}(u), w \in N_{in}(a^\star)$, we are interested in the following non-convex problem

$$\begin{aligned}
\min_{\mu', G' > 0} \quad & n_v \mathrm{kl}(q_{v,u}, q'_{v,u}) + n_w \mathrm{kl}(q_{w,a^\star}, q'_{w,a^\star}) \\
\text{s.t.} \quad & q'_{u,v} = G'_{u,v} \mu'_v \qquad \forall (u,v) \in V^2, \\
& \mu'_u \geq \mu'_{a^\star}.
\end{aligned}$$

However, the solution, as one may expect, is 0. Simply note one can take $q'_{v,u} = q_{v,u}$ by choosing $\mu'_u = \mu_{a^\star}$ and $G'_{u,v} = q_{v,u}/\mu_{a^\star}$ (which is ≤ 1). Similarly, we have $q'_{w,a^\star} = q_{w,a^\star}$ by choosing $\mu'_{a^\star} = \mu_{a^\star}$ and $G'_{w,a^\star} = q_{w,a^\star}/\mu_{a^\star} = G_{w,a^\star}$. Henceforth, in the Bernoulli case we have that $(T^\star(\nu))^{-1} = 0$.

□

C.2 The informed setting

We now give a proof of theorem 2, which follows closely the one in the uninformed case.

Proof of theorem 2. First note that knowing the set of edges that was activated is equivalent to knowing the value of $(Y_{t,(V_t,u)})_u$. Then, denote the density associated to $Y_{t,(V_t,u)}$ by $g_{v,u}$ under ν (sim. $g'_{v,u}$ for ν'). We can write the log-likelihood ratio as

$$\begin{aligned}
L_t &= \sum_{n=1}^{t} \sum_{u \in V} \sum_{v \in N_{in}(u)} \mathbf{1}_{\{V_n = v\}} \ln\left(\frac{\mathrm{d}\mathbb{P}_\nu(Z_{n,u}, Y_{n,(v,u)} V_n)}{\mathrm{d}\mathbb{P}_{\nu'}(Z_{n,u}, Y_{n,(v,u)}, V_n)} \right), \\
&= \sum_{n=1}^{t} \sum_{u \in V} \sum_{v \in N_{in}(u)} \mathbf{1}_{\{V_n = v\}} \Bigg[\mathbf{1}_{\{Y_{n,(v,u)} = 1\}} \ln\left(\frac{f_u(Z_{n,u}) g_{v,u}(1)}{f'_u(Z_{n,u}) g'_{v,u}(1)} \right) \\
&\qquad\qquad + \mathbf{1}_{\{Y_{n,(v,u)} = 0\}} \ln\left(\frac{g_{v,u}(0)}{g'_{v,u}(0)} \right) \Bigg], \\
&= \sum_{u \in V} \sum_{v \in N_{in}(u)} \sum_{n=1}^{N_v(t)} \left[\mathbf{1}_{\{Y_{n,(v,u)} = 1\}} \ln\left(\frac{f_u(W_{n,u}) g_{v,u}(1)}{f'_u(W_{n,u}) g'_{v,u}(1)} \right) + \mathbf{1}_{\{Y_{n,(v,u)} = 0\}} \ln\left(\frac{g_{v,u}(0)}{g'_{v,u}(0)} \right) \right].
\end{aligned}$$

where $(W_{n,u})_n$ is a sequence of i.i.d. random variables distributed according to ν_u. Hence

$$\begin{aligned}
\mathbb{E}_\nu[L_\tau] = \sum_{u \in V} \sum_{v \in N_{in}(u)} \mathbb{E}_\nu \Bigg[\sum_{n=1}^{\infty} \mathbf{1}_{\{N_v(\tau) \geq n\}} \Bigg[& \mathbf{1}_{\{Y_{n,(v,u)} = 1\}} \ln\left(\frac{f_u(W_{n,u})}{f'_u(W_{n,u})} \right) \\
& \mathbf{1}_{\{Y_{n,(v,u)} = 1\}} \ln\left(\frac{g_{v,u}(1)}{g'_{v,u}(1)} \right) + \mathbf{1}_{\{Y_{n,(v,u)} = 0\}} \ln\left(\frac{g_{v,u}(0)}{g'_{v,u}(0)} \right) \Bigg] \Bigg].
\end{aligned}$$

Note that $\{N_v(\tau) \geq n\} = \{N_v(\tau) \leq n-1\} \in \mathcal{F}_{n-1}$, therefore $W_{n,u}$ and $Y_{n,(v,u)}$ are independent of that event. Hence

$$\begin{aligned}
\mathbb{E}_\nu[L_\tau] &= \sum_{u \in V} \sum_{v \in N_{in}(u)} \sum_{n=1}^{\infty} \mathbb{P}_\nu(N_v(\tau) \geq n) \left[G_{v,u} \mathrm{KL}(\nu_u, \nu'_u) + \mathrm{kl}(G_{v,u}, G'_{v,u}) \right], \\
&= \sum_{u \in V} \sum_{v \in N_{in}(u)} \mathbb{E}_\nu[N_v(\tau)] \left[G_{v,u} \mathrm{KL}(\nu_u, \nu'_u) + \mathrm{kl}(G_{v,u}, G'_{v,u}) \right].
\end{aligned}$$

Therefore, applying (Kaufmann et al., 2016, Lemma 1) at $t = \tau$, $\mathbb{E}_\nu[L_\tau] \geq \mathrm{kl}(\delta, 1-\delta)$.

Consider the set of confusing models $\mathrm{Alt}(\nu) = \{\nu' = (\mu', G') : a^\star(\mu) \neq a^\star(\mu'), \nu \ll \nu'\}$, and define the selection rate of a vertex v as $\omega_v = \mathbb{E}_\nu[N_v(\tau)]/\mathbb{E}_\nu[\tau]$. Then, by minimizing over the set of confusing models, and then optimizing over $\omega = (\omega_v)_{v\in V}$ over the simplex $\Delta(V)$, we obtain

$$\mathbb{E}_\nu[\tau] \underbrace{\sup_{w\in\Delta(V)} \inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{u\in V} \sum_{v\in N_{in}(u)} \omega_v \left[G_{v,u}\mathrm{KL}(\nu_u, \nu'_u) + \mathrm{kl}(G_{v,u}, G'_{v,u})\right]}_{=:(T^\star(\nu))^{-1}} \geq \mathrm{kl}(\delta, 1-\delta).$$

Hence, consider now the expression $\inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{u\in V} \sum_{v\in N_{in}(u)} w_v \left[G_{v,u}\mathrm{KL}(\nu_u, \nu'_u) + \mathrm{kl}(G_{v,u}, G'_{v,u})\right]$ and observe that it simplifies as in the proof of theorem 1:

$$\begin{aligned}
&\inf_{\nu'\in\mathrm{Alt}(\nu)} \sum_{u\in V} \sum_{v\in N_{in}(u)} \omega_v \left[G_{v,u}\mathrm{KL}(\nu_u, \nu'_u) + \mathrm{kl}(G_{v,u}, G'_{v,u})\right] \\
&\quad= \min_{u\neq a^\star} \inf_{G',\nu'_u,\nu'_{a^\star}:\mu'_u\geq\mu'_{a^\star}} \sum_{v\in N_{in}(u)} \omega_v G_{v,u}\mathrm{KL}(\nu_u, \nu'_u) + \sum_{v\in N_{in}(a^\star)} \omega_v G_{v,a^\star}\mathrm{KL}(\nu_{a^\star}, \nu'_{a^\star}), \\
&\quad= \min_{u\neq a^\star} \inf_{G',\nu'_u,\nu'_{a^\star}:\mu'_u\geq\mu'_{a^\star}} m_u\mathrm{KL}(\nu_u, \nu'_u) + m_{a^\star}\mathrm{KL}(\nu_{a^\star}, \nu'_{a^\star}),
\end{aligned}$$

as in the proof of theorem 1. For the known graph case, note that the set of confusing models becomes $\mathrm{Alt}'(\nu) \coloneqq \{\nu' = (\mu', G) : a^\star(\mu) \neq a^\star(\mu'), \nu_u \ll \nu'_u\ \forall u \in V\}$, from which one can conclude the same result. □

C.3 Scaling properties

To gain a better intuition of the characteristic time in theorem 1, we can focus on the Gaussian case where $\nu_u = \mathcal{N}(\mu_u, \sigma^2)$. For this case we have that $\mathrm{KL}(\nu_u, \nu_v) = (\mu_u - \mu_v)^2/(2\sigma^2)$, and $I_\alpha(\nu_u, \nu_v) = \dfrac{\alpha(1-\alpha)(\mu_u - \mu_v)^2}{2\sigma^2}$. Therefore $T^\star(\nu)$ can be computed by solving the following convex problem

$$\begin{aligned}
T^\star(\nu) = &\inf_{m\in\mathbb{R}_+^K,\omega\in\Delta(V)} \max_{u\neq a^\star} \left(m_u^{-1} + m_{a^\star}^{-1}\right)\frac{2\sigma^2}{\Delta_u^2} \\
&\text{s.t. } m_u = \sum_{v\in N_{in}(u)} \omega_v G_{v,u} \quad \forall u \in V.
\end{aligned} \tag{13}$$

C.3.1 General case

We restate here proposition 1 and provide a proof.

Proposition 7. *For an observable model $\nu = (\{\nu_u\}_u, G)$ with a graph G satisfying $\delta(G)+\sigma(G) > 0$ and Gaussian random rewards $\nu_u = \mathcal{N}(\mu_u, \sigma^2)$ we can upper bound $T^\star$ as*

$$T^\star(\nu) \leq \frac{4\left[\delta(G) + \sigma(G) - \left\lfloor\frac{\sigma(G)}{\alpha(G)+1}\right\rfloor\right]\sigma^2}{\min_{u\neq a^\star}\min(\bar{G}_u, \bar{G}_{a^\star})\Delta_u^2},$$

where $\bar{G}_u \coloneqq \max\left(\max_{v\in D(G)} G_{v,u}, \min_{v\in L(G):G_{v,u}>0} G_{v,u}\right)$ (sim. $\bar{G}_{a^\star}$).

Proof. By the condition $\sigma(G) + \delta(G) > 0$ the graph is either weakly observable, or has strongly observable nodes with self-loops, or both.

Using the expression of $T^\star(\nu)$ in eq. (13) note that for each weakly observable vertex u there exists $v \in N_{in}(u) \cap D(G)$, where $D(G)$ is the smallest set dominating the set of weakly observable vertices (see definition 3).

For any strongly observable vertex by lemma 7 we need at-most $\sigma(G) - \left\lfloor\frac{\sigma(G)}{\alpha(G)+1}\right\rfloor$ vertices in $L(G)$ to dominate $|SO(G)|$ if $\sigma(G) > 0$.

If $\sigma(G) = 0$, then we have $\delta(G) > 0$: hence any vertex $u \in D(G)$ also dominates any $v \in SO(G)$. And thus $D(G)$ dominates the entire graph.

Therefore we need at-most $\kappa = \delta(G) + \sigma(G) - \left\lfloor \frac{\sigma(G)}{\alpha(G)+1} \right\rfloor$ vertices to dominate the graph. Using this information, we allocate probability mass uniformly across these vertices.

1. For any vertex $u \in W(G)$ let $\omega_v = 1/\kappa$ for all $v \in D(G)$. This fact allows us to lower bound m_u as

$$\begin{aligned} m_u &= \sum_{v \in N_{in}(u) \setminus D(G)} \omega_v G_{v,u} + \sum_{w \in D(G) \cap N_{in}(u)} \omega_w G_{w,u}, \\ &\geq \sum_{w \in D(G) \cap N_{in}(u)} \omega_w G_{w,u}, \\ &\geq \sum_{w \in D(G) \cap N_{in}(u)} \frac{1}{\kappa} G_{w,u}, \\ &\geq \max_{w \in D(G) \cap N_{in}(u)} \frac{1}{\kappa} G_{w,u}, \\ &\geq \max_{w \in D(G)} \frac{1}{\kappa} G_{w,u}. \end{aligned}$$

2. For any vertex $u \in L(G)$ with a self-loop we can lower bound m_u as

$$m_u \geq \min_{v \in L(G): G_{v,u} > 0} \frac{1}{\kappa} G_{v,u}.$$

Therefore, for any $u \in V$ we have

$$\frac{1}{\kappa m_u} \leq \begin{cases} \dfrac{1}{\min_{v \in L(G): G_{v,u} > 0} G_{v,u}} & \text{if } u \in L(G), \\ \dfrac{1}{\max_{v \in D(G)} G_{v,u}} & \text{otherwise,} \end{cases}$$

Then, let $\bar{G}_u \coloneqq \max\left(\max_{v \in D(G)} G_{v,u}, \min_{v \in L(G): G_{v,u} > 0} G_{v,u}\right)$. We conclude that

$$\begin{aligned} T^\star(\nu) &\leq \max_{u \neq a^\star} \kappa \left(\bar{G}_u^{-1} + \bar{G}_{a^\star}^{-1}\right) \frac{2\sigma^2}{\Delta_u^2}, \\ &\leq \frac{4\kappa\sigma^2}{\min_{u \neq a^\star} \min(\bar{G}_u, \bar{G}_{a^\star}) \Delta_u^2}, \end{aligned}$$

where in the last expression we used $a + b \leq 2\max(a, b)$. □

C.3.2 The loopless clique

We now consider the scaling for the case case $\delta(G) + \sigma(G) = 0$, which corresponds to the loopless clique. We restate here proposition 2 and provide a proof.

Proposition 8. *For an observable model $\nu = (\{\nu_u\}_u, G)$ with $\delta(G) + \sigma(G) = 0$, and Gaussian random rewards $\nu_u = \mathcal{N}(\mu_u, \sigma^2)$, we can upper bound $T^\star$ as*

$$T^\star(\nu) \leq \frac{4\sigma^2}{\bar{G}\Delta_{\min}^2},$$

where $\bar{G} \coloneqq \min_{(v,w) \in V^2 : v \neq w} \max_{u \neq a^\star} \frac{1}{G_{v,w}(u)} + \frac{1}{G_{v,w}(a^\star)}$ and $G_{v,w}(u) \coloneqq G_{v,u} + G_{w,u}$.

Proof. First, note that since $\delta(G) = 0$ there are no weakly observable vertices. Since also $\sigma(G) = 0$, we must have that $V = E(G)$, where $E(G)$ is the set of strongly observable vertices without self-loops. Hence, by definition, it is the loopless clique. By lemma 7 we need 2 vertices to dominate the graph. Define the following set

$$\mathcal{A} \coloneqq \operatorname*{arg\,min}_{(v,w) \in V^2 : v \neq w} \max_{u \neq a^\star} G_{v,w}(u)^{-1} + G_{v,w}(a^\star)^{-1}, \text{ where } G_{v,w}(u) \coloneqq G_{v,u} + G_{w,u},$$

and denote by $\bar{G} \coloneqq \min_{(v,w)\in V^2: v\neq w} \max_{u\neq a^\star} G_{v_0,w_0}(u)^{-1} + G_{v_0,w_0}(a^\star)^{-1}$ the optimal value for any $(v_0, w_0) \in \mathcal{A}$. Then, let $\omega_{v_0} = \omega_{w_0} = 1/2$ for a generic pair $(v_0, w_0) \in \mathcal{A}$. For any $u \in V$ we have m_u

$$\begin{aligned} m_u &= \sum_{v\in V} w_v G_{v,u} \\ &\geq (G_{v_0,u} + G_{w_0,u})/2. \end{aligned}$$

Note that by Lemma 7, $m_u > 0$ for any $u \in V$. Therefore,

$$m_u^{-1} \leq 2G_{v_0,w_0}(u)^{-1}.$$

Hence, we conclude that

$$\begin{aligned} T^*(\nu) &\leq \max_{u\neq a^*}(m_u^{-1} + m_{a^*}^{-1})\frac{2\sigma^2}{\Delta_u^2} \\ &\leq \max_{u\neq a^*}\left(G_{v_0,w_0}(u)^{-1} + G_{v_0,w_0}(a^\star)^{-1}\right)\frac{4\sigma^2}{\Delta_{\min}^2} \\ &= \frac{4\sigma^2}{\bar{G}\Delta_{\min}^2}. \end{aligned}$$

□

C.3.3 Heuristic solution: scaling and spectral properties

In general we find it hard to characterize the scaling of $T^\star(\nu)$ in terms of the spectral properties of G without a more adequate analysis. Furthermore, we also wonder if it is possible to find a closed-form solution that can be easily used.

Intuition suggests that, by exploiting the underlying topology of the graph, a good solution ω should be sparse (which, in turns, helps minimizing the sample complexity). However, it may be hard to find a simple sparse solution.

To that aim, we can gain some intuition from the BAI problem for the classical multi-armed bandit setup. From the analysis in Garivier and Kaufmann (2016, Appendix A.4), an approximately optimal solution in the Gaussian case is given by $\omega_u^\star \propto 1/\Delta_u^2$, with $\Delta_{a^\star} = \Delta_{\min}$.

Hence, we propose also that $m_u \propto 1/\Delta_u^2$, with $m = G^\top\omega$. At this point we could try to minimize the MSE loss between m and the vector $\Delta^{-2} \coloneqq (1/\Delta_u^2)_{u\in V}$, subject to $\|w\|_1 = 1$. However, in this problem we are more interested in the directional alignment between m and Δ^{-2}, rather than in their magnitude, since the magnitude of ω is constrained [3]. Therefore, one may be interested in maximizing teh similarity $m^\top\Delta^{-2}$, or rather

$$\max_{\omega:\|\omega\|_2\leq\alpha} \omega^\top G\Delta^{-2}$$

for some constraint α. Obviously the solution is $\omega = \alpha G\Delta^{-2}/\|G\Delta^{-2}\|_2$ in the classical Euclidean space with the ℓ_2 norm. However, such solution is not a distribution. To that aim, we project $G\Delta^{-2}$ on the closest distribution in the KL sense, defined as $\mathrm{Proj}_{\mathrm{KL}}(x) \coloneqq \min_{p\in\mathcal{P}} \mathrm{KL}(p, x)$ for any x such that $x_u \geq 0$ for every u.

Lemma 8. *The projection of $G\Delta^{-2}$ in the KL sense is given by $\omega_{\text{heur}} = G\Delta^{-2}/\|G\Delta^{-2}\|_1$.*

Proof. For some vector x satisfying $x \geq 0$, write the Lagrangian of $\min_{p\in\mathcal{P}} \mathrm{KL}(p, x)$:

$$\mathcal{L}(p, \lambda) = \sum_u p_u \ln\left(\frac{p_u}{x_u}\right) + \lambda\left(1 - \sum_u p_u\right).$$

Then, we check the first order condition $\partial\mathcal{L}/\partial p_u = 0 \Rightarrow \ln\left(\frac{p_u}{x_u}\right) + 1 - \lambda = 0$, implying that $p_u = x_u e^{\lambda-1}$. Using the fact that $\sum_u p_u = 1$ we also obtain $e^{1-\lambda} = \sum_u x_u = \|x\|_1$. Therefore $\lambda = 1 - \ln(\|x\|_1)$, from which we conclude that $p_u = x_u/\|x\|_1$. □

[3] Note that also the MSE problem $\arg\min_\beta \|y - A\beta\|_2^2 = \arg\min_\beta -2y^\top A\beta + \|A\beta\|_2^2$ also tries to solve a problem of alignment through the term $2y^\top A\beta$, while the second term $\|A\beta\|_2^2$ can be considered a form of regularization.

Such allocation $w_{\text{heur}} \coloneqq \frac{G\Delta^{-2}}{\|G\Delta^{-2}\|_1}$ makes intuitively sense: a vertex u will be chosen with probability proportional to $\sum_{v\in V} G_{uv}\Delta_v^{-2}$, thus assigning higher preference to vertices that permit the learner to sample arms with small sub-optimality gaps.

For this heuristic allocation w_{heur} we can provide the following upper bound on its scaling.

Lemma 9. *For a model $\nu = (\{\nu_u\}_u, G)$ with an observable graph G and Gaussian random rewards $\nu_u = \mathcal{N}(\mu_u, \sigma^2)$ we can upper bound $T(\omega_{\text{heur}}; \nu)$ as*

$$T^\star(\nu) \le T(\omega_{\text{heur}}; \nu) \le \frac{\|G\Delta^{-2}\|_1}{\sigma_{\min}(G)^2} \cdot 4\sigma^2 \le \frac{K \min\left(\sqrt{K}\sigma_{\max}(G), \sum_i \sigma_i(G)\right)}{\Delta_{\min}^2 \sigma_{\min}(G)^2} \cdot 4\sigma^2.$$

where $\sigma_{\min}(G), \sigma_{\max}(G), \sigma_i(G)$ are, respectively the minimum singular value of G, the maximum singular value of G and the i-th singular value of G.

Proof. Again, we prove this corollary by lower bound each m_u. Note that we have $m = G^\top G\omega_{\text{heur}}$. Denote by G_u the u-th row of G, then

$$\begin{aligned} m_u &= \frac{\sum_{v\in V} (G^T G)_{u,v}\Delta_v^{-2}}{\|G\Delta^{-2}\|_1} \\ &\ge \frac{\|G_u\|_2^2 \Delta_u^{-2}}{\|G\Delta^{-2}\|_1} \\ &\ge \frac{\|G_u\|_2^2 \Delta_u^{-2}}{\|G\Delta^{-2}\|_1}. \end{aligned}$$

And therefore, using that $\|G_u\|_2^2 \ge \sigma_{\min}(G)^2$, we observe that

$$\begin{aligned} T(\omega_{\text{heur}}; \nu) &\le \max_{u\neq a^\star} \left(\frac{\Delta_u^2}{\|G_u\|_2^2} + \frac{\Delta_{\min}^2}{\|G_{a^\star}\|_2^2}\right) \frac{2\|G\Delta^{-2}\|_1 \sigma^2}{\Delta_u^2}, \\ &\le \max_{u\neq a^\star} \left(\Delta_u^2 + \Delta_{\min}^2\right) \frac{2\|G\Delta^{-2}\|_1 \sigma^2}{\sigma_{\min}(G)^2 \Delta_u^2}, \\ &\le \frac{\|G\Delta^{-2}\|_1}{\sigma_{\min}(G)^2} \cdot 4\sigma^2. \end{aligned}$$

Lastly, denoting by G_u the u-th row of G, we obtain that $\|G\Delta^{-2}\|_1 \le \sum_u |G_u^\top \Delta^{-2}| \le \|\Delta^{-2}\|_2 \sum_u \|G_u\|_2 \le \frac{\sqrt{K}}{\Delta_{\min}^2} \sum_u \|G_u\|_2$. We conclude by noting that $\sum_u \|G_u\|_2 \le K\sigma_{\max}(G)$, and thus $\|G\Delta^{-2}\|_1 \le \frac{K^{3/2}\sigma_{\max}(G)}{\Delta_{\min}^2}$

Additionally, we also that $\|G\Delta^{-2}\|_1 \le \sum_u |G_u^\top \Delta^{-2}| \le \|\Delta^{-2}\|_\infty \sum_u \|G_u\|_1$ by Holder's inequality. Now, let $\|\cdot\|_*$ denote the Schatten-1 norm. Using that $\|\text{vec}(G)\|_1 \le K\|G\|_* = K\sum_i \sigma_i(G)$ we have $\|G\Delta^{-2}\|_1 \le \frac{K\sum_i \sigma_i(G)}{\Delta_{\min}^2}$. □

An alternative approach that is sparse. We note that the above analysis does not take fully advantage of the graph structure. An alternative approach that yields a better scaling is to instead consider the similarity problem

$$\max_{\omega:\|\omega\|_1=1} \omega^\top G\Delta^{-2}.$$

The optimal solution then is simply $\omega_u = \mathbf{1}_{\{u\in\mathcal{G}\}}/|\mathcal{G}|$, where $\mathcal{G} = \arg\max_v (G\Delta^{-2})_v$. This is an efficient allocation, since it scales as $O\left(\frac{|\mathcal{G}|}{\Delta_{\min}^2 \max_{u\in\mathcal{G}} \min_v G_{u,v}}\right)$. However, such solution is admissible only if it guarantees that $V \ll \mathcal{G}$, i.e., this set of vertices $\mathcal{G}$ dominates the graph.

Alternatively, one can choose the top k vertices $\mathcal{U} = \{u_1, u_2, \dots, u_k\}$ ordered according to $(G\Delta^{-2})_{u_1} \ge \dots \ge (G\Delta^{-2})_{u_k} \ge \dots (G\Delta^{-2})_{u_K}$ satisfying $V \ll \{u_1, \dots, u_k\}$ (i.e., these vertices dominate the graph). Then, one can simply let $\omega_u = \mathbf{1}_{\{u\in\mathcal{U}\}}/|\mathcal{U}|$. Since these are also the vertices that maximize the average information collected from the graph, we believe this solution to be sample efficient. A simple analysis, shows that the worst case scaling in this scenario is $O\left(\frac{|\mathcal{U}|}{\Delta_{\min}^2 \max_{u\in\mathcal{U}} \min_v G_{u,v}}\right)$.

D Analysis of TaS-FG

In this section we provide an analysis of TaS-FG for an observable model in the uninformed case (with continuous rewards) or in the informed case. In appendix D.1 we analyse the sampling rule. In appendix D.2 we analyse the stopping rule. Lastly, in appendix D.3, we analyse the sample complexity.

D.1 Sampling Rule

The proof of the tracking proposition proposition 5 is inspired by D-tracking Garivier and Kaufmann (2016); Jedra and Proutiere (2020). In Degenne and Koolen (2019) they show that classical D-tracking Garivier and Kaufmann (2016) may fail to converge when $C^\star(\nu)$ is a convex set of possible optimal allocations. However, using a modified version it is possible to prove the convergence. We take inspiration from Jedra and Proutiere (2020), where they applied a modified D-tracking to the linear bandit case and showed convergence of $w^\star(t)$.

The intuition behind the proof is that tracking the average of the converging sequence $(\omega^\star(t))_t$, which is a convex combination, converges to a stable point in the convex set $C^\star(\nu)$. The proof makes use of the following result, from Berge (1963).

Theorem 5 (Maximum theorem Berge (1963))**.** *Let $C^\star(\nu) = \arg\inf_{\omega\in\Delta(V)} T(\omega;\nu)$. Then $T^\star(\nu) \coloneqq T(\omega^\star;\nu), \omega^\star \in C^\star(\nu)$, is continuous at ν (in the sense of $(G, \{\mu_u\}_u) \in [0,1]^{K\times K} \times [0,1]^K$), $C^\star(\nu)$ is convex, compact and non-empty. Furthermore, we have that for any open neighborhood $\mathcal{V}$ of $C^\star(\nu)$, there exists an open neighborhood $\mathcal{U}$ of ν, such that for all $\nu' \in \mathcal{U}$ we have $C^\star(\nu') \subseteq \mathcal{V}$.*

Here we state, and prove, a more general version, of proposition 5 (which follows by taking $\alpha_{t,n} = 1/t$ in the next proposition).

Proposition 9. *Let $S_t = \{u \in V : N_u(t) < \sqrt{t} - K/2\}$. The D-tracking rule, defined as*

$$V_t \in \begin{cases} \arg\min_{u\in S_t} N_u(t) & S_t \neq \emptyset \\ \arg\min_{u\in V} N_u(t) - t\sum_{n=1}^t \alpha_{t,n}\omega_u^\star(n) & \textit{otherwise} \end{cases}, \tag{14}$$

where, for every $t \geq 1$, the sequence $\alpha_t = (\alpha_{t,n})_{n=1}^t$ satisfies: (1) $\alpha_{t,n} \in [0,1]$; (2) for every fixed $n \in \{1,\dots,t\}$ we have $\alpha_{t,n} = o(1)$ in t ; (3) for all t, $\sum_{n=1}^t \alpha_{t,n} = 1$.

Such tracking rule ensures that for all $\epsilon > 0$ there exists $t(\epsilon)$ such that for all $t \geq t(\epsilon)$ we have

$$\|N(t)/t - \bar{v}(t)\|_\infty \leq 5(K-1)\epsilon,$$

where $\bar{v}(t) \coloneqq \arg\inf_{\omega\in C^\star(\nu)} \|\omega - \sum_{n=1}^t \alpha_n\omega^\star(n)\|_\infty$, and $\lim_{t\to\infty} \inf_{\omega\in C^\star(\nu)} \|N(t)/t - \omega\|_\infty \to 0$ almost surely.

Proof. Define the projection of $x \in \mathbb{R}^n$ onto C as $\mathrm{Proj}_C(x) \coloneqq \arg\inf_{\omega\in C} \|x - w\|_\infty$, which is guaranteed to exists if C is convex and compact.

Let $n(t) \coloneqq \mathrm{Proj}_{C^\star(\nu)}(N_t/t)$. The proof lies showing that $\|N_t/t - n(t)\|_\infty \to 0$ as $t \to \infty$. To that aim, we first need to show that $\inf_{w\in C^\star(\nu)} \|\bar{\omega}^\star(t) - w\|_\infty \to 0$, where $\bar{\omega}^\star(t) \coloneqq \sum_{n=1}^t \alpha_n\omega^\star(n)$ is a convex combination of the estimated optimal allocations up to time t.

Begin by defining the following quantities

$$\bar{v}(t) \coloneqq \mathrm{Proj}_{C^\star(\nu)}(\bar{\omega}^\star(t)) \text{ and } v(t) \coloneqq \mathrm{Proj}_{C^\star(\nu)}(\omega^\star(t)),$$

which are, respectively, the projection onto $C^\star(\nu)$ of the average estimated allocation and the projection of the last estimated allocation.

By the forced exploration step, we have that $N_u(t) \to \infty$ for every $u \in V$. Since the model is observable, we can invoke the law of large number and guarantee that $\mathbb{P}_\nu(\lim_{t\to\infty} \hat{\nu}(t) = \nu) = 1$ in the sense that $(\hat{G}(t), \hat{\mu}(t)) \to (G, \mu)$ almost surely. Then, by continuity of the problem (see theorem 5) we have that $\forall\epsilon > 0 \exists t_0(\epsilon) : \sup_{\omega\in C^\star(\hat{\nu}(t))} \|\omega - \mathrm{Proj}_{C^\star(\nu)}(\omega)\|_\infty \leq \epsilon$ for all $t \geq t_0(\epsilon)$.

Henceforth, for $t \geq t_0(\epsilon)$ we have $\|\omega^\star(t) - v(t)\|_\infty \leq \sup_{w\in C^\star(\hat{\nu}(t))} \|\omega - \mathrm{Proj}_{C^\star(\nu)}(\omega)\|_\infty \leq \epsilon$, thus we derive

$$\begin{aligned}\left\|\sum_{n=1}^{t} \alpha_{t,n} v(n) - \bar{\omega}^\star(n)\right\|_\infty &\leq \sum_{n=1}^{t_0(\epsilon)} \alpha_{t,n}\|v(n) - \omega^\star(n)\|_\infty + \sum_{n=t_0(\epsilon)+1}^{t} \alpha_{t,n}\|v(n) - \omega^\star(n)\|_\infty,\\ &\leq t_0(\epsilon)\bar{\alpha}_{t,t_0(\epsilon)} + \epsilon,\end{aligned}$$

where $\bar{\alpha}_{t,t_0(\epsilon)} = \max_{1\leq n\leq t_0(\epsilon)} \alpha_{t,n}$ and we used the fact that $\sum_n \alpha_{t,n} = 1$. Hence, for any $x \in C^\star(\nu)$ note that $\|\bar{\omega}^\star(t) - \bar{v}(t)\|_\infty \leq \|\bar{\omega}^\star(t) - x\|_\infty$. For a fixed t, one can choose $x = \sum_{n=1}^t \alpha_{t,n} v(n)$ since every $v(n) \in C^\star(\nu)$, and also a convex combination belongs to $C^\star(\nu)$ by convexity. Henceforth

$$\|\bar{\omega}^\star(t) - \bar{v}(t)\|_\infty \leq \left\|\bar{\omega}^\star(t) - \sum_{n=1}^{t} \alpha_{t,n} v(n)\right\|_\infty \leq t_0(\epsilon)\bar{\alpha}_{t,t_0(\epsilon)} + \epsilon.$$

Since for every fixed n we have $\bar{\alpha}_{t,n} = o(1)$ in t, there exists $t_1(\epsilon)$ such that for $t \geq t_1(\epsilon)$ we have $\bar{\alpha}_{t,t_0(\epsilon)} \leq \epsilon/t_0(\epsilon)$, which implies that $\|\bar{\omega}^\star(t) - \bar{v}(t)\|_\infty \leq 2\epsilon$ for $t \geq \max(t_0(\epsilon), t_1(\epsilon))$.

Now we prove that $\|N_t/t - n(t)\|_\infty \to 0$ as $t \to \infty$. Define $t_2(\epsilon) := \max(t_0(\epsilon), t_1(\epsilon))$, and observe that $\|N_t/t - n(t)\|_\infty \leq \|N_t/t - \bar{v}(t)\|_\infty$.

Therefore we are interested in bounding the quantity $\|N_t/t - \bar{v}^\star(t)\|_\infty$, and we use similar arguments as in (Garivier and Kaufmann, 2016, Lemma 17). Let $t \geq t_2(\epsilon)$. Define $E_u(t) = N_u(t) - t\bar{v}_u(t)$ for every $u \in V$ and note that $\sum_u E_u(t) = 0 \Rightarrow \min_u E_u(t) \leq 0$. We want to show that $\sup_u |E_u(t)/t|$ is bounded.

We begin by showing the following

$$\{U_{t+1} = u\} \subseteq \mathcal{E}_1(t) \cup \mathcal{E}_2(t) \subseteq \{E_u(t) \leq 2t\epsilon\},$$

where $\mathcal{E}_1(t) = \{u = \arg\min_{u\in V} N_u(t) - t\bar{\omega}_u^\star(t)\}$ and $\mathcal{E}_2(t) = \{N_u(t) \leq g(t)\}$ with $g(t) = \max(0, (\sqrt{t} - K/2)) - 1$.

For the first part, if $\{U_{t+1} = u\} \subseteq \mathcal{E}_1(t)$, since $t \geq t_2(\epsilon)$ we have

$$\begin{aligned}E_u(t) &= N_u(t) - t\bar{v}_u(t) \pm t\bar{\omega}_u^\star(t),\\ &\leq N_u(t) - t\bar{\omega}_u^\star(t) + 2t\epsilon,\\ &= \min_v N_v(t) - t\bar{\omega}_v^\star(t) + 2t\epsilon,\\ &\leq \min_v E_v(t) + 4t\epsilon,\\ &\leq 4t\epsilon.\end{aligned}$$

where in the second equality we used the fact that $\{U_{t+1} = u\} \subset \mathcal{E}_1(t)$ and in the last inequality that $\min_v E_v(t) \leq 0$.

For the second part, as shown in (Garivier and Kaufmann, 2016, Lemma 17), there exists $t_3(\epsilon)$ such that for $t \geq t_3(\epsilon)$ then $g(t) \leq 4t\epsilon$ and $1/t \leq \epsilon$. Then if $\{U_{t+1} = u\} \subseteq \mathcal{E}_2(t)$ we have that

$$E_u(t) \leq g(t) - t\bar{v}_u(t) \leq 4t\epsilon.$$

Therefore, as in (Garivier and Kaufmann, 2016, Lemma 17), one can conclude that for $t \geq t' := \max(t_2(\epsilon), t_3(\epsilon))$

$$E_u(t) \leq \max(E_u(t'), 4t\epsilon + 1).$$

Using that $\sum_u E_u(t) = 0$, and that for all $t \geq t', E_u(t') \leq t'$ and $1/t \leq \epsilon$ we have that

$$\sup_i |E_u(t)/t| \leq (K-1)\max(t'/t, 4\epsilon + 1/t) \leq (K-1)\max(5\epsilon, t'/t).$$

Hence, there exists $t'' \geq t'$ such that for all $t \geq t''$ we have $\sup_i |E_u(t)/t| \leq 5(K-1)\epsilon$. Letting $\epsilon \to 0$ concludes the proof. □

Hence, proposition 5 follows by choosing $\alpha_{t,n} = 1/t$ in the previous proposition. Another possible choice is the exponential smoothing factor $\alpha_{t,n} = \kappa_t \lambda^{t-n}$ with $\lambda \in (0,1)$ and $\kappa_t = \frac{1-\lambda}{1-\lambda^t}$. In the next subsection we investigate which choice of $\alpha_{t,n}$ is better.

D.1.1 Is the average of the allocations the best convex combination?

A natural question that arises is which factor $\alpha_{t,n}$ to use. Why is $\alpha_{t,n} = 1/t$ a good choice? This question is related to the fluctuations of the underlying process, and to the stability of the exploration process. We try to give an answer by looking at the variance of the resulting allocation $\bar{w}^\star(t)$.

The i.i.d. case. We begin by considering the i.i.d. case, which supports the fact that a simple average is a good approach to minimize variance.

Lemma 10. *Consider an i.i.d. sequence of Gaussian random variables $\{X_n\}_n$ with 0 mean and variance σ^2. Let $\bar{X}_t(w) = \sum_{n=1}^t w_n X_n$, with $\{w_n\} \in \Delta(\{1, \dots, t\}) = \Delta([t])$. Then*

$$\min_{w \in \Delta([t])} \mathrm{Var}(X_t(w)) = \frac{\sigma^2}{t},$$

which is achieved for $w_i = 1/t$, for all $i \in [t]$.

Proof. Note that $\mathrm{Var}(X_t(w)) = \sum_{n=1}^t w_n^2 \sigma^2$. Introduce the Lagrangian $\mathcal{L}(w, \lambda) = \sum_{n=1}^t w_n^2 \sigma^2 + \lambda(1 - \sum_{n=1}^t w_n)$. Checking the first order condition yields $d\mathcal{L}/dw_n = 2w_n\sigma^2 = \lambda$, hence $w_n = \lambda/(2\sigma^2)$. Since $\sum_n w_n = 1$ we must have $t\lambda/(2\sigma^2) = 1 \Rightarrow \lambda = 2\sigma^2/t$. Therefore $w_n = 1/t$. We conclude that $\min_w \mathrm{Var}(X_t(w)) = \sigma^2/t$. □

More in general, the weighting should be inversely proportional to the variance of the underlying random variable, according to the inverse variance weighting principle Hartung et al. (2011). That is, one can use the same approach as in the previous lemma to easily derive that in case $X_n \sim \mathcal{N}(0, \sigma_n^2)$, then the optimal weighting is $w_n = \frac{1}{\sigma_n^2}\left(\sum_{k=1}^t \frac{1}{\sigma_k^2}\right)^{-1}$.

A more complex case. While the i.i.d. case seems to indicate that taking a simple average is a good approach to minimize the variance, it may not always be the case. In fact, we may expect the random variable $w^\star(t)$ to have smaller fluctuations as t grows larger. We try to give a more complete picture by also looking at a more complex case.

Consider a process $X_n = X_{n-1} + \xi_n$, where ξ_n is an i.i.d. zero-mean process with variance $\mathbb{E}[\xi_n^2] \leq C/n^{1+\alpha}$ for some $\alpha, C > 0$ and $n \geq 2$. And let $X_1 = \xi_1$, with $\mathbb{E}[\xi_1] \leq \sigma^2$.

We define S_t^{avg} to be be the simple average

$$S_t^{avg} = \frac{1}{t}(X_1 + X_2 + \cdots + X_t)$$

Similarly, we define the exponentially smoothed average with $\kappa_t = (1-\lambda)/(1-\lambda^t)$:

$$S_t^{exp} = \kappa_t \sum_{n=1}^t \lambda^{t-n} X_n.$$

Then, we obtain the following result on the variance of the two averages.

Lemma 11. *Consider the simple average S_t^{avg} and the exponentially smoothed average S_t^{exp} with factor $\lambda \in (0,1)$. Then $\mathrm{Var}(S_t^{avg}) \leq \mathrm{Var}(S_t^{exp})$.*

Proof. Let S_t^{avg} be the simple average, and note the following rewriting:

$$S_t^{avg} = \frac{1}{t}(X_1 + X_2 + \cdots + X_t) = X_1 + \frac{(t-1)}{t}\xi_2 + \cdots + \xi_t = \sum_{n=1}^t \frac{t-n+1}{t}\xi_n.$$

Also rewrite the exponentially smoothed average as follows

$$\begin{aligned}
S_t^{exp} &= \kappa_t \sum_{n=1}^t \lambda^{t-n} X_n, \\
&= \kappa_t \sum_{n=1}^t \lambda^{t-n} \sum_{i=1}^n \xi_i, \\
&= \kappa_t \sum_{i=1}^t \xi_i \sum_{n=i}^t \lambda^{t-n}, \\
&= \kappa_t \sum_{i=1}^t \xi_i \sum_{n=0}^{t-i} \lambda^n, \\
&= \kappa_t \sum_{i=1}^t \xi_i \frac{1-\lambda^{t-i+1}}{1-\lambda}, \\
&= \frac{1}{1-\lambda^t} \sum_{i=1}^t (1-\lambda^{t-i+1})\xi_i.
\end{aligned}$$

Due to the properties of X_n we have that $\mathbb{E}[\xi_n] = \mathbb{E}[\mathbb{E}[\xi_n|\mathcal{F}_{n-1}]] = 0$ and $\mathbb{E}[\xi_j\xi_n] = 0$. Therefore, we can write the variance of the averages as follows:

$$\begin{aligned}
\mathrm{Var}(S_t^{avg}) &\leq \sigma^2 + C\sum_{n=2}^t \frac{(t-n+1)^2}{t^2 n^{1+\alpha}}, \\
\mathrm{Var}(S_t^{exp}) &\leq \sigma^2 + \frac{C}{(1-\lambda^t)^2} \sum_{n=2}^t \frac{(1-\lambda^{t-n+1})^2}{n^{1+\alpha}}.
\end{aligned}$$

To show that $\mathrm{Var}(S_t^{avg}) \leq \mathrm{Var}(S_t^{exp})$ we can check if the following inequality holds for all $n \in \{2, \dots, t\}$:

$$\frac{t-n+1}{t} \leq \frac{1-\lambda^{t-n+1}}{1-\lambda^t}.$$

We can prove that the function $h(x) = \frac{1-\lambda^x}{x}$ is decreasing in $x \in [1, t]$. To that aim, compute the derivative $h'(x) = \frac{-\lambda^x x \ln(\lambda) - 1 + \lambda^x}{x^2}$. We are interested in checking if the numerator is negative. Then

$$-\lambda^x x \ln(\lambda) - 1 + \lambda^x \leq 0 \Rightarrow \lambda^x(1 - x\ln(\lambda)) \leq 1.$$

Rewrite as $e^{x\ln(\lambda)}(1 - x\ln(\lambda)) \leq 1$ and let $y = -x\ln(\lambda)$. Then

$$e^{-y}(1+y) \leq 1 \Rightarrow 1 + y \leq e^y,$$

which is always true for $y \geq 0$.

Hence, we have shown that $h(x) \geq h(t)$ for $x \leq t$. Letting $x = t-n+1$, with $n = 2, \dots, t$, concludes the proof. □

Unfortunately considering an approach that minimizes the variance is rather difficult. However, numerical experiments seem to suggest that a simple empirical average is an effective approach.

D.2 Stopping Rule

The (fixed-confidence) Best Arm Identification problem in multi-armed bandit models can be seen as a hypothesis testing problem, where we are testing if $\mu \in \mathcal{H}_k$, where $\mathcal{H}_k = \{\mu' : \mu_k' > \max_{j\neq k} \mu_j'\}$. That is, we are testing if the optimal action in μ is k.

Denoting by $n_a(t)$ the number of times the outcome of a certain action a is observed, the generalized likelihood ratio statistics (GLR) for such problem can be written as

$$\inf_{\lambda \in \mathrm{Alt}(\hat{\mu}(t))} \sum_a n_a(t) \mathrm{KL}(\hat{\nu}_a(t), \lambda_a),$$

where $\nu_a(t)$ is the estimated distribution of rewards for arm a, which depends solely on $\hat{\mu}_a(t)$ due to the assumption that ν_a is a single-parameter exponential distribution, and $\mathrm{Alt}(\hat{\mu}(t)) = \{\lambda : \arg\max_a \lambda_a \neq \arg\max_a\}$ is the set of confusing model.

Taking inspiration from such approach, define the following GLR statistic

$$\Lambda(t) \coloneqq \min_{u \neq \hat{a}_t} \inf_{\lambda : \lambda_u \geq \lambda_{\hat{a}_t}} \sum_{v \in V} M_v(t) \mathrm{KL}(\hat{\nu}_v(t), \lambda_v),$$

where λ is an alternative reward parameter. Then, note that $\Lambda(t)$ can be conveniently rewritten as

$$\begin{aligned}\Lambda(t) &= \min_{u \neq \hat{a}_t} (M_u(t) + M_{\hat{a}_t}(t)) I_{\frac{M_{\hat{a}_t}(t)}{M_u(t)+M_{\hat{a}_t}(t)}}(\hat{\nu}_{\hat{a}_t}, \hat{\nu}_u(t)),\\ &= \min_{u \neq \hat{a}_t} M_{\hat{a}_t}(t)\mathrm{KL}(\hat{\nu}_{\hat{a}_t}(t), \hat{\nu}_{\hat{a}_t,u}) + M_u(t)\mathrm{KL}(\hat{\nu}_u(t), \hat{\nu}_{\hat{a}_t,u}),\end{aligned}$$

where $\hat{\nu}_{\hat{a}_t,u}$ is a distribution of rewards depending on the parameter $\hat{\mu}_{\hat{a}_t,u}(t)$ defined as

$$\hat{\mu}_{a,b}(t) = \frac{M_a(t)}{M_a(t) + M_b(t)} \hat{\mu}_a(t) + \frac{M_b(t)}{M_a(t) + M_b(t)} \hat{\mu}_b(t).$$

One can then show that $tT(N_t/t; \hat{\nu}(t))^{-1}$ is equivalent to $\Lambda(t)$. First, observe that

$$T(N_t/t; \hat{\nu}(t))^{-1} = \min_{u \neq \hat{a}_t} (m_u(t) + m_{\hat{a}_t}(t)) I_{\frac{m_{\hat{a}_t}(t)}{m_u(t)+m_{\hat{a}_t}(t)}}(\hat{\nu}_{\hat{a}_t}(t), \hat{\nu}_u(t)),$$

where $m_u(t) = \sum_v \hat{G}_{v,u}(t) \frac{N_v(t)}{t} = \sum_v \frac{N_{v,u}(t)}{N_v(t)} \frac{N_v(t)}{t} = M_u(t)/t$. Using this latter fact, and noting that $\frac{m_{\hat{a}_t}(t)}{m_u(t)+m_{\hat{a}_t}(t)} = \frac{M_{\hat{a}_t}(t)}{M_u(t)+M_{\hat{a}_t}(t)}$, we get

$$tT(N_t/t; \hat{\nu}(t))^{-1} = \min_{u \neq \hat{a}_t} (M_u(t) + M_{\hat{a}_t}(t)) I_{\frac{M_{\hat{a}_t}(t)}{M_u(t)+M_{\hat{a}_t}(t)}}(\hat{\nu}_{\hat{a}_t}(t), \hat{\nu}_u(t)) = \Lambda(t).$$

We can now provide the proof of the stopping rule.

Proof of proposition 6. First note that the event $\{\hat{a}_\tau \neq a^\star(\mu)\} \subset \{\nu \in \mathrm{Alt}(\hat{\nu}(\tau))\}$. From the discussion above, using the notation $\Lambda(t) = tT(N_t/t; \hat{\nu}(t))^{-1}$, we observe that under $\{\nu \in \mathrm{Alt}(\hat{\nu}(\tau))\}$ then the following inequalities hold

$$\begin{aligned}\mathbb{P}_\nu(\tau < \infty, \hat{a}_\tau \neq a^\star(\mu)) &\leq \mathbb{P}_\nu(\exists t \in \mathbb{N} : \hat{a}_t \neq a^\star(\mu), tT(N_t/t; \hat{\nu}(t))^{-1} > \beta(t, \delta)),\\ &\leq \mathbb{P}_\nu\left(\exists t \in \mathbb{N} : \hat{a}_t \neq a^\star(\mu), \min_{u \neq \hat{a}_t} \inf_{\lambda : \lambda_u \geq \lambda_{\hat{a}_t}} \sum_{v \in V} M_v(t)\mathrm{KL}(\hat{\nu}_v(t), \lambda_v) > \beta(t, \delta)\right),\\ &\leq \mathbb{P}_\nu\left(\exists t \in \mathbb{N}, \exists u \neq \hat{a}_t : \hat{a}_t \neq a^\star(\mu), \inf_{\lambda : \lambda_u \geq \lambda_{\hat{a}_t}} \sum_{v \in V} M_v(t)\mathrm{KL}(\hat{\nu}_v(t), \lambda_v) > \beta(t, \delta)\right),\\ &\leq \mathbb{P}_\nu\left(\exists t \in \mathbb{N}, \exists u \neq \hat{a}_t : \sum_{v \in \{u, \hat{a}_t\}} M_v(t)\mathrm{KL}(\hat{\nu}_v(t), \nu_v) > \beta(t, \delta)\right).\end{aligned}$$

Now, from (Kaufmann and Koolen, 2021, Theorem 7), we know that

$$\mathbb{P}_\nu\left(\exists t \in \mathbb{N}, \exists u \neq \hat{a}_t : \sum_{v \in \{u, \hat{a}_t\}} M_v(t)\mathrm{KL}(\hat{\nu}_v(t), \nu_v) > 2\mathcal{C}_{\exp}\left(\frac{\ln\left(\frac{K-1}{\delta}\right)}{2}\right) + 3 \sum_{v \in \{\hat{a}_t, u\}} \ln(1 + \ln(M_v(t)))\right) \leq \delta.$$

where we applied (Kaufmann and Koolen, 2021, Theorem 7) over $K - 1$ subsets $\{\underbrace{(u, \hat{a}_t)}_{\mathcal{S}_u}\}_{u \neq \hat{a}_t}$ of size 2 each, and took a union bound. Finally, using Jensen's inequality we also have that

$$\begin{aligned}\ln(1+\ln(M_v(t)))+\ln(1+\ln(M_u(t))) &\le 2\ln\left(\frac{1+\ln(M_v(t))}{2}+\frac{1+\ln(M_u(t))}{2}\right),\\ &= 2\ln\left(1+\frac{\ln(M_v(t))+\ln(M_u(t))}{2}\right),\\ &\le 2\ln\left(1+\ln\left(\frac{M_v(t)+M_u(t)}{2}\right)\right).\end{aligned}$$

Note that $M_v(t)+M_u(t)$ cannot exceed $2t$ (just consider a full feedback graph where $G_{u,v}=1$ so that $M_v(t)=t$ for every v). Hence, this implies that the GLR statistics is δ-PC with the threshold

$$\beta(t,\delta)=2\mathcal{C}_{\exp}\left(\frac{\ln\left(\frac{K-1}{\delta}\right)}{2}\right)+6\ln(1+\ln(t)).$$

□

Definition of $\mathcal{C}_{\exp}(x)$. Last, but not least, we briefly explain the definition of $\mathcal{C}_{\exp}(x)$. We define $\mathcal{C}_{\exp}(x)$ as (Kaufmann and Koolen, 2021, Theorem 7) $\mathcal{C}_{\exp}(x) \coloneqq 2\tilde{h}_{3/2}\left(\frac{h^{-1}(1+x)+\ln(2\zeta(2))}{2}\right)$, where: $\zeta(s)=\sum_{n\ge 1}n^{-s}$; $h(u)=u-\ln(u)$ for $u\ge 1$; lastly, for for any $z\in[1,e]$ and $x\ge 0$:

$$\tilde{h}_z(x)=\begin{cases}h^{-1}(x)e^{1/h^{-1}(x)} & \text{if } x\ge h(1/\ln z),\\ z(x-\ln\ln z) & \text{otherwise.}\end{cases}$$

D.3 Sample Complexity Analysis

The following sample complexity analysis follows the analysis of Garivier and Kaufmann (2016) while adopting necessary changes for our problem setup. We first prove part (1) and (2) of theorem 3, and prove part (3) of theorem 3 separately. In the end of this section, we provide the proof for corollary 1.

Proof of part (1) and (2) of theorem 3: Let $\mathcal{E}$ be the event:

$$\mathcal{E}=\left\{\inf_{w\in C^\star(\nu)}\left\|\frac{N(t)}{t}-\omega\right\|_\infty \xrightarrow{t\to\infty} 0, \hat{\nu}(t)\xrightarrow{t\to\infty}\nu\right\}.$$

By Proposition 5 and the law of large number, we have $\mathcal{E}$ holds with probability 1. On $\mathcal{E}$, with the continuity property of function $T(\omega,\nu)^{-1}$ at $(\omega^\star(\nu),\nu)$ and proposition 5, for every $\omega^\star(\nu)\in C^\star(\nu)$, we have for all $\epsilon>0$ there exists $t_0\in\mathbb{N}$ such that for all $t\ge t_0$:

$$T(N(t)/t;\hat{\nu}(t))^{-1}\ge\frac{1}{1+\epsilon}T^\star(\nu)^{-1}.$$

Therefore, for $t\ge t_0$:

$$L(t)=tT(N(t)/t;\hat{\nu}(t))^{-1}\ge\frac{t}{1+\epsilon}T^\star(\nu)^{-1}.$$

Hence,

$$\begin{aligned}\tau&=\inf\{t\in\mathbb{N}:L(t)\ge\beta(t,\delta)\},\\ &\le t_0\vee\inf\left\{t\in\mathbb{N}:\frac{t}{1+\epsilon}T^\star(\nu)^{-1}\ge\beta(t,\delta)\right\}.\end{aligned}$$

Recall that $\beta(t,\delta)\coloneqq 2\mathcal{C}_{\exp}\left(\frac{\ln\left(\frac{K-1}{\delta}\right)}{2}\right)+6\ln(1+\ln(t))$. Note that there exists a universal constant B such that $\beta(t,\delta)\le\ln(Bt/\delta)$. Hence,

$$\tau\le t_0\vee\inf\left\{t\in\mathbb{N}:\frac{t}{1+\epsilon}T^\star(\nu)^{-1}\ge\ln(Bt/\delta)\right\}.$$

Applying Lemma 18 of Garivier and Kaufmann (2016) by letting $\alpha = 1$:

$$\tau \leq t_0 \vee (1+\epsilon)T^\star(\nu)\left[\ln\left(\frac{Be(1+\epsilon)T^\star(\nu)}{\delta}\right) + \ln\ln\left(\frac{B(1+\epsilon)T^\star(\nu)}{\delta}\right)\right].$$

Thus, τ is finite with probability 1. And

$$\limsup_{\delta\to 0}\frac{\tau}{\ln(1/\delta)} \leq (1+\epsilon)T^*(\nu).$$

We conclude the proof of part (2) by letting $\epsilon \to 0$. □

Proof of part (3) of theorem 3: Let $T \in \mathbb{N}$, for $\epsilon > 0$, define $\mathcal{E}_T := \bigcap_{t=T^{\frac{1}{4}}}^{T}(\hat{\nu}(t) \in \mathcal{I}_\epsilon)$, where $\mathcal{I}_\epsilon := \{\nu' : \|\nu' - \nu\|_\infty \leq \epsilon\}$ and $\|\nu' - \nu\|_\infty := \max\{\|G' - G\|_\infty, \|\mu' - \mu\|_\infty\}$. Following the same argument as Lemma 19 of Garivier and Kaufmann (2016), one can show that there exist two constant B and C (that depend on ν and ϵ) such that $\mathbb{P}_\nu(\mathcal{E}_T^c) \leq BT\exp(-CT^{\frac{1}{8}})$. Denote

$$C_\epsilon^\star(\nu) = \inf_{\substack{\omega' : \|\omega' - \mathrm{Proj}_{C^\star(\nu)}(\omega')\|_\infty \leq 5(K-1)\epsilon \\ \nu' : \|\nu' - \nu\|_\infty \leq \epsilon}} T(\omega', \nu')^{-1}.$$

By proposition 9, for any ϵ, there exists $T(\epsilon)$ such that for any $T \geq T(\epsilon)$ and $t \geq \sqrt{T}$, $\|N(t)/t - \mathrm{Proj}_{C^\star(\nu)}(N(t)/t)\|_\infty \leq 5(K-1)\epsilon$. With this fact, on $\mathcal{E}_T$, for $T \geq T(\epsilon)$ and $t \geq \sqrt{T}$, one has:

$$L(t) = tT(N(t)/t; \hat{\nu}_t)^{-1} \geq tC_\epsilon^\star(\nu).$$

Therefore, let $T \geq T(\epsilon)$, on $\mathcal{E}_T$,

$$\begin{aligned}
\min\{\tau_\delta, T\} &\leq \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbf{1}_{(\tau_\delta > t)}, \\
&= \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbf{1}_{(L(t) \leq \beta(t,\delta))}, \\
&\leq \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbf{1}_{(tC_\epsilon^\star(\nu) \leq \beta(t,\delta))}, \\
&\leq \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbf{1}_{(tC_\epsilon^\star(\nu) \leq \beta(T,\delta))}, \\
&\leq \sqrt{T} + \frac{\beta(T,\delta)}{C_\epsilon^\star(\nu)}.
\end{aligned}$$

Denote $T_0 = \inf\left\{T \in \mathbb{N} : \sqrt{T} + \frac{\beta(T,\delta)}{C_\epsilon^\star(\nu)} \leq T\right\}$. Thus, one has for $T \geq \max\{T_0, T(\epsilon)\}$, $\mathcal{E}_T \subseteq (\tau_\delta \leq T)$. Hence,

$$\begin{aligned}
\mathbb{E}[\tau_\delta] &\leq \max\{T_0, T(\epsilon)\} + \sum_{T=\max\{T_0, T_\epsilon\}}^{\infty} \mathbb{P}(\tau_\delta > T), \\
&\leq T_0 + T(\epsilon) + \sum_{T=\max\{T_0, T_\epsilon\}}^{\infty} BT\exp(-CT^{\frac{1}{8}}).
\end{aligned}$$

We then upper bound T_0. By introducing a constant $C(\eta) = \inf\left\{T \in \mathbb{N} : T - \sqrt{T} \geq \frac{T}{1+\eta}\right\}$, one has

$$\begin{aligned}
T_0 &\leq C(\eta) + \inf\left\{T \in \mathbb{N} : \frac{\beta(T,\delta)}{C_\epsilon^\star(\nu)} \leq \frac{T}{1+\eta}\right\}, \\
&\leq C(\eta) + \inf\left\{T \in \mathbb{N} : \frac{C_\epsilon^\star(\nu)T}{1+\eta} \geq \ln(BT/\delta)\right\}.
\end{aligned}$$

Applying Lemma 18 of Garivier and Kaufmann (2016) again:

$$T_0(\delta) \leq C(\eta) + (1+\eta) C_\epsilon^\star(\nu)^{-1} \left[\ln\left(\frac{Be(1+\eta)}{C_\epsilon^\star(\nu)\delta} \right) + \ln\ln\left(\frac{B(1+\eta)}{C_\epsilon^\star(\nu)\delta} \right) \right].$$

Therefore,

$$\limsup_{\delta\to 0} \frac{\mathbb{E}[\tau_\delta]}{\ln(1/\delta)} \leq \frac{(1+\eta)}{C_\epsilon^\star(\nu)}.$$

From the continuity property of function $T(\omega,\nu)^{-1}$ at $(\omega^\star(\nu),\nu)$ for each $\omega^\star(\nu)$ in $C^\star(\nu)$, one has

$$\lim_{\epsilon\to 0} C_\epsilon^\star(\nu) = T^\star(\nu)^{-1},$$

Letting η go to 0:

$$\limsup_{\delta\to 0} \frac{\mathbb{E}[\tau_\delta]}{\ln(1/\delta)} \leq T^\star(\nu).$$

□

Proof of Corollary 1. The proof relies on the following proposition, which is a direct application of Proposition 9.

Proposition 10. *Let $S_t = \{u \in V : N_u(t) < \sqrt{t} - K/2\}$. The D-tracking rule, defined as*

$$V_t \in \begin{cases} \arg\min_{u\in S_t} N_u(t) & S_t \neq \emptyset \\ \arg\min_{u\in V} N_u(t) - t\sum_{n=1}^t \alpha_{t,n}\omega_{\text{heur}}(n) & \textit{otherwise} \end{cases}, \tag{15}$$

where, for every $t \geq 1$, the sequence $\alpha_t = (\alpha_{t,n})_{n=1}^t$ satisfies: (1) $\alpha_{t,n} \in [0,1]$; (2) for every fixed $n \in \{1,\dots,t\}$ we have $\alpha_{t,n} = o(1)$ in t ; (3) for all t, $\sum_{n=1}^t \alpha_{t,n} = 1$.

Such tracking rule ensures that for all $\epsilon > 0$ there exists $t(\epsilon)$ such that for all $t \geq t(\epsilon)$ we have

$$\|N(t)/t - \omega_{\text{heur}}\|_\infty \leq 5(K-1)\epsilon,$$

and $\lim_{t\to\infty} \|N(t)/t - \omega_{\text{heur}}\|_\infty \to 0$ almost surely.

Proof. Proposition 10 can be proved using the same analysis as in Proposition 9, except that one needs to replace $C^\star(\nu)$ with $\{\omega_{\text{heur}}\}$. We thus skip the full proof. □

Corollary 1 can then be proved using the same analysis as in Theorem 3 combined with Proposition 10. □

We also provide an almost-surely lower bound for this heuristic algorithm, as established in the following Proposition 11.

Proposition 11. *TaS-FG with $\omega^\star(t) = \omega_{\text{heur}}(t)$ satisfies that $\mathbb{P}_\nu\left(\liminf_{\delta\to 0} \frac{\tau}{\ln(1/\delta)} \geq T^\star(\nu)\right) = 1$.*

Proof. By the continuity property of function $T(\omega,\nu)^{-1}$ at $(\omega_{\text{heur}},\nu)$ and Proposition 10, for any $\epsilon > 0$, with probability 1, there exists t_1 such that for any $t \geq t_1$,

$$L(t) = tT(N(t)/t; \hat{\nu}(t))^{-1} \leq \frac{t}{1-\epsilon} T(\omega_{\text{heur}},\nu)^{-1} \leq \frac{t}{1-\epsilon} T^\star(\nu)^{-1}.$$

By the definition of the stopping time,

$$\begin{aligned} \tau &= \inf\{t \in \mathbb{N} : L(t) \geq \beta(t,\delta)\}, \\ &\geq \inf\{t \in \mathbb{N} : L(t) \geq \ln(1/\delta)\}. \end{aligned}$$

When $\delta \to 0$, $\inf\{t \in \mathbb{N} : L(t) \geq \ln(1/\delta)\} > t_1$. Therefore,

$$\begin{aligned} \liminf_{\delta\to 0} \frac{\tau}{\ln(1/\delta)} &\geq \liminf_{\delta\to 0} \frac{\inf\left\{t\in\mathbb{N} : tT^\star(\nu)^{-1} \geq (1-\epsilon)\ln(1/\delta)\right\}}{\ln(1/\delta)}, \\ &\geq T^\star(\nu)(1-\epsilon). \end{aligned}$$

Letting ϵ go to 0 concludes the proof. □