Graph-Triggered Rising Bandits

G. Genalti, M. Mussi, N. Gatti, M. Restelli, M. Castiglioni and A. M. Metelli

{gianmarco.genalti, marco.mussi, nicola.gatti, marcello.restelli, matteo.castiglioni, albertomaria.metelli}@polimi.it

POLITECNICO MILANO 1863

Stochastic Rising Bandits

Evolving Rewards Bandits

- The reward obtained by choosing action $I_t \sim \pi$ at round t is a **random variable with mean** $\mu_i(n^\pi_{I_t,t})$, for every $n^\pi_{I_t,t} \in \mathbb{N}$ and $I_t \in [k]$
- n^π_t is a **quantity that depends on the history**, e.g., the number of pulls $N_{i,t}$ or the current round t
- The goal is to **maximize** the **expected cumulative reward**:

$$J_T(\pi) = \mathbb{E}\left[\sum_{t=1}^{T} \mu_{I_t}(n^\pi_{I_t,t})\right], \text{ where } I_t \sim \pi$$

  or, equivalently, **minimize** the **regret**:

$$R_T(\pi) = \max_{\tilde{\pi}} J_T(\tilde{\pi}) - J_T(\pi)$$

- **Notable examples** are:

$$n^\pi_{i,t} = N^\pi_{i,t} \text{ (number of pulls)} \implies \textbf{Rested Bandits}$$

$$n^\pi_{i,t} = t \text{ (time)} \implies \textbf{Restless Bandits}$$

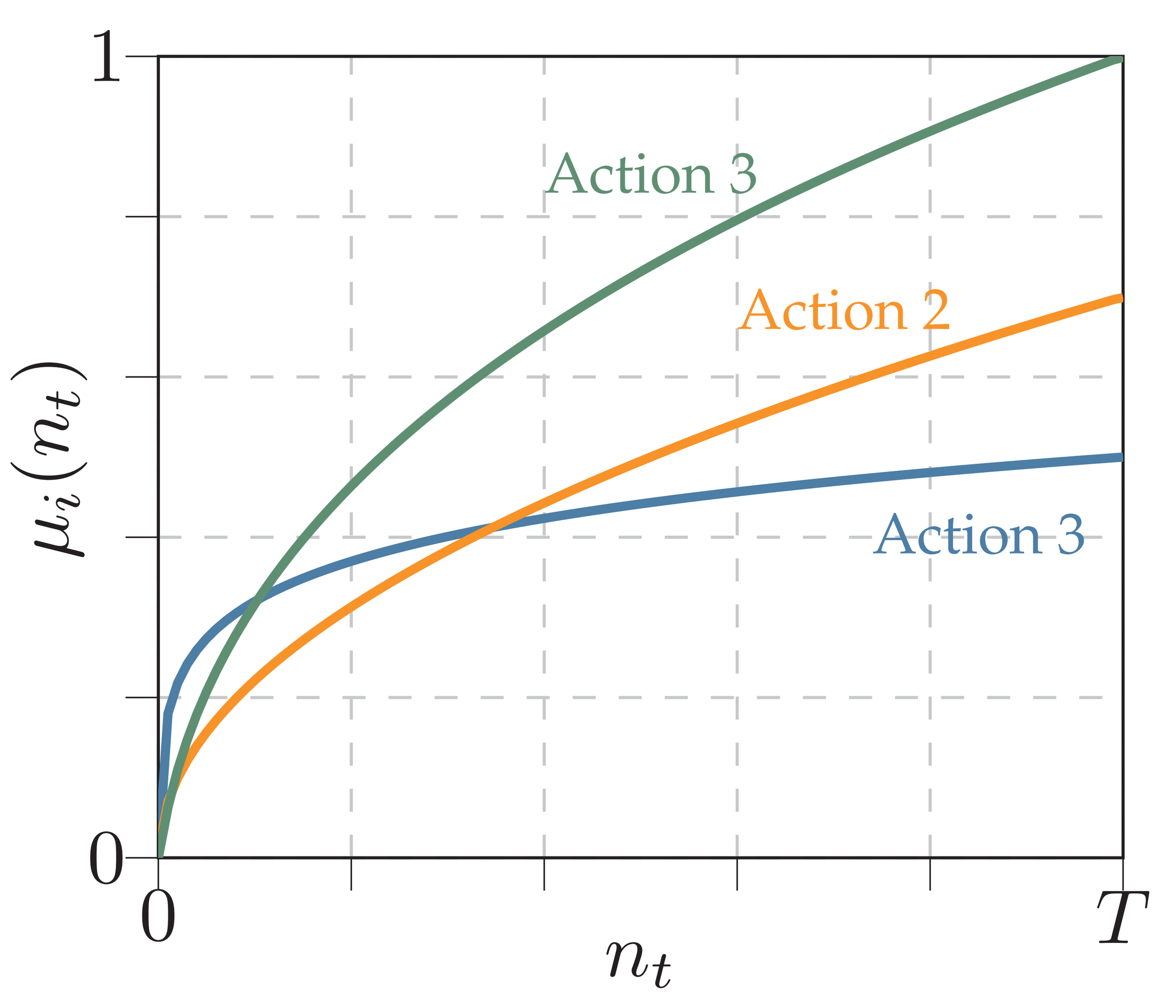
Rising Bandits

Rising Bandits are a **special class of evolving rewards bandits** where $\mu_i : n \mapsto [0,1]$ satisfy the following assumption for every $i \in [k]$ and $n \in [T]$:

$$\textbf{\textit{Non-decreasing:}} \quad \gamma_i(n) \geq 0$$

$$\textbf{\textit{Concave:}} \quad \gamma_i(n-1) \geq \gamma_i(n)$$

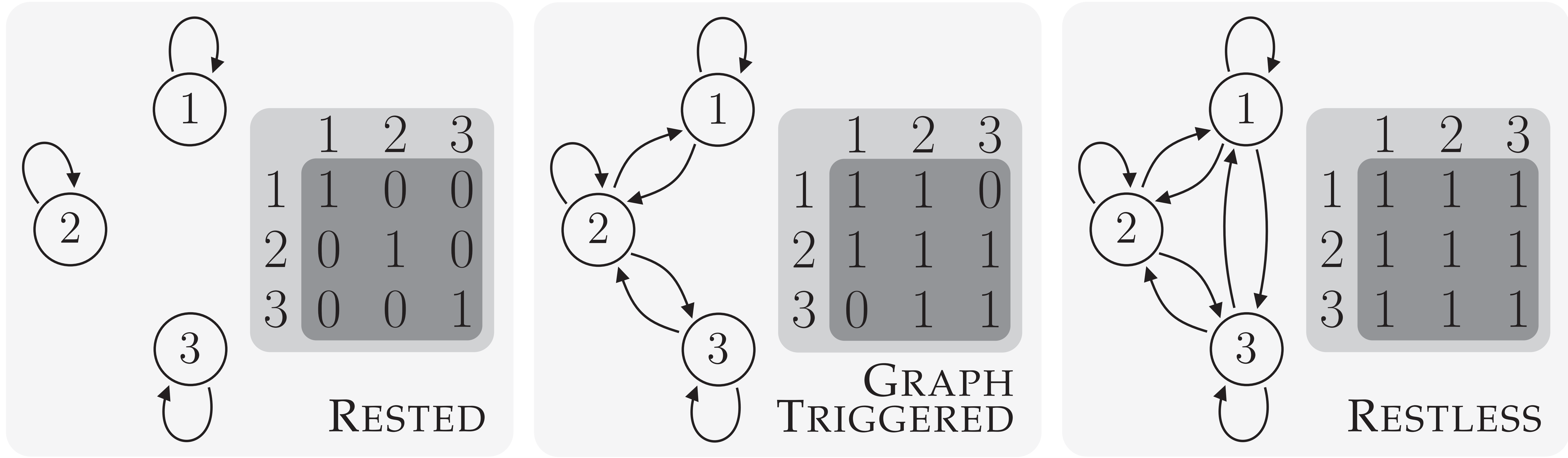
where $\gamma_i(n) := \mu_i(n+1) - \mu_i(n)$



Graph-Triggered Bandits

In **Graph-Triggered Bandits**, actions are related by the means of an **undirected graph**

$$n^\pi_{i,t} = \widetilde{N}^\pi_{i,t} := \sum_{j \text{ connected to } i} N^\pi_{j,t} \quad \text{(number of \textbf{triggers})} \implies \textbf{Graph-Triggered Bandits}$$



In **Rested** Rising Bandits, optimal policy commits on **one action**

Optimal policy is **NP-Hard to compute!**

In **Restless** Rising Bandits, optimal policy is **greedy**

Rested and **Restless** Bandits $\subset$ **Graph-Triggered Bandits** $\subset$ Evolving Rewards Bandits

Graph-Triggered Rising Bandits with Cluster Graph

For graph-triggered **rising** bandits, **if the graph can be partitioned into a set $\mathcal{C}$ of disjoint cliques**, given a time horizon T the optimal action I^*_t can be computed as

$$\underbrace{I^*_t \in \arg\max_{i \in C^*_T} \mu_i(t)}_{\text{Greedy behavior like in \textbf{restless} rising bandits}} \qquad \text{where} \qquad \underbrace{C^*_T \in \arg\max_{C \in \mathcal{C}} \sum_{t \in [T]} \max_{j \in C} \mu_j(t)}_{\text{Commitment on one clique like in \textbf{rested} rising bandits}} .$$

Solving Graph-Triggered Rising Bandits

Let $\Upsilon_{\boldsymbol{\nu}}(M,q) := \sum_{t \in [M-1]} \max_{i \in [k]} \gamma_i(t)^q$, for every $M \in [T]$ and $q \in [0,1]$

Regret Bound for Deterministic GTRBs with Cluster Graph

$$R_{\boldsymbol{\nu},\mathbf{G}}(\texttt{BR-BG-UB}) \leq \widetilde{\mathcal{O}}\left(\inf_{q \in [0,1]} \left\{ \underbrace{T^q \sum_{C_m \in \mathcal{C}} |C_m| \Upsilon_{\boldsymbol{\nu}}\left(\left\lceil \frac{\widetilde{N}_{C_m,T}}{|C_m|} \right\rceil, q\right)}_{\text{Rested Contribution}} + \underbrace{\sum_{C_m \in \mathcal{C}} |C_m| \widetilde{N}^{\frac{q}{1+q}}_{C_m,T} \Upsilon_{\boldsymbol{\nu}}\left(\left\lceil \frac{\widetilde{N}_{C_m,T}}{|C_m|} \right\rceil, q\right)^{\frac{1}{1+q}}}_{\text{Restless Contribution}} \right\}\right)$$

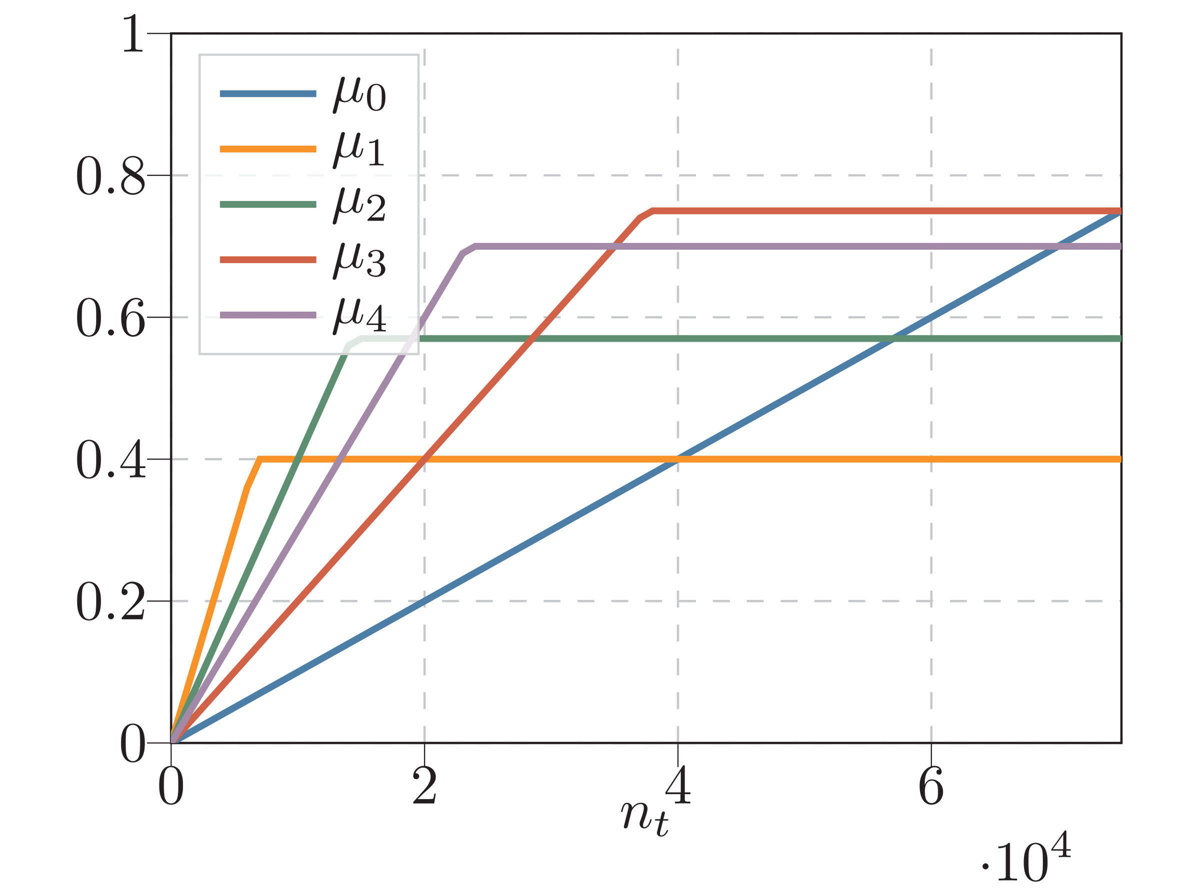
Regret Bound for Stochastic GTRBs with Cluster Graph

$$R_{\boldsymbol{\nu},\mathbf{G}}(\texttt{R-}\square\texttt{-UCB}) \leq \widetilde{\mathcal{O}}\left(\min_{q \in [0,1]} \left\{ \underbrace{(\sigma T)^{\frac{2}{3}}}_{\text{Noise Contribution}} + \underbrace{\bar{k}_1 T^q \Upsilon_{\boldsymbol{\nu}}\left(\left\lceil \frac{T}{\bar{k}_1} \right\rceil, q\right)}_{\text{Rested Contribution}} + \underbrace{T^{\frac{2q}{1+q}} \sum_{C_m \in \mathcal{C}_{\mathbf{G}} : |C_m| > 1} |C_m| \Upsilon_{\boldsymbol{\nu}}\left(\left\lceil \frac{T}{|C_m|} \right\rceil, q\right)^{\frac{1}{1+q}}}_{\text{Restless Contribution}} \right\}\right),$$

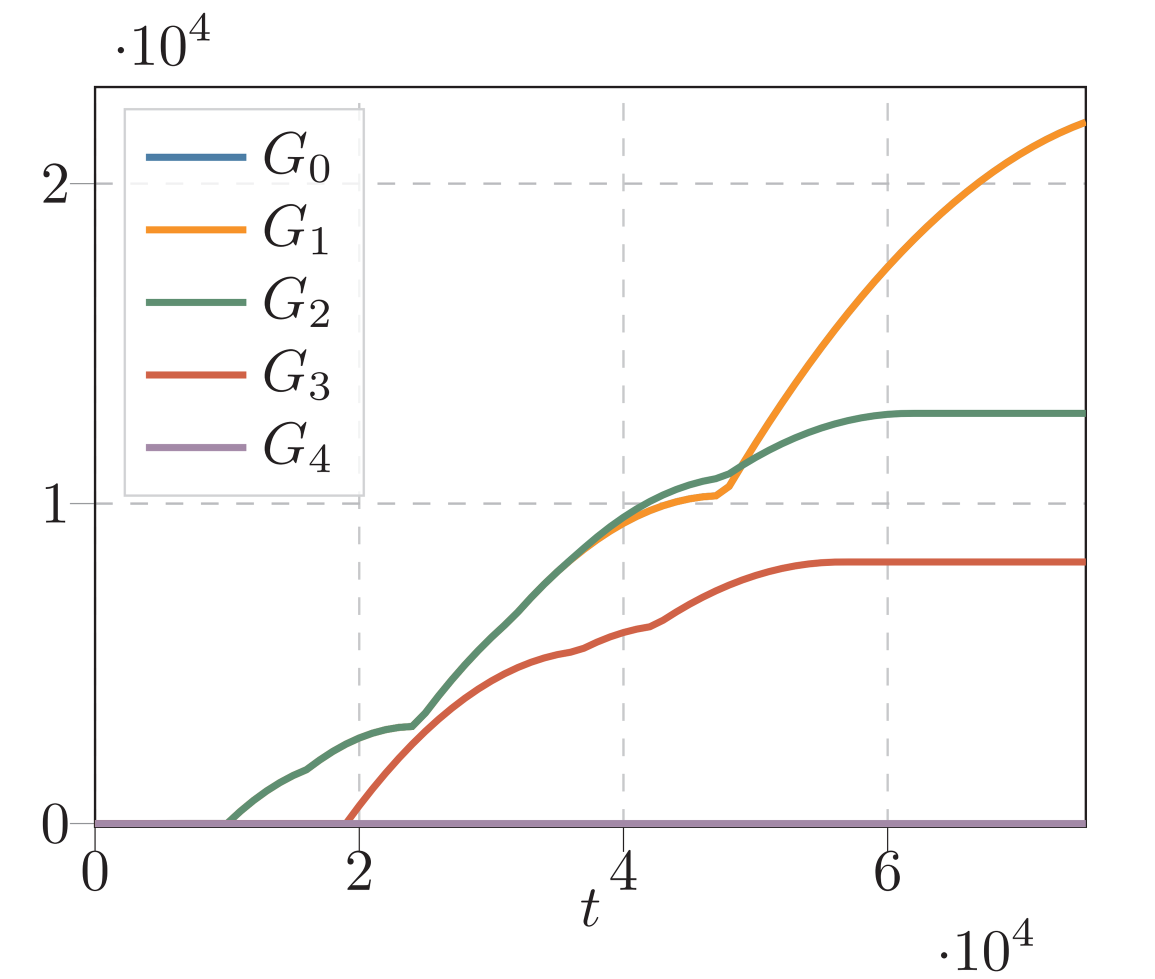
Results for GTRBs with cluster graph can be used to obtain **regret guarantees for general graphs!**

Empirical Validation

We consider a **rising bandit** with $k = 5$ and reward functions from the family $\mu_i = \min\{\kappa_i n_t, m_i\}$



Using **cluster graphs only**, we show the regret bounds attained by `BR-BG-UB` when varying the adjacency matrix in a **deterministic** setting



As the graph becomes sparser, the regret becomes higher, which is concordant with theory

References

R. M. Karp. Reducibility among combinatorial problems. *Complexity of Computer Computations*, 1972.

A. M. Metelli, F. Trovò, M. Pirola, and M. Restelli. Stochastic rising bandits. In *International Conference on Machine Learning*, 2022.

C. Tekin and M. Liu. Online learning of rested and restless bandits. *IEEE Transactions on Information Theory*, 2012.

P. Whittle. Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability*, 1988.