Pure Exploration and Regret Minimization in Matching Bandits

Flore Sentenac, Jialin Yi, Clément Calauzenes, Vianney Perchet, Milan Vojnovic

Motivation: collaborative activities

Competitive game of tic-tac-toe:

Pool of users

→

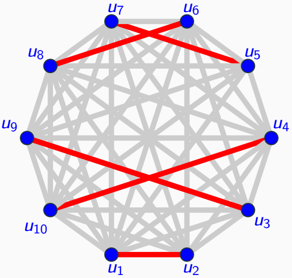
Selected Players

→

Play if both of them want to.

- Gaming apps e.g. Go, competitive quizzes or drawing...
- Teamwork
- Online labor platforms

Matching (Semi-)Bandit

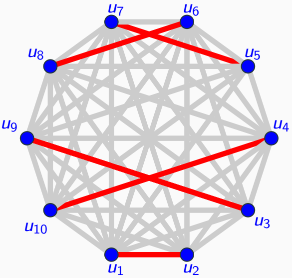


Set of arms : edges of the graph $(\mathcal{U}, \mathcal{E})$

Combinatorial constraint : the selected arms form a matching (no vertex is selected twice)

+ Rank one structure: Expected reward for arm (u_i, u_j), $\mathbb{E}[x_{ij}(t)] = u_i u_j$

Matching (Semi-)Bandit



At round t:

Select matching m_t

Receive reward $\sum_{(i,j)\in m_t} x_{ij}(t)$

Observe $\{x_{ij}(t)\}_{(i,j)\in m_t}$

- Regret minimization: maximize expected cumulated reward
- Pure exploration: identify best super-arm w.h.p. as fast as possible

Regret minimisation with maximum matching sampling

Δ_{min} := gap between best super arm and second best super-arm.

Combinatorial semi-bandits:

→ Generic Regret scales as:

$$\frac{N^2 \log^2(N)}{\Delta_{min}} \log(T)$$

Matching/Rank One structure:

→ Regret of Adaptive Matching scales as:

$$\frac{N \log(N)}{\Delta_{min}} \log(T)$$

Adaptive Matching

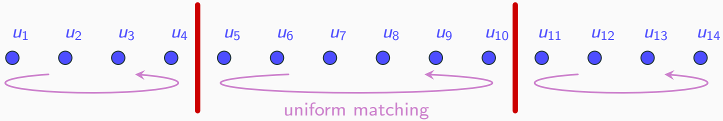
w.l.o.g $u_1 \geq u_2 \geq \ldots \geq u_N$.

$\rightarrow$ The optimal matching pairs the elements in decreasing order

$$\{(u_1, u_2), (u_3, u_4), \ldots, (u_{N-1}, u_N)\}$$

Algorithm 1: Adaptive Matching

```
in: set of items [N];
for t = 1, ... do
    Uniformly match the items within the same cluster;
    Partition items into ranked clusters;
end
```



General Results

	Regret Minimisation (horizon T)	Pure Exploration (target precision δ)
Pair	$\sum_{\Delta_{2,i}>0} \frac{1}{u_1 \Delta_{2,i}} \log(T)$ (tight up to mul. cst)	$\sum_{\Delta_{2,i}>0} \frac{1}{(u_1 \Delta_{2,i})^2} \log(\frac{1}{\delta})$ (tight up to mul. cst)
Maximum Matching	$\frac{N \log(N)}{\Delta_{\min}} \log(T)$	tight bounds up to mul. cst in interesting parameter regimes

$\Delta_{2,i} := u_2 - u_i$

Thank You !