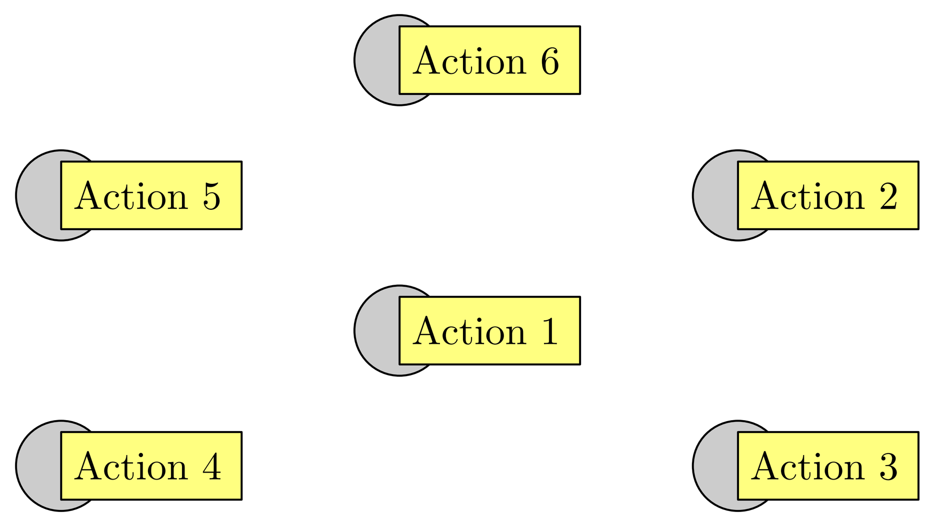
Online learning with Erdős-Rényi side-observation graphs

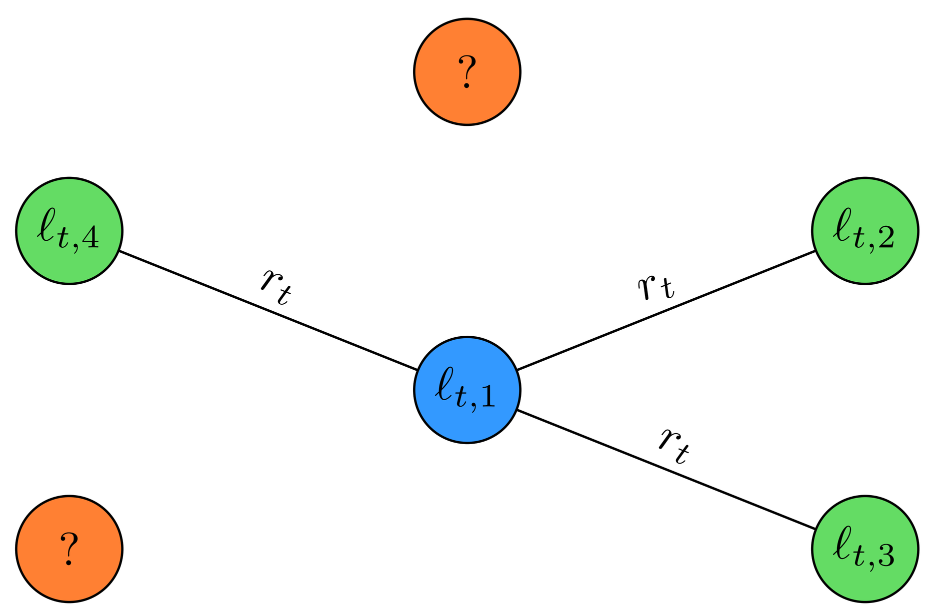
Tomas.Kocak@inria.fr, Neu.Gergely@gmail.com, and Michal.Valko@inria.fr

Problem illustration

- **Select one of the actions** (for example action 1):



- **Nature generates Erdős-Rényi graph** with parameter r_t



- **Incur loss** of the selected action
- **Observe losses** of the neighbors of the selected action

Problem formalization

Learning process

- N **actions** (nodes of a graph)
- T rounds:
  - Environment (adversary) **sets losses** for actions
  - Environment **choses observation probability** r_t
  - Learner **picks an action** I_t to play
  - Learner **incurs the loss** ℓ_{t,I_t} of the action I_t
  - Learner **observes loss** $\ell_{t,j}$ with probability $r_t\ \forall j \neq i$

Goal of the learner

- **Minimize cumulative regret** R_t defined as

$$R_t = \underbrace{\sum_{t=1}^{T} \ell_{t,I_t}}_{\text{learner}} - \underbrace{\min_{j\in[N]} \left[\sum_{t=1}^{T} \ell_{t,j}\right]}_{\text{best action}}$$

Previous work

Previous papers:
(Mannor and Shamir, 2011; Alon et al., 2013; Kocák et al., 2014, 2016)

+ Can handle general (possibly adversarial) graphs
− Need to know the second neighborhood of selected action

Current paper:

+ Need to know only the first neighborhood of selected action
− Can handle only Erdős Rényi graphs

No assumptions (general graph with the first neighborhood):

- Problem as hard as MAB problem (Cohen et al., 2016)
- Learner can ignore side observations

Exp3-Res algorithm

- **Compute exponential weights** using loss estimates $\widehat{\ell}_{s,i}$

$$w_{t,i} = \exp\left(-\eta_t \sum_{s=1}^{t-1} \widehat{\ell}_{s,i}\right)$$

- **Create a probability distribution** such that $p_{t,i} \propto w_{t,i}$
- **Play action** I_t such that

$$\mathbb{P}(I_t = i) p_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{N} w_{t,j}}$$

- **Create loss estimates** $\ell_{t,i}$ using side observations
  - Estimates compensate for incomplete observations
  - Good loss approximation: $\mathbb{E}[\widehat{\ell}_{t,i}] \approx \ell_{t,i}$

Loss estimates

Ideal (unbiased) loss estimate:

$$\widehat{\ell}_{t,i} = \frac{\ell_{t,i}\mathbb{1}\{\text{loss } \ell_{t,i} \text{ is observed}\}}{\mathbb{P}\{\text{loss } \ell_{t,i} \text{ is observed}\}} = \frac{\ell_{t,i} O_{t,i}}{o_{t,i}}$$

where

$$\mathbb{P}(\text{loss } \ell_{t,i} \text{ is observed}) = \underbrace{p_{t,i}}_{I_t=i} + \underbrace{(1-p_{t,i})}_{I_t\neq i} \underbrace{r_t}_{\text{side observation}} = o_{t,i}$$

Problem:

- No access to $o_{t,i}$ (unknown r_t)

Solution:

- Use other side observations to compensate for unknown r_t
- Design $G_{t,i}$ such that $\mathbb{E}[G_{t,i}] \approx \frac{1}{o_{t,i}}$
- Define low-biased loss estimates as

$$\widehat{\ell}_{t,i} = \ell_{t,i} O_{t,i} G_{t,i}$$

Geometric resampling

Independent random variables

- $\{R_k\}_{k=1}^{\infty}$ - Bernoulli random variables with parameter r_t
- $\{P_k\}_{k=1}^{\infty}$ - Bernoulli random variables with parameter $p_{t,i}$

Combining $\{R_k\}_{k=1}^{\infty}$ **and** $\{P_k\}_{k=1}^{\infty}$ **we get**

$$O_k = P_k + (1 - P_k) R_k$$
$$\mathbb{E}[O_k] = p_{t,i} + (1 - p_{t,i}) r_t = o_{t,i}$$

- $\{O_k\}_{k=1}^{\infty}$ - Bernoulli random variables with parameter $o_{t,i}$

$$G^*_{t,i} = \min\{k :\ O_k = 1\}$$

- $G^*_{t,i}$ - Geometric random variable with parameter $o_{t,i}$

$$\mathbb{E}[G^*_{t,i}] = \frac{1}{o_{t,i}}$$

Challenge:

- r_t is unknown for learner
- Access to $\{R_k\}_{k=1}^{\infty}$ is limited by number of actions
  - $N - 2$ samples (all actions except I_t and i)

Solution:

- Setting R_k to 1 for all $k > N - 2$ (Introducing bias to $G^*_{t,i}$)
  - Bias is optimistic and controlled.

$$G_{t,i} = \min\{\{k \le N-2 :\ O_k = 1\} \cup \{N-1\}\}$$

$$\mathbb{E}[G_{t,i}] = \frac{1 - (1 - o_{t,i})^{N-1}}{o_{t,i}}$$

Limitations:

- Bias term $(1 - o_{t,i})^{N-1}$ appears in the regret bound

Assupmtion:

- Let $r_t \ge \log(T)/(2N-2)$. Then bias term is negligible

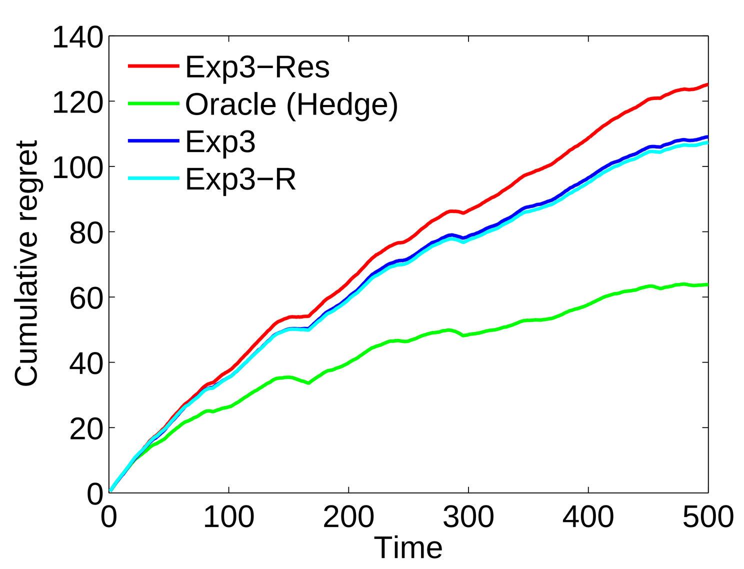
$$(1 - o_{t,i})^{N-1} \le (1 - r_t)^{N-1} \le e^{-r_t(N-1)} \le \frac{1}{\sqrt{T}}$$

Theoretical guaranties

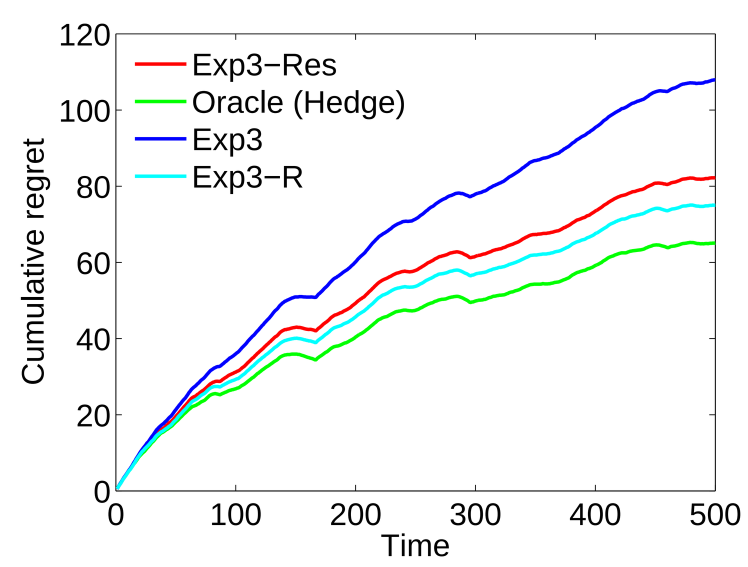
Theorem 1. *Assume that* $r_t \ge \log(T)/(2N-2)$ *for all* t*. Then, the expected regret of* Exp3-Res *satisfies*

$$\mathbb{E}[R_T] \le 2\sqrt{\left(N^2 + \sum_{t=1}^{T} \frac{1}{r_t}\right) \log N} + \sqrt{T} = \widetilde{\mathcal{O}}\left(\sqrt{\frac{T}{\bar{r}}}\right)$$

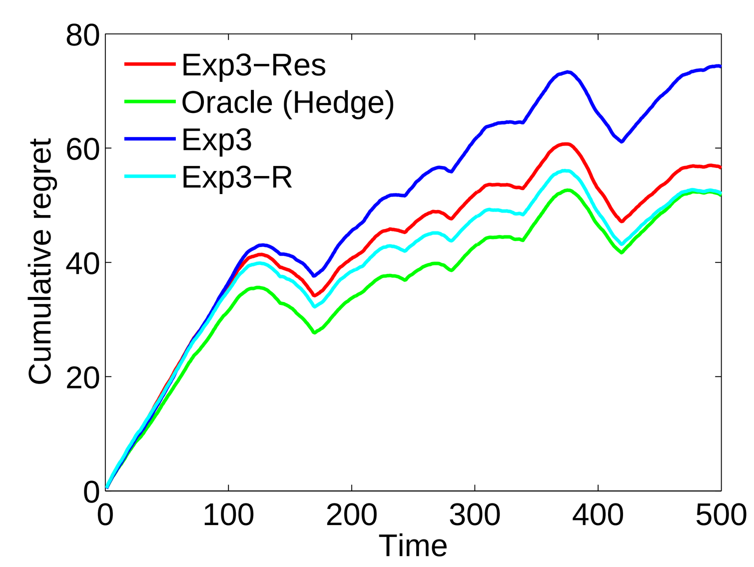
Empirical results



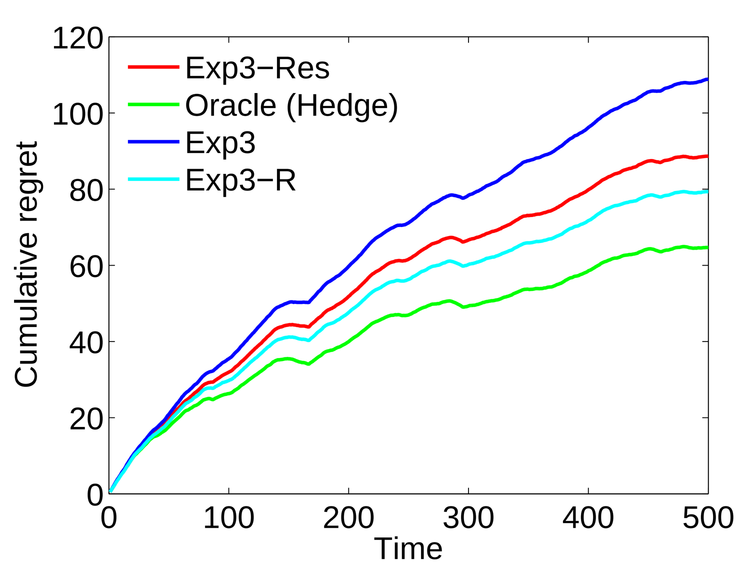
(a) Static sequence $(r_t)_t^T$, $r_t = 0$



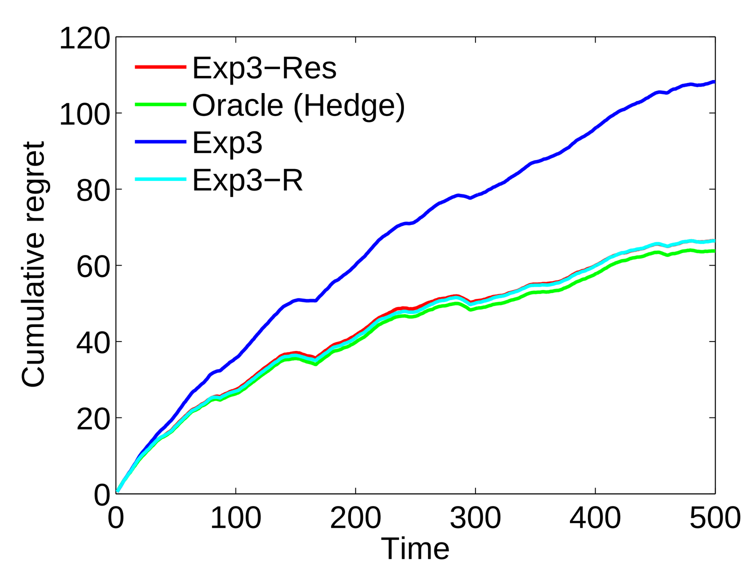
(b) Static sequence $(r_t)_t^T$, $r_t = 0.06 \approx \log(T)/(2N-2)$



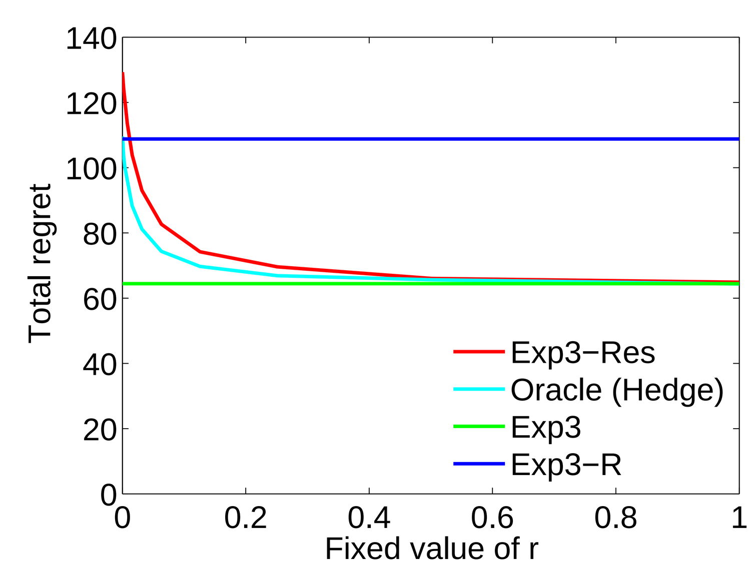
(c) Changing sequence $(r_t)_t^T$ with uniformly distributed r_t on $[0, 0.2]$



(d) Sequence $(r_t)_t^T$ generated as a random walk on $[0, 0.1]$



(e) Sequence $(r_t)_t^T$ generated as a random walk on $[0, 1]$



(f) Total regret for different values of static $(r_t)_t^T$

Figure 1: Comparison of algorithms for different amount of side information sequences (different sequences $(r_t)_t^T$)

Remark and future work

- Exp3-Res performs almost as well as the algorithm (Exp3-R) which knows exact values of r_t at any time t

Open problems:

- Is there an algorithm for $r_t < \log(T)/(2N-2)$?
- IS there a way to generalize the algorithm for different graph models?

References

Alon, N., Cesa-Bianchi, N., Gentile, C., and Mansour, Y. (2013). From bandits to experts: A tale of domination and independence. In *Neural Information Processing Systems*.

Cohen, A., Hazan, T., and Koren, T. (2016). Online learning with feedback graphs without the graphs. In *International Conference on Machine Learning (to appear)*.

Kocák, T., Neu, G., and Valko, M. (2016). Online learning with noisy side observations. In *International Conference on Artificial Intelligence and Statistics*, pages 1186–1194.

Kocák, T., Neu, G., Valko, M., and Munos, R. (2014). Efficient learning by implicit exploration in bandit problems with side observations. In *Neural Information Processing Systems*, pages 613–621.

Mannor, S. and Shamir, O. (2011). From bandits to experts: On the value of side-observations. In *Neural Information Processing Systems*.