Seminár z teoretickej informatiky
Fakulta matematiky, fyziky a informatiky UK, Bratislava
September 22, 2016

WHERE IS JUSTIN BIEBER?

Alexandra Carpentier, Institut für Mathematik an der Universität Potsdam
Michal Valko, SequeL, Inria Lille - Nord Europe

HOW TO RULE THE WORLD?

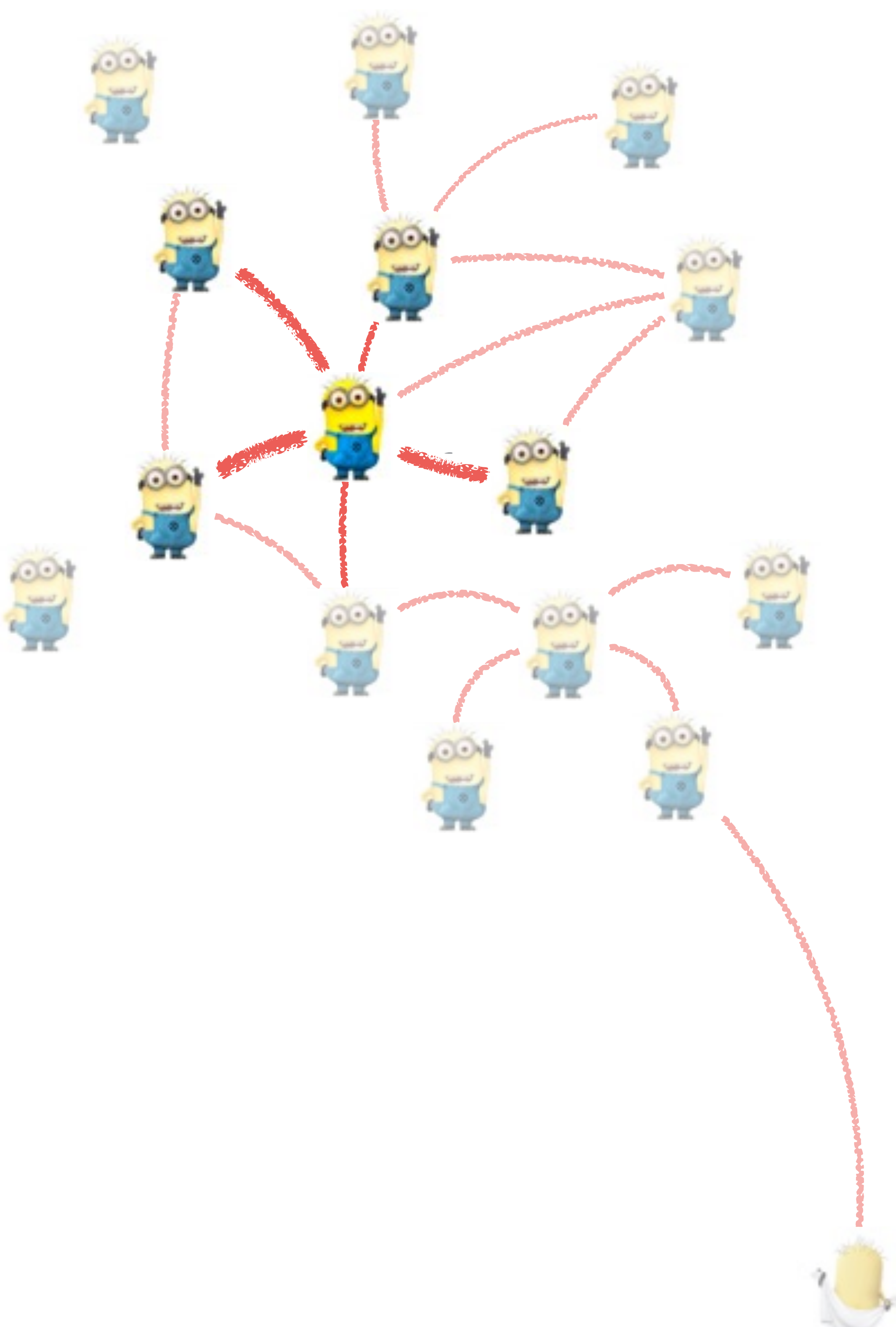
Influence the influential!

Religion

Politics

Culture

MAXIMIZING INFLUENCE



Product placement

- dispatch few to sell more
- target influential people

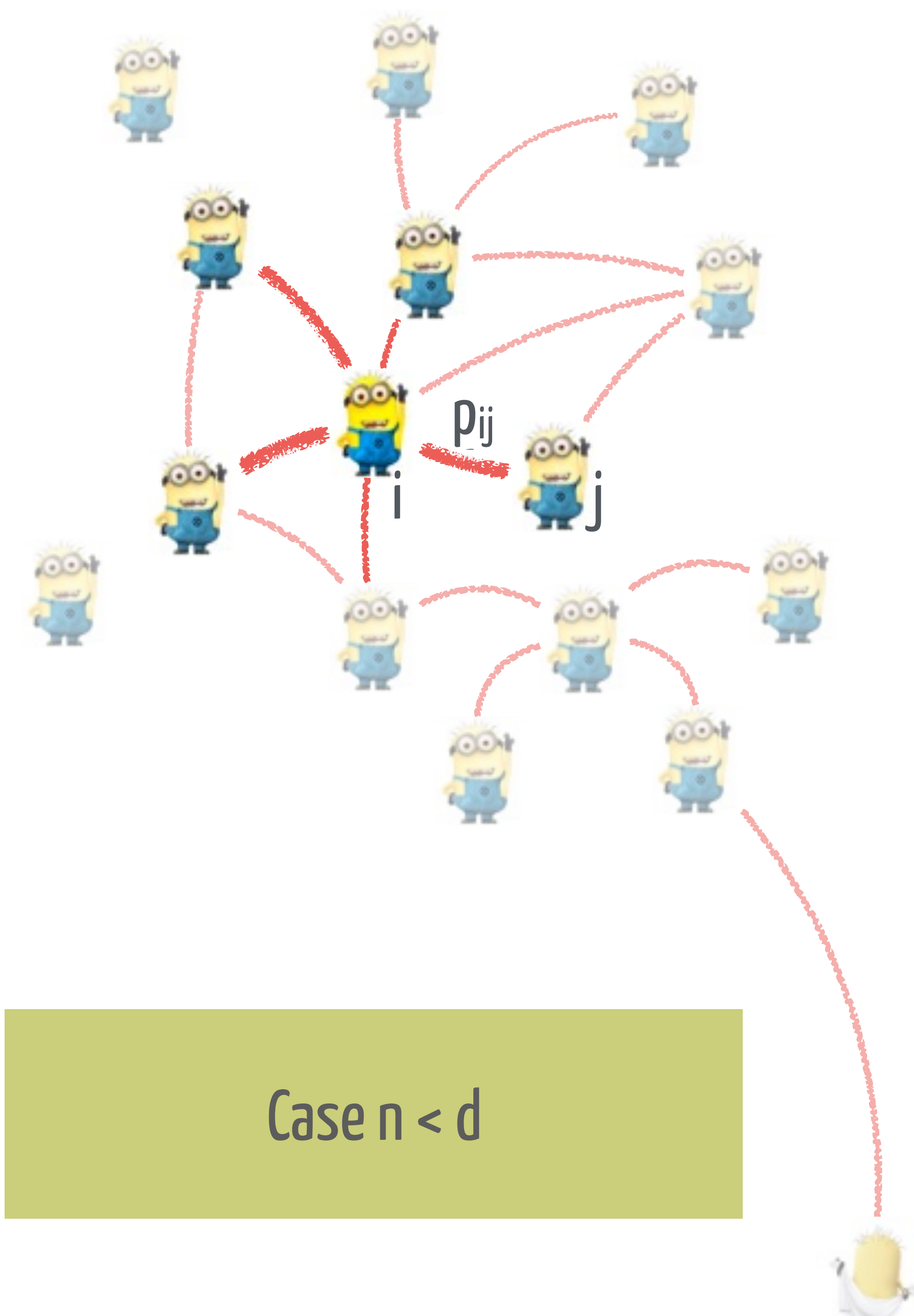
Gathering the information?

- likes on FB
- promotional codes

Unknown graphs

- all prior work needed to know the graph
- here: provably learning faster without it

REVEALING BANDITS FOR LOCAL INFLUENCE



Unknown $(p_{ij})_{ij}$ — (symmetric) probability of influences

In each time step t = 1, …., n

learner picks a node k_t

environment **reveals** the set of influenced node S_{k_t}

Select influential people = Find the strategy maximising

$$L_n = \sum_{t=1}^{n} |S_{k_t,t}|$$

What this is a bandit problem?

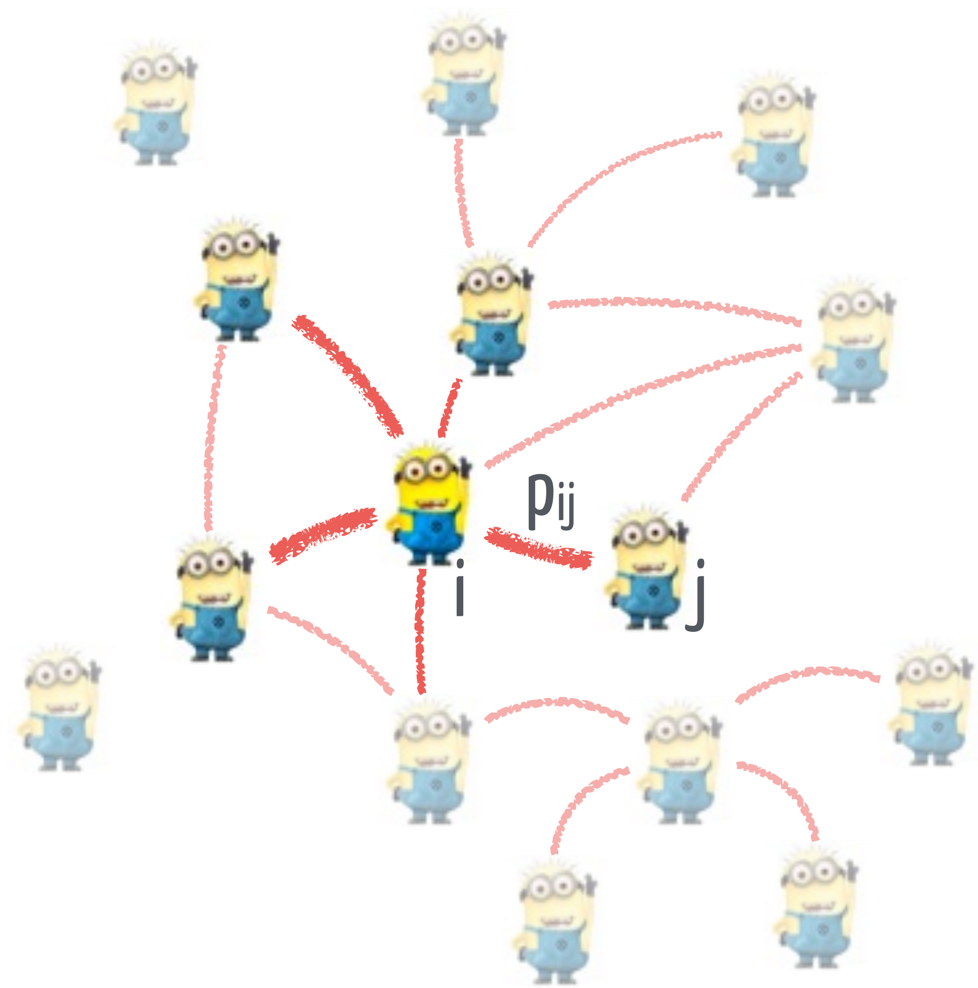
Case n < d

What are bandits anyway?

PERFORMANCE CRITERION

The number of expected influences of node **k** is by definition

$$r_k = \mathbb{E}\left[|S_{k,t}|\right] = \sum_{j \leq d} p_{k,j}$$



Oracle strategy always selects the best

$$k^\star = \arg\max_k \mathbb{E}\left[\sum_{t=1}^{n} |S_{k,t}|\right] = \arg\max_k n r_k$$

Expected regret of the oracle strategy

$$\mathbb{E}\left[L_n^\star\right] = n r_\star$$

Expected regret of any adaptive strategy unaware of $(p_{ij})_{ij}$

$$\mathbb{E}\left[R_n\right] = \mathbb{E}\left[L_n^\star\right] - \mathbb{E}\left[L_n\right]$$

BASELINE

- We **only** receive |S| instead of S
- Can be mapped to **multi-arm** bandits
  - rewards are 0, ..., d
  - variance bounded with r_{kt}
- We adapt **MOSS** to GraphMOSS
- Regret upper bound of GraphMOSS

$$\mathbb{E}\left[R_n\right] \leq U \min\left(r_\star n, r_\star d + \sqrt{r_\star n d}\right)$$

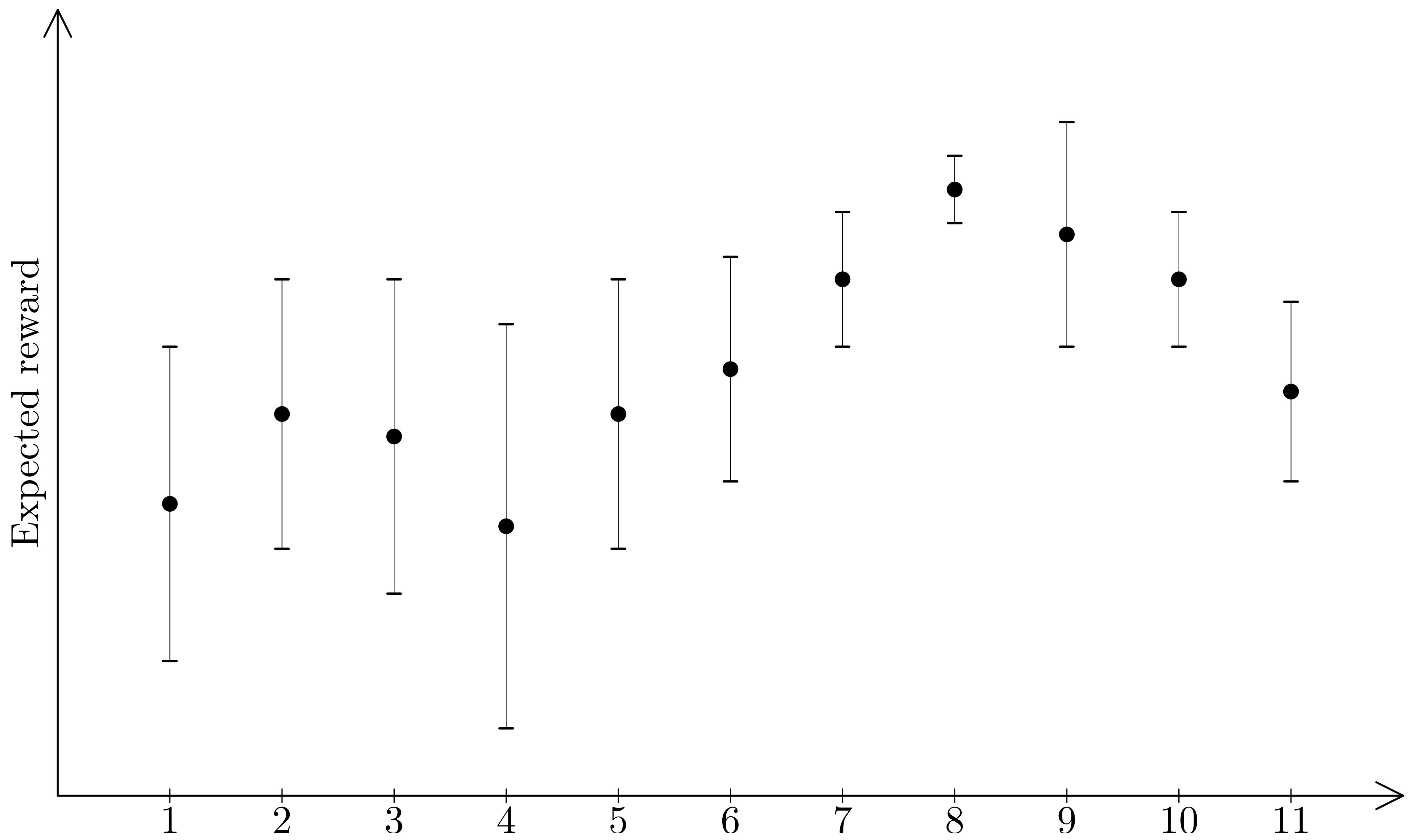
each node at least once

unlearnable case n ≤ d

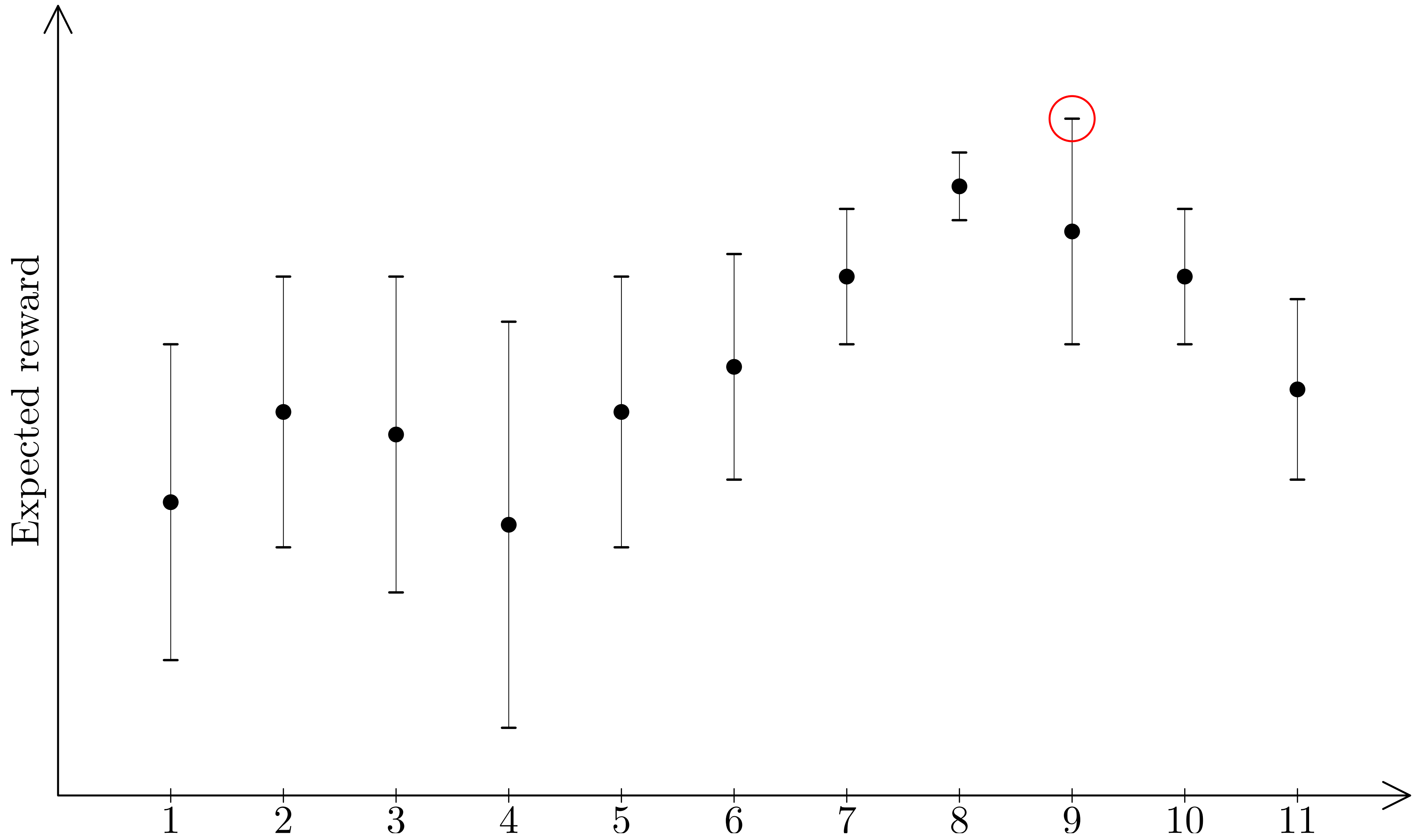
- matching lower bound

Crash course on stochastic bandits?

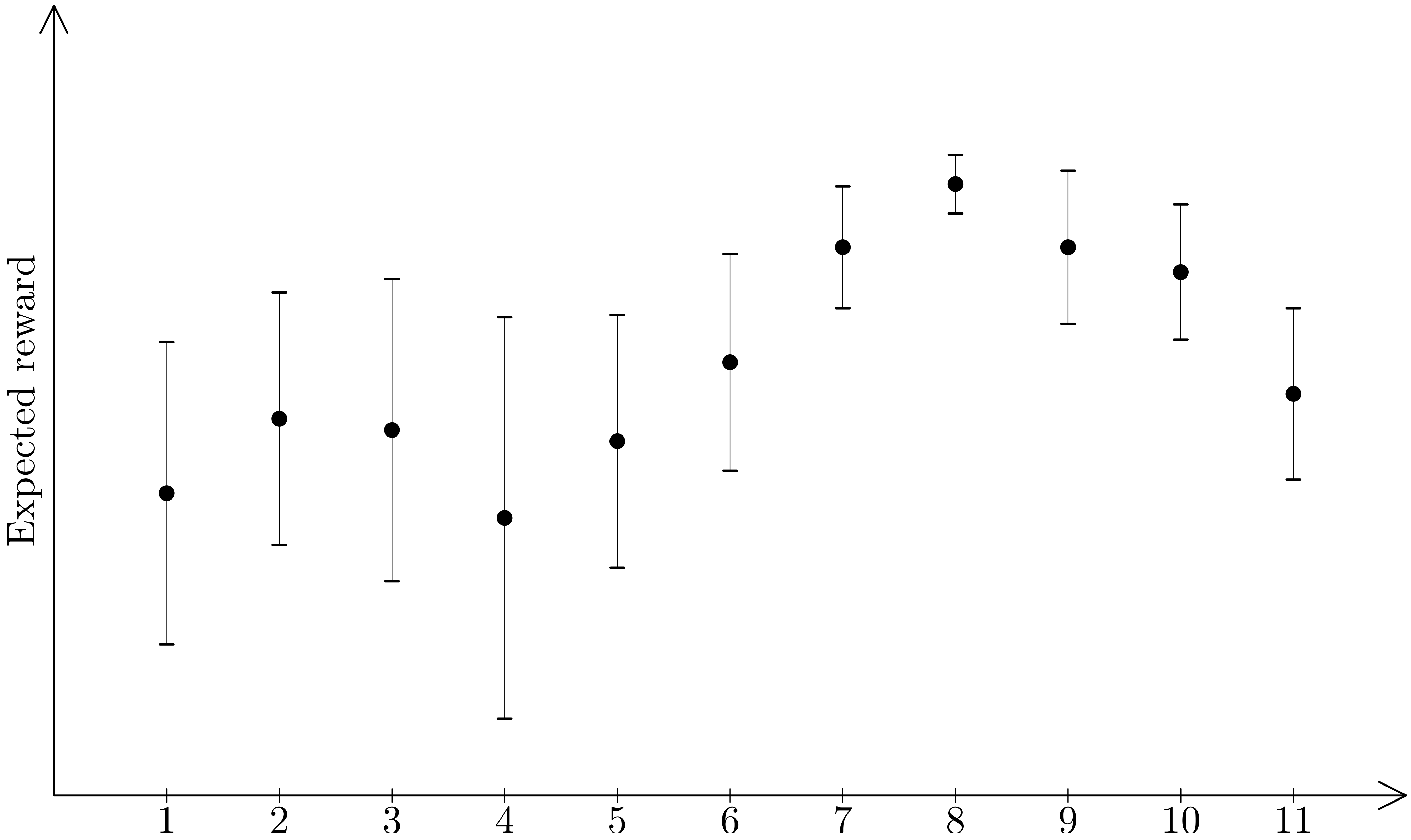
UPPER CONFIDENCE BOUND BASED ALGOS



UPPER CONFIDENCE BOUND BASED ALGOS



UPPER CONFIDENCE BOUND BASED ALGOS



GRAPHMOSS FOR THE RESTRICTED SETTING

GraphMOSS

Input

d: the number of nodes

n: time horizon

Initialization

Sample each arm twice

Update $\widehat{r}_{k,2d}$, $\widehat{\sigma}_{k,2d}$, and $T_{k,2d} \leftarrow 2$, for $\forall k \leq d$

for $t = 2d+1, \ldots, n$ **do**

$$C_{k,t} \leftarrow 2\widehat{\sigma}_{k,t}\sqrt{\frac{\max(\log(n/(dT_{k,t})),0)}{T_{k,t}}} + \frac{2\max(\log(n/(dT_{k,t})),0)}{T_{k,t}}, \text{ for } \forall k \leq d$$

$k_t \leftarrow \arg\max_k \widehat{r}_{k,t} + C_{k,t}$

Sample node k_t and receive $|S_{k_t,t}|$

Update $\widehat{r}_{k,t+1}$, $\widehat{\sigma}_{k,t+1}$, and $T_{k,t+1}$, for $\forall k \leq d$

end for

$$\mathbb{E}[R_n] \leq U \min\left(r_\star n, r_\star d + \sqrt{r_\star n d}\right)$$

each node at least once

unlearnable case n ≤ d

BACK TO THE REAL SETTING

- Can we actually do better?
  - Well, not really.....
  - Minimax optimal rate is still the same
- But the bad cases are somehow pathological
  - isolated nodes
  - uncorrelated being influenced and being influential
  - Barabási–Albert etc tell us that the real-world graphs are not like that
- Let's think of some measure of difficulty
  - to define some non-degenerate cases
  - ideas?

DETECTABLE DIMENSION

- number of nodes we can efficiently extract in less than n rounds
- function D controls number of nodes given a gap

$$D(\Delta) \stackrel{\text{def}}{=} |\{i \leq d : r_\star^\circ - r_i^\circ \leq \Delta\}|$$

- D(r) = d for r≥ r_* and D(0) = number of most influenced nodes
- **Detectable dimension D_*** = $D(\Delta_*)$
- Detectable gap Δ_* constants coming from the analysis and the Bernstein inequality

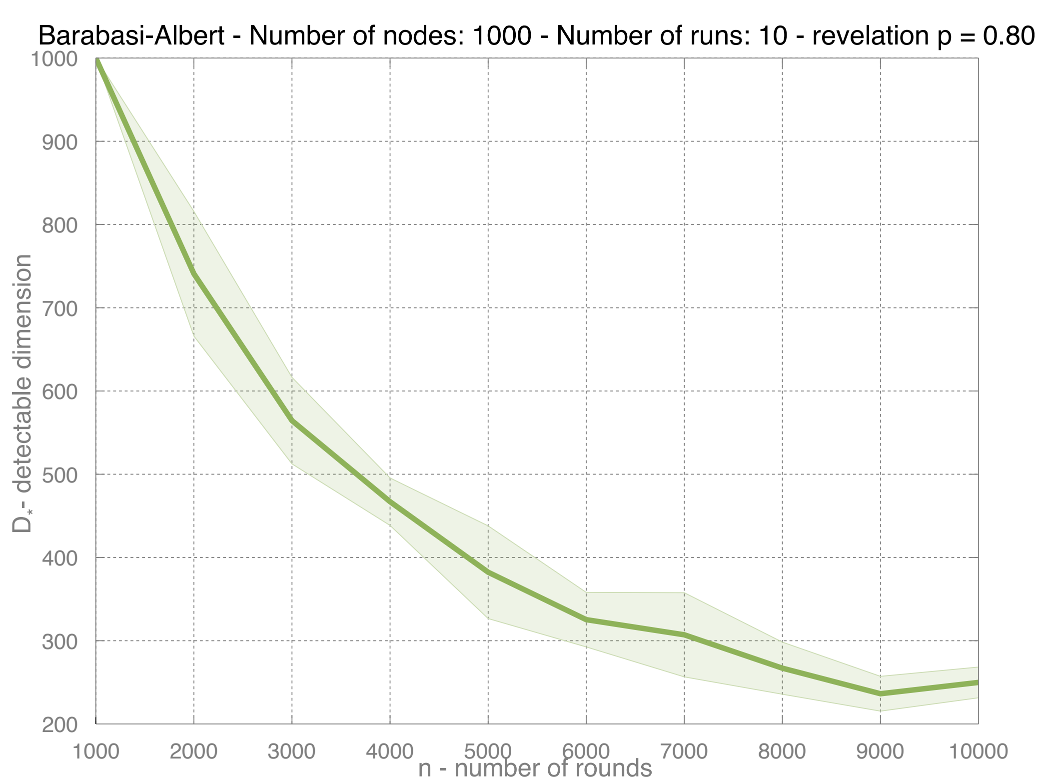
$$\Delta_\star \stackrel{\text{def}}{=} 16\sqrt{\frac{r_\star^\circ d \log(nd)}{T_\star}} + \frac{80 d \log(nd)}{T_\star}$$

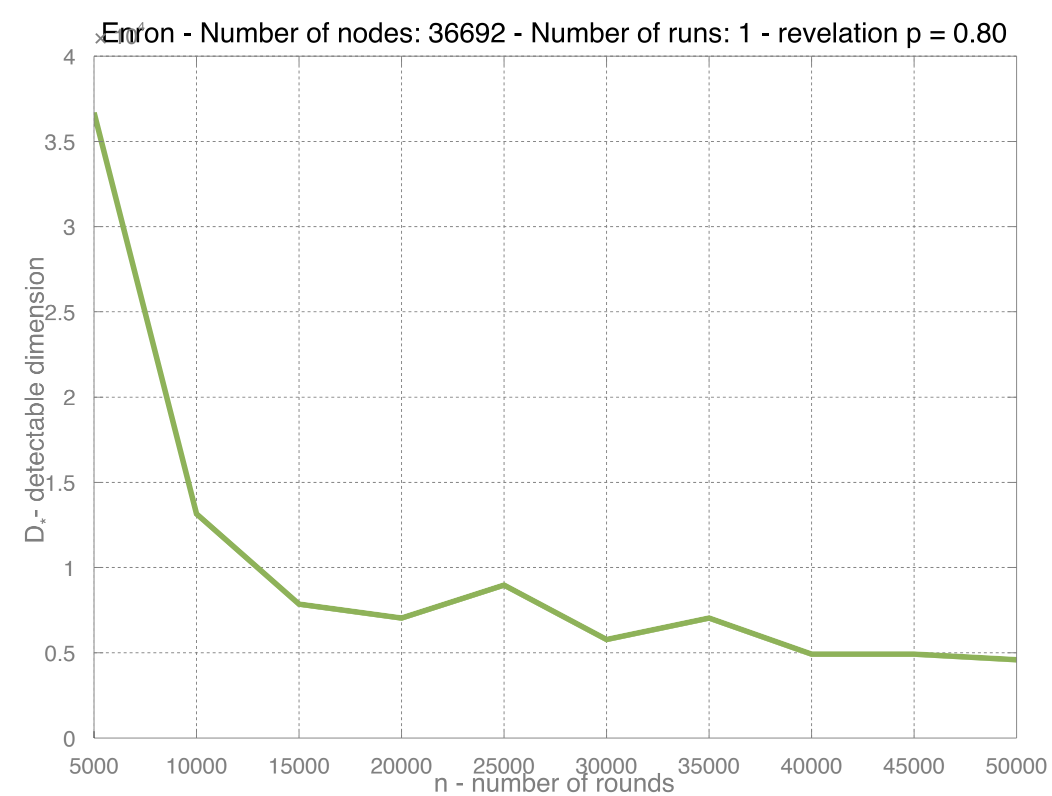
- Detectable horizon T_*, smallest integer s.t. $T_\star r_\star^\circ \geq \sqrt{D_\star n r_\star^\circ}$
- Equivalently: D_* corresponding to smallest T_* such that

$$T_\star r_\star^\circ \geq \sqrt{D\left(16\sqrt{\frac{r_\star^\circ d \log(nd)}{T_\star}} + \frac{80 d \log(nd)}{T_\star}\right) n r_\star^\circ}$$

HOW DOES D* BEHAVE?

- For (easy, structured) star graphs $D_* = 1$ even for small n (big gain)
- For (difficult) empty graphs $D_* = d$ even for large n (no gain)
- In general: D_* roughly decreases with n and it is small when D decreases quickly
- For n large enough D_* is the number of the most influences nodes
- Example: D_* for Barabási–Albert model & Enron graph as a function of n





OUR SOLUTION

BAndit REvelator: 2-phase algorithm

- **global** exploration phase
  - super-efficient exploration 😸
  - linear regret 😿 — needs to be short!
  - extracts **D*** nodes
- **bandit** phase
  - uses a minimax-optimal bandit algorithm
  - GraphMOSS is a little brother of MOSS
  - has a “square root” regret on **D*** nodes
- **D* realizes the optimal trade-off !**
  - different from exploration/exploitation tradeoff

Algorithm

BARE - BAndit REvelator

Input

d: the number of nodes

n: time horizon

Initialization

$T_{k,t} \leftarrow 0$, for $\forall k \leq d$

$\widehat{r^{\circ}_{k,t}} \leftarrow 0$, for $\forall k \leq d$

$t \leftarrow 1$, $\widehat{T}_{\star} \leftarrow 0$, $\widehat{D}_{\star,t} \leftarrow d$, $\widehat{\sigma}_{\star,1} \leftarrow d$

Global exploration phase

while $t\left(\widehat{\sigma}_{\star,t} - 4\sqrt{d\log(dn)/t}\right) \leq \sqrt{\widehat{D}_{\star,t} n}$ **do**

Influence a node at random (choose k_t uniformly at random) and get $S_{k_t,t}$ from this node

$\widehat{r^{\circ}_{k,t+1}} \leftarrow \frac{t}{t+1}\widehat{r^{\circ}_{k,t}} + \frac{d}{t+1} S_{k_t,t}(k)$

$\widehat{\sigma}_{\star,t+1} \leftarrow \max_{k'} \sqrt{\widehat{r^{\circ}_{k',t+1}} + 8d\log(nd)/(t+1)}$

$w_{\star,t+1} \leftarrow 8\widehat{\sigma}_{\star,t+1}\sqrt{\frac{d\log(nd)}{t+1}} + \frac{24d\log(nd)}{t+1}$

$\widehat{D}_{\star,t+1} \leftarrow \left|\left\{k : \max_{k'} \widehat{r^{\circ}_{k',t+1}} - \widehat{r^{\circ}_{k,t+1}} \leq w_{\star,t+1}\right\}\right|$

$t \leftarrow t+1$

end while

$\widehat{T}_{\star} \leftarrow t$.

Bandit phase

Run minimax-optimal bandit algorithm on the $\widehat{D}_{\star,\widehat{T}_{\star}}$ chosen nodes (e.g., Algorithm 1)

EMPIRICAL RESULTS

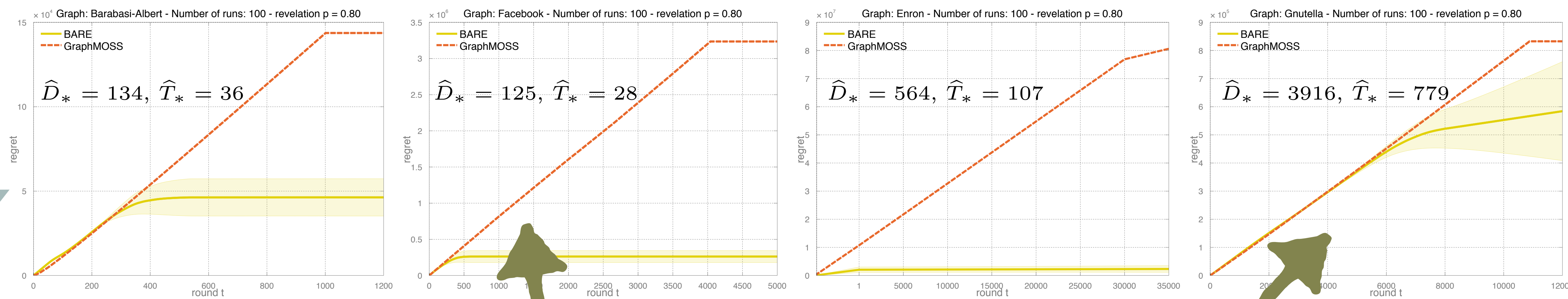
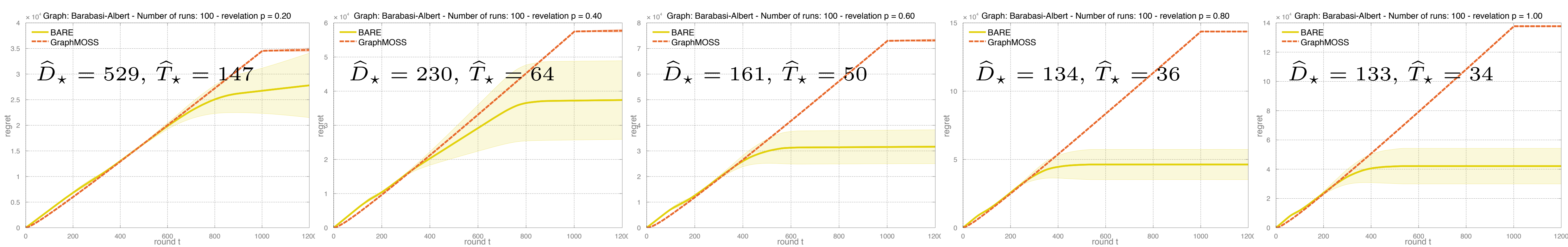


Figure 1: *Left*: Barabási-Albert. *Middle left*: Facebook. *Middle right*: Enron. *Right*: Gnutella.

Enron and Facebook **vs.** Gnutella (decentralised)



Varying a (constant) probability of influence

REVEALING BANDITS: WHAT DO YOU MEAN?

- Ignoring the structure again? $\tilde{\mathcal{O}}\left(\sqrt{r_* T N}\right)$

reward of the best node

- **BAndit REvelator:** 2-phase algorithm
- **global** exploration phase
  - super-efficient exploration
  - linear regret — needs to be short!
  - extracts D_* nodes
- **bandit** phase
  - uses a minimax-optimal bandit algorithm (GraphMOSS)
  - has a "square root" regret on D_* nodes
- D_* realizes the optimal trade-off !
  - different from exploration/exploitation tradeoff

Regret of BARE

$$\tilde{\mathcal{O}}\left(\sqrt{r_* T D_*}\right)$$

PROOF TIME!

- **D*** - detectable dimension (depends on T and the structure)
  - **good case**: star-shaped graph can have D* = 1
  - **bad case:** a graph with many small cliques.
  - **the worst case:** all nodes are disconnected except 2

PROOF: ESTIMATES IN THE GLOBAL PHASE

Reward estimate

$$\widehat{r^\circ_{k,t}} = \frac{1}{t}\sum_{t'=1}^{t} dS_{k_t,t'}(k)$$

Bernstein: w.p. $1-1/n^2$

$$\left|\widehat{r^\circ_{k,t}} - r^\circ_k\right| \leq 4\sqrt{\frac{dr^\circ_k \log(nd)}{t}} + \frac{4d\log(nd)}{t}$$

On this event (ξ)

$$\left|\sqrt{r^\circ_k} - \sqrt{\widehat{r^\circ_{k,t}} + \frac{8d\log(nd)}{t}}\right| \leq 4\sqrt{\frac{d\log(nd)}{t}}$$

And in particular

$$\left|\widehat{\sigma}_{\star,t} - \sqrt{r^\circ_\star}\right| \leq 4\sqrt{\frac{d\log(nd)}{t}}$$

PROOF: ESTIMATES IN THE GLOBAL PHASE

Similarly (from the same Bernstein bound)

$$\left|\left(\max_{k'} \widehat{r^\circ_{k',t}} - \widehat{r^\circ_{k,t}}\right) - (r^\circ_\star - r^\circ_k)\right| \leq 8\sqrt{\frac{d r^\circ_\star \log(nd)}{t}} + \frac{8d\log(nd)}{t}$$

Plugging in the upper bound ("variance") on the best reward

$$\left|\left(\max_{k'} \widehat{r^\circ_{k',t}} - \widehat{r^\circ_{k,t}}\right) - (r^\circ_\star - r^\circ_k)\right| \leq 8\widehat{\sigma}_{\star,t}\sqrt{\frac{d\log(nd)}{t}} + \frac{40d\log(nd)}{t}$$

On ξ, BARE will keep the most influenced nodes (and maybe some more)

$$\widehat{D}_{\star,t} \leq D\left(16\widehat{\sigma}_{\star,t}\sqrt{\frac{d\log(nd)}{t}} + \frac{80d\log(nd)}{t}\right) \leq D\left(16\sqrt{\frac{d r^\circ_\star \log(nd)}{t}} + \frac{144d\log(nd)}{t}\right).$$

PROOF: WHEN DOES THE GLOBAL PHASE END?

CASE 1: **Before 3T***

By BARE

$$3T_\star\sqrt{r^\circ_\star} \geq \widehat{T}_\star\sqrt{r^\circ_\star} \geq \sqrt{\widehat{D}_{\star,\widehat{T}_\star} n}.$$

By def of D*

$$\sqrt{D_\star n r^\circ_\star} \geq (T_\star - 1) r^\circ_\star \geq T_\star r^\circ_\star / 2,$$

Together

$$\widehat{D}_{\star,\widehat{T}_\star} \leq 36 D_\star$$

... and the optimal arm is among the kept ones.

PROOF: WHEN DOES THE GLOBAL PHASE END?

CASE 2: **After 3T*** (we show that this cannot happen, by contradiction)

By BARE (the "if" part)
$$3T_\star\left(\widehat{\sigma}_{\star,T^\star} - 4\sqrt{\frac{d\log(nd)}{T_\star}}\right) \le \sqrt{\widehat{D}_{\star,T_\star} n}$$

As before, we have
$$3T_\star\left(\sqrt{r_\star^\circ} - 8\sqrt{\frac{d\log(nd)}{T_\star}}\right) \le \sqrt{\widehat{D}_{\star,T_\star} n}$$

Which implies
$$\frac{3T_\star\sqrt{r_\star^\circ}}{3} \le \sqrt{\widehat{D}_{\star,T_\star} n}$$

$$T_\star\sqrt{r_\star^\circ} \le \sqrt{D\left(16\sqrt{\frac{d r_\star^\circ \log(nd)}{T_\star}} + \frac{144 d\log(nd)}{T_\star}\right) n}$$

$\le \sqrt{D_\star n},$ which is false by definition of D* and T*

PROOF: FINISH

In both cases:

$$\widehat{T}_\star \leq 3T_\star \quad \text{and} \quad \widehat{D}_{\star,\widehat{T}_\star} \leq 36D_\star$$

We upper bound the regret on the remaining rounds:

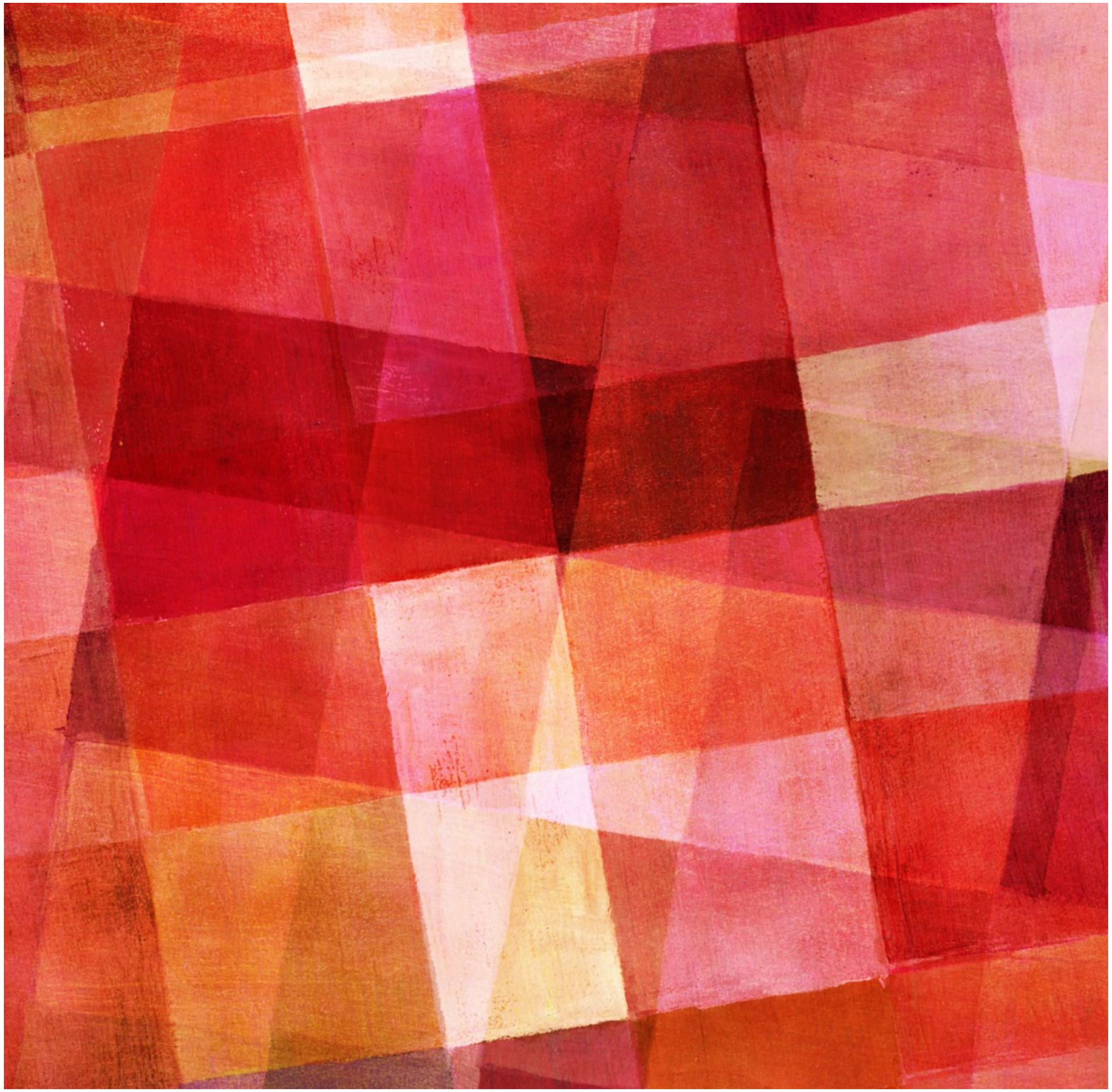
$$C'D_\star r_\star + C'\sqrt{r_\star \widehat{D}_{\star,\widehat{T}_\star}\left(n - \widehat{T}_\star\right)} \leq C'D_\star r_\star + C'\sqrt{r_\star D_\star n}$$

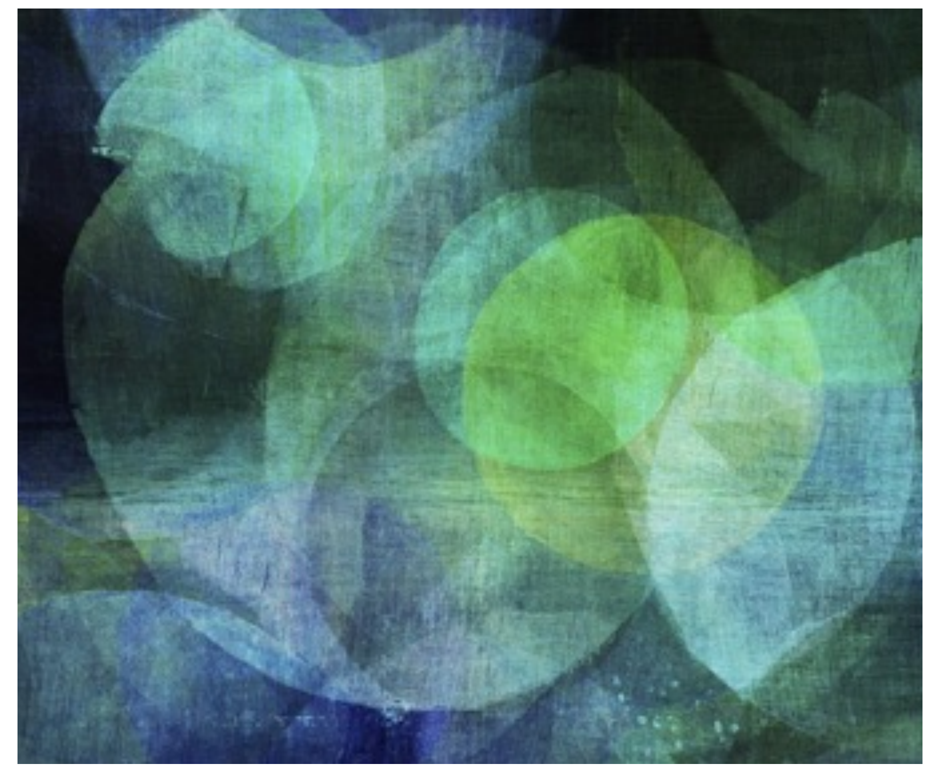
Final regret

$$\begin{aligned}\mathbb{E}\left[R_n\right] &\leq T_\star r_\star + C'D_\star r_\star + C'\sqrt{r_\star D_\star n} + n\varepsilon_\star + \frac{r_\star}{n} \\ &\leq (C' + 2)\sqrt{r_\star D_\star n} + 2C'r_\star D_\star + n\varepsilon_\star \\ &\leq C\left(\sqrt{r_\star D_\star n} + r_\star D_\star\right) + n\varepsilon_\star.\end{aligned}$$

NEXT: GLOBAL INFLUENCE MODELS

- Kempe, Kleinberg, Tárdos, 2003, 2015: **Independence Cascades**, Linear Threshold models
  - **global** and **multiple-source** models
- Different feed-back models
  - **Full bandit** (only the number of influenced nodes)
  - **Node-level semi-bandit** (identities of influenced nodes)
  - **Edge-level semi-bandit** (identities of influenced edges)
    - http://arxiv.org/abs/1605.06593 (Wen, Kveton, MV)
    - IMLinUCB with linear parametrization of edge weights
    - Regret analysis for subset of graphs (forests, ...)





ExtraLearn

Michal Valko, SequeL, Inria Lille - Nord Europe, michal.valko@inria.fr
http://researchers.lille.inria.fr/~valko/hp/