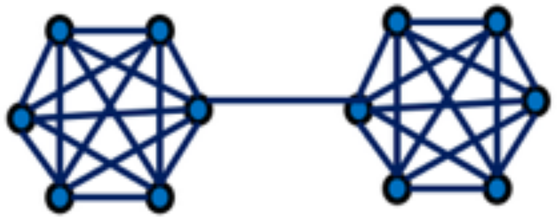
GRAPHS IN MACHINE LEARNING

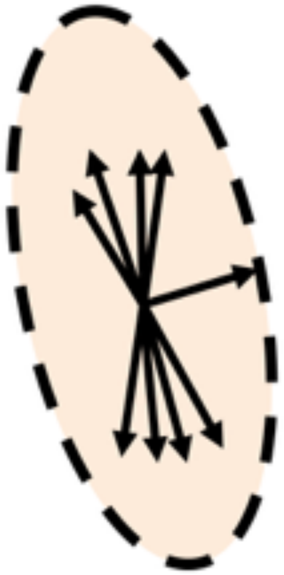
Michal Valko, SequeL, Inria Lille - Nord Europe

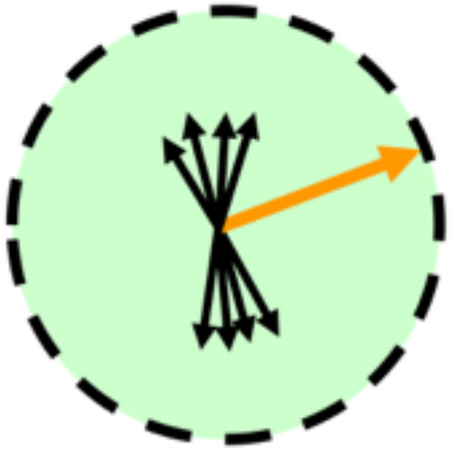
TA: **Pierre Perrault**

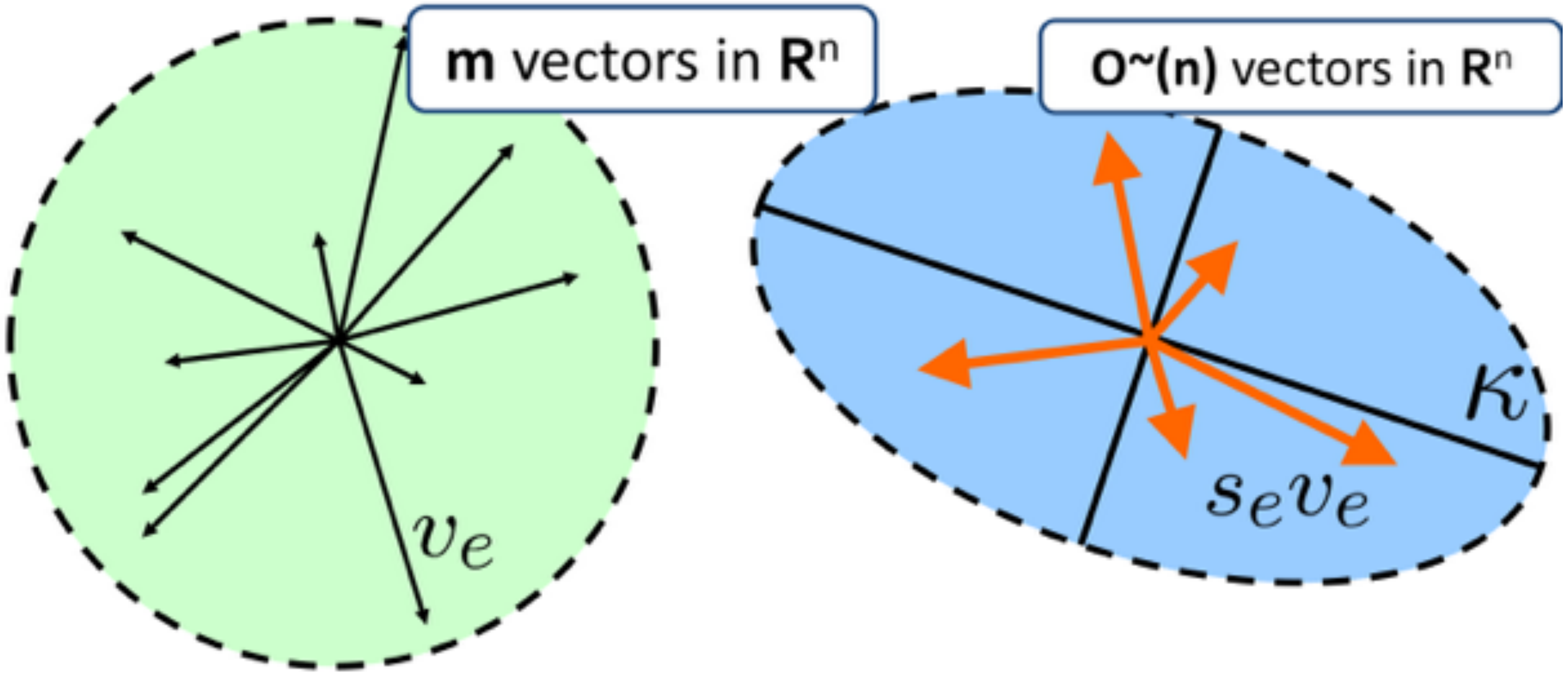
MVA 2017/2018 Partially based on material by Tomáš Kocák

PREVIOUS LECTURE BY PIERRE









Questions?

image credits: Srivastava

THIS LECTURE LAST LECTURE OF THE COURSE

- Graph bandits
  - Spectral bandits
  - Observability graphs
  - Side information
  - Influence Maximization

RL/BANDITS ~ SEQUENTIAL DECISION-MAKING

unsupervised - supervised-semisupervised-active

MULTI-ARM BANDITS IN LAS VEGAS
DECEMBER 2017

ps: several course projects are on this topic

PREVIOUS LAB SESSION

27. 11. 2017 by Pierre Perrault

Content (this time lecture in class + coding at home)

- Large-scale graph construction and processing (in class
- Scalable algorithms:
  - Online face recognizer (to code in Matlab)
  - Iterative label propagation (to code in Matlab)
  - Graph sparsification (presented in class)

AR: record a video with faces

Short written report

Questions to piazza

Deadline: 11. 12. 2017 (today)

FINAL CLASS PROJECTS

- time and formatting description on the class website
- grade: report + short presentation of the **team**
- deadlines
  - **8. 1. 2018** final report (for all projects)
  - from 9. 1. 2018, presentations (mostly over Skype/Hangout)
  - AR: sign-up for the presentation (info already there)
- project report: 5-10 pages in NIPS format
- presentation: 15+5 minutes (**time it!**)
- everybody has to present
- book presentation time slot on the website
- **explicitly state your contributions** report + talk

PHD POSITIONS AT INRIA LILLE – MAGNET

Open **PhD positions at Inria Lille** (Magnet team). Lille is 1 hour away from Paris, 30 minutes from Brussels, 1.5 hours from London and 2.5 hours from Amsterdam.

- The topic is **decentralized machine learning.** Consider a P2P network with many devices, each with a local dataset. How can we design/analyze algorithms allowing the devices to learn from the union of their datasets without leaking too much sensitive information about individual data points?
- We also have **Master internship positions (**in decentralized/private machine learning, and natural language processing)
- Check https://team.inria.fr/magnet/job-offers/ or contact aurelien.bellet@inria.fr

PHD + INTERNSHIP AT INRIA LILLE – SEQUEL

Dynamically Evolving Long-Term Autonomy

- join project between 4 partners, UPF Barcelona, MUL Austria, ULG Belgium, and Inria
- Jonsson, Neu, Gomez, Valko, Kaufmann, Lazaric, Auer, Ortner, Cornelusse, Ernst
- **PhD position at SequeL team at Inria**
- project starts on 1.1.2018, PhD student expected to start September/October 2018
- 4 postdocs, one in each center
- Inria will lead the effort on adaptive planning with a model that can adapt to changes. Inria will work with MUL on the hierarchical state partitioning
- contact: (Emilie Kaufmann & Michal Valko) @ SequeL @ Inria

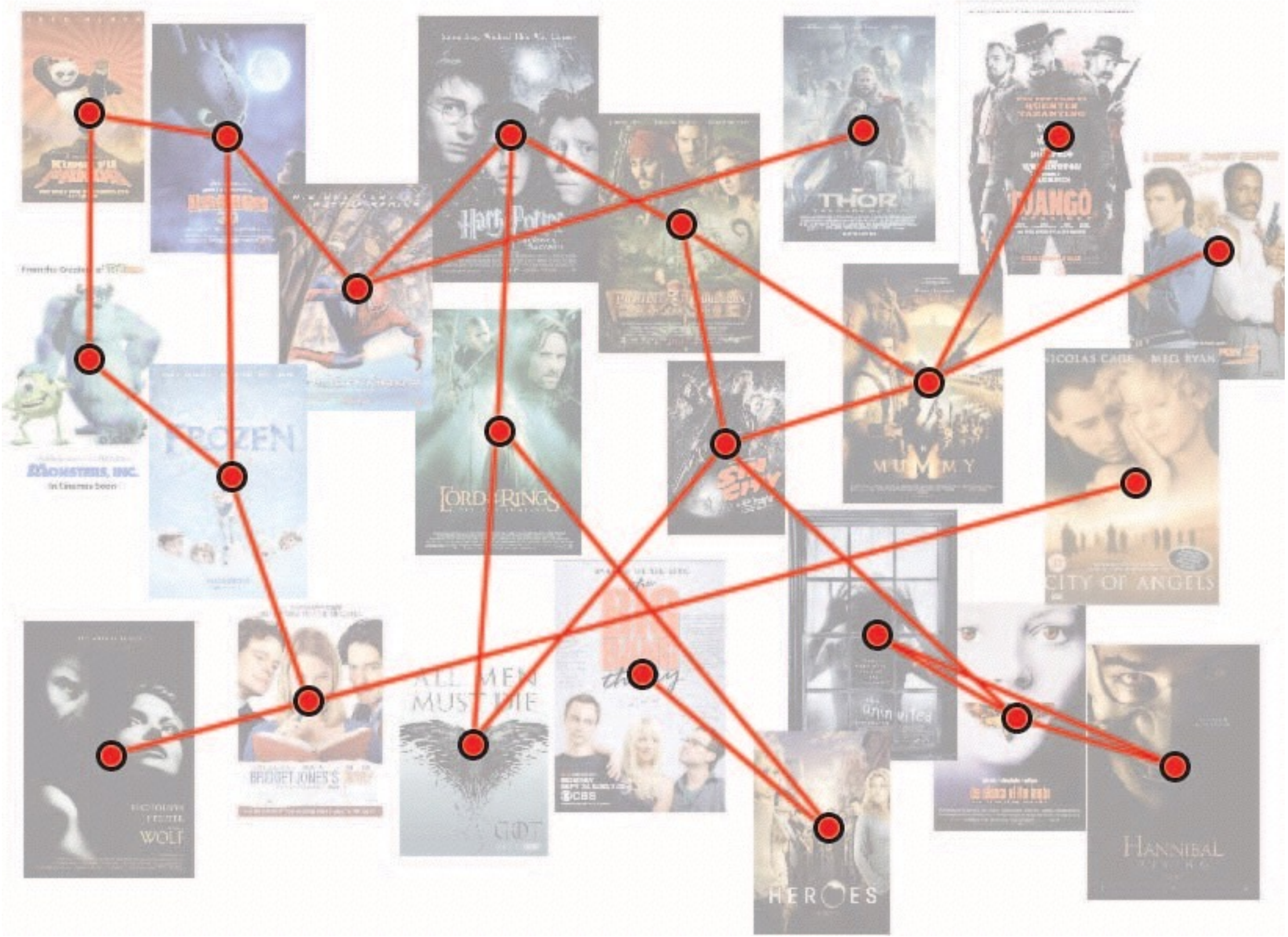
Internship position: extending **TrailBlazer** with **BAI-MCTS**

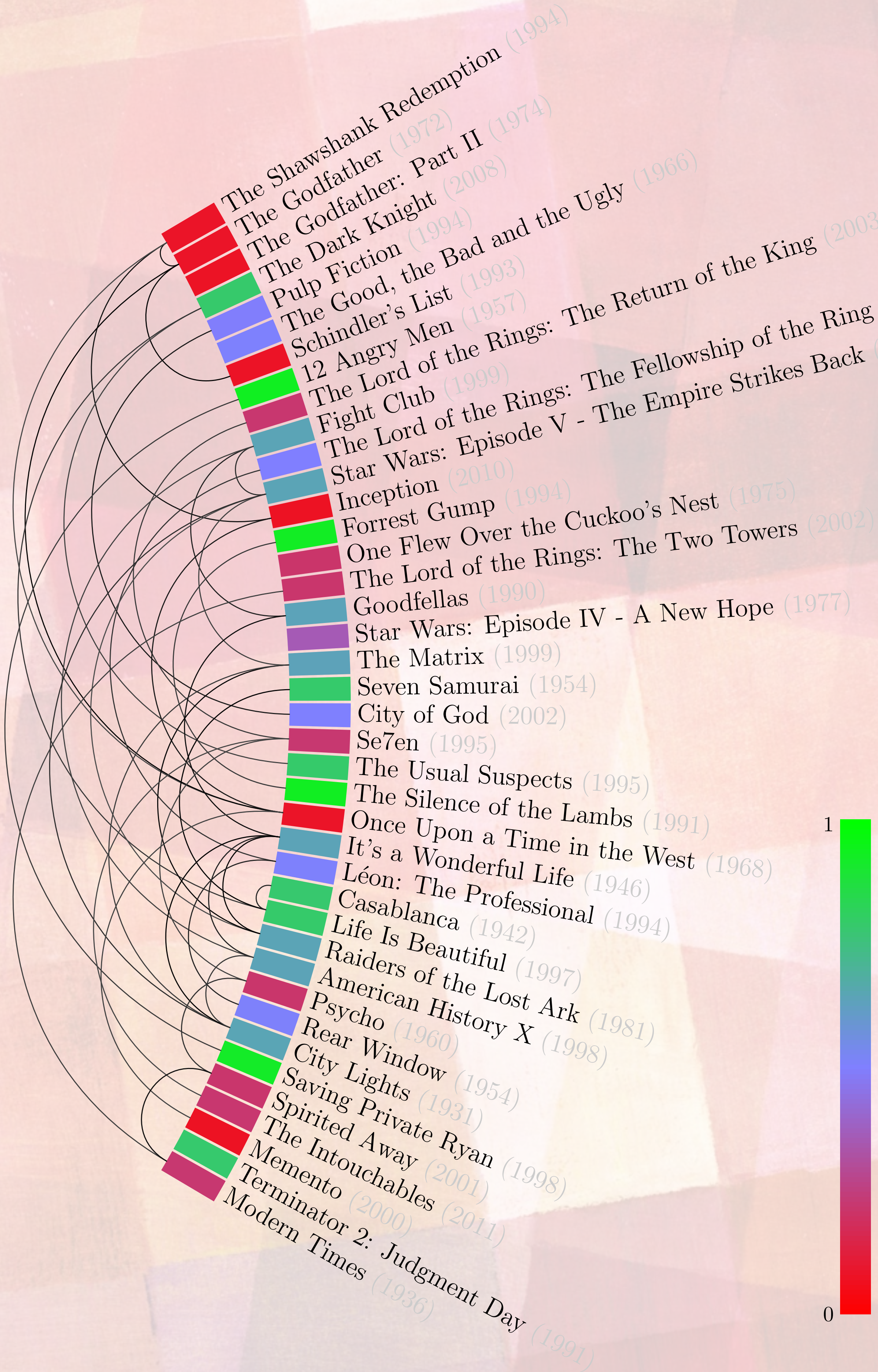
BADASS INTERNSHIPS AT LILLE AND SACLAY

https://project.inria.fr/badass/students/

- Structured set of Bandits
- Structured Bandits: Should Optimism strike back?
- Optimal exploration in Multi-armed bandits
- Computational complexity in Multi-armed Bandits
- Sensitivity analysis and intrinsic horizons in Markov Decision Processes
- Reinforcement Learning with Predictive State Representations

Please, send a message directly to the contact email provided in the document detailing each proposal.


Kung Fu Panda
Monsters, Inc.
Frozen
Harry Potter
Lord of the Rings
Thor
Django
City of Angels
Mummy
All Men Must Die
GOT
Bridget Jones's
Wolf
Heroes
Hannibal


The Shawshank Redemption (1994)
The Godfather (1972)
The Godfather: Part II (1974)
The Dark Knight (2008)
Pulp Fiction (1994)
The Good, the Bad and the Ugly (1966)
Schindler's List (1993)
12 Angry Men (1957)
The Lord of the Rings: The Return of the King (2003)
Fight Club (1999)
The Lord of the Rings: The Fellowship of the Ring (2001)
Star Wars: Episode V - The Empire Strikes Back (1980)
Inception (2010)
Forrest Gump (1994)
One Flew Over the Cuckoo's Nest (1975)
The Lord of the Rings: The Two Towers (2002)
Goodfellas (1990)
Star Wars: Episode IV - A New Hope (1977)
The Matrix (1999)
Seven Samurai (1954)
City of God (2002)
Se7en (1995)
The Usual Suspects (1995)
The Silence of the Lambs (1991)
Once Upon a Time in the West (1968)
It's a Wonderful Life (1946)
Léon: The Professional (1994)
Casablanca (1942)
Life Is Beautiful (1997)
Raiders of the Lost Ark (1981)
American History X (1998)
Psycho (1960)
Rear Window (1954)
City Lights (1931)
Saving Private Ryan (1998)
Spirited Away (2001)
The Intouchables (2011)
Memento (2000)
Terminator 2: Judgment Day (1991)
Modern Times (1936)
1
0

Example of a graph bandit problem

movie recommendation

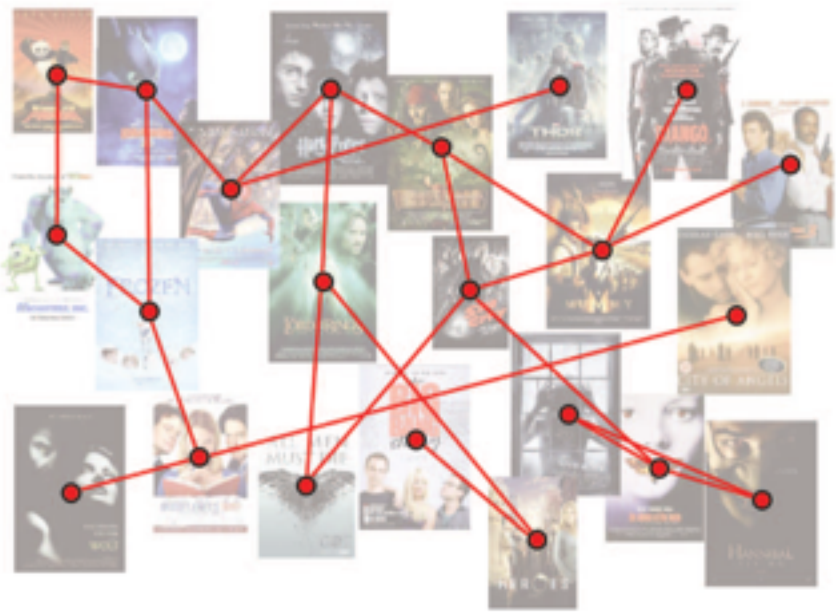
- recommend movies to a **single** user
- **goal:** maximise the sum of the ratings (minimise regret)
- good prediction after just a few steps

$$T \ll N$$

- extra information
  - ratings are **smooth** on a graph
- main question: can we learn **faster**?

GETTING REAL

Let’s be lazy and ignore the structure



→

Multi-armed bandit problem!

#actions

#rounds

Worst case regret (to the best fixed strategy)

$$R_T = \mathcal{O}\left(\sqrt{NT}\right)$$

Matching lower bound (Auer, Cesa-Bianchi, Freund, Schapire 2002)

How big is N? Number of movies on **http://www.imdb.com/stats**: 4,029,967

Problem: Too many actions!

LEARNING FASTER

$$R_T = \mathcal{O}\left(\sqrt{NT}\right)$$

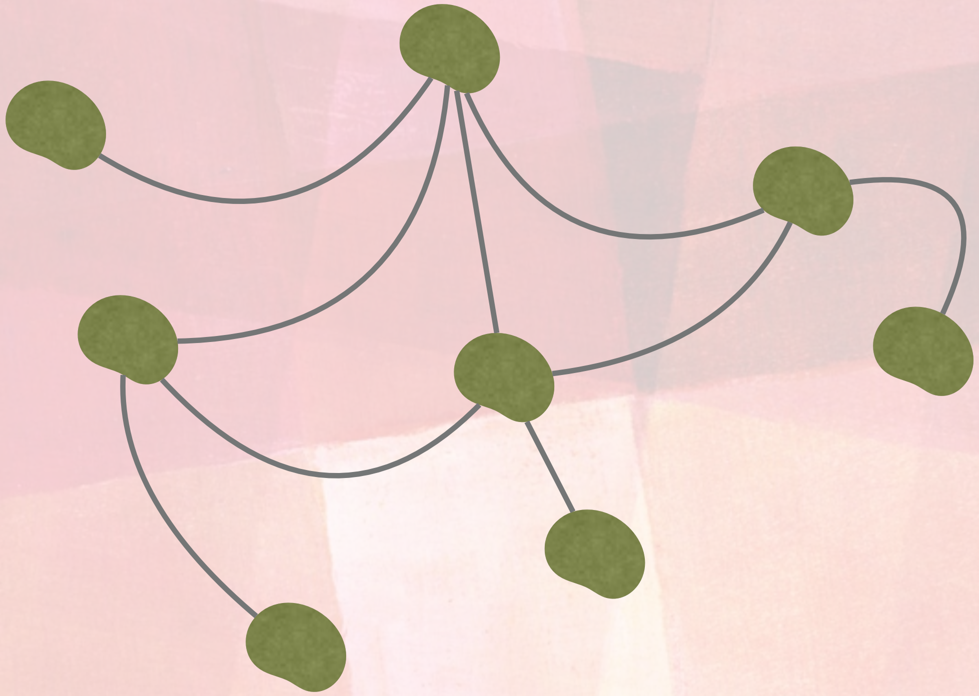
#actions

#rounds

- Arm independence is too strong and unnecessary
- Replace **N** with something much smaller
  - problem/instance/data dependent
  - example: linear bandits **N** to **D**

#dimensions

- **Today: Graph** Bandits!
  - sequential problems where **actions are nodes** on a graph
  - find strategies that replace **N** with a smaller **graph-dependent** quantity



GRAPH BANDITS: GENERAL SETUP

Every round **t** the learner

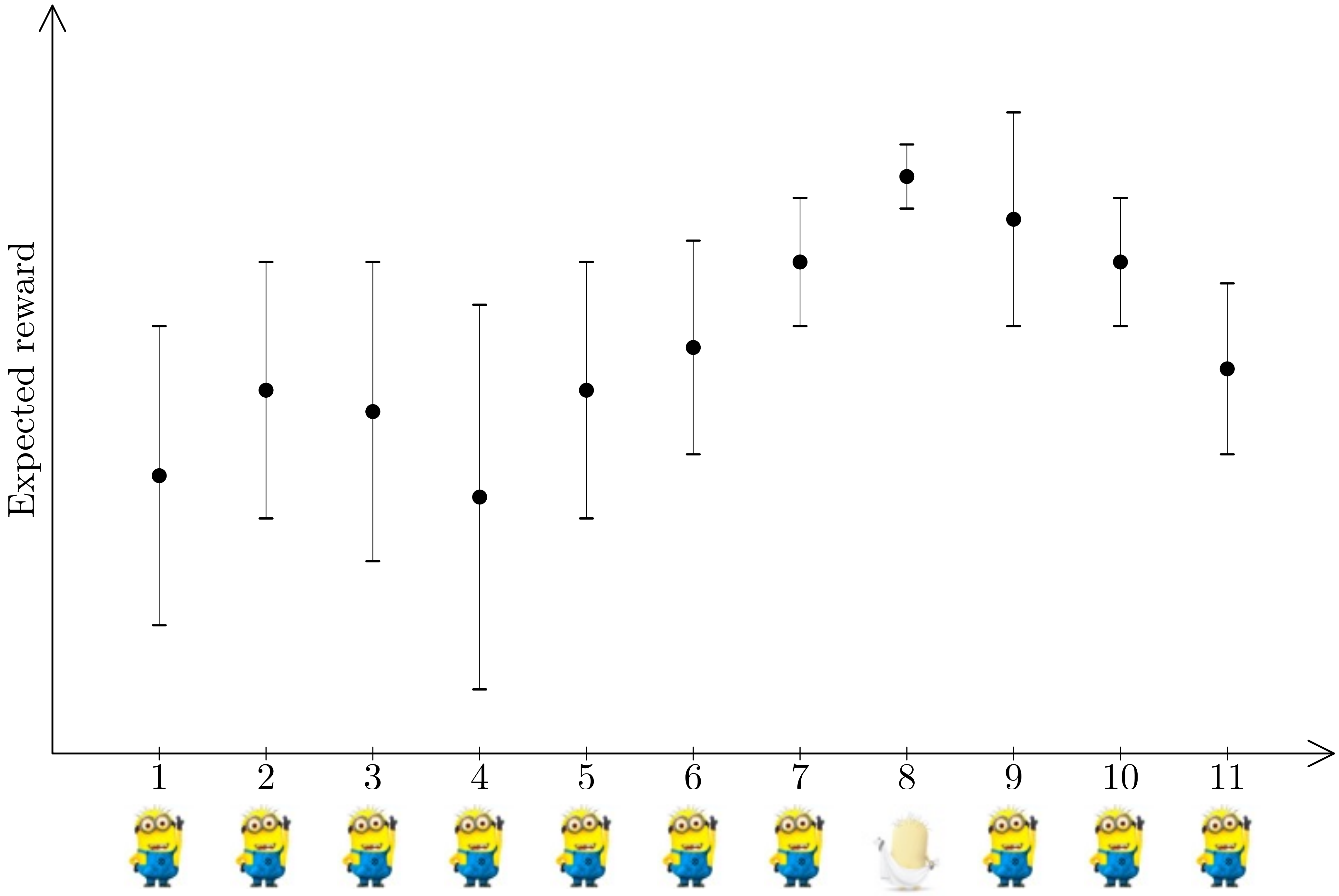
- picks a node $I_t \in [N]$
- incurs a loss ℓ_{t,I_t}
- optional feedback

The performance is total expected regret

$$R_T = \max_{i \in [N]} \mathbb{E}\left[\sum_{t=1}^{T} (\ell_{t,I_t} - \ell_{t,i})\right]$$

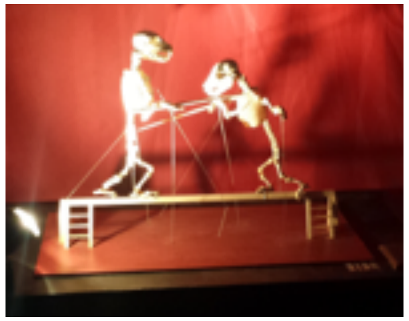
Specific problems differ in
1. loss
2. feedback
3. guarantees

UPPER CONFIDENCE BOUND BASED ALGOS

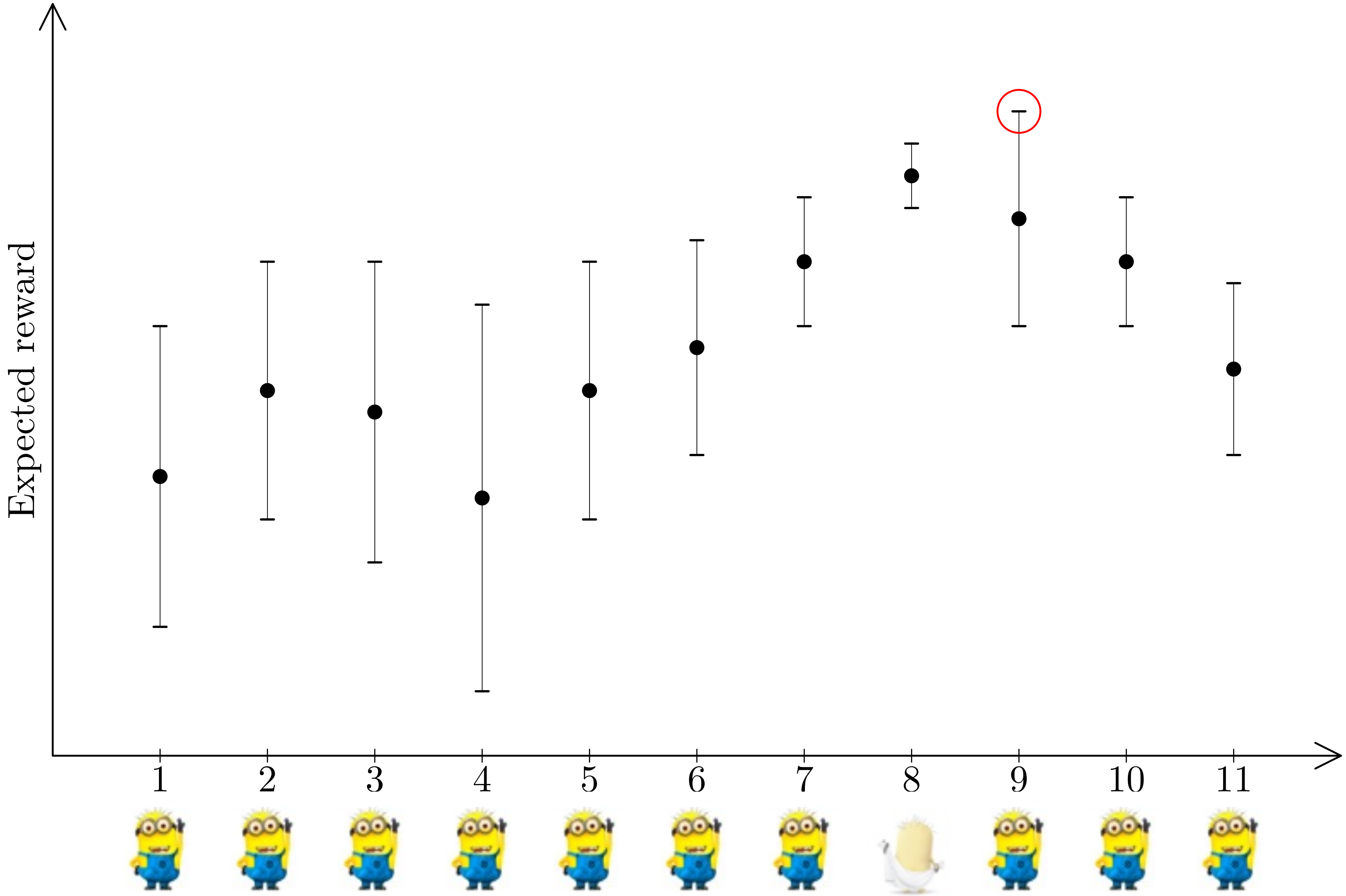


MULTI-ARM BANDITS IN CAFÉ CULTURE

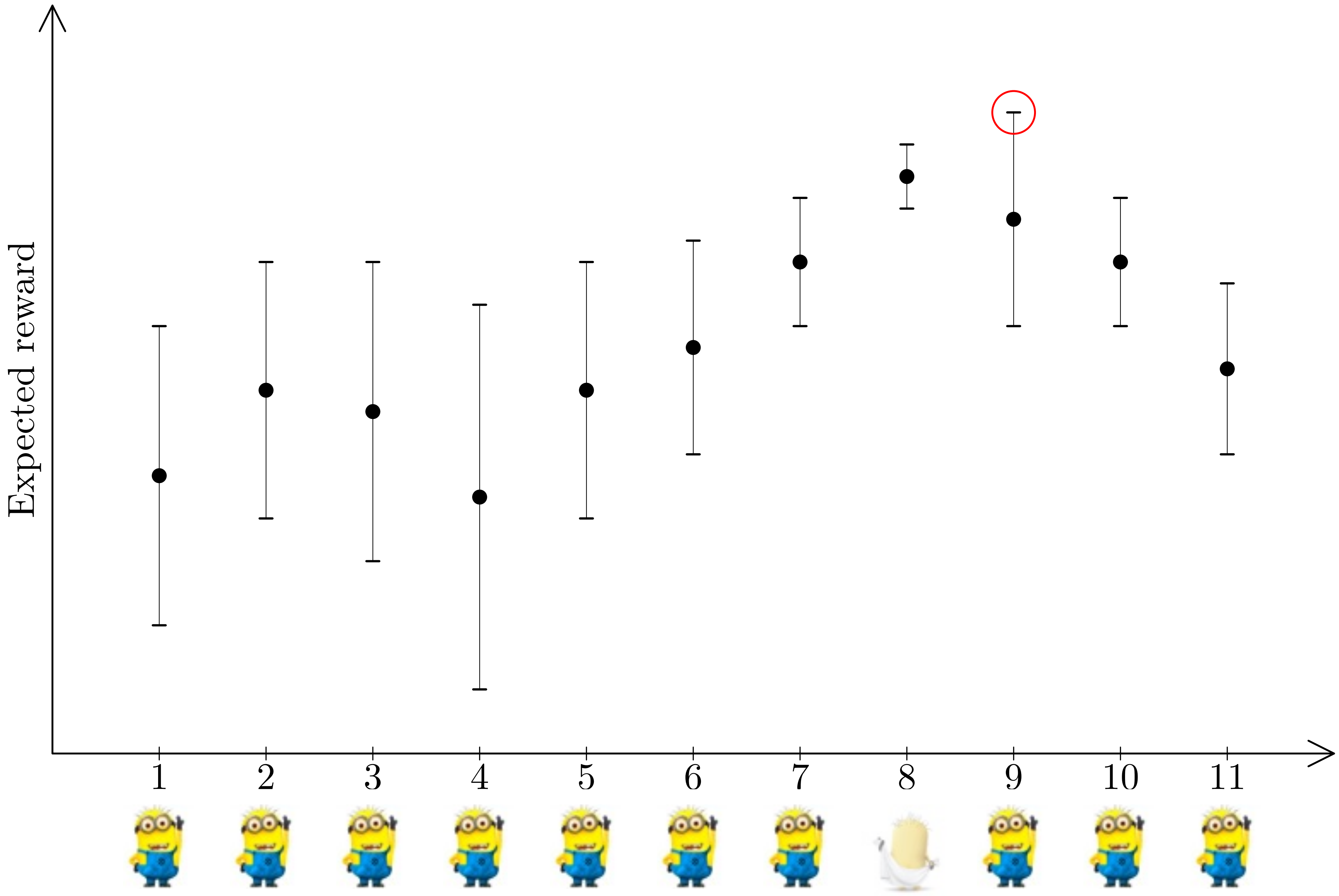
Video recorded March 30th, 2015, 13h50,
Université de Lille, Susie & the Piggy Bones Band



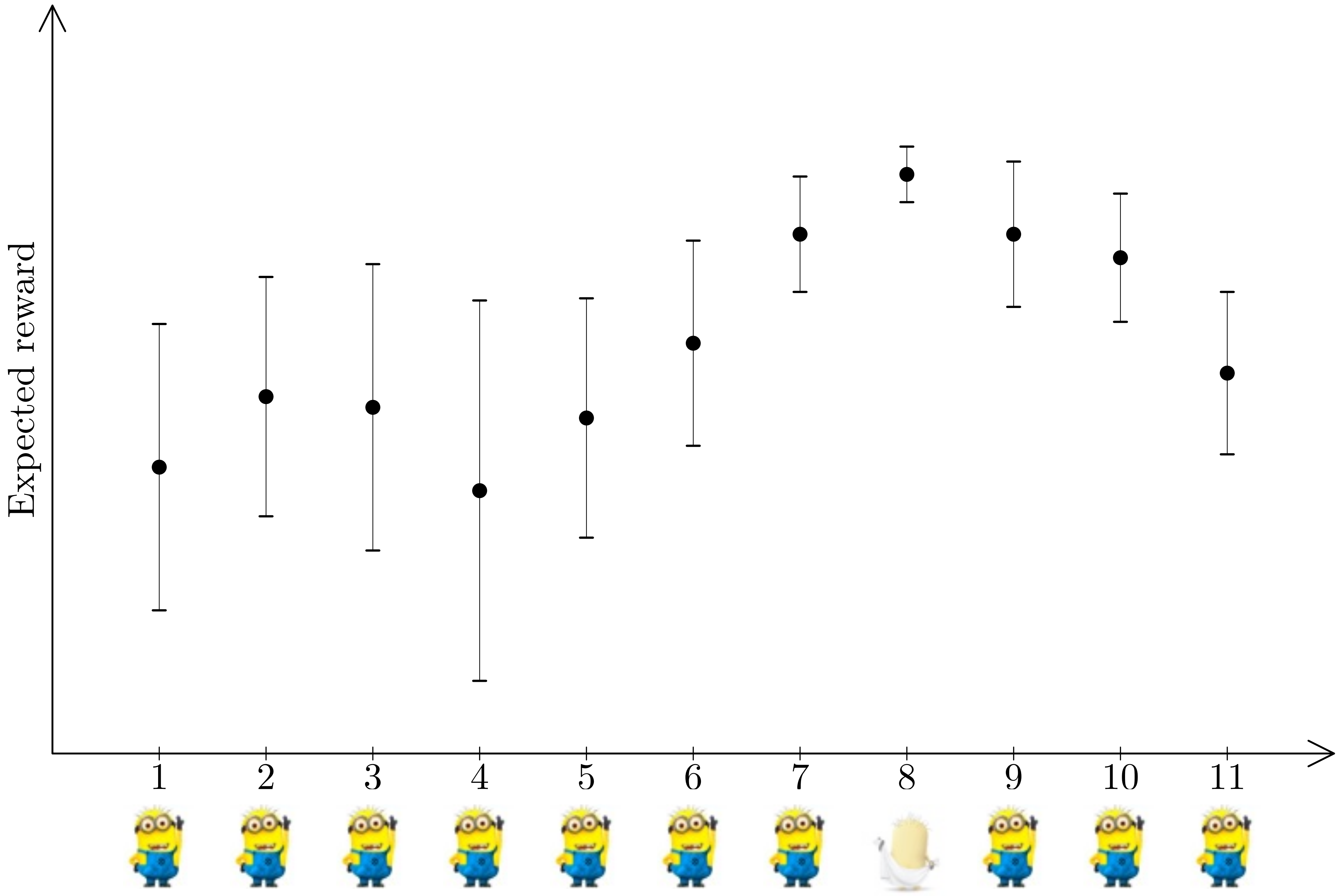
UPPER CONFIDENCE BOUND BASED ALGOS



UPPER CONFIDENCE BOUND BASED ALGOS



UPPER CONFIDENCE BOUND BASED ALGOS



STRUCTURES IN BANDIT PROBLEMS

GRAPHS

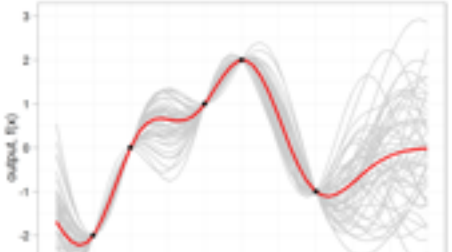
KERNELS

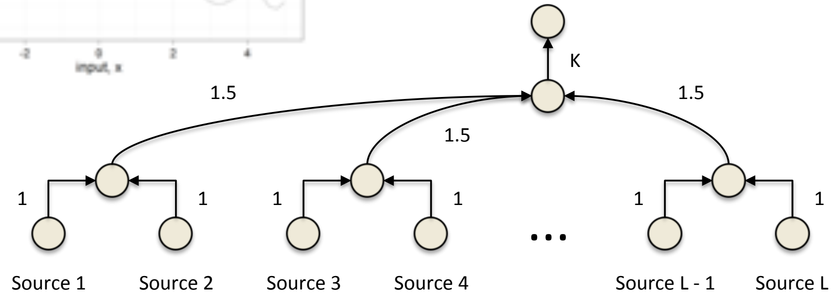
POLYMATROIDS

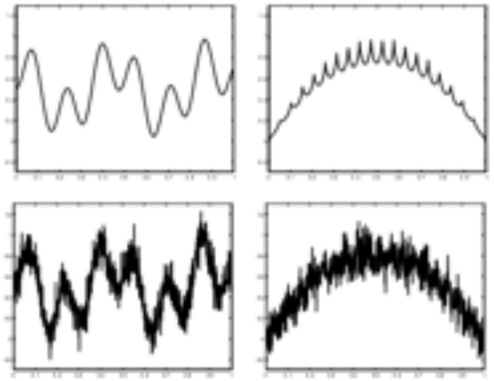
BLACK-BOX FUNCTIONS

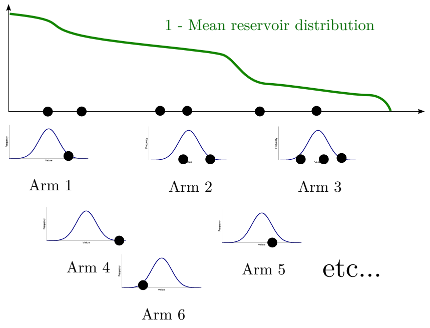
STRUCTURES WITHOUT TOPOLOGY

. . .

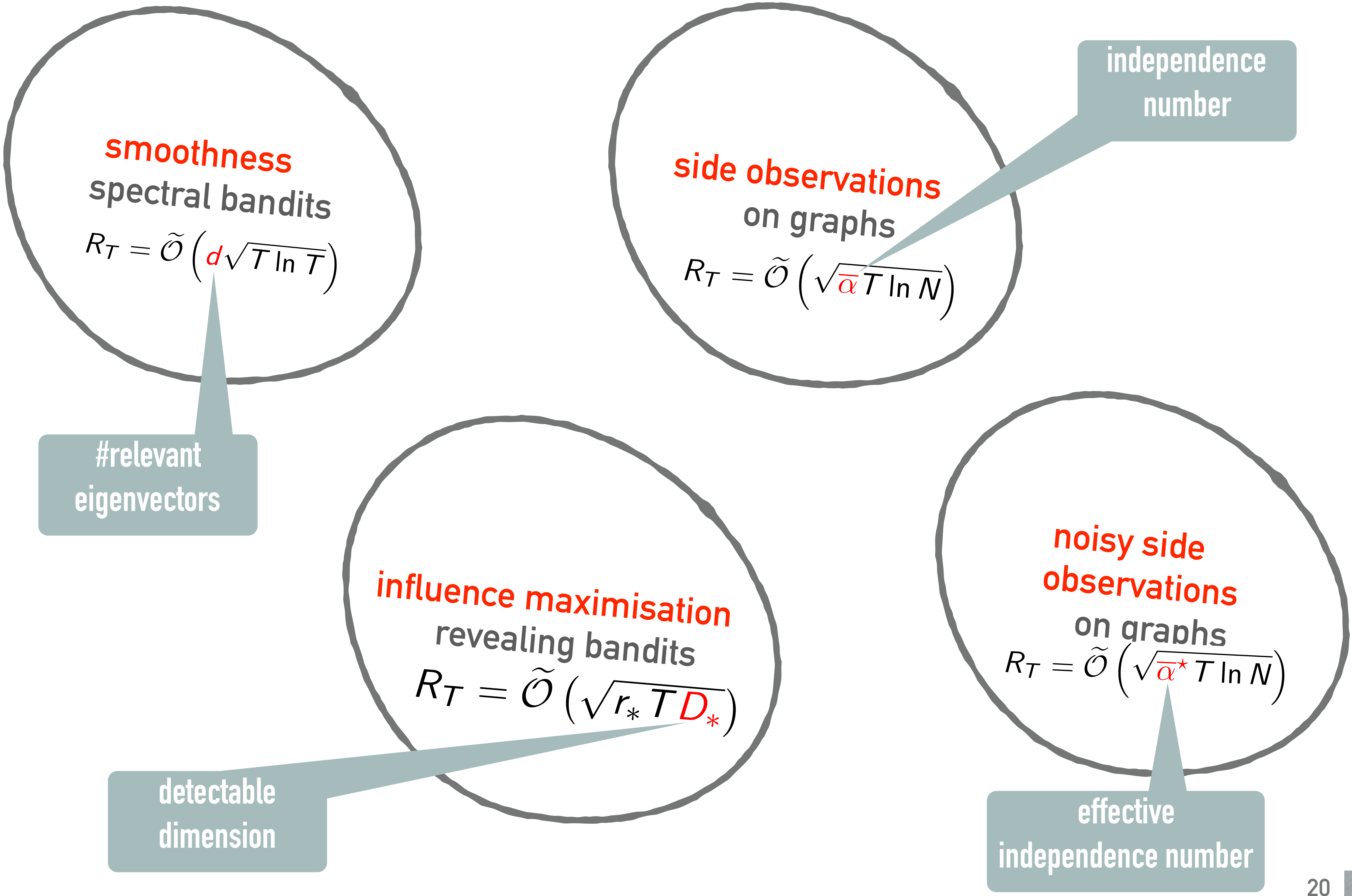








SPECIFIC GRAPH BANDIT SETTINGS



smoothness
spectral bandits

$R_T = \widetilde{\mathcal{O}}\left(d\sqrt{T \ln T}\right)$

#relevant eigenvectors

side observations
on graphs

$R_T = \widetilde{\mathcal{O}}\left(\sqrt{\alpha T \ln N}\right)$

independence number

influence maximisation
revealing bandits

$R_T = \widetilde{\mathcal{O}}\left(\sqrt{r_* T D_*}\right)$

detectable dimension

noisy side observations
on graphs

$R_T = \widetilde{\mathcal{O}}\left(\sqrt{\alpha^\star T \ln N}\right)$

effective independence number

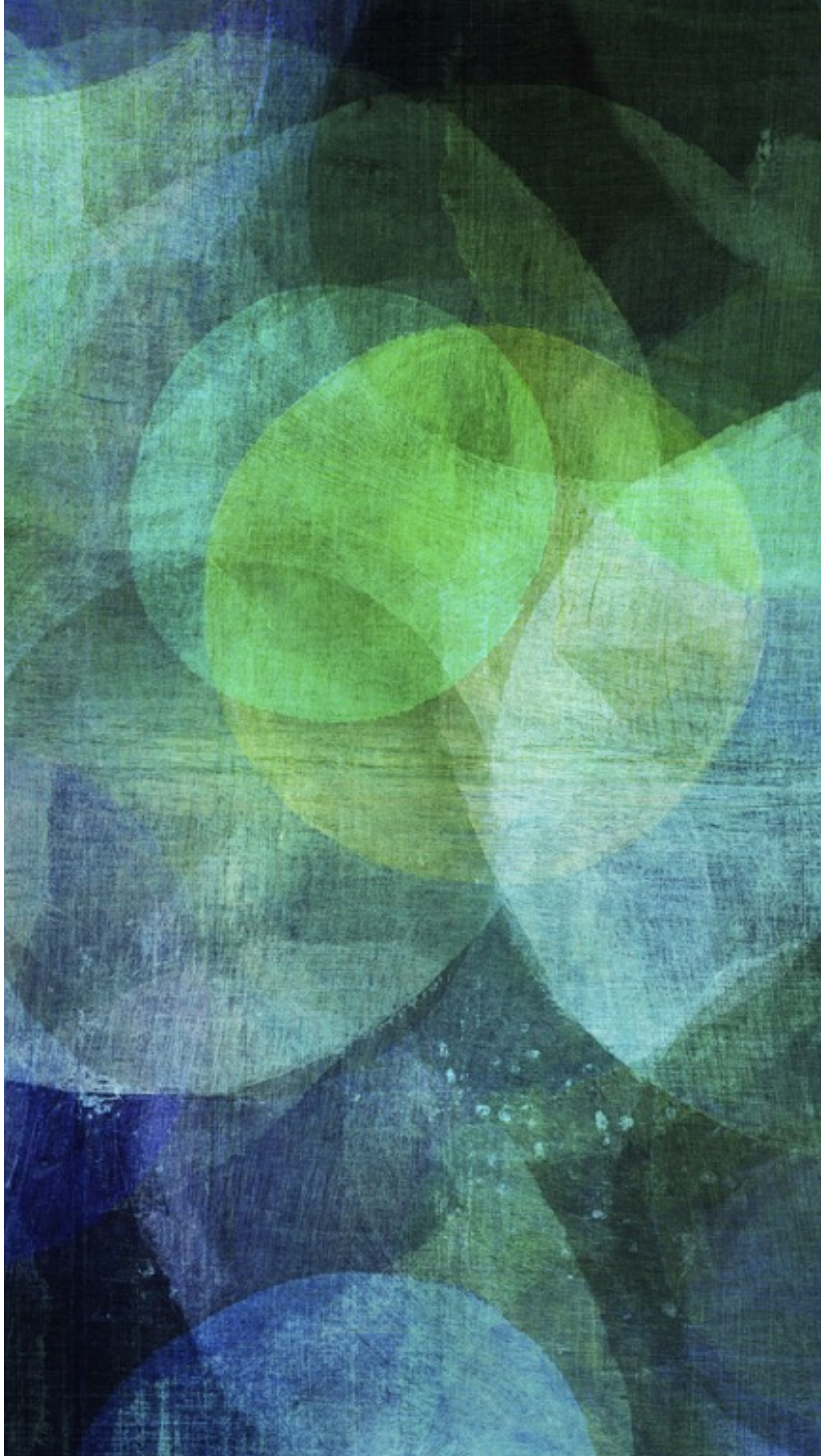
MV, Munos, Kveton, Kocák: **Spectral Bandits for Smooth Graph Functions**, ICML 2014

Kocák, MV, Munos, Agrawal: **Spectral Thompson Sampling**, AAAI 2014

Hanawal, Saligrama, MV, Munos: **Cheap Bandits**, ICML 2015

SPECTRAL BANDITS

exploiting smoothness of rewards on graphs



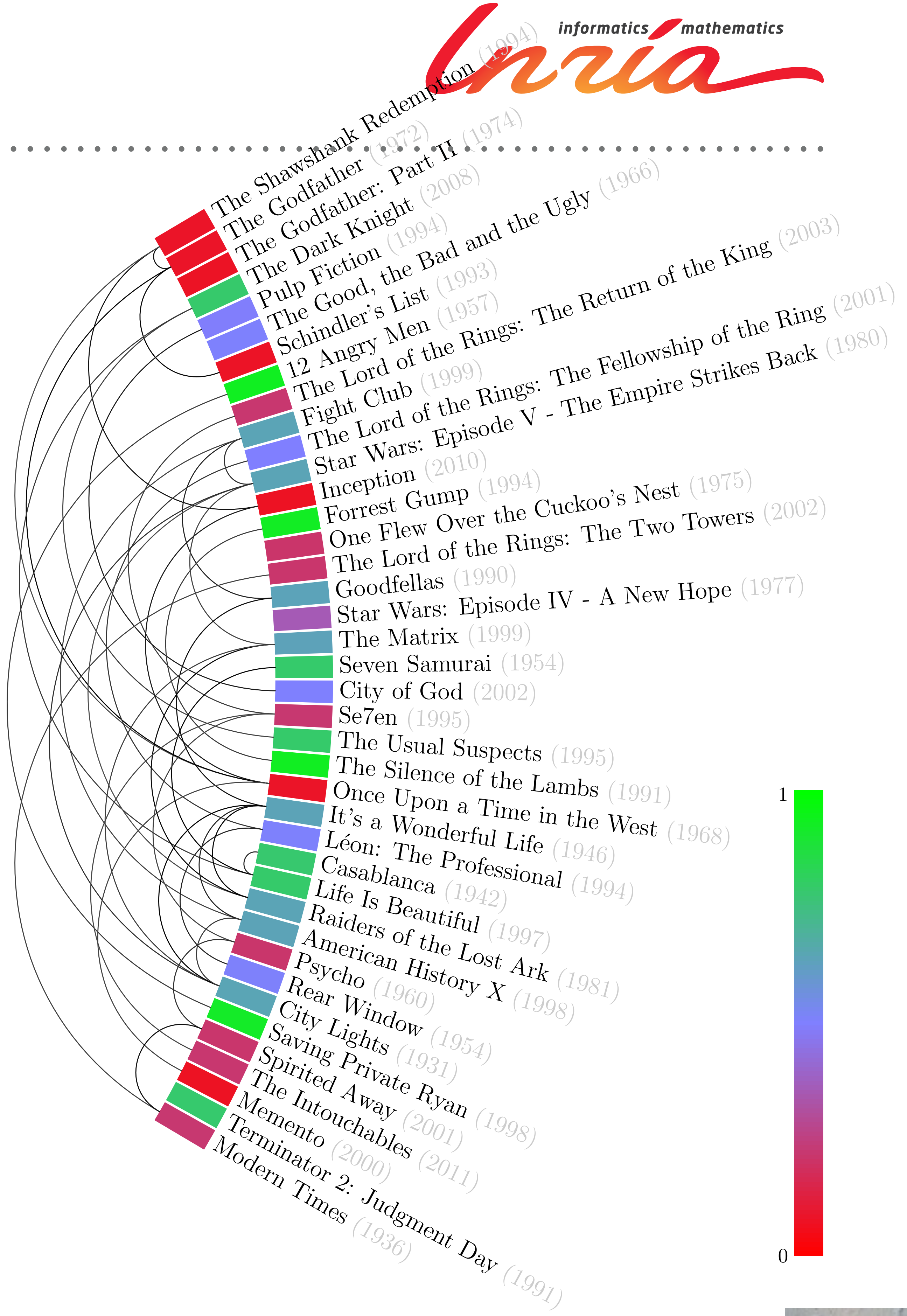
SPECTRAL BANDITS

Assumptions

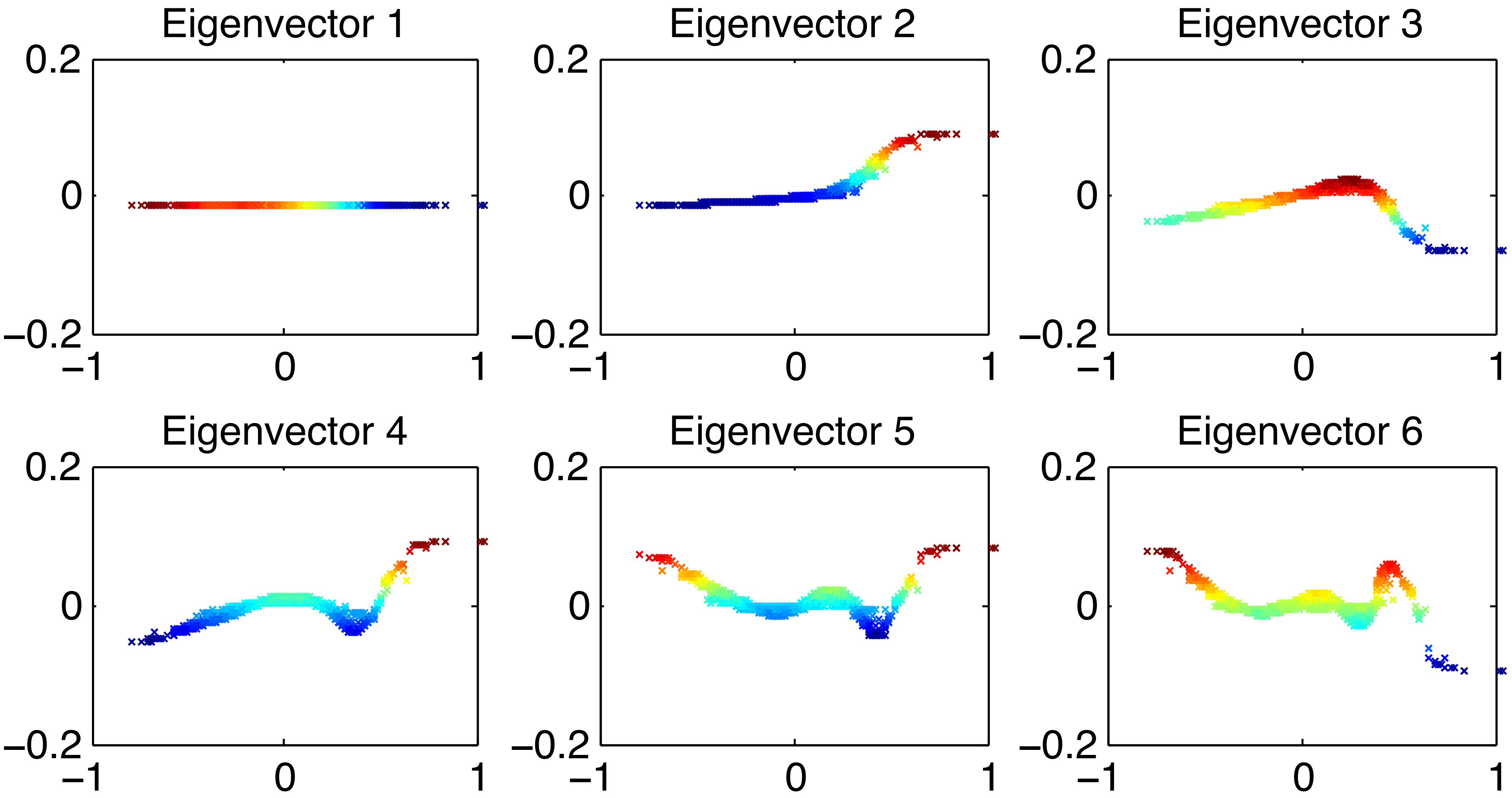
- Unknown reward function $f : V(G) \to \mathbb{R}$.
- Function f is **smooth** on a graph.
- Neighboring movies $\Rightarrow$ similar preferences.
- Similar preferences $\not\Rightarrow$ neighboring movies.

Desiderata

An algorithm useful in the case $T \ll N$!


The Shawshank Redemption (1994)
The Godfather (1972)
The Godfather: Part II (1974)
The Dark Knight (2008)
Pulp Fiction (1994)
The Good, the Bad and the Ugly (1966)
Schindler's List (1993)
12 Angry Men (1957)
The Lord of the Rings: The Return of the King (2003)
Fight Club (1999)
The Lord of the Rings: The Fellowship of the Ring (2001)
Star Wars: Episode V - The Empire Strikes Back (1980)
Inception (2010)
Forrest Gump (1994)
One Flew Over the Cuckoo's Nest (1975)
The Lord of the Rings: The Two Towers (2002)
Goodfellas (1990)
Star Wars: Episode IV - A New Hope (1977)
The Matrix (1999)
Seven Samurai (1954)
City of God (2002)
Se7en (1995)
The Usual Suspects (1995)
The Silence of the Lambs (1991)
Once Upon a Time in the West (1968)
It's a Wonderful Life (1946)
Léon: The Professional (1994)
Casablanca (1942)
Life Is Beautiful (1997)
Raiders of the Lost Ark (1981)
American History X (1998)
Psycho (1960)
Rear Window (1954)
City Lights (1931)
Saving Private Ryan (1998)
Spirited Away (2001)
The Intouchables (2011)
Memento (2000)
Terminator 2: Judgment Day (1991)
Modern Times (1936)
1
0

FLIXSTER DATA



Eigenvectors from the Flixster data corresponding to the smallest few eigenvalues of the graph Laplacian projected onto the first principal component of data. Colors indicate the values.

SPECTRAL BANDIT: LEARNING SETTING

Learning setting for a bandit algorithm π

- In each time t step choose a node $\pi(t)$.
- the $\pi(t)$-th row $\mathbf{x}_{\pi(t)}$ of the matrix $\mathbf{Q}$ corresponds to the arm $\pi(t)$.
- Obtain noisy reward $r_t = \mathbf{x}_{\pi(t)}^{\mathsf{T}}\boldsymbol{\alpha}^* + \varepsilon_t$. Note: $\mathbf{x}_{\pi(t)}^{\mathsf{T}}\boldsymbol{\alpha}^* = f_{\pi(t)}$
    - ε_t is R-sub-Gaussian noise. $\forall\xi \in \mathbb{R},\ \mathbb{E}[e^{\xi\varepsilon_t}] \leq \exp\left(\xi^2 R^2/2\right)$
- Minimize cumulative regret

$$R_T = T\max_a\left(\mathbf{x}_a^{\mathsf{T}}\boldsymbol{\alpha}^*\right) - \sum_{t=1}^{T}\mathbf{x}_{\pi(t)}^{\mathsf{T}}\boldsymbol{\alpha}^*.$$

Can we just use linear bandits?

LINEAR VS. SPECTRAL BANDITS

- **Linear bandit algorithms**
  - **LinUCB** (Li et al., 2010)
    - Regret bound $\approx D\sqrt{T \ln T}$
  - **LinearTS** (Agrawal and Goyal, 2013)
    - Regret bound $\approx D\sqrt{T \ln N}$

Note: D is ambient dimension, in our case N, length of x_i.
Number of actions, e.g., all possible movies → **HUGE!**

- **Spectral bandit algorithms**
  - **SpectralUCB** (Valko et al., ICML 2014)
    - Regret bound $\approx d\sqrt{T \ln T}$
    - Operations per step: $D^2 N$
  - **SpectralTS** (Kocák et al., AAAI 2014)
    - Regret bound $\approx d\sqrt{T \ln N}$
    - Operations per step: $D^2 + DN$

Note: d is effective dimension, usually much smaller than D.

SPECTRAL BANDITS – EFFECTIVE DIMENSION

- **Effective dimension:** Largest d such that

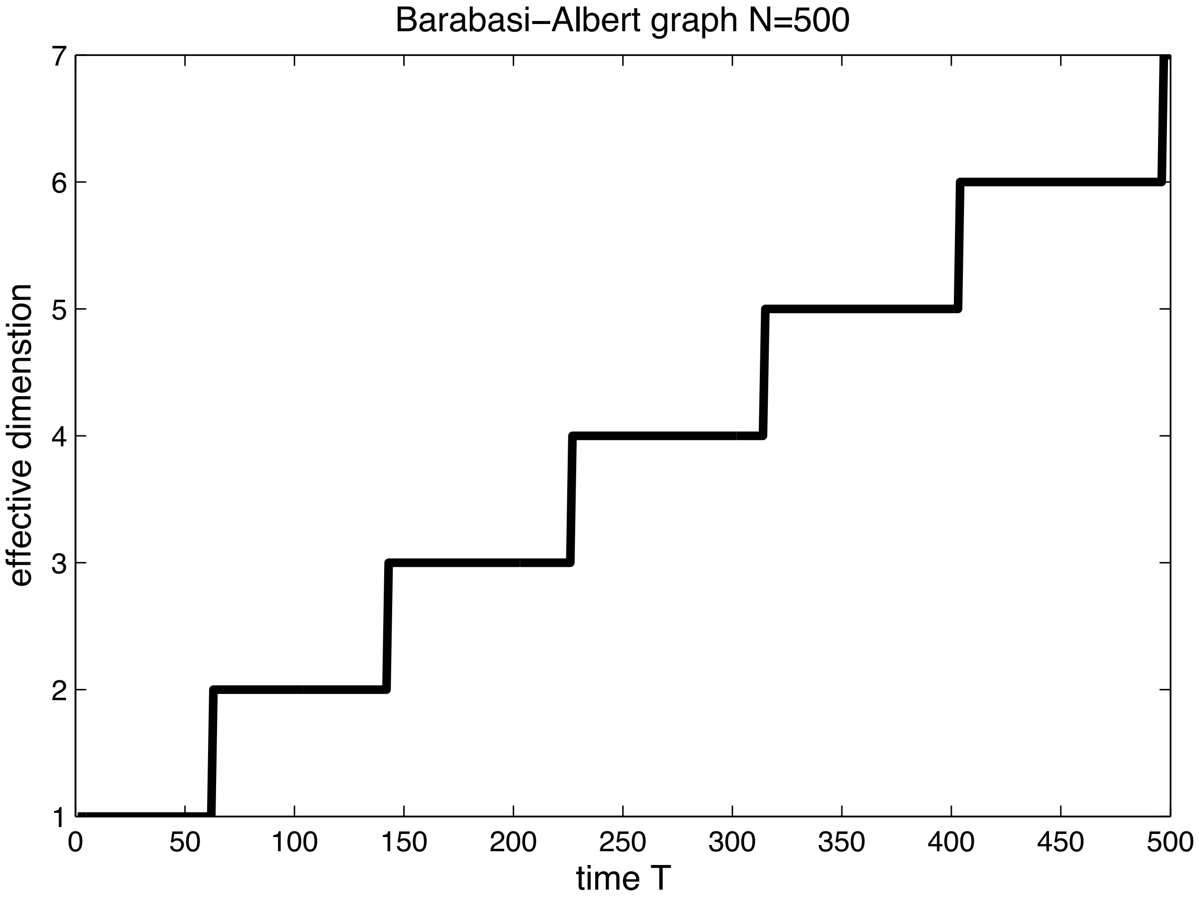
$$(d-1)\lambda_d \leq \frac{T}{\log(1+T/\lambda)}.$$

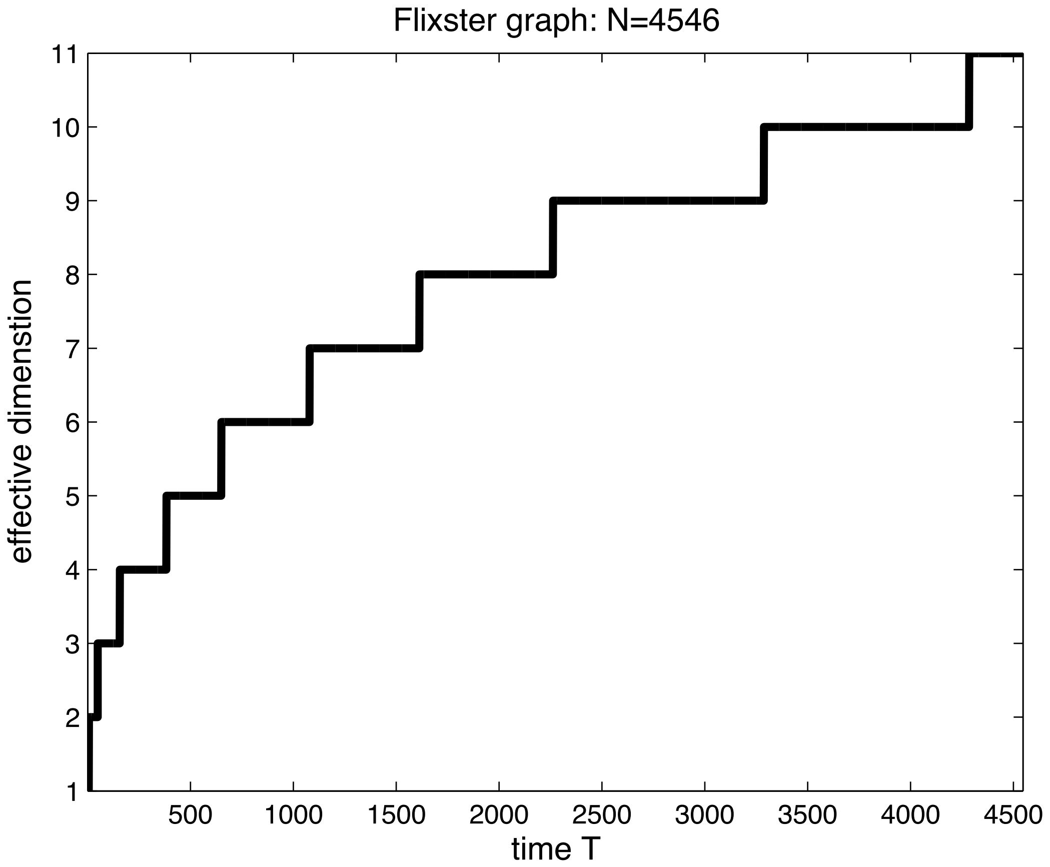
- Function of time horizon and graph properties
- λ_i : i-th smallest eigenvalue of $\mathbf{\Lambda}$.
- λ: Regularization parameter of the algorithm.

Properties:

- d is small when the coefficients λ_i grow rapidly above time.
- d is related to the number of "non-negligible" dimensions.
- Usually d is much smaller than D in real world graphs.
- Can be computed beforehand.

SPECTRAL BANDITS - EFFECTIVE DIMENSION





$$d \ll D$$

Note: In our setting $T < N = D$.

SPECTRAL UCB

Given a vector of weights $\boldsymbol{\alpha}$, we define its $\boldsymbol{\Lambda}$ norm as

$$\|\boldsymbol{\alpha}\|_{\boldsymbol{\Lambda}} = \sqrt{\sum_{k=1}^{N} \lambda_k \alpha_k^2} = \sqrt{\boldsymbol{\alpha}^{\mathsf{T}} \boldsymbol{\Lambda} \boldsymbol{\alpha}},$$

and fit the ratings r_v with a (regularized) least-squares estimate

$$\widehat{\boldsymbol{\alpha}}_t = \arg\min_{\boldsymbol{\alpha}} \left(\sum_{v=1}^{t} \left[\langle \mathbf{x}_v, \boldsymbol{\alpha} \rangle - r_v \right]^2 + \|\boldsymbol{\alpha}\|_{\boldsymbol{\Lambda}}^2 \right).$$

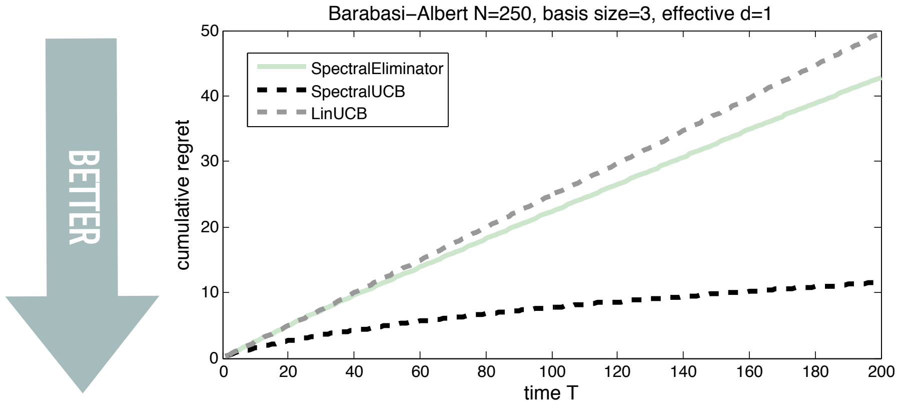
$\|\boldsymbol{\alpha}\|_{\boldsymbol{\Lambda}}$ is a penalty for non-smooth combinations of eigenvectors.

SPECTRAL UCB

```
Input:
    $N$, $T$, $\{\mathbf{\Lambda_L}, \mathbf{Q}\}$, $\lambda$, $\delta$, $R$, $C$
Run:
    $\mathbf{\Lambda} \leftarrow \mathbf{\Lambda_L} + \lambda \mathbf{I}$
    $d \leftarrow \max\{d : (d-1)\lambda_d \leq T/\ln(1+T/\lambda)\}$
for $t = 1$ to $T$ do
    Update the basis coefficients $\widehat{\alpha}$:
    $\mathbf{X}_t \leftarrow [\mathbf{x}_{\pi(1)}, \ldots, \mathbf{x}_{\pi(t-1)}]^{\mathsf{T}}$
    $\mathbf{r} \leftarrow [r_1, \ldots, r_{t-1}]^{\mathsf{T}}$
    $\mathbf{V}_t \leftarrow \mathbf{X}_t \mathbf{X}_t^{\mathsf{T}} + \mathbf{\Lambda}$
    $\widehat{\alpha}_t \leftarrow \mathbf{V}_t^{-1} \mathbf{X}_t^{\mathsf{T}} \mathbf{r}$
    $c_t \leftarrow 2R\sqrt{d\ln(1+t/\lambda) + 2\ln(1/\delta)} + C$
    $\pi(t) \leftarrow \arg\max_a \left(\mathbf{x}_a^{\mathsf{T}} \widehat{\alpha} + c_t \|\mathbf{x}_a\|_{\mathbf{V}_t^{-1}}\right)$
    Observe the reward $r_t$
end for
```

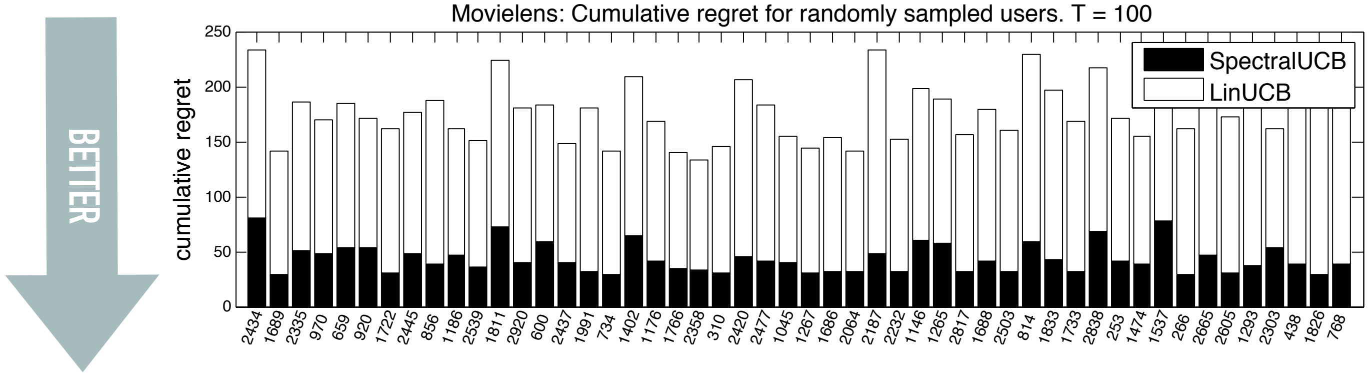
SPECTRALUCB REGRET BOUND

- d: Effective dimension.
- λ: Minimal eigenvalue of $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_\mathbf{L} + \lambda\mathbf{I}$.
- C: Smoothness upper bound, $\|\boldsymbol{\alpha}^*\|_{\boldsymbol{\Lambda}} \leq C$.
- $\mathbf{x}_i^\mathsf{T}\boldsymbol{\alpha}^* \in [-1, 1]$ for all i.

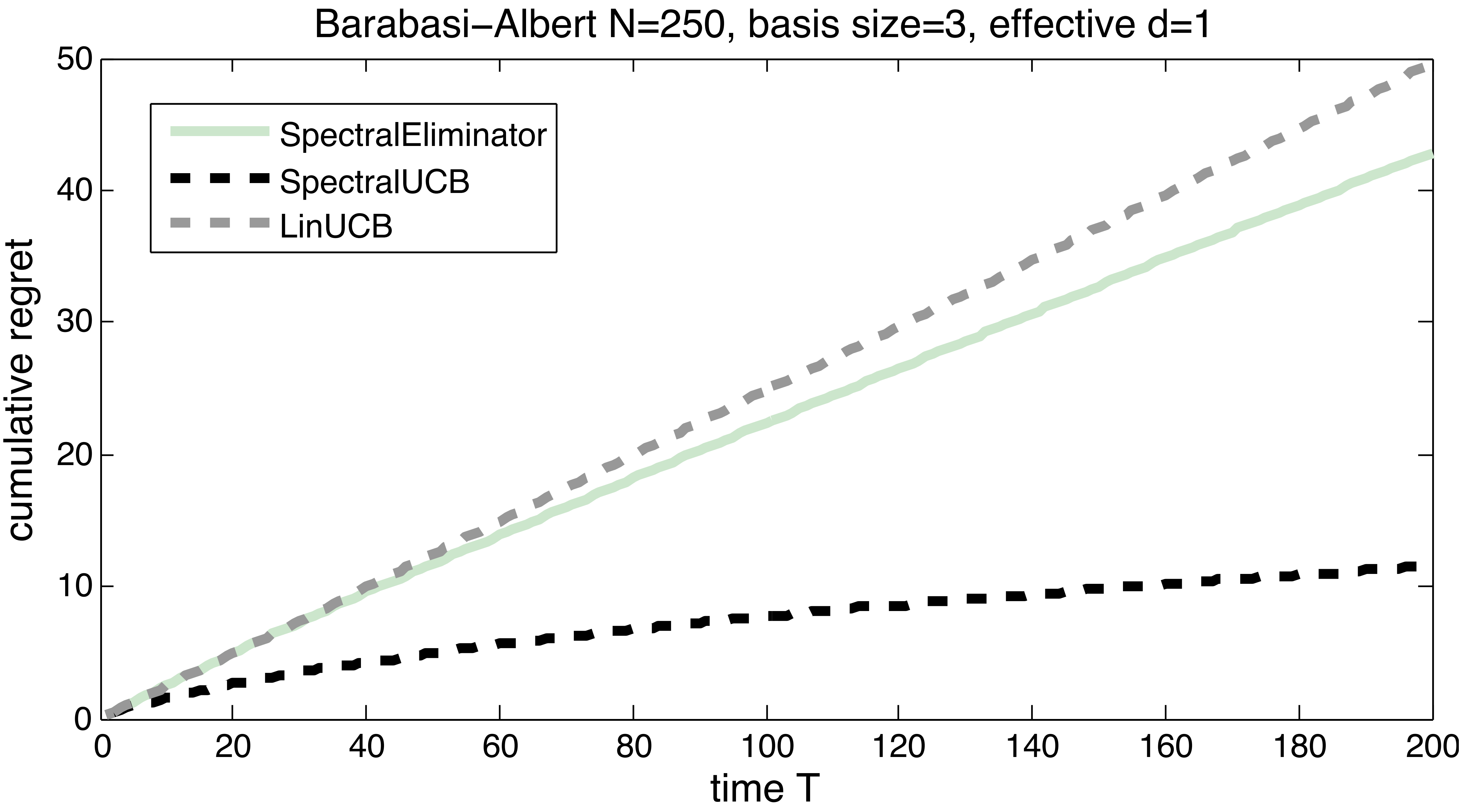


The **cumulative regret** R_T of **SpectralUCB** is with probability $1 - \delta$ bounded as

$$R_T \leq \left(8R\sqrt{d \ln \frac{\lambda + T}{\lambda} + 2 \ln \frac{1}{\delta}} + 4C + 4\right) \sqrt{dT \ln \frac{\lambda + T}{\lambda}}.$$

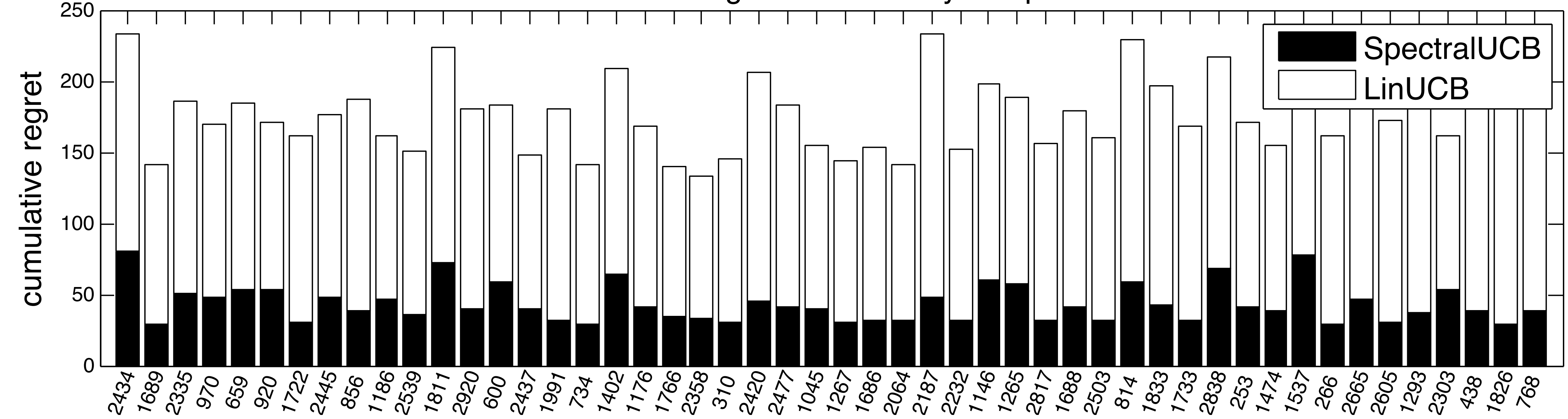


SPECTRAL UCN ON BA GRAPH

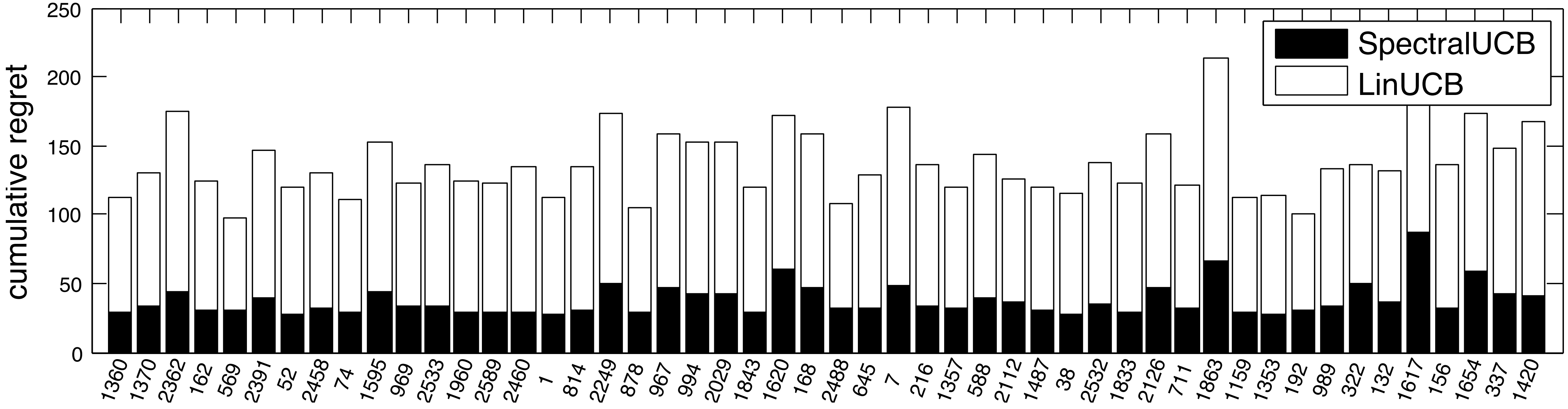


SPECTRAL UCN ON REAL DATA

Movielens: Cumulative regret for randomly sampled users. T = 100



Flixster: Cumulative regret for randomly sampled users. T = 100



SPECTRAL UCB: REGRET ANALYSIS

- Derivation of the confidence ellipsoid for $\widehat{\boldsymbol{\alpha}}$ with probability $1-\delta$.
    - Using analysis of OFUL (Abbasi-Yadkori et al., 2011)

$$|\mathbf{x}^\mathsf{T}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^*)| \le \|\mathbf{x}\|_{\mathbf{V}_t^{-1}} \left(R\sqrt{2\ln\left(\frac{|\mathbf{V}_t|^{1/2}}{\delta|\boldsymbol{\Lambda}|^{1/2}}\right)}+C\right)$$

- Regret in one time step: $r_t = \mathbf{x}_*^\mathsf{T}\boldsymbol{\alpha}^* - \mathbf{x}_{\pi(t)}^\mathsf{T}\boldsymbol{\alpha}^* \le 2c_t\|\mathbf{x}_{\pi(t)}\|_{\mathbf{V}_t^{-1}}$
- Cumulative regret:

$$R_T=\sum_{t=1}^{T} r_t \le \sqrt{T\sum_{t=1}^{T} r_t^2} \le 2(c_T+1)\sqrt{2T\ln\frac{|\mathbf{V}_T|}{|\boldsymbol{\Lambda}|}}$$

- Upperbound for $\ln(|\mathbf{V}_t|/|\boldsymbol{\Lambda}|)$

$$\ln\frac{|\mathbf{V}_t|}{|\boldsymbol{\Lambda}|} \le \ln\frac{|\mathbf{V}_T|}{|\boldsymbol{\Lambda}|} \le 2d\ln\left(\frac{\lambda+T}{\lambda}\right)$$

SPECTRAL UCB: REGRET ANALYSIS

Sylvester's determinant theorem:

$$|\mathbf{A}+\mathbf{x}\mathbf{x}^{\mathsf{T}}| = |\mathbf{A}||\mathbf{I}+\mathbf{A}^{-1}\mathbf{x}\mathbf{x}^{\mathsf{T}}| = |\mathbf{A}|(1+\mathbf{x}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{x})$$

Goal:

- Upperbound determinant $|\mathbf{A}+\mathbf{x}\mathbf{x}^{\mathsf{T}}|$ for $\|\mathbf{x}\|_2 \leq 1$
- Upperbound $\mathbf{x}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{x}$

$$\mathbf{x}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{x} = \mathbf{x}^{\mathsf{T}}\mathbf{Q}\boldsymbol{\Lambda}^{-1}\mathbf{Q}^{\mathsf{T}}\mathbf{x} = \mathbf{y}^{\mathsf{T}}\boldsymbol{\Lambda}^{-1}\mathbf{y} = \sum_{i=1}^{N}\lambda_i^{-1}y_i^2$$

- $\|\mathbf{y}\|_2 \leq 1$.
- $\mathbf{y}$ is a canonical vector.
- $\mathbf{x} = \mathbf{Q}\mathbf{y}$ is an eigenvector of $\mathbf{A}$.

SPECTRAL UCB: REGRET ANALYSIS

Corollary: Determinant $|\mathbf{V}_T|$ of $\mathbf{V}_T = \mathbf{\Lambda} + \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^{\mathsf{T}}$ is maximized when all $\mathbf{x}_t$ are aligned with axes.

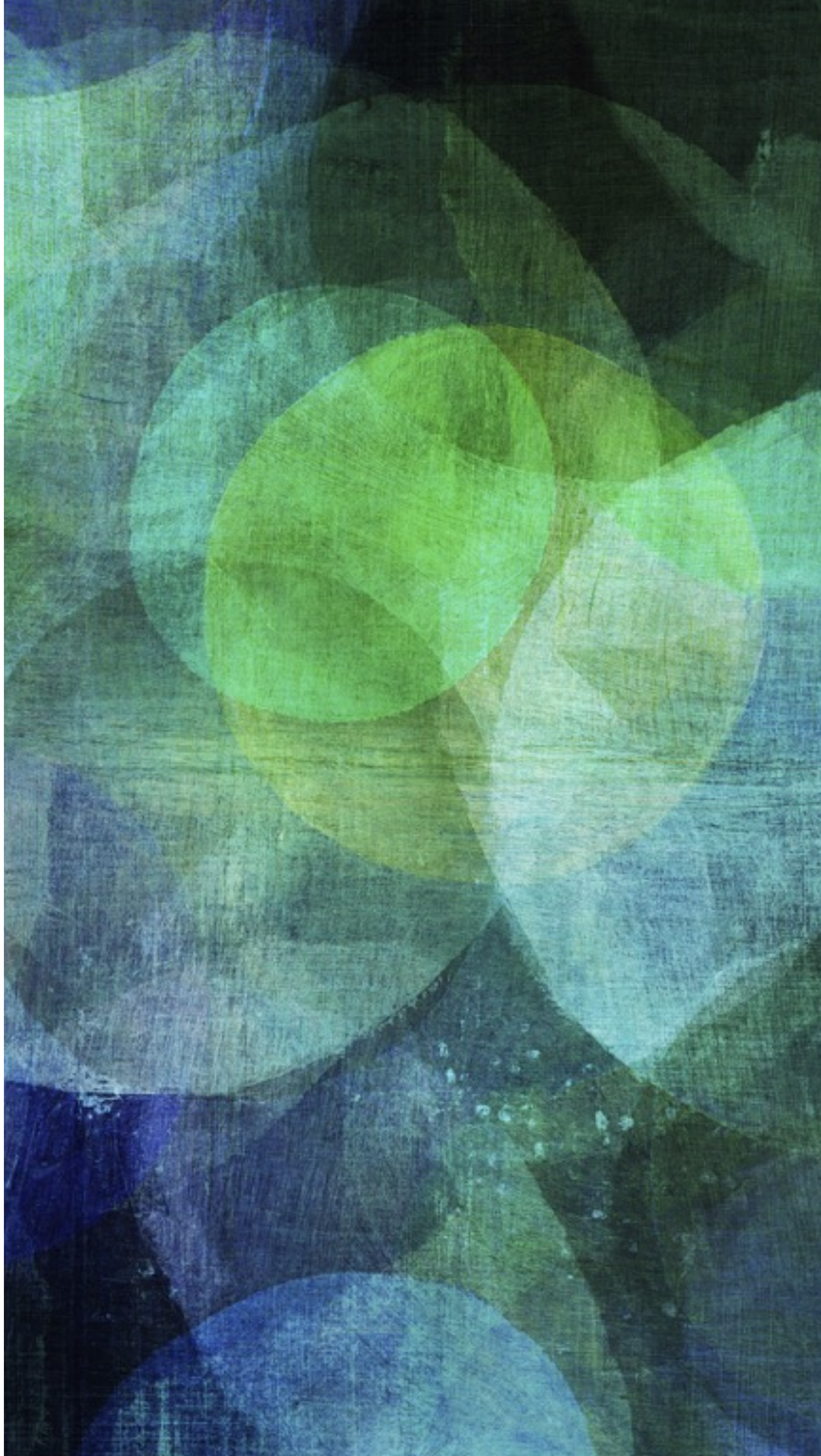
$$
\begin{aligned}
|\mathbf{V}_T| &\le \max_{\sum t_i = T} \prod (\lambda_i + t_i) \\
\ln \frac{|\mathbf{V}_T|}{|\mathbf{\Lambda}|} &\le \max_{\sum t_i = T} \sum \ln\left(1 + \frac{t_i}{\lambda_i}\right) \\
\ln \frac{|\mathbf{V}_T|}{|\mathbf{\Lambda}|} &\le \sum_{i=1}^{d} \ln\left(1 + \frac{T}{\lambda}\right) + \sum_{i=d+1}^{N} \ln\left(1 + \frac{t_i}{\lambda_{d+1}}\right) \\
&\le d \ln\left(1 + \frac{T}{\lambda}\right) + \frac{T}{\lambda_{d+1}} \\
&\le 2d \ln\left(1 + \frac{T}{\lambda}\right)
\end{aligned}
$$

Carpentier, MV: **Revealing Graph Bandits for Maximising Local Influence**, AISTATS 2016

Wen, Kveton, MV: **Influence Maximization with Semi-Bandit Feedback**, (arXiv:1605.06593)

INFLUENCE MAXIMISATION

looking for the influential nodes **while** exploring the graph



HOW TO RULE THE WORLD?

Influence the influential!

JULY 18, 2016

Religion

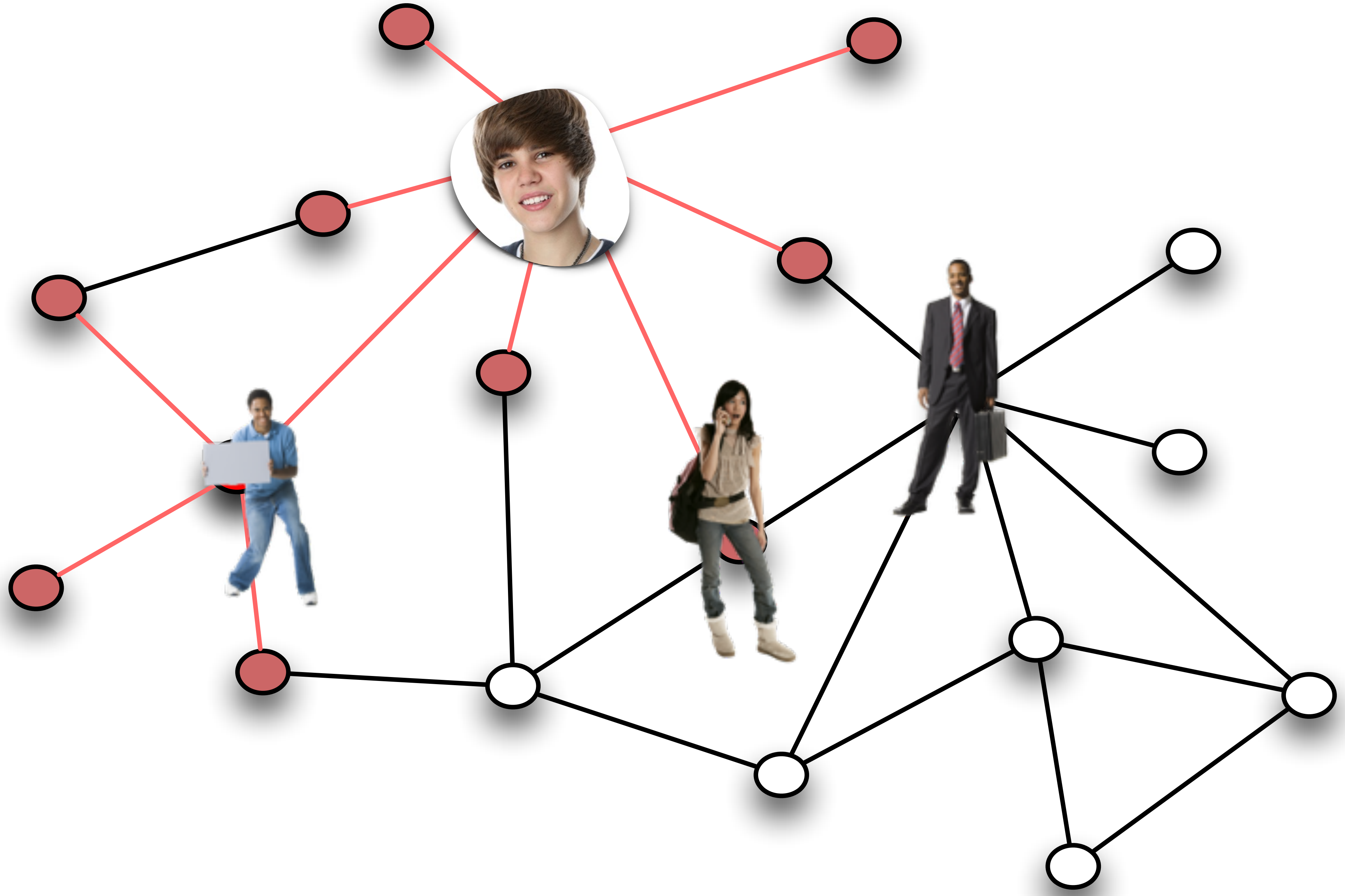
March 26, 2017

Politics

September 1, 2009

Culture

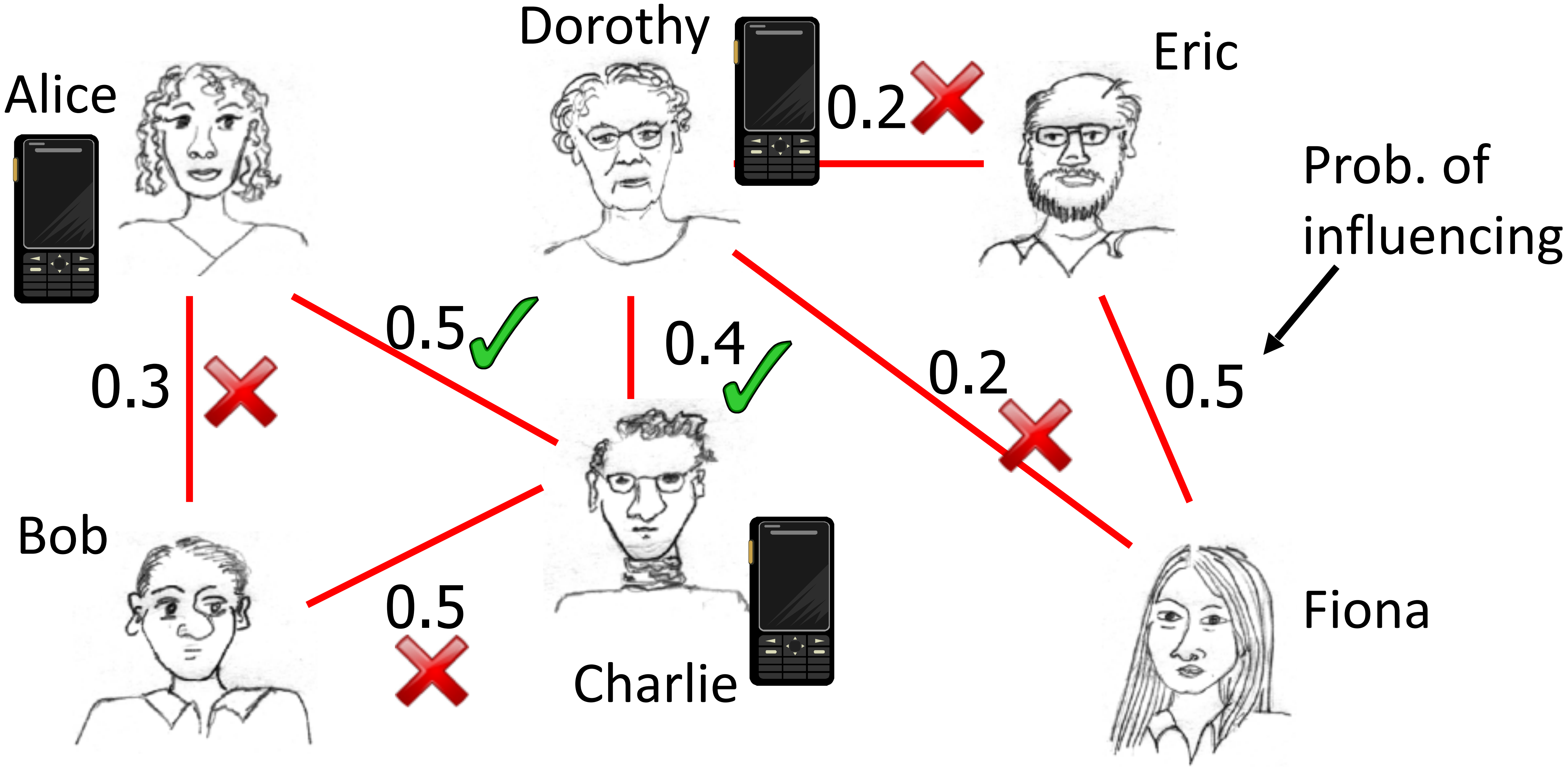
338 ET 200



$$F(S) = \text{spread}$$

slide from Stefanie Jegelka

EXAMPLE: INFLUENCE IN SOCIAL NETWORKS [KEMPE, KLEINBERG, TARDOS KDD '03]

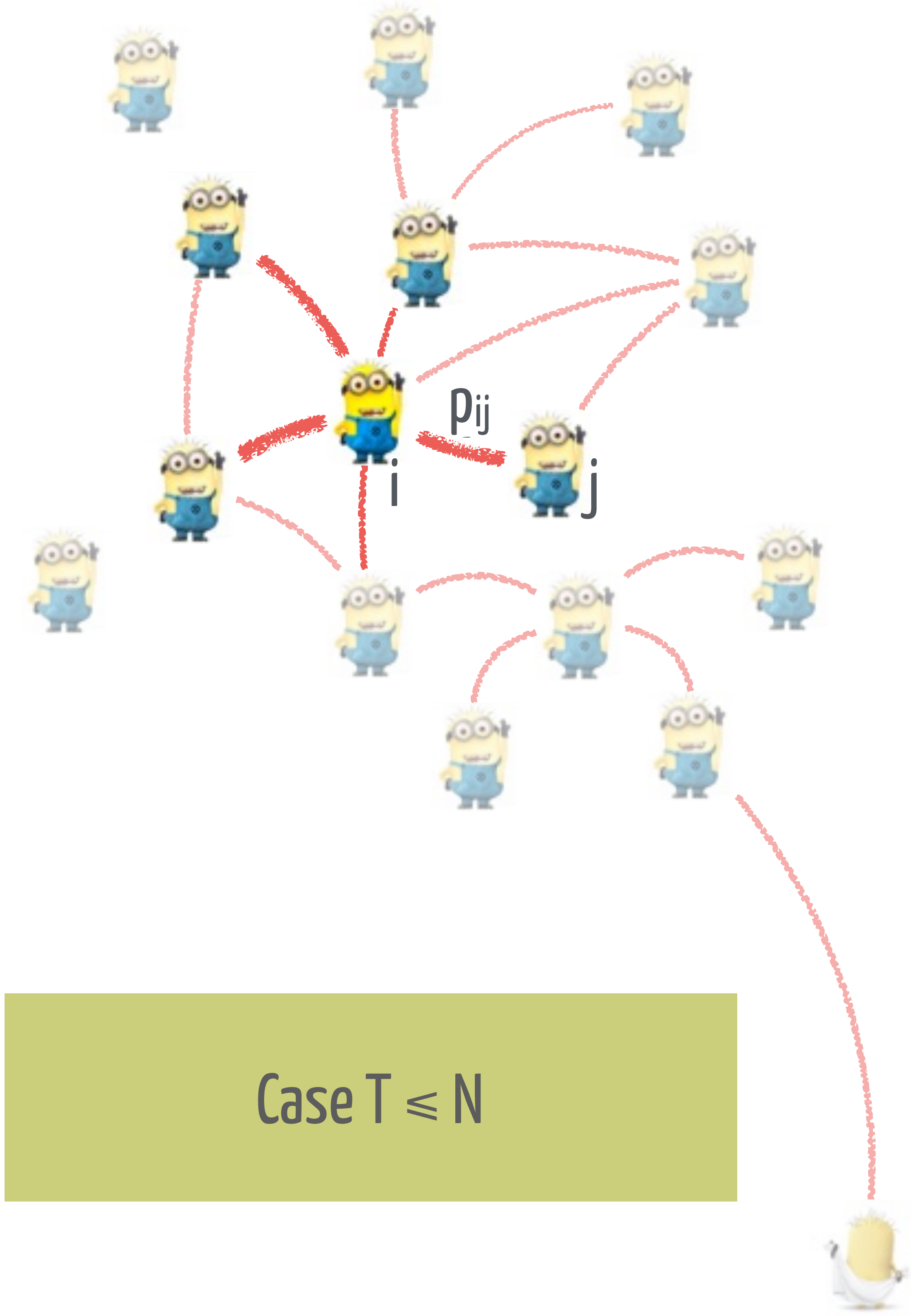


Who should get free cell phones?

V = {Alice,Bob,Charlie,Dorothy,Eric,Fiona}

F(A) = Expected number of people influenced when targeting A

REVEALING BANDITS FOR LOCAL INFLUENCE



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

In each time step t = 1, …., T

learner picks a node k_t

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

Select influential people = Find the strategy maximising

$$L_T = \sum_{t=1}^{T} |S_{k_t,t}|$$

Why this is a bandit problem?

Case T ≤ N

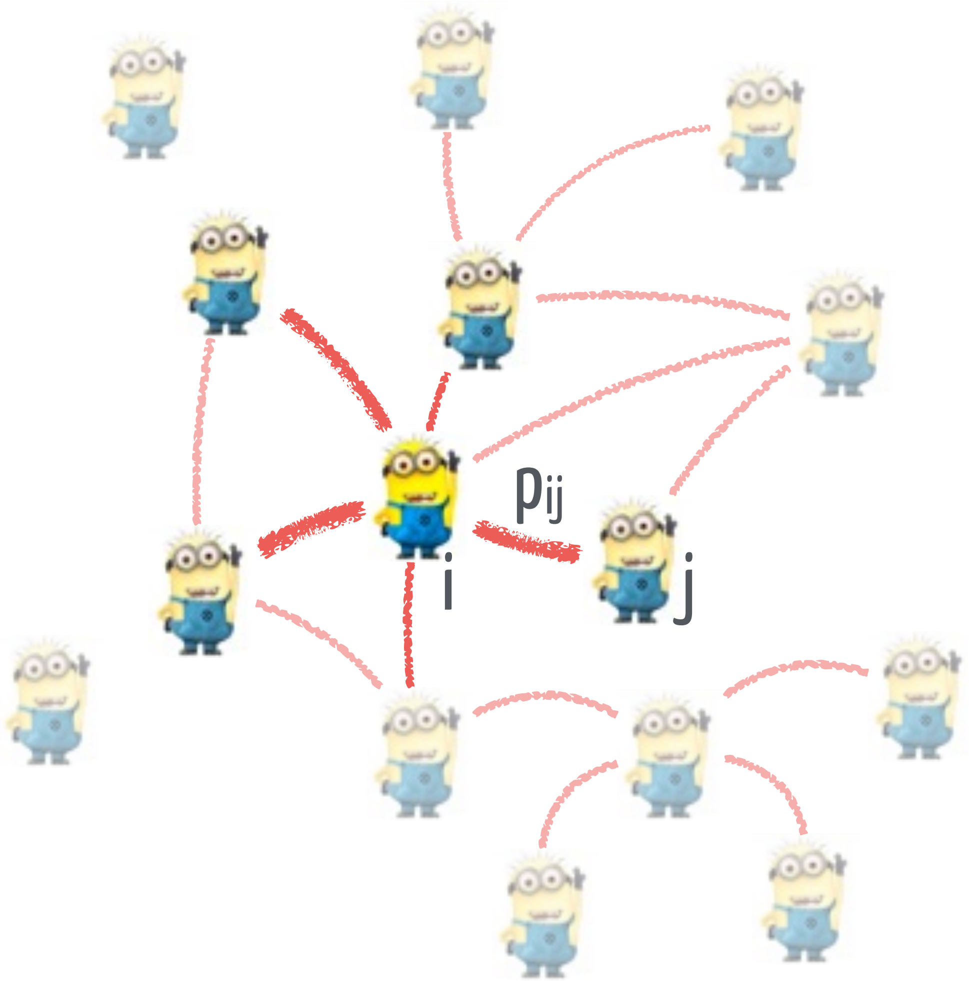
PERFORMANCE CRITERION

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

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

Oracle strategy always selects the best

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



Expected regret of the oracle strategy

$$\mathbb{E}\left[L_T^\star\right] = T r_\star$$

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

$$\mathbb{E}\left[R_T\right] = \mathbb{E}\left[L_T^\star\right] - \mathbb{E}\left[L_T\right]$$

BASELINE

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

$$\mathbb{E}\left[R_T\right] \leq U \min\left(r_\star T, r_\star N + \sqrt{r_\star T N}\right)$$

each node at least once

unlearnable case T ≤ N

- matching lower bound

Crash course on stochastic bandits?

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

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

$$D(\Delta) = |\{i \leq N : r_\star^\circ - r_i^\circ \leq \Delta\}|$$

- $D(r) = N$ for $r \geq r_*$ 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

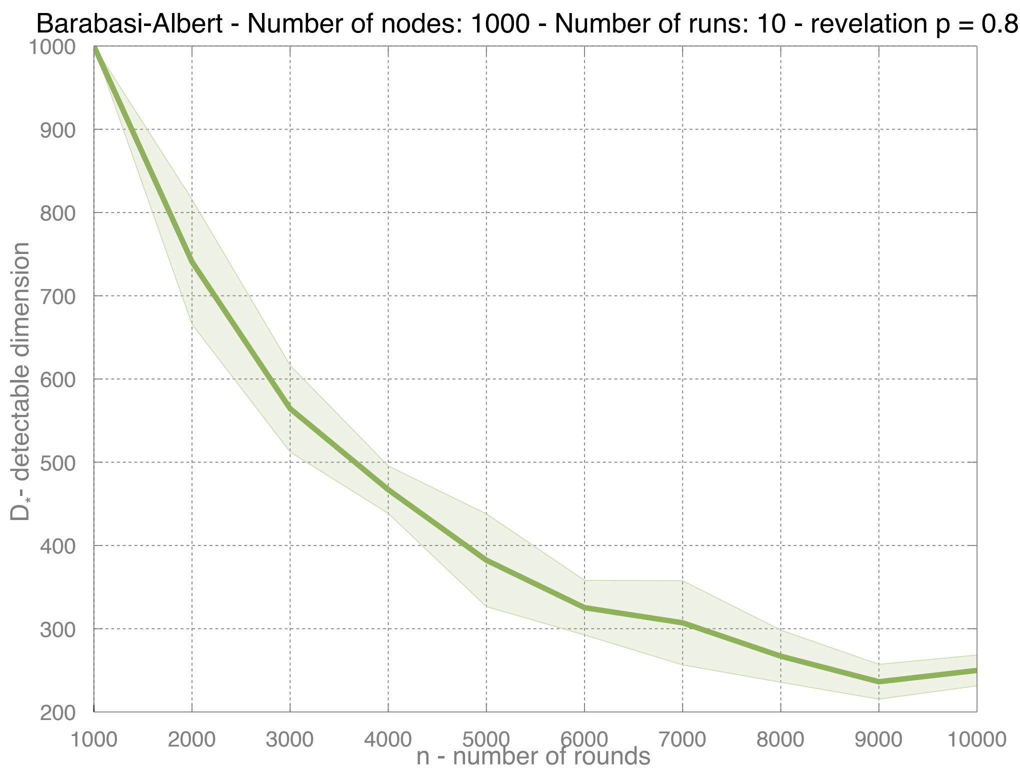
$$\Delta_\star = 16\sqrt{\frac{r_\star^\circ N \log(TN)}{T_\star}} + \frac{144 N \log(TN)}{T_\star}$$

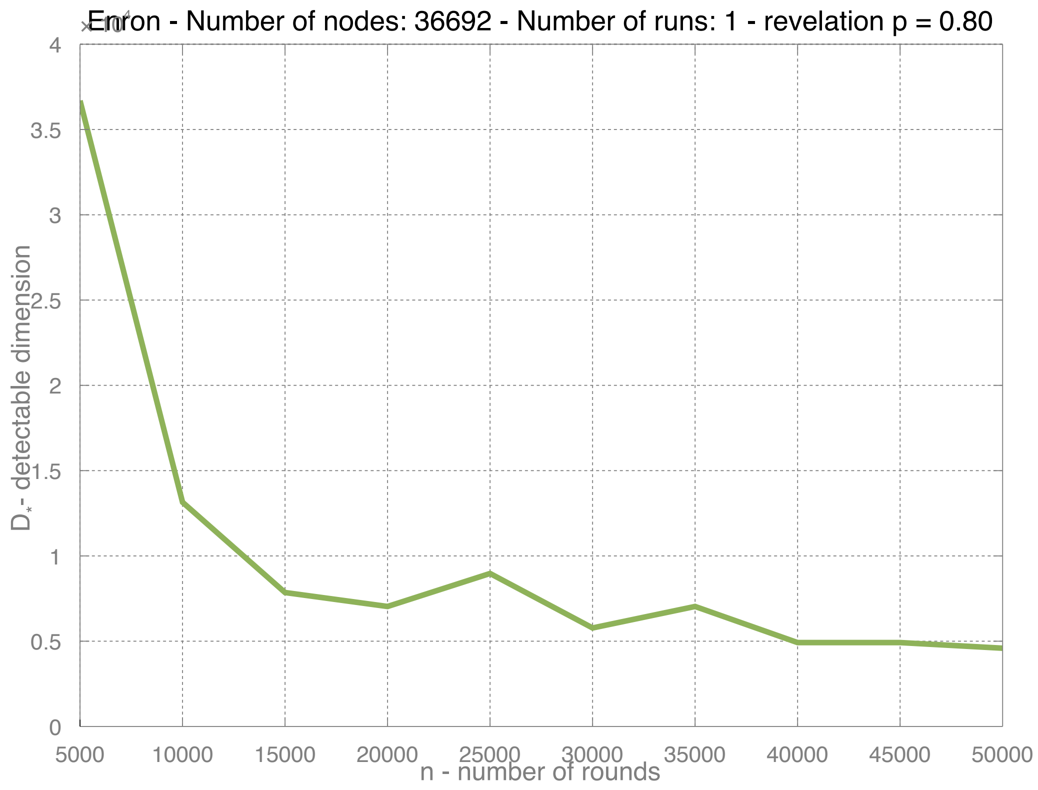
- Detectable horizon T_*, smallest integer s.t. $T_\star r_\star^\circ \geq \sqrt{D_\star T r_\star^\circ},$
- Equivalently: D_* corresponding to smallest T_* such that

$$T_\star r_\star^\circ \geq \sqrt{D\left(16\sqrt{\frac{r_\star^\circ N \log(TN)}{T_\star}} + \frac{144 N \log(TN)}{T_\star}\right) T r_\star^\circ}$$

HOW DOES D* BEHAVE?

- For (easy, structured) star graphs $D_* = 1$ even for small n (big gain)
- For (difficult) empty graphs D_*= N even for large T (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 T





BARE 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

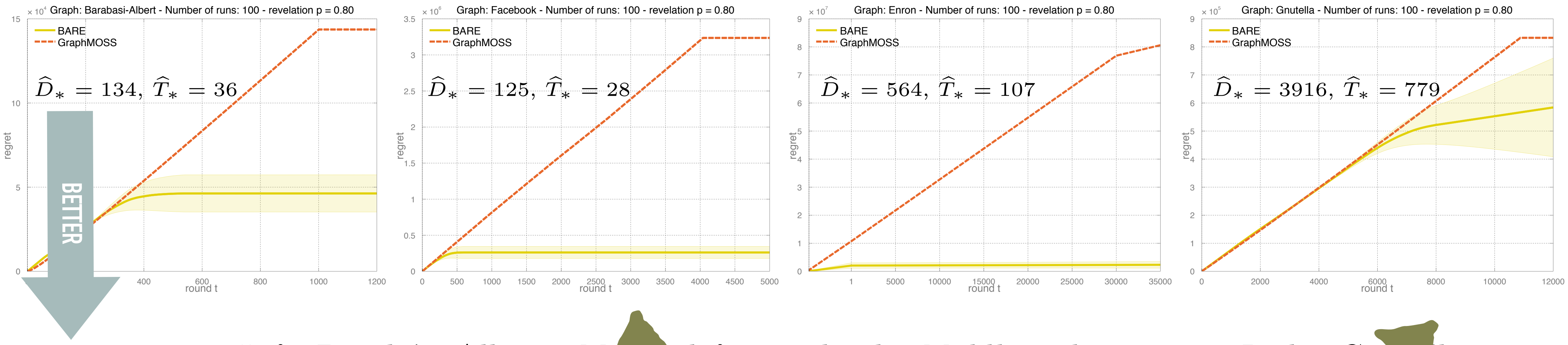
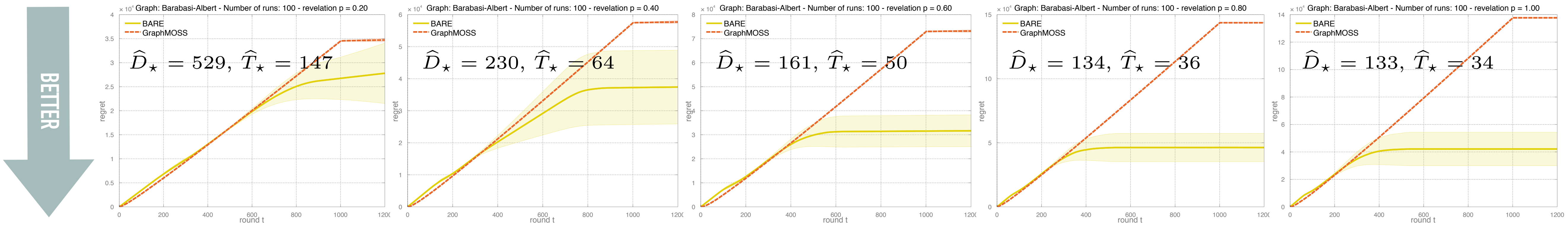


Figure 1: *Left*: Barabási-Albert. *Middle left*: Facebook. *Middle right*: Enron. *Right*: Gnutella.

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



Varying a (constant) probability of influence

REVEALING BANDITS: WHAT DO YOU MEAN?

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

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

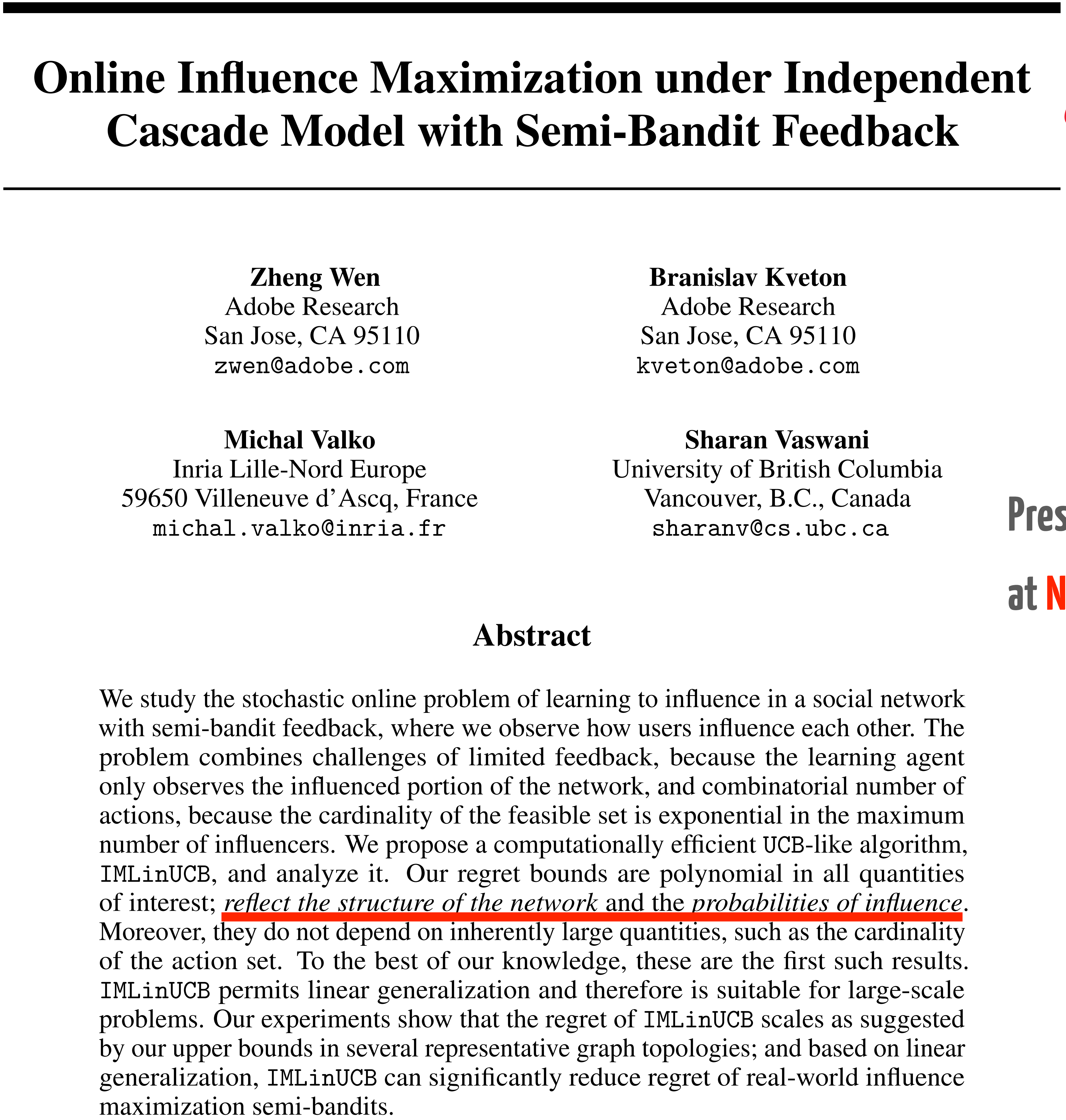
Regret of BARE

$$\mathcal{O}(\sqrt{r_\star T D_\star})$$

- **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

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)
    - Wen, Kveton, Valko, Vaswani, **NIPS 2017**
    - IMLinUCB with linear parametrization of edge weights
    - Regret analysis for **general graphs, cascading model, and multiple-sources**



Online Influence Maximization under Independent Cascade Model with Semi-Bandit Feedback

Zheng Wen
Adobe Research
San Jose, CA 95110
zwen@adobe.com

Branislav Kveton
Adobe Research
San Jose, CA 95110
kveton@adobe.com

Michal Valko
Inria Lille-Nord Europe
59650 Villeneuve d'Ascq, France
michal.valko@inria.fr

Sharan Vaswani
University of British Columbia
Vancouver, B.C., Canada
sharanv@cs.ubc.ca

Presented 5 days ago

at NIPS 2017, Long Beach, CA

Abstract

We study the stochastic online problem of learning to influence in a social network with semi-bandit feedback, where we observe how users influence each other. The problem combines challenges of limited feedback, because the learning agent only observes the influenced portion of the network, and combinatorial number of actions, because the cardinality of the feasible set is exponential in the maximum number of influencers. We propose a computationally efficient UCB-like algorithm, IMLinUCB, and analyze it. Our regret bounds are polynomial in all quantities of interest; *reflect the structure of the network* and the *probabilities of influence*. Moreover, they do not depend on inherently large quantities, such as the cardinality of the action set. To the best of our knowledge, these are the first such results. IMLinUCB permits linear generalization and therefore is suitable for large-scale problems. Our experiments show that the regret of IMLinUCB scales as suggested by our upper bounds in several representative graph topologies; and based on linear generalization, IMLinUCB can significantly reduce regret of real-world influence maximization semi-bandits.

CHALLENGES AND SOLUTIONS

seed set

seed size

$$\max_{\mathcal{S}:\,|\mathcal{S}|=K} f(\mathcal{S}, \overline{w})$$

- **Already the offline problem is NP hard**
  - solution: **approximation/randomized algorithms**
- **Lots of edges**
  - lots of parameters to learn, if we want to scale, we need to reduce this complexity
  - solution: **linear approximation of probabilities**
- **Combinatorial size of possible seed-sets**
  - Combinatorial Bandits: IMLinUCB
- **Understanding what's going on?**
  - known analyses VERY loose (e.g., scaling with 1/$p_{\min}$, or only assymptotic)

APPROXIMATION ORACLE

- the optimal offline solution

seed size

$$\max_{\mathcal{S}:\,|\mathcal{S}|=K} f(\mathcal{S}, \overline{w})$$

- the oracle solution that is γ-optimal with probability at least α

$$\mathcal{S}^{*} = \mathtt{ORACLE}(\mathcal{G}, K, \overline{w})$$

- γ-optimal

$$f(\mathcal{S}^{*}, \overline{w}) \geq \gamma f(\mathcal{S}^{\text{opt}}, \overline{w})$$

- γ-optimal with probability at least α

$$\mathbb{E}\left[f(\mathcal{S}^{*}, \overline{w})\right] \geq \alpha\gamma f(\mathcal{S}^{\text{opt}}, \overline{w})$$

- Our problem is a triple:

unknown to the agent

$$(\mathcal{G}, K, \overline{w})$$

topology

seed size

LINEAR GENERALIZATION

— learning the only network (weights) is VERY impractical

$$\rho \triangleq \max_{e\in\mathcal{E}} \left|\overline{w}(e) - x_e^{\top}\theta^*\right|$$

this is small

true weights

linear approximation

— by choosing the dimension (size of $\theta*$) we can reduce this complexity

— if we do not want to lose generality we set **d** to the number of edges

ALGORITHM AND PERFORMANCE MEASURE

Algorithm 1 `IMLinUCB`: Influence Maximization Linear UCB

```
Input: graph $\mathcal{G}$, source node set cardinality $K$, oracle ORACLE, feature vector $x_e$'s, and algorithm parameters $\sigma, c > 0$,
Initialization: $B_0 \leftarrow 0 \in \Re^d$, $\mathbf{M}_0 \leftarrow I \in \Re^{d \times d}$
for $t = 1, 2, \dots, n$ do
    1. set $\bar{\theta}_{t-1} \leftarrow \sigma^{-2}\mathbf{M}_{t-1}^{-1} B_{t-1}$ and the UCBs as $U_t(e) \leftarrow \mathrm{Proj}_{[0,1]}\left( x_e^\top \bar{\theta}_{t-1} + c\sqrt{x_e^\top \mathbf{M}_{t-1}^{-1} x_e} \right)$
    for all $e \in \mathcal{E}$
    2. choose $\mathcal{S}_t \in$ ORACLE$(\mathcal{G}, K, U_t)$, and observe the edge-level semi-bandit feedback
    3. update statistics:
        (a) initialize $\mathbf{M}_t \leftarrow \mathbf{M}_{t-1}$ and $B_t \leftarrow B_{t-1}$
        (b) for all observed edges $e \in \mathcal{E}$, update $\mathbf{M}_t \leftarrow \mathbf{M}_t + \sigma^{-2} x_e x_e^\top$ and $B_t \leftarrow B_t + x_e \mathbf{w}_t(e)$
```

$$R^\eta(n) = \sum_{t=1}^n \mathbb{E}\left[R_t^\eta\right]$$

$$R_t^\eta = f(\mathcal{S}^{\mathrm{opt}}, \mathbf{w}_t) - \frac{1}{\eta} f(\mathcal{S}_t, \mathbf{w}_t)$$

MAXIMUM OBSERVED RELEVANCE

$N_{\mathcal{S},e} \triangleq \sum_{v\in\mathcal{V}\setminus\mathcal{S}} \mathbf{1}\{e \text{ is relevant to } v \text{ under } \mathcal{S}\}$ and $P_{\mathcal{S},e} \triangleq \mathbb{P}\left(e \text{ is observed} \mid \mathcal{S}\right)$

only depends on topology

depends on both

$$C_* \triangleq \max_{\mathcal{S}:\,|\mathcal{S}|=K} \sqrt{\sum_{e\in\mathcal{E}} N^2_{\mathcal{S},e} P_{\mathcal{S},e}}$$

max (over) 2-norm of N weighted by P

- Worst-case upper bound:

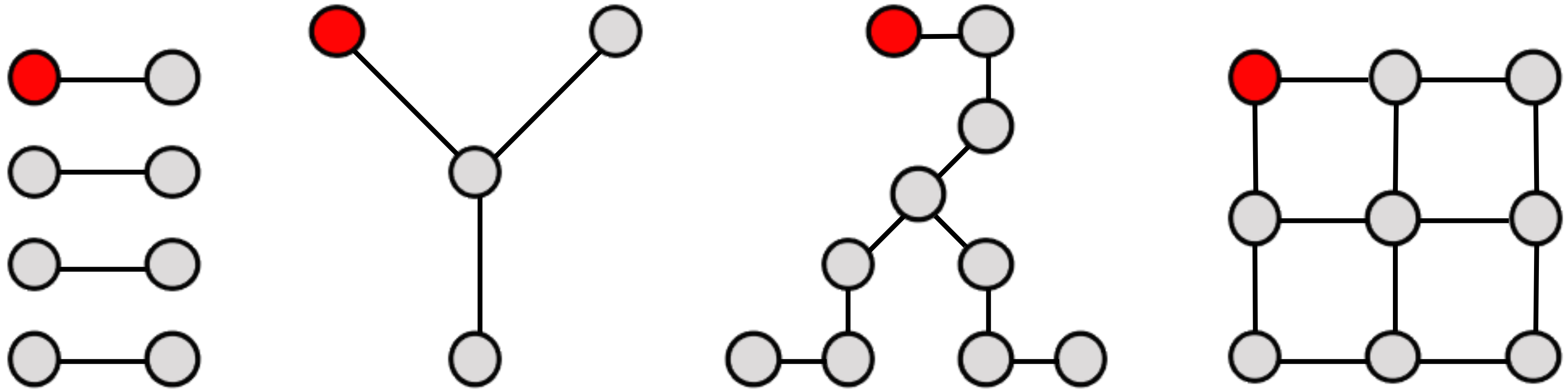
#nodes

#edges

$$C_* \le C_{\mathcal{G}} \triangleq \max_{\mathcal{S}:\,|\mathcal{S}|=K} \sqrt{\sum_{e\in\mathcal{E}} N^2_{\mathcal{S},e}} \le (L-K)\sqrt{|\mathcal{E}|} = \mathcal{O}\left(L\sqrt{|\mathcal{E}|}\right) = \mathcal{O}\left(L^2\right)$$

seed size

WORST-CASE BOUNDS



(a) (b) (c) (d)

L = number of nodes, K = size of seed set, n = number of rounds, d = dimension

topology	$C_{\mathcal{G}}$ (worst-case C_*)	$R^{\alpha\gamma}(n)$ for general $\mathbf{X}$	$R^{\alpha\gamma}(n)$ for $\mathbf{X} = \mathbf{I}$
bar graph	$\mathcal{O}(\sqrt{K})$	$\widetilde{\mathcal{O}}\left(dK\sqrt{n}/(\alpha\gamma)\right)$	$\widetilde{\mathcal{O}}\left(L\sqrt{Kn}/(\alpha\gamma)\right)$
star graph	$\mathcal{O}(L\sqrt{K})$	$\widetilde{\mathcal{O}}\left(dL^{\frac{3}{2}}\sqrt{Kn}/(\alpha\gamma)\right)$	$\widetilde{\mathcal{O}}\left(L^2\sqrt{Kn}/(\alpha\gamma)\right)$
ray graph	$\mathcal{O}(L^{\frac{5}{4}}\sqrt{K})$	$\widetilde{\mathcal{O}}\left(dL^{\frac{7}{4}}\sqrt{Kn}/(\alpha\gamma)\right)$	$\widetilde{\mathcal{O}}\left(L^{\frac{9}{4}}\sqrt{Kn}/(\alpha\gamma)\right)$
tree graph	$\mathcal{O}(L^{\frac{3}{2}})$	$\widetilde{\mathcal{O}}\left(dL^2\sqrt{n}/(\alpha\gamma)\right)$	$\widetilde{\mathcal{O}}\left(L^{\frac{5}{2}}\sqrt{n}/(\alpha\gamma)\right)$
grid graph	$\mathcal{O}(L^{\frac{3}{2}})$	$\widetilde{\mathcal{O}}\left(dL^2\sqrt{n}/(\alpha\gamma)\right)$	$\widetilde{\mathcal{O}}\left(L^{\frac{5}{2}}\sqrt{n}/(\alpha\gamma)\right)$
complete graph	$\mathcal{O}(L^2)$	$\widetilde{\mathcal{O}}\left(dL^3\sqrt{n}/(\alpha\gamma)\right)$	$\widetilde{\mathcal{O}}\left(L^4\sqrt{n}/(\alpha\gamma)\right)$

Table 1: $C_{\mathcal{G}}$ and *worst-case* regret bounds for different graph topologies

RESULTS

$$R^{\alpha\gamma}(n) \leq \frac{2cC_*}{\alpha\gamma}\sqrt{dn|\mathcal{E}|\log_2\left(1+\frac{n|\mathcal{E}|}{d}\right)} + 1 = \widetilde{\mathcal{O}}\left(dC_*\sqrt{|\mathcal{E}|n}/(\alpha\gamma)\right)$$

$$\leq \widetilde{\mathcal{O}}\left(d(L-K)|\mathcal{E}|\sqrt{n}/(\alpha\gamma)\right).$$

How good (tight) is this?

- comparison with linear bandits
- comparison with general combinatorial bandits
- (L-K) factor
- **How good is C*?**

PROOF SKETCH?

- when are our upper bounds on the estimates right?

$$\xi_{t-1} = \{|x_e^\mathsf{T}(\overline{\theta}_{\tau-1} - \theta^*)| \le c\sqrt{x_e^\mathsf{T}\mathbf{M}_{\tau-1}^{-1}x_e}, \forall e \in \mathcal{E}, \forall \tau \le t\}$$

- .... decomposes the regret at round t

$$\mathbb{E}[R_t^{\alpha\gamma}] \le \mathbb{P}\left(\xi_{t-1}\right)\mathbb{E}\left[R_t^{\alpha\gamma}|\xi_{t-1}\right] + \mathbb{P}\left(\overline{\xi}_{t-1}\right)[L-K]$$

- monotonicity of f

$$\mathbb{E}\left[R_t^{\alpha\gamma}|\xi_{t-1}\right] \le \mathbb{E}\left[f(\mathcal{S}_t, U_t) - f(\mathcal{S}_t, \overline{w})|\xi_{t-1}\right]/(\alpha\gamma)$$

decomposed into nodes

- studying the Markov process of propagation
  - consider non-overlaping layers
  - random stopping time

$$f(\mathcal{S}_t, U_t, v) - f(\mathcal{S}_t, \overline{w}, v) \le \sum_{e\in\mathcal{E}_{\mathcal{S}_t,v}} \mathbb{E}\left[\mathbf{1}\left\{O_t(e)\right\}\left[U_t(e) - \overline{w}(e)\right]|\mathcal{H}_{t-1}, \mathcal{S}_t\right]$$

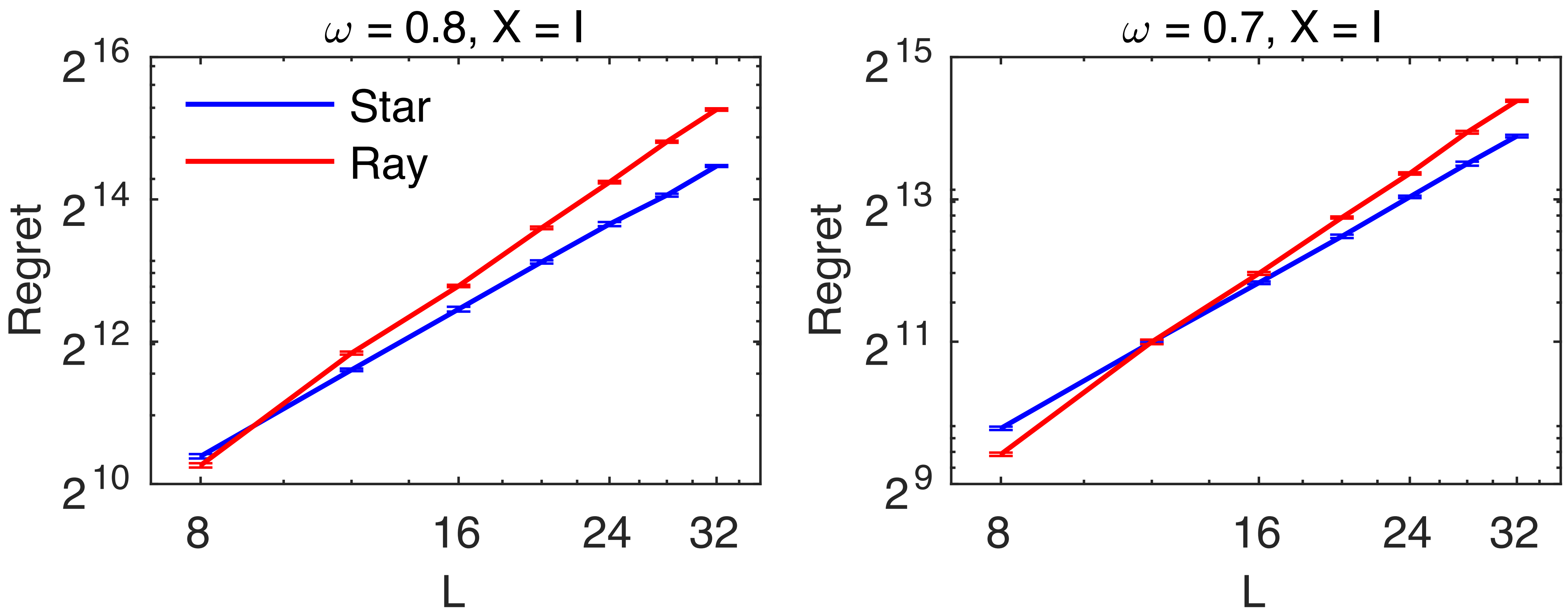
probability that node v is influences

EXPERIMENTS

- **Objective:** "Check" how good is our C*
- Tabular case, K = 1, exact comparison possible, all weights are same = ω

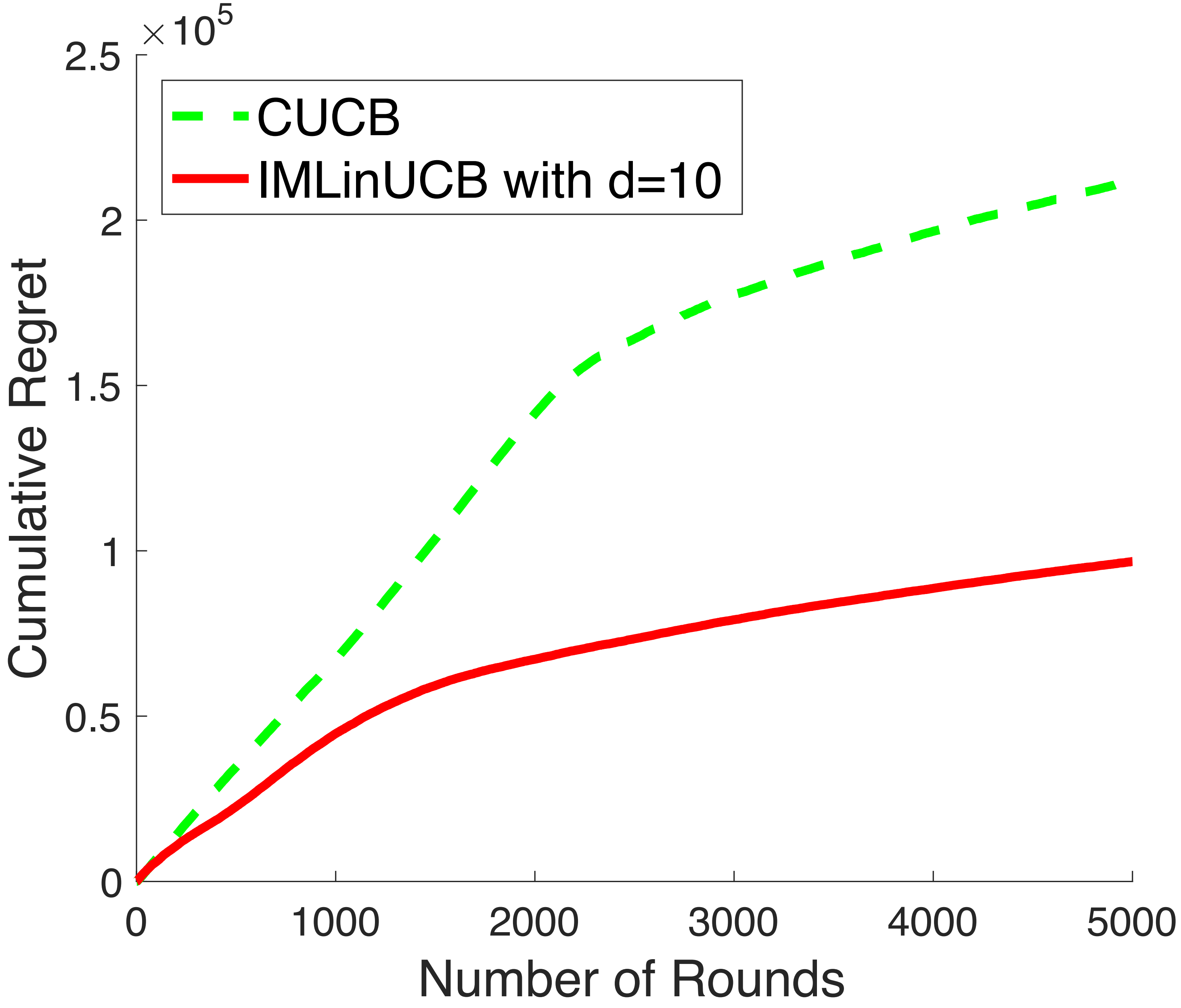
Star $\widetilde{\mathcal{O}}(L^2)$ vs. $\mathcal{O}(L^{2.040})$ and $\mathcal{O}(L^{2.056})$

Ray $\widetilde{\mathcal{O}}(L^{\frac{9}{4}})$ vs. $\mathcal{O}(L^{2.488})$ and $\mathcal{O}(L^{2.467})$



- **Conclusion:** evidence that our C* is a reasonable complexity measure

FACEBOOK EXPERIMENT



- real Facebook (a small subgraph)
- weights from U(0,0.1)
- **nodetovec** with d=10
  - imperfect
- K = 10
- CUCB with no linear generalisation

CONCLUSION AND NEXT STEPS

- **Active learning on graphs**
  - learning the graph **while** acting on it optimal
  - **difficulty of the problem** and scaling with it
  - online influence maximization
    - local model (minimax optimal algorithm)
    - global cascading model
- **What is next?**
  - dynamic/evolving graphs
  - realistic accessibility constraints

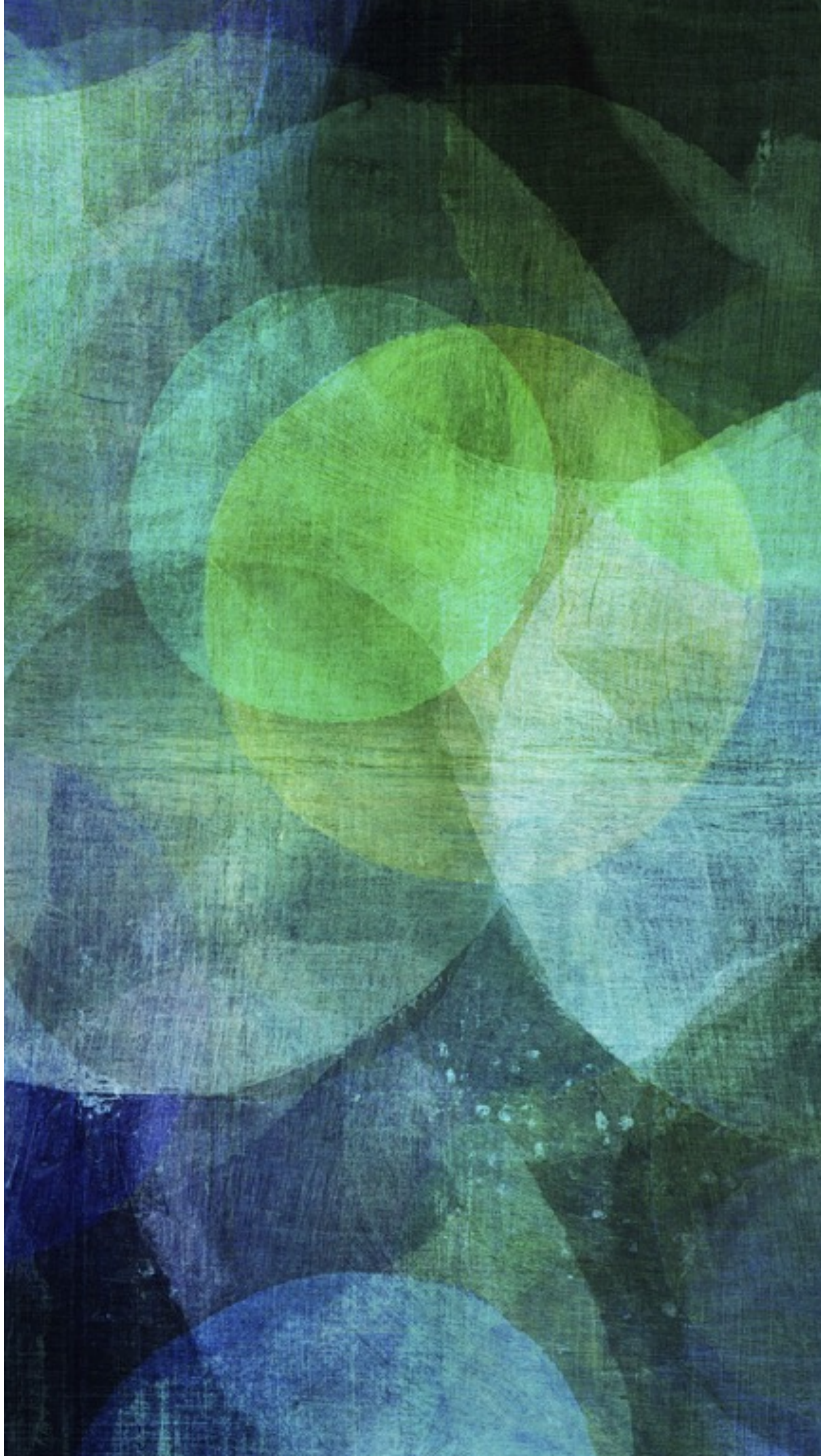
Kocák, Neu, MV, Munos: **Efficient learning by implicit exploration in bandit problems with side observations**, NIPS 2014

Kocák, Neu, MV: **Online learning with Erdos-Rényi side-observation graphs** UAI 2016

Kocák, Neu, MV: **Online learning with noisy side observations**, AISTATS 2016

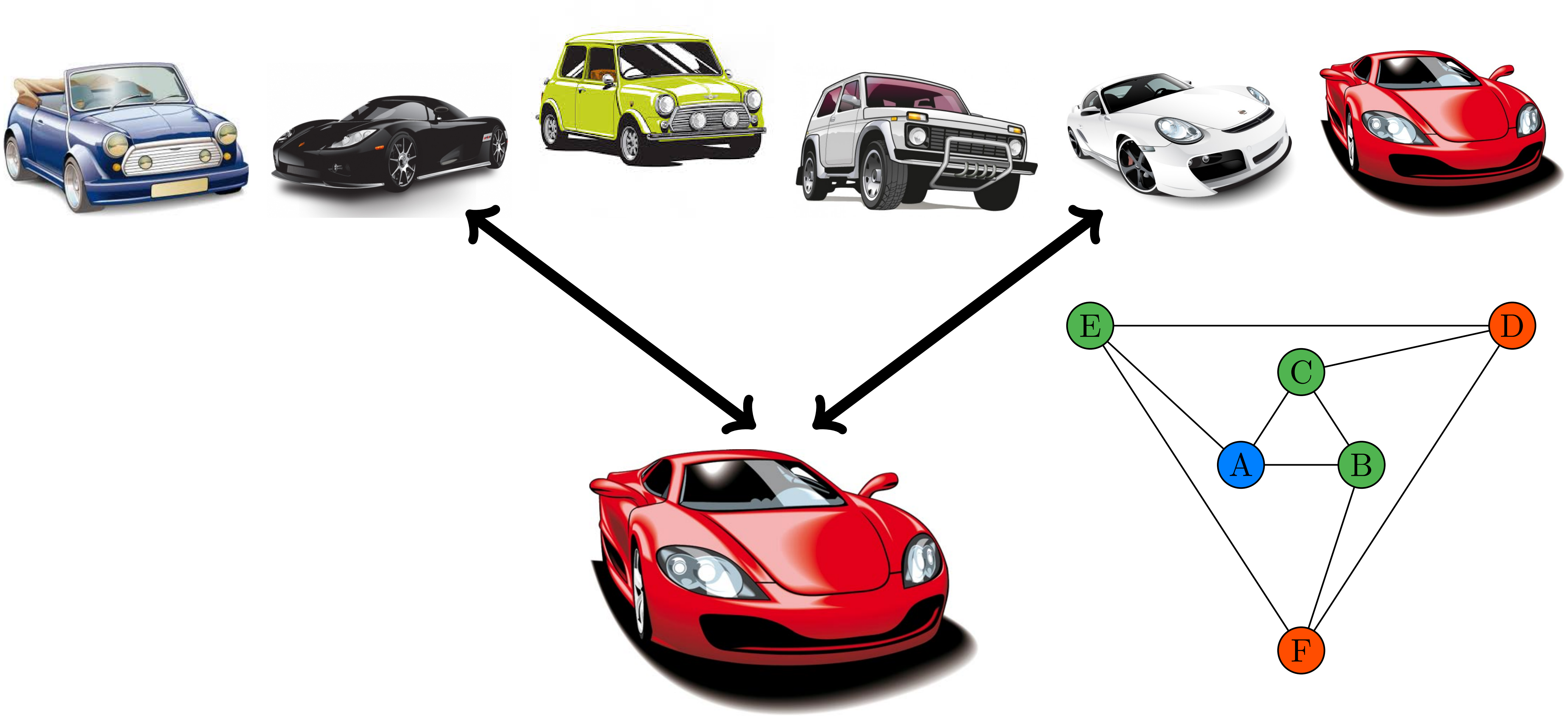
GRAPH BANDITS WITH SIDE OBSERVATIONS

exploiting **free** observations from neighbouring nodes



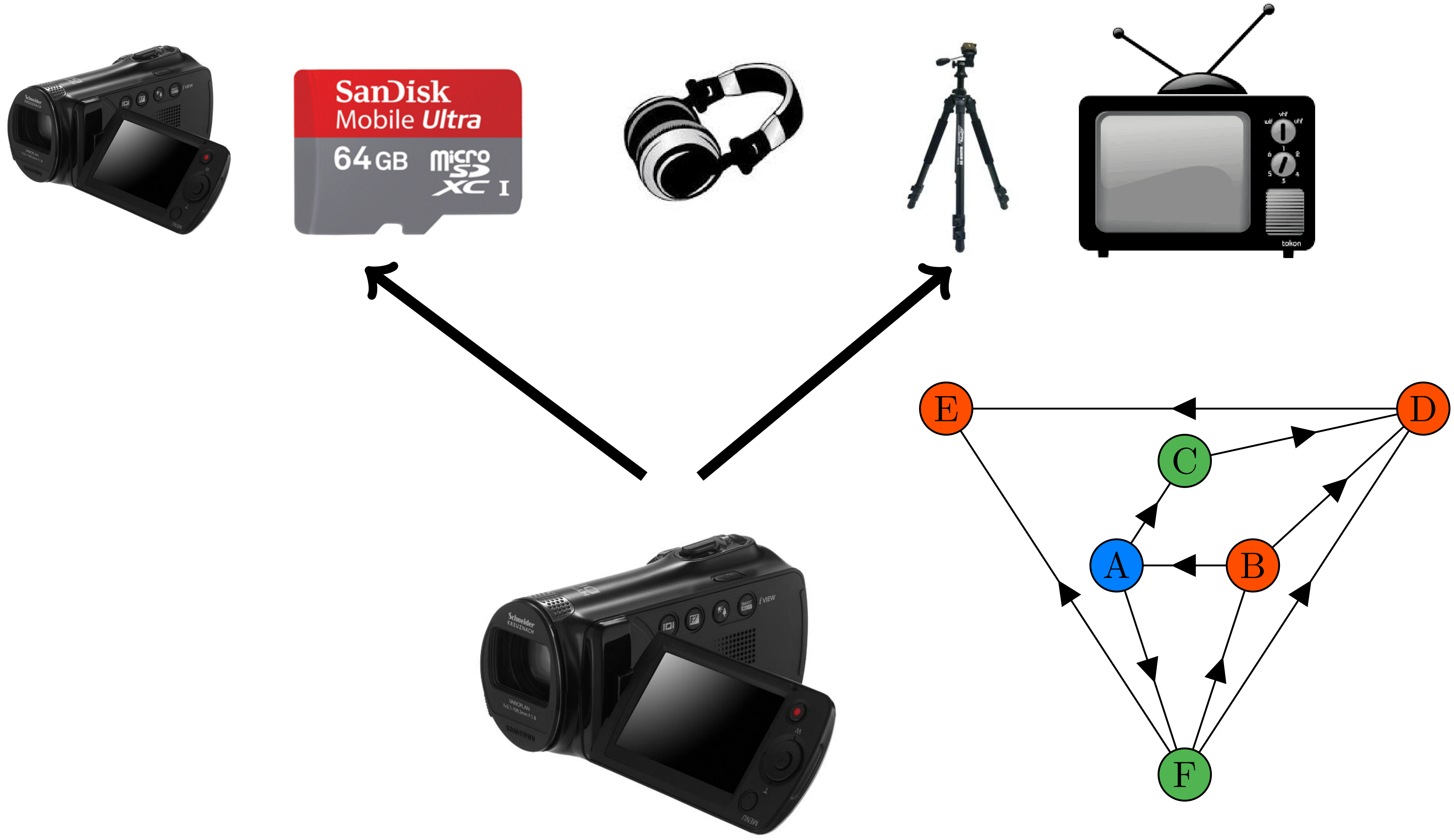
SIDE OBSERVATIONS: UNDIRECTED

Example 1: undirected observations



SIDE OBSERVATIONS: DIRECTED

Example 2: Directed observation



SIDE OBSERVATIONS – AN INTERMEDIATE GAME

Full-information

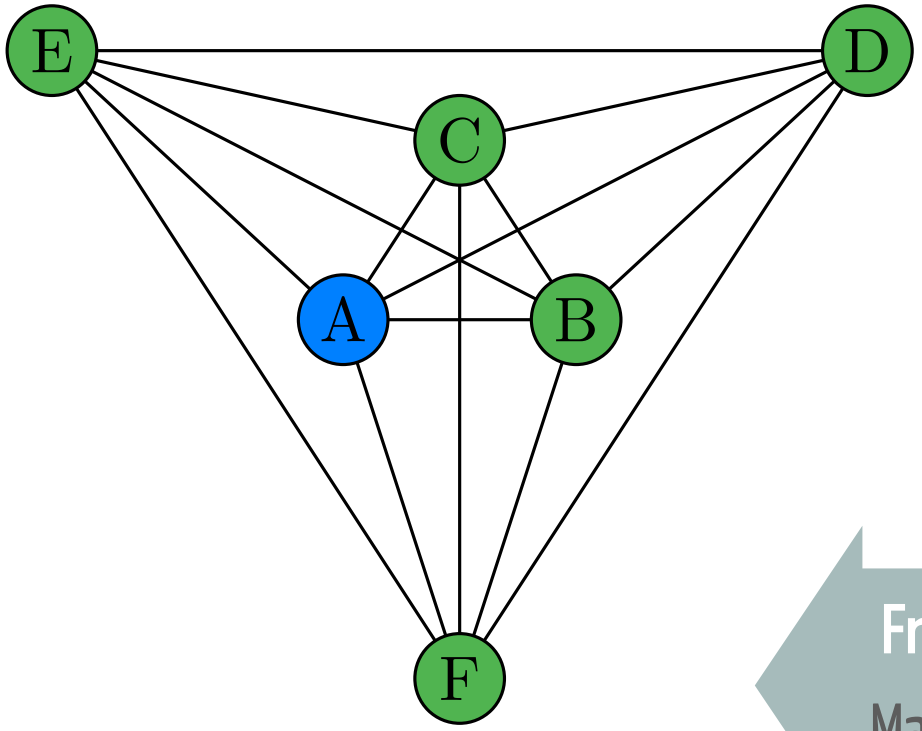
- observe losses of **all** actions
- example: Hedge

$$R_T = \widetilde{\mathcal{O}}(\sqrt{T})$$

Bandits

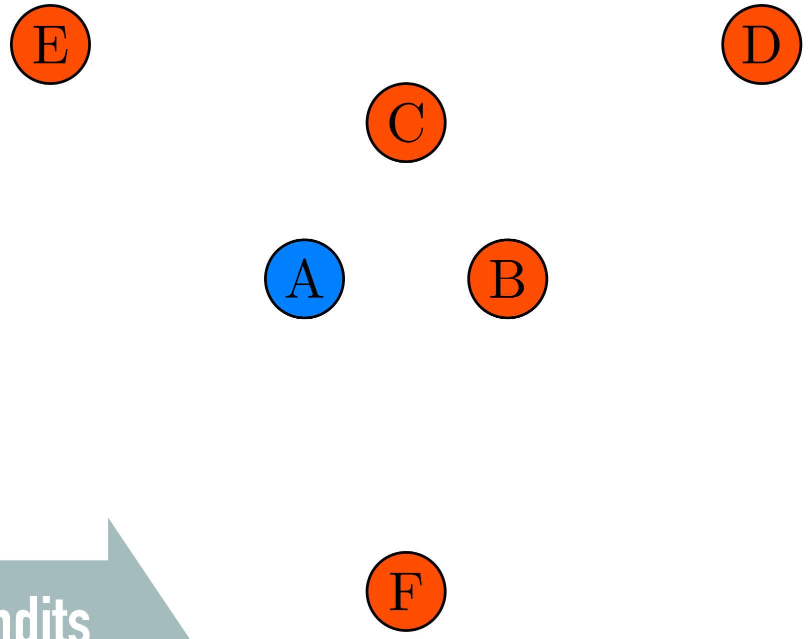
- observe losses of **the chosen** action
- example: EXP3

$$R_T = \widetilde{\mathcal{O}}(\sqrt{NT})$$



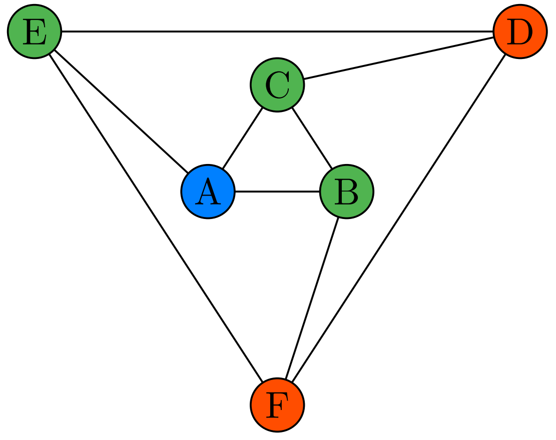
From Experts to Bandits

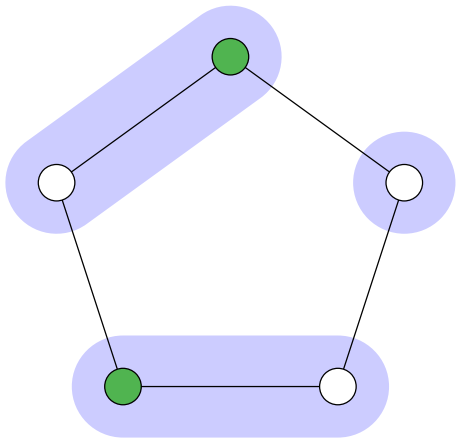
Mannor and Shamir 2011

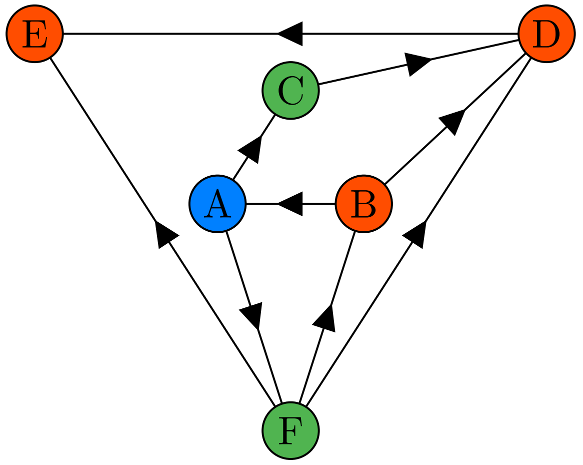


KNOWLEDGE OF OBSERVATION GRAPHS

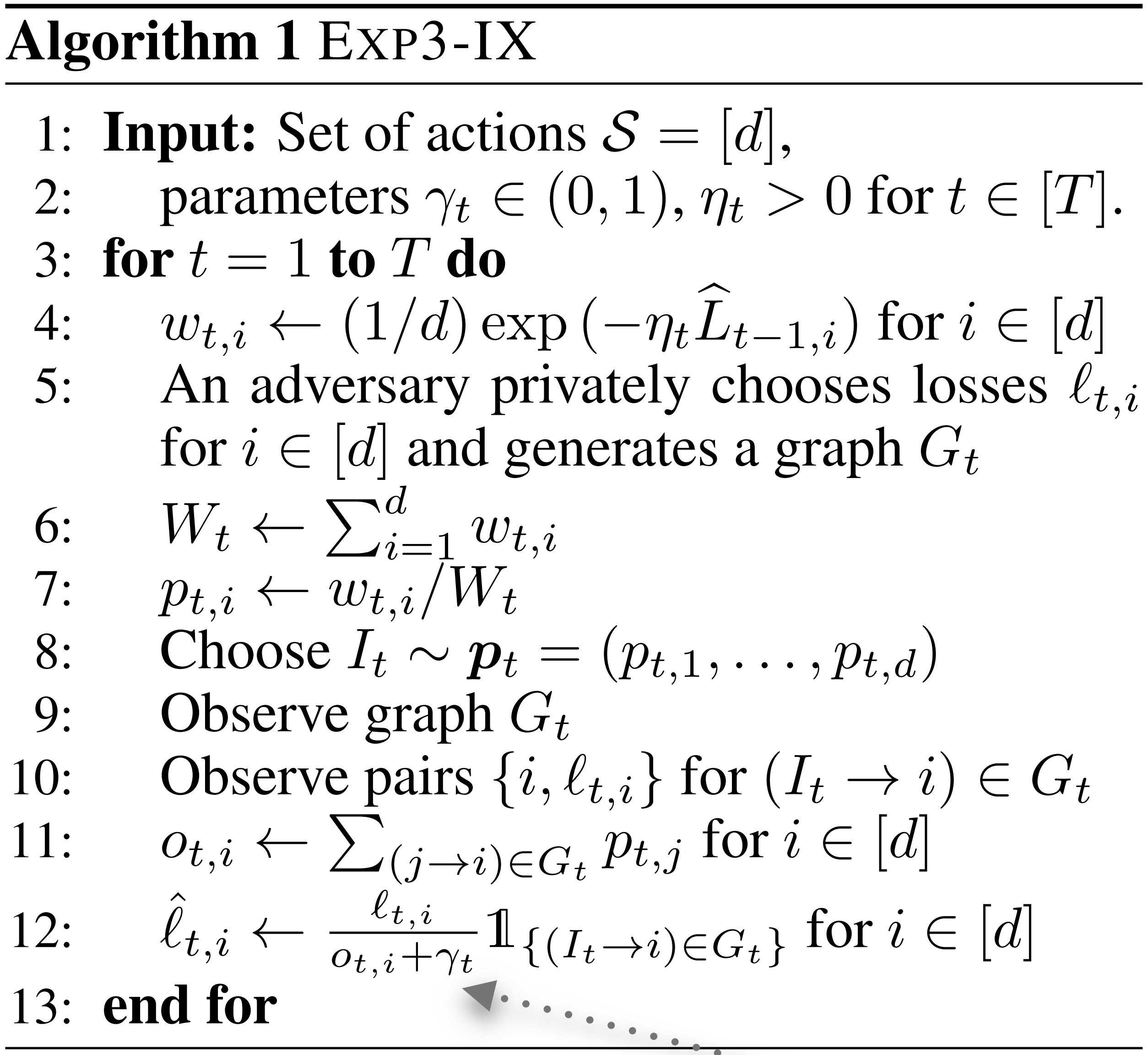
- ELP (Mannor and Shamir 2011)
  - **EXP3** - with **“LP balanced exploration”**
  - undirected $O(\sqrt{(\alpha T)})$ ✅ – needs to know G_t
  - directed case $O(\sqrt{(cT)})$ – needs to know G_t
- EXP3-SET (Alon, Cesa-Bianchi, Gentile, Mansour, 2013)
  - undirected $O(\sqrt{(\alpha T)})$ ✅ does not need to know G_t ✅
- EXP3-DOM (Alon, Cesa-Bianchi, Gentile, Mansour, 2013)
  - directed $O(\sqrt{(\alpha T)})$ ✅ – need to know G_t
  - **calculates dominating set**







EXP3-IX: IMPLICIT EXPLORATION



Algorithm 1 EXP3-IX

```
Input: Set of actions $\mathcal{S} = [d]$,
    parameters $\gamma_t \in (0,1)$, $\eta_t > 0$ for $t \in [T]$.
for $t = 1$ to $T$ do
    $w_{t,i} \leftarrow (1/d)\exp\left(-\eta_t \widehat{L}_{t-1,i}\right)$ for $i \in [d]$
    An adversary privately chooses losses $\ell_{t,i}$
    for $i \in [d]$ and generates a graph $G_t$
    $W_t \leftarrow \sum_{i=1}^{d} w_{t,i}$
    $p_{t,i} \leftarrow w_{t,i}/W_t$
    Choose $I_t \sim \boldsymbol{p}_t = (p_{t,1}, \ldots, p_{t,d})$
    Observe graph $G_t$
    Observe pairs $\{i, \ell_{t,i}\}$ for $(I_t \to i) \in G_t$
    $o_{t,i} \leftarrow \sum_{(j \to i) \in G_t} p_{t,j}$ for $i \in [d]$
    $\hat{\ell}_{t,i} \leftarrow \frac{\ell_{t,i}}{o_{t,i} + \gamma_t} \mathbb{1}_{\{(I_t \to i) \in G_t\}}$ for $i \in [d]$
end for
```

Benefits of the **implicit exploration**

- **no need to know the graph before**
- no need to estimate dominating set
- no need for doubling trick
- no need for aggregation

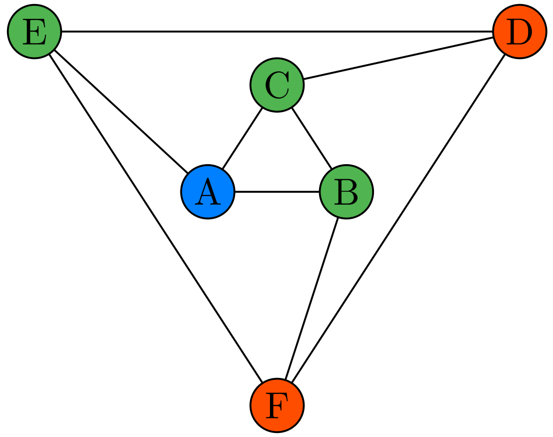
$$R_T = \widetilde{\mathcal{O}}\left(\sqrt{\overline{\alpha}\, T \ln N}\right)$$

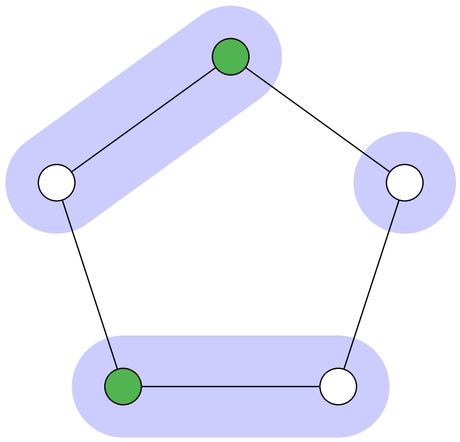
Optimistic bias for the loss estimates

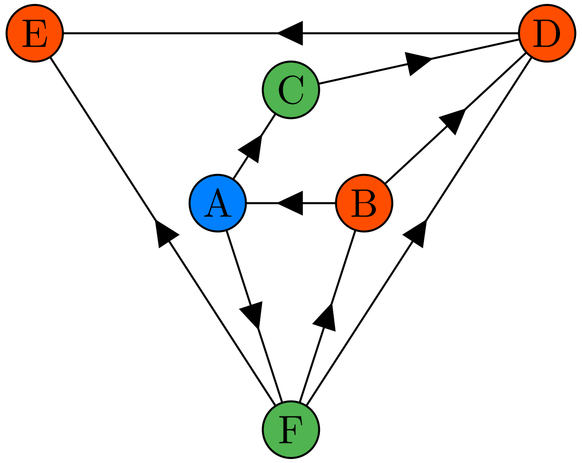
$$\mathbb{E}[\hat{\ell}_{t,i}] = \frac{\ell_{t,i}}{o_{t,i} + \gamma} o_{t,i} + 0(1 - o_{t,i}) = \ell_{t,i} - \ell_{t,i} \frac{\gamma}{o_{t,i} + \gamma} \leq \ell_{t,i}$$

FOLLOW UPS

- EXP3-IX (Kocák, Neu, MV, Munos, 2014)
  - directed $O(\sqrt{(\alpha T)})$ ✅ does not need to know G_t ✅
- EXP3.G (Alon, Cesa-Bianchi, Dekel, Koren, 2015)
  - directed $O(\sqrt{(\alpha T)})$ ✅ does not need to know G_t ✅
  - mixes uniform distribution
  - more general algorithm for settings **beyond bandits**
  - high-probability bound
- Neu 2015: high-probability bound for EXP3-IX

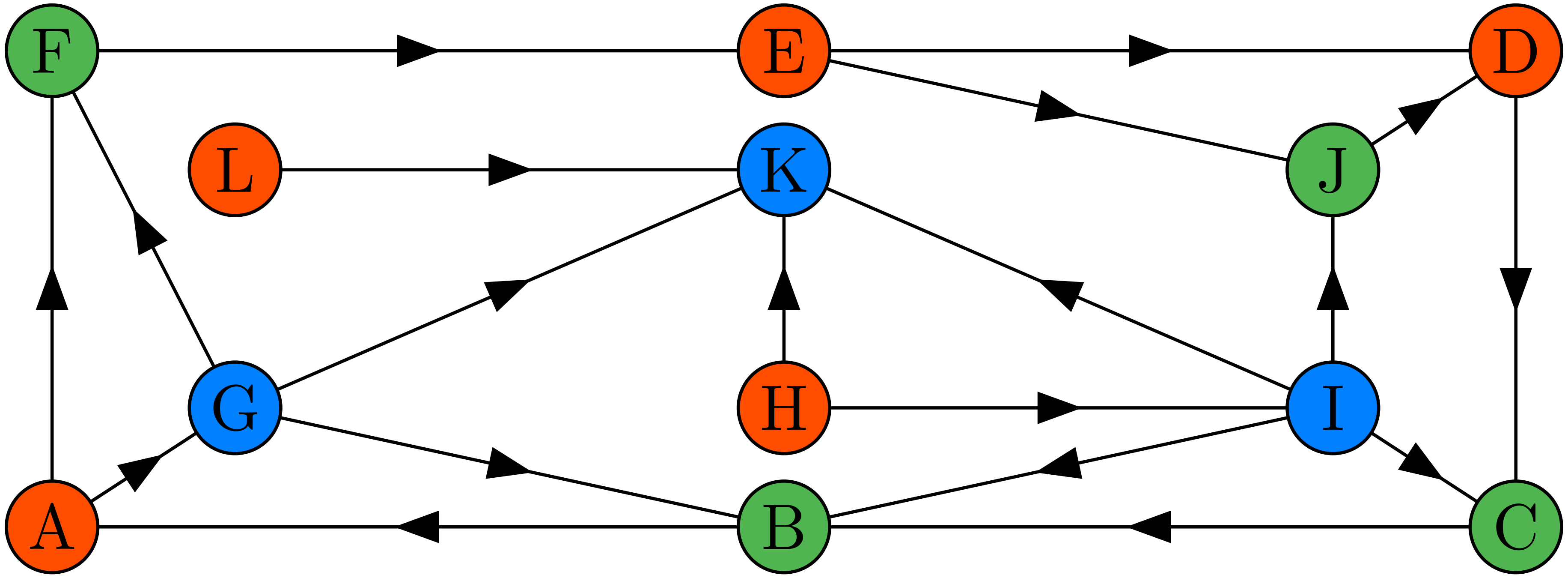






COMPLEX GRAPH ACTIONS

Example: online shortest path semi-bandits with observing traffic on the side streets



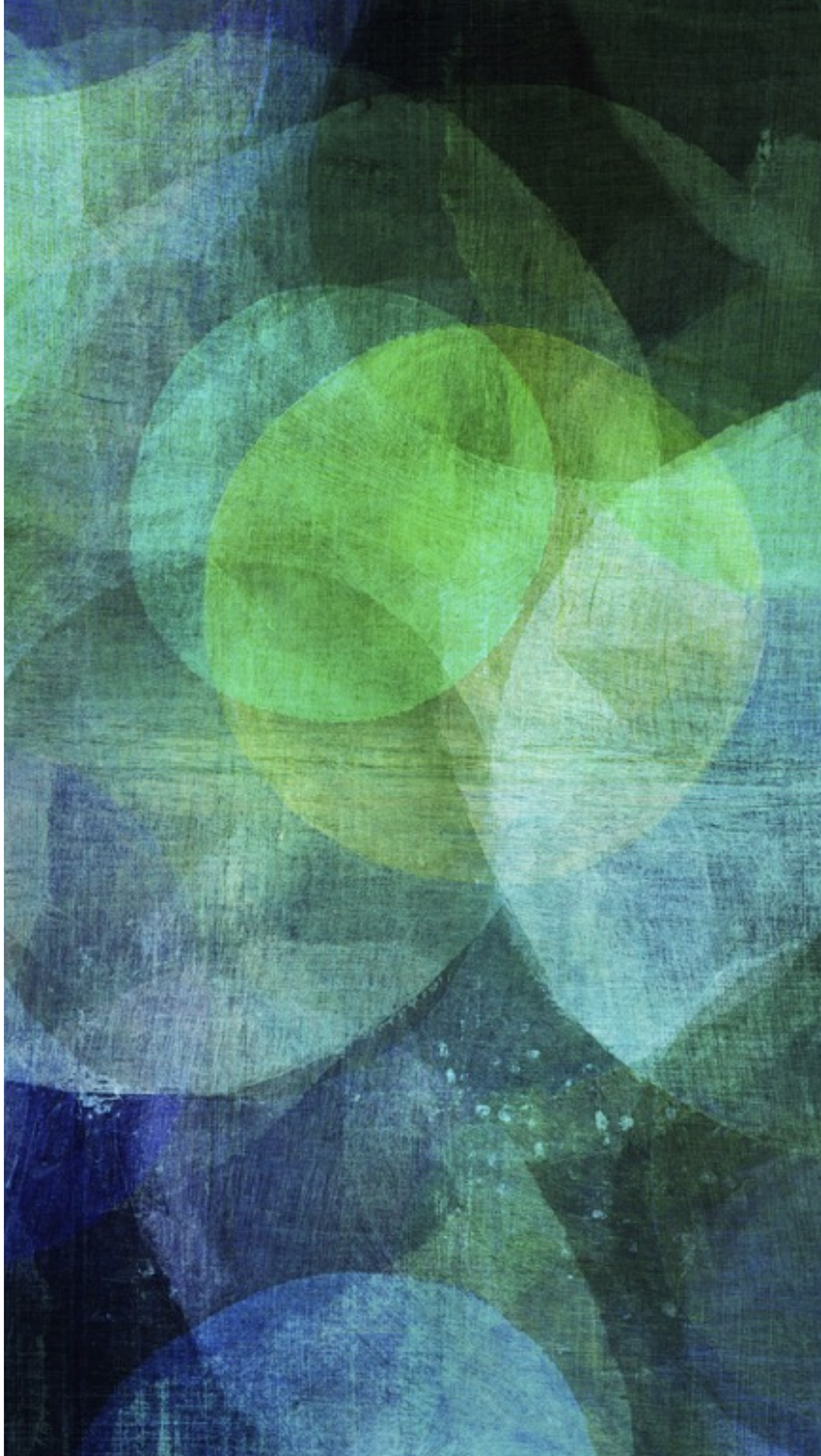
- Play action $\mathbf{V}_t \in S \subset \{0,1\}^N$, $\|\mathbf{v}\|_1 \leq m$ from all $\mathbf{v} \in S$
- Obtain losses $\mathbf{V}_t^\mathsf{T} \boldsymbol{\ell}_t$
- Observe additional losses according to the graph

$$R_T = \widetilde{\mathcal{O}}\left(m^{3/2}\sqrt{\sum_{t=1}^{T}\alpha_t}\right) = \widetilde{\mathcal{O}}\left(m^{3/2}\sqrt{\alpha}T\right)$$

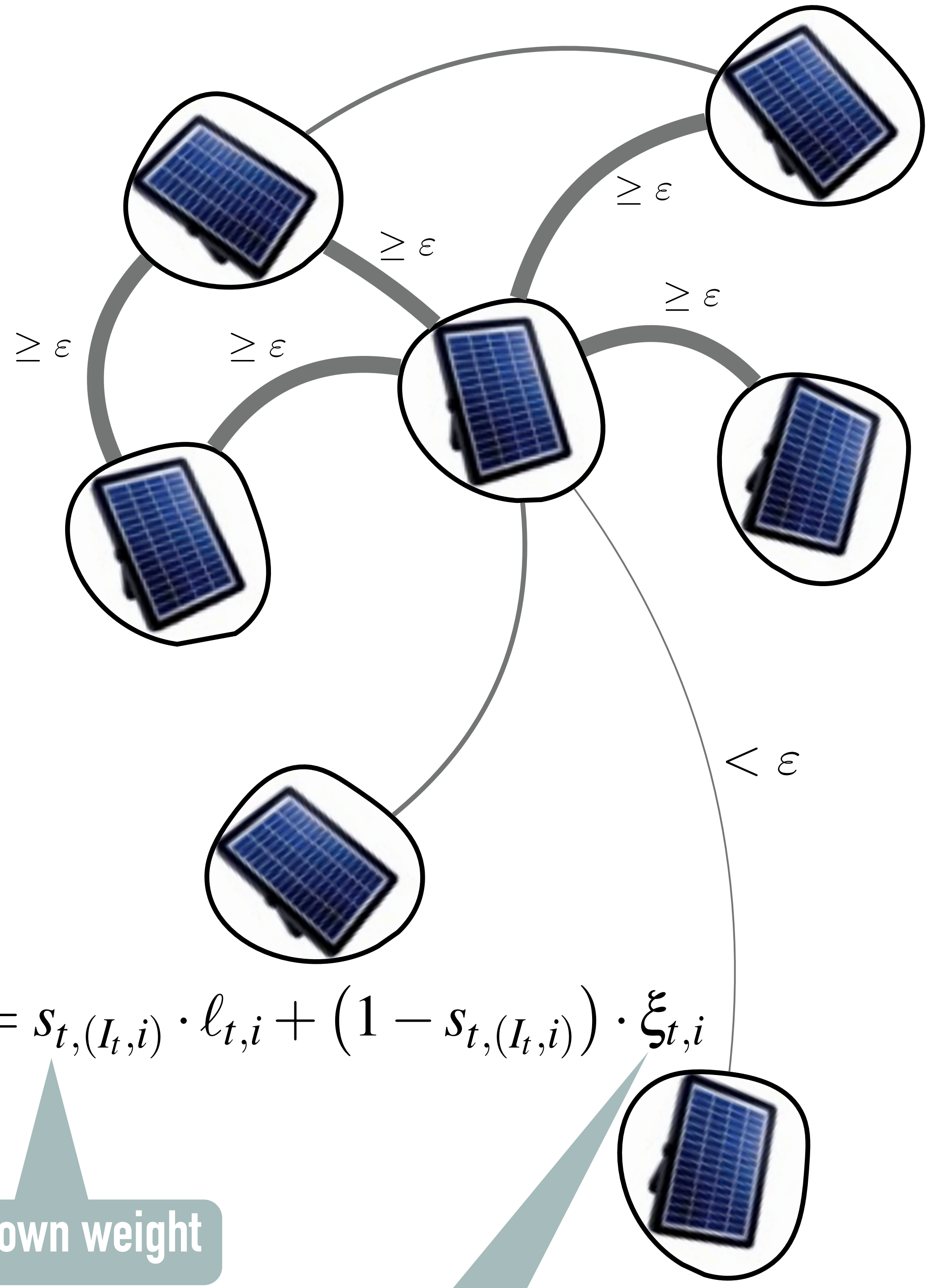
Kocák, Neu, MV: **Online learning with noisy side observations**, AISTATS 2016

GRAPH BANDITS WITH NOISY SIDE OBSERVATIONS

exploiting side observations that can be perturbed by certain level of noise



NOISY SIDE OBSERVATIONS



$$c_{t,i} = s_{t,(I_t,i)} \cdot \ell_{t,i} + \left(1 - s_{t,(I_t,i)}\right) \cdot \xi_{t,i}$$

known weight

zero-mean noise

Want: only reliable information!

1) If we know the perfect cutoff ε

- reliable: use as exact
- unreliable: rubbish

then we can improve over pure bandit setting!

2) Treating noisy observation induces bias

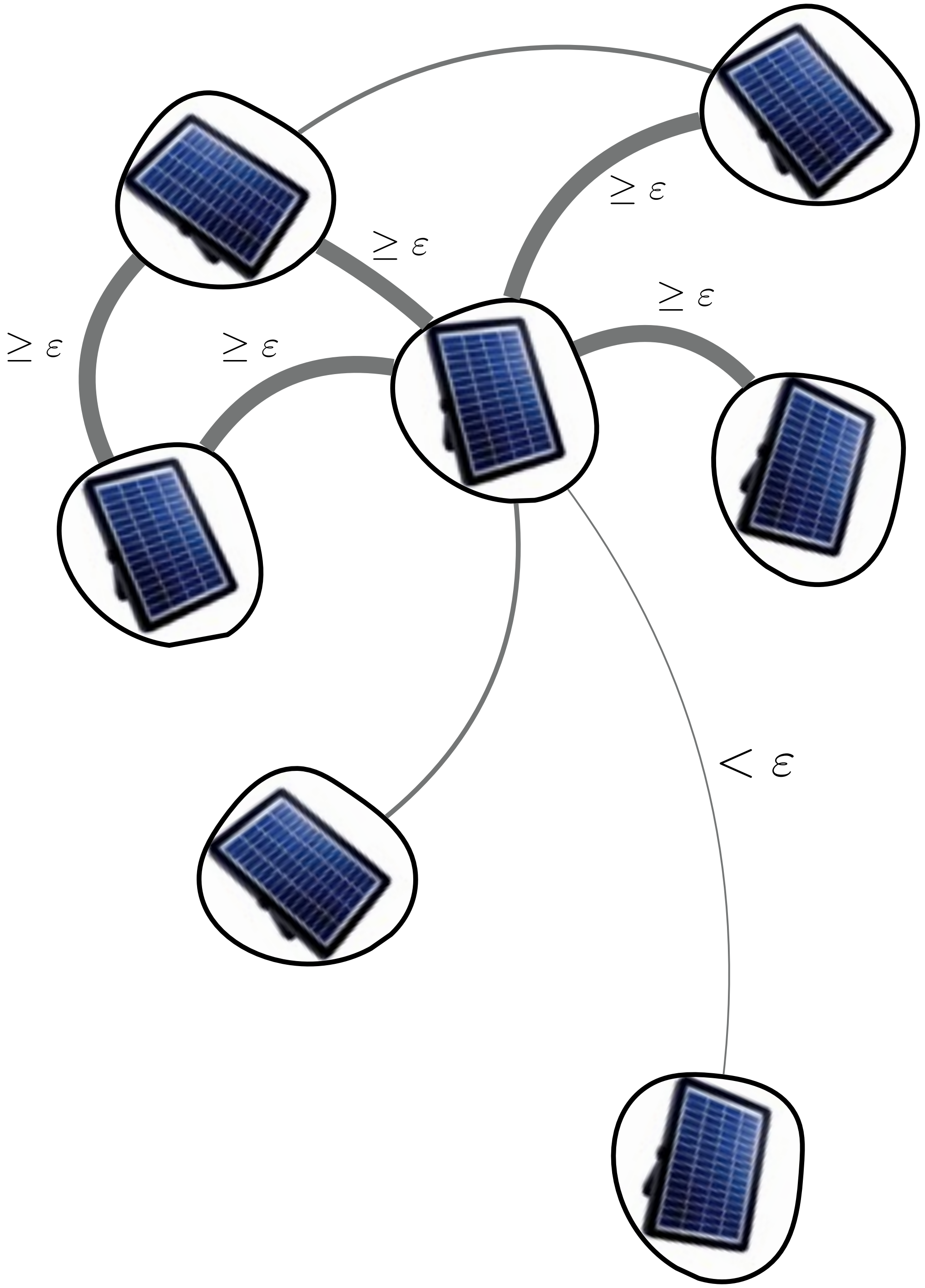
What can we hope for?

$$\tilde{\mathcal{O}}\left(\sqrt{1T}\right) \le \qquad \le \tilde{\mathcal{O}}\left(\sqrt{NT}\right)$$

effective independence number

Can we learn without knowing either ε or α^* ?

NOISY SIDE OBSERVATIONS



Threshold estimate $R_T = \widetilde{\mathcal{O}}\left(\sqrt{\alpha^\star T}\right)$

$$\widehat{\ell}^{(\mathrm{T})}_{t,i} = \frac{c_{t,i}\mathbb{I}_{\{s_{t,(I_t,i)}\geq\varepsilon_t\}}}{\sum_{j=1}^N p_{t,j}s_{t,(j,i)}\mathbb{I}_{\{s_{t,(j,i)}\geq\varepsilon_t\}}+\gamma_t}$$

WIX estimate $R_T = \widetilde{\mathcal{O}}\left(\sqrt{\alpha^\star T}\right)$

$$\widehat{\ell}_{t,i} = \frac{s_{t,(I_t,i)}\cdot c_{t,i}}{\sum_{j=1}^N p_{t,j}s^2_{t,(j,i)}+\gamma_t}$$

Since $\alpha^* \leq \alpha(1)/1 \leq N$

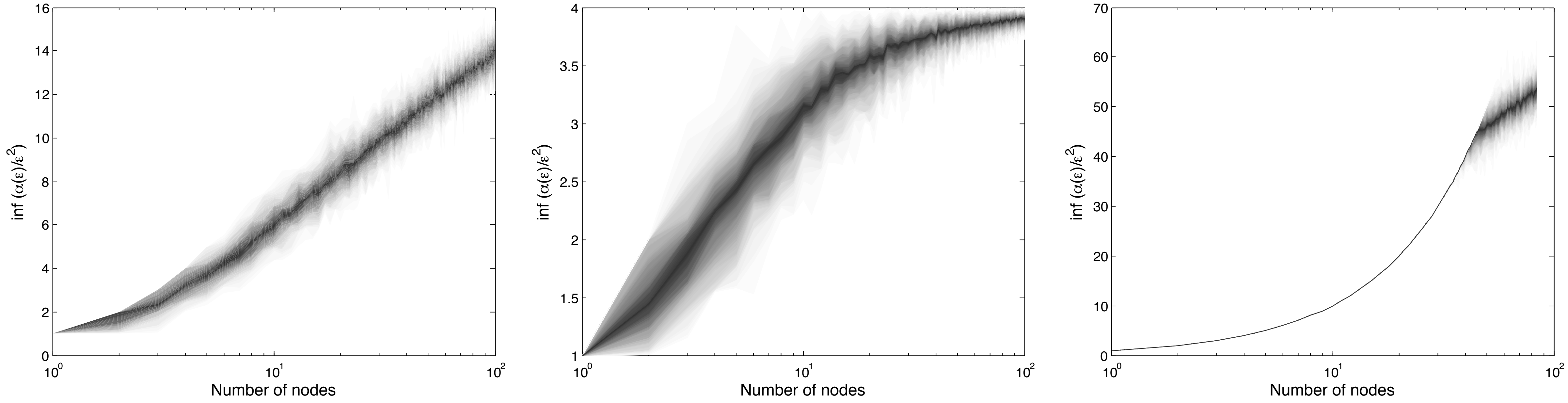
incorporating noisy observations does not hurt

$$\widetilde{\mathcal{O}}\left(\sqrt{\alpha^\star T}\right) \leq \widetilde{\mathcal{O}}\left(\sqrt{NT}\right)$$

But how much does it help?

EMPIRICAL RESULTS

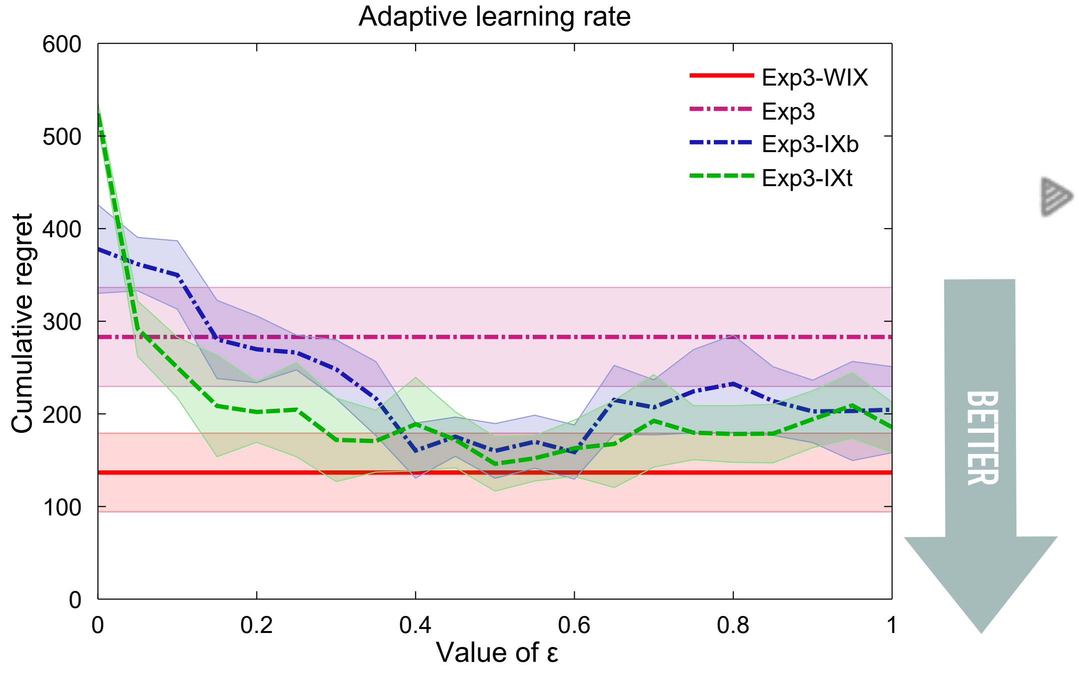
EMPIRICAL α^* FOR RANDOM GRAPHS WITH IID WEIGHTS



(a) $U(0,1)$ weights

(b) $U(\frac{1}{2},1)$ weights

(c) $U(0,\frac{1}{2})$ weights



- **special case:** if s_{ij} is either 0 or ε than $\alpha^* = \alpha/\varepsilon^2$
  - For this special case, there is a matches $\Theta(\sqrt{(\alpha T)}/\varepsilon)$ by Wu, György, Szepesvári, 2015.

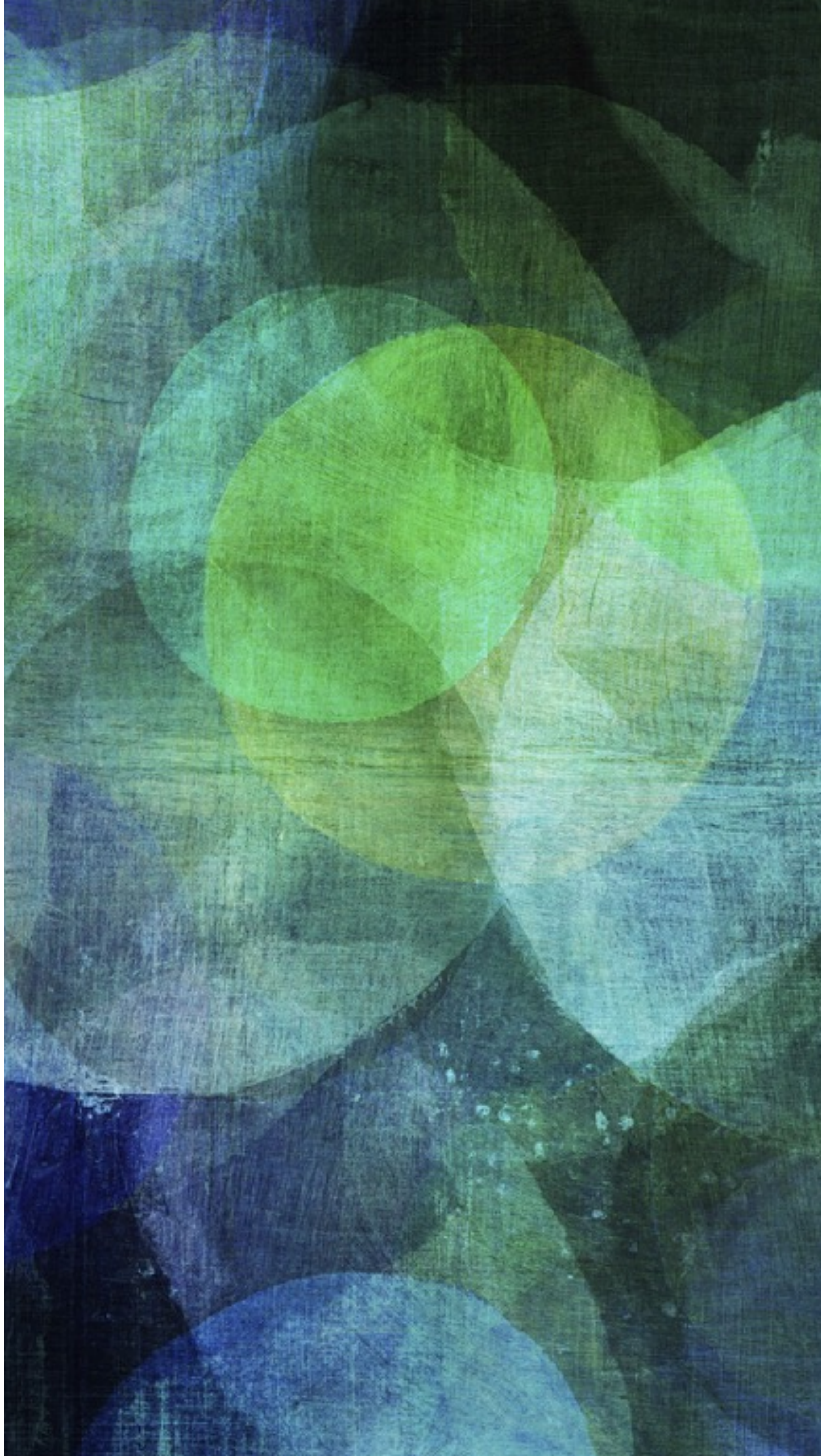
NEW DIRECTIONS: UNKNOWN GRAPHS!

- **Learning on the graph while learning the graph?**
  - **most of algorithms require (some) knowledge of the graph**
  - not always available to the learner
- Question: Can we learn faster without knowing the graphs?
  - example: social network provider has little incentive to reveal the graphs to advertisers
- Answer: **Cohen, Hazan, and Koren:** Online learning with **feedback** graphs without the graphs (ICML June 19-24, 2016)
  - **NO!** (in general we cannot, but possible in the stochastic case)
- Coming up next:
  - **Erdös-Rényi side observation graphs** (UAI June 25-26, 2016)

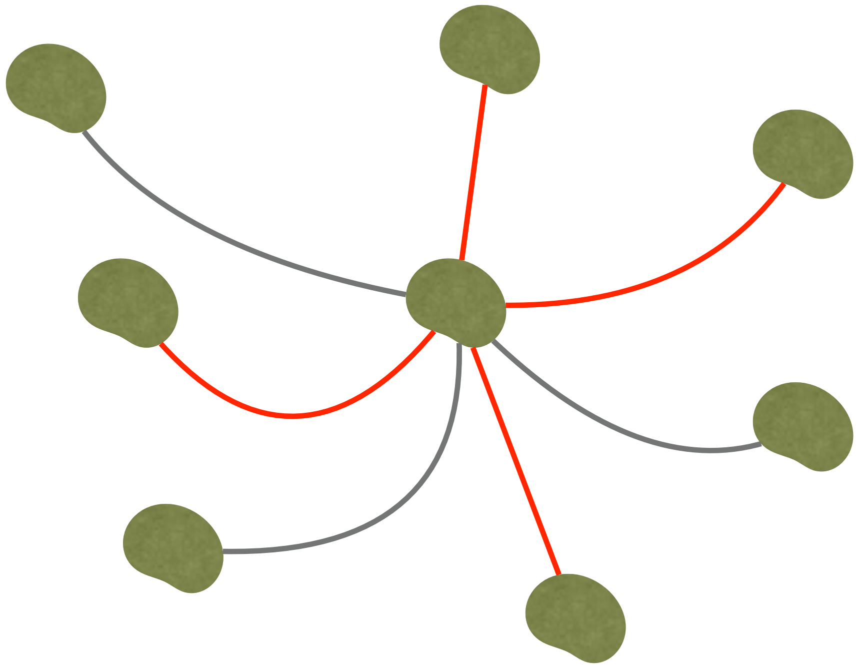
Kocák, Neu, MV: **Online learning with Erdos-Rényi side-observation graphs**
UAI 2016

GRAPH BANDITS WITH ERDÖS-RÉNYI OBSERVATIONS

side observations from graph generators



PROTOCOL FOR ERDÖS-RÉNYI GRAPHS



Every round **t** the learner

- picks a node I_t
- suffers loss for I_t
- receives feedback
  - for I_t
  - for every other node with probability r_t

is loss of i observed?

true loss

$$\widehat{\ell}^{\star}_{t,i} = \frac{O_{t,i}\ell_{t,i}}{p_{t,i} + (1-p_{t,i})r_t}$$

probability of picking i

probability of side observation

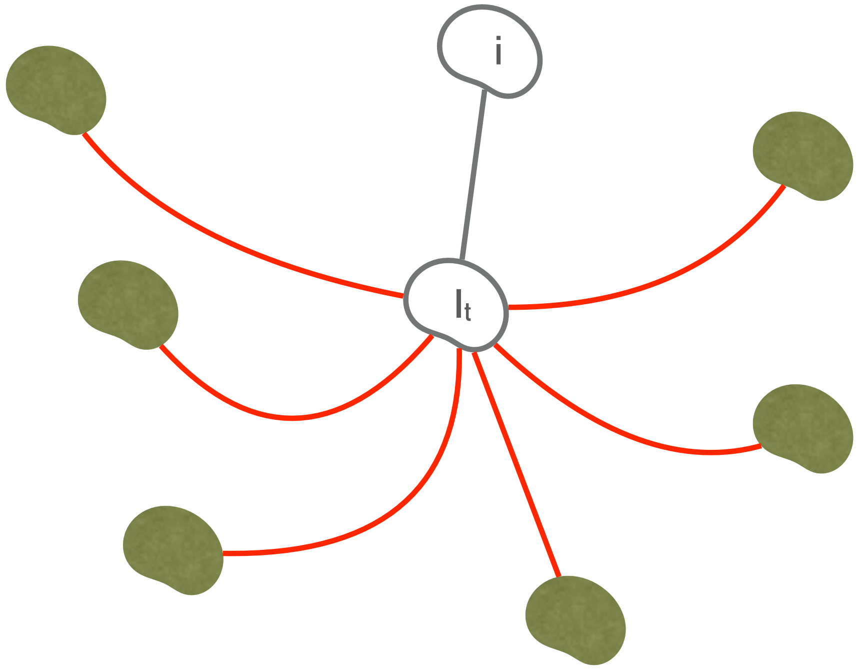
Regret of Exp3-SET (Alon et al. 2013):

$$\mathcal{O}\left(\sqrt{\textstyle\sum_t (1/r_t)(1-(1-r_t)^N)\log N}\right)$$

How to estimate r_t in every round when it is changing?

How to estimate losses without the knowledge of r_t?

PROTOCOL FOR ERDÖS-RÉNYI GRAPHS



- N-2 samples from Bernoulli(r_t) ... R(k)
- N-2 samples from p_{ti} ... P(k)
- O'(k) = P(k) + (1-P(k))R(k)
- G_{ti} = min{k : O'(k) = 1} U {N-1}
- E[G_{ti}] ≈ 1/(p_{ti} +(1-p_{ti})r_t)

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

If r_t ≥ (log T)/(2N-2) then

$$\mathcal{O}\left(\sqrt{\log N \sum_{t=1}^{T} \frac{1}{r_t}}\right)$$

Lower bound (Alon et al. 2013) $\Omega(\sqrt{T/r})$

Get rid of r_t ≥ (log T)/(2N-2)?

is loss of i observed?

true loss

$$\widehat{\ell}^{\star}_{t,i} = \frac{O_{t,i}\ell_{t,i}}{p_{t,i} + (1 - p_{t,i})r_t}$$

probability of picking i

probability of side observation

MORE GRAPH BANDITS AND BEYOND!

Noga Alon et al. (2015) Beyond bandits. Complete characterization: Bártok et al. (2014)

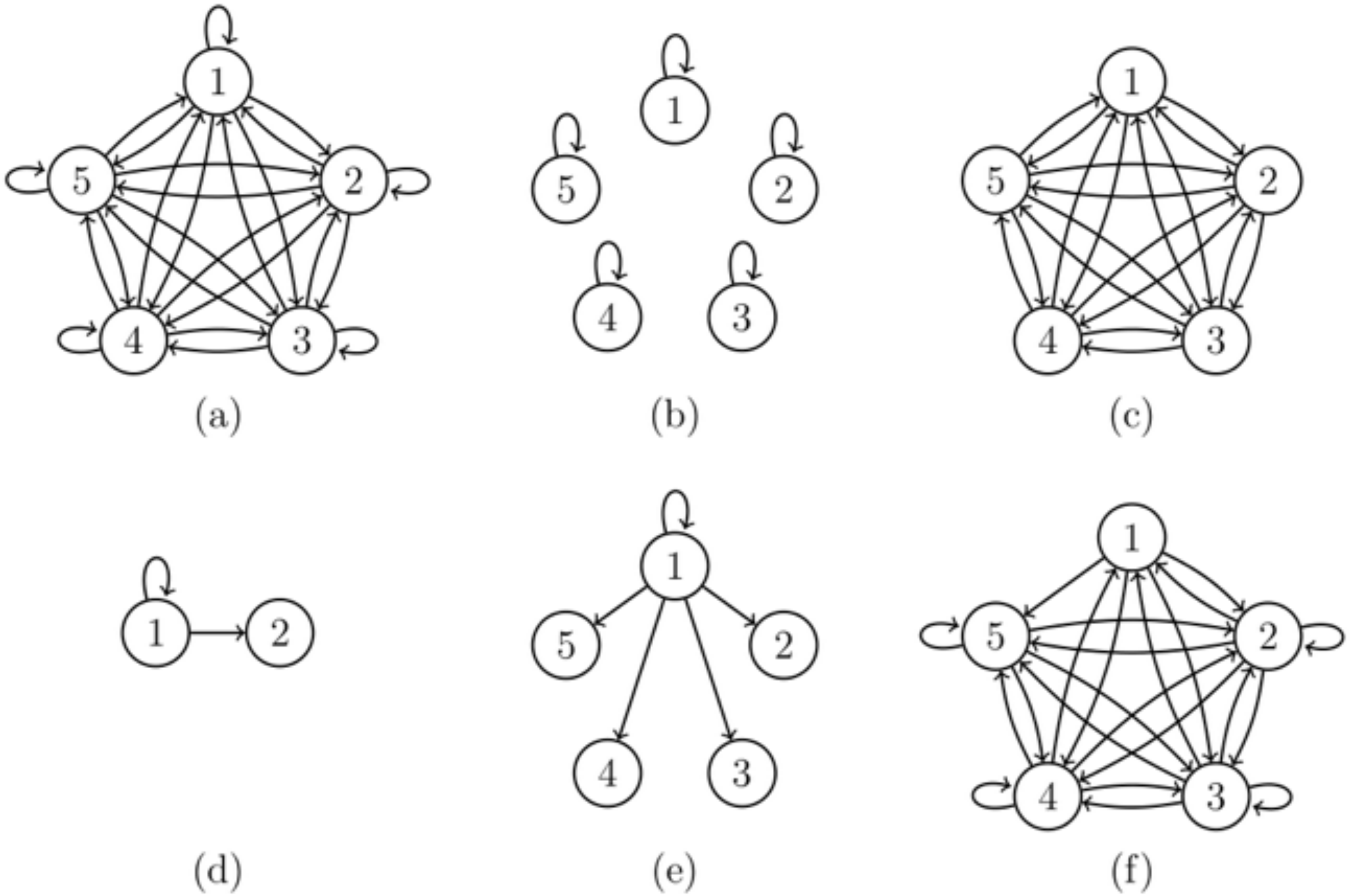
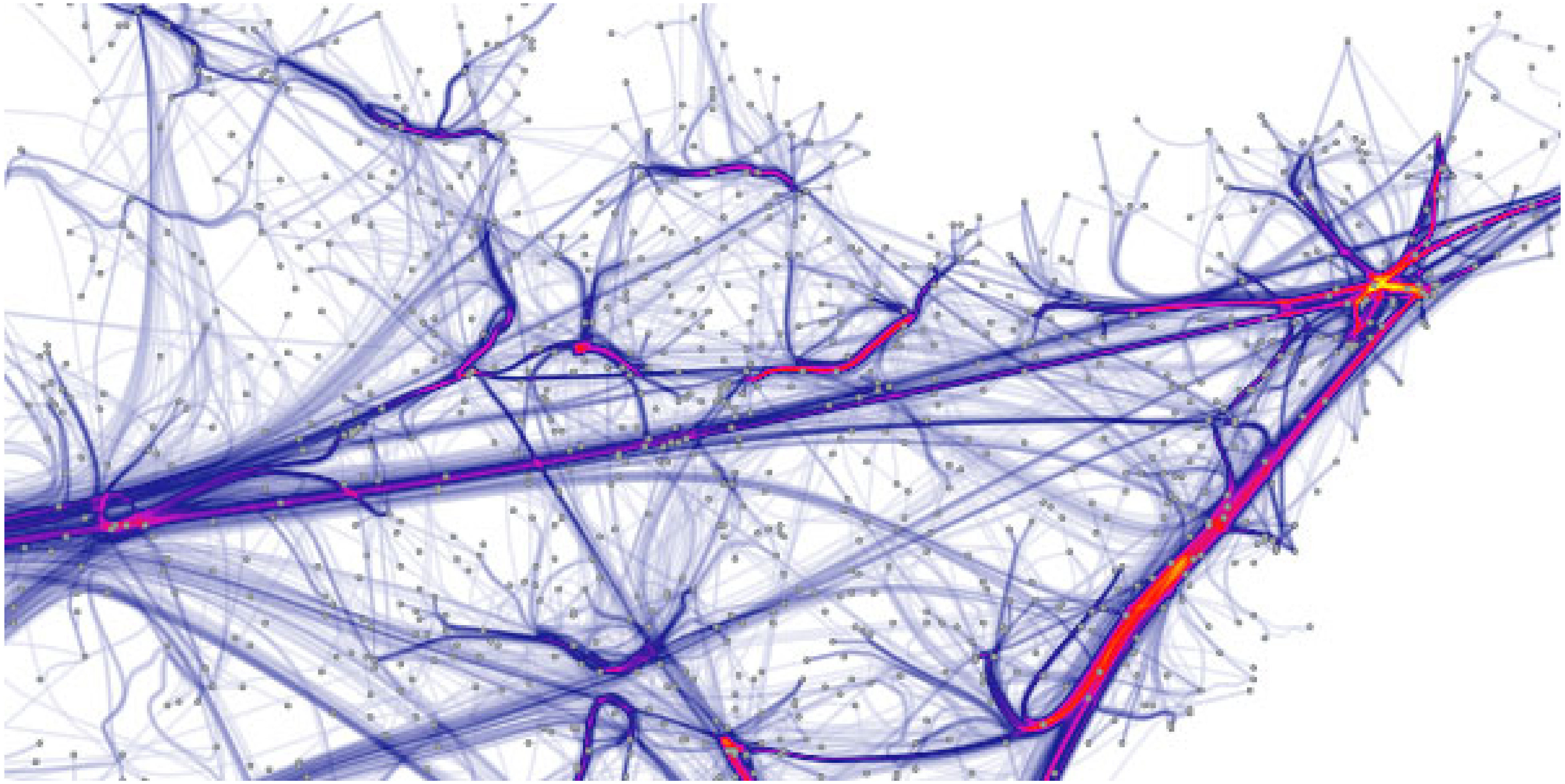


Figure 1: Examples of feedback graphs: (a) *full feedback*, (b) *bandit feedback*, (c) *loopless clique*, (d) *apple tasting*, (e) *revealing action*, (f) a clique minus a self-loop and another edge.

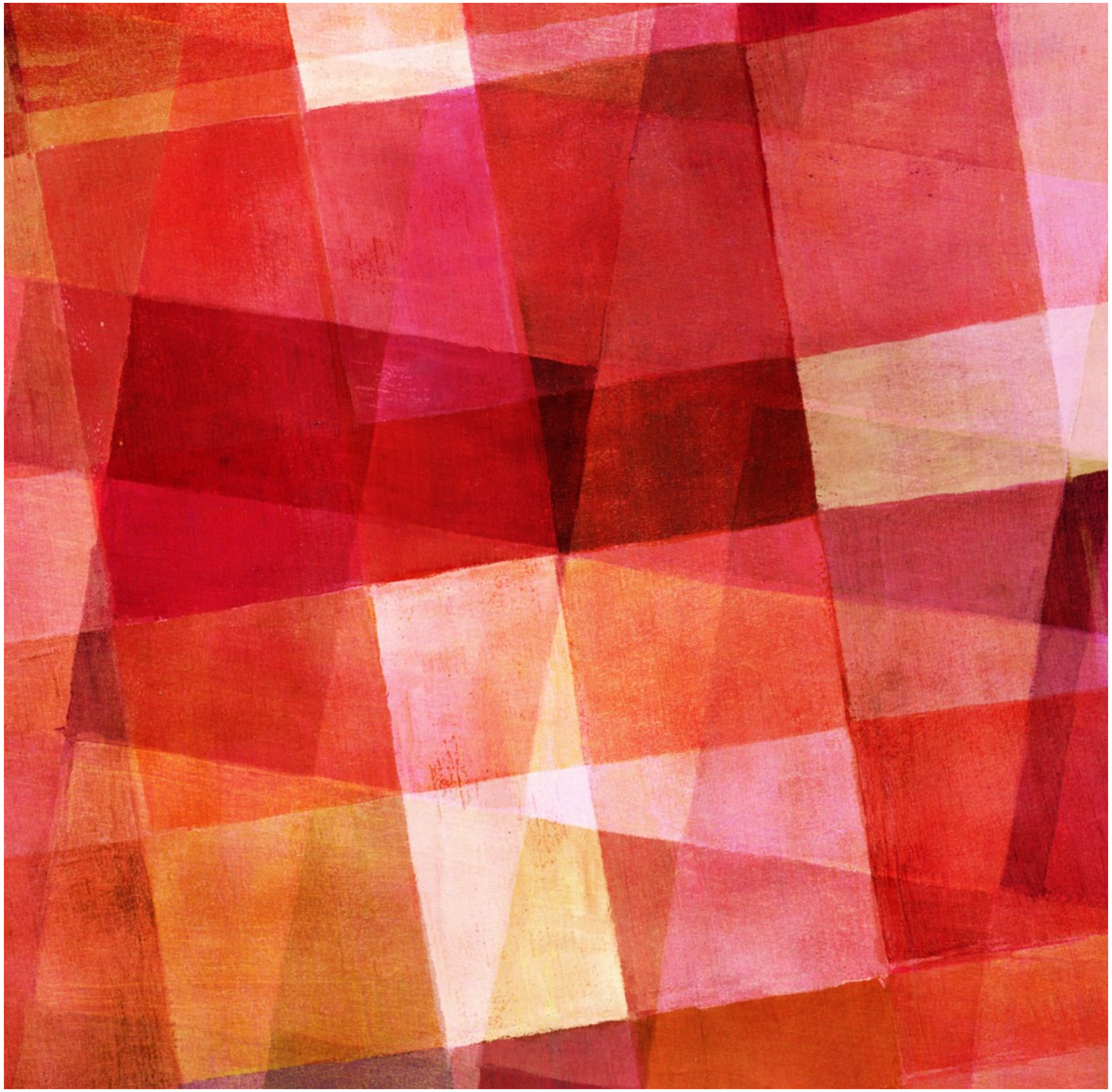
LAST WORDS …

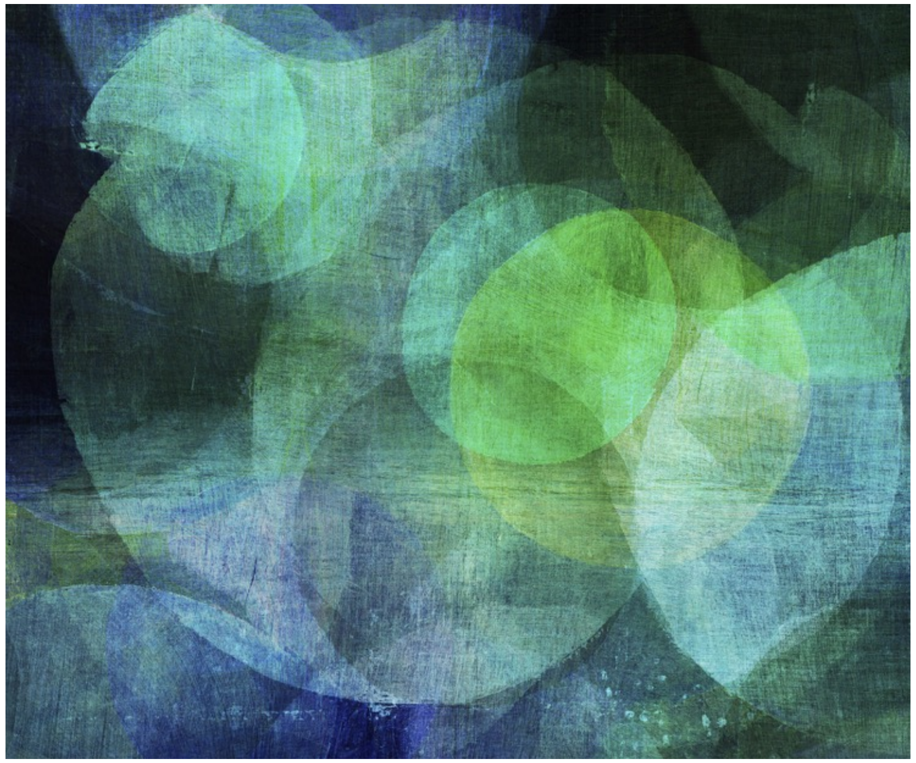
Survey: http://researchers.lille.inria.fr/~valko/hp/publications/valko2016bandits.pdf **(Part I)**

1) good luck with the projects 2) AlteGrad follows this course 3) see you at projects talks



THAT'S ALL – THANK YOU!





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