

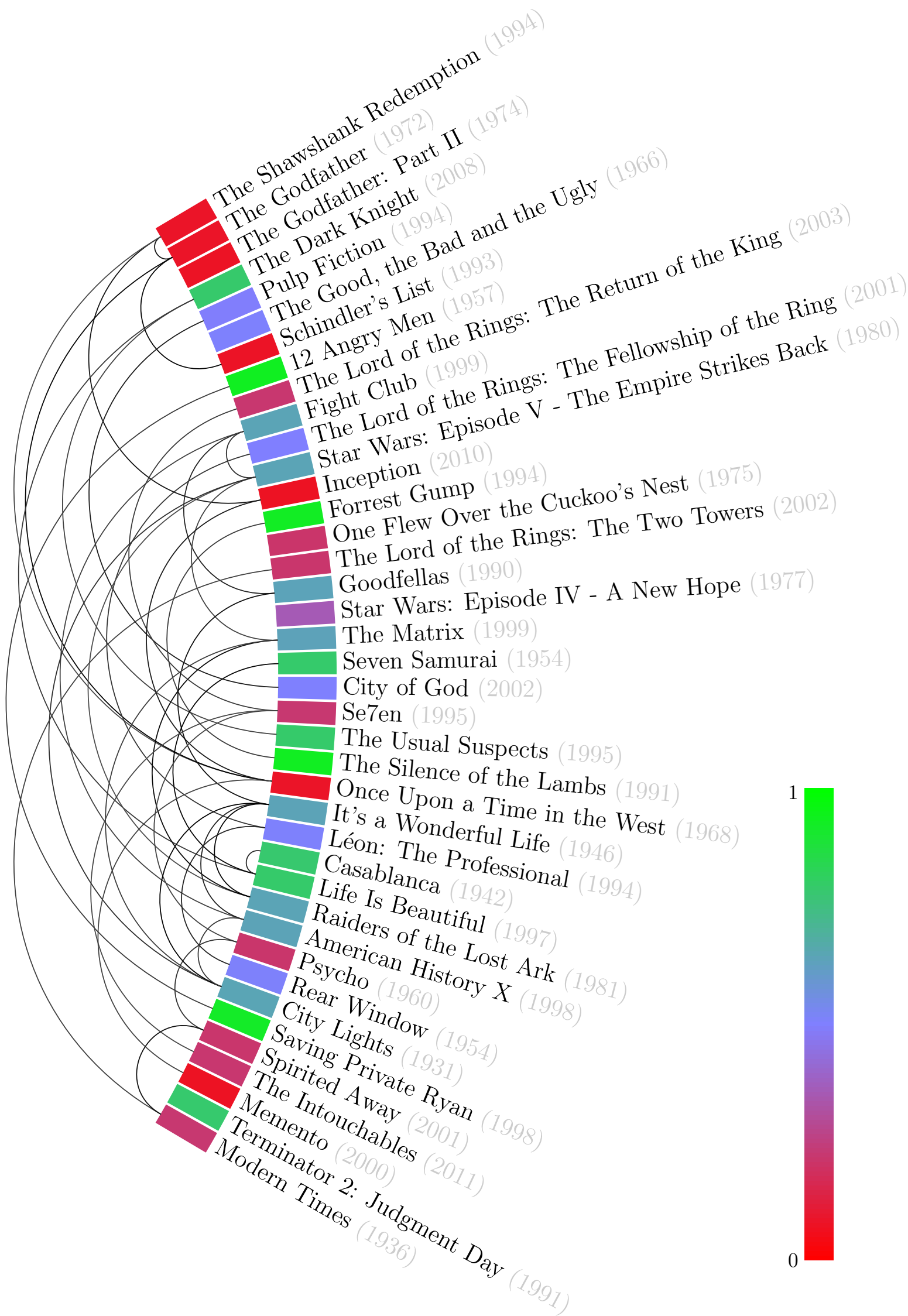
SPECTRAL BANDITS FOR SMOOTH GRAPH FUNCTIONS

TOMAS.KOCAK@INRIA.FR, MICHAL.VALKO@INRIA.FR, REMI.MUNOS@INRIA.FR, AND BRANISLAV.KVETON@TECHNICOLOR.COM



MOTIVATION - MOVIE RECOMMENDATION

- **Goal:** Movie recommendation based on similarities
- **Challenges:** Good prediction after just a few steps ($T \ll N$)
- **Prior knowledge:** The preferences of movies are smooth over a given weighted similarity graph



- Colors represent *single-user preferences*.
- Connected (similar) movies have similar user ratings

SMOOTH GRAPH FUNCTIONS

- **Graph function:** mapping from set of the graph vertices $V(G)$ into real numbers

- **Smoothness of a graph function $S_G(f)$:**

– eigendecomposition of graph Laplacian: $\mathcal{L} = \mathbf{Q}\mathbf{A}\mathbf{Q}^T$

$$S_G(f) = \frac{1}{2} \sum_{u,v \in V(G)} w_{u,v} (f(u) - f(v))^2 = \mathbf{f}^T \mathcal{L} \mathbf{f}$$

$$= \mathbf{f}^T \mathbf{Q} \mathbf{A} \mathbf{Q}^T \mathbf{f} = \alpha^{*T} \mathbf{A} \alpha^* = \|\alpha^*\|_{\mathbf{A}} = \sum_{i=1}^N \lambda_i \alpha_i^2$$

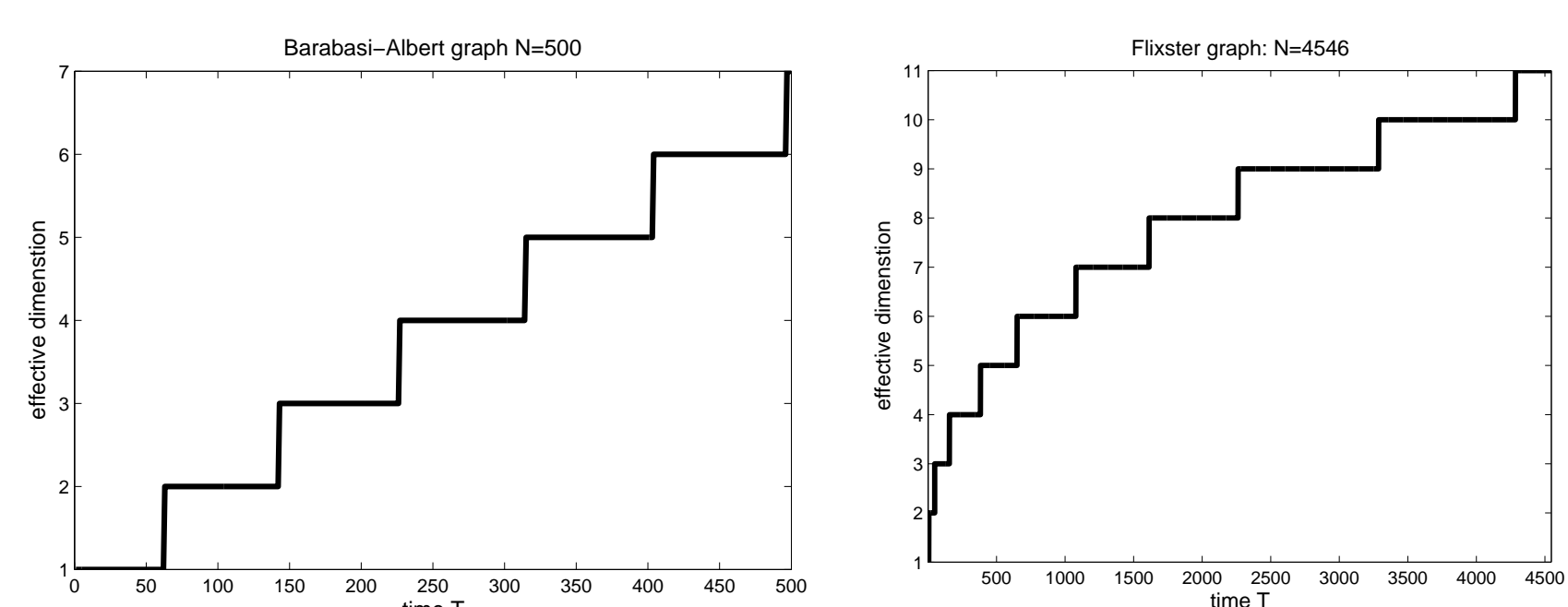
- **Observation:** $S_G(\mathbf{q}_i) = \lambda_i$

- **Smoothness and regularization:** Small value of (a) $S_G(f)$ (b) \mathbf{A} norm of α^* (c) α_i for large λ_i

EFFECTIVE DIMENSION

Definition 1. Let the effective dimension d be the largest d such that

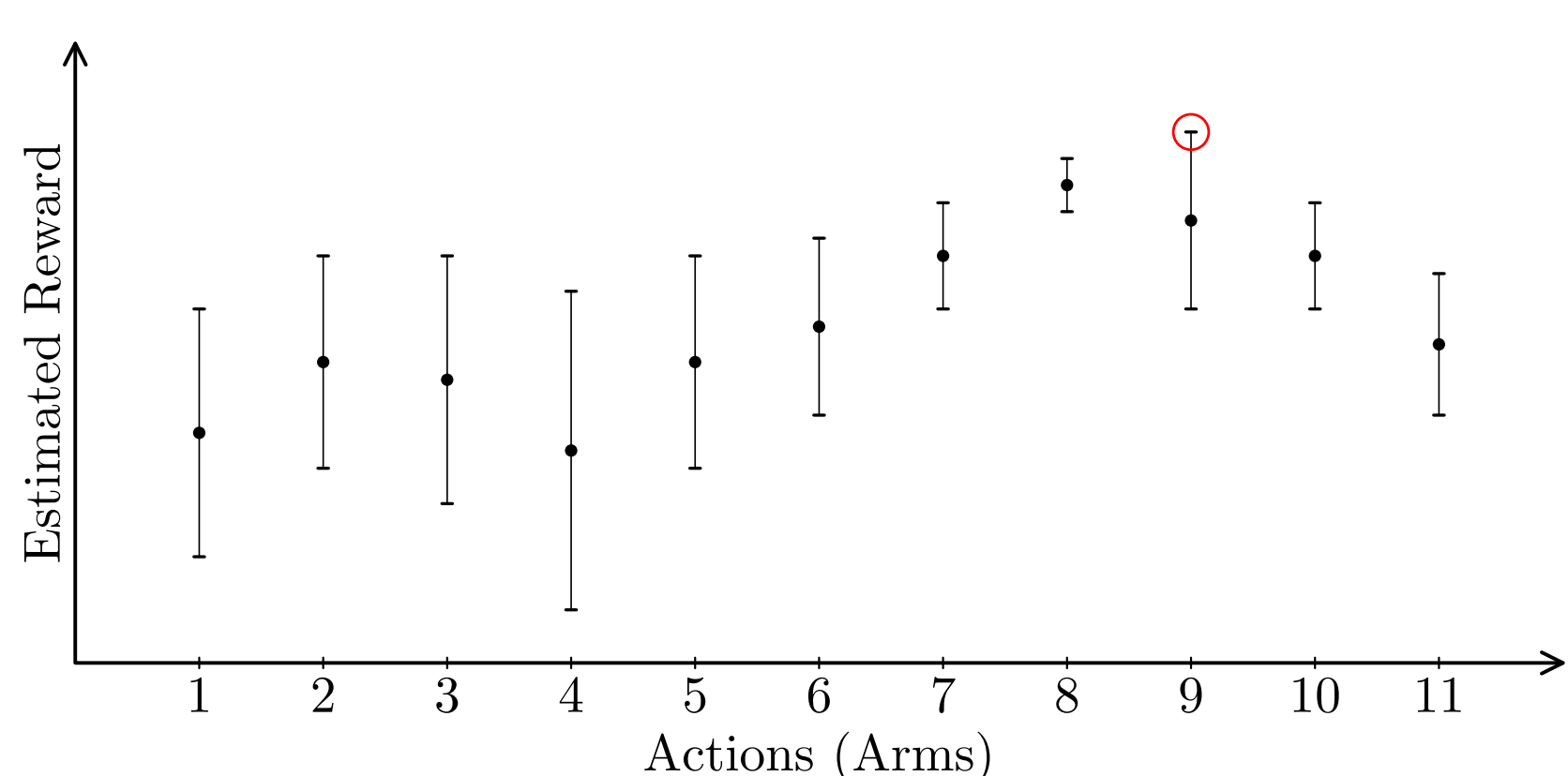
$$(d-1)\lambda_d \leq \frac{T}{\log(1+T/\lambda)}$$



- d is small when the coefficients λ_i grow rapidly above time.
- d is related to the number of “non-negligible” dimensions

UPPER CONFIDENCE BOUND ALGORITHMS

- Pick an arm with the highest upper confidence bound.
- Update estimates and confidence intervals.



SETTING

- **Task:** Each time t , pick an action (node) to get a reward.
- **Reward:** $\mathbf{x}_t^T \alpha^* + \varepsilon_t$ (with unknown parameter α^*)
 - \mathbf{x}_i is the i -th row of \mathbf{Q}
 - reward is a combination of smooth eigenvectors
- **Goal:** Minimize the cumulative regret w.r.t. the best node

$$R_T = T \max_v f_{\alpha^*}(v) - \sum_{t=1}^T f_{\alpha^*}(\pi(t))$$

ALGORITHM 1 - SPECTRALUCB

Input:

N : the number of nodes, T : the number of pulls
 $\{\mathbf{A}_{\mathcal{L}}, \mathbf{Q}\}$ spectral basis of \mathcal{L}
 λ, δ : regularization and confidence parameters
 R, C : upper bounds on the noise and $\|\alpha^*\|_{\mathbf{A}}$

Run:

$\mathbf{A} \leftarrow \mathbf{A}_{\mathcal{L}} + \lambda \mathbf{I}$
 $d \leftarrow \max\{d : (d-1)\lambda_d \leq T/\log(1+T/\lambda)\}$

for $t = 1$ to T do

Update the basis coefficients $\hat{\alpha}$:

$\mathbf{X}_t \leftarrow [\mathbf{x}_1, \dots, \mathbf{x}_{t-1}]^T$

$\mathbf{r} \leftarrow [r_1, \dots, r_{t-1}]^T$

$\mathbf{V}_t \leftarrow \mathbf{X}_t \mathbf{X}_t^T + \mathbf{A}$

$\hat{\alpha}_t \leftarrow \mathbf{V}_t^{-1} \mathbf{X}_t^T \mathbf{r}$

$c_t \leftarrow 2R\sqrt{d \log(1+T/\lambda)} + 2 \log(1/\delta) + C$

Choose the node v_t (\mathbf{x}_{v_t} -th row of \mathbf{Q}):

$v_t \leftarrow \arg \max_v (f_{\hat{\alpha}_t}(v) + c_t \|\mathbf{x}_{v_t}\|_{\mathbf{V}_t^{-1}})$

Observe the reward r_t

end for

ALGORITHM 2 - SPECTRAL ELIMINATOR

Input:

N : the number of nodes, T : the number of pulls
 $\{\mathbf{A}_{\mathcal{L}}, \mathbf{Q}\}$ spectral basis of \mathcal{L}
 λ : regularization parameter
 $\beta, \{t_j\}_J^J$ parameters of the elimination and phases
 $\mathbf{A}_1 \leftarrow \{\mathbf{x}_1, \dots, \mathbf{x}_K\}$

for $j = 1$ to J do

$\mathbf{V}_{t_j} \leftarrow \gamma \mathbf{A}_{\mathcal{L}} + \lambda \mathbf{I}$

for $t = t_j$ to $\min(t_{j+1} - 1, T)$ do

Play $\mathbf{x}_t \in \mathbf{A}_j$ with the largest width to observe r_t :

$\mathbf{x}_t \leftarrow \arg \max_{\mathbf{x} \in \mathbf{A}_j} \|\mathbf{x}\|_{\mathbf{V}_t^{-1}}$

$\mathbf{V}_{t+1} \leftarrow \mathbf{V}_t + \mathbf{x}_t \mathbf{x}_t^T$

end for

Eliminate the arms that are not promising:

$\hat{\alpha}_t \leftarrow \mathbf{V}_t^{-1} [\mathbf{x}_{t_j}, \dots, \mathbf{x}_t] [r_{t_j}, \dots, r_t]^T$

$p \leftarrow \max_{\mathbf{x} \in \mathbf{A}_j} \langle \hat{\alpha}_t, \mathbf{x} \rangle - \|\mathbf{x}\|_{\mathbf{V}_t^{-1}} \beta$

$\mathbf{A}_{j+1} \leftarrow \{\mathbf{x} \in \mathbf{A}_j, \langle \hat{\alpha}_t, \mathbf{x} \rangle + \|\mathbf{x}\|_{\mathbf{V}_t^{-1}} \beta \geq p\}$

end for

MAIN RESULTS

SpectralUCB regret bound

Theorem 1. Let be the minimum eigenvalue of \mathbf{A} . If $\|\alpha^*\|_{\mathbf{A}} \leq C$ and for all $\mathbf{x}_v, \langle \mathbf{x}_v, \alpha^* \rangle \in [-1, 1]$, then the cumulative regret of SpectralUCB is with probability at least $1 - \delta$ bounded as

$$R_T \leq \left[8R\sqrt{d \log(1+T/\lambda)} + 2 \log(1/\delta) + 4C + 4 \right] \times \sqrt{dT \log(1+T/\lambda)} \approx d\sqrt{T}$$

Setting $\mathbf{A} = \mathbf{I}$ we recover LinUCB. Since $\log(|\mathbf{V}_T|/|\mathbf{A}|)$ can be upper-bounded by $D \log T$ [1], we obtain $\tilde{O}(D\sqrt{T})$ for LinUCB.

SpectralEliminator regret bound

Theorem 2. Choose the phases starts as $t_j = 2^{j-1}$. Assume all rewards are in $[0, 1]$ and $\|\alpha^*\|_{\mathbf{A}} \leq C$. For any $\delta > 0$, with probability at least $1 - \delta$, the cumulative regret of SpectralEliminator algorithm run with parameter $\beta = 2R\sqrt{14 \log(2K \log_2 T/\delta)} + C$ is bounded as:

$$R_T \leq \frac{4}{\log 2} \left(2R\sqrt{14 \log \frac{2K \log_2 T}{\delta}} + C \right) \sqrt{dT \log \left(1 + \frac{T}{\lambda} \right)} \approx \sqrt{dT}$$

If $\mathbf{A} = \mathbf{I}$, we get a competitor to SupLinRel [2], with $\tilde{O}(\sqrt{DT})$ regret, with significantly more elegant algorithm and analysis.

Linear vs. Spectral bandits

Linear	Spectral
LinUCB $D\sqrt{T \ln T}$	SpectralUCB $d\sqrt{T \ln T}$
SupLinRel $\sqrt{DT \ln T}$	SpectralEliminator $\sqrt{dT \ln T}$

ANALYSES SKETCH

- Derivation of the confidence ellipsoid for estimate $\hat{\alpha}$.

By self-normalized bound of [1]: w. p. $1 - \delta$:

$$|\mathbf{x}^T (\hat{\alpha} - \alpha^*)| \leq \|\mathbf{x}\|_{\mathbf{V}_t^{-1}} \left(R\sqrt{2 \log \left(\frac{|\mathbf{V}_t|^{1/2}}{\delta |\mathbf{A}|^{1/2}} \right)} + C \right)$$

- Our key result coming from spectral properties of \mathbf{V}_t :

$$\log \frac{|\mathbf{V}_t|}{|\mathbf{A}|} \leq 2d \log \left(1 + \frac{T}{\lambda} \right)$$

SpectralUCB

- Regret in one time step: $\mathbf{x}_t^T \alpha^* - \mathbf{x}_{\pi(t)}^T \alpha^* \leq 2c_t \|\mathbf{x}_{\pi(t)}\|_{\mathbf{V}_t^{-1}}$

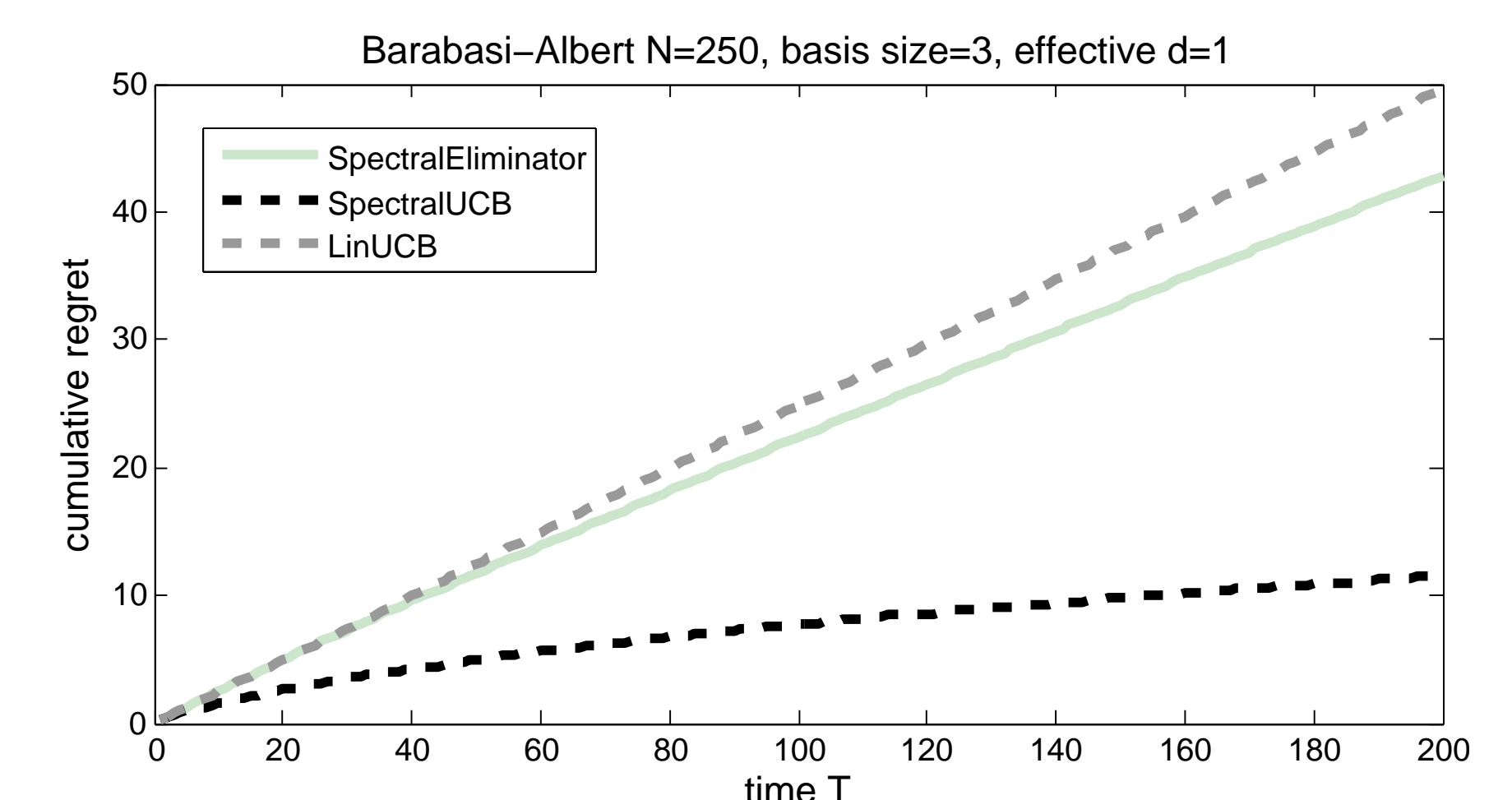
SpectralEliminator

- Divide time into sets ($t_1 = 1 \leq t_2 \leq \dots$) to introduce independence for Azuma-Hoeffding inequality and observe $R_T \leq \sum_{j=0}^J (t_{j+1} - t_j) [\langle \mathbf{x}^* - \mathbf{x}_t, \hat{\alpha}_j \rangle + (\|\mathbf{x}^*\|_{\mathbf{V}_j^{-1}} + \|\mathbf{x}_t\|_{\mathbf{V}_j^{-1}}) \beta]$
- Bound $\langle \mathbf{x}^* - \mathbf{x}_t, \hat{\alpha}_j \rangle$ for each phase
- No bad arms: $\langle \mathbf{x}^* - \mathbf{x}_t, \hat{\alpha}_j \rangle \leq (\|\mathbf{x}^*\|_{\mathbf{V}_j^{-1}} + \|\mathbf{x}_t\|_{\mathbf{V}_j^{-1}}) \beta$
- By algorithm: $\|\mathbf{x}_s\|_{\mathbf{V}_j^{-1}}^2 \leq \frac{1}{t_j - t_{j-1}} \sum_{s=t_{j-1}+1}^{t_j} \|\mathbf{x}_s\|_{\mathbf{V}_{s-1}}^2$
- $\sum_{s=t_{j-1}+1}^{t_j} \min \left(1, \|\mathbf{x}_s\|_{\mathbf{V}_{s-1}}^2 \right) \leq \log \frac{|\mathbf{V}_j|}{|\mathbf{A}|}$

EXPERIMENTS

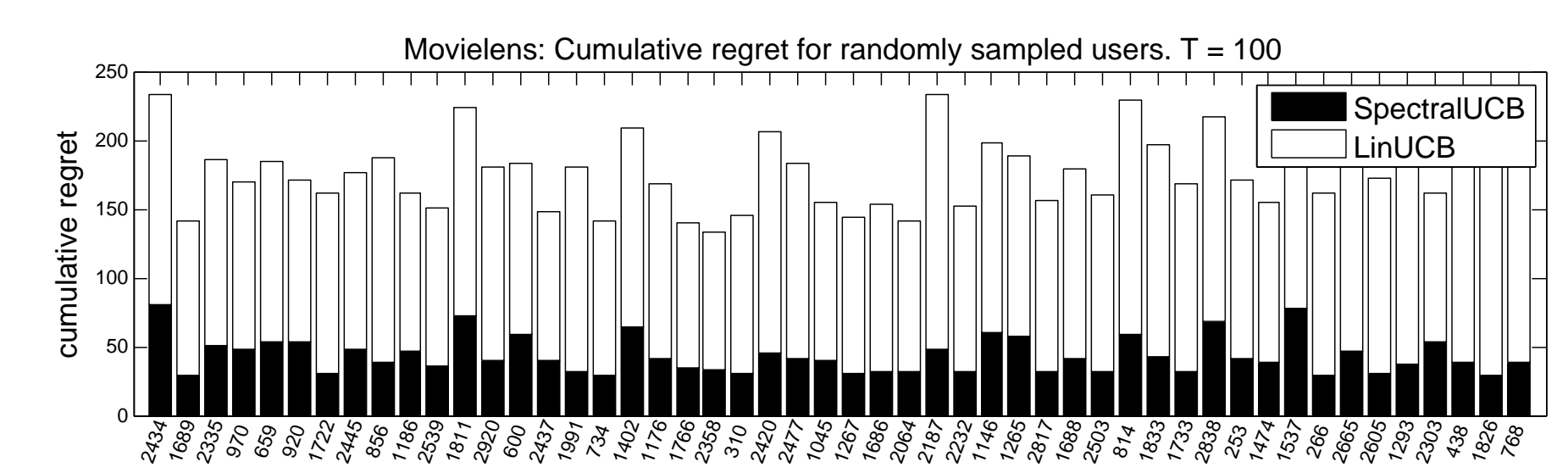
Synthetic Experiment

Barabasi-Albert (BA) model with the degree parameter 3.



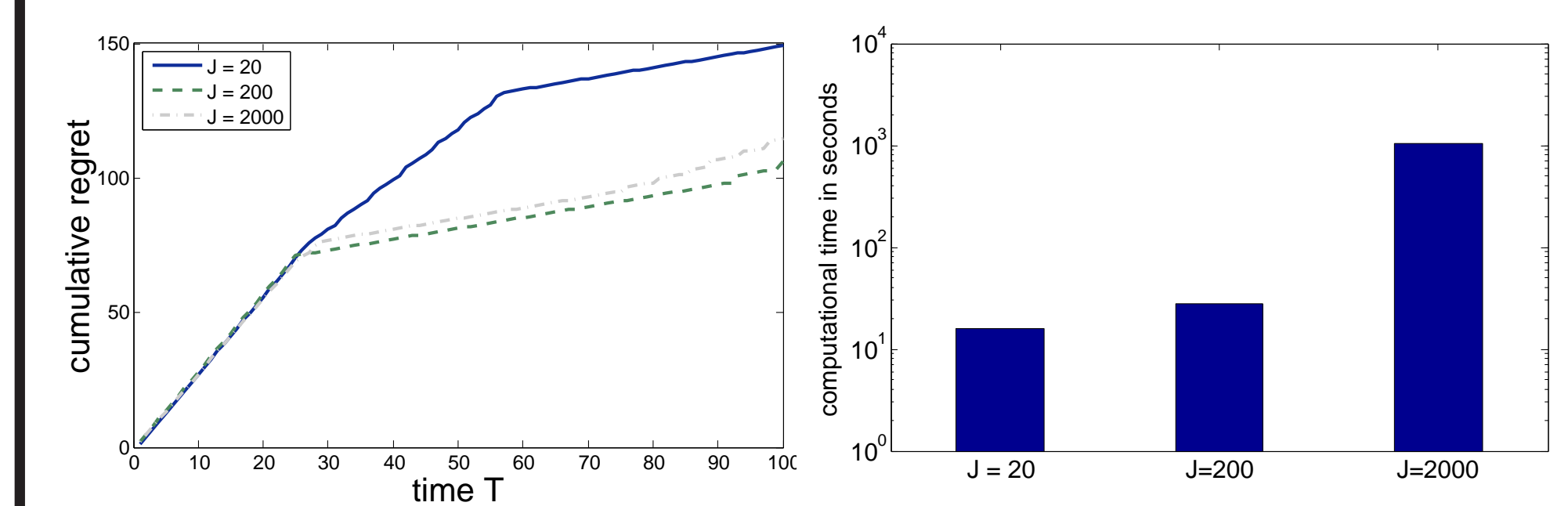
Movie Experiment

MovieLens dataset of 6k users who rated one million movies.



Improving the running time: reduced eigenbasis

- **Reduced basis:** We only need first few eigenvectors.
- **Getting J eigenvectors:** $\mathcal{O}(Jm \log m)$ time for m edges
- Computationally less expensive, comparable performance.



REFERENCES

- [1] Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Improved Algorithms for Linear Stochastic Bandits. In *Neural Information Processing Systems*. 2011.
- [2] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3:397–422, March 2002.

ACKNOWLEDGMENTS

