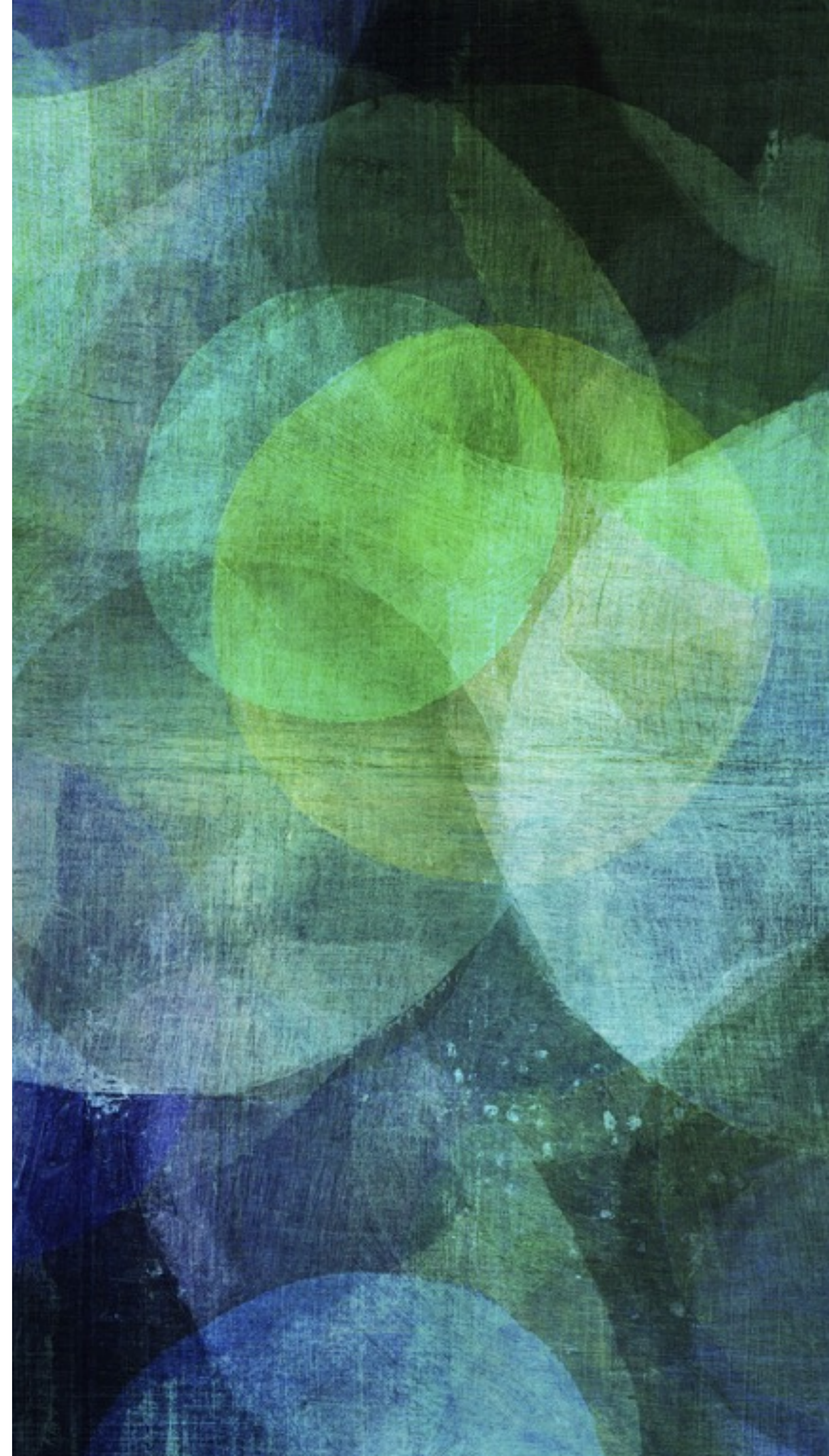
MV, Munos, Kveton, Kocák: **Spectral Bandits for Smooth Graph Functions**, ICML 2014

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

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

SPECTRAL BANDITS

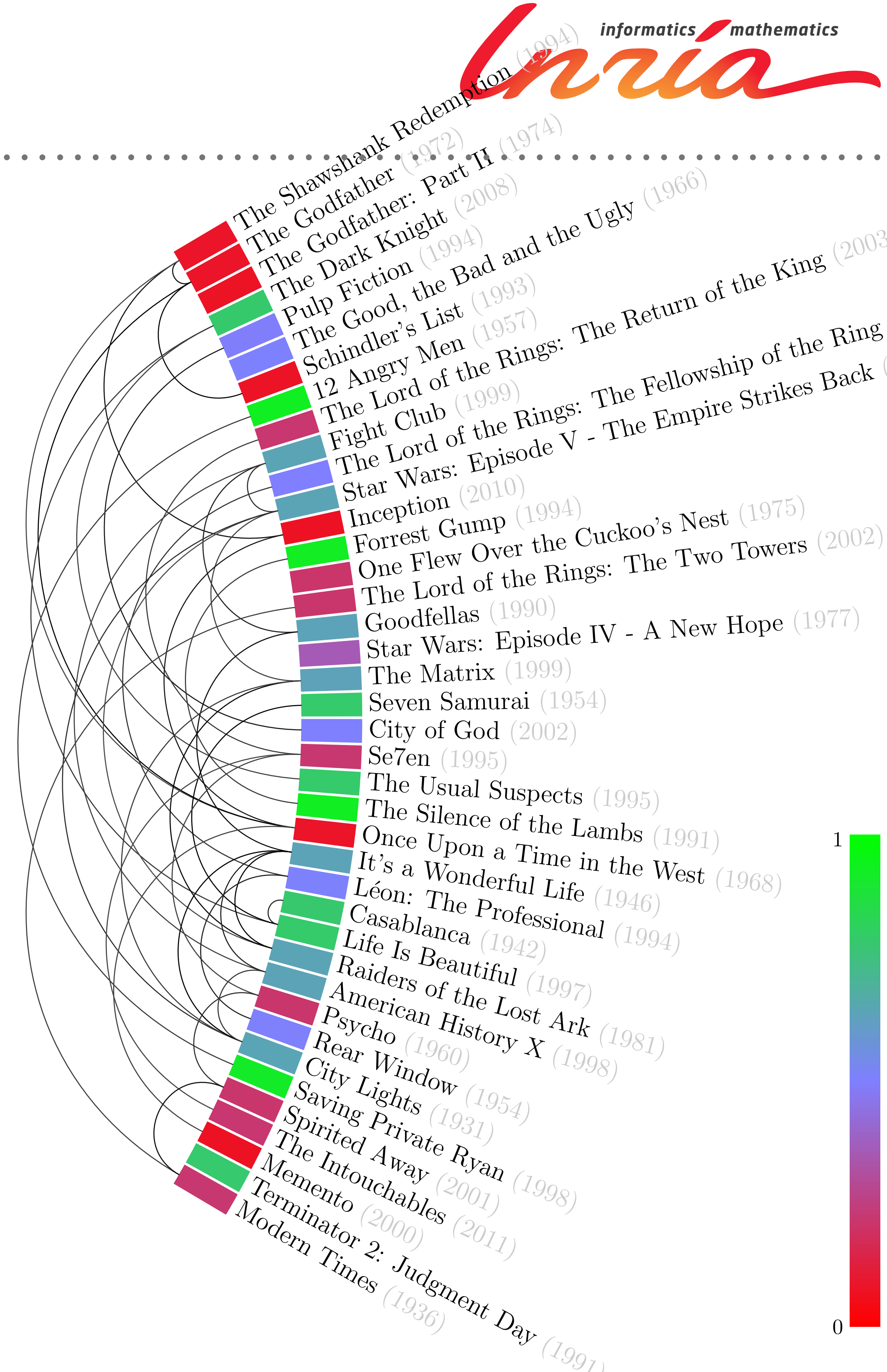
exploiting smoothness of rewards on graphs



SPECTRAL BANDITS

Assumptions

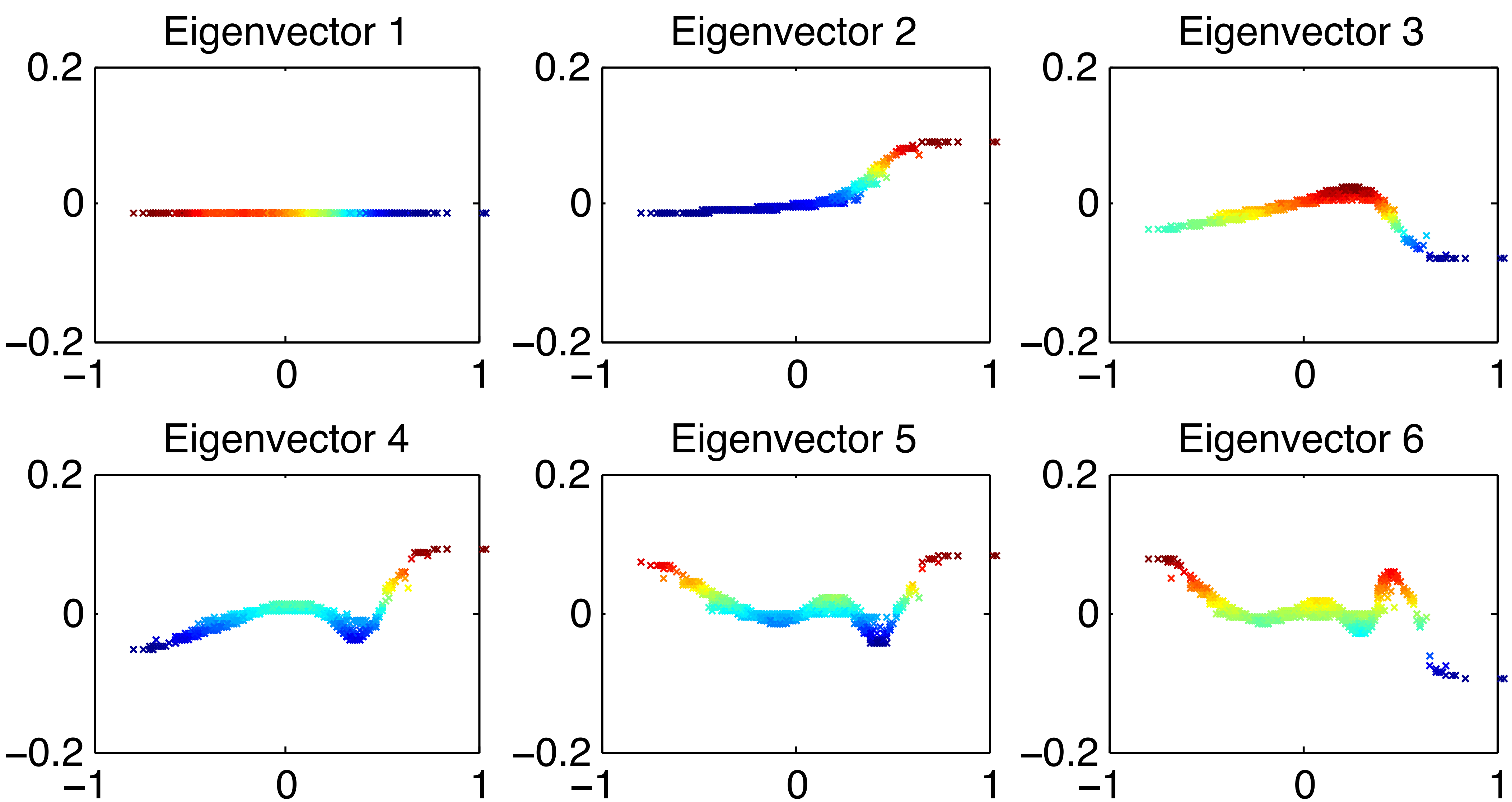
- Unknown reward function $f : V(G) \to \mathbb{R}$.
- Function f is **smooth** on a graph.
- Neighboring movies $\Rightarrow$ similar preferences.
- Similar preferences $\not\Rightarrow$ neighboring movies.



Desiderata

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

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 $\rightarrow$ **HUGE!**

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

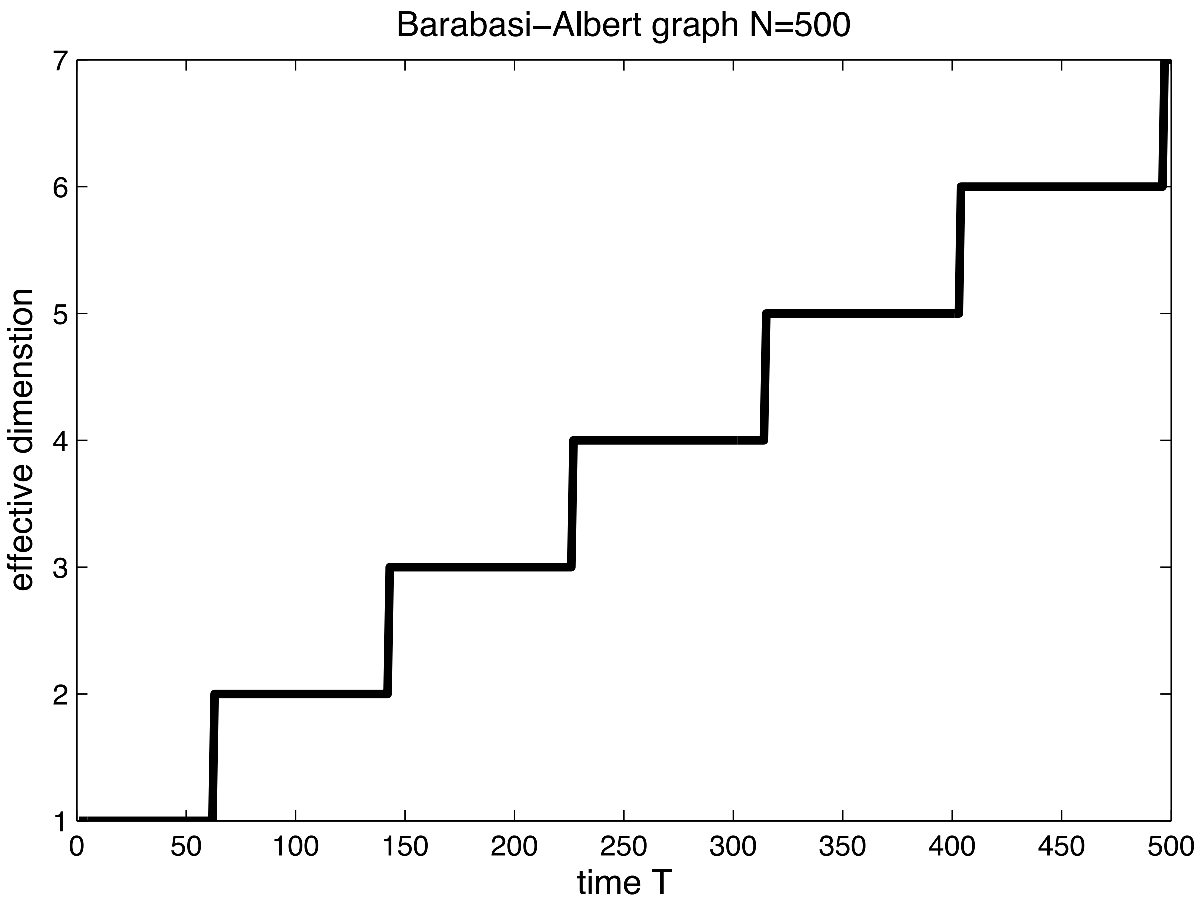
$$(d-1)\lambda_d \leq \frac{T}{\log(1+T/\lambda)}.$$

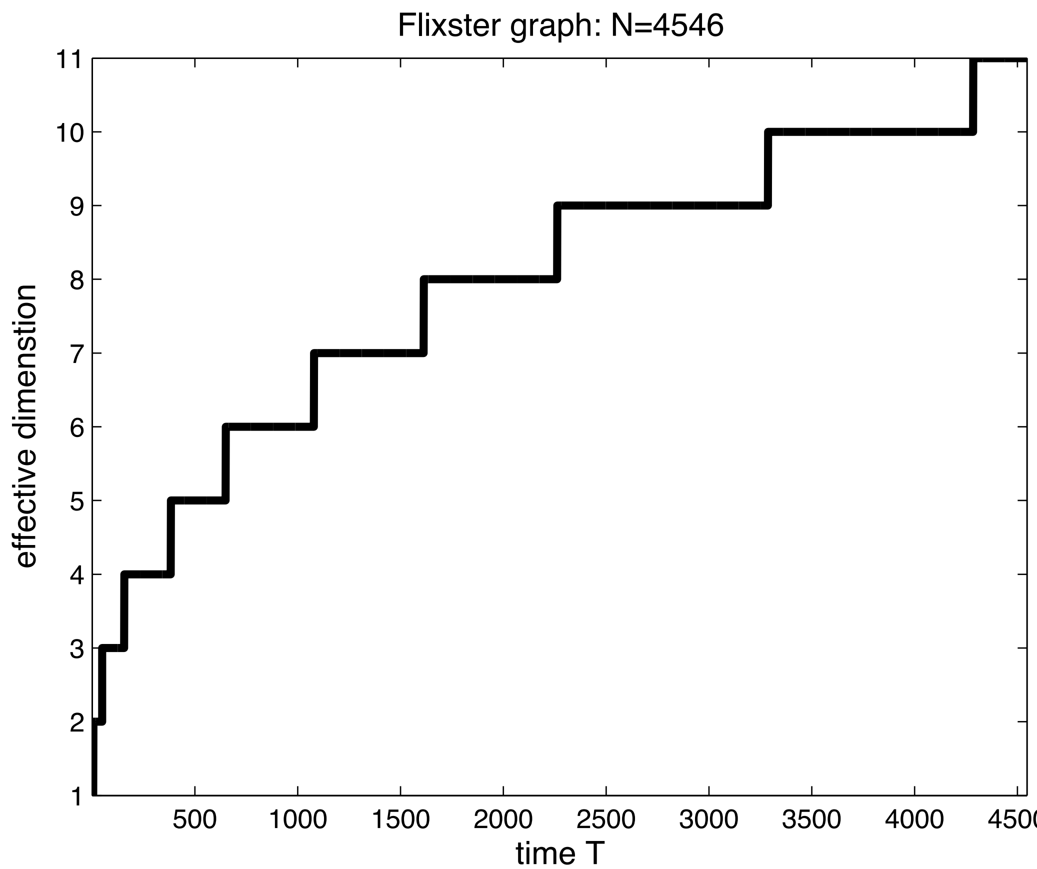
- Function of time horizon and graph properties
- λ_i : i-th smallest eigenvalue of $\mathbf{\Lambda}$.
- λ: Regularization parameter of the algorithm.

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

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

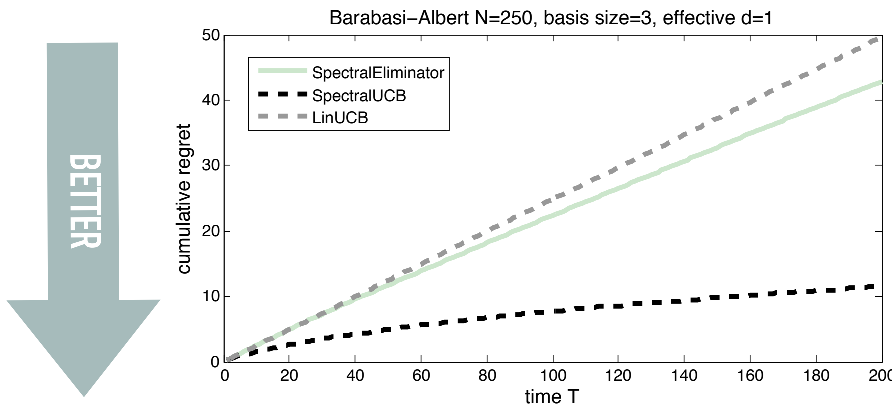
$\|\boldsymbol{\alpha}\|_{\boldsymbol{\Lambda}}$ is a penalty for non-smooth combinations of eigenvectors.

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

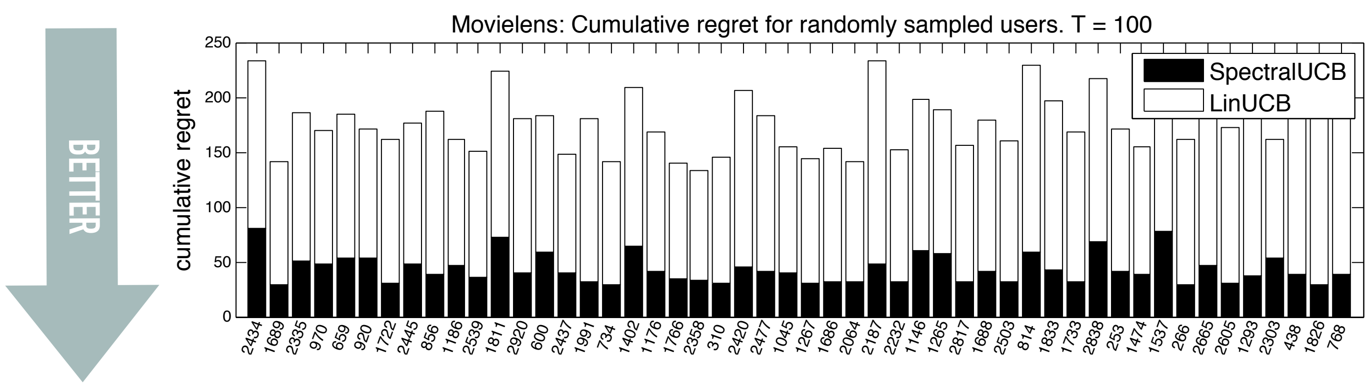
SPECTRALUCB REGRET BOUND

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

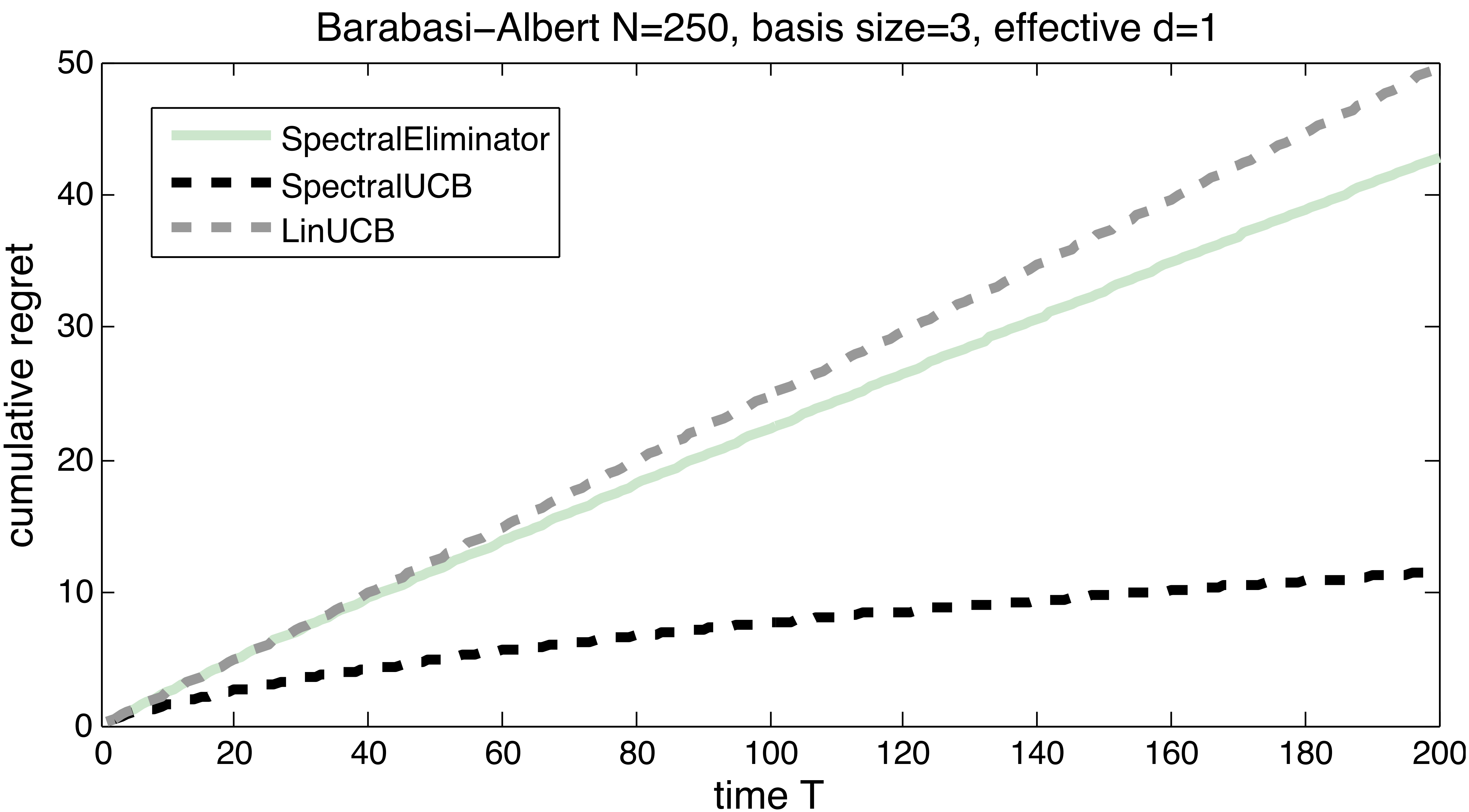


The **cumulative regret** R_T of **SpectralUCB** is with probability $1-\delta$ bounded as

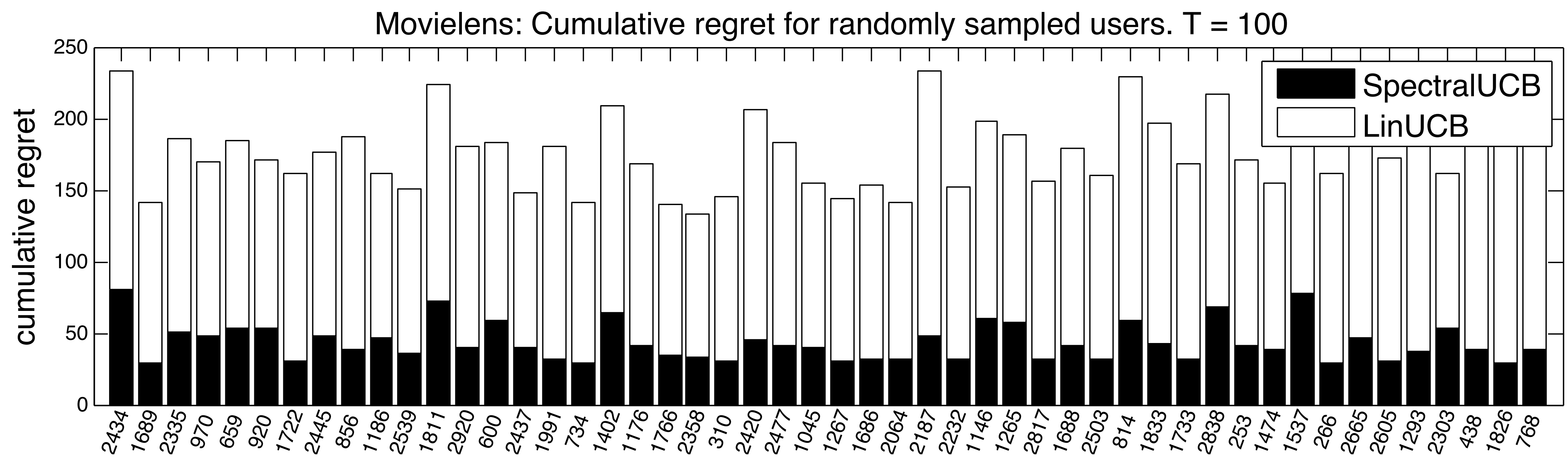
$$R_T \leq \left(8R\sqrt{d \ln \frac{\lambda + T}{\lambda} + 2 \ln \frac{1}{\delta}} + 4C + 4 \right) \sqrt{dT \ln \frac{\lambda + T}{\lambda}}.$$

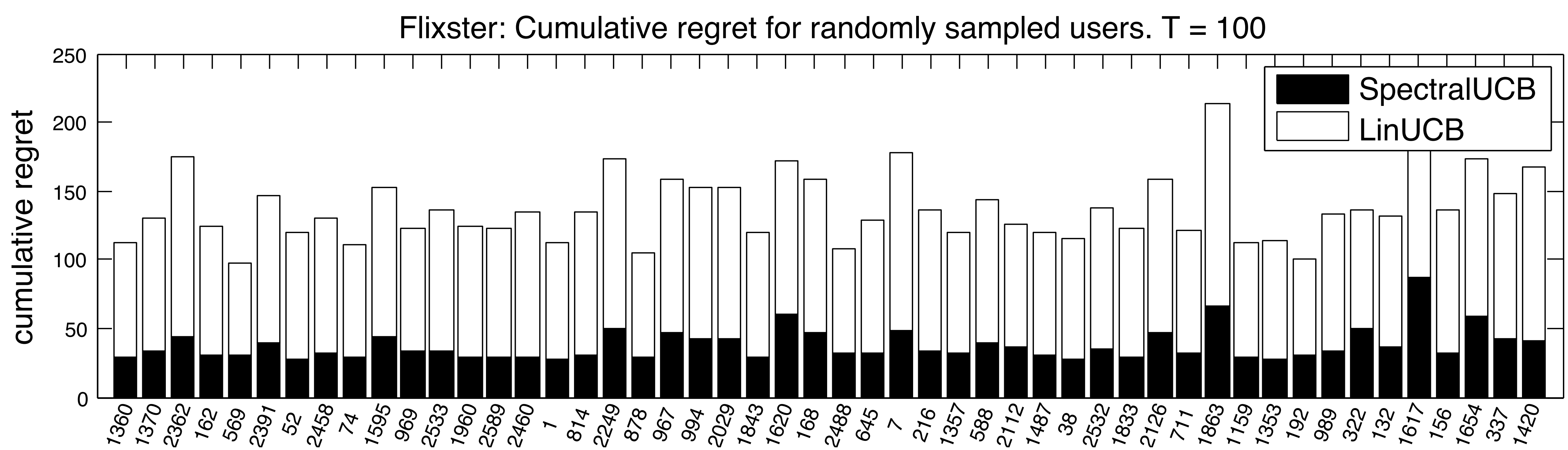


SPECTRAL UCN ON BA GRAPH



SPECTRAL UCN ON REAL DATA





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.

$$|\mathbf{V}_T| \leq \max_{\sum t_i = T} \prod (\lambda_i + t_i)$$

$$\ln \frac{|\mathbf{V}_T|}{|\mathbf{\Lambda}|} \leq \max_{\sum t_i = T} \sum \ln \left(1 + \frac{t_i}{\lambda_i}\right)$$

$$\ln \frac{|\mathbf{V}_T|}{|\mathbf{\Lambda}|} \leq \sum_{i=1}^{d} \ln \left(1 + \frac{T}{\lambda}\right) + \sum_{i=d+1}^{N} \ln \left(1 + \frac{t_i}{\lambda_{d+1}}\right)$$

$$\leq d \ln \left(1 + \frac{T}{\lambda}\right) + \frac{T}{\lambda_{d+1}}$$

$$\leq 2d \ln \left(1 + \frac{T}{\lambda}\right)$$