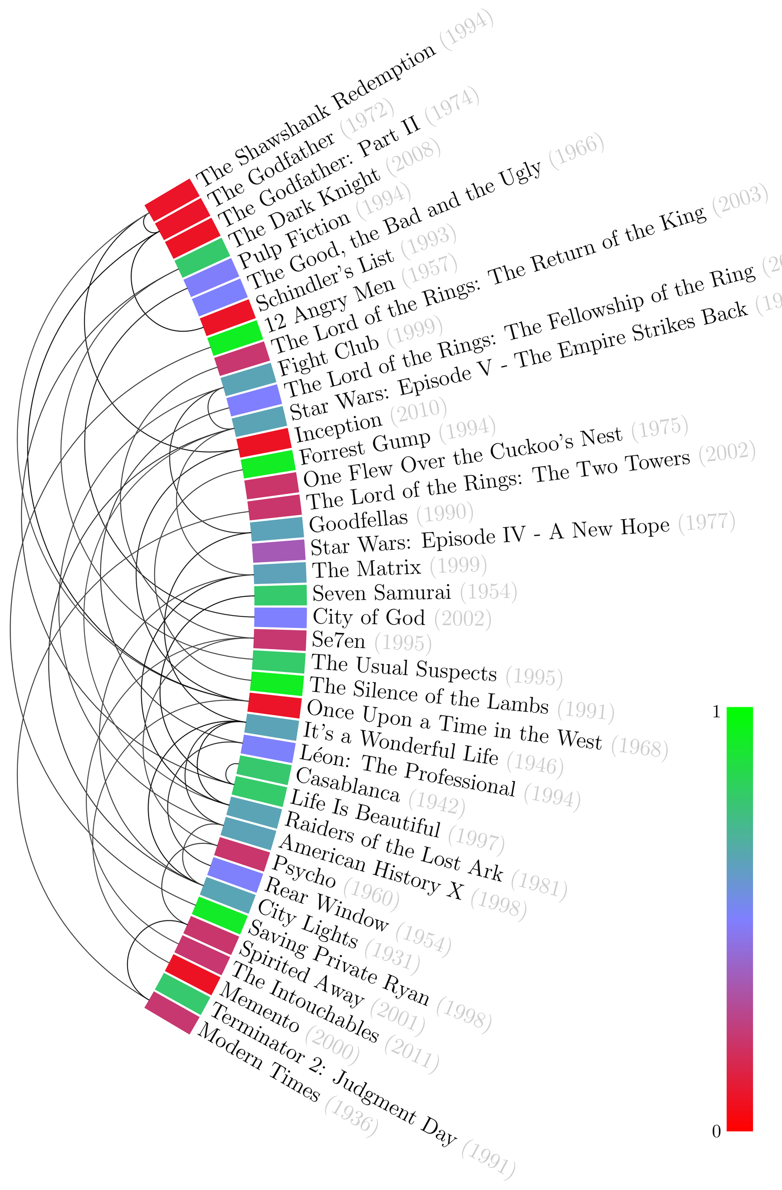
Spectral Thompson Sampling

Tomas.Kocak@inria.fr, Michal.Valko@inria.fr, Remi.Munos@inria.fr, and Shipra@microsoft.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
- **Existing solution:** SpectralUCB algorithm [3]
- **New solution:** SpectralTS (computationally more efficient)

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: $\mathcal{L} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\mathsf{T}$

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

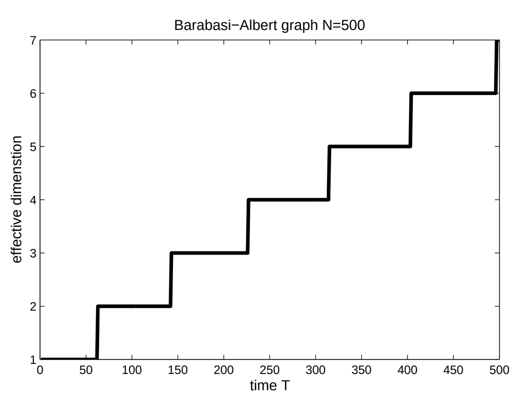
$$= \mathbf{f}^\mathsf{T}\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\mathsf{T}\mathbf{f} = \boldsymbol{\mu}^\mathsf{T}\boldsymbol{\Lambda}\boldsymbol{\mu} = \|\boldsymbol{\mu}\|_{\boldsymbol{\Lambda}} = \sum_{i=1}^N \lambda_i\mu_i^2$$

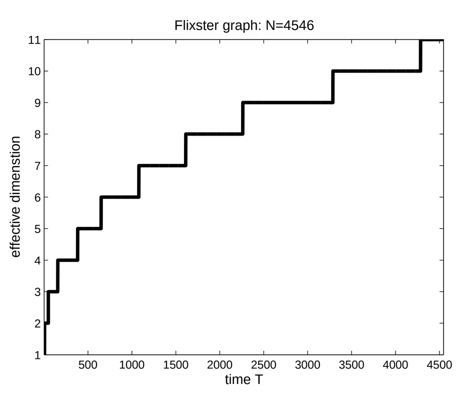
- **Observation:** $S_G(\mathbf{q}_i) = \lambda_i$
- **Smoothness and regularization:** Small value of **(a)** $S_G(f)$ **(b)** $\boldsymbol{\Lambda}$ norm of $\boldsymbol{\mu}$ **(c)** μ_i for large λ_i

Effective Dimension

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

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





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

Setting

- **Task:** Each time t, pick an action (node) to get a reward.
- **Reward:** $\mathbf{b}_i^\mathsf{T}\boldsymbol{\mu} + \varepsilon_t$ (with unknown parameter $\boldsymbol{\mu}$)
  - $\mathbf{b}_i$ 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

$$R_T = T\max_v \mathbf{b}_v^\mathsf{T}\boldsymbol{\mu} - \sum_{t=1}^T \mathbf{b}_{a(t)}^\mathsf{T}\boldsymbol{\mu}$$

Spectral Thompson Sampling

- Play arm which maximizes posterior probability of being the best
  - Sample $\tilde{\mu}$ from the distribution $\mathcal{N}(\hat{\mu}, v^2\mathbf{B}^{-1})$
  - Play arm which maximizes $\mathbf{b}^\mathsf{T}\tilde{\mu}$ and observe reward
- Compute posterior distribution according to reward received

Input:
N: number of arms, T: number of pulls
$\{\boldsymbol{\Lambda}_\mathcal{L}, \mathbf{Q}\}$: spectral basis of graph Laplacian $\mathcal{L}$
λ, δ: regularization and confidence parameters
R, C: upper bounds on noise and $\|\boldsymbol{\mu}\|_{\boldsymbol{\Lambda}}$
Initialization:
$v = R\sqrt{6d\log((\lambda+T)/\delta\lambda)} + C$
$\hat{\boldsymbol{\mu}} = 0_N$, $\mathbf{f} = 0_N$, $\mathbf{B} = \boldsymbol{\Lambda}_\mathcal{L} + \lambda\mathbf{I}_N$
Run:
for $t = 1$ **to** T **do**
Sample $\tilde{\boldsymbol{\mu}} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, v^2\mathbf{B}^{-1})$
$a(t) \leftarrow \arg\max_a \mathbf{b}_a^\mathsf{T}\tilde{\boldsymbol{\mu}}$
Observe a noisy reward $r(t) = \mathbf{b}_{a(t)}^\mathsf{T}\boldsymbol{\mu} + \varepsilon_t$
$\mathbf{f} \leftarrow \mathbf{f} + \mathbf{b}_{a(t)}r(t)$
Update $\mathbf{B} \leftarrow \mathbf{B} + \mathbf{b}_{a(t)}\mathbf{b}_{a(t)}^\mathsf{T}$
Update $\hat{\boldsymbol{\mu}} \leftarrow \mathbf{B}^{-1}\mathbf{f}$
end for

Main result

SpectralTS regret bound

Theorem 1. *Let d be the effective dimension and λ be the minimum eigenvalue of $\boldsymbol{\Lambda}$. If $\|\boldsymbol{\mu}\|_{\boldsymbol{\Lambda}} \le C$ and for all $\mathbf{b}_i$, $|\mathbf{b}_i^\mathsf{T}\boldsymbol{\mu}| \le 1$, then the cumulative regret of Spectral Thompson Sampling is with probability at least $1-\delta$ bounded as*

$$\mathcal{R}(T) \le \frac{11g}{p}\sqrt{\frac{4+4\lambda}{\lambda}dT\log\frac{\lambda+T}{\lambda}} + \frac{1}{T} + \frac{g}{p}\left(\frac{11}{\sqrt{\lambda}}+2\right)\sqrt{2T\log\frac{2}{\delta}},$$

where $p = 1/(4e\sqrt{\pi})$ and

$$g = \sqrt{4\log TN}\left(R\sqrt{6d\log\left(\frac{\lambda+T}{\delta\lambda}\right)}+C\right) + R\sqrt{2d\log\left(\frac{(\lambda+T)T^2}{\delta\lambda}\right)} + C.$$

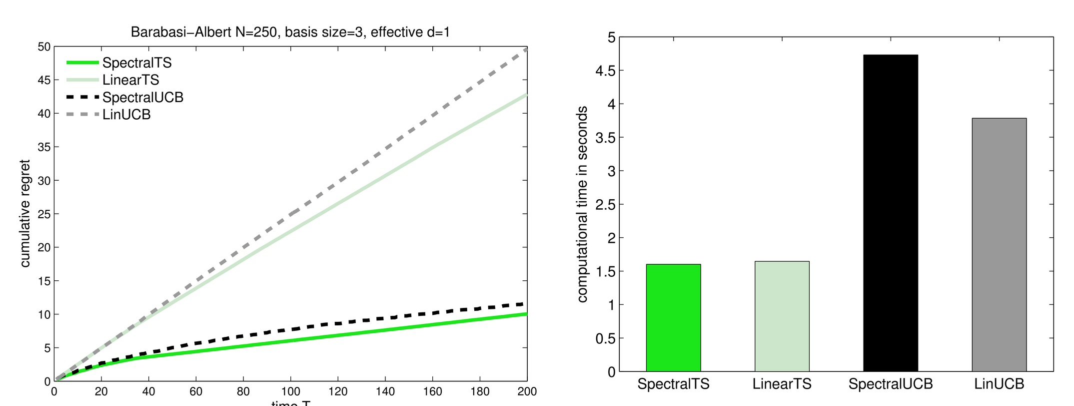
Setting $\boldsymbol{\Lambda} = \mathbf{I}$ we recover LinearTS. Since $\log(|\mathbf{B}_T|/|\boldsymbol{\Lambda}|)$ can be upperbounded by $D\log T$ [1], we obtain $\tilde{\mathcal{O}}(D\sqrt{T})$ for LinearTS.

Linear vs. Spectral Bandits

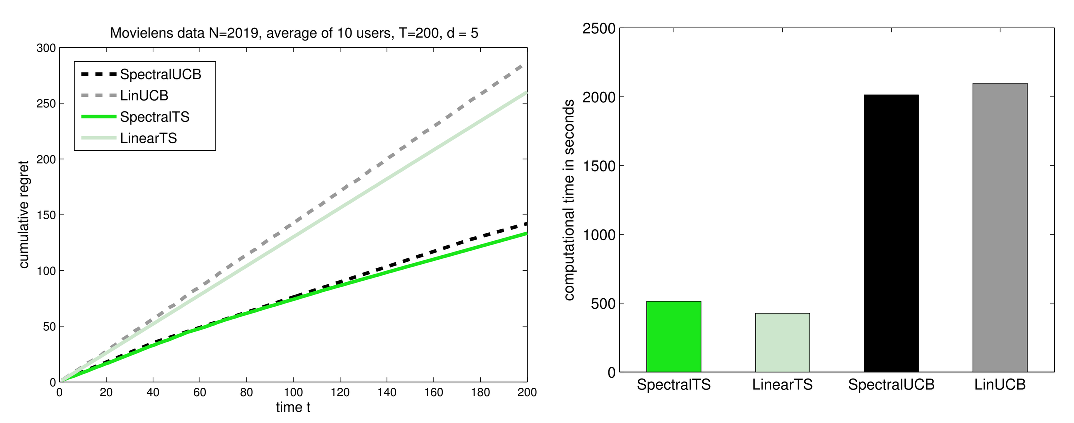
	Linear	*Spectral*
Optimistic Approach D^2N per step update	**LinUCB** $D\sqrt{T}\log T$	**SpectralUCB** $d\sqrt{T}\log T$
Thompson Sampling D^2+DN per step update	**LinearTS** $D\sqrt{T\log N}$	***SpectralTS*** $d\sqrt{T\log N}$

Experiments

Synthetic Experiment
Barabási-Albert (BA) model with the degree parameter 3.



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



Spectral Thompson Sampling

- Better regret than LinearTS and LinUCB
- Better run time than LinUCB and SpectralUCB

Analysis sketch

Divide arms into two groups

- $\Delta_i = \mathbf{b}_*^\mathsf{T}\boldsymbol{\mu} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu} \le g\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$ arm i is **unsaturated**
- $\Delta_i = \mathbf{b}_*^\mathsf{T}\boldsymbol{\mu} - \mathbf{b}_i^\mathsf{T}\boldsymbol{\mu} > g\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$ arm i is **saturated**

Saturated arm

- Small standard deviation and high regret
- Low probability of picking

Unsaturated arm

- Low regret bounded by a factor of standard deviation
- High probability of picking

Regret on playing unsaturated arm
By self-normalized bound of [1], with probability $1-\delta/T^2$:

$$|\mathbf{b}_i^\mathsf{T}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})| \le \|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}\left(R\sqrt{2d\log\left(\frac{|\mathbf{B}_t|T^2}{|\boldsymbol{\Lambda}|\delta}\right)}+C\right)$$

With probability at least $1-1/T^2$:

$$|\mathbf{b}_i^\mathsf{T}(\tilde{\boldsymbol{\mu}}-\hat{\boldsymbol{\mu}})| \le \left(R\sqrt{6d\log\left(\frac{|\mathbf{B}_t|}{|\boldsymbol{\Lambda}|\delta}\right)}+C\right)\|\mathbf{x}_i\|_{\mathbf{B}_t^{-1}}\sqrt{4\ln(TN)}$$

Our key result coming from spectral properties of $\mathbf{B}_t$:

$$\log\frac{|\mathbf{B}_t|}{|\boldsymbol{\Lambda}|} \le 2d\log\left(1+\frac{T}{\lambda}\right)$$

Together we get:

$$|\mathbf{b}_i^\mathsf{T}(\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu})| \le g\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}$$

Super-martingale process

$$\text{regret}'(t) = \text{regret}(t)\cdot\mathbb{1}\{|\mathbf{b}_i^\mathsf{T}\hat{\boldsymbol{\mu}}(t)-\mathbf{b}_i^\mathsf{T}\boldsymbol{\mu}| \le l\|\mathbf{b}_i\|_{\mathbf{B}_t^{-1}}\}$$

$$\mathcal{F}_t = \{a(\tau), r(\tau), \tau = 1,\dots,t\} \cup \{\mathbf{b}_i, i = 1,\dots,N\}$$

$$X_t = \text{regret}'(t) - \frac{11g}{p}\|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} - \frac{1}{T^2}$$

$$Y_t = \sum_{w=1}^t X_w.$$

where $l = R\sqrt{2d\log((\lambda+T)T^2/(\delta\lambda))} + C$.
$(Y_t;\ t = 0,\dots,T)$ is a super-martingale process w.r.t. filtration $\mathcal{F}_t$.
With probability $1-\delta/2$:

$$\text{regret}(t) = \text{regret}'(t)$$

Azuma-Hoeffding inequality for super-martingale, w. p. $1-\delta/2$:

$$\sum_{t=1}^T \text{regret}'(t) \le \frac{11g}{p}\sum_{t=1}^T\|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} + \frac{1}{T} + \frac{g}{p}\left(\frac{11}{\sqrt{\lambda}}+2\right)\sqrt{2T\ln\frac{2}{\delta}}.$$

By Cauchy-Schwartz inequality:

$$\sum_{t=1}^T\|\mathbf{b}_{a(t)}\|_{\mathbf{B}_t^{-1}} \le \sqrt{T\sum_{t=1}^T\|\mathbf{b}_{a(t)}\|^2_{\mathbf{B}_t^{-1}}} \le \sqrt{T\left(2+\frac{2}{\lambda}\right)\ln\frac{|\mathbf{B}_T|}{|\boldsymbol{\Lambda}|}} \le \sqrt{\frac{4+4\lambda}{\lambda}dT\ln\frac{\lambda+T}{\lambda}}.$$

References

[1] Yasin Abbasi-Yadkori, David Pál, and Csaba Szepesvári. Improved Algorithms for Linear Stochastic Bandits. In *Neural Information Processing Systems*. 2011.

[2] Shipra Agrawal and Navin Goyal. Thompson Sampling for Contextual Bandits with Linear Payoffs. In *International Conference on Machine Learning*, 2013.

[3] Michal Valko, Rémi Munos, Branislav Kveton, and Tomáš Kocák. Spectral Bandits for Smooth Graph Functions. In *31th International Conference on Machine Learning*, 2014.

Acknowledgments

intel
CompLACS