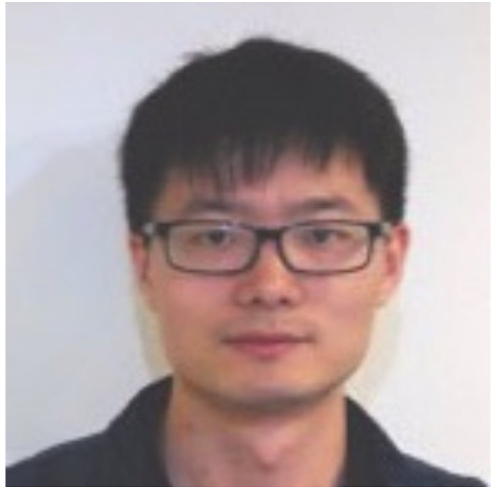
Laplacian-regularized Graph Bandits

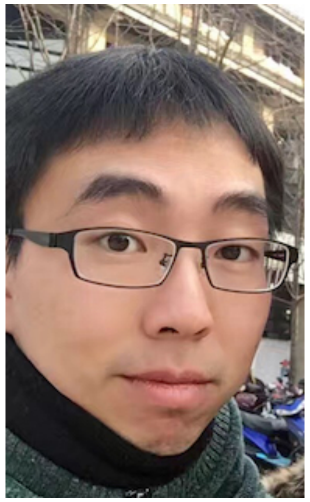
Laura Toni
University College London

25 September 2019

Laplacian-regularized Graph Bandits

Laura Toni





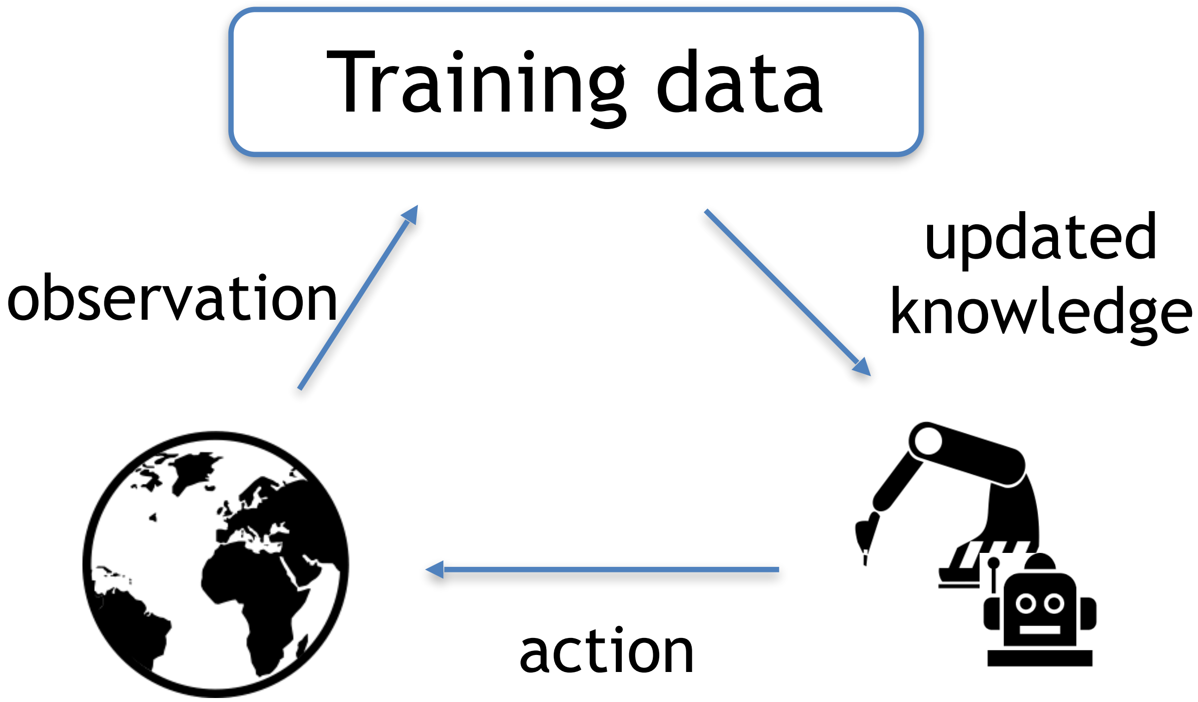
Kaige Yang
UCL

Xiaowen Dong
University of Oxford

Outline

- Graphs and Bandit
- Importance of Graphs in Decision-Making
- A Laplacian Perspective
- Output and Intuitions
- Conclusions

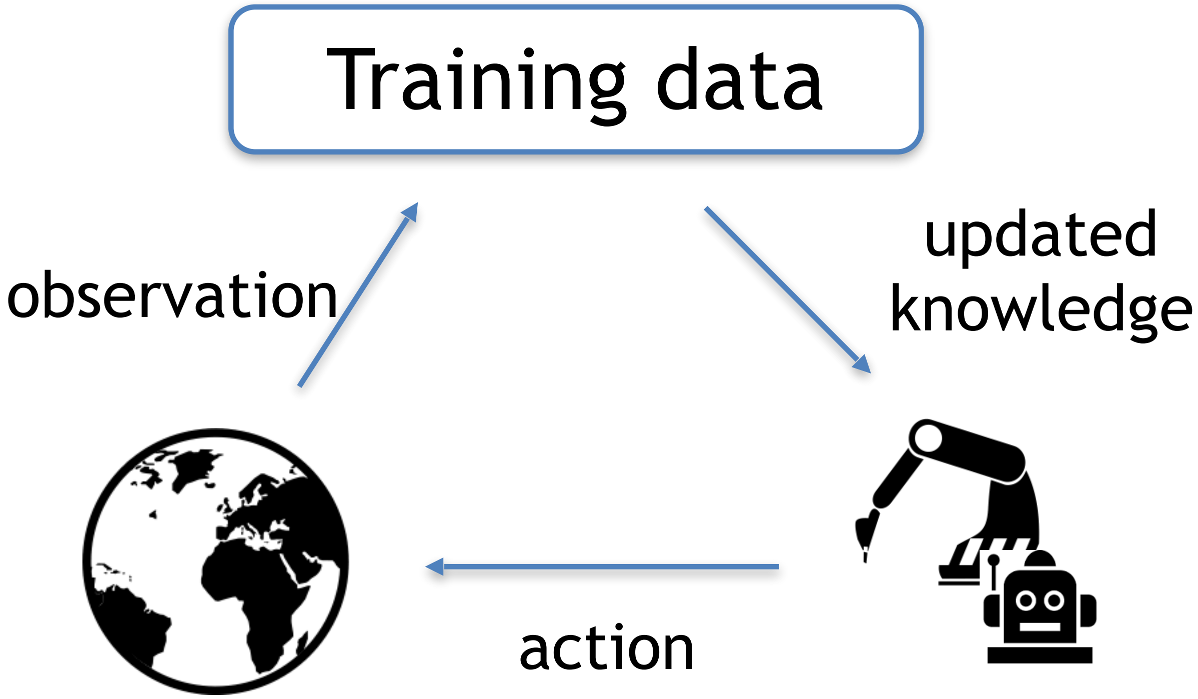
Main Challenges in DMS



Theoretically addressed by

- Multi-arm bandit problem
- Reinforcement Learning

Main Challenges in DMS



Theoretically addressed by

- Multi-arm bandit problem
- Reinforcement Learning

- Find the optimal trade-off between **exploration** and **exploitation** ⇨ bandit and RL problems
- **Sampling-efficiency**: the learning performance does not scale with the ambient dimension (number of arms, states, etc) ⇨ structured learning

Structured DMS - Main Challenges

- In DMSs, context or action payoffs (data) have **semantically reach information**

> Structured problems obviate the curse of dimensionality by exploiting the data structure

Structured DMS - Main Challenges

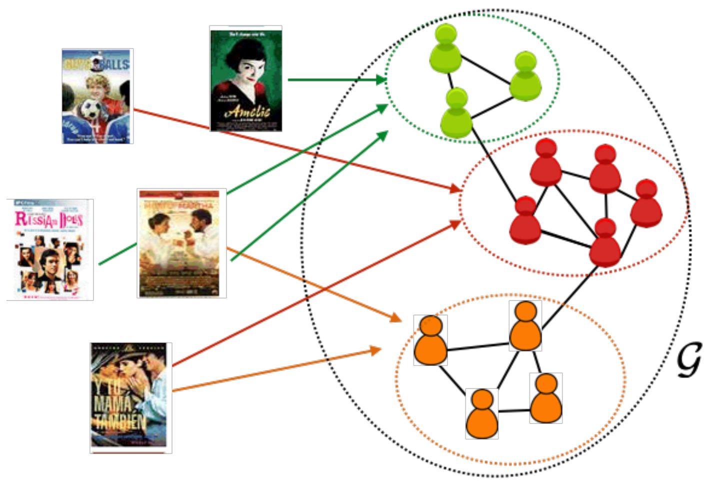
- In DMSs, context or action payoffs (data) have **semantically reach information**

> Structured problems obviate the curse of dimensionality by exploiting the data structure

Structured DMS - Main Challenges

- In DMSs, context or action payoffs (data) have **semantically reach information**

Structured problems obviate the curse of dimensionality by exploiting the data structure



Graph Clustering

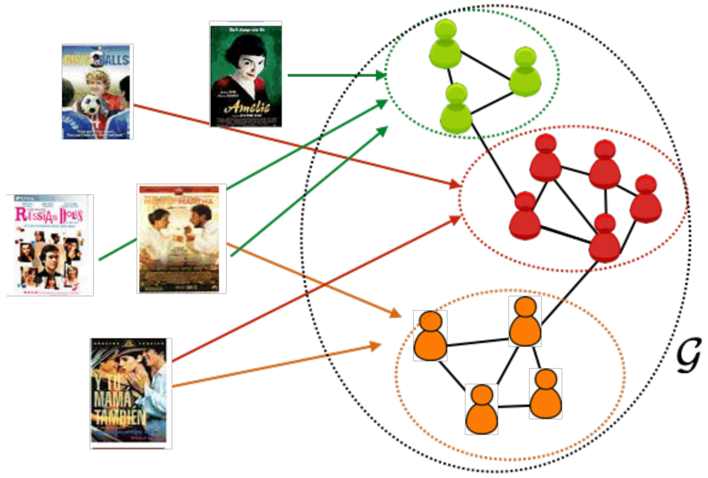
- reducing the curse of dimensionality
- degradation in real-world data

C. Gentile et al., "Online Clustering of Bandits", ICML 2014

Structured DMS - Main Challenges

- In DMSs, context or action payoffs (data) have **semantically reach information**

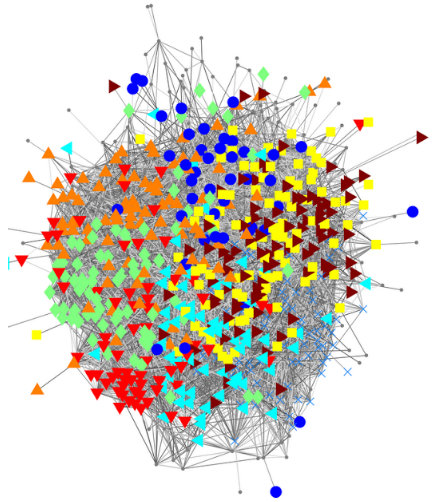
> Structured problems obviate the curse of dimensionality by exploiting the data structure



Graph Clustering

- reducing the curse of dimensionality
- degradation in real-world data

Need for more sophisticated frameworks (than clustering) to handle high-dimensional and structured data



C. Gentile et al., “Online Clustering of Bandits”, ICML 2014

Structured DMS - Main Challenges

- In DMSs, context or action payoffs (data) have **semantically reach information**

- It is important to identify and **leverage the structure** underneaths these data

Many works on Bandit are graph based, see overview [1]

- **data-structure in bandits:**
  - Gentile, C., Li, S., and Zappella, G. "*Online clustering of bandits*", *ICML 2014*
  - Korda, N., Szorenyi, B., and Shuai, L. "*Distributed clustering of linear bandits in peer to peer networks*", *JMLR, 2016*
  - Yang, K. and Toni, L., "*Graph-based recommendation system*", IEEE GlobalSIP, 2018

[1] Michal Valko: Bandits on graphs and structures, habilitation thesis, École normale supérieure de Cachan (ENS Cachan 2016)

Structured DMS - Main Challenges

- In DMSs, context or action payoffs (data) have **semantically reach information**

- It is important to identify and **leverage the structure** underneaths these data

Many works on Bandit are graph based, see overview [1]

- **data-structure in bandits:**
  - Gentile, C., Li, S., and Zappella, G. "*Online clustering of bandits*", *ICML 2014*
  - Korda, N., Szorenyi, B., and Shuai, L. "*Distributed clustering of linear bandits in peer to peer networks*", *JMLR, 2016*
  - Yang, K. and Toni, L., "*Graph-based recommendation system*", IEEE GlobalSIP, 2018

can we capture the external information beyond data-structure?

[1] Michal Valko: Bandits on graphs and structures, habilitation thesis, École normale supérieure de Cachan (ENS Cachan 2016)

Structured DMS - Main Challenges

- In DMSs, context or action payoffs (data) have **semantically reach information**

- It is important to identify and **leverage the structure** underneaths these data

Many works on Bandit are graph based, see overview [1]

- **spectral bandits:**
  - N. Cesa-Bianchi, et al., *"A gang of bandits"*, *NeurIPS* 2013
  - M. Valko, et al., "*Spectral bandits for smooth graph functions*", ICML 2014
  - S. Vaswani, M Schmidt, and L. Lakshmanan, "*Horde of bandits using gaussian markov random fields*", arXiv, 2017.
  - other recent works on asynchronous and decentralized network bandits

[1] Michal Valko: Bandits on graphs and structures, habilitation thesis, École normale supérieure de Cachan (ENS Cachan 2016)

Structured DMS - Main Challenges

- In DMSs, context or action payoffs (data) have **semantically reach information**

- It is important to identify and **leverage the structure** underneaths these data

Many works on Bandit are graph based, see overview [1]

- **spectral bandits:**
  - N. Cesa-Bianchi, et al., *"A gang of bandits"*, *NeurIPS* 2013
  - M. Valko, et al., "*Spectral bandits for smooth graph functions*", ICML 2014
  - S. Vaswani, M Schmidt, and L. Lakshmanan, "*Horde of bandits using gaussian markov random fields*", arXiv, 2017.
  - other recent works on asynchronous and decentralized network bandits

    - single user bandit
    - no per-user error bound —> coarse regret upper bounds scaling linearly with the number of users
    - high computational complexity

[1] Michal Valko: Bandits on graphs and structures, habilitation thesis, École normale supérieure de Cachan (ENS Cachan 2016)

Structured DMS - Main Challenges

- In DMSs, context or action payoffs (data) have **semantically reach information**

- It is important to identify and **leverage the structure** underneaths these data

- Highly interesting studies on graph-bandit already published, but most of them work in the **graph spatial (vertex) domain**

Structured DMS - Main Challenges

- In DMSs, context or action payoffs (data) have **semantically reach information**

- It is important to identify and **leverage the structure** underneaths these data

- Highly interesting studies on graph-bandit already published, but most of them work in the **graph spatial (vertex) domain**

- Data can be high-dimensional, time-varying, and composition of superimposed phenomena.
- Need proper framework to capture both data-structure and external-geometry information (graphs)

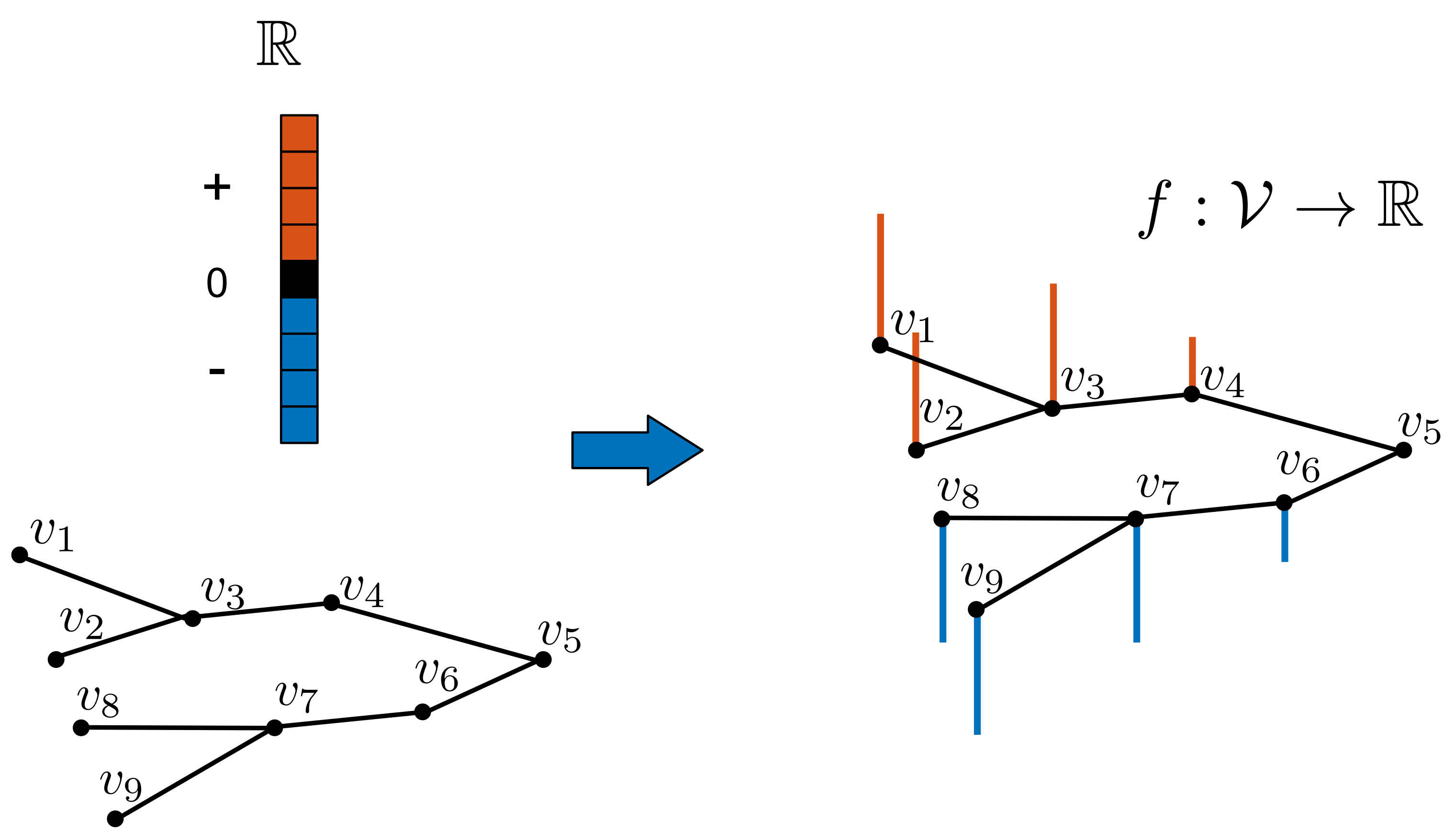
Structured DMS - Main Challenges

- In DMSs, context or action payoffs (data) have **semantically reach information**
- It is important to identify and **leverage the structure** underneaths these data
- Highly interesting studies on graph-bandit already published, but most of them work in the **graph spatial (vertex) domain**
- Data can be high-dimensional, time-varying, and composition of superimposed phenomena.
- Need proper framework to capture both data-structure and external-geometry information (graphs)

Graph signal processing (GSP) can be applied to DMSs to address the above challenges and needs

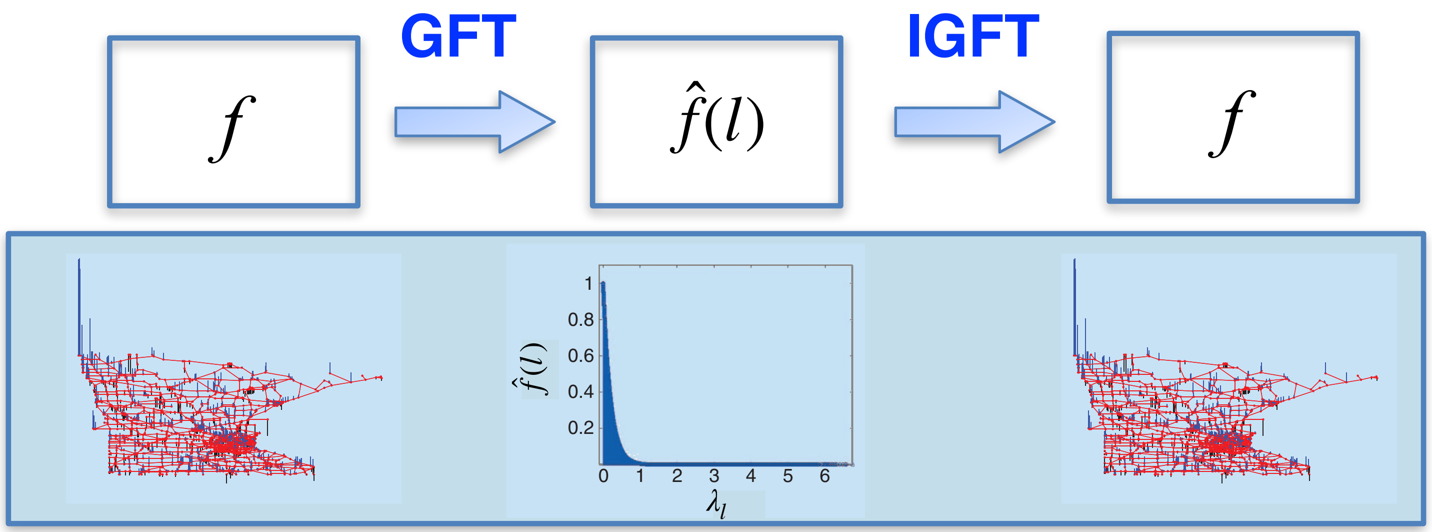
Graph Signal Processing

Structured but irregular data can be represented by graph signals



Goal: to capture both structure (edges) and data (values at vertices)

Frequency Analysis

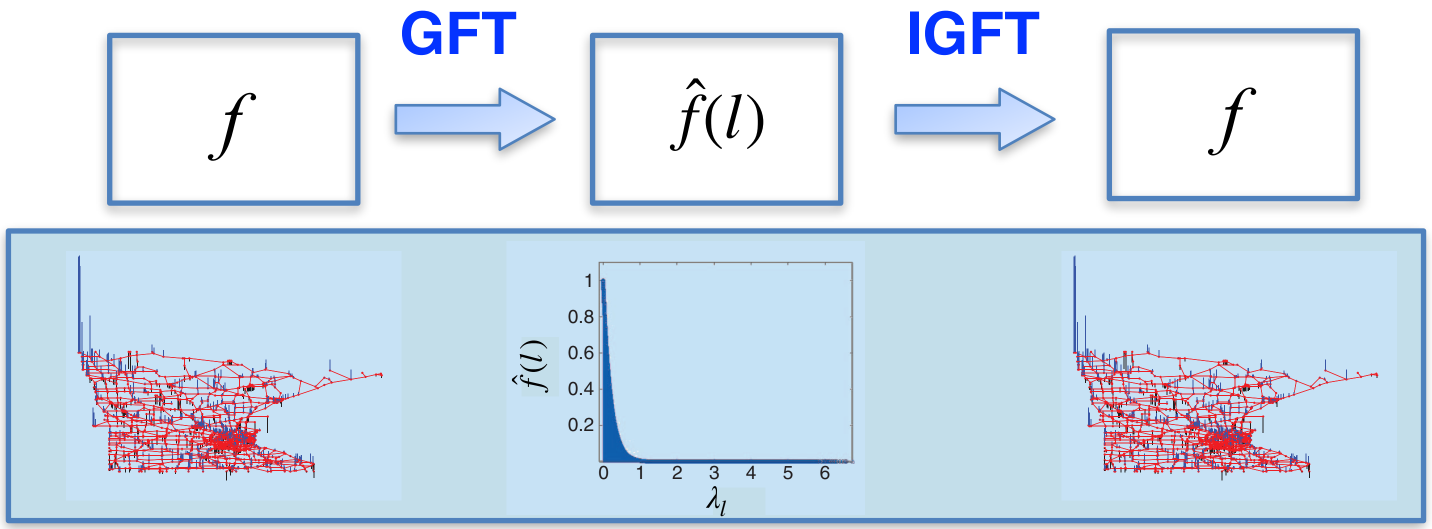


$$\hat{f}(l) = \langle f, \chi_l \rangle = \sum_{n=1}^{N} f(n)\chi_l^*(n)$$

$$f(n) = \sum_{l=0}^{N-1} \hat{f}(l)\chi_l(n), \ \forall n \in$$

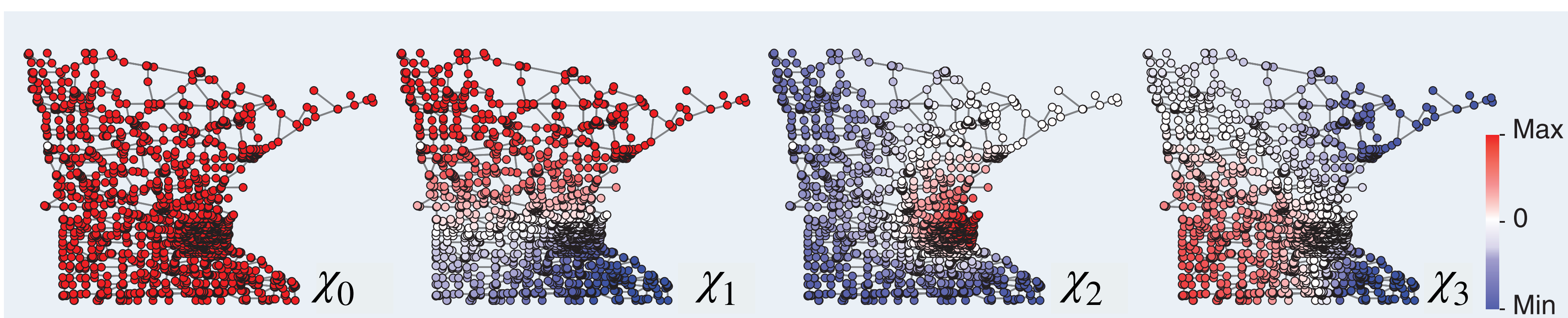
- Shuman, David I., et al. "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains." *IEEE signal processing magazine* 30.3 (2013): 83-98
- Bronstein, Michael M., et al. "Geometric deep learning: going beyond euclidean data." *IEEE Signal Processing Magazine* 34.4 (2017): 18-42.

Frequency Analysis



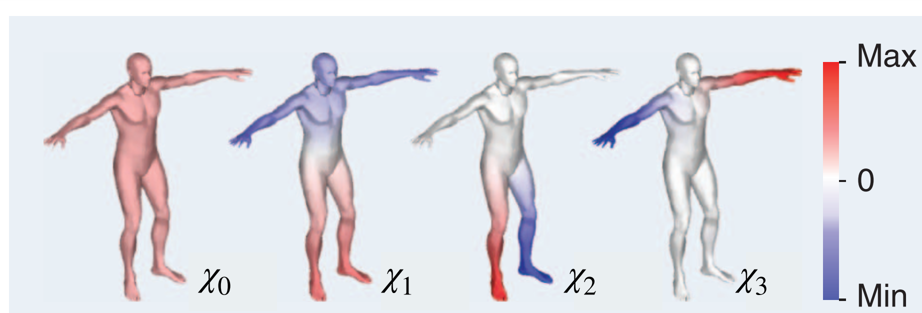
$$\hat{f}(l) = \langle f, \chi_l \rangle = \sum_{n=1}^{N} f(n)\chi_l^*(n)$$

$$f(n) = \sum_{l=0}^{N-1} \hat{f}(l)\chi_l(n), \ \forall n \in$$



low frequency → high frequency

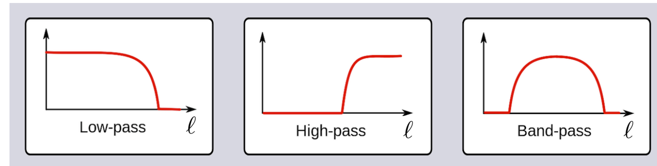
$$\chi_0^T L \chi_0 = \lambda_0$$



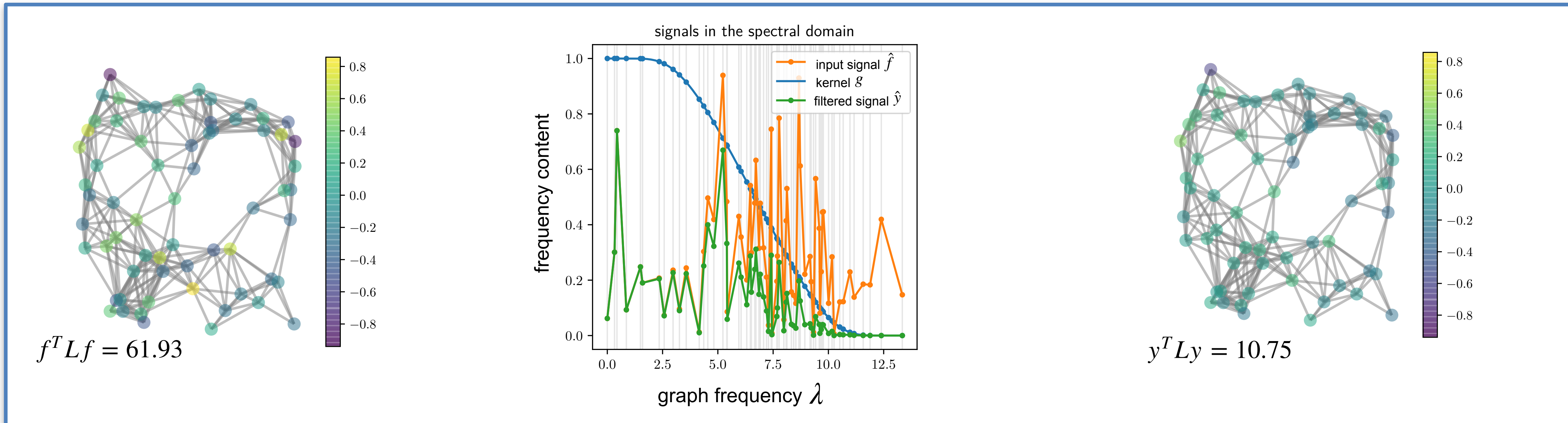
- Shuman, David I., et al. "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains." *IEEE signal processing magazine* 30.3 (2013): 83-98
- Bronstein, Michael M., et al. "Geometric deep learning: going beyond euclidean data." *IEEE Signal Processing Magazine* 34.4 (2017): 18-42.

Filtering and Smoothness

Filtering



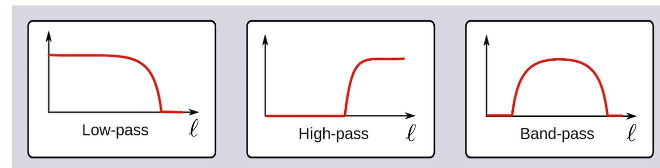
f $\xrightarrow{\text{GFT}}$ $\hat{f}(l)$ $\xrightarrow{\hat{g}(\lambda_l)}$ $\hat{g}(\lambda_l)\hat{f}(l)$ $\xrightarrow{\text{IGFT}}$ $y = \sum_{l=0}^{N-1} \hat{g}(\lambda_l)\hat{f}(l)\chi_l(n)$



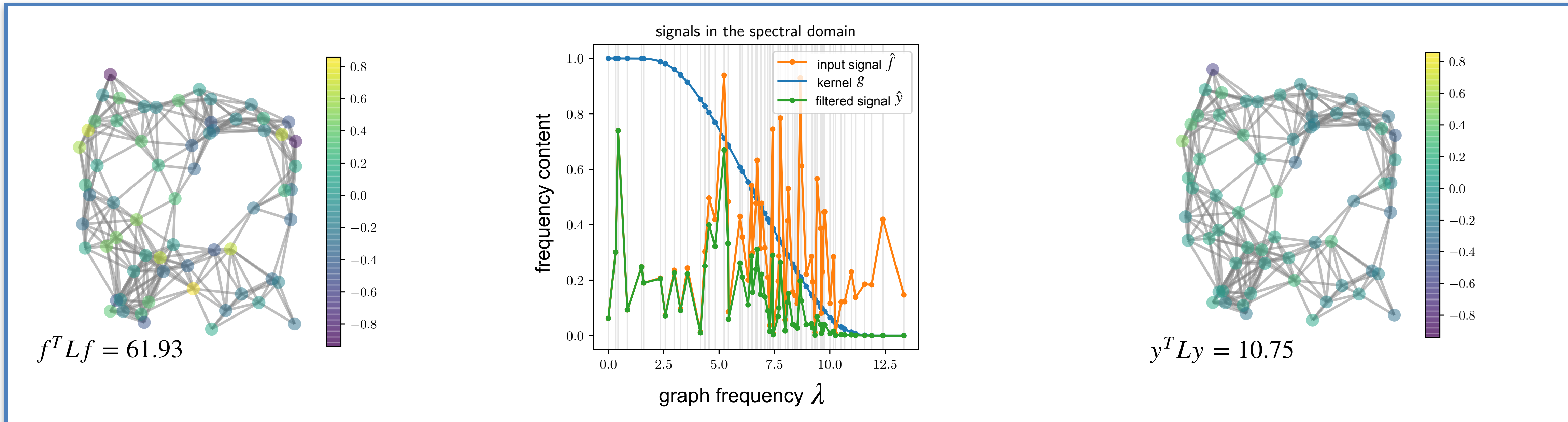
M. Defferrard, Deep Learning on Graphs: a journey from continuous manifolds to discrete networks (KCL/UCL Junior Geometry Seminar)

Filtering and Smoothness

Filtering



$$f \xrightarrow{\text{GFT}} \hat{f}(l) \xrightarrow{\hat{g}(\lambda_l)} \hat{g}(\lambda_l)\hat{f}(l) \xrightarrow{\text{IGFT}} y = \sum_{l=0}^{N-1} \hat{g}(\lambda_l)\hat{f}(l)\chi_l(n)$$



Denoising problem

$$y^\star = \arg\min_y \left\{ ||y - f||_2^2 + \gamma y^T L y \right\}$$

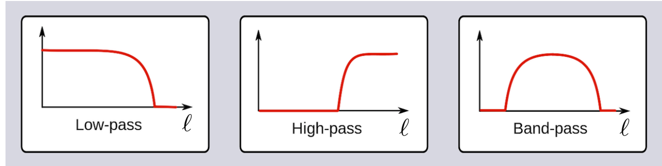
$$y^\star = (I + \gamma L)^{-1} f = \chi \left(I + \gamma \Lambda\right)^{-1} \chi^T f$$

remove noise by low-pass filtering in the graph spectral domain

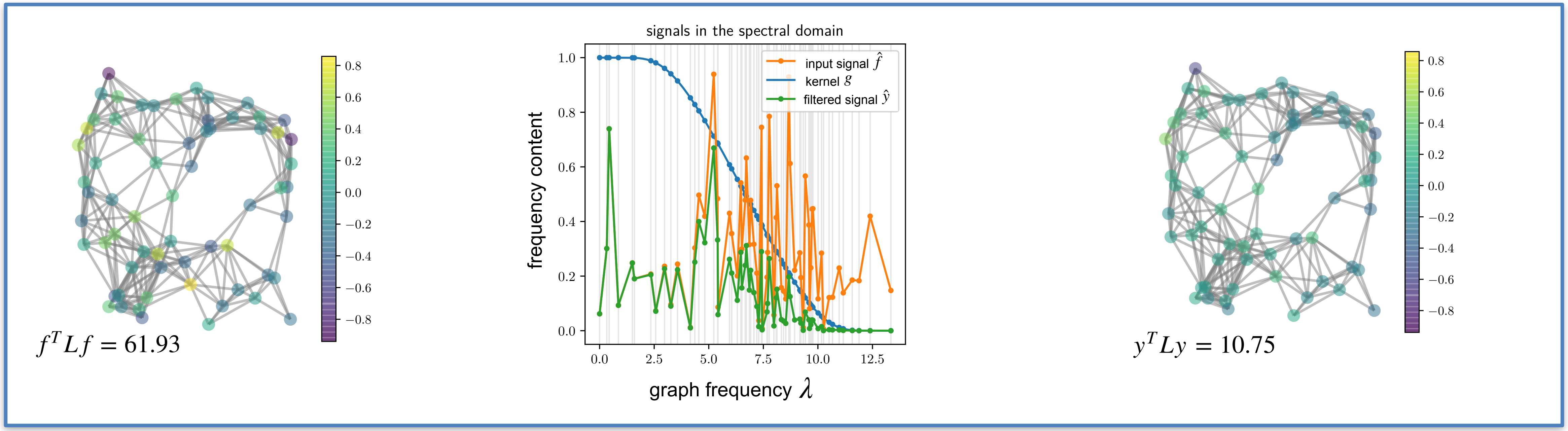
M. Defferrard, Deep Learning on Graphs: a journey from continuous manifolds to discrete networks (KCL/UCL Junior Geometry Seminar)

Filtering and Smoothness

Filtering



$$f \xrightarrow{\text{GFT}} \hat{f}(l) \xrightarrow{\hat{g}(\lambda_l)} \hat{g}(\lambda_l)\hat{f}(l) \xrightarrow{\text{IGFT}} y = \sum_{l=0}^{N-1} \hat{g}(\lambda_l)\hat{f}(l)\chi_l(n)$$



Denoising problem

$$y^\star = \arg\min_y \left\{ ||y - f||_2^2 + \gamma y^T L y \right\}$$

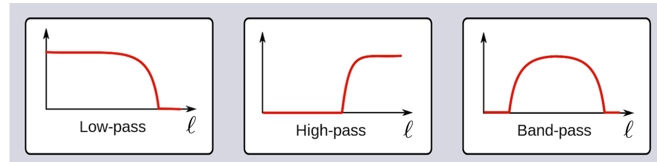
$$y^\star = \underbrace{\boxed{(I + \gamma L)^{-1}}}_{g(L)} f = \chi \left(I + \gamma \Lambda\right)^{-1} \chi^T f$$

remove noise by low-pass filtering in the graph spectral domain

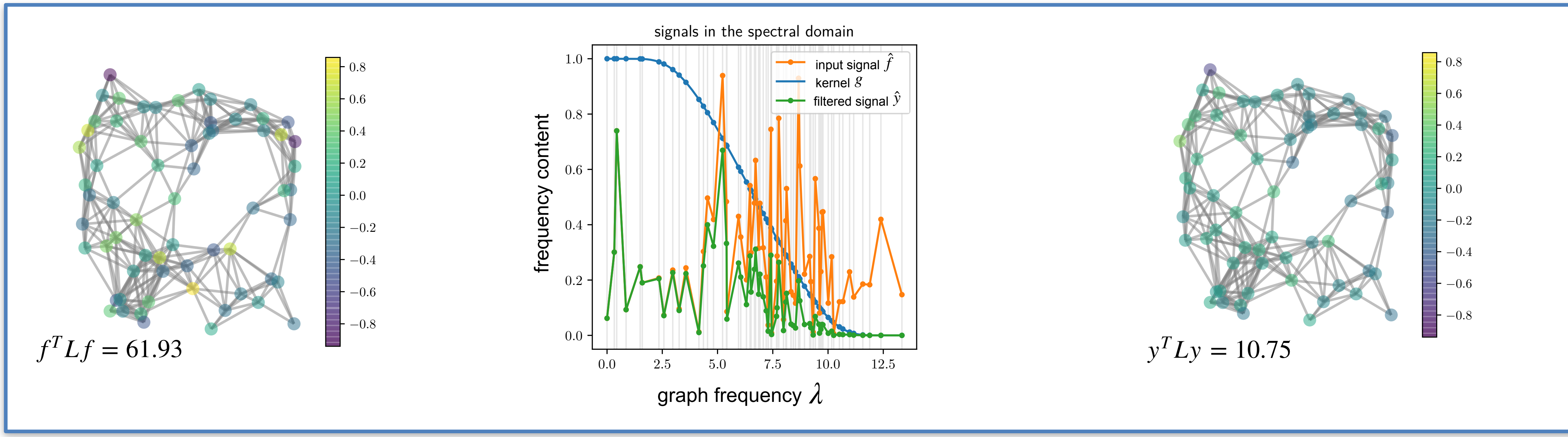
M. Defferrard, Deep Learning on Graphs: a journey from continuous manifolds to discrete networks (KCL/UCL Junior Geometry Seminar)

Filtering and Smoothness

Filtering



$$f \xrightarrow{\text{GFT}} \hat{f}(l) \xrightarrow{\hat{g}(\lambda_l)} \hat{g}(\lambda_l)\hat{f}(l) \xrightarrow{\text{IGFT}} y = \sum_{l=0}^{N-1} \hat{g}(\lambda_l)\hat{f}(l)\chi_l(n)$$



Denoising problem

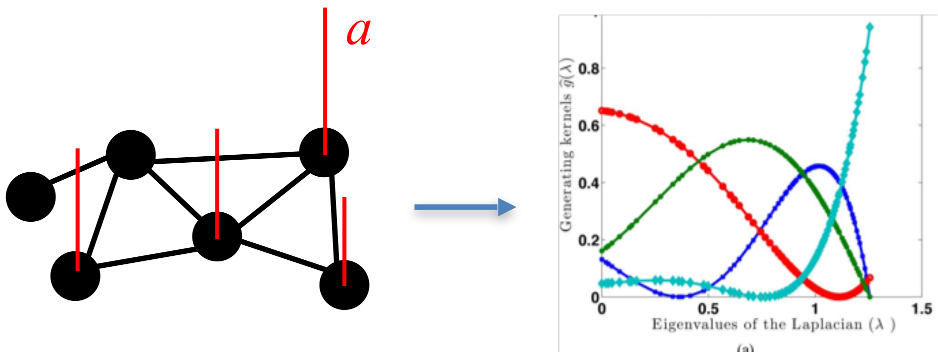
$$y^\star = \arg\min_y \left\{ ||y - f||_2^2 + \gamma y^T L y \right\}$$

$$y^\star = \underbrace{\boxed{(I + \gamma L)^{-1}}}_{g(L)} f = \chi \underbrace{\boxed{(I + \gamma \Lambda)^{-1} \underbrace{\boxed{\chi^T f}}_{\hat{f}(l)}}}_{\hat{y}(l)}$$

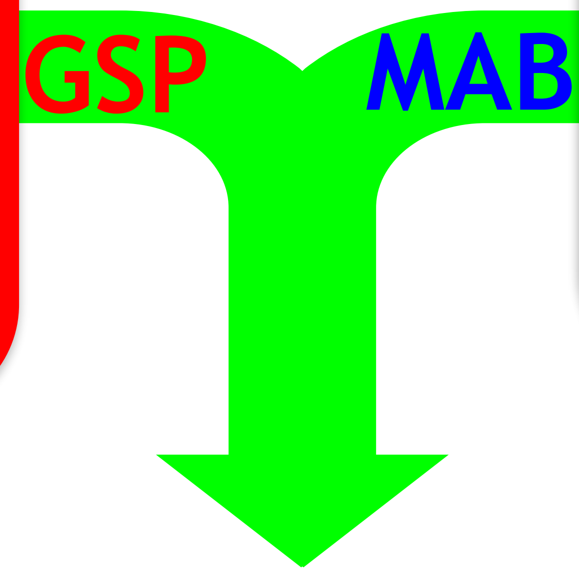
remove noise by low-pass filtering in the graph spectral domain

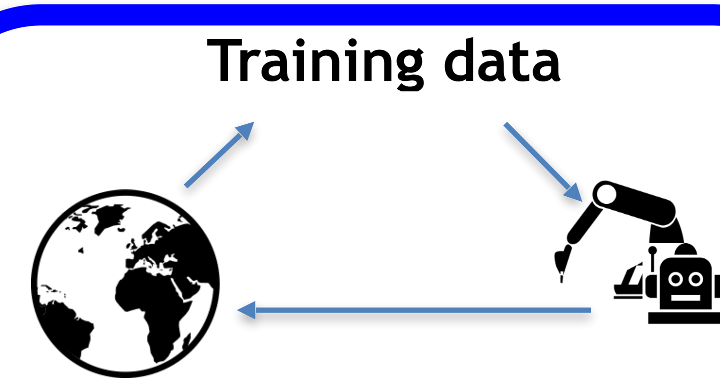
M. Defferrard, Deep Learning on Graphs: a journey from continuous manifolds to discrete networks (KCL/UCL Junior Geometry Seminar)

GSP for Online DMS



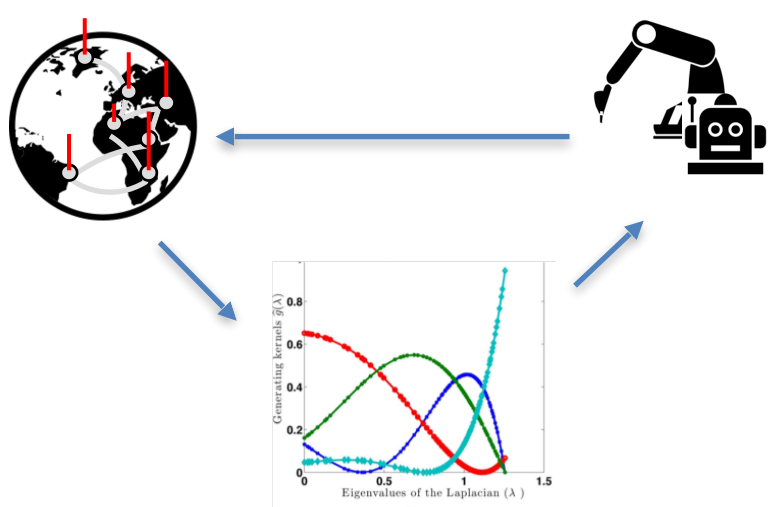
GSP to exploit spectral properties





Exploration exploitation trade-off

GSP-Based MAB



- **Data-efficiency:** learn in a sparse domain
- **Accuracy:** learning representation that preserves the geometry of the problem
- **Mathematical framework** is missing
- Not many works **beyond smoothness**

Outline

- Graphs and Bandit
- Importance of Graphs in Decision-Making
- A Laplacian Perspective
- Output and Intuitions
- Conclusions

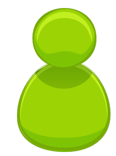
Recommendation Model

Aim: Infer the best item by running a sequence of trials

item 1 

item 2 

?



$$y = \boldsymbol{x}^T \boldsymbol{\theta} + \eta$$

$\boldsymbol{x} \in \mathbb{R}^d$: item feature vector

$\boldsymbol{\theta} \in \mathbb{R}^d$: user parameter vector

y : linear payoff

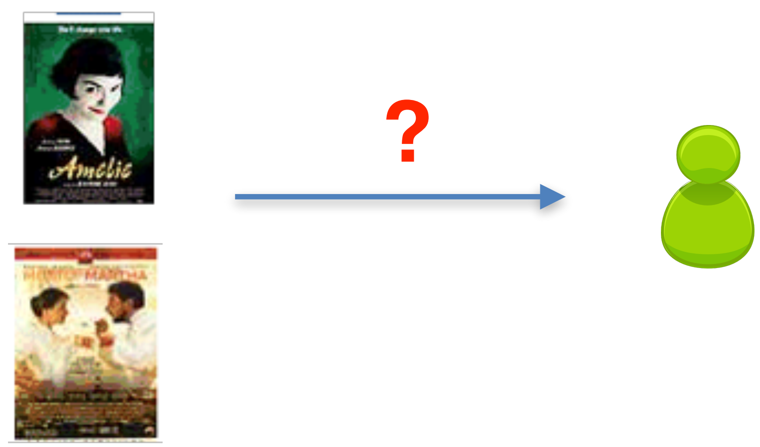
η : σ – sub-Gaussian noise

Recommendation Model

Aim: Infer the best item by running a sequence of trials

item 1

item 2



$$y = \boldsymbol{x}^T\boldsymbol{\theta} + \eta$$

Well known **bandit problem** with assumptions:
(i) stochasticity, (ii) i.i.d.,
(iii) stationarity

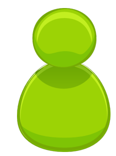
Recommendation Model

Aim: Infer the best item by running a sequence of trials

item 1 

item 2 

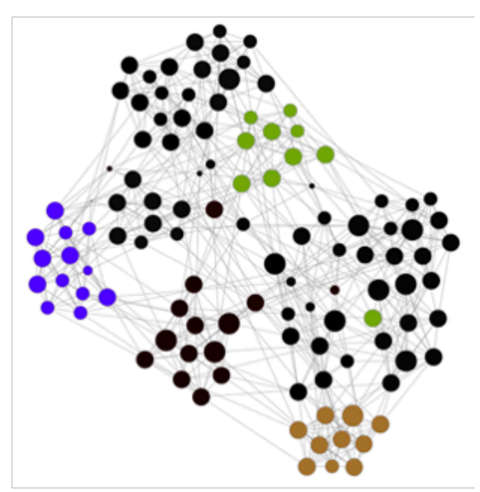
?



$$y = \boldsymbol{x}^T \boldsymbol{\theta} + \eta$$

Well known **bandit problem** with assumptions:
(i) stochasticity, (ii) i.i.d.,
(iii) stationarity

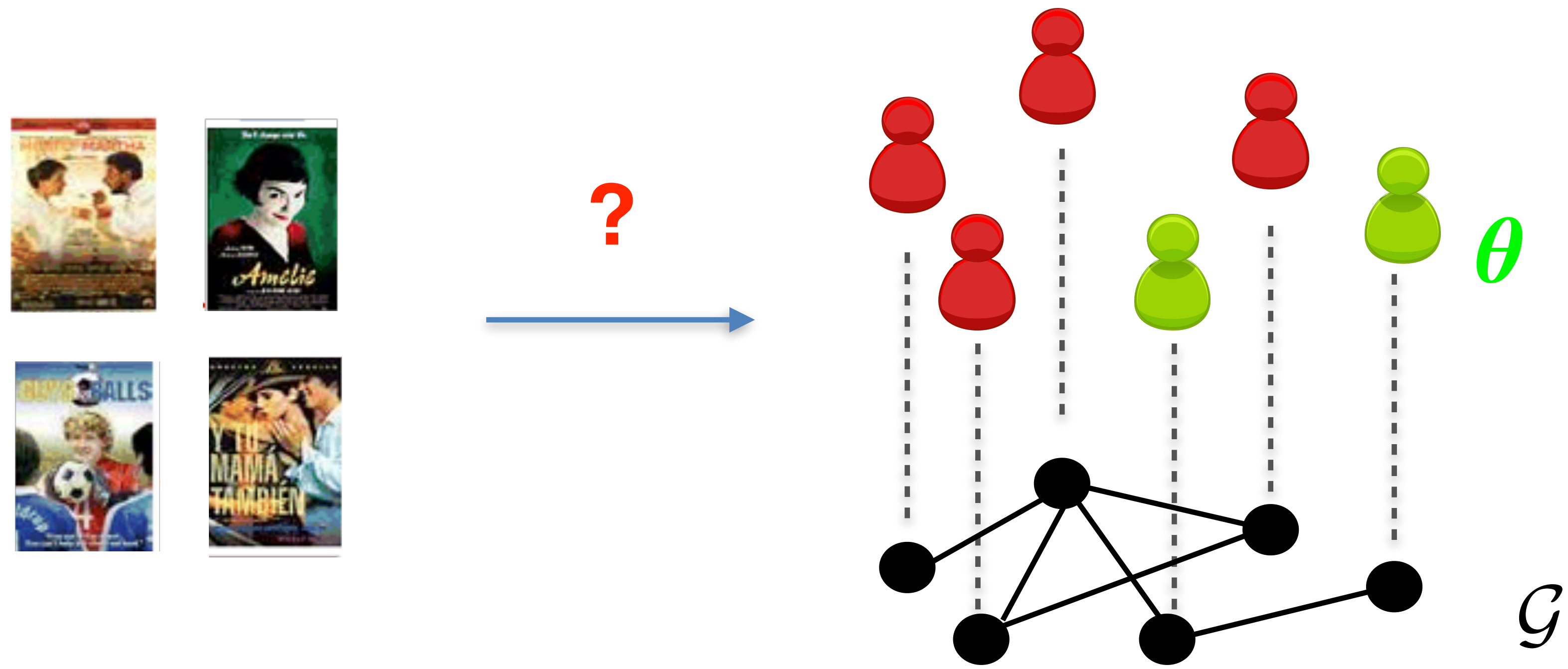
?



Our interest: multi-user (high-dimensional) case

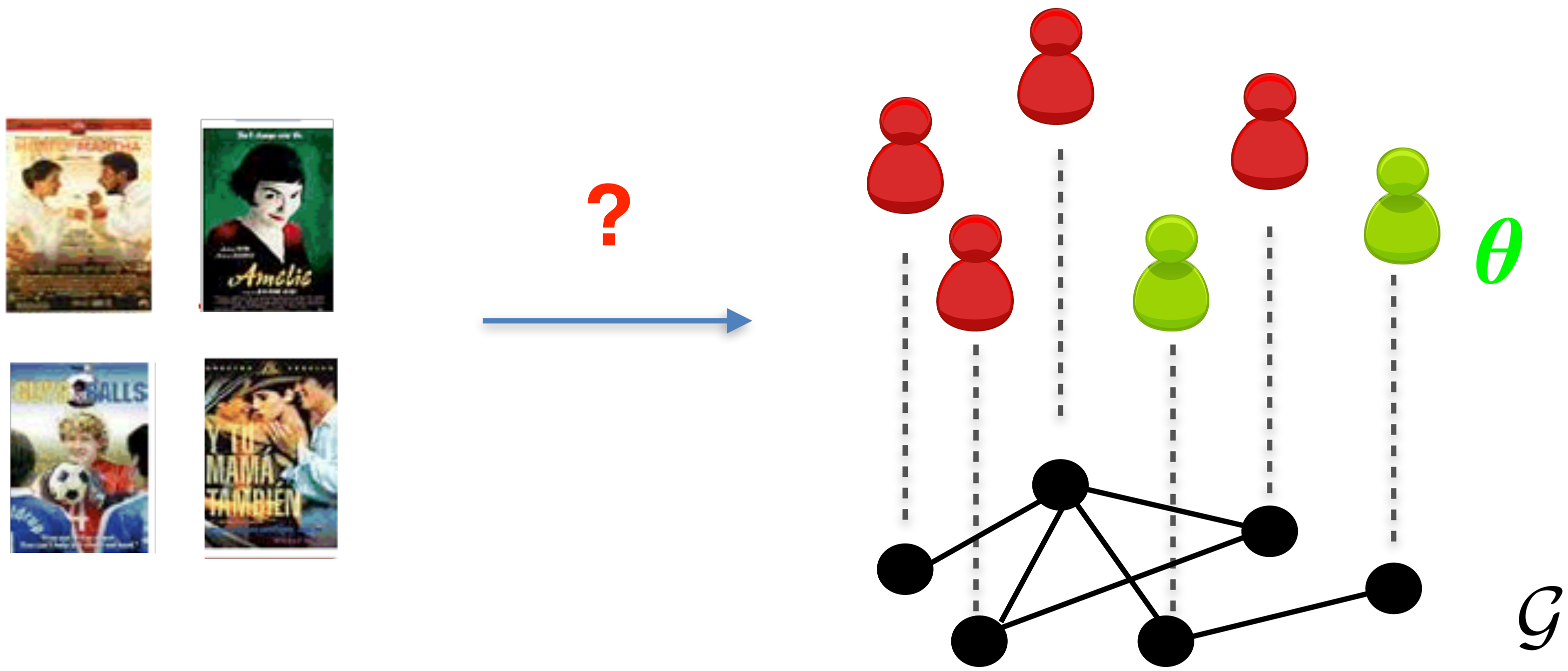
Today's talk

Settings



- centralized agent
- m arms and n users
- users appearing uniformly at random
- At round t, user i_t appears, and an agent
  - chooses an arm a_t
  - receives a reward $y_t = \boldsymbol{x}_{a_t}^T \boldsymbol{\theta}_{i_t} + \eta_t$

Settings

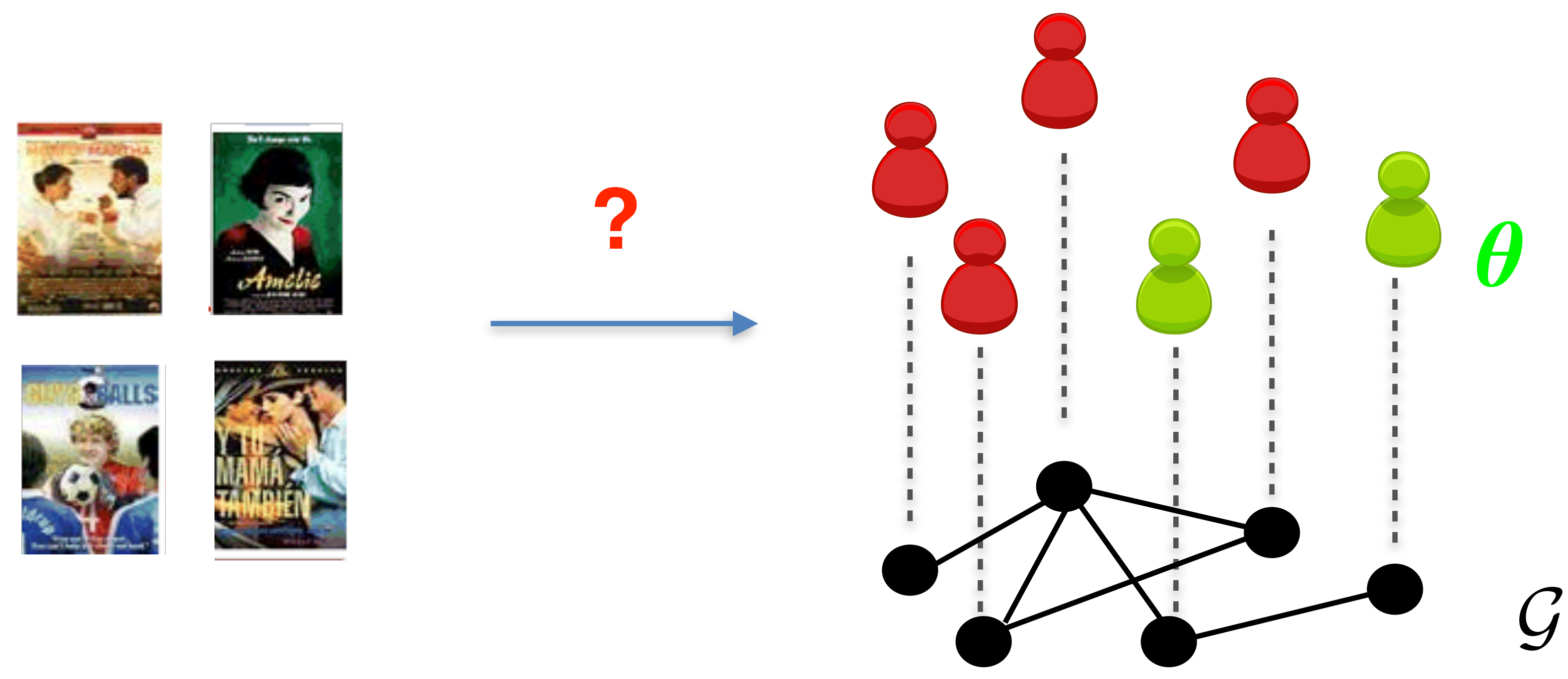


- centralized agent
- m arms and n users
- users appearing uniformly at random
- At round t, user i_t appears, and an agent
  - chooses an arm a_t
  - receives a reward $y_t = \boldsymbol{x}_{a_t}^T \boldsymbol{\theta}_{i_t} + \eta_t$
- Sequential sampling strategy (**bandit algorithm**)

$$a_{t+1} = F_t(i_1, a_1, y_1, \ldots, i_t, a_t, y_t \mid i_{t+1})$$

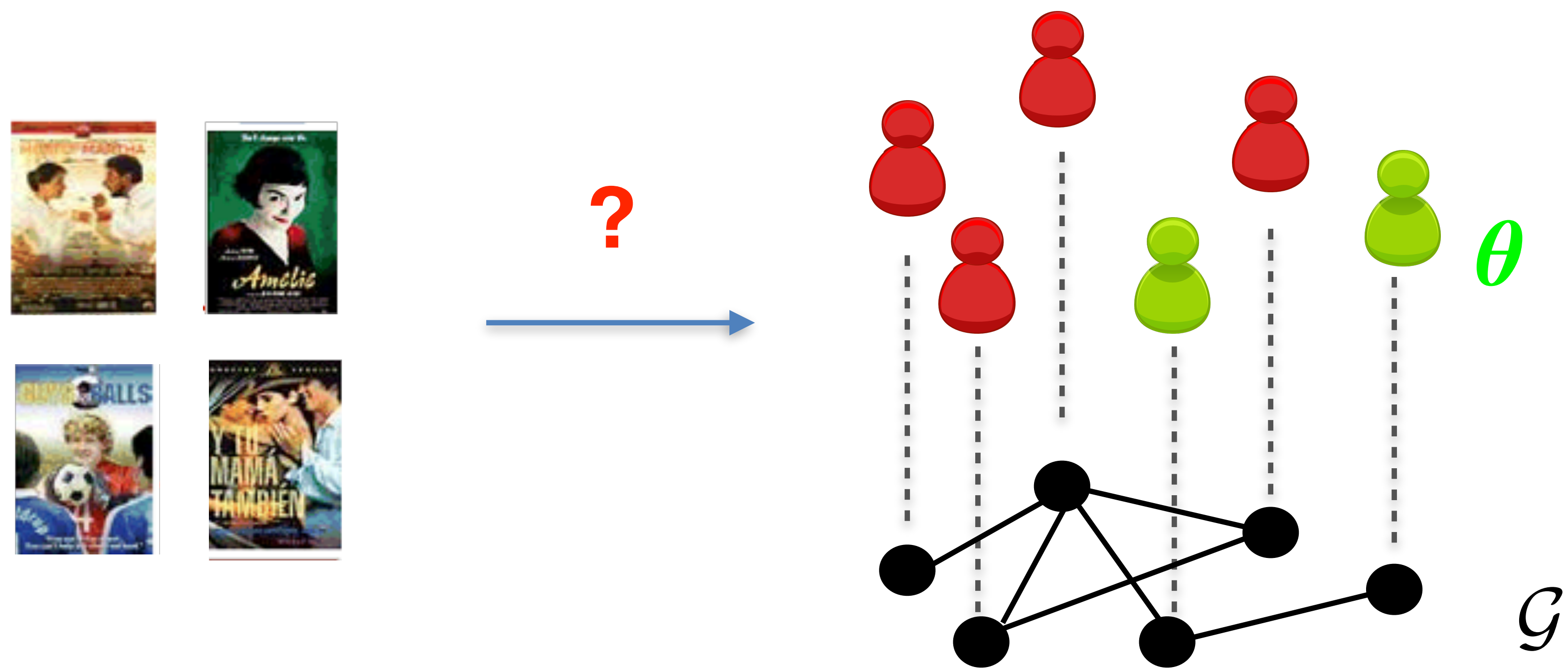
- Goal: Maximize sum of rewards $\mathbb{E}\left[\sum_{t=1}^{T} y_t\right]$

Settings



- $\mathcal{G} = (V, E, \boldsymbol{W})$: undirected-weighted graph
- $W_{i,j} = W_{j,i}$: captures similarity between users i and j (i.e,$\boldsymbol{\theta}_{i,j} = \boldsymbol{\theta}_{j,i}$)
- $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$:combinatorial Laplacian of $\mathcal{G}$

Settings



- $\mathcal{G} = (V, E, \boldsymbol{W})$: undirected-weighted graph
- $W_{i,j} = W_{j,i}$: captures similarity between users i and j (i.e,$\boldsymbol{\theta}_{i,j} = \boldsymbol{\theta}_{j,i}$)
- $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$:combinatorial Laplacian of $\mathcal{G}$

Similarity captured in the latent space

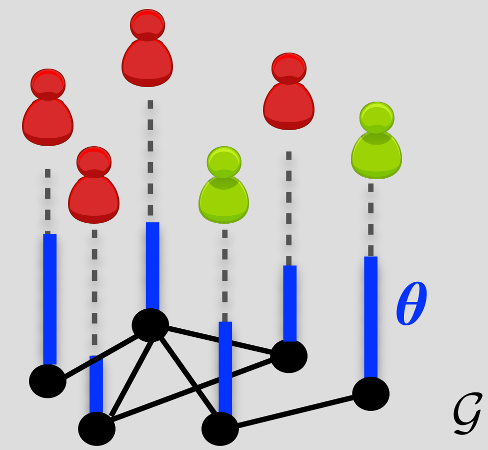
Assumptions

- User preferences mapped into a graph of similarities

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n]^T \in \mathbb{R}^{n \times d}$$: signal on graph

- Exploitation of smoothness prior

$$tr(\boldsymbol{\Theta}^T \mathscr{L} \boldsymbol{\Theta}) = \frac{1}{4} \sum_{k=1}^{d} \sum_{i \sim j} \left(\frac{W_{ij}}{D_{ii}} + \frac{W_{ji}}{D_{jj}} \right) \left(\Theta_{ik} - \Theta_{jk} \right)^2$$

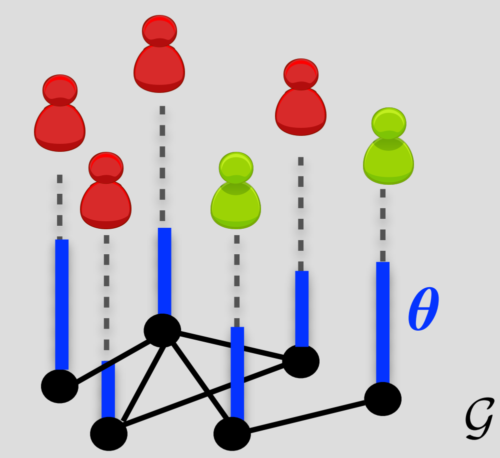


Assumptions

- User preferences mapped into a graph of similarities

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n]^T \in \mathbb{R}^{n \times d}$$: signal on graph

- Exploitation of smoothness prior



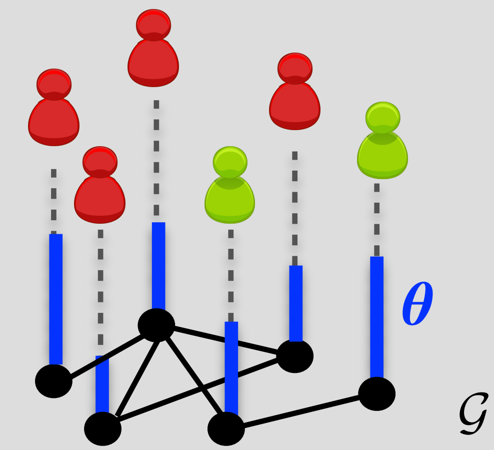
smoothness measure

$$\boxed{tr(\boldsymbol{\Theta}^T \mathscr{L} \boldsymbol{\Theta})} = \frac{1}{4} \sum_{k=1}^{d} \sum_{i \sim j} \left(\frac{W_{ij}}{D_{ii}} + \frac{W_{ji}}{D_{jj}} \right) (\Theta_{ik} - \Theta_{jk})^2$$

Assumptions

- User preferences mapped into a graph of similarities

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n]^T \in \mathbb{R}^{n \times d}$$: signal on graph



- Exploitation of smoothness prior

smoothness measure

$$\boxed{tr(\boldsymbol{\Theta}^T \mathscr{L} \boldsymbol{\Theta})} = \frac{1}{4} \sum_{k=1}^{d} \sum_{i \sim j} \left(\frac{W_{ij}}{D_{ii}} + \frac{W_{ji}}{D_{jj}} \right) (\Theta_{ik} - \Theta_{jk})^2$$

- Smoothness of $\boldsymbol{\Theta}$ over graph $\mathscr{G}$ can be quantified using the Laplacian quadratic form

- We express smoothness as a function of the **random-walk Laplacian**

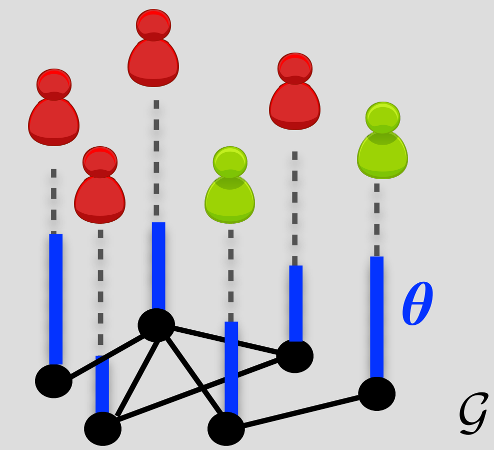
$$\mathscr{L} = \boldsymbol{D}^{-1} \boldsymbol{L}$$ with $\mathscr{L}_{ii} = 1$ and $\sum_{j \neq i} \mathscr{L}_{ji} = -1$

Assumptions

- User preferences mapped into a graph of similarities

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n]^T \in \mathbb{R}^{n \times d}$$: signal on graph

- Exploitation of smoothness prior



smoothness measure $\boxed{tr(\boldsymbol{\Theta}^T \mathscr{L} \boldsymbol{\Theta})} = \frac{1}{4} \sum_{k=1}^{d} \sum_{i \sim j} \left(\frac{W_{ij}}{D_{ii}} + \frac{W_{ji}}{D_{jj}} \right) (\Theta_{ik} - \Theta_{jk})^2$

- Smoothness of $\boldsymbol{\Theta}$ over graph $\mathscr{G}$ can be quantified using the Laplacian quadratic form

- We express smoothness as a function of the **random-walk Laplacian**

$$\mathscr{L} = \boldsymbol{D}^{-1}\boldsymbol{L} \qquad \text{with } \mathscr{L}_{ii} = 1 \text{ and } \sum_{j \neq i} \mathscr{L}_{ji} = -1$$

- avoiding a regret scaling with D_{ii}
- achieving convexity property needed to bound the estimation error

Problem Formulation

Given

- the users graph $\mathscr{G}$
- arm feature vector $\boldsymbol{x}_a, a \in \{1,2,\ldots,m\}$
- no information about the user $\boldsymbol{\theta}_i, i \in \{1,2,\ldots,n\}$?

The agent seeks the optimal selection strategy that minimizes the cumulative (pseudo) regret

$$R_T = \sum_{t=1}^{T} \left((\mathbf{x}_t^*)^T \boldsymbol{\theta}_{i_t} - \mathbf{x}_t^T \boldsymbol{\theta}_{i_t} \right)$$

Problem Formulation

Given

- the users graph $\mathcal{G}$
- arm feature vector $\boldsymbol{x}_a, a \in \{1,2,\ldots,m\}$
- no information about the user $\boldsymbol{\theta}_i, i \in \{1,2,\ldots,n\}$?

The agent seeks the optimal selection strategy that minimizes the cumulative (pseudo) regret

$$R_T = \sum_{t=1}^{T} \left((\mathbf{x}_t^*)^T \boldsymbol{\theta}_{i_t} - \mathbf{x}_t^T \boldsymbol{\theta}_{i_t} \right)$$

Under smoothness prior, the users parameter vector is estimated as

$$\hat{\boldsymbol{\Theta}}_t = \arg \min_{\boldsymbol{\Theta} \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \sum_{\tau \in t_i} (\mathbf{x}_\tau^T \boldsymbol{\theta}_i - y_{i,\tau})^2 + \alpha \ tr(\boldsymbol{\Theta}^T \mathcal{L} \boldsymbol{\Theta})$$

Problem Formulation

Given

- the users graph $\mathcal{G}$
- arm feature vector $\boldsymbol{x}_a, a \in \{1,2,\ldots,m\}$
- no information about the user $\boldsymbol{\theta}_i, i \in \{1,2,\ldots,n\}$?

The agent seeks the optimal selection strategy that minimizes the cumulative (pseudo) regret

$$R_T = \sum_{t=1}^{T} \left((\mathbf{x}_t^*)^T \boldsymbol{\theta}_{i_t} - \mathbf{x}_t^T \boldsymbol{\theta}_{i_t} \right)$$

Under smoothness prior, the users parameter vector is estimated as

$$\hat{\boldsymbol{\Theta}}_t = \arg \min_{\boldsymbol{\Theta} \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \sum_{\tau \in t_i} \underbrace{(\mathbf{x}_\tau^T \boldsymbol{\theta}_i - y_{i,\tau})^2}_{\text{fidelity term}} + \alpha \; tr(\boldsymbol{\Theta}^T \mathcal{L} \boldsymbol{\Theta})$$

Problem Formulation

Given

- the users graph $\mathcal{G}$
- arm feature vector $\boldsymbol{x}_a, a \in \{1,2,\ldots,m\}$
- no information about the user $\boldsymbol{\theta}_i, i \in \{1,2,\ldots,n\}$?

The agent seeks the optimal selection strategy that minimizes the cumulative (pseudo) regret

$$R_T = \sum_{t=1}^{T} \left((\mathbf{x}_t^*)^T \boldsymbol{\theta}_{i_t} - \mathbf{x}_t^T \boldsymbol{\theta}_{i_t} \right)$$

Under smoothness prior, the users parameter vector is estimated as

$$\hat{\boldsymbol{\Theta}}_t = \arg \min_{\boldsymbol{\Theta} \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \sum_{\tau \in t_i} \underbrace{(\mathbf{x}_\tau^T \boldsymbol{\theta}_i - y_{i,\tau})^2}_{\text{fidelity term}} + \alpha \underbrace{tr(\boldsymbol{\Theta}^T \mathcal{L} \boldsymbol{\Theta})}_{\text{smoothness regularizer}}$$

Problem Formulation

Given

- the users graph $\mathcal{G}$
- arm feature vector $\boldsymbol{x}_a, a \in \{1,2,\ldots,m\}$
- no information about the user $\boldsymbol{\theta}_i, i \in \{1,2,\ldots,n\}$?

The agent seeks the optimal selection strategy that minimizes the cumulative (pseudo) regret

$$R_T = \sum_{t=1}^{T} \left((\mathbf{x}_t^*)^T \boldsymbol{\theta}_{i_t} - \mathbf{x}_t^T \boldsymbol{\theta}_{i_t} \right)$$

Under smoothness prior, the users parameter vector is estimated as

$$\hat{\boldsymbol{\Theta}}_t = \arg \min_{\boldsymbol{\Theta} \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \sum_{\tau \in t_i} (\mathbf{x}_\tau^T \boldsymbol{\theta}_i - y_{i,\tau})^2 + \alpha \, tr(\boldsymbol{\Theta}^T \mathcal{L} \boldsymbol{\Theta})$$

fidelity term smoothness regularizer

The agent selects sequential actions as follows

$$\mathbf{x}_{i,t} = \arg \max_{(\mathbf{x},\boldsymbol{\theta}) \in (\mathcal{D}, \mathcal{C}_{i,t})} \mathbf{x}^T \boldsymbol{\theta}$$

Problem Formulation

Given

- the users graph $\mathcal{G}$
- arm feature vector $\boldsymbol{x}_a, a \in \{1,2,\ldots,m\}$
- no information about the user $\boldsymbol{\theta}_i, i \in \{1,2,\ldots,n\}$?

The agent seeks the optimal selection strategy that minimizes the cumulative (pseudo) regret

$$R_T = \sum_{t=1}^{T} \left((\mathbf{x}_t^*)^T \boldsymbol{\theta}_{i_t} - \mathbf{x}_t^T \boldsymbol{\theta}_{i_t} \right)$$

Under smoothness prior, the users parameter vector is estimated as

$$\hat{\boldsymbol{\Theta}}_t = \arg \min_{\boldsymbol{\Theta} \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \sum_{\tau \in t_i} \underbrace{(\mathbf{x}_\tau^T \boldsymbol{\theta}_i - y_{i,\tau})^2}_{\text{fidelity term}} + \alpha \underbrace{tr(\boldsymbol{\Theta}^T \mathcal{L} \boldsymbol{\Theta})}_{\text{smoothness regularizer}}$$

The agent selects sequential actions as follows

$$\boxed{\mathbf{x}_{i,t} = \arg \max_{(\mathbf{x},\boldsymbol{\theta}) \in (\mathcal{D}, \mathcal{C}_{i,t})} \mathbf{x}^T \boldsymbol{\theta}}$$

confident set ?

Problem Formulation

Main Challenges

- smoothness not imposed in the observation domain but in the representation one
- no theoretical error bound for Laplacian regularized estimate
- computational complexity

Main Novelties

- derivation single-user estimation error bound
- proposed single-user UCB in bandit problem
- low-complexity (local) algorithm
- cumulative regret bound as a function of graph properties

Laplacian-regularized Estimator

Closed form solution

$$\hat{\boldsymbol{\Theta}}_t = \arg \min_{\boldsymbol{\Theta} \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \sum_{\tau \in t_i} (\mathbf{x}_\tau^T \boldsymbol{\theta}_i - y_{i,\tau})^2 + \alpha \; tr(\boldsymbol{\Theta}^T \mathscr{L} \boldsymbol{\Theta})$$

$$vec(\hat{\boldsymbol{\Theta}}_t) = (\boldsymbol{\Phi}_t \boldsymbol{\Phi}_t^T + \alpha \mathscr{L} \otimes \mathbf{I})^{-1} \boldsymbol{\Phi}_t \mathbf{Y}_t \qquad \boldsymbol{\Phi}_t = [\phi_1, \phi_2, ..., \phi_t] \in \mathbb{R}^{nd \times t}$$

where $\otimes$ is the Kronecker product, and $vec(\hat{\boldsymbol{\Theta}}_t)$ is a concatenation of column of $\hat{\boldsymbol{\Theta}}_t$

Laplacian-regularized Estimator

Closed form solution

$$\hat{\boldsymbol{\Theta}}_t = \arg\min_{\boldsymbol{\Theta}\in\mathbb{R}^{n\times d}} \sum_{i=1}^{n} \sum_{\tau\in t_i} (\mathbf{x}_\tau^T \boldsymbol{\theta}_i - y_{i,\tau})^2 + \alpha\ tr(\boldsymbol{\Theta}^T \mathscr{L} \boldsymbol{\Theta})$$

$$vec(\hat{\boldsymbol{\Theta}}_t) = (\boldsymbol{\Phi}_t \boldsymbol{\Phi}_t^T + \alpha \mathscr{L} \otimes \mathbf{I})^{-1} \boldsymbol{\Phi}_t \mathbf{Y}_t \qquad \boldsymbol{\Phi}_t = [\phi_1, \phi_2, ..., \phi_t] \in \mathbb{R}^{nd\times t}$$

where $\otimes$ is the Kronecker product, and $vec(\hat{\boldsymbol{\Theta}}_t)$ is a concatenation of column of $\hat{\boldsymbol{\Theta}}_t$

decoupling estimates

Lemma 1. $\hat{\boldsymbol{\Theta}}_t$ *is obtained from Eq. 5, let* $\hat{\boldsymbol{\theta}}_{i,t}$ *be the* i*-th row of* $\hat{\boldsymbol{\Theta}}_t$ *which is the estimate of* $\boldsymbol{\theta}_i$. $\hat{\boldsymbol{\theta}}_{i,t}$ *can be approximated by :*

$$\hat{\boldsymbol{\theta}}_{i,t} \approx \mathbf{A}_{i,t}^{-1} \mathbf{X}_{i,t} \mathbf{Y}_{i,t} - \alpha \mathbf{A}_{i,t}^{-1} \sum_{j=1}^{n} \mathcal{L}_{ij} \mathbf{A}_{j,t}^{-1} \mathbf{X}_{j,t} \mathbf{Y}_{j,t} \quad (7)$$

where $\mathbf{A}_{i,t} = \sum_{\tau\in t_i} \mathbf{x}_\tau \mathbf{x}_\tau^T \in \mathbb{R}^{d\times d}$ *is the Gram matrix of user* i, $\mathcal{L}_{ij}$ *is the* (i,j)*-th element in* $\mathcal{L}$, $\mathbf{Y}_{i,t} = [y_{i,1}, ..., y_{i,t_i}]$ *are the collection of payoffs associated with user* i *up to time* t.

Laplacian-regularized Estimator

$$\mathcal{C}_{i,t} = \{\boldsymbol{\theta}_{i,t} : ||\hat{\boldsymbol{\theta}}_{i,t} - \boldsymbol{\theta}_{i,t}||_{\boldsymbol{\Lambda}_{i,t}} \leq \beta_{i,t}\} \tag{8}$$

precision matrix of $vec(\hat{\boldsymbol{\Theta}}_t)$

$$\boldsymbol{\Lambda}_{i,t} = \mathbf{A}_{i,t} + 2\alpha\mathcal{L}_{ii}\mathbf{I} + \alpha^2 \sum_{j=1}^{n} \mathcal{L}_{ij}^2 \mathbf{A}_{j,t}^{-1} \tag{10}$$

Laplacian-regularized Estimator

$$\mathcal{C}_{i,t} = \{\boldsymbol{\theta}_{i,t} : ||\hat{\boldsymbol{\theta}}_{i,t} - \boldsymbol{\theta}_{i,t}||_{\mathbf{\Lambda}_{i,t}} \leq \beta_{i,t}\} \tag{8}$$

$$\mathbf{\Lambda}_{i,t} = \mathbf{A}_{i,t} + 2\alpha\mathcal{L}_{ii}\mathbf{I} + \alpha^2 \sum_{j=1}^{n} \mathcal{L}_{ij}^2 \mathbf{A}_{j,t}^{-1} \tag{10}$$

Laplacian-regularized Estimator

$$\mathcal{C}_{i,t} = \{\boldsymbol{\theta}_{i,t} : ||\hat{\boldsymbol{\theta}}_{i,t} - \boldsymbol{\theta}_{i,t}||_{\boldsymbol{\Lambda}_{i,t}} \leq \beta_{i,t}\} \qquad (8)$$

$$\boldsymbol{\Lambda}_{i,t} = \mathbf{A}_{i,t} + 2\alpha\mathcal{L}_{ii}\mathbf{I} + \alpha^2 \sum_{j=1}^{n} \mathcal{L}_{ij}^2 \mathbf{A}_{j,t}^{-1} \qquad (10)$$

error bound

Lemma 2. *t_i is the set of time at which user i is served up to time t. $\mathbf{A}_{i,t} = \sum_{\tau \in t_i} \mathbf{x}_\tau \mathbf{x}_\tau^T$, $\mathbf{V}_{i,t} = \mathbf{A}_{i,t} + \alpha\mathcal{L}_{ii}\mathbf{I}$, $\boldsymbol{\xi}_{i,t} = \sum_{\tau \in t_i} \mathbf{x}_{i,\tau}\eta_{i,\tau}$, $\mathbf{I} \in \mathbb{R}^{d \times d}$ is the identity matrix. $\boldsymbol{\Lambda}_{i,t}$ is defined in Eq. 10. Denote $\boldsymbol{\Delta}_i = \sum_{j=1}^{n} \mathcal{L}_{ij}\boldsymbol{\theta}_j$, the size of the confidence set defined in Eq. 8 satisfies the following upper bound with probability $1 - \delta$ with $\delta \in [0, 1]$.*

$$\beta_{i,t} = \sigma \sqrt{2 \log \frac{|\mathbf{V}_{i,t}|^{1/2}}{\delta |\alpha \mathbf{I}|^{1/2}}} + \sqrt{\alpha} ||\boldsymbol{\Delta}_i||_2$$

Laplacian-regularized Estimator

$$\mathcal{C}_{i,t} = \{\boldsymbol{\theta}_{i,t} : ||\hat{\boldsymbol{\theta}}_{i,t} - \boldsymbol{\theta}_{i,t}||_{\boldsymbol{\Lambda}_{i,t}} \leq \beta_{i,t}\} \quad (8)$$

$$\boldsymbol{\Lambda}_{i,t} = \mathbf{A}_{i,t} + 2\alpha\mathcal{L}_{ii}\mathbf{I} + \alpha^2 \sum_{j=1}^{n} \mathcal{L}_{ij}^2 \mathbf{A}_{j,t}^{-1} \quad (10)$$

error bound

Lemma 2. *t_i is the set of time at which user i is served up to time t. $\mathbf{A}_{i,t} = \sum_{\tau \in t_i} \mathbf{x}_\tau \mathbf{x}_\tau^T$, $\mathbf{V}_{i,t} = \mathbf{A}_{i,t} + \alpha\mathcal{L}_{ii}\mathbf{I}$, $\boldsymbol{\xi}_{i,t} = \sum_{\tau \in t_i} \mathbf{x}_{i,\tau}\eta_{i,\tau}$, $\mathbf{I} \in \mathbb{R}^{d \times d}$ is the identity matrix. $\boldsymbol{\Lambda}_{i,t}$ is defined in Eq. 10. Denote $\boldsymbol{\Delta}_i = \sum_{j=1}^{n} \mathcal{L}_{ij}\boldsymbol{\theta}_j$, the size of the confidence set defined in Eq. 8 satisfies the following upper bound with probability $1-\delta$ with $\delta \in [0,1]$.*

$$\beta_{i,t} = \sigma\sqrt{2\log\frac{|\mathbf{V}_{i,t}|^{1/2}}{\delta|\alpha\mathbf{I}|^{1/2}}} + \sqrt{\alpha}\boxed{||\boldsymbol{\Delta}_i||_2}$$

$$\boxed{\boldsymbol{\Delta}_i = \sum_{i=1}^{n} \mathscr{L}_{ij}\boldsymbol{\theta}_j = \boldsymbol{\theta}_i - \sum_{j \neq i} (-\mathscr{L}_{ij}\boldsymbol{\theta}_j)}$$

graph information

Laplacian-regularized Estimator

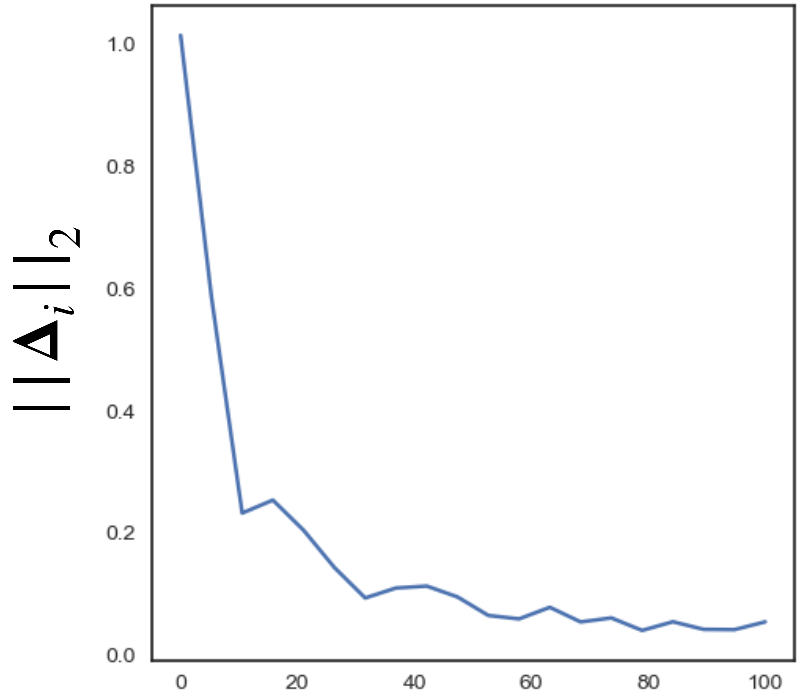
$$\mathcal{C}_{i,t} = \{\boldsymbol{\theta}_{i,t} : ||\hat{\boldsymbol{\theta}}_{i,t} - \boldsymbol{\theta}_{i,t}||_{\boldsymbol{\Lambda}_{i,t}} \leq \beta_{i,t}\} \qquad (8)$$

$$\boldsymbol{\Lambda}_{i,t} = \mathbf{A}_{i,t} + 2\alpha\mathcal{L}_{ii}\mathbf{I} + \alpha^2 \sum_{j=1}^{n} \mathcal{L}_{ij}^2 \mathbf{A}_{j,t}^{-1} \qquad (10)$$

error bound

Lemma 2. *t_i is the set of time at which user i is served up to time t. $\mathbf{A}_{i,t} = \sum_{\tau \in t_i} \mathbf{x}_\tau \mathbf{x}_\tau^T$, $\mathbf{V}_{i,t} = \mathbf{A}_{i,t} + \alpha\mathcal{L}_{ii}\mathbf{I}$, $\boldsymbol{\xi}_{i,t} = \sum_{\tau \in t_i} \mathbf{x}_{i,\tau}\eta_{i,\tau}$, $\mathbf{I} \in \mathbb{R}^{d\times d}$ is the identity matrix. $\boldsymbol{\Lambda}_{i,t}$ is defined in Eq. 10. Denote $\boldsymbol{\Delta}_i = \sum_{j=1}^{n} \mathcal{L}_{ij}\boldsymbol{\theta}_j$, the size of the confidence set defined in Eq. 8 satisfies the following upper bound with probability $1-\delta$ with $\delta \in [0,1]$.*

$$\beta_{i,t} = \sigma\sqrt{2\log\frac{|\mathbf{V}_{i,t}|^{1/2}}{\delta|\alpha\mathbf{I}|^{1/2}}} + \sqrt{\alpha}\boxed{||\boldsymbol{\Delta}_i||_2}$$



smoothness

$||\boldsymbol{\Delta}_i||_2 \in [0, ||\boldsymbol{\theta}_i||_2]$

$$\boxed{\boldsymbol{\Delta}_i = \sum_{i=1}^{n} \mathcal{L}_{ij}\boldsymbol{\theta}_j = \boldsymbol{\theta}_i - \sum_{j\neq i}(-\mathcal{L}_{ij}\boldsymbol{\theta}_j)}$$

graph information

GraphUCB

Algorithm 1: GraphUCB

Input : α, T, $\boldsymbol{\mathcal{L}}$, δ

Initialization : For any $i \in \{1, 2, ..., n\}$

$\hat{\boldsymbol{\theta}}_{0,i} = \mathbf{0} \in \mathbb{R}^d$, $\boldsymbol{\Lambda}_{0,i} = \mathbf{0} \in \mathbb{R}^{d \times d}$,

$\mathbf{A}_{0,i} = \mathbf{0} \in \mathbb{R}^{d \times d}$, $\beta_{i,t} = 0$.

for $t \in [1, T]$ **do**

User index i_t is selected

1. $\mathbf{A}_{i,t} \leftarrow \mathbf{A}_{i,t-1} + \mathbf{x}_{i,t-1}\mathbf{x}_{i,t-1}^T$ if $i = i_t$.
2. $\mathbf{A}_{j,t} \leftarrow \mathbf{A}_{j,t-1}$, $\forall j \neq i_t$.
3. Update $\boldsymbol{\Lambda}_{i,t}$
4. Select $\mathbf{x}_{i,t}$ $\arg\max_{\mathbf{x} \in \mathcal{D}} \mathbf{x}^T \hat{\boldsymbol{\theta}}_{i,t} + \beta_{i,t} ||\mathbf{x}||_{\boldsymbol{\Lambda}_{i,t}^{-1}}$
5. Receive the payoff $y_{i,t}$.
6. Update $\hat{\boldsymbol{\Theta}}_t$

end

Analysis

Lemma 3. *Define* $\Psi_{i,t_i} = \frac{\sum_{t=1}^{t_i} ||\mathbf{x}_{i,t}||^2_{\mathbf{\Lambda}^{-1}_{i,t}}}{\sum_{t=1}^{t_i} ||\mathbf{x}_{i,t}||^2_{\mathbf{V}^{-1}_{i,t}}}$, *where* $\mathbf{V}_{i,t_i} = \mathbf{A}_{i,t_i} + \alpha\mathcal{L}_{ii}\mathbf{I}$ *and* $\mathbf{\Lambda}_{i,t_i}$ *defined*[1] *in Eq. 10. Without loss of generality, assume* $||\mathbf{x}_{i,t}||_2 \leq 1$ *for any* t, t_i *and* i, *then*

$$\Psi_{i,t_i} \in (0,1] \tag{14}$$

Furthermore, denser connected graph leads to smaller Ψ_{i,t_i}. *Empirical evidence is provided in Fig. 1-b.*

$$\mathbf{\Lambda}_{i,t} = \mathbf{A}_{i,t} + 2\alpha\mathcal{L}_{ii}\mathbf{I} + \alpha^2 \sum_{j=1}^{n} \mathcal{L}^2_{ij}\mathbf{A}^{-1}_{j,t} \tag{10}$$

Analysis

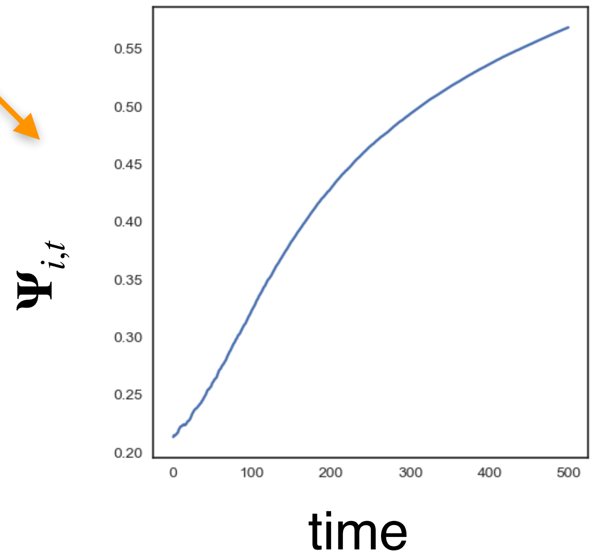
Lemma 3. *Define* $\Psi_{i,t_i} = \frac{\sum_{t=1}^{t_i} ||\mathbf{x}_{i,t}||^2_{\mathbf{\Lambda}^{-1}_{i,t}}}{\sum_{t=1}^{t_i} ||\mathbf{x}_{i,t}||^2_{\mathbf{V}^{-1}_{i,t}}}$, *where* $\mathbf{V}_{i,t_i} = \mathbf{A}_{i,t_i} + \alpha\mathcal{L}_{ii}\mathbf{I}$ *and* $\mathbf{\Lambda}_{i,t_i}$ *defined*[1] *in Eq. 10. Without loss of generality, assume* $||\mathbf{x}_{i,t}||_2 \leq 1$ *for any* t, t_i *and* i, *then*

$$\Psi_{i,t_i} \in (0, 1] \tag{14}$$

Furthermore, denser connected graph leads to smaller Ψ_{i,t_i}. *Empirical evidence is provided in Fig. 1-b.*

$$\mathbf{\Lambda}_{i,t} = \mathbf{A}_{i,t} + 2\alpha\mathcal{L}_{ii}\mathbf{I} + \alpha^2 \sum_{j=1}^{n} \mathcal{L}^2_{ij}\mathbf{A}^{-1}_{j,t} \tag{10}$$

it provides a comparison with no-graph UCB



Regret Analysis

Single User Regret

The cumulative regret over t_i of user i satisfies the following upper bound with probability $1 - \delta$

$$\mathcal{O}\left(\left(\sqrt{d \log(t_i)} + \sqrt{\alpha}\,||\boldsymbol{\Delta}_i||_2\right) \Psi_{i,t_i} \sqrt{d t_i \log(t_i)} \right) = \mathcal{O}\left(d \sqrt{t_i} \Psi_{i,t_i} \right)$$

Network Regret

Assuming users are served uniformly, then, over the time horizon T, the total cumulative regret $R_T = \sum_{i=1}^{n} R_{i,t_i}$ experienced by all users satisfies the following upper bound with probability $1 - \delta$

$$\mathcal{O}\left(d \sqrt{Tn} \max_i \Psi_{i,t_i} \right)$$

Single User Comparison

Single user

LinUCB $\mathcal{O}\left(\left(\sqrt{d\log(t_i)} + \sqrt{\alpha}\,||\boldsymbol{\theta}_i||_2\right)\sqrt{dt_i\log(t_i)} \right)$

GraphUCB $\mathcal{O}\left(\left(\sqrt{d\log(t_i)} + \sqrt{\alpha}\,||\boldsymbol{\Delta}_i||_2\right)\Psi_{i,t_i}\sqrt{dt_i\log(t_i)} \right)$

$||\boldsymbol{\Delta}_i||_2 \in [0, ||\boldsymbol{\theta}_i||_2]$ $\Psi_{i,t_i} \in [0,1]$

- Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). "A contextual-bandit approach to personalized news article recommendation", In Proceedings of the 19th international conference on World wide web, pages 661–670.
- Cesa-Bianchi, N., Gentile, C., and Zappella, G. "A gang of bandits", NeurIPS 2013

Single User Comparison

Single user

LinUCB
$$\mathcal{O}\left(\left(\sqrt{d\log(t_i)}+\sqrt{\alpha}\,||\boldsymbol{\theta}_i||_2\right)\sqrt{dt_i\log(t_i)}\right)$$

GraphUCB
$$\mathcal{O}\left(\left(\sqrt{d\log(t_i)}+\sqrt{\alpha}\,\boxed{||\boldsymbol{\Delta}_i||_2\right)\Psi_{i,t_i}}\sqrt{dt_i\log(t_i)}\right)$$

$||\boldsymbol{\Delta}_i||_2 \in [0, ||\boldsymbol{\theta}_i||_2]$

$\Psi_{i,t_i} \in [0,1]$

smoothness and connectivity reduce the regret

- Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). "A contextual-bandit approach to personalized news article recommendation", In Proceedings of the 19th international conference on World wide web, pages 661–670.
- Cesa-Bianchi, N., Gentile, C., and Zappella, G. "A gang of bandits", NeurIPS 2013

Single User Comparison

Single user

LinUCB

$$\mathcal{O}\left(\left(\sqrt{d\log(t_i)}+\sqrt{\alpha}\,||\boldsymbol{\theta}_i||_2\right)\sqrt{dt_i\log(t_i)}\right)$$

GraphUCB

$$\mathcal{O}\left(\left(\sqrt{d\log(t_i)}+\sqrt{\alpha}\,||\boldsymbol{\Delta}_i||_2\right)\Psi_{i,t_i}\sqrt{dt_i\log(t_i)}\right)$$

All users

GOB.Lin

$$\mathcal{O}\left(nd\sqrt{T}\right)$$

GraphUCB

$$\mathcal{O}\left(d\sqrt{Tn}\max_i \Psi_{i,t_i}\right)$$

- Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). "A contextual-bandit approach to personalized news article recommendation", In Proceedings of the 19th international conference on World wide web, pages 661–670.
- Cesa-Bianchi, N., Gentile, C., and Zappella, G. "A gang of bandits", NeurIPS 2013

Single User Comparison

Single user

LinUCB $$\mathcal{O}\left(\left(\sqrt{d \log(t_i)} + \sqrt{\alpha}\,||\boldsymbol{\theta}_i||_2\right)\sqrt{dt_i \log(t_i)} \right)$$

GraphUCB $$\mathcal{O}\left(\left(\sqrt{d \log(t_i)} + \sqrt{\alpha}\,||\boldsymbol{\Delta}_i||_2\right)\Psi_{i,t_i}\sqrt{dt_i \log(t_i)} \right)$$

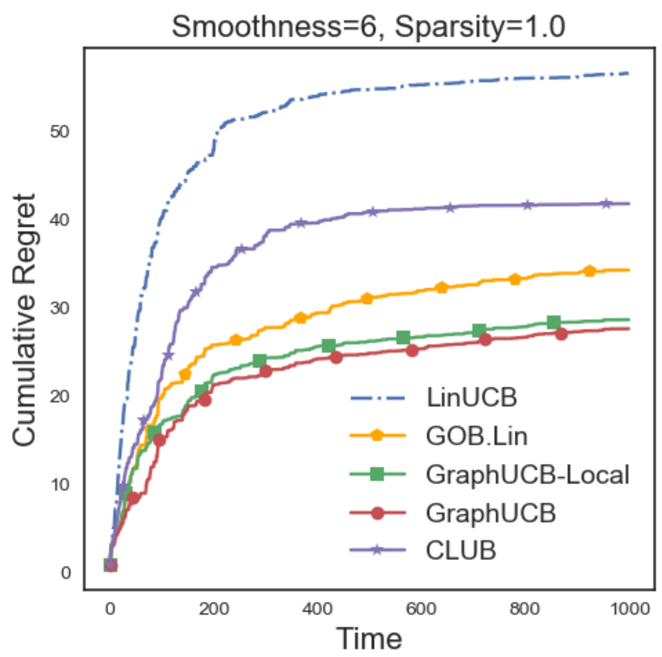
All users

GOB.Lin $$\mathcal{O}\left(nd\sqrt{T} \right)$$

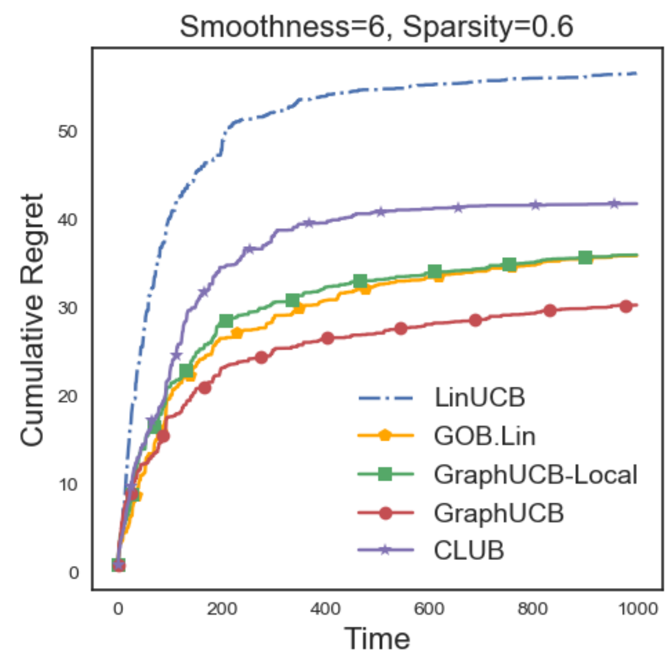
GraphUCB $$\mathcal{O}\left(d\sqrt{Tn}\,\boxed{\max_i \Psi_{i,t_i}} \right)$$

- Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). "A contextual-bandit approach to personalized news article recommendation", In Proceedings of the 19th international conference on World wide web, pages 661–670.
- Cesa-Bianchi, N., Gentile, C., and Zappella, G. "A gang of bandits", NeurIPS 2013

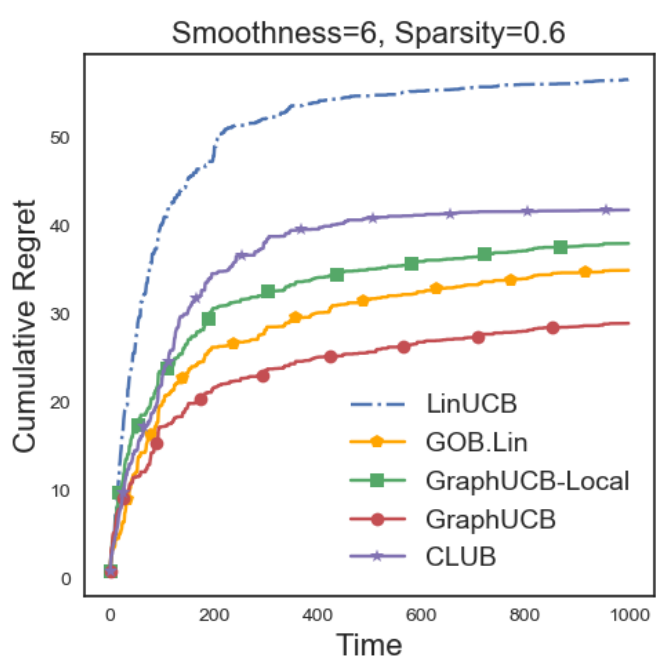
Results - Synthetic



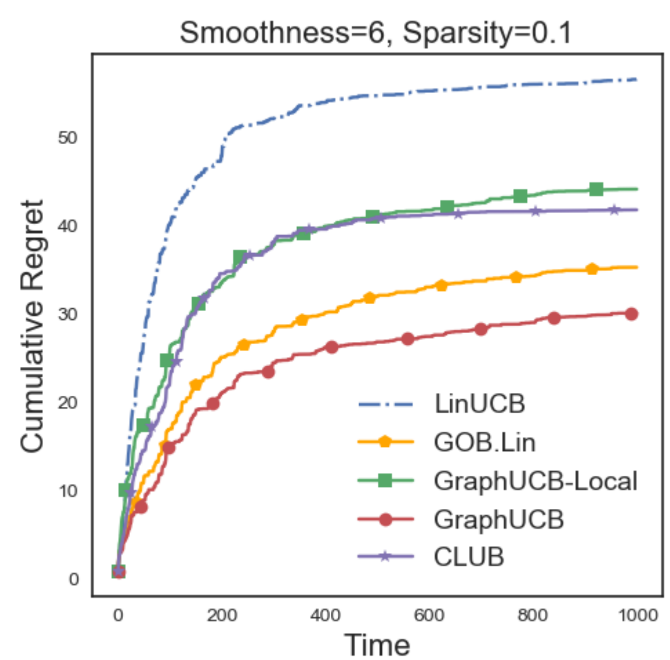
(a) RBF



(b) RBF-Sparse (0.5)

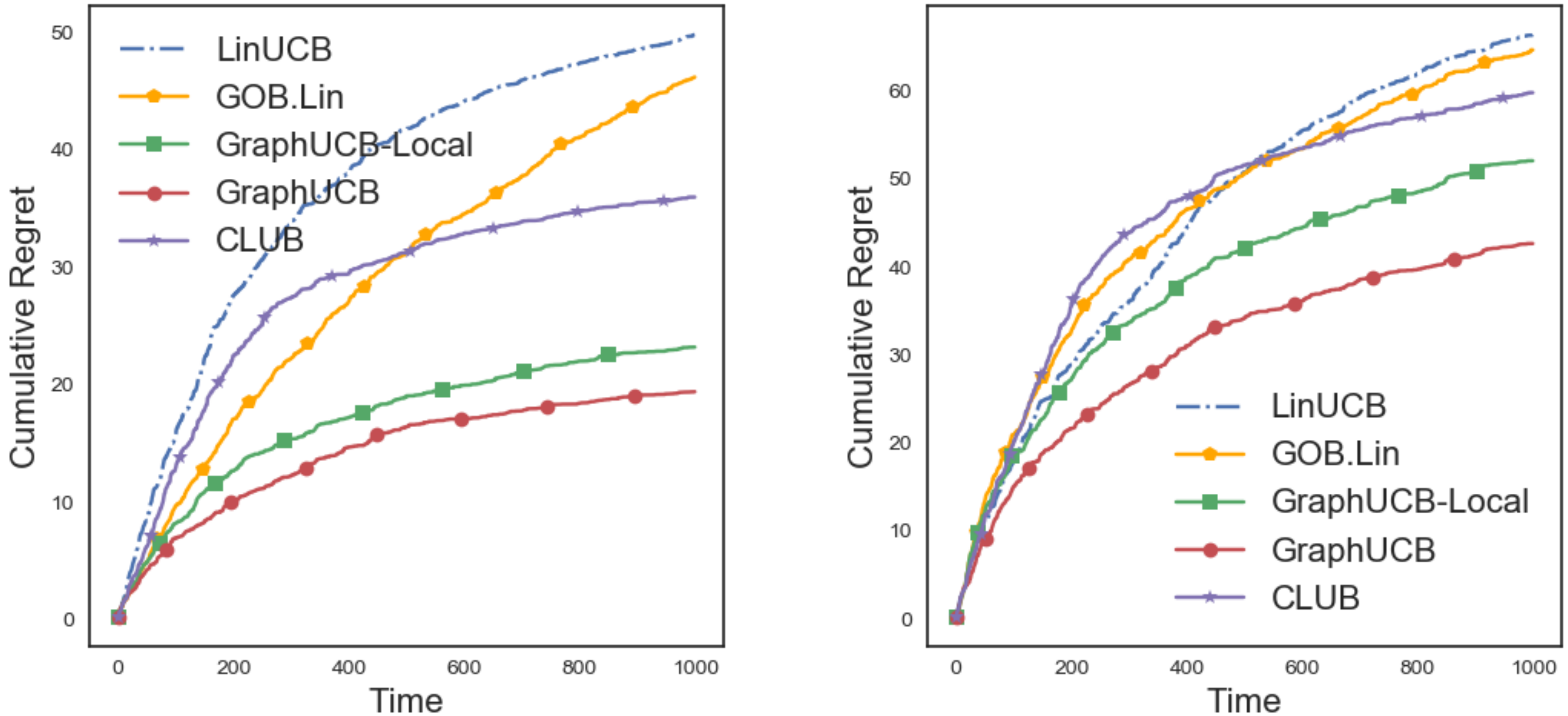


(c) ER (p=0.2)



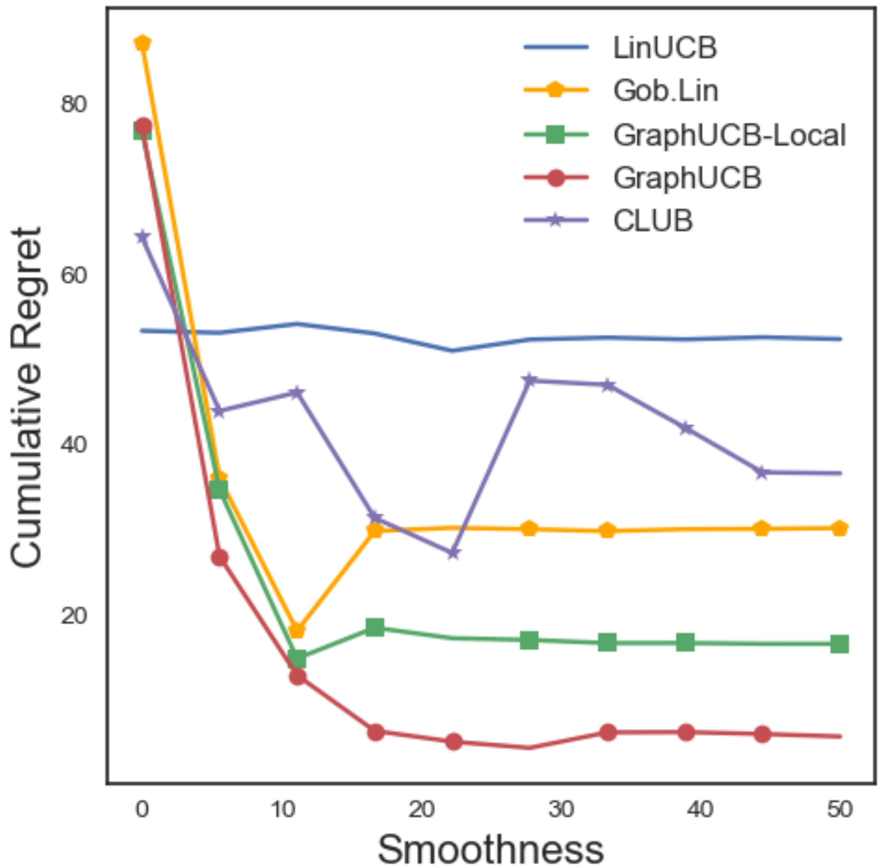
(d) BA (m=1)

Results - Real World Data

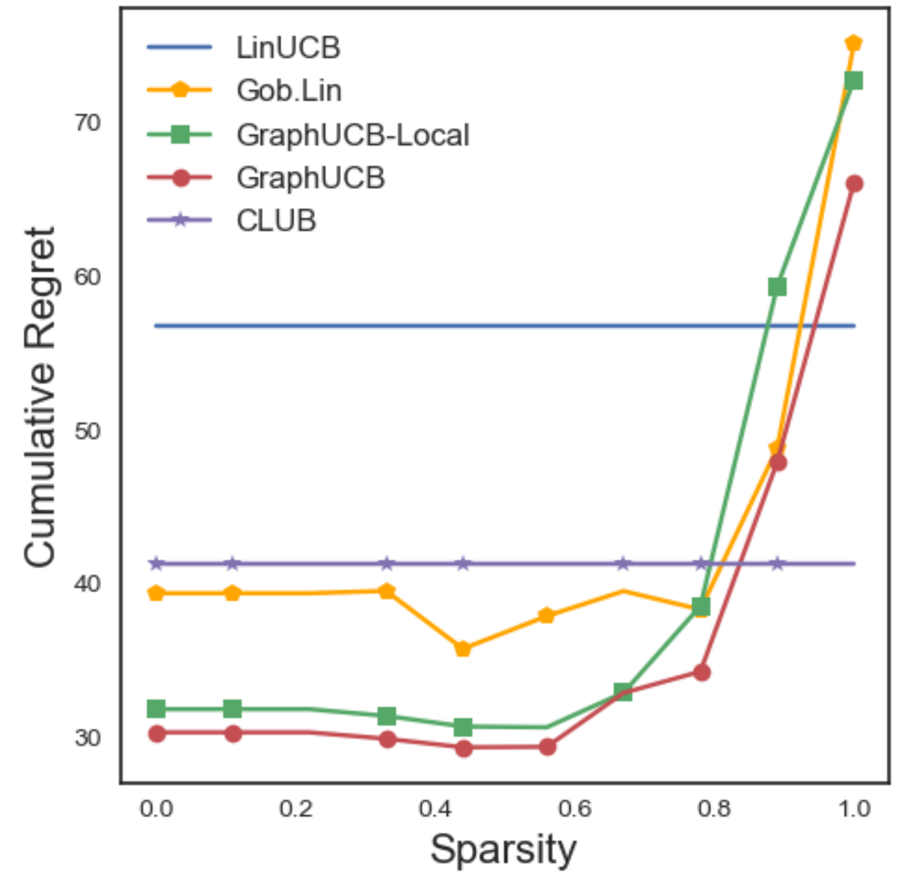


(a) MovieLens (b) Netflix

Results - Graph Features



(a) Smoothness: γ in Eq. 21



(b) RBF (Sparsity)

Conclusions

- Proposed GraphUCB to solve the stochastic linear bandit problem with multiple users - known user graph
- Single-user UCB
- GraphUCB leads to lower cumulative regret as compared to algorithms which ignore user graph
- Proposed local-GraphUCB - need further investigation

Conclusions

- Proposed GraphUCB to solve the stochastic linear bandit problem with multiple users - known user graph

- Single-user UCB

- GraphUCB leads to lower cumulative regret as compared to algorithms which ignore user graph
- Proposed local-GraphUCB - need further investigation

- Next?
  - better understanding of the effect of the graph
  - bandit optimality as function of graph features

  - graph learning and other GSP properties applied to MABs?

References

- Cesa-Bianchi, Nicolò, Tommaso R. Cesari, and Claire Monteleoni. "Cooperative Online Learning: Keeping your Neighbors Updated" *arXiv preprint arXiv:1901.08082* (2019)
- K Yang, X Dong, L Toni, "Laplacian-regularized graph bandits: Algorithms and theoretical analysis", arXiv preprint arXiv:1907.05632, 2019
- K Yang, X Dong, L Toni , "Error Analysis on Graph Laplacian Regularized Estimator", arXiv preprint arXiv:1902.03720, 2019
- Yang, K. and Toni, L., *Graph-based recommendation system*, IEEE GlobalSIP, 2018
- A. Carpentier, and M. Valko, *"Revealing graph bandits for maximizing local influence",* International Conference on Artificial Intelligence and Statistics. 2016
- E. E. Gad, et al. "Active learning on weighted graphs using adaptive and non-adaptive approaches" ICASSP, 2016
- Li, S., Karatzoglou, A., and Gentile, C. *Collaborative filtering bandits,* ACM SIGIR 2016
- Korda, N., Szorenyi, B., and Shuai, L. *Distributed clustering of linear bandits in peer to peer networks, JMLR, 2016*
- Gentile, C., Li, S., and Zappella, G. *Online clustering of bandits, ICML 2014*
- M. Valko et al., "*Spectral Bandits for Smooth Graph Functions*", JMLR 2014
- Q. Gu and J. Han, "*Online spectral learning on a graph with bandit feedback*", in Proc. IEEE Int. Conf. on Data Mining, 2014
- D. Thanou, D. I. Shuman, and P. Frossard. "*Learning parametric dictionaries for signals on graphs*", IEEE Trans. on Signal Processing, 2014
- Cesa-Bianchi, N., Gentile, C., and Zappella, G., *A gang of bandits, NeurIPS* 2013
- W. Chu, L. Li, L. Reyzin, and R. E. Schapire, "*Contextual bandits with linear payoff functions*", in AISTATS, 2011
- Vaswani, S., Schmidt, M., and Lakshmanan, L. V., Horde of bandits using gaussian markov random fields, AISTATS 2017

Thank You! Questions?

Learning and Signal Processing Lab
UCL
https://laspucl2016.com