Learning Parametric Dictionaries for Signals on Graphs

Dorina Thanou

Signal Processing Lab (LTS4)

March 25, 2015

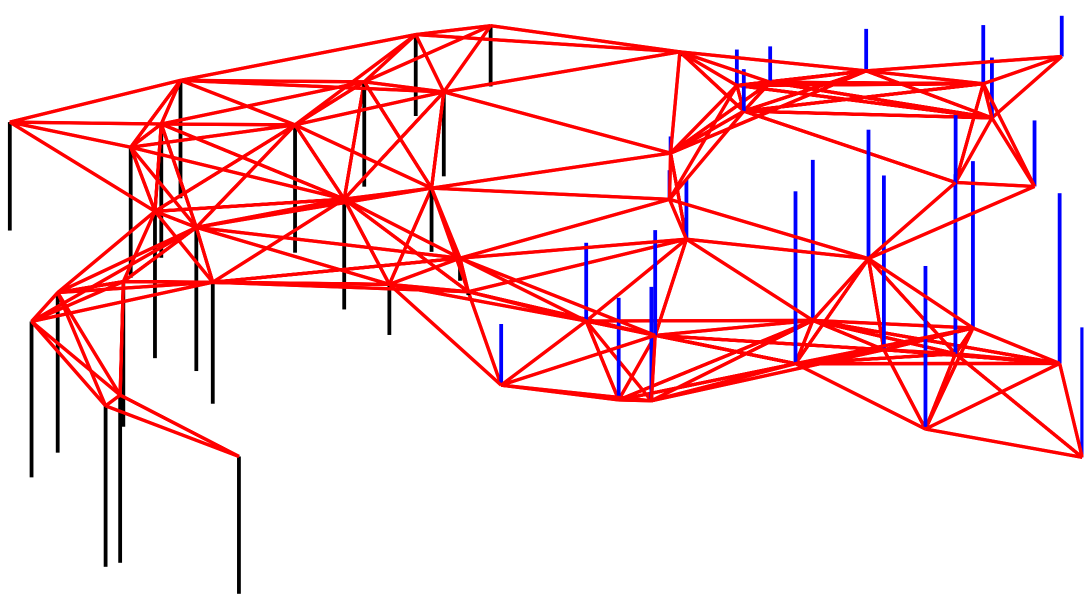
(joint work with D. I Shuman and P. Frossard)

EPFL – Signal Processing Laboratory (LTS4)
http://lts4.epfl.ch

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Signals on Graphs

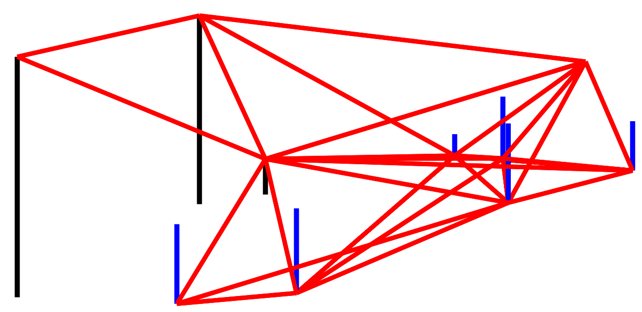
- **Graphs:** flexible tools to represent the geometric structure of signals defined on irregular domains



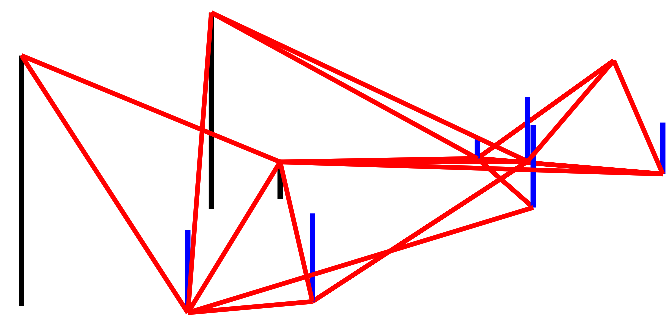
- **Vertices:** discrete data domain
- **Edge weights:** pairwise relationships between vertices
- **Graph signal:** function that assigns a real value to each vertex

Interplay Between Topology and Signals

- The dependencies that arise from the connectivity of the graph define the graph signals



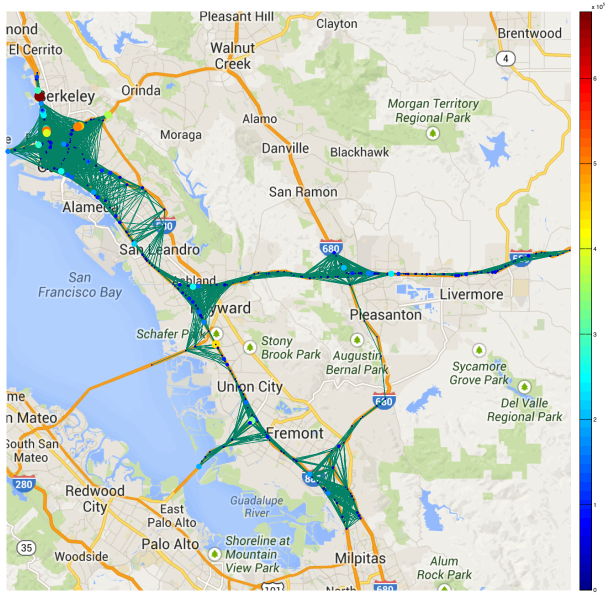
$\mathcal{G}_1$



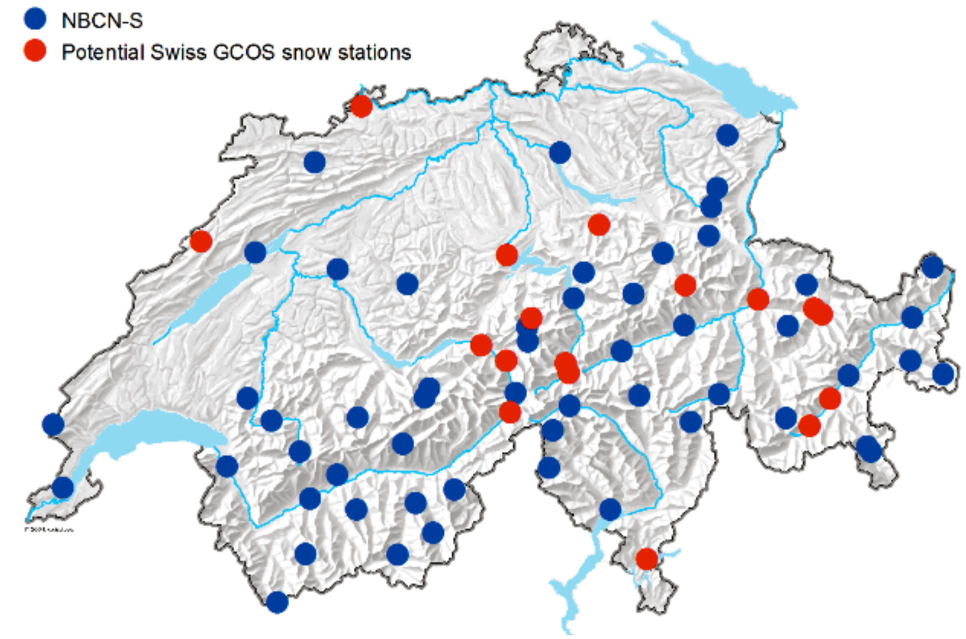
$\mathcal{G}_2$

- The same signal is smoother with respect to the intrinsic structure of $\mathcal{G}_1$

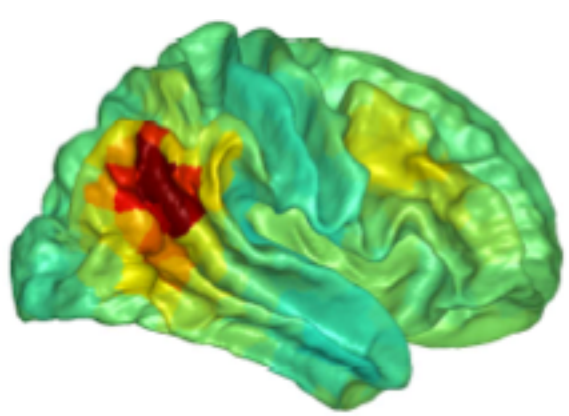
Examples



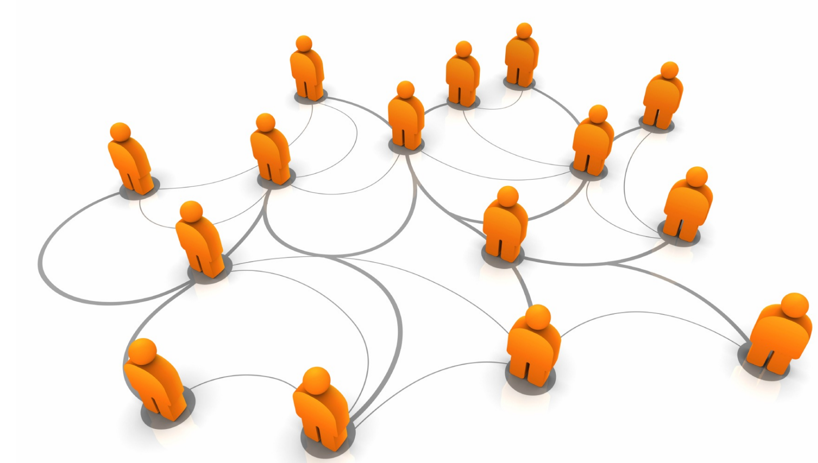
Transportation Networks



Sensor Networks (Source: www.meteoswiss.ch)



Biological Networks



Social Networks

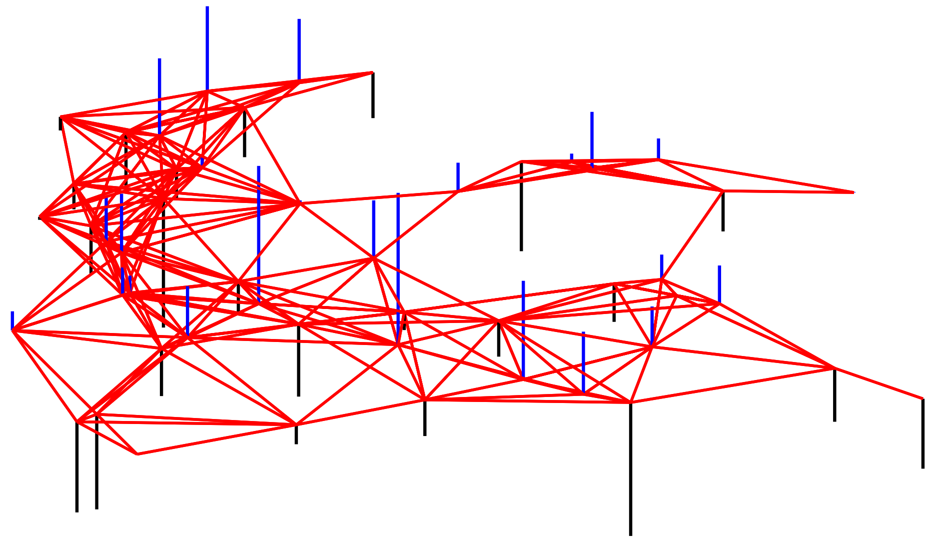
Need for efficient tools to identify and exploit structure in these signals

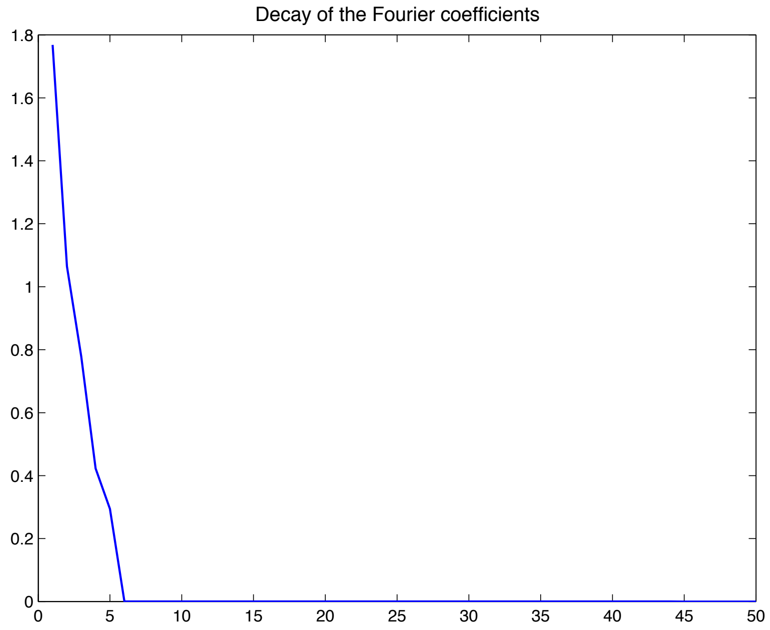
Outline

- Motivation
- Preliminaries on Signal Processing on Graphs
- Parametric Dictionary Learning on Graphs
  - Dictionary Structure
  - Dictionary Learning Algorithm
- Extension to Multiple Graphs
- Conclusion

Processing on Graphs

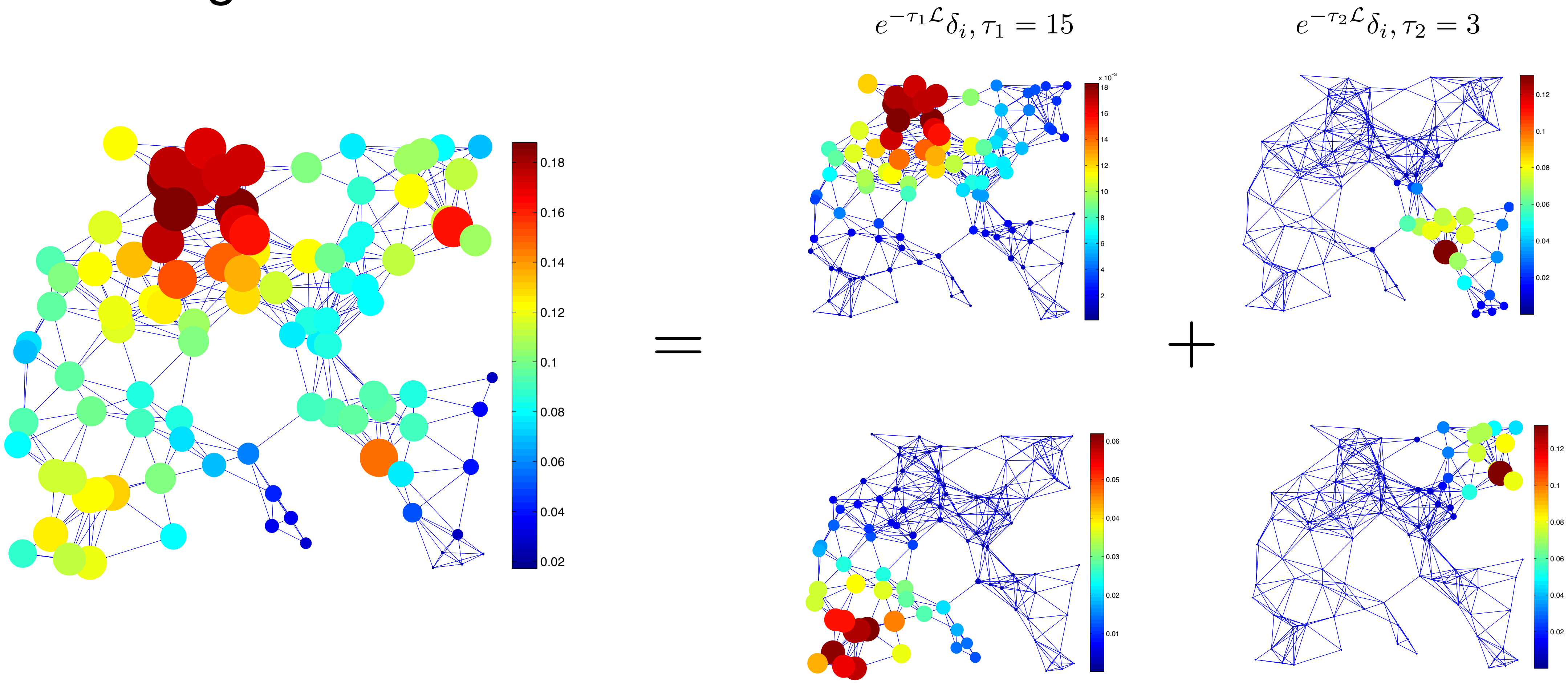
- Require the extraction of core features
  - Compression/Storage





Processing on Graphs

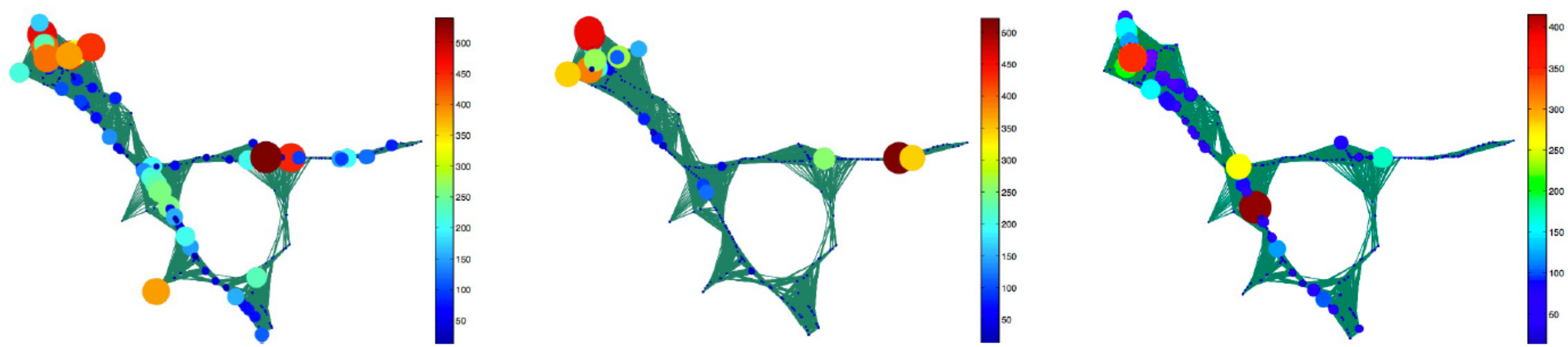
- Require the extraction of core features
  - Compression/Storage
  - Recognition/Identification of common features



Sparsity can reveal signal's structure

Dictionary for Graph Signals

- Need for new, meaningful graph signal representations that
  - ✓ reveal relevant structural properties of the graph signals/extract important features on graphs
  - ✓ sparsely represent different classes of signals on graphs

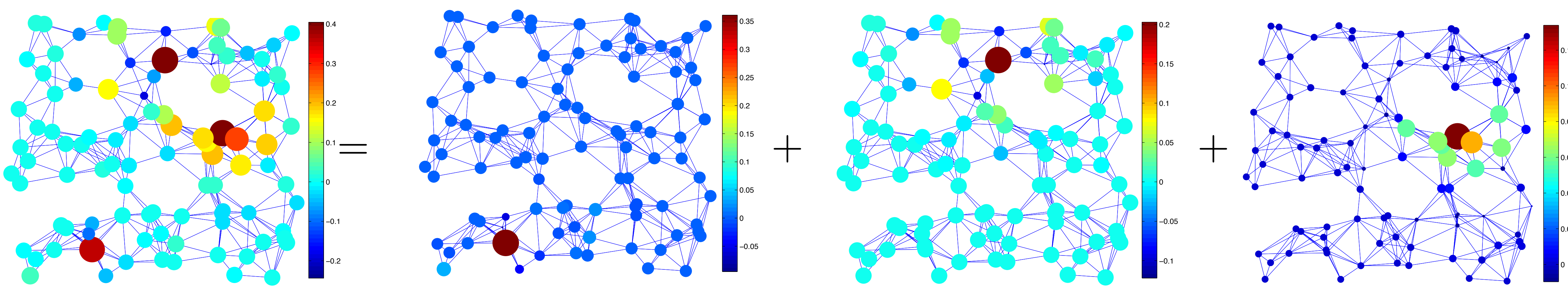


(a) Day 1 (b) Day 2 (c) Day 3

How can we define sparsity on graphs?

Our approach: Parametric graph dictionary

- We consider a general class of graph signals that are linear combination of overlapping local patterns
- The patterns can be translated in different nodes of the graphs.



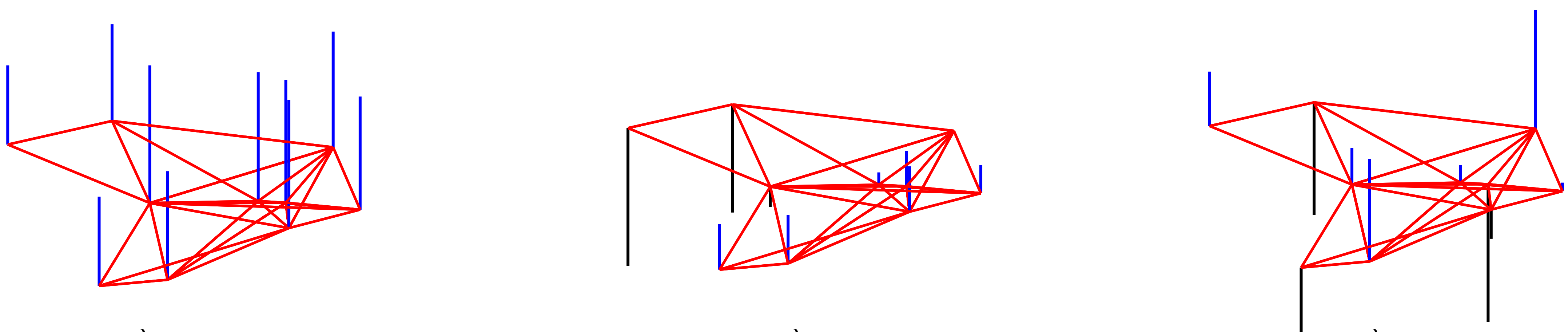
Given a set of training signals living on a graph, learn an overcomplete, parametric dictionary that incorporates the graph structure and can be efficiently implemented

Notations

- Connected, undirected, weighted graph $\mathcal{G} = (V, E, W)$
- Graph signal: a function $y : V \to \mathbb{R}$ that assigns real values to each vertex of the graph
- Normalized Laplacian $\mathcal{L} = I - D^{-1/2} W D^{-1/2}$
  - Complete set of orthonormal eigenvectors $\chi = [\chi_0, \chi_1, ..., \chi_{N-1}]$
  - Real, non-negative eigenvalues $0 = \lambda_0 < \lambda_1 <= \lambda_2 <= ... <= \lambda_{N-1} <= 2$

Graph Fourier Transform

- The eigenvectors of the Laplacian provide a harmonic analysis of graph signals



λ_0 λ_1 λ_8

GFT:
$$\hat{y}(\lambda_\ell) = \langle y, \chi_\ell \rangle = \sum_{n=1}^{N} y(n)\chi_\ell^*(n), \quad \ell = 0, 1, ..., N-1$$

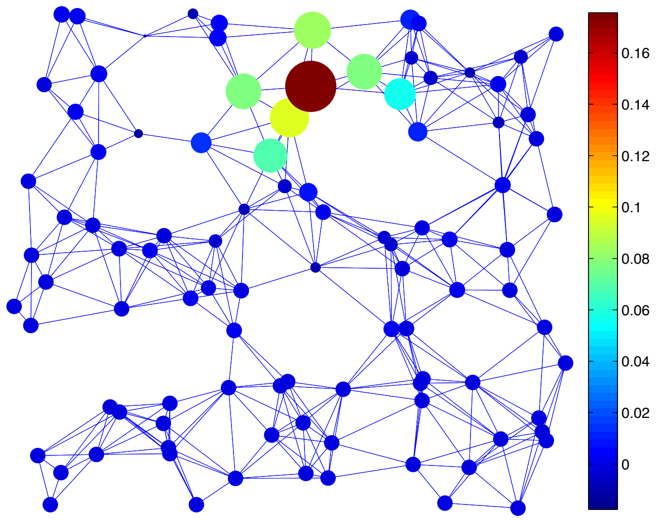
IGFT:
$$y(n) = \sum_{\ell=0}^{N-1} \hat{y}(\lambda_\ell)\, \chi_\ell(n), \quad \forall n \in \mathcal{V}$$

Translation of graph signals

- Given a signal g defined in the vertex domain, the translation to node n is defined as

$$T_n g = \sqrt{N}(g * \delta_n) = \sqrt{N} \sum_{\ell=0}^{N-1} \hat{g}(\lambda_\ell)\chi_\ell^*(n)\chi_\ell$$

- Smoothness of g controls the localization of $T_n g$ around the vertex n

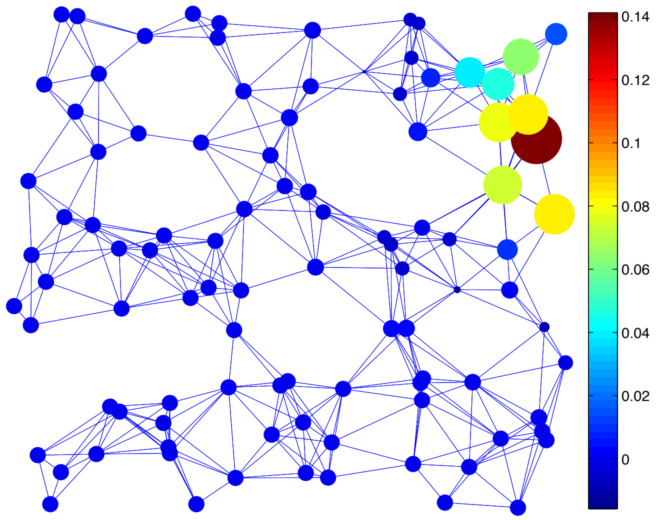


Translation of graph signals

- Given a signal g defined in the vertex domain, the translation to node n is defined as

$$T_n g = \sqrt{N}(g * \delta_n) = \sqrt{N} \sum_{\ell=0}^{N-1} \hat{g}(\lambda_\ell) \chi_\ell^*(n) \chi_\ell$$

- Smoothness of g controls the localization of $T_n g$ around the vertex n

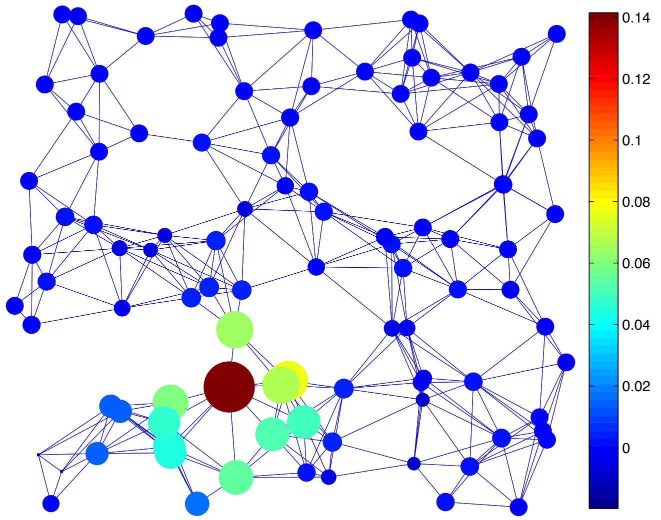


Translation of graph signals

- Given a signal g defined in the vertex domain, the translation to node n is defined as

$$T_n g = \sqrt{N}(g * \delta_n) = \sqrt{N} \sum_{\ell=0}^{N-1} \hat{g}(\lambda_\ell)\chi_\ell^*(n)\chi_\ell$$

- Smoothness of g controls the localization of $T_n g$ around the vertex n



Parametric Dictionary Structure

- A set of generating kernels $\{\widehat{g_s}(\cdot)\}_{s=1,2,\ldots,S}$ capture the spectral characteristics of the signals
- The atoms $T_n g$ are localized around a node n by choosing the kernel $\widehat{g}(\cdot)$ to be a smooth polynomial kernel of degree K

$$\hat{g}(\lambda_\ell) = \sum_{k=0}^{K} \alpha_k \lambda_\ell^k, \quad \ell = 0, \ldots, N-1$$

- A set of localized atoms is obtained by

$$Tg = \sqrt{N}\hat{g}(\mathcal{L}) = \sqrt{N}\sum_{k=0}^{K} \alpha_k \mathcal{L}^k$$

Parametric Dictionary Structure

- The structured graph dictionary $\mathcal{D} = [\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_S]$ is a concatenation of S subdictionaries of the form

$$\mathcal{D}_s = \widehat{g_s}(\mathcal{L}) = \sum_{k=0}^{K} \alpha_{sk} \mathcal{L}^k$$

  - Each kernel corresponds to a local pattern which is placed in different areas of the graph

  - The resulting atom given by column n of $\mathcal{D}_s$ has a support concentrated in the K-hop neighbourhood of vertex n

Dictionary Learning Algorithm

- Given a set of training signals $Y = [y_1, y_2, ..., y_M] \in \Re^N$ on the graph $\mathcal{G}$, solve

$$\begin{aligned}
\underset{\alpha \in \mathbb{R}^{(K+1)S},\ X \in \mathbb{R}^{SN \times M}}{\operatorname{argmin}} \quad & \{||Y - \mathcal{D}X||_F^2 + \mu\|\alpha\|_2^2\} \\
\text{subject to} \quad & \|x_m\|_0 \leq T_0, \quad \forall m \in \{1, ..., M\}, \\
& \mathcal{D}_s = \sum_{k=0}^{K} \alpha_{sk} \mathcal{L}^k, \quad \forall s \in \{1, 2, ..., S\} \\
& 0 \preceq \mathcal{D}_s \preceq c, \quad \forall s \in \{1, 2, ..., S\} \\
& (c - \epsilon_1)I \preceq \sum_{s=1}^{S} \mathcal{D}_s \preceq (c + \epsilon_2)I,
\end{aligned}$$

The spectral constraints guarantee that:
1. The learned kernels cover the whole spectrum
2. The dictionary is a frame

Outcome

- A structured dictionary that can sparsely represent graph signals and can be efficiently applied:
  1. Compact and easy to store (only $(K+1)S$ parameters)
  2. Fast application of dictionary forward and adjoint operators when the graph is sparse.

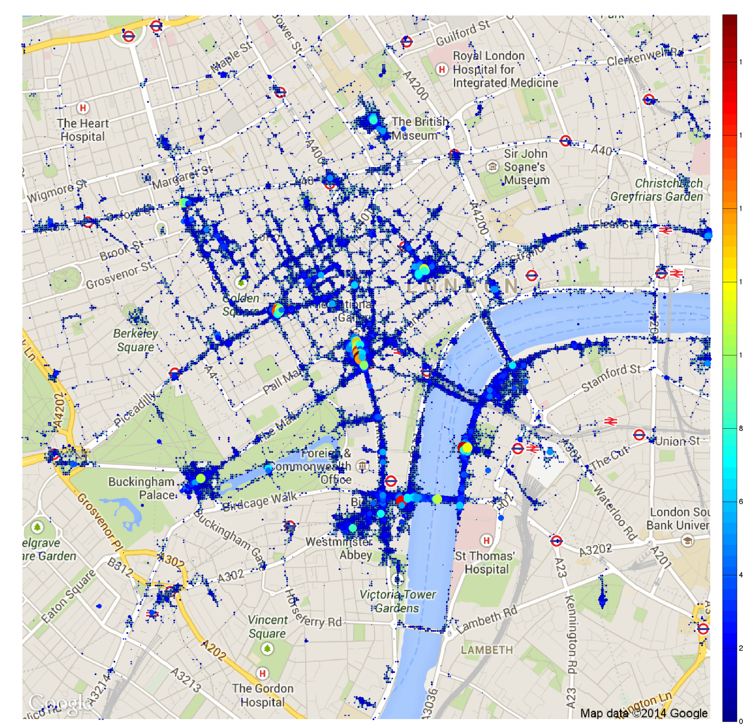
Example $$\mathcal{D}^T y = \sum_{s=1}^{S} \sum_{k=0}^{K} \alpha_{sk} \mathcal{L}^k y$$

- The computational cost of $\{\mathcal{L}^k y\}_{k=1,2,\ldots,K}$ is $O(K|\mathcal{E}|)$
- The total computational cost is $O(K|\mathcal{E}| + NSK)$

  3. Implementable in distributed settings

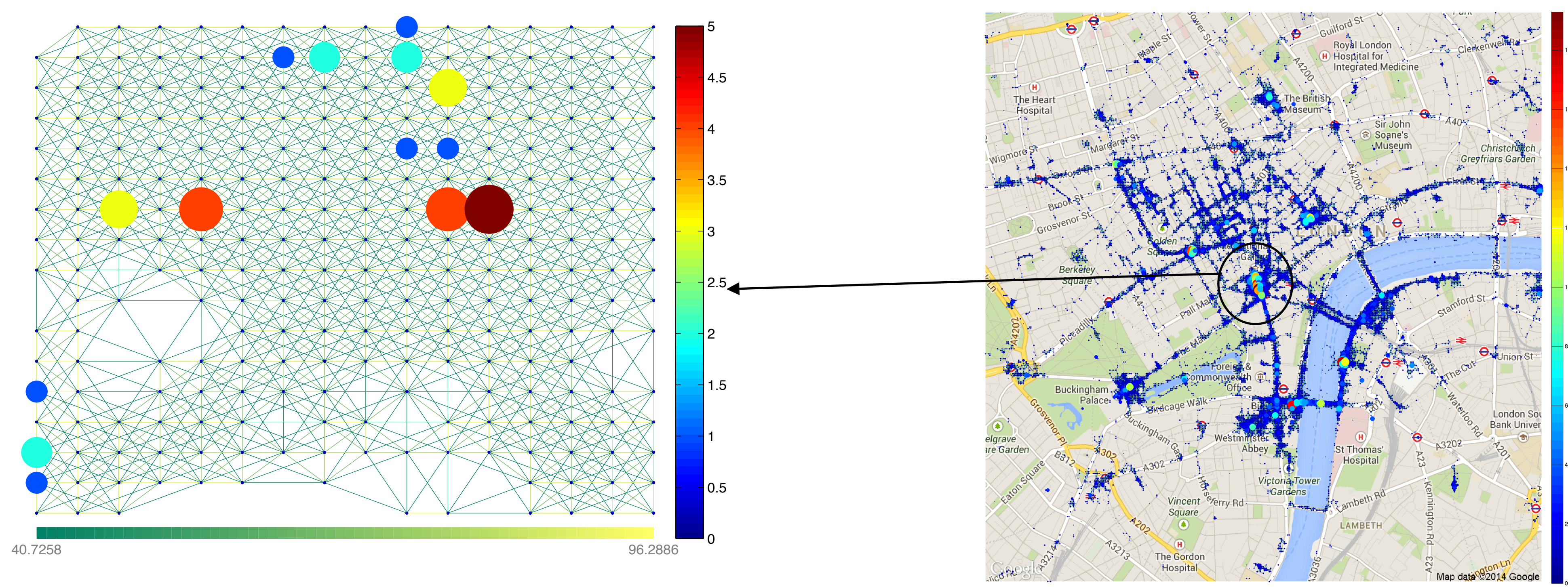
Flickr dataset

- Nodes: 245 vertices in the Trafalgar Square (London), each representing a geographical area 10x10m^2
- Assign edges when distance < 30m
- Graph Signals: Daily number of distinct users that took photos between Jan. 2010 and June 2012



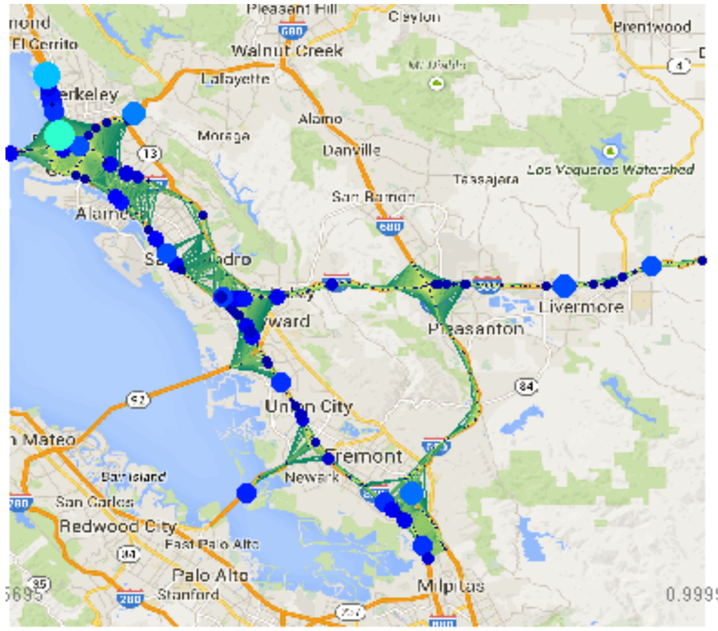
Flickr dataset

- Nodes: 245 vertices in the Trafalgar Square (London), each representing a geographical area 10x10m^2
- Assign edges when distance < 30m
- Graph Signals: Daily number of distinct users that took photos between Jan. 2010 and June 2012



Traffic dataset

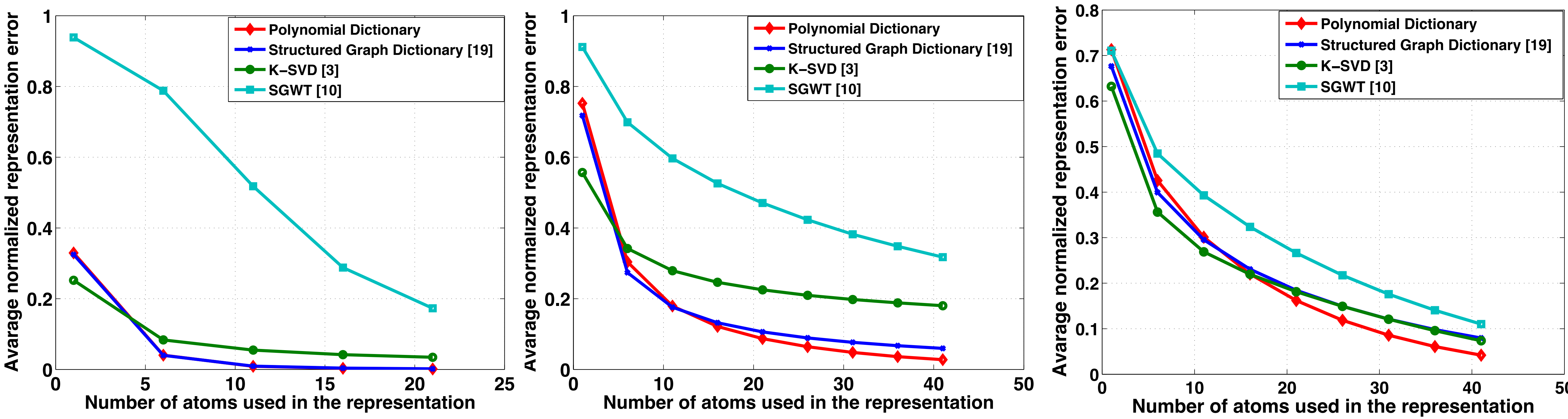
- Nodes: 439 detector stations in Alameda County, CA
- Assign edge when distance < 13km
- Graph Signals: Daily number of bottlenecks (in minutes) between Jan. 2007 to May. 2013



Brain dataset

- Nodes: 90 brain regions of contiguous voxels
- Edges assigned if anatomical distance < 40 mm
- Graph Signals: fMRI signals acquired on five subjects, in different states - 1290 signals per subject

Approximation performance

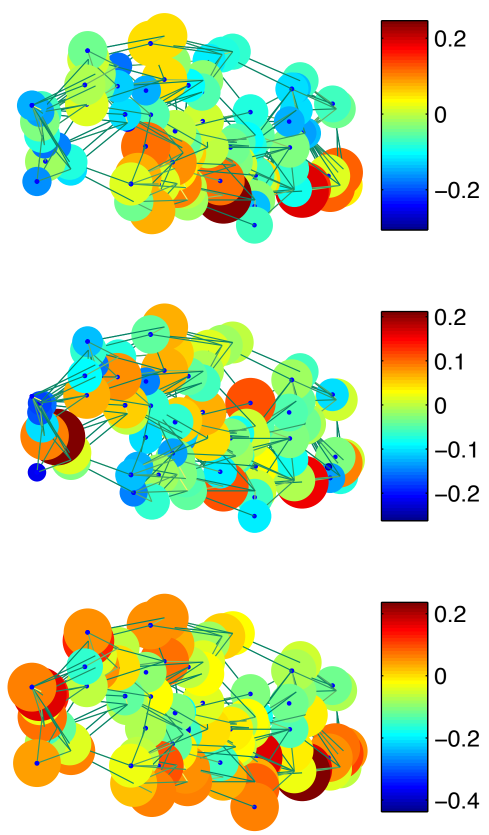


(a) Flickr dataset **(b) Traffic dataset** **(c) Brain dataset**

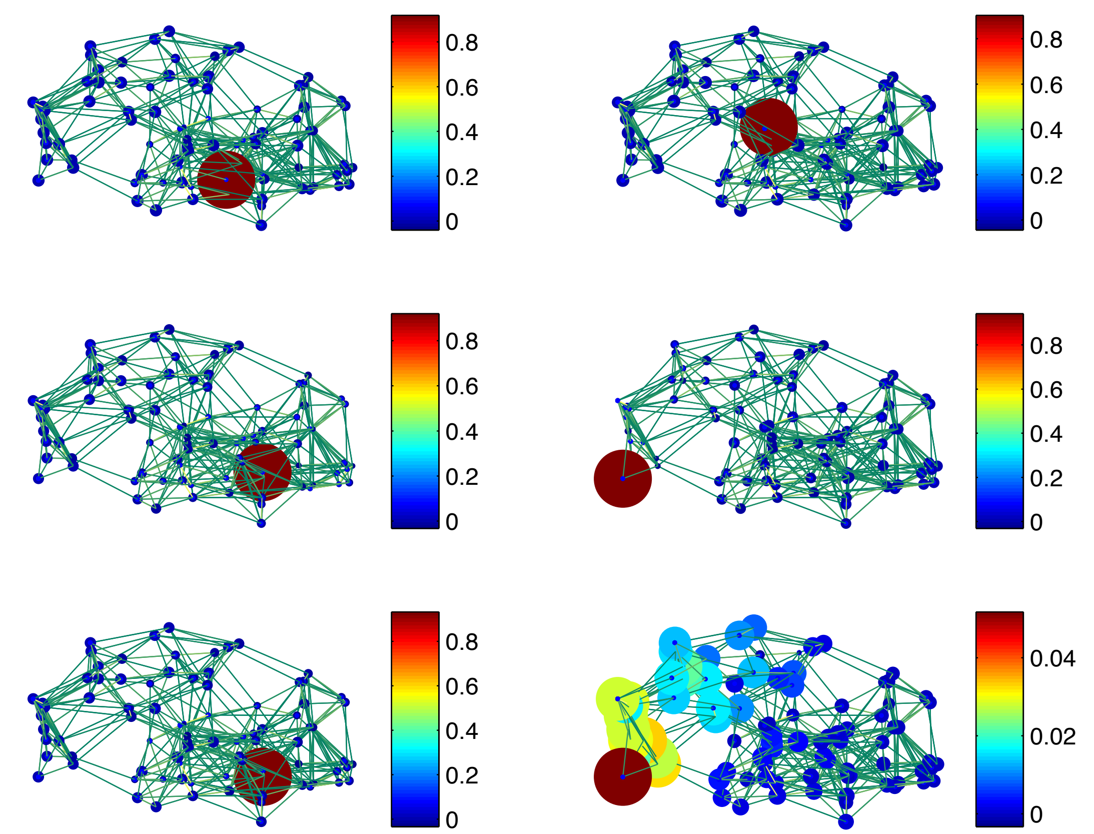
- As the sparsity level increases, the localization property becomes beneficial
- The polynomial dictionary is able to learn local patterns in areas of the graph that do not show up in the training signals.

Examples of Learned Atoms

- Six most commonly included atoms by OMP



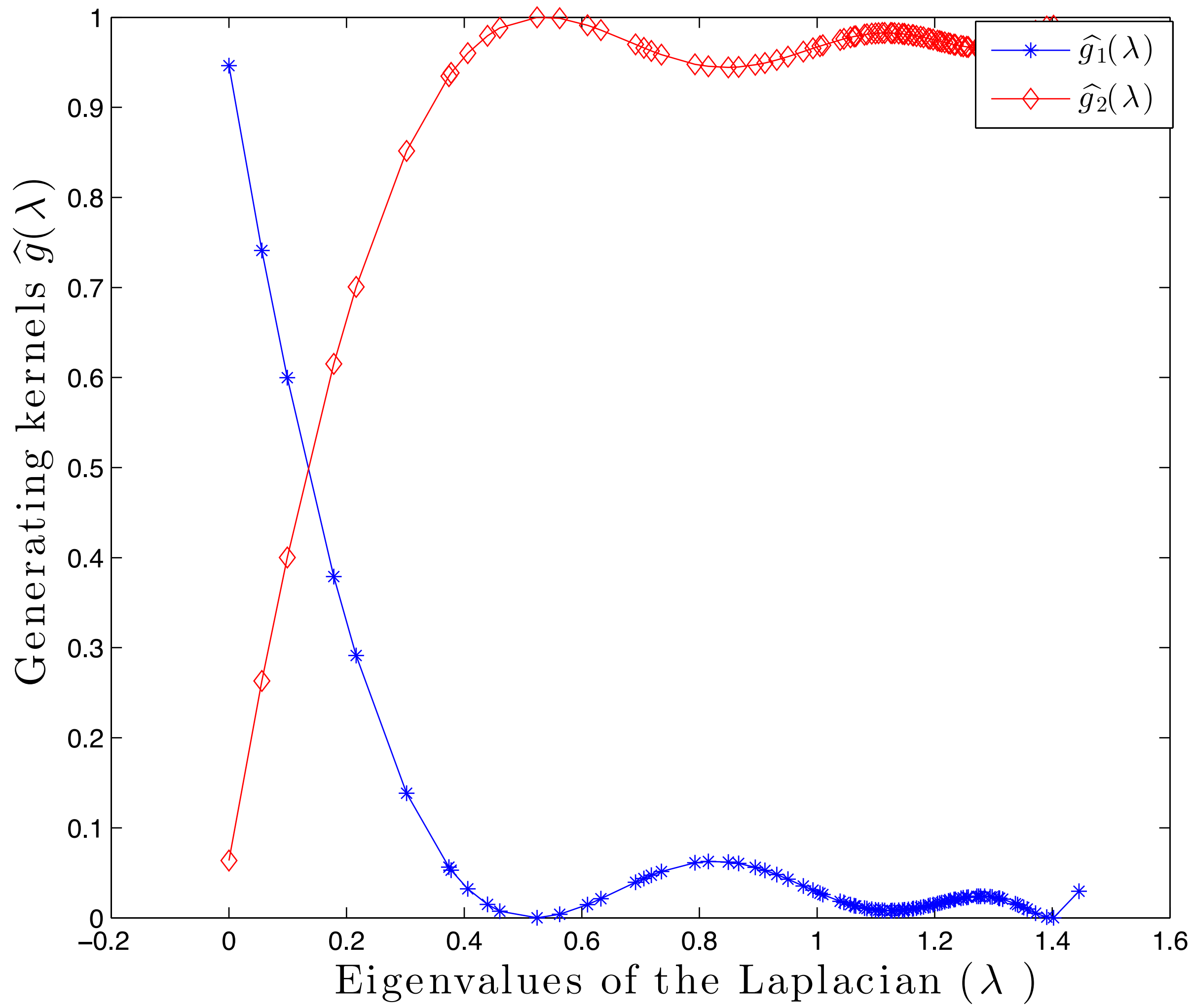
K-SVD Dictionary



Polynomial Graph Dictionary

Example of Learned Kernels

- Brain dataset:



- The learned kernels correspond to a low-pass and a high-pass filter

Application: Image Segmentation

- For each pixel (node), extrapolate 5x5 patches
  - Patch binary graph: connect each pixel to its horizontal and vertical neighbors
  - Signal on the graph: patch intensity
- Learn a dictionary from patch signals with $S = 4, K = 15$
- Filter each signal with the learned filters i.e.,

$$\mathcal{D}_s^T y_j = \sum_{\ell=0}^{N-1} \widehat{y_j}(\lambda_\ell)\widehat{g_s}(\lambda_\ell)\chi_\ell$$

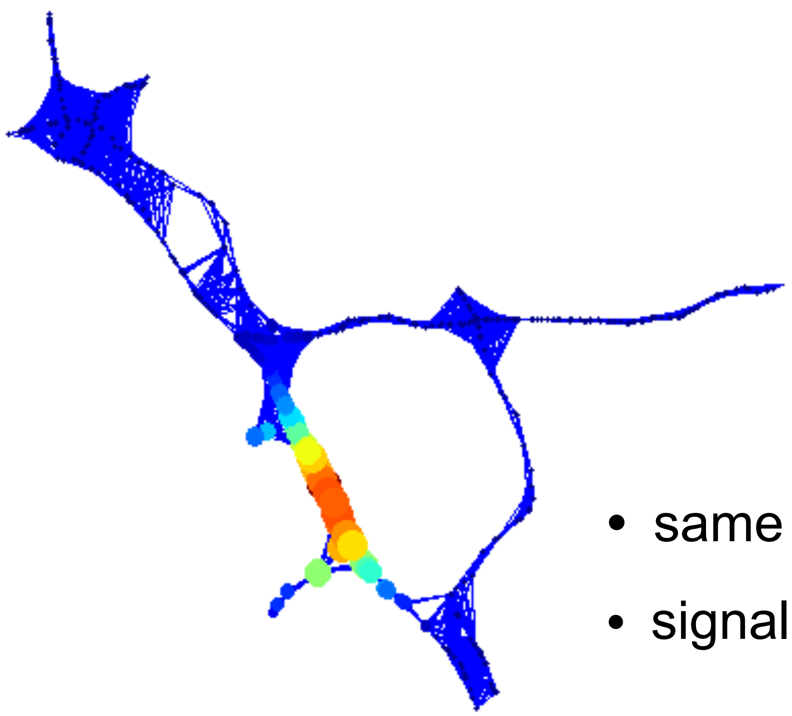
- Node feature: mean and variance of the filtered signals
- Clustering: K-means on the feature vectors

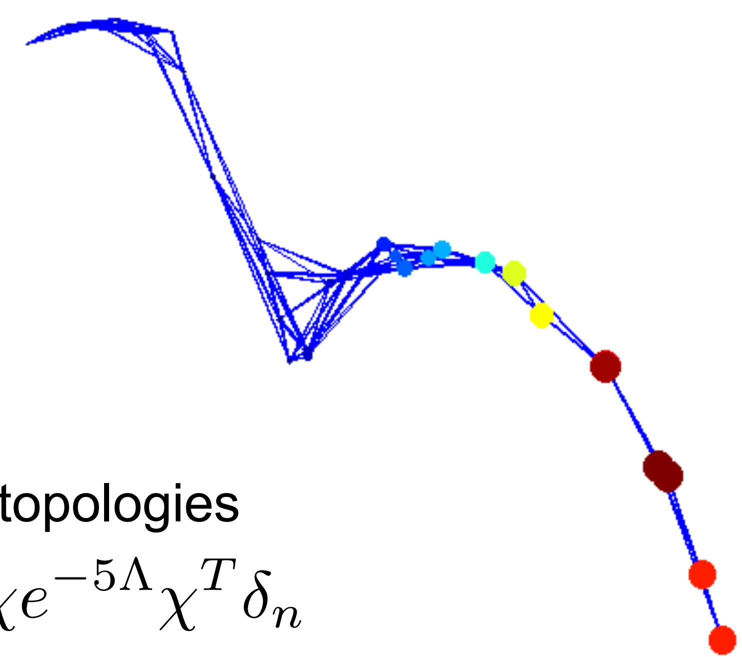
Application: Image Segmentation

Multi-graph dictionary learning

- **Problem:** Learn atoms for effective representation of signals, that are collected on different graph topologies

- **Main assumption:** Signals on different topologies may share similar spectral characteristics





- same process evolving in two different topologies
- signal observation: $y = e^{-5\mathcal{L}}\delta_n = \chi e^{-5\Lambda}\chi^T\delta_n$

Multi-graph dictionary learning

- Learn generating kernels that capture the common information across the graphs in the spectral domain

- Capture the common spectral components through the polynomial coefficients

$$
\begin{aligned}
\underset{\alpha \in \mathbb{R}^{(K+1)S},\ X_t \in \mathbb{R}^{SN \times M_t}}{\operatorname{argmin}} \quad & \left\{ \sum_{t=1}^{T} \frac{1}{M_t} ||Y_t - \mathcal{D}_t X_t||_F^2 + \mu \|\alpha\|_2^2 \right\} \\
\text{subject to} \quad & \|X_t^m\|_0 \leq T_0, \quad \forall m \in \{1, ..., M_t\}, \\
& \mathcal{D}_t^s = \sum_{k=0}^{K} \alpha_{sk} \mathcal{L}_t^k, \forall s \in \{1, 2, ..., S\}, \\
& 0 \preceq \mathcal{D}_t^s \preceq c, \forall s \in \{1, 2, ..., S\}
\end{aligned}
$$

Example of graph processes

1. Heat diffusion kernel:

$$\widehat{g}_\tau(\lambda_k) = e^{-\tau\lambda_k}$$

2. Wave kernel:

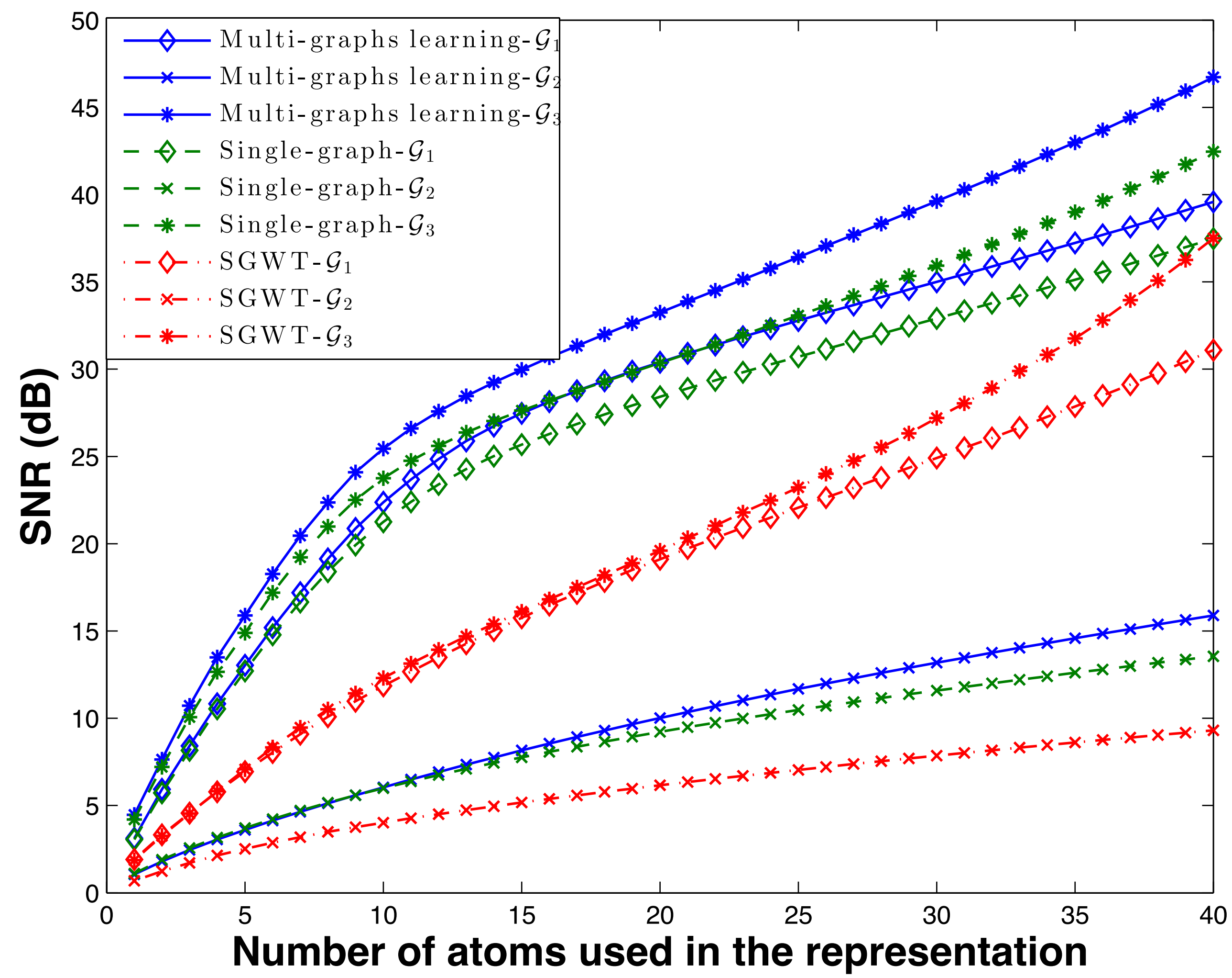
$$\widehat{g}_\tau(\lambda_k) = e^{-\frac{(\tau - \log \lambda_k)^2}{2\sigma^2}}$$

3. Spectral graph wavelet kernel:

$$\widehat{g}_\tau(\lambda_k) = g(\tau\lambda_k)$$

Preliminary results

- Consider bottleneck signals from Jan. 2007-Aug.2014 on three different graphs:
  - San Francisco $(\mathcal{G}_1)$
  - Alameda $(\mathcal{G}_2)$
  - Santa Barbara $(\mathcal{G}_3)$



Summary

- Take-home messages:
  - Localization is important
  - Polynomial matrix functions of the graph Laplacian seems to be a flexible structure for sparsely representing graph signals

- Still many open questions:
  - Regularization on graphs with the learned kernels
  - Applications where the kernel information could be beneficial, such as classification, coding etc
  - Definition of the optimal graph topology

References

- D. I Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, “The emerging field of signal processing on graphs: Extending high - dimensional data analysis to networks and other irregular domains,” IEEE Signal Process. Mag., vol. 30, no. 3, pp. 83–98, May 2013
- **D. Thanou, D. I Shuman, and P. Frossard, “Learning Parametric Dictionaries for Signals on Graphs”, IEEE Trans. Signal Process., vol. 62, no. 15, Aug. 2014**
- **D. Thanou, and P. Frossard, “Multi-graph learning of spectral graph dictionaries”, accepted in ICASSP 2015.**
- D. Hammond, P. Vandergheynst, and R. Gribonval, “Wavelets on graphs via spectral graph theory,” Appl. Comput. Harmon. Anal., vol. 30, no. 2, pp. 129–150, March 2010.
- R. R. Coifman and M. Maggioni, “Diffusion wavelets,” Appl. Comput. Harmon. Anal., vol. 21, pp. 53–94, March 2006.
- X. Zhu and M. Rabbat, “Approximating signals supported on graphs,” in Proc. IEEE Int. Conf. Acc., Speech, and Signal Process., Kyoto, Japan, Mar. 2012, pp. 3921–3924.
- S. K. Narang and A. Ortega, “Perfect reconstruction two-channel wavelet filter banks for graph structured data,” IEEE Trans. Signal Process., vol. 60, no. 6, pp. 2786–2799, June 2012.
- S. K. Narang and A. Ortega, “Compact Support Biorthogonal Wavelet Filterbanks for Arbitrary Undirected Graphs,” IEEE Trans. Signal Process., vol. 61, no. 19, pp. 4673–4685, Oct. 2013.

References

- R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," Proc. of the IEEE, vol. 98, no. 6, pp. 1045 –1057, Apr. 2010.
- M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, 2006.
- X. Zhang, X. Dong, and P. Frossard, "Learning of structured graph dictionaries," in Proc. IEEE Int. Conf. Acc., Speech, and Signal Process., Kyoto, Japan, Mar. 2012, pp. 3373 – 3376.