ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

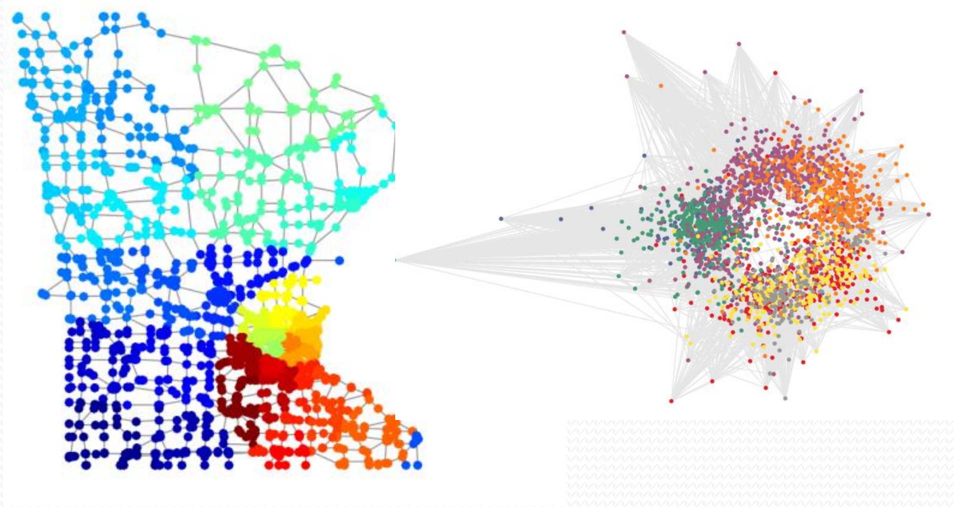
ESTIMATION OF PARTIALLY OBSERVED GRAPH SIGNALS BY LEARNING SPECTRALLY MATCHED GRAPH DICTIONARIES

Osman Furkan KAR

August 28, 2023
METU, Department of Electrical and Electronics Engineering
Room: D115

Graph Models

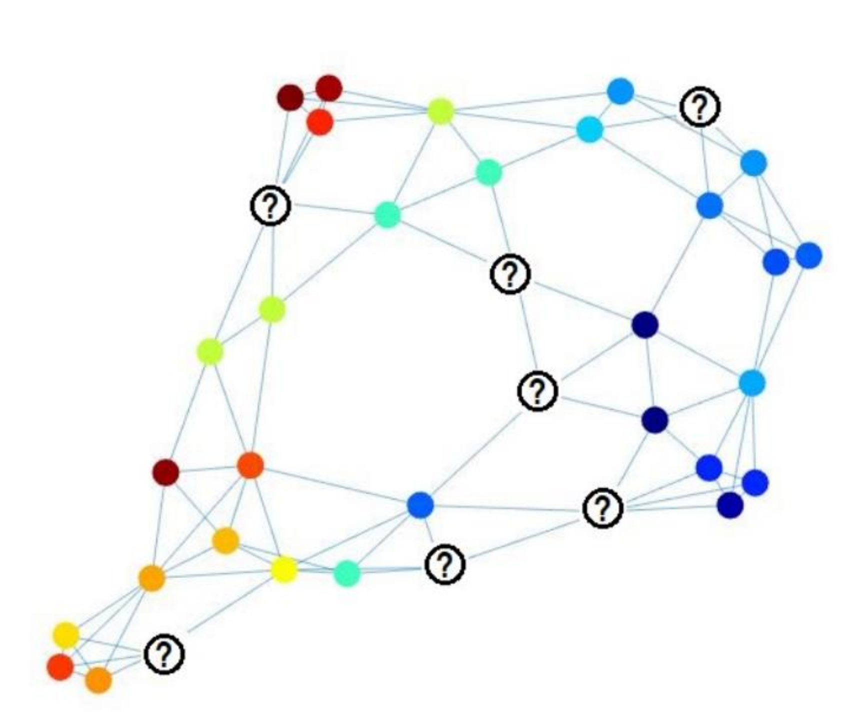
Social Network



Sensor Network

- Many modern applications involve data acquired on an irregular network topology.

Graph Models



Estimation/Inpainting of graph signals

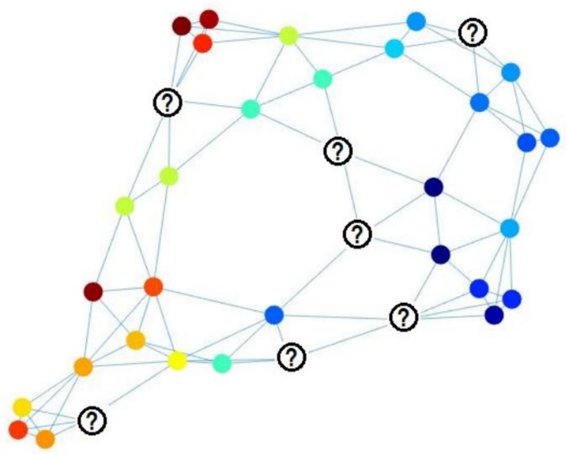
OUTLINE

- Foundations of GSP
- Literature
- Proposed Method (SGKL)
  - Novelties
  - Signal Model, Objective function
  - Algorithm
- Experiments
  - Comparative
  - Performance and Sensitivity Analysis
- Conclusion

OUTLINE

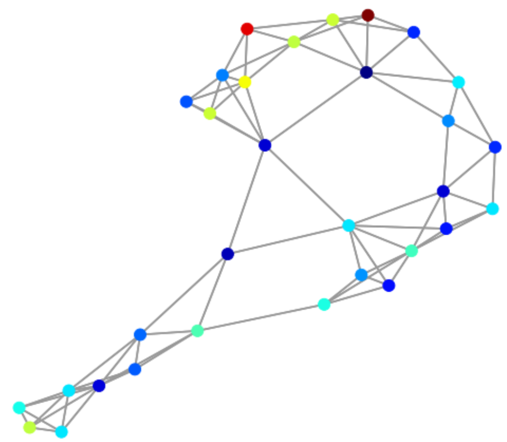
- **Foundations of GSP**
- Literature
- Proposed Method (SGKL)
    - Novelties
    - Signal Model, Objective function
    - Algorithm
- Experiments
    - Comparative
    - Performance and Sensitivity Analysis
- Conclusion

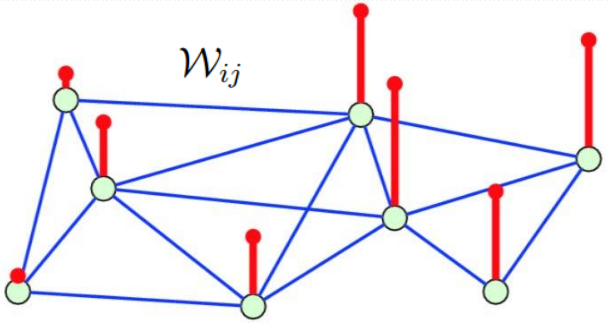
Introduction: Graph Signal Processing



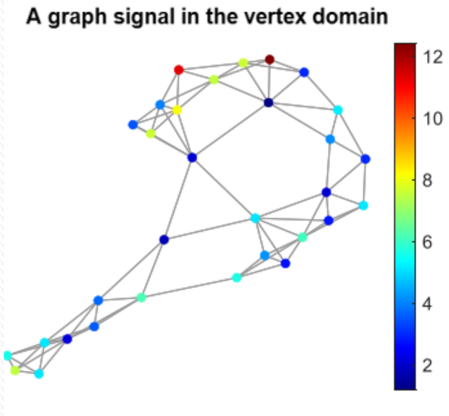
$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$$

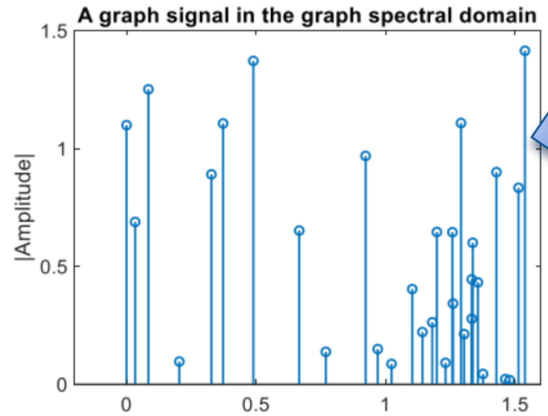
$f : \mathcal{V} \rightarrow \mathbb{R}$: A graph signal, $f \in \mathbb{R}^N$





Introduction: Graph Signal Processing





$$D_{ii} = \sum_j \mathcal{W}_{ij}$$

$$\mathcal{L}_{\mathcal{D}} := \mathcal{D} - \mathcal{W}.$$

$$\mathcal{L} := \mathcal{I} - \mathcal{D}^{-1/2}\mathcal{W}\mathcal{D}^{-1/2}$$

$$\mathcal{L}\, u_k = \lambda_k u_k$$

Classical Signal Processing

$$\hat{f}(\xi) := \langle f, e^{2\pi j \xi t} \rangle = \int_{\mathbb{R}} f(t) e^{-2\pi j \xi t} dt.$$

Graph Signal Processing

$$\hat{f}\,(\lambda_\ell) := \langle \mathbf{f}, \mathbf{u}_\ell \rangle = \sum_{i=1}^{N} f(i) u_\ell^*(i).$$

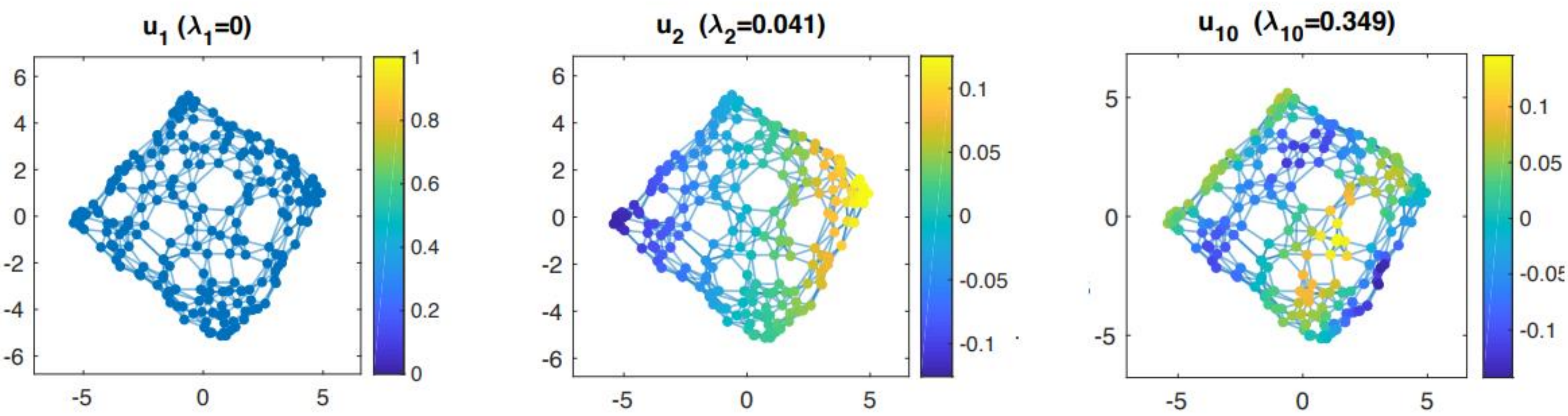
Introduction: Graph Signal Processing

Classical Signal Processing

$$\hat{f}(\xi) := \langle f, e^{2\pi j \xi t} \rangle = \int_{\mathbb{R}} f(t) e^{-2\pi j \xi t} dt.$$

Graph Signal Processing

$$\hat{f}(\lambda_\ell) := \langle \mathbf{f}, \mathbf{u}_\ell \rangle = \sum_{i=1}^{N} f(i) u_\ell^*(i).$$



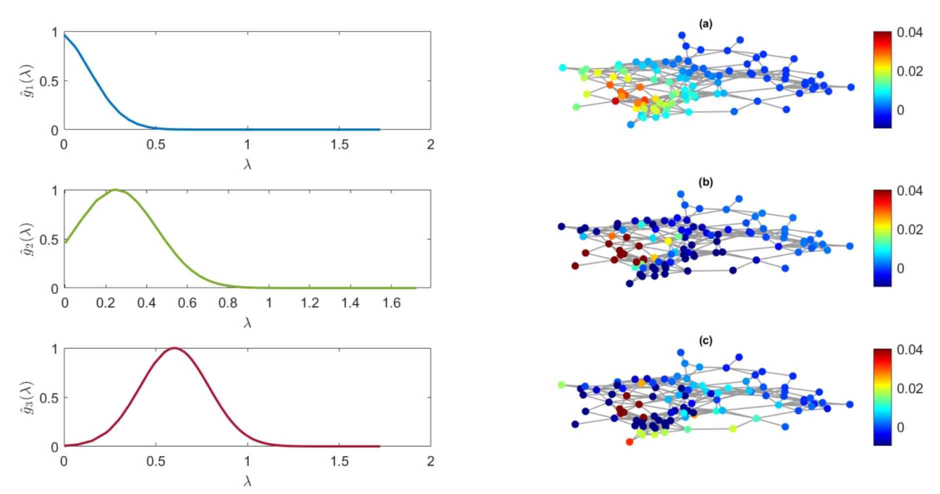
Introduction: Graph Signal Processing

Graph signal processing: Irregular topologies

Define:

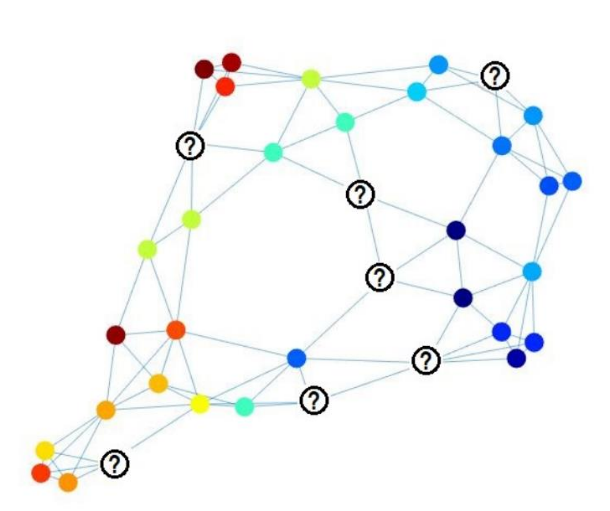
- Inverse GFT
- Translation & Localization
- Modulation
- Filtering
- Convolution
- …

in Spectral Domain



OUTLINE

- Foundations of GSP
- **Literature**
- Proposed Method (SGKL)
    - Novelties
    - Signal Model, Objective function
    - Algorithm
- Experiments
    - Comparative
    - Performance and Sensitivity Analysis
- Conclusion

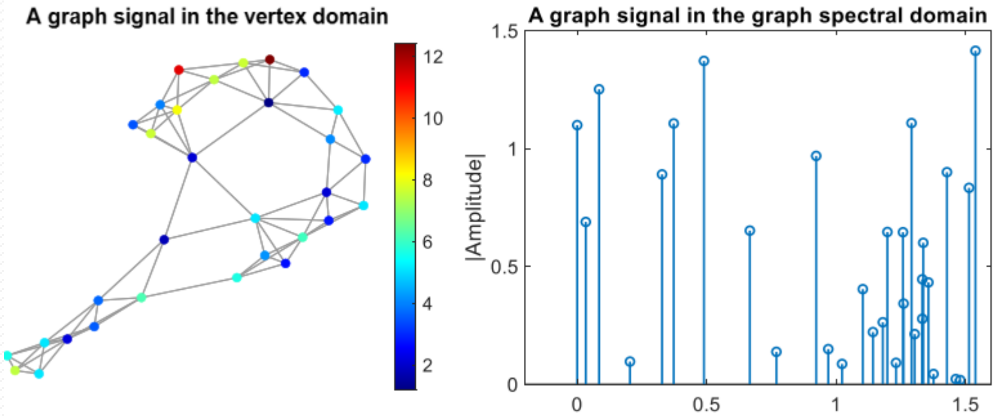


Literature

- Inference of graph signals based on multi-scale dictionary learning techniques (GEMS) (Yankelevsky, 2019)
- Tikhonov regularization (Graph Based Tik.) (Zhou, 2004)
- Total Variation Minimization (TVMin) (Jung, 2019)
- Interpolation of non-smooth graph signals (LSEM) (Mazarguil, 2022)
- Iterative graph signal reconstruction methods (O-PGIR) (Brugnoli, 2020)
- Learning spectrally concentrated kernels method (SCGDL) (Turhan, 2021)
- Graph neural networks and deep algorithm unrolling based methods

Literature

- Smoothness,
- Bandlimitedness,



Wind-Speed Measurement, Molene Dataset

- Lack of capacity to generalize

OUTLINE

- Foundations of GSP
- Literature
- **Proposed Method (SGKL)**
  - Novelties
  - Signal Model, Objective function
  - Algorithm
- Experiments
  - Comparative
  - Performance and Sensitivity Analysis
- Conclusion

Proposed Method

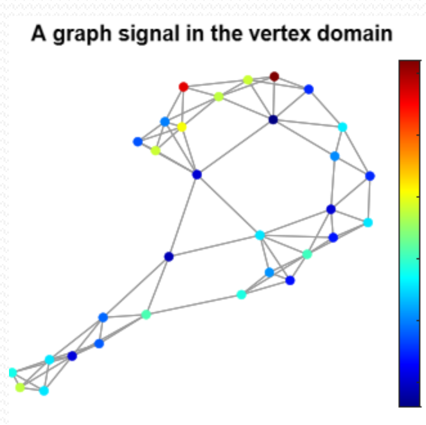
- Estimation of partially observed graph signals

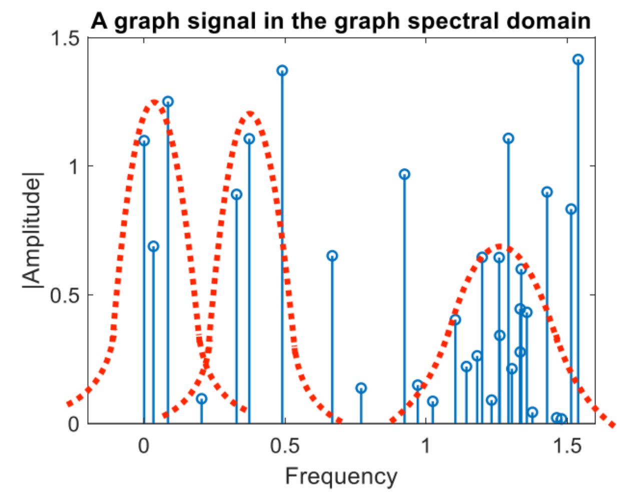
•Narrowband Spectral Graph Kernels
•Spectral Graph Dictionaries

Dictionary

$$Y^{m}_{\text{estimated}} = D^{m} X^{m}$$

Sparse Codes

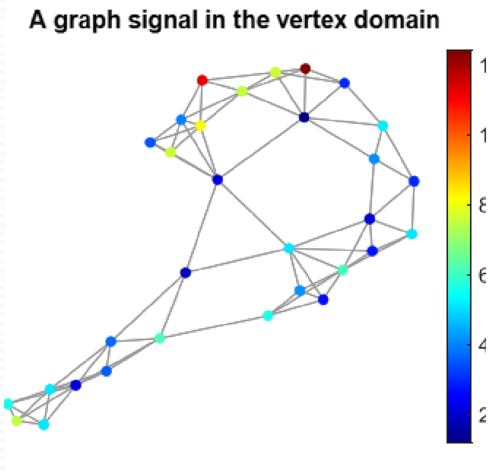


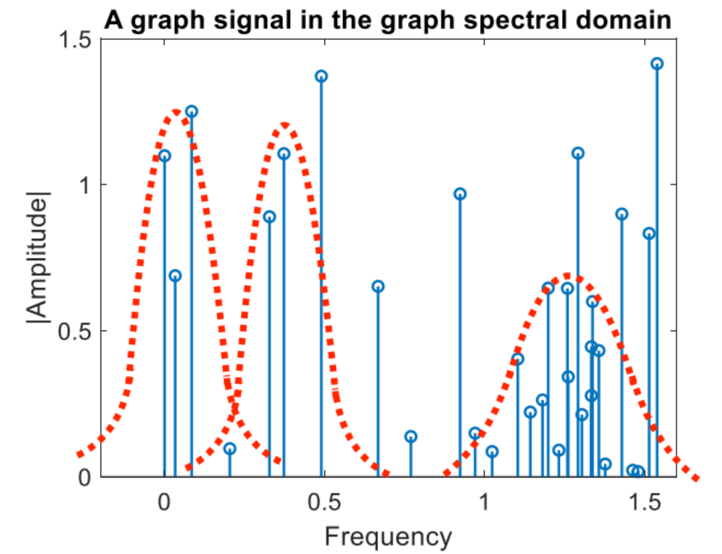


Novelties

- Explicitly model the spectral components of the data and fits them to the parameters of narrowband spectral graph kernels

- Data have missing entries while learning where missing entries of each signal in the dataset differ from each other

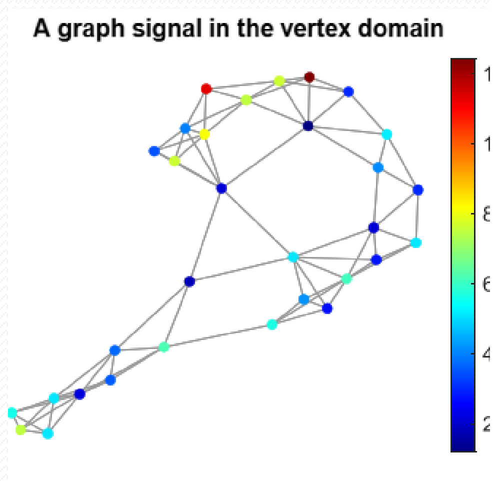


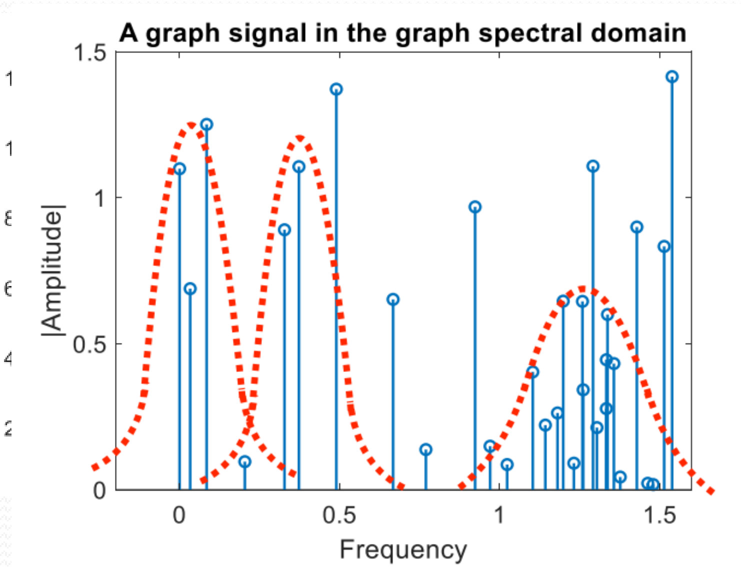


Novelties

- The capacity to generalize the model to any given graph by learning the graph dictionaries solely in the spectral domain

- Ability to jointly learn graph dictionaries and their sparse codings for multiple graphs and data sets.





Signal Model and Notation

$$\hat{g}_j(\lambda) = \exp\left(-\frac{\|\lambda-\mu_j\|^2}{s_j^2}\right)$$ → Gaussian Spectral Kernels

$$D_j^m = U^m \hat{g}_j(\Lambda^m)(U^m)^T \in \mathbb{R}^{N^m \times N^m}$$ → Spectral Sub-Dictionaries

$$D^m = \begin{bmatrix} D_1^m & D_2^m & \cdots & D_J^m \end{bmatrix} \in \mathbb{R}^{N^m \times JN^m}$$ → Spectral Dictionaries

$$y_i^m = D^m x_i^m + w_i^m$$

Proposed Method

Two Step Minimization. Not Jointly Convex.

$$
\begin{aligned}
\min_{\{X^m\},\psi} \; & \overbrace{\sum_{j=1}^{J} (\mu_j)^2 + \eta_s \sum_{j=1}^{J} (s_j - s_0)^2}^{\text{Spectral Kernel Parameters}} + \eta_x \sum_{m=1}^{M} \|X^m\|_1 \\
& + \eta_w \sum_{m=1}^{M} \sum_{i=1}^{K^m} \left\| S^{m,i} y_i^m - S^{m,i} D^m x_i^m \right\|^2 \\
& + \underbrace{\eta_y \sum_{m=1}^{M} \operatorname{tr}((X^m)^T (D^m)^T L^m D^m X^m)}_{\text{Smoothly Varying Reconstructed Signal}} + \eta_c \sum_{m=1}^{M} \operatorname{tr}((X^m) \widetilde{L}^m (X^m)^T)
\end{aligned}
$$

Sparse Representation over Graph Dictionary

Coherency with partial observations

Similar Reconstructed Signals with Similar Dictionary Atoms

Proposed Method

Two Step Minimization. Not Jointly Convex.

- Fix D, Optimize X
  - ADMM
- Fix X, Optimize D
  - Gradient Descent

$$\min_{\{X^m\}} \eta_x \sum_{m=1}^{M} \|X^m\|_1 + \eta_w \sum_{m=1}^{M} \sum_{i=1}^{K^m} \left\|S^{m,i} y_i^m - S^{m,i} D^m x_i^m\right\|^2$$
$$+ \eta_y \sum_{m=1}^{M} ((X^m)^T (D^m)^T \mathcal{L}^m D^m X^m)$$
$$+ \eta_c \sum_{m=1}^{M} ((X^m) \widetilde{\mathcal{L}}^m (X^m)^T).$$

$$\min_{\psi} f(\psi) = \min_{\psi} \sum_{j=1}^{J} (\mu_j)^2 + \eta_s \sum_{j=1}^{J} (s_j - s_0)^2$$
$$+ \eta_w \sum_{m=1}^{M} \sum_{i=1}^{K^m} \left\|S^{m,i} y_i^m - S^{m,i} D^m x_i^m\right\|^2$$
$$+ \eta_y \sum_{m=1}^{M} ((X^m)^T (D^m)^T \mathcal{L}^m D^m X^m).$$

OUTLINE

- Foundations of GSP
- Literature
- Proposed Method (SGKL)
  - Novelties
  - Signal Model, Objective function
  - Algorithm
- **Experiments**
  - **Comparative**
  - Performance and Sensitivity Analysis
- Conclusion

Experiments: Baseline Methods

- Spectral Graph Wavelet Dictionaries (tight frame, Meyer, Mexican hat, and ab-spline) (Hammond, 2011)
- Inference of graph signals based on multi-scale dictionary learning (GEMS) (Yankelevsky, 2019)
- Tikhonov regularization (Graph Based Tik.) (Zhou, 2004)
- Total Variation Minimization (TVMin) (Jung, 2019)
- Interpolation of non-smooth graph signals (LSEM) (Mazarguil, 2022)
- Iterative graph signal reconstruction methods (O-PGIR) (Brugnoli, 2020)
- Learning spectrally concentrated kernels method (SCGDL) (Turhan, 2021)

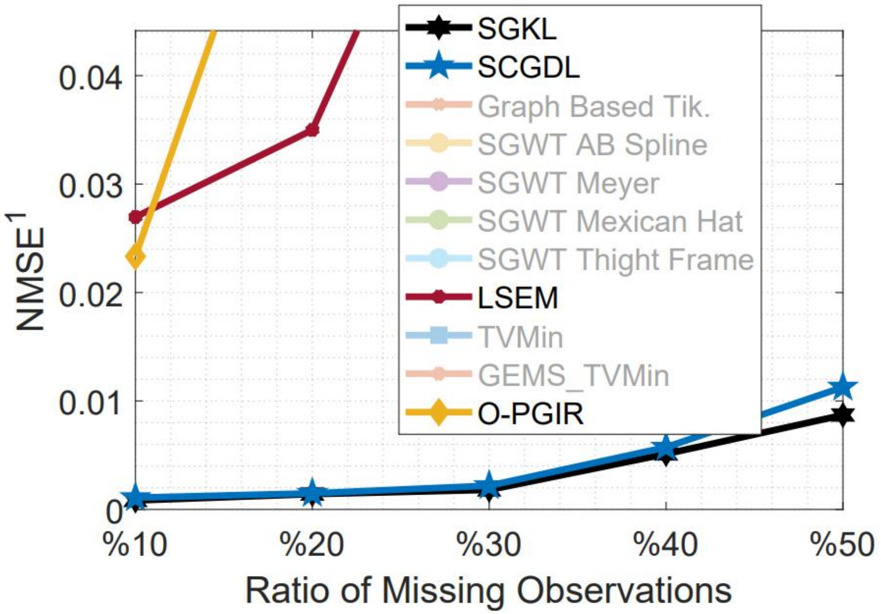
Experiments: Data Set

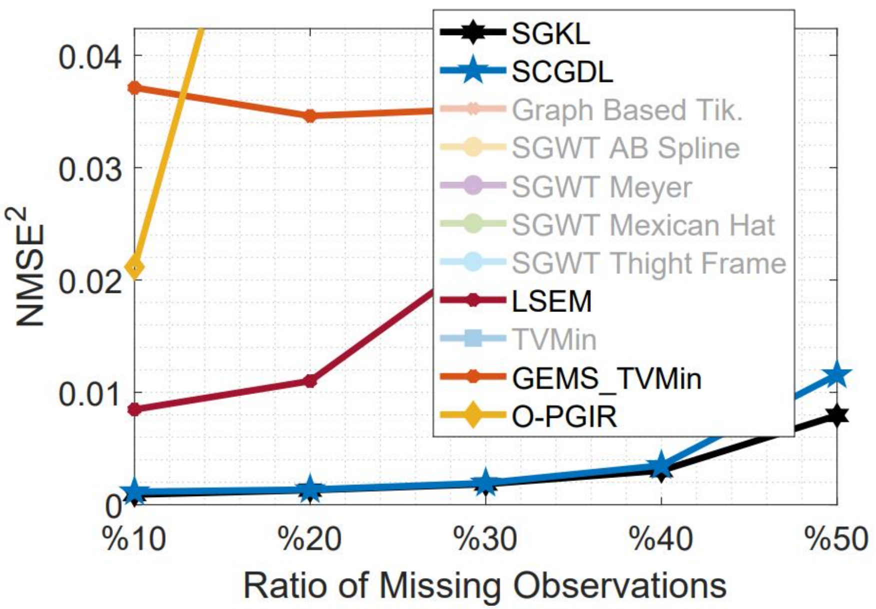
Synthetic Data Set:

- G^1 and G^2 10-NN graphs with 100 Nodes.
- J=4 Spectral Kernels
- K^1 = 200 and K^2 = 400 Signals

Experiments: Results

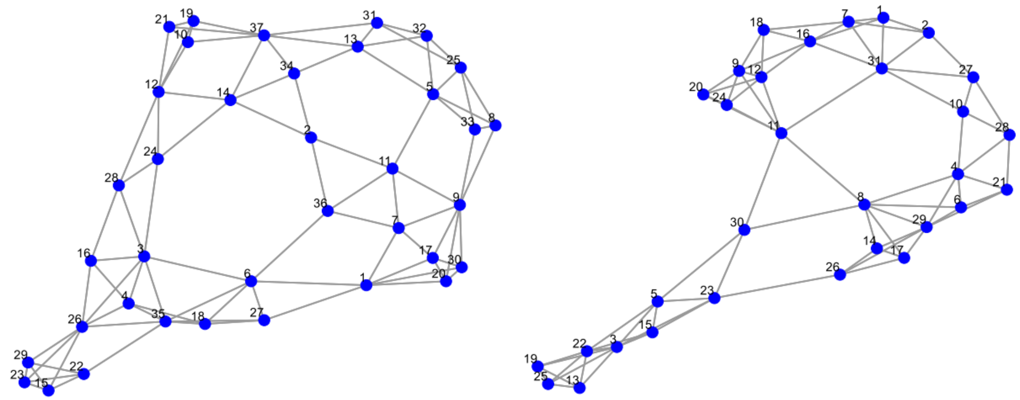
$$NMSE^m = \left\| Y_u^m - \tilde{Y}_u^m \right\|^2 / \| Y_u^m \|^2$$





Experiments: Data Set

Molene Data Set:
Released by the French national meteorological service which consists of temperature and wind speed measurements taken at different locations in the Brittany region of France.

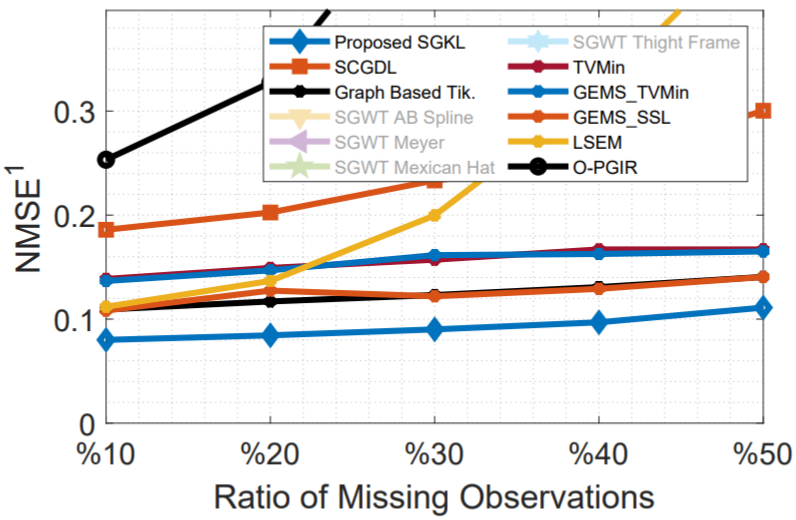


(a) Temperature (b) Wind-Speed

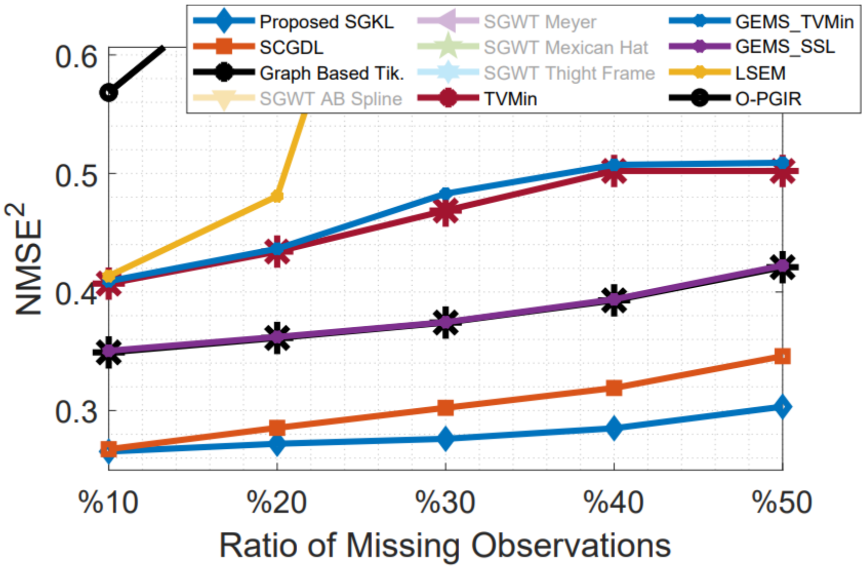
$N^1 = 37$ $N^2 = 31$, 5-NN
$K^1 = K^2 = 744$

Experiments: Results

$$NMSE^m = \left\| Y_u^m - \tilde{Y}_u^m \right\|^2 / \| Y_u^m \|^2$$



(a) G1-Temperature



(b) G2-Wind-Speed

Experiments: Data Set

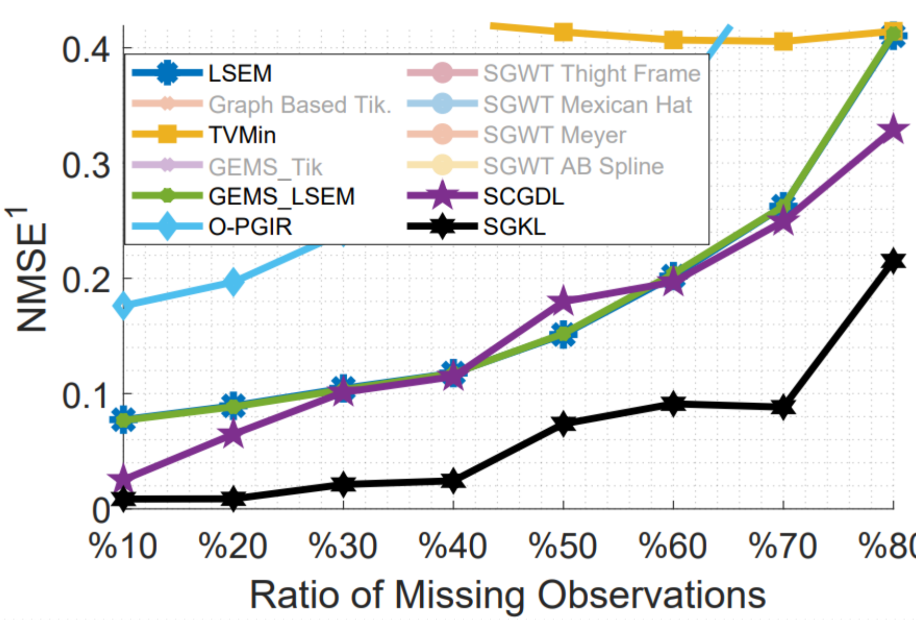
Covid Data Set:
a publicly available online dataset which contains the number of daily new cases per country

(a) Europe (b) US

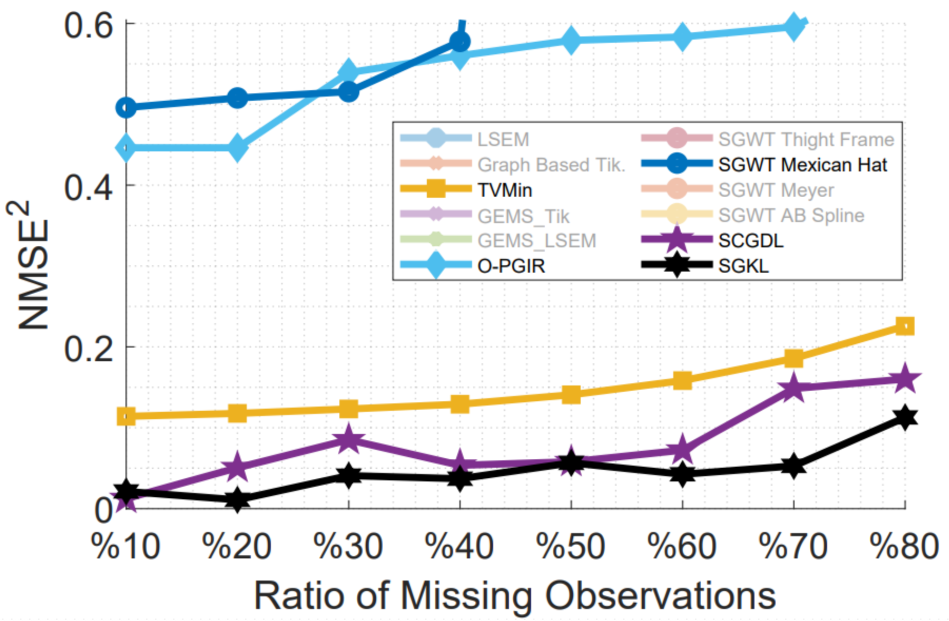
$N^1 = 37$ $N^2 = 50$, 4-NN, 3-NN
$K^1 = K^2 = 483$

Experiments: Results

$$NMSE^m = \left\|Y_u^m - \tilde{Y}_u^m\right\|^2 / \|Y_u^m\|^2$$

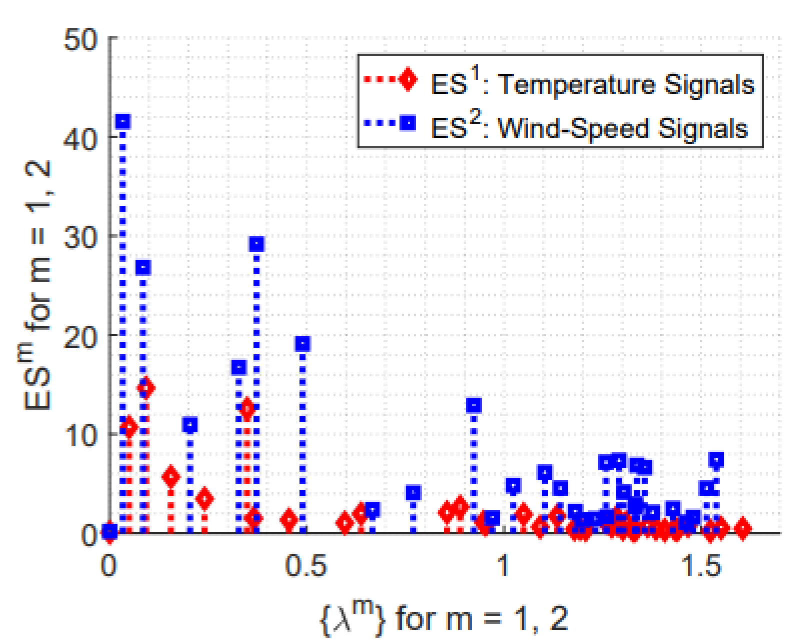


(a) G[1]-Europe

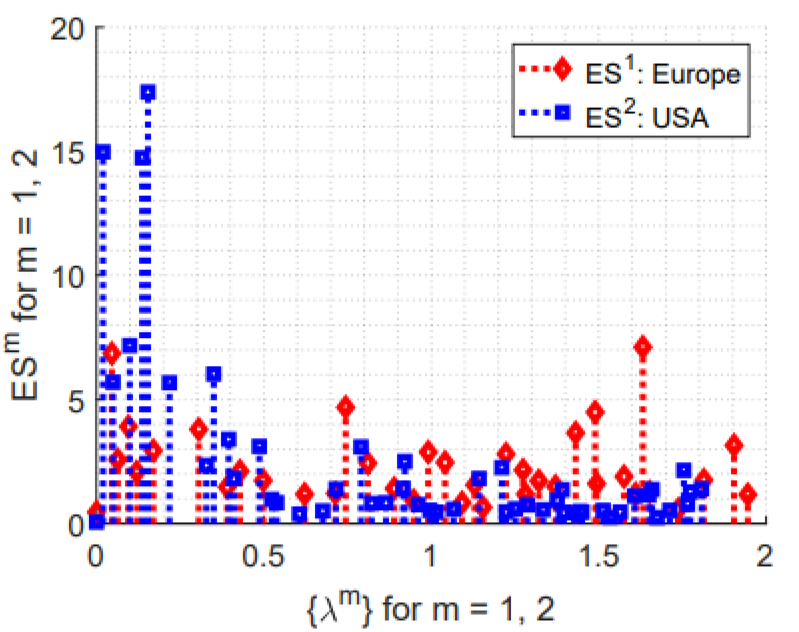


(b) G[2]-US

Experiments: Energy Spectrum of Datasets



Molene



Covid

OUTLINE

- Foundations of GSP
- Literature
- Proposed Method (SGKL)
    - Novelties
    - Signal Model, Objective function
    - Algorithm
- **Experiments**
    - Comparative
    - **Performance and Sensitivity Analysis**
- Conclusion

Experiments: Performance and Sensitivity Analysis

Effect of:

- Noise Level
- Learning different numbers of spectral kernels
- Ground-truth sparsity
- Learning
    - Jointly on multiple graphs or
    - Independently on each single graph
- Sensitivity analysis of the hyperparameters
    - η_ω
    - η_y
    - η_c
    - η_x
    - η_s

Experiments: Noise Level

$$NMSE^m = \left\| Y_u^m - \tilde{Y}_u^m \right\|^2 / \|Y_u^m\|^2$$

SNR, dB	$\mathcal{G}^1$	$\mathcal{G}^2$
-8	1.61	1.49
-6	1.38	1.31
-5	1.02	1.02
-3	0.61	0.59
-1	0.43	0.42
6	0.13	0.12
9	0.01	0.01
15	0.003	0.004
21	$4{\cdot}10^{-4}$	$5{\cdot}10^{-4}$
35	$4{\cdot}10^{-4}$	$4{\cdot}10^{-4}$

NMSE vs SNR for $\mathcal{G}^1$ and $\mathcal{G}^2$

$$NMSE^m = \left\| Y_u^m - \tilde{Y}_u^m \right\|^2 / \| Y_u^m \|^2$$

Experiments: Number of Kernels

J	$\mathcal{G}^1$	$\mathcal{G}^2$
1	0.0027	0.0016
2	$4.38 \cdot 10^{-4}$	$3.71 \cdot 10^{-4}$
3	$3.36 \cdot 10^{-4}$	$2.53 \cdot 10^{-4}$
4	$4.06 \cdot 10^{-4}$	$3.6 \cdot 10^{-4}$
5	0.0039	0.0051

NMSE vs J for $\mathcal{G}^1$ and $\mathcal{G}^2$

$$NMSE^m = \left\|Y_u^m - \tilde{Y}_u^m\right\|^2 / \|Y_u^m\|^2$$

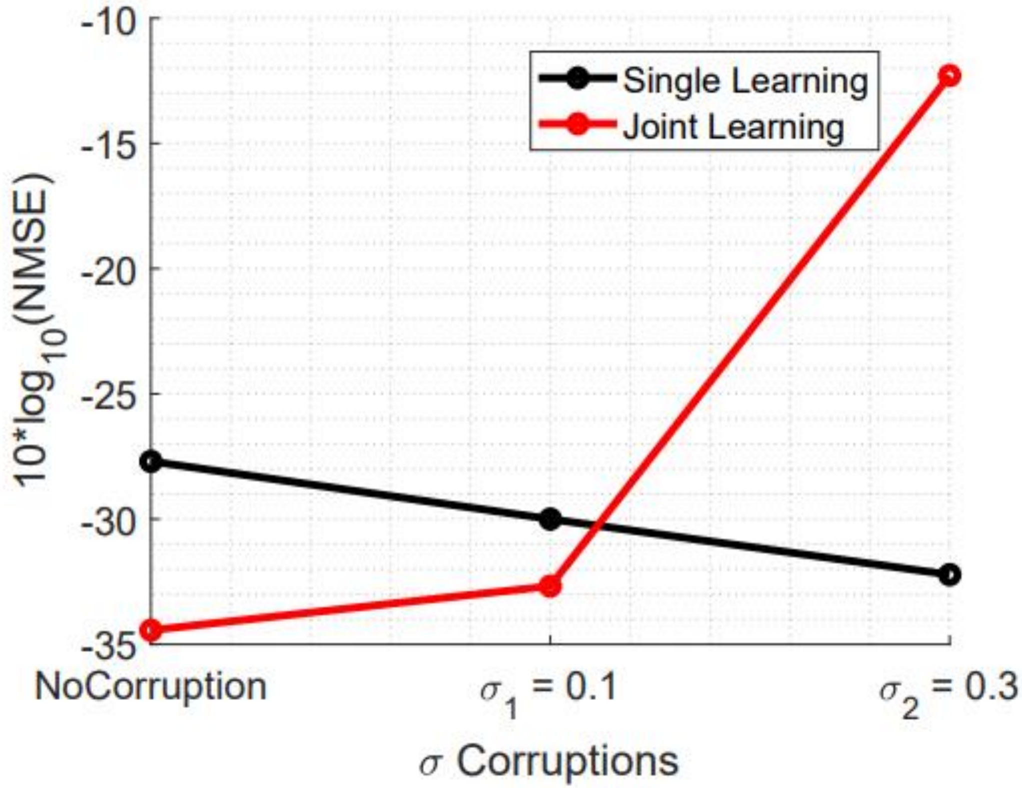
Experiments: Ground-truth sparsity

$T_{0,GT}$	$\mathcal{G}^1$	$\mathcal{G}^2$
10	$62 \cdot 10^{-4}$	$58 \cdot 10^{-4}$
20	$17 \cdot 10^{-4}$	$18 \cdot 10^{-4}$
30	$18 \cdot 10^{-4}$	$22 \cdot 10^{-4}$
40	$4.06 \cdot 10^{-4}$	$3.6 \cdot 10^{-4}$
50	$4 \cdot 10^{-4}$	$5 \cdot 10^{-4}$
60	$14 \cdot 10^{-4}$	$15 \cdot 10^{-4}$
70	$43 \cdot 10^{-4}$	$47 \cdot 10^{-4}$
80	$25 \cdot 10^{-4}$	$27 \cdot 10^{-4}$
90	$33 \cdot 10^{-4}$	$29 \cdot 10^{-4}$
100	$25 \cdot 10^{-4}$	$25 \cdot 10^{-4}$

NMSE vs $T_{0,GT}$ for $\mathcal{G}^1$ and $\mathcal{G}^2$

$$NMSE^m = \left\| Y_u^m - \tilde{Y}_u^m \right\|^2 / \|Y_u^m\|^2$$

Experiments: Joint vs Independent Learning



Single vs joint learning of graph dictionaries in the case of mismatches in the ground-truth kernels.

Experiments: η_y

$$NMSE^m = \left\|Y_u^m - \tilde{Y}_u^m\right\|^2 / \|Y_u^m\|^2$$

η_y	$\mathcal{G}^1$	$\mathcal{G}^2$
10^0	0.0042	0.0052
10^1	0.0091	0.010
10^2	$4.06 \cdot 10^{-4}$	$3.6 \cdot 10^{-4}$
10^3	0.0034	0.041
10^4	0.013	0.014
10^5	0.022	0.022

NMSE vs η_y for $\mathcal{G}^1$ and $\mathcal{G}^2$

$$\eta_y \sum_{m=1}^{M} ((X^m)^T (D^m)^T \mathcal{L}^m D^m X^m)$$

Experiments: η_c

$$NMSE^m = \left\|Y_u^m - \tilde{Y}_u^m\right\|^2 / \|Y_u^m\|^2$$

η_c	$\mathcal{G}^1$	$\mathcal{G}^2$
10^0	0.01	0.01
10^1	0.006	0.006
10^2	0.002	0.002
10^3	$4.06 \cdot 10^{-4}$	$3.6 \cdot 10^{-4}$
10^4	0.003	0.002
10^5	0.003	0.003

NMSE vs η_c for $\mathcal{G}^1$ and $\mathcal{G}^2$

$$\eta_c \sum_{m=1}^{M} ((X^m)\tilde{\mathcal{L}}^m(X^m)^T)$$

OUTLINE

- Foundations of GSP
- Literature
- Proposed Method (SGKL)
  - Novelties
  - Signal Model, Objective function
  - Algorithm
- Experiments
  - Comparative
  - Performance and Sensitivity Analysis
- **Conclusion**

Conclusion

- We propose a graph signal model based on representations using narrowband graph kernel prototypes.
- The initially unknown observations of the signals are then estimated based on this learnt model.
- Experiments on synthetic and real graph signal sets show that the proposed method provides promising signal estimation performance compared to baseline solutions.

Conclusion: Future Remarks

- Exploring more challenging scenarios with missing entries on the graph in regional, grid, or patterned forms.

- Developing approximate solutions for faster computation and enhanced scalability.

Thank You