Dictionary Design for Graph Signal Processing

David Shuman

January 15, 2019

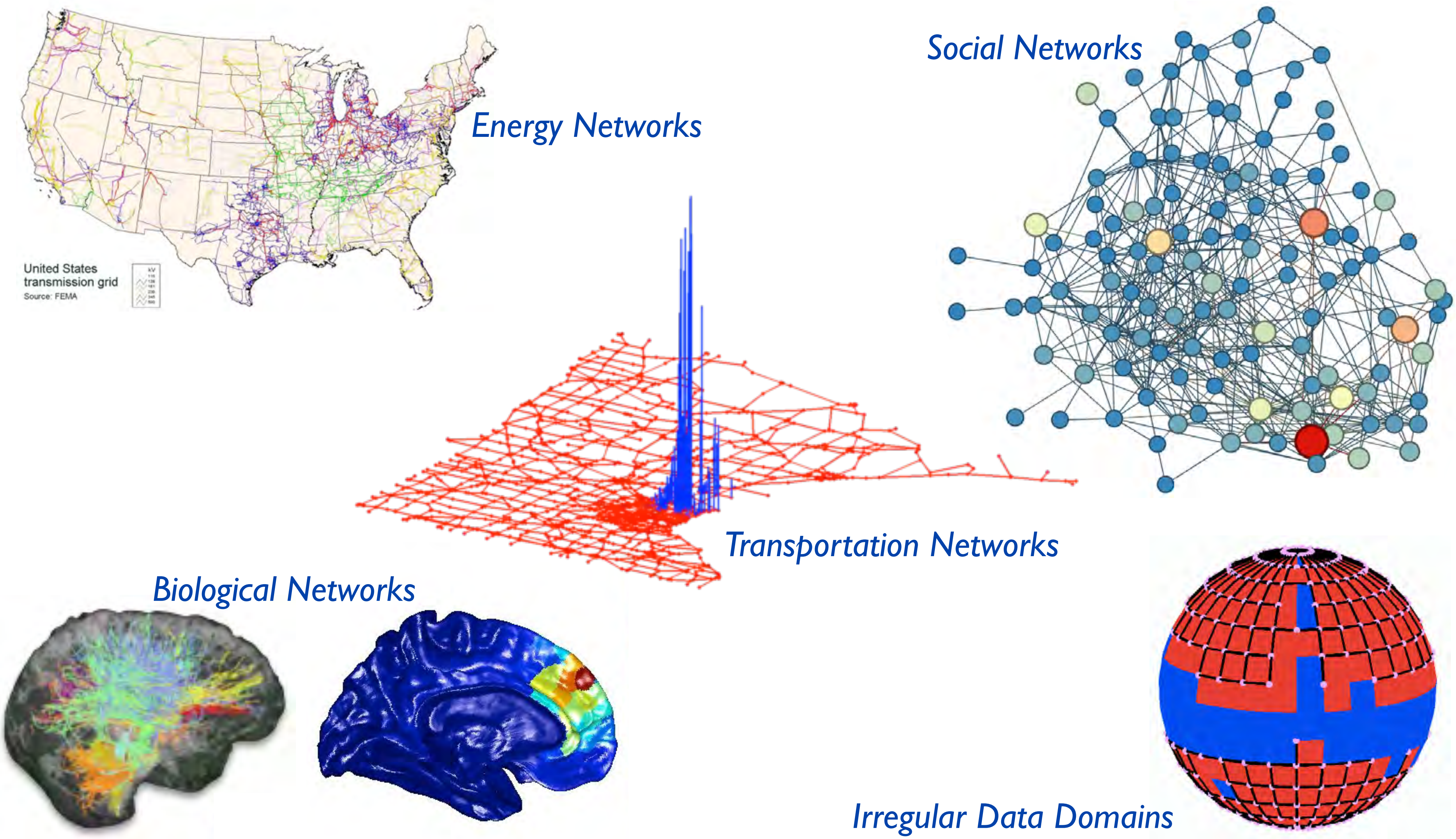
Applied Mathematics Seminar

Yale University

Special thanks and acknowledgement to my collaborators:
Andre Archer, Sam Armon, Andrew Beveridge, Xiaowen Dong, Tiffany Fan, Mohammad Javad Faraji, Stefan Faridani, Pascal Frossard, Tom Halverson, Nicki Holighaus, Jason McEwen, Daniel Kressner, Yan Jin, Shuni Li, Zhaoqi Li, Sunil Narang, Antonio Ortega, Javier Pérez-Trufero, Nathanaël Perraudin, Benjamin Ricaud, Dorina Thanou, Shashanka Ubaru, Pierre Vandergheynst, Elle Weeks, and Christoph Wiesmeyr

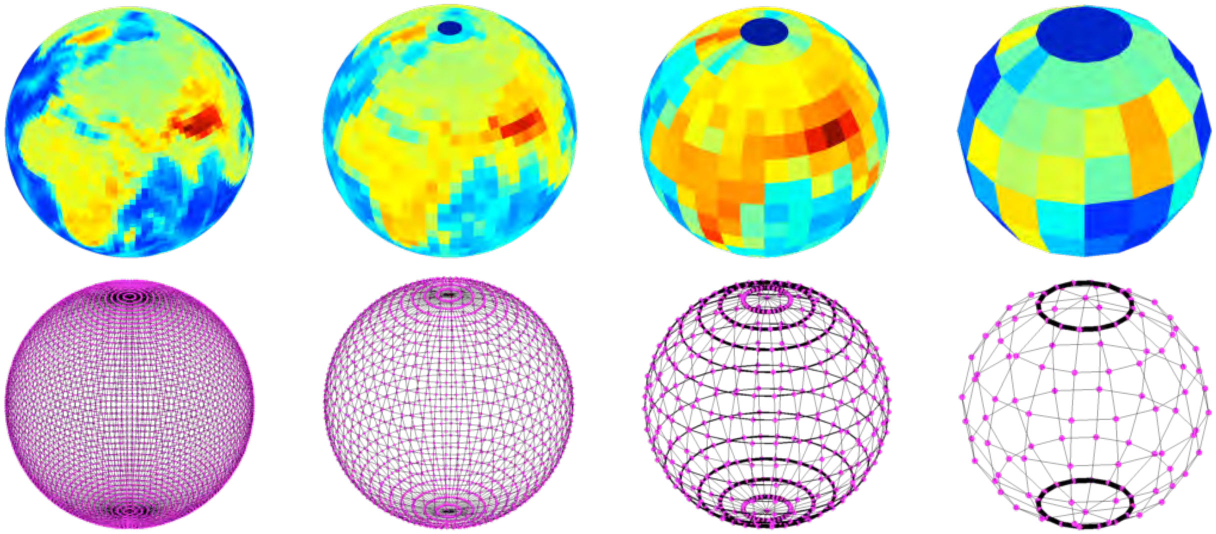
MACALESTER

Signal Processing on Graphs

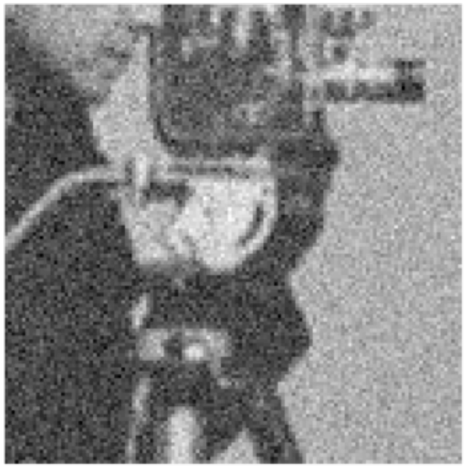


Some Typical Graph Signal Processing Problems

Compression / Visualization

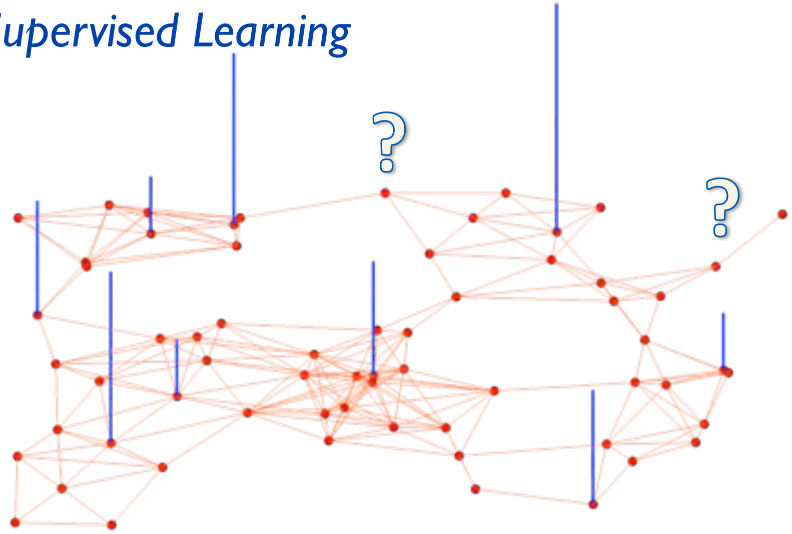


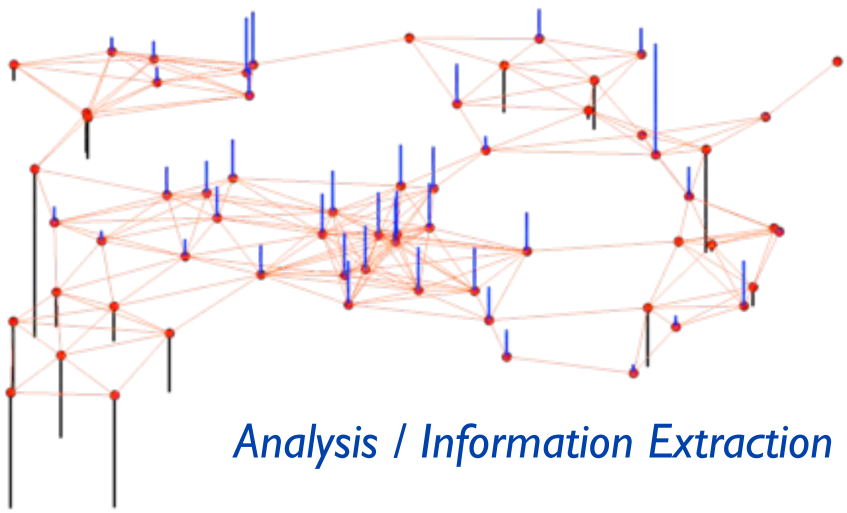
Earth data source: Frederik Simons



Denoising

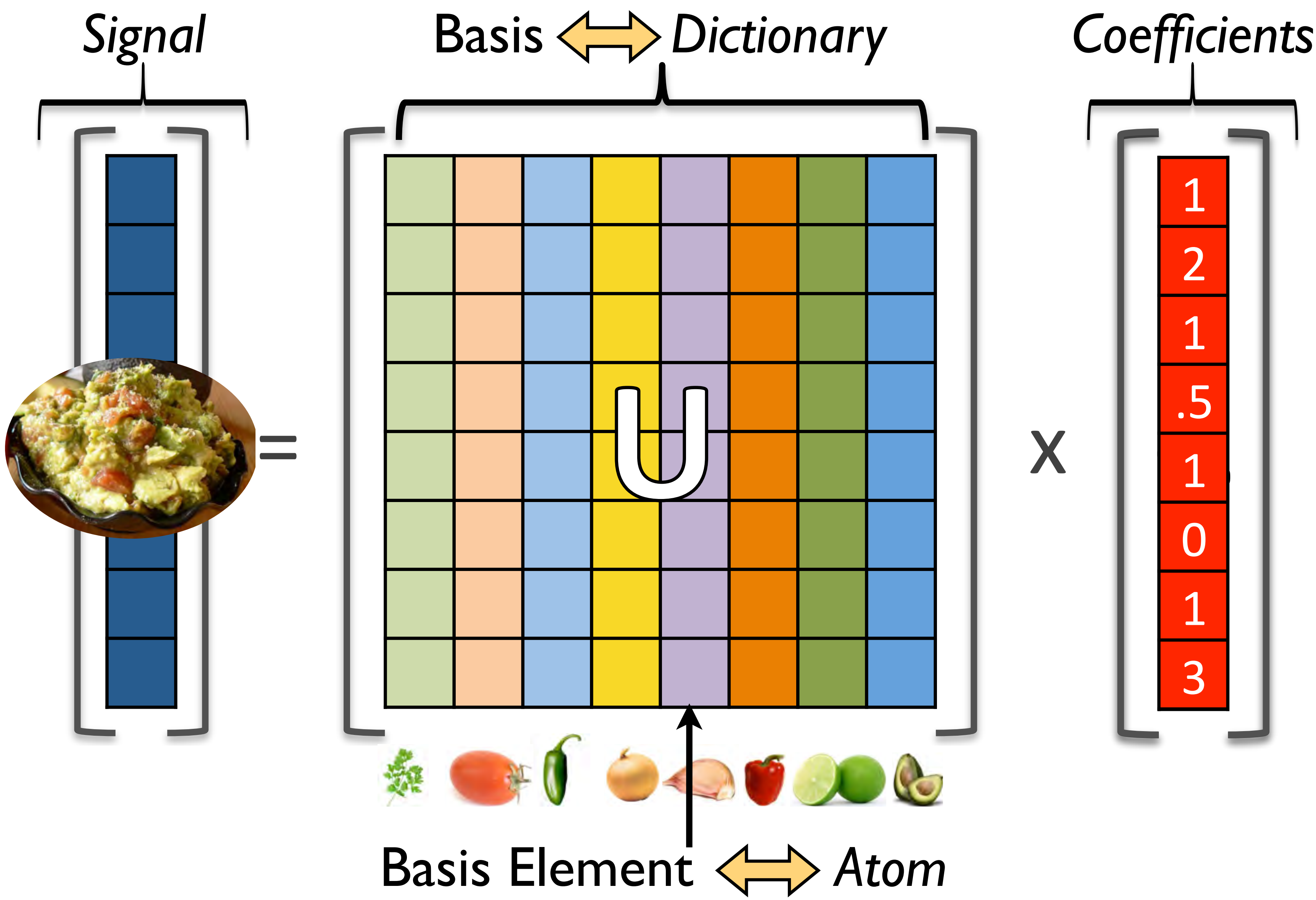
Semi-Supervised Learning



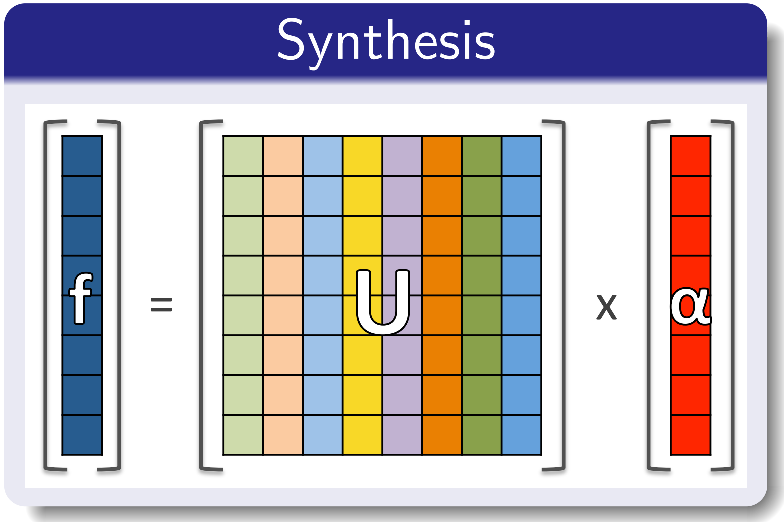


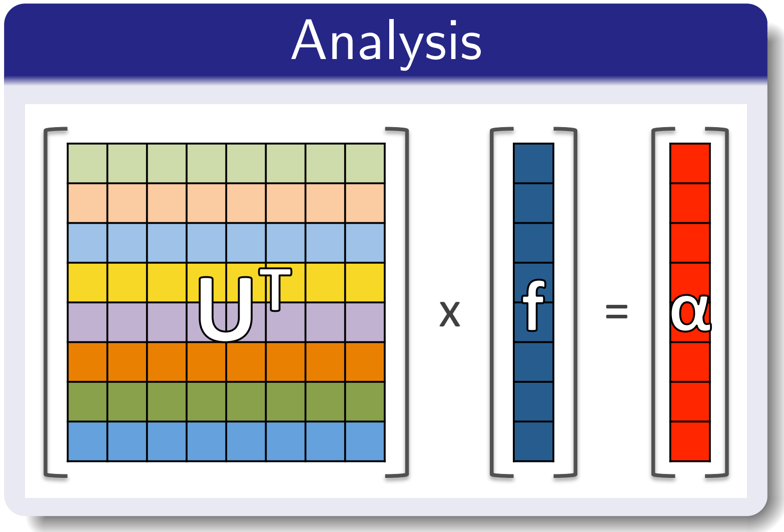
Analysis / Information Extraction

Orthonormal Dictionaries



Orthonormal Dictionaries (cont.)





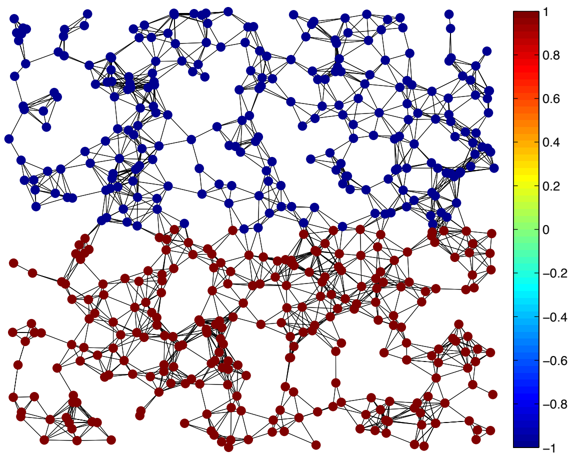
$$f = \sum_{\ell} \alpha_{\ell} u_{\ell} = \sum_{\ell} \langle f, u_{\ell} \rangle u_{\ell}$$

Overcomplete Dictionaries and Sparsity

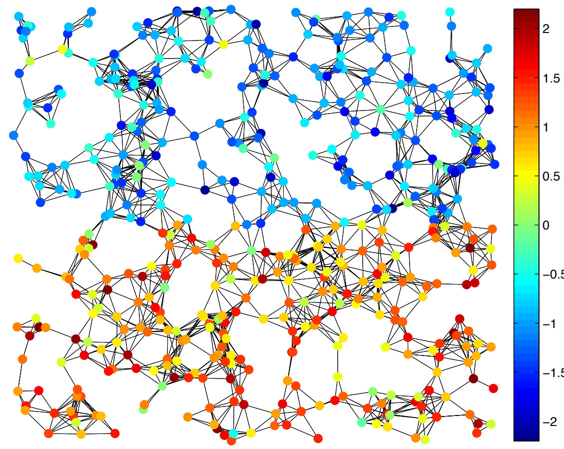
- Given an overcomplete Φ, there are infinitely many choices of α that lead to the same signal f
- Useful to *sparsely* represent signals —> few non-zero coefficients in α

Motivating Example: Denoising

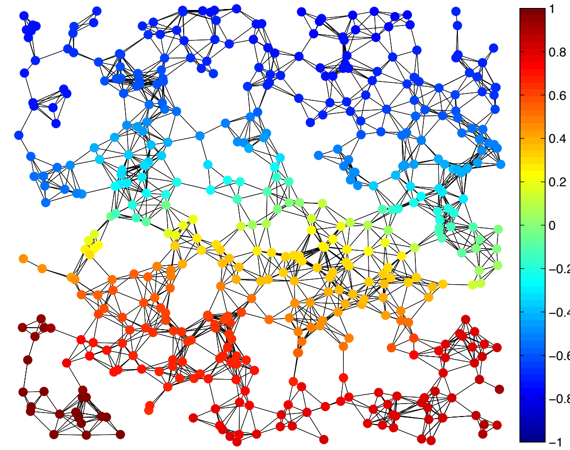
- Tikhonov regularization for denoising: $\operatorname{argmin}_f \left\{ ||f-y||_2^2 + \gamma f^T \mathcal{L} f \right\}$



Original

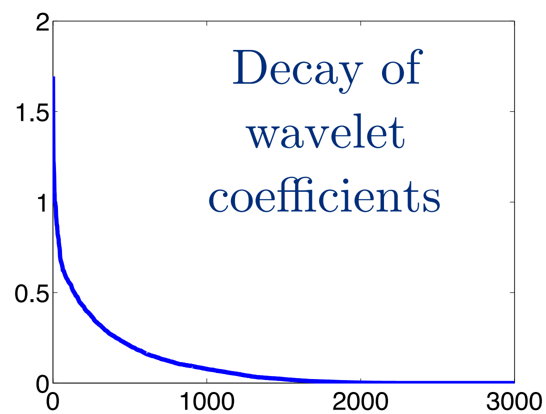


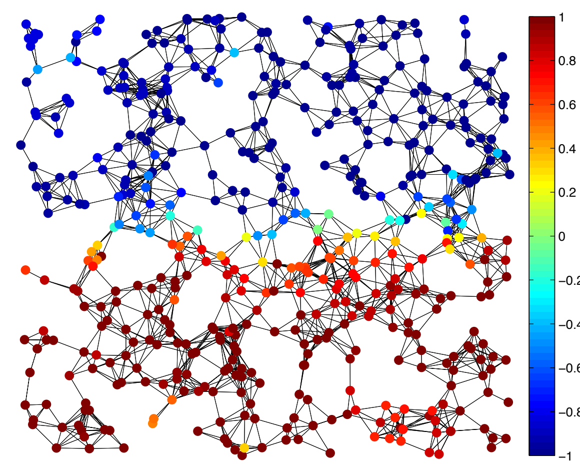
Noisy



Denoised

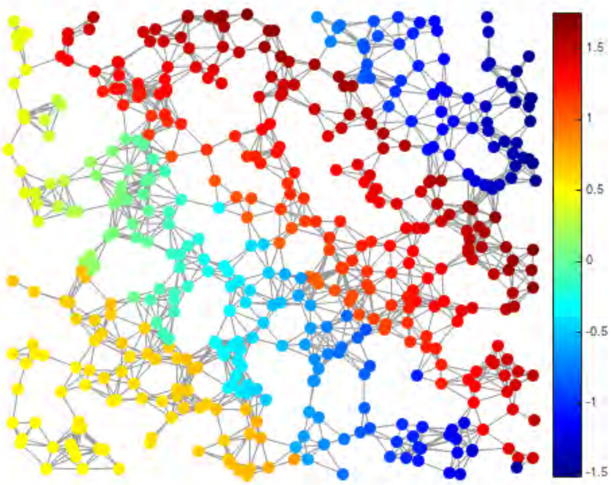
- Wavelet denoising: $\operatorname{argmin}_a \left\{ ||f-W^*a||_2^2 + \gamma ||a||_{1,\mu} \right\}$



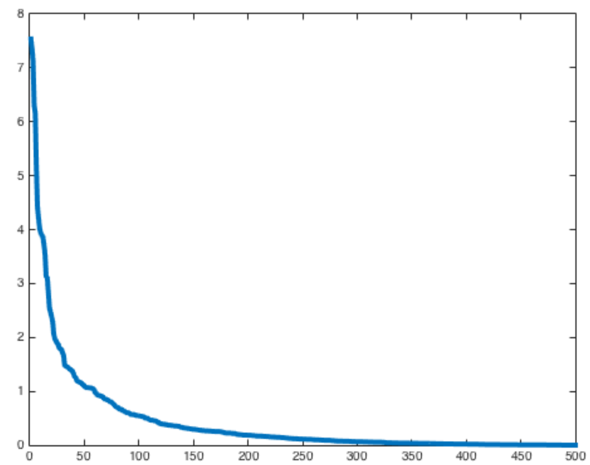


Motiving Example: Compression

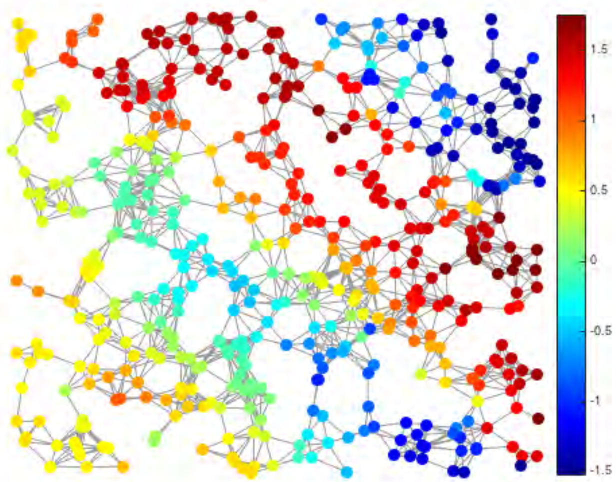
Piecewise-Smooth Signal with Discontinuities



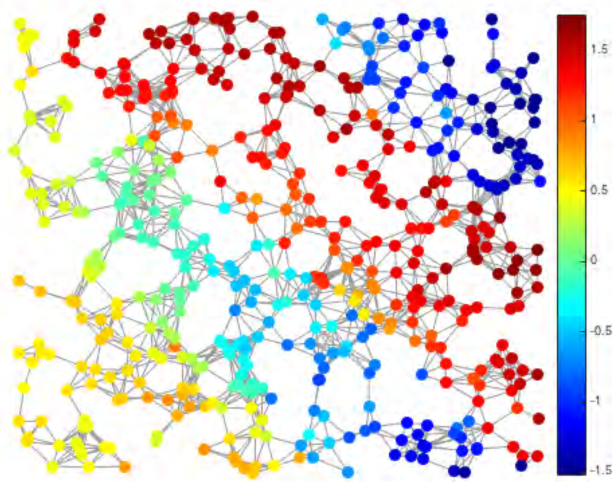
Diffusion Wavelet Coefficients, Sorted by Magnitude



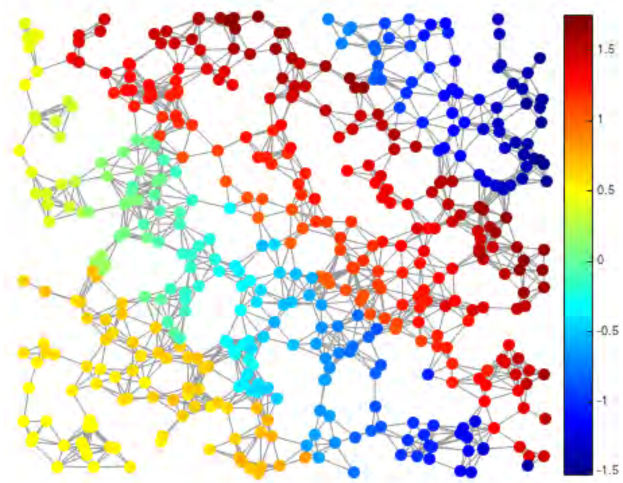
Reconstruction from 10% of Coefficients



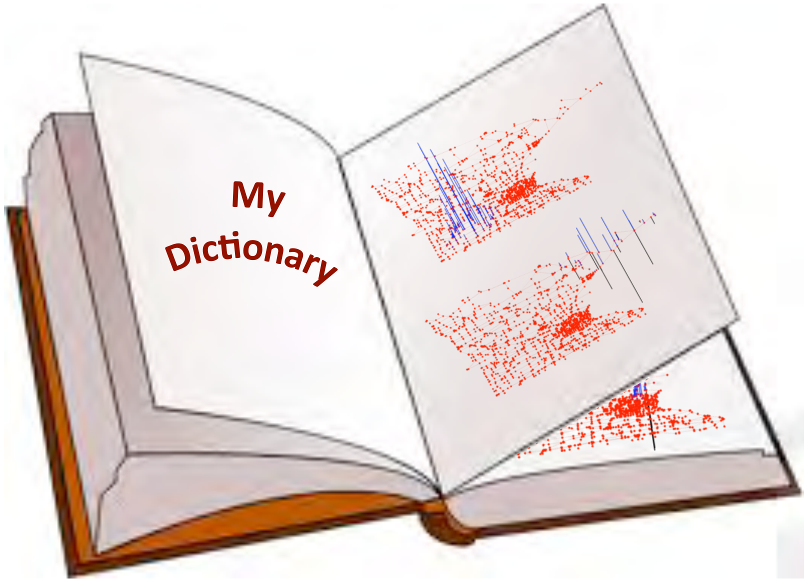
Reconstruction from 20% of Coefficients



Reconstruction from 50% of Coefficients



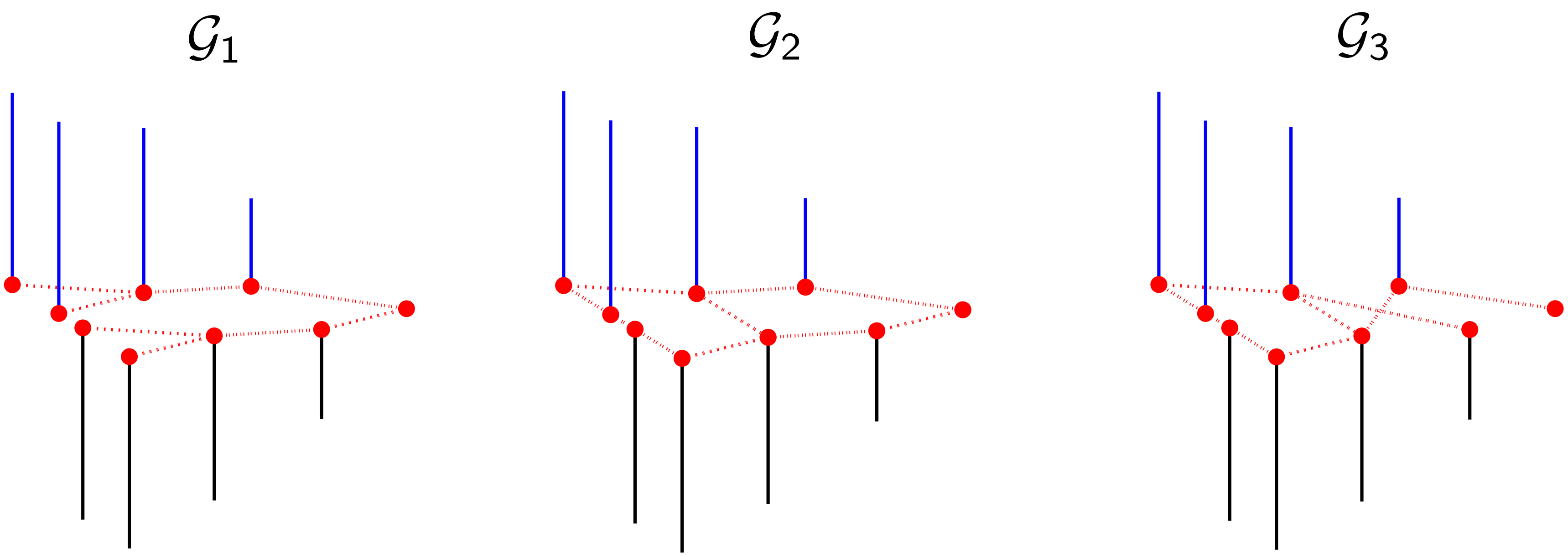
Dictionary Design for Signals on Graphs



Desirable Characteristics

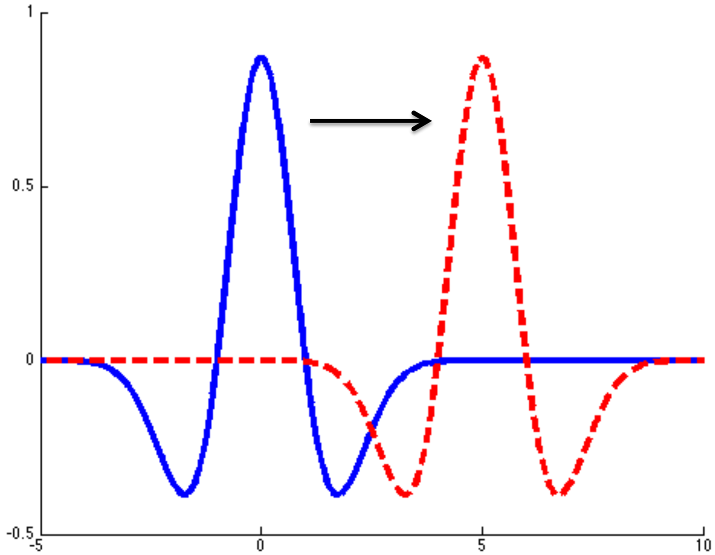
- Ability to *sparsely* represent signals — few non-zero coefficients in α
- Ability to capture the relevant characteristics of signals to extract information
- Computationally efficient to apply Φ and Φ^T
- Tight frames

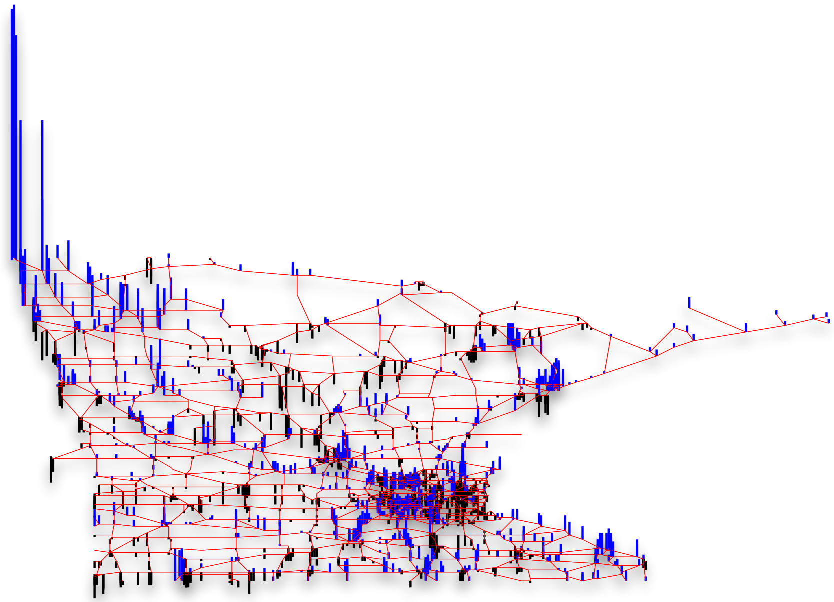
Why Do We Need New Dictionaries?



To identify and exploit structure in the data, we need to account for the intrinsic geometric structure of the underlying graph data domain

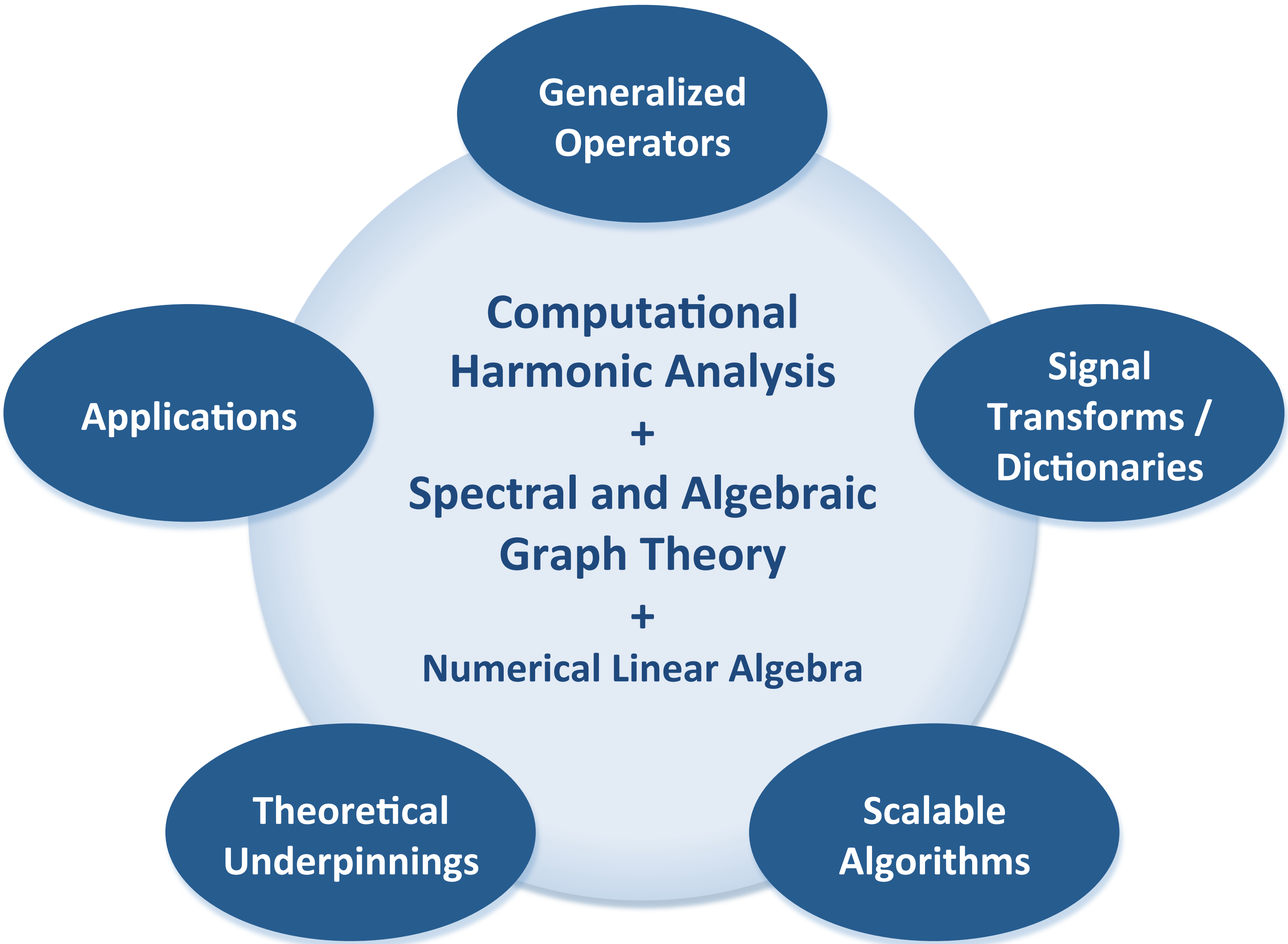
The Essence of the Problem





- Weighted graphs are irregular structures that lack a shift-invariant notion of translation
- Many simple yet fundamental concepts that underlie classical signal processing techniques become significantly more challenging in the graph setting

Approach: Leverage Intuition from Euclidean Settings to Develop New Mathematical Tools for the Graph Setting



Generalized Operators

Combinatorial Graph Laplacian

- Connected, undirected, weighted graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, W\}$

- Degree matrix D: zeros except diagonals, which are sums of weights of edges incident to corresponding node

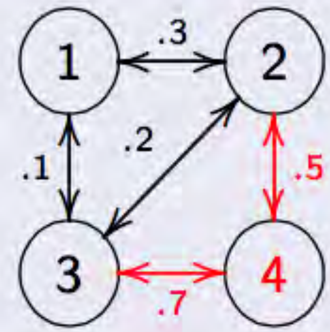
- Non-normalized graph Laplacian: $\mathcal{L} := D - W$

- Complete set of orthonormal eigenvectors and associated real, non-negative eigenvalues:

$$\mathcal{L}u_\ell = \lambda_\ell u_\ell,$$

ordered w.l.o.g. s.t.

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 ... \leq \lambda_{N-1} := \lambda_{\max}$$



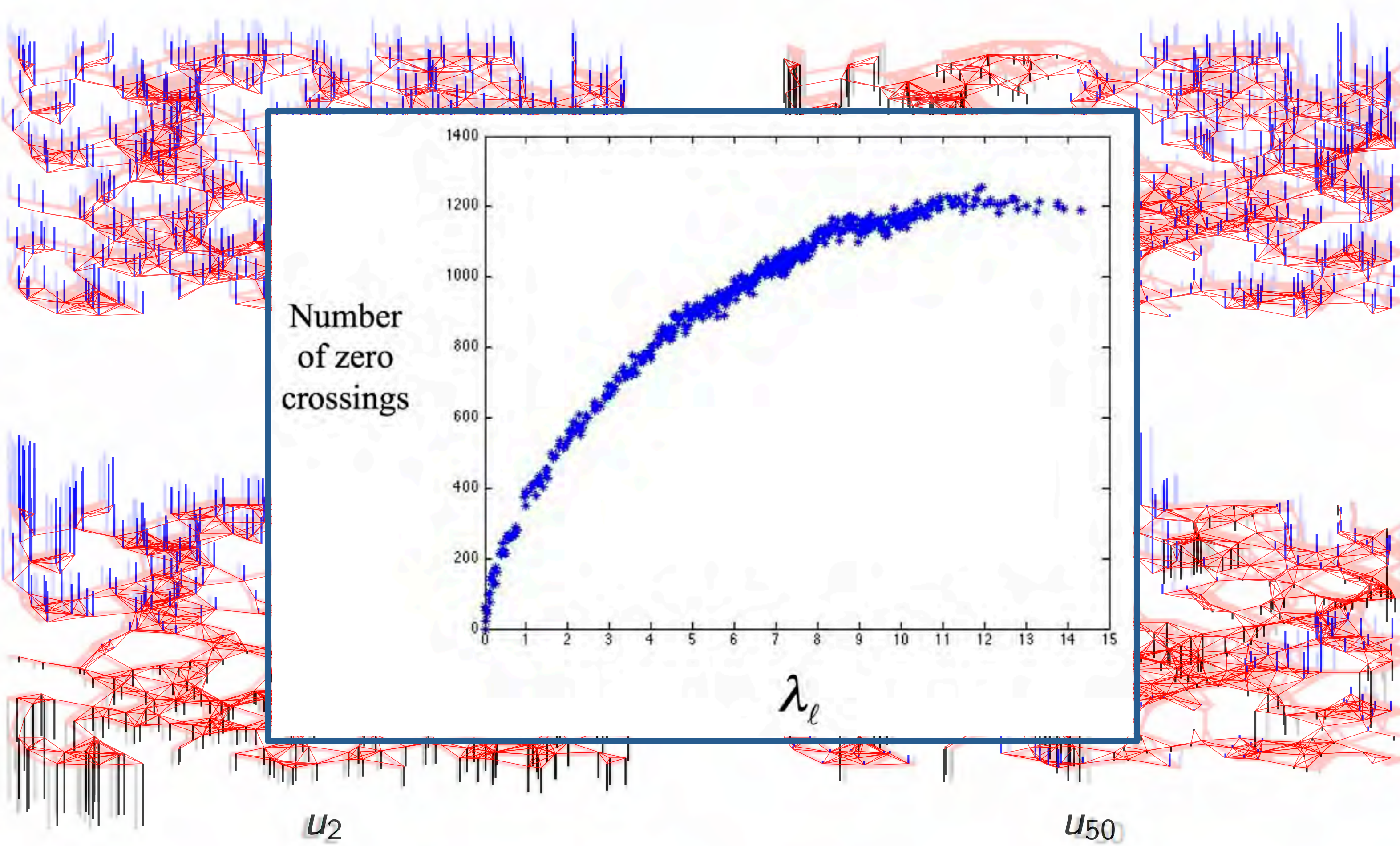
$$W = \begin{bmatrix} 0 & .3 & .1 & 0 \\ .3 & 0 & .2 & .5 \\ .1 & .2 & 0 & .7 \\ 0 & .5 & .7 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} .4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1.2 \end{bmatrix}$$

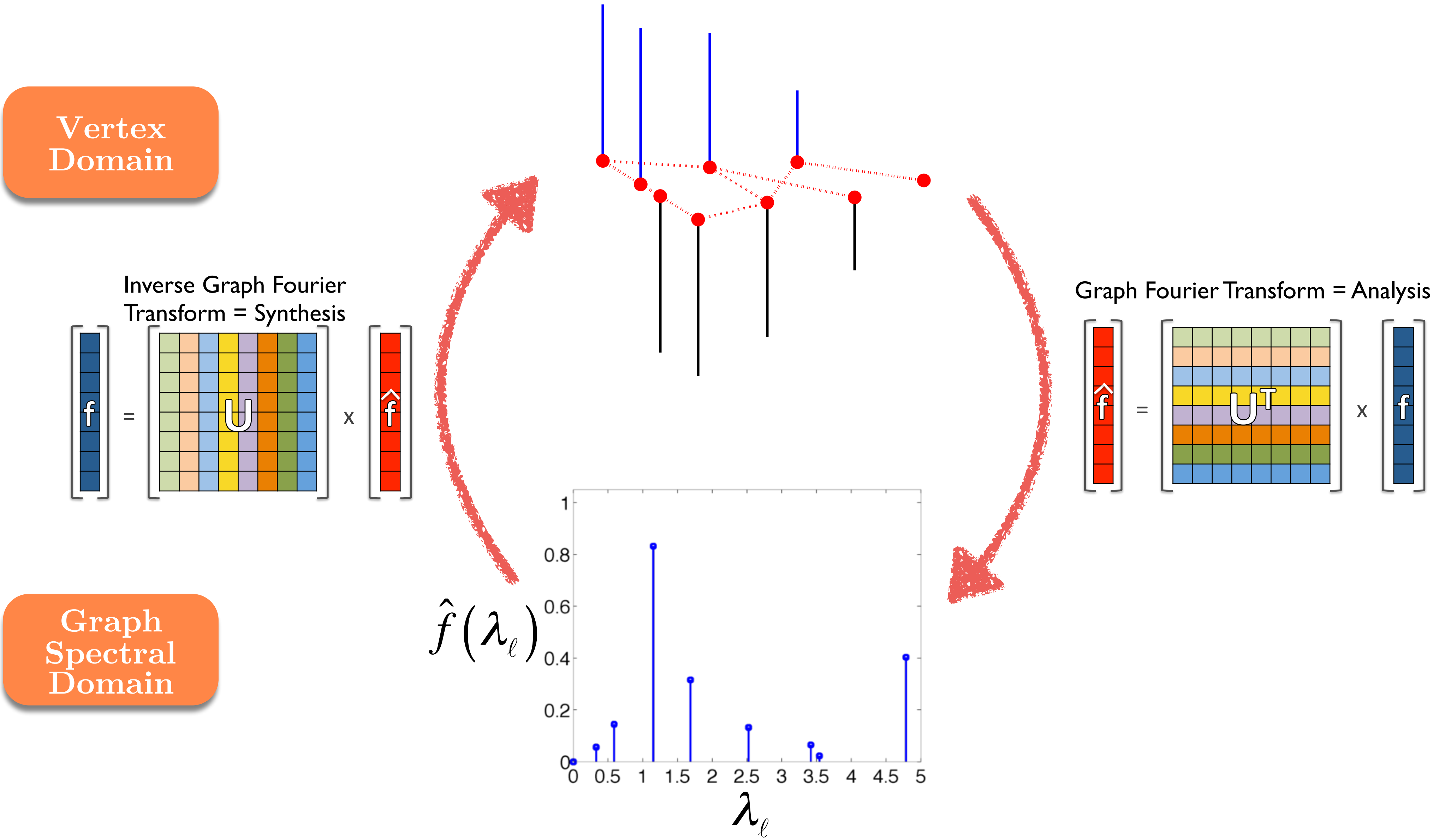
- Discrete difference operator: $(\mathcal{L}f)(i) = \sum_{j \in \mathcal{N}_i} W_{i,j}[f(i) - f(j)]$

Graph Fourier Transform

- Graph Laplacian eigenvectors are the analog of complex exponentials: Values of the eigenvectors associated with low eigenvalues change less rapidly across connected vertices
- Different choices of graph Fourier basis include combinatorial/normalized/random walk Laplacian eigenbasis or generalized eigenbasis of adjacency matrix

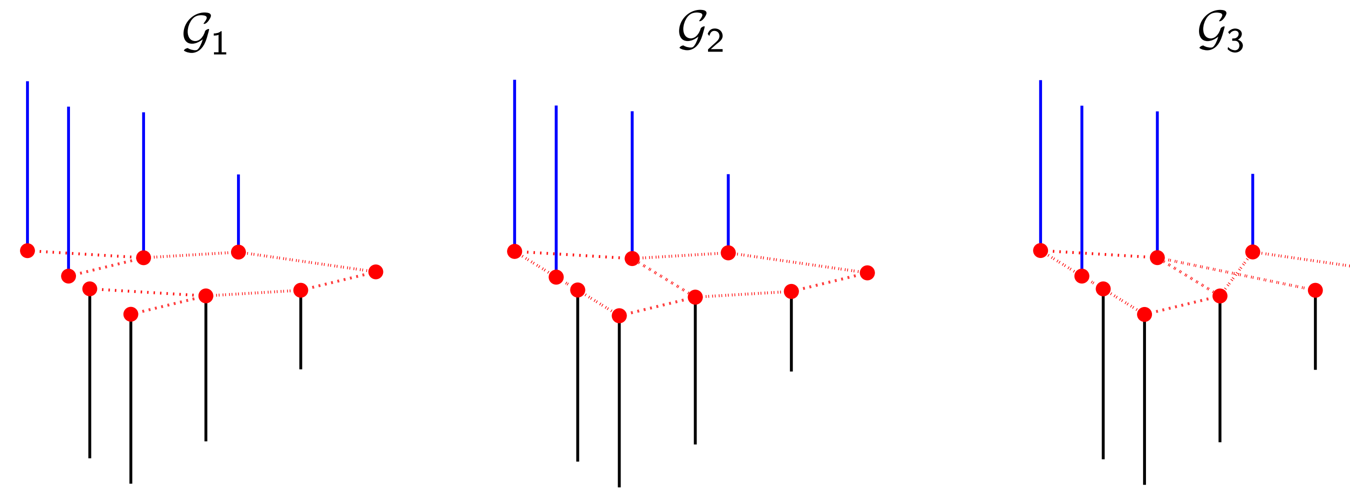


The GFT Incorporates the Graph Structure

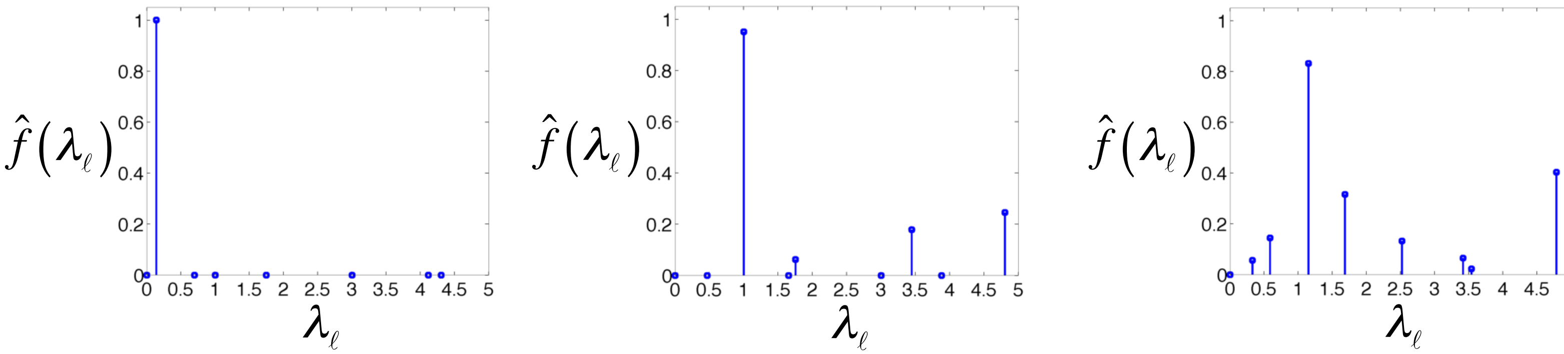


The GFT Incorporates the Graph Structure

Vertex Domain

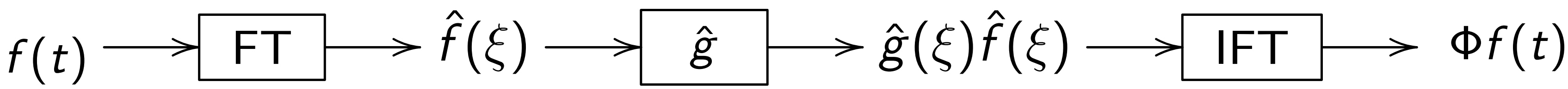


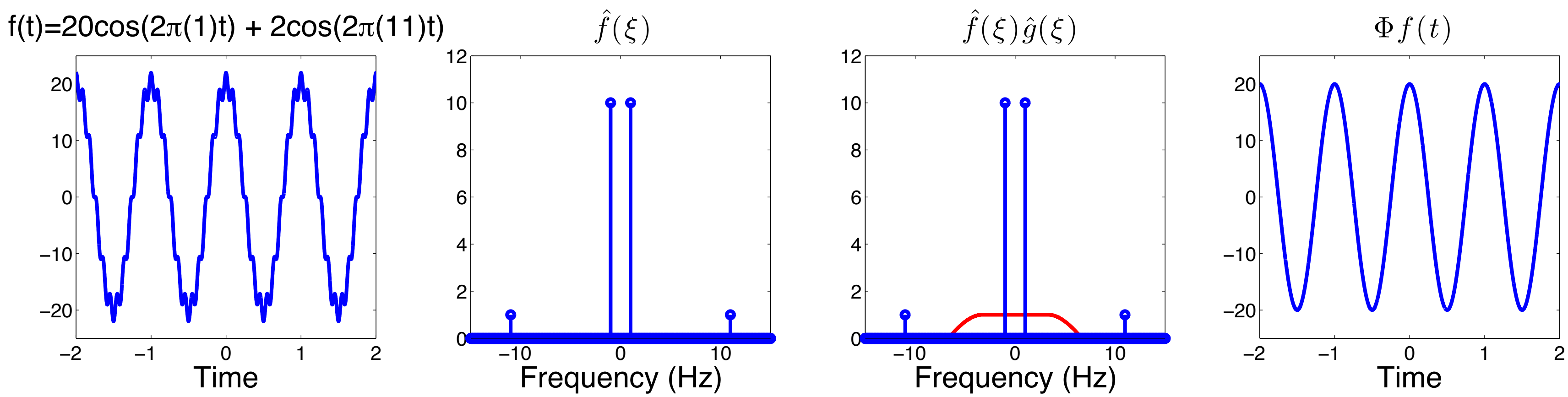
Graph Spectral Domain



Graph Spectral Filtering

- Filtering: represent an input signal as a combination of other signals, and amplify or attenuate the contributions of some of the component signals
- In classical signal processing, the most common choice of basis the complex exponentials, which results in frequency filtering

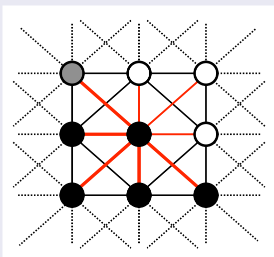




Example: Image Denoising by Low-Pass Graph Filtering

$$f(n) \longrightarrow \boxed{\text{GFT}} \longrightarrow \hat{f}(\lambda_\ell) \longrightarrow \boxed{\hat{g}} \longrightarrow \hat{g}(\lambda_\ell)\hat{f}(\lambda_\ell) \longrightarrow \boxed{\text{IGFT}} \longrightarrow \Phi f(n)$$

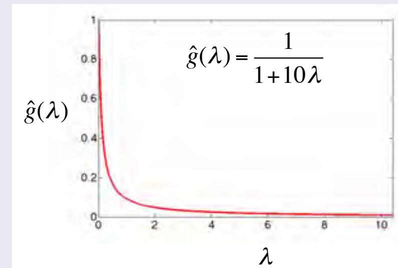
Semi-Local Graph



Tikhonov Regularization

$$\underset{f}{\text{argmin}} \left\{ \|f - y\|_2^2 + \gamma f^{\mathrm{T}} \mathcal{L} f \right\}$$

$$\Longrightarrow \hat{g}(\lambda_\ell) = \frac{1}{1 + \gamma \lambda_\ell}$$



Original Image | Noisy Image | Gaussian-Filtered (Std. Dev. = 1.5) | Gaussian-Filtered (Std. Dev. = 3.5) | Graph-Filtered

Shuman et al., The emerging field of signal processing on graphs, SPM, 2013

Approximating a Matrix Function Times a Vector

- Filtering: $g(\mathcal{L})f = Ug(\Lambda)U^*f$
- Too expensive to compute U and Λ for large graphs
- Common approach: estimate $\lambda_{\max}$ and approximate the filter g on the interval $[0, \lambda_{\max}]$ by a polynomial, rational, or spline function
- Example: Truncated Chebyshev polynomial approximation

$$g(\mathcal{L})f = \frac{1}{2}c_0 f + \sum_{k=1}^{\infty} c_k \bar{T}_k(\mathcal{L})f \approx \frac{1}{2}c_0 f + \sum_{k=1}^{K} c_k \bar{T}_k(\mathcal{L})f =: \tilde{g}(\mathcal{L})f$$

- Use the three-term recurrence relation to compute $\bar{T}_k(\mathcal{L})f$ from $\bar{T}_{k-1}(\mathcal{L})f$ and $\bar{T}_{k-2}(\mathcal{L})f$, at the cost of one sparse matrix-vector multiplication by $\mathcal{L}$
- Pros: Fast for large, sparse graphs $[\mathcal{O}(K|\mathcal{E}|)]$; convergence guarantees when the filter g is analytic/smooth; distributable

Druskin and Knizhnerman, "Two polynomial methods of calculating functions of symmetric matrices," 1989

Generalized Translation/Localization

- Define a generalized convolution by imposing that convolution in the vertex domain is multiplication in the graph spectral domain
- Define generalized translation via generalized convolution with a delta (i.e., filter a delta)

Functions on the Real Line

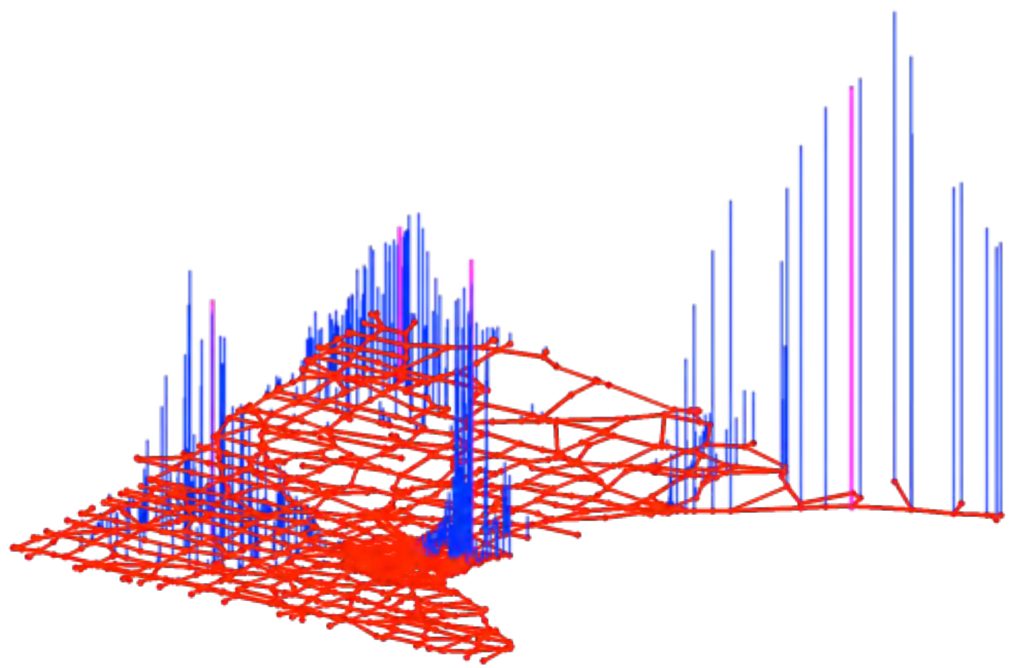
For $f \in L^2(\mathbb{R})$, in the weak sense

$$
\begin{aligned}
(T_s f)(t) &:= f(t-s) \\
&= (f * \delta_s)(t) \\
&= \int_{\mathbb{R}} \hat{f}(\xi) e^{-2\pi i \xi s} e^{2\pi i \xi t} d\xi
\end{aligned}
$$

Functions on the Vertices of a Graph

For $f \in \mathbb{R}^N$, we define

$$
\begin{aligned}
(T_i f)(n) &:= \sqrt{N}(f * \delta_i)(n) \\
&= \sqrt{N} \sum_{\ell=0}^{N-1} \hat{f}(\lambda_\ell) u_\ell^*(i) u_\ell(n)
\end{aligned}
$$



Properties of Generalized Translation/ Localization

- Warning 1: Do not have the group structure of classical translation:

$$T_i T_j \neq T_{i+j}$$

- Warning 2: Unlike the classical case, generalized translation operators are not unitary, so $\|T_i g\|_2 \neq \|g\|_2$ in general

- However, the mean is preserved: $\sum_n (T_i g)(n) = \sum_n g(n)$

Theorem (Smoothness of $\hat{g}$ leads to localization of $T_i g$ around vertex i)

Let $\hat{g} : [0, \lambda_{\max}] \to \mathbb{R}$ be a kernel and define $d_{in} := d_{\mathcal{G}}(i, n)$. Then

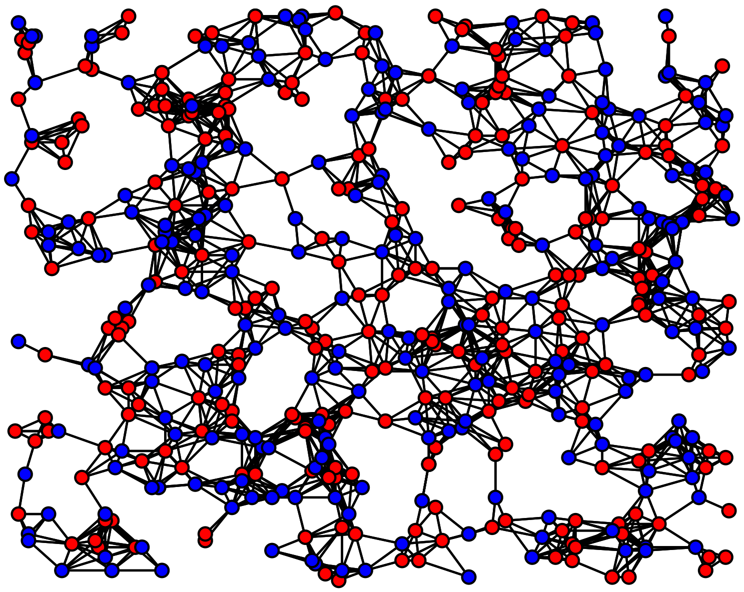
$$|(T_i g)(n)| \leq \sqrt{N} B_{\hat{g}}(d_{in} - 1),$$

where $B_{\hat{g}}(K)$ is the minimax polynomial approximation error over all polynomials of degree K:

$$B_{\hat{g}}(K) := \inf_{\widehat{p_K}} \left\{ \sup_{\lambda \in [0, \lambda_{\max}]} |\hat{g}(\lambda) - \widehat{p_K}(\lambda)| \right\}.$$

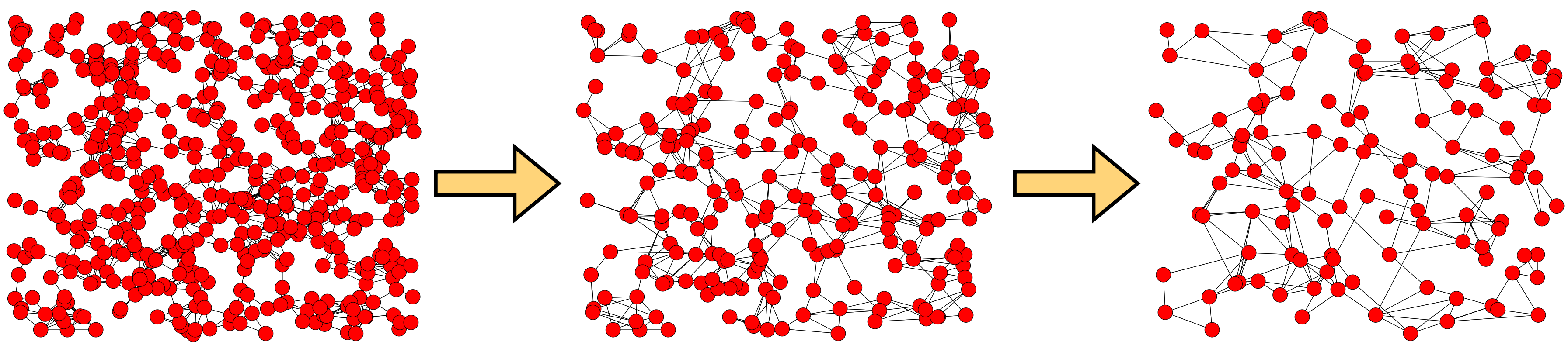
Downsampling and Graph Reduction

Downsampling



- Downsampling + graph reduction = a multiresolution of graphs
- Methods used here:
  - Graph downsampling by polarity of Laplacian eigenvector associated with largest eigenvalue
  - Kron reduction with spectral sparsification
- Alternative: coarse graining

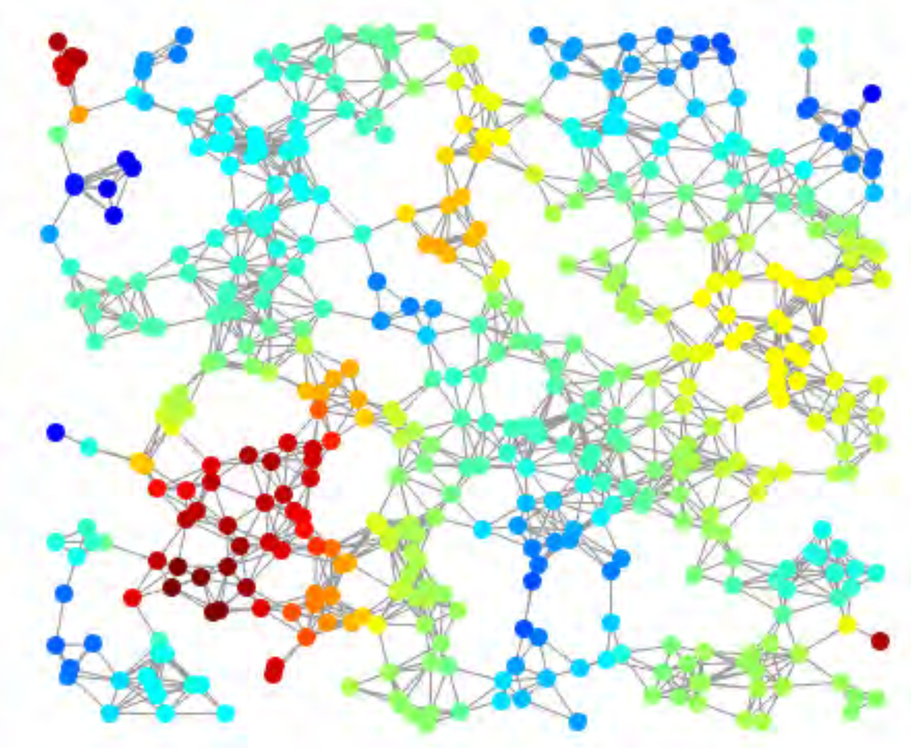
Graph Reduction



Sampling and Interpolation

- How to sample a graph signal and interpolate from the samples?
- How to choose the samples depends on your prior knowledge of the data
- Subset V_s of vertices is a <u>uniqueness set</u> for a subspace P iff:

  If two signals in the subspace P have the same values on the vertices in the uniqueness set, then they are the same signal

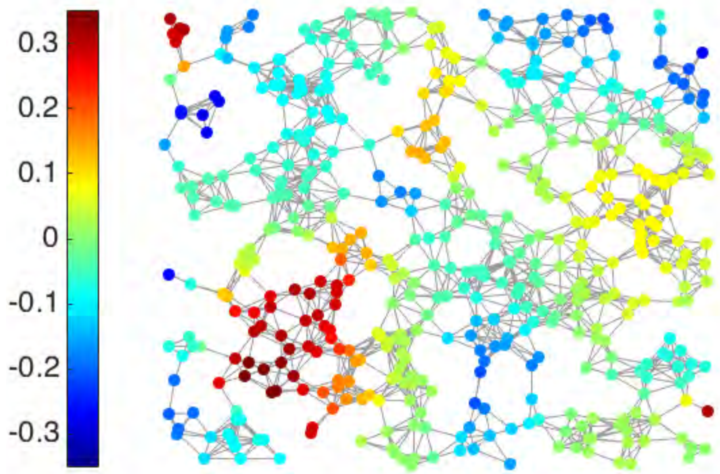


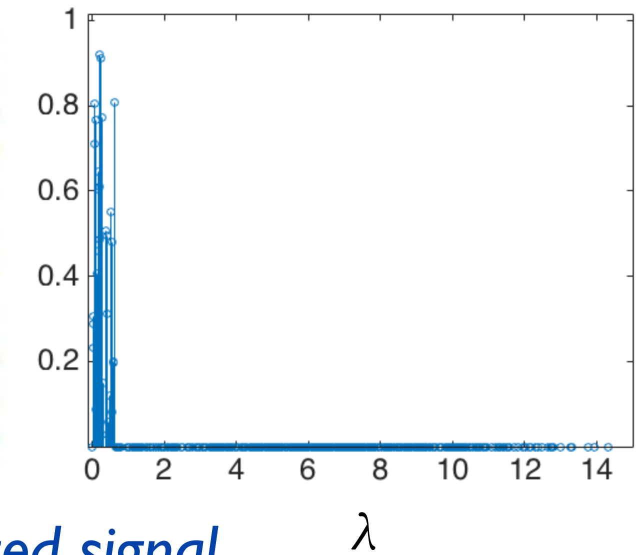
Can we recover all 500 values of this signal from 30 measurements? If so, where should we take those measurements?

Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," Trans. AMS, 2008

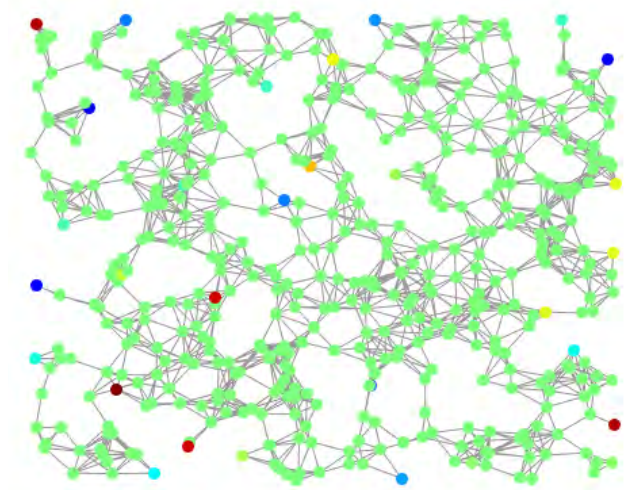
Sampling and Interpolation: Signals Concentrated on Spectral Bands

Example: subspace of globally smooth signals with band limit λ_{29}





Bandlimited signal



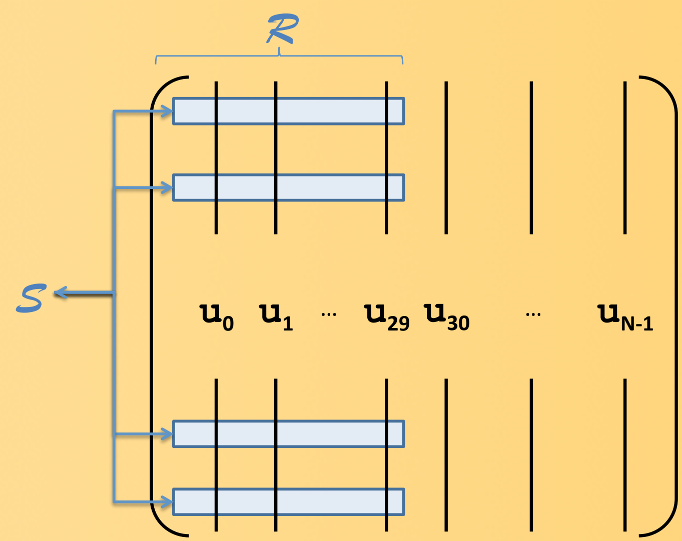
Uniqueness set

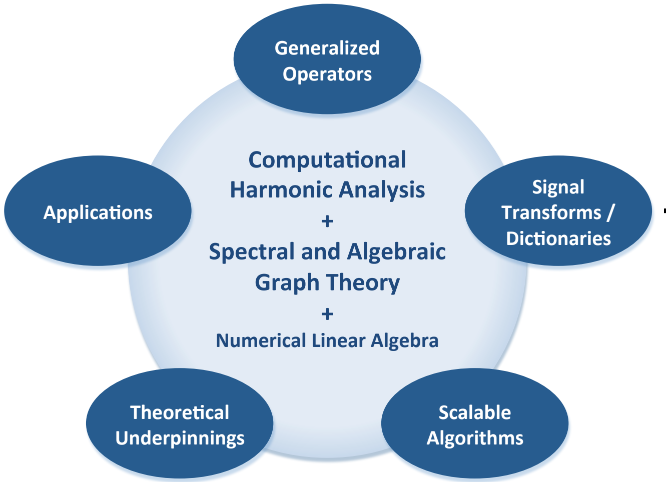
1. Recover graph Fourier coefficients:

$$U_{\mathcal{S},\mathcal{R}}x = f_{\mathcal{S}}$$

2. Interpolation / reconstruction:

$$\tilde{f} = U_{:,\mathcal{R}}x$$





Approaches to Graph Signal Dictionary Design

Analytic Versus Trained Dictionaries

Rubinstein et al., Dictionaries for sparse representation modeling, Proc. IEEE, 2010

- Analytic dictionaries: adapted to graph structure, but not to any specific training signals

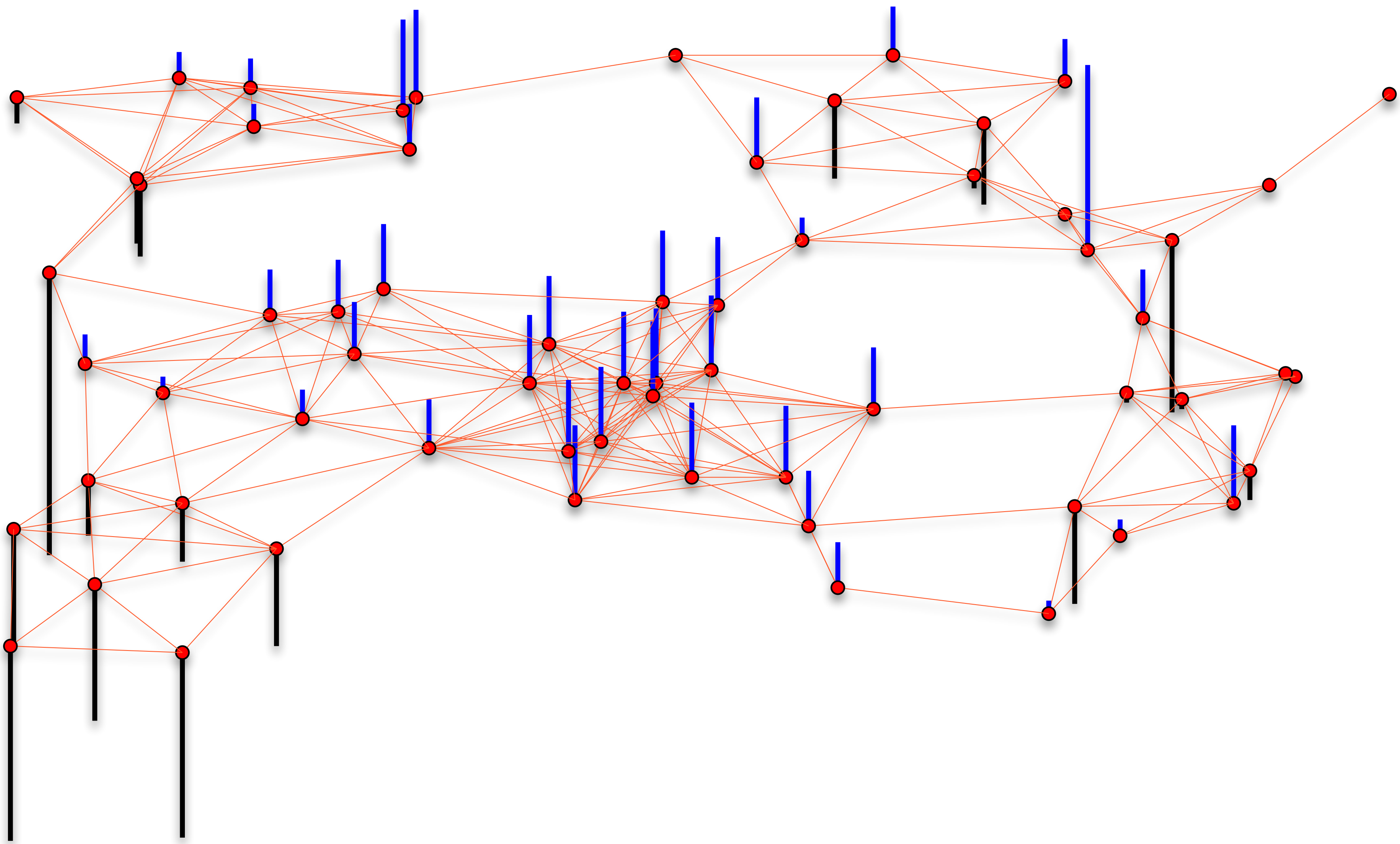
- Dictionary learning: adapt dictionary to training data
  - Aharon et al., The K-SVD, TSP, 2003
  - Engan et al., Method of optimal directions for frame design, ICASSP, 1999
  - These general methods do not explicitly account for graph structure

- Parametric training: force some structure upon the dictionary (e.g., to incorporate graph topology, ensure an efficient computational implementation), but use training signals to learn parameters

Survey of Approaches to Graph Signal Dictionary Design

- Graph Fourier transform
- Vertex domain designs
- Diffusion-based designs
- **Windowed graph Fourier transform**
- **Spectral domain designs**
- **Generalized filter banks**

Motivating Example: Any Structure?



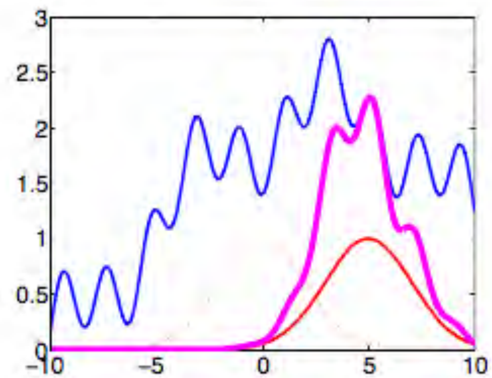
Classical Windowed Fourier Transform

- Localized Fourier analysis – joint descriptions of signals' temporal and spectral behavior

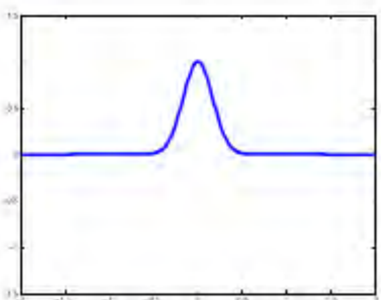
  - Localized oscillations appear frequently in audio processing, vibration analysis, radar detection, etc.
  - e.g., identify musical notes and melody at different times

- Windowed (short-time) Fourier transform of $f \in L^2(\mathbb{R})$:

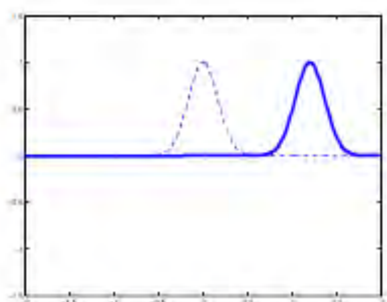
$$Sf(s,\xi) := \langle f, g_{s,\xi} \rangle = \int_{-\infty}^{\infty} f(t)\overline{g(t-s)}e^{-2\pi i \xi t}\,dt$$



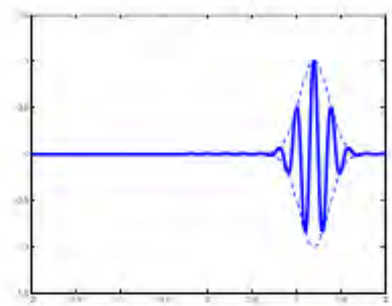
- The atoms $g_{s,\xi}$ are localized in time and frequency:



Translation T_s $\Longrightarrow$

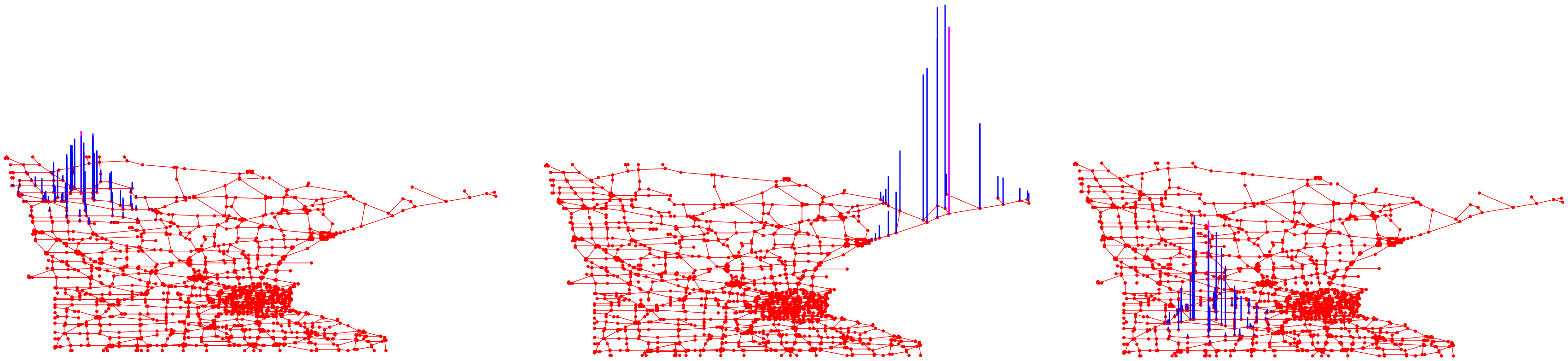


Modulation M_ξ $\Longrightarrow$



Windowed Graph Fourier Transform

1. Translate a window g to each vertex of the graph



$T_{100}g$ $T_{200}g$ $T_{2000}g$

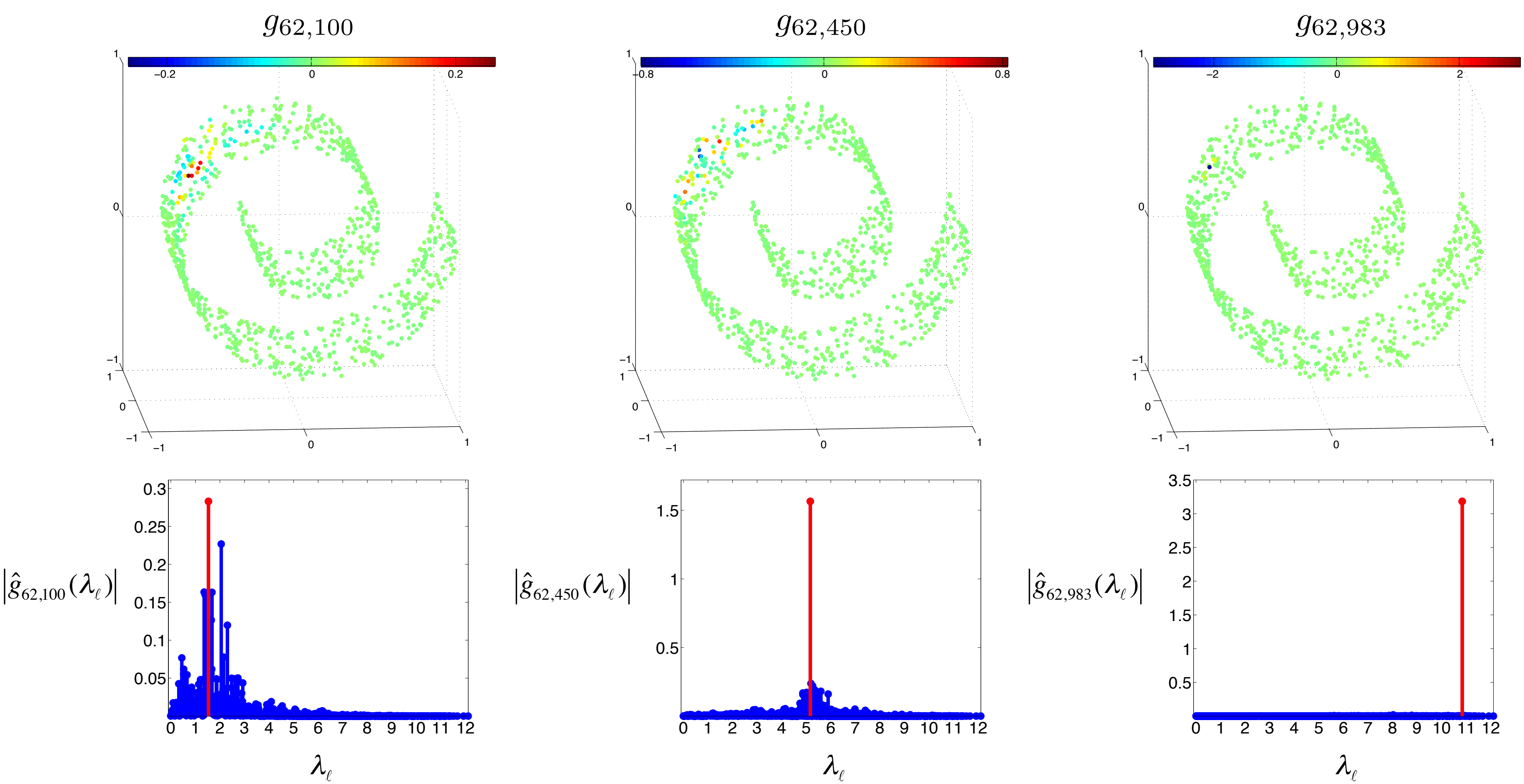
2. Multiply each component of the graph signal f of interest by the corresponding component of the translated window $T_i g$

3. Take the graph Fourier transform of $f .* T_i g$ (recall analysis)

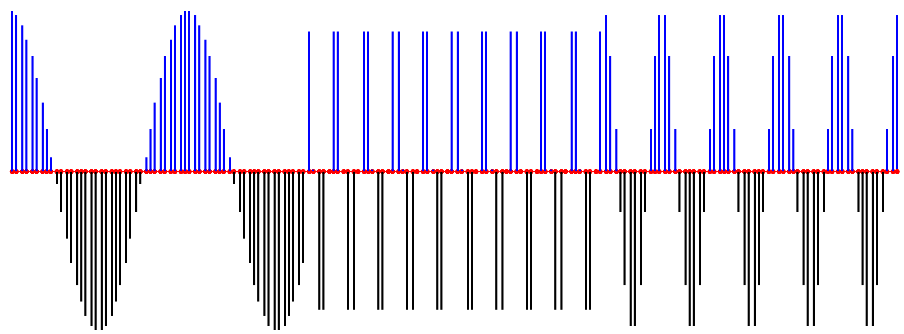
Shuman et al., Vertex-frequency analysis on graphs, ACHA, 2016

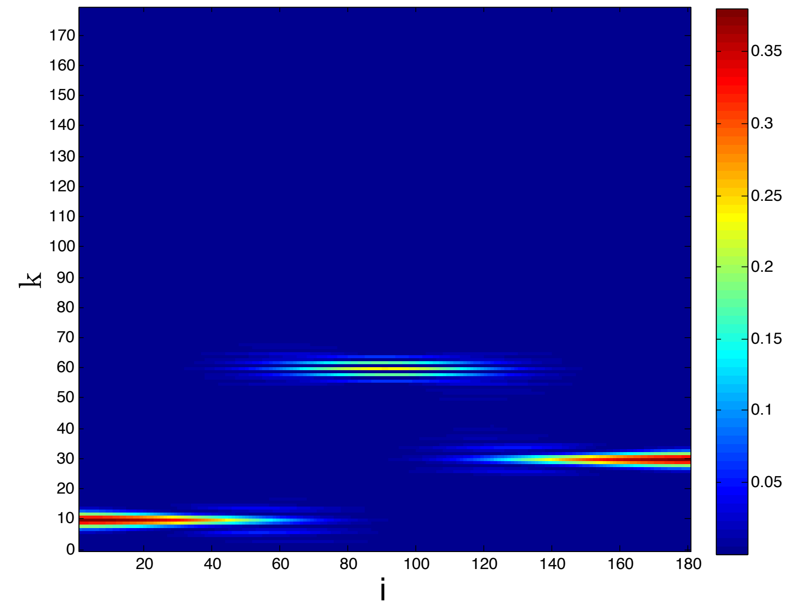
Windowed Graph Fourier Transform (cont.)

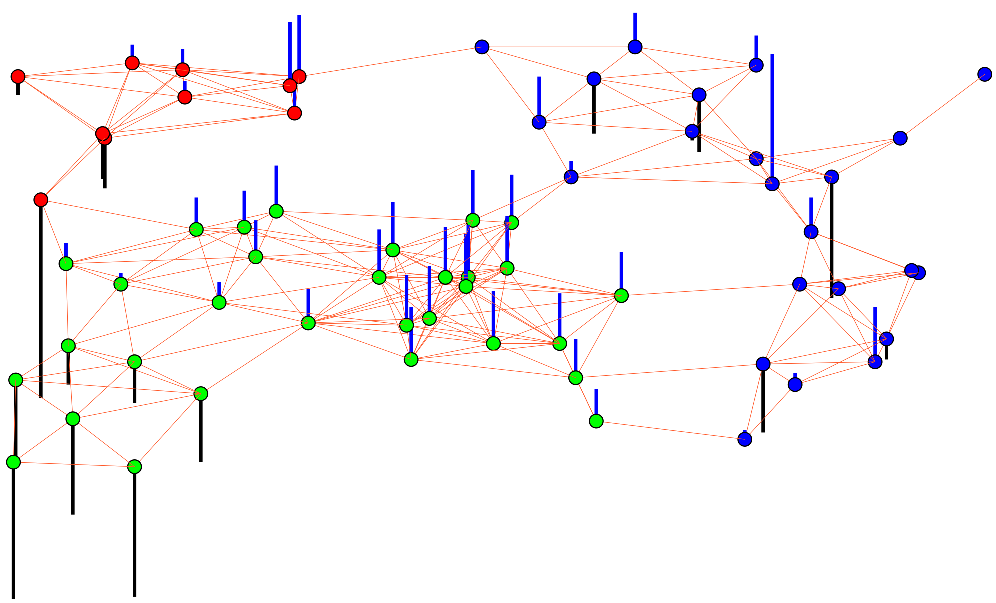
- Windowed graph Fourier atoms: $g_{i,k} := M_k T_i g$

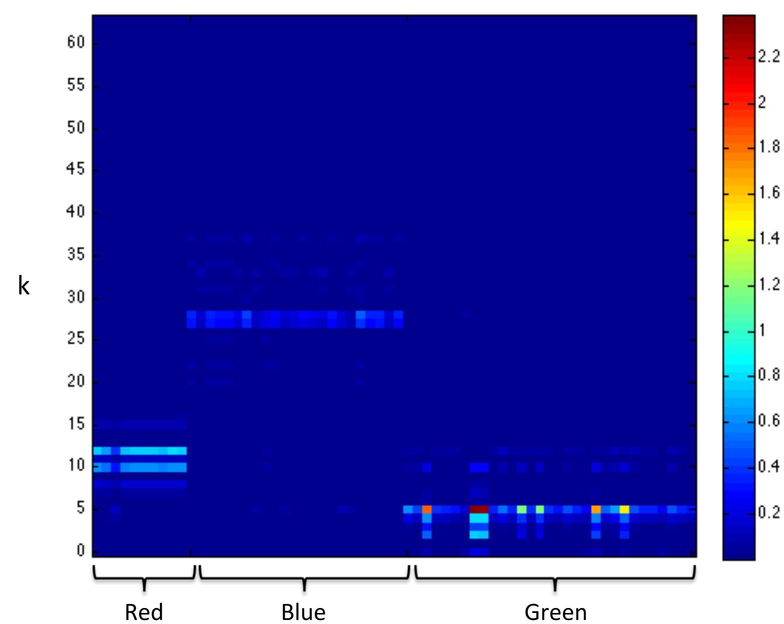


Spectrogram Examples



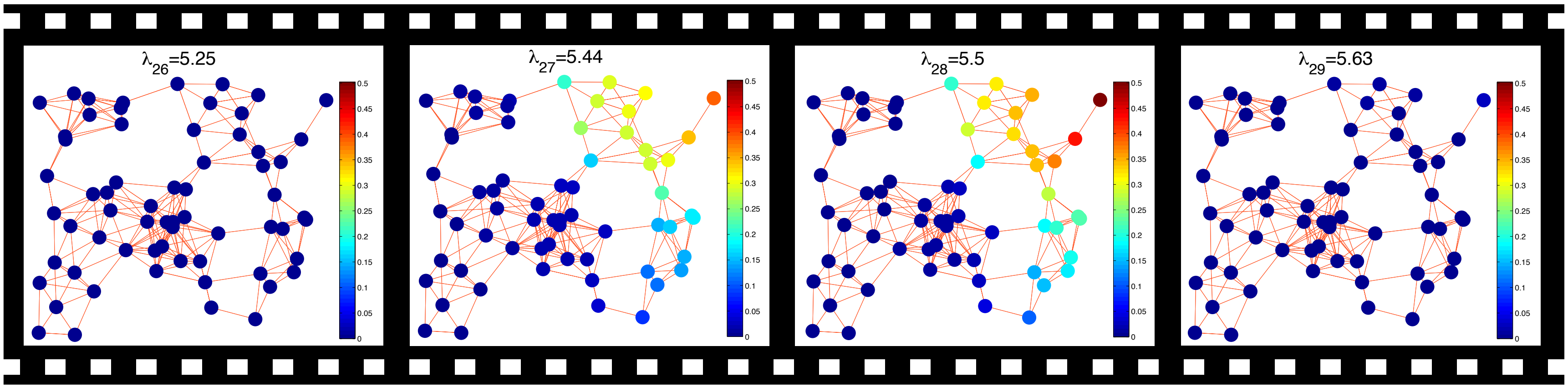


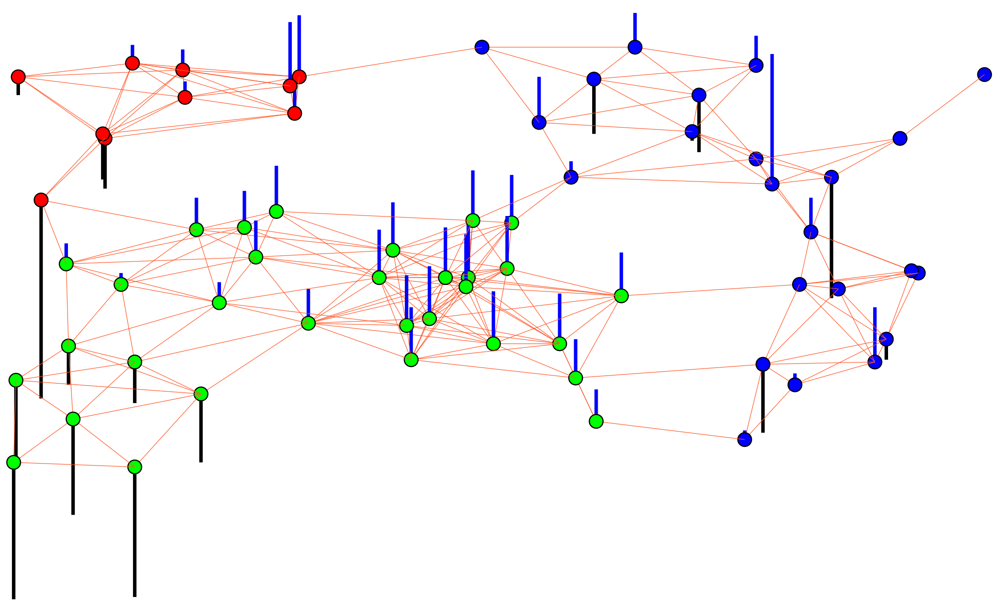


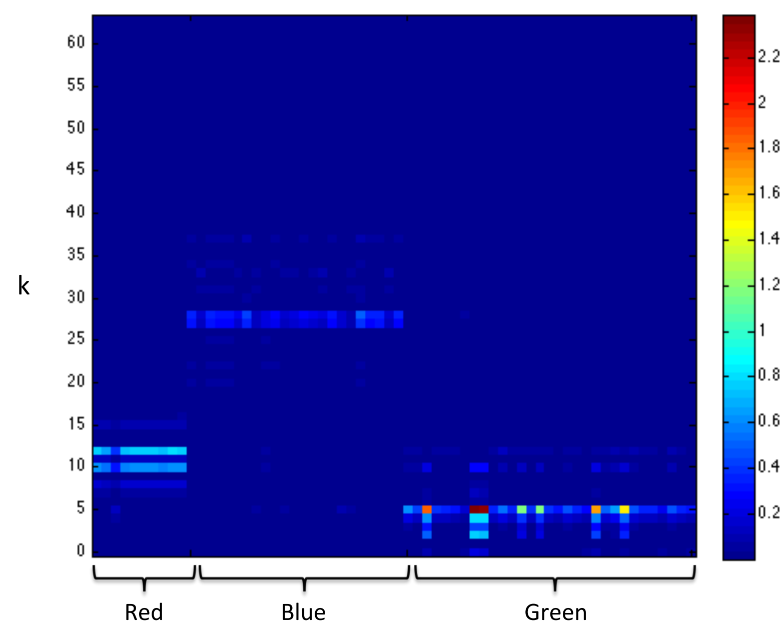


Spectrogram Examples

- Spectrogram = frequency-lapse video



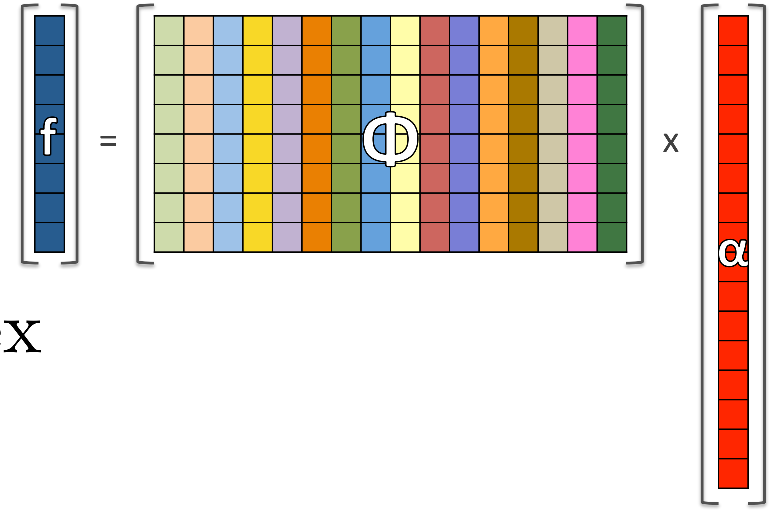




Survey of Approaches to Graph Signal Dictionary Design

- Windowed graph Fourier transform
- **Spectral domain designs**
- Generalized filter banks

Dictionary of Localized Kernels



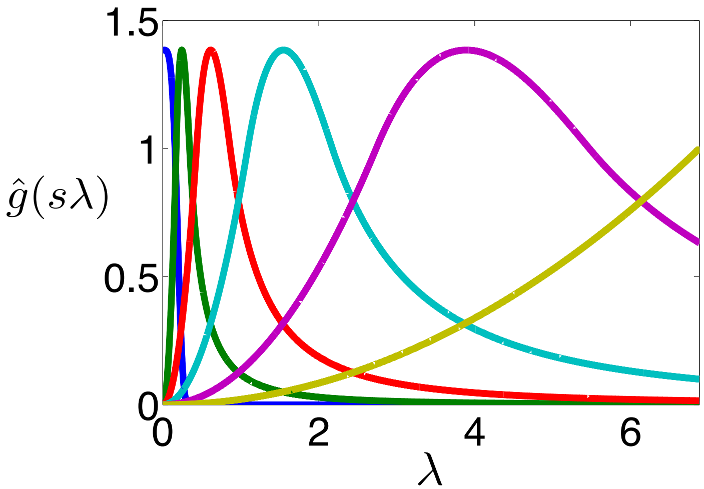
- M kernels/patterns $\{g_m\}$
- Localize each pattern to each vertex
  - Atoms of the form $T_i g_m = g_m(\mathcal{L})\delta_i$
  - $\Phi = [g_1(\mathcal{L}), g_2(\mathcal{L}), \ldots, g_M(\mathcal{L})]$
- Dictionary is overcomplete with MN atoms
- Approximate kernels with polynomials
  - Ensures joint localization in both domains
  - Fast computations with dictionary and its adjoint

Example: Spectral Graph Wavelets

Hammond et al., Wavelets on graphs via spectral graph theory, ACHA, 2011

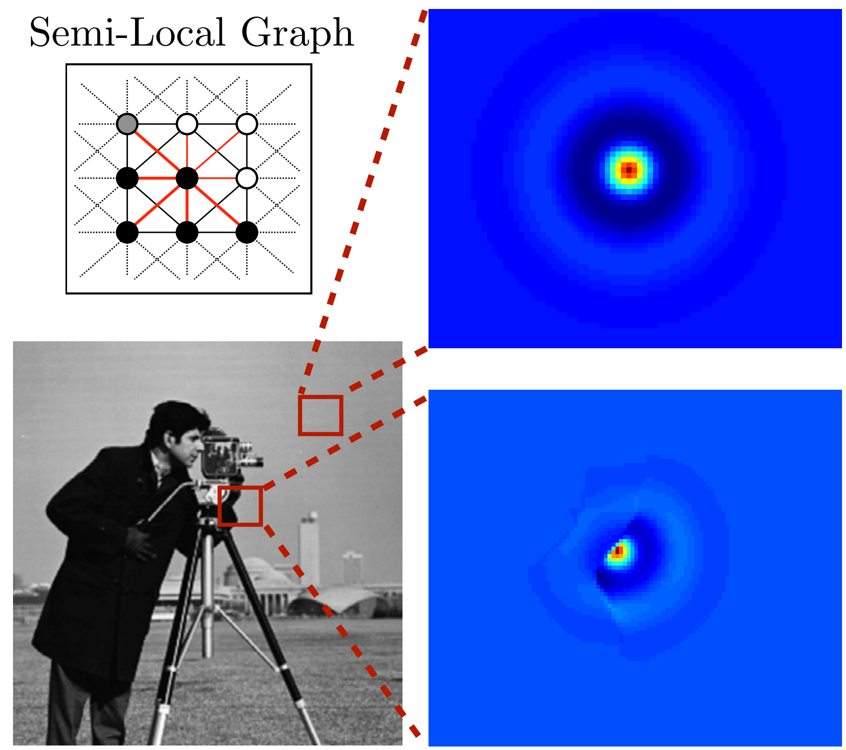
- Generalized dilation:

$$\widehat{\mathcal{D}_s g}(\lambda) = \hat{g}(s\lambda)$$

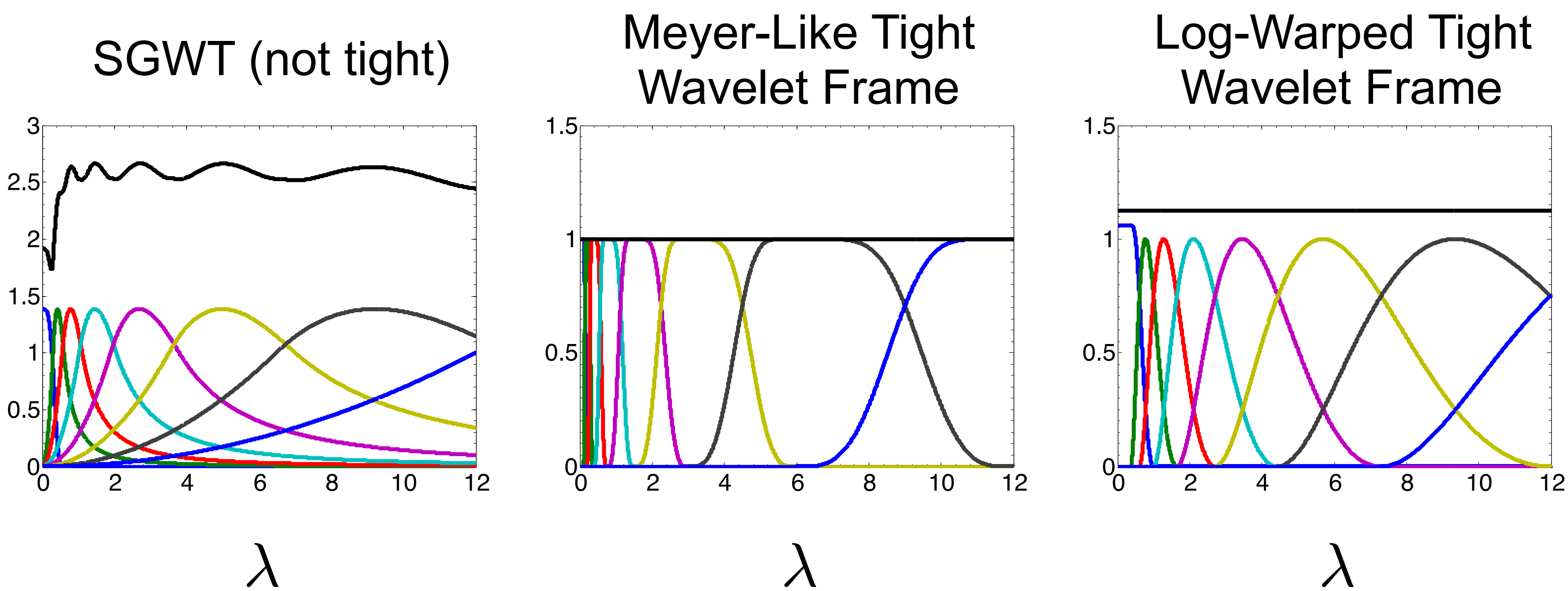


- Spectral graph wavelet at scale s, centered at vertex n:

$$\psi_{s,n}(i) := (T_n D_s g)(i) = \sum_{\ell=0}^{N=1} \hat{g}(s\lambda_\ell) u_\ell^*(n) u_\ell(i)$$



Improvement 1: Energy Conservation

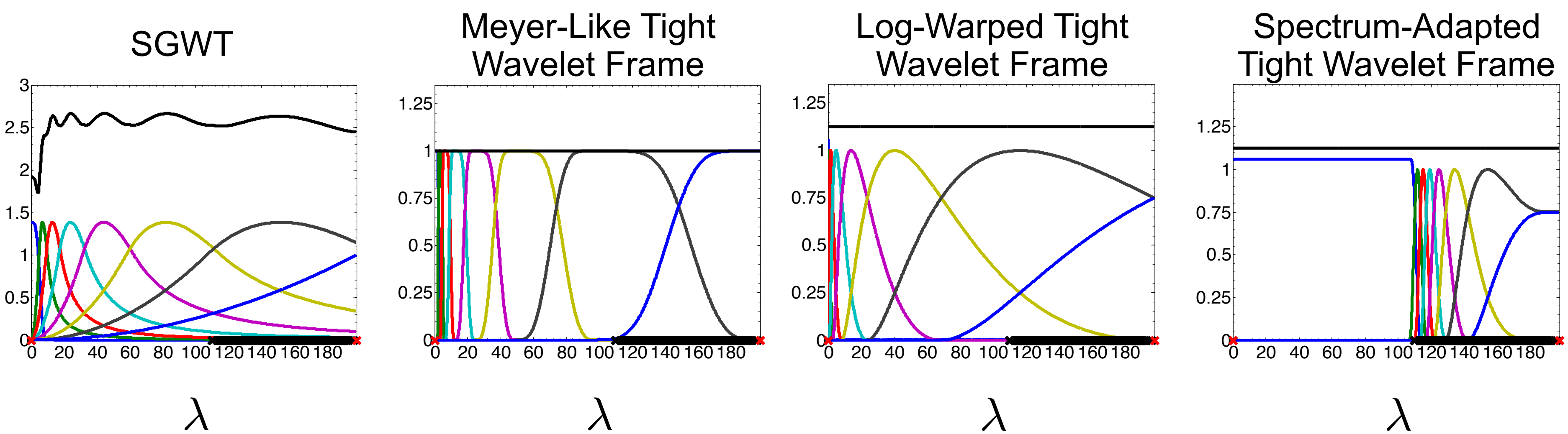


$$||\Phi^* f||_2^2 = \sum_{i=1}^{N} \sum_{m=1}^{M} |\langle f, g_{i,m} \rangle|^2 = A||f||_2^2$$

- Leonardi and Van De Ville, Tight wavelet frames on multislice graphs, TSP, 2013
- Shuman et al., Spectrum-adapted tight graph wavelet and vertex-frequency frames, TSP, 2015

Improvement 2: Discrimination Power

- Ideally, atoms should not be too correlated with each other
- An extreme example:



- Cumulative coherence for a given sparsity level k

$$\mu_1(k) := \max_{|\Theta|=k} \max_{\psi \in \mathcal{D}_{\{1,2,\ldots,N \cdot M\}} \setminus \Theta} \sum_{\theta \in \Theta} \frac{|\langle \psi, \mathcal{D}_\theta \rangle|}{||\psi||_2 ||\mathcal{D}_\theta||_2}$$

Aside: Fast Estimation of the Spectral Distribution with the Kernel Polynomial Method

$$F_{\mathcal{L}}(\lambda)$$

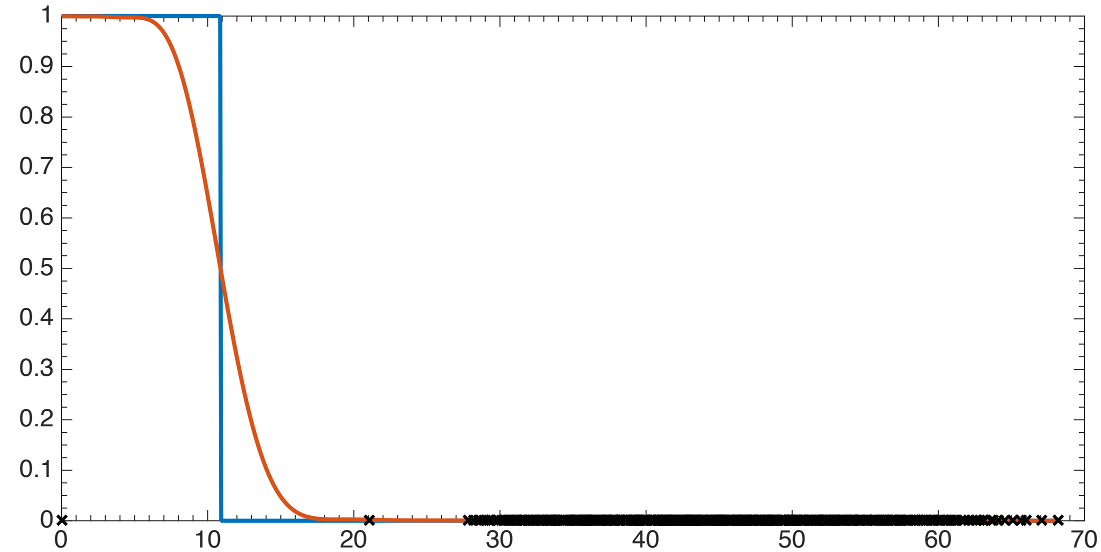
$$= \frac{1}{N}\sum_{\ell=0}^{N-1} \mathbb{1}_{\{\lambda_\ell \leq \lambda\}}$$

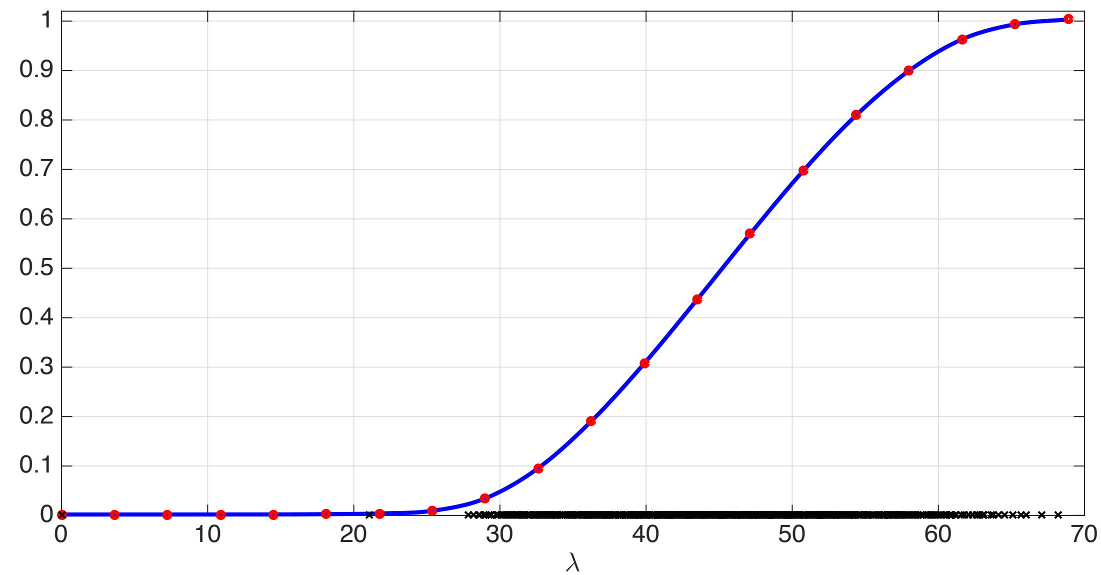
$$= \frac{1}{N}\mathrm{tr}(\phi_\lambda(\mathcal{L}))$$

$$= \frac{1}{N}\mathbb{E}[r^\top \phi_\lambda(\mathcal{L}) r]$$

$$\approx \frac{1}{N}\left(\frac{1}{J}\sum_{j=1}^{J} r^{(j)\top} \phi_\lambda(\mathcal{L}) r^{(j)}\right)$$

Polynomial filter

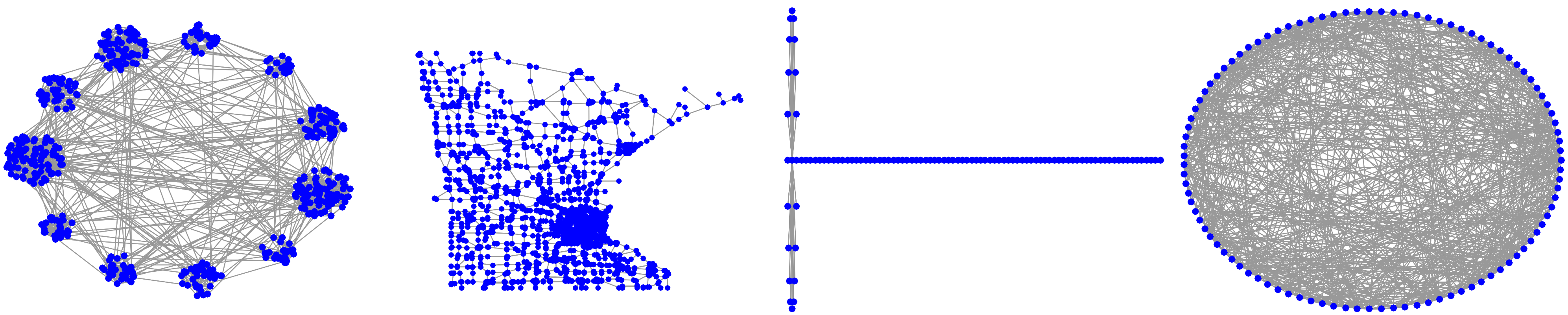


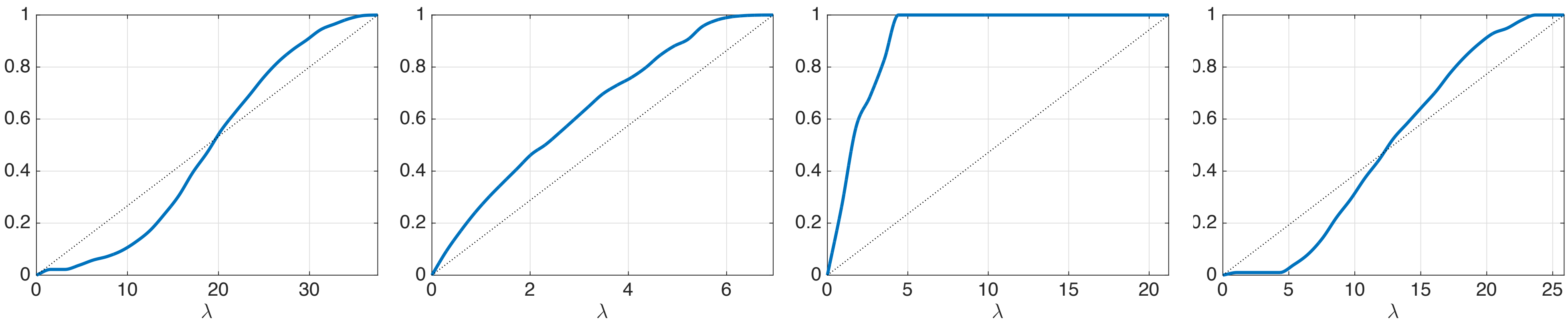


- Monotonic cubic interpolation

Lin et al., "Approximating spectral densities of large matrices," *SIAM Review*, 2016

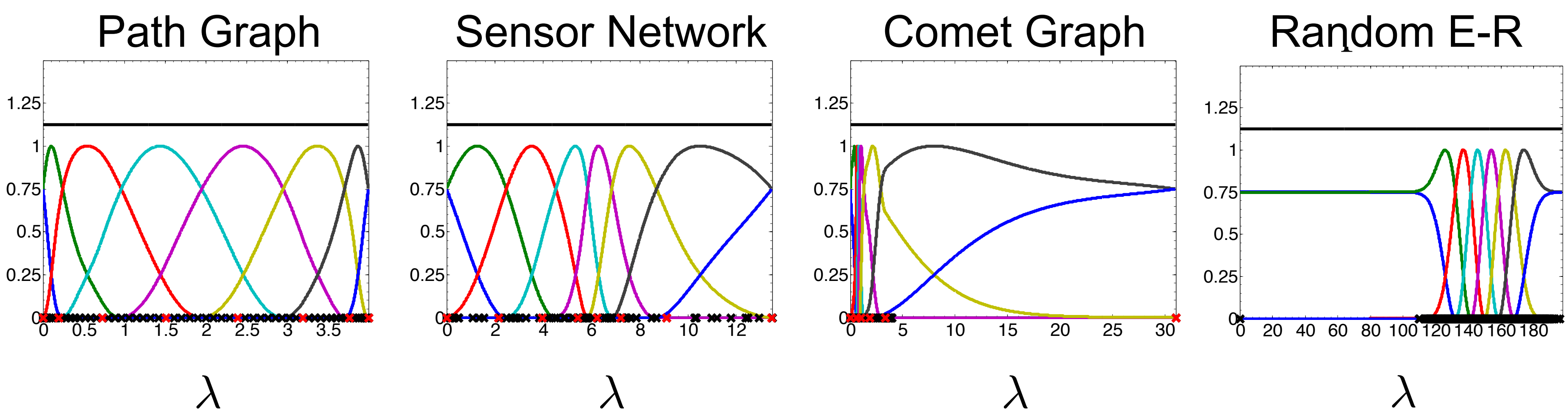
Spectral Distribution Examples





Improvement 2: Discrimination Power

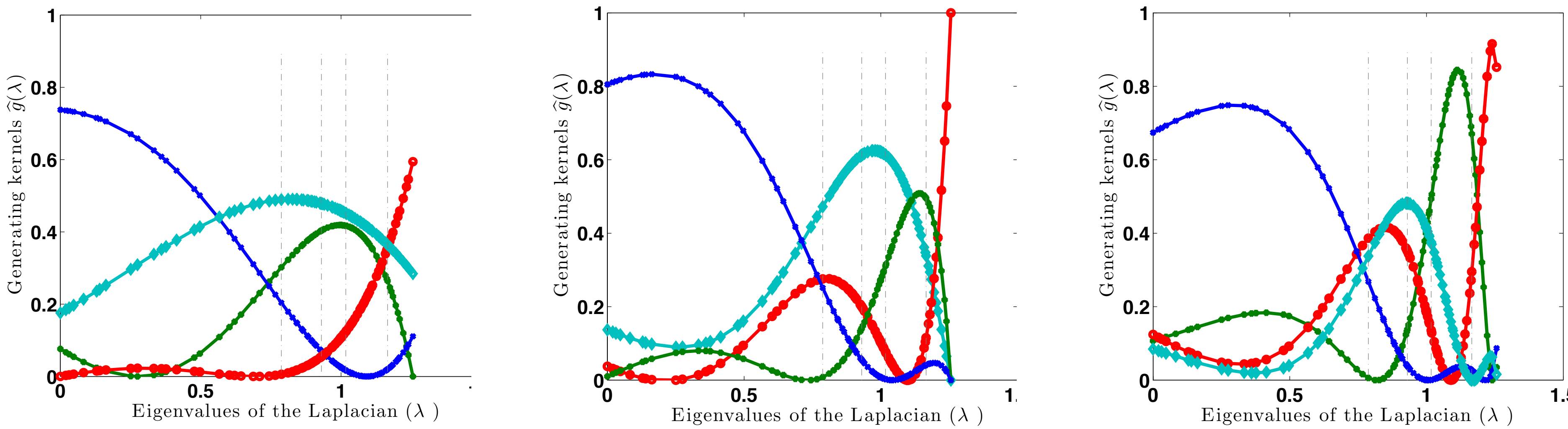
- With access to a rough estimate of the spectral density, we can adapt the filters to the spectrum via warping



Shuman et al., Spectrum-adapted tight graph wavelet and vertex-frequency frames, TSP, 2015

Variant: Parametric Learning

- Restrict kernels to be polynomials of a given degree, and learn the polynomial coefficients from a training data set



(a) $K = 5$ (b) $K = 10$ (c) $K = 20$

Zhang et al., Learning of structured graph dictionaries, ICASSP, 2012

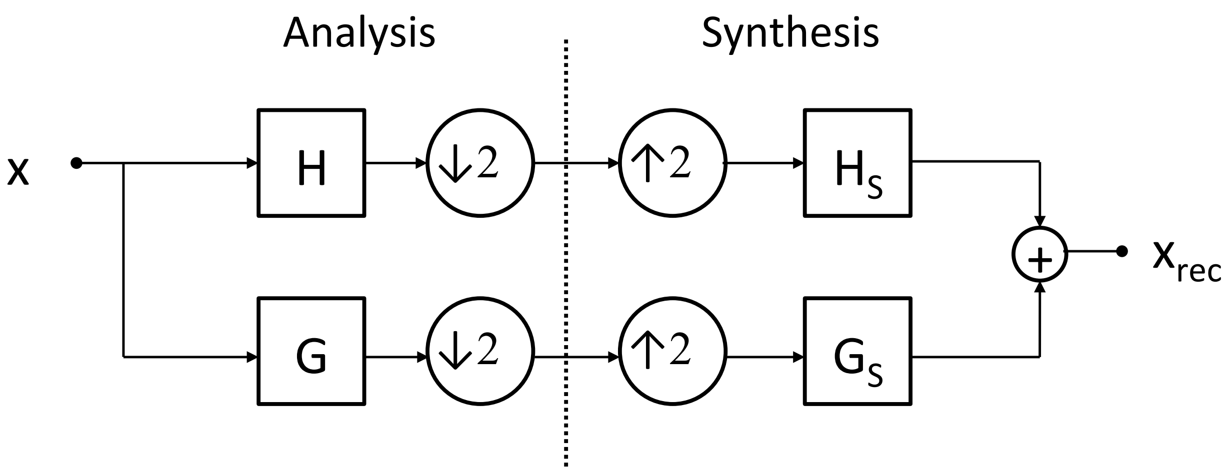
Thanou et al., Learning parametric dictionaries for signals on graphs, TSP, 2014

Survey of Approaches to Graph Signal Dictionary Design

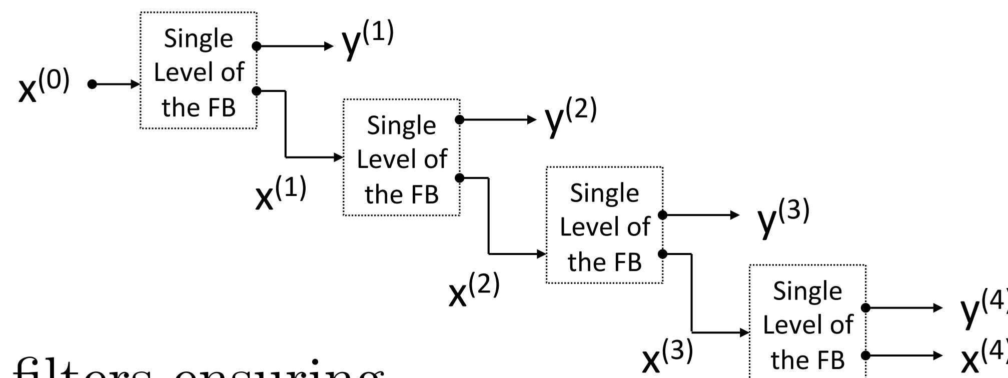
- Windowed graph Fourier transform
- Spectral domain designs
- **Generalized filter banks**

1D Wavelets Via Filter Banks

Classical 2-Channel Critically Sampled Filter Bank



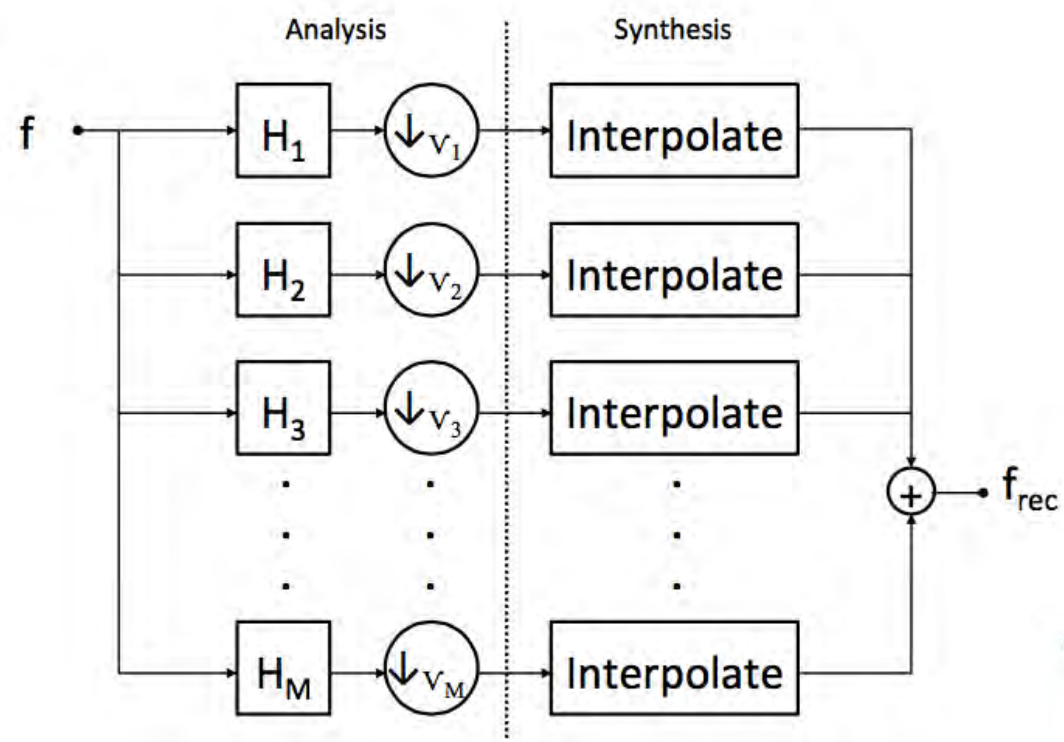
Iterating Low Pass Branch Yields Wavelets



- To extend to the graph setting, we need appropriate notions of downsampling, upsampling, filtering, graph reduction
- Some issues that arise:
  - Difficulty generalizing conditions on filters ensuring properties such as perfect reconstruction, orthogonality
  - Preserving a meaningful correspondence between filtering at different resolution levels

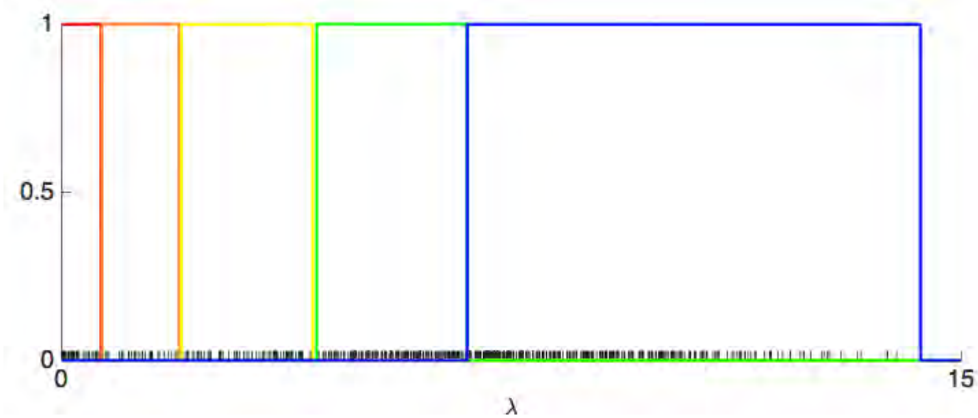
M-Channel Critically Sampled Graph Filter Bank

Architecture



- Number of vertices in V_i is equal to the number of eigenvalues in the support of the corresponding filter

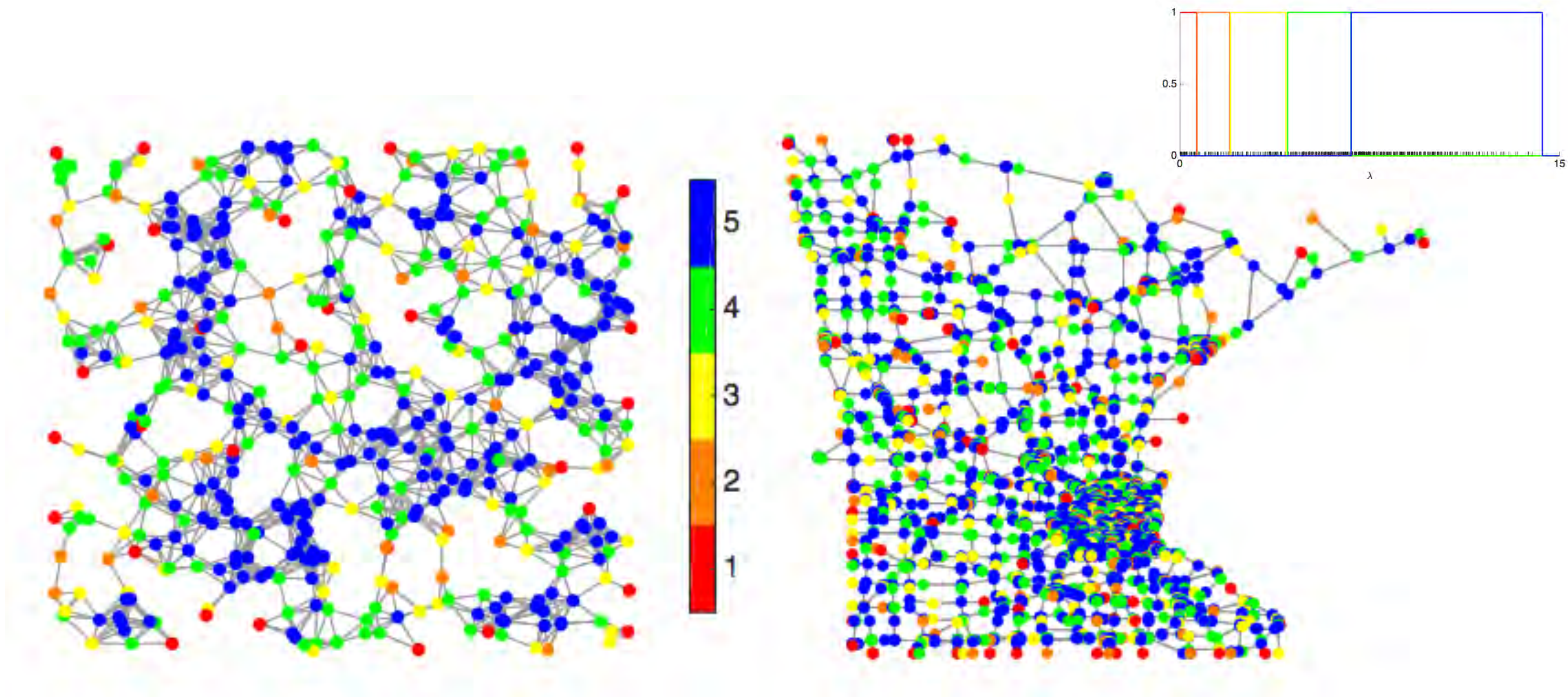
Ideal Filter Bank



Jin and Shuman., "An M-channel critically sampled filter bank for graph signals," ICASSP, 2017

M-Channel Critically Sampled Graph Filter Bank

- Partition into uniqueness sets for ideal filter bank subspaces:

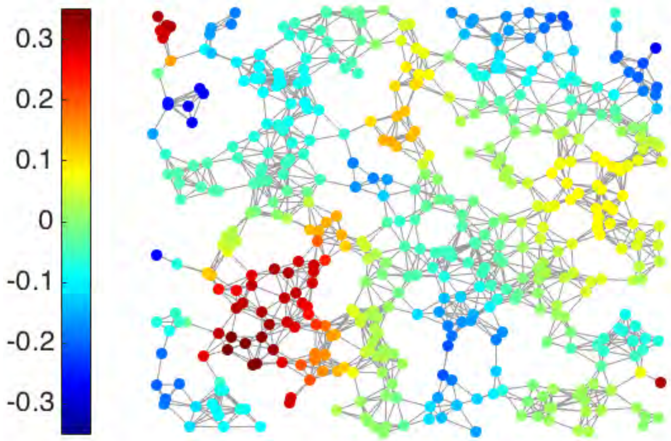


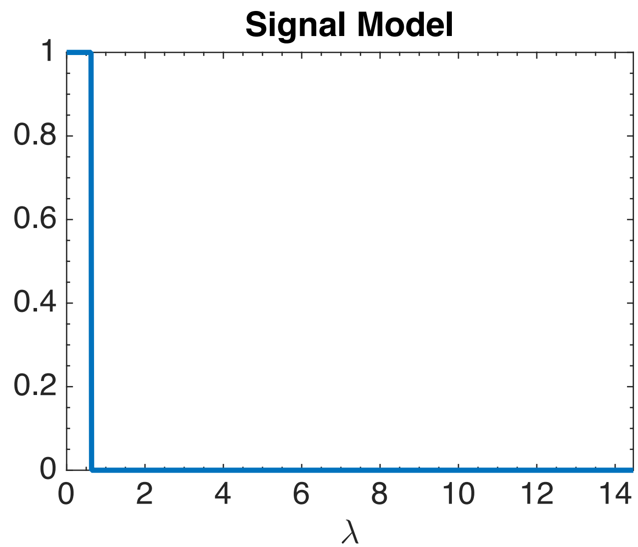
- To avoid a full eigendecomposition, we would like to use random, non-uniform sampling and fast, approximate reconstruction methods

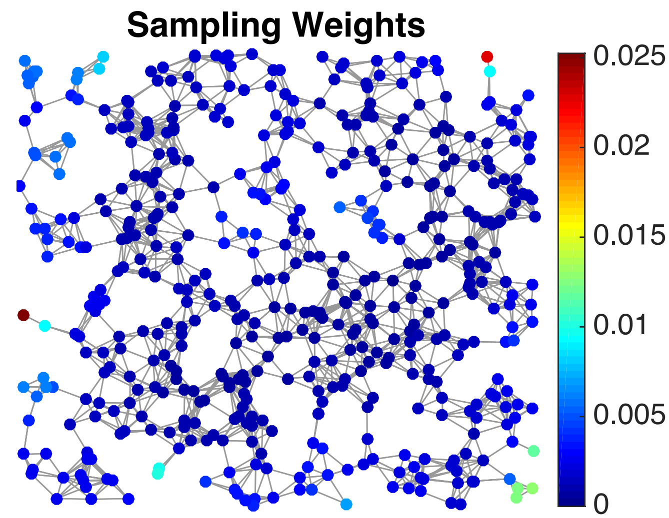
Greene and Magnanti, "Some abstract pivot algorithms," *SIAM J. Appl. Math.*, 1975

Non-Uniform Random Sampling

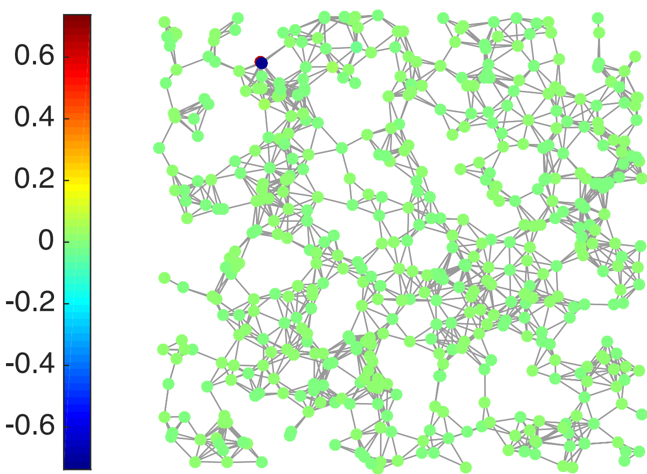
Lowpass (smooth) signals

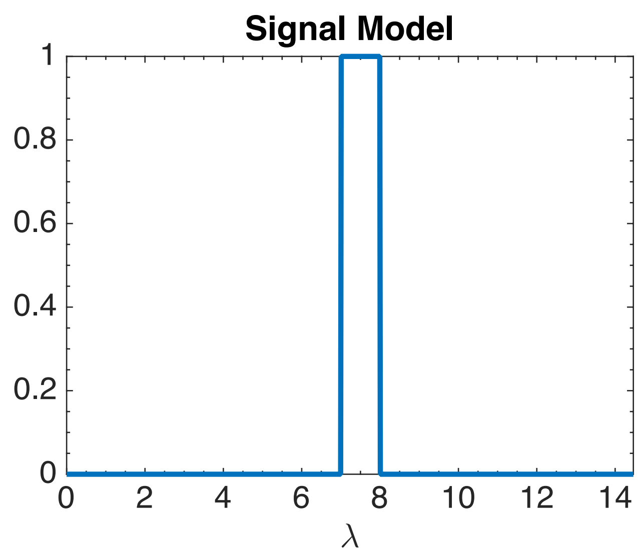


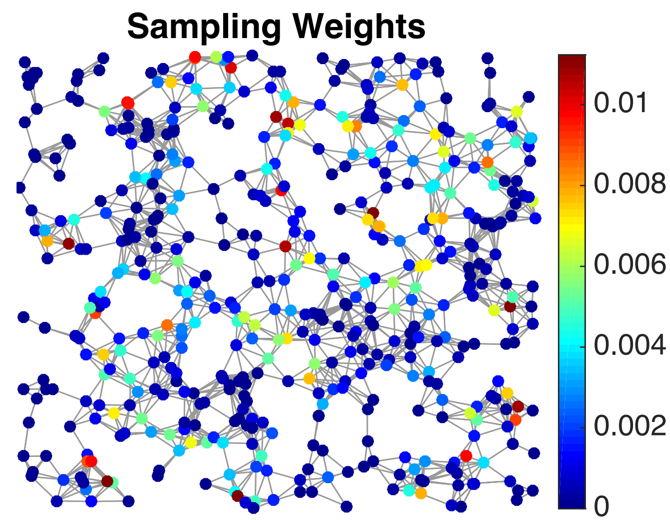




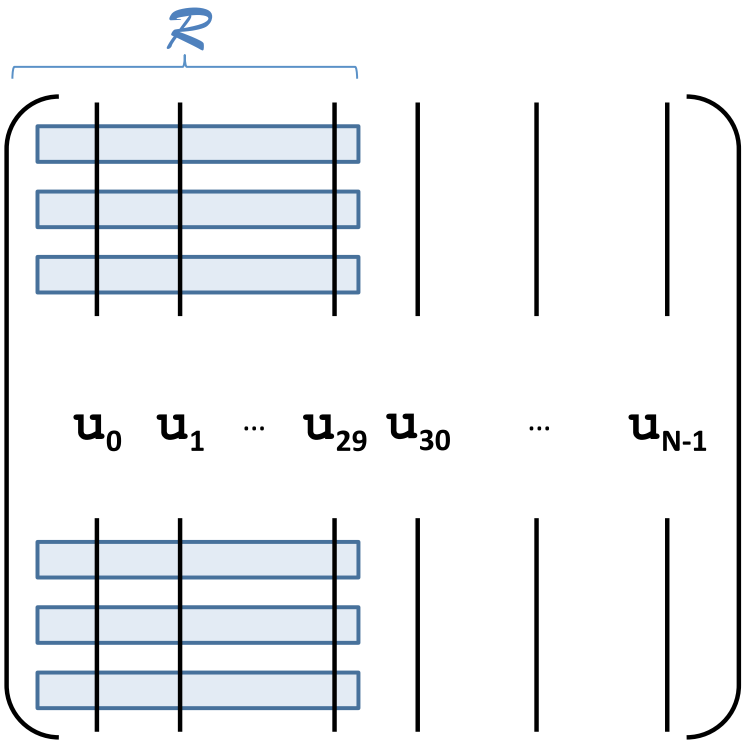
Midpass signals

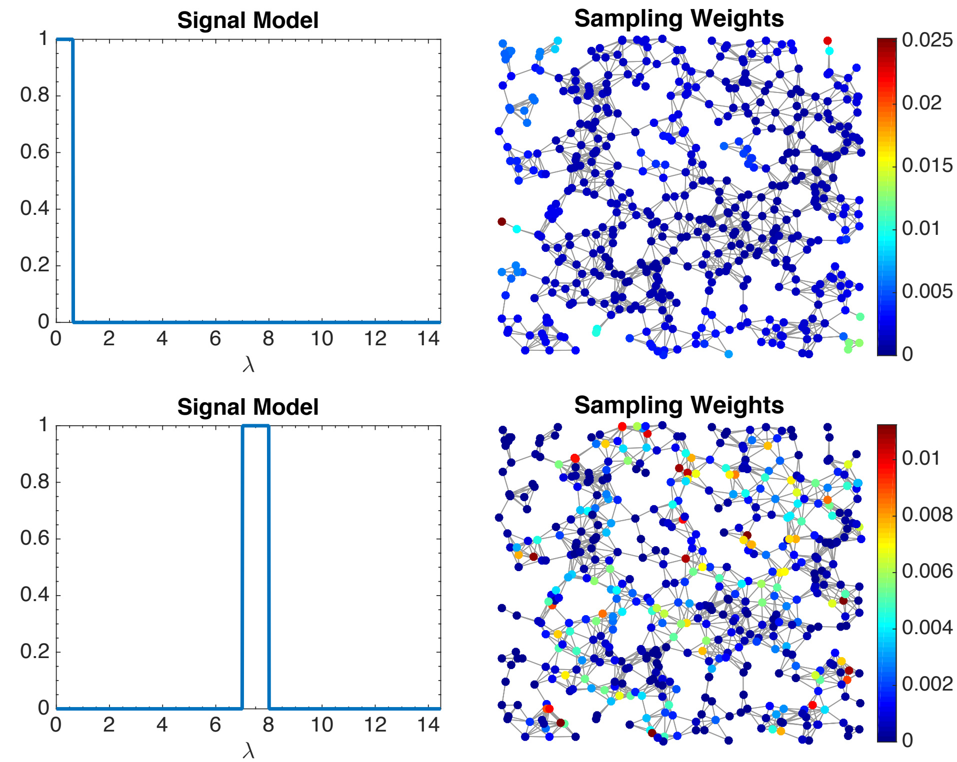






Non-Uniform Random Sampling





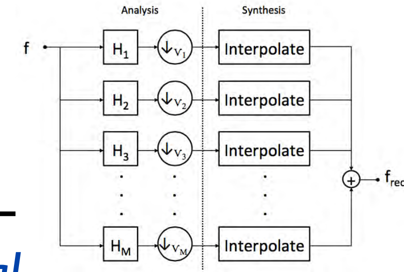
$$p_i \sim ||U_{:,\mathcal{R}}^\top \delta_i||_2^2$$

$$\approx \frac{1}{T}\sum_{t=1}^{T}[(\tilde{h}(\mathcal{L})r^{(t)})(i)]^2$$

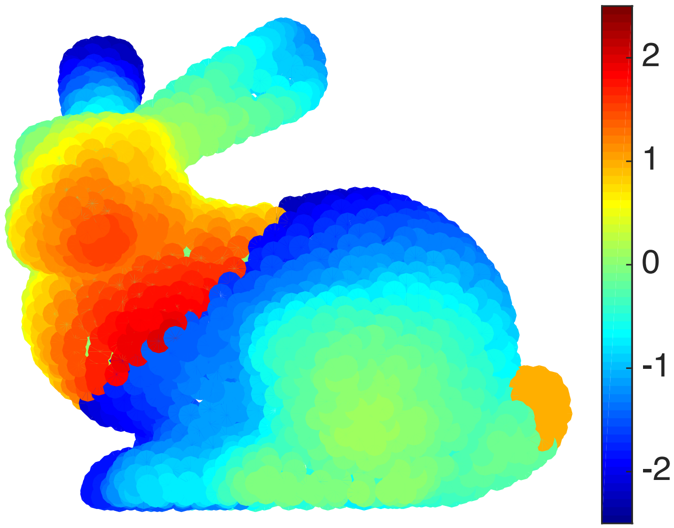
approximation of the filter describing the signal model

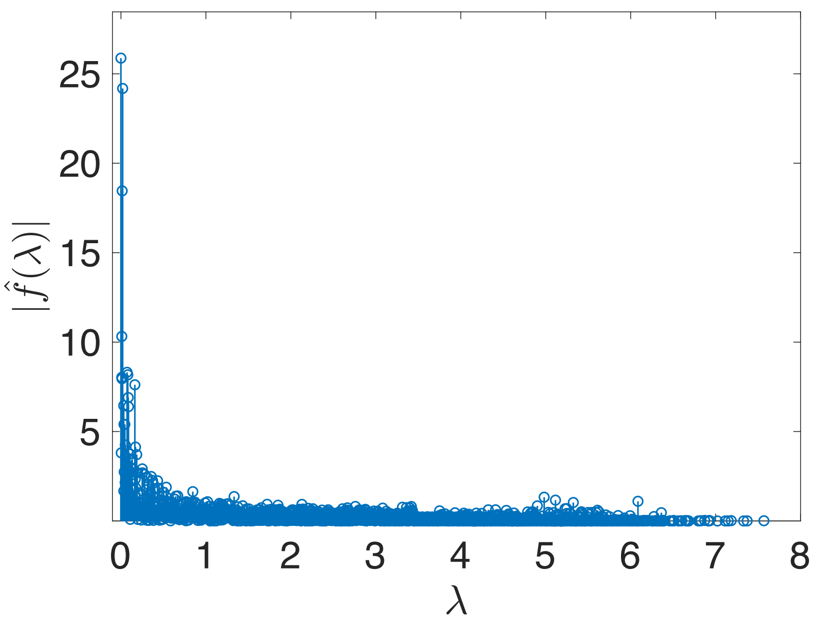
independent random entries that follow a standard normal dist.

Signal-Adapted M-CSFB

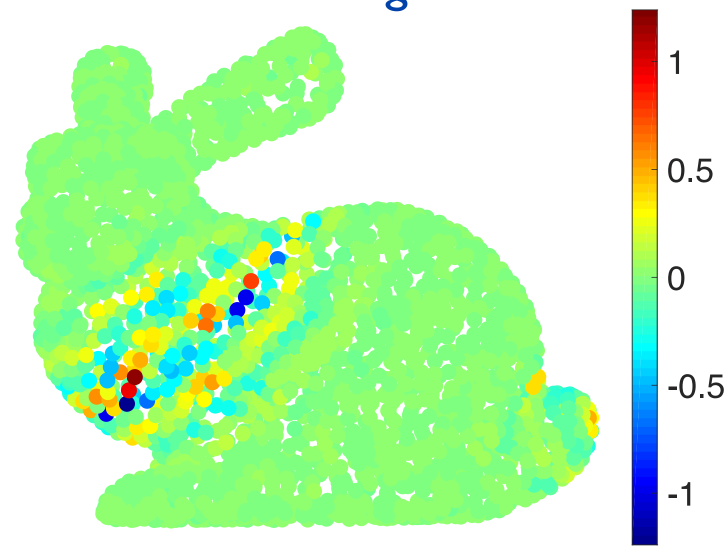


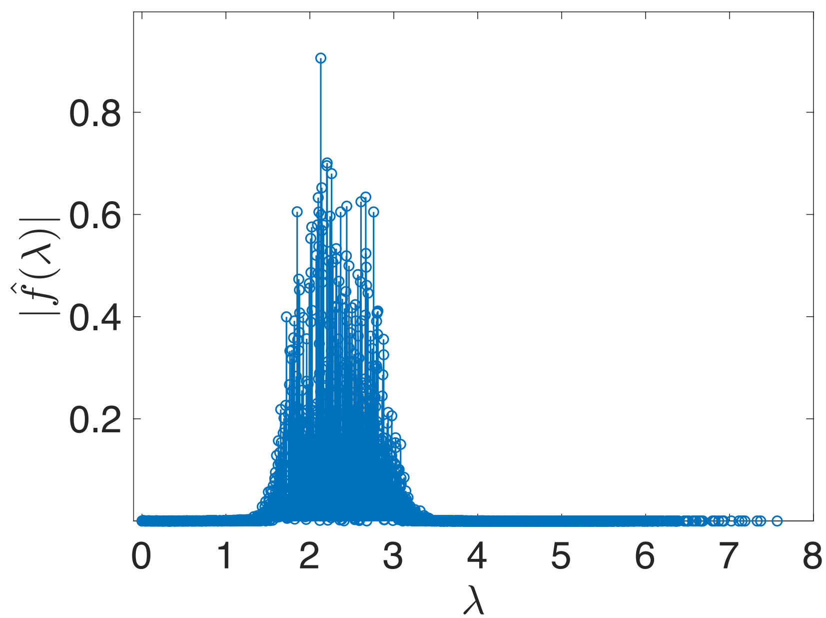
Signal



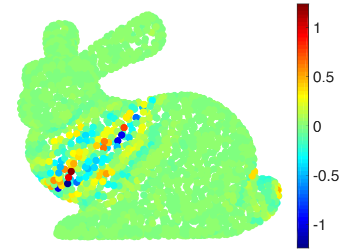


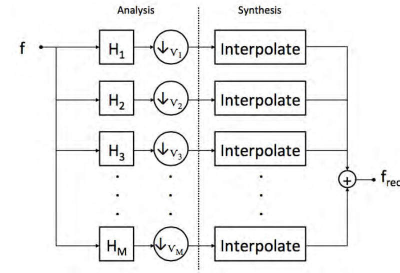
Filtered Signal

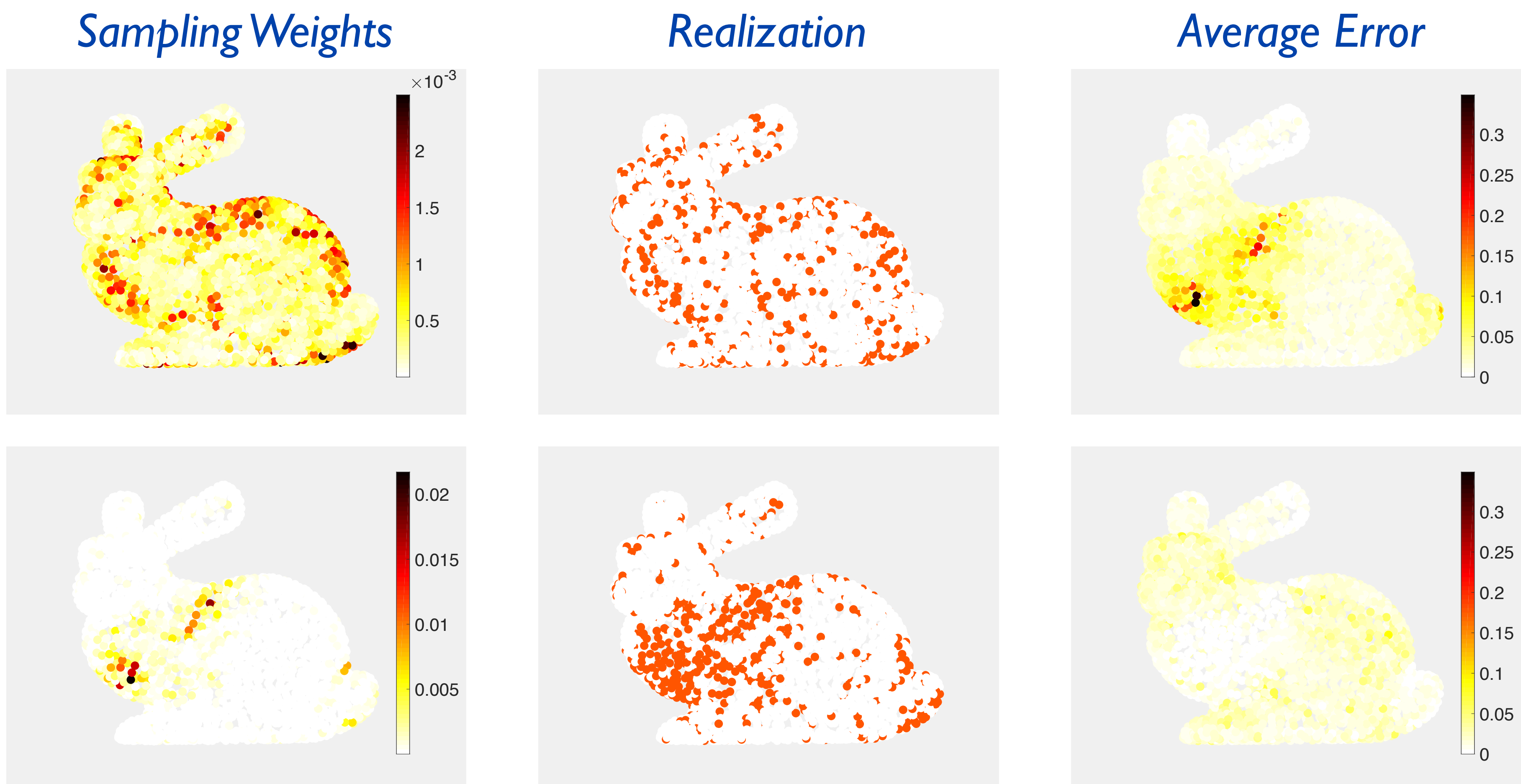




Signal-Adapted M-CSFB





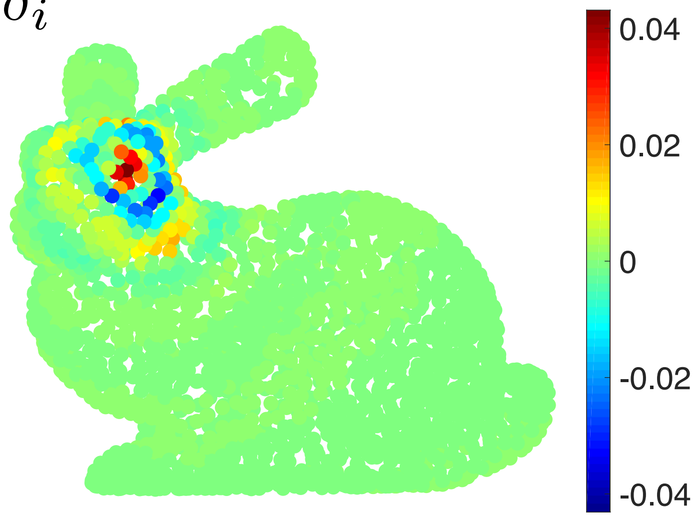


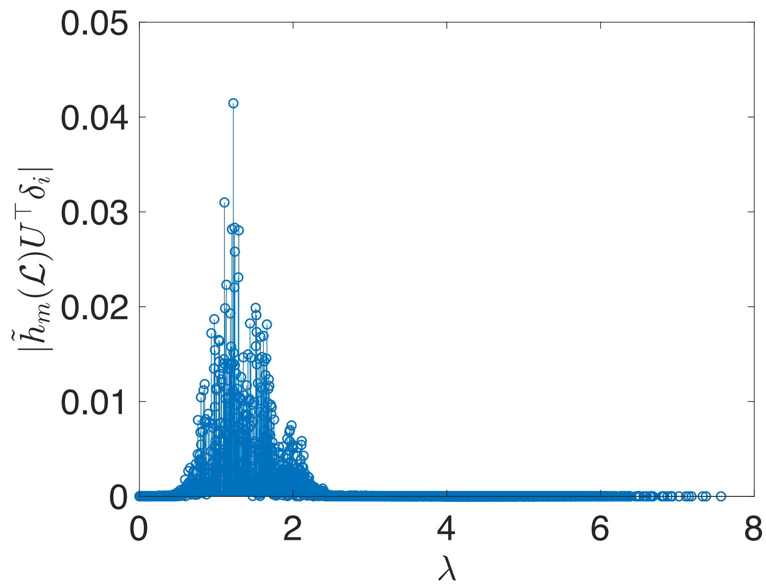
- Also reallocate samples across bands based on signal's energy

Li, Jin, and Shuman, "Scalable M-channel critically sampled filter banks," 2018

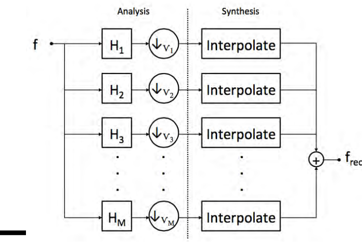
Joint Localization of Atoms

- Dictionary atoms are of the form $\tilde{h}_m(\mathcal{L})\delta_i$
- Localized within K hops of center vertex
- As K increases, become more concentrated in spectral domain





Efficient Interpolation

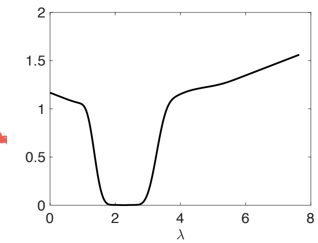


downsampling operator

approximate by convex optimization problem

$$\min_{z \in \mathrm{col}(U_{:,\mathcal{R}_m})} ||\Omega_{m,\mathcal{V}_m}^{-1/2}(M_m z - y_{\mathcal{V}_m})||_2^2$$

signal model space



$$\min_{z \in \mathbb{R}^N} \left\{ \gamma ||\Omega_{m,\mathcal{V}_m}^{-1/2}(M_m z - y_{\mathcal{V}_m})||_2^2 + z^\top \varphi_m(\mathcal{L}) z \right\}$$

optimality condition

$$\left(\gamma M_m^\top \Omega_{m,\mathcal{V}_m}^{-1} M_m + \varphi_m(\mathcal{L})\right) z = \gamma M_m^\top \Omega_{m,\mathcal{V}_m}^{-1} y_{\mathcal{V}_m}$$

solve with preconditioned conjugate gradient

preconditioner:

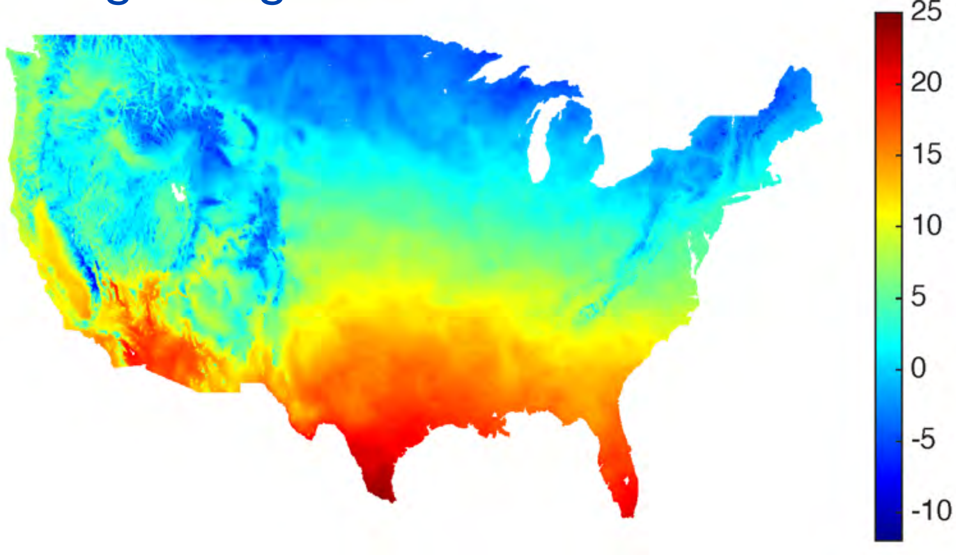
$$\mathrm{diag}\left(1 + \frac{\gamma}{p_i} \mathbb{1}_{\{i \in \mathcal{V}_m\}}\right)$$

Li, Jin, and Shuman, "Scalable M-channel critically sampled filter banks for graph signals," 2018

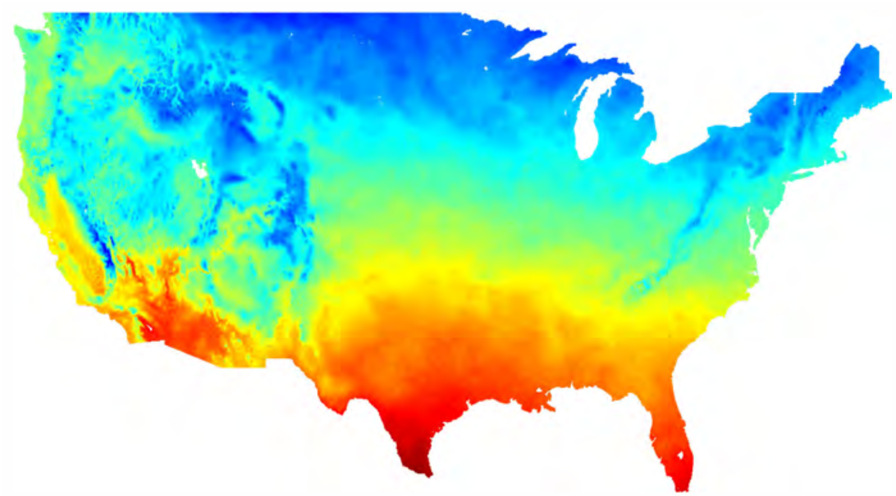
Puy et al., "Random sampling of band limited signals on graphs," ACHA, 2016

Compression Example

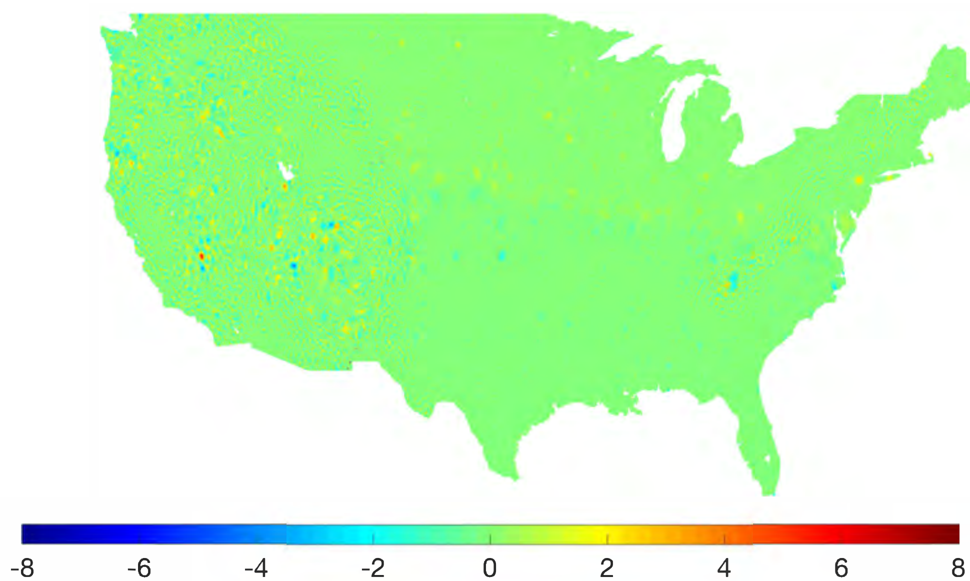
Original Signal



Reconstruction from 10% of Coefficients

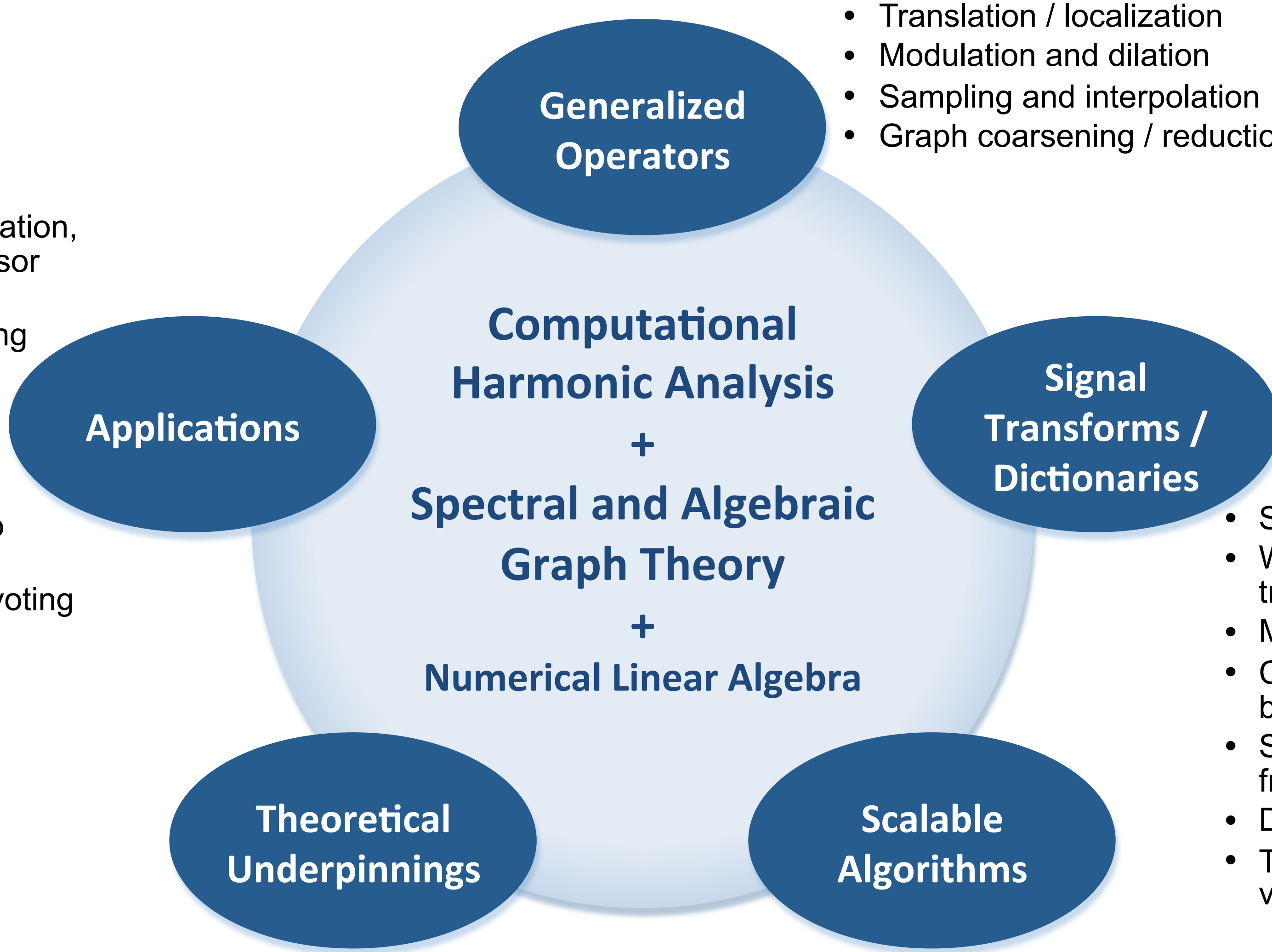


Reconstruction Error



- N=469,404
- Computation times
  - Analysis: 45-90 sec
  - Synthesis: 90-1000 sec

Recent, Ongoing, and Future Work



Generalized Operators

- Filtering / convolution
- Translation / localization
- Modulation and dilation
- Sampling and interpolation
- Graph coarsening / reduction

Applications

- Social, transportation, energy, and sensor networks
- Statistical learning
- Astrophysics
  - CMB
  - Asteroids
- Brain networks
- Image and video processing
- Ranked choice voting
- …

Signal Transforms / Dictionaries

- Spectral graph wavelets
- Windowed graph Fourier transform
- Multiscale pyramids
- Critically-sampled filter banks
- Spectrum-adapted tight frames
- Dictionary learning
- Time series data at each vertex

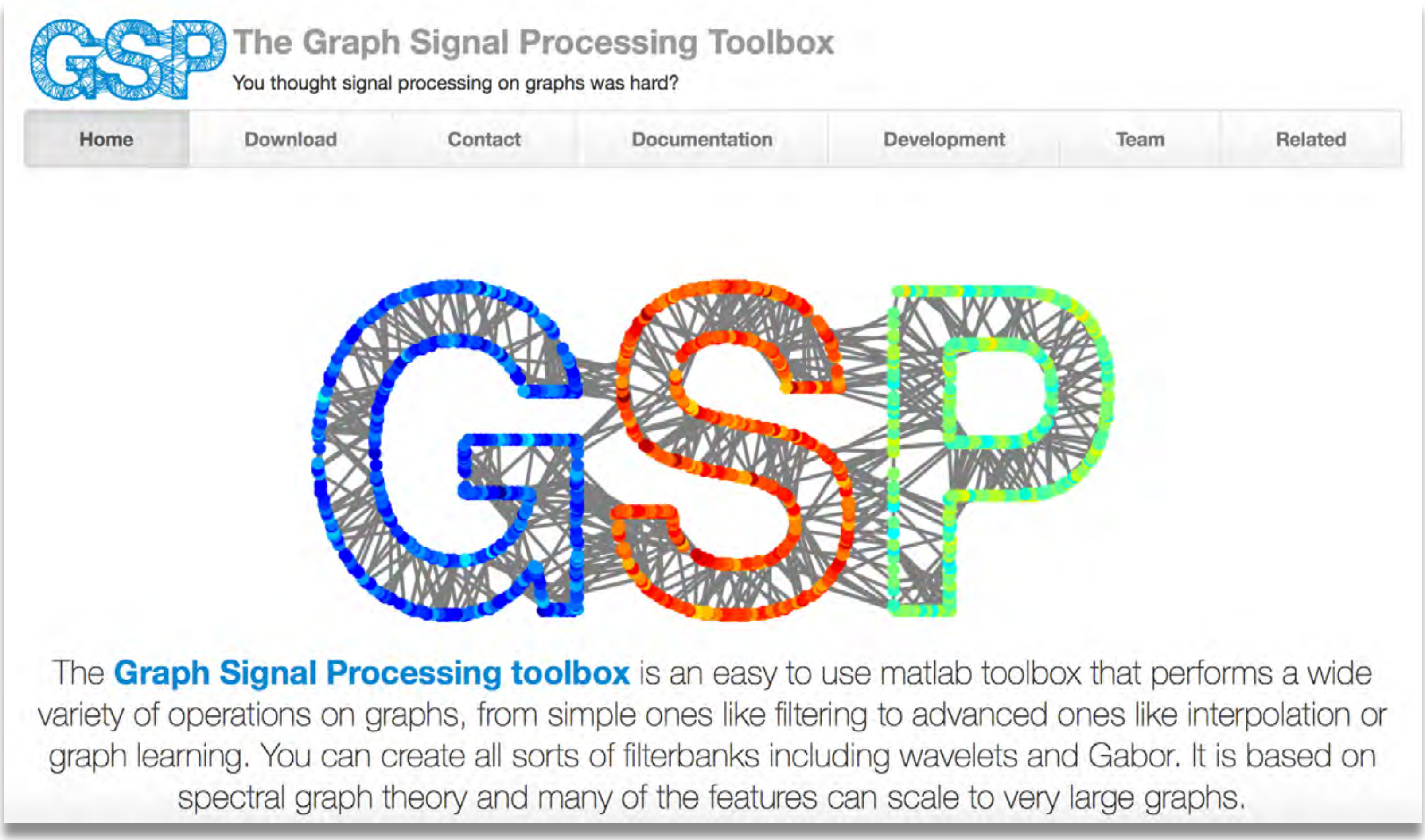
Theoretical Underpinnings

- Connections between sparsity and smoothness
- Uncertainty principles
- Graph Laplacian eigenvector localization

Scalable Algorithms

- Goal: Avoid full eigen-decomposition
- Polynomial approximation algorithms
  - Centralized (incl. GPU)
  - Distributed
- Spectrum-adapted approximation algorithms

Explore



- https://lts2.epfl.ch/gsp/
- https://www.macalester.edu/~dshuman1/publications.html