Numerical Linear Algebra Problems Arising in Graph Signal Processing

---

David Shuman

March 19, 2018

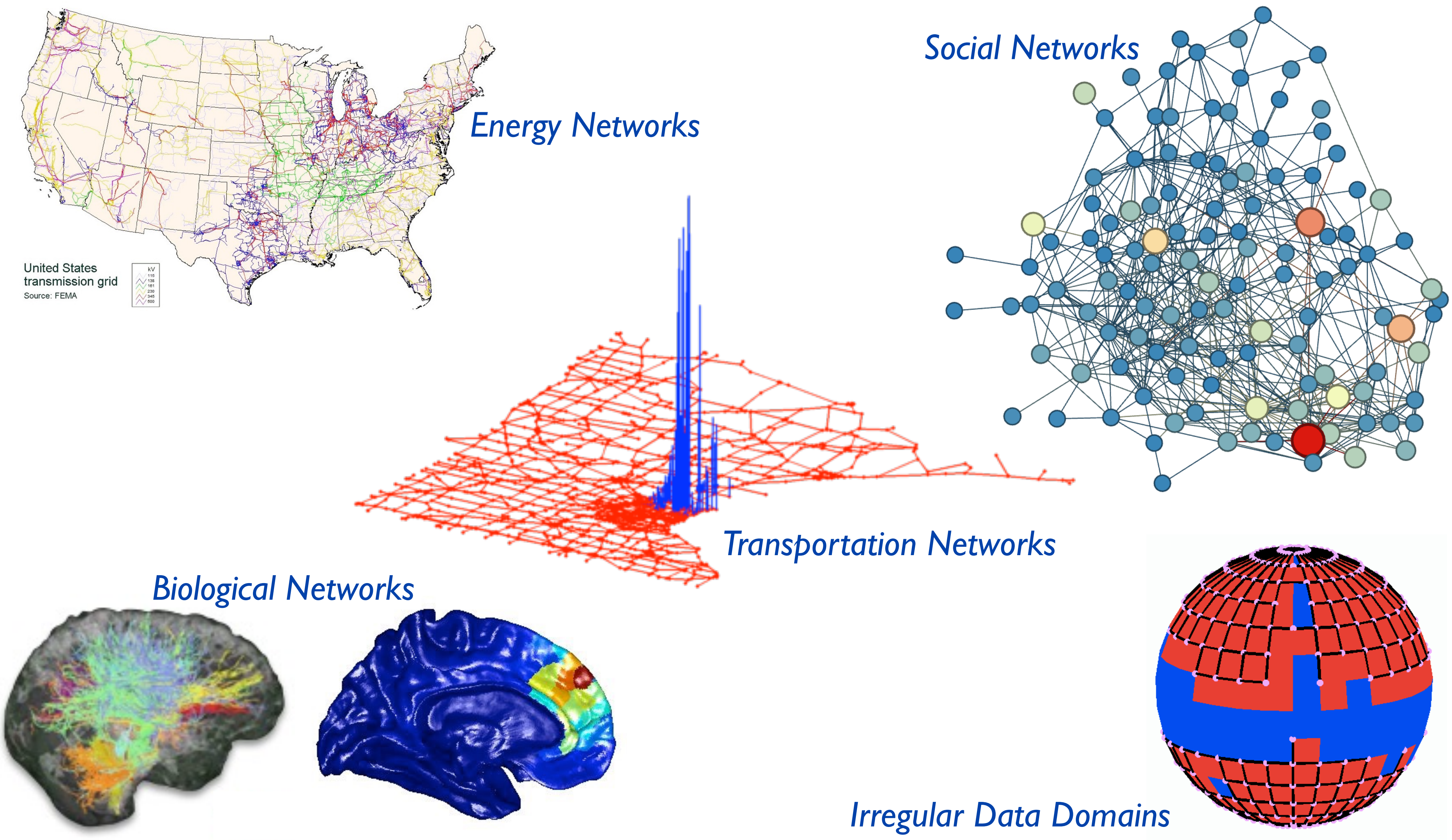
Department of Computer Science and Engineering

University of Minnesota

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

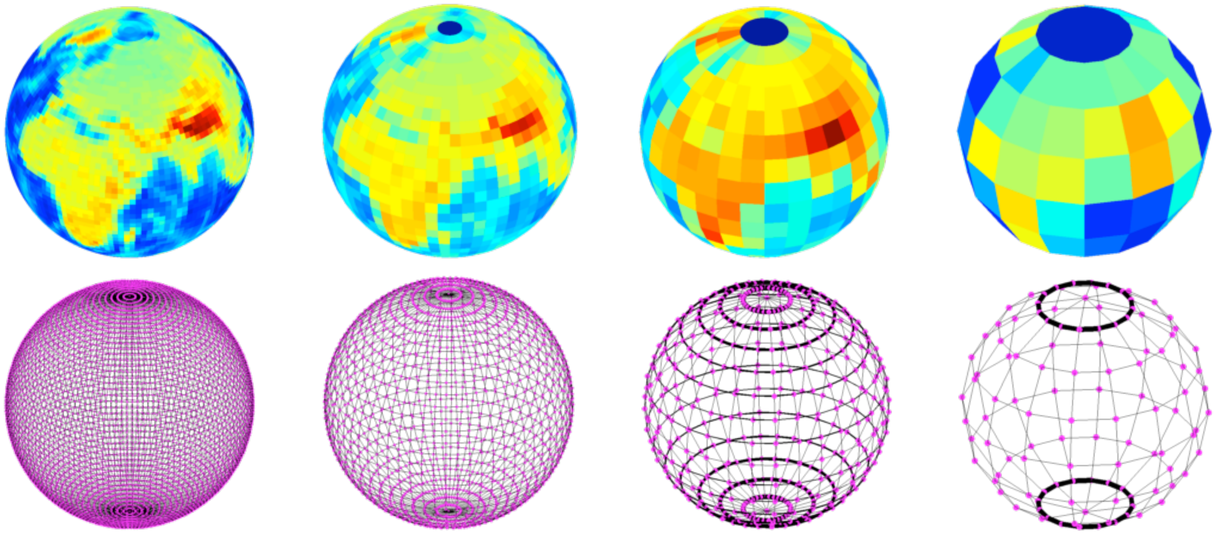
MACALESTER

Signal Processing on Graphs

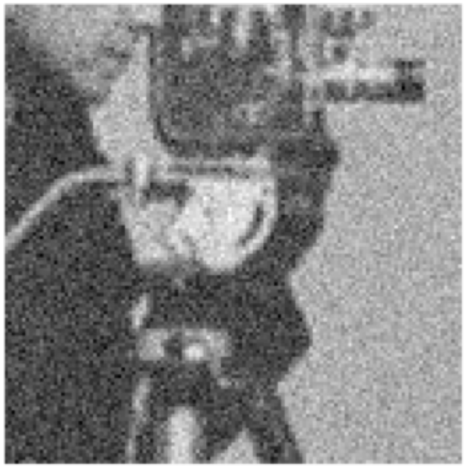


Some Typical Graph Signal Processing Problems

Compression / Visualization

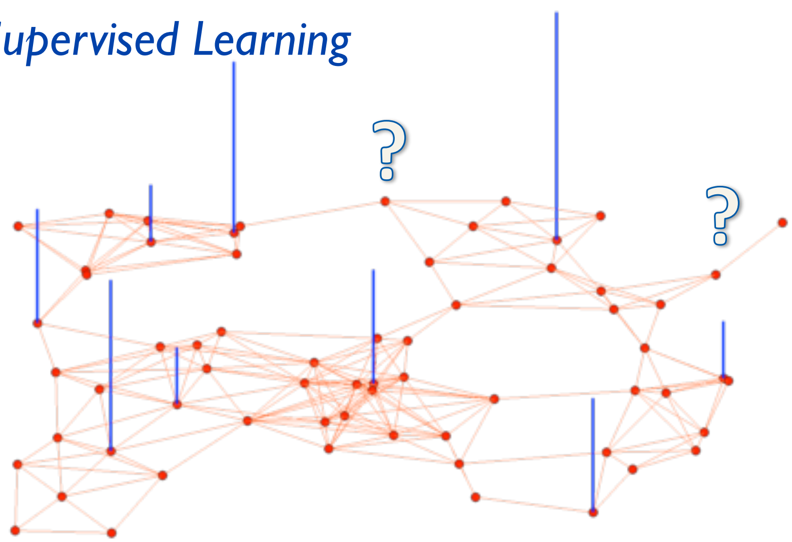


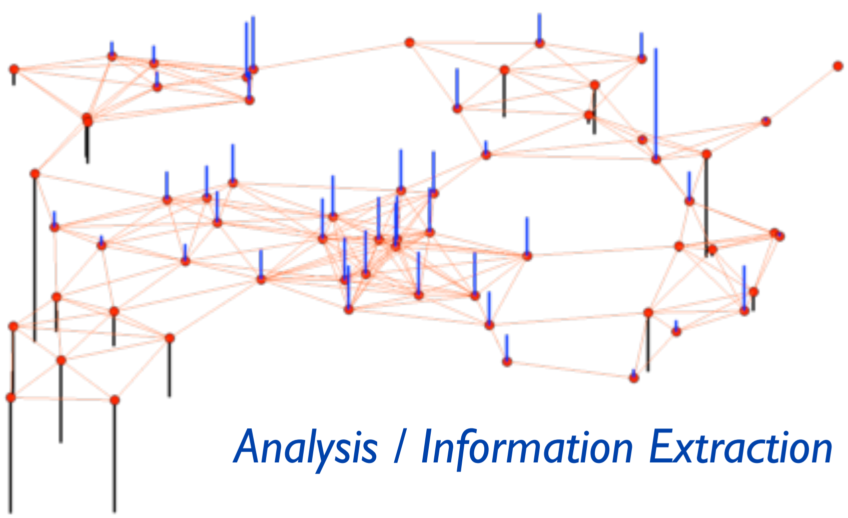
Earth data source: Frederik Simons



Denoising

Semi-Supervised Learning

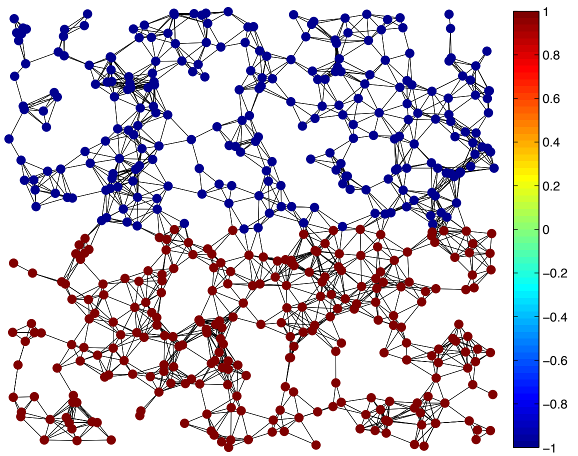




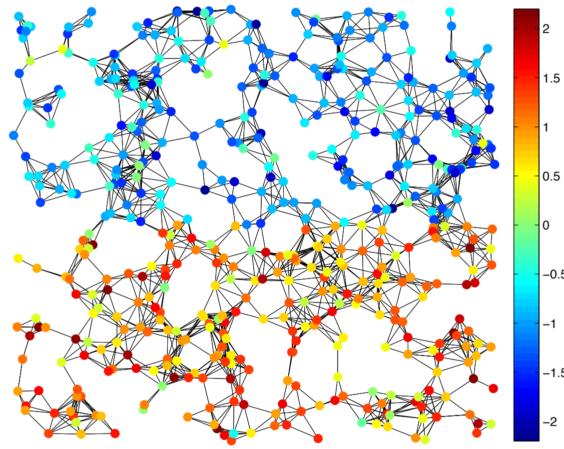
Analysis / Information Extraction

Motivating Example: Denoising

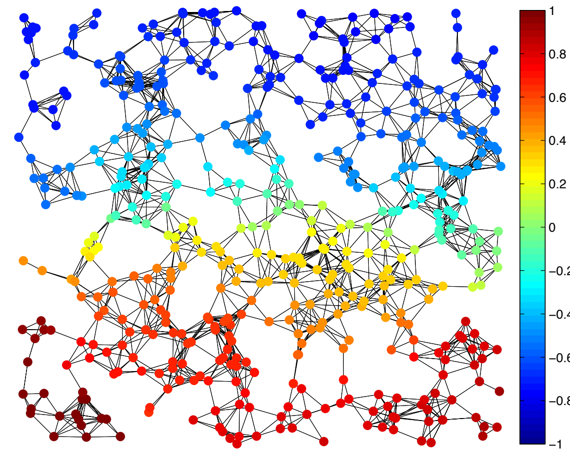
- Tikhonov regularization for denoising: $\text{argmin}_f \left\{ ||f - y||_2^2 + \gamma f^T \mathcal{L} f \right\}$



Original

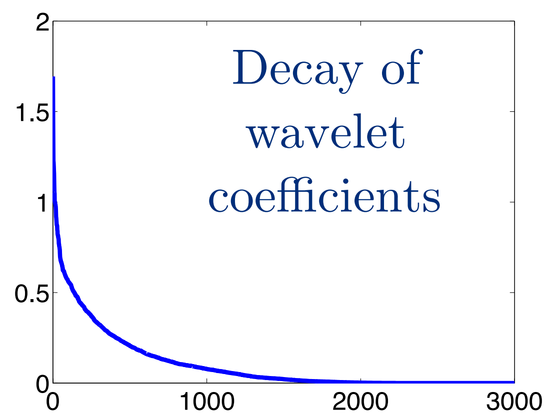


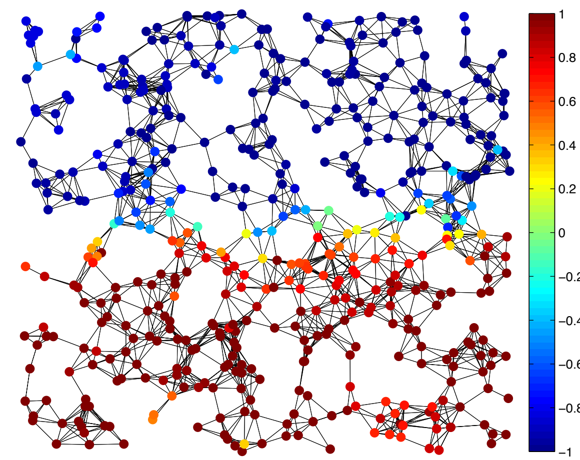
Noisy



Denoised

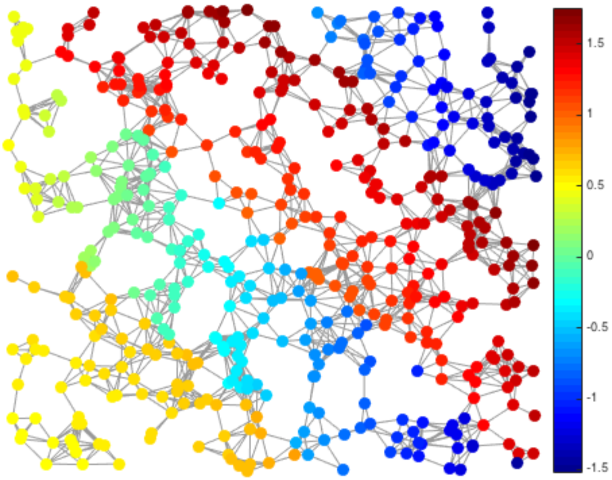
- Wavelet denoising: $\text{argmin}_a \left\{ ||f - W^* a||_2^2 + \gamma ||a||_{1,\mu} \right\}$



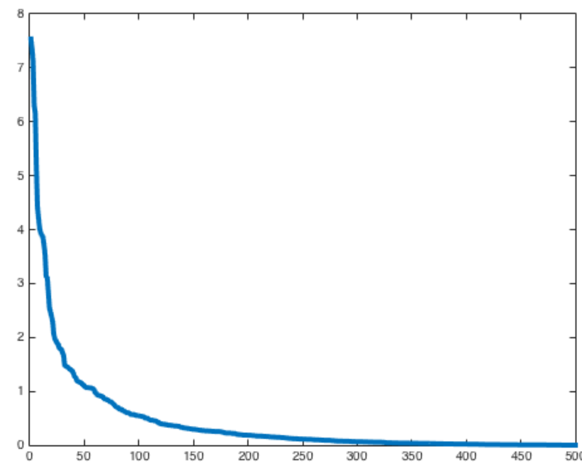


Motiving Example: Compression

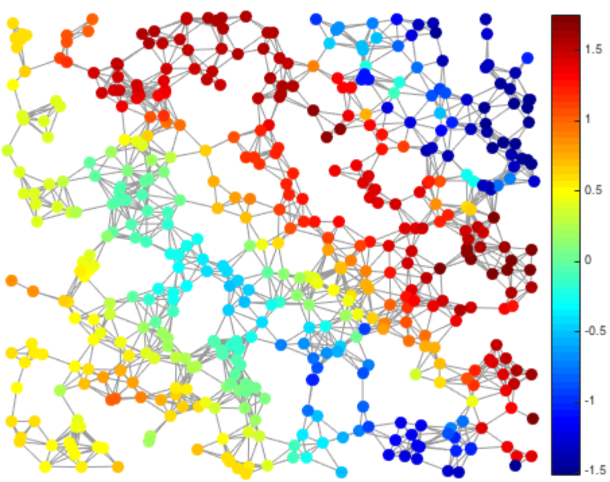
Piecewise-Smooth Signal with Discontinuities



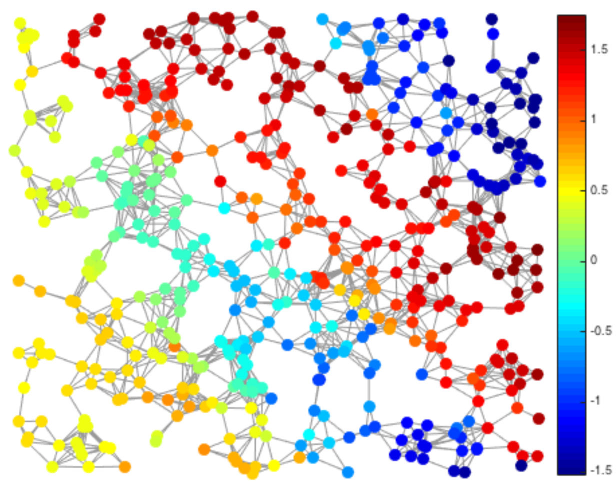
Diffusion Wavelet Coefficients, Sorted by Magnitude



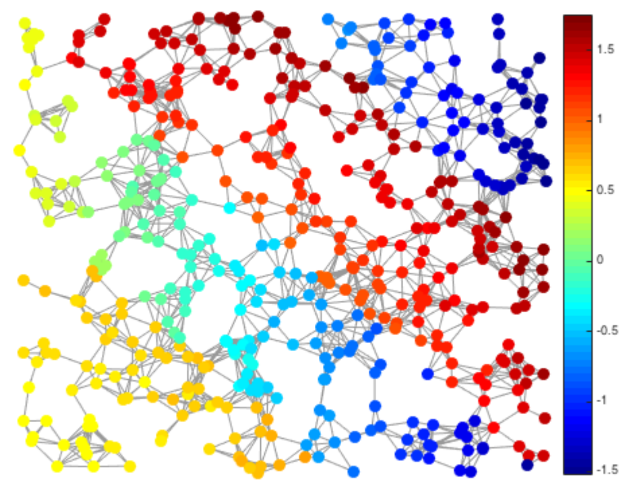
Reconstruction from 10% of Coefficients



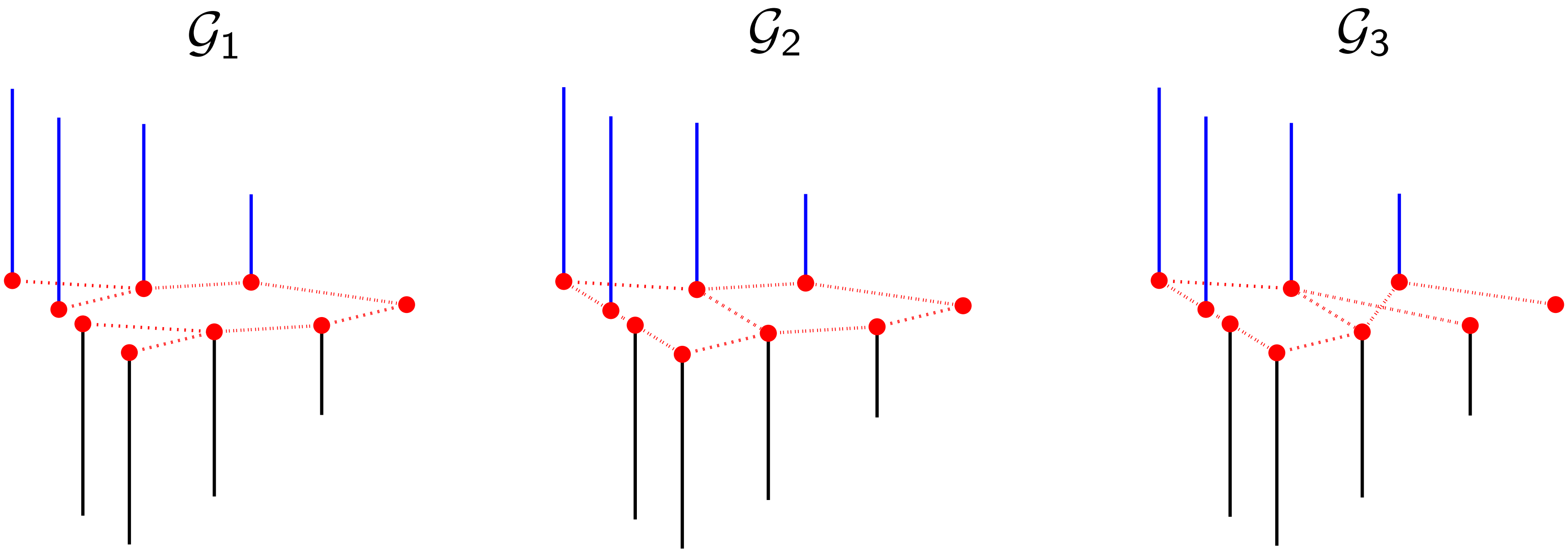
Reconstruction from 20% of Coefficients



Reconstruction from 50% of Coefficients

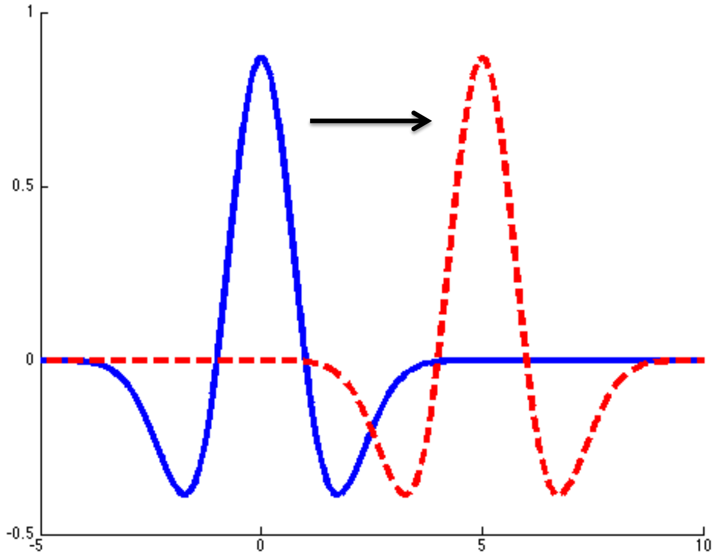


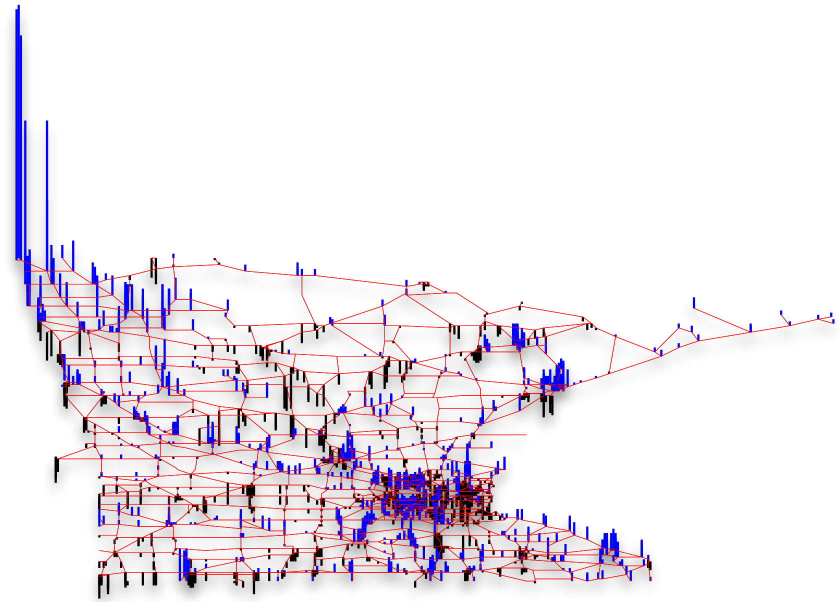
Why Do We Need New Dictionaries?



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

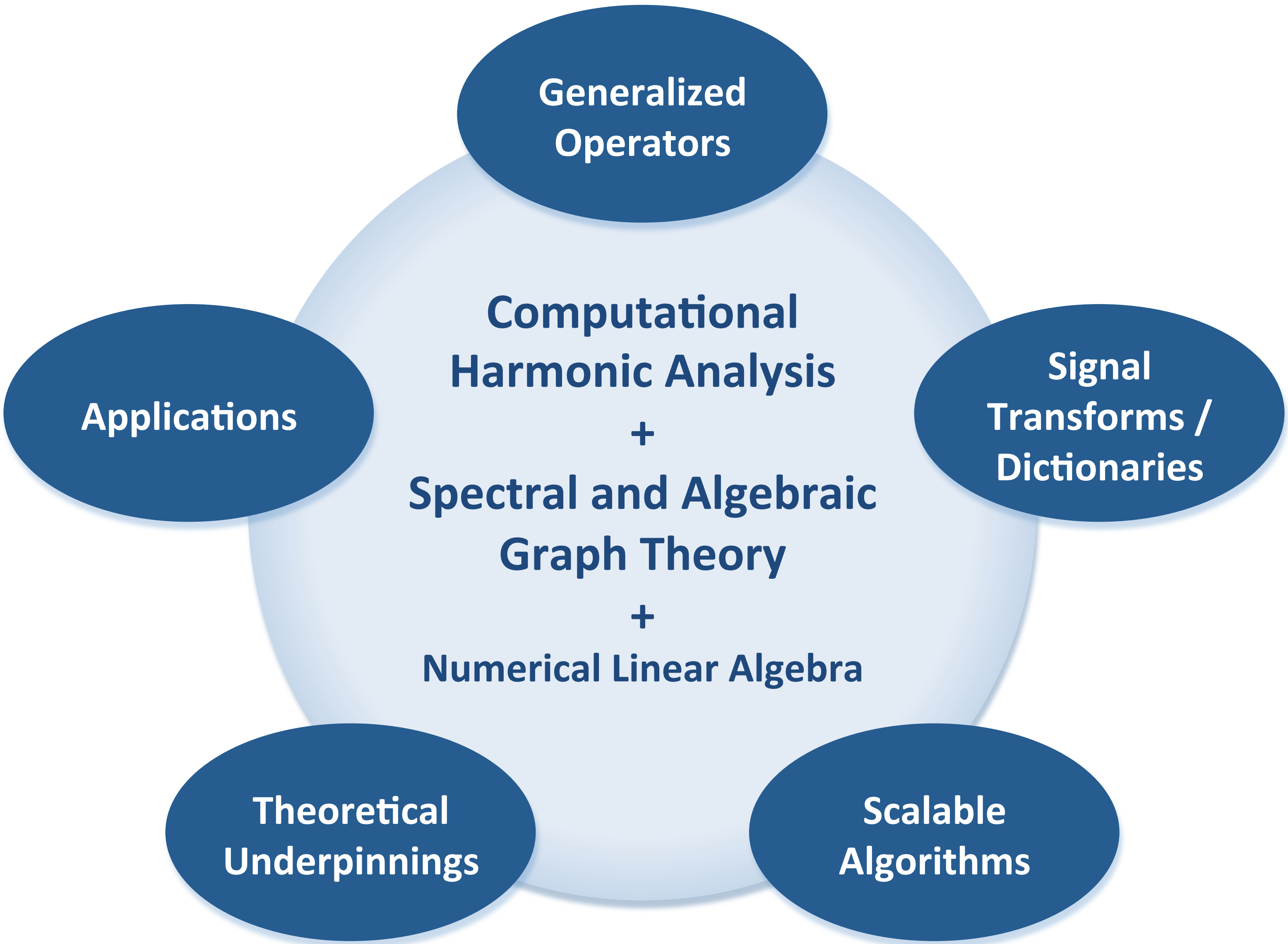
The Essence of the Problem





- 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

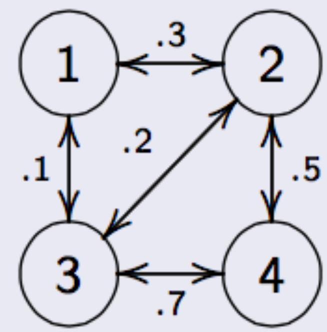
Approach: Leverage Intuition from Euclidean Settings to Develop New Mathematical Tools for the Graph Setting



The Graph Spectral Domain(s)

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

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

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

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

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

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

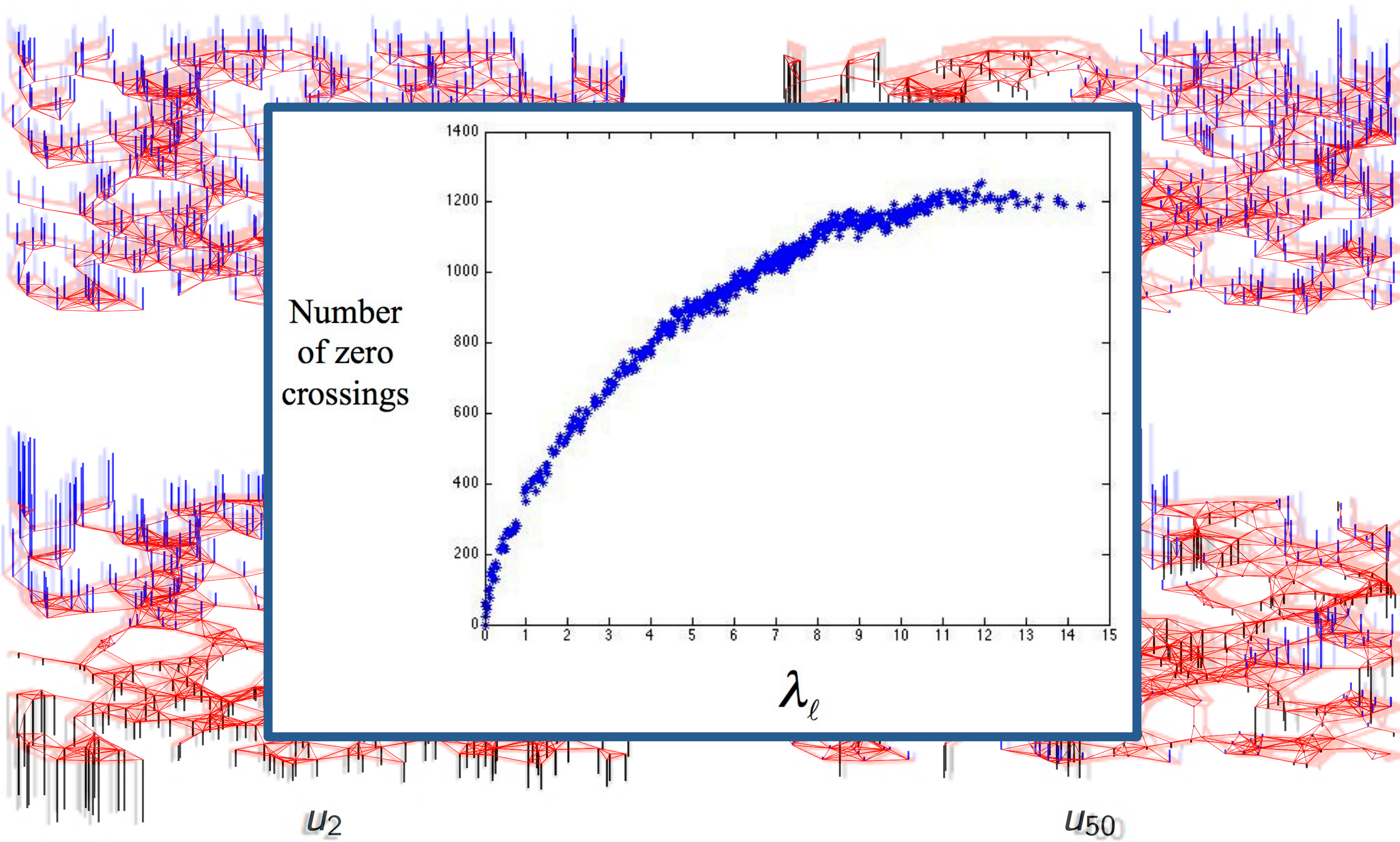
  ordered w.l.o.g. s.t.

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

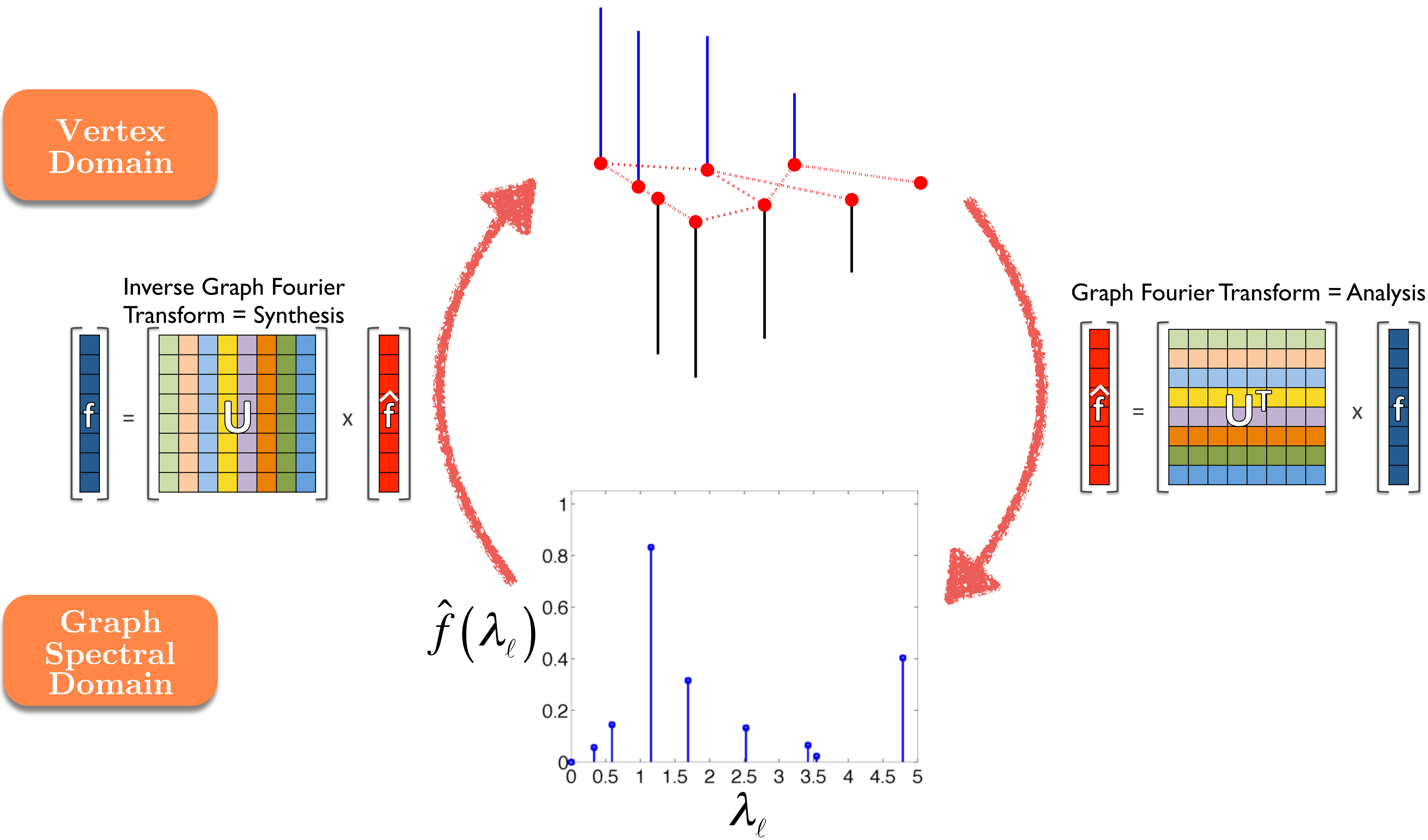
- Discrete difference operator: $(\mathcal{L}f)(i) = \sum_{j \in \mathcal{N}_i} W_{i,j}[f(i) - f(j)]$

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

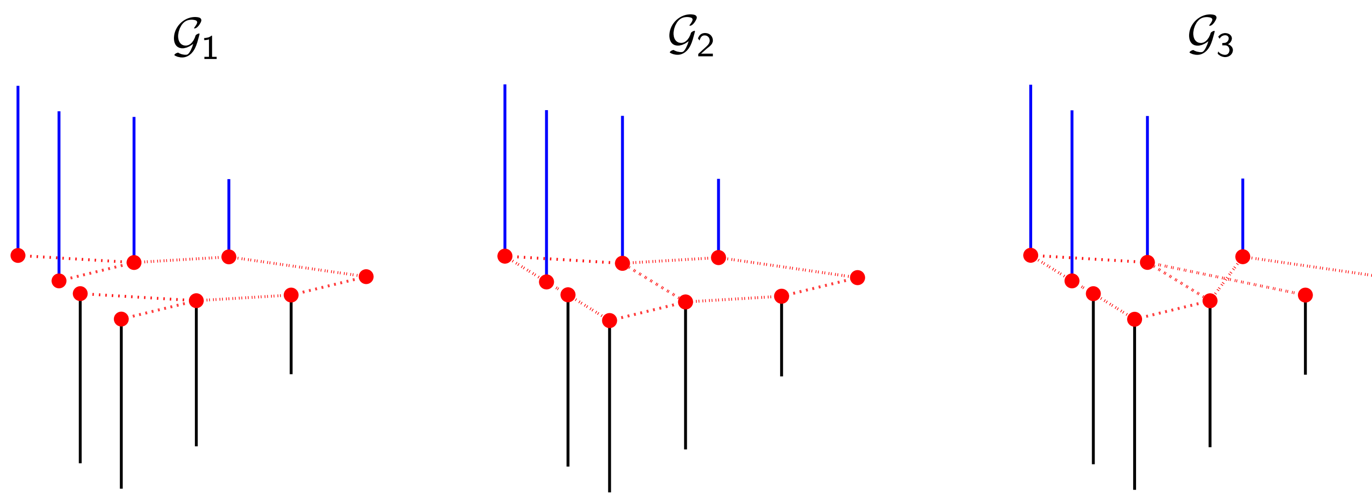


Graph Signals in Two Domains

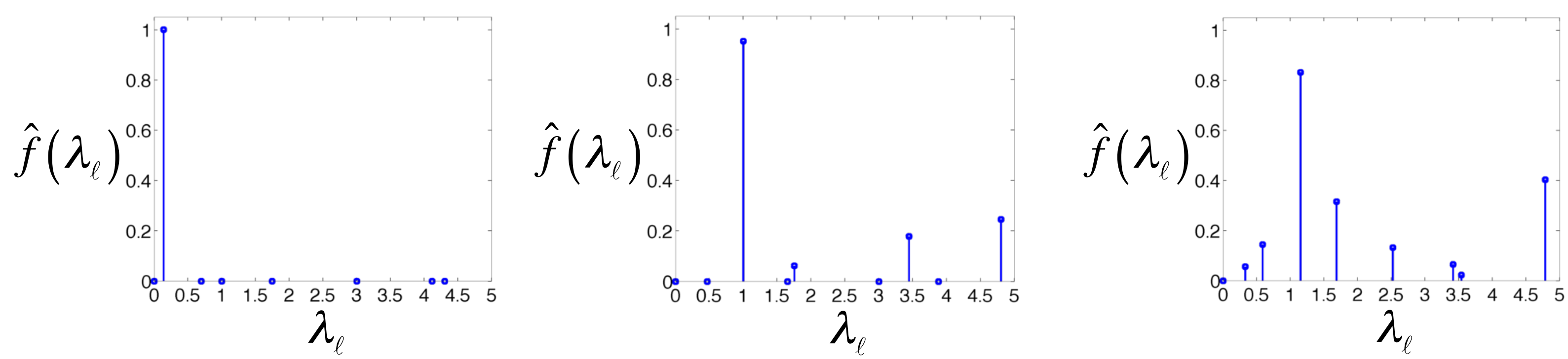


The GFT Incorporates the Graph Structure

Vertex Domain



Graph Spectral Domain

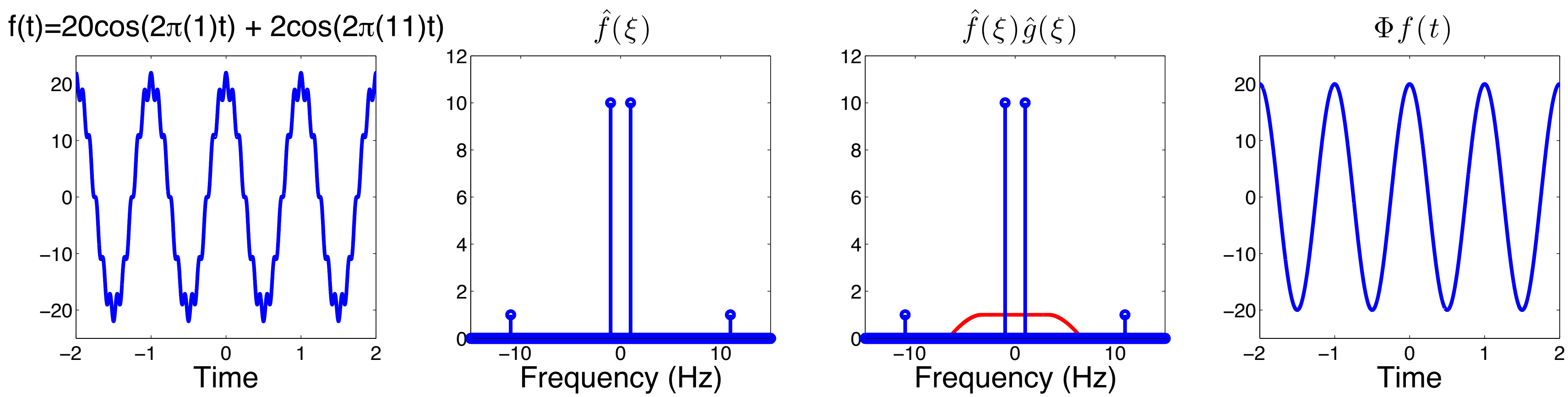


Problem 1:
Efficient Graph Spectral Filtering

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

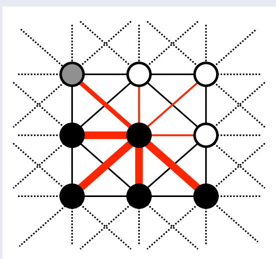
$$f(t) \longrightarrow \boxed{\text{FT}} \longrightarrow \hat{f}(\xi) \longrightarrow \boxed{\hat{g}} \longrightarrow \hat{g}(\xi)\hat{f}(\xi) \longrightarrow \boxed{\text{IFT}} \longrightarrow \Phi f(t)$$



Example: Image Denoising by Low-Pass Graph Filtering

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

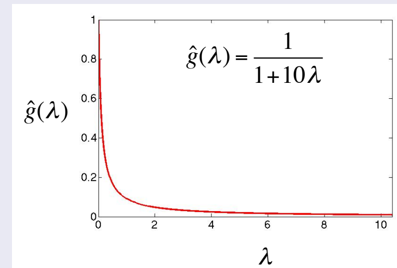
Semi-Local Graph



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

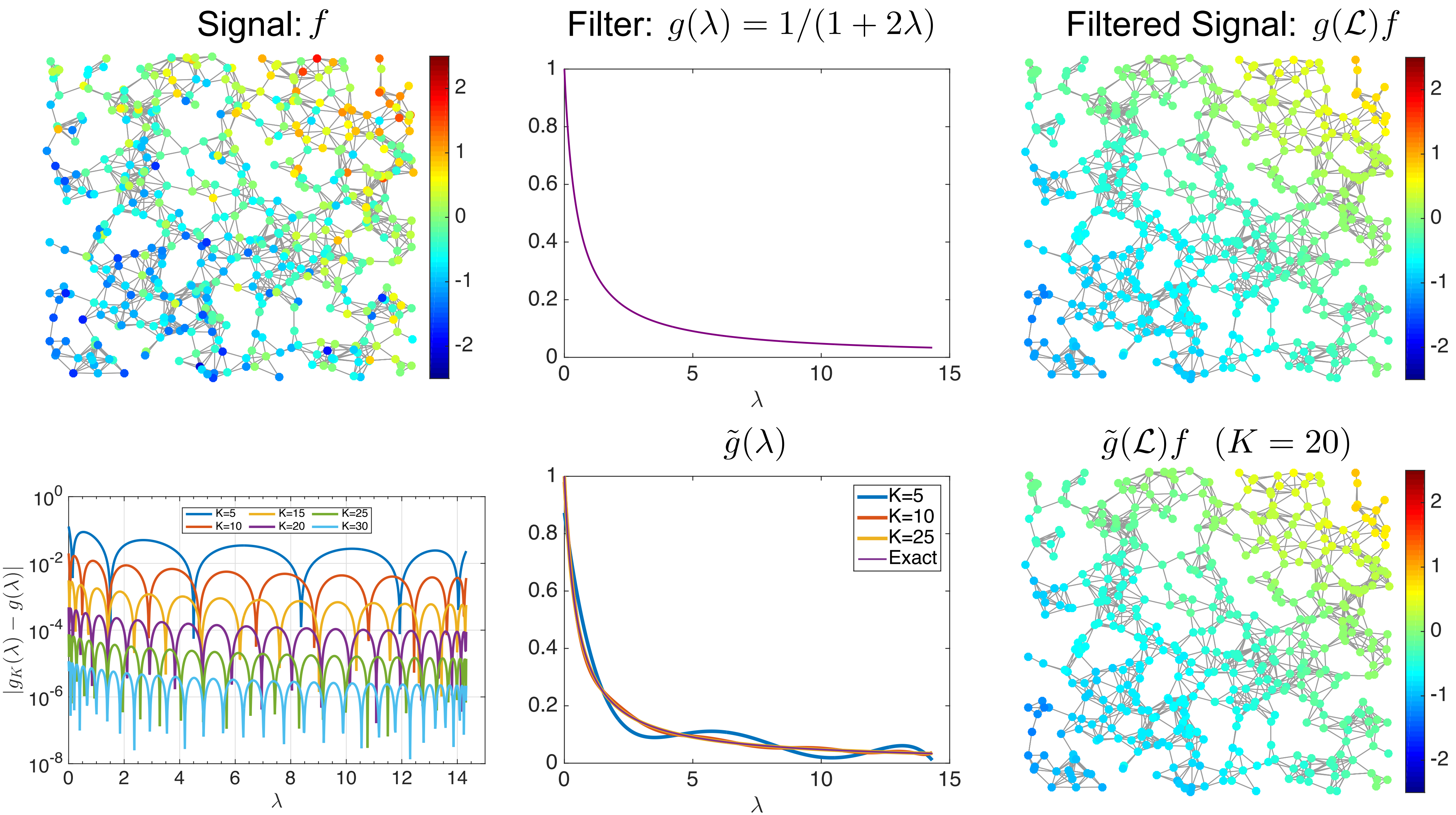
- Filtering: $g(\mathcal{L})f = Ug(\Lambda)U^* f$
- Too expensive to compute U and Λ for large graphs
- Common approach: estimate $\lambda_{\max}$ and approximate the filter g on the interval $[0, \lambda_{\max}]$ by a polynomial, rational, or spline function
- Example: Truncated Chebyshev polynomial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

Chebyshev Polynomial Approximation Example



Shuman et al., "Chebyshev polynomial approximation for distributed signal processing," DCOSS, 2011

Shuman et al., "Distributed signal processing via Chebyshev polynomial approximation," SIPN, 2018

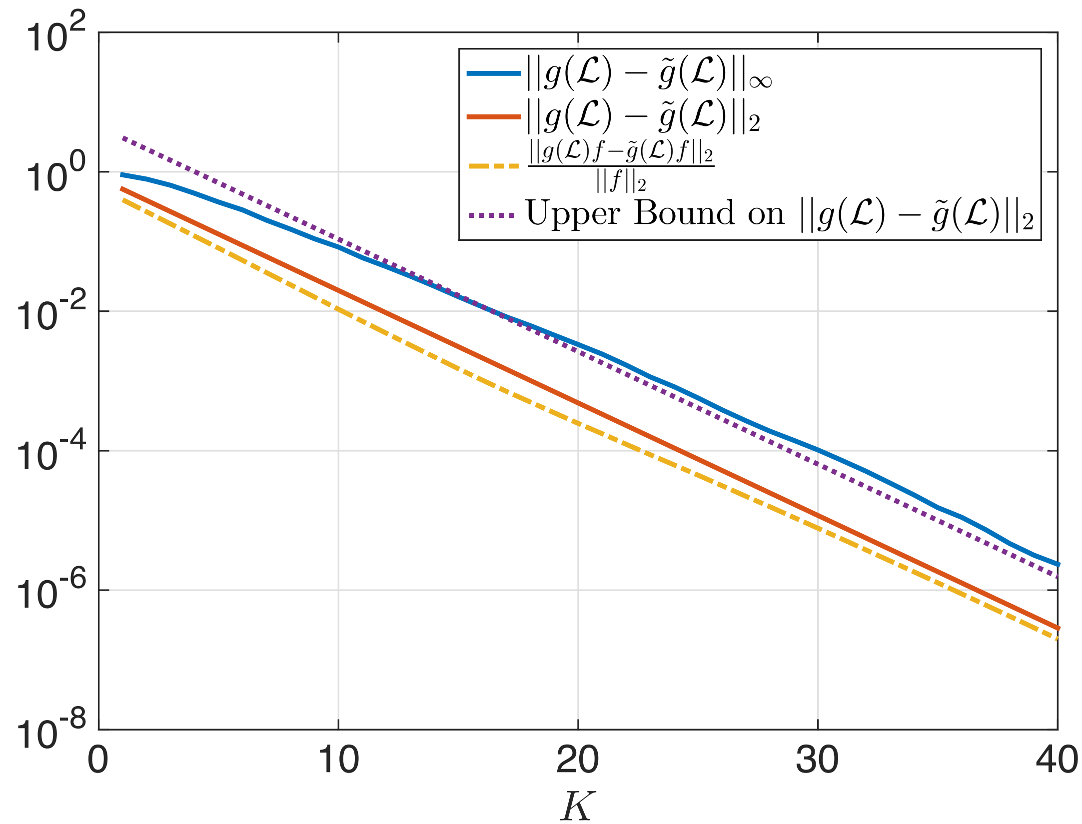
Chebyshev Polynomial Approximation Example

$$||g(\mathcal{L}) - \tilde{g}(\mathcal{L})||_2 = \boxed{\max_{\ell=0,1,\ldots,N-1} |g(\lambda_\ell) - \tilde{g}(\lambda_\ell)|}$$

$$\leq \boxed{\sup_{\lambda \in [0,\lambda_{\max}]} |g(\lambda) - \tilde{g}(\lambda)|} = \mathcal{O}(\rho^{-K})$$

For analytic filters

- Classical polynomial approximation theory focuses on the whole interval
- Error actually only depends on the error at the eigenvalues
- If we knew the eigenvalues, we could do a discrete least squares approximation with a fixed order polynomial

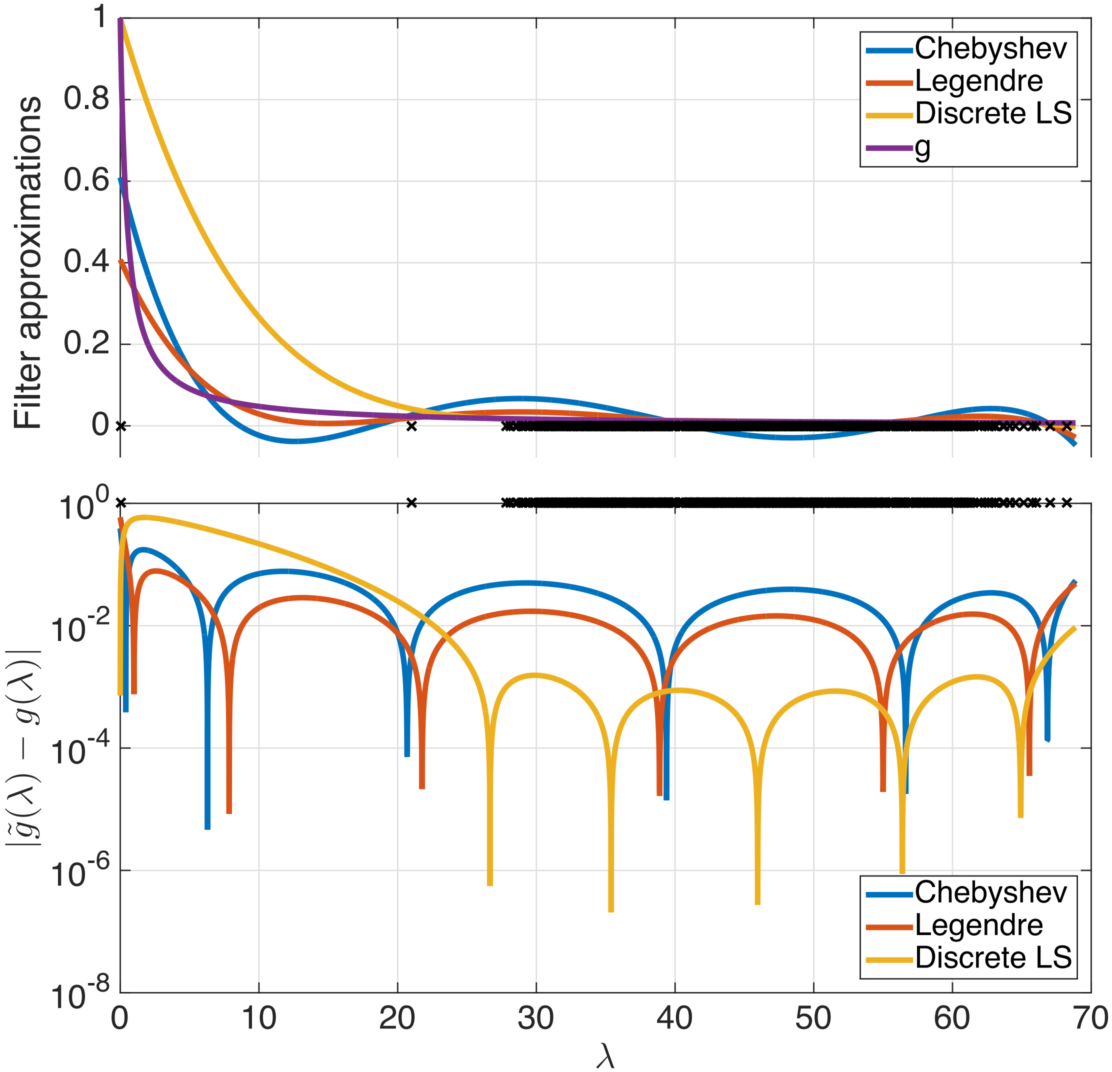


Shuman et al., "Chebyshev polynomial approximation for distributed signal processing," DCOSS, 2011

Shuman et al., "Distributed signal processing via Chebyshev polynomial approximation," SIPN, 2018

Spectrum-Adapted Function Approximation

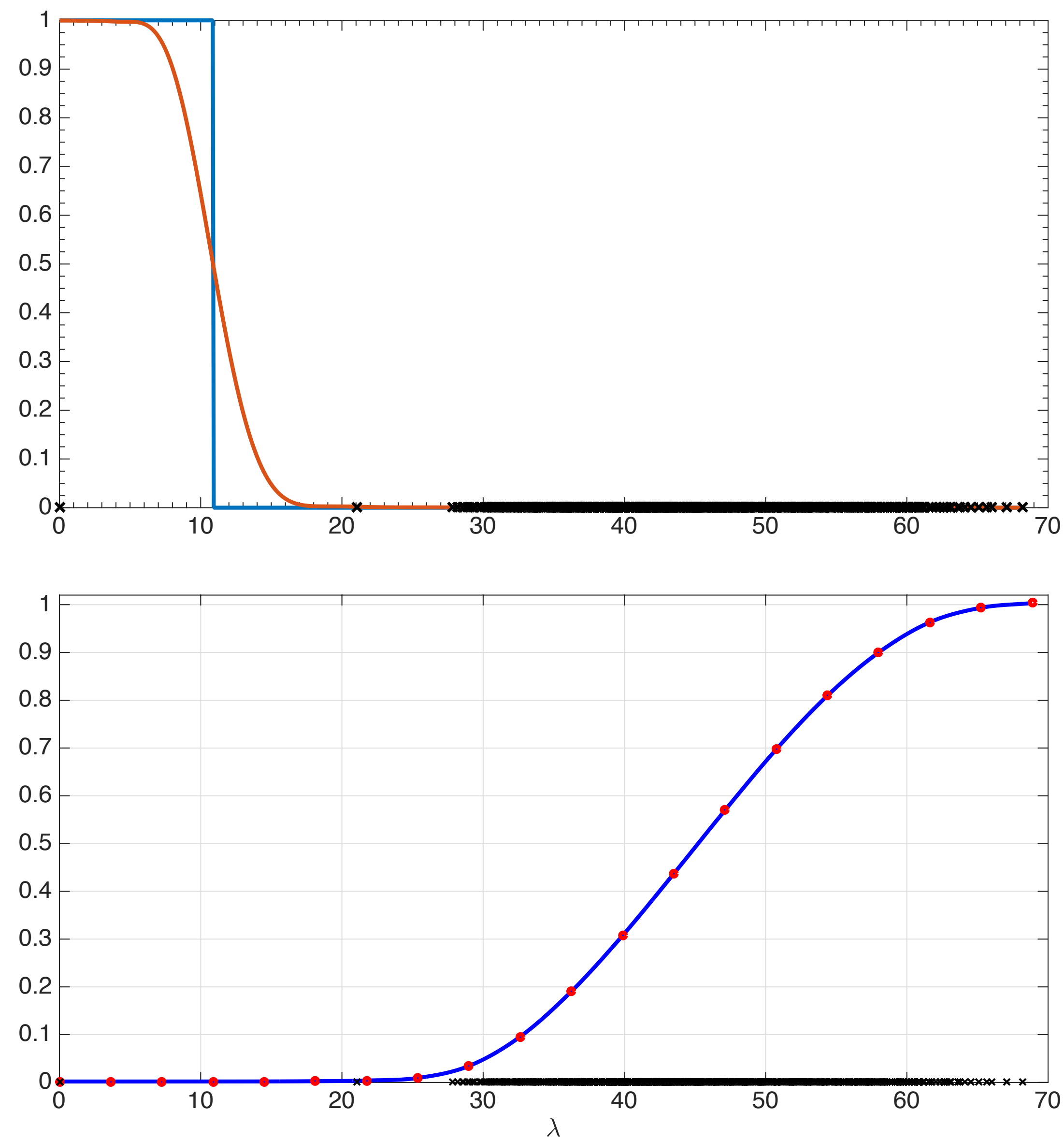
Degree 5 polynomial approximation to g(x)=1/(1+2x) on a random Erdös-Renyi graph with 500 vertices and edge probability 0.2



Estimating the Spectral Distribution Example: Kernel Polynomial Method

$$
\begin{aligned}
F_{\mathcal{L}}(\lambda) &= \frac{1}{N}\sum_{\ell=0}^{N-1} \mathbb{1}_{\{\lambda_\ell \le \lambda\}} \\
&= \frac{1}{N}\operatorname{tr}(\phi_\lambda(\mathcal{L})) \\
&= \frac{1}{N}\mathbb{E}[r^\top \phi_\lambda(\mathcal{L}) r] \\
&\approx \frac{1}{N}\left(\frac{1}{J}\sum_{j=1}^{J} r^{(j)\top} \boxed{\phi_\lambda(\mathcal{L}) r^{(j)}}\right)
\end{aligned}
$$

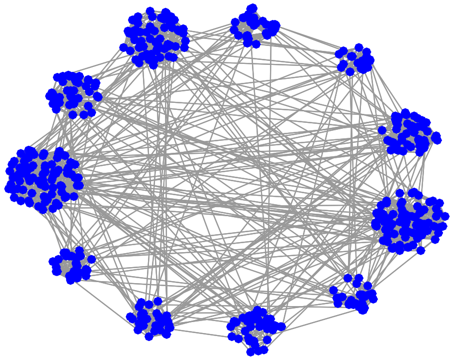
Polynomial filter

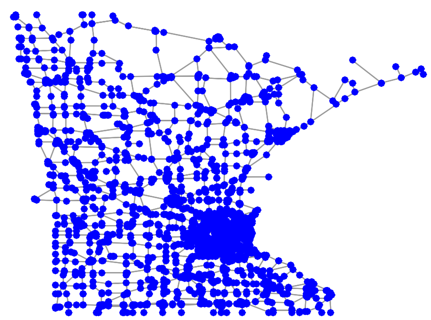


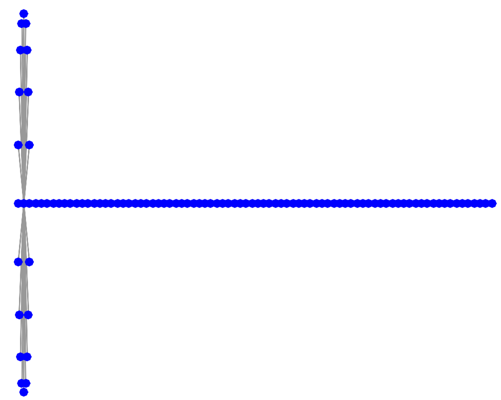
- Monotonic cubic interpolation
- Iterate, updating the filter approximation

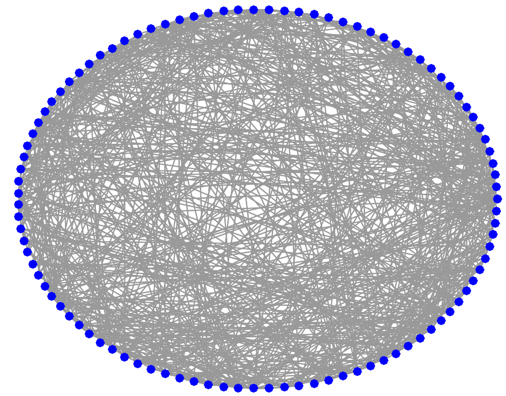
Lin et al., "Approximating spectral densities of large matrices," *SIAM Review*, 2016

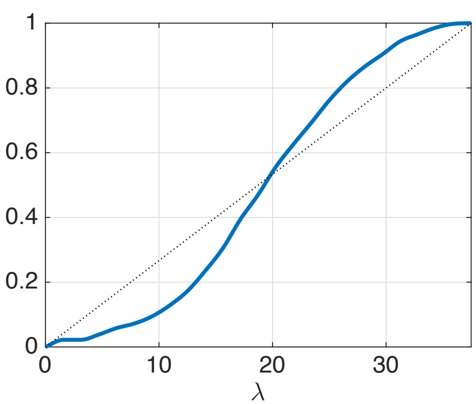
Spectral Distribution Examples

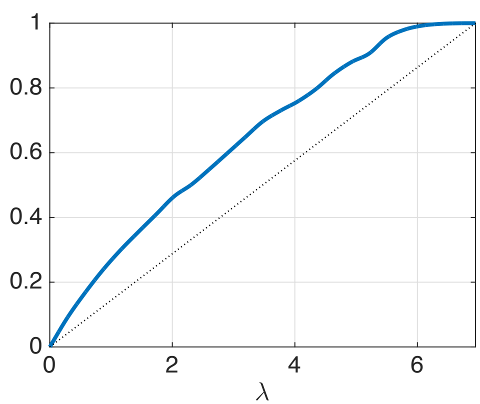


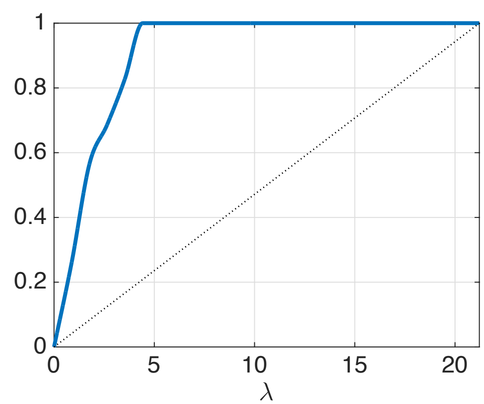


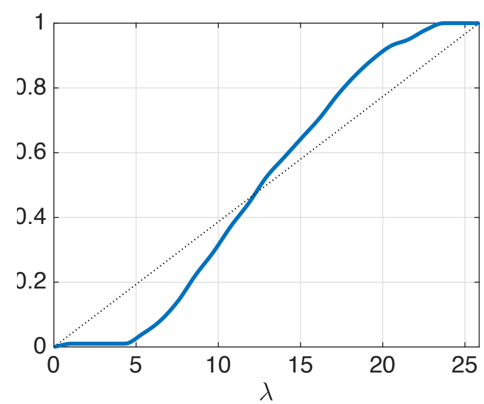




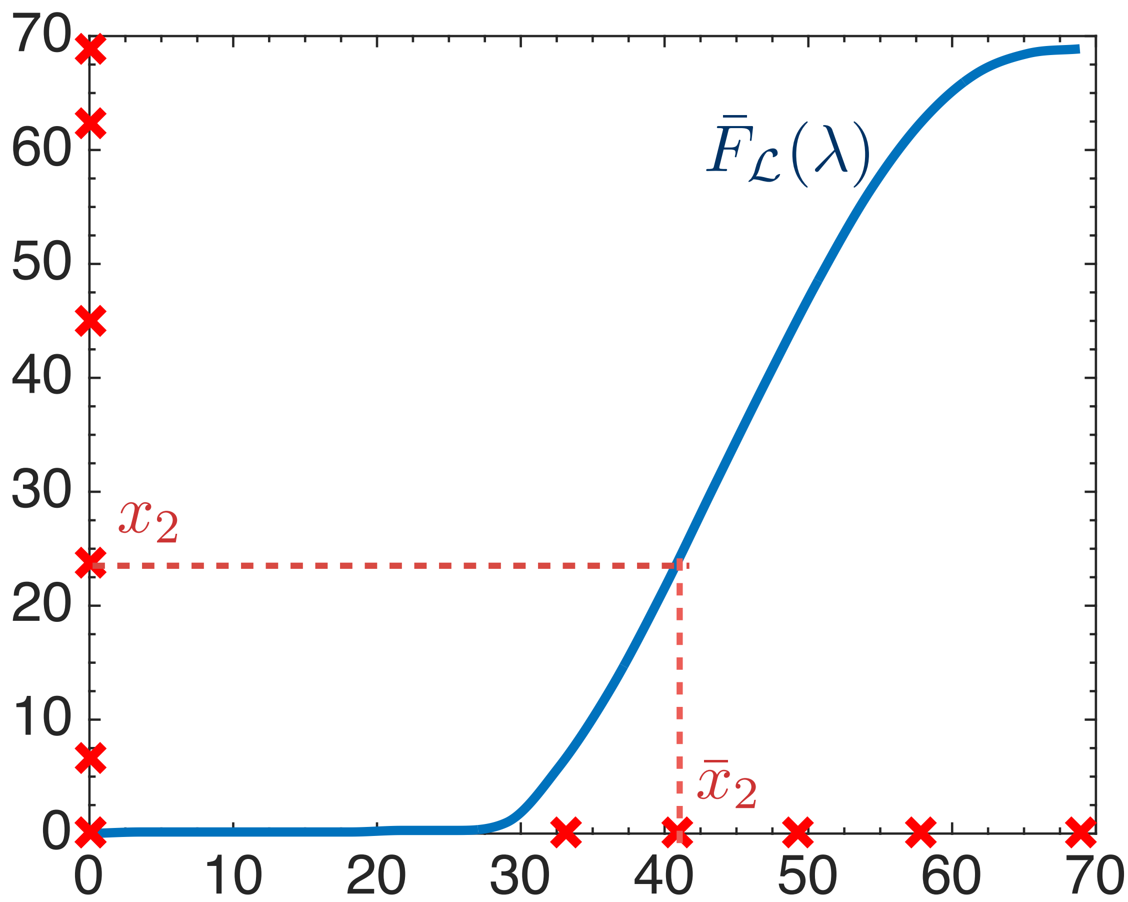








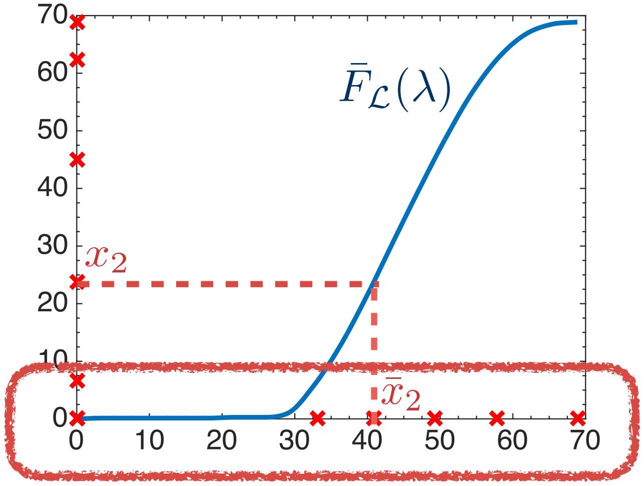
Approach 1: Warped Interpolation Points

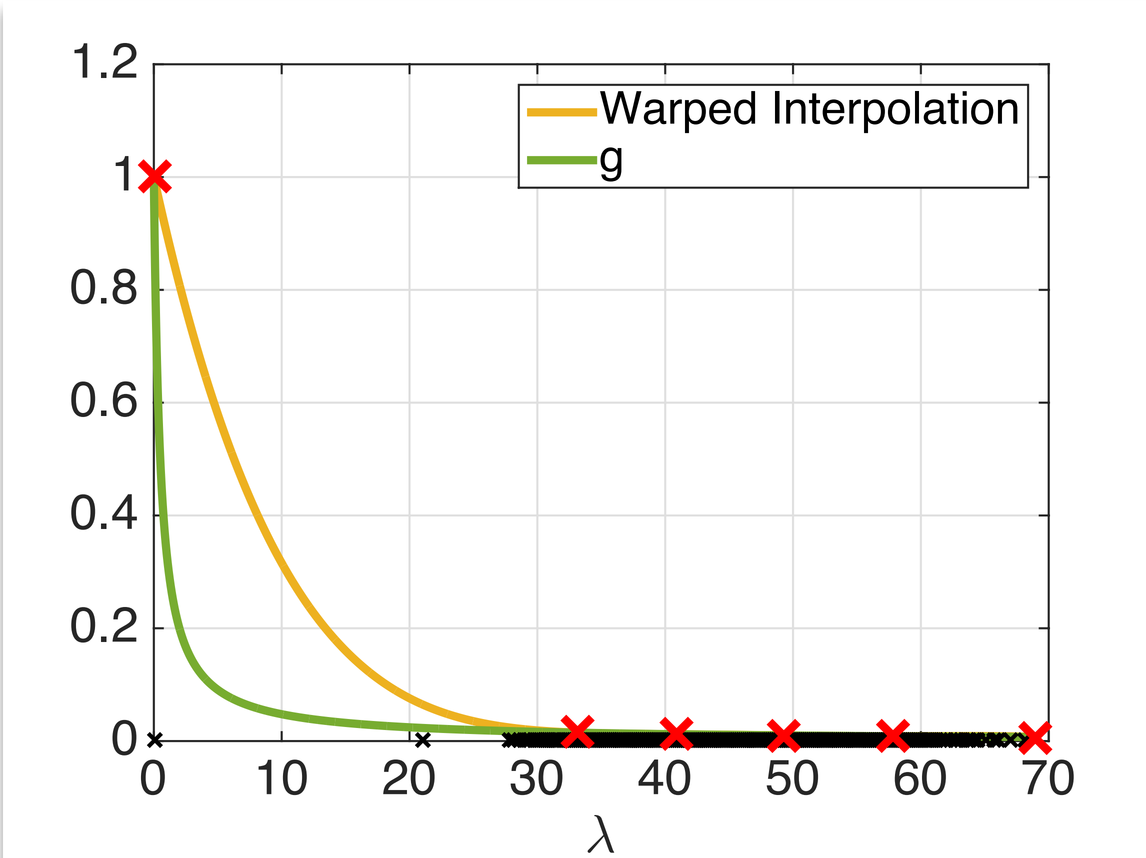


1. Compute the spectrum-adapted interpolation points:

$$\{\bar{x}_k = \bar{F}_{\mathcal{L}}^{-1}(x_k)\}_{k=0}^{K}$$

Approach 1: Warped Interpolation Points

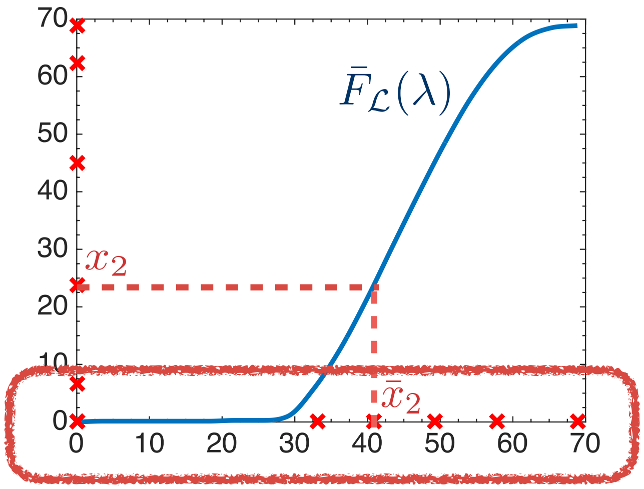


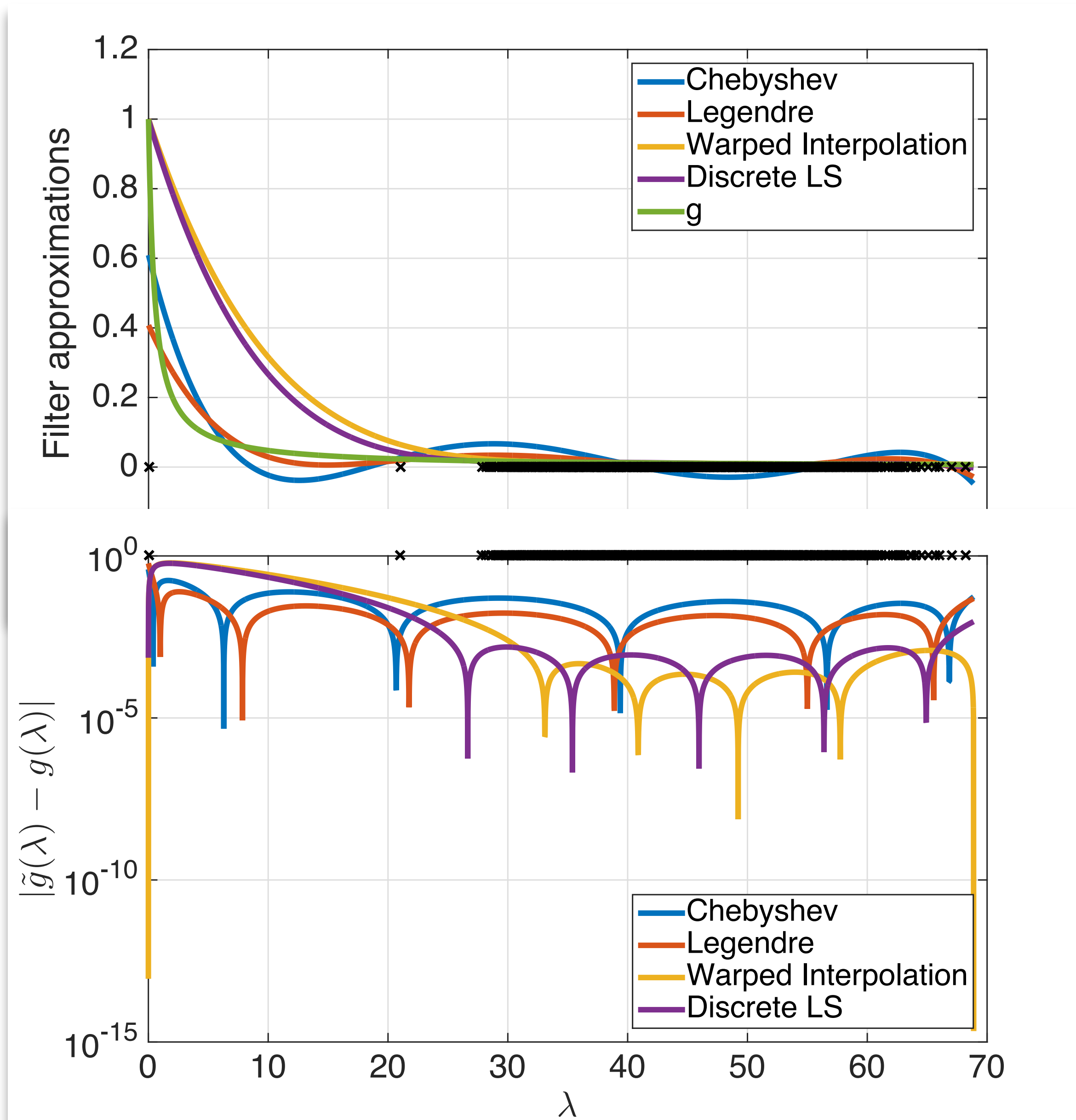


2. Perform barycentric interpolation on $\{\bar{x}_k, g(\bar{x}_k)\}_{k=0}^{K}$ to evaluate $\tilde{g}(x)$ at a cost of $\mathcal{O}(K^2)$ (once) and $\mathcal{O}(K)$ (for each x)

Trefethen, *Approximation Theory and Approximation Practice*, 2013

Approach 1: Warped Interpolation Points





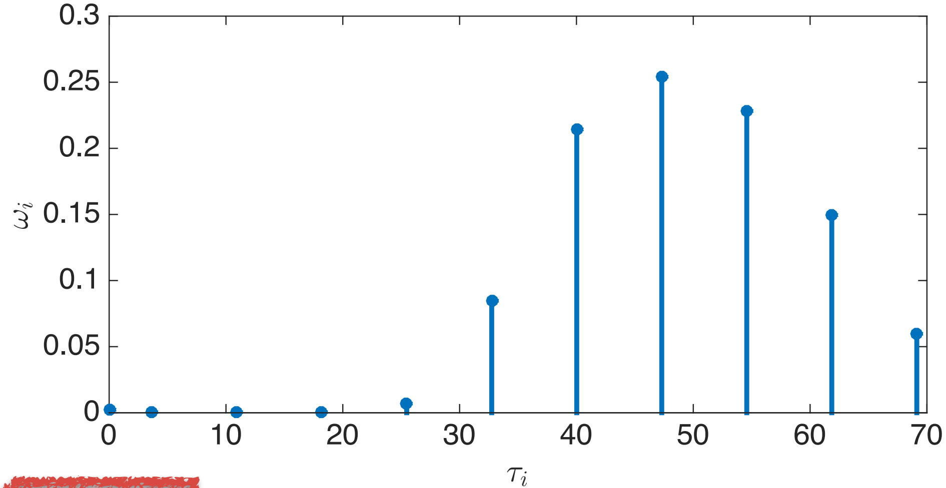
Questions?

Reactions?

Suggestions?

Approach 2: Matrix-Adapted Orthogonal Polynomials

- Polynomials orthogonal with respect to a given measure:



- Inner product $\langle g, \pi_k \rangle = \sum_{i=1}^{M} \omega_i g(\tau_i) \boxed{\pi_k(\tau_i)}$ and truncated expansion:

$$g(\lambda) = \sum_{k=0}^{\infty} \frac{\langle g, \pi_k \rangle}{\langle \pi_k, \pi_k \rangle} \pi_k(\lambda) \approx \sum_{k=0}^{K} \frac{\langle g, \pi_k \rangle}{\langle \pi_k, \pi_k \rangle} \pi_k(\lambda) =: \tilde{g}(\lambda)$$

Computable from the discrete measure via a Lanczos type algorithm

- Three-term recurrence:

$$\pi_{k+1}(\lambda) = (\lambda - \boxed{\alpha_k})\pi_k(\lambda) - \boxed{\beta_k}\pi_{k-1}(\lambda)$$

Can recursively evaluate at the abscissae $\{\tau_i\}$ up through π_K

Gautschi, *Orthogonal Polynomials: Computation and Approximation*, 2004

Problem 2:
Non-Uniform Random Sampling and Reconstruction

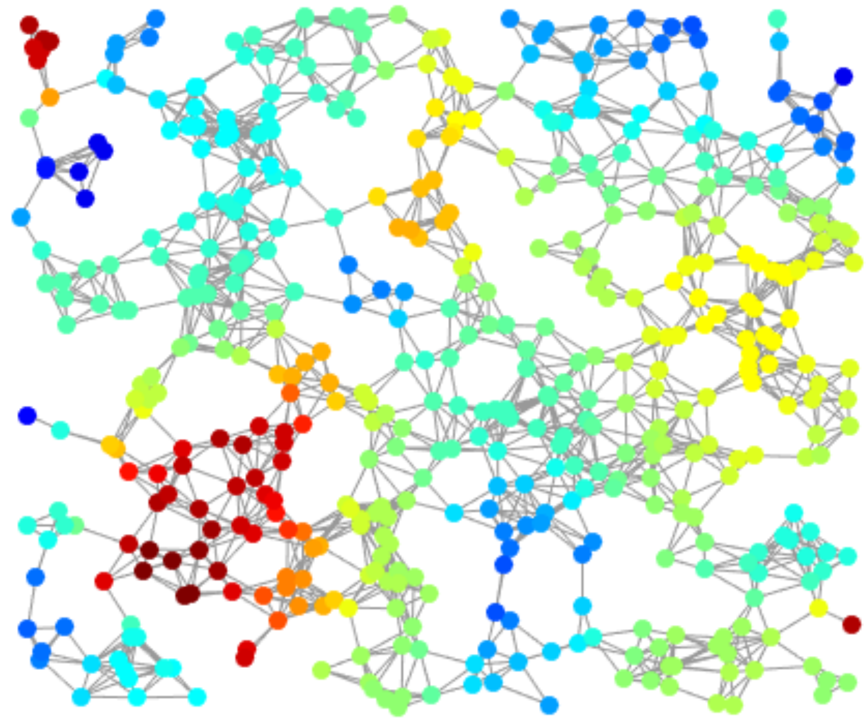
Sampling and Interpolation

- 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

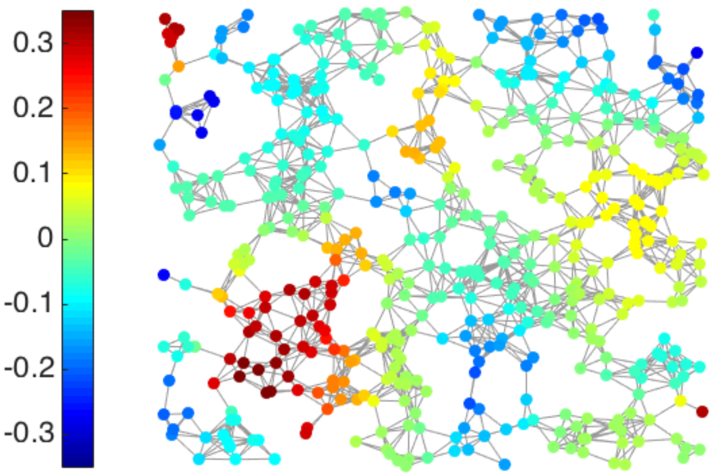


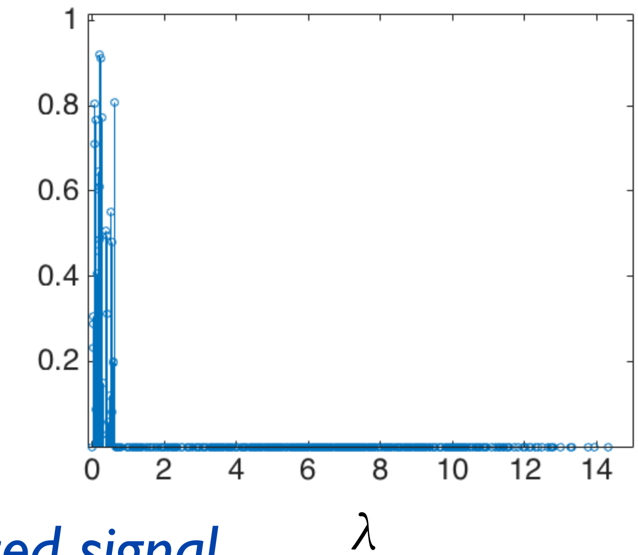
Can we recover all 500 values of this signal from 30 measurements? If so, where should we take those measurements?

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

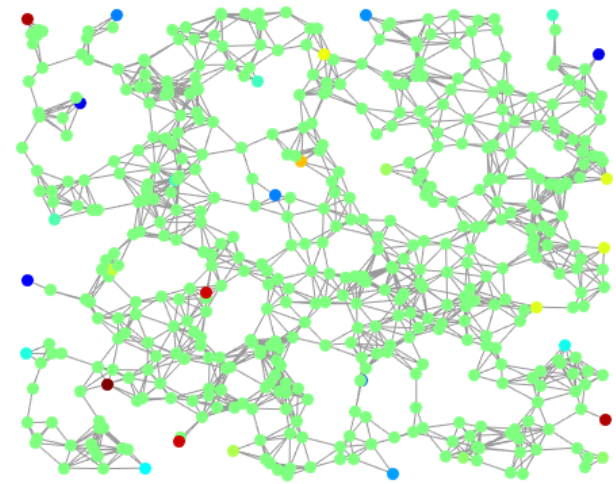
Sampling and Interpolation: Signals Concentrated on Spectral Bands

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





Bandlimited signal



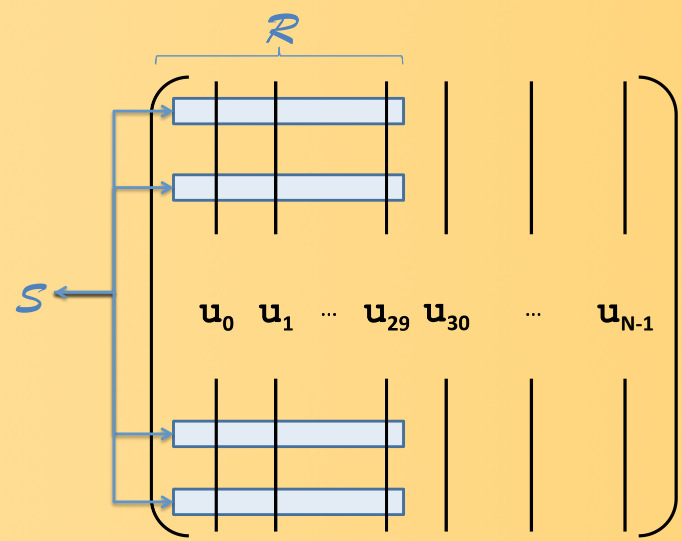
Uniqueness set

1. Recover graph Fourier coefficients:

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

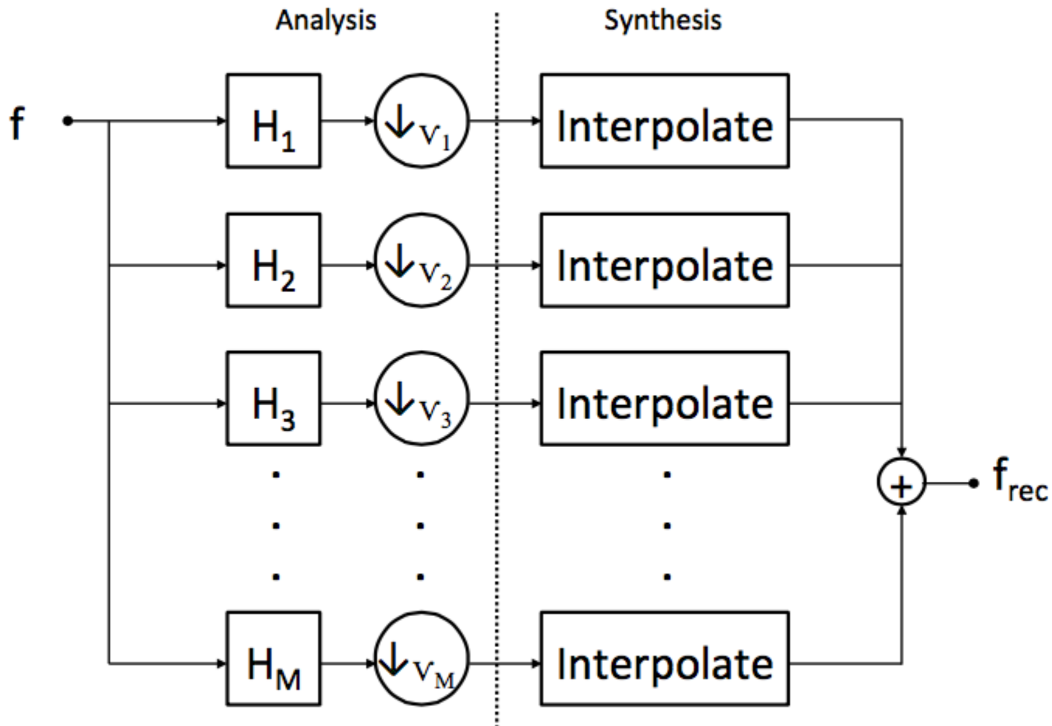
2. Interpolation / reconstruction:

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



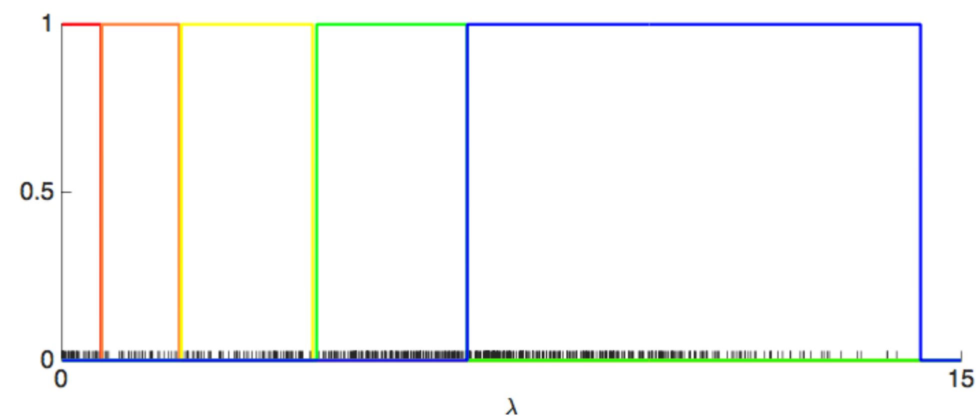
M-Channel Critically Sampled Graph Filter Bank

Architecture



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

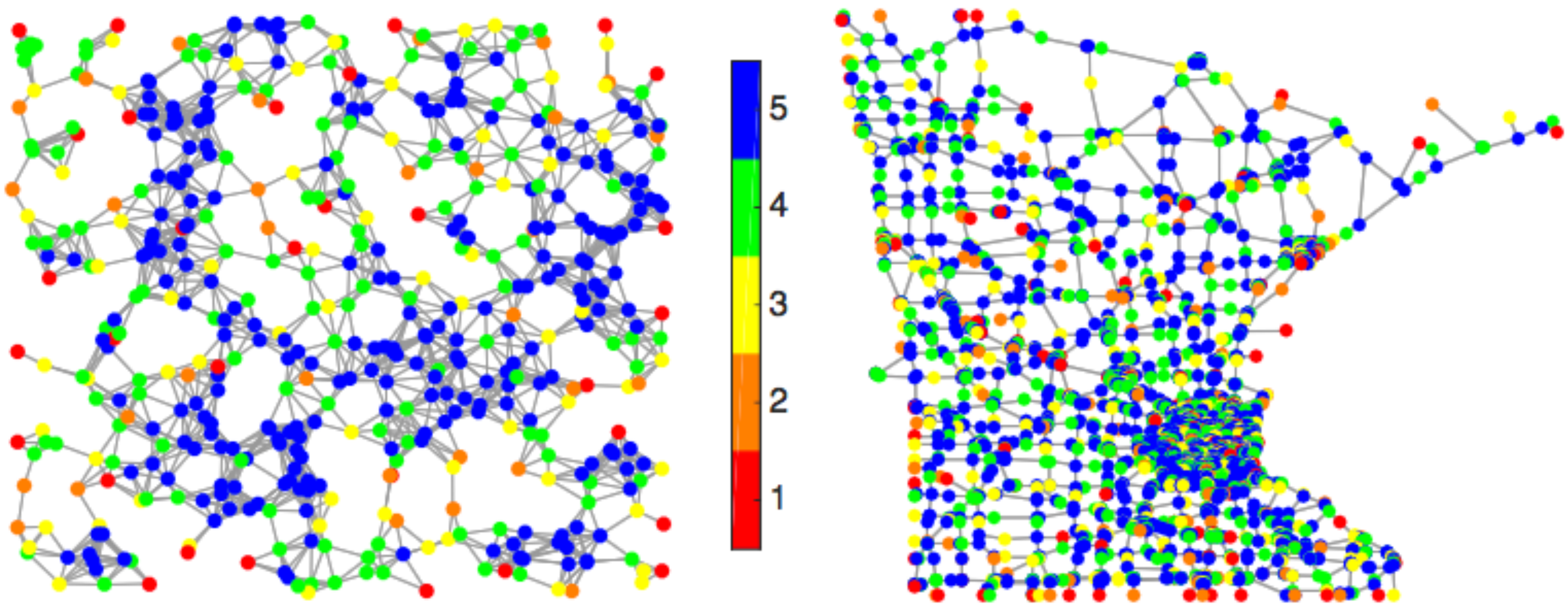
Ideal Filter Bank



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:

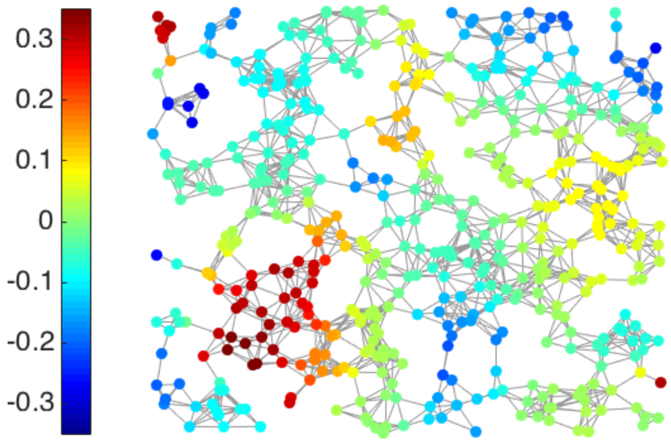


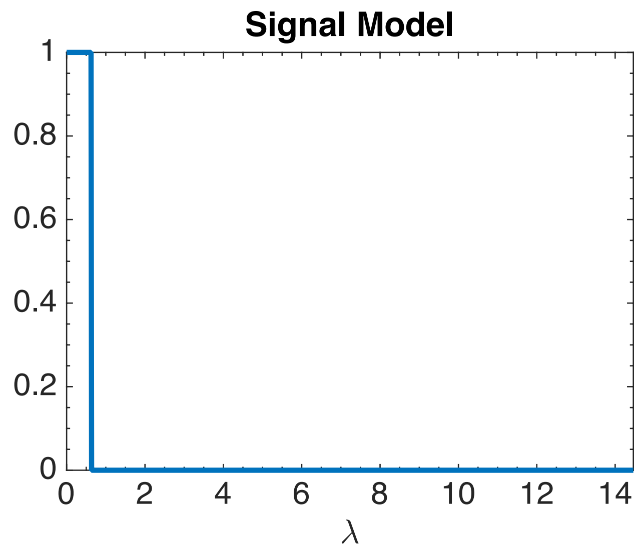
- To avoid a full eigendecomposition, we would like to use random, non-uniform sampling and fast, approximate reconstruction methods

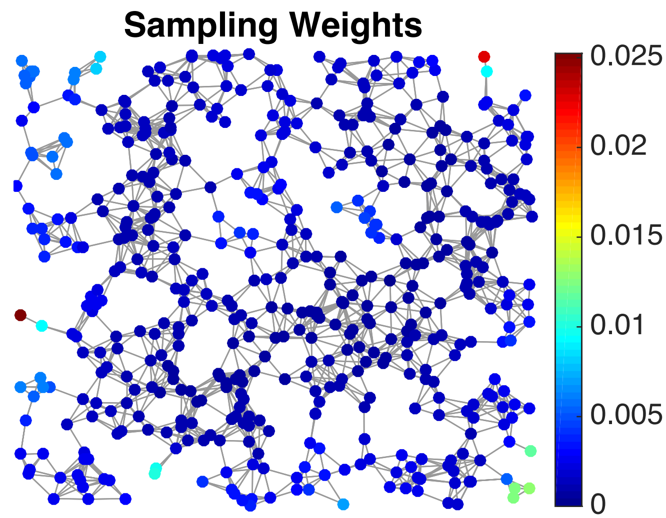
Greene and Magnanti, “Some abstract pivot algorithms,” *SIAM J. Appl. Math.*, 1975

Non-Uniform Random Sampling

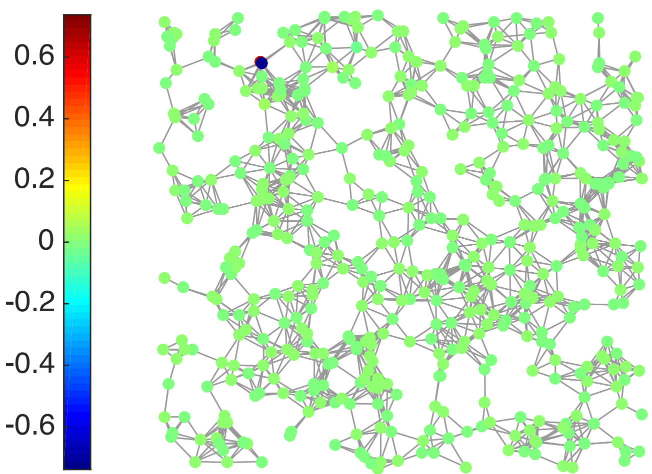
Lowpass (smooth) signals

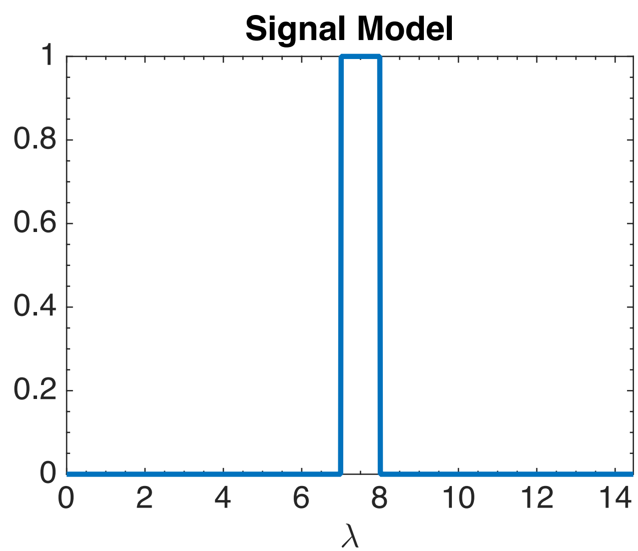


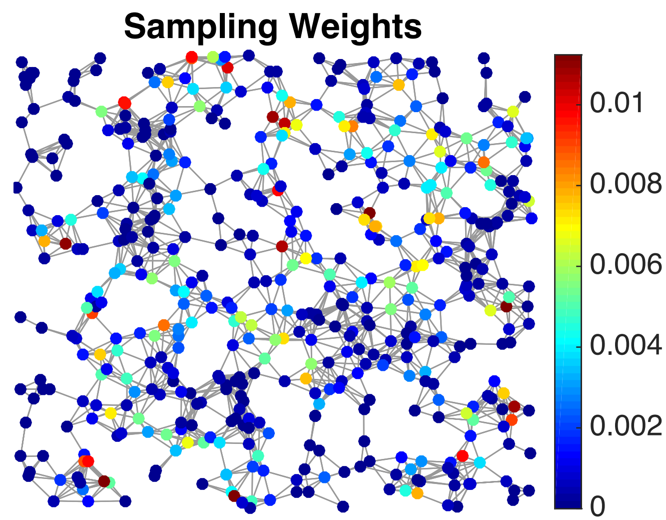




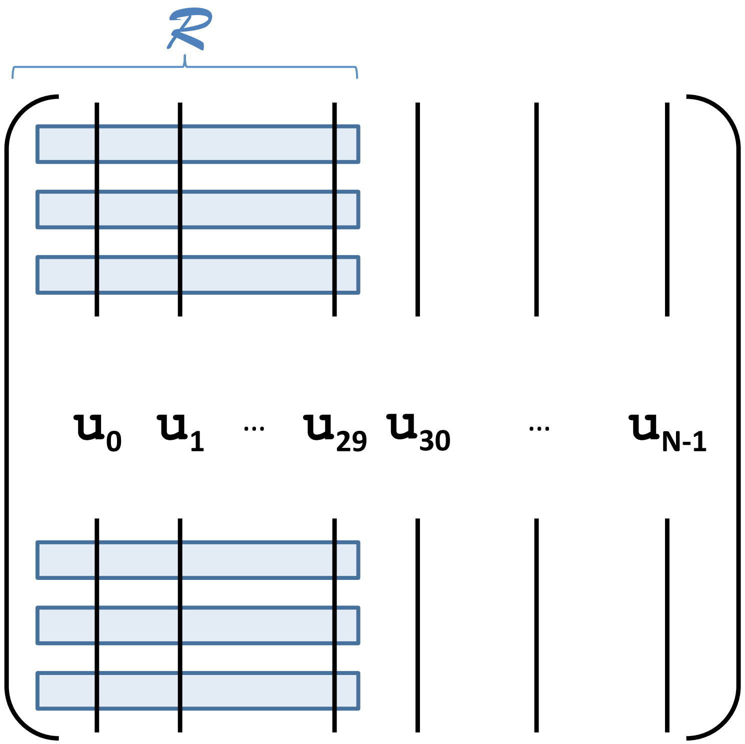
Midpass signals

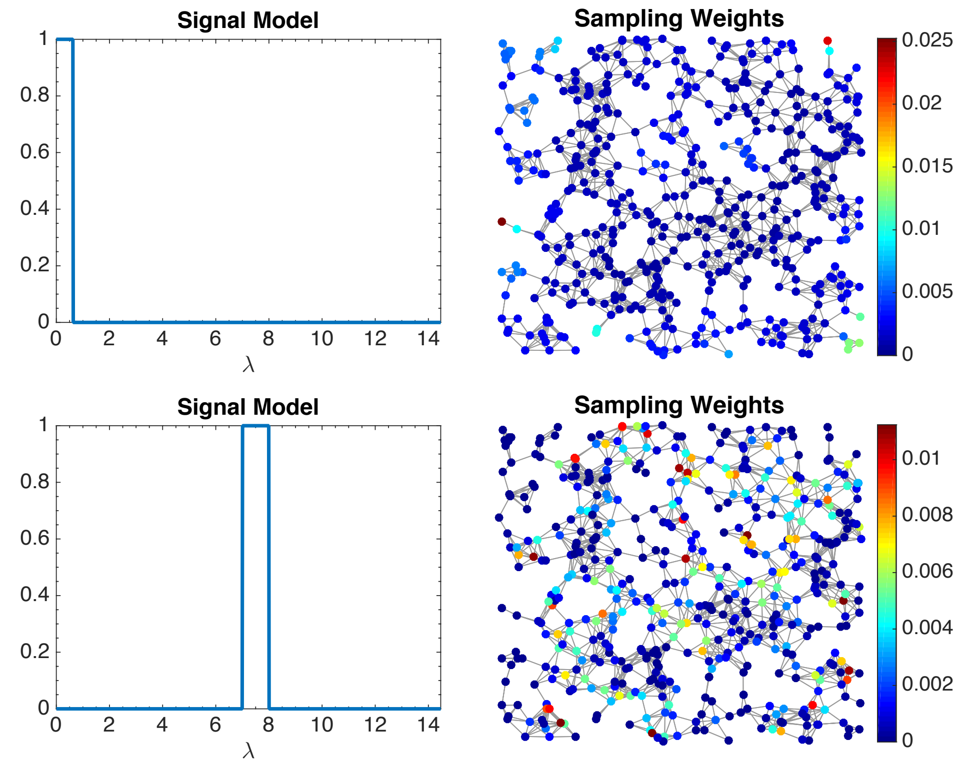






Non-Uniform Random Sampling





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

approximation of the filter describing the signal model

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

independent random entries that follow a standard normal dist.

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

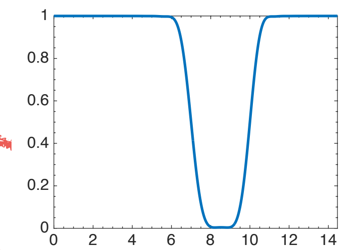
Efficient Reconstruction

downsampling operator

approximate by convex optimization problem

$$\min_{z \in \mathrm{col}(U_{:,\mathcal{R}})} \|P_\Omega^{-1/2}(Mz - f_{\mathcal{S}})\|_2$$

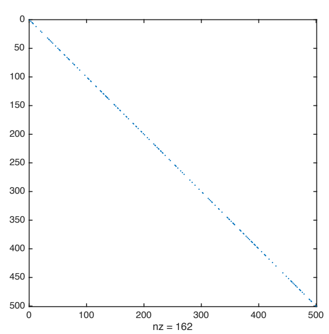
signal model space



$$\min_{z \in \mathbb{R}^N} \left\{ \|P_\Omega^{-1/2}(Mz - f_{\mathcal{S}})\|_2^2 + \gamma z^\top (1 - \tilde{h})(\mathcal{L}) z \right\}$$

optimality condition

$$\left(M^\top P_\Omega^{-1} M + \gamma (1 - \tilde{h})(\mathcal{L}) \right) z = M^\top P_\Omega^{-1} f_{\mathcal{S}}$$



solve with preconditioned conjugate gradient??

preconditioner:

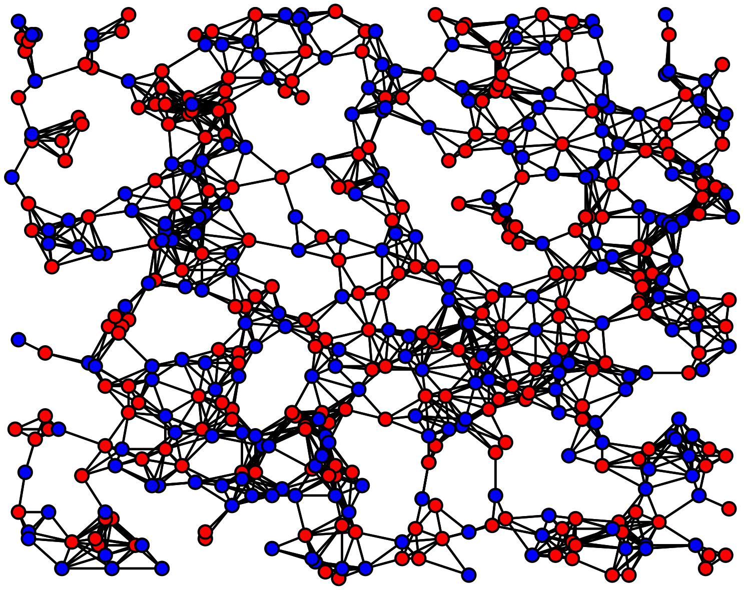
$$\mathrm{diag}(1./(p_i + \gamma))$$

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

Zero to Four Additional Problems

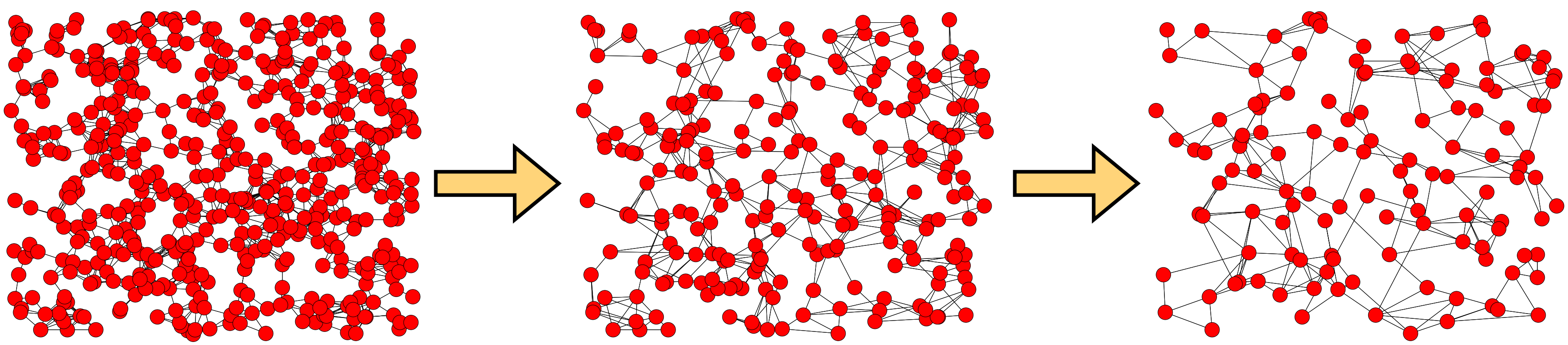
Multiresolutions of Graphs

Downsampling

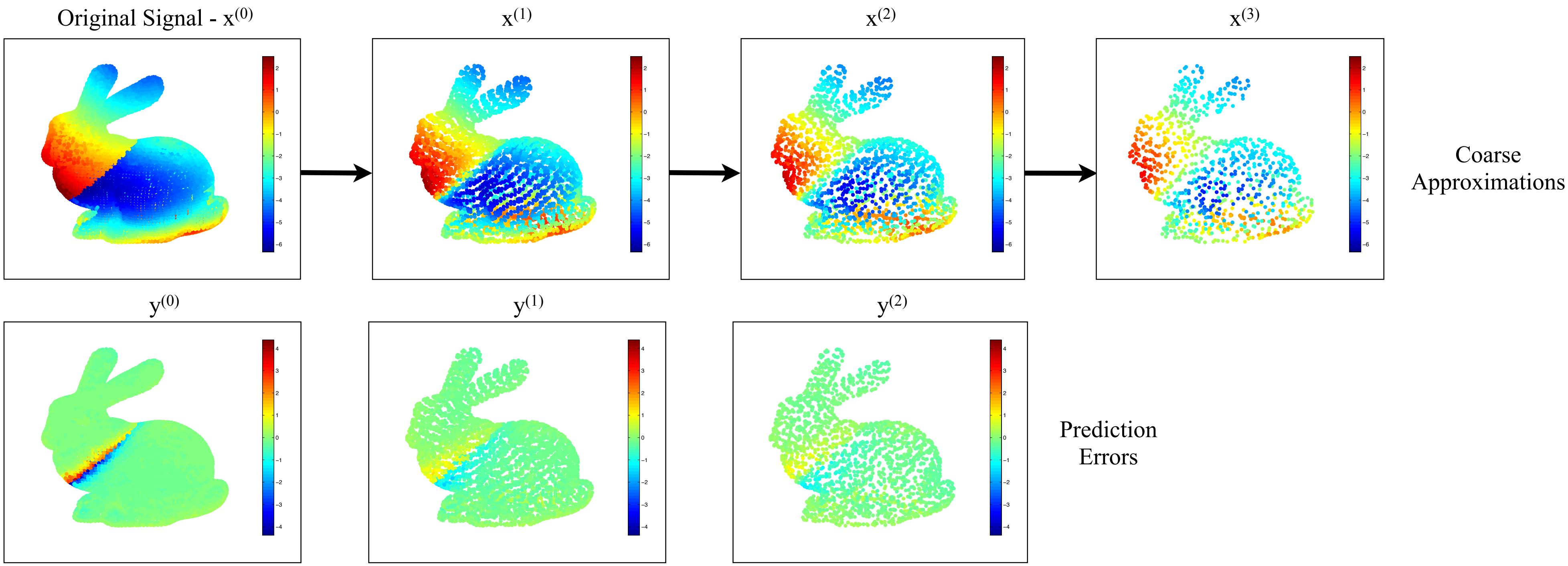


- Downsampling + graph reduction = a multiresolution of graphs
- Methods used here:
  - Graph downsampling by polarity of Laplacian eigenvector associated with largest eigenvalue
  - Kron reduction with spectral sparsification
- Alternative: coarse graining

Graph Reduction

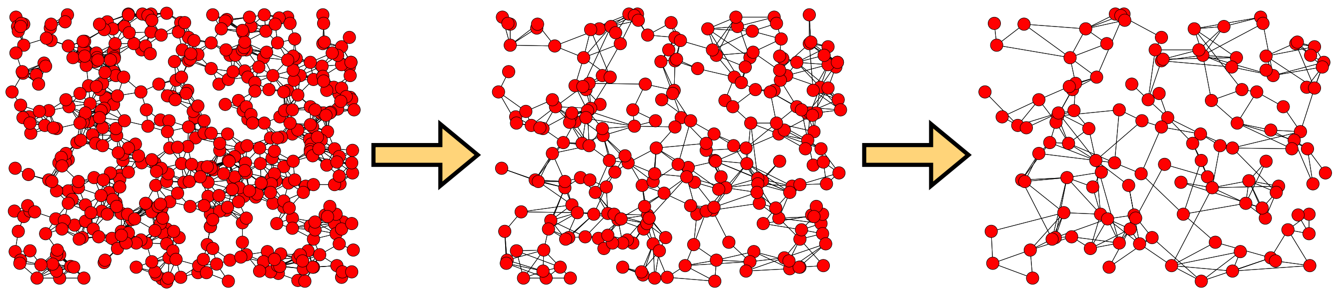
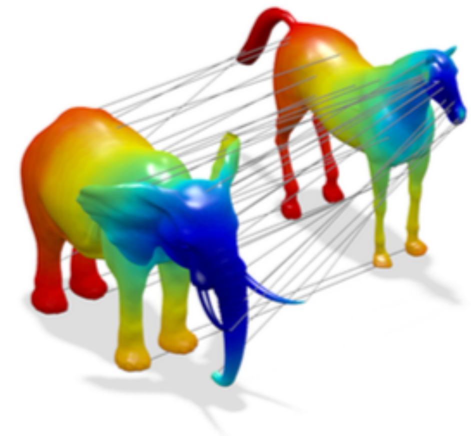


A Multiscale Pyramid Transform for Graph Signals



Shuman et al., "A multiscale pyramid transform for graph signals," TSP, 2016

Problem 3: Graph Reduction to Preserve Eigenstructure

- How do we perform the downsampling and rewiring in such a way that the Laplacian eigenspaces lifted from the coarsened graphs are closely aligned with the Laplacian eigenspaces associated with the smallest eigenvalues of the finer graphs?

- Closely related to multigrid techniques, functional correspondence in computer graphics

- Applications in spectral clustering, visualization, dictionary design, machine learning, . . .

Kovnatsky et al., “Coupled quasi-harmonic bases,” EUROGRAPHICS, 201

Problem 4: Approximation of Dense Matrices by Functions of Sparse Matrices

Given dense $A_1, A_2 \in \mathbb{R}^{N \times N}$,

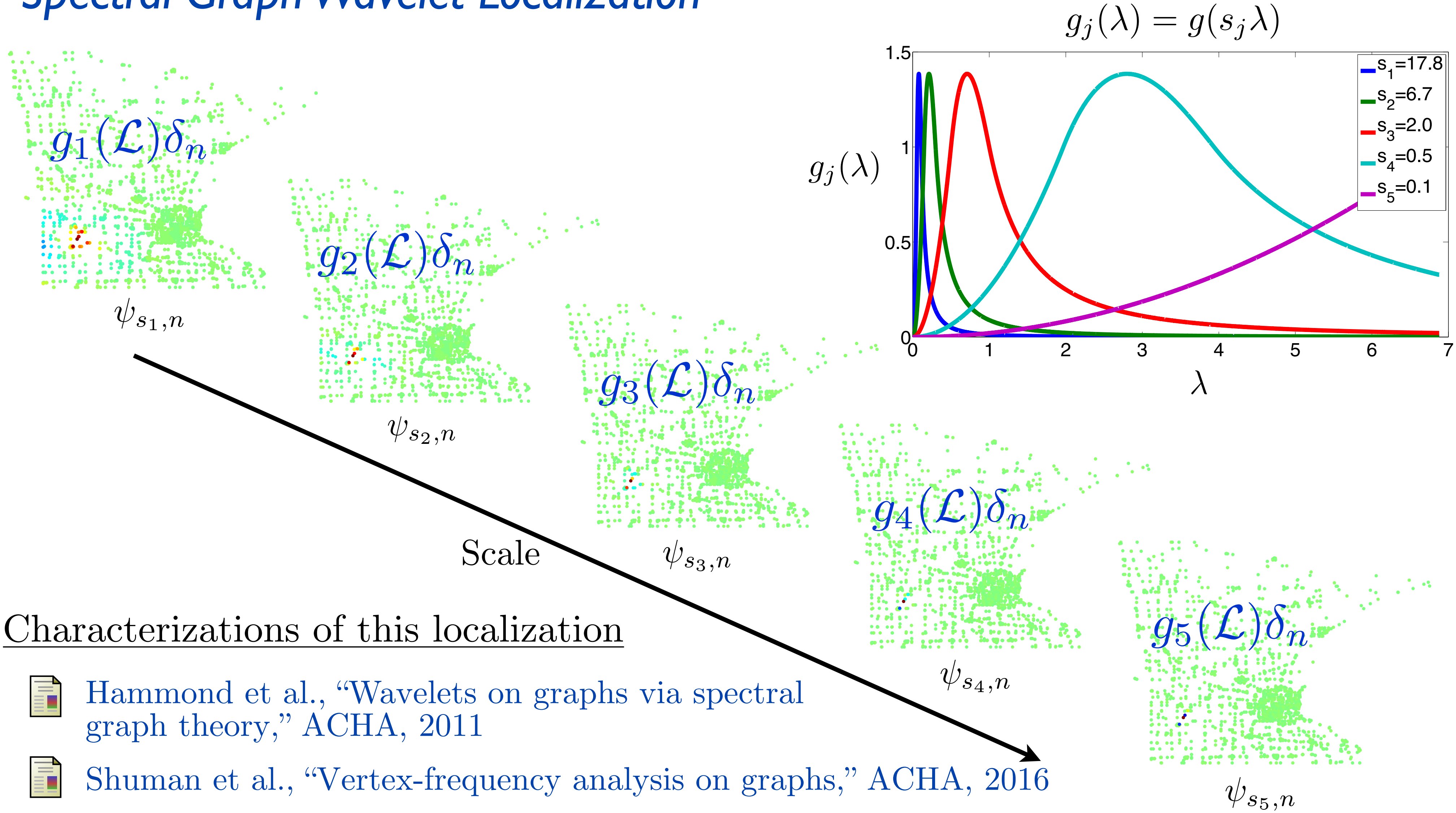
find $g_1(\cdot), g_2(\cdot)$, and a sparse $A_0 \in \mathbb{R}^{N \times N}$

such that

$g_1(A_0) \approx A_1$ and $g_2(A_0) \approx A_2$

Problem 5: "Off-Diagonal" Decay of Matrix Functions

Spectral Graph Wavelet Localization

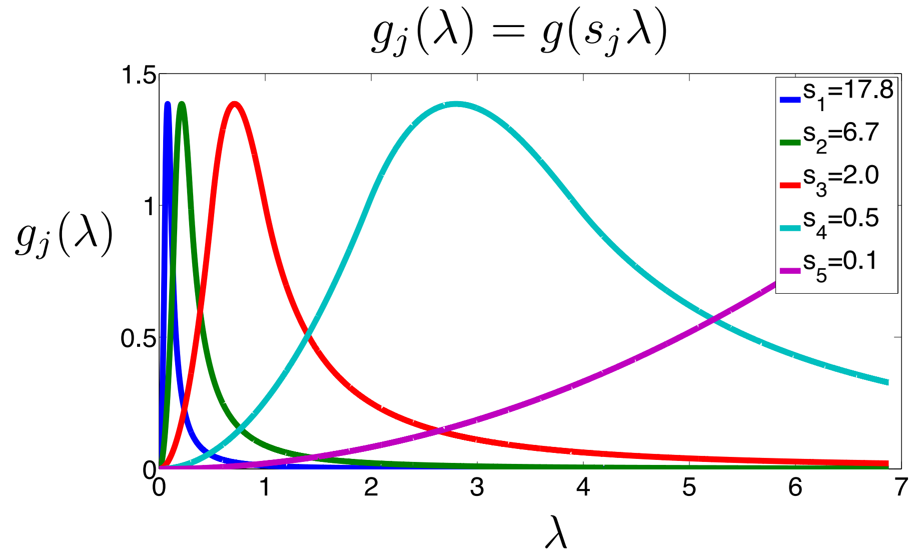


<u>Characterizations of this localization</u>

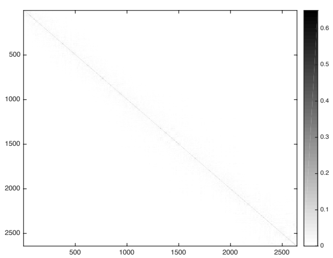
- Hammond et al., "Wavelets on graphs via spectral graph theory," ACHA, 2011
- Shuman et al., "Vertex-frequency analysis on graphs," ACHA, 2016
- Benzi and Razouk, "Decay bounds and O(n) algorithms for approximating functions of sparse matrices", ETNA, 2007

Problem 5: "Off-Diagonal" Decay of Matrix Functions

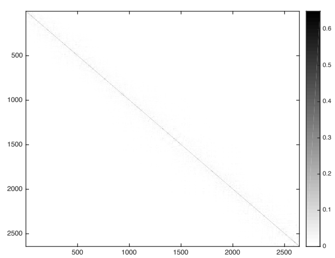
Spectral Graph Wavelet Localization



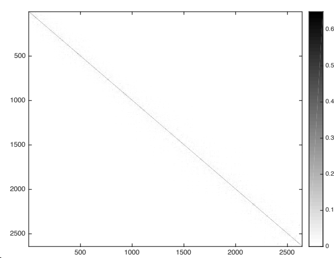
$g_1(\mathcal{L})$



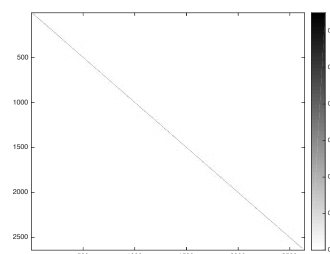
$g_2(\mathcal{L})$



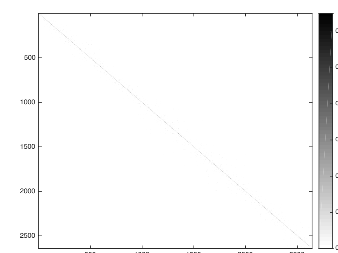
$g_3(\mathcal{L})$



$g_4(\mathcal{L})$



$g_5(\mathcal{L})$

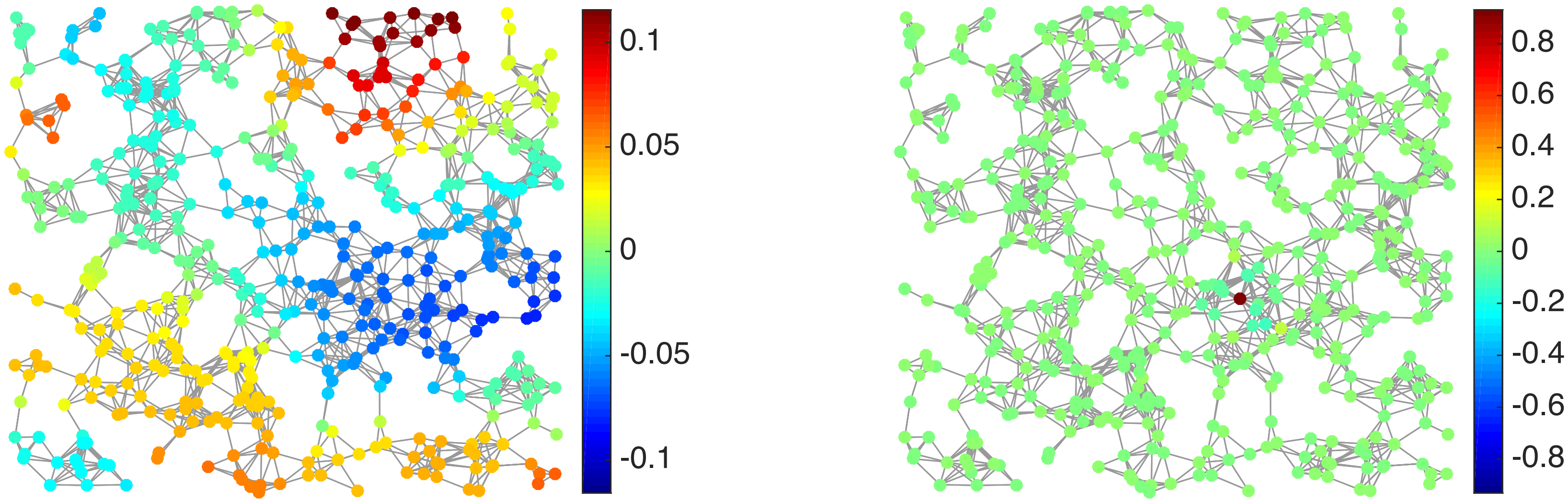


Scale

Characterizations of this localization

- Hammond et al., "Wavelets on graphs via spectral graph theory," ACHA, 2011
- Shuman et al., "Vertex-frequency analysis on graphs," ACHA, 2016
- Benzi and Razouk, "Decay bounds and O(n) algorithms for approximating functions of sparse matrices", ETNA, 2007

Problem 6: Graph Laplacian Eigenvector Localization



- How are structural properties of weighted graphs such as the regularity, clustering, modularity, and other spectral properties theoretically related to the (non-)localization of the graph Laplacian eigenvectors?