Dictionary Design for Graph Signal Processing

David Shuman

October 24, 2017

Inverse Problems and Analysis Seminar

Department of Mathematical Sciences

University of Delaware

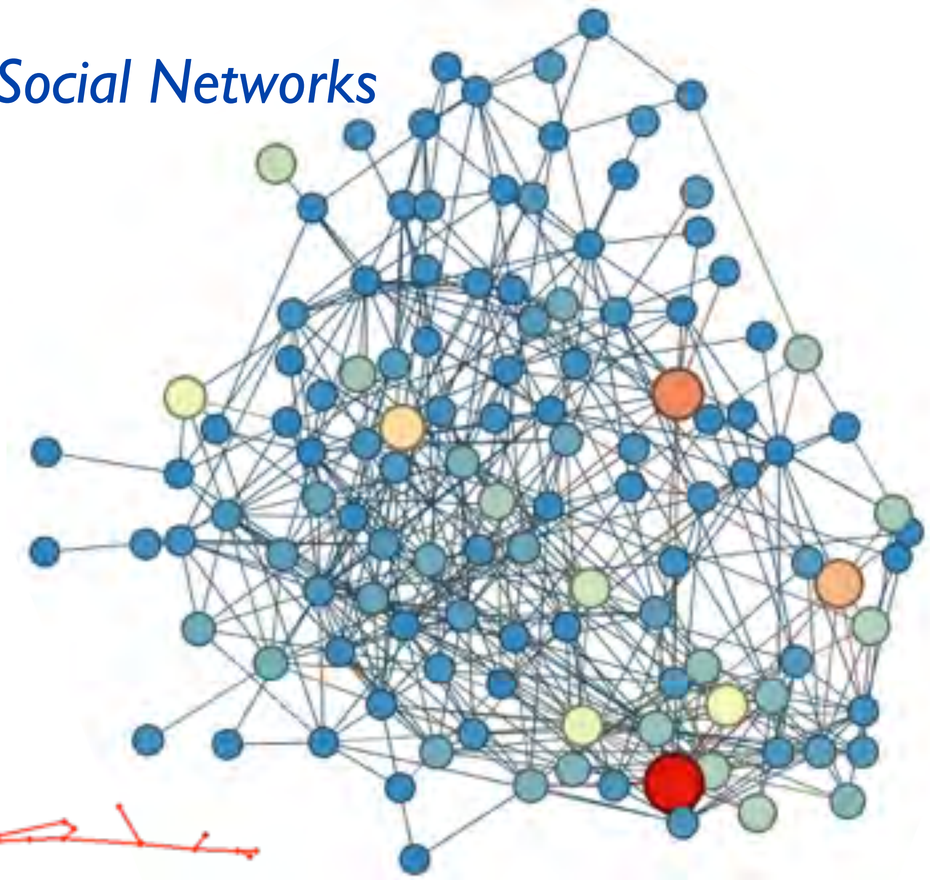
Special thanks and acknowledgement to my collaborators:
Andre Archer, Sam Armon, Andrew Beveridge, Xiaowen Dong, Mohammad Javad Faraji, Stefan Faridani, Pascal Frossard, Nicki Holighaus, Jason McEwen, Daniel Kressner, Yan Jin, Shuni Li, Sunil Narang, Antonio Ortega, Javier Pérez-Trufero, Nathanaël Perraudin,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

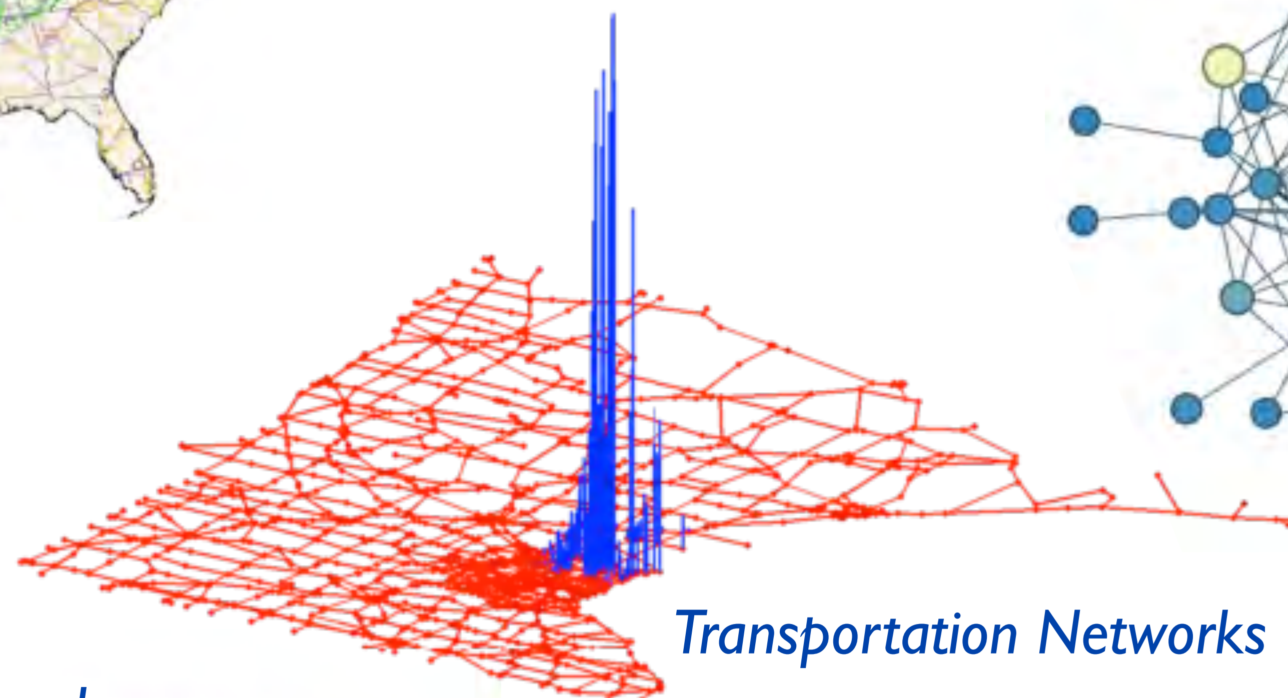
MACALESTER

Signal Processing on Graphs

Energy Networks

Social Networks





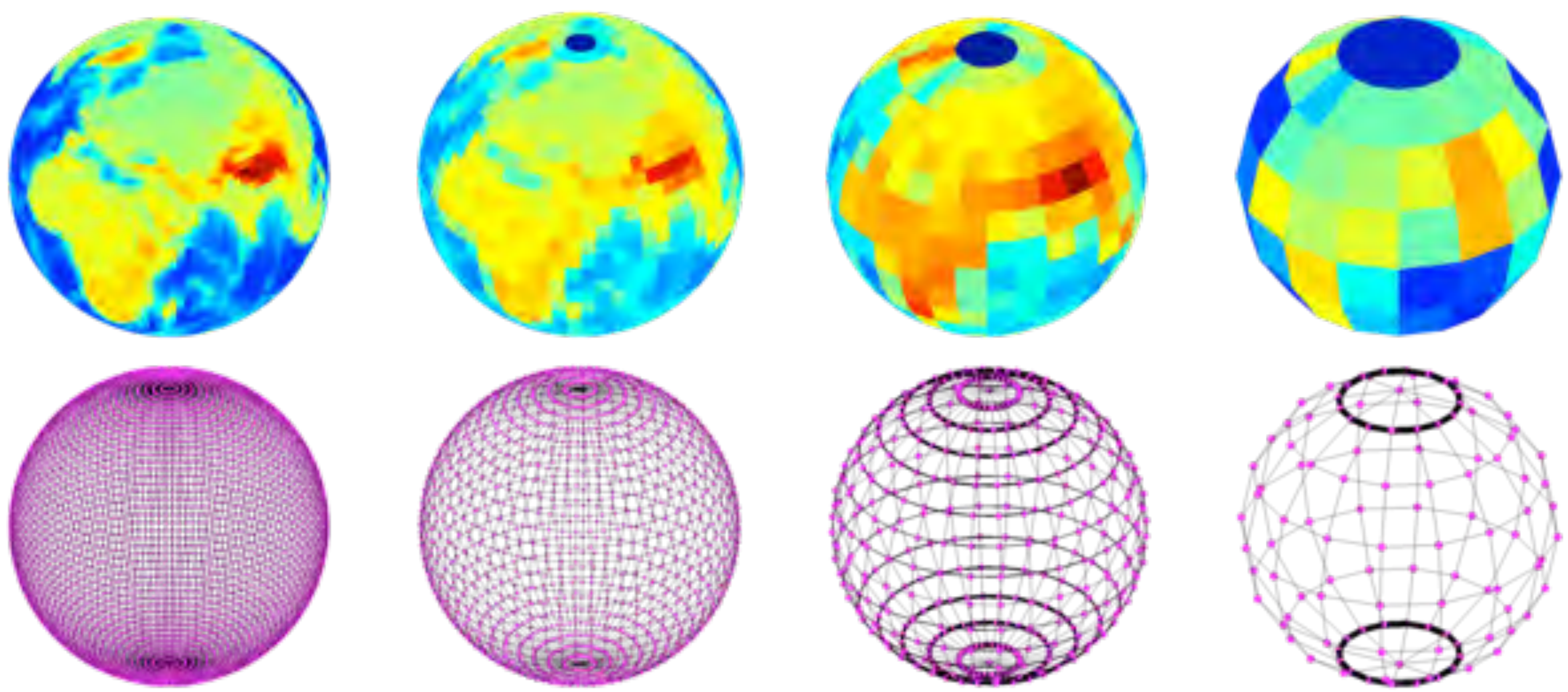
Transportation Networks

Biological Networks

Irregular Data Domains

Some Typical Graph Signal Processing Problems

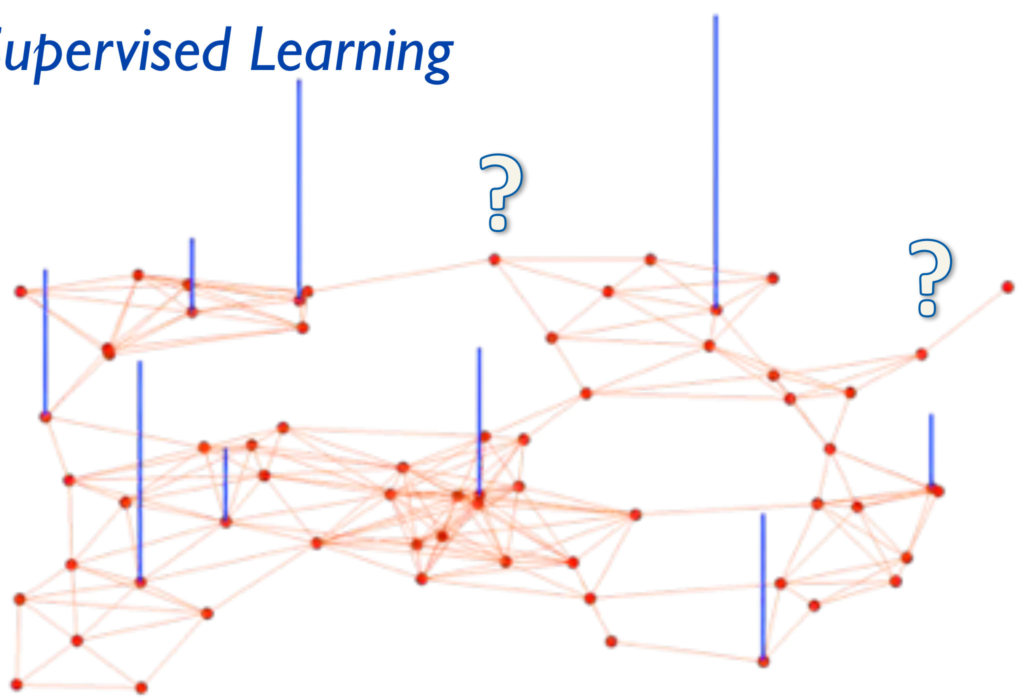
Compression / Visualization

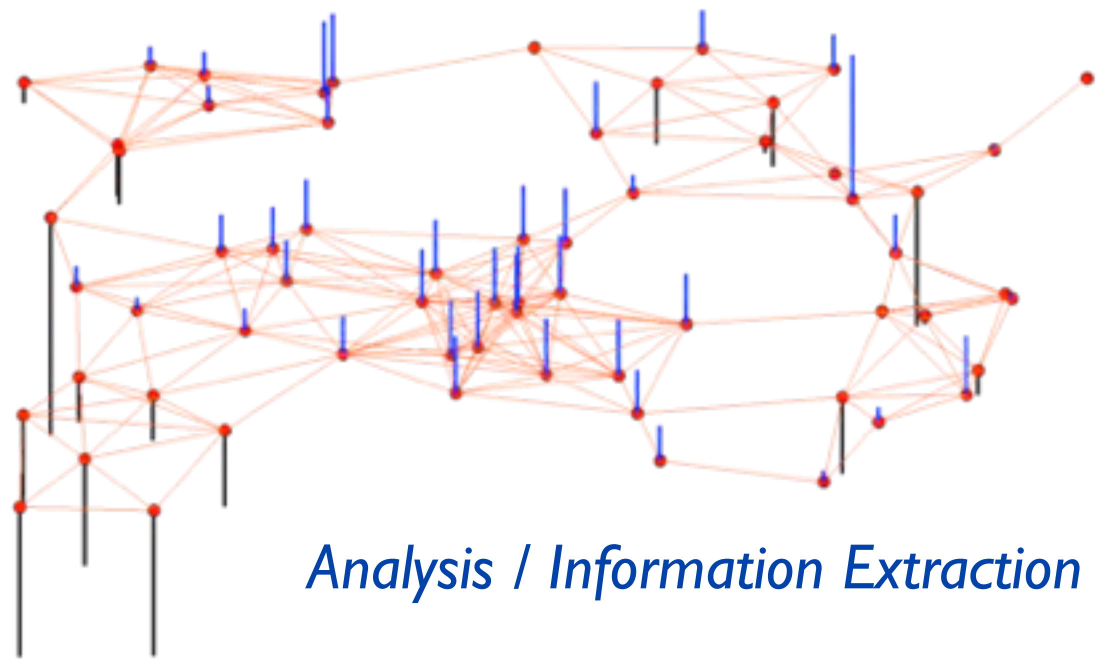


Earth data source: Frederik Simons

Denoising

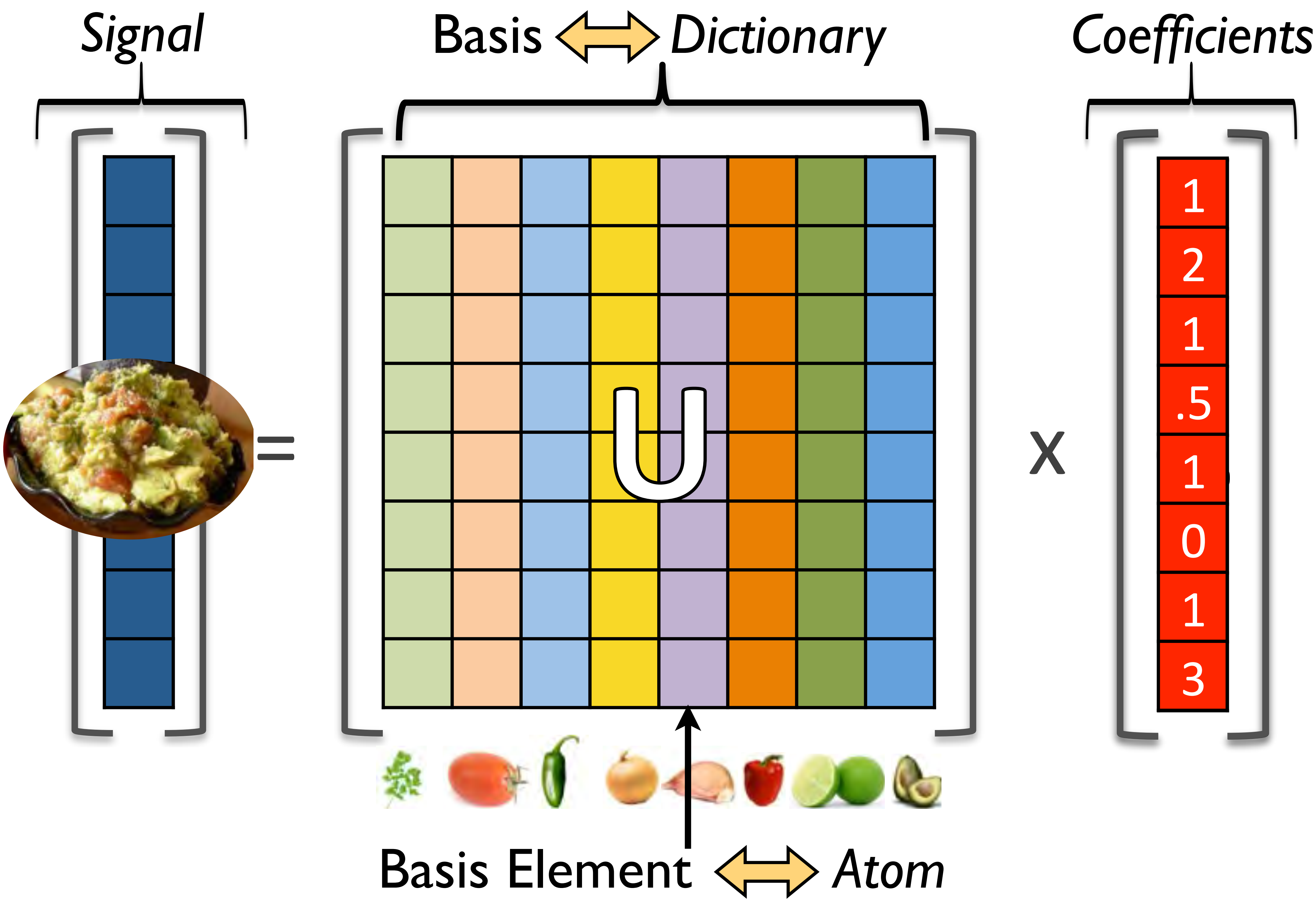
Semi-Supervised Learning



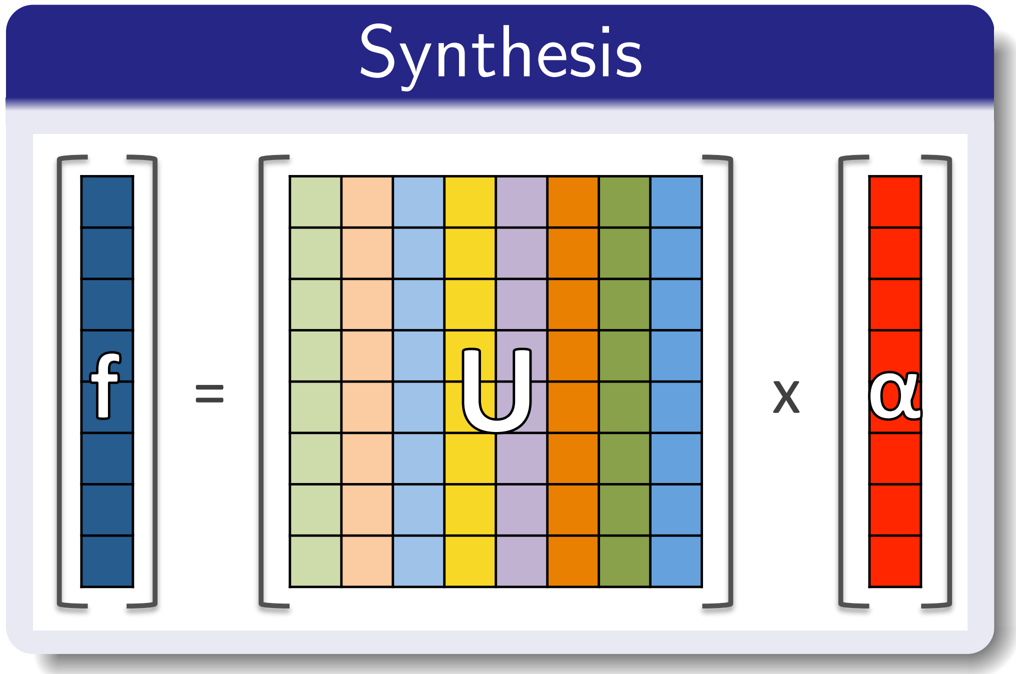


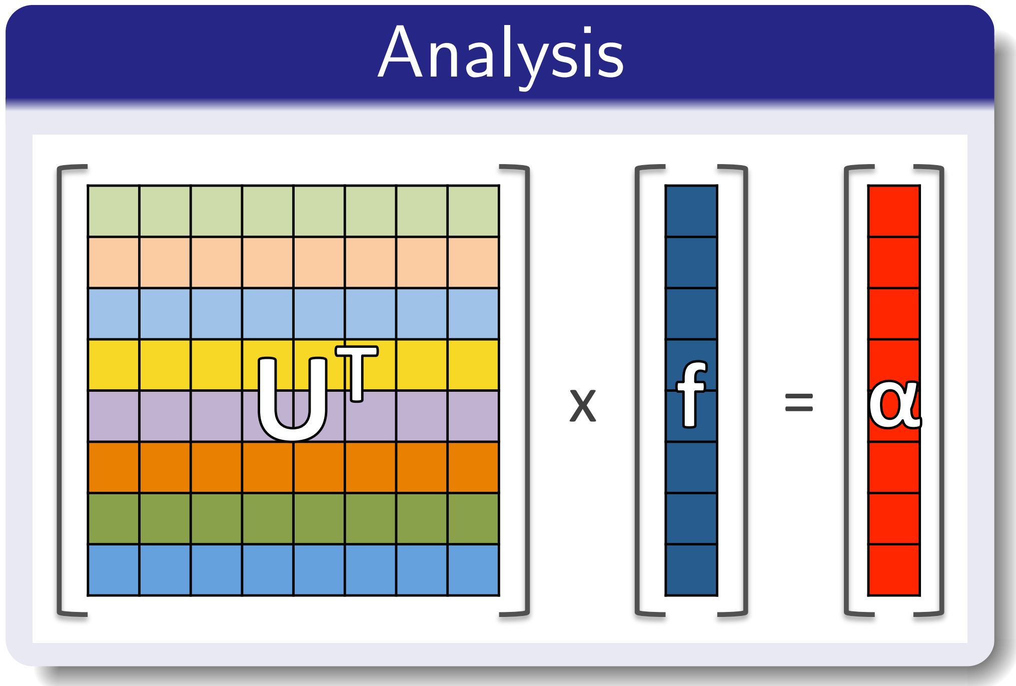
Analysis / Information Extraction

Orthonormal Dictionaries



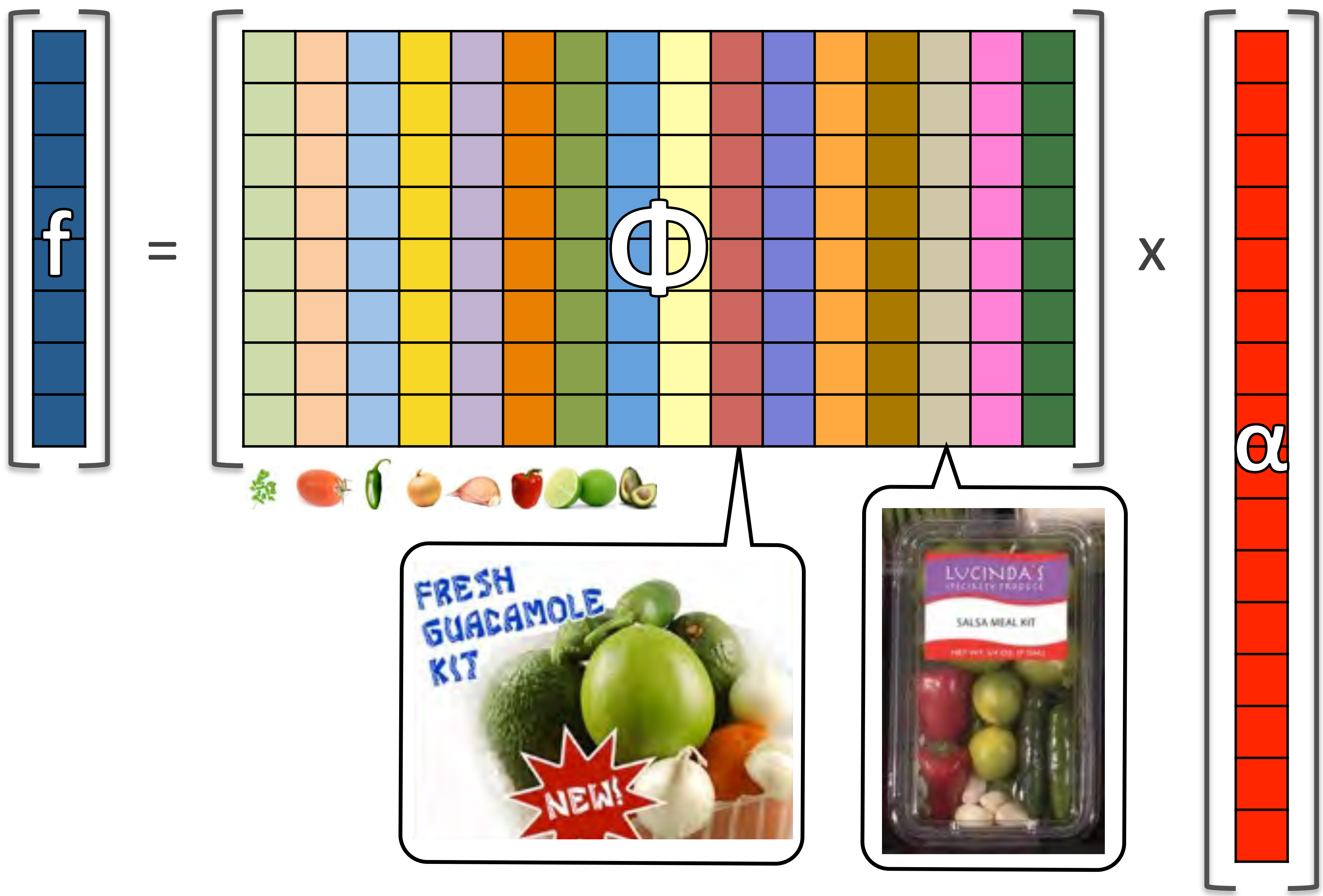
Orthonormal Dictionaries (cont.)





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

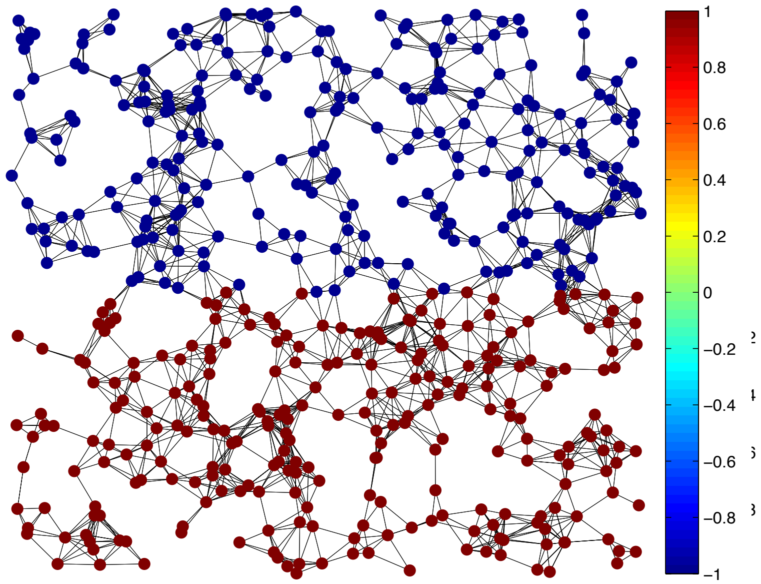
Overcomplete Dictionaries and Sparsity



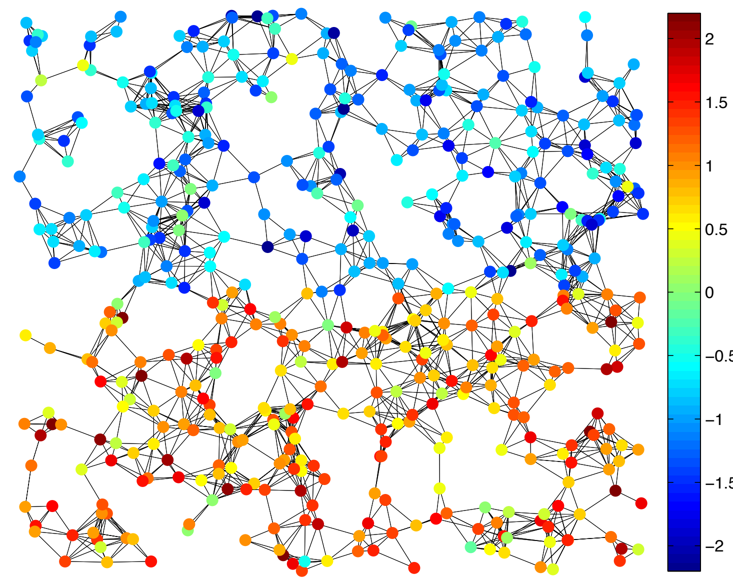
- Given an overcomplete Φ, there are infinitely many choices of α that lead to the same signal f
- Useful to *sparsely* represent signals —> few non-zero coefficients in α

Motivating Example: Denosing

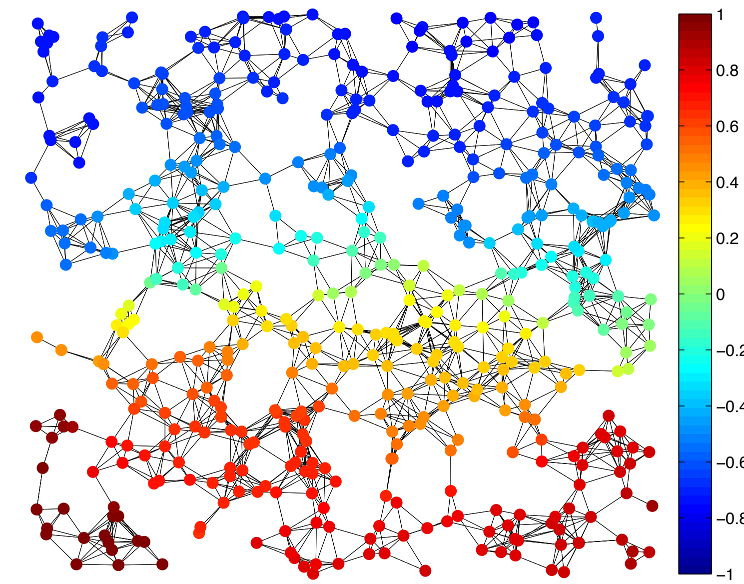
- Tikhonov regularization for denoising: $\text{argmin}_f \left\{ ||f - y||_2^2 + \gamma f^T \mathcal{L} f \right\}$



Original

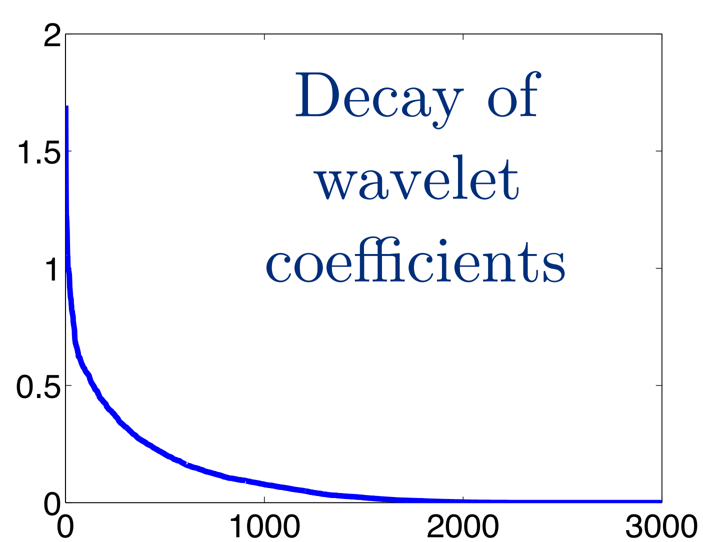


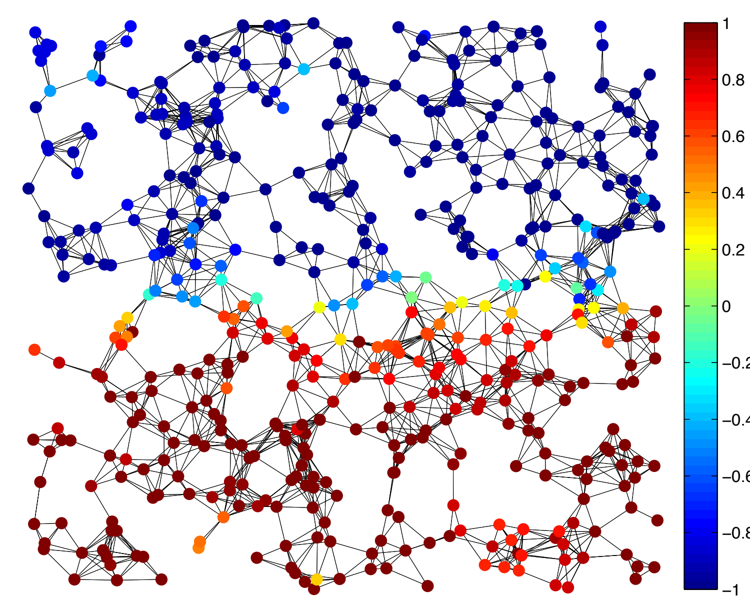
Noisy



Denoised

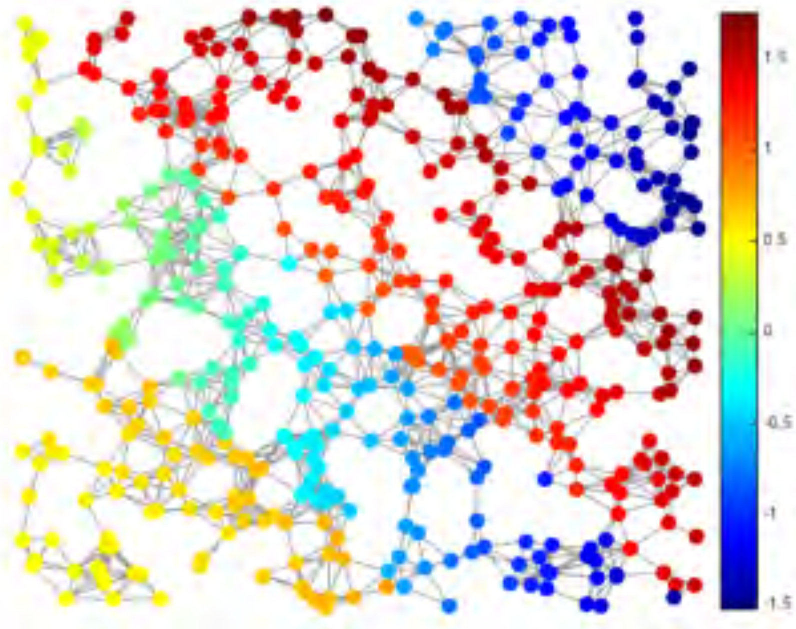
- Wavelet denoising: $\text{argmin}_a \left\{ ||f - W^* a||_2^2 + \gamma ||a||_{1,\mu} \right\}$



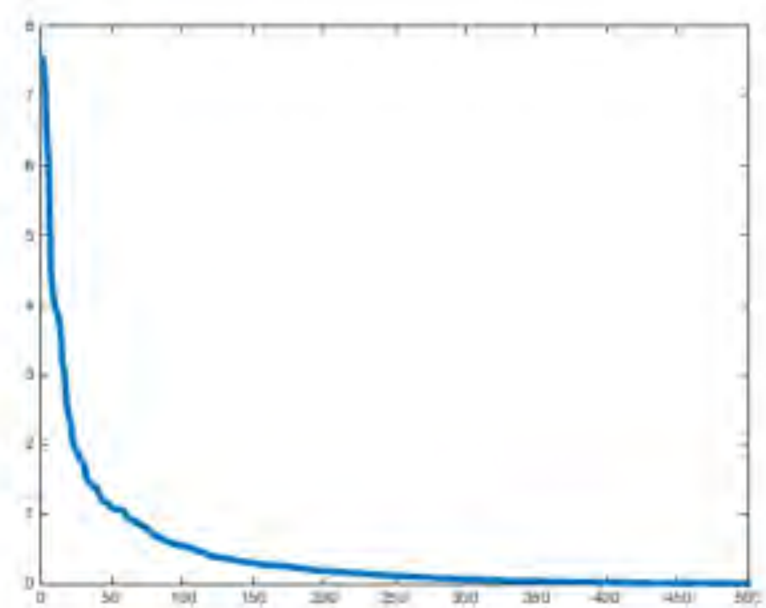


Motiving Example: Compression

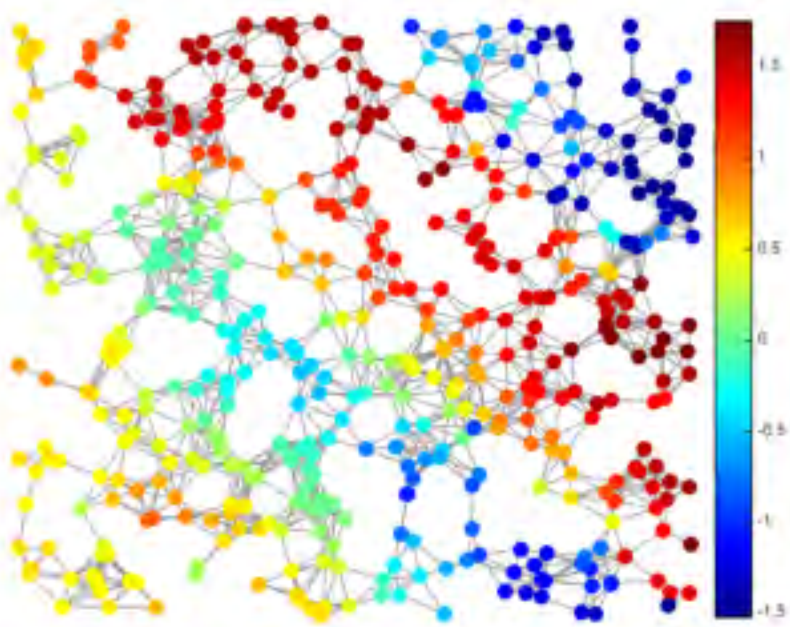
Piecewise-Smooth Signal with
Discontinuities



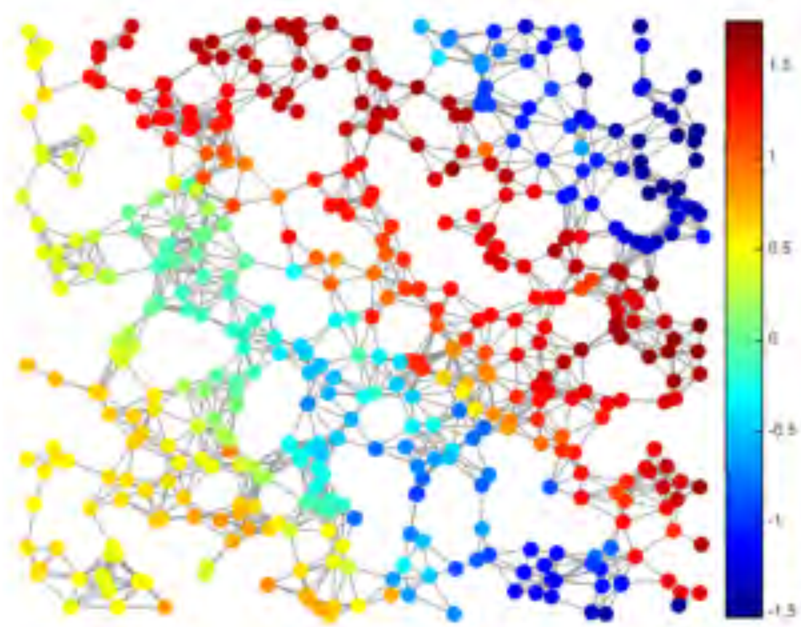
Diffusion Wavelet Coefficients,
Sorted by Magnitude



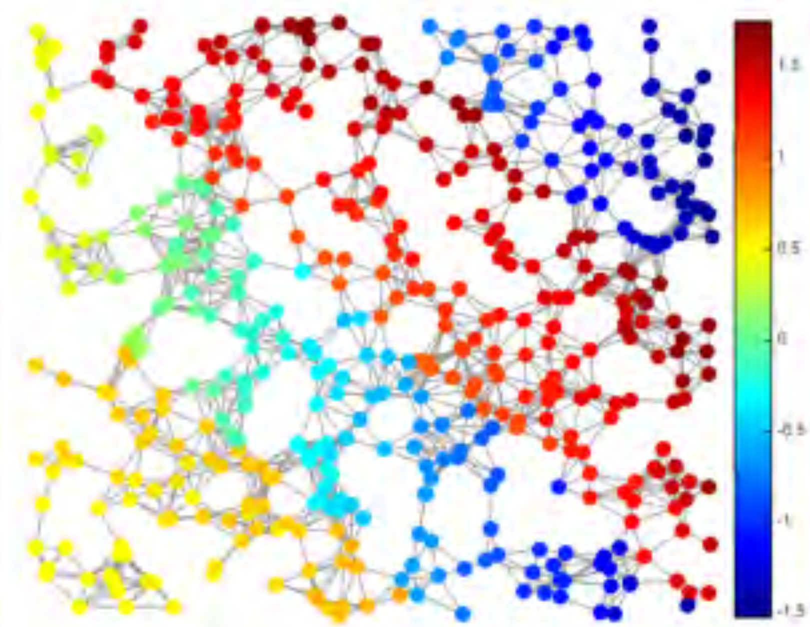
Reconstruction from 10%
of Coefficients



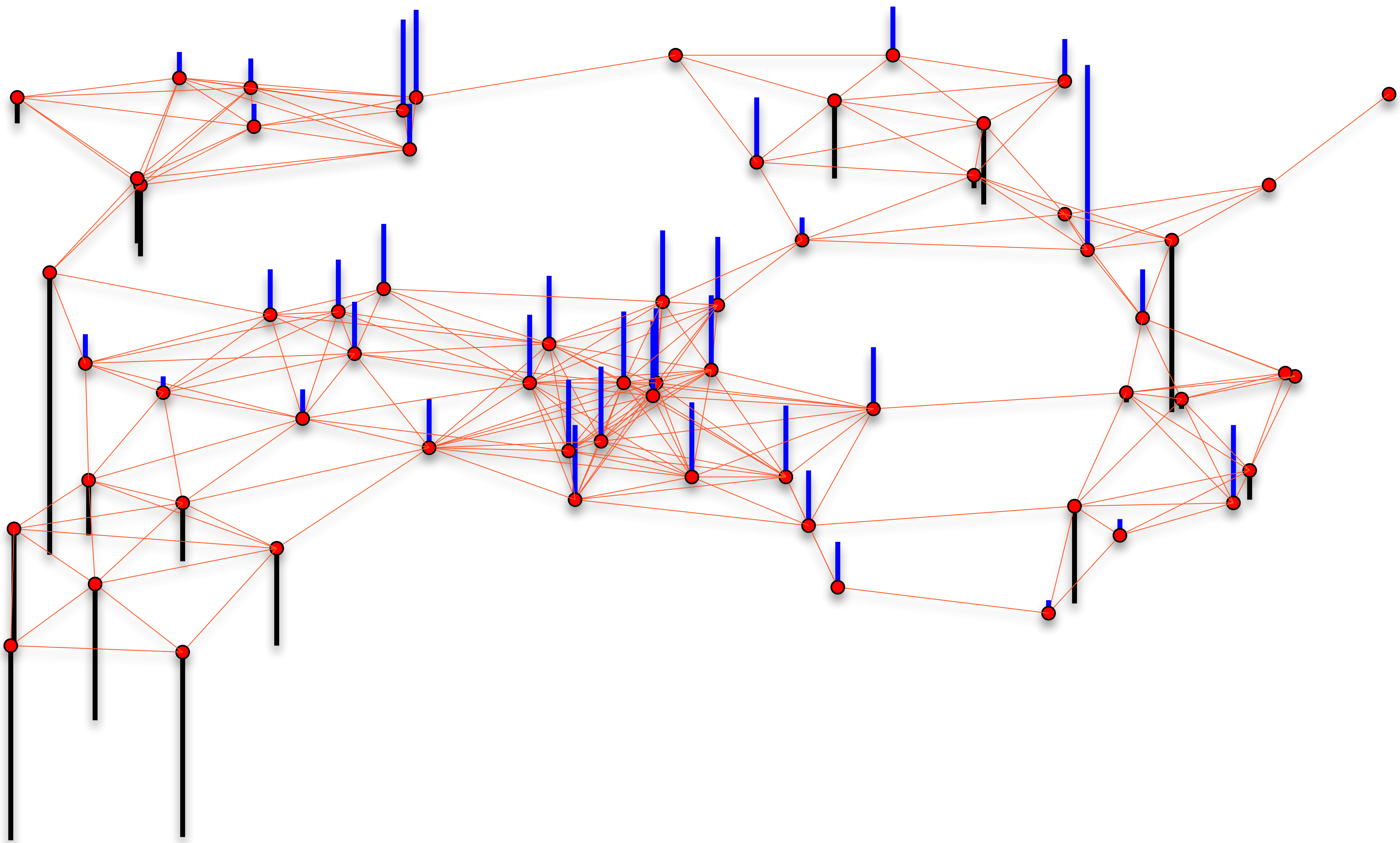
Reconstruction from 20%
of Coefficients



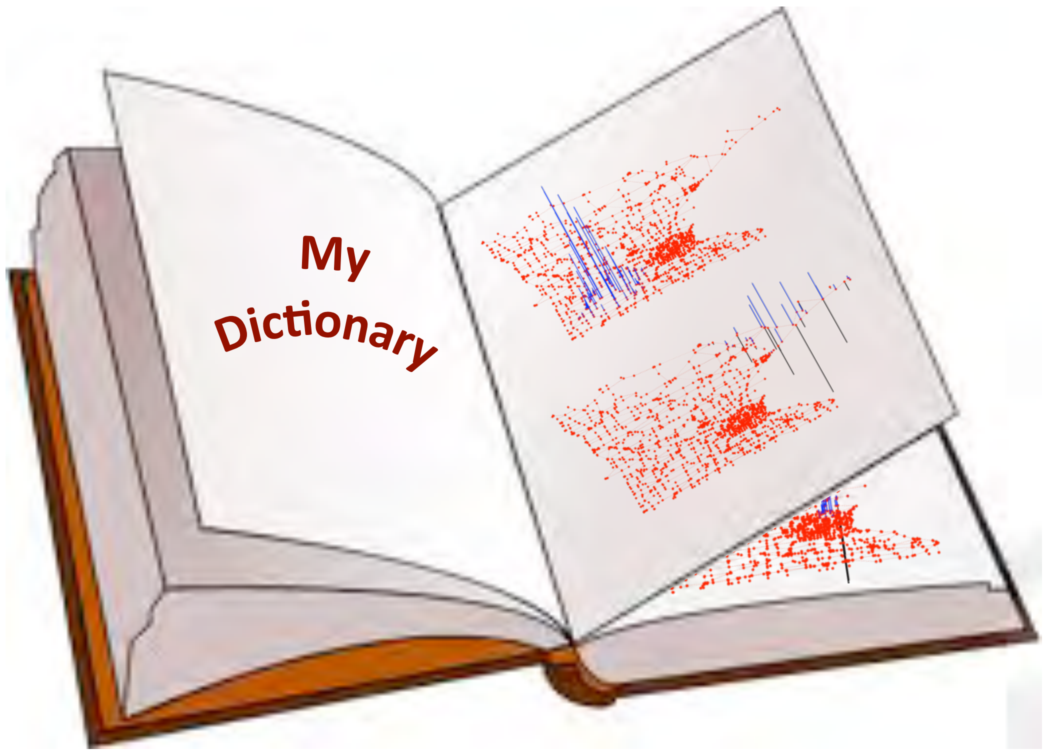
Reconstruction from 50%
of Coefficients



Motivating Example: Any Structure?



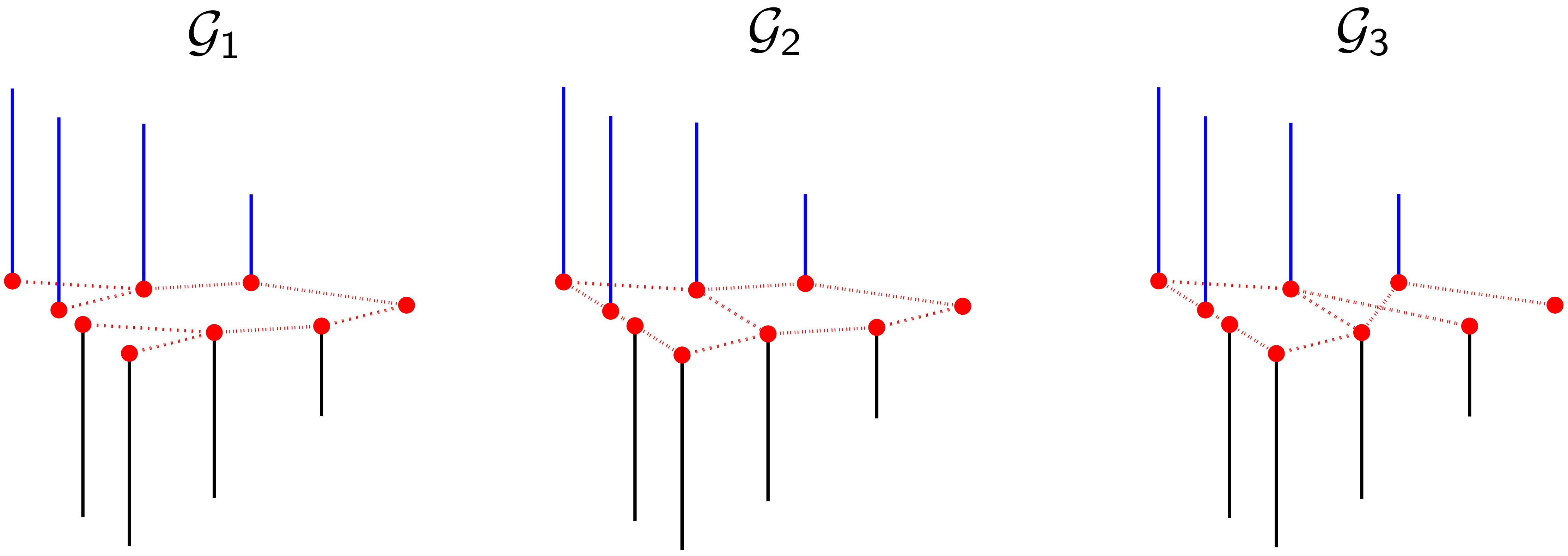
Dictionary Design for Signals on Graphs



Desirable Characteristics

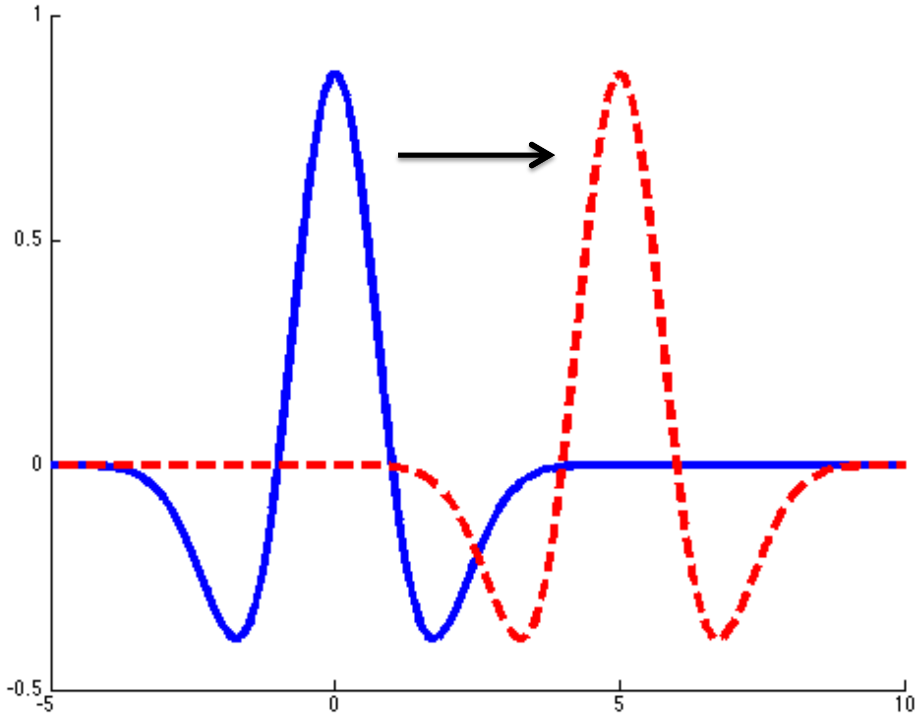
- Ability to *sparsely* represent signals — few non-zero coefficients in α
- Ability to capture the relevant characteristics of signals to extract information
- Computationally efficient to apply Φ and Φ^T
- Tight frames

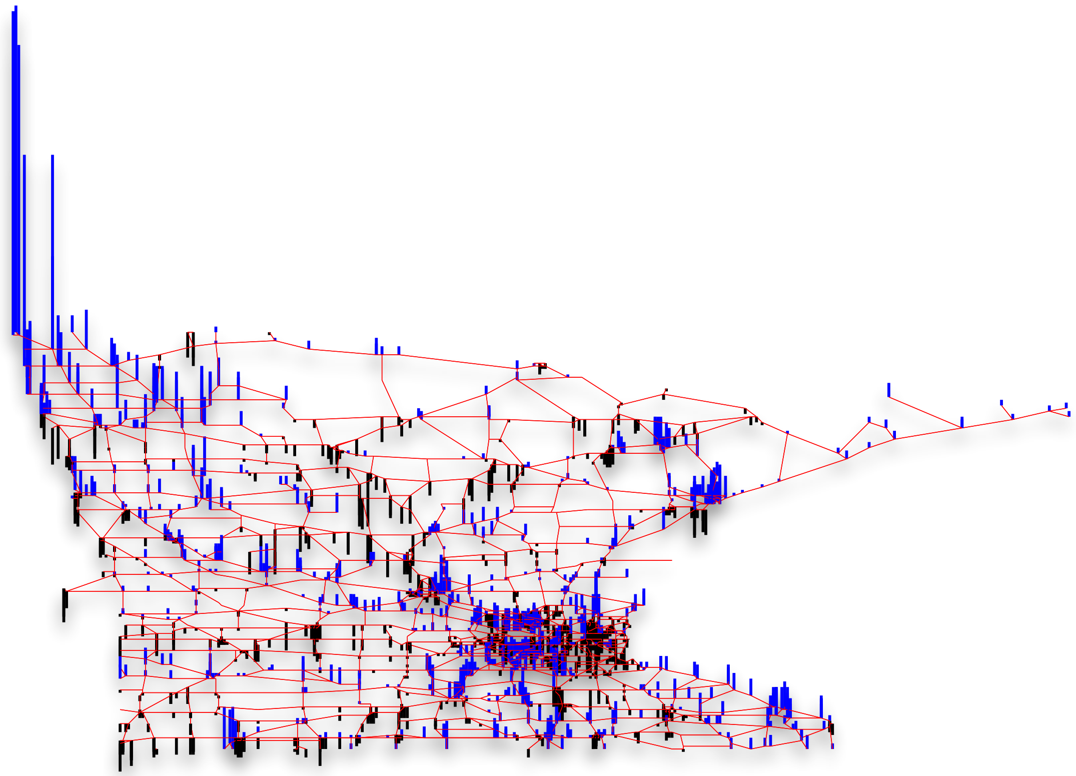
Why Do We Need New Dictionaries?



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

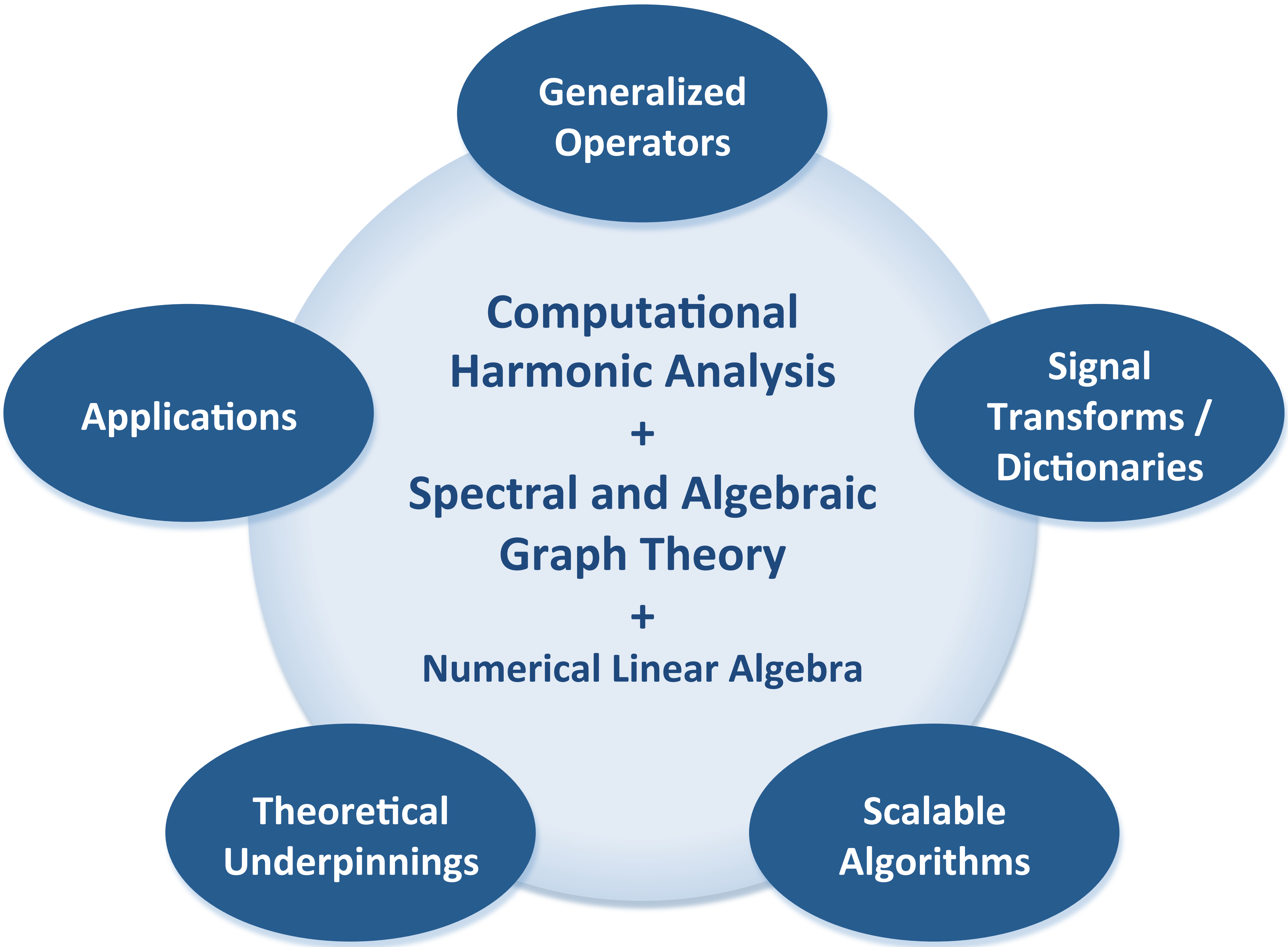
The Essence of the Problem





- 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

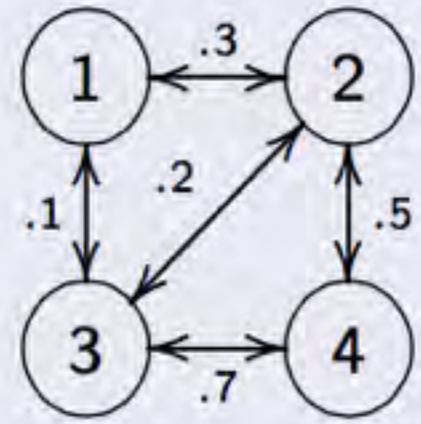
Approach: Leverage Intuition from Euclidean Settings to Develop New Mathematical Tools for the Graph Setting



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

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

- 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:

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

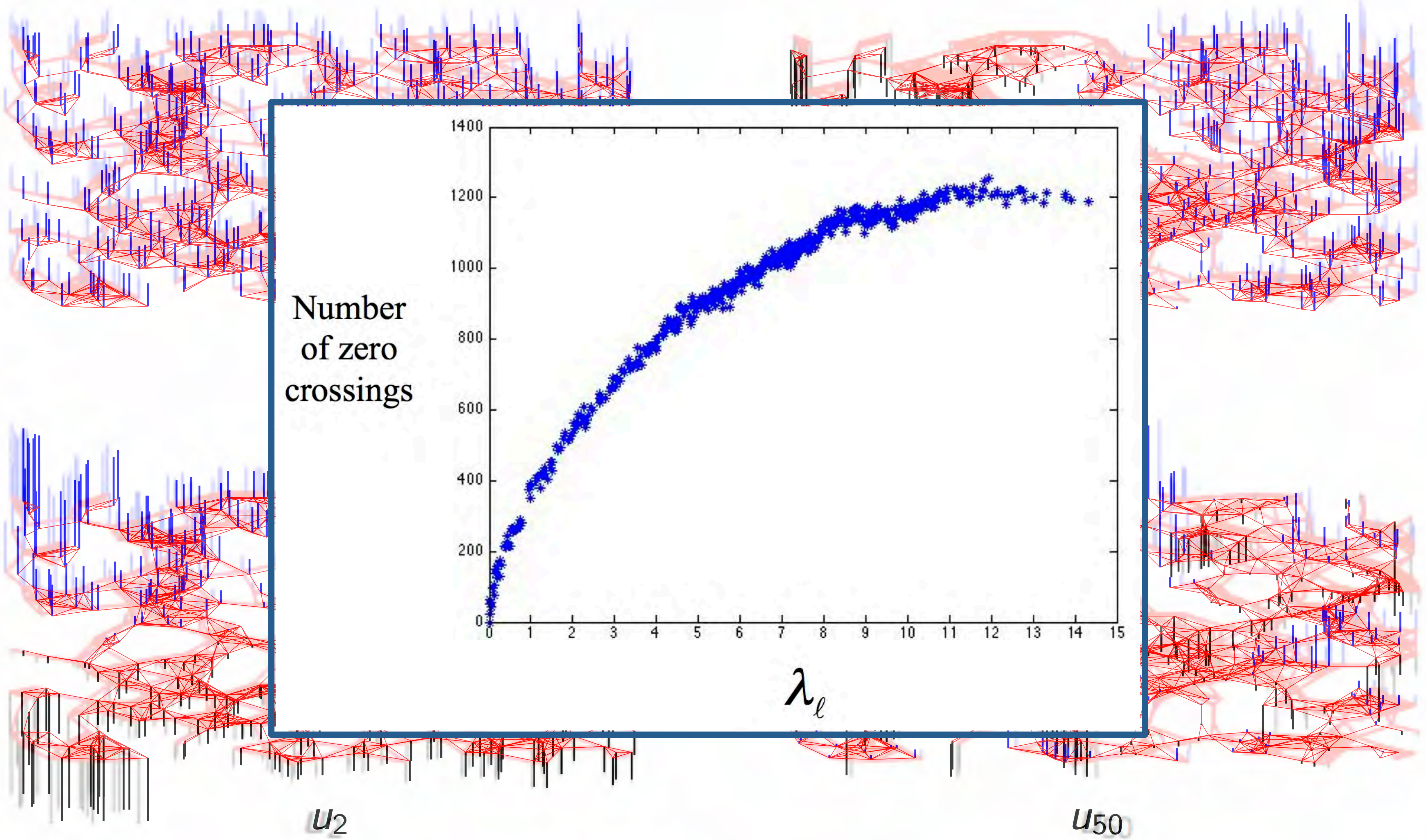
  ordered w.l.o.g. s.t.

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

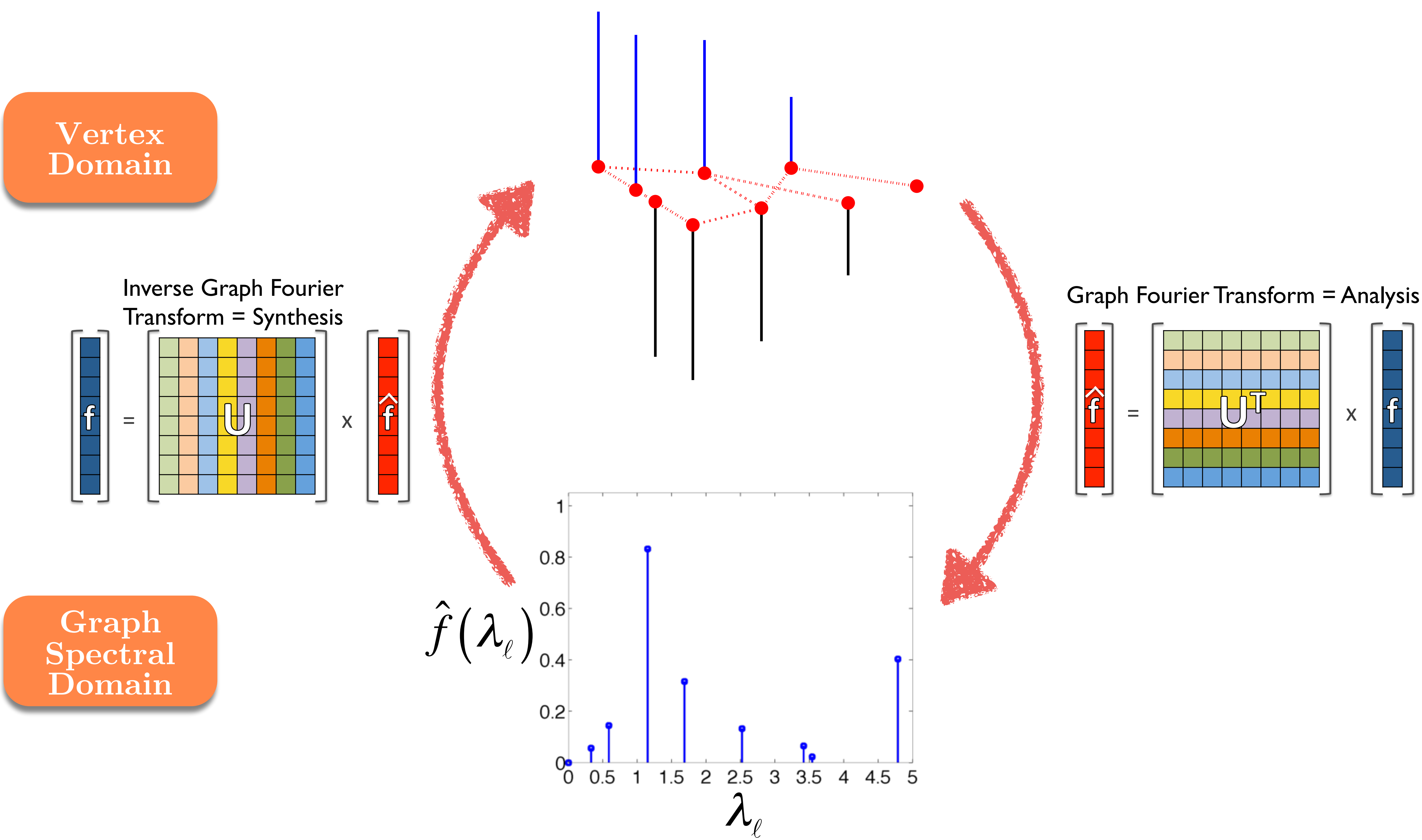
- Discrete difference operator: $(\mathcal{L}f)(i) = \sum_{j \in \mathcal{N}_i} W_{i,j}[f(i) - f(j)]$

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

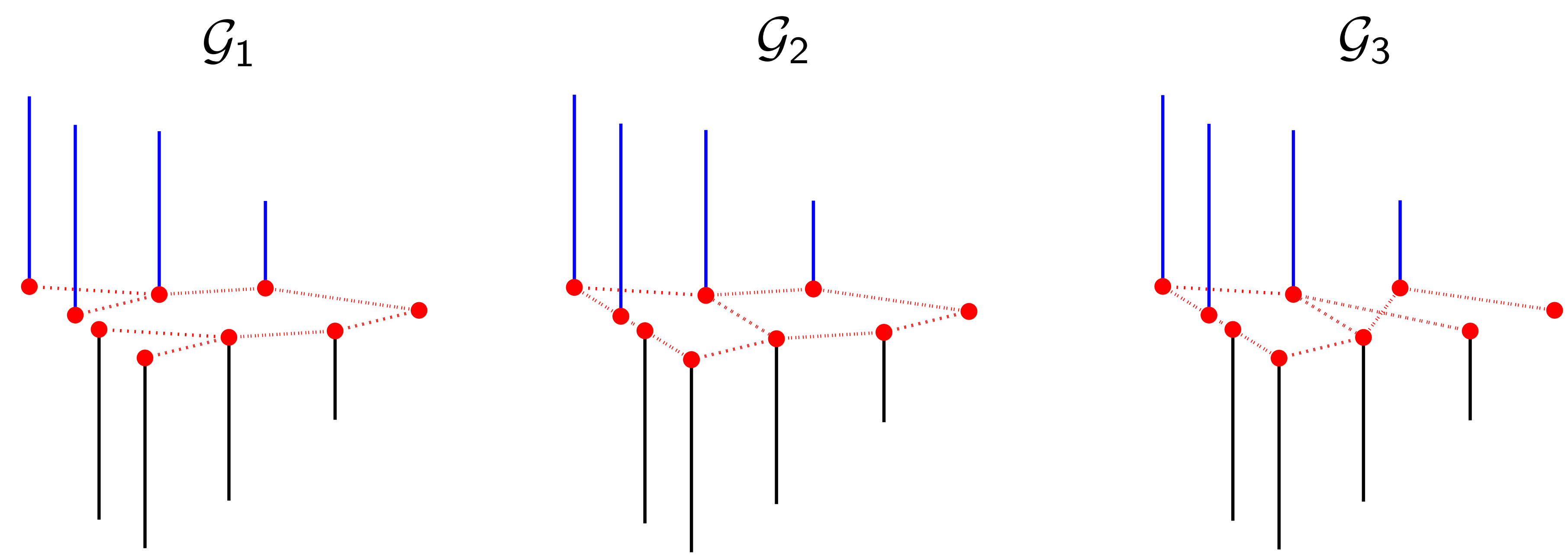


The GFT Incorporates the Graph Structure

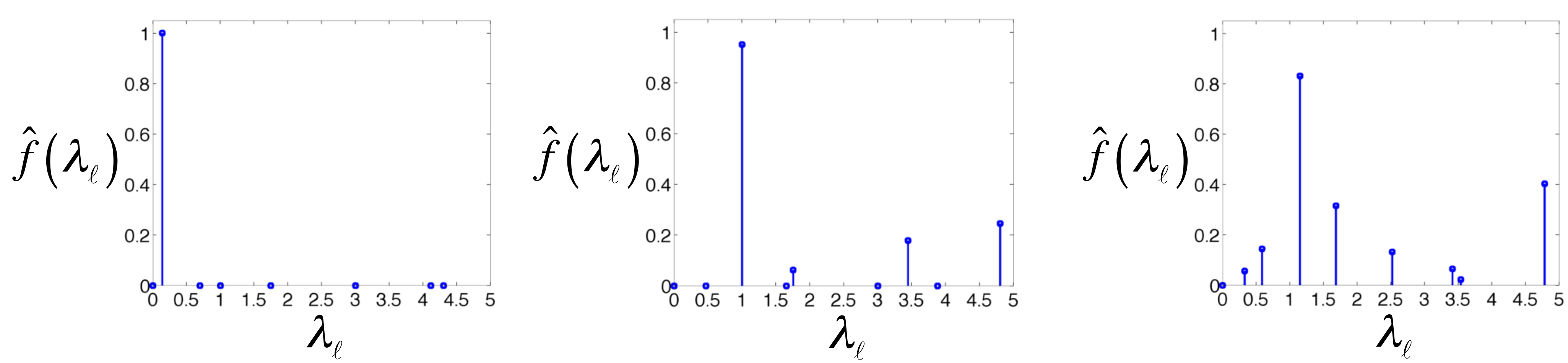


The GFT Incorporates the Graph Structure

Vertex Domain

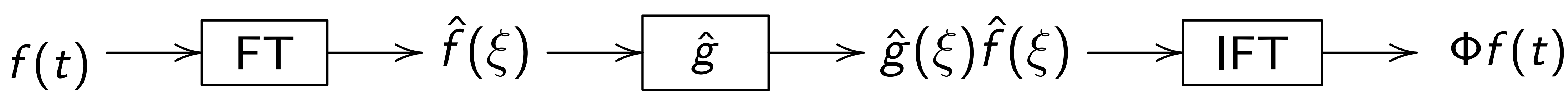


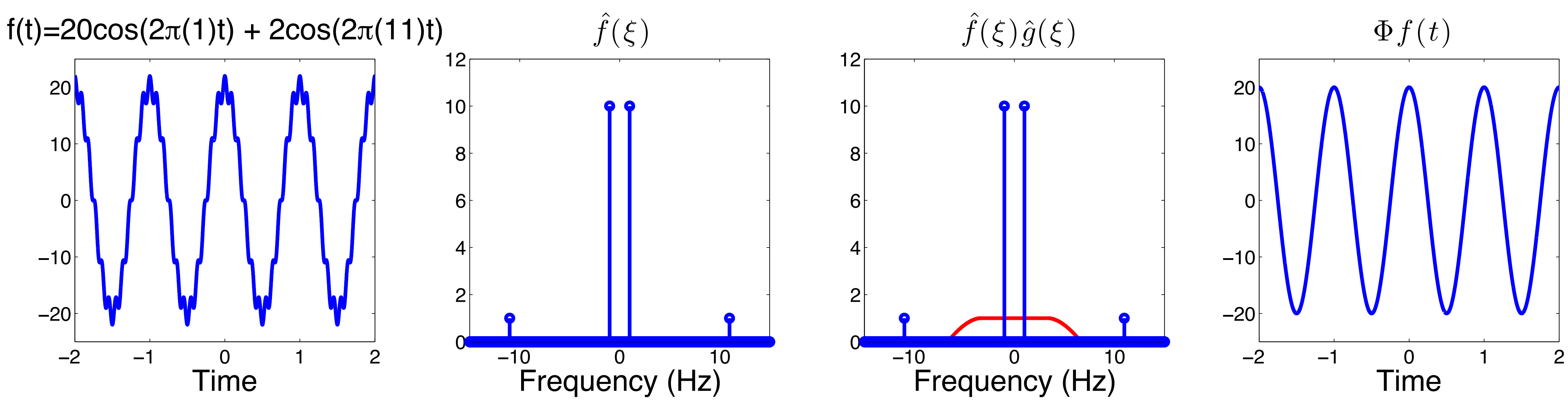
Graph Spectral Domain



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

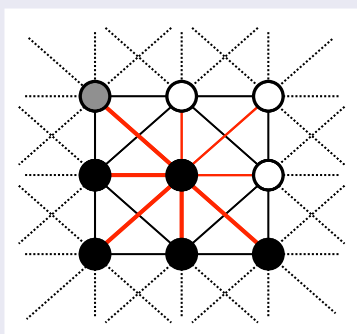




Example: Image Denoising by Low-Pass Graph Filtering

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

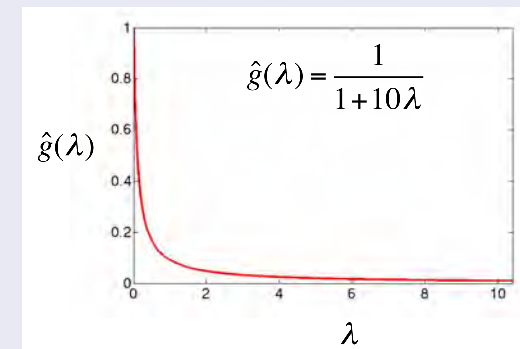
Semi-Local Graph

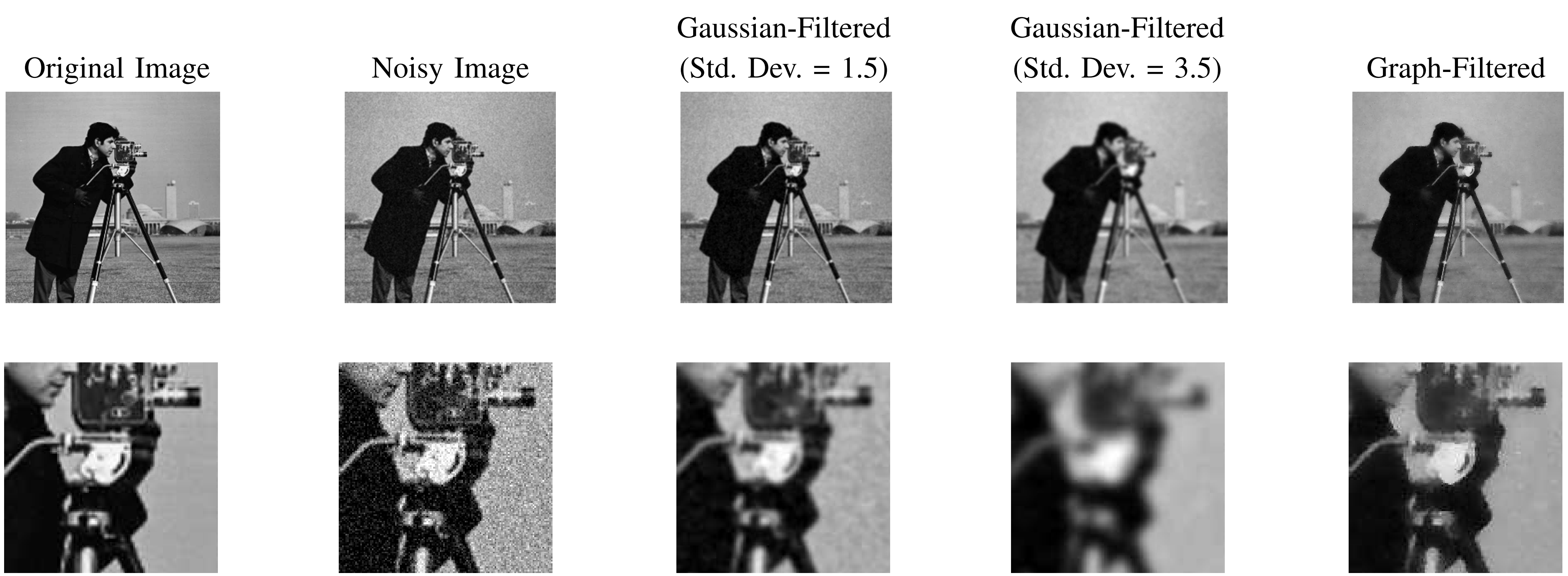


Tikhonov Regularization

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

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



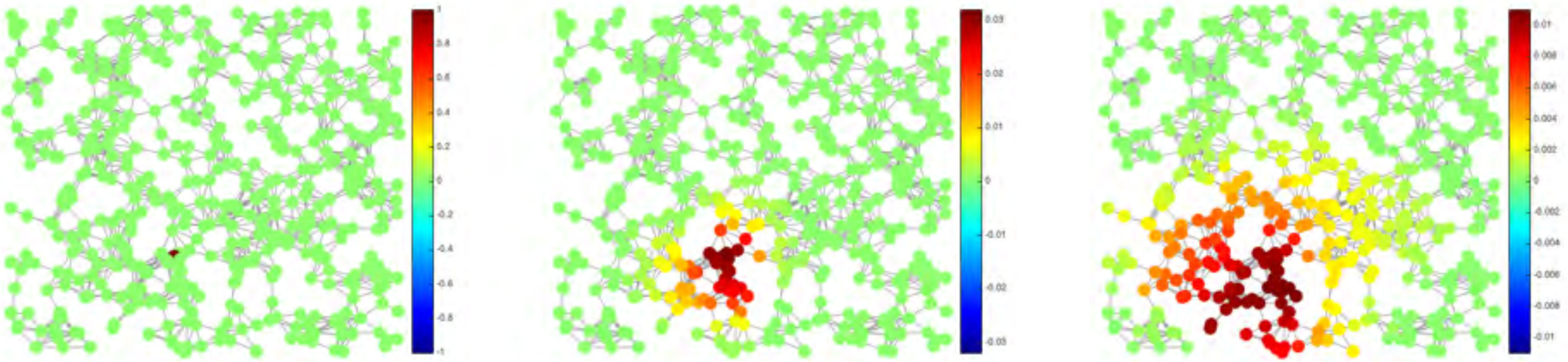


Shuman et al., The emerging field of signal processing on graphs, SPM, 2013

Diffusion: $e^{-\tau \mathcal{L}} \delta_i$

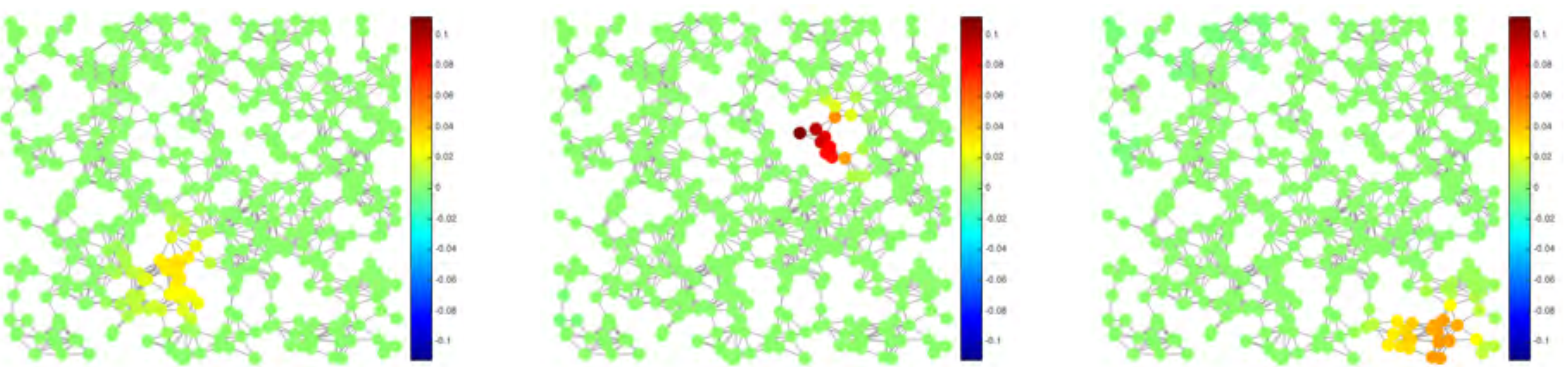
- Start with a unit of energy at a single vertex and let it diffuse:

τ=0 τ=1 τ=5



- How much it diffuses over a fixed time depends on the graph structure:

τ=1 τ=1 τ=1



Coifman and Lafon, Diffusion maps, ACHA, 2006

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

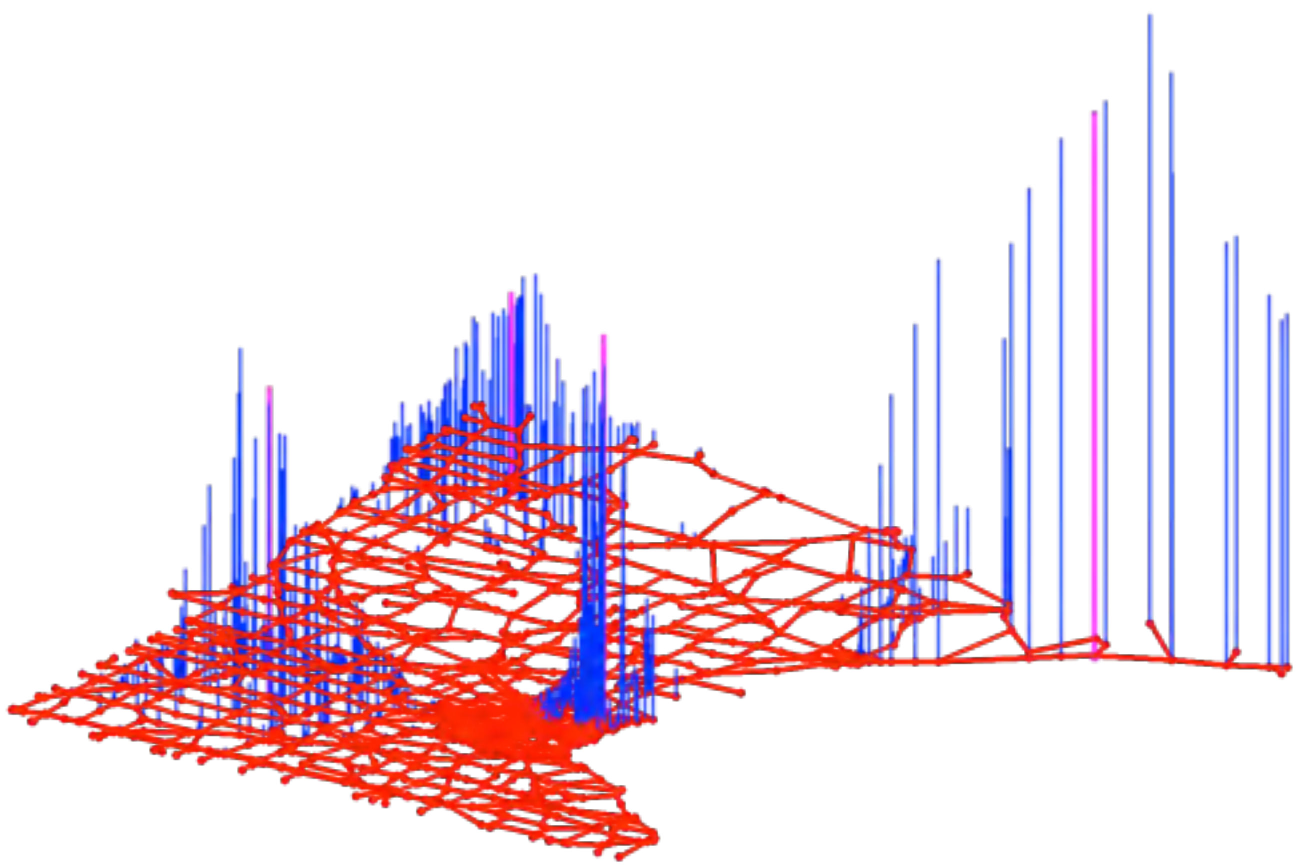
For $f \in L^2(\mathbb{R})$, in the weak sense

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

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

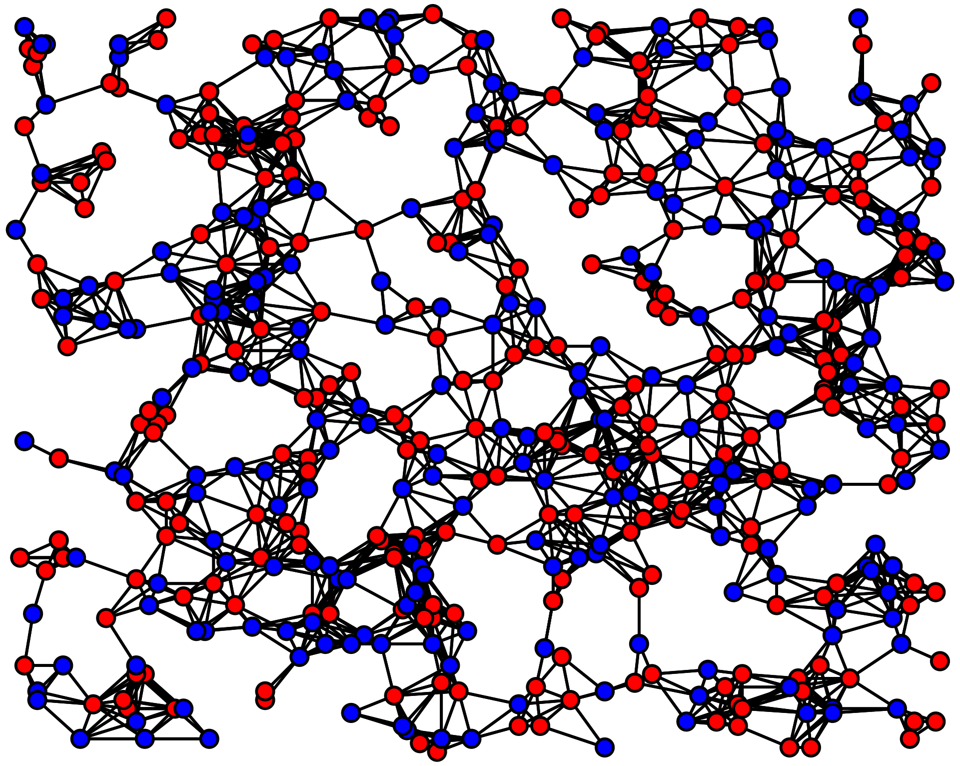
$$|(T_i g)(n)| \leq \sqrt{N} B_{\hat{g}}(d_{in} - 1),$$

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

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

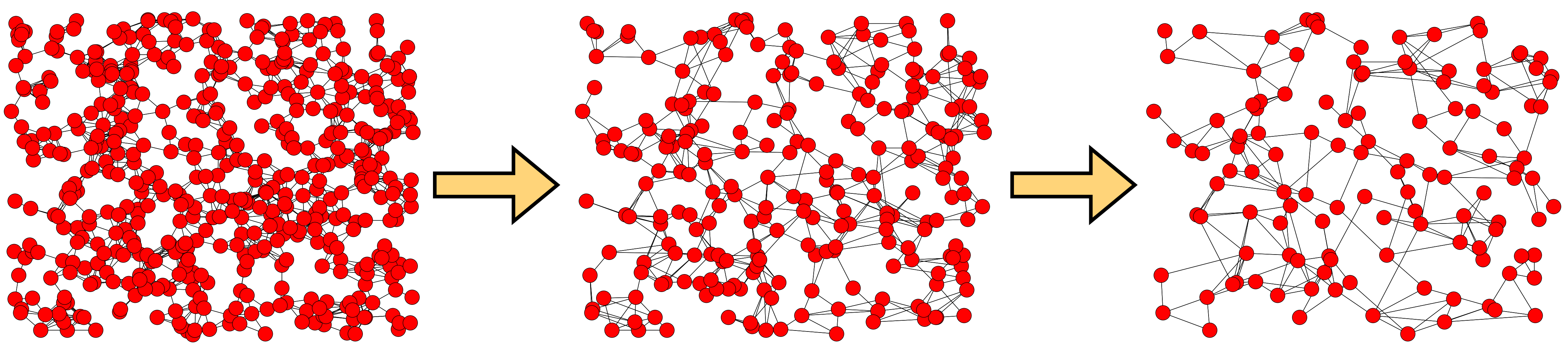
Downsampling and Graph Reduction

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

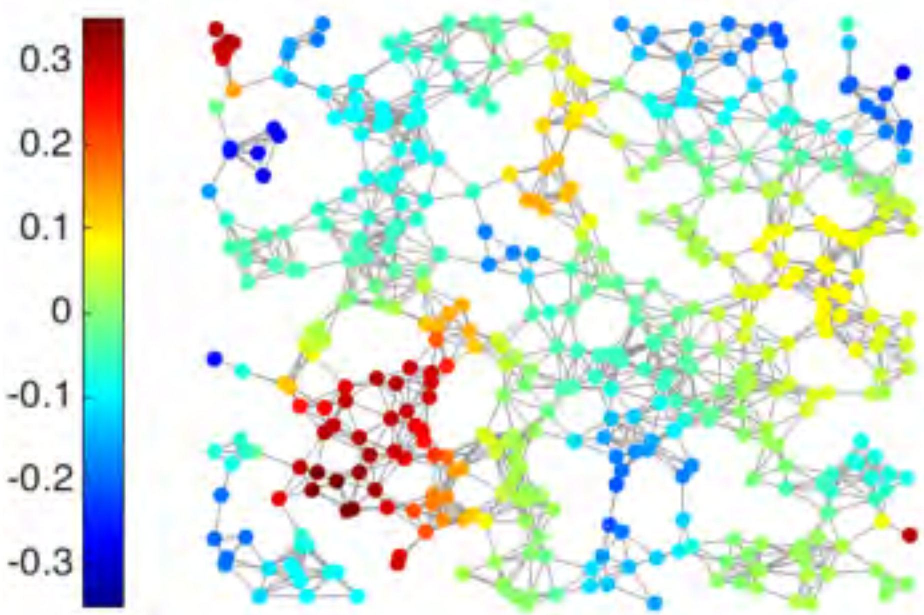
Graph Reduction

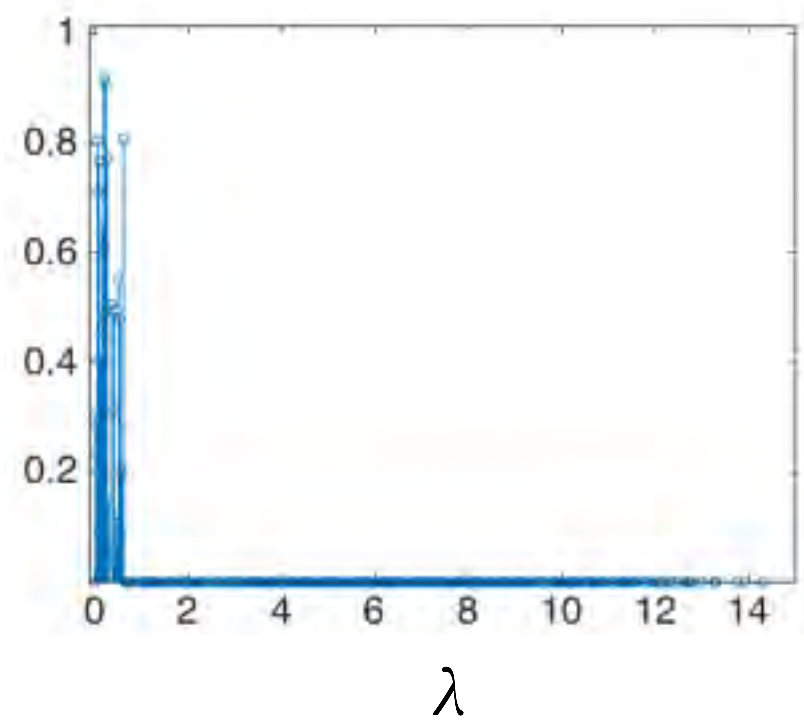


Sampling and Interpolation

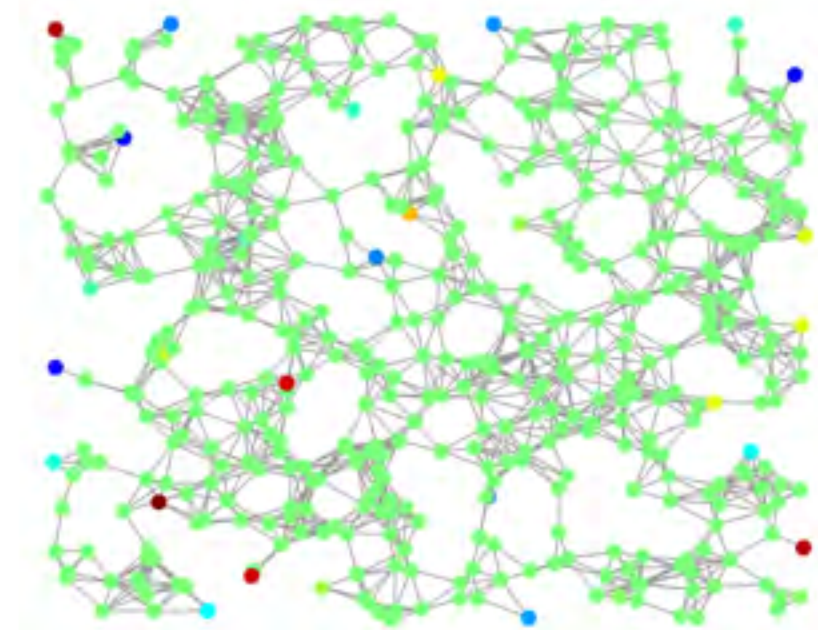
- How to sample a graph signal and interpolate from the samples?
- Subset V_s of vertices is a <u>uniqueness set</u> for a subspace P iff:

  If two signals in the subspace P have the same values on the vertices in the uniqueness set, then they are the same signal
- Example: subspace of globally smooth signals with band limit λ_{29}



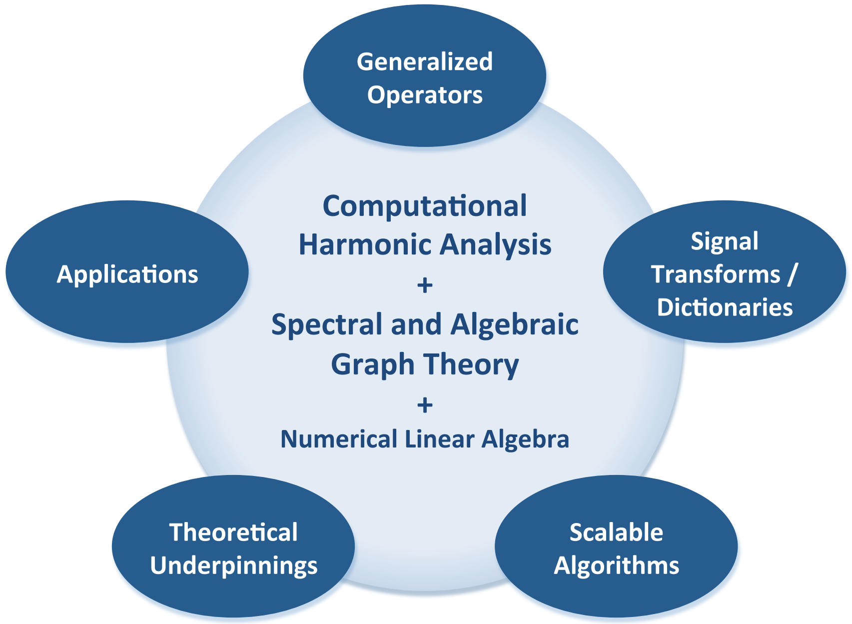


Example signal

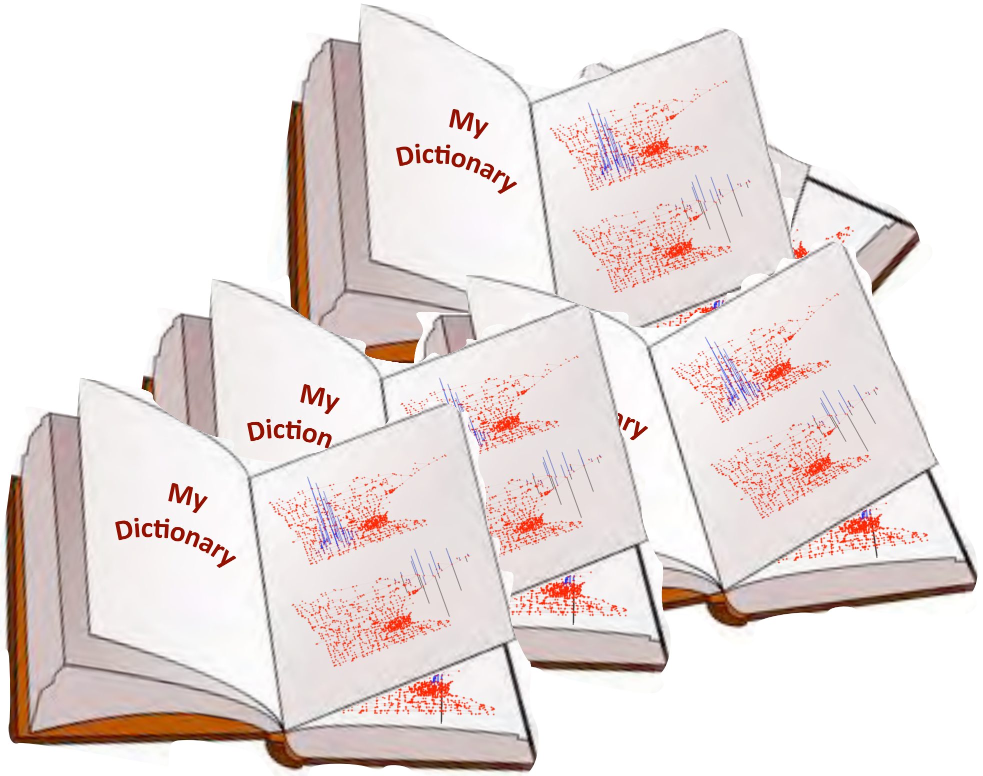


Uniqueness set

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



Survey of 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

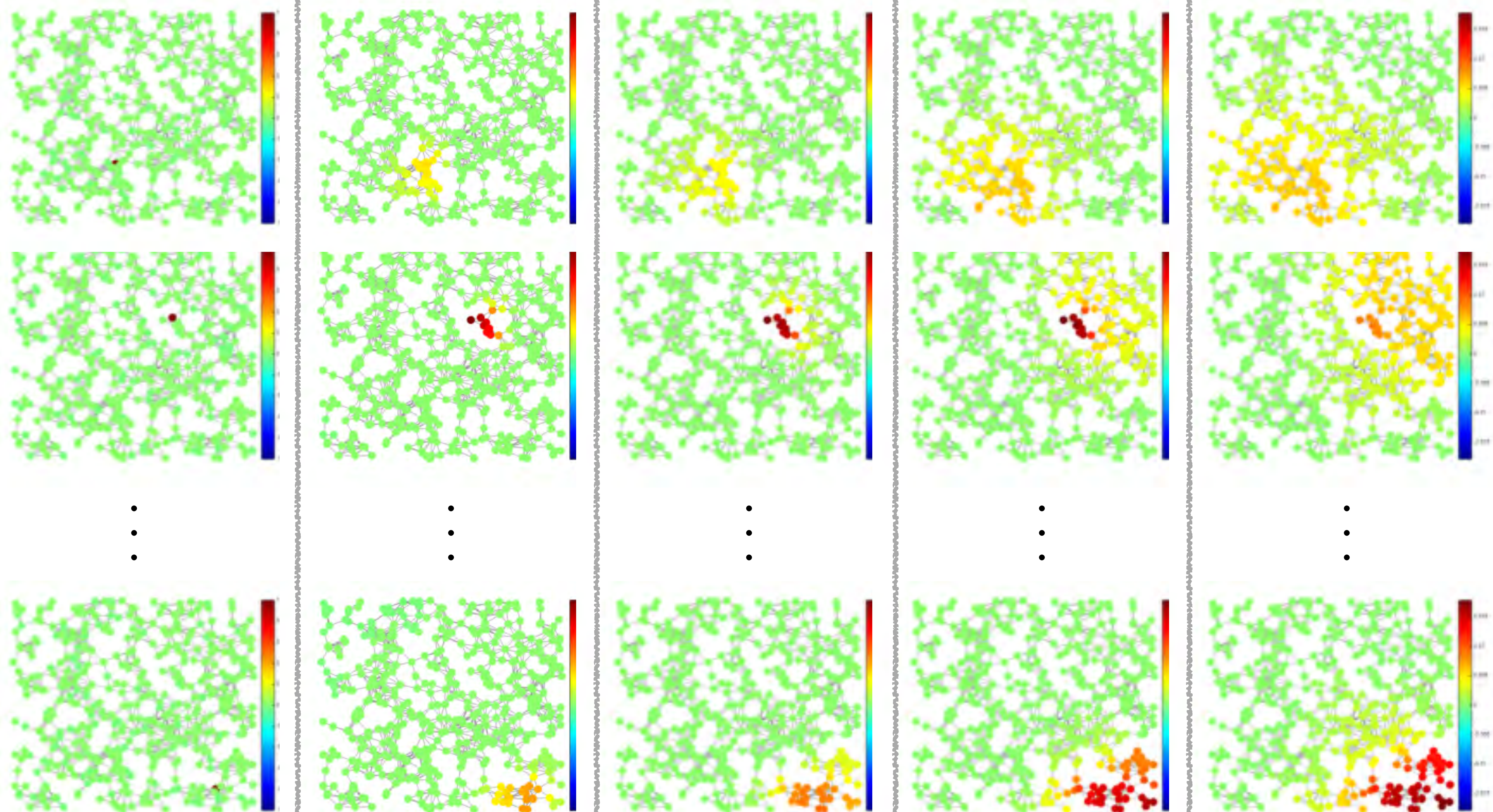
Survey of Approaches to Graph Signal Dictionary Design

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

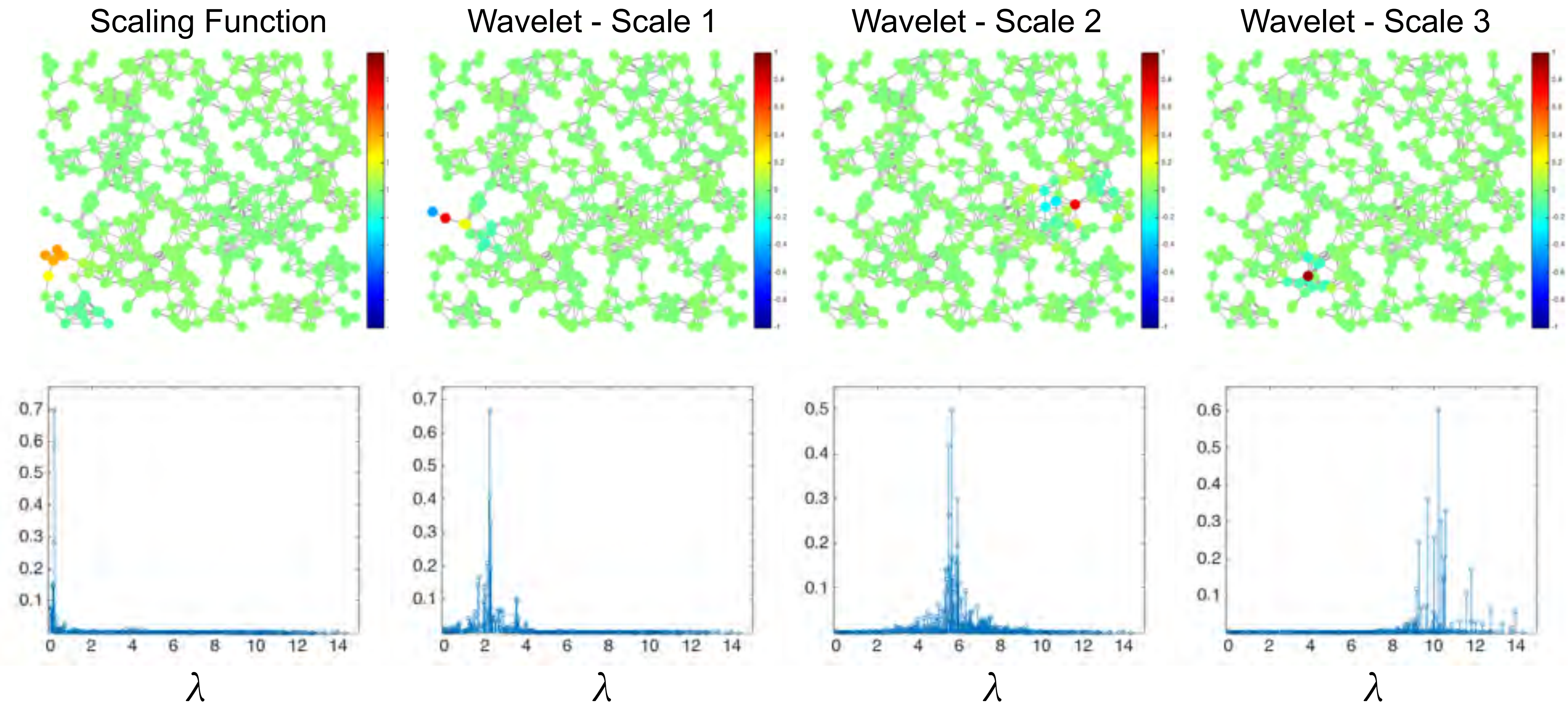
Multiresolution Scaling Function Spaces (Approximation Spaces)

	S^0	S^1	S^3	S^7	S^{15}
Numerical rank	500	402	109	52	32

Columns of S^{2^j-1} (500 in each matrix)



Diffusion Wavelet Atoms



- Coifman and Maggioni, Diffusion wavelets, ACHA, 2006
- Maggioni et al., Biorthogonal diffusion wavelets, SPIE, 2005
- Bremer et al., Diffusion wavelet packets, ACHA, 2006

Survey of Approaches to Graph Signal Dictionary Design

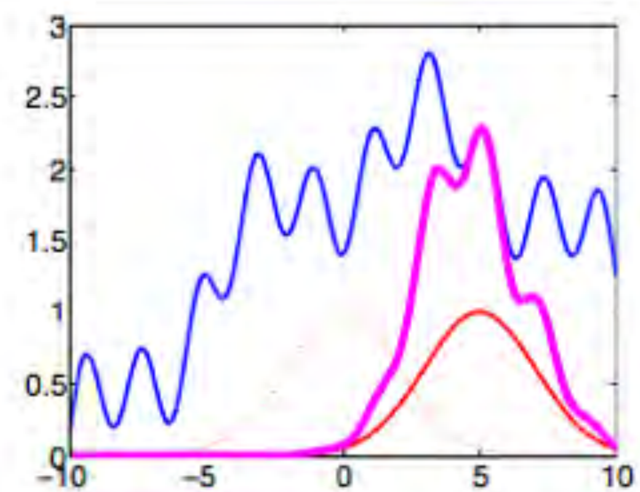
- Diffusion-based designs
- **Windowed graph Fourier transform**
- Spectral domain designs
- Generalized filter banks

Classical Windowed Fourier Transform

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

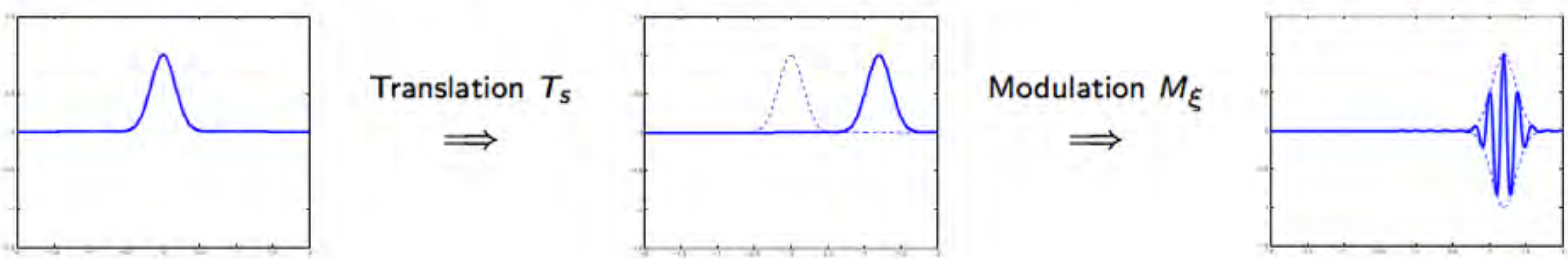
    - Localized oscillations appear frequently in audio processing, vibration analysis, radar detection, etc.
    - e.g., identify musical notes and melody at different times

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



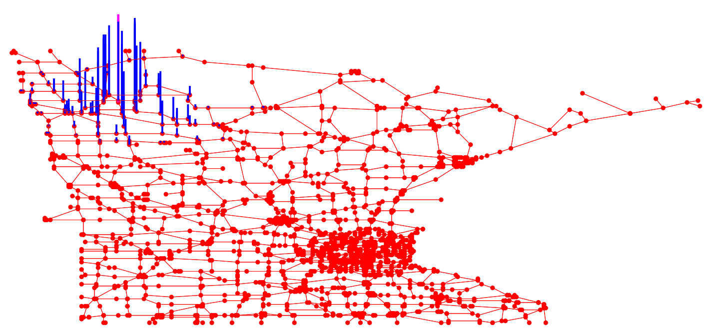
$$Sf(s,\xi) := \langle f, g_{s,\xi} \rangle = \int_{-\infty}^{\infty} f(t)\overline{g(t-s)}e^{-2\pi i \xi t}\,dt$$

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

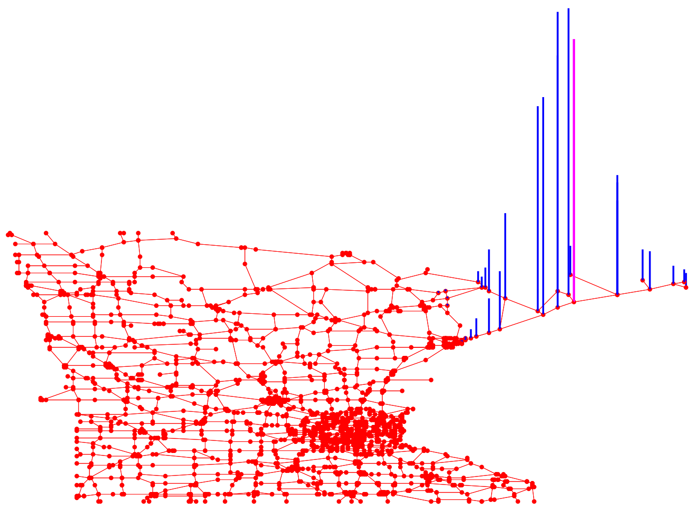


Windowed Graph Fourier Transform

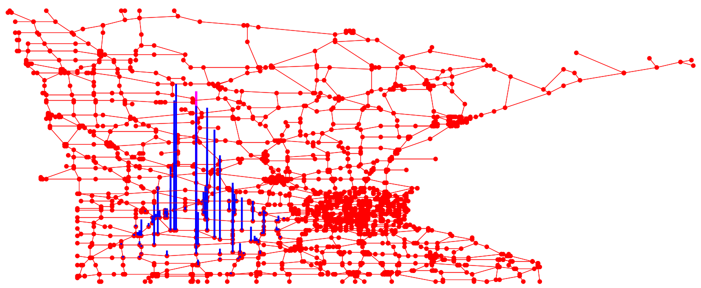
1. Translate a window g to each vertex of the graph



$T_{100}g$



$T_{200}g$



$T_{2000}g$

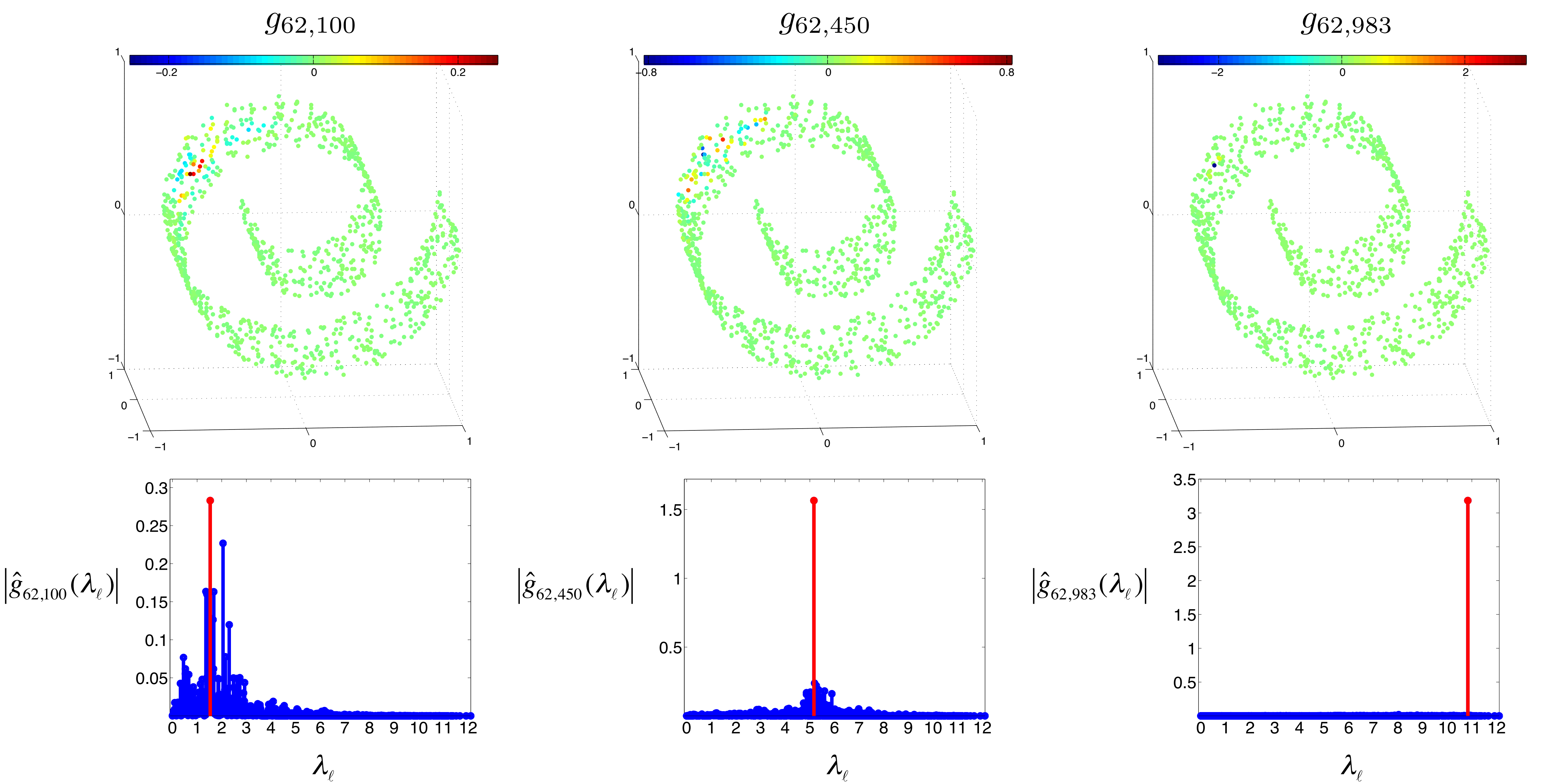
2. Multiply each component of the graph signal f of interest by the corresponding component of the translated window $T_i g$

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

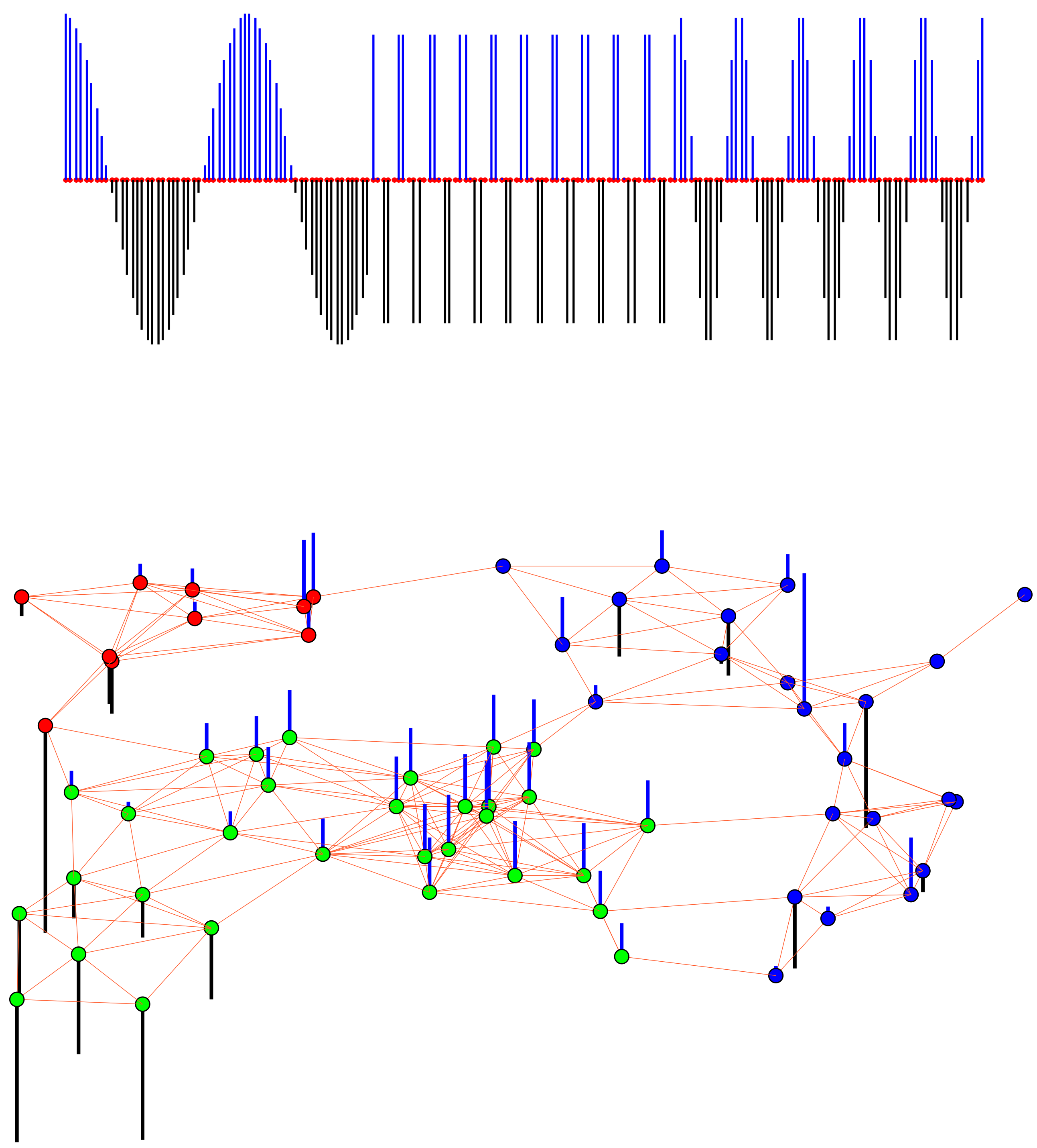
Shuman et al., Vertex-frequency analysis on graphs, ACHA, 2016

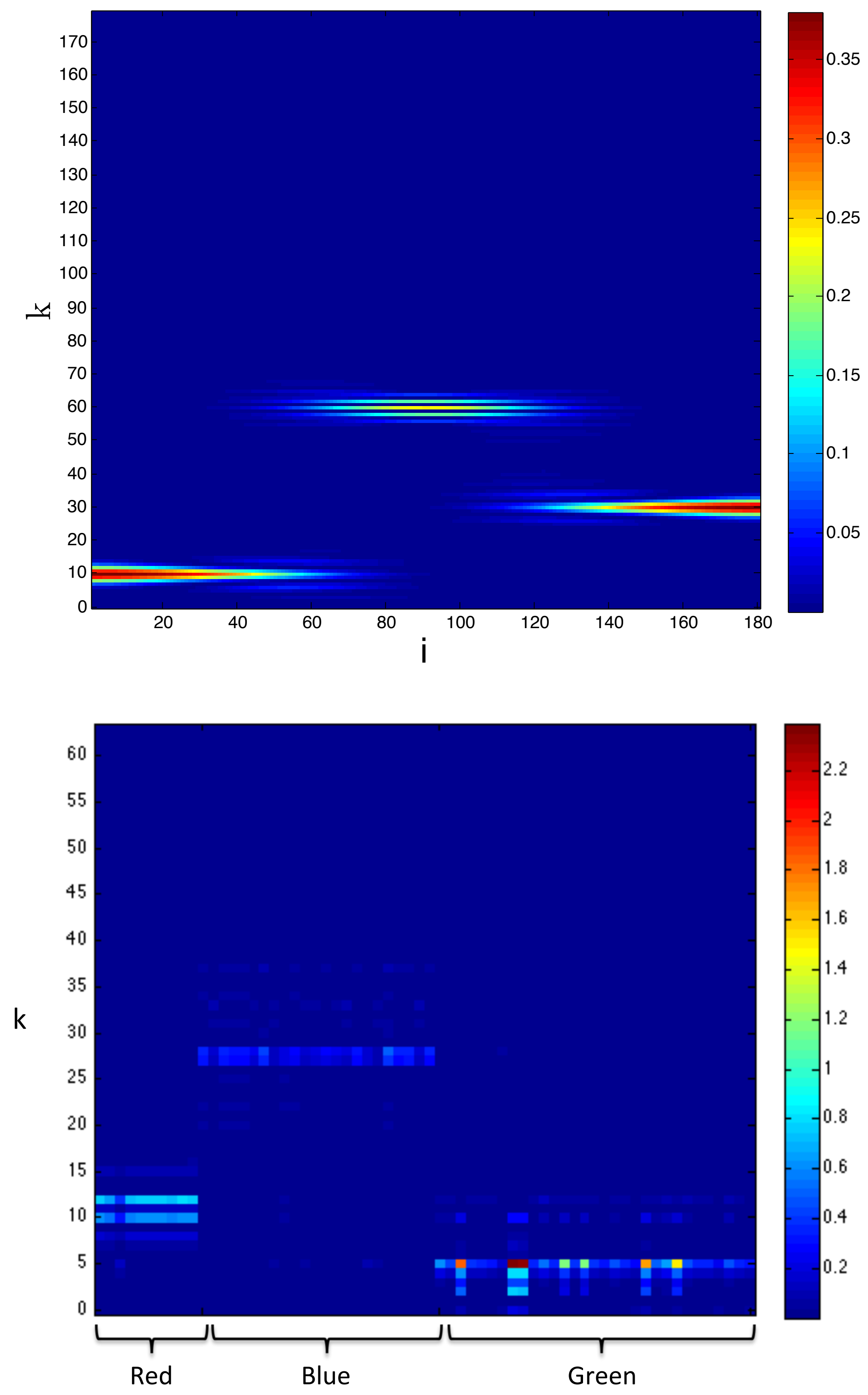
Windowed Graph Fourier Transform (cont.)

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



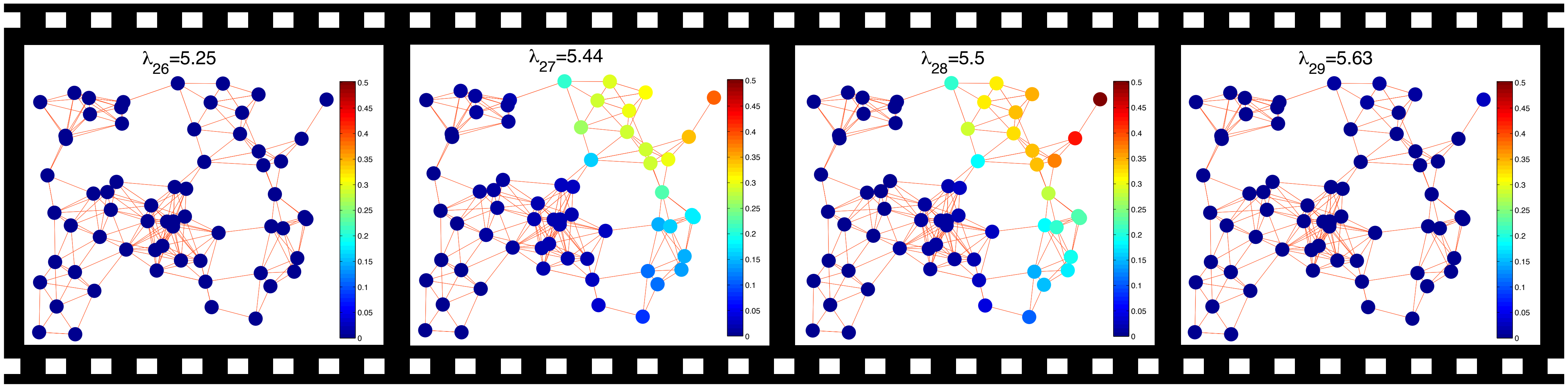
Spectrogram Examples

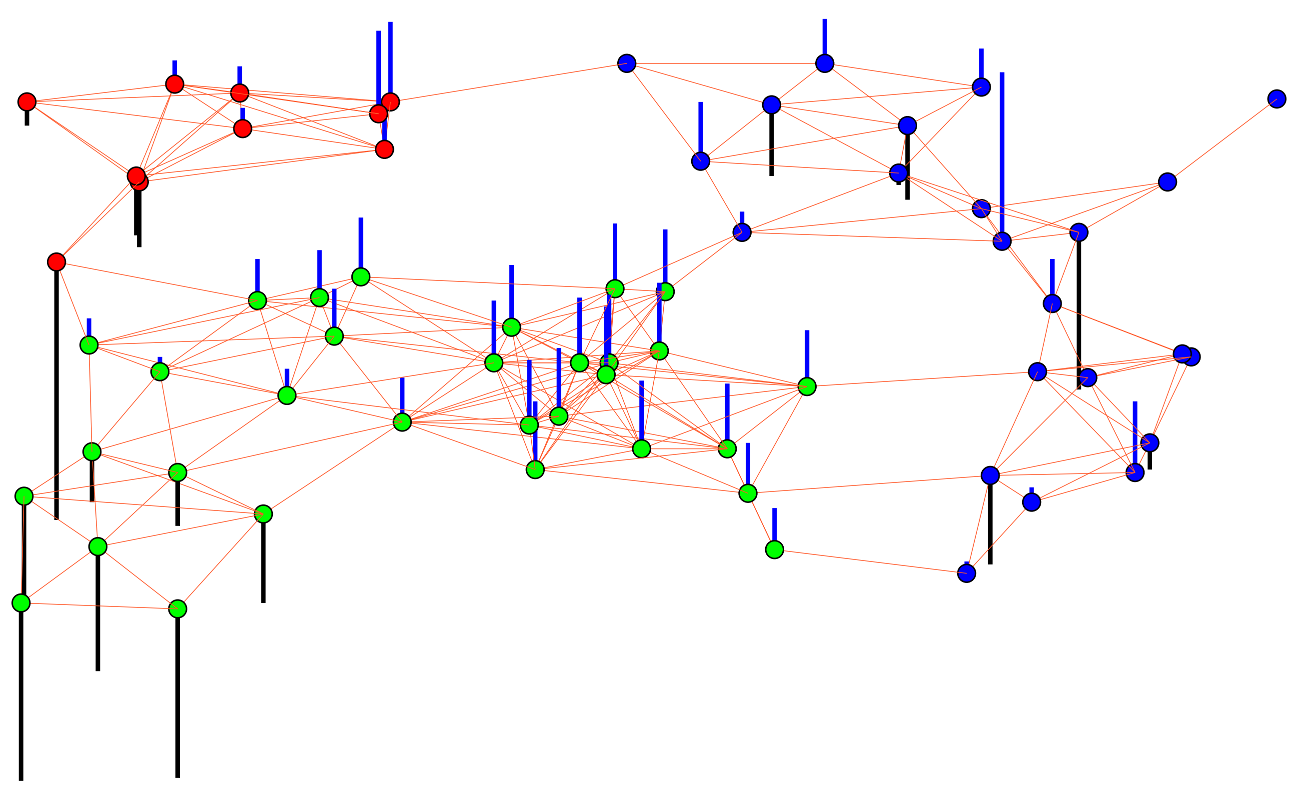


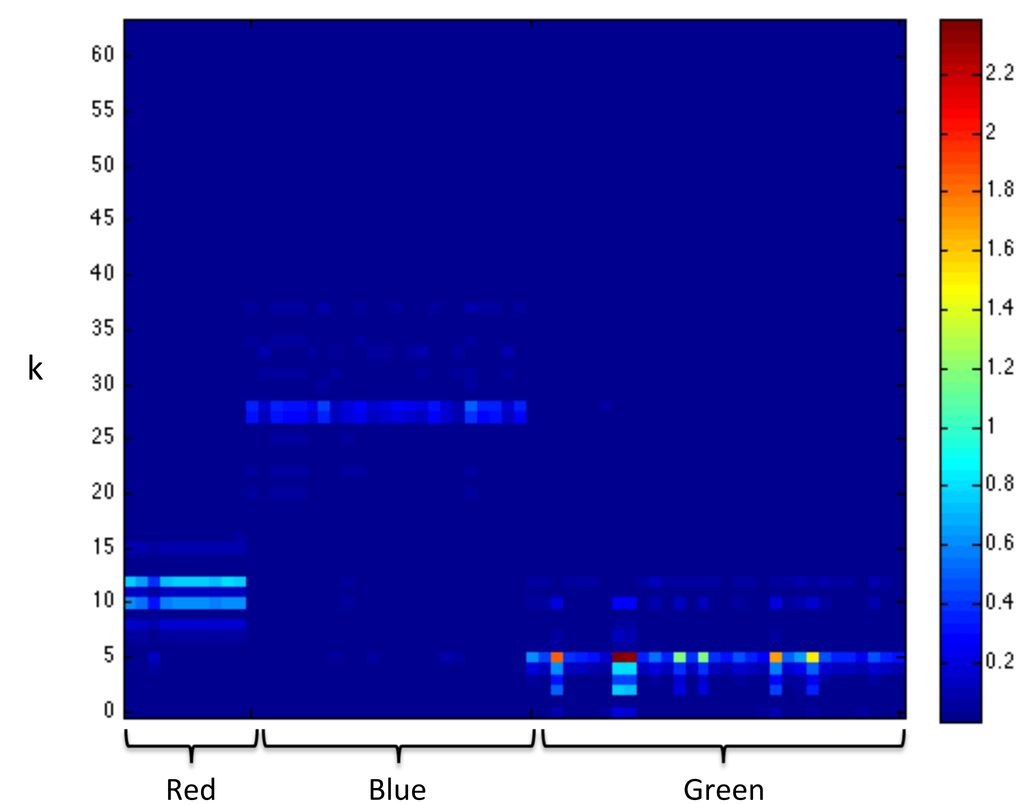


Spectrogram Examples

- Spectrogram = frequency-lapse video







Survey of Approaches to Graph Signal Dictionary Design

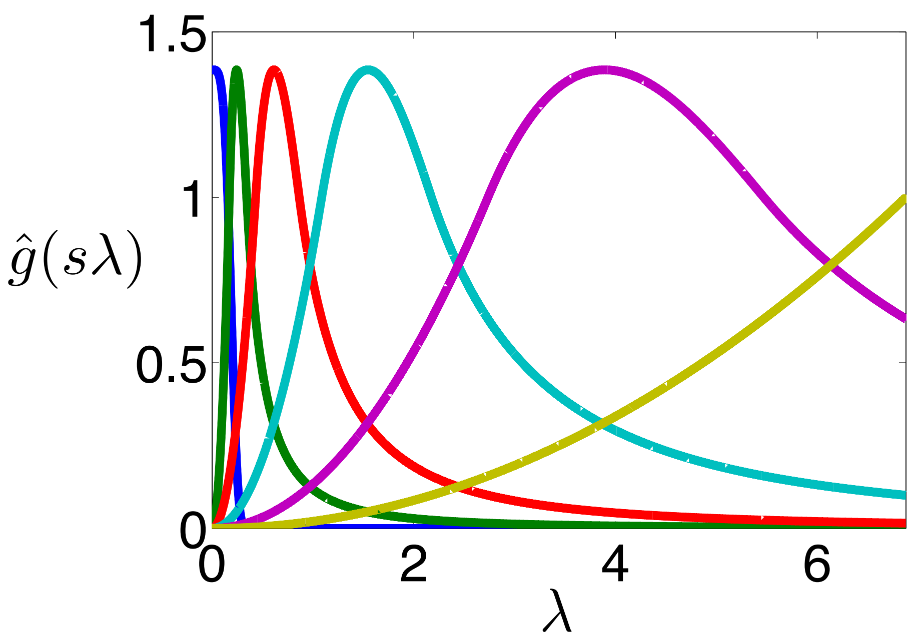
- Diffusion-based designs
- Windowed graph Fourier transform
- **Spectral domain designs**
- Generalized filter banks

Spectral Graph Wavelets

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

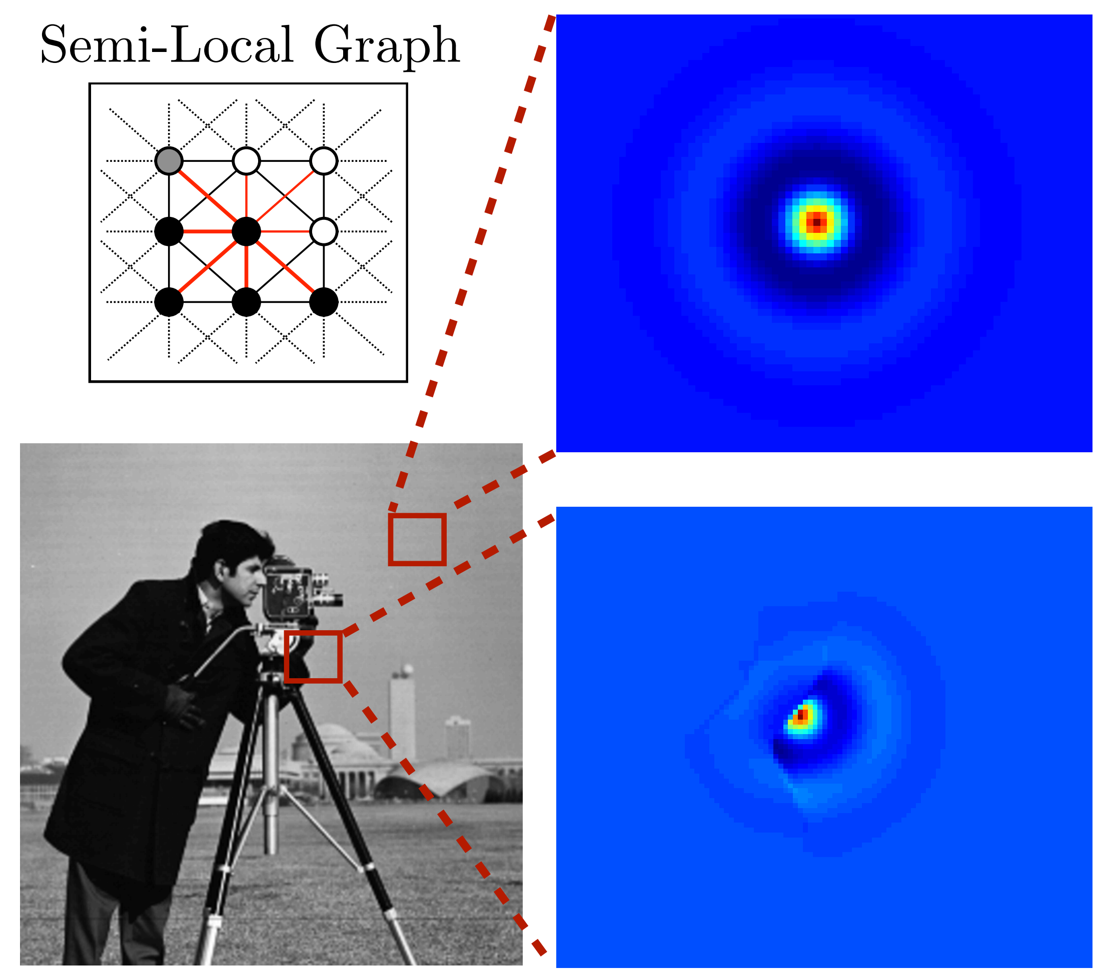
- Generalized dilation:

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

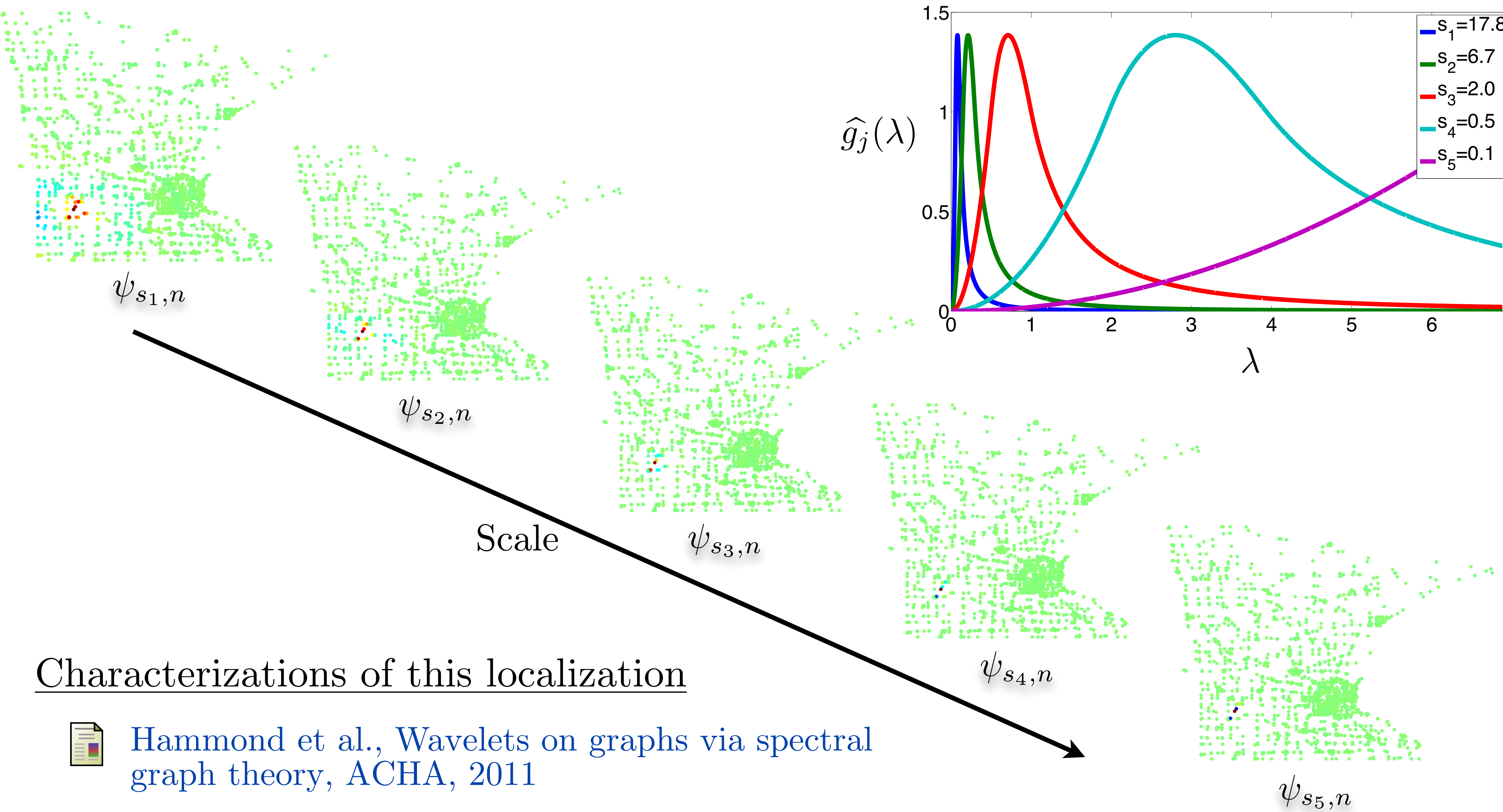


- Spectral graph wavelet at scale s, centered at vertex n:

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



Spectral Graph Wavelet Localization

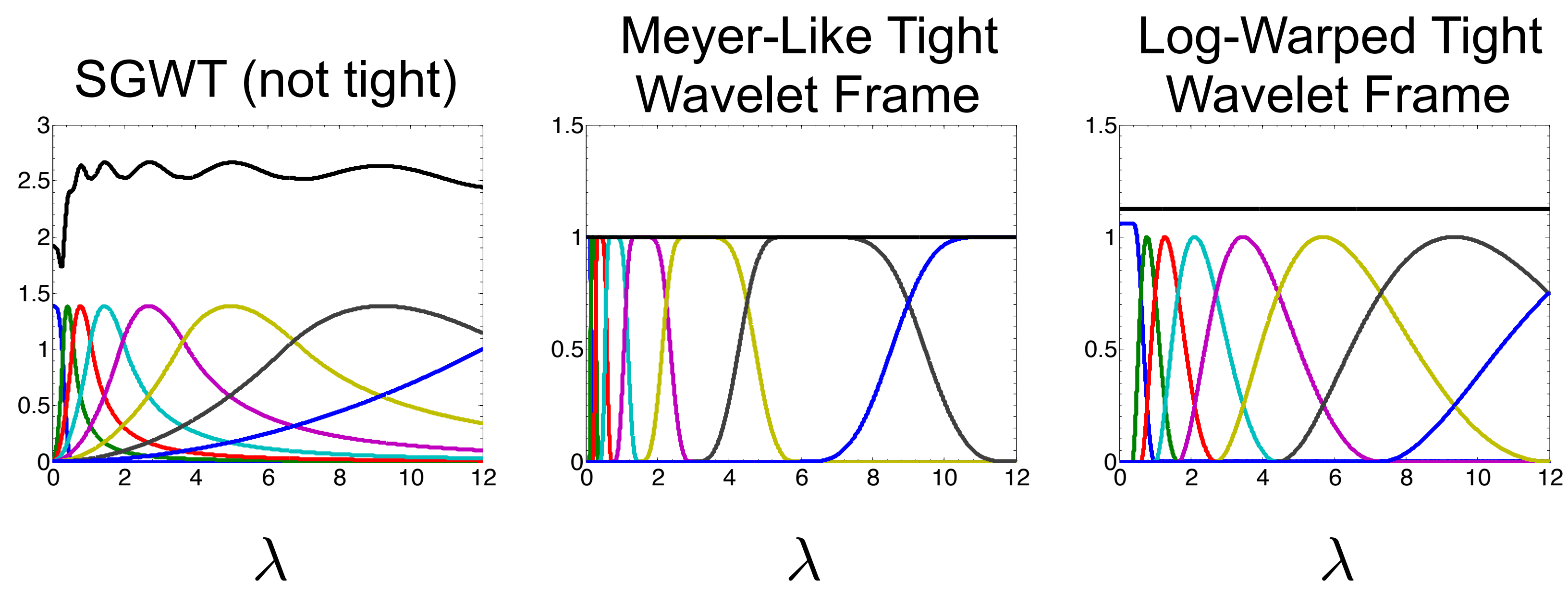


Characterizations of this localization

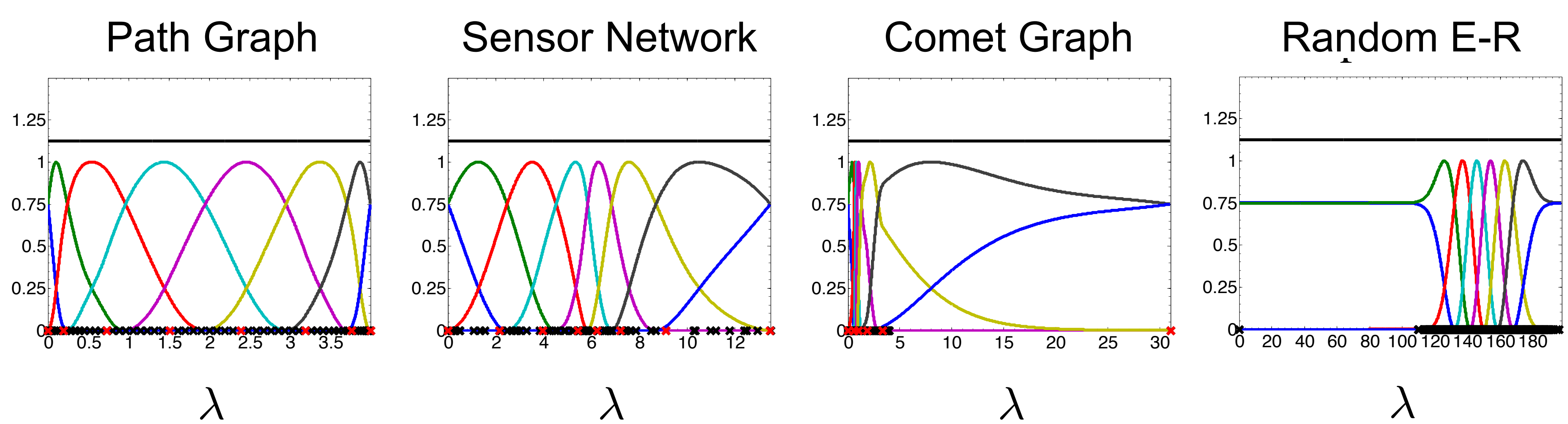
- Hammond et al., Wavelets on graphs via spectral graph theory, ACHA, 2011
- Shuman et al., Vertex-frequency analysis on graphs, ACHA, 2016

Translated Kernel Variants

Tight Wavelet Frames



Spectrum-Adapted

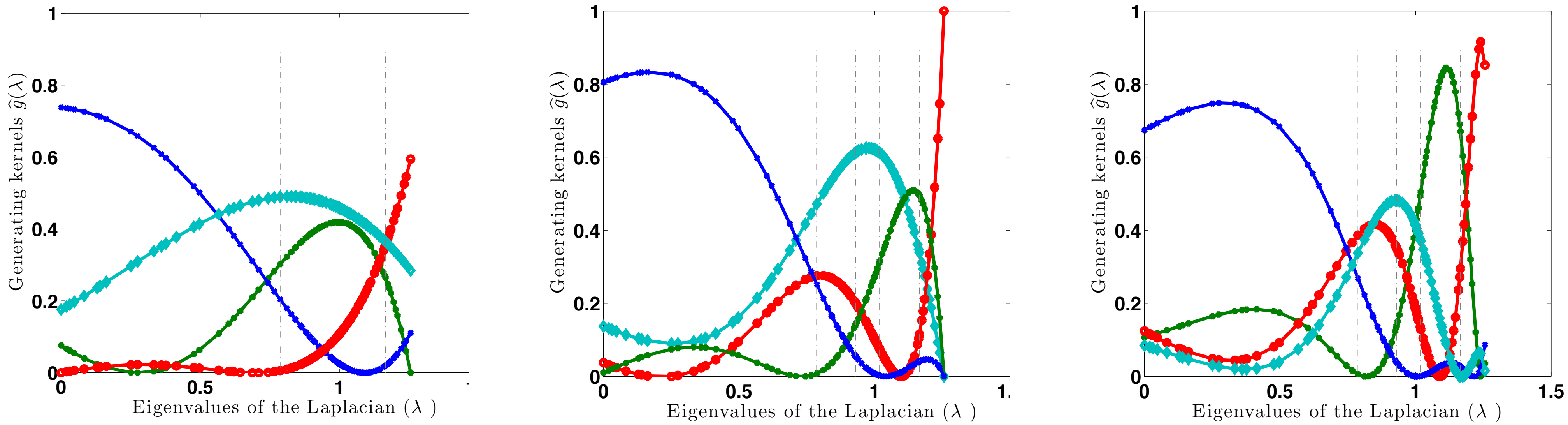


- Leonardi and Van De Ville, Tight wavelet frames on multislice graphs, TSP, 2013
- Shuman et al., Spectrum-adapted tight graph wavelet and vertex-frequency frames, TSP, 2015

Translated Kernel Variants (cont.)

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

Parametric Learning



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

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

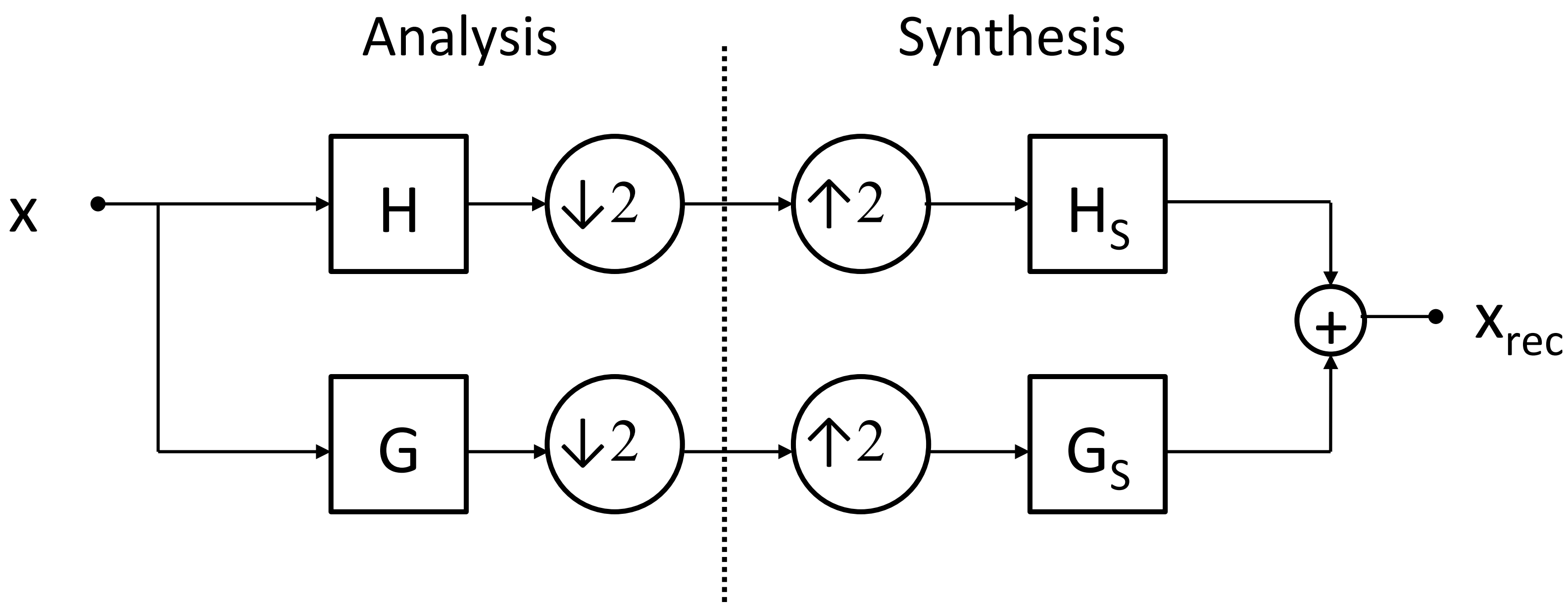
Thanou et al., Learning parametric dictionaries for signals on graphs, TSP, 2014

Survey of Approaches to Graph Signal Dictionary Design

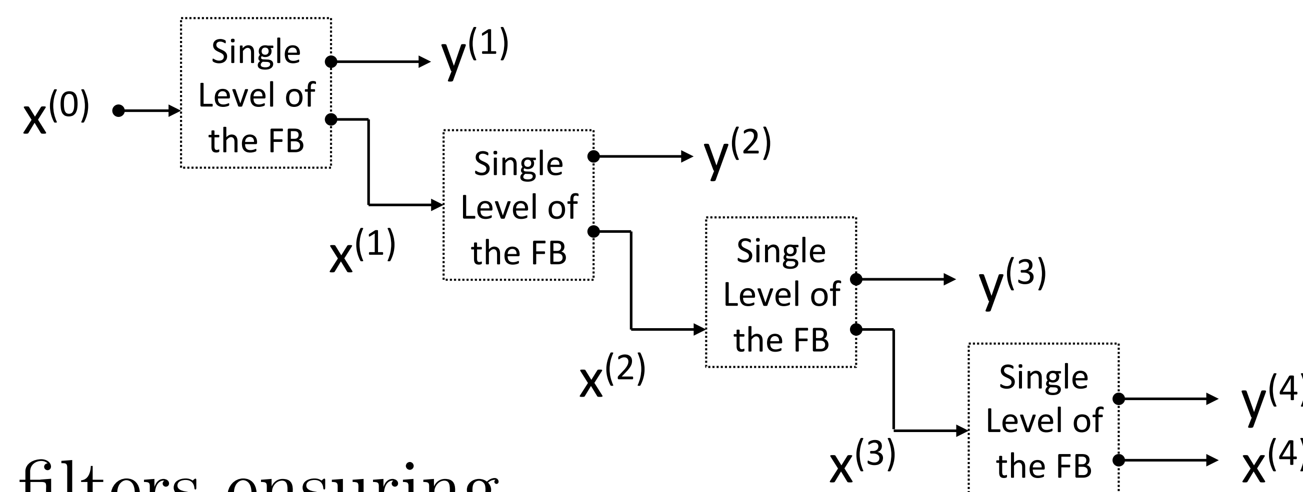
- Diffusion-based designs
- Windowed graph Fourier transform
- Spectral domain designs
- **Generalized filter banks**

1D Wavelets Via Filter Banks

Classical 2-Channel Critically Sampled Filter Bank

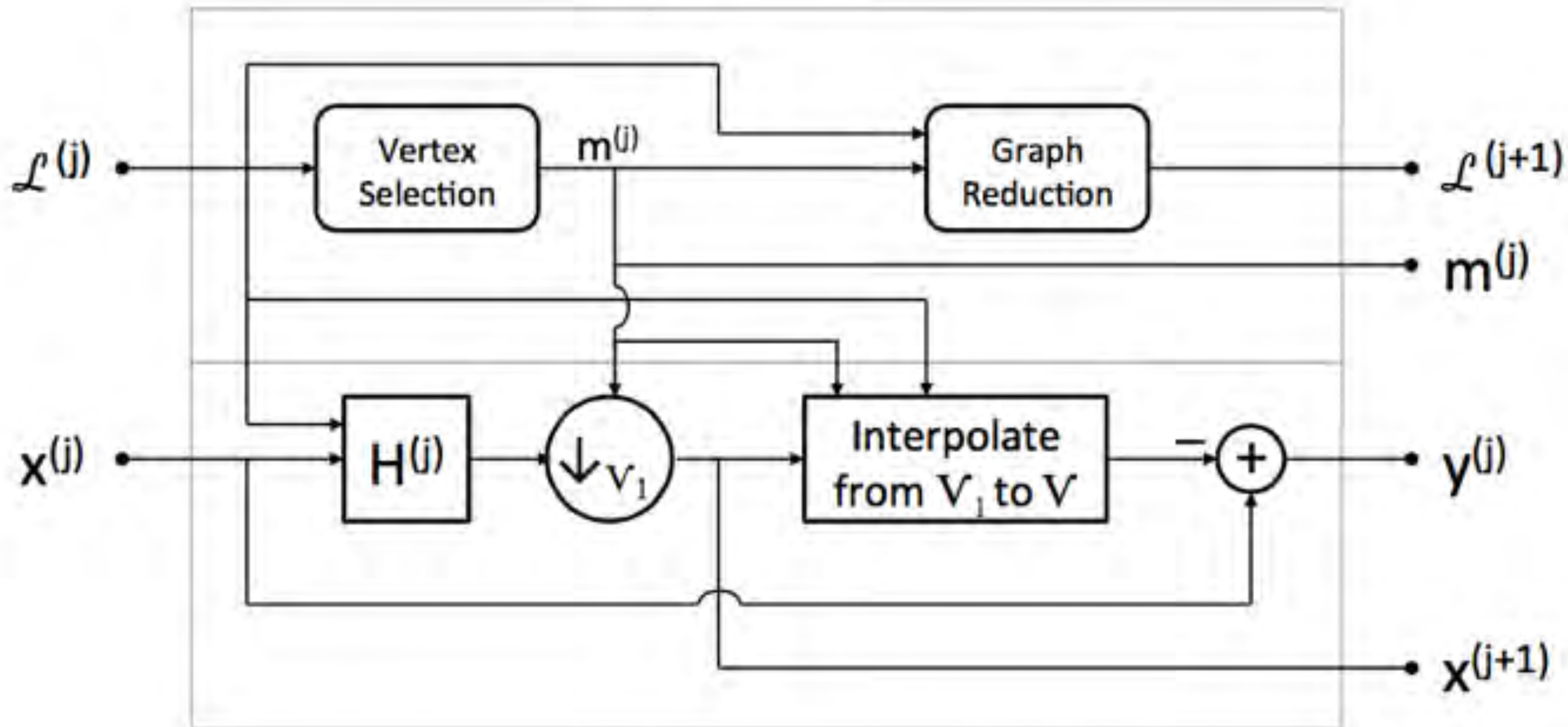


Iterating Low Pass Branch Yields Wavelets



- 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

Architecture Example 1: A Multiscale Pyramid Transform for Graph Signal

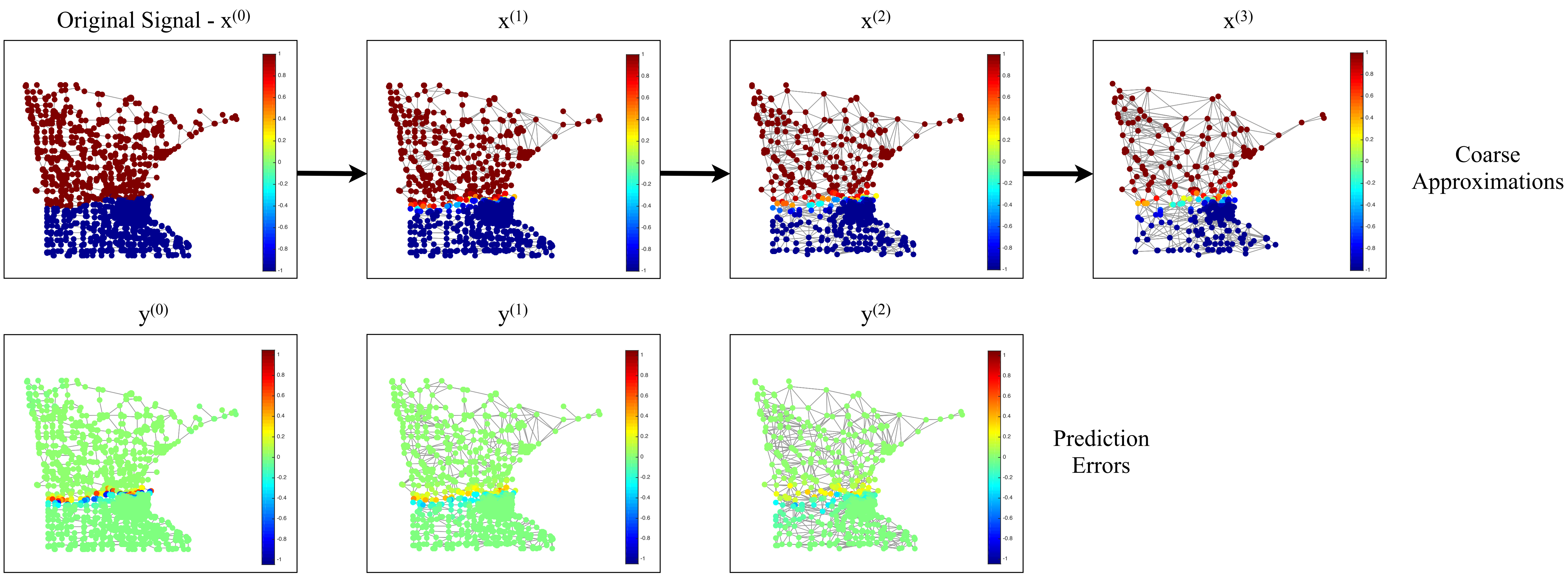


- Generalization of classical Laplacian pyramid of Burt and Adelson
- Overcomplete transform
- Replace classical prediction step (upsample then low pass filter) with a graph interpolation operator
- Iterate on $x^{(j+1)}$: Yields a multi-resolution of the underlying graph and a multi-resolution approximation of the the graph signal

Burt and Adelson, The Laplacian pyramid as a compact image code, TCOM, 1983

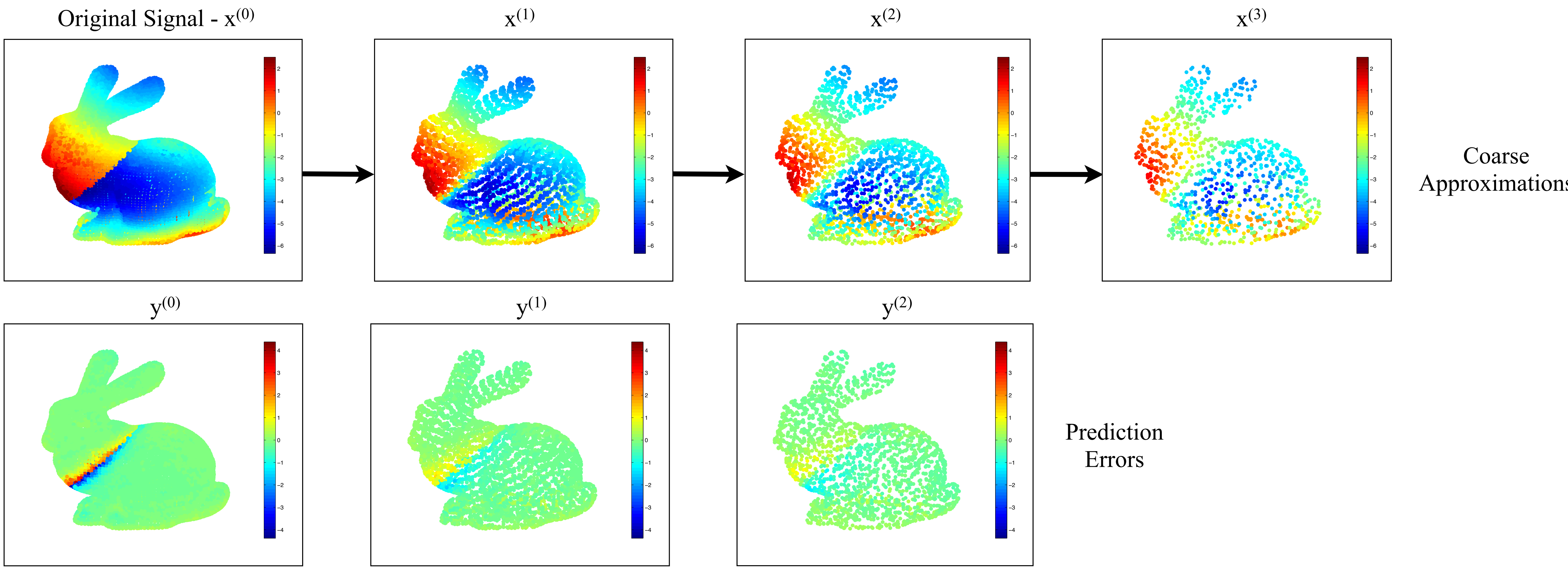
Shuman et al., A multiscale pyramid transform for graph signals, TSP, 2016

A Multiscale Pyramid Transform for Graph Signals
Multiresolution Examples



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

A Multiscale Pyramid Transform for Graph Signals
Multiresolution Examples



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

A Multiscale Pyramid Transform for Graph Signals
Compression Example

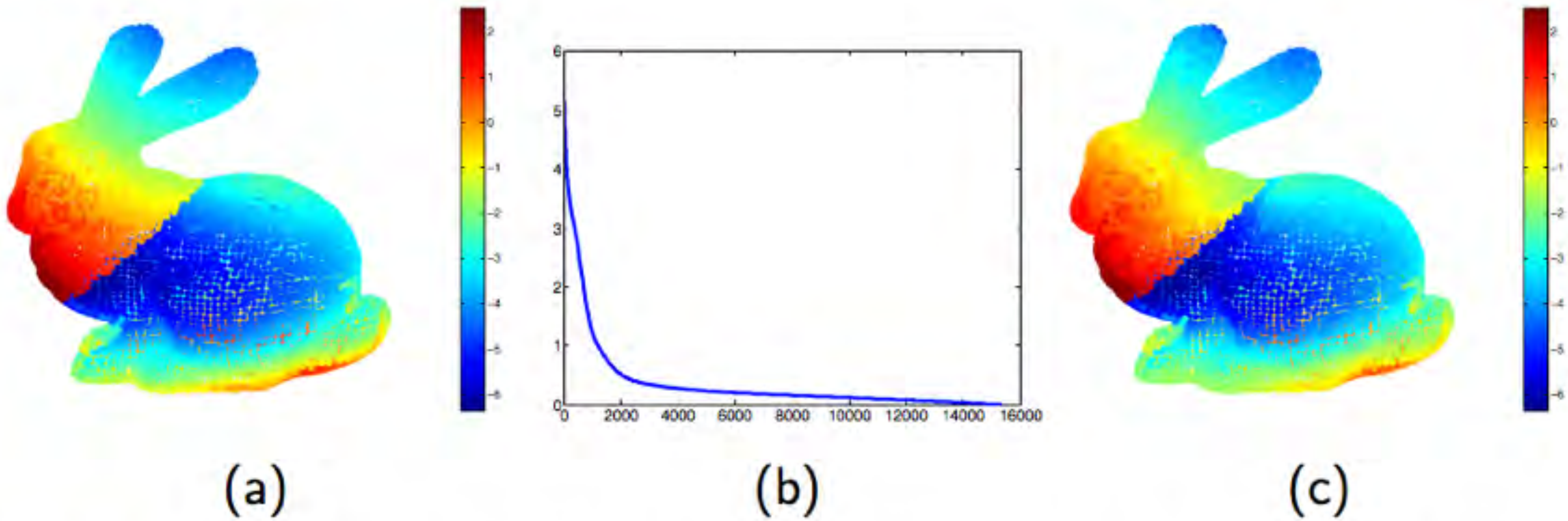
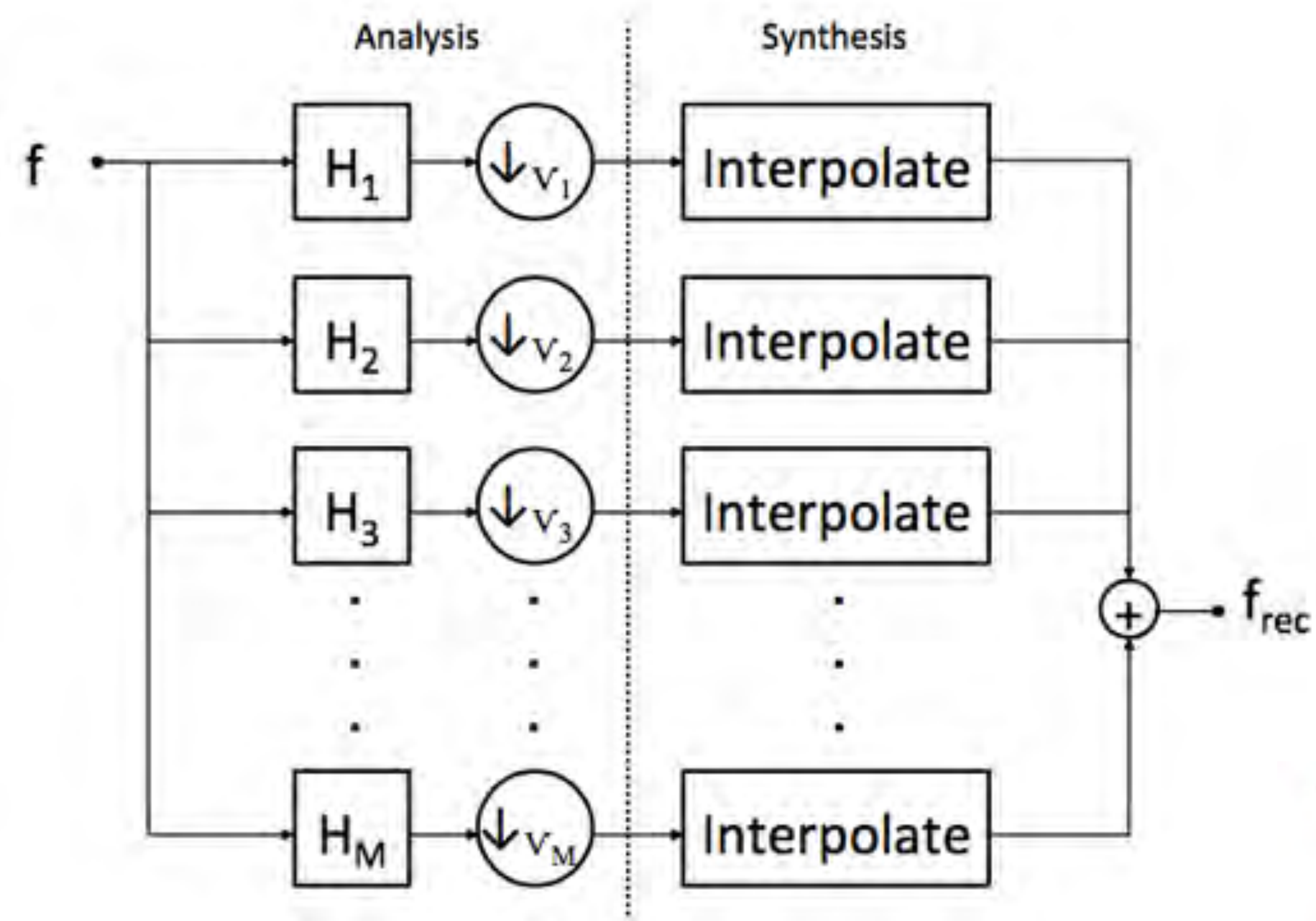


Figure: Compression example. (a) The original piecewise-smooth signal with a discontinuity on the Stanford bunny. (b) The sorted magnitudes of the 15346 pyramid transform coefficients. (c) The reconstruction from the 2724 coefficients with the largest magnitudes, using the least squares synthesis.

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

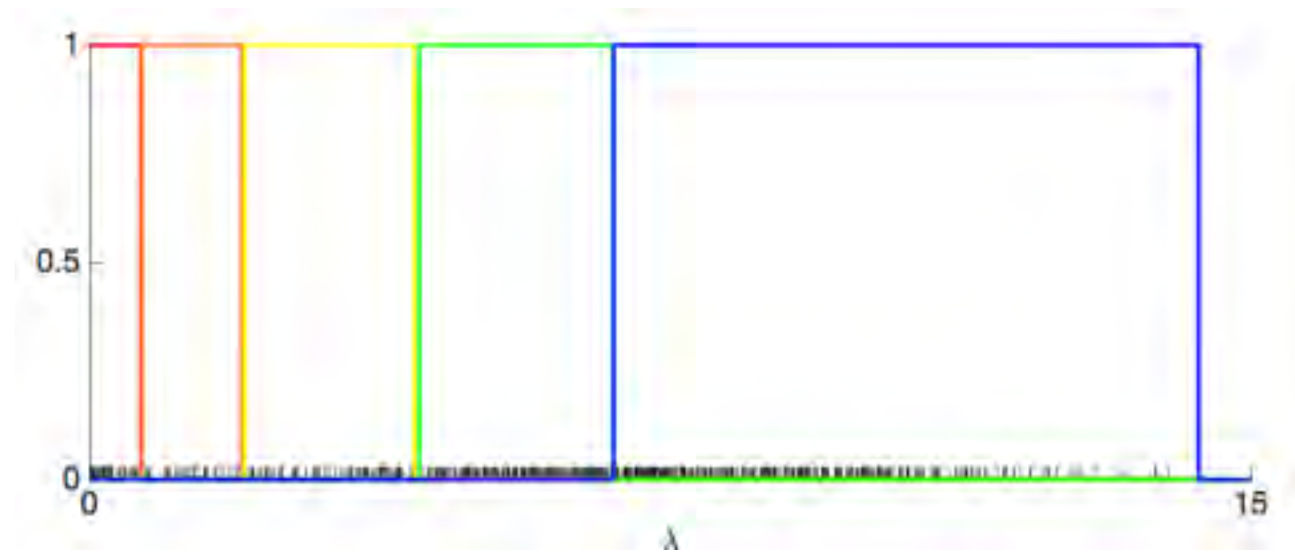
Architecture Example 2: M-Channel Critically Sampled Graph Filter Bank

Architecture



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

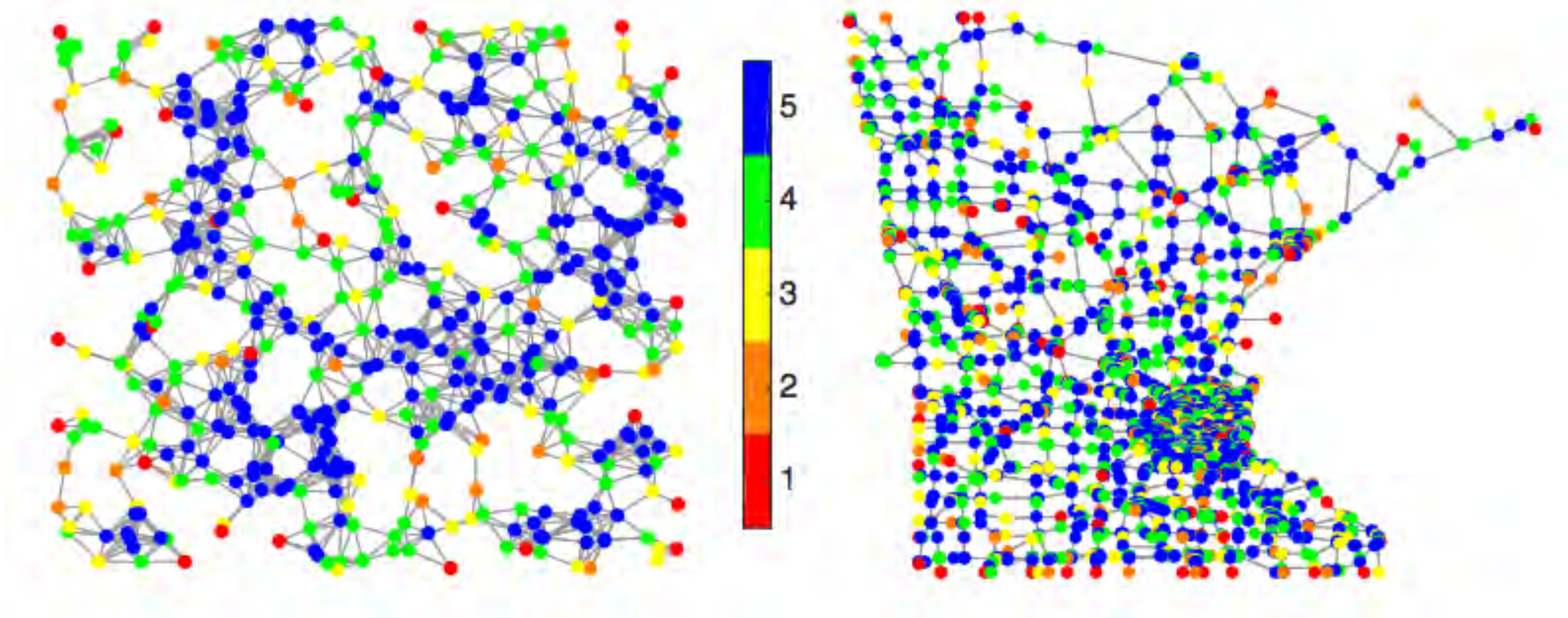
Ideal Filter Bank



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

Sampling and Interpolation

- How to choose which vertices to sample for each band?
- Partition into uniqueness sets for ideal filter bank subspaces:



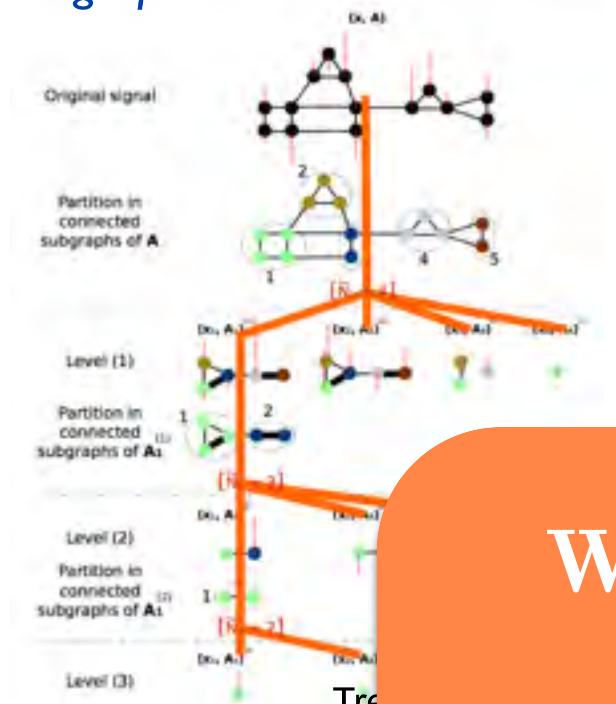
- To avoid a full eigendecomposition, we can use random, non-uniform sampling and fast, approximate reconstruction methods

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

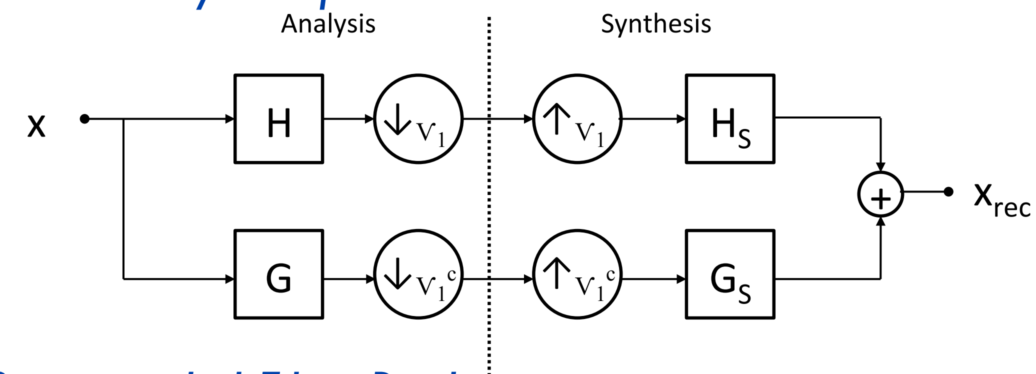
How to Evaluate Dictionaries / Open Research Questions

Dictionaries Galore

Subgraph Filter Bank

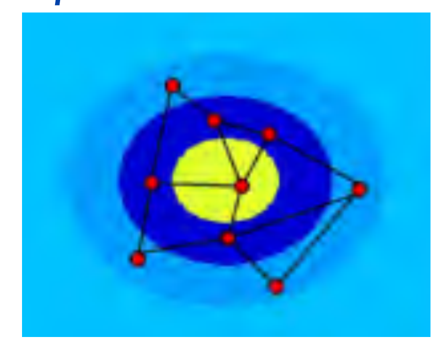


Critically-Sampled Filter Bank



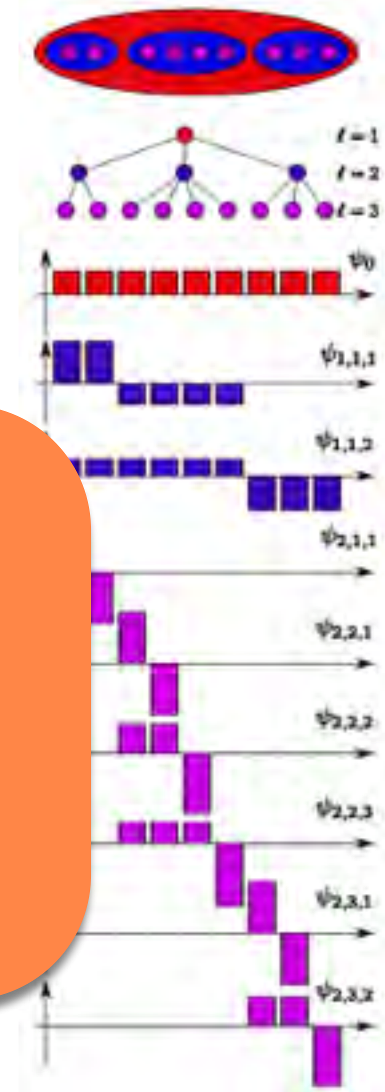
Oversampled Filter Bank

Spatial Wavelets



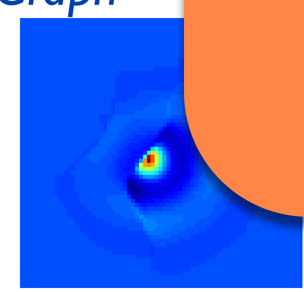
Crovella and Kolaczyk, 2003

Hierarchical Trees

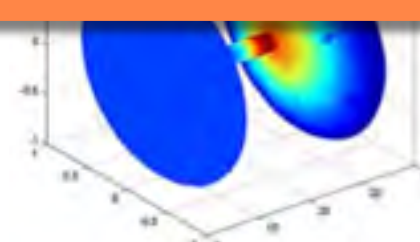


Gavish et al, 2010

Spectral Graph Wavelets



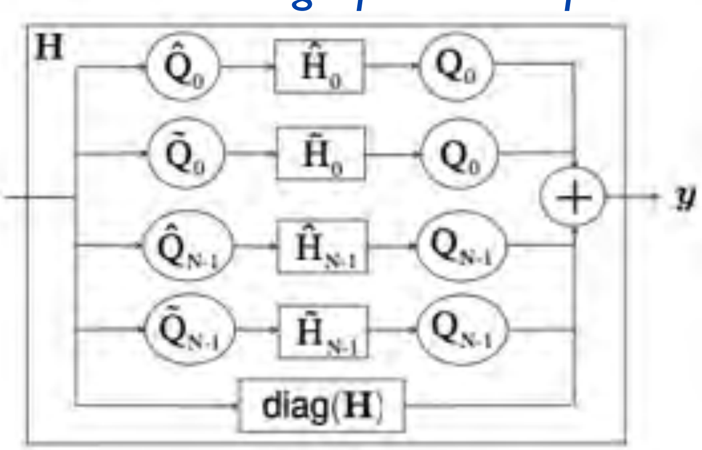
Which multiscale transforms for signals on graphs are well-suited for which signal processing tasks, which classes of signals, and which types of graphs?



Coifman and Maggioni, 2006

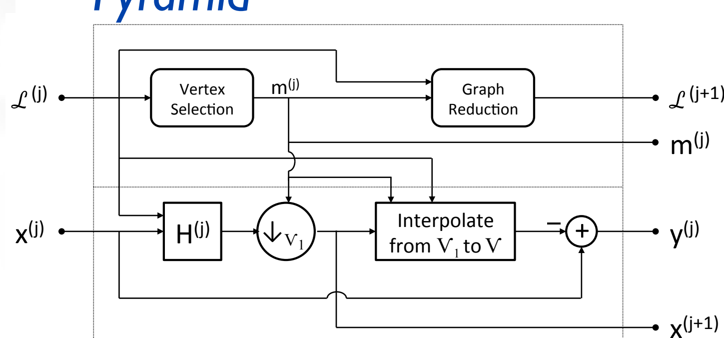
Narang and Ortega, TSP, 2012

Circulant Subgraph Decomposition

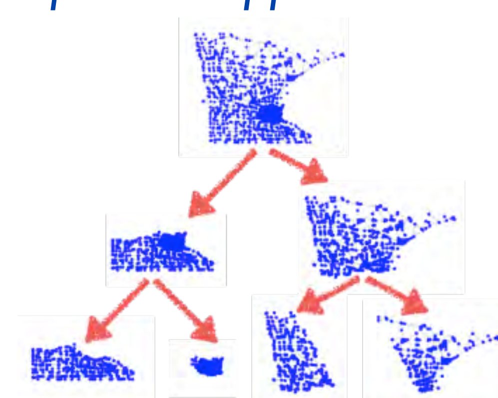


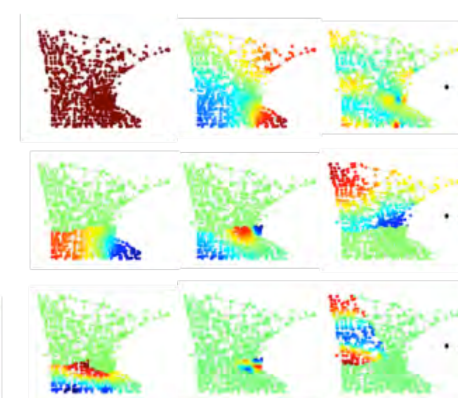
Ekambaram, Ph.D. Thesis, 2013

Pyramid



Top-Down Approaches





1. Signal Models and Sparsity

- For signals on Euclidean data domains, we have results characterizing classes of signals that are well-approximated by different transforms
  - e.g., piecewise-smooth 1D signals by wavelets, 2D cartoons with curvilinear discontinuities by curvelets/shearlets

- Empirically, many of the proposed transforms sparsely represent smooth and piecewise smooth graph signals, but there is little in the way of theoretical guarantees to date

- Theoretical connections between properties of graph signals, the graph structure, and the decay of transform coefficients?

2. Application-Driven Developments

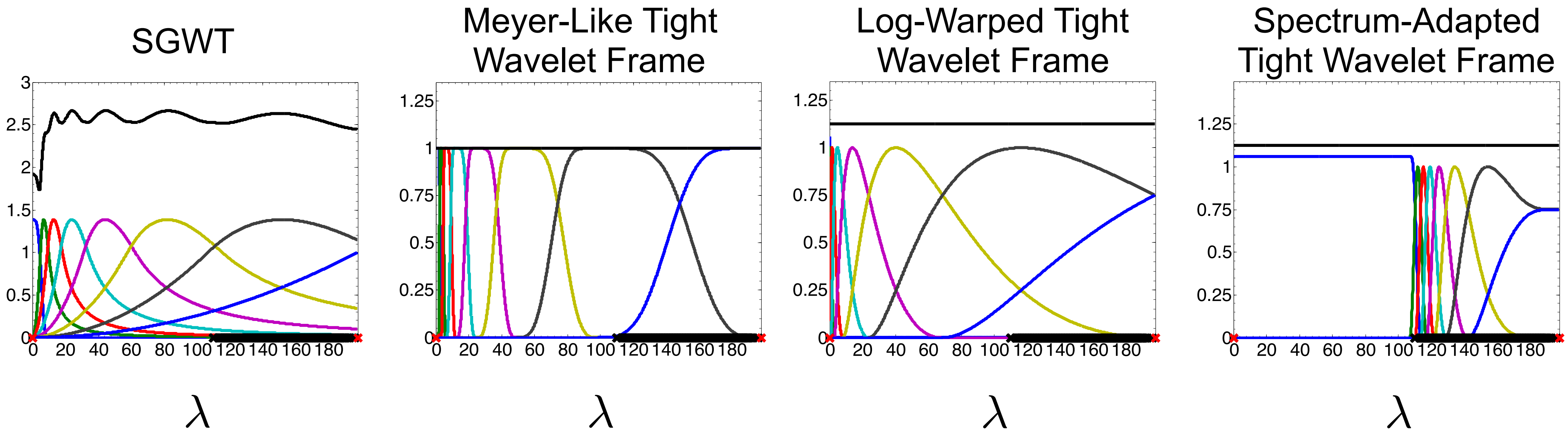
- Recent applications include brain signals, road traffic, video compression, epidemic outbreaks, climate data, and social networks

- Which mathematical models actually match graph signals found in applications?

- How can signal processing tasks arising in certain applications inform dictionary design?

3. Cumulative Coherence of Atoms

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



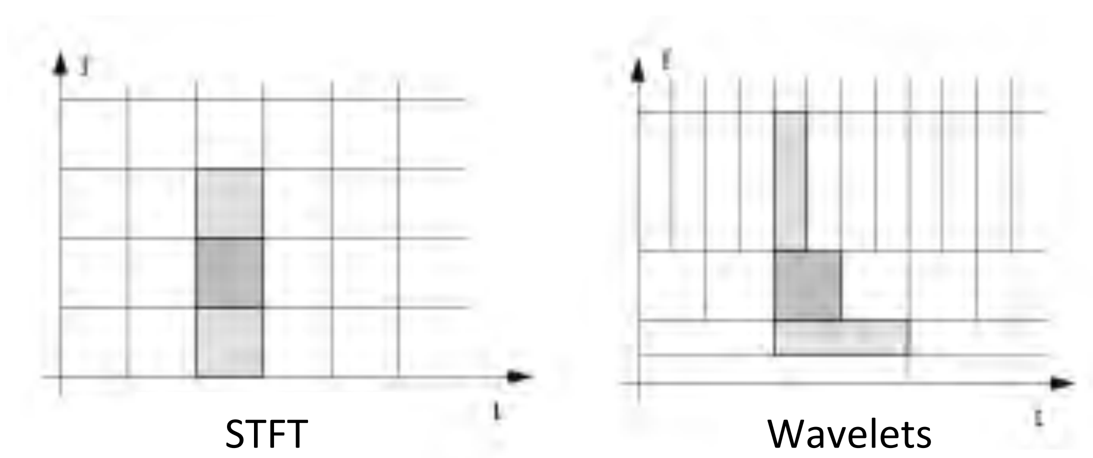
- Cumulative coherence for a given sparsity level k

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

4. Vertex-Frequency Tiling

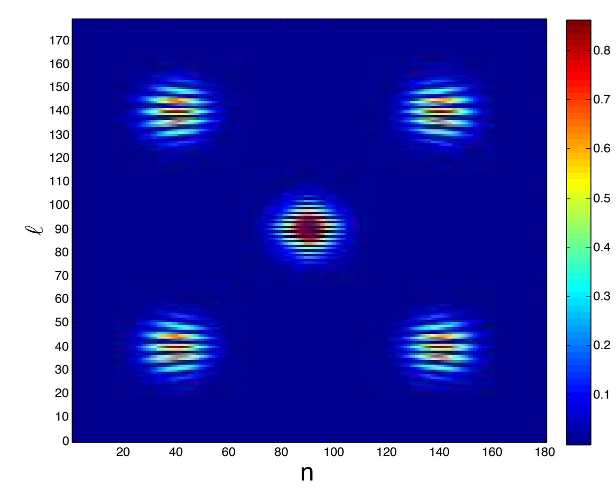
- To sparsely represent large classes of signals, it can be desirable for dictionary atoms to be jointly localized in vertex (time) and graph spectral (frequency) domains
- For signals on the real line, the Heisenberg uncertainty principle characterizes the tradeoff in resolution between the two domains

Signals on the Real Line

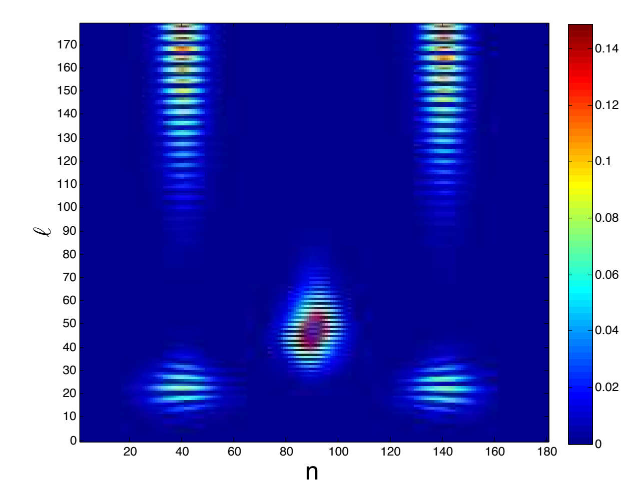


Source: Vetterli and Kovačević, 1995

Graph Signals on the Path Graph



WGFT

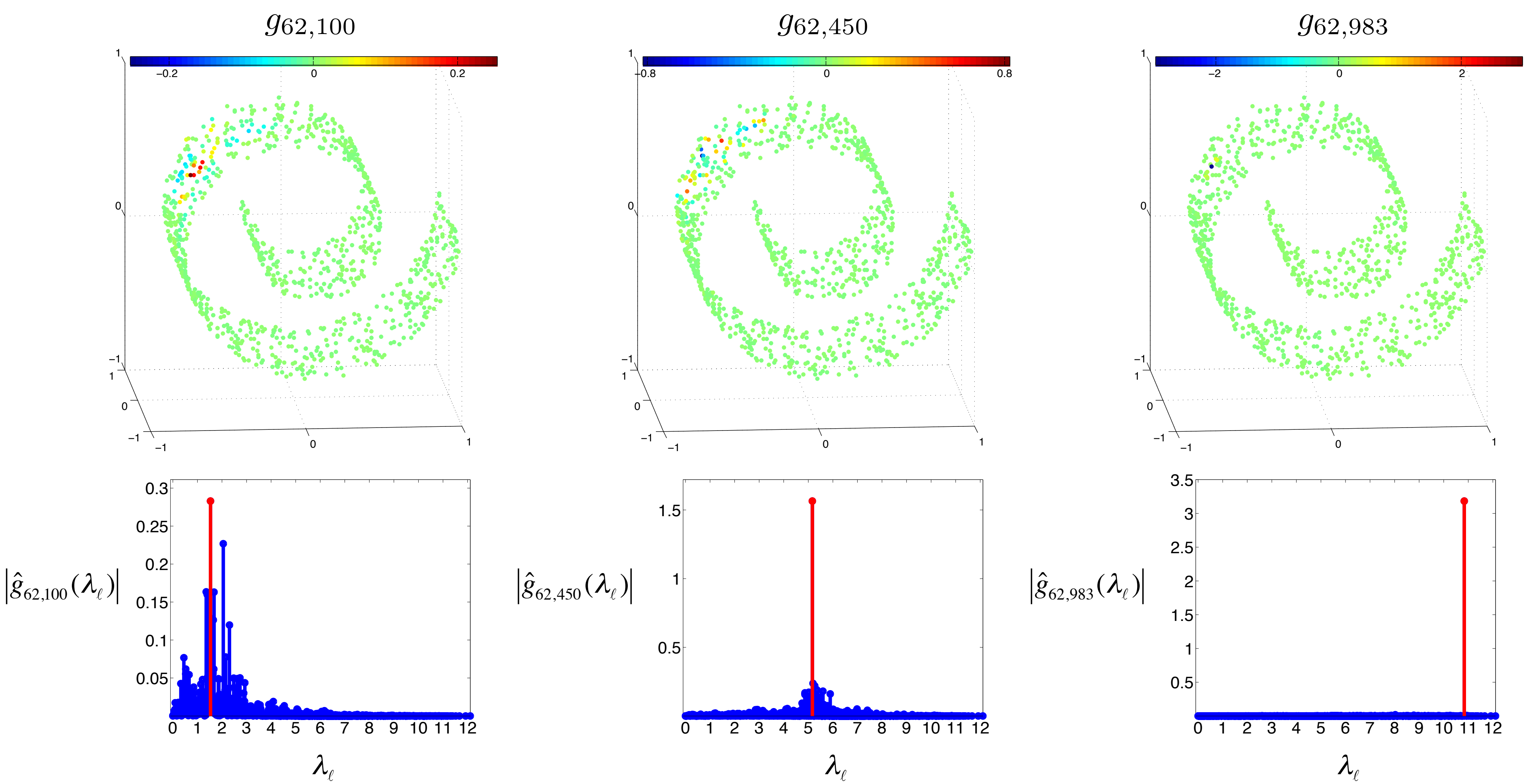


SGWT

4. Vertex-Frequency Tiling (cont.)

- Unlike the complex exponentials, the graph Laplacian eigenvectors can be localized (highly concentrated on a small region of the graph) Saito and Woei, RIMS Kokyuoku, 2011

- As a result, some graph signals may be simultaneously localized in both the vertex and graph spectral domains



4. Vertex-Frequency Tiling: Open Questions

- New uncertainty principles?
  - Different ways to measure spreads in the two domains?
  - Uncertainty principles can be used to show unexpected things are *possible*
  - Example: partial, noisy observation of a bandlimited signal recoverable because a bandlimited signal cannot be concentrated on missing values (provided few enough values are missing and/or bandlimit is low enough)

Donoho and Stark, Uncertainty principles and signal recovery, 1989

Theoretical results characterizing fundamental limits of graph signals such as uncertainty principles inform dictionary design

- How are structural properties of weighted graphs theoretically related to the (non-)localization of the graph Laplacian eigenvectors?

Agaskar and Lu, A spectral graph uncertainty principle, T. Info. Theory, 2013

Pasdeloup et al., Toward an uncertainty principle for weighted graphs, 2015

Tsitsvero et al., Signals on graphs: Uncertainty principle and sampling, 2015

Perraudin et al., Global and local uncertainty principles, 2016

5. Scalable/Distributed Implementations

- Routines that avoid full eigendecompositions
  - e.g., polynomial approximations for graph spectral filtering
  - fast graph Fourier transforms?
- Reduce storage and communication requirements in distributed settings
- Leverage numerical linear algebra literature / form collaborations with researchers from that area
- Connections with solving symmetric, diagonally-dominant systems of equations

Spielman, http://www.cs.yale.edu/homes/spielman/precon/precon.html
Saad, Iterative methods for sparse linear systems, 2003
Livne and Brandt, Lean algebraic multigrid: Fast graph Laplacian linear solver, 2012
Vishnoi, Lx=b Laplacian solvers and their algorithmic applications, 2013

6. Graph Construction and Choice of Graph Fourier Basis

- Different choices of graph construction (choosing edges and weights, directed/undirected)
- Different notions of distance (geodesic/shortest path, resistance, diffusion, algebraic)
- Different choices of graph Fourier basis
- Recent flurry of work on graph topology identification/learning

Summary

- Weighted graphs are a flexible tool to represent a wide variety of topologically-complicated data domains
- To identify and exploit structure in the data, we need to design dictionaries that incorporate the intrinsic geometric structure of the underlying data domain
- Try to leverage intuition from computational harmonic analysis of signals on Euclidean domains
  - Some ideas generalize relatively straightforwardly (e.g., notion of frequency)
  - However, signals and transforms on graphs can have surprising properties due to the irregularity of the data domains (e.g., uncertainty principle)
- Field is emerging / recently emerged
  - Requires more connections/iterations between dictionary design, theory, algorithms, and applications
  - Application of these techniques to real science and engineering problems is in its infancy

Explore

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