

# Localized Spectral Graph Filter Frames

## *A Unifying Framework, Survey of Design Considerations, and Numerical Comparison (Extended Cut)*

David I Shuman

Macalester College, Department of Mathematics, Statistics, and Computer Science

dshuman1@macalester.edu

**Abstract**—Representing data residing on a graph as a linear combination of building block signals can enable efficient and insightful visual or statistical analysis of the data, and such representations prove useful as regularizers in signal processing and machine learning tasks. Designing collections of building block signals – or more formally, dictionaries of atoms – that specifically account for the underlying graph structure as well as any available representative training signals has been an active area of research over the last decade. In this article, we survey a particular class of dictionaries called localized spectral graph filter frames, whose atoms are created by localizing spectral patterns to different regions of the graph. After showing how this class encompasses a variety of approaches from spectral graph wavelets to graph filter banks, we focus on the two main questions of how to design the spectral filters and how to select the center vertices to which the patterns are localized. Throughout, we emphasize computationally efficient methods that ensure the resulting transforms and their inverses can be applied to data residing on large, sparse graphs. We demonstrate how this class of transform methods can be used in signal processing tasks such as denoising and non-linear approximation, and provide code for readers to experiment with these methods in new application domains.

### I. INTRODUCTION: DICTIONARIES OF GRAPH SIGNALS

A major line of work in graph signal processing [1], [2] over the past ten years has been to design new transform methods that account for the underlying graph structure in order to identify and exploit structure in data residing on a connected, weighted, undirected graph. The most common approach is to construct a dictionary of atoms (building block signals), and represent the graph signal of interest as a linear combination of these atoms. Such representations enable visual analysis of data, statistical analysis of data, and data compression, and can also be leveraged as regularizers in machine learning and ill-posed inverse problems such as inpainting, denoising, and classification.

In general, desirable properties when designing dictionaries for graph signals include: (i) the atoms have an interpretable form that accounts for the underlying graph structure, so that the inner products between a graph signal and each atom are informative; (ii) the dictionary comprises an orthonormal basis or tight frame for the signal space, so that the contribution of each atom can be computed via an inner product with the graph signal, and the energy of the graph signal is equal to a constant multiple of the energy of the transform coefficients;

(iii) it is numerically efficient to apply the dictionary analysis and synthesis operators (forward and inverse transforms); and (iv) signals of certain mathematical classes can be represented exactly or approximately as *sparse* linear combinations of a subset of the dictionary atoms.

By our count, approximately 100 conference and journal articles written in the last decade have introduced new dictionaries for graph signals. These include designs for analytic dictionaries that are adapted to the graph structure but not any specific training data, as well as techniques for learning dictionaries from training data. Some broad classes of dictionaries include graph Fourier transforms; windowed graph Fourier transforms (e.g., [3]); vertex domain designs including spatial wavelets (e.g., [4], [5]), hierarchical trees (e.g., [6]), lifting transforms (e.g., [7]), and top-down approaches (e.g., [8], [9]); diffusion-based designs (e.g., [10]); spectral domain designs (e.g., [11]); pyramid transforms (e.g., [12]); and generalized filter banks (e.g., [13]–[16]). Despite, or perhaps because of, the number of new dictionary designs for graph signals, it remains difficult to identify which dictionary might be best suited for a specific task, or to understand subtle qualitative tradeoffs when specifying the parameters of a given dictionary construction.

In this survey, we restrict our attention to *localized spectral graph filter frames*, whose atoms are created by localizing patterns (spectral filters) to different regions of the graph. The seminal example of a dictionary of such atoms is spectral graph wavelets [11]. However, localized spectral graph filter frames are *broad in their scope*, including more recently proposed methods such as single-level filter banks for graph signals [14]–[25], variational or interpolating splines [26], [27], frames adapted to training data [28], [29], frame constructions for general graph signals [30]–[33], frame constructions tailored to specific applications such as fMRI data analysis [34]–[36] or community mining [37], “natural” wavelets [38], and even some vertex domain constructions [4], [5].<sup>1</sup> Our motivations for examining these dictionaries include (i) the design framework is flexible - it can yield highly redundant dictionaries to sparsely represent graph signals or new bases to efficiently extract structure from data on graphs, and it can also incorporate representative training signals when they are

<sup>1</sup>While diffusion wavelets [10] likely inspired many of these dictionaries, diffusion wavelets do not technically fit into the localized spectral graph filter dictionary framework outlined in Sec. II due to the additional step of numerically orthogonalizing the atoms.

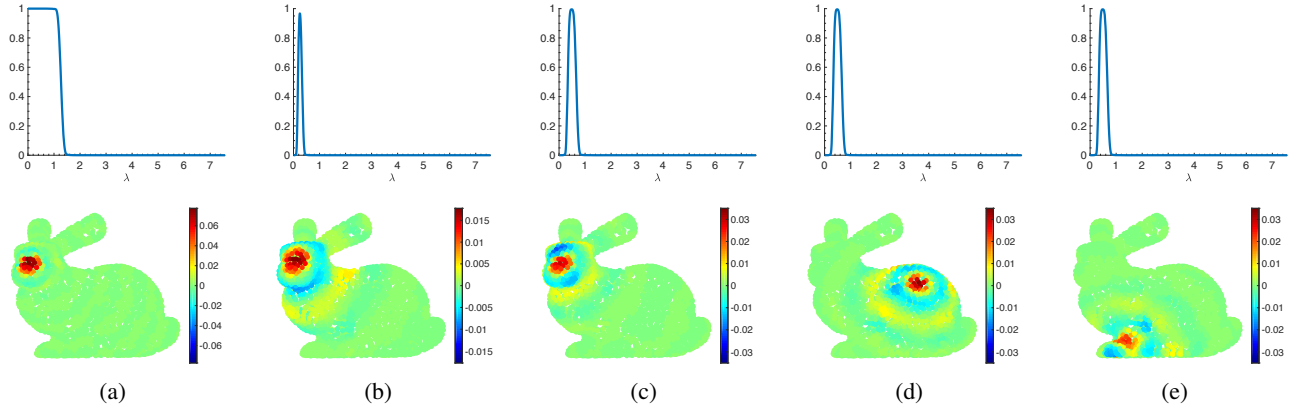


Fig. 1. Localized spectral graph filter frame atoms. (a)-(c) Three different filters/patterns localized to the same center vertex. (c)-(e) The same filter/pattern localized to three different center vertices.

available; (ii) the atoms have a physically interpretable structure and their closed form definition opens the door to formal mathematical analysis; and (iii) fast numerical approximations exist to efficiently apply these dictionary transforms and their inverses to data residing on large, sparse graphs, which are increasingly common in signal processing and machine learning applications. Due to the multiscale and localized structure of their atoms, these dictionaries are particularly relevant for applications where interesting phenomena are expressed in discontinuities or quick changes in signal values in smaller regions of the graph, analogous to edges in images.

The organization of the article is as follows. In the next section, we detail a unifying framework for localized spectral graph filter dictionaries. We survey the key design considerations for this class of dictionaries in Section III (design of the spectral filters) and Section IV (selection of the center vertices). In Section V, we review recent work on theoretical considerations and metrics that inform the design of these dictionaries. In Sections VI and VII, we pose specific high-level questions that get to the heart of the “where do we start when specifying a dictionary?” issue, attempt to answer these questions with new numerical comparisons and qualitative insights, and discuss how these comparisons inform future work in the area.

## II. DICTIONARIES OF LOCALIZED SPECTRAL PATTERNS

Keeping with the notation of [1], we consider data residing on a connected, weighted, undirected graph  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$  characterized by a finite set of vertices  $\mathcal{V}$  with  $|\mathcal{V}| = N$ , a set of edges  $\mathcal{E}$ , and a weighted adjacency matrix  $\mathbf{W}$ . A signal or function  $f: \mathcal{V} \rightarrow \mathbb{R}$  defined on the vertices of the graph may be represented as a vector  $\mathbf{f} \in \mathbb{R}^N$ , where the  $i$ th element of the vector  $\mathbf{f}$  represents the graph signal value at vertex  $i$  in  $\mathcal{V}$ .

The dictionaries we consider feature atoms of the form

$$\varphi_{i,j} := T_i g_j := \hat{g}_j(\mathcal{L}) \delta_i = \mathbf{U} \hat{g}_j(\mathbf{\Lambda}) \mathbf{U}^* \delta_i. \quad (1)$$

In (1),  $\delta_i$  is a graph signal with a value of 1 at vertex  $i$  and 0 elsewhere,  $\mathcal{L} = \mathbf{D} - \mathbf{W} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^*$  is the (combinatorial) graph Laplacian, the columns of  $\mathbf{U}$  are the orthogonal eigenvectors of  $\mathcal{L}$ , the  $*$  symbol denotes conjugate transpose, and  $\mathbf{\Lambda}$  is a diagonal matrix whose  $\ell$ th diagonal element  $\lambda_\ell$  is the eigenvalue

of  $\mathcal{L}$  associated with the eigenvector  $\mathbf{u}_\ell$ , the  $\ell$ th column of  $\mathbf{U}$ .<sup>2</sup> Each *spectral graph filter* or *kernel*  $\hat{g}_j(\cdot)$  is a function from  $\sigma(\mathcal{L}) = \{\lambda_0, \lambda_1, \dots, \lambda_{N-1}\}$ , the set of Laplacian eigenvalues, to the real numbers. Thus,  $\hat{g}_j(\mathbf{\Lambda})$  is a diagonal matrix with the  $\ell$ th diagonal entry equal to  $\hat{g}_j(\lambda_\ell)$ . In practice, these filter functions are often defined on the continuous range  $[0, \bar{\lambda}]$ , where  $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{N-1} = \lambda_{\max} \leq \bar{\lambda}$ .

At a more intuitive level, we can think of each spectral graph filter  $\hat{g}_j(\cdot)$  as defining a spectral pattern that is localized to different regions of the graph, with vertex  $i$  being the center of the localized pattern  $\varphi_{i,j} = T_i g_j$ . As an extreme example, if the spectral pattern is  $\hat{g}(\lambda_\ell) = 1$  for all  $\lambda_\ell$ , then the localized pattern centered at vertex  $i$  is  $T_i g = \mathbf{U} \mathbf{U}^* \delta_i = \delta_i$ . Fig. 1 displays more localized spectral patterns on the Stanford bunny graph [39].

We refer to a collection of atoms of the form (1) as a *localized spectral graph filter dictionary* (LSGFD), denoted by

$$\mathcal{D} = \{T_i g_j\}_{j=1,2,\dots,J; i \in \mathcal{V}_j}. \quad (2)$$

In (2),  $\mathcal{V}_j \subseteq \mathcal{V}$  is the set of center vertices to which the  $j$ th spectral pattern  $\hat{g}_j(\cdot)$  is localized, and each atom  $\varphi_{i,j} = T_i g_j$  is a graph signal in  $\mathbb{R}^N$ . Therefore, to fully specify a LSGFD  $\mathcal{D}$ , we need to answer the following questions, which are the focus of the next two sections, respectively:

- 1) How many spectral patterns should we use, and what should those patterns be? That is, we must specify the number of filters,  $J$ , and the form of the filters  $\{\hat{g}_1(\cdot), \hat{g}_2(\cdot), \dots, \hat{g}_J(\cdot)\}$ .
- 2) For each spectral pattern  $\hat{g}_j(\cdot)$ , how many center vertices should the pattern be localized to, and which vertices should those be? That is, we must specify the sets  $\mathcal{V}_j$  for each  $j$ .

In specifying the spectral patterns and sets of center vertices for LSGFDs, it is also important to keep in mind (i) what information is available, and (ii) the graph size. In all cases

<sup>2</sup>While we use the combinatorial (non-normalized) graph Laplacian  $\mathcal{L}$  throughout, the ideas we discuss apply to dictionaries comprised of atoms of the form (1) with the graph Fourier basis  $\mathbf{U}$  chosen as the eigenvectors of other symmetric generalized graph Laplacian operators such as the normalized graph Laplacian  $\mathcal{L}_{\text{norm}} = \mathbf{D}^{-\frac{1}{2}} \mathcal{L} \mathbf{D}^{-\frac{1}{2}}$ .

in this survey, we assume the underlying graph structure  $\mathcal{G}$  is known, although learning graph structures is a vibrant area of ongoing research (see, e.g., [40], [41] and references therein). In terms of data available in the design of the dictionary, there are three possibilities: (i) no data is available (the default unless otherwise specified), (ii) the design of the dictionary atoms may also take into account a set of one or more specific graph signals that are to be analyzed by the dictionary (we refer to the resulting dictionaries as *signal-adapted*), and (iii) a set of training data is available to learn parameters of the LSGFD, but the dictionary is then used to analyze a different set of (presumably similar) graph signals.

For small to medium sized graphs (say on the order of 10,000 or fewer vertices), the full Laplacian eigendecomposition  $\mathcal{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$  can be computed, and therefore the exact Laplacian eigenvectors and eigenvalues can be used in the dictionary design. For larger graphs, however, it may not be tractable to perform this decomposition, and we therefore put an emphasis in the next two sections on methods that do not require these quantities. Without the Laplacian eigenvectors, we almost always need an estimate of the maximum eigenvalue  $\lambda_{\max}$  via, e.g., a few steps of the Lanczos algorithm [42] or a closed form upper bound on it. For example,  $\bar{\lambda}$  can be taken to be the maximum sum of the degrees of any two vertices connected by an edge,  $\lambda_{\max} \leq \max_{\{(m,n) \in \mathcal{E}\}} \{d(m) + d(n)\}$ , where  $d(n)$  is the degree of vertex  $n$  [43], [44, Cor. 3.2].

In addition to obtaining a fast estimate for the spectral range  $[0, \lambda_{\max}]$ , it is often also beneficial to estimate the distribution of the Laplacian eigenvalues over the spectral range. Specifically, the *cumulative spectral density function* or *empirical spectral cumulative distribution* of  $\mathcal{L}$ , defined as

$$P_\lambda(z) := \frac{1}{N} \sum_{\ell=0}^{N-1} \mathbb{1}_{\{\lambda_\ell \leq z\}}, \quad (3)$$

can be efficiently estimated by different methods [45]. We use a variant of the kernel polynomial method [46] detailed in [16, Alg. 2] that leverages Hutchinson's stochastic trace estimator to estimate the number of eigenvalues below linearly spaced points between 0 and  $\lambda_{\max}$ , and then interpolates these values via monotonic piecewise cubic interpolation to generate an estimate of the cumulative spectral density function (3). The computational cost is proportional to the number of edges in the graph. As an order of magnitude example, for a sparse graph with more than 469,000 vertices, estimates for the maximum eigenvalue and the density function can be computed on a laptop in approximately 1 second and 16 seconds, respectively. In Fig. 2, we show examples of exact and approximate cumulative spectral density functions on six different graphs. In summary, while the full Laplacian eigendecomposition is necessary to exactly compute the atoms in (1) and their inner products with a graph signal, the spectral range and density function can be computed inexpensively and leveraged in the design of the filters, the selection of the center vertices, and the *approximate* computation of the inner products between the graph signal and each dictionary atom. We discuss these details further in the next two sections.

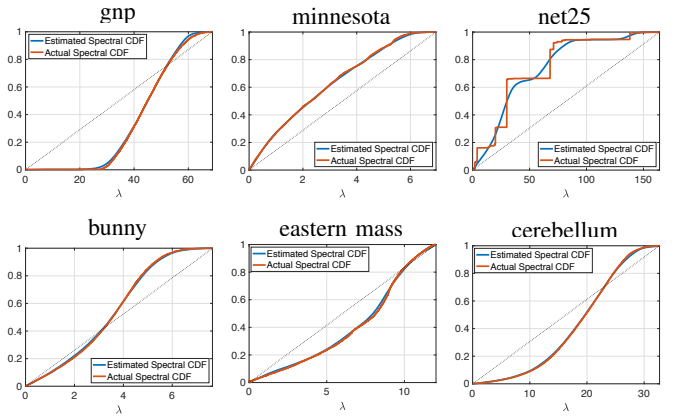


Fig. 2. Estimated and actual cumulative spectral density functions (3) for six graph Laplacians: a random Erdős-Renyi graph with  $N = 500$  vertices and edge probability 0.2; the Minnesota traffic network ( $N = 2642$ ) [47]; the Laplacian of the Andrianov net25 matrix ( $N = 9520$ ) from the SuiteSparse Matrix Collection [48]; the Stanford bunny graph ( $N = 2503$ ) [39]; an 8-neighbor local graph for Eastern Massachusetts ( $N = 877$ ) [16]; and a graph of the cerebellum region of the brain ( $N = 4465$ ) [36].

We represent the *synthesis operator* with the matrix  $\Phi \in \mathbb{R}^{N \times M}$ , where the columns of  $\Phi$  are the  $M = \sum_{j=1}^J |\mathcal{V}_j|$  dictionary atoms in  $\mathcal{D}$ . We refer to its adjoint  $\Phi^*$  as the *analysis operator*; this conjugate transpose matrix maps a graph signal to the analysis coefficients  $\{\langle \mathbf{f}, \varphi_{i,j} \rangle\}$ . If the dictionary  $\mathcal{D}$  satisfies the *frame condition* for all vectors  $\mathbf{f}$  in some subspace  $\mathcal{S}$  of  $\mathbb{R}^N$  (or all of  $\mathbb{R}^N$ ),

$$A \|\mathbf{f}\|_2^2 \leq \|\Phi^* \mathbf{f}\|_2^2 = \sum_{j=1}^J \sum_{i \in \mathcal{V}_j} |\langle \mathbf{f}, \varphi_{i,j} \rangle|^2 \leq B \|\mathbf{f}\|_2^2, \quad (4)$$

then any graph signal in the subspace can be exactly recovered from its analysis coefficients  $\Phi^* \mathbf{f}$ . Moreover, if  $A = B$  in (4), the dictionary is said to be a *tight frame*, and

$$\mathbf{f} = \frac{1}{A} \sum_{j=1}^J \sum_{i \in \mathcal{V}_j} \langle \mathbf{f}, \varphi_{i,j} \rangle \varphi_{i,j} = \frac{1}{A} \Phi \Phi^* \mathbf{f}.$$

A tight frame with frame bounds  $A = B = 1$  is called a *Parseval frame*, and has the added benefit that  $\|\Phi^* \mathbf{f}\|_2 = \|\mathbf{f}\|_2$ ; i.e., the energy of the analysis coefficients is the same as the energy of the graph signal. For more properties of frames, see [49], [50].

Finally, we mention the connection between the aforementioned analysis coefficients and graph spectral filter banks. As shown in Fig. 3, in a  $J$ -channel graph filter bank (e.g., [13]–[16]),  $J$  different filters are applied to the signal, and the values of  $\hat{g}_j(\mathcal{L})\mathbf{f}$ , the filtered signal in the  $j$ th channel, at a specified set of downsampled vertices  $\mathcal{V}_j$  are stored. The set of analysis coefficients  $\{\langle \mathbf{f}, \varphi_{i,j} \rangle\}_{i \in \mathcal{V}_j}$  derived from the atoms generated by localizing the filter  $\hat{g}_j(\cdot)$  to each of the center vertices in  $\mathcal{V}_j$  corresponds exactly to the downsampled values in the  $j$ th channel of the filter bank.

### III. DESIGN OF THE SPECTRAL FILTERS

There are three broad classes of spectral filter designs: (i) those adapted only to the spectral range  $[0, \lambda_{\max}]$  (e.g., [11],

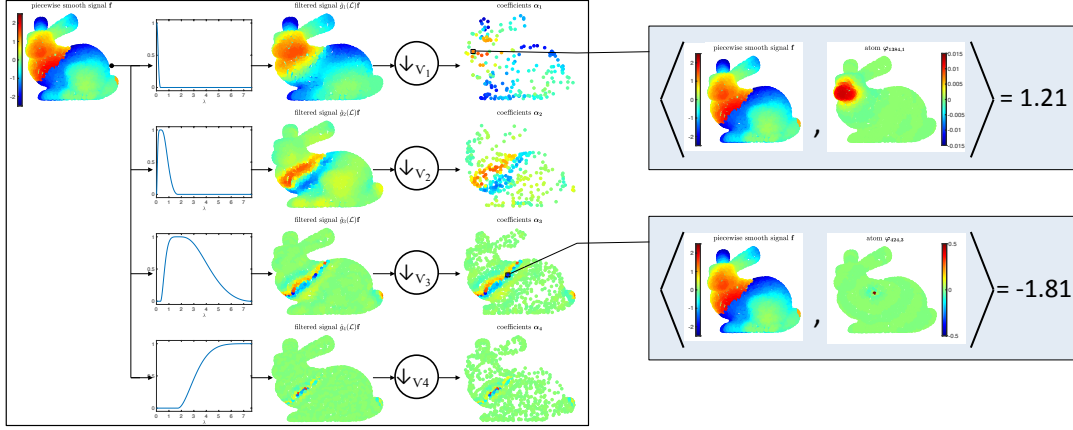


Fig. 3. Equivalence between single-level graph spectral filter banks and localized spectral graph filter dictionaries. In the 4-channel filter bank on the left, the graph signal is filtered by each of the  $J = 4$  spectral filters and then downsampled on the corresponding vertex sets  $\{\mathcal{V}_j\}$  to yield the filter bank coefficients  $\{\alpha_j\}$ . Each coefficient  $\alpha_{i,j} = [\hat{g}_j(\mathcal{L})\mathbf{f}](i) = \delta_i^* \hat{g}_j(\mathcal{L})\mathbf{f} = \mathbf{f}^* \hat{g}_j(\mathcal{L})\delta_i$  in the final column of the left box corresponds exactly to the inner product between the graph signal and the dictionary atom  $\varphi_{i,j}$ , as  $\langle \mathbf{f}, \varphi_{i,j} \rangle = \langle \mathbf{f}, T_i g_j \rangle = \langle \mathbf{f}, \hat{g}_j(\mathcal{L})\delta_i \rangle = \mathbf{f}^* \hat{g}_j(\mathcal{L})\delta_i$ .

[14], [18], [20], [23], [30], [32], [33]]<sup>3</sup>; (ii) those adapted to an estimate of the cumulative spectral density function of the graph (e.g., [16], [31]); and (iii) those adapted to both the graph and training signals residing on the graph (e.g., [28], [29]). Examples of all three of these classes are shown in Examples 2-4. In this section, we survey a number of considerations when designing the filters for localized spectral graph filter dictionaries. All but the final of the following considerations apply to all three of the aforementioned classes of filter designs.

#### A. Localization of the atoms in the vertex domain

Whether the underlying graph represents a traffic network, a social network, a biological network, or some other type of network, interesting phenomena in the data often occur at a *local scale*, particularly for extremely large graphs. To find or make inferences about such localized patterns, it can be helpful to have dictionary atoms whose energies are concentrated in smaller regions of the graph. One method to guarantee that each atom's energy is strictly localized in a small neighborhood of radius  $K$  around its center vertex  $i$  is to choose the spectral filters to be degree  $K$  polynomials.

*Theorem 1 (Lemma 5.2 of [11], Lemma 2 of [3]):* Let  $d_G(i, n)$  be the *geodesic* or *shortest path* distance between vertices  $i$  and  $n$ ; i.e., the minimum number of edges in any path connecting the two vertices. Let  $\widehat{p}_K$  be a polynomial kernel with degree  $K$ ; i.e.,

$$\widehat{p}_K(\lambda) = \sum_{k=0}^K a_k \lambda^k \quad (5)$$

for some coefficients  $\{a_k\}_{k=0,1,\dots,K}$ . If  $d_G(i, n) > K$ , then  $(T_i \widehat{p}_K)(n) = 0$ .

More generally, the localization of the dictionary atoms in the vertex (spatial) domain is closely related to the smoothness of the filters. There are multiple ways to measure both

“localization” and “smoothness,” but one is to examine how the magnitude of the localized pattern at vertex  $n$  decays as the distance between  $n$  and the center vertex  $i$  increases, depending on how close the filter  $\hat{g}_j(\cdot)$  is to a degree  $K$  polynomial.

*Theorem 2 (Theorem 1 of [3], Theorem 5.16 of [51], Theorem 8.2 of [52]):* Let  $\hat{g}_j : [0, \lambda_{\max}] \rightarrow \mathbb{R}$  be a spectral filter and define  $K_{in} := d_G(i, n) - 1$ . Then

$$\begin{aligned} |(T_i g_j)(n)| &\leq \inf_{\widehat{p}_{K_{in}}} \left\{ \sup_{\lambda \in \sigma(\mathcal{L})} |\hat{g}_j(\lambda) - \widehat{p}_{K_{in}}(\lambda)| \right\} \\ &\leq \inf_{\widehat{p}_{K_{in}}} \left\{ \sup_{\lambda \in [0, \lambda_{\max}]} |\hat{g}_j(\lambda) - \widehat{p}_{K_{in}}(\lambda)| \right\}, \quad (6) \end{aligned}$$

where the infimum in (6) is taken over all polynomial kernels of degree  $K_{in}$ , as defined in (5). If  $\hat{g}_j(\cdot)$  is real analytic on  $[0, \lambda_{\max}]$ , the upper bound in (6) converges geometrically to 0 as  $d_G(i, n)$  increases.

In short, and less precisely mathematically, the smoother the filter  $\hat{g}_j(\cdot)$  is in the spectral domain, the more concentrated is the energy of the atom  $\varphi_{i,j} = T_i g_j$  around the center vertex  $i$ ; compare, e.g., the first two atoms shown in Fig. 1.

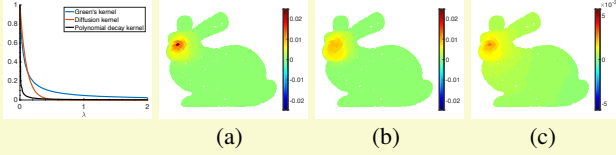
#### B. Eigenvector groupings

Recall from the introduction that in order for the inner products between a graph signal and each atom to be informative, the atoms should have interpretable structural features that account for the underlying graph. The localization in the vertex domain described above is one such structural feature. The shape of the filter  $\hat{g}(\cdot)$  in the graph spectral domain leads to another: smoothness in terms of how much the atom's values vary between neighboring vertices, particularly those connected by a high edge weight. The unit-norm Laplacian eigenvectors satisfy

$$\lambda_\ell = \mathbf{u}_\ell^* \mathcal{L} \mathbf{u}_\ell = \sum_{(m,n) \in \mathcal{E}} W_{m,n} [u_\ell(m) - u_\ell(n)]^2, \quad (7)$$

<sup>3</sup>The tight wavelet frames of [30] are also adapted to the maximum degree of the graph.

*Example 1 (Variational/interpolating splines):* Variational or interpolating splines on graphs, pioneered in [26] and further studied in [12], [27], [53], are atoms of the form (1) with a single low pass filter  $\hat{g}(\cdot)$  localized to a subset  $\mathcal{V}_1$  of the vertices. They are used as an interpolation basis to interpolate an entire graph signal from its sample values at the vertices in  $\mathcal{V}_1$ . In the images below, we show three examples of interpolating kernels and an atom generated from each on the Stanford bunny graph. The three filters are (a) a Green's kernel [26]  $\hat{g}(\lambda_\ell) = \frac{\epsilon}{(\lambda_\ell + \epsilon)^s}$  with  $\epsilon = .05$  and  $s = 1$ ; (b) a diffusion kernel [54]  $\hat{g}(\lambda_\ell) = e^{-\tau\lambda_\ell}$  with  $\tau = 10$ ; and (c) a polynomial decay kernel [27]  $\hat{g}(\lambda_\ell) = \frac{1}{(\ell+1)^s}$  with  $s = 1$ .



and therefore the eigenvectors associated with the lower eigenvalues vary less from vertex to neighboring vertex. Moreover, the eigenvectors are roughly ordered in terms of the number of zero crossings, defined as edges where the values of the eigenvector at the two connected vertices have opposite signs [1, Fig. 3].

Based off the analogy between this smoothness of Laplacian eigenvectors and the frequency of complex exponentials in one-dimensional signal processing, the most common spectral design approach in the graph signal processing literature is to choose filters concentrated on one part of the graph spectrum, grouping together eigenvectors with similar levels of total variation with respect to the graph, as defined in (7). In particular, when localized to different center vertices via (1), filters whose support is concentrated on the eigenvectors associated with small eigenvalues lead to *scaling functions* or *windows* around the center vertex (c.f., Fig. 1(a) and Example 1). The inner products between such atoms and a graph signal provide information about the trend or local average of the signal in the neighborhood of the center vertex. On the other hand, all Laplacian eigenvectors associated with eigenvalues greater than 0 sum to zero, because they are orthogonal to  $\mathbf{u}_0$ , which is constant across all vertices. Thus, any filter with  $\hat{g}_j(0) = 0$  yields atoms  $\varphi_{i,j}$  that have a mean of zero and feature some oscillation (c.f., Fig. 1(b)-(c)).

Noting that the Laplacian eigenvalues correspond to the total variation of the eigenvectors but not necessarily the directions of their oscillations on the graph (see Fig. 4), some more recent works [38], [55]–[57] investigate other ways to group the eigenvectors. For example, [38] suggests to view the eigenvectors as probability distributions on the graph, quantify the distances between eigenvectors using optimal transport theory, cluster the eigenvectors based on their distances, and construct a spectral filter for each cluster, with the support of the filter set to match the eigenvectors of that cluster.

### C. Orthogonality or near orthogonality

To reduce the correlation between atoms (and in turn improve the discriminatory power of taking inner products between each atom and a target signal, as discussed below in Sec. V), it may be desirable for all atoms that are generated from a single filter to be orthogonal or near orthogonal to all atoms that are generated from all other filters; i.e.,  $\langle T_i g_j, T_{i'} g_{j'} \rangle \approx 0$  for all  $j' \neq j$ . This can be ensured via the filter design, with a sufficient condition for the orthogonality of atoms generated from different patterns being that  $\hat{g}_j(\lambda_\ell) \hat{g}_{j'}(\lambda_\ell) = 0$  for all  $j' \neq j$  and all  $\lambda_\ell$ .

For the specific case when  $J = 2$  and  $\mathcal{G}$  is a bipartite graph with the normalized Laplacian eigenvectors as the graph Fourier basis<sup>4</sup>, it is possible to go a step further and generate  $N$  atoms that are not only orthogonal to atoms generated from the other filter, but also from the same filter.

*Theorem 3 ([14]):* Let  $\mathcal{G}$  be a bipartite graph with a bipartition  $\{\mathcal{V}_1, \mathcal{V}_2\}$ , and consider an LGSFD  $\mathcal{D}$  of the form (2) with  $J = 2$  (i.e.,  $\hat{g}_i(\cdot)$  is localized to the center vertices in  $\mathcal{V}_i$  for  $i = 1, 2$ ), using the normalized Laplacian graph Fourier basis. Then necessary and sufficient conditions on the filters to ensure that the  $N$  atoms of  $\mathcal{D}$  form an orthogonal basis for  $\mathbb{R}^N$  are that  $\hat{g}_1(\lambda_\ell) \hat{g}_1(2 - \lambda_\ell) = \hat{g}_2(\lambda_\ell) \hat{g}_2(2 - \lambda_\ell)$  and  $|\hat{g}_1(\lambda_\ell)|^2 + |\hat{g}_2(\lambda_\ell)|^2 = c^2$  for all  $\lambda_\ell$  and any constant  $c$ .

### D. Coverage of the spectrum

A necessary condition for  $\mathcal{D}$  to be a frame for all graph signals in  $\mathbb{R}^N$  is that for all  $\lambda \in \sigma(\mathcal{L})$ ,  $\hat{g}_j(\lambda) \neq 0$  for some  $j \in \{1, 2, \dots, J\}$ . If this is not true for some  $\lambda_\ell$ , then  $\langle \mathbf{u}_\ell, \varphi_{i,j} \rangle = 0$  for all  $i$  and  $j$  and thus  $\|\Phi^* \mathbf{u}_\ell\|_2^2 = 0$ , contradicting the frame condition (4). Thus, choosing a set of filters that covers the full spectral range  $[0, \lambda_{\max}]$  (or at least the portion of it whose Laplacian eigenvectors span the subspace of signals of interest) is a good place to start.

In the case that  $\mathcal{V}_j = \mathcal{V}$  for all  $j$  (i.e., every spectral pattern is localized to every vertex), which is often referred to as

<sup>4</sup>These conditions can be adapted for a regular bipartite graph with the non-normalized Laplacian eigenvectors as the graph Fourier basis.

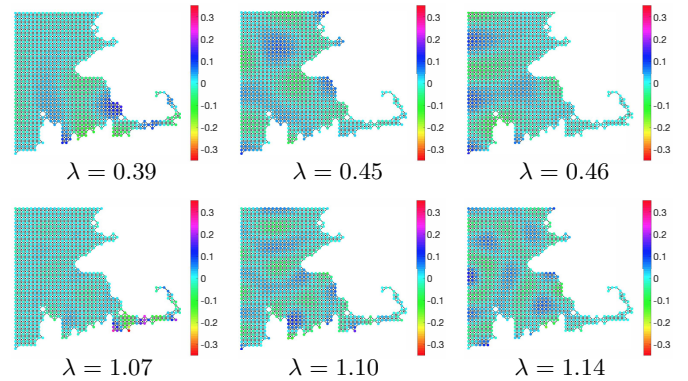


Fig. 4. Six Laplacian eigenvectors of the Eastern Massachusetts 8-neighbor graph, labeled by the corresponding eigenvalues. While the three in the top row and the three in the bottom row are more similar to each other in terms of total variation (captured by the eigenvalue), the directions of the oscillations and regions where the eigenvectors' energies are concentrated are not necessarily in linear order. For example, both  $\mathbf{u}_{0.39}$  and  $\mathbf{u}_{1.07}$  have more of their energies concentrated on the vertices corresponding to Cape Cod.

complete sampling or an undecimated filter bank, it is possible to design the spectral filters such that  $\mathcal{D}$  is a (tight) Parseval frame.

*Theorem 4 (Theorem 5.6 of [11], [30], and Lemma 1 of [31]):* Let  $\mathcal{V}_j = \mathcal{V}$  for all  $j$  and  $\mathcal{D} := \{\varphi_{i,j}\}_{i=1,2,\dots,N; j=1,2,\dots,J}$  be a dictionary of atoms with  $\varphi_{i,j} := T_i g_j$ . Define  $G(\lambda) := \sum_{j=1}^J |\hat{g}_j(\lambda)|^2$ . If  $G(\lambda) > 0$  for all  $\lambda \in \sigma(\mathcal{L})$ , then the frame condition (4) is satisfied for all  $\mathbf{f} \in \mathbb{R}^N$ , with frame bounds  $A = \min_{\lambda \in \sigma(\mathcal{L})} G(\lambda)$  and  $B = \max_{\lambda \in \sigma(\mathcal{L})} G(\lambda)$ . In particular, if  $G(\lambda)$  is constant on  $\sigma(\mathcal{L})$ ,  $\mathcal{D}$  is a *tight frame* with  $A = B$ . Moreover, if  $\sum_{j=1}^J |\hat{g}_j(\lambda)|^2 = 1, \forall \lambda \in \sigma(\mathcal{L})$ , then  $\mathcal{D}$  is a *Parseval frame*; i.e.,  $\|\Phi^* \mathbf{f}\|_2^2 = \sum_{j=1}^J \sum_{i=1}^N |\langle \mathbf{f}, \varphi_{i,j} \rangle|^2 = \|\mathbf{f}\|_2^2, \forall \mathbf{f} \in \mathbb{R}^N$ . Example 2 contains filters that satisfy the sufficient condition of Theorem 4 for a Parseval frame.

It is important to distinguish between coverage of the spectral range and coverage of the spectrum. One subtlety about Theorem 4 is that while the filters are often designed over the interval  $[0, \lambda_{\max}]$  or  $[0, \bar{\lambda}]$ , the condition for a tight frame is that  $G(\lambda)$  only needs to be constant on the set of actual Laplacian eigenvalues  $\sigma(\mathcal{L})$ , as these are the values of the filter that contribute to the definition of the atom in (1). Related to this point, if a filter is defined on the interval  $[0, \lambda_{\max}]$ , but  $\hat{g}_j(\lambda) = 0$  for all  $\lambda$  in  $\sigma(\mathcal{L})$ , then any atoms  $T_i g_j$  derived from this filter are equal to the zero vector, and therefore do not provide any useful information about the graph signal. One way to avoid such non-informative atoms is to adapt the filter design not only to the spectral width, but also to the estimated spectral density function (3). By leveraging the spectral density approximation, the *spectrum-adapted* designs in [31] warp a set of non-adapted filters in order that the support of each filter approximately contains a desired number of Laplacian eigenvalues (e.g., each filter has the same number of eigenvalues or they satisfy a dyadic structure with twice as many in each subsequent filter, moving from the low end of the spectrum to the high end). See Example 3 for examples of spectrum-adapted filter designs.

### E. Computational efficiency and approximation

As mentioned in Sec. II, exactly computing the graph Laplacian eigenvectors is only feasible for small to medium graphs, implying that for large graphs, the computation of the analysis coefficients  $\langle \mathbf{f}, \varphi_{i,j} \rangle = \delta_i^* \hat{g}_j(\mathcal{L}) \mathbf{f}$  must be efficiently approximated. Methods for approximating a *matrix function* times a vector (i.e.,  $\hat{g}_j(\mathcal{L}) \mathbf{f}$ ) include Krylov subspace methods such as the Lanczos method, contour integral methods, conjugate gradient, algebraic multigrid methods, rational approximations (also referred to as *infinite impulse response filters* in the graph signal processing community [59], [60]), spline approximations, and polynomial approximations (see [61]–[63] for surveys of these methods in centralized and distributed settings).

We focus our attention and numerical experiments in this survey on degree  $K$  polynomial approximations  $\hat{p}_{j,K}(\lambda)$  of

the form (5) to each filter  $\hat{g}_j(\cdot)$  (also referred to as *finite impulse response (FIR) filters* in the graph signal processing community). The approximation  $\mathbf{x}_j^{(K)} = \hat{p}_{j,K}(\mathcal{L}) \mathbf{f}$  to  $\hat{g}_j(\mathcal{L}) \mathbf{f}$  can be computed recursively, either through a three-term recurrence for specific types of polynomials (e.g., Chebyshev) or through the nested multiplication iteration [67, Section 9.2.4]

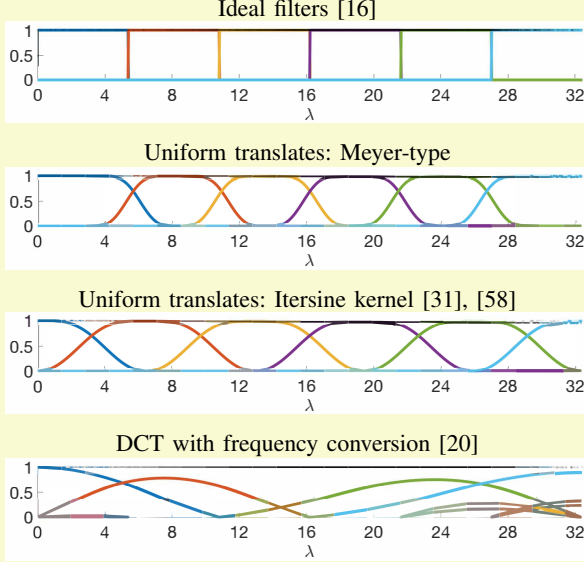
$$\mathbf{x}_j^{(l)} = a_{j,K-l} \mathbf{f} + \mathcal{L} \mathbf{x}_j^{(l-1)}, \quad l = 1, 2, \dots, K, \quad (8)$$

with  $\mathbf{x}_j^{(0)} = a_{j,K} \mathbf{f}$ . The computational complexity of computing  $\mathbf{x}_j^{(K)} = \hat{g}_j(\mathcal{L}) \mathbf{f}$  through (8) or a three-term recurrence is  $\mathcal{O}(K|\mathcal{E}|)$ , which for a large, sparse graph is approximately linear in the number of vertices, as opposed to the  $\mathcal{O}(N^3)$  required to naively compute the full eigendecomposition of  $\mathcal{L}$ . Additional advantages of the polynomial approximations include (i) the atoms are strictly localized as described in Theorem 1, (ii) in addition to the analysis operator, the synthesis operator  $\Phi$  can be applied efficiently, and (iii) both the analysis and synthesis computations can be performed in a distributed setting where each vertex only knows its own signal value and can only communicate with its neighboring vertices [63], [68]. Polynomial approximation methods commonly used in the graph signal processing literature include Chebyshev [11], [65] and Jackson-Chebyshev [66]. Example 4 shows filters resulting from these polynomial approximation methods. For the specific case of approximating an ideal low pass filter with a small degree polynomial, [69] introduces energy compaction filters that maximize the energy of the polynomial filter that is concentrated on the specified band.

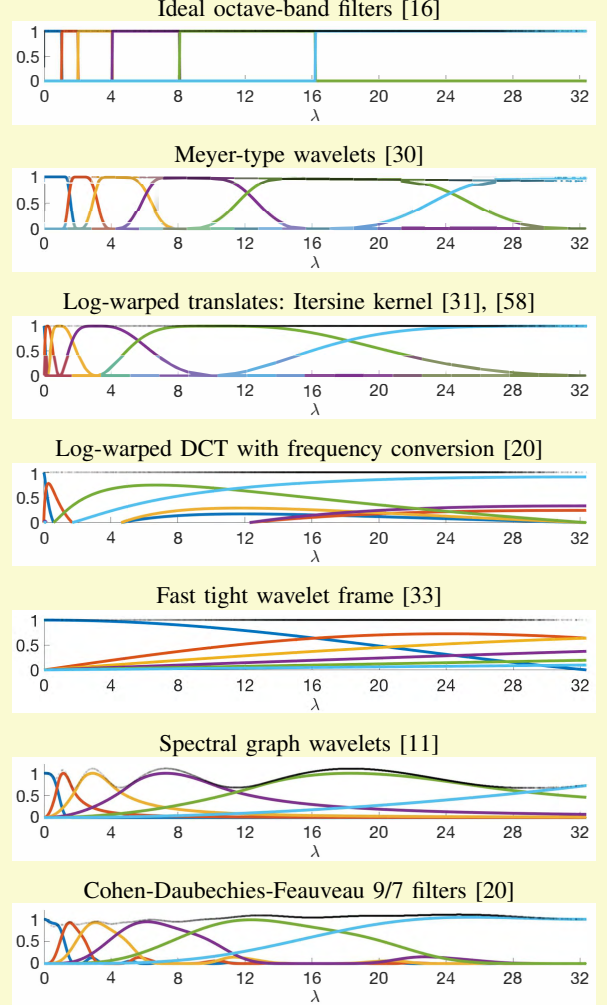
However, there are also tradeoffs to using polynomial filters. First, polynomial approximations to filter designs that meet the orthogonality or tight frame criteria may no longer satisfy these conditions. In fact, [23] shows that it is not possible to find  $J$  polynomial filters with the property that  $G(\lambda)$  is constant for all  $\lambda$  in the interval  $[0, \lambda_{\max}]$ . Although it may be possible to satisfy this condition for all  $\lambda$  in  $\sigma(\mathcal{L})$ , it is not usually tractable to do so. When using ideal filters, one option to mitigate the approximation error at the Laplacian eigenvalues (recall that the filter values at these eigenvalues are the only filter values that actually affect the form of the dictionary atoms) is to attempt to place the endpoints of the subband filters in areas of the spectrum with low density (or even better, in spectral gaps), as the error is typically highest near the endpoints (c.f., middle row of Example 4) [16]. A second option for mitigating the approximation error is to choose polynomials that control the error in specific parts of the spectrum, such as transformed linear phase multirate filters [23], which reduce the error near the eigenvalue 0 (no DC leakage) or spectrum-adapted polynomial approximation [70], which can reduce the error in high density areas of the spectrum. A third option is to directly choose the initial set of filters to be polynomials, or at least choose them to be smoother functions that are more accurately approximated by polynomials (e.g., [20]).

*Example 2 (Spectral filter designs that only use the spectral range):* We show 11 different sets of six filter patterns for the cerebellum graph [36], whose spectral range is  $[0, \lambda_{\max}] = [0, 32.4]$ . In all images, the vertical axis represents the value of the filter, and the shaded gray/black circles represent the values of  $G(\lambda) = \sum_{j=1}^J |\hat{g}_j(\lambda)|^2$  at each of the Laplacian eigenvalues, with darker areas denoting regions of higher spectral density. For each of the filters  $\hat{g}_j(\lambda)$  in the set of uniform translates, the corresponding filter in the log-warped set (right) is given by  $\hat{h}_j(\lambda) = \hat{g}_j\left(\frac{\lambda_{\max}}{\omega(\lambda_{\max})}\omega(\lambda)\right)$ , where the warping function is  $\omega(\lambda) = \log(1 + \nu\lambda)$  for a parameter  $\nu > 0$  ( $\nu = 10$  here).

### Uniform Translates



### Wavelets (Octave-Band)



Let us highlight some of the design considerations mentioned in this section:

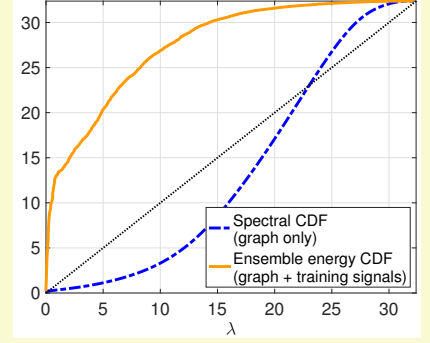
- 1) All of these designs cover the entire spectrum; i.e.,  $G(\lambda) > 0$  for all  $\lambda \in [0, \lambda_{\max}]$ . Thus, each design yields a frame when every filter is localized to be centered at every vertex.
- 2) The last two sets of wavelet filters are the only two amongst those shown that do not satisfy the Parseval frame condition,  $G(\lambda) = 1$  for all  $\lambda \in \sigma(\mathcal{L})$ , from Theorem 4.
- 3) Because the ideal filters in the top row do not overlap, the atoms generated from a filter are orthogonal to the atoms generated from any other filter. On the other hand, the atoms generated by localizing filters with significant overlap (e.g., the yellow and red filters of the fast tight wavelet frame) to the same center vertex are likely to be highly correlated.
- 4) As detailed in Theorem 2, filters that are smooth (well approximated by low order polynomials) yield dictionary atoms that are more localized around the center vertex; i.e., most of their energy is close to the center. In particular, for the octave-band designs such as the Meyer-type wavelets, log-warped translates, and spectral graph wavelets, the filters that cover the upper end of the spectrum yield more localized atoms than the scaling and wavelet filters at the low end of the spectrum.

### F. The number of filters

For sets of filters that cover the whole spectrum, typical choices of  $J$  in the literature are in the 4-8 range; however, we are not aware of theoretical analysis concerning the choice of  $J$ . In general, increasing  $J$  may lead to sparser representations by increasing the number of atoms and the ability to distinguish between signals by capturing the behavior of the signal across smaller spectral regions. On the other hand, at some point, the benefits may saturate as the atoms become more correlated (see, e.g., [29, Fig. 13]). Moreover,

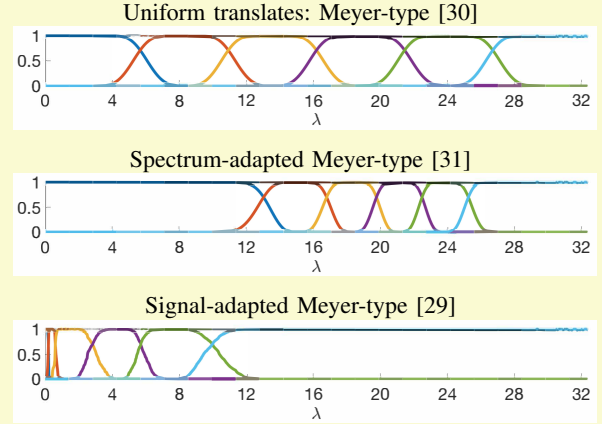
as the filters become narrower, they are more difficult to approximate by polynomials. If the dictionary is critically sampled ( $M = \sum_{j=1}^J |\mathcal{V}_j| = N$ ), increasing  $J$  also leads to fewer center vertices for each filter, making accurate reconstruction from the analysis coefficients more difficult. Narrower filters, especially those whose support is at the high end of the spectrum, can also lead to dictionary atoms that are more sensitive to small perturbations in the graph weights, an important consideration in applications where the graph is estimated. We explore the choice of  $J$  further in Section VI.

*Example 3 (Spectral filter designs that are adapted to the spectral density of the graph or ensemble energy density of training signals):* The ideal filters and uniform translates from Example 2 are designed to cover equal portions of the spectral range  $[0, \lambda_{\max}]$ , but are the same for any two graphs with the same value of  $\lambda_{\max}$ . One option to further adapt the filters to the specific graph structure is to warp them so that each filter contains roughly the same number of Laplacian eigenvalues [31]. This is accomplished by defining spectrum-adapted filters  $\tilde{g}_j(\lambda) = \hat{g}_j(\lambda_{\max} P_\lambda(\lambda))$ , where  $\{\hat{g}_j(\cdot)\}$  are the uniform translates from Example 2 and  $P_\lambda(\cdot)$  is an estimate of the cumulative spectral density defined in (3). For the cerebellum graph, the Laplacian eigenvalues are concentrated in the middle of the spectral range (right), and therefore the spectrum-adapted uniform Meyer-type filters shown in the middle row below are narrower in this region of



high spectral density. The same idea can be used to generate spectrum-adapted wavelets  $\tilde{h}_j(\lambda) = \hat{g}_j\left(\frac{\lambda_{\max}}{\omega(\lambda_{\max})}\omega(\lambda)\right)$ , by using a warping function  $\omega(\lambda) = \log(1 + \nu\lambda_{\max}P_\lambda(\lambda))$ .

When training data is available on the graph, a second option is to adapt the filters to be narrower in the regions of the spectrum where the energy of the training signals is concentrated [29]. This can again be achieved via warping, using an estimate of the ensemble energy cumulative spectral density function  $P_Y(z)$  defined in (9) in place of the spectral density estimate  $P_\lambda(z)$  in the warping function. The plot of the density functions (above right) shows that despite the Laplacian eigenvalues being more heavily concentrated in the middle of the spectrum, the energies of 292 fMRI training signals on the cerebellum graph are heavily concentrated in the lower end of the spectrum. Therefore, the signal-adapted design that aims to have roughly the same signal energy in each filter band (bottom row, right) features narrower filters at lower eigenvalues.



### G. Available data

The graph spectral filters can also be adapted to an ensemble of  $T$  training signals,  $\{\mathbf{y}_t\}_{t=1,2,\dots,T}$ , that are exemplary of the data to be analyzed by the dictionary  $\mathcal{D}$ , when such training signals are available. We briefly review two *parametric dictionary learning* approaches, both of which set  $\mathcal{V}_j = \mathcal{V}$  for all  $j$  and adapt the filters  $\{\hat{g}_j(\cdot)\}$  to the training data. The first approach, presented in [29] and shown in Example 3, is to design the filters so that (i) each filter captures a roughly equal amount of the energy of the ensemble of training signals, and (ii) the filters satisfy the tight frame condition  $G(\lambda) = 1$  for all  $\lambda \in [0, \lambda_{\max}]$ . Similar to the spectrum-adapted filter design of [31], this *signal-adapted filter design* constructs a set of prototype filters that uniformly cover the spectrum, and then warps/transforms the filters. Whereas the warping function in [31] is an estimate of the cumulative distribution of the Laplacian eigenvalues (3), the warping function in [29] estimates the *ensemble energy cumulative spectral density*

$$P_Y(z) := \frac{\sum_{\{\ell: 0 < \lambda_\ell \leq z\}} \frac{1}{T} \sum_{t=1}^T \left| \left\langle \frac{\mathbf{y}_t}{\|\mathbf{y}_t\|_2}, \mathbf{u}_\ell \right\rangle \right|^2}{\sum_{\ell=1}^{N-1} \frac{1}{T} \sum_{t=1}^T \left| \left\langle \frac{\mathbf{y}_t}{\|\mathbf{y}_t\|_2}, \mathbf{u}_\ell \right\rangle \right|^2}, \quad (9)$$

which can also be efficiently approximated [71]. An example where a signal-adapted spectral design is particularly beneficial is in functional magnetic resonance imaging (fMRI), where the energy of the data tends to be highly concentrated

at the low end of the spectrum of the cerebellum graph even though there are more eigenvalues at the upper end of the spectrum [29].

A second approach to incorporate the training data, presented in [28], is to force the  $J$  spectral filters to be polynomials, and through optimization, find polynomial coefficients that (i) lead to sparse representations of the training data, and (ii) yield filters that cover the spectrum so that the frame is close to being tight (i.e., the ratio of frame bounds  $\frac{B}{A}$  is close to 1). Specifically, [28] suggests to alternate between (a) a sparse approximation step that fixes the dictionary (i.e., fixes the polynomial filters) and uses orthogonal matching pursuit to find the coefficient matrix  $\mathbf{X}$  that minimizes  $\|\mathbf{Y} - \mathbf{\Phi}\mathbf{X}\|_F^2$  subject to  $\|\mathbf{x}_t\|_0 \leq K_0$  for all  $t$ , where the columns of  $\mathbf{Y}$  are the training signals,  $\mathbf{\Phi}$  is the current dictionary with normalized atoms, and  $K_0$  is a fixed sparsity level; and (b) a dictionary update step that fixes that coefficient matrix  $\mathbf{X}$  and updates the polynomial filter coefficients by minimizing  $\|\mathbf{Y} - \mathbf{\Phi}\mathbf{X}\|_F^2 + \mu \sum_{j=1}^J \|\mathbf{a}_j\|_2^2$ , where  $\mathbf{a}_j$  is a vector of the  $K+1$  polynomial coefficients in (5) for the  $j$ th filter, subject to constraints ensuring that the learned polynomial filters are nonnegative and cover the whole spectrum ( $c - \epsilon \leq \sum_{j=1}^J \hat{g}(\lambda) \leq c + \epsilon$  for some constants  $c$  and  $\epsilon$ ).

### IV. SELECTION OF THE CENTER VERTICES

When selecting the center vertices  $\mathcal{V}_j$  for the  $j$ th filter, four broad options are most commonly used: (i) take  $\mathcal{V}_j = \mathcal{V}$



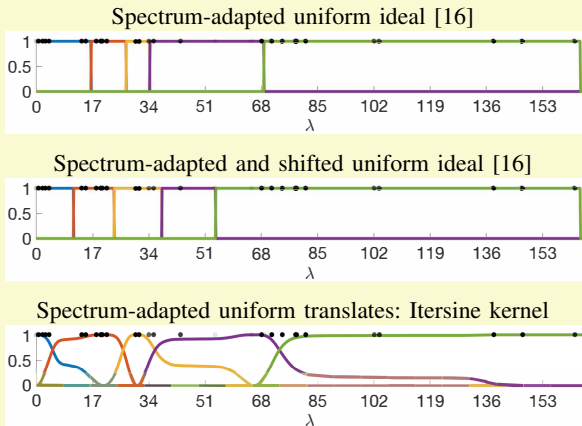
*Example 4 (Fast transforms and inverse transforms via polynomial approximation):* Approximating each spectral filter  $\hat{g}_j(\cdot)$  by a degree  $K$  polynomial  $\hat{p}_{j,K}(\cdot)$  reduces the complexity of applying the dictionary analysis operator from  $\mathcal{O}(N^3)$  to  $\mathcal{O}(K|\mathcal{E}|)$ , which for large, sparse graphs is  $\mathcal{O}(N)$ . The drawback of this scalable approximation is that the tight Parseval frame condition  $G(\lambda) = 1$  for all  $\lambda \in \sigma(\mathcal{L})$  of Theorem 4 is not typically satisfied by the polynomial approximants. Let  $\tilde{\Phi}^*$  be the approximate analysis operator with the polynomial filters, and  $\alpha = \tilde{\Phi}^* \mathbf{f}$  be the resulting analysis coefficients. There are three common options for fast, approximate inverse transforms. The first is to solve  $\tilde{\Phi} \tilde{\Phi}^* \mathbf{f}_{\text{rec}} = \tilde{\Phi} \alpha$  via the conjugate gradient method [11]; and the second is the frame inversion algorithm [64, Ch. 3] that sets  $\mathbf{f}_{\text{rec}}^{(0)} = \frac{2}{A+B} \tilde{\Phi} \alpha$  and iterates  $\mathbf{f}_{\text{rec}}^{(t)} = \mathbf{f}_{\text{rec}}^{(0)} + \mathbf{f}_{\text{rec}}^{(t-1)} - \frac{2}{A+B} \tilde{\Phi} \tilde{\Phi}^* \mathbf{f}_{\text{rec}}^{(t-1)}$ . Both of these iterative methods have complexity  $\mathcal{O}(2TK|\mathcal{E}|)$ , where the number of iterations  $T$  is typically small (5-10), and the speed of convergence depends on how close the ratio of frame bounds  $\frac{B}{A}$  is to 1 (recall that when  $\mathcal{V}_j = \mathcal{V}$  for all  $j$ , the lower frame bound is  $A = \min_{\lambda \in \sigma(\mathcal{L})} \sum_{j=1}^J |\hat{p}_{j,K}(\lambda)|^2$  and the upper frame bound is  $B = \max_{\lambda \in \sigma(\mathcal{L})} \sum_{j=1}^J |\hat{p}_{j,K}(\lambda)|^2$ ). Thus, for the non-tight frame generated from the polynomial filters, near perfect reconstruction is still possible at the same  $\mathcal{O}(N)$  complexity, but the inverse transform may require 10-20 times the number of computations as the fast analysis operator. A third, faster ( $\mathcal{O}(K|\mathcal{E}|)$ ) but less accurate option is to just take  $\mathbf{f}_{\text{rec}} = \frac{2}{A+B} \tilde{\Phi} \alpha$  (i.e., stop the frame inversion algorithm after the initial guess). The high-level intuition is that  $\frac{2}{A+B} \tilde{\Phi} \tilde{\Phi}^*$  is close to the identity matrix  $\mathbf{I}_N$  if  $\frac{B}{A}$  is close to 1 [64, Ch. 3]. For this faster synthesis operator, the squared reconstruction error can be upper bounded by

$$\begin{aligned} \|\mathbf{f} - \mathbf{f}_{\text{rec}}\|_2^2 &= \left\| \mathbf{f} - \frac{2}{A+B} \tilde{\Phi} \alpha \right\|_2^2 \\ &= \left\| \left( \mathbf{I}_N - \frac{2}{A+B} \sum_{j=1}^J \hat{p}_{j,K}^2(\mathcal{L}) \right) \mathbf{f} \right\|_2^2 = \sum_{\ell=0}^{N-1} |f(\lambda_\ell)|^2 \left[ 1 - \frac{2}{A+B} \sum_{j=1}^J |\hat{p}_{j,K}(\lambda_\ell)|^2 \right]^2 \leq \left( \frac{r}{2+r} \right)^2 \|\mathbf{f}\|_2^2, \end{aligned}$$

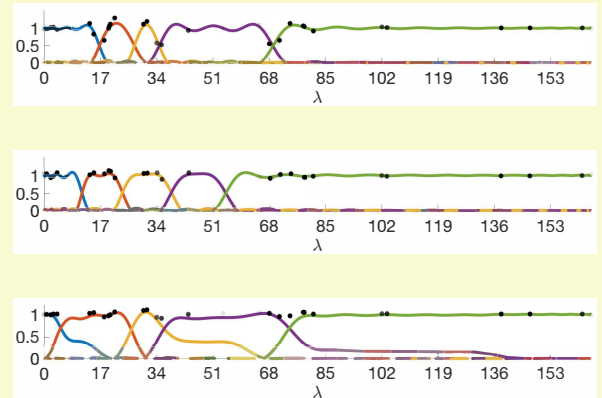
where  $r = \frac{B}{A} - 1$  [11], [64, Ch. 3]. So, regardless of the choice of fast inverse transform, it is desirable for  $\tilde{G}(\lambda) := \sum_{j=1}^J |\hat{p}_{j,K}(\lambda_\ell)|^2$  to be close to 1 for each  $\lambda_\ell$ , yielding a ratio of frame bounds  $\frac{B}{A}$  close to 1 and a small value of  $r$ .

In the figures below, we show degree 40 Chebyshev polynomial approximations to three different sets of five filters on the net25 graph, which features many repeated eigenvalues and therefore has areas of the spectrum with high density. The polynomial approximants to the ideal filter bank in the top row yield a  $\tilde{G}(\lambda)$  (black points in the right column of images) that fluctuates across  $\lambda \in \sigma(\mathcal{L})$ , and the ratio of frame bounds  $\frac{B}{A}$  is equal to 2.78. By shifting the filter end points slightly to be in regions of lower spectral density (middle row), the frame bound ratio drops to 1.85. The smoother spectrum-adapted translates (bottom row) are more amenable to polynomial approximation; the polynomial filters for this design have a frame bound ratio of 1.16.

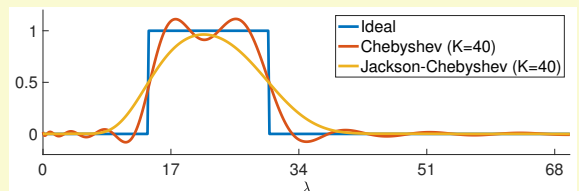
#### Initial Design



#### Degree 40 Chebyshev Polynomial Approximation



While the Chebyshev polynomial approximations [11], [65] are good general choices as they are near optimal in terms of minimizing the maximum approximation error across the spectrum, they may not be the most desirable in certain situations, such as approximating a series of ideal bandpass filters. This is because the oscillations of the Chebyshev polynomials may lead to more energy at the spectral values farther from the bandpass region. The Jackson-Chebyshev polynomial approximations [66] damp these Gibbs oscillations, resulting in less energy farther from the bandpass region (right).



for every filter (i.e., localize every pattern to every vertex as done in [11]); (ii) select the center vertex sets such that  $\sum_{j=1}^J |\mathcal{V}_j| = N$  (i.e., the total number of atoms is equal to  $N$ , the number of vertices in the graph), which is typically referred to as *critical sampling* (e.g., [16]); (iii) do not localize every filter to every vertex but do not restrict the total number of atoms, resulting in a frame (overcomplete dictionary) with more than  $N$  atoms, but fewer than  $JN$  atoms; and (iv) localize a single filter  $\hat{g}_1(\cdot)$  to a strict subset  $\mathcal{V}_1 \subset \mathcal{V}$  of vertices to generate an interpolation basis for a subspace of graph signals, as discussed in Example 1. For option (i), there is no choice to be made regarding the selection of center vertices; for options (ii)-(iv), the specific center vertices for each filter  $\hat{g}_j(\cdot)$  must be chosen, and this selection process may also involve deciding how many center vertices to use for each filter. We now outline the main considerations when choosing between these four options and/or selecting the specific center vertices.

#### A. Frame Bounds and Reconstruction Error

Recall that when  $\mathcal{V}_j = \mathcal{V}$  for every  $j$ , Theorem 4 outlines the computation of the frame bounds and provides a sufficient condition on the filters to yield a (tight) Parseval frame. When  $\mathcal{D}$  is a Parseval frame,

$$\mathbf{f} = \sum_{j=1}^J \sum_{i=1}^N \langle \mathbf{f}, \varphi_{i,j} \rangle \varphi_{i,j} = \sum_{j=1}^J \hat{g}_j(\mathcal{L}) \alpha_j, \quad (10)$$

and so the graph signal  $\mathbf{f}$  can be perfectly reconstructed from the vectors of analysis coefficients,  $\alpha_j = \{\langle \mathbf{f}, \varphi_{i,j} \rangle\}_{i \in \mathcal{V}_j}$ .

When  $\mathcal{V}_j = \mathcal{V}$  for every  $j$  and  $G(\lambda) > 0$  for all  $\lambda \in \sigma(\mathcal{L})$ , the atoms form a frame, but not necessarily a tight frame. This is also the case when each filter is not centered to each vertex, as long as the dictionary atoms span the space of graph signals under consideration (typically  $\mathbb{R}^N$ ). Example 4 details three options for inverse transforms in these situations where the dictionary atoms form a frame, but not a tight frame. The least squares solution via conjugate gradient and iterative frame inversion algorithm are accurate but converge slower when the frame is farther from being tight ( $\frac{B}{A} \gg 1$ ). We are not aware of any investigations into how to select the center vertices from general weighted graphs in a manner that explicitly controls the ratio of frame bounds.

#### B. Band-By-Band Reconstruction and Connections to Sampling and Interpolation Theory

When the filters are not localized to every vertex, but (i) do not overlap too much and (ii) evenly cover the whole spectrum ( $\sum_{j=1}^J \hat{g}_j(\lambda) \approx 1$  for all  $\lambda$ ), an alternative approach to synthesis is to try to reconstruct the signal from each band separately and add them up, similar to (10), except that we replace  $\hat{g}_j(\mathcal{L})$  with a different synthesis operator for each band. The main idea is that  $\mathbf{f} \approx \sum_{j=1}^J \hat{g}_j(\mathcal{L}) \mathbf{f}$ , where each filtered signal  $\hat{g}_j(\mathcal{L}) \mathbf{f}$  belongs to the subspace  $\mathcal{U}_j$  spanned by the eigenvectors  $\{\mathbf{u}_\ell\}_{\ell: \hat{g}_j(\lambda_\ell) \neq 0}$ . Thus,  $\dim(\mathcal{U}_j)$  provides an estimate for the number of center vertices,  $|\mathcal{V}_j|$ , required to recover  $\hat{g}_j(\mathcal{L}) \mathbf{f}$  from the analysis coefficients  $\alpha_j$ .

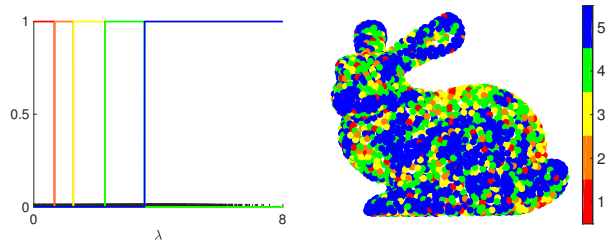


Fig. 5. Selection of center vertices for a critically sampled LSGFD with five spectrum-adapted ideal octave-band filters on the bunny graph. Exactly one of the five filters is localized to each vertex  $i$ , according to the mapping shown in the right image, yielding a total of  $N$  atoms [16].

The problem of first selecting the center vertices  $\mathcal{V}_j$  and then specifying a method to recover  $\hat{g}_j(\mathcal{L}) \mathbf{f}$  from  $\alpha_j$  falls into the broader category of *sampling and interpolation of graph signals*. Generally speaking, algorithm development in this area proceeds as follows: (i) define a signal model and, if appropriate, a measurement noise model; (ii) specify a reconstruction method that maps a given set of (possibly noisy) sample values of the signal to the entire signal in a way that optimizes a specified error criteria, accounting for the signal model (and noise model) and graph structure; and (iii) given the signal model and reconstruction method, select vertices (often constrained to a fixed number of them) on which to sample the graph signal values. See [72]–[75] for more detailed literature reviews and theoretical results on sampling and interpolation of graph signals. The majority of sampling selection methods (i) focus on smooth or lowpass graph signals, and (ii) require the Laplacian eigenvectors associated with the eigenvalues  $\lambda_\ell$  for which  $\hat{g}_j(\lambda_\ell) \neq 0$ . We focus our review in the remainder of this section on sampling strategies where at least one of these conditions is not met, as these strategies are particularly relevant for LSGFD design.

For the special case of a set of ideal filters that partition the spectrum, as shown, e.g., in Example 2, Example 4, and Fig. 5, the critically sampled sets of center vertices can be constructed so as to ensure the signal can be recovered perfectly from the  $N$  analysis coefficients.

*Theorem 5 (Prop. 2 and Cor. 1 of [16]):* Let  $\hat{g}_1(\cdot), \hat{g}_2(\cdot), \dots, \hat{g}_J(\cdot)$  be a set of spectral filters with the properties that for all  $\lambda \in \sigma(\mathcal{L})$ ,  $G(\lambda) = 1$  and  $\hat{g}_j(\lambda) \hat{g}_{j'}(\lambda) = 0$  for all  $j \neq j'$  (i.e., the filters form a partition of the spectrum  $[0, \lambda_{\max}]$  such that each eigenvalue is in the support of exactly one filter). Then there exists a partition  $\{\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_J\}$  of  $\mathcal{V}$  with  $|\mathcal{V}_j| = \sum_{\ell=0}^{N-1} \mathbb{1}_{\{\hat{g}_j(\lambda_\ell)=1\}}$  such that the resulting dictionary  $\mathcal{D}$  of the form (2) is a basis. Each atom in the basis is orthogonal to all atoms generated from a different filter. Ref. [16] provides a constructive algorithm for finding the center vertex sets  $\{\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_J\}$  in Theorem 5, the choice of which is not unique. This algorithm requires a full eigen-decomposition and therefore is only applicable to small or medium graphs. Each set  $\mathcal{V}_j$  is a *uniqueness set* [76] for the subspace  $\mathcal{U}_j$  spanned by the eigenvectors  $\{\mathbf{u}_\ell\}_{\ell: \hat{g}_j(\lambda_\ell)=1}$ . That is, any graph signal in  $\mathcal{U}_j$  can be uniquely recovered from its values at the vertices in  $\mathcal{V}_j$ . In Fig. 5, we display a set of five spectrum-adapted ideal octave-band filters on the bunny graph, and the corresponding partition of the vertices into the uniqueness sets  $\mathcal{V}_1$  to  $\mathcal{V}_5$ .

Which sampling and interpolation techniques are applicable to smoother approximations to ideal filter banks (e.g., middle row, right in Example 4) for large, sparse graphs, where computation of the Laplacian eigenvectors is not tractable? The fastest methods are random, not necessarily uniform, sampling methods. For example, leveraging the literature on compressed sensing, [16], [77] propose to sample according to a non-uniform distribution with the weight at vertex  $i$  proportional to an approximation of  $\|\mathbf{U}_{\mathcal{R}_j}^* \delta_i\|_2^2$ , where  $\mathbf{U}_{\mathcal{R}_j}$  is the submatrix of  $\mathbf{U}$  containing the columns corresponding to the support of the ideal filter  $\hat{g}_j$ . The  $j$ th filtered signal is then reconstructed via the optimization problem

$$\min_{\mathbf{z}_j \in \mathbb{R}^N} \left\{ \mathbf{z}_j^* \phi_j(\mathcal{L}) \mathbf{z}_j + \kappa \|\Omega_j^{-\frac{1}{2}} (\mathbf{M}_j \mathbf{z}_j - \alpha_j)\|_2^2 \right\}, \quad (11)$$

where  $\Omega_j$  is a  $|\mathcal{V}_j| \times |\mathcal{V}_j|$  diagonal matrix with the  $j^{\text{th}}$  channel sampling weights of  $\mathcal{V}_j$  along the diagonal,  $\mathbf{M}_j \in \mathbb{R}^{|\mathcal{V}_j| \times N}$  is a downsampling matrix that maps a full graph signal to a vector of its values on  $\mathcal{V}_j$ , and  $\kappa > 0$  is a parameter to trade off the two optimization objectives: the regularization term  $\mathbf{z}_j^* \phi_j(\mathcal{L}) \mathbf{z}_j$  in (11) penalizes reconstructions with support outside of the desired spectral band, and the data fidelity term  $\|\Omega_j^{-\frac{1}{2}} (\mathbf{M}_j \mathbf{z}_j - \alpha_j)\|_2^2$  penalizes reconstructions that do not match the analysis coefficients (filtered signal values). From the first-order optimality conditions, the solution to (11) is the solution to the linear system of equations

$$\left( \kappa \mathbf{M}_j^* \Omega_j^{-1} \mathbf{M}_j + \phi_j(\mathcal{L}) \right) \mathbf{z}_j = \kappa \mathbf{M}_j^* \Omega_j^{-1} \alpha_j, \quad (12)$$

which can be solved, for example, with the preconditioned conjugate gradient method (see [16] for more on the choice of the preconditioner).

The other more scalable options are efficient greedy methods that do not rely on the Laplacian eigendecomposition, such as [73], [78]. For example, for each  $j$ , the eigendecomposition-free (ED-free) method of [73] attempts to select center vertices  $i$  such that (i)  $\|T_i g_j\|_1 = \|\varphi_{i,j}\|_1$  is large, and (ii) the atoms  $\varphi_{i,j}$  do not overlap too much in the vertex domain. The binary search Gershgorin disc alignment (BS-GDA) method of [78] aims to select vertices in a way that reduces the condition number of the matrix  $\kappa \mathbf{M}_j^* \Omega_j^{-1} \mathbf{M}_j + \phi_j(\mathcal{L})$  in (12) (with  $\Omega_j = \mathbf{I}_N$  and a regularization term of  $\phi(\mathcal{L}) = \mathcal{L}$ ), improving the reconstruction stability. These methods may lead to better reconstruction for lowpass signals, but are slower than random sampling and are either not applicable to or may lead to worse reconstruction for bandpass signals (see Sec. IV-D and Example 5).

### C. Allocation of Center Vertices Across the Filters

When the sum of the cardinalities  $\{|\mathcal{V}_j|\}$  is constrained (e.g., to  $N$ ), how should we allocate the center vertices (samples) across the filters? One option is to estimate the spectral cumulative density and match the number of samples to the estimated number of eigenvalues in each band. Another option is to adjust the distribution of samples to the signal  $\mathbf{f}$  by multiplying the initial allocation of samples to  $\mathcal{V}_j$  by a factor that increases with the amount of energy in the filtered signal  $\hat{g}_j(\mathcal{L}) \mathbf{f}$  [16].

### D. Computational Complexity

We briefly analyze the computational complexity implications of the choice of method for selecting the center vertices on large, sparse graphs, the most typical and important case for complexity considerations. We assume the filters are degree  $K$  polynomials throughout this analysis. First, for selecting the center vertices, the costs of the random sampling methods (either signal-adapted or not) are negligible if the spectral density of the Laplacian has been estimated (or  $\mathcal{O}(K|\mathcal{E}|)$  if it has not), at least an order of magnitude faster than the  $\mathcal{O}(KN|\mathcal{E}|) = \mathcal{O}(N^2)$  complexity of the greedy ED-free method. Second, regardless of the selection of center vertices, applying the analysis operator (forward transform) has the same complexity  $\mathcal{O}(K|\mathcal{E}|)$  as the case where we localize each filter to every vertex; i.e., there is not a significant computational savings in the analysis step from subsampling the center vertices. Third, as detailed in Example 4, the complexity of the inverse transform is  $\mathcal{O}(2TK|\mathcal{E}|)$  if either the least squares least squares solution via conjugate gradient or iterative frame inversion algorithm is used, and  $\mathcal{O}(K|\mathcal{E}|)$  if the inverse is approximated by a constant multiple of the analysis operator. The band-by-band reconstruction method of (12) has a similar complexity of  $\mathcal{O}(T\tilde{K}|\mathcal{E}|)$ , where  $\tilde{K}$  is the degree of the polynomial penalty function  $\phi$ . These inverse transform complexities do not depend on either the method used to select the center vertices or the number of center vertices for each filter. In summary, for large, sparse graphs, the complexities of the design, forward transform, and inverse transform are all no more than linear in the number of vertices, as long as each filter is localized to every vertex or the center vertices are selected through non-uniform random sampling.

### E. Memory

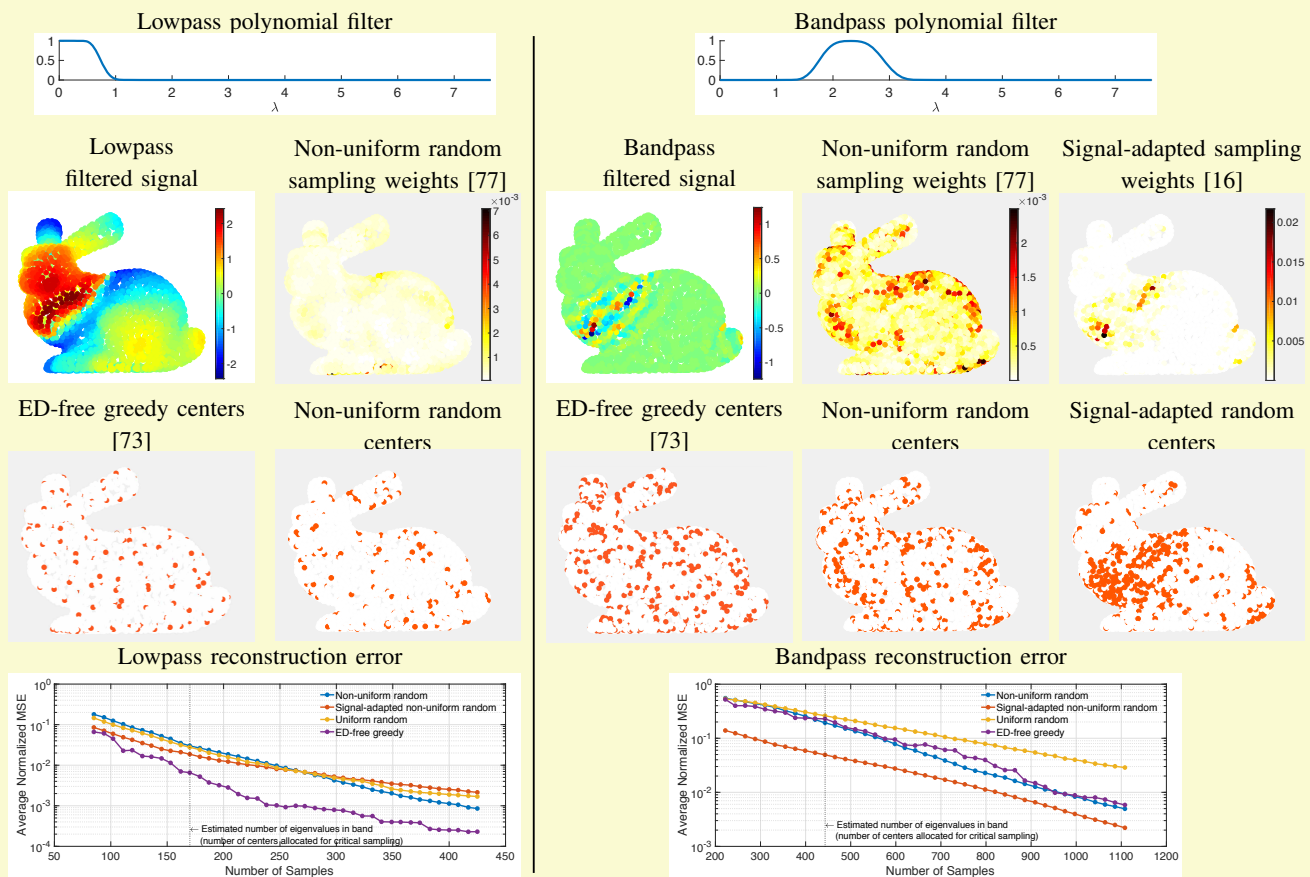
The total number of analysis coefficients to store for a graph signal of length  $N$  is  $\sum_{j=1}^J |\mathcal{V}_j|$ . Thus, the complete sampling of option (ii) yields  $NJ$  coefficients to store, while the critical sampling of option (iii), e.g., reduces that number to  $N$ . Except for special filters such as the heat kernel, the ED-free method requires storing the entire matrices  $\{\hat{g}_j(\mathcal{L})\}$ ; thus, when memory is an important concern and a critically sampled dictionary is desired, random sampling should be used to select the center vertex sets.

## V. THEORETICAL CONSIDERATIONS AND METRICS

Whether the dictionary of atoms is critically sampled (a basis for the vector space of signals) or redundant (a frame for the vector space of signals), different mathematical characteristics can be beneficial for different applications. Desirable characteristics of dictionaries may include, for example:

- The atoms are not too correlated, in order to enhance the discriminatory power of taking inner products between each atom and a target signal;
- The atoms are jointly localized in the vertex domain and the spectral domain.
- Classes of signals on the graph (e.g., globally smooth or locally smooth signals), can be represented as sparse combinations of the dictionary atoms.

*Example 5 (Scalable sampling strategies for LSGFD design):* We compare the sampling and reconstruction of a lowpass and a bandpass filtered signal using (i) greedy eigendecomposition-free (ED-free) sampling [73]; (ii) uniform random sampling; (iii) non-uniform random sampling [16], [77]; and (iv) signal-adapted non-uniform random sampling [16]. For each polynomial filter type, we show the random sampling distributions, examples of sets of center vertices selected by these methods, and the normalized mean square reconstruction error (NMSE)  $\frac{\|\mathbf{z}_j^* - \hat{g}_j(\mathcal{L})\mathbf{f}\|_2^2}{\|\hat{g}_j(\mathcal{L})\mathbf{f}\|_2^2}$  between the filtered signal and the reconstruction  $\mathbf{z}_j^*$ , computed via (12), averaged over 50 trials of choosing the random center locations (samples). The greedy ED-free method explicitly prioritizes choosing samples that are not too close to the previously chosen samples. The non-uniform random sampling weights are derived from a computationally-efficient approximation to the ideal distribution, for which the probability of sampling vertex  $i$  is proportional to  $\|\mathbf{U}_{\mathcal{R}_j}^* \delta_i\|_2^2$ . This approach has a close connection to *leverage score* sampling in the statistics and numerical linear algebra literature [79], [80]. The signal-adapted non-uniform random sampling distribution (shown in the third column of the right group of images) is computed by multiplying the initial non-uniform weight associated with vertex  $i$  (shown in the second column of the right group of images) by  $\log(1 + |(\hat{g}_j(\mathcal{L})\mathbf{f})(i)|)$  and then renormalizing. The intuition is that it is beneficial to take additional samples in regions of the graph where the filtered signal has the most energy. For this particular bandpass signal, the regions of highest energy are around the midsection and tail of the bunny. Indeed, the signal-adapted method leads to more samples in these areas, and, in turn, to better reconstruction performance [16]. The average NMSE of the signal-adapted sampling method at the dashed vertical line represents a 79%, 78%, and 83% reduction of the errors of the ED-free, non-uniform random sampling, and uniform random sampling methods, respectively.



Key insights highlighted by this example include:

- 1) For the lowpass filter, the results are consistent with the common insight from the graph sampling literature that the scalable greedy methods are slower than random sampling, but can yield better reconstruction performance by forcing the samples to be more spread out [75].
- 2) For bandpass or highpass filters, however, forcing or incentivizing the samples to be more spread out does not necessarily (or even usually) improve reconstruction performance.
- 3) When performing random sampling to choose the center vertices, using a non-uniform sampling distribution is more important for bandpass and highpass filtered signals. For lowpass filtered signals, the non-uniform random sampling distribution is much closer to uniform as the energy distributions of the eigenvectors at the low end of the spectrum tend to be more evenly spread across the graph than those associated with higher Laplacian eigenvalues.
- 4) In the classical sampling and interpolation problem, the complete signal is not usually available when selecting the sample locations. However, in the context of subsampled LSGFDs, it is often feasible and highly beneficial to adapt the non-uniform random selection of the center vertices to the energy distribution of each filtered signal.

We briefly describe each of these considerations, in order.

One common metric for capturing the correlation between dictionary atoms is the *cumulative coherence* [81] of the dictionary  $\mathcal{D}$ , which, for a given sparsity level  $k$ , is defined as

$$\mu_1(k) := \max_{\{\Theta \subset \mathcal{D}: |\Theta|=k\}} \max_{\psi \in \mathcal{D} \setminus \Theta} \sum_{\theta \in \Theta} \frac{|\langle \psi, \theta \rangle|}{\|\psi\|_2 \|\theta\|_2}.$$

Uncertainty principles for signals on graphs (e.g., [82]–[86]) characterize the degree to which graph signals can be jointly localized in both the vertex (spatial) domain and the spectral (frequency) domain. Particularly interesting for guiding the selection of the center vertices of LGSFDs are the uncertainty principles developed in [82], [86]. Let  $\gamma \in \mathbb{R}^N$  be a *spatial filter or window function*; i.e., a set of weights assigned to the vertices, with  $0 \leq \gamma_i \leq 1$  for all  $i \in \mathcal{V}$  and  $\max_{i \in \mathcal{V}} \gamma_i = 1$ . Typical examples of such spatial filters include (i) setting the weights equal to 1 for all vertices in a certain region of the graph and 0 elsewhere, or (ii) choosing each  $\gamma_i$  to be a measure of the distance from vertex  $i$  to a fixed center vertex  $n$ . Then for any graph signal  $\mathbf{f}$ , the quantity  $\bar{\mathbf{m}}_\gamma(\mathbf{f}) := \frac{\mathbf{f}^* \Gamma \mathbf{f}}{\|\mathbf{f}\|_2^2}$ , where  $\Gamma$  is a diagonal matrix with diagonal elements equal to the weights  $\gamma$ , captures the portion of the energy of  $\mathbf{f}$  that is located in the specified region of the graph (i.e., those vertices with high weights  $\gamma_i$ ). Similarly, for each filter satisfying  $0 \leq \hat{g}_j(\lambda_\ell) \leq 1$  for all  $\ell$  and  $\max_\ell \hat{g}_j(\lambda_\ell) = 1$ , the quantity  $\bar{\mathbf{c}}_{\hat{g}_j}(\mathbf{f}) := \frac{\mathbf{f}^* \hat{g}_j(\mathcal{L}) \mathbf{f}}{\|\mathbf{f}\|_2^2}$  captures the portion of the energy of  $\mathbf{f}$  that is located in the region of the spectrum specified by the filter. The uncertainty principles in [82], [86] characterize and provide algorithms to approximate the sets  $\mathcal{W}_{\gamma, \hat{g}_j} := \{(\bar{\mathbf{m}}_\gamma(\mathbf{f}), \bar{\mathbf{c}}_{\hat{g}_j}(\mathbf{f})) : \|\mathbf{f}\|_2 = 1\} \subseteq [0, 1]^2$ . For example, for a given filter pattern, such a principle can inform how localized a dictionary atom of the form (1) can be in the vertex domain (typically around the center vertex  $i$ ). Due to the irregularity of general graphs and the possibility of highly localized Laplacian eigenvectors, uncertainty does not always exist, in which case  $\mathcal{W}_{\gamma, \hat{g}_j}$  may be equal to  $[0, 1]^2$ .

With regard to the third desirable characteristic above, nearly a decade after being listed as an open issue in [1], relatively little progress has been made in developing a mathematical theory of approximation linking structural properties of graph signals and their underlying graphs to the sparsity of the analysis coefficients  $\{\langle \mathbf{f}, \varphi_{i,j} \rangle\}$ , analogous to the theory for wavelet transform coefficients in Euclidean domains (see, e.g., [87]). In [88], vertex domain dictionary designs are proposed that sparsely represent defined classes of piecewise constant and piecewise smooth graph signals. For the special case of signals on circulant graphs, [25] defines a family of complex exponential polynomial graph signals and designs a class of filters that annihilate graph signals from this class; i.e.,  $\langle \mathbf{f}, T_i g_j \rangle = 0$  for all  $i$ . Reference [89] defines notions of global and local regularity for graph signals, and begins to connect the regularities of the signals and the degree of the polynomial filters to the decay of the magnitudes of spectral graph wavelet analysis coefficients.

## VI. APPLICATION EXAMPLES AND COMPARISON VIA NUMERICAL EXPERIMENTS

In this section, we first describe LSGFD transform methods for two signal processing tasks – denoising and non-linear approximation (compression) – and then perform a set of targeted numerical experiments that attempt to answer high-level design questions and help focus the community’s research going forward. Our objective is not to determine whether a specific dictionary outperforms other dictionaries in a specific task; for that type of analysis, we encourage readers to experiment on their own data with the publicly available code used to generate all figures and tables in this article.

### A. Denoising

We consider the denoising problem of recovering a graph signal  $\mathbf{f}$  from a noisy observation,  $\mathbf{y} = \mathbf{f} + \boldsymbol{\xi}$ , where  $\boldsymbol{\xi} \in \mathbb{R}^N$  is a white Gaussian noise vector whose entries are independent and identically distributed (i.i.d.) normal random variables with mean 0 and known variance  $\sigma^2$ . We use the common wavelet denoising method of performing soft thresholding on the LSGFD transform coefficients, and then resynthesizing the signal with the inverse transform. Specifically, we take  $\mathbf{f}_{\text{denoised}} = \Phi^{*-1}(\bar{\boldsymbol{\alpha}})$ , where each soft thresholded coefficient in the vector  $\bar{\boldsymbol{\alpha}}$  is set to

$$\bar{\alpha}_{i,j} = \text{sgn}(\langle \mathbf{y}, \varphi_{i,j} \rangle) \cdot \max\{0, |\langle \mathbf{y}, \varphi_{i,j} \rangle| - \mathcal{Y}_{i,j}\}. \quad (13)$$

If the dictionary used to transform the noisy signal is a Parseval frame, then  $\Phi^{*-1}(\bar{\boldsymbol{\alpha}}) = \Phi \bar{\boldsymbol{\alpha}}$ ; otherwise (e.g., not every filter is localized to every vertex or the filters do not satisfy the tight frame condition of Theorem 4 due to polynomial approximation),  $\Phi^{*-1}$  can be taken to be any of the three approximate inverse transforms discussed in Example 4.

We set the soft thresholds  $\mathcal{Y}_{i,j}$  in (13) to 0 for the scaling functions (atoms generated from filters satisfying  $\hat{g}_j(0) > 0$ ) since these coefficients are not expected to be sparse. For the other thresholds, as in [32], to account for the fact that the dictionary atoms have different norms, we use atom-adapted thresholds of the form  $\mathcal{Y}_{i,j} = \mathcal{Y}_j \sigma \|\varphi_{i,j}\|_2$ , where  $\{\mathcal{Y}_j\}$  are the  $J - 1$  scalar parameters. The optimal value of each  $\mathcal{Y}_j$  is estimated with Stein’s Unbiased Risk Estimator (SURE). As detailed in [90], for soft thresholding on a dictionary with a single lowpass filter ( $j = 1$ ), this amounts to solving the following for each  $j = 2, 3, \dots, J$ :

$$\underset{\mathcal{Y}_j}{\text{argmin}} \sum_{i=1}^N \left\{ \min\{|\langle \mathbf{y}, \varphi_{i,j} \rangle|^2, \mathcal{Y}_j^2 \sigma^2 \|\varphi_{i,j}\|_2^2\} + 2\sigma^2 \|\varphi_{i,j}\|_2^2 \mathbb{1}_{\{|\langle \mathbf{y}, \varphi_{i,j} \rangle| \geq \mathcal{Y}_j \sigma \|\varphi_{i,j}\|_2\}} \right\}. \quad (14)$$

Importantly, the objective in (14) does not depend on the unknown signal  $\mathbf{f}$ . The thresholds  $\{\mathcal{Y}_{i,j}\}$  and the objective in (14) depend on the atom norms  $\{\|\varphi_{i,j}\|_2\}$ . For small to medium graphs, these can be computed exactly; for polynomial filter designs on large, sparse graphs, they can be efficiently estimated as  $\|\varphi_{i,j}\|_2 \approx \frac{1}{\sigma} \text{s.d.}(\{\delta_i^* \hat{p}_{j,K}(\mathcal{L}) \boldsymbol{\eta}_l\}_{l=1}^L)$ , where  $\hat{p}_{j,K}$  is a polynomial approximation to  $\hat{g}_j$  and  $\{\boldsymbol{\eta}_l\}_{l=1}^L$  is a sequence of i.i.d. random vectors, each with i.i.d. entries normally distributed with mean 0 and variance  $\sigma^2$ .

$\sigma/\sigma_f$	Sensor Network				Bunny				Minnesota				Cerebellum			
	1/8	1/4	1/2	1	1/8	1/4	1/2	1	1/8	1/4	1/2	1	1/8	1/4	1/2	1
Spectral graph wavelets [11]	1.56	2.29	<b>4.04</b>	6.37	<b>4.99</b>	<b>5.92</b>	<b>7.69</b>	<b>10.46</b>	5.81	6.69	8.50	10.93	0.53	1.74	3.76	6.85
Cohen-Daubechies-Feauveau 9/7 filters [20]	1.50	2.04	3.83	5.85	4.76	5.68	7.44	10.42	5.51	6.52	8.61	11.86	0.50	1.60	4.03	6.86
Uniform ideal filters	1.00	1.89	3.52	5.75	3.51	4.54	6.48	9.33	3.50	4.72	6.45	7.20	0.66	<b>2.06</b>	4.02	7.05
Octave-band (wavelet) ideal filters	0.96	1.82	3.61	6.12	3.67	4.81	6.97	10.16	4.37	5.81	8.23	11.61	0.44	1.49	3.62	6.55
DCT with frequency conversion [20]	<b>1.77</b>	<b>2.30</b>	3.94	6.26	4.87	5.88	7.61	10.32	4.86	5.83	7.51	8.73	0.63	1.89	<b>4.14</b>	<b>7.10</b>
Log-warped DCT with frequency conversion [20]	1.39	2.07	3.85	6.26	4.51	5.76	7.60	10.37	<b>6.29</b>	<b>7.08</b>	<b>9.10</b>	<b>12.06</b>	0.41	1.39	3.59	6.31
Uniform translates: Itersine kernel [31]	1.33	2.14	3.93	6.37	4.61	5.69	7.36	10.41	4.63	5.89	7.94	9.57	<b>0.69</b>	2.03	4.10	7.09
Log-warped translates (wavelets): Itersine kernel [31]	1.42	2.15	3.98	6.38	4.40	5.76	7.48	10.44	5.81	6.76	8.86	11.94	0.55	1.70	3.90	6.99
Spectrum-adapted uniform translates: Itersine kernel [31]	1.48	2.11	3.97	6.27	4.35	5.46	7.14	10.32	5.11	6.23	8.48	11.22	0.68	2.05	4.05	7.03
Spectrum-adapted wavelets: Itersine kernel [31]	1.32	2.12	4.00	<b>6.61</b>	4.46	5.60	7.59	10.45	5.69	6.65	8.94	12.02	0.63	1.95	4.05	7.10
Signal-adapted Meyer-type [29]													0.39	1.38	3.53	6.85

TABLE I

DENOISING PERFORMANCE FOR DIFFERENT GRAPH SIGNALS, NOISE LEVELS, AND FILTER DESIGNS, SHOWN IN SNR IMPROVEMENT:

$$\Delta_{\text{SNR}} = 10 \log_{10} \left( \frac{\|\mathbf{f}\|_2^2}{\|\mathbf{f}_{\text{DENOISED}} - \mathbf{f}\|_2^2} \right) - 10 \log_{10} \left( \frac{\|\mathbf{f}\|_2^2}{\|\boldsymbol{\xi}\|_2^2} \right) = 10 \log_{10} \left( \frac{\|\boldsymbol{\xi}\|_2^2}{\|\mathbf{f}_{\text{DENOISED}} - \mathbf{f}\|_2^2} \right)$$

### B. Non-linear approximation

One approach to compression of smooth and piecewise-smooth graph signals is to represent them as sparse linear combinations of LSGFD atoms. To find such sparse representations for a graph signal  $\mathbf{f}$ , the sparse coding optimization

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\mathbf{f} - \Phi\boldsymbol{\alpha}\|_2^2 \quad \text{subject to } \|\boldsymbol{\alpha}\|_0 \leq T_0,$$

where  $T_0$  is a predefined sparsity level, can be approximately solved, e.g., by normalizing the dictionary atoms and then applying the greedy orthogonal matching pursuit (OMP) algorithm [81]. When the graph is very large and OMP becomes impractical computationally, a common approximation method is to hard threshold the analysis coefficients (normalized by an estimate of the corresponding atom norm), and then resynthesize the signal from the  $T_0$  largest magnitude coefficients, via one of the inverse transform methods described in Example 4.

### C. Design considerations revisited

We use five different test signals in this section: the piecewise-smooth signal on a sensor network shown in Fig. 6 [12]; the piecewise-smooth signal on the bunny graph shown in Fig. 3 [12]; the piecewise constant signal on the Minnesota road network from [14]; the average of 292 fMRI signals on the cerebellum graph [29], [36]; and the average temperatures for March 2018 at  $N = 469, 404$  locations in the United States on an eight-neighbor local graph from [16]. We preprocess all signals by normalizing them to have mean zero.

1) *How should we design the filters? Are there dictionary metrics that can inform this decision?*: In Table I, we examine the denoising performance for different graph signals, noise levels ( $\sigma_f$  is the standard deviation of the signal values and  $\sigma$  is the standard deviation of the noise values), and filter designs. For each filter design type, we report the best SNR improvement over the range of 3-12 filters, each localized to every vertex in the graph to create the dictionary. In Fig. 6, we use OMP to compress a piecewise-smooth signal on a sensor network with nine different dictionaries. For both of these application examples, we use exact computations throughout.

What are the design takeaways from these examples? First, there is no clear “best” filter design method across applications

and setup parameters, which is not surprising but worth stating. Second, using redundant dictionaries generally enables sparser representations of the data and better compression performance than the bases considered in this example. Third, with exact computations, there is not a significant drawback from using a non-tight frame; e.g., the spectral graph wavelets have excellent performance in many of the denoising setups and the compression example. Fourth, in this setting, the cumulative coherence is not necessarily a good prediction of compression (sparse approximation) performance. For example,  $\mu_1(25)$  is equal to 24.96, 24.77, 24.57, 15.71, 13.12, and 4.12 for the translated Hann wavelets, uniform Hann translates, spectral graph wavelets, uniform DCT filters, fast tight wavelet frame, and  $M$ -channel critically sampled filter bank, respectively; that is, higher, not lower, cumulative coherence is correlated with better performance. Indeed, the identification of dictionary metrics that correlate with application performance metrics is very much an open issue.

2) *How should we select the number of filters?*: There is not an easy answer for best practices in choosing the number of filters in an LSGFD. For small and medium graphs where exact filtering is tractable, the choice that yields the best performance usually depends on the specific application and signal model. For example, in Table II, for each type of filter design, we provide the number of filters that leads to the best performance in the corresponding entry of the denoising

$\sigma/\sigma_f$	Sensor Network				Bunny				Minnesota				Cerebellum			
	1/8	1/4	1/2	1	1/8	1/4	1/2	1	1/8	1/4	1/2	1	1/8	1/4	1/2	1
SGWT	7	12	8	10	9	10	8	8	12	6	6	12	4	4	8	9
CDF 9/7	12	8	10	6	7	12	9	6	9	8	12	9	12	9	12	5
UnifIdeal	6	10	11	12	6	7	11	12	9	12	12	12	11	11	11	11
WavIdeal	4	8	4	6	6	11	9	6	9	8	12	9	12	9	6	5
UnifDCT	4	6	6	11	4	8	12	12	12	12	12	11	12	12	12	12
WavDCT	6	12	6	12	6	8	6	8	4	6	6	10	10	12	12	12
UnifTr	4	6	6	10	4	5	8	12	10	12	12	12	11	11	12	9
WavTr	3	4	12	4	4	4	5	6	3	4	5	9	7	11	12	12
SpAUnifTr	3	4	6	7	5	5	9	11	6	11	12	12	10	12	10	11
SpAWavTr	4	4	7	4	4	4	4	7	4	4	4	5	12	12	12	9
SigAUnif													12	12	12	12

TABLE II

NUMBER OF FILTERS THAT ACHIEVES THE BEST DENOISING RESULT SHOWN IN THE CORRESPONDING ENTRY IN TABLE I. THE RANGE CONSIDERED FOR EACH DESIGN IS 3 TO 12 FILTERS.

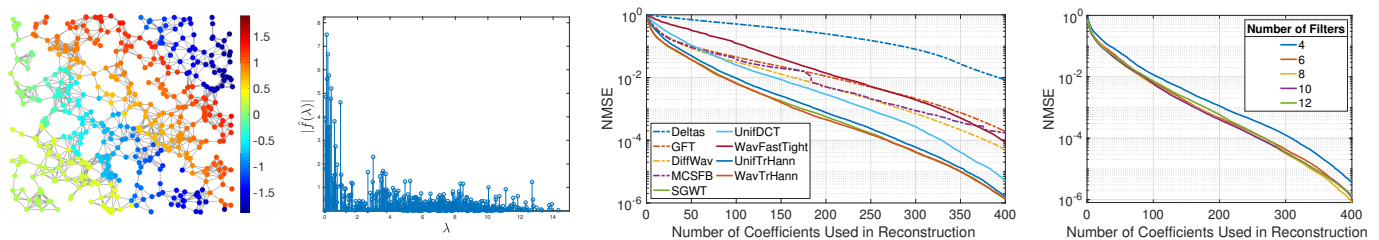


Fig. 6. Non-linear approximation of the piecewise-smooth signal on a random sensor network with 500 vertices from [12]. The left two images show the signal in the vertex domain and spectral domain, respectively. The third image shows the normalized mean square reconstruction error incurred by applying the OMP algorithm to the analysis coefficients from nine different dictionary transforms. The first four are bases, with errors shown in dashed lines, and the next five are frames generated from six filters, with all but the spectral graph wavelet frame being Parseval frames. The final image shows the same error for tight frames with different numbers of uniform translates of the Hann kernel as the filters.

results of Table I. The best number of filters to use varies widely depending on the type of filters, level of noise, and underlying graph signal, most likely according to how well some of the resulting atoms can capture the energy of the signal. Similarly, the last image in Fig. 6 shows that the choice of the number of filters in the compression of the sensor signal does not have an outsized impact when using an exactly computed dictionary generated (in this case, based on octave-band translates of a Hann kernel).

The drawbacks of using more filters are exacerbated when the data resides on a large graph, necessitating approximations for computational tractability. In Fig. 7, we show the denoising performance on the bunny graph signal of tight wavelet frames generated from different numbers of uniform translates. With exact computations, the performance continues to slowly increase with additional filters (although it has nearly saturated at  $J = 10$  filters); however, as the number of filters increases, they become narrower and more difficult to approximate by low degree polynomials, leading to worse performance with approximate computations. This is part of the reason a general rule of thumb in practice has been to use 4-8 filters.

3) *When is it beneficial to use fewer center vertices?:* It appears to be most beneficial to include fewer center vertices in applications where either (i) memory is a critical issue, or (ii) the graph signals under consideration reside on very large, sparse graphs, necessitating approximations at all stages of the signal processing or machine learning pipelines. As an example, in Fig. 8, we compress the average temperature

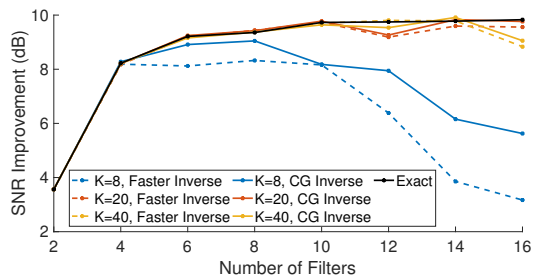


Fig. 7. The wider filters arising in the filter designs with a lower number of filters are more amenable to approximation by low order polynomials, which in turn keeps the ratio of frame bounds closer to 1 and can lead to better reconstruction and denoising performance when fast numerical approximations are used.

signal, reconstructing it from  $\frac{N}{2}$  of the analysis coefficients. We compare the hard thresholding method using a five filter spectrum-adapted tight wavelet dictionary with complete sampling ( $\mathcal{V}_j = \mathcal{V}$  for all  $j$ ) to the band-by-band reconstruction method outlined in Section IV-B, using the 5-channel critically sampled filter bank ( $M$ -CSFB) design detailed in [16] with the center vertices for each filter selected via signal-adapted nonuniform random sampling. The hard thresholding method tends to keep the coefficients associated with the scaling functions ( $j = 1$ ) and center vertices in the northern and southern parts of the country. Combined with the fact that the atoms are localized within  $K = 50$  hops of the center vertices due to the necessary polynomial approximation of the filters on a graph this size, it incurs more reconstruction error in areas of the graph where either the center vertices were not selected or there are sharper changes in the signal values, which would be captured by the discarded wavelet coefficients ( $j > 1$ ). Although it is possible to mitigate this issue to some extent via more clever reconstruction methods, the problem also becomes more pronounced as the compression ratio increases and fewer analysis coefficients are stored. The  $M$ -CSFB dictionary, on the other hand, only has 28,022 atoms generated from the scaling filter, leading to many of the wavelet coefficients being used in the reconstruction. Moreover, the random sampling selection of center vertices leads to preserved coefficients associated with atoms centered across the country. A high-level difference between these two approaches is that the  $M$ -CSFB already includes some of the compression in the process of choosing the strict subsets  $\mathcal{V}_j \subset \mathcal{V}$  to be the center vertices.

## VII. SUMMARY AND FUTURE DIRECTIONS

In summary, localized spectral graph filter frames feature structured atoms with analytically tractable properties such as localization around the center vertex and spectral patterns that carry a notion of smoothness with respect to the graph. At the same time, efficient numerical approximations exist to the forward and inverse transforms, rendering these dictionaries useful in myriad applications on large, sparse graphs. For small and medium graphs, it is typical to perform an exact eigendecomposition, localize each filter to every vertex, and choose the filters to satisfy the condition of Theorem 4, guaranteeing that the frame is tight. For large, sparse graphs, we reviewed fast techniques to approximate or bound the

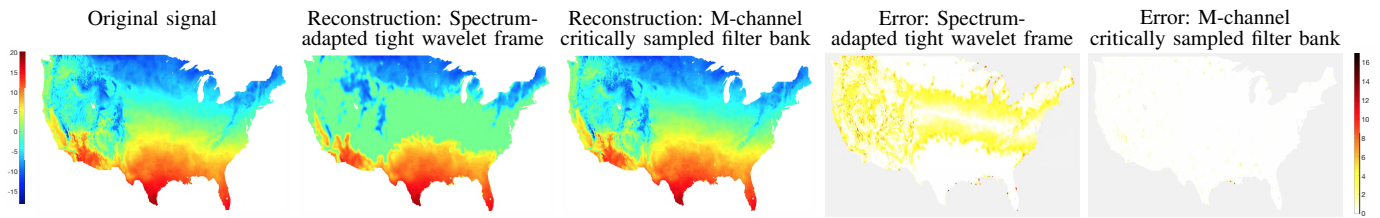


Fig. 8. Compression of the average temperature signal from [16]: the initial signal, the reconstructions from  $\frac{N}{2} = 234,702$  analysis coefficients of two different dictionaries, and the absolute values of the corresponding reconstruction errors. The NMSEs are 0.099 and 0.001 for the redundant tight wavelet frame and the critically sampled filter bank dictionary, respectively.

maximum graph Laplacian eigenvalue and the cumulative spectral density function, and subsequently, how to leverage these approximations in the design and application of the spectral filters, the selection of center vertices via signal-adapted non-uniform random sampling, and fast reconstruction methods.

Open issues and future directions in this line of research include:

- 1) Continued investigation of algorithms that use LSGFDs to efficiently extract information in the context of specific application domains and signal processing and machine learning tasks; as just one example, [91] begins to investigate how to leverage the structured sparsity of LSGFD analysis coefficients to recover piecewise-smooth graph signals in the semi-supervised learning problem.
- 2) New connections between theoretical metrics, dictionary design, and applications. Many unanswered or partially answered questions remain on this front: What mathematical classes of graph signals are sparsely represented by LSGFDs with specific spectral patterns? Are there easily computable metrics on the dictionary that are good predictors of performance in application tasks, as demonstrated either empirically or via mathematical analysis? In what applications is it most beneficial to adapt the filters to the spectrum or the energy density of training signals?
- 3) The development of faster greedy or hybrid greedy/random graph sampling methods that are applicable to both smooth and non-smooth graph signals, as well as accompanying fast, scalable interpolation methods.
- 4) Extensions of the dictionary design principles reviewed here to the settings of data on directed graphs, time series data on graphs, and deep learning / convolutional neural networks on graphs.

#### ACKNOWLEDGMENTS

The author would like to thank the anonymous reviewers and Hamid Behjat for constructive feedback on earlier versions of this article.

#### REFERENCES

- [1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [2] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.
- [3] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 260–291, Mar. 2016.
- [4] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *Proc. IEEE INFOCOM*, vol. 3, Mar. 2003, pp. 1848–1857.
- [5] W. Wang and K. Ramchandran, "Random multiresolution representations for arbitrary sensor network graphs," in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, vol. 4, May 2006, pp. 161–164.
- [6] M. Gavish, B. Nadler, and R. R. Coifman, "Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2010, pp. 367–374.
- [7] S. K. Narang and A. Ortega, "Lifting based wavelet transforms on graphs," in *Proc. APSIPA ASC*, Oct. 2009, pp. 441–444.
- [8] A. D. Szlam, M. Maggioni, R. R. Coifman, and J. C. Bremer, Jr., "Diffusion-driven multiscale analysis on manifolds and graphs: top-down and bottom-up constructions," in *Proc. SPIE Wavelets*, vol. 5914, Aug. 2005, pp. 445–455.
- [9] J. Iriou and N. Saito, "Hierarchical graph Laplacian eigen transforms," *SIAM Letters*, vol. 6, pp. 21–24, 2014.
- [10] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 53–94, 2006.
- [11] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.
- [12] D. I. Shuman, M. J. Faraji, and P. Vandergheynst, "A multiscale pyramid transform for graph signals," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2119–2134, Apr. 2016.
- [13] S. K. Narang and A. Ortega, "Local two-channel critically sampled filterbanks on graphs," in *Proc. Int. Conf. Image Process.*, Sep. 2010, pp. 333–336.
- [14] —, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012.
- [15] V. N. Ekambaram, G. Fanti, B. Ayazifar, and K. Ramchandran, "Critically-sampled perfect-reconstruction spline-wavelet filterbanks for graph signals," in *Proc. IEEE Glob. Conf. Signal and Inform. Process.*, 2013, pp. 475–478.
- [16] S. Li, Y. Jin, and D. I. Shuman, "Scalable  $M$ -channel critically sampled filter banks for graph signals," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 3954–3969, Aug. 2019.
- [17] S. K. Narang and A. Ortega, "Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4673–4685, Oct. 2013.
- [18] D. B. Tay and Z. Lin, "Design of near orthogonal graph filter banks," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 701–704, Jun. 2015.
- [19] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.
- [20] A. Sakiyama, K. Watanabe, and Y. Tanaka, "Spectral graph wavelets and filter banks with low approximation error," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 3, pp. 230–245, 2016.
- [21] A. Anis and A. Ortega, "Critical sampling for wavelet filterbanks on arbitrary graphs," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, 2017, pp. 3889–3893.
- [22] D. B. Tay, Y. Tanaka, and A. Sakiyama, "Critically sampled graph filter banks with polynomial filters from regular domain filter banks," *Signal Process.*, vol. 131, pp. 66–72, 2017.
- [23] —, "Almost tight spectral graph wavelets with polynomial filters," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 812–824, 2017.



- [24] O. Teke and P. P. Vaidyanathan, "Extending classical multirate signal processing theory to graphs – Part II: M-channel filter banks," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 423–437, 2016.
- [25] M. S. Kotzagiannidis and P. L. Dragotti, "Splines and wavelets on circulant graphs," *Appl. Comput. Harmon. Anal.*, vol. 47, no. 2, pp. 481–515, Sep. 2019.
- [26] I. Pesenson, "Variational splines and Paley-Wiener spaces on combinatorial graphs," *Constr. Approx.*, vol. 29, no. 1, pp. 1–21, Feb. 2009.
- [27] W. Erb, "Graph signal interpolation with positive definite graph basis functions," *arXiv preprint arXiv:1912.02069*, 2019.
- [28] D. Thanou, D. I Shuman, and P. Frossard, "Learning parametric dictionaries for signals on graphs," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3849–3862, Aug. 2014.
- [29] H. Behjat, U. Richter, D. Van De Ville, and L. Sörnmo, "Signal-adapted tight frames on graphs," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 6017–6029, Nov. 2016.
- [30] N. Leonardi and D. Van De Ville, "Tight wavelet frames on multislice graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3357–3367, Jul. 2013.
- [31] D. I Shuman, C. Wiesmeyr, N. Holighaus, and P. Vandergheynst, "Spectrum-adapted tight graph wavelet and vertex-frequency frames," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4223–4235, Aug. 2015.
- [32] F. Göbel, G. Blanchard, and U. von Luxburg, "Construction of tight frames on graphs and application to denoising," in *Handbook of Big Data Analytics*. Springer, 2018, pp. 503–522.
- [33] B. Dong, "Sparse representation on graphs by tight wavelet frames and applications," *Appl. Comput. Harmon. Anal.*, vol. 42, no. 3, pp. 452–479, 2017.
- [34] H. Behjat, N. Leonardi, and D. Van De Ville, "Statistical parametric mapping of functional MRI data using wavelets adapted to the cerebral cortex," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2013, pp. 1070–1073.
- [35] H. Behjat, N. Leonardi, L. Sörnmo, and D. Van De Ville, "Canonical cerebellar graph wavelets and their application to fMRI activation mapping," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, 2014, pp. 1039–1042.
- [36] —, "Anatomically-adapted graph wavelets for improved group-level fMRI activation mapping," *NeuroImage*, vol. 123, pp. 185–199, 2015.
- [37] N. Tremblay and P. Borgnat, "Graph wavelets for multiscale community mining," *IEEE Trans. Signal Process.*, vol. 62, pp. 5227–5239, 2014.
- [38] N. Saito, "The first steps toward building natural graph wavelets." Presented at the Graph Signal Processing Workshop, Jun. 2019, <http://math.ucdavis.edu/~saito/talks/gsp19.pdf>.
- [39] Stanford University Computer Graphics Laboratory, "The Stanford 3D Scanning Repository," <http://graphics.stanford.edu/data/3Dscanrep/>.
- [40] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 44–63, May 2019.
- [41] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 16–43, May 2019.
- [42] Y. Zhou and R. C. Li, "Bounding the spectrum of large Hermitian matrices," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 480–493, 2011.
- [43] W. N. Anderson and T. D. Morley, "Eigenvalues of the Laplacian of a graph," *Linear Multilinear Algebra*, vol. 18, no. 2, pp. 141–145, 1985.
- [44] K. C. Das and R. P. Bapat, "A sharp upper bound on the largest Laplacian eigenvalue of weighted graphs," *Lin. Alg. Appl.*, vol. 409, pp. 153–165, Nov. 2005.
- [45] L. Lin, Y. Saad, and C. Yang, "Approximating spectral densities of large matrices," *SIAM Review*, vol. 58, no. 1, pp. 34–65, 2016.
- [46] R. N. Silver, H. Röder, A. F. Voter, and J. D. Kress, "Kernel polynomial approximations for densities of states and spectral functions," *J. Comput. Phys.*, vol. 124, no. 1, pp. 115–130, 1996.
- [47] D. Gleich, "The MatlabBGL Matlab library," [http://www.cs.purdue.edu/homes/dgleich/packages/matlab\\_bgl/index.html](http://www.cs.purdue.edu/homes/dgleich/packages/matlab_bgl/index.html).
- [48] T. A. Davis and Y. Hu, "The University of Florida sparse matrix collection," *ACM Trans. Math. Softw.*, vol. 38, no. 1, pp. 1:1–1:25, 2011.
- [49] J. Kovačević and A. Chebira, "Life beyond bases: The advent of frames (part I)," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 86–104, Jul. 2007.
- [50] O. Christensen, *Frames and Bases*. Birkhauser, 2008.
- [51] J. C. Mason and D. C. Handscomb, *Chebyshev Polynomials*. Chapman and Hall, 2003.
- [52] L. N. Trefethen, *Approximation Theory and Approximation Practice*. SIAM, 2013.
- [53] J. P. Ward, F. J. Narcowich, and J. D. Ward, "Interpolating splines on graphs for data science applications," *arXiv preprint arXiv:1806.10695*, Oct. 2018.
- [54] R. I. Kondor and J. Lafferty, "Diffusion kernels on graphs and other discrete structures," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2002, pp. 315–22.
- [55] N. Saito, "How can we naturally order and organize graph Laplacian eigenvectors?" in *Proc. IEEE Stat. Signal Process. Wkshp.*, 2018, pp. 483–487.
- [56] A. Cloninger and S. Steinerberger, "On the dual geometry of Laplacian eigenfunctions," *Exp. Math.*, pp. 1–11, 2018.
- [57] H. Li and N. Saito, "Metrics of graph Laplacian eigenvectors," in *SPIE Wavelets and Sparsity*, Aug. 2019.
- [58] N. Perraudin, J. Paratte, D. I Shuman, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," <https://lts2.epfl.ch/gsp/>.
- [59] X. Shi, H. Feng, M. Zhai, T. Yang, and B. Hu, "Infinite impulse response graph filters in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1113–1117, Aug. 2015.
- [60] J. Liu, E. Isufi, and G. Leus, "Filter design for autoregressive moving average graph filters," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 1, pp. 47–60, 2019.
- [61] N. J. Higham, *Functions of Matrices*. Society for Industrial and Applied Mathematics, 2008.
- [62] A. Frommer and V. Simoncini, "Matrix functions," in *Model Order Reduction: Theory, Research Aspects and Applications*. Springer, 2008, pp. 275–303.
- [63] D. I Shuman, P. Vandergheynst, D. Kressner, and P. Frossard, "Distributed signal processing via Chebyshev polynomial approximation," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 4, pp. 736–751, 2018.
- [64] I. Daubechies, *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992, vol. 61.
- [65] V. L. Druskin and L. A. Knizhnerman, "Two polynomial methods of calculating functions of symmetric matrices," *U.S.S.R. Comput. Maths. Math. Phys.*, vol. 29, no. 6, pp. 112–121, 1989.
- [66] E. Di Napoli, E. Polizzi, and Y. Saad, "Efficient estimation of eigenvalue counts in an interval," *Numer. Linear Algebra Appl.*, vol. 23, no. 4, pp. 674–692, Aug. 2016.
- [67] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 2013.
- [68] A. Loukas, A. Simonetto, and G. Leus, "Distributed autoregressive moving average graph filters," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1931–1935, Nov. 2015.
- [69] O. Teke and P. Vaidyanathan, "Energy compaction filters on graphs," in *Proc. IEEE Glob. Conf. Signal and Inform. Process.*, 2018, pp. 783–787.
- [70] L. Fan, D. I Shuman, S. Ubaru, and Y. Saad, "Spectrum-adapted polynomial approximation for matrix functions," in *Proc. IEEE Int. Conf. Acc., Speech, and Signal Process.*, May 2019.
- [71] H. Behjat and D. Van De Ville, "Spectral design of signal-adapted tight frames on graphs," in *Vertex-Frequency Analysis of Graph Signals*. Springer, 2019, pp. 177–206.
- [72] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3775–3789, Jul. 2016.
- [73] A. Sakiyama, Y. Tanaka, T. Tanaka, and A. Ortega, "Eigendecomposition-free sampling set selection for graph signals," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2679–2692, Mar. 2019.
- [74] P. Di Lorenzo, S. Barbarossa, and P. Banelli, "Sampling and recovery of graph signals," in *Cooperative and Graph Signal Processing*, 2018, pp. 261–282.
- [75] Y. Tanaka, Y. C. Eldar, A. Ortega, and G. Cheung, "Sampling on graphs: From theory to applications," *arXiv preprint arXiv:2003.03957*, 2020.
- [76] I. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, 2008.
- [77] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Appl. Comput. Harmon. Anal.*, vol. 44, no. 2, pp. 446–475, Mar. 2018.
- [78] Y. Bai, F. Wang, G. Cheung, Y. Nakatsukasa, and W. Gao, "Fast graph sampling set selection using Gershgorin disc alignment," *IEEE Trans. Signal Process.*, vol. 68, pp. 2419–2434, Mar. 2020.
- [79] P. Drineas, M. Magdon-Ismail, M. W. Mahoney, and D. P. Woodruff, "Fast approximation of matrix coherence and statistical leverage," *J. Mach. Learn. Res.*, vol. 13, no. Dec., pp. 3475–3506, 2012.
- [80] M. W. Mahoney and P. Drineas, "CUR matrix decompositions for improved data analysis," *Proc. Natl. Acad. Sci.*, vol. 106, no. 3, pp. 697–702, 2009.
- [81] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

- [82] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4845–4860, 2016.
- [83] N. Perraudin, B. Ricaud, D. I Shuman, and P. Vandergheynst, "Global and local uncertainty principles for signals on graphs," *APSIPA Trans. Signal Inf. Process.*, Apr. 2018.
- [84] D. Van De Ville, R. Demesmaecker, and M. G. Preti, "When Slepian meets Fiedler: Putting a focus on the graph spectrum," *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 1001–1004, 2017.
- [85] O. Teke and P. P. Vaidyanathan, "Uncertainty principles and sparse eigenvectors of graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5406–5420, 2017.
- [86] W. Erb, "Shapes of uncertainty in spectral graph theory," *arXiv preprint arXiv:1909.10865*, Sep. 2019.
- [87] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *Appl. Comput. Harmon. Anal.*, vol. 1, no. 1, pp. 100–115, Dec. 1993.
- [88] S. Chen, A. Singh, and J. Kovačević, "Multiresolution representations for piecewise-smooth signals on graphs," *arXiv preprint arXiv:1803.02944*, Mar. 2018.
- [89] B. Ricaud, D. I Shuman, and P. Vandergheynst, "On the sparsity of wavelet coefficients for signals on graphs," in *SPIE Wavelets and Sparsity*, Aug. 2013.
- [90] B. de Loynes, F. Navarro, and B. Olivier, "Data-driven thresholding in denoising with spectral graph wavelet transform," *arXiv preprint arXiv:1906.01882*, 2019.
- [91] D. I Shuman, M. J. Faraji, and P. Vandergheynst, "Semi-supervised learning with spectral graph wavelets," in *Proc. Int. Conf. Samp. Theory and Appl.*, May 2011.



**David I Shuman** received the B.A. degree in economics and the M.S. degree in engineering-economic systems and operations research from Stanford University, Stanford, CA, in 2001 and the M.S. degree in electrical engineering: systems, the M.S. degree in applied mathematics, and the Ph.D. degree in electrical engineering: systems from the University of Michigan, Ann Arbor, in 2006, 2009, and 2010, respectively.

He is currently an Associate Professor in the Department of Mathematics, Statistics, and Computer Science, Macalester College, St. Paul, Minnesota, which he joined in January 2014. From 2010 to 2013, he was a Postdoctoral Researcher at the Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. His research interests include signal processing on graphs, computational harmonic analysis, and stochastic scheduling and resource allocation problems.

Dr. Shuman has served as an Associate Editor for the IEEE Transactions on Signal and Information Processing Over Networks (2019-) and the IEEE Signal Processing Letters (2017-2019), and on the Technical Program Committee for the IEEE Global Conference on Signal and Information Processing (2015-2018). He received the 2016 IEEE Signal Processing Magazine Best Paper Award, and was a 2014-2015 Project NEXt Fellow of the Mathematical Association of America.