

Discrete Signal Processing on Graphs: Sampling Theory

Siheng Chen, Rohan Varma, Aliaksei Sandryhaila, Jelena Kovačević

arXiv:1503.05432v2 [cs.IT] 8 Aug 2015

Abstract—We propose a sampling theory for signals that are supported on either directed or undirected graphs. The theory follows the same paradigm as classical sampling theory. We show that perfect recovery is possible for graph signals bandlimited under the graph Fourier transform. The sampled signal coefficients form a new graph signal, whose corresponding graph structure preserves the first-order difference of the original graph signal. For general graphs, an optimal sampling operator based on experimentally designed sampling is proposed to guarantee perfect recovery and robustness to noise; for graphs whose graph Fourier transforms are frames with maximal robustness to erasures as well as for Erdős-Rényi graphs, random sampling leads to perfect recovery with high probability. We further establish the connection to the sampling theory of finite discrete-time signal processing and previous work on signal recovery on graphs. To handle full-band graph signals, we propose a graph filter bank based on sampling theory on graphs. Finally, we apply the proposed sampling theory to semi-supervised classification on online blogs and digit images, where we achieve similar or better performance with fewer labeled samples compared to previous work.

Index Terms—Discrete signal processing on graphs, sampling theory, experimentally designed sampling, compressed sensing

I. INTRODUCTION

With the explosive growth of information and communication, signals are generated at an unprecedented rate from various sources, including social, citation, biological, and physical infrastructure [1], [2], among others. Unlike time-series signals or images, these signals possess complex, irregular structure, which requires novel processing techniques leading to the emerging field of signal processing on graphs [3], [4].

Signal processing on graphs extends classical discrete signal processing to signals with an underlying complex, irregular structure. The framework models that underlying structure by a graph and signals by graph signals, generalizing concepts and tools from classical discrete signal processing to graph signal processing. Recent work includes graph-based filtering [5], [6], [7], graph-based transforms [5], [8], [9], sampling and interpolation on graphs [10], [11], [12], uncertainty principle on graphs [13], semi-supervised classification on graphs [14], [15], [16], graph dictionary learning [17], [18], denoising [6], [19], community detection and clustering on graphs [20], [21], [22], graph signal recovery [23], [24], [25] and distributed algorithms [26], [27].

S. Chen and R. Varma are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA. Email: sihengc@andrew.cmu.edu, rohanv@andrew.cmu.edu. A. Sandryhaila is with HP Vertica, Pittsburgh, PA, 15203 USA. Email: aliaksei.sandryhaila@hp.com. J. Kovačević is with the Departments of Electrical and Computer Engineering and Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA. Email: jelenak@cmu.edu.

Two basic approaches to signal processing on graphs have been considered: The first is rooted in the *spectral graph theory* and builds upon the *graph Laplacian matrix* [3]. Since the standard graph Laplacian matrix is restricted to be symmetric and positive semi-definite, this approach is applicable only to undirected graphs with real and nonnegative edge weights. The second approach, *discrete signal processing on graphs* (DSP_G) [5], [28], is rooted in the *algebraic signal processing theory* [29], [30] and builds upon the graph shift operator, which works as the elementary operator that generates all linear shift-invariant filters for signals with a given structure. The graph shift operator is the adjacency matrix and represents the relational dependencies between each pair of nodes. Since the graph shift is not restricted to be symmetric, the corresponding framework is applicable to arbitrary graphs, those with undirected or directed edges, with real or complex, nonnegative or negative weights. Both frameworks analyze signals with complex, irregular structure, generalizing a series of concepts and tools from classical signal processing, such as graph filters, graph Fourier transform, to diverse graph-based applications.

In this paper, we consider the classical signal processing task of sampling and interpolation within the framework of DSP_G [31], [32]. As the bridge connecting sequences and functions, classical sampling theory shows that a bandlimited function can be perfectly recovered from its sampled sequence if the sampling rate is high enough [33]. More generally, we can treat any decrease in dimension via a linear operator as sampling, and, conversely, any increase in dimension via a linear operator as interpolation [31], [34]. Formulating a sampling theory in this context is equivalent to moving between higher- and lower-dimensional spaces.

A sampling theory for graphs has interesting applications. For example, given a graph representing friendship connectivity on Facebook, we can sample a fraction of users and query their hobbies and then recover all users' hobbies. The task of sampling on graphs is, however, not well understood [11], [12], because graph signals lie on complex, irregular structures. It is even more challenging to find a graph structure that is associated with the sampled signal coefficients; in the Facebook example, we sample a small fraction of users and an associated graph structure would allow us to infer new connectivity between those sampled users, even when they are not directly connected in the original graph.

Previous works on sampling theory [10], [12], [35] consider graph signals that are uniquely sampled onto a given subset of nodes. This approach is hard to apply to directed graphs. It also does not explain which graph structure supports these sampled coefficients.

In this paper, we propose a sampling theory for signals that

are supported on either directed or undirected graphs. Perfect recovery is possible for graph signals bandlimited under the graph Fourier transform. We also show that the sampled signal coefficients form a new graph signal whose corresponding graph structure is constructed from the original graph structure. The proposed sampling theory follows Chapter 5 from [31] and is consistent with classical sampling theory.

We call a sampling operator that leads to perfect recovery a *qualified sampling operator*. We show that for general graphs, an optimal sampling operator based on experimentally designed sampling is proposed to guarantee perfect recovery and robustness to noise; for graphs whose graph Fourier transforms are frames with maximal robustness to erasures as well as for Erdős-Rényi graphs, random sampling leads to perfect recovery with high probability. We further establish the connection to sampling theory of finite discrete-time signal processing and previous works on sampling theory on graphs. To handle full-band graph signals, we propose graph filter banks to force graph signals to be bandlimited. Finally, we apply the proposed sampling theory to semi-supervised classification of online blogs and digit images, where we achieve similar or better performance with fewer labeled samples compared to the previous works.

Contributions. The main contributions of the paper are as follows:

- A novel framework for sampling a graph signal, which solves complex sampling problems by using simple tools from linear algebra;
- A novel approach for sampling a graph by preserving the first-order difference of the original graph signal;
- A novel approach for designing a sampling operator on graphs.

Outline of the paper. Section II formulates the problem and briefly reviews DSP_G, which lays the foundation for this paper; Section III describes the proposed sampling theory for graph signals, and the proposed construction of graph structures for the sampled signal coefficients; Section IV studies the qualified sampling operator, including random sampling and experimentally designed sampling; Section V discusses the relations to previous works and extends the sampling framework to the design of graph filter banks; Section VI shows the application to semi-supervised learning; Section VII concludes the paper and provides pointers to future directions.

II. DISCRETE SIGNAL PROCESSING ON GRAPHS

In this section, we briefly review relevant concepts of discrete signal processing on graphs; a thorough introduction can be found in [4], [28]. It is a theoretical framework that generalizes classical discrete signal processing from regular domains, such as lines and rectangular lattices, to irregular structures that are commonly described by graphs.

A. Graph Shift

Discrete signal processing on graphs studies signals with complex, irregular structure represented by a graph $G = (\mathcal{V}, \mathcal{A})$, where $\mathcal{V} = \{v_0, \dots, v_{N-1}\}$ is the set of nodes and $\mathcal{A} \in \mathbb{C}^{N \times N}$ is the *graph shift*, or a weighted adjacency matrix.

It represents the connections of the graph G , which can be either directed or undirected (note that the standard graph Laplacian matrix can only represent undirected graphs [3]). The edge weight $A_{n,m}$ between nodes v_n and v_m is a quantitative expression of the underlying relation between the n th and the m th node, such as similarity, dependency, or a communication pattern. To guarantee that the shift operator is properly scaled, we normalize the graph shift \mathcal{A} to satisfy $|\lambda_{\max}(\mathcal{A})| = 1$.

B. Graph Signal

Given the graph representation $G = (\mathcal{V}, \mathcal{A})$, a *graph signal* is defined as the map on the graph nodes that assigns the signal coefficient $x_n \in \mathbb{C}$ to the node v_n . Once the node order is fixed, the graph signal can be written as a vector

$$\mathbf{x} = [x_0 \ x_1 \ \dots \ x_{N-1}]^T \in \mathbb{C}^N, \quad (1)$$

where the n th signal coefficient corresponds to node v_n .

C. Graph Fourier Transform

In general, a Fourier transform corresponds to the expansion of a signal using basis elements that are invariant to filtering; here, this basis is the eigenbasis of the graph shift \mathcal{A} (or, if the complete eigenbasis does not exist, the Jordan eigenbasis of \mathcal{A}). For simplicity, assume that \mathcal{A} has a complete eigenbasis and the spectral decomposition of \mathcal{A} is [31]

$$\mathcal{A} = \mathbf{V} \Lambda \mathbf{V}^{-1}, \quad (2)$$

where the eigenvectors of \mathcal{A} form the columns of matrix \mathbf{V} , and $\Lambda \in \mathbb{C}^{N \times N}$ is the diagonal matrix of corresponding eigenvalues $\lambda_0, \dots, \lambda_{N-1}$ of \mathcal{A} . These eigenvalues represent frequencies on the graph [28]. We do not specify the ordering of graph frequencies here and we will explain why later.

Definition 1. The *graph Fourier transform* of $\mathbf{x} \in \mathbb{C}^N$ is

$$\widehat{\mathbf{x}} = \mathbf{V}^{-1} \mathbf{x}. \quad (3)$$

The *inverse graph Fourier transform* is

$$\mathbf{x} = \mathbf{V} \widehat{\mathbf{x}}.$$

The vector $\widehat{\mathbf{x}}$ in (3) represents the signal's expansion in the eigenvector basis and describes the frequency content of the graph signal \mathbf{x} . The inverse graph Fourier transform reconstructs the graph signal from its frequency content by combining graph frequency components weighted by the coefficients of the signal's graph Fourier transform.

III. SAMPLING ON GRAPHS

Previous works on sampling theory of graph signals is based on spectral graph theory [12]. The bandwidth of graph signals is defined based on the value of graph frequencies, which correspond to the eigenvalues of the graph Laplacian matrix. Since each graph has its own graph frequencies, it is hard in practice to specify a general cut-off graph frequency; it is also computationally inefficient to compute all the values of graph frequencies, especially for large graphs.

In this section, we propose a novel sampling framework for graph signals. Here, the bandwidth definition is based on the

Symbol	Description	Dimension
A	graph shift	$N \times N$
x	graph signal	N
V^{-1}	graph Fourier transform matrix	$N \times N$
\hat{x}	graph signal in the frequency domain	N
Ψ	sampling operator	$M \times N$
Φ	interpolation operator	$N \times M$
\mathcal{M}	sampled indices	$N \times M$
$x_{\mathcal{M}}$	sampled signal coefficients of x	M
$\hat{x}_{(K)}$	first K coefficients of \hat{x}	K
$V_{(K)}$	first K columns of V	$N \times K$

TABLE I: Key notation used in the paper.

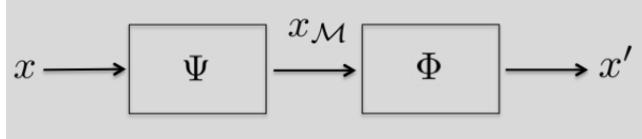


Fig. 1: Sampling followed by interpolation.

number of non-zero signal coefficients in the graph Fourier domain. Since each signal coefficient in the graph Fourier domain corresponds to a graph frequency, the bandwidth definition is also based on the number of graph frequencies. This makes the proposed sampling framework strongly connected to linear algebra, that is, we are allowed to use simple tools from linear algebra to perform sampling on complex, irregular graphs.

A. Sampling & Interpolation

Suppose that we want to sample M coefficients of a graph signal $x \in \mathbb{C}^N$ to produce a sampled signal $x_{\mathcal{M}} \in \mathbb{C}^M$ ($M < N$), where $\mathcal{M} = (\mathcal{M}_0, \dots, \mathcal{M}_{M-1})$ denotes the sequence of *sampled* indices, and $\mathcal{M}_i \in \{0, 1, \dots, N-1\}$. We then interpolate $x_{\mathcal{M}}$ to get $x' \in \mathbb{C}^N$, which recovers x either exactly or approximately. The sampling operator Ψ is a linear mapping from \mathbb{C}^N to \mathbb{C}^M , defined as

$$\Psi_{i,j} = \begin{cases} 1, & j = \mathcal{M}_i; \\ 0, & \text{otherwise}, \end{cases} \quad (4)$$

and the interpolation operator Φ is a linear mapping from \mathbb{C}^M to \mathbb{C}^N (see Figure 1),

$$\begin{aligned} \text{sampling : } & x_{\mathcal{M}} = \Psi x \in \mathbb{C}^M, \\ \text{interpolation : } & x' = \Phi x_{\mathcal{M}} = \Phi \Psi x \in \mathbb{C}^N, \end{aligned}$$

where $x' \in \mathbb{R}^N$ recovers x either exactly or approximately. We consider two sampling strategies: *random sampling* means that sample indices are chosen from $\{0, 1, \dots, N-1\}$ independently and randomly; and *experimentally design sampling* means that sample indices can be chosen beforehand. It is clear that random sampling is a subset of experimentally design sampling.

Perfect recovery happens for all x when $\Phi\Psi$ is the identity matrix. This is not possible in general because $\text{rank}(\Phi\Psi) \leq M < N$; it is, however, possible to do this for signals with specific structure that we will define as bandlimited graph signals, as in classical discrete signal processing.

B. Sampling Theory for Graph Signals

We now define a class of bandlimited graph signals, which makes perfect recovery possible.

Definition 2. A graph signal is called *bandlimited* when there exists a $K \in \{0, 1, \dots, N-1\}$ such that its graph Fourier transform \hat{x} satisfies

$$\hat{x}_k = 0 \quad \text{for all } k \geq K.$$

The smallest such K is called the *bandwidth* of x . A graph signal that is not bandlimited is called a *full-band graph signal*.

Note that the bandlimited graph signals here do not necessarily mean low-pass, or smooth. Since we do not specify the ordering of frequencies, we can reorder the eigenvalues and permute the corresponding eigenvectors in the graph Fourier transform matrix to choose any band in the graph Fourier domain. The bandlimited graph signals are smooth only when we sort the eigenvalues in a descending order. The bandlimited restriction here is equivalent to limiting the number of non-zero signal coefficients in the graph Fourier domain with known supports. This generalization is potentially useful to represent non-smooth graph signals.

Definition 3. The set of graph signals in \mathbb{C}^N with bandwidth of at most K is a closed subspace denoted $\text{BL}_K(V^{-1})$, with V^{-1} as in (2).

When defining the bandwidth, we focus on the number of graph frequencies, while previous works [12] focus on the value of graph frequencies. There are two shortcomings to using the values of graph frequencies: (a) When considering the values of graph frequencies, we ignore the discrete nature of graphs; because graph frequencies are discrete, two cut-off graph frequencies on the same graph can lead to the same bandlimited space. For example, assume a graph has graph frequencies 0, 0.1, 0.4, 0.6 and 2; when we set the cut-off frequency to either 0.2 or 0.3, they lead to the same bandlimited space; (b) The values of graph frequencies cannot be compared between different graphs. Since each graph has its own graph frequencies, a same value of the cut-off graph frequency on two graphs can mean different things. For example, one graph has graph frequencies as 0, 0.1, 0.2, 0.4 and 2, and another has graph frequencies 0, 1.1, 1.6, 1.8, and 2; when we set the cut-off frequency to 1, that is, we preserve all the graph frequencies that are no greater than 1, first graph preserves three out of four graph frequencies and the second graph only preserves one out of four. The values of graph frequencies thus do not necessarily give a direct and intuitive understanding about the bandlimited space. Another key advantage of using the number of graph frequencies is to build a strong connection to linear algebra allowing for the use of simple tools from linear algebra in sampling and interpolation of bandlimited graph signals.

In Theorem 5.2 in [31], the authors show the recovery for vectors via projection, which lays the theoretical foundation for the classical sampling theory. Following the theorem, we obtain the following result, the proof of which can be found in [34].

Theorem 1. Let Ψ satisfy

$$\text{rank}(\Psi V_{(K)}) = K,$$

where $V_{(K)} \in \mathbb{R}^{N \times K}$ denotes the first K columns of V . For all $x \in \text{BL}_K(V^{-1})$, perfect recovery, $x = \Phi\Psi x$, is achieved by choosing

$$\Phi = V_{(K)} U,$$

with $U\Psi V_{(K)}$ a $K \times K$ identity matrix.

Theorem 1 is applicable for all graph signals that have a few non-zero elements in the graph Fourier domain with known supports, that is, $K < N$.

Similarly to the classical sampling theory, the sampling rate has a lower bound for graph signals as well, that is, the sample size M should be no smaller than the bandwidth K . When $M < K$, $\text{rank}(U\Psi V_{(K)}) \leq \text{rank}(U) \leq M < K$, and thus, $U\Psi V_{(K)}$ can never be an identity matrix. For $U\Psi V_{(K)}$ to be an identity matrix, U is the inverse of $\Psi V_{(K)}$ when $M = K$; it is a pseudo-inverse of $\Psi V_{(K)}$ when $M > K$, where the redundancy can be useful for reducing the influence of noise. For simplicity, we only consider $M = K$ and U invertible. When $M > K$, we simply select K out of M sampled signal coefficients to ensure that the sample size and the bandwidth are the same.

From Theorem 1, we see that an arbitrary sampling operator may not lead to perfect recovery even for bandlimited graph signals. When the sampling operator Ψ satisfies the full-rank assumption (5), we call it a *qualified sampling operator*. To satisfy (5), the sampling operator should select at least one set of K linearly-independent rows in $V_{(K)}$. Since V is invertible, the column vectors in V are linearly independent and $\text{rank}(V_{(K)}) = K$ always holds; in other words, at least one set of K linearly-independent rows in $V_{(K)}$ always exists. Since the graph shift A is given, one can find such a set independently of the graph signal. Given such a set, Theorem 1 guarantees perfect recovery of bandlimited graph signals. To find linearly-independent rows in a matrix, fast algorithms exist, such as QR decomposition; see [36], [31]. Since we only need to know the graph structure to design a *qualified sampling operator*, this follows the experimentally designed sampling. We will expand this topic in Section IV.

C. Sampled Graph Signal

We just showed that perfect recovery is possible when the graph signal is bandlimited. We now show that the sampled signal coefficients form a new graph signal, whose corresponding graph shift can be constructed from the original graph shift.

Although the following results can be generalized to $M > K$ easily, we only consider $M = K$ for simplicity. Let the sampling operator Ψ and the interpolation operator Φ satisfy the conditions in Theorem 1. For all $x \in \text{BL}_K(V^{-1})$, we have

$$\begin{aligned} x &= \Phi\Psi x = \Phi x_M \stackrel{(a)}{=} V_{(K)} U x_M \\ &\stackrel{(b)}{=} V_{(K)} \widehat{x}_{(K)}, \end{aligned}$$

where $\widehat{x}_{(K)}$ denotes the first K coefficients of \widehat{x} , (a) follows from Theorem 1 and (b) from Definition 2. We thus get

$$\widehat{x}_{(K)} = U x_M,$$

and

$$x_M = U^{-1} U x_M = U^{-1} \widehat{x}_{(K)}.$$

From what we have seen, the sampled signal coefficients x_M and the frequency content $\widehat{x}_{(K)}$ form a Fourier pair because x_M can be constructed from $\widehat{x}_{(K)}$ through U^{-1} and $\widehat{x}_{(K)}$ can also be constructed from x_M through U . This implies that, according to Definition 1 and the spectral decomposition (2), x_M is a graph signal associated with the graph Fourier transform matrix U and a new graph shift

$$A_M = U^{-1} \Lambda_{(K)} U \in \mathbb{C}^{K \times K},$$

where $\Lambda_{(K)} \in \mathbb{C}^{K \times K}$ is a diagonal matrix that samples the first K eigenvalues of Λ . This leads to the following theorem.

Theorem 2. Let $x \in \text{BL}_K(V^{-1})$ and let

$$x_M = \Psi x \in \mathbb{C}^K$$

be its sampled version, where Ψ is a qualified sampling operator. Then, the graph shift associated with the graph signal x_M is

$$A_M = U^{-1} \Lambda_{(K)} U \in \mathbb{C}^{K \times K}, \quad (5)$$

with $U = (\Psi V_{(K)})^{-1}$.

From Theorem 2, we see that the graph shift A_M is constructed by sampling the rows of the eigenvector matrix and sampling the first K eigenvalues of the original graph shift A . We simply say that A_M is sampled from A , preserving certain information in the graph Fourier domain.

Since the bandwidth of x is K , the first K coefficients in the frequency domain are $\widehat{x}_{(K)} = \widehat{x}_M$, and the other $N - K$ coefficients are $\widehat{x}_{(-K)} = 0$; in other words, the frequency contents of the original graph signal x and the sampled graph signal x_M are equivalent after performing their corresponding graph Fourier transforms.

Similarly to Theorem 1, by reordering the eigenvalues and permuting the corresponding eigenvectors in the graph Fourier transform matrix, Theorem 2 is applicable to all graph signals that have limited support in the graph Fourier domain.

D. Property of A Sampled Graph Signal

We argued that $A_M = U^{-1} \Lambda_{(K)} U$ is the graph shift that supports the sampled signal coefficients x_M following from a mathematical equivalence between the graph Fourier transform for the sampled graph signal and the graph shift. We, in fact, implicitly proposed an approach to sampling graphs. Since sampled graphs always lose information, we now study which information A_M preserves.

Theorem 3. For all $x \in \text{BL}_K(V^{-1})$,

$$x_M - A_M x_M = \Psi(x - Ax).$$

Proof.

$$\begin{aligned} x_M - A_M x_M &= U^{-1} \widehat{x}_M - U^{-1} \Lambda_{(K)} U U^{-1} \widehat{x}_M \\ &= \Psi V_{(K)} (I - \Lambda_{(K)}) \widehat{x}_{(K)} \\ &= \Psi(x - Ax), \end{aligned}$$

where the last equality follows from $x \in \text{BL}_K(V^{-1})$. \square

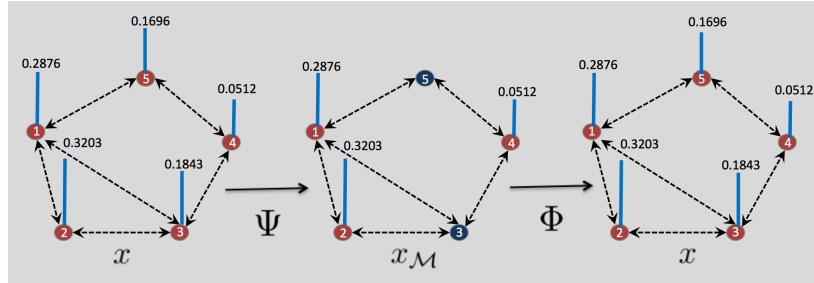


Fig. 2: Sampling followed by interpolation. The arrows indicate that the edges are directed.

The term $x - Ax$ measures the difference between the original graph signal and its shifted version. This is also called the first-order difference of x , while the term $\Psi(x - Ax)$ measures the first-order difference of x at sampled indices. Furthermore, $\|x - Ax\|_p^p$ is the graph total variation based on the ℓ_p -norm, which is a quantitative characteristic that measures the smoothness of a graph signal [28]. When using a sampled graph to represent the sampled signal coefficients, we lose the connectivity information between the sampled nodes and all the other nodes; despite this, $A_{\mathcal{M}}$ still preserves the first-order difference at sampled indices. Instead of focusing on preserving connectivity properties as in prior work [37], we emphasize the interplay between signals and structures.

E. Example

We consider a five-node directed graph with graph shift

$$A = \begin{bmatrix} 0 & \frac{2}{5} & \frac{2}{5} & 0 & \frac{1}{5} \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}.$$

The corresponding inverse graph Fourier transform matrix is

$$V = \begin{bmatrix} 0.45 & 0.19 & 0.25 & 0.35 & -0.40 \\ 0.45 & 0.40 & 0.16 & -0.74 & 0.18 \\ 0.45 & 0.08 & -0.56 & 0.29 & 0.36 \\ 0.45 & -0.66 & -0.41 & -0.47 & -0.57 \\ 0.45 & -0.60 & 0.66 & 0.13 & 0.59 \end{bmatrix},$$

and the frequencies are

$$\Lambda = \text{diag}[1 \ 0.39 \ -0.12 \ -0.44 \ -0.83].$$

Let $K = 3$; generate a bandlimited graph signal $x \in \text{BL}_3(V^{-1})$ as

$$\hat{x} = [0.5 \ 0.2 \ 0.1 \ 0 \ 0]^T,$$

with

$$x = [0.29 \ 0.32 \ 0.18 \ 0.05 \ 0.17]^T,$$

and the first-order difference of x is

$$x - Ax = [0.05 \ 0.07 \ -0.05 \ -0.13 \ 0.0002]^T.$$

We can check the first three columns of V to see that all sets of three rows are independent. According to the sampling theorem, we can then recover x perfectly by sampling any three of its coefficients; for example, sample the first, second

and the fourth coefficients. Then, $\mathcal{M} = (1, 2, 4)$, $x_{\mathcal{M}} = [0.29 \ 0.32 \ 0.05]^T$, and the sampling operator

$$\Psi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

is qualified. We recover x by using the following interpolation operator (see Figure 2)

$$\Phi = V_{(3)}(\Psi V_{(3)})^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2.7 & 2.87 & 0.83 \\ 0 & 0 & 1 \\ 5.04 & -3.98 & -0.05 \end{bmatrix}.$$

The inverse graph Fourier transform matrix for the sampled signal is

$$U^{-1} = \Psi V_{(3)} = \begin{bmatrix} 0.45 & 0.19 & 0.25 \\ 0.45 & 0.40 & 0.16 \\ 0.45 & -0.66 & -0.41 \end{bmatrix},$$

and the sampled frequencies are

$$\Lambda_{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.39 & 0 \\ 0 & 0 & -0.12 \end{bmatrix}.$$

The sampled graph shift is then constructed as

$$A_{\mathcal{M}} = U^{-1} \Lambda_{(3)} U = \begin{bmatrix} 0.07 & 0.75 & 0.32 \\ -0.23 & 0.96 & 0.28 \\ 1.17 & -0.56 & 0.39 \end{bmatrix}.$$

The first-order difference of $x_{\mathcal{M}}$ is

$$x_{\mathcal{M}} - A_{\mathcal{M}} x_{\mathcal{M}} = [0.05 \ 0.07 \ -0.13]^T = \Psi(x - Ax).$$

We see that the sampled graph shift contains self-loops and negative weights and seems to be dissimilar to A , but $A_{\mathcal{M}}$ preserves a part of the frequency content of A because U^{-1} is sampled from V and $\Lambda_{(3)}$ is sampled from A . $A_{\mathcal{M}}$ also preserves the first-order difference of x , which validates Theorem 3.

IV. SAMPLING WITH A QUALIFIED SAMPLING OPERATOR

As shown in Section III, only a qualified sampling operator (5) can lead to perfect recovery for bandlimited graph signals. Since a qualified sampling operator (5) is designed via the graph structure, it belongs to experimentally designed sampling. The design consists in finding K linearly independent rows in $V_{(K)}$, which gives multiple choices. In this section, we

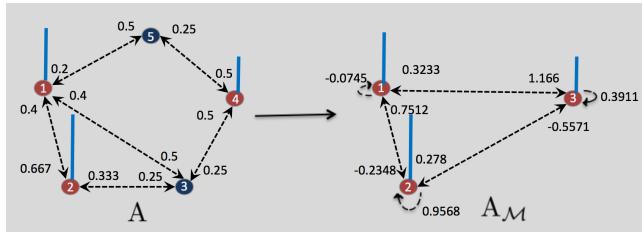


Fig. 3: Sampling a graph.

propose an optimal approach to designing a qualified sampling operators by minimizing the effect of noise for general graphs. We then show that for some specific graphs, random sampling also leads to perfect recovery with high probability.

A. Experimentally Designed Sampling

We now show how to design a qualified sampling operator on any given graph that is robust to noise. We then compare this optimal sampling operator with a random sampling operator on a sensor network.

1) Optimal Sampling Operator: As mentioned in Section III-B, at least one set of K linearly-independent rows in $V_{(K)}$ always exists. When we have multiple choices of K linearly-independent rows, we aim to find the optimal one to minimize the effect of noise.

We consider a model where noise e is introduced during sampling as follows,

$$x_{\mathcal{M}} = \Psi x + e,$$

where Ψ is a qualified sampling operator. The recovered graph signal, x'_e , is then

$$x'_e = \Phi x_{\mathcal{M}} = \Phi \Psi x + \Phi e = x + \Phi e.$$

To bound the effect of noise, we have

$$\begin{aligned} \|x' - x\|_2 &= \|\Phi e\|_2 = \|V_{(K)} U e\|_2 \\ &\leq \|V_{(K)}\|_2 \|U\|_2 \|e\|_2, \end{aligned}$$

where the inequality follows from the definition of the spectral norm. Since $\|V_{(K)}\|_2$ and $\|e\|_2$ are fixed, we want U to have a small spectral norm. From this perspective, for each feasible Ψ , we compute the inverse or pseudo-inverse of $\Psi V_{(K)}$ to obtain U ; the best choice comes from the U with the smallest spectral norm. This is equivalent to maximizing the smallest singular value of $\Psi V_{(K)}$,

$$\Psi^{opt} = \arg \max_{\Psi} \sigma_{\min}(\Psi V_{(K)}), \quad (6)$$

where σ_{\min} denotes the smallest singular value. The solution of (6) is optimal in terms of minimizing the effect of noise; we simply call it *optimal sampling operator*. Since we restrict the form of Ψ in (4), (6) is non-deterministic polynomial-time hard. To solve (6), we can use a greedy algorithm as shown in Algorithm 1. In a previous work, the authors solved a similar optimization problem for matrix approximation and showed that the greedy algorithm gives a good approximation to the global optimum [38]. Note that \mathcal{M} is the sampling sequence, indicating which rows to select, and $(V_{(K)})_{\mathcal{M}}$ denotes the sampled rows from $V_{(K)}$. When increasing the number of

samples, the smallest singular value of $\Psi V_{(K)}$ grows, and thus, redundant samples make the algorithm robust to noise.

Algorithm 1 Optimal Sampling Operator via Greedy Algorithm

Input $V_{(K)}$ the first K columns of V
Output M the number of samples
 \mathcal{M} sampling set

Function

```

while  $|\mathcal{M}| < M$ 
     $m = \arg \max_i \sigma_{\min}((V_{(K)})_{\mathcal{M} + \{m\}})$ 
     $\mathcal{M} \leftarrow \mathcal{M} + \{m\}$ 
end
return  $\mathcal{M}$ 

```

2) Simulations: We consider 150 weather stations in the United States that record local temperatures [5]. The graph representing these weather stations is obtained by assigning an edge when the geodesic distance between each pair of weather stations is smaller than 500 miles, that is, the graph shift A is formed as

$$A_{i,j} = \begin{cases} 1, & \text{when } 0 < d_{i,j} < 500; \\ 0, & \text{otherwise,} \end{cases}$$

where $d_{i,j}$ is the geodesic distance between the i th and the j th weather stations.

We simulate a graph signal with bandwidth of 3 as

$$x = \mathbf{v}_1 + 0.5\mathbf{v}_2 + 2\mathbf{v}_3,$$

where \mathbf{v}_i is the i th column in V . We design two sampling operators to recover x : an arbitrary qualified sampling operator and the optimal sampling operator. We know that both of them recover x perfectly given 3 samples. To validate the robustness to noise, we add Gaussian noise with mean zero and variance 0.01 to each sample. We recover the graph signal from its samples by using the interpolation operator (5).

Figure 4 (f) and (h) show the recovered graph signal from each of these two sampling operators. We see that an arbitrary qualified sampling operator is not robust to noise and fails to recover the original graph signal, while the optimal sampling operator is robust to noise and approximately recovers the original graph signal.

B. Random Sampling

Previously, we showed that we need to design a qualified sampling operator to achieve perfect recovery. We now show that, for some specific graphs, random sampling leads to perfect recovery with high probabilities.

1) Frames with Maximal Robustness to Erasures: A frame $\{f_0, f_1, \dots, f_{N-1}\}$ is a generating system for \mathbb{C}^K with $N \geq K$, when there exist two constants $a > 0$ and $b < \infty$, such that for all $x \in \mathbb{C}^N$,

$$a \|x\|^2 \leq \sum_k |f_k^T x|^2 \leq b \|x\|^2.$$

In finite dimensions, we represent the frame as an $N \times K$ matrix with rows f_k^T . The frame is *maximally robust to erasures* when every $K \times K$ submatrix (obtained by deleting $N - K$

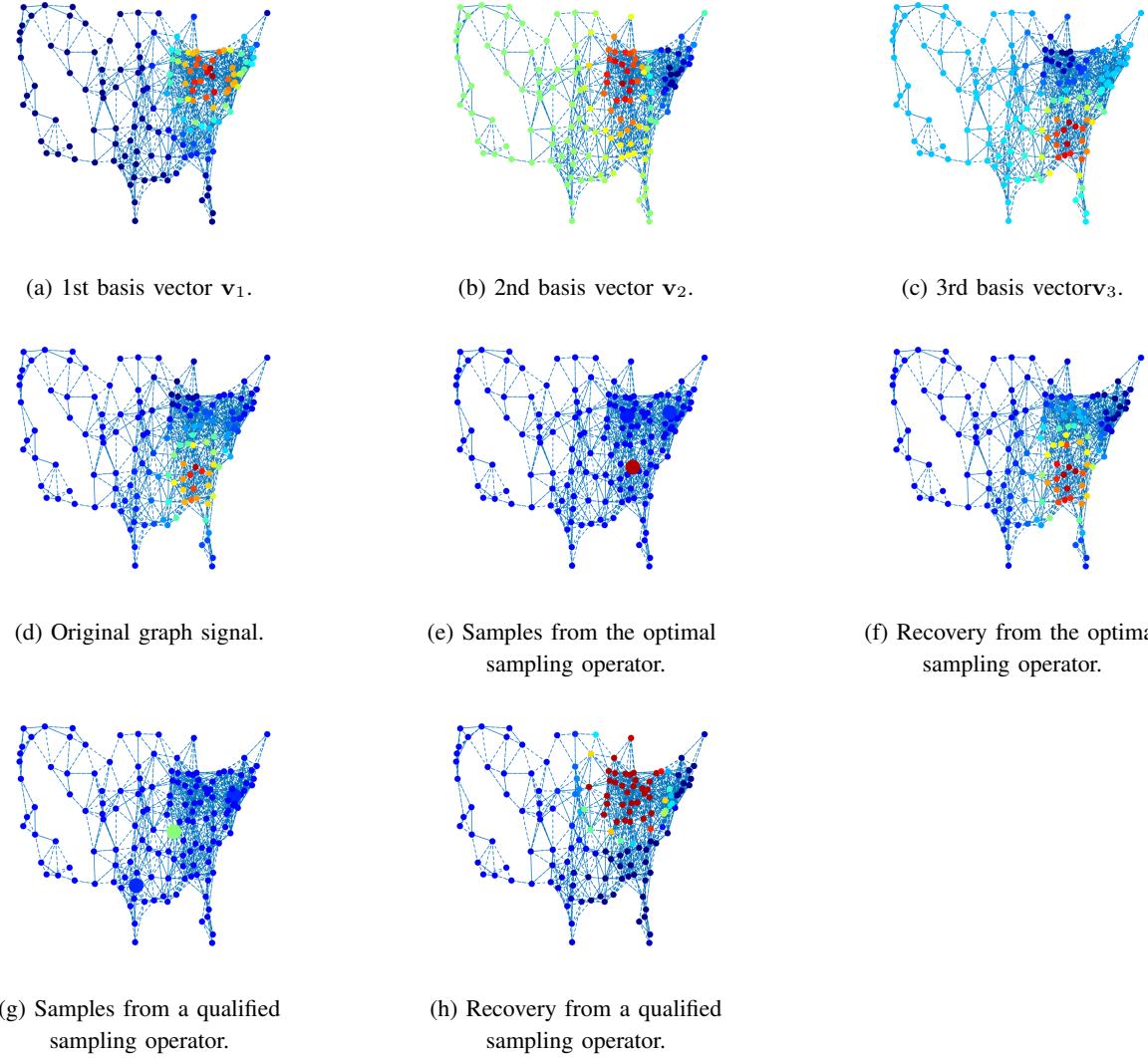


Fig. 4: Graph signals on the sensor network. Colors indicate values of signal coefficients. Red represents high values, blue represents low values. Big nodes in (e) and (g) represent the sampled nodes in each case.

rows) is invertible [39]. In [39], the authors show that a polynomial transform matrix is one example of a frame maximally robust to erasures; in [40], the authors show that many lapped orthogonal transforms and lapped tight frame transforms are also maximally robust to erasures. It is clear that if the inverse graph Fourier transform matrix V as in (2) is maximally robust to erasures, any sampling operator that samples at least K signal coefficients guarantees perfect recovery; in other words, when a graph Fourier transform matrix happens to be a polynomial transform matrix, sampling any K signal coefficients leads to perfect recovery.

For example, a circulant graph is a graph whose adjacency matrix is circulant [41]. The circulant graph shift, C , can be represented as a polynomial of the cyclic permutation matrix, A . The graph Fourier transform of the cyclic permutation matrix is the discrete Fourier transform, which is again a

polynomial transform matrix. As described above, we have

$$\begin{aligned} C &= \sum_{i=0}^{L-1} h_i A^i = \sum_{i=0}^{L-1} h_i (F^* \Lambda F)^i \\ &= F^* \left(\sum_{i=0}^{L-1} h_i \Lambda^i \right) F, \end{aligned}$$

where L is the order of the polynomial, and h_i is the coefficient corresponding to the i th order. Since the graph Fourier transform matrix of a circulant graph is the discrete Fourier transform matrix, we can perfectly recover a circulant-graph signal with bandwidth K by sampling any $M \geq K$ signal coefficients as shown in Theorem 1. In other words, perfect recovery is guaranteed when we randomly sample a sufficient number of signal coefficients.

2) *Erdős-Rényi Graph*: An Erdős-Rényi graph is constructed by connecting nodes randomly, where each edge is included in the graph with probability p independent of any other edge [1], [2]. We aim to show that by sampling K signal

coefficients randomly, the singular values of the corresponding $\Psi V_{(K)}$ are bounded.

Lemma 1. Let a graph shift $A \in \mathbb{R}^{N \times N}$ represent an Erdős-Rényi graph, where each pair of vertices is connected randomly and independently with probability $p = g(N) \log(N)/N$, and $g(\cdot)$ is some positive function. Let V be the eigenvector matrix of A with $V^T V = NI$, and let the number of sampled coefficients satisfy

$$M \geq K \frac{\log^{2.2} g(N) \log(N)}{p} \max(C_1 \log K, C_2 \log \frac{3}{\delta}),$$

for some positive constants C_1, C_2 . Then,

$$P \left(\left\| \frac{1}{M} (\Psi V_{(K)})^T (\Psi V_{(K)}) - I \right\|_2 \leq \frac{1}{2} \right) \leq 1 - \delta \quad (7)$$

is satisfied for all sampling operators Ψ that sample M signal coefficients.

Proof. For an Erdős-Rényi graph, the eigenvector matrix V satisfies

$$\max_{i,j} |V_{i,j}| = O \left(\sqrt{\log^{2.2} g(N) \log N / p} \right)$$

for $p = g(N) \log(N)/N$ [42]. By substituting V into Theorem 1.2 in [43], we obtain (7). \square

Theorem 4. Let A, V, Ψ be defined as in Lemma 1. With probability $(1 - \delta)$, $\Psi V_{(K)}$ is a frame in \mathbb{C}^K with lower bound $M/2$ and upper bound $3M/2$.

Proof. Using Lemma 1, with probability $(1 - \delta)$, we have

$$\left\| \frac{1}{M} (\Psi V_{(K)})^T (\Psi V_{(K)}) - I \right\|_2 \leq \frac{1}{2}.$$

We thus obtain for all $x \in \mathbb{C}^K$,

$$\begin{aligned} -\frac{1}{2} x^T x &\leq x^T \left(\frac{1}{M} (\Psi V_{(K)})^T (\Psi V_{(K)}) - I \right) x \leq \frac{1}{2} x^T x, \\ \frac{M}{2} x^T x &\leq x^T (\Psi V_{(K)})^T (\Psi V_{(K)}) x \leq \frac{3M}{2} x^T x. \end{aligned}$$

\square

From Theorem 4, we see that the singular values of $\Psi V_{(K)}$ are bounded with high probability. This shows that $\Psi V_{(K)}$ has full rank with high probability; in other words, with high probability, perfect recovery is achieved for Erdős-Rényi graph signals when we randomly sample a sufficient number of signal coefficients.

3) *Simulations:* We verify Theorem 4 by checking the probability of satisfying the full-rank assumption by random sampling on Erdős-Rényi graphs. Once the full-rank assumption is satisfied, we can find a qualified sampling operator to achieve perfect recovery, thus, we call this probability the *success rate of perfect recovery*.

We check the success rate of perfect recovery with various sizes and connection probabilities. We vary the size of an Erdős-Rényi graph from 50 to 500 and the connection probabilities with an interval of 0.01 from 0 to 0.5. For each given size and connection probability, we generate 100 graphs randomly. Suppose that for each graph, the corresponding graph signal is of fixed bandwidth $K = 10$. Given a graph shift, we randomly

sample 10 rows from the first 10 columns of the graph Fourier transform matrix and check whether the 10×10 matrix is of full rank. Based on Theorem 1, if the 10×10 matrix is of full rank, the perfect recovery is guaranteed. For each given graph shift, we run the random sampling for 100 graphs, and count the number of successes to obtain the success rate.

Figure 5 shows success rates for sizes 50 and 500 averaged over 100 random tests. When we fix the graph size, in Erdős-Rényi graphs, the success rate increases as the connection probability increases, that is, more connections lead to higher probability of getting a qualified sampling operator. When we compare the different sizes of the same type of graph, the success rate increases as the size increases, that is, larger graph sizes lead to higher probabilities of getting a qualified sampling operator. Overall, with a sufficient number of connections, the success rates are close to 100%. The simulation results suggest that the full-rank assumption is easier to satisfy when there exist more connections on graphs. The intuition is that in a larger graph with a higher connection probability, the difference between nodes is smaller, that is, each node has a similar connectivity property and there is no preference to sampling one rather than the other.

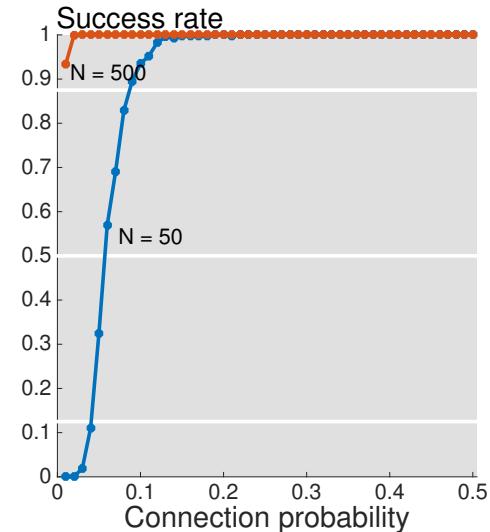


Fig. 5: Success rates for Erdős-Rényi graphs. The blue curve represents an Erdős-Rényi graph of size 50 and the red curve represents an Erdős-Rényi graph of size 500.

V. RELATIONS & EXTENSIONS

We now discuss three topics: relation to the sampling theory for finite discrete-time signals, relation to compressed sensing, and how to handle a full-band graph signal.

A. Relation to Sampling Theory for Finite Discrete-Time Signals

We call the graph that supports a finite discrete-time signal a *finite discrete-time graph*, which specifies the time-ordering from the past to future. The finite discrete-time graph can be

represented by the cyclic permutation matrix [31], [28],

$$A = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 1 & 0 \end{bmatrix} = V \Lambda V^{-1}, \quad (8)$$

where the eigenvector matrix

$$V = [v_0 \ v_1 \ \cdots \ v_{N-1}] = \left[\frac{1}{\sqrt{N}} (W^{jk})^* \right]_{j,k=0,\dots,N-1} \quad (9)$$

is the Hermitian transpose of the N -point discrete Fourier transform matrix, $V = F^*$, V^{-1} is the N -point discrete Fourier transform matrix, $V^{-1} = F$, and the eigenvalue matrix is

$$\Lambda = \text{diag}[W^0 \ W^1 \ \cdots \ W^{N-1}], \quad (10)$$

where $W = e^{-2\pi j/N}$. We see that Definitions 2, 3 and Theorem 1 are immediately applicable to finite discrete-time signals and are consistent with sampling of such signals [31].

Definition 4. A discrete-time signal is called *bandlimited* when there exists $K \in \{0, 1, \dots, N-1\}$ such that its discrete Fourier transform \hat{x} satisfies

$$\hat{x}_k = 0 \quad \text{for all } k \geq K.$$

The smallest such K is called the *bandwidth* of x . A discrete-time signal that is not bandlimited is called a *full-band discrete-time signal*.

Definition 5. The set of discrete-time signals in \mathbb{C}^N with bandwidth of at most K is a closed subspace denoted $\text{BL}_K(F)$, with F the discrete Fourier transform matrix.

With this definition of the discrete Fourier transform matrix, the highest frequency is in the middle of the spectrum (although this is just a matter of ordering). From Definitions 4 and 5, we can permute the rows in the discrete Fourier transform matrix to choose any frequency band. Since the discrete Fourier transform matrix is a Vandermonde matrix, any K rows of $F_{(K)}^*$ are independent [36], [31]; in other words, $\text{rank}(\Psi F_{(K)}^*) = K$ always holds when $M \geq K$. We apply now Theorem 1 to obtain the following result.

Corollary 1. Let Ψ satisfy that the sampling number is no smaller than the bandwidth, $M \geq K$. For all $x \in \text{BL}_K(F)$, perfect recovery, $x = \Phi \Psi x$, is achieved by choosing

$$\Phi = F_{(K)}^* U,$$

with $U \Psi F_{(K)}^*$ a $K \times K$ identity matrix, and $F_{(K)}^*$ denotes the first K columns of F^* .

From Corollary 1, we can perfectly recover a discrete-time signal when it is bandlimited.

Similarly to Theorem 2, we can show that a new graph shift can be constructed from the finite discrete-time graph. Multiple sampling mechanisms can be used to sample a new graph shift; an intuitive one is as follows: let $x \in \mathbb{C}^N$ be a finite discrete-time signal, where N is even. Reorder the frequencies in (10), by putting the frequencies with even indices first,

$$\tilde{\Lambda} = \text{diag}[\lambda_0 \ \lambda_2 \ \cdots \ \lambda_{N-2} \ \lambda_1 \ \lambda_3 \ \cdots \ \lambda_{N-1}].$$

Similarly, reorder the columns of V in (9) by putting the columns with even indices first

$$\tilde{V} = [v_0 \ v_2 \ \cdots \ v_{N-2} \ v_1 \ v_3 \ \cdots \ v_{N-1}].$$

One can check that $\tilde{V} \tilde{\Lambda} \tilde{V}^{-1}$ is still the same cyclic permutation matrix. Suppose we want to preserve the first $N/2$ frequency components in $\tilde{\Lambda}$; the sampled frequencies are then

$$\tilde{\Lambda}_{(N/2)} = \text{diag}[\lambda_0 \ \lambda_2 \ \cdots \ \lambda_{N-2}].$$

Let a sampling operator Ψ choose the first $N/2$ rows in $\tilde{V}_{(N/2)}$,

$$\Psi \tilde{V}_{(N/2)} = \left[\frac{1}{\sqrt{N}} (W^{2jk})^* \right]_{j,k=0,\dots,N/2-1},$$

which is the Hermitian transpose of the discrete Fourier transform of size $N/2$ and satisfies $\text{rank}(\Psi \tilde{V}_{(N/2)}) = N/2$ in Theorem 2. The sampled graph Fourier transform matrix $U = (\Psi \tilde{V}_{(N/2)})^{-1}$ is the discrete Fourier transform of size $N/2$. The sampled graph shift is then constructed as

$$A_M = U^{-1} \tilde{\Lambda}_{(N/2)} U,$$

which is exactly the $N/2 \times N/2$ cyclic permutation matrix. Hence, we have shown that by choosing an appropriate sampling mechanism, a smaller finite discrete-time graph is obtained from a larger finite discrete-time graph by using Theorem 2. We note that using a different ordering or sampling operator would result in a graph shift that can be different and non-intuitive. This is simply a matter of choosing different frequency components.

B. Relation to Compressed Sensing

Compressed sensing is a sampling framework to recover sparse signals with a few measurements [44]. The theory asserts that a few samples guarantee the recovery of the original signals when the signals and the sampling approaches are well-defined in some theoretical aspects. To be more specific, given the sampling operator $\Psi \in \mathbb{R}^{M \times N}$, $M \ll N$, and the sampled signal $x_M = \Psi x$, a sparse signal $x \in \mathbb{R}^N$ is recovered by solving

$$\min_x \|x\|_0, \quad \text{subject to } x_M = \Psi x. \quad (11)$$

Since the l_0 norm is not convex, the optimization is a non-deterministic polynomial-time hard problem. To obtain a computationally efficient algorithm, the l_1 -norm based algorithm, known as basis pursuit or basis pursuit with denoising, recovers the sparse signal with small approximation error [45].

In the standard compressed sensing theory, signals have to be sparse or approximately sparse to guarantee accurate recovery properties. In [46], the authors proposed a general way to perform compressed sensing with non-sparse signals using dictionaries. Specifically, a general signal $x \in \mathbb{R}^N$ is recovered by

$$\min_x \|Dx\|_0, \quad \text{subject to } x_M = \Psi x, \quad (12)$$

where D is a dictionary designed to make Dx sparse. When specifying x to be a graph signal, and D to be the appropriate graph Fourier transform of the graph on which the signal resides, Dx represents the frequency content of x , which is

sparse when x is of limited bandwidth. Equation (12) recovers a bandlimited graph signal from a few sampled signal coefficients via an optimization approach. The proposed sampling theory deals with the cases where the frequencies corresponding to non-zero elements are known, and can be reordered to form a bandlimited graph signal. Compressed sensing deals with the cases where the frequencies corresponding to non-zero elements are unknown, which is a more general and harder problem. If we have access to the position of the non-zero elements, the proposed sampling theory uses the smallest number of samples to achieve perfect recovery.

C. Relation to Signal Recovery on Graphs

Signal recovery on graphs attempts to recover graph signals that are assumed to be smooth with respect to underlying graphs, from noisy, missing, or corrupted samples [23]. Previous works studied signal recovery on graphs from different perspectives. For example, in [47], the authors considered that a graph is a discrete representation of a manifold, and aimed at recovering graph signals by regularizing the smoothness functional; in [48], the authors aimed at recovering graph signals by regularizing the combinatorial Dirichlet; in [23], the authors aimed at recovering graph signals by regularizing the graph total variation; in [49], the authors aimed at recovering graph signals by finding the empirical modes; in [18], the authors aimed at recovering graph signals by training a graph-based dictionary. These works focused on minimizing the empirical recovery error, and dealt with general graph signal models. It is thus hard to show when and how the missing signal coefficients can be exactly recovered.

Similarly to signal recovery on graphs, sampling theory on graphs also attempts to recover graph signals from incomplete samples. The main difference is that sampling theory on graphs focuses on a subset of smooth graph signals, that is, bandlimited graph signals, and theoretically analyzes when and how the missing signal coefficients can be exactly recovered. The authors in [10], [50], [12] considered a similar problem to ours, that is, recovery of bandlimited graph signals by sampling a few signal coefficients in the vertex domain. The main differences are as follows: (1) we focus on the graph adjacency matrix, not the graph Laplacian matrix; (2) when defining the bandwidth, we focus on the number of frequencies, not the values of frequencies. Some recent extensions of sampling theory on graphs include [51], [52], [53].

D. Graph Downsampling & Graph Filter Banks

In classical signal processing, sampling refers to sampling a continuous function and downsampling refers to sampling a sequence. Both concepts use fewer samples to represent the overall shape of the original signal. Since a graph signal is discrete in nature, sampling and downsampling are the same. Previous works implemented graph downsampling via graph coloring [6] or minimum spanning tree [54].

The proposed sampling theory provides a family of qualified sampling operators (5) with an optimal sampling operator as in (6). To downsample a graph by 2, one can set the bandwidth to a half of the number of nodes, that is, $K = N/2$, and use (6)

to obtain an optimal sampling operator. An example for the finite discrete-time signals was shown in Section V-A.

As shown in Theorem 1, perfect recovery is achieved when graph signals are bandlimited. To handle full-band graph signals, we propose an approach based on graph filter banks, where each channel does not need to recover perfectly but in conjunction they do.

Let x be a full-band graph signal, which, without loss of generality, we can express without loss of generality as the addition of two bandlimited signals supported on the same graph, that is, $x = x^l + x^h$, where

$$x^l = P^l x, \quad x^h = P^h x,$$

and

$$P^l = V \begin{bmatrix} I_K & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} V^{-1}, \quad P^h = V \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{N-K} \end{bmatrix} V^{-1}.$$

We see that x^l contains the first K frequencies, x^h contains the other $N - K$ frequencies, and each is bandlimited. We do sampling and interpolation for x^l and x^h in two channels, respectively. Take the first channel as an example. Following Theorems 1 and 2, we use a qualified sampling operator Ψ^l to sample x^l , and obtain the sampled signal coefficients as $x_{\mathcal{M}^l}^l = \Psi^l x^l$, with the corresponding graph as $A_{\mathcal{M}^l}$. We can recover x^l by using interpolation operator Φ^l as $x^l = \Phi^l x_{\mathcal{M}^l}^l$. Finally, we add the results from both channels to obtain the original full-band graph signal (also illustrated in Figure 6). The main benefit of working with a graph filter bank is that, instead of dealing with a long graph signal with a large graph, we are allowed to focus on the frequency bands of interests and deal with a shorter graph signal with a small graph in each channel.

We do not restrict the samples from two bands, $x_{\mathcal{M}^l}^l$ and $x_{\mathcal{M}^h}^h$ the same size because we can adaptively design the sampling and interpolation operators based on their own sizes. This is similar to the filter banks in the classical literature where the spectrum is not evenly partitioned between the channels [55]. We see that the above idea can easily be generalized to multiple channels by splitting the original graph signal into multiple bandlimited graph signals; instead of dealing with a huge graph, we work with multiple small graphs, which makes computation easier.

Simulations. We now show an example where we analyze graph signals by using the proposed graph filter banks. Similarly to Section IV-A2, we consider that the weather stations across the U.S. form a graph and temperature values measured at each weather station in one day form a graph signal. Suppose that a high-frequency component represents some pattern of weather change; we want to detect this pattern given the temperature values. We can decompose a graph signal of temperature values into a low-frequency channel (largest 15 frequencies) and a high-frequency channel (smallest 5 frequencies). In each channel, we sample the bandlimited graph signal to obtain a sparse and lossless representation. Figure 7 shows a comparison between temperature values on January 1st, 2013 and May 1st, 2013. We intentionally added some high-frequency component to the temperature values on January 1st, 2013. We see that the high-frequency channel

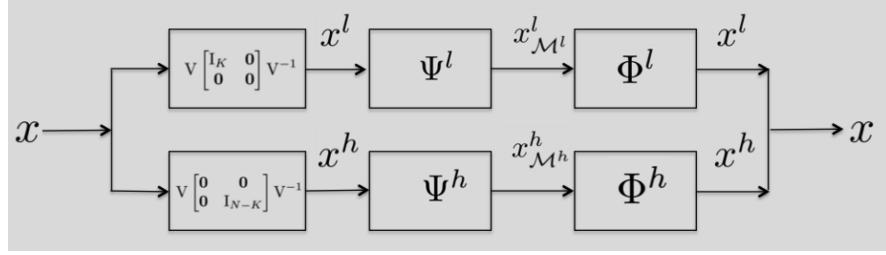


Fig. 6: Graph filter bank that splits the graph signal into two bandlimited graph signals. In each channel, we perform sampling and interpolation, following Theorem 1. Finally, we add the results from both channels to obtain the original full-band graph signal.

in Figure 7 (a) detects the change, while the high-frequency channel in Figure 7 (b) does not.

Directly using the graph frequency components can also detect the high-frequency components, but the graph frequency components cannot be easily visualized. Since the graph structure and the decomposed channels are fixed, the optimal sampling operator and corresponding interpolation operator in each channel can be designed in advance, which means that we just need to look at the sampled coefficients of a fixed set of nodes to check whether a channel is activated. The graph filter bank is thus fast and visualization friendly.

VI. APPLICATIONS

The proposed sampling theory on graphs can be applied to semi-supervised learning, whose goal is to classify data with a few labeled and a large number of unlabeled samples [56]. One approach is based on graphs, where the labels are often assumed to be smooth on graphs. From a perspective of signal processing, smoothness can be expressed as lowpass nature of the signal. Recovering smooth labels on graphs is then equivalent to interpolating a low-pass graph signal. We now look at two examples, including classification of online blogs and handwritten digits.

1) *Sampling Online Blogs:* We first aim to investigate the success rate of perfect recovery by using random sampling, and then classify the labels of the online blogs. We consider a dataset of $N = 1224$ online political blogs as either conservative or liberal [57]. We represent conservative labels as $+1$ and liberal ones as -1 . The blogs are represented by a graph in which nodes represent blogs, and directed graph edges correspond to hyperlink references between blogs. The graph signal here is the label assigned to the blogs, called the labeling signal. We use the spectral decomposition in (2) for this online-blog graph to get the graph frequencies in a descending order and the corresponding graph Fourier transform matrix. The labeling signal is a full-band signal, but approximately bandlimited.

To investigate the success rate of perfect recovery by using random sampling, we vary the bandwidth K of the labeling signal with an interval of 1 from 1 to 20, randomly sample K rows from the first K columns of the graph Fourier transform matrix, and check whether the $K \times K$ matrix has full rank. For each bandwidth, we randomly sample 10,000 times, and count the number of successes to obtain the success rate. Figure 8

(a) shows the resulting success rate. We see that the success rate decreases as we increase the bandwidth; it is above 90%, when the bandwidth is no greater than 20. It means that we can achieve perfect recovery by using random sampling with a fairly high probability. As the bandwidth increases, even if we get an equal number of samples, the success rate still decreases, because when we take more samples, it is easier to get a sample correlated with the previous samples.

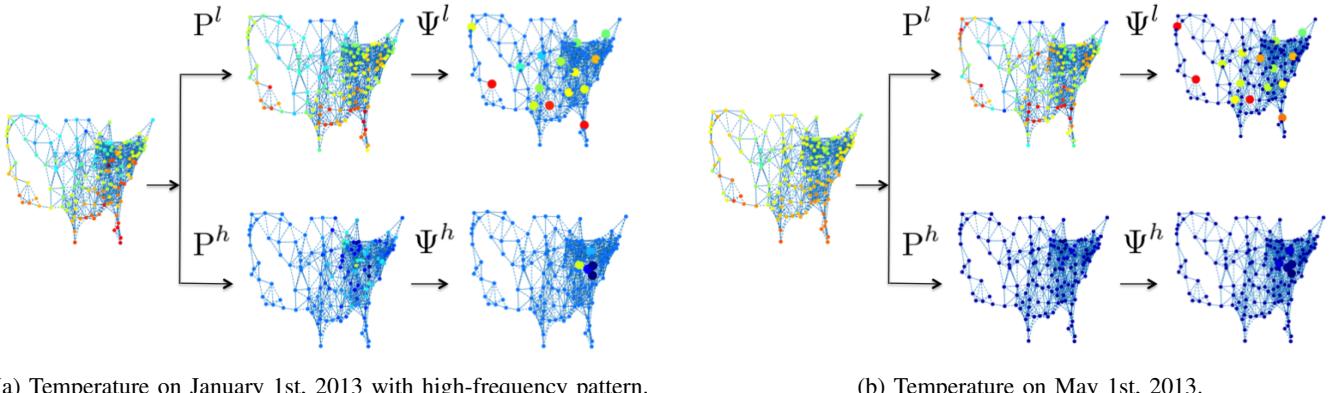
Since a qualified sampling operator is independent of graph signals, we precompute the qualified sampling operator for the online-blog graph, as discussed in Section III-B. When the labeling signal is bandlimited, we can sample M labels from it by using a qualified sampling operator, and recover the labeling signal by using the corresponding interpolation operator. In other words, we can design a set of blogs to label before querying any label. Most of the time, however, the labeling signal is not bandlimited, and it is not possible to achieve perfect recovery. Since we only care about the sign of the labels, we use only the low frequency content to approximate the labeling signal; after that, we set a threshold to assign labels. To minimize the influence from the high-frequency content, we can use the optimal sampling operator in Algorithm 1.

We solve the following optimization problem to recover the low frequency content,

$$\hat{x}_{(K)}^{opt} = \arg \min_{\hat{x}_{(K)} \in \mathbb{R}^K} \|\text{sgn}(\Psi V_{(K)} \hat{x}_{(K)}) - x_{\mathcal{M}}\|_2^2, \quad (13)$$

where $\Psi \in \mathbb{R}^{M \times N}$ is a sampling operator, $x_{\mathcal{M}} \in \mathbb{R}^M$ is a vector of the sampled labels whose element is either $+1$ or -1 , and $\text{sgn}(\cdot)$ sets all positive values to $+1$ and all negative values to -1 . Note that without $\text{sgn}(\cdot)$, the solution of (13) is $(\Psi V_{(K)})^{-1} x_{\mathcal{M}}$ in Theorem 1, which perfectly recovers the labeling signal when it is bandlimited. When the labeling signal is not bandlimited, the solution of (13) approximates the low-frequency content. The ℓ_2 norm (13) can be relaxed by the logit function and solved by logistic regression [58]. The recovered labels are then $x^{opt} = \text{sgn}(V_{(K)} \hat{x}_{(K)}^{opt})$.

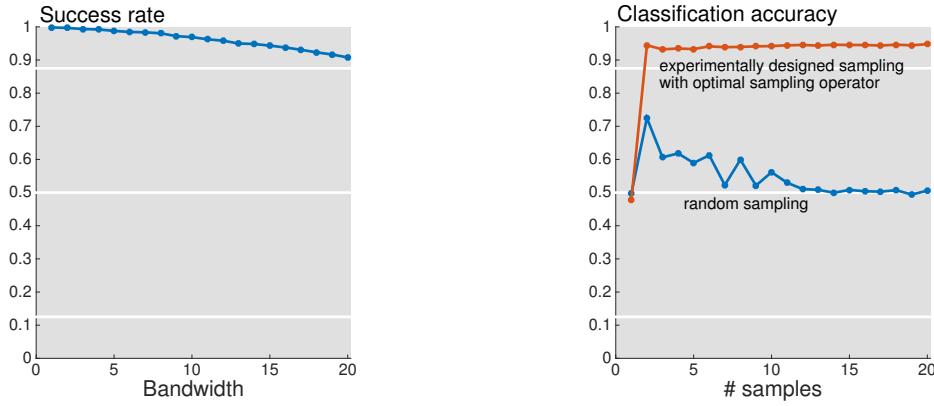
Figure 8 (b) compares the classification accuracies between optimal sampling and random sampling by varying the sample size with an interval of 1 from 1 to 20. We see that the optimal sampling significantly outperforms random sampling, and random sampling does not improve with more samples, because the interpolation operator (13) assumes that the sampling operator is qualified, which is not always true for random



(a) Temperature on January 1st, 2013 with high-frequency pattern.

(b) Temperature on May 1st, 2013.

Fig. 7: Graph filter banks analysis.



(a) Success rate as a function of bandwidth. (b) Classification accuracy as a function of the number of samples.

Fig. 8: Classification for online blogs. When increasing the bandwidth, it is harder to find a qualified sampling operator. The experimentally designed sampling with the optimal sampling operator outperforms random sampling.

sampling as shown in Figure 8 (a). Note that classification accuracy for the optimal sampling is as high as 94.44% by only sampling two blogs, and the classification accuracy gets slightly better as we increase the number of samples. Compared with the previous results [24], to achieve around 94% classification accuracy,

- harmonic functions on graph samples 120 blogs;
- graph Laplacian regularization samples 120 blogs;
- graph total variation regularization samples 10 blogs; and
- the proposed optimal sampling operator (6) samples 2 blogs.

The improvement comes from the fact that, instead of sampling randomly as in [24], we use the optimal sampling operator to choose samples based on the graph structure.

2) *Classification for Handwritten Digits*: We aim to use the proposed sampling theory to classify handwritten digits and achieve high classification accuracy with fewer samples.

We work with two handwritten digit datasets, the MNIST [59] and the USPS [60]. Each dataset includes ten classes (0-9 digit characters). The MNIST dataset includes 60,000 samples in total. We randomly select 1000 samples for each digit character, for a total of $N = 10,000$ digit images; each image is normalized to $28 \times 28 = 784$ pixels. The USPS dataset

includes 11,000 samples in total. We use all the images in the dataset; each image is normalized to $16 \times 16 = 256$ pixels.

Since same digits produce similar images, it is intuitive to build a graph to reflect the relational dependencies among images. For each dataset, we construct a 12-nearest neighbor graph to represent the digit images. The nodes represent digit images and each node is connected to 12 other nodes that represent the most similar digit images; the similarity is measured by the Euclidean distance. The graph shift is constructed as $A_{i,j} = P_{i,j} / \sum_i P_{i,j}$, with

$$P_{i,j} = \exp\left(\frac{-N^2 \|f_i - f_j\|_2}{\sum_{i,j} \|f_i - f_j\|_2}\right),$$

with f_i a vector representing the digit image. The graph shift is asymmetric, representing a directed graph, which cannot be handled by graph Laplacian-based methods.

Similarly to Section VI-1, we aim to label all the digit images by actively querying the labels of a few images. To handle 10-class classification, we form a ground-truth matrix X of size $N \times 10$. The element $X_{i,j}$ is +1, indicating the membership of the i th image in the j th digit class, and is -1 otherwise. We obtain the optimal sampling operator Ψ as shown in Algorithm 1. The querying samples are then

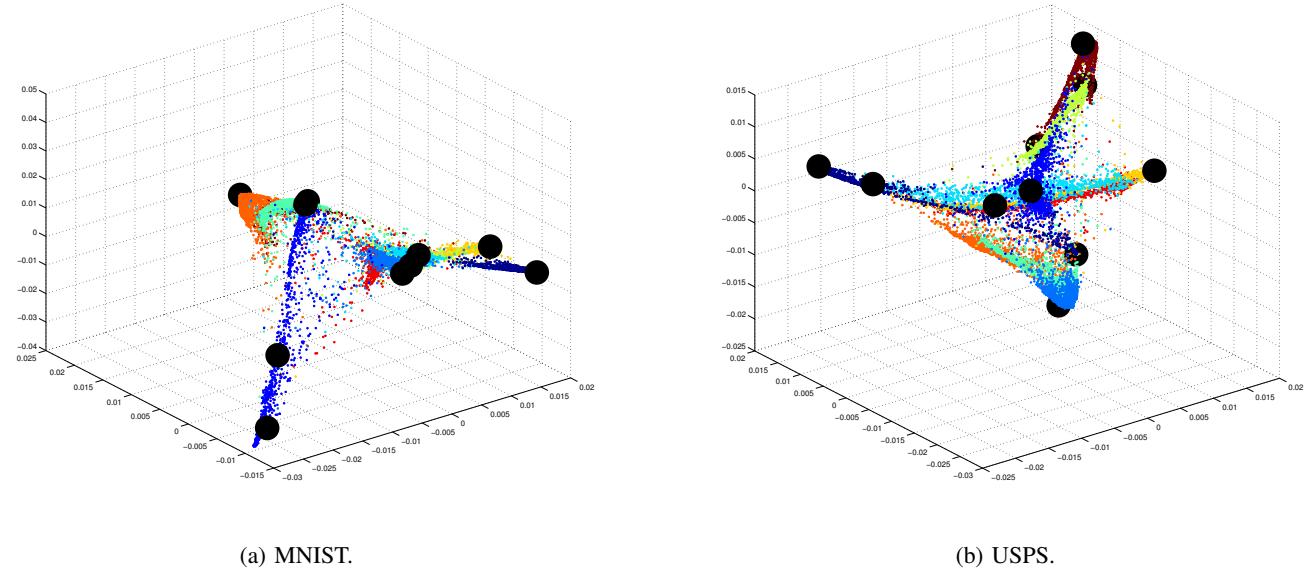


Fig. 9: Graph representations of the MNIST and USPS datasets. For both datasets, the nodes (digit images) with the same digit characters are shown in the same color and the big black dots indicate 10 sampled nodes by using the optimal sampling operators in Algorithm 1.

$X_{\mathcal{M}} = \Psi X \in \mathbb{R}^{M \times 10}$. We recover the low frequency content as

$$\hat{X}_{(K)}^{opt} = \arg \min_{\hat{X}_{(K)} \in \mathbb{R}^{K \times 10}} \left\| \text{sgn}(\Psi V_{(K)} \hat{X}_{(K)}) - X_{\mathcal{M}} \right\|_2^2. \quad (14)$$

We solve (14) approximately by using logistic regression and then obtain the estimated label matrix $X^{opt} = V_{(K)} \hat{X}_{(K)}^{opt} \in \mathbb{R}^{N \times 10}$, whose element $(X^{opt})_{i,j}$ shows a confidence of labeling the i th image as the j th digit. We finally label each digit image by choosing the one with largest value in each row of X^{opt} .

The graph representations of the MNIST and USPS datasets, and the optimal sampling sets are shown in Figure 9. The coordinates of nodes come from the corresponding rows of the first three columns of the inverse graph Fourier transform. We see that the images with the same digit characters form clusters, and the optimal sampling operator chooses representative samples from different clusters.

Figure 10 shows the classification accuracy by varying the sample size with an interval of 10 from 10 to 100 for both datasets. For the MNIST dataset, we query 0.1%–1% images; for the USPS dataset, we query 0.09% – 0.9% images. We achieve around 90% classification accuracy by querying only 0.5% images for both datasets. Compared with the previous results [35], in the USPS dataset, given 100 samples,

- local linear reconstruction is around 65%;
- normalized cut based active learning is around 70%;
- graph sampling based active semi-supervised learning is around 85%; and
- the proposed optimal sampling operator (6) with the interpolation operator (14) achieves 91.69%.

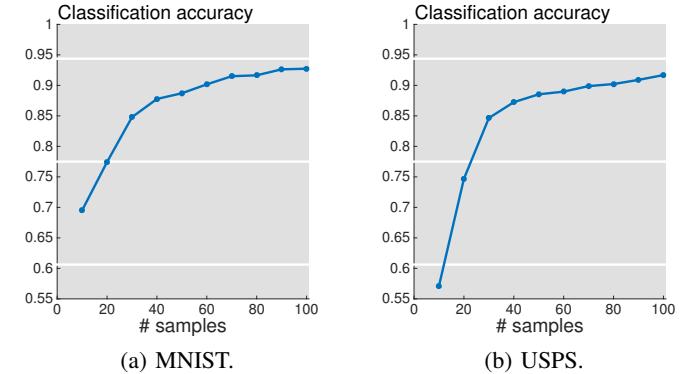


Fig. 10: Classification accuracy of the MNIST and USPS datasets as a function of the number of querying samples.

VII. CONCLUSIONS

We proposed a novel sampling framework for graph signals that follows the same paradigm as classical sampling theory and strongly connects to linear algebra. We showed that perfect recovery is possible when graph signals are bandlimited. The sampled signal coefficients then form a new graph signal, whose corresponding graph structure is constructed from the original graph structure, preserving the first-order difference of the original graph signal. We studied a qualified sampling operator for both random sampling and experimentally designed sampling. We further established the connection to the sampling theory for finite discrete-time signal processing and previous works on the sampling theory on graphs, and showed how to handle full-band graphs signals by using graph filter banks. We showed applications to semi-supervised classification of online blogs and digit images, where the proposed

sampling and interpolation operators perform competitively.

REFERENCES

- [1] M. Jackson, *Social and Economic Networks*, Princeton University Press, 2008.
- [2] M. Newman, *Networks: An Introduction*, Oxford University Press, 2010.
- [3] 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, pp. 83–98, May 2013.
- [4] A. Sandryhaila and J. M. F. Moura, “Big data processing with signal processing on graphs,” *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, 2014.
- [5] A. Sandryhaila and J. M. F. Moura, “Discrete signal processing on graphs,” *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
- [6] S. K. Narang and A. Ortega, “Perfect reconstruction two-channel wavelet filter banks for graph structured data,” *IEEE Trans. Signal Process.*, vol. 60, pp. 2786–2799, June 2012.
- [7] 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.
- [8] D. K. Hammond, P. Vandergheynst, and R. Gribonval, “Wavelets on graphs via spectral graph theory,” *Appl. Comput. Harmon. Anal.*, vol. 30, pp. 129–150, Mar. 2011.
- [9] S. K. Narang, G. Shen, and A. Ortega, “Unidirectional graph-based wavelet transforms for efficient data gathering in sensor networks,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Dallas, TX, Mar. 2010, pp. 2902–2905.
- [10] I. Z. Pesenson, “Sampling in Paley-Wiener spaces on combinatorial graphs,” *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, May 2008.
- [11] S. K. Narang, A. Gadde, and A. Ortega, “Signal processing techniques for interpolation in graph structured data,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, May 2013, pp. 5445–5449.
- [12] A. Anis, A. Gadde, and A. Ortega, “Towards a sampling theorem for signals on arbitrary graphs,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2014, pp. 3864–3868.
- [13] A. Agaskar and Y. M. Lu, “A spectral graph uncertainty principle,” *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4338–4356, July 2013.
- [14] S. Chen, F. Cerdà, P. Rizzo, J. Bielak, J. H. Garrett, and J. Kovačević, “Semi-supervised multiresolution classification using adaptive graph filtering with application to indirect bridge structural health monitoring,” *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2879–2893, June 2014.
- [15] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, “Adaptive graph filtering: Multiresolution classification on graphs,” in *IEEE GlobalSIP*, Austin, TX, Dec. 2013, pp. 427 – 430.
- [16] V. N. Ekmabaram, B. Ayazifar G. Fanti, and K. Ramchandran, “Wavelet-regularized graph semi-supervised learning,” in *Proc. IEEE Glob. Conf. Signal Information Process.*, Austin, TX, Dec. 2013, pp. 423 – 426.
- [17] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, “Learning graphs from observations under signal smoothness prior,” *IEEE Trans. Signal Process.*, 2014, Submitted.
- [18] D. Thanou, D. I. Shuman, and P. Frossard, “Learning parametric dictionaries for signals on graphs,” *IEEE Trans. Signal Process.*, vol. 62, pp. 3849–3862, June 2014.
- [19] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, “Signal denoising on graphs via graph filtering,” in *Proc. IEEE Glob. Conf. Signal Information Process.*, Atlanta, GA, Dec. 2014.
- [20] N. Tremblay and P. Borgnat, “Graph wavelets for multiscale community mining,” *IEEE Trans. Signal Process.*, vol. 62, pp. 5227–5239, Oct. 2014.
- [21] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov, “Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds,” *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 905–918, Feb. 2014.
- [22] P.-Y. Chen and A. O. Hero, “Local Fiedler vector centrality for detection of deep and overlapping communities in networks,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Florence, 2014, pp. 1120 – 1124.
- [23] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, “Signal recovery on graphs: Variation minimization,” *IEEE Trans. Signal Process.*, 2014, To appear.
- [24] S. Chen, A. Sandryhaila, G. Lederman, Z. Wang, J. M. F. Moura, P. Rizzo, J. Bielak, J. H. Garrett, and J. Kovačević, “Signal inpainting on graphs via total variation minimization,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Florence, May 2014, pp. 8267 – 8271.
- [25] X. Wang, P. Liu, and Y. Gu, “Local-set-based graph signal reconstruction,” *IEEE Trans. Signal Process.*, 2015, To appear.
- [26] X. Wang, M. Wang, and Y. Gu, “A distributed tracking algorithm for reconstruction of graph signals,” *IEEE Journal of Selected Topics on Signal Processing*, 2015, To appear.
- [27] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, “Distributed algorithm for graph signals,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brisbane, Apr. 2015.
- [28] A. Sandryhaila and J. M. F. Moura, “Discrete signal processing on graphs: Frequency analysis,” *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, June 2014.
- [29] M. Püschel and J. M. F. Moura, “Algebraic signal processing theory: Foundation and 1-D time,” *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3572–3585, Aug. 2008.
- [30] M. Püschel and J. M. F. Moura, “Algebraic signal processing theory: 1-D space,” *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3586–3599, Aug. 2008.
- [31] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of Signal Processing*, Cambridge University Press, 2014, <http://www.fourierandwavelets.org/>.
- [32] J. Kovačević and M. Püschel, “Algebraic signal processing theory: Sampling for infinite and finite 1-D space,” *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 242–257, Jan. 2010.
- [33] M. Unser, “Sampling – 50 years after Shannon,” *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.
- [34] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, “Sampling theory for graph signals,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brisbane, Australia, Apr. 2015.
- [35] A. Gadde, A. Anis, and A. Ortega, “Active semi-supervised learning using sampling theory for graph signals,” in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, New York, USA, 2014, KDD ’14, pp. 492–501.
- [36] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
- [37] D. Spielman and N. Srivastava, “Graph sparsification by effective resistances,” *SIAM Journal on Computing*, vol. 40, no. 6, pp. 1913–1926, 2011.
- [38] H. Avron and C. Boutsidis, “Faster subset selection for matrices and applications,” *SIAM J. Matrix Analysis Applications*, vol. 34, no. 4, pp. 1464–1499, 2013.
- [39] M. Püschel and J. Kovačević, “Real, tight frames with maximal robustness to erasures,” in *Proc. Data Compr. Conf.*, Snowbird, UT, Mar. 2005, pp. 63–72.
- [40] A. Sandryhaila, A. Chebira, C. Milo, J. Kovačević, and M. Püschel, “Systematic construction of real lapped tight frame transforms,” *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2256–2567, May 2010.
- [41] V. N. Ekmabaram, G. C. Fanti, B. Ayazifar, and K. Ramchandran, “Multiresolution graph signal processing via circulant structures,” in *2013 IEEE DSP/SPE*, Napa, CA, Aug. 2013, pp. 112–117.
- [42] L. V. Tran, V. H. Vu, and K. Wang, “Sparse random graphs: Eigenvalues and eigenvectors,” *Random Struct. Algorithms*, vol. 42, no. 1, pp. 110–134, 2013.
- [43] E. J. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, pp. 110–134, 2007.
- [44] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [45] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
- [46] E. J. Candès, Y. Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries,” *Applied and Computational Harmonic Analysis*, vol. 31, pp. 59–73, 2010.
- [47] M. Belkin and P. Niyogi, “Semi-supervised learning on riemannian manifolds,” *Machine Learning*, vol. 56, no. 1-3, pp. 209–239, 2004.
- [48] L. Grady and E. Schwartz, “Anisotropic interpolation on graphs: The combinatorial dirichlet problem,” *Technical Report CAS/CNS-2003-014*, July 2003.
- [49] N. Tremblay, P. Borgnat, and P. Flandrin, “Graph empirical mode decomposition,” in *Proc. Eur. Conf. Signal Process.*, Lisbon, Sept. 2014, pp. 2350–2354.
- [50] I. Z. Pesenson, “Variational splines and paley-wiener spaces on combinatorial graphs,” *Constr. Approx.*, vol. 29, pp. 1–21, 2009.
- [51] H. Führ and I. Z. Pesenson, “Poincaré and plancherel-polya inequalities in harmonic analysis on weighted combinatorial graphs,” *SIAM J. Discrete Math.*, vol. 27, no. 4, pp. 2007–2028, 2013.
- [52] S. Chen, R. Varma, A. Singh, and J. Kovačević, “Signal recovery on graphs: Random versus experimentally designed sampling,” in *SampTA*, Washington, DC, May 2015.

- [53] A. Gadde and A. Ortega, "A probabilistic interpretation of sampling theory of graph signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brisbane, Australia, Apr. 2015.
- [54] H. Q. Nguyen and M. N. Do, "Downsampling of signals on graphs via maximum spanning trees," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 182–191, Jan. 2015.
- [55] M. Vetterli, "A theory of multirate filter banks," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 3, pp. 356–372, Mar. 1987.
- [56] X. Zhu, "Semi-supervised learning literature survey," Tech. Rep. 1530, Univ. Wisconsin-Madison, 2005.
- [57] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 U.S. election Divided they blog," in *Proc. LinkKDD*, 2005, pp. 36–43.
- [58] C. M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics. Springer, 2006.
- [59] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Intelligent Signal Processing*. 2001, pp. 306–351, IEEE Press.
- [60] J. Hull, "A database for handwritten text recognition research," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 550–554, May 1994.