Discrete Signal Processing on Graphs: Sampling Theory

Siheng Chen, *Student Member, IEEE*, Rohan Varma, *Student Member, IEEE*, Aliaksei Sandryhaila, *Member, IEEE*, Jelena Kovačević, *Fellow, IEEE*

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 the perfect recovery is possible for graph signals bandlimited under the graph Fourier transform, and the sampled signal coefficients form a new graph signal, whose corresponding graph structure is constructed from the original graph structure, preserving frequency contents. By imposing a specific structure on the graph, graph signals reduce to finite discrete-time signals and the proposed sampling theory works reduces to classical signal processing. We further establish the connection to frames with maximal robustness to erasures as well as compressed sensing, and show how to choose the optimal sampling operator, how random sampling works on circulant graphs and Erdős-Rényi graphs, and how to handle full-band graph signals by using graph filter banks. We validate the proposed sampling theory on the simulated datasets of Erdős-Rényi graphs and small-world graphs, and a real-world dataset of online blogs. We show that for each case, the proposed sampling theory achieves perfect recovery with high probability. 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 work.

Index Terms—Sampling theorem, discrete signal processing on graphs, frames, 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 involves graph-based filtering [5], [6], [7], graph-based transforms [5], [8], [9], sampling and interpolation on graphs [10], [11], [12], uncertainty principle

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.

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], graphs signal recovery [23], [24], [25] and distributed algorithm [26], [27].

Two basic approaches to signal processing on graphs have been considered: The first is rooted in spectral graph theory and builds upon the *graph Laplacian matrix* [3]. Since the 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 on 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 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 implications and applications. For example, given a graph representing friendship connectivity in 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 structure. 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 not consistent with classical sampling theory and applies to undirected graphs only. It also does not explain how a 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 further establish the connection to frames with maximal robustness to erasures as well as compressed sensing, show how to choose the optimal sampling operator, and how random sampling works on circulant graphs and Erdős-Rényi graphs. To handle full-band graphs signals, we propose graph filter banks to force graphs signals to be bandlimited. We validate the proposed sampling theory on the simulated datasets of Erdős-Rényi graphs, small-world graphs, and a real-world dataset of online blogs. We show that for each case, the proposed sampling theory achieves perfect recovery with high probability. 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 work.

Contributions. The contributions of the paper are as follows:

- A novel sampling theory for graph signals, which follows the same paradigm as classical sampling theory;
- A novel approach to construct a graph structure that supports the sampled signal coefficients;
- A novel principle to choose the optimal sampling operator;
- A novel approach to construct graph filter banks to analyze full-band graph signal;
- A novel framework to do semi-supervised learning.

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. The proposed sampling theory is evaluated in in Section V. Section VI 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 $G = (\mathcal{V}, A)$, where $\mathcal{V} = \{v_0, \ldots, v_{N-1}\}$ is the set of nodes and

$A \in \mathbb{C}^{N \times N}$ 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 nth and the mth node, such as a similarity, a dependency, or a communication pattern.

B. Graph Signal

Given the graph representation $G = (\mathcal{V}, A)$, 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

$$x = \begin{bmatrix} x_0 & x_1 & \ldots & x_{N-1} \end{bmatrix}^T \in \mathbb{C}^N, \qquad (1)$$

where the nth 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 functions that are invariant to filtering; here, this basis is the eigenbasis of the graph shift A (or, if the complete eigenbasis does not exist, the Jordan eigenbasis of A). For simplicity, assume that A has a complete eigenbasis and the spectral decomposition of A is [31]

$$A = V \Lambda V^{-1}, \qquad (2)$$

where the eigenvectors of A form the columns of matrix V, and $\Lambda \in \mathbb{C}^{N \times N}$ is the diagonal matrix of corresponding eigenvalues $\lambda_0, \ldots, \lambda_{N-1}$ of 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 $x \in \mathbb{C}^N$ is

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

The *inverse graph Fourier transform* is

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

The vector $\widehat{x}$ in (3) represents the signal's expansion in the eigenvector basis and describes the frequency content of the graph signal 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

In this section, we propose a sampling theory for graph signals. We show that perfect recovery is possible for graph signals bandlimited under the graph Fourier transform, and a new graph shift for the sampled signal coefficients is constructed from the original graph shift. Parts of this section have appeared in [34].

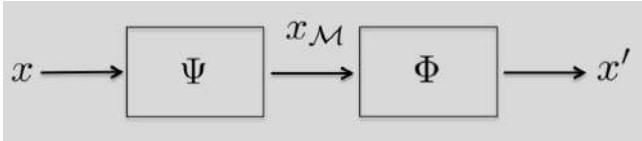| Symbol	Description	Dimension
A	graph shift	$N \times N$
x	graph signal	N
V^{-1}	graph Fourier transform matrix	$N \times N$
$\widehat{x}$	graph signal in the frequency domain	N
Ψ	sampling operator	$M \times N$
Φ	interpolation operator	$N \times M$
$\mathcal{M}$	sampled indices	
$x_{\mathcal{M}}$	sampled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

A. Sampling Theory for Graph Signals

Suppose that we want to sample M coefficients in a graph signal $x \in \mathbb{C}^N$ to produce a sampled signal $x_{\mathcal{M}} \in \mathbb{C}^M$ ($M < N$), where $\mathcal{M} = (\mathcal{M}_0, \cdots, \mathcal{M}_{M-1})$ denotes the sequence of *sampled* indices, and $\mathcal{M}_i \in \{0, 1, \cdots, N-1\}$. 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),

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

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 bandlimited graph signals as in the classical sampling theory,

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

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

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

When defining the bandwidth, we focus on the number of frequencies, while the previous work [12] focuses on the value of frequencies.

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).

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]. Denote $V_{(K)} \in \mathbb{R}^{N \times K}$ as the first K columns of V, $\Psi \in \mathbb{R}^{M \times N}$ as a sampling operator, and $\Phi \in \mathbb{R}^{N \times M}$ as an interpolation operator.

Theorem 1. Let Ψ satisfy

$$\text{rank}(\Psi V_{(K)}) = K. \quad (5)$$

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.

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 restriction is equivalent to limiting the number of nonzero elements in the graph Fourier domain. Theorem 1 is thus applicable for all graph signals that have a few nonzero elements in the graph Fourier domain, 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. Since $U \Psi V_{(K)}$ is 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].

B. Sampled Coefficients as A New 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 new 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 $x \in \mathrm{BL}_K(\mathrm{V}^{-1})$, we have

$$x = \Phi \Psi x = \Phi x_{\mathcal{M}} \overset{(a)}{=} \mathrm{V}_{(K)} \mathrm{U} x_{\mathcal{M}}$$
$$\overset{(b)}{=} \mathrm{V}_{(K)} \widehat{x}_{(K)}.$$

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

$$\widehat{x}_{(K)} = \mathrm{U} x_{\mathcal{M}}.$$

and

$$x_{\mathcal{M}} = \mathrm{U}^{-1} \mathrm{U} x_{\mathcal{M}} = \mathrm{U}^{-1} \widehat{x}_{(K)}.$$

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

$$\mathrm{A}_{\mathcal{M}} = \mathrm{U}^{-1} \Lambda_{(K)} \mathrm{U} \in \mathbb{C}^{K \times K},$$

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 $x \in \mathrm{BL}_K(\mathrm{V}^{-1})$ and let

$$x_{\mathcal{M}} = \Psi x \in \mathbb{C}^K$$

be its sampled version, where Ψ is a qualified sampling operator. Then, the graph shift associated with the graph signal $x_{\mathcal{M}}$ is

$$\mathrm{A}_{\mathcal{M}} = \mathrm{U}^{-1} \Lambda_{(K)} \mathrm{U} \in \mathbb{C}^{K \times K}, \qquad (6)$$

where U be $(\Psi \mathrm{V}_{(K)})^{-1}$. The graph Fourier transform of $x_{\mathcal{M}}$ is

$$\widehat{x}_{\mathcal{M}} = \mathrm{U} x_{\mathcal{M}} \in \mathbb{C}^K,$$

and the inverse graph Fourier transform is

$$x_{\mathcal{M}} = \mathrm{U}^{-1} \widehat{x}_{\mathcal{M}} \in \mathbb{C}^K.$$

From Theorem 2, we see that the graph shift $\mathrm{A}_{\mathcal{M}}$ 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 $\mathrm{A}_{\mathcal{M}}$ 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 $\widehat{x}_{(K)} = \widehat{x}_{\mathcal{M}}$, 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 $x_{\mathcal{M}}$ 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.

C. Example

We consider a five-node directed graph with graph shift

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

The corresponding inverse graph Fourier transform matrix is

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

and the frequencies are

$$\Lambda = \mathrm{diag} \begin{bmatrix} 1 & 0.39 & -0.12 & -0.44 & -0.83 \end{bmatrix}.$$

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

$$x = \begin{bmatrix} 0.29 & 0.32 & 0.18 & 0.05 & 0.17 \end{bmatrix}^T,$$

and

$$\widehat{x} = \begin{bmatrix} 0.5 & 0.2 & 0.1 & 0 & 0 \end{bmatrix}^T,$$

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, $\mathcal{M} = (1, 2, 4)$, $x_{\mathcal{M}} = \begin{bmatrix} 0.29 & 0.32 & 0.05 \end{bmatrix}^T$, 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)

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

The inverse graph Fourier transform matrix for the sampled signal is

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

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

$$\mathrm{A}_{\mathcal{M}} = \mathrm{U}^{-1} \Lambda_{(3)} \mathrm{U} = \begin{bmatrix} 0.39 & 0.31 & 0.24 \\ -0.62 & -0.06 & -0.49 \\ 1.56 & 0.26 & 0.95 \end{bmatrix}.$$
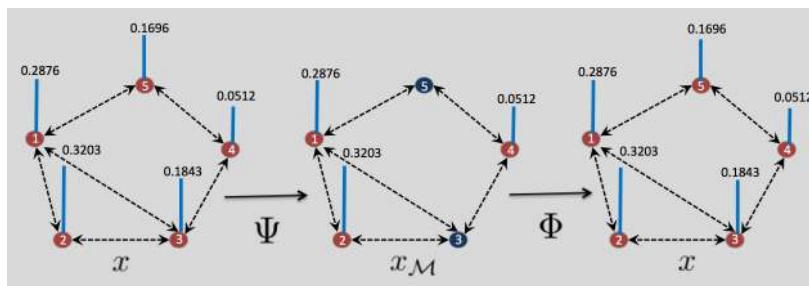
Fig. 2: Sampling followed by interpolation. The arrows indicate that the edges are directed.

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

IV. RELATIONS & EXTENSIONS

We now discuss four topics: relation to the sampling theory for finite discrete-time signals how to choose a sampling operator, how random sampling works, and how to handle graph filter banks.

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 indicates the time-ordering from the past to future. The finite discrete-time graph can be represented by the cyclic permutation matrix [31], [28],

$$A = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 1 & 0 \end{bmatrix} = V \Lambda V^{-1}, \qquad (7)$$

where the eigenvector matrix

$$V = \begin{bmatrix} v_0 & v_1 & \cdots & v_{N-1} \end{bmatrix} = \left[\frac{1}{\sqrt{N}} (W^{jk})^* \right]_{j,k=0,\cdots N-1}, \qquad (8)$$

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

$$\Lambda = \text{diag} \begin{bmatrix} W^0 & W^1 & \cdots & W^{N-1} \end{bmatrix}, \qquad (9)$$

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 $K \in \{0, 1, \cdots, N-1\}$ such that its discrete Fourier transform $\widehat{x}$ satisfies

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

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 $\text{BL}_K(\text{F})$, with F as 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. Denote $F^*_{(K)}$ be the first K columns of F^*, $\Psi \in \mathbb{R}^{M \times N}$ as a sampling operator, and $\Phi \in \mathbb{R}^{N \times M}$ as an interpolation operator.

Theorem 3. Let Ψ satisfy that the sampling number is no less than the bandwidth, $M \geq K$. For all $x \in \text{BL}_K(\text{F})$, 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.

From Theorem 3, 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 done 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 (9), by putting the frequencies with even indices first,

$$\widetilde{\Lambda} = \text{diag} \begin{bmatrix} \lambda_0 & \lambda_2 & \cdots & \lambda_{N-2} & \lambda_1 & \lambda_3 & \cdots & \lambda_{N-1} \end{bmatrix}.$$

Correspondingly, reorder the columns of V in (8) by putting the columns with even indices first

$$\widetilde{V} = \begin{bmatrix} v_0 & v_2 & \cdots & v_{N-2} & v_1 & v_3 & \cdots & v_{N-1} \end{bmatrix}.$$

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

$$\widetilde{\Lambda}_{(N/2)} = \text{diag} \begin{bmatrix} \lambda_0 & \lambda_2 & \cdots & \lambda_{N-2} \end{bmatrix}.$$

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

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

which is the Hermitian transpose of the discrete Fourier transform of size $N/2$ and satisfies $\text{rank}(\Psi \widetilde{V}_{(N/2)}) = N/2$ in Theorem 2. The sampled graph Fourier transform matrix
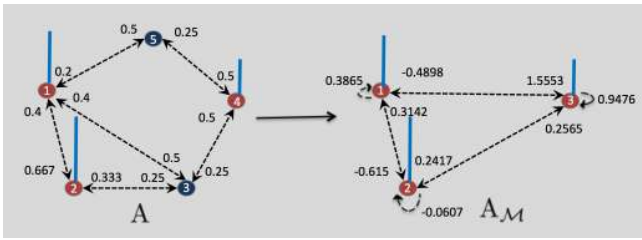
Fig. 3: Sampling a graph.

$U = (\Psi \widetilde{V}_{(N/2)})^{-1}$ is the discrete Fourier transform of size $N/2$. The sampled graph shift is then constructed as

$$A_{\mathcal{M}} = U^{-1} \widetilde{\Lambda}_{(N/2)} U,$$

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 contents.

B. Relation to Graph Laplacian Based Sampling Theory

The authors in [12] consider a similar problem to ours, that is, to recover bandlimited graph signals by sampling a few signal coefficients in the vertex domain. The differences are as follows: (1) when defining the bandwidth, we focus on the number of frequencies, not the value of frequencies; (2) the proposed sampling theory works for both directed and undirected graphs and is consistent with the classical sampling theory when handling finite discrete-time signals; (3) to find a qualified sampling operator, we need independent rows in a matrix, which can be solved by fast algorithms, such as QR decomposition; in [12], a non-convex optimization problem needs to be solved by using a greedy algorithm.

C. Relation to Frames with Maximal Robustness to Erasures

A frame $\{f_0, f_2, \cdots, f_{N-1}\}$ is a generating system for $\mathbb{C}^K$, where $N \geq K$, 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 $N - K$ rows) is invertible [37]. In [37], the authors show that a polynomial transform matrix is a frame maximally robust to erasures; in [38], 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 [39]. 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 ith 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 3. In other words, perfect recovery is guaranteed when we randomly sample a sufficient number signal coefficients.

D. Relation to Compressed Sensing

Compressed sensing is a sampling framework to recover sparse signals in a few measurements [40]. 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 $\Psi \in \mathbb{R}^{M \times N}, M << N$ and the sampled signal $x_{\mathcal{M}} = \Psi x$, a sparse signal $x \in \mathbb{R}^N$ is recovered by solving

$$\min_x \|x\|_0, \quad \text{subject to } x_{\mathcal{M}} = \Psi x. \tag{10}$$

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 [41].

In the standard compressed sensing theory, the signals have to be sparse or approximately sparse to gurantee accurate recovery properties. In [42], the authors proposed a general way to perform compressed sensing with non-sparse signals using dictionaries. To be more specific, a general signal $x \in \mathbb{R}^N$, is recovered by

$$\min_x \|D x\|_0, \quad \text{subject to } x_{\mathcal{M}} = \Psi x, \tag{11}$$

where D is a dictionary designed to make $D x$ 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, $D x$ represents the frequency content of x, which is sparse when x is of limited bandwidth. Equation (11) 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 nonzero frequencies are known, and can be reordered to form a bandlimited graph signal. Compressed sensing deals with the cases where the nonzero

frequencies are unknown, which is a more general and harder problem. If we have access to the position of the nonzero frequencies, the proposed sampling theory uses the smallest number of samples to achieve perfect recovery.

E. Optimal Sampling Operator

As mentioned in Section III-A, 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 an 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 smallest spectral norm. This is equivalent to maximizing the smallest singular value of $\Psi V_{(K)}$,

$$\Psi^{opt} = \arg \max_{\Psi} \sigma_{\min}(\Psi V_{(K)}), \qquad (12)$$

where $\sigma_{\min}$ denotes the smallest singular value. Since we restrict the form of Ψ in (4), (12) is non-deterministic polynomial-time hard. To solve (12), we can use a greedy algorithm as shown in Algorithm 1. Note that $\mathcal{M}$ is the sampling sequence, indicating which row 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 $\Psi V_{(K)}$ is growing, 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
	M	the number of samples
Output	$\mathcal{M}$	sampling set

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

F. Random Sampling

In Section IV-A, we saw that when sampling enough signal coefficients, any sampling operator leads to perfect recovery for discrete-time signals. Here we show that similar results apply to Erdős-Rényi random graphs.

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

Lemma 1. Let a graph shift $A \in \mathbb{C}^{N \times N}$ 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 function. Let V be the eigenvector matrix of A. Let the sampling number satisfy

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

for some positive constants C_1, C_2. Then,

$$P\left(\left\|\frac{1}{M}(\Psi V_{(K)})^T (\Psi V_{(K)}) - I\right\|_2 \leq \frac{1}{2}\right) \leq 1 - \delta, \quad (13)$$

for all sampling operators Ψ that sample M signal coefficients.

Proof. Since the graph shift A is real and symmetric, the eigenvector matrix V is unitary and satisfies

$$\max_{i,j} |V_{i,j}| = O\left(\sqrt{\log^{2.2} g(N) \log N / (N^2 p)}\right)$$

for $p = g(N) \log(N)/N$ [43]. By substituting V into Theorem 1.2 in [44] and obtain (13). □

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 $x \in \mathbb{C}^K$,

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

□

From Theorem 4, we see that the singular values of $\Psi V_{(K)}$ are well bounded with high probability. It 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 sufficient signal coefficients.

G. Graph Downsampling & Graph Filter Banks

In classical signal processing, sampling refers to sample a continuous function and downsampling refers to sample a sequence. Both concepts consider to use fewer samples to represent the overall shape of the original signal. Since a graph signal is discrete in nature, sampling and downsampling are

actually the same thing. Previous works implemented graph downsampling via graph coloring [6], or minimum spanning tree [45]. Those algorithms, however, do not have solid justifications from the perspective of signal processing.

The proposed sampling theory provides a family of qualified sampling operators (5) with an optimal one (12). To do graph downsampling by 2, one can set the bandwidth as a half of the number of nodes, that is, $K = N/2$, and use (12) to obtain an optimal sampling operator. An example for the finite discrete-time signals has been shown in Section IV-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 graph downsampling is a key component.

Let x be a full-band graph signal, which we can express as the addition of two bandlimited signals supported on the same graph, i. e., $x = x^l + x^h$, where

$$x^l = V \begin{bmatrix} I_K & 0 \\ 0 & 0 \end{bmatrix} V^{-1} x,$$

and

$$x^h = V \begin{bmatrix} 0 & 0 \\ 0 & I_{N-K} \end{bmatrix} V^{-1} x.$$

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 the paradigm in Theorems 1 and 2, we use a feasible sampling operator Ψ^l to sample x^l, and obtain the sampled signal coefficients as $x^l_{\mathcal{M}^l} = \Psi^l x^l$, with the corresponding graph as $A_{\mathcal{M}^l}$. We can recover x^l by using interpolation operator Φ^l as $x^l = \Phi^l x^l_{\mathcal{M}^l}$. Finally, we add the results from both channels to obtain the original full-band graph signal (also illustrated in Figure 4). We do not restrict that the samples from two bands, $x^l_{\mathcal{M}^l}$ and $x^h_{\mathcal{M}^h}$, have the same size because we can adaptively design the sampling and interpolation operators based on the their own sizes. This is similar to the filter banks in the classical literature that channels need not to evenly split the spectrum [46].

We see that the above idea can easily be generalized to multiple channels by splitting the original graphs signal into multiple bandlimited graph signals; instead of dealing with a huge graph, we work with multiple small graphs, which makes computation easier.

V. Experiments

In this section, we validate the proposed sampling theory on two classical types of graphs, Erdős-Rényi graphs and small-world graphs. We show that the perfect recovery is achieved in each type of graph with high probability. We then validate it on a real-world dataset of online blogs, where the perfect recovery is achieved again with high probability. We also apply the proposed sampling theory to the classification of online blogs and digit images, where we achieve similar or better performance with fewer labeled samples compared to the previous work.

A. Simulations

We aim to validate the full-rank assumption (5) of the proposed sampling theory on Erdős-Rényi graphs and small-world graphs, investigating the probability of satisfying the full-rank assumption by random sampling. Since once the full-rank assumption is satisfied, we can find a qualified sampling operator to achieve perfect recovery, we call this probability as *success rate of perfect recovery*.

1) Experimental Setup: Suppose that for each graph, we deal with the corresponding graph signals with fixed bandwidth $K = 10$. Given a graph shift, we randomly sample 10 rows from the first 10 columns of graph Fourier transform matrix, and check if 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.

Erdős-Rényi graphs. As shown in Section IV.C, with high probability, perfect recovery is achieved for Erdős-Rényi graph signals when we randomly sample a sufficient number of signal coefficients. We verify this result experimentally by randomly sampling Erdős-Rényi graphs with various sizes and connection probabilities. We vary the size 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.

Small-world graphs. A small-world graph is a graph where any node can reach most other nodes within a small number of steps [1], [2]. In the context of a social network, this results in the small world phenomenon of people being linked by a small number of mutual acquaintances or connections. Many empirical graphs that we encounter in the real world show this small-world phenomenon; online social networks and gene networks are examples. We use the Watts-Strogatz model to generate such graphs, which includes three variables, size, connection probability, and rewiring probability [47]. We vary the size from 50 to 500; the connection probabilities with an interval of 0.01 from 0 to 0.5, and fix the rewiring probability to be 0.1. For each given size and connection probability, we generate 100 graphs randomly.

2) Results: Figure 5 shows success rates for size averaged over 100 random tests for each of three types of graphs. When we fix the size of graphs, in Erdős-Rényi graphs, the success rates increase as the connection probability increases, i.e., more connections lead to higher probability to get a qualified sampling operator; in small-worlds graphs, the success rates increase as the connection probability increases, i.e., more connections lead to higher probability to get a qualified sampling operator. The simulation results suggest that the full-rank assumption is easier to satisfy when there exist more connections on graphs. There is no deep understanding of the eigenvectors of random graphs, it is unclear how the connections on random graphs influence the corresponding graph Fourier transform, the intuition is that more connections leads to more information in the graph Fourier transform matrix, making the rows less dependent, and easier to satisfy the full-rank assumption.

When we compare the different sizes of the same type of graph, the success rate increases as the size increases,
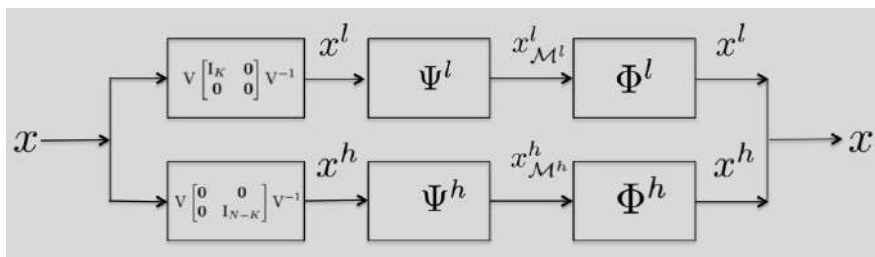
Fig. 4: Graph filter bank that splits the graph signal into two bandlimited graph signals. In each channel, we perform sampling and interpolation, following the paradigm in Theorem 1. Finally, we add the results from both channels to obtain the original full-band graph signal.
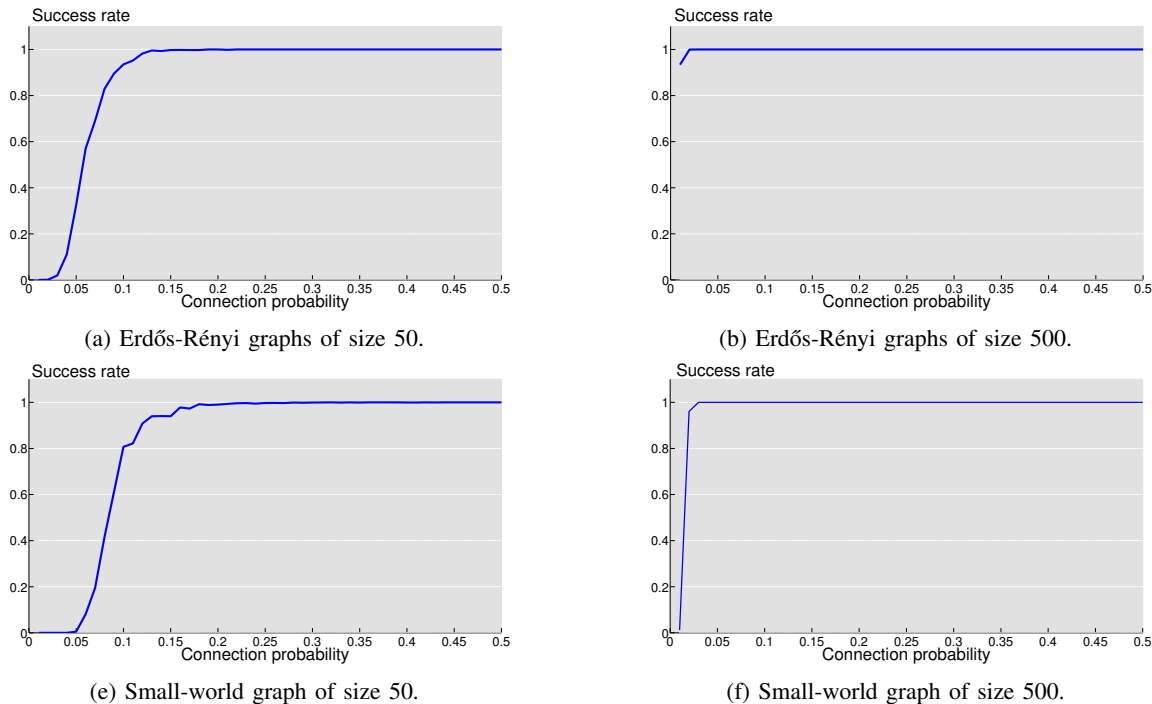

(a) Erdős-Rényi graphs of size 50.

(b) Erdős-Rényi graphs of size 500.

(e) Small-world graph of size 50.

(f) Small-world graph of size 500.

Fig. 5: Success rates.

i.e., larger sizes of graphs lead to higher probabilities of getting a qualified sampling operator. Overall, with sufficient connections, the success rates are close to 100% for each type of graphs. The intuition is that larger size of the graph leads to more information in the graph Fourier transform matrix, which makes the rows more independent. It is then easier to satisfy the full-rank assumption.

B. Sampling Online Blogs

We aim to validate the full-rank assumption of the proposed sampling theory on online blogs, investigating the success rate of perfect recovery using random sampling, and further classifying the labels of the online blogs.

1) Dataset: We consider a dataset of $N = 1224$ online political blogs as either conservative or liberal [49]. 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. The main information is preserved in the low frequencies.

2) Experimental Setup & Results: To investigate the success rate of perfect recovery 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 if 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 6 (a) shows the resulting success rate. We see that the success rates decrease as we increase the bandwidth, but the success rates are all above 90% when the bandwidth is no greater than 20. As the bandwidth increases, even if we get an equal number of samples, the success rate still decreases, because more information is embedded in a graph signal with

wider bandwidth, it is thus harder to find a qualified sampling operator.

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-A. 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 label all the blogs by actively querying the labels of a few blogs, which is a task of semi-supervised classification [50]. Note that, instead of updating the next query based on previous responses, all the queries here are designed in advance.

Most of the time, however, the labeling signal is not bandlimited, and it is infeasible 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 use the optimal sampling operator in Algorithm 1, and solve the following optimization problem to recover the low frequency content,

$$\widehat{x}_{(K)}^{opt} = \arg\min_{\widehat{x}_{(K)}\in\mathbb{R}^K} \left\| \mathrm{sgn}(\Psi\,\mathrm{V}_{(K)}\,\widehat{x}_{(K)}) - x_{\mathcal{M}} \right\|_2^2, \quad (14)$$

where $\Psi \in \mathbb{R}^{M\times N}$ is the optimal sampling operator, $x_{\mathcal{M}} \in \mathbb{R}^M$ is a vector of the sampled labels whose element is either $+1$ or -1, and $\mathrm{sgn}(*)$ sets all positive values to $+1$, and all negative values to -1. Note that without $\mathrm{sgn}(*)$, the solution of (14) is $(\Psi\,\mathrm{V}_{(K)})^{-1}x_{\mathcal{M}}$ in Theorem 1, which perfectly recovers the labeling signal when it is bandlimited. When the labeling signal is not bandlimited, the solution of (14) approximates the low frequency content. The ℓ_2 norm (14) can be relaxed by the logit function and solved by logistic regression [51]. The recovered labels are then $x^{opt} = \mathrm{sgn}(\mathrm{V}_{(K)}\,\widehat{x}_{(K)}^{opt})$.

Figure 6 (b) shows the classification accuracy by varying the sample size with an interval of 1 from 1 to 20. We see that the classification accuracy is as high as 94.44% by sampling only two blogs. The classification accuracy gets slightly better as we increases the number of samples. Compared to the previous results [24], harmonic functions achieve 94.68% by sampling 120 blogs, graph Laplacian regularization achieves 94.62% by sampling 120 blogs, graph total variation minimization achieves 94.76% by sampling 10 blogs, and graph total variation regularization achieves 94.68% by sampling 10 blogs. The improvement is from that, instead of sampling randomly as in [24], we use the optimal sampling operator to choose samples based on the graph structure actively.

C. Semi-Supervised Classification for Handwritten Digits

We aim to use the proposed sampling theory to classify the handwritten digits, and achieve high classification accuracy with fewer samples.

1) Dataset: We work with two handwritten digit datasets, the MNIST [52] and the USPS [53]. 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 of image is normalized to the size of $28 \times 28 = 784$

pixels. The USPS dataset includes 11,000 samples in total. We use all the images in the dataset; each image is normalized to the size of $16 \times 16 = 256$ 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 $\mathrm{A}_{i,j} = \mathrm{P}_{i,j} / \sum_i \mathrm{P}_{i,j}$, with

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

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

2) Experimental Setup & Results: Similarly to Section V-B, 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 $\mathrm{X}_{i,j}$ is $+1$, indicating the membership of the ith image in the jth digit, and is -1 otherwise. We obtain the optimal sampling operator Ψ as shown in Algorithm 1. The querying samples are then $\mathrm{X}_{\mathcal{M}} = \Psi\,\mathrm{X} \in \mathbb{R}^{M\times 10}$. We recover the low frequency content as

$$\widehat{\mathrm{X}}_{(K)}^{opt} = \arg\min_{\widehat{\mathrm{X}}_{(K)}\in\mathbb{R}^{K\times 10}} \left\| \mathrm{sgn}(\Psi\,\mathrm{V}_{(K)}\,\widehat{\mathrm{X}}_{(K)}) - \mathrm{X}_{\mathcal{M}} \right\|_2^2. \tag{15}$$

We solve (15) approximately by using logistic regression and then obtain the estimated label matrix $\mathrm{X}^{opt} = \mathrm{V}_{(K)}\,\widehat{\mathrm{X}}_{(K)}^{opt} \in \mathbb{R}^{N\times 10}$, whose element $(\mathrm{X}^{opt})_{i,j}$ shows a confidence of labeling the ith image as the jth 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 7. 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 the optimal sampling operator chooses representative samples from different clusters.

Figure 8 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 to the previous results [35], in the USPS dataset, given 100 samples, the accuracy of local linear reconstruction is around 65%, the accuracy of METIS graph partitioning based heuristic is around 70%, and the accuracy of graph sampling based active semi-supervised learning is around 85%, while the proposed method achieves 91.69%.

VI. CONCLUSIONS

We propose a sampling theory for graph signals that follows the same paradigm as classical sampling theory. We
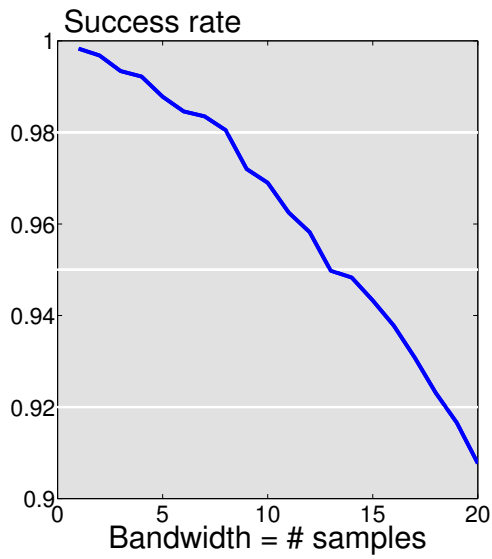
(a) Success rate as a function of the bandwidth.
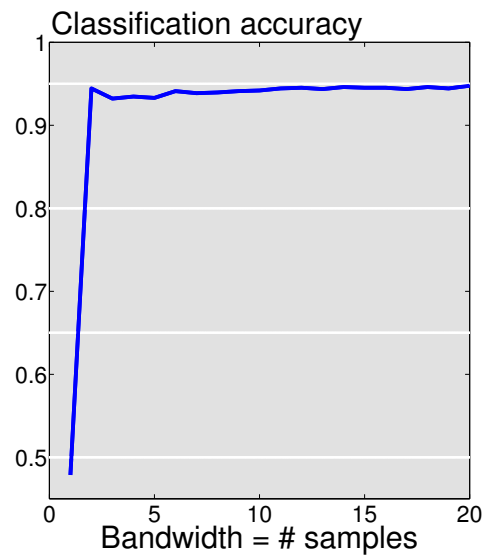
(b) Classification accuracy as a function of the bandwidth. .

Fig. 6: Online blogs When increasing the bandwidth, more information is embedded, it is thus harder to find a qualified sampling operator, but we can use more frequencies to approximate the labeling signals by taking more samples, obtaining higher classification accuracy.
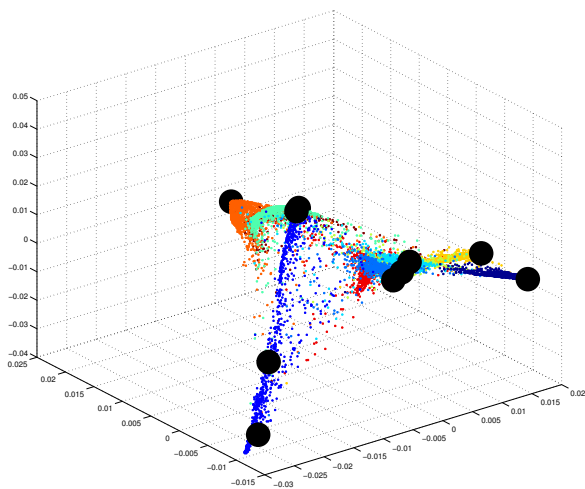
(a) MNIST.
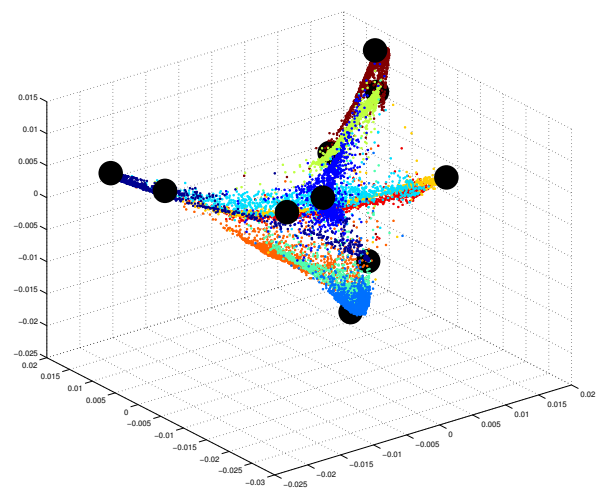
(b) USPS.

Fig. 7: 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 the optimal sampling operators in Algorithm 1.
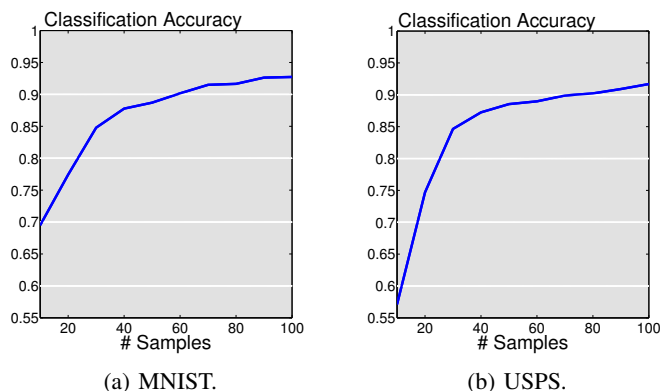
(a) MNIST.   (b) USPS.

Fig. 8: Classification accuracy of the MNIST and USPS datasets as a function of the number of querying samples.

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 frequency content. By imposing a specific structure on the graph, graph signals reduce to finite discrete-time signals, effectively ensuring that the proposed sampling theory is consistent with existing theory. We also established the connection to the theories of frames with maximal robustness to erasures and compressed sensing. We showed a principle to choose the optimal sampling operator, how random sampling works on circulant graphs and Erdős-Rényi graphs, and how to handle full-band graphs signals by using graph filter banks. We validated the proposed sampling theory on the simulated datasets of Erdős-Rényi graphs, small-world graphs, and a real-world dataset of online blogs. We showed that for each case, the proposed sampling theory achieves the perfect recovery with a high probability. Finally, we apply the proposed sampling theory to semi-supervised classification of online blogs and digit images. Compared to the previous work, we achieve similar or better performance with fewer labeled samples. Some open issues are to study the graph Fourier transform of various graphs, design optimal sampling operators efficiently, and study the recovery of noise graph signals.

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. Cerda, 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. Ekambaram, 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," *IEEE Trans. Signal Process.*, 2014, Submitted.

[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] 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.

[38] 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.

[39] V. N. Ekambaram, 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.

[40] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[41] 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.

[42] 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.

[43] 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.

[44] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 110–134, 2007.

[45] H. Q. Nguyen and M. N. Do, "Downsampling of signals on graphs via maximum spanning trees," *IEEE Trans. Signal Process.*, 2015, Submitted.

[46] M. Vetterli, "A theory of multirate filter banks," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 3, pp. 356–372, Mar. 1987.

[47] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440442, June 1998.

[48] A-L Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509512, Oct. 1999.

[49] 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.

[50] X. Zhu, "Semi-supervised learning literature survey," Tech. Rep. 1530, Univ. Wisconsin-Madison, 2005.

[51] C. M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics. Springer, 2006.

[52] 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.

[53] J. Hull, "A database for handwritten text recognition research," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 550554, May 1994.