

A NOVEL SCHEME FOR SUPPORT IDENTIFICATION AND ITERATIVE SAMPLING OF BANDLIMITED GRAPH SIGNALS

Abolfazl Hashemi[†], Rasoul Shafipour[‡], Haris Vikalo[†], and Gonzalo Mateos[‡]

[†]Dept. of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX, USA

[‡]Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627, USA

ABSTRACT

We study the problem of sampling and reconstruction of bandlimited graph signals where the objective is to select a node subset of prescribed cardinality that ensures interpolation of the original signal with the lowest reconstruction error. We propose an efficient iterative selection sampling approach and show that in the noiseless case the original signal is exactly recovered from the set of selected nodes. In the case of noisy measurements, a bound on the reconstruction error of the proposed algorithm is established. We further address the support identification of the bandlimited signal with unknown support and show that under a pragmatic sufficient condition, the proposed framework requires minimal number of samples to perfectly identify the support. The efficacy of the proposed methods is illustrated through numerical simulations on synthetic and real-world graphs.

Index Terms— graph signal processing, sampling, reconstruction, iterative algorithms.

1. INTRODUCTION

Network data naturally supported on the vertices of a graph are becoming increasingly ubiquitous, with examples ranging from measurements of neural activities at different regions of the brain [1], to vehicle trajectories over road networks [2]. Under the assumption that properties of the network process relate to the underlying graph, the goal of graph signal processing (GSP) is to broaden the scope of traditional signal processing tasks and develop algorithms that fruitfully exploit this relational structure [3, 4].

A cornerstone generalization of traditional signal processing approaches which has recently drawn considerable attention pertains to sampling and reconstruction of graph signals [5–13]. The task of finding an exact sampling set to perform reconstruction with minimal information loss is known to be NP-hard and conditions for exact reconstruction of graph signals from noiseless samples were put forth in [5–7, 13]. Existing approaches for sampling and reconstruction of graph signals can typically be categorized to two main groups of selection sampling [13] and aggregation sampling [9], where the focus of this paper is on the former.

Sampling of noise-corrupted signals using randomized schemes including uniform and leverage score sampling is studied in [13, 14], for which optimal selection sampling distributions and performance bounds are derived. Reference [15] borrows the variable density sampling strategy from compressed sensing to derive a random selection sampling scheme with optimal distribution. State-of-the-art random selection sampling schemes typically require sampling more

than k nodes of a k -band-limited graph signal to achieve near perfect recovery and their performance deteriorates for reconstruction of graph signals with relatively large bandwidth.

A main challenge in sampling and reconstruction is the problem of identifying the support of bandlimited graph signals [7, 9, 16–18]. In [7, 18], support identification of smooth graph signals is studied. However, Approaches in [7, 18] rely solely on a user-defined sampling strategy and the graph Laplacian, and disregard the availability of observations of the graph signal. A similar scheme is developed in [9] for aggregation sampling where under established assumptions on topology of the graph, conditions for exact support identification from noiseless measurements are put forth. An alternating minimization approach is proposed in [16] that jointly recovers the unknown support of the signal and designs a sampling strategy in an iterative fashion. However, the convergence of the alternating scheme in [16] is not guaranteed and the conditions for exact support identifications are unknown [16].

In this work, we consider the task of selection sampling and reconstruction of bandlimited graph signals with unknown support. Following ideas from compressed sensing, we propose a novel and efficient iterative selection sampling approach that exploits the low-cost selection criterion of the orthogonal matching pursuit algorithm [19] to recursively select a subset of nodes of the graph. We theoretically demonstrate that in the noiseless case the original k -bandlimited signal is exactly recovered from the set of selected nodes with cardinality k . When the measurements are noisy, we establish a worst-case performance bound on the reconstruction error of the proposed algorithm. We further extend our results to the case of bandlimited signals with unknown supports, and demonstrate that under a pragmatic SNR condition, the proposed framework still requires k samples to ensure exact recovery of signals with unknown supports from historical samples of the graph signal. Simulation studies show the proposed sampling algorithm compares favorably to competing random selection sampling alternatives.¹

2. PRELIMINARIES

Consider a network represented by a graph \mathcal{G} consisting of a node set \mathcal{N} of cardinality N and a weighted adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ whose (i, j) entry, \mathbf{A}_{ij} , denotes weight of the edge connecting node i to node j . A graph signal $\mathbf{x} \in \mathbb{R}^N$ is a vertex-valued network process that can be represented by a vector of size N supported on \mathcal{N} , where its i^{th} component denotes the signal value at node i . Let \mathbf{x} be a graph signal which is k -bandlimited in a given basis $\mathbf{V} \in \mathbb{R}^{N \times N}$. This means that the signal's so-called graph Fourier transform (GFT) $\bar{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$ is k -sparse. There are several choices for \mathbf{V} in the lit-

Work in this paper was supported in part by the NSF awards CCF-1750428 and ECCS-1809356.

¹Proofs of the theoretical results in this paper are omitted for brevity and stated in the extended manuscript [20].

erature with most aiming to decompose a graph signal into different modes of variation with respect to the graph topology. For instance, $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N]$ can be defined via the Jordan decomposition of the adjacency matrix [21, 22], through the eigenvectors of the Laplacian when \mathcal{G} is undirected [3], or it can be obtained as the result of an optimization procedure [23, 24]. In this paper, we assume the adjacency matrix $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ is normal which in turn implies \mathbf{V} is unitary and $\mathbf{V}^{-1} = \mathbf{V}^\top$.

Recall that since \mathbf{x} is bandlimited, $\tilde{\mathbf{x}}$ is sparse with at most k nonzero entries. Let \mathcal{K} be the support set of $\tilde{\mathbf{x}}$, where $|\mathcal{K}| = k$. Then, one can write $\mathbf{x} = \mathbf{U}\tilde{\mathbf{x}}_{\mathcal{K}}$, where $\mathbf{U} = \mathbf{V}_{\mathcal{K},c}$ and $\mathbf{V}_{\mathcal{K},c}(\mathbf{x}_{\mathcal{K}})$ is a submatrix (subvector) of $\mathbf{V}(\mathbf{x})$ that contains columns (elements) indexed by the set \mathcal{K} . Also, notation $\mathbf{V}_{\mathcal{K},r}$ will represent the submatrix that contains rows indexed by \mathcal{K} . In the sequel, we first assume that the support set \mathcal{K} is known. In section 4, we discuss how to tackle sampling scenarios where \mathcal{K} is unknown.

3. PROPOSED FRAMEWORK

In this section, we consider sampling of bandlimited signals with known support. Specifically, we assume that a graph signal \mathbf{x} is sparse given a basis \mathbf{V} . Let $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$ be the normal decomposition of \mathbf{A} , the adjacency matrix of the undirected graph \mathcal{G} . We first consider the noise-free scenario and then extend our results to the case of sampling and reconstruction from noisy measurements.

3.1. Sampling bandlimited graph signals

In selection sampling (see, e.g. [13]), sampling a graph signal amounts to finding a sampling matrix $\mathbf{C} \in \{0, 1\}^{k \times N}$, such that $\tilde{\mathbf{x}} = \mathbf{C}\mathbf{x}$, where $\tilde{\mathbf{x}}$ is the sampled graph signal. Since \mathbf{x} is bandlimited with support \mathcal{K} , and $\mathbf{x} = \mathbf{U}\tilde{\mathbf{x}}_{\mathcal{K}}$, it holds that $\tilde{\mathbf{x}} = \mathbf{C}\mathbf{U}\tilde{\mathbf{x}}_{\mathcal{K}}$. The original signal can then be reconstructed according to

$$\hat{\mathbf{x}} = \mathbf{U}\tilde{\mathbf{x}}_{\mathcal{K}} = \mathbf{U}(\mathbf{C}\mathbf{U})^{-1}\tilde{\mathbf{x}}. \quad (1)$$

According to (1), a necessary and sufficient condition for perfect reconstruction, i.e. $\hat{\mathbf{x}} = \mathbf{x}$ is invertibility of matrix $\mathbf{C}\mathbf{U}$. However, as argued in [5, 9] (see, e.g., Section III-A in [9]), current selection sampling approaches cannot construct a sampling matrix to ensure $\mathbf{C}\mathbf{U}$ is invertible for an arbitrary graph, and invertibility of $\mathbf{C}\mathbf{U}$ is checked by inspection which requires intensive computational costs for large graphs. To overcome this issue, motivated by the well-known OMP algorithm in compressed sensing [19], we propose a simple iterative scheme with complexity $\mathcal{O}(Nk^2)$ that guarantees perfect recovery of \mathbf{x} from the sampled signal $\tilde{\mathbf{x}}$.

The proposed approach (see Algorithm 1) works as follows. First, the algorithm chooses an arbitrary node of the graph with index ℓ for some $\ell \in [N]$ as a *residual node*. Then, in i^{th} iteration, the algorithm identifies a node with index j_s to be included in the sampling set \mathcal{S} according to

$$j_s = \arg \max_{j \in \mathcal{N} \setminus \mathcal{S}} \frac{|\mathbf{r}_{i-1}^\top \mathbf{u}_j|^2}{\|\mathbf{u}_j\|_2^2}, \quad (2)$$

where \mathbf{u}_j is the j^{th} row of \mathbf{U} , $\mathbf{r}_i = \mathbf{P}_{\mathcal{S}}^\perp \mathbf{u}_\ell$ is a residual vector initialized at $\mathbf{r}_0 = \mathbf{u}_\ell$, $\mathbf{P}_{\mathcal{S}}^\perp \mathbf{I}_N - \mathbf{U}_{\mathcal{S},r}^\top (\mathbf{U}_{\mathcal{S},r}^\top)^\dagger$ is the projection operator onto the orthogonal complement of the subspace spanned by the rows of $\mathbf{U}_{\mathcal{S},r}$, and $\mathbf{U}_{\mathcal{S},r}^\dagger = (\mathbf{U}_{\mathcal{S},r}^\top \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{U}_{\mathcal{S},r}$. This procedure is repeated for k iterations to construct \mathcal{S} . Theorem 1 demonstrates that Algorithm 1 returns a sampling set that ensure perfect recovery of the graph signal \mathbf{x} under noise-free scenario.

Algorithm 1 Iterative Selection Sampling

- 1: **Input:** \mathbf{U} , k , number of samples $m \geq k$.
 - 2: **Output:** Subset $\mathcal{S} \subseteq \mathcal{N}$ with $|\mathcal{S}| = m$.
 - 3: Initialize $\mathcal{S} = \emptyset$, $\mathbf{r}_0 = \mathbf{u}_\ell$ for some $\ell \in [N]$, $i = 0$.
 - 4: **while** $|\mathcal{S}| < k$
 - 5: $i \leftarrow i + 1$
 - 6: $i_s = \arg \max_{j \in \mathcal{N} \setminus \mathcal{S}} \frac{|\mathbf{r}_{i-1}^\top \mathbf{u}_j|^2}{\|\mathbf{u}_j\|_2^2}$
 - 7: Set $\mathcal{S} \leftarrow \mathcal{S} \cup \{i_s\}$
 - 8: $\mathbf{r}_i = \mathbf{P}_{\mathcal{S}}^\perp \mathbf{u}_\ell$
 - 9: **end while**
 - 10: **return** \mathcal{S} .
-

Theorem 1. *Let \mathcal{S} be the sampling set constructed by Algorithm 1 and let \mathbf{C} be the corresponding sampling matrix such that $|\mathcal{S}| = k$. Then, the matrix $\mathbf{C}\mathbf{U}$ is invertible.*

Theorem 1 states that as long as the adjacency matrix \mathbf{A} is normal and the basis \mathbf{V} exists, the proposed selection sampling scheme guarantees perfect reconstruction of the original signal from its noiseless samples. This condition guarantees recovery for a wide class of graph structures in compared to, for example, the aggregation sampling scheme [9] that requires eigenvalues of \mathbf{A} that are indexed by \mathcal{K} to be distinct.

3.2. Sampling in the presence of noise

We now provide an extended analysis of the proposed Algorithm 1 for scenarios where only noisy samples of the graph signal are available. Note that because of the noise, a perfect reconstruction is not possible. Nonetheless, we provide an upperbound on the reconstruction error of the proposed sampling scheme as a function of the noise covariance and the sampling matrix \mathbf{C} . Another different aspect of sampling and reconstruction under noise is that it might be desirable to select $m \geq k$ nodes as the sampling set to achieve better reconstruction accuracy. This is in contrast to the noiseless case where only $m = k$ is sufficient for perfect reconstruction if the sampling set is constructed by Algorithm 1 as stated in Theorem 1.

To further understand, let $\mathbf{y} = \mathbf{x} + \mathbf{n}$ be the noise-corrupted signal, where $\mathbf{n} \in \mathbb{R}^N$ is the zero-mean noise vector with covariance matrix $\mathbb{E}[\mathbf{n}\mathbf{n}^\top] = \mathbf{Q}$. Therefore, since $\mathbf{x} = \mathbf{U}\tilde{\mathbf{x}}_{\mathcal{K}}$, the samples $\tilde{\mathbf{x}}$ and the non-zero frequency components of \mathbf{x} are related via the linear model

$$\tilde{\mathbf{x}} = \mathbf{y}_{\mathcal{S}} = \mathbf{U}_{\mathcal{S},r} \tilde{\mathbf{x}}_{\mathcal{K}} + \mathbf{n}_{\mathcal{S}}, \quad (3)$$

where $\mathbf{U}_{\mathcal{S},r} = \mathbf{C}\mathbf{U}$, $\mathbf{y}_{\mathcal{S}} = \mathbf{C}\mathbf{y}$, and $\mathbf{n}_{\mathcal{S}} = \mathbf{C}\mathbf{n}$. The reconstructed signal in the Fourier domain (denoted by $\hat{\mathbf{x}}$) satisfies the normal equation [25],

$$\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r} \hat{\mathbf{x}} = \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \tilde{\mathbf{x}}, \quad (4)$$

where $\mathbf{Q}_{\mathcal{S}} = \mathbf{C}\mathbf{Q}\mathbf{C}^\top$ is the covariance of $\mathbf{n}_{\mathcal{S}}$. A necessary condition for recovery is invertibility of the matrix $\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r}$. Indeed, as stated in the following theorem, if \mathcal{S} is selected using Algorithm 1, $\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r}$ is invertible and we can recover the original graph signal, up to an error term.

Proposition 1. *Let \mathcal{S} be the sampling set constructed by Algorithm 1, \mathbf{C} be the corresponding sampling matrix and denote $\mathbf{U}_{\mathcal{S},r} = \mathbf{C}\mathbf{U}$. Then, matrix $\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r}$ is invertible and if $\|\mathbf{n}\|_2 \leq \epsilon_{\mathbf{n}}$, then reconstruction error of the signal reconstructed from \mathcal{S} satisfies*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \sigma_{\max}((\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1}) \epsilon_{\mathbf{n}}, \quad (5)$$

where $\sigma_{\max}(\cdot)$ outputs the maximum singular value of a matrix.

Proposition 1 states an explicit bound on the reconstruction error of the proposed sampling scheme for general noise models with bounded l_2 -norm. Also, the proposed selection sampling scheme preserves the structural properties of noise's statistics. More specifically, if \mathbf{n} is white with $\mathbf{Q} = \sigma^2 \mathbf{I}_N$, the effective noise remains unchanged and hence white which is in contrast to, for example, aggregation sampling that the effective noise becomes correlated; see Section IV-A in [9].

4. SUPPORT RECOVERY FROM HISTORICAL SIGNALS

So far, we have assumed that the support of the bandlimited graph signal \mathbf{x} is known. However, in many practical applications the support of the signal might be unknown and one needs to recover the support prior to or concurrent with the sampling step. In this section, we address the problem of identifying the support of bandlimited graph signals from a minimal number of (fully) observed signals at all nodes prior to sampling. Note that cost of evaluating the signal value at all nodes of the graph stems from accommodating identification of unknown supports without making any specific assumptions on the structure of the graph.

To that end, for a given graph and its Fourier basis \mathbf{V} , suppose that we observe P signals collected in a matrix $\mathbf{Y} = [\mathbf{y}^1, \dots, \mathbf{y}^P] \in \mathbb{R}^{N \times P}$. The goal is to perform support recovery of the underlying bandlimited signals collected in $\mathbf{X} = [\mathbf{x}^1, \dots, \mathbf{x}^P] \in \mathbb{R}^{N \times P}$ which amounts to estimating sparse GFTs $\bar{\mathbf{X}} = [\bar{\mathbf{x}}^1, \dots, \bar{\mathbf{x}}^P] = \mathbf{V}^T \mathbf{X}$. Pragmatically, we assume that \mathbf{V} has full rank and the GFTs $\bar{\mathbf{x}}^i$ share a common support. The latter assumption specifies matrix $\bar{\mathbf{X}}$ in a block-sparse model which has entire rows as zeros or nonzeros. Recall that \mathcal{K} is the support set of GFTs $\bar{\mathbf{x}}^i$ with $|\mathcal{K}| = k$. Then, upon defining

$$\Psi_{\mathcal{K}} := \{ \bar{\mathbf{X}} = [\bar{\mathbf{x}}^1, \dots, \bar{\mathbf{x}}^P] \in \mathbb{R}^{N \times P} \text{ such that } \bar{\mathbf{X}}_{\mathcal{K}, r} = \mathbf{0} \text{ for } i \notin \mathcal{K}, \mathcal{K} \subset [N], |\mathcal{K}| = k \}, \quad (6)$$

the bandlimited signal recovery (thus support \mathcal{K}) boils down to solving

$$\min_{\bar{\mathbf{X}}} \|\bar{\mathbf{X}} - \bar{\mathbf{Y}}\|_F^2 \text{ s.t. } \bar{\mathbf{X}} \in \Psi_{\mathcal{K}}, \quad (7)$$

where $\bar{\mathbf{Y}} = \mathbf{V}^T \mathbf{Y}$. Note that support recovery is a byproduct of finding the optimum in (7). Inspired by [26], we propose to use (2, 0) mixed norm to reformulate (7) as

$$\min_{\bar{\mathbf{X}}} \|\bar{\mathbf{X}} - \bar{\mathbf{Y}}\|_F^2 \text{ s.t. } \|\bar{\mathbf{X}}\|_{2,0} \leq k, \quad (8)$$

where the (p, q) mixed norm of matrix $\bar{\mathbf{X}}$ is defined as

$$\|\bar{\mathbf{X}}\|_{p,q} := \left(\sum_{i=1}^N \|\bar{\mathbf{x}}^i\|_p^q \right)^{1/q}. \quad (9)$$

Note that (2, 0) norm counts the number of nonzero rows in $\bar{\mathbf{X}}$.

It is immediately apparent that the solution $\bar{\mathbf{X}}^*$ of (8) is obtained by the row-wise l_2 norm thresholding on $\bar{\mathbf{Y}}$. Specifically, upon calculating the l_2 norm of the rows of $\bar{\mathbf{Y}}$ (i.e., $\|\bar{\mathbf{Y}}(i, :)\|_2$ for all i , where $\bar{\mathbf{Y}}(i, :)$ denotes the i th row of $\bar{\mathbf{Y}}$), let ζ be the k^{th} largest l_2 norm. Then, the solution $\bar{\mathbf{X}}^*$ has rows given by

$$\bar{\mathbf{X}}(i, :)^* = \begin{cases} \bar{\mathbf{Y}}(i, :), & \|\bar{\mathbf{Y}}(i, :)\|_2 \geq \zeta \\ \mathbf{0} & \text{o.w.} \end{cases} \quad (10)$$

The running time of finding the solution of (8) according to (10) is $\mathcal{O}(NP + N \log N)$. This comprises of finding the l_2 norm of the rows

in $\mathcal{O}(NP)$ and then performing an off-the-shelf sorting algorithm (e.g., merge sort) in $\mathcal{O}(N \log N)$.

Analyzing support identifiability using the proposed scheme is challenging in the presence of noise. To that end, we impose the energy constraint on the noise signals so not to exceed a threshold ϵ_n . Specifically, in the following theorem, we provide sufficient conditions to exactly recover the support from the noisy observations under the energy-constrained noise model.

Theorem 2. Consider support recovery of P bandlimited graph signals $\mathbf{X} = \mathbf{V}\bar{\mathbf{X}} = \mathbf{V}[\bar{\mathbf{x}}^1, \dots, \bar{\mathbf{x}}^P] \in \mathbb{R}^{N \times P}$ with unknown shared support \mathcal{K} with $|\mathcal{K}| = k$; i.e., $\bar{\mathbf{X}} \in \Psi_{\mathcal{K}}$. Let $\mathbf{Y} = [\mathbf{y}^1, \dots, \mathbf{y}^P] = \mathbf{X} + \mathbf{N}$ denote the noisy measurements of graph signals \mathbf{X} , where $\mathbf{N} = [\mathbf{n}^1, \dots, \mathbf{n}^P] \in \mathbb{R}^{N \times P}$ models the corruption noise signals. Assume \mathbf{V} is orthogonal and \mathbf{n}^i 's have bounded l_2 norm; in particular $\|\mathbf{n}^i\|_2 \leq \epsilon_n$ for all $i \in [P]$. Then, GFTs $\bar{\mathbf{X}}$ and, as a byproduct, the support \mathcal{K} are identifiable in (7) if

$$\min_{i \in \mathcal{K}} \|\bar{\mathbf{X}}(i, :)\|_2 > 2\epsilon_n \sqrt{P}. \quad (11)$$

Inspection of (11) shows that in the noiseless scenario (i.e., $\epsilon_n = 0$), the support is identifiable even for $P = 1$. Moreover, in the presence of noise, it is conceivable that as P grows, the chance of (11) being satisfied increases and the support recovery problem is rendered identifiable for more observations; see also the numerical tests in Section 5. Naturally, another insightful identifiability result for stochastic noise models would be valuable, but left as future work.

Compared to existing sampling schemes capable of support identification in noiseless scenario, e.g. [9] that require twice as many samples as the bandwidth of the graph signal (i.e., k) for perfect reconstruction of \mathbf{x} and further conditions on the structure of the graph – e.g., distinct nonzero eigenvalues of the graph shift operator (adjacency \mathbf{A} or the Laplacian \mathbf{L} defined as $\mathbf{L} = \text{diag}(\mathbf{A}\mathbf{1}_N) - \mathbf{A}$, for example) – the proposed framework [cf. (10)] needs only k samples of the nodes of the graph to achieve perfect recovery and identification of \mathcal{K} . Additionally, the proposed framework is capable of support identification for the broad range of graphs which have normal graph shift operator. Adaptive sampling and reconstruction of graph signals with unknown supports can be achieved as a result of combining the proposed support identification framework with the proposed sampling scheme outlined in Algorithm 1, and is left as a subject of our future studies.

5. SIMULATION RESULTS

We study the recovery of signals supported on synthetic and real-world graphs to assess performance of the proposed support recovery and sampling algorithms. To this end, we first consider undirected Erdős-Rényi random graphs \mathcal{G} of size $N = 1000$ and edge probability 0.15 or 0.25 [27]. Bandlimited graph signals $\mathbf{x} = \mathbf{U}\bar{\mathbf{x}}_{\mathcal{K}}$ are generated by taking \mathbf{U} as the k randomly selected eigenvectors \mathbf{V} of the graph adjacency matrix, where $k = 150$ or 300. The non-zero frequency components $\bar{\mathbf{x}}_{\mathcal{K}}$ are drawn independently from a zero-mean Gaussian distribution with standard deviation 100. We also corrupt the signals (measurements) by additive Gaussian noise with 20dB power. We first start by observing P signals across all nodes when the frequency support is unknown and try to recover the support using the proposed formulation in (8), (10). Fig. 1(a) depicts the support recovery error as the ratio of number of elements common in the ground truth and inferred frequency support to the bandwidth (k) as a function of P , where the results are obtained by averaging over 100 Monte-Carlo simulations. We notice that as P increases, the recovery error decreases monotonically [cf. Theorem 2] for different

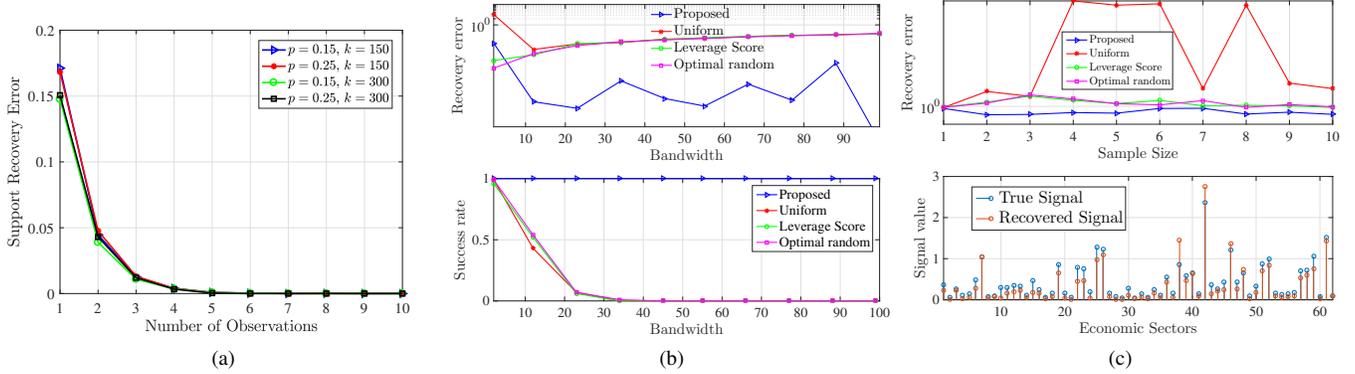


Fig. 1: (a) Support recovery error of the proposed scheme versus number of observations for different degrees of connectivities (p) and bandwidths (k). (b) Recovery error (top) and success rate (bottom) comparison of Algorithm 1 and different random selection sampling schemes versus bandwidth (k) for undirected Erdős-Rényi random graphs. (c) Top: Recovery error comparison of different selection sampling schemes as a function of sample size for the economy network. Bottom: Recovered and true graph signal at different economic sectors using Algorithm 1.

degrees of connectivity (p) and bandwidths (k). As expected, the recovery performance does not rely on the edge probabilities, since we only need \mathbf{V} to be full rank which is satisfied in all undirected graphs considered in this simulation study. Moreover, since with higher bandwidth the energy in GFT components and the chance of satisfying (11) increases, the support recovery task becomes easier, as predicted by the results of Theorem 2.

Next, we consider the task of sampling and reconstruction of noise-corrupted bandlimited graph signals with known support. Specifically, we consider undirected Erdős-Rényi random graphs \mathcal{G} of size $N = 100$ and edge probability 0.2. We generate $\mathbf{x} = \mathbf{U}\bar{\mathbf{x}}_K$ by taking \mathbf{U} the first k eigenvectors \mathbf{V} of the graph adjacency matrix, where we vary k linearly from 2 to 99. The non-zero frequency components $\bar{\mathbf{x}}_K$ are drawn independently from a zero-mean Gaussian distribution with standard deviation 100 and the signal is corrupted with a Gaussian noise term with $\mathbf{Q} = 0.02^2\mathbf{I}_N$. We compare the recovery performance of the proposed scheme in Algorithm 1 with the state-of-the-art uniform, leverage score, and optimal random sampling schemes [13–15]. We define the recovery error as the ratio of the error energy to the true signal’s energy. Furthermore, *success rate* [13] is defined as the fraction of instances where \mathbf{CU} is invertible [cf. (1)]. The results, averaged over 100 independent instances, are illustrated in Fig 1(b). As we can see from Fig 1(b)(top), the proposed scheme consistently achieves a lower recovery error compared to competing schemes. Moreover, as Fig 1(b)(bottom) illustrates, when the bandwidth increases success rate of random sampling schemes decreases while the success rate of proposed scheme is always one, as we proved in Theorem 1.

Finally, we use the data from Bureau of Economic Analysis of the U.S. Department of Commerce which publicizes an annual table of input and outputs organized by economic sectors². Specifically, we use 62 industrial sectors as defined by the North American Industry Classification System as nodes and construct the weighted edges and the graph signal similar to [9]. To that end, the (undirected) edge weight between the two nodes represents the average total production of the sectors, the first sector being used as the input of the other sector, expressed in trillions of dollars per year. This edge weight is averaged over the years 2008, 2009, and 2010. Also, two artificial nodes are connected to all 62 nodes as the added value generated and the level of production destined to the market of final users. Thus, the final graph has $N = 64$ nodes. The weights lower than 0.01 are then thresholded to zero and the eigenvalue decomposition of the

corresponding adjacency matrix $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$ is obtained. A graph signal $\mathbf{x} \in \mathbb{R}^{64}$ can be regarded as a unidimensional total production – in trillion of dollars – of each sector during the year 2011. Signal \mathbf{x} is shown to be approximately (low-pass) bandlimited in [9, Fig. 4(a)(top)] with a bandwidth of 4.

We try to interpolate sectors’ production by observing a few nodes using Algorithm 1 assuming that the signal is low-pass (i.e., with smooth variations over the built network). Then, we vary the sample size and compare the recovery performance of the proposed scheme with the state-of-the-art uniform, leverage score, and optimal random sampling schemes [13–15] averaged over 1000 Monte-Carlo simulations as shown in Fig. 1(c)(top). As the figure indicates, the proposed algorithm outperforms uniform, leverage score, and optimal random sampling schemes [13–15]. However, we are not experiencing perfect recovery for our proposed Algorithm 1 in this noiseless scenario because the signal is not purely bandlimited. Moreover, Fig. 1(c)(bottom) shows a realization of the graph signal \mathbf{x} superimposed with the reconstructed signal obtained using Algorithm 1 with $k = 2$ for all nodes excluding two artificial ones. The recovery error of the reconstructed signal is approximately 1.32% and as Fig. 1(c)(bottom) illustrates, $\hat{\mathbf{x}}$ closely approximates \mathbf{x} .

6. ACKNOWLEDGMENT

We would like to thank the authors in [9] for providing the data used for the economy network in the last experiment.

7. CONCLUSION

We considered the task of sampling and reconstruction of k -bandlimited graph signals. We proposed an efficient iterative sampling approach that exploits the low-cost selection criterion of the orthogonal matching pursuit algorithm to recursively select a subset of nodes of the graph. We also theoretically showed that in the noiseless case the original k -bandlimited signal is exactly recovered from the set of selected nodes with cardinality k . In the case of noisy measurements, we established a worst-case performance bound on the reconstruction error of the proposed algorithm. We further extended our results to the case where the support of the bandlimited signal is unknown and demonstrated under a mild SNR condition, the proposed framework still requires k samples to ensure exact recovery of signals with unknown supports from historical samples of the graph signal. Simulation studies showed that the proposed sampling algorithm compares favorably to competing alternatives.

²Dataset from <https://www.bea.gov>

8. REFERENCES

- [1] W. Huang, L. Goldsberry, N. F. Wymbs, S. T. Grafton, D. S. Bassett, and A. Ribeiro, "Graph frequency analysis of brain signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 7, pp. 1189–1203, Oct. 2016.
- [2] J. A. Deri and J. M. Moura, "New York City taxi analysis with graph signal processing," in *Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1275–1279, IEEE, Dec. 2016.
- [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," *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 83–98, May 2013.
- [4] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
- [5] H. Shomorony and A. S. Avestimehr, "Sampling large data on graphs," in *Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 933–936, IEEE, 2014.
- [6] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.
- [7] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, Jul. 2016.
- [8] S. P. Chepuri and G. Leus, "Subsampling for graph power spectrum estimation," in *Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 1–5, IEEE, 2016.
- [9] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1832–1843, Apr. 2016.
- [10] F. Gama, A. G. Marques, G. Mateos, and A. Ribeiro, "Rethinking sketching as sampling: Linear transforms of graph signals," in *Asilomar Conference on Signals, Systems and Computers*, pp. 522–526, IEEE, 2016.
- [11] L. F. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 34–47, Jan. 2018.
- [12] A. Hashemi, R. Shafipour, H. Vikalo, and G. Mateos, "Sampling and reconstruction of graph signals via weak submodularity and semidefinite relaxation," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4179–4183, IEEE, Apr. 2018.
- [13] S. Chen, R. Varma, A. Sandryhaila, and J. Kovacevic, "Discrete signal processing on graphs: Sampling theory," *IEEE Transactions on Signal Processing*, vol. 24, no. 63, pp. 6510–6523, Dec. 2015.
- [14] S. Chen, R. Varma, A. Singh, and J. Kovacevic, "Signal recovery on graphs: Fundamental limits of sampling strategies," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 539–554, Dec. 2016.
- [15] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Applied and Computational Harmonic Analysis*, Mar. 2018.
- [16] P. Di Lorenzo, S. Barbarossa, P. Banelli, and S. Sardellitti, "Adaptive least mean squares estimation of graph signals," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 555–568, Dec. 2016.
- [17] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 764–778, Feb. 2017.
- [18] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5445–5449, IEEE, May 2013.
- [19] Y. C. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Asilomar Conference on Signals, Systems and Computers*, pp. 40–44, IEEE, 1993.
- [20] A. Hashemi, R. Shafipour, H. Vikalo, and G. Mateos, "Efficient sampling of bandlimited graph signals," *arXiv preprint*, Jun. 2018.
- [21] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Transactions on Signal Processing*, vol. 62, pp. 3042–3054, June Jun. 2014.
- [22] J. A. Deri and J. M. Moura, "Spectral projector-based graph fourier transforms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 785–795, Sep. 2017.
- [23] R. Shafipour, A. Khodabakhsh, G. Mateos, and E. Nikolova, "A directed graph fourier transform with spread frequency components," *arXiv preprint arXiv:1804.03000*, 2018.
- [24] S. Sardellitti, S. Barbarossa, and P. Di Lorenzo, "On the graph Fourier transform for directed graphs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 796–811, Sep. 2017.
- [25] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.
- [26] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
- [27] M. Newman, *Networks: an introduction*. Oxford university press, 2010.