Signal Processing 195 (2022) 108505

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Towards accelerated greedy sampling and reconstruction of bandlimited graph signals

Abolfazl Hashemi[a,1,*], Rasoul Shafipour[b,2], Haris Vikalo[c], Gonzalo Mateos[d]

[a] School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA
[b] Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, USA
[c] Microsoft, Redmond, WA 98052, USA
[d] Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627, USA

ARTICLE INFO

Article history:
Received 7 September 2021
Revised 3 February 2022
Accepted 11 February 2022
Available online 13 February 2022

Keywords:
Graph signal processing
Sampling
Reconstruction
Weak submodularity
Iterative algorithms

ABSTRACT

We study the problem of sampling and reconstructing spectrally sparse graph signals where the objective is to select a subset of nodes of prespecified cardinality that ensures interpolation of the original signal with the lowest possible reconstruction error. This task is of critical importance in Graph signal processing (GSP) and while existing methods generally provide satisfactory performance, they typically entail a prohibitive computational cost when it comes to the study of large-scale problems. Thus, there is a need for accelerated and efficient methods tailored for high-dimensional and large-scale sampling and reconstruction tasks. To this end, we first consider a non-Bayesian scenario and propose an efficient iterative node sampling procedure that in the noiseless case enables exact recovery of the original signal from the set of selected nodes. In the case of noisy measurements, a bound on the reconstruction error of the proposed algorithm is established. Then, we consider the Bayesian scenario where we formulate the sampling task as the problem of maximizing a monotone weak submodular function, and propose a randomized-greedy algorithm to find a sub-optimal subset of informative nodes. We derive worst-case performance guarantees on the mean-square error achieved by the randomized-greedy algorithm for general non-stationary graph signals.

1. Introduction

Network data that are naturally supported on vertices of a graph are becoming increasingly ubiquitous, with examples ranging from the measurements of neural activities in different regions of the brain [3] to vehicle trajectories over road networks [4]. Predicated on the assumption that the properties of a 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 [5,6].

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 ith component denotes the signal value at node i.

A cornerstone problem in GSP that has drawn considerable attention in recent years pertains to sampling and reconstruction of graph signals [7–15]. The task of selecting a subset of nodes whose signals enable reconstruction of the information in the entire graph with minimal loss is known to be NP-hard. Conditions for exact reconstruction of graph signals from noiseless samples were put forth in [7–10]. Existing approaches for sampling and reconstruction of graph signals can be categorized in two main groups – selection sampling [10] and aggregation sampling [12]. The focus of the current paper is on the former.

1.1. Related work

Sampling of noise-corrupted signals using randomized schemes including uniform and leverage score sampling is studied in [16]; there, optimal sampling distributions and performance bounds are

* Corresponding author.
E-mail address: abolfazl@purdue.edu (A. Hashemi).
[1] Work done while with the Department of ECE, University of Texas at Austin.
[2] Work done while with the Department of ECE, University of Rochester.

https://doi.org/10.1016/j.sigpro.2022.108505
0165-1684/

Table 1
Properties of sampling schemes for spectrally sparse signals in scenarios where the basis matrix $\mathbf{U}$ is known.

Assumption	Optimality criteria	Algorithms
noise-free samples, non-Bayesian	full rank $\mathbf{U}_S$	Gaussian elimination, greedy [7], random [16,17]
noisy samples, non-Bayesian	$\min \mathrm{Tr}(\mathbb{E}[(\mathbf{x}-\hat{\mathbf{x}})((\mathbf{x}-\hat{\mathbf{x}})^\top])$	Gaussian elimination, greedy [10], random [16,17]
noisy samples, Bayesian	$\min \mathrm{Tr}(\mathbb{E}[(\mathbf{x}-\hat{\mathbf{x}})((\mathbf{x}-\hat{\mathbf{x}})^\top])$	greedy [15], convex optimization [31]

derived. Building on the ideas of variable density sampling from compressed sensing, [17] derives random sampling schemes and proves that $\mathcal{O}(k \log k)$ samples are sufficient to recover all k-spectrally sparse signals with high probability. Moreover, [17] provides a fast technique for accurate estimation of the optimal sampling distribution. Recent work [18] relies on loop-erased random walks on graphs to speed up sampling of bandlimited signals. In [11,15], reconstruction of graph signals and their power spectrum density was studied and schemes based on the greedy sensor selection algorithm [19,20] were developed. However, the performance guarantees in [15,16] are restricted to the case of stationary graph signals, i.e., the covariance matrix in the nodal or spectral domains is required to have a certain structure (e.g., diagonal; see also [21–23]).

An influential work [24] presents a method that enables recovery of some bandlimited functions on a simple undirected unweighted graph using signal values observed on the so-called uniqueness sets of vertices; see also [25] and [26]. An iterative local set-based algorithm that relies on graph partitioning to improve convergence rate of bandlimited graph signals reconstruction is proposed in [27].

The sampling approach in [12] relies on collecting observations at a single node instead of a subset of nodes via successive applications of the so-called graph shift operator and aggregating the results. Specifically, shifted versions of the signal are sampled at a single node which, under certain conditions, enables recovery of the signal at all nodes. While the aggregation sampling in [12] reduces to the classical sampling of time signals, the required inspection of the invertibility of the submatrix of eigenvectors is computationally expensive. Moreover, the recovery of graph signals from their partial samples collected via the aggregation scheme requires the first k components (signal bandwidth) to be distinct, which may not be the case in certain applications. Table 1 summarizes properties of a few

A main challenge in sampling and reconstruction of spectrally sparse graph signals is the problem of identifying their support [9,12,25,28,29]. In [9,26], support identification of smooth graph signals is studied. However, the techniques in [9,25] 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 [12] for aggregation sampling where under established assumptions on the topology of a graph, conditions for the exact support identification from noiseless measurements are established. In particular, the aggregation sampling method of [12] requires twice as many samples as the bandwidth of the graph signal (i.e., k) to guarantee perfect recovery in the noiseless setting. An alternating minimization approach that jointly recovers unknown support of the signal and designs a sampling strategy in an iterative fashion is proposed in [28]. However, convergence of the alternating scheme in [28] is not guaranteed and the conditions for exact support identification are unknown [28].

1.2. Contribution

Although tremendous efforts have been made to address fundamental theoretical and algorithmic questions in sampling and reconstruction of bandlimited graph signals, the high computational costs of existing methods that deliver competitive reconstruction performance typically render their applicability challenging, especially in applications dealing with large-scale and high-dimensional graphs. Therefore, developing scalable, efficient, and accelerated sampling and reconstruction algorithms with provable performance is highly desired.

In this paper, we consider the task of sampling and reconstruction of spectrally sparse graph signals in various settings. We first study the non-Bayesian scenario where no prior information about signal covariance is available. Based on ideas from compressed sensing, we develop a novel and efficient iterative sampling approach that exploits the low-cost selection criterion of the orthogonal matching pursuit algorithm [30] to recursively select a subset of nodes of the graph. We theoretically demonstrate that in the noiseless case the original k-spectrally sparse signal can be recovered exactly from the set of selected nodes with cardinality k. In the case of ℓ_2-norm bounded noise, we establish a bound on the worst-case reconstruction error of the proposed algorithm that turns out to be proportional to the bound on the ℓ_2-norm of the noise term. The proposed scheme requires only that the graph adjacency matrix is normal, a typical assumption in prior works on the sampling of graph signals. Therefore, the proposed iterative algorithm guarantees recovery for a wide class of graph structures.

Next, we study a Bayesian scenario where the graph signal is a non-stationary network process with a known non-diagonal covariance matrix. Following [15,19,20], we formulate the sampling task as the problem of maximizing a monotone weak submodular function that is directly related to the mean square error (MSE) of the linear estimator of the original graph signal. To find a sub-optimal solution to this combinatorial optimization problem, we propose a randomized-greedy algorithm that is significantly faster than the greedy sampling method in [15,19,20]. We theoretically analyze performance of the proposed randomized-greedy algorithm and demonstrate that the resulting MSE is a constant factor away from the MSE of the optimal sampling set. Unlike the prior work in [15], our results do not require stationarity of the graph signal. Furthermore, in contrast to the existing theoretical works, we do not restrict our study to the case of additive white noise. Instead, we assume that the noise coefficients are independent and allow the power of noise to vary across individual nodes of the graph.

Simulation studies on both synthetic and real world graphs verify our theoretical findings and illustrate that the proposed sampling framework compares favorably to competing alternatives in terms of both accuracy and runtime.

Preliminary results of this work is published in [1,2]. In addition to providing the details of proofs which were missing from Hashemi et al. [1], 2] and discussing the computational complexity of the proposed algorithms, for the Bayesian setting, we extend the scope of our study to provide *high probability* error bounds for the achievable mean-square error performance of the proposed randomized greedy sampling schemes. Finally, in our extensive experimental study, we discuss two new applications of graph sampling, namely, localization of UAVs under power constraints and semi-supervised face clustering via subspace learning. We further demonstrate the efficacy of our randomized greedy algorithm on a large-scale preferential attachment graph with 10,000 nodes.

1.3. Organization

The rest of the paper is organized as follows. Section 2 reviews the relevant background and concepts. In Section 3, we formally state the sampling problem and develop the proposed iterative selection sampling method. In Section 4, we study the Bayesian setting, introduce the randomized-greedy algorithm for the sampling task and theoretically analyze its performance. Section 5 presents simulation results while the concluding remarks are stated in Section 6.

2. Preliminaries

In this section, we overview notation, concepts, and definitions that are used in the development of the proposed algorithmic and theoretical frameworks.

2.1. Notations

Bold capital letters denote matrices while bold lowercase letters represent vectors. Sets are denoted by calligraphic letters and $|\mathcal{S}|$ denotes the cardinality of set $\mathcal{S}$. $\mathbf{A}_{ij}$ denotes the (i,j) entry of $\mathbf{A}$, $\mathbf{a}_j$ ($\mathbf{a}^j$) is the jth row (column) of $\mathbf{A}$, $\mathbf{A}_{\mathcal{S},r}$ ($\mathbf{A}_{\mathcal{S},c}$) is the submatrix of $\mathbf{A}$ that contains rows (columns) indexed by the set $\mathcal{S}$, and $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ are the largest and smallest eigenvalues of $\mathbf{A}$, respectively. $\mathbf{P}_{\mathcal{S}}^{\perp} = \mathbf{I}_n - \mathbf{A}_{\mathcal{S},r}^{\top}(\mathbf{A}_{\mathcal{S},r}^{\top})^{\dagger}$ is the projection operator onto the orthogonal complement of the subspace spanned by the rows of $\mathbf{A}_{\mathcal{S},r}$, where $\mathbf{A}^{\dagger} = \left(\mathbf{A}^{\top}\mathbf{A}\right)^{-1}\mathbf{A}^{\top}$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{A}$ and $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ is the identity matrix. Finally, supp($\mathbf{x}$) returns the support of $\mathbf{x}$ and $[n] := \{1, 2, \ldots, n\}$.

2.2. Spectrally sparse graph signals

Let $\mathbf{x}$ be a graph signal which is k-spectrally sparse 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 literature with most aiming to decompose a graph signal into different modes of variation with respect to the graph topology. For instance, $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_N]$ can be defined via the Jordan decomposition of the adjacency matrix [32,33], through the eigenvectors of the Laplacian when $\mathcal{G}$ is undirected [5], or it can be obtained as the result of an optimization procedure [34,35]. In this paper, we assume that 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 spectrally sparse, $\bar{\mathbf{x}}$ is sparse with at most k nonzero entries. Let $\mathcal{K}$ be the support set of $\bar{\mathbf{x}}$, where $|\mathcal{K}| = k$. Then one can write $\mathbf{x} = \mathbf{U}\bar{\mathbf{x}}_{\mathcal{K}}$, where $\mathbf{U} = \mathbf{V}_{\mathcal{K},c}$. In the sequel, without loss of generality we assume $\mathbf{U}$ does not contain all-zero rows; otherwise, one could omit the all-zero rows of $\mathbf{U}$ and their corresponding nodes from the graph as they provide no meaningful information about the graph signals. Moreover, we proceed by assuming that the support set $\mathcal{K}$ is known.

Remark 1. As in the prior work on sampling graph signals [10,12,13,15–17], our proposed schemes require the graph Fourier transform (GFT) bases (i.e., $\mathbf{V}$) as input; this involves eigenvalue decomposition of $\mathbf{A}$ which may be computationally intensive for large graphs. The focus of this paper, however, is not on the preprocessing step of finding $\mathbf{V}$ but rather on developing efficient sampling algorithms with theoretical performance guarantees on the achievable reconstruction error in a variety of settings.

2.3. Submodularity and weak submodular functions

An important concept in contemporary combinatorial optimization is the notion of submodular functions that has recently found applications in many signal processing tasks. Relevant concepts are formally defined below.

Definition 1 (Submodularity and monotonicity)**.** Let $\mathcal{X}$ be a ground set. Set function $f : 2^{\mathcal{X}} \to \mathbb{R}$ is submodular if

$$f(\mathcal{S}\cup\{j\}) - f(\mathcal{S}) \geq f(\mathcal{T}\cup\{j\}) - f(\mathcal{T})$$

for all subsets $\mathcal{S} \subseteq \mathcal{T} \subset \mathcal{X}$ and $j \in \mathcal{X}\backslash\mathcal{T}$. The term $f_j(\mathcal{S}) := f(\mathcal{S}\cup\{j\}) - f(\mathcal{S})$ is the marginal value of adding element j to set $\mathcal{S}$. Furthermore, f is monotone if $f(\mathcal{S}) \leq f(\mathcal{T})$ for all $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{X}$.

In many applications, the objective function of a combinatorial optimization problem of interest is not submodular. The notion of set functions with bounded curvature captures these scenarios by generalizing the concept of submodularity.

Definition 2 (Curvature)**.** The maximum element-wise curvature of a monotone non-decreasing function f is defined as

$$\mathcal{C}_f = \max_{l\in[N-1]} \max_{(\mathcal{S},\mathcal{T},i)\in\mathcal{X}_l} f_i(\mathcal{T})/f_i(\mathcal{S}),$$

where $\mathcal{X}_l = \{(\mathcal{S},\mathcal{T},i) | \mathcal{S}\subset\mathcal{T}\subset\mathcal{X}, i\in\mathcal{X}\backslash\mathcal{T}, |\mathcal{T}\backslash\mathcal{S}| = l, |\mathcal{X}| = N\}$.

The maximum element-wise curvature essentially quantifies how close the set function is to being submodular. It is worth noting that a set function $f(\mathcal{S})$ is submodular if and only if its maximum element-wise curvature satisfies $\mathcal{C}_f \leq 1$. When $\mathcal{C}_f > 1$, $f(\mathcal{S})$ is called a *weak submodular* function.

3. Sampling of spectrally sparse graph signals

In this section, we study the problem of sampling spectrally sparse signals with known support. In particular, we assume that a graph signal $\mathbf{x}$ is sparse given a basis $\mathbf{V}$ and that $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\top}$, where $\mathbf{A}$ is the adjacency matrix of the undirected graph $\mathcal{G}$; alternatively, we may use the Laplacian matrix $\mathbf{L}$ to characterize the undirected graph. We can also consider any orthogonal basis for general directed graphs; see e.g., [34]. We first consider the noise-free scenario (Section 3.1) and then extend our results to the case of sampling and reconstruction from noisy measurements (Section 3.3).

3.1. Sampling strategy

In selection sampling (see, e.g. [10]), sampling a graph signal amounts to finding a matrix $\mathbf{C} \in \{0,1\}^{k\times N}$ such that $\tilde{\mathbf{x}} = \mathbf{C}\mathbf{x}$, where $\tilde{\mathbf{x}}$ denotes the sampled graph signal. Since $\mathbf{x}$ is spectrally sparse with support $\mathcal{K}$ and $\mathbf{x} = \mathbf{U}\bar{\mathbf{x}}_{\mathcal{K}}$, it holds that $\tilde{\mathbf{x}} = \mathbf{C}\mathbf{U}\bar{\mathbf{x}}_{\mathcal{K}}$. The original signal can then be reconstructed as

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

According to (1), a necessary and sufficient condition for perfect reconstruction (i.e., $\hat{\mathbf{x}} = \mathbf{x}$) from noiseless observations is guaranteed by the invertibility of matrix $\mathbf{C}\mathbf{U}$. However, as argued in [7,12] (see, e.g. Section III-A in [12]), current random selection sampling approaches cannot construct a sampling matrix to ensure $\mathbf{C}\mathbf{U}$ is invertible for an arbitrary graph; moreover, invertibility of $\mathbf{C}\mathbf{U}$ is checked by inspection which for large graphs requires intensive computational effort. To overcome these issues, motivated by the well-known OMP algorithm in compressed sensing [30], we propose a simple iterative scheme with complexity $\mathcal{O}(Nk^2)$ that guarantees perfect recovery of $\mathbf{x}$ from the sampled signal $\tilde{\mathbf{x}}$. OMP is an iterative scheme that aims to build a full-rank matrix from a set of feature vectors by identifying vectors that add a higher expressiveness power to the current selection; the expressiveness is captured by the notion of the *residual vector* $\mathbf{r}$. Therefore, since our aim relies on invertibility of $\mathbf{C}\mathbf{U}$, we can apply OMP to select a subset of $\mathbf{U}$'s column, which by construction, will be full-rank.

The proposed approach (see Algorithm 1) works as follows. First, the algorithm chooses a node of the graph with index ℓ as a

Algorithm 1 Iterative Selection Sampling.

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

residual node such that $\ell = \arg\min_{j\in[N]} \|\mathbf{u}_j\|$. Intuitively, this node has weaker expressiveness power compared to other point and since in Algorithm 1 the residual node is excluded from the selection procedure, this choice empirically leads to smaller reconstruction error in noisy scenario. Next, in the ith iteration the algorithm identifies a node – excluding the residual node – with index s_j to be included in the sampling set $\mathcal{S}$ according to

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

where $\mathbf{r}_i = \mathbf{P}_{\mathcal{S}}^{\perp}\mathbf{u}_\ell$ is a residual vector initialized as $\mathbf{r}_0 = \mathbf{u}_\ell$, and $\mathbf{P}_{\mathcal{S}}^{\perp} = \mathbf{I}_n - \mathbf{U}_{\mathcal{S},r}^\top (\mathbf{U}_{\mathcal{S},r}^\top)^\dagger$. Note that (2) is exactly the selection criterion of the OMP algorithm. This procedure is then repeated for k iterations to construct the sampling set $\mathcal{S}$. Once $\mathcal{S}$ is found, $\mathbf{CU}$ would then be an invertible matrix, ensuring a necessary and sufficient condition for perfect recovery.

Remark 2. Optimization (2) is related to the greedy column subset selection approach in [36]. Specifically, both methods attempt to identify a subset of the rows/columns that best represent the entire matrix. However, they focus on different applications which in turn results in different definitions of the residuals. In [36], the residual is defined as the original matrix itself. Hence the computational complexity of the greedy approach in [36] is significantly higher than that of Algorithm 1 where the residual is merely a vector.

Theorem 1 demonstrates that Algorithm 1 returns a sampling set which ensures perfect recovery of the graph signal $\mathbf{x}$ in the noise-free scenario.

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

Proof. See Appendix A. □

Theorem 1 states that as long as the adjacency matrix $\mathbf{A}$ is normal, the proposed selection scheme guarantees perfect reconstruction of the original signal from its noiseless samples. Therefore, in contrast to existing random selection sampling and aggregation sampling schemes [10,12,17] that require strong conditions on $\mathbf{A}$ (e.g., eigenvalues of $\mathbf{A}$ to be distinct), Algorithm 1 guarantees recovery for a wider class of graphs.

3.2. Complexity analysis

The worst-case computational complexity of Algorithm 1 is analyzed next. In the ith iteration, step 6 costs $\mathcal{O}(k(N-i))$ as one needs to search over $N-i$ rows of $\mathbf{U}$ and compute inner-products of k-dimensional vectors in order to evaluate the selection criterion. Step 8 is a matrix-vector product whose complexity is $\mathcal{O}(k^2)$.

Table 2
Computational complexity comparison between the proposed algorithms and the existing methods.

Algorithm	Setting	Complexity
Proposed Algorithm 1	non-Bayesian	$\mathcal{O}(k^2N)$
Greedy [10]	non-Bayesian	$\mathcal{O}(k^4N)$
Greedy [15]	Bayesian	$\mathcal{O}(Nk^3)$
Proposed Algorithm 2	Bayesian	$\mathcal{O}(Nk^2)$

Note that in our implementation we use the modified Gram-Schmidt (MGS) algorithm to update the residual vector with a significantly lower complexity of $\mathcal{O}(ki)$. Thus, the total cost of the ith iteration is $\mathcal{O}(k(N-i)+ki) = \mathcal{O}(k(N-i))$. Since $i \leq k$ and there are k iterations, the overall complexity of Algorithm 1 is $\mathcal{O}(Nk^2)$. Please refer to Table 2 for a comparison between computational costs of proposed schemes in this paper to the existing methods.

3.3. Sampling in the presence of noise

Here we provide an extension of the proposed selection sampling scheme to the scenarios where only noisy observations of the graph nodes are available. Note that due to noise, perfect reconstruction is no longer possible. Nevertheless, we provide an upper bound on the reconstruction error of the proposed sampling scheme as a function of the noise covariance and the sampling matrix $\mathbf{C}$. Another distinguishing aspect of sampling and reconstruction in the presence of noise is that, to achieve better reconstruction accuracy, it may be desirable to select $m \geq k$ nodes as the sampling set. This stands in contrast to the noiseless case where, as we proved, $m = k$ sampling nodes are sufficient for perfect reconstruction if the sampling set is constructed by Algorithm 1.

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

$$\tilde{\mathbf{x}} = \mathbf{y}_{\mathcal{S}} = \mathbf{U}_{\mathcal{S},r}\bar{\mathbf{x}}_{\mathcal{K}} + \mathbf{n}_{\mathcal{S}}, \tag{3}$$

where $\mathbf{U}_{\mathcal{S},r} = \mathbf{CU}$, $\mathbf{y}_{\mathcal{S}} = \mathbf{Cy}$, and $\mathbf{n}_{\mathcal{S}} = \mathbf{Cn}$. The reconstructed signal in the Fourier domain is found by seeking the least square solution and satisfies the normal equation [37],

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

where $\mathbf{Q}_{\mathcal{S}} = \mathbf{CQC}^\top$ is the covariance of $\mathbf{n}_{\mathcal{S}}$.

If $\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r}$ is invertible, we can recover the original graph signal up to an error term as stated in the following proposition.

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

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

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

Proof. See Appendix B. □

Compared to the noiseless scenario where the main challenge is to ensure that $\mathbf{CU}$ is invertible, in the presence of noise we are interested in finding a sampling scheme with the lowest reconstruction error. Although Proposition 1 provides a performance bound for any sampling matrix $\mathbf{C}$ constructed by Algorithm 1, our specific choice of the residual node, $\ell = \arg\min_{j\in[N]} \mathbf{u}_j$, is not exploited in the proof of Proposition 1 and further analysis along those lines is left as part of the future work. We empirically observed that with

the proposed choice of the residual, the matrix product on the right-hand side of (5) has smaller maximum singular value than if the residual node is selected uniformly at random. We also note that the statistics of noise is not exploited when constructing $\mathbf{C}$. This is similar to state-of-the-art random selection sampling and aggregation sampling schemes [10,12,17] where one needs to rely on exhaustive search over the space of all sampling matrices to find the one that results in the lowest MSE. In the Bayesian setting studied in Section 4 where one assumes a prior distribution on $\mathbf{x}$, the original signal can be reconstructed up to an error term for any $\mathbf{C} \in \mathbb{R}^{m\times N}$ with $m \geq k$. Therefore, invertibility of $\mathbf{CU}$ is not a concern in the Bayesian case where we focus on the construction of a sampling set $\mathcal{S}$ with the lowest reconstruction error.

Note that the Gaussian elimination scheme also finds a full-rank submatrix $\mathbf{U}_{\mathcal{S}}$. According to Anis et al. [9], the sampling set found by Gaussian elimination with partial row pivoting corresponds to indices of the pivot rows. Therefore, in contrast to Algorithm 1 that takes into accounts representative power of each node in all frequency components (by considering the ℓ_2 norm of $\mathbf{u}_j$'s and their correlation with the residual), Gaussian elimination with partial row pivoting only considers individual frequency components when forming the sampling set. Hence, the signal reconstructed by such a scheme may not be robust to noise statistics. On the other hand, by choosing the residual node according to $\ell = \arg\min_{j\in[N]} \mathbf{u}_j$, Algorithm 1 finds an invertible submatrix and further finds a subset of rows of $\mathbf{U}$ with strong representation capability.

4. Bayesian sampling of graph signals

So far we have considered the problem of sampling in scenarios where the graph signal is not stochastic. In this section, we consider the problem of sampling and interpolation in a Bayesian setting where the graph signal is a non-stationary network process. To this end, we adopt the following definition of stationarity, recently proposed in [22].

Definition 3. A stochastic graph signal $\mathbf{x}$ is graph wide-sense stationary (GWSS) if and only if the matrix

$$\mathbb{E}[\bar{\mathbf{x}}\bar{\mathbf{x}}^\top] = \mathbf{V}^\top \mathbb{E}[\mathbf{x}\mathbf{x}^\top]\mathbf{V} \tag{6}$$

is diagonal.

In addition to our novel algorithmic contributions, the setting we consider in this section is more general than those considered in [15,38–40]. Specifically, unlike the prior work [15], we assume that the signal in not necessarily stationary with respect to $\mathcal{G}$ and that $\bar{\mathbf{x}}$ is a zero-mean random vector with generally non-diagonal covariance matrix $\mathbb{E}[\bar{\mathbf{x}}\bar{\mathbf{x}}^\top] = \mathbf{W}$. Furthermore, we do not restrict our study to the case of additive white noise. Rather, we consider a more practical setting where the noise terms are independent but the noise power varies across individual nodes of the graph. That is, if $\mathbf{y} = \mathbf{x} + \mathbf{n}$ denotes the noise-corrupted signal, $\mathbf{n} \in \mathbb{R}^N$ is a zero-mean noise vector with covariance matrix $\mathbb{E}[\mathbf{n}\mathbf{n}^\top] = \mathbf{Q} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)$. Note that this particular scenario is not explored in [15] or the related sensor selection and experimental design schemes [38–40].

Let $\mathcal{S}$ denote a sampling set of $m \geq k$ graph nodes. Since $\mathbf{x} = \mathbf{U}\bar{\mathbf{x}}_{\mathcal{K}}$, the samples $\mathbf{y}_{\mathcal{S}}$ and the non-zero frequency components of $\mathbf{x}$ are related via the Bayesian linear model

$$\mathbf{y}_{\mathcal{S}} = \mathbf{U}_{\mathcal{S},r}\bar{\mathbf{x}}_{\mathcal{K}} + \mathbf{n}_{\mathcal{S}}. \tag{7}$$

As before, in order to find $\hat{\mathbf{x}}$ it suffices to estimate $\bar{\mathbf{x}}_{\mathcal{K}}$ based on $\mathbf{y}_{\mathcal{S}}$. The least mean-square estimator of $\bar{\mathbf{x}}_{\mathcal{K}}$, denoted by $\hat{\bar{\mathbf{x}}}_{\mathcal{K}}$, is the Bayesian counterparts of the normal equations in the Gauss-Markov theorem (see, e.g. [37, Ch. 10]). In other words, it is given by

$$\hat{\bar{\mathbf{x}}}_{\mathcal{K}} = \bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1}\mathbf{y}_{\mathcal{S}}, \tag{8}$$

where

$$\begin{aligned}\bar{\mathbf{\Sigma}}_{\mathcal{S}} &= \left(\mathbf{W}^{-1} + \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1}\mathbf{U}_{\mathcal{S},r}\right)^{-1} \\ &= \left(\mathbf{W}^{-1} + \sum_{j\in\mathcal{S}} \frac{1}{\sigma_j^2}\mathbf{u}_j\mathbf{u}_j^\top\right)^{-1}\end{aligned} \tag{9}$$

is the error covariance matrix of $\hat{\bar{\mathbf{x}}}_{\mathcal{K}}$. Therefore, $\hat{\mathbf{x}} = \mathbf{U}\hat{\bar{\mathbf{x}}}_{\mathcal{K}}$ and its error covariance matrix is $\mathbf{\Sigma}_{\mathcal{S}} = \mathbf{U}\bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{U}^\top$.

The problem of sampling for near-optimal reconstruction can now be formulated as the task of choosing $\mathcal{S}$ so as to minimize the MSE of the estimator $\hat{\mathbf{x}}$. Since the MSE is defined as the trace of the error covariance matrix, we arrive at the following optimization problem,

$$\min_{\mathcal{S}} \quad \mathrm{Tr}(\mathbf{\Sigma}_{\mathcal{S}}) \text{ s.t.} \quad \mathcal{S} \subseteq \mathcal{N}, \quad |\mathcal{S}| \leq m. \tag{10}$$

Using trace properties and the fact that $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_m$, (10) simplifies to

$$\min_{\mathcal{S}} \quad \mathrm{Tr}\left(\bar{\mathbf{\Sigma}}_{\mathcal{S}}\right) \text{ s.t.} \quad \mathcal{S} \subseteq \mathcal{N}, \quad |\mathcal{S}| \leq m. \tag{11}$$

The optimization problem (11) is NP-hard and evaluating all $\binom{N}{m}$ possibilities to find the exact solution is intractable even for relatively small graphs. To this end, we propose an alternative to find a near-optimal solution in polynomial time. In [15], similar to the greedy sensor selection approach of [19,20], a greedy algorithm is proposed for the described Bayesian setting and its performance is analyzed under the assumption that the graph signal is stationary and the noise is white. The greedy algorithm aims to form a sampling set $\mathcal{S}$ iteratively, by greedily choosing the nodes one at a time, according to a specific selection criterion. Motivated by greedy maximization of submodular approach, the selection criterion is the marginal of gain of adding a new node when the objective function of (11) is treated as a submodular function of $\mathcal{S}$.

In applications dealing with extremely large graphs, the greedy algorithm in [15] might be computationally infeasible. Moreover, the graph signal is not necessary stationary and, perhaps more importantly, different nodes of a graph may experience different levels of noise. To address these challenges, motivated by the algorithm recently developed in [41] for maximization of *strictly* submodular functions, we develop a randomized-greedy algorithm for Bayesian sampling of graph signals that is significantly faster than the greedy algorithm. In addition, by leveraging the notion of weak submodularity, we establish performance bounds for the general setting of non-stationary graph signals.

4.1. Randomized-greedy selection sampling

Following [15,19,20], we start by formulating (11) as a set function maximization task. Let $f(\mathcal{S}) = \mathrm{Tr}(\mathbf{W} - \bar{\mathbf{\Sigma}}_{\mathcal{S}})$. Then (11) can equivalently be written as

$$\max_{\mathcal{S}} \quad f(\mathcal{S}) \quad \text{s.t.} \quad \mathcal{S} \subseteq \mathcal{N}, \quad |\mathcal{S}| \leq m. \tag{12}$$

In Proposition 2 below, by applying the matrix inversion lemma [42] we establish that $f(\mathcal{S})$ is monotone and weakly submodular. Moreover, we derive an efficient recursion to find the marginal gain of adding a new node to the sampling set $\mathcal{S}$. Given that we use the marginal gain as the selection criterion, the following proposition will greatly reduce the computational cost of evaluating the selection criterion.

Proposition 2. *$f(\mathcal{S}) = \mathrm{Tr}(\mathbf{W} - \bar{\mathbf{\Sigma}}_{\mathcal{S}})$ is a weak submodular, monotonically increasing set function, $f(\emptyset) = 0$, and for all $j \in \mathcal{N}\backslash\mathcal{S}$*

$$f(\mathcal{S}\cup\{j\}) - f(\mathcal{S}) = \frac{\mathbf{u}_j^\top \bar{\mathbf{\Sigma}}_{\mathcal{S}}^2 \mathbf{u}_j}{\sigma_j^2 + \mathbf{u}_j^\top \bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{u}_j}, \quad \text{and} \tag{13}$$

$$\bar{\mathbf{\Sigma}}_{\mathcal{S}\cup\{j\}} = \bar{\mathbf{\Sigma}}_{\mathcal{S}} - \frac{\bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{u}_j\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}}{\sigma_j^2+\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{u}_j}. \tag{14}$$

Proof. See Appendix C. □

Proposition 2 enables efficient construction of the sampling set in an iterative fashion. To further reduce the computational cost, we propose a randomized-greedy algorithm for selection sampling with minimal MSE that selects a sampling set iteratively.

Our aim will be to perform greedy selection in an iterative fashion by identifying nodes that maximize the selection criterion, i.e., the marginal gain given by (13). However, to reduce the cost of greedy search, we incorporate random sampling. In particular starting with $\mathcal{S}=\emptyset$, at iteration $(i+1)$ of the algorithm, a subset $\mathcal{R}$ of size s is sampled uniformly at random and without replacement from $\mathcal{N}\backslash\mathcal{S}$, the set of all un-sampled nodes. The marginal gain of each node in $\mathcal{R}$ is then found using (13), and the one corresponding to the highest marginal gain is added to $\mathcal{S}$. Then, the algorithm employs (14) to update $\bar{\mathbf{\Sigma}}_{\mathcal{S}}$ for the subsequent iteration (i.e., to be used in calculating the marginal gain/selection criterion in the next iteration). This procedure is repeated until some stopping criteria, e.g., a condition on the cardinality of $\mathcal{S}$ is met. Regarding s, the size of the randomly sampled subset $\mathcal{R}$, we follow the suggestion in [41] and set $s=\frac{N}{m}\log\frac{1}{\epsilon}$ where $e^{-m}\le\epsilon<1$ is a predetermined parameter that controls the trade-off between the computational cost and MSE of the reconstructed signal; randomized-greedy algorithm with smaller ϵ produces sampling solutions with lower MSE while the one with larger ϵ requires lower computational cost. Note that if $\epsilon=e^{-m}$, the randomized-greedy algorithm in each iteration considers all the available nodes and hence matches the greedy scheme in [15]. However, as we illustrate in our simulation studies, the proposed randomized-greedy algorithm is significantly faster than the greedy method in [15] for large ϵ while returning essentially the same sampling solution. The randomized-greedy algorithm is formalized as Algorithm 2.

Algorithm 2 Randomized-greedy Graph Sampling.

1: **Input:** ¶, **U**, m, ϵ.
2: **Output:** Subset $S\subseteq\mathcal{N}$ with $|S|=m\ge k$.
3: Initialize $\mathcal{S}=\emptyset$, $\bar{\mathbf{\Sigma}}_{\mathcal{S}}=$ ¶.
4: **while** $|S|<m$ **do**
5: Choose $\mathcal{R}$ by sampling $s=\frac{N}{m}\log(1/\epsilon)$ indices uniformly at random from $\mathcal{N}\backslash\mathcal{S}$
6: $j_s=\arg\max_{j\in\mathcal{R}}\frac{\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}^2\mathbf{u}_j}{\sigma_j^2+\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{u}_j}$
7: $\bar{\mathbf{\Sigma}}_{\mathcal{S}\cup\{j_s\}}=\bar{\mathbf{\Sigma}}_{\mathcal{S}}-\frac{\bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{u}_j\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}}{\sigma_j^2+\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{u}_j}$
8: Set $\mathcal{S}\leftarrow\mathcal{S}\cup\{j_s\}$
9: **end while**
10: **return** $\mathcal{S}$.

4.2. Complexity analysis

To take a closer look at computational complexity of Algorithm 2, note that step 6 costs $\mathcal{O}(\frac{N}{m}k^2\log(\frac{1}{\epsilon}))$ since one needs to compute $\frac{N}{m}\log(\frac{1}{\epsilon})$ marginal gains, each requiring $\mathcal{O}(k^2)$ operations. Furthermore, step 7 requires $\mathcal{O}(k^2)$ arithmetic operations. Since there are m such iterations, running time of Algorithm 2 is $\mathcal{O}(Nk^2\log(\frac{1}{\epsilon}))$. Please refer to Table 2 for a comparison between computational costs of proposed schemes in this paper to the existing methods.

4.3. Theoretical analysis

In this section, we analyze performance of the proposed randomized-greedy algorithm in a range of scenarios.

Theorem 2 below states that if $f(\mathcal{S})$ is characterized by a bounded maximum element-wise curvature, Algorithm 2 returns a sampling subset yielding an MSE that is on average within a multiplicative factor of the MSE associated with the optimal sampling set.

Theorem 2. *Let $\mathcal{C}_f$ denote the maximum element-wise curvature of $f(\mathcal{S})=\mathrm{Tr}(\mathbf{W}-\bar{\mathbf{\Sigma}}_{\mathcal{S}})$, the objective function in (12). Let $\alpha=(1-e^{-\frac{1}{c}}-\frac{\epsilon^\beta}{c})$, where $c=\max\{1,\mathcal{C}_f\}$, $e^{-m}\le\epsilon<1$, and $\beta=1+\max\{0,\frac{s}{2N}-\frac{1}{2(N-s)}\}$. Let $\mathcal{S}_{rg}$ be the sampling set returned by the randomized greedy algorithm and let $\mathcal{O}$ denote the optimal solution of (11). Then*

$$\mathbb{E}\Big[\mathrm{Tr}(\bar{\mathbf{\Sigma}}_{\mathcal{S}_{rg}})\Big]\le\alpha\mathrm{Tr}(\bar{\mathbf{\Sigma}}_{\mathcal{O}})+(1-\alpha)\mathrm{Tr}(\mathbf{W}). \tag{15}$$

Proof. The proof of Theorem 2 relies on the argument that if $s=\frac{N}{m}\log\frac{1}{\epsilon}$, then with high probability the random set $\mathcal{R}$ in each iteration of Algorithm 2 contains at least one node from $\mathcal{O}$. See Appendix D for the complete proof. □

Compared to the results of [41] where the maximization of *strictly* submodular and monotone functions is considered, Theorem 2 relaxes this assumption and states that submodularity is not required for near-optimal performance of the randomized greedy algorithm. In particular, if the set function is *weak submodular*, Algorithm 2 still selects a sampling set with an MSE near that achieved by the optimal sampling set. In addition, even if the function is submodular (e.g., when the objective is $\log\det(.)$ function instead of the MSE), the approximation factor in Theorem 2 is tighter than that of [41] as the result of the analysis presented in the proof of Theorem 2. Moreover, a major assumption in [41] is that $\mathcal{R}$ is constructed by sampling with replacement. In contrast, we assume $\mathcal{R}$ is constructed by sampling without replacement and carry out the analysis in this setting.

Next, we study the performance of the randomized greedy algorithm using the tools of probably approximately correct (PAC) learning theory [43,44]. That is, not only the sampling set selected by Algorithm 2 is on expectation near optimal, but the MSE associated with the selected sampling set is with high probability close to the smallest achievable MSE. The randomization of Algorithm 2 can be interpreted as an approximation of the marginal gains of the nodes selected by the greedy scheme [15,19,20]. More specifically, following this interpretation for the ith iteration we have $f_{j_{rg}}(\mathcal{S}_{rg}):=\eta_i f_{j_g}(\mathcal{S}_g)$, where subscripts rg and g indicate the sampling sets and nodes selected by the randomized greedy (Algorithm 2) and the greedy algorithm in [15], respectively, and $0<\eta_i\le1$ for all $i\in[m]$ are random variables. Following this argument and by employing the Bernstein inequality [45], we arrive Theorem 3 which states that the randomized greedy algorithm selects a near-optimal sampling set with high probability.

Theorem 3. *Instate the notation and hypotheses of Theorem 2. Assume $\{\eta_i\}_{i=1}^m$ is a collection of random variables such that $\mathbb{E}[\eta_i]\ge\mu_\epsilon$, for all $i\in[m]$. Then, it holds that*

$$\mathrm{Tr}(\bar{\mathbf{\Sigma}}_{\mathcal{S}_{rg}})\le\Big(1-e^{-\sum_{i=1}^m\frac{\eta_i}{mc}}\Big)\mathrm{Tr}(\bar{\mathbf{\Sigma}}_{\mathcal{O}})+e^{-\sum_{i=1}^m\frac{\eta_i}{mc}}\mathrm{Tr}(\mathbf{W}). \tag{16}$$

Moreover, if $\{\eta_i\}_{i=1}^m$ are independent, for all $0<q<1$ with probability at least $1-e^{-Cm}$ it holds that

$$\mathrm{Tr}(\bar{\mathbf{\Sigma}}_{\mathcal{S}_{rg}})\le\Big(1-e^{-\frac{(1-q)\mu_\epsilon}{c}}\Big)\mathrm{Tr}(\bar{\mathbf{\Sigma}}_{\mathcal{O}})+e^{-\frac{(1-q)\mu_\epsilon}{c}}\mathrm{Tr}(\mathbf{W}) \tag{17}$$

for some $C>0$.

Proof. See Appendix E. □

In our simulation studies (see Section 5), we empirically verify the results of Theorems 2 and 3 and illustrate that Algorithm 2 performs favorably compared to the competing greedy scheme both on average and for each individual sampling task.

Finally, in Theorem 4 we extend the results of [15] derived for stationary graph signals and show that the maximum element-wise curvature of $f(\mathcal{S}) = \mathrm{Tr}(\mathbf{W} - \tilde{\mathbf{\Sigma}}_{\mathcal{S}})$ is bounded even for non-stationary graph signals and in the scenario where the statistics of the noise varies across nodes of the graph.

Theorem 4. *Let $\mathcal{C}_f$ be the maximum element-wise curvature of $f(\mathcal{S}) = \mathrm{Tr}(\mathbf{W} - \tilde{\mathbf{\Sigma}}_{\mathcal{S}})$. Then it holds that*

$$\mathcal{C}_{\max} \leq \max_{j \in \mathcal{N}} \frac{\lambda^2_{\max}(\mathbf{W})}{\lambda^2_{\min}(\mathbf{W})} \left(1 + \frac{\lambda_{\max}(\mathbf{W})}{\sigma_j^2}\right)^3. \tag{18}$$

Proof. See Appendix F. □

It was shown in [15] that if $\mathbf{x}$ is stationary and $\mathbf{W} = \sigma^2_{\mathbf{x}} \mathbf{I}_N$ for some $\sigma^2_{\mathbf{x}} > 0$ and $\sigma_j^2 = \sigma^2$ for all $j \in \mathcal{N}$, then the curvature of the MSE objective is bounded. However, Theorem 4 holds even in the scenarios where the signal is non-stationary and the noise is not white.

5. Numerical examples

To assess the proposed support recovery and sampling algorithms, we study their performance in recovery of signals supported on synthetic and real-world graphs. In the first two subsections, we benchmark the performance of Algorithm 1, while in the rest of the subsections, we focus on evaluating the efficacy of the proposed randomized greedy algorithm.

5.1. Synthetic Erdős-Rényi random graphs I

We first 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 forming $\mathbf{U}$ using the first k eigenvectors of the graph adjacency matrix, where k is varied 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. The signal is corrupted by 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 state-of-the-art uniform, leverage score, and optimal random sampling schemes [10,16,17]. We define the recovery error as the ratio of the error energy to the true signal's energy. Furthermore, the *success rate* [10] is defined as the fraction of instances where $\mathbf{CU}$ is invertible [cf. (1)]. The results, averaged over 100 independent instances, are shown in Fig 1(a). As we can see from Fig 1(a) (top), the proposed scheme consistently achieves lower recovery error than competing schemes. Moreover, as shown in Fig 1(a) (bottom), when the bandwidth increases the success rate of random sampling schemes decreases while the success rate of the proposed scheme is always one, as formally established in Theorem 1. Additionally, note that the reconstruction error of Algorithm 1 is given in Proposition 1. There, we establish an upper-bound on the Reconstruction error. Note that Proposition 1 establishes perfect success rate and bounds the reconstruction error in the noisy scenario; it does not necessary establish a monotonically decreasing error. Indeed, a signal has a higher bandwidth, then a larger sampling set is required to ensure perfect success rate. This discussion then justifies then non-monotonic recovery error of the proposed algorithm as well as the benchmarking schemes.

Next, we compare the proposed sampling algorithm with Algorithm 1 of [10] (see Fig 2) for undirected Erdős-Rényi random graphs where we consider smaller bandwidth here to accommodate the computational cost of Algorithm 1 of [10]. A that disadvantage of Algorithm 1 of [10] compared to our method is that the iterative method of [10] needs to perform singular value decomposition in each iteration to identify the sampling operator (see step 2 of Algorithm 1 in [10]). Additionally, similar to our scheme which requires a residual node for initialization, [10] also needs an initial node. However, the selection of such an initial node is unclear in Algorithm 1 of [10]. One major benefit of our method is that, as we show in Theorem 1, the proposed scheme achieves perfect recovery while Algorithm 1 of [10] does not have this important property. In terms of the empirical comparison, as Fig 2 shows, the proposed iterative algorithm achieves a lower reconstruction error while consistently achieving success rate of one.

5.2. Real graph: interpolation of industrial sectors' production

Next, we analyze data from the Bureau of Economic Analysis of the U.S. Department of Commerce which publicizes an annual table of input and outputs organized by economic sectors.[3] Specifically, we represent by nodes 62 industrial sectors as defined by the North American Industry Classification System, and construct weighted edges and the graph signal similar to Marques et al. [12]. The (undirected) edge weight between two nodes represents the average total production of the sectors, the first sector being used as the input to 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 thresholded to zero and the eigenvalue decomposition of the corresponding adjacency matrix $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$ is performed. 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 [12, Fig. 4(a)(top)] with a bandwidth of 4.

We interpolate sectors' production by observing a few nodes using Algorithm 1 and 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 state-of-the-art uniform, leverage score, and optimal random sampling schemes [10,16,17] averaged over 1000 Monte-Carlo simulations as shown in Fig. 1(b) (top). As the figure indicates, the proposed algorithm outperforms uniform, leverage score, and optimal random sampling schemes [10,16,17]. However, Algorithm 1 does not achieve perfect recovery in this noiseless scenario because the signal is not truly bandlimited. Moreover, Fig. 1(b) (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%; as Fig. 1(b) (bottom) illustrates, $\hat{\mathbf{x}}$ closely approximates $\mathbf{x}$.

5.3. Synthetic graph: localization of UAVs

We now tackle a UAV localization problem in which the goal is to estimate absolute positions of robots from on-board sensor measurements. Specifically, consider a network of N UAVs moving in a 2D plane and assume that each UAV is equipped with two systems: a laser scanner that measures the relative position of other UAVs within a sensing radius, and a GPS system that finds the absolute position of the UAV.[4] While the laser system can find relative po-

[3] Dataset from https://www.bea.gov.

[4] Notice that the graph structure in this application is essentially the time-varying communication network between the UAVs. In our simulation studies, we

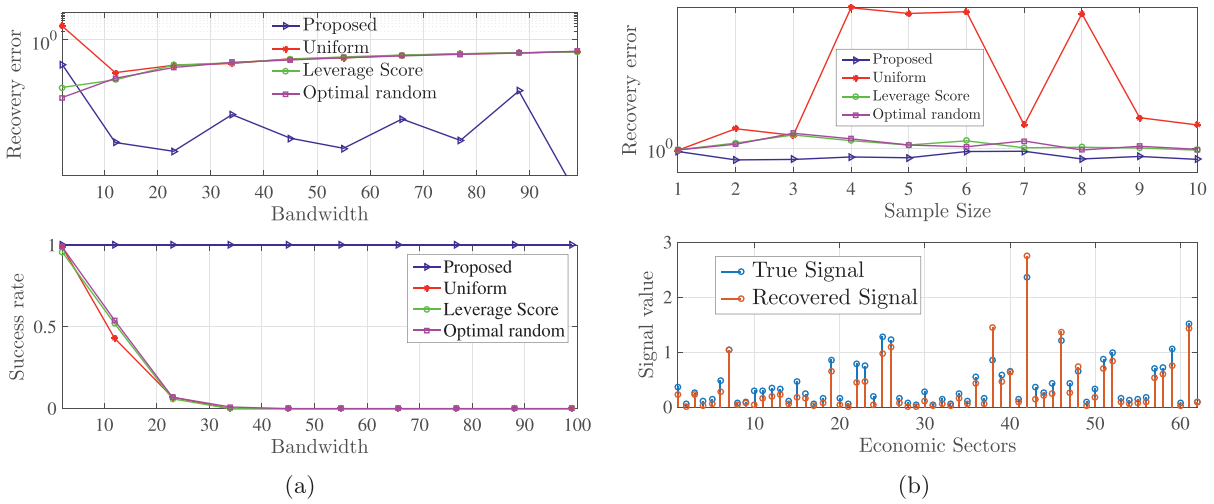

Fig. 1. (a) Recovery error (top) and success rate (bottom) of Algorithm 1 and various random selection sampling schemes versus bandwidth (k) for undirected Erdős-Rényi random graphs. (b) Top: Recovery error comparison of different selection sampling schemes as a function of the sample size for the economy network. Bottom: Recovered and true graph signals for various economic sectors using Algorithm 1.

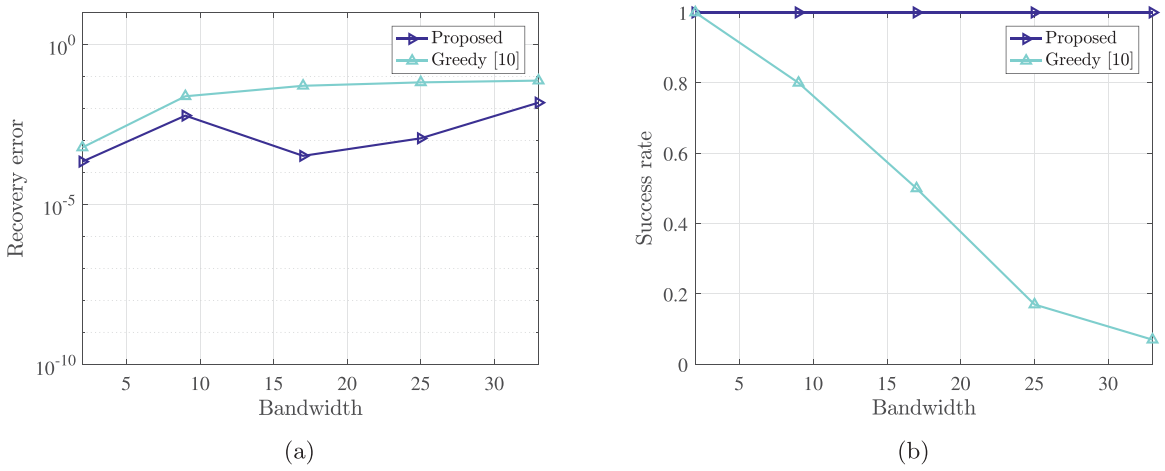

Fig. 2. (a) Recovery error and (b) success rate (bottom) of Algorithm 1 and the Greedy method of [10] versus bandwidth (k) for undirected Erdős-Rényi random graphs.

sitions of the nearby UAVs with minimal power consumption, the GPS system requires intensive power to receive the location of the UAV from the control unit located potentially far from the network of UAVs. We consider the scenario where such inherent energy constraints prevent some UAVs to collect GPS data, i.e., only a subset of the UAVs can use the GPS. The objective is to compute the most representative subset of the UAVs so to minimize the MSE of the estimated global positions of all UAVs. To this end, we employ the proposed randomized-greedy scheme in Algorithm 2 with various values of ϵ to find a sampling set (a subset of UAVs) and compare its recovery error to that of the greedy sampling scheme [15]. Note that two graph signals, namely the x and y coordinates of UAVs, are supported on the network. Further, since UAVs that are close to each other have similar locations, both of these graph signals are smooth and hence bandlimited. It is worth to note that the collection of UAVs, typically referred to as UAV swarm, has a *swarm leader* that is task with handling costlier computations and is capable of communicating with the control unit that guides the swarm in moving in the environment.

We run Monte Carlo simulations with 1000 instances where we consider 1000 UAVs distributed uniformly on a 10×10 grid; the range of the laser system is set to 0.3 and the power of noise affecting laser measurements is set to 10^{-2}. The recovery error and running time results as a function of signals' bandwidth – which is also the size of the sampling set – are shown in Fig. 4(a) and (b), respectively. As we see in Fig. 4(a), performance of the proposed scheme and the greedy algorithm are fairly similar; as bandwidth increases, the recovery error decreases. Furthermore, as ϵ gets smaller, the gap between the performance of the proposed scheme and the greedy algorithm reduces until becoming negligible. The running time comparison illustrated in Fig. 4(b) reveals that for the largest sampling set considered (i.e. $k = 50$), the proposed scheme is more than 2x faster than the greedy method. Additionally, the complexity of the proposed scheme is linear in k, while that of the greedy method is quadratic, as predicted by our theoretical results; see also [1] for additional MSE performance and runtime comparisons with the greedy sampling algorithm in [15].

5.4. Real graph: semi-supervised face clustering

Clustering faces of individuals is an important task in computer vision [46–49]. In real-world settings, labeling all images is practically infeasible. However, acquiring labels even for a small sub-

consider the localization task for only a single time-step. Nonetheless, the proposed sampling scheme can be employed in every time step where identification UAVs with GPS turned-on is required.

Fig. 3. Face clustering: given images of multiple subjects, the goal is to find images that belong to the same subject (examples from the EYaleB dataset [46]).

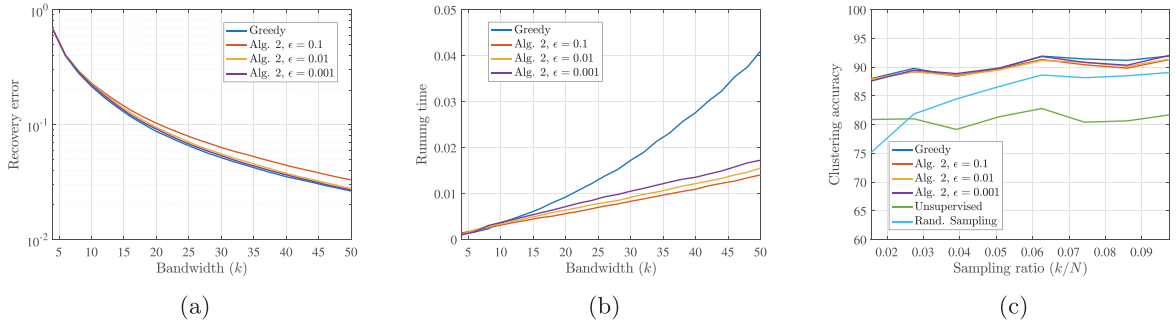

Fig. 4. (a) Recovery error comparison of the greedy scheme [15] and Algorithm 2 as a function of bandwidth for the UAV localization problem. (b) Running time comparison of the greedy scheme [15] and Algorithm 2 as a function of bandwidth for the UAV localization problem. (c) Clustering accuracy of greedy [15], Algorithm 2, random sampling, and unsupervised methods as a function of the sampling ratio for the face clustering application.

set of data that can represent all images may drastically improve the clustering accuracy. The proposed randomized-greedy selection sampling framework can be employed in this setting to acquire labels for a small number of images to achieve improved clustering accuracy. To this end, we test the randomized-greedy algorithm on EYaleB dataset [46] (see Fig. 3) which contains frontal face images of 38 individuals under 64 different illumination conditions. Similar to the prior works (see, e.g., [47–49]), in our studies the images are down-sampled to 48×42 from the original size of 192×168. In each of 100 independent instances of the Monte Carlo simulation we randomly pick 8 subjects and all of their images as the data points to be clustered; this results in a clustering problem with $N = 512$ data points. To construct the underlying graph signal and capture similarity of the data points, we employ the sparse subspace clustering (SSC) scheme recently proposed in [47] to find the adjacency matrix $\mathbf{A}$ and the Laplacian matrix $\mathbf{L}$. The graph signal support on the constructed similarity graph is discrete valued, i.e., the value of each node is an integer in $\{1, \ldots, 8\}$. Note that the graph signal supported on the constructed similarity graph is smooth and bandlimited as similar images are unlikely to correspond to different individuals. The performance comparison of Algorithm 2 with various values for ϵ, greedy sampling method, random sampling schemes, and the unsupervised clustering method are illustrated in Fig. 4(c) as a function of the sampling ratio (k/N). For the sake of clarity of presentation, we only show the result of the best method among uniform, leverage score, and optimal random sampling approaches [10,17]. As we see in Fig. 4(c), the greedy and randomized-greedy schemes deliver the best clustering performance; as we increase size of the sampling set, the accuracy of semi-supervised schemes improves and the gap between the performance of random sampling methods and the proposed scheme decreases. Furthermore, our simulation studies reveal that acquiring labels of only 8 data points using the proposed scheme results in more than 12% improvements in clustering accuracy as compared to the unsupervised method.

5.5. Synthetic Erdős-Rényi random graphs II

Since Algorithm 2 is a randomized scheme, in this section we study the performance of Algorithm 2 for each individual sampling tasks (i.e. each Monte-Carlo realizations). To this end, we again consider the Erdős-Rényi random graphs, similar to those in Section 5.1. Here, we study the setting where $N = 10$ and $k = 4$. Bandlimited graph signals are generated as before except that this time we take $\mathbf{U}$ as the first 4 eigenvectors of the adjacency matrix. Fig. 5(a) depicts superimposed MSE histograms of Algorithm 2 and the greedy sampling scheme [15] for 100 realizations per method and fixed $|S| = 4$. As the figure illustrates, the proposed randomized greedy schemes performs well and is comparable with the greedy approach, not just on average but also for majority of individual sampling tasks.

5.6. Real graph: minnesota road network

Next, we consider the Minnesota road network[5] with $N = 2642$ nodes in order to showcase scalability of the proposed graph sampling method. To that end, Bandlimited graph signals are generated by taking the first $k = 600$ eigenvectors of the graph Laplacian matrix, where the non-zero frequency components are drawn from a zero-mean, multivariate Gaussian distribution with randomly chosen PSD covariance matrix $\mathbf{W}$. The signals are corrupted with additive white Gaussian noise with $\sigma^2 = 10^{-2}\mathbf{I}_N$. As expected, Fig. 5(b) and (c) depict trends of decreasing MSE and increasing running time versus $|S|$, respectively. The results are averaged over 1000 Monte-Carlo simulations run. Remarkably, the proposed randomized greedy procedure achieves an order-of-magnitude speedup over the state-of-the-art algorithm in [15] while showing only a marginal degradation in the MSE performance. Note that the time of performing eigenvalue decomposition to find the graph shift operator $\mathbf{U}$ in MATLAB was less than 2 s on a typical laptop. Fig. 5(d) depicts the runtime comparison of the proposed scheme versus the benchmark by accounting for the time of computing the eigenvalue decomposition.

5.7. Synthetic graph: large-scale preferential attachment random graph

Finally, we consider a large-scale preferential attachment random graph [50] with $N = 10,000$ nodes to show the superiority of the proposed Algorithm 2 over existing methods. In particular, similar to the previous random graph simulations, we generate random band-limited Gaussian graph signals using the first 500 eigenvectors of the preferential attachment graph adjacency matrix (see Fig. 6(c) for the structure of the sparse adjacency matrix). The results are illustrated in Fig. 6 where as we see, Algorithm 2 achieves

[5] https://sparse.tamu.edu/Gleich/minnesota

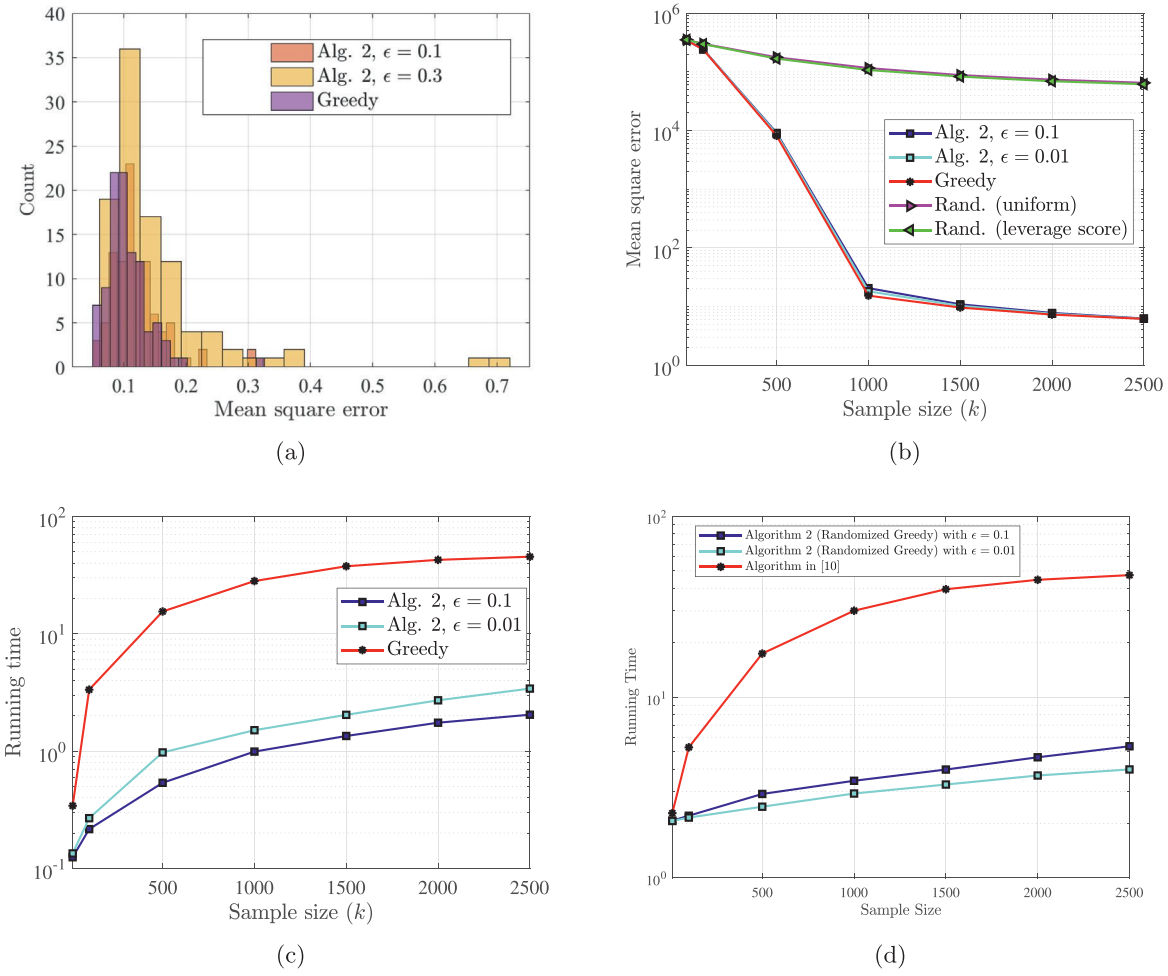

Fig. 5. (a) Histogram of MSE values for 100 realizations and fixed sampling set size in simulated Erdős-Rényi graphs. (b) MSE comparison of greedy [15], Algorithm 2, and random sampling schemes on Minnesota road network. (c) Running time comparison of the greedy scheme [15] and Algorithm 2 on Minnesota road network, excluding the time of eigenvalue decomposition. (d) Running time comparison of the greedy scheme [15] and Algorithm 2 on Minnesota road network, including the time of eigenvalue decomposition.

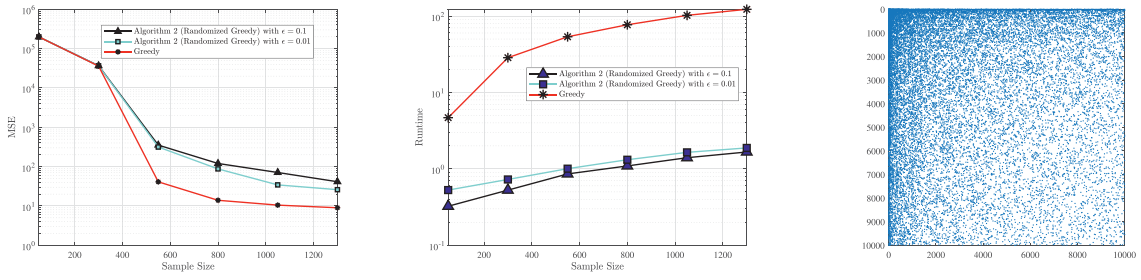

Fig. 6. Performance comparison of greedy scheme [15] and Algorithm 2 on a large-scale preferential attachment random graph with $N = 10,000$ nodes.

the same performance as that of the greedy scheme [15] while incurring orders of magnitude lower running time.

6. Conclusion

We considered the task of sampling and reconstruction of spectrally sparse graph signals. where the goal is to interpolate a (non-stationary) graph signal from a small subset of the nodes with the lowest reconstruction error. First, we studied the non-Bayesian scenario and 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 then theoretically showed that in the noiseless case the original k-spectrally sparse signal is perfectly 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. In the Bayesian scenario where the graph signal is a non-stationary random process, we formulated the sampling task as the problem of maximizing a monotone weak submodular function that is directly related to the mean square error (MSE) of the linear estimator of

the original signal. We proposed a randomized-greedy algorithm to find a sub-optimal subset of sampling nodes. By analyzing the performance of the randomized-greedy algorithm, we showed that the resulting MSE is a constant factor away from the MSE of the optimal sampling set. Unlike prior work, our guarantees do not require stationarity of the graph signal and the study is not restricted to the case of additive white noise. Instead, the noise coefficients are assumed to be independent but the power of noise varies across individual nodes of the graph. Extensive simulations on synthetic and real-world graphs with applications in economics, localization, and clustering showed that the proposed iterative and randomized-greedy selection sampling algorithms outperform the competing alternatives in terms of accuracy and runtime.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Abolfazl Hashemi: Conceptualization, Methodology, Formal analysis, Software, Writing – original draft. **Rasoul Shafipour:** Conceptualization, Methodology, Formal analysis, Software, Writing – original draft. **Haris Vikalo:** Writing – review & editing, Funding acquisition. **Gonzalo Mateos:** Writing – review & editing, Funding acquisition.

Acknowledgment

We would like to thank the authors of [12] for providing the data used for the economy network analysis. Work in this paper was supported in part by the NSF award CCF-1750428 and ECCS-1809327.

Appendix A. Proof of Theorem 1

To prove the theorem, it suffices to show that Algorithm 1 selects a subset of rows of $\mathbf{U}$ which are linearly independent. Consider the ith iteration where $\mathbf{u}_{s_i}$ is identified and assume that until this iteration $\mathcal{S}$ contains indices of a collection of linearly independent vectors $\{\mathbf{u}_{s_1},\ldots,\mathbf{u}_{s_{(i-1)}}\}$. If $|\mathbf{r}_{i-1}^\top \mathbf{u}_{s_i}| \neq 0$, since $\mathbf{r}_{i-1}$ is orthogonal to the span of $\{\mathbf{u}_{s_1},\ldots,\mathbf{u}_{s_{(i-1)}}\}$, $\mathbf{u}_{s_i}$ is not in the span of these vectors. Hence, $\{\mathbf{u}_{s_1},\ldots,\mathbf{u}_{s_i}\}$ is also a collection of linearly independent vectors and by an inductive argument we conclude that rows of $\mathbf{U}_{\mathcal{S},r}$ are linearly independent. Now assume $|\mathbf{r}_{i-1}^\top \mathbf{u}_{s_i}| = 0$ for some $i \leq k$. Since $\mathbf{U}$ does not have all-zero rows, this condition implies $\mathbf{r}_{i-1} = \mathbf{0}$.[6] Therefore, all the remaining rows of $\mathbf{U}$ which are not selected lie in the span of $\{\mathbf{u}_{s_1},\ldots,\mathbf{u}_{s_{(i-1)}}\}$. Since by assumption $i \leq k$, this condition implies that the rank of $\mathbf{U}$ is at most $k-1$ which contradicts the fact that $\mathbf{V}$ is a basis and $\mathbf{U}$ has full column-rank. Therefore, $\mathbf{r}_{i-1} = \mathbf{0}$ holds only for $i > k$ and thus rows of $\mathbf{U}_{\mathcal{S},r}$ are linearly independent. This completes the proof.

Appendix B. Proof of Proposition 1

According to Theorem 1, if $m = k$, $\mathbf{U}_{\mathcal{S},r} = \mathbf{C}\mathbf{U}$ is invertible. Therefore, since $\mathbf{Q}_{\mathcal{S}}$ is positive definite and invertible it is easy to see that $\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r}$ is also invertible. Now consider the case $m \geq k$ where $\mathbf{U}_{\mathcal{S},r} \in \mathbb{R}^{m\times k}$ is a tall full rank matrix. Let $\mathbf{Q}_{\mathcal{S}}^{-1} = \mathbf{L}\mathbf{L}^\top$ be the Cholesky decomposition of $\mathbf{Q}_{\mathcal{S}}^{-1}$. Since $\mathbf{Q}_{\mathcal{S}}^{-1}$ is a positive definite matrix, $\mathbf{L} \in \mathbb{R}^{m\times m}$ is full rank and invertible. Therefore, $\mathbf{L}_u = \mathbf{L}^\top \mathbf{U}_{\mathcal{S},r}$ is also a full rank matrix. Thus, $\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r} = \mathbf{L}_u^\top \mathbf{L}_u \in \mathbb{R}^{k\times k}$ is full rank and invertible. Hence, for any $m \geq k$ given a $\mathbf{C}$ constructed by Algorithm 1, (4) simplifies to

$$\hat{\tilde{\mathbf{x}}} = (\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \tilde{\mathbf{x}}. \tag{19}$$

Since $\mathbf{x} = \mathbf{U}\bar{\mathbf{x}}_{\mathcal{K}}$, the reconstructed signal $\hat{\mathbf{x}}$ can be obtained according to

$$\begin{aligned}
\hat{\mathbf{x}} &= \mathbf{U}(\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \tilde{\mathbf{x}} \\
&= \mathbf{U}(\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} (\mathbf{U}_{\mathcal{S},r}\bar{\mathbf{x}}_{\mathcal{K}} + \mathbf{n}_{\mathcal{S}}) \\
&= \mathbf{U}\bar{\mathbf{x}}_{\mathcal{K}} + \mathbf{U}(\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{C}\mathbf{n} \\
&= \mathbf{x} + \mathbf{U}(\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{n}_{\mathcal{S}}.
\end{aligned} \tag{20}$$

Therefore,

$$\begin{aligned}
\|\hat{\mathbf{x}} - \mathbf{x}\|_2 &\leq \|\mathbf{U}(\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{n}_{\mathcal{S}}\|_2 \\
&\overset{(a)}{\leq} \|\mathbf{U}(\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{n}\|_2 \\
&\overset{(b)}{\leq} \sigma_{\max}(\mathbf{U}(\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1}) \epsilon_{\mathbf{n}} \\
&\overset{(c)}{\leq} \sigma_{\max}(\mathbf{U}) \sigma_{\max}((\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1}) \epsilon_{\mathbf{n}} \\
&\overset{(d)}{\leq} \sigma_{\max}((\mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1} \mathbf{U}_{\mathcal{S},r})^{-1} \mathbf{U}_{\mathcal{S},r}^\top \mathbf{Q}_{\mathcal{S}}^{-1}) \epsilon_{\mathbf{n}}
\end{aligned} \tag{21}$$

where (a) and (b) follow by the assumption $\|\mathbf{n}_{\mathcal{S}}\|_2 \leq \|\mathbf{n}\|_2 \leq \epsilon_{\mathbf{n}}$, (c) stems from submultiplicative property of ℓ_2-norm, and (d) is by the fact that $\sigma_{\max}(\mathbf{U}) = 1$ as it is a submatrix of an orthogonal matrix.

Appendix C. Proof of Proposition 2

We first verify that

$$f(\emptyset) = \mathrm{Tr}\big(\mathbf{W} - \bar{\mathbf{\Sigma}}_{\emptyset}\big) = \mathrm{Tr}(\mathbf{W} - \mathbf{W}) = 0.$$

Next, to show monotonicity, we establish a recursive relation for the marginal gain of selecting a new node on graph. More specifically, for $j \in [n]\backslash\mathcal{S}$ it holds that

$$\begin{aligned}
f_j(\mathcal{S}) &= \mathrm{Tr}\big(\mathbf{W} - \bar{\mathbf{\Sigma}}_{\mathcal{S}\cup\{j\}}\big) - \mathrm{Tr}\big(\mathbf{W} - \bar{\mathbf{\Sigma}}_{\mathcal{S}}\big) \\
&= \mathrm{Tr}\big(\bar{\mathbf{\Sigma}}_{\mathcal{S}}\big) - \mathrm{Tr}\big(\bar{\mathbf{\Sigma}}_{\mathcal{S}\cup\{j\}}\big) \\
&= \mathrm{Tr}\big(\bar{\mathbf{\Sigma}}_{\mathcal{S}}\big) - \mathrm{Tr}\Big(\big(\bar{\mathbf{\Sigma}}_{\mathcal{S}}^{-1} + \sigma_j^{-2}\mathbf{u}_j\mathbf{u}_j^\top\big)^{-1}\Big) \\
&\overset{(a)}{=} \mathrm{Tr}\left(\frac{\bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{u}_j\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}}{\sigma_j^2 + \mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{u}_j}\right) \overset{(b)}{=} \frac{\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}^2\mathbf{u}_j}{\sigma_j^2 + \mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{u}_j}
\end{aligned} \tag{22}$$

where (a) easily follows by applying Sherman-Morrison formula [42] on matrix $(\bar{\mathbf{\Sigma}}_{\mathcal{S}}^{-1} + \sigma_j^{-2}\mathbf{u}_j\mathbf{u}_j^\top)^{-1}$, and (b) is due to properties of the trace of a matrix. Finally, since $\bar{\mathbf{\Sigma}}_{\mathcal{S}}$ is the error covariance matrix, it is symmetric and positive definite. Hence, $f_j(\mathcal{S}) > 0$, which in turn implies monotonicity.

Appendix D. Proof of Theorem 2

To prove the stated results, we first we state Lemma 1 [51] that upper-bounds the difference between the values of the objective corresponding to two sets having different cardinalities.

Lemma 1. *[51] Let f denote a monotone set function with the maximum element-wise curvatures $\mathcal{C}_{\max}$. Let $\mathcal{S}$ and $\mathcal{T}$ be any two sampling sets such that $\mathcal{S} \subset \mathcal{T} \subseteq \mathcal{N}$ with $|\mathcal{T}\backslash\mathcal{S}| = r$. Then, it holds that*

$$f(\mathcal{T}) - f(\mathcal{S}) \leq C(r) \sum_{j\in\mathcal{T}\backslash\mathcal{S}} f_j(\mathcal{S}), \tag{23}$$

[6] We note that if $|\mathbf{r}_{i-1}^\top \mathbf{u}_{s_i}| = 0$ for $i \leq k$, $\mathbf{r}_{i-1} \neq \mathbf{0}$, and $\mathbf{r}_{i-1}$ and $\mathbf{u}_{s_i}$ are orthogonal, then since s_i is the optimizer of the selection criterion in step 6, $\mathbf{r}_{i-1}$ is orthogonal to all $\mathbf{u}_j$ with $j \in \mathcal{N}\backslash\{\ell\}\backslash\mathcal{S}$. Now, since by definition $\mathbf{r}_{i-1}$ is orthogonal to the subspace spanned by nodes indexed by $\mathcal{S}$, we conclude that $\mathbf{r}_{i-1} \in \mathbb{R}^k$ is orthogonal to the subspace spanned by all $\mathbf{u}_j$, i.e. $\mathbb{R}^k$. However, this can only hold for $\mathbf{r}_{i-1} = \mathbf{0}$.

where $C(r) = \frac{1}{r}(1 + (r-1)\mathcal{C}_f)$. Moreover, $C(r)$ is decreasing (increasing) with respect to $\mathcal{R}$ if $\mathcal{C}_f < 1$ ($\mathcal{C}_f > 1$).

To prove the theorem, we first establish a bound on the expected value of the marginal gains of adding new nodes to the sampling set. Then, using the results of Lemma 1, we reduce the proof of approximation factor to that of the classical greedy algorithm introduced in [52]. More specifically, consider the ith iteration of Algorithm 2 and let $\mathcal{S}$ and $(i+1)_s$ denote the current sampling set and the index of node selected at the $(i+1)$st iteration of Algorithm 2. A necessary condition to achieve the optimal MSE is that set $\mathcal{R}$ at each iteration must contain at least one node from the optimal sampling set $\mathcal{O}$. Let $\Phi = \mathcal{R} \cap (\mathcal{O} \backslash \mathcal{S})$. Since $\mathcal{R}$ is generated via sampling without replacement, it holds that

$$\begin{aligned}\Pr\{\Phi = \emptyset\} &= \prod_{l=0}^{s-1}\left(1 - \frac{|\mathcal{O}\backslash\mathcal{S}|}{|\mathcal{N}\backslash\mathcal{S}| - l}\right)\\ &\overset{(a)}{\leq} \left(1 - \frac{|\mathcal{O}\backslash\mathcal{S}|}{s}\sum_{l=0}^{s-1}\frac{1}{N-l}\right)^s\\ &\overset{(b)}{\leq} (1 - \frac{|\mathcal{O}\backslash\mathcal{S}|}{s}(H_N - H_{N-s}))^s \end{aligned} \tag{24}$$

where (a) is by the inequality between arithmetic and geometric means and the fact that $|\mathcal{N}\backslash\mathcal{S}| \leq N$, and

$$H_p = \sum_{l=1}^{p}\frac{1}{p} = \log p + \gamma + \zeta_p \tag{25}$$

in (b) is the pth harmonic number. The object γ in (25) is the Euler-Mascheroni constant, and $\zeta_p = \frac{1}{2p} - \mathcal{O}(\frac{1}{p^4})$ is a monotonically decreasing sequence related to Hurwitz zeta function that satisfies $\zeta_p - \zeta_{p-q} = \frac{1}{2p} - \frac{1}{2(p-q)} + \mathcal{O}(\frac{1}{(p-q)^4})$ for all integers $p > q$ [53]. Therefore, using the identity (25) and the fact that $(1+x)^y \leq e^{xy}$ for any real number $y > 0$, we obtain

$$\Pr\{\Phi = \emptyset\} \leq \left((1 - \tfrac{s}{N})e^{\frac{s}{2N(N-s)}}\right)^{|\mathcal{O}\backslash\mathcal{S}|}. \tag{26}$$

Let $\beta_1 = 1 + (\frac{s}{2N} - \frac{1}{2(N-s)})$. Employing the inequality $\log(1-x) \leq -x - \frac{x^2}{2}$ for $0 < x < 1$ yields $\Pr\{\Phi = \emptyset\} \leq e^{-\frac{\beta_1 s}{N}|\mathcal{O}\backslash\mathcal{S}|}$. Following a similar argument one can obtain $\Pr\{\Phi = \emptyset\} \leq e^{-\frac{s}{N}|\mathcal{O}\backslash\mathcal{S}|}$.

Let $\beta = \max\{1, \beta_1\}$. Then

$$\Pr\{\Phi \neq \emptyset\} \geq 1 - e^{-\frac{\beta s}{N}|\mathcal{O}\backslash\mathcal{S}|} \geq \frac{1-\epsilon^\beta}{m}(|\mathcal{O}\backslash\mathcal{S}|) \tag{27}$$

from the definition of $s = \frac{N}{m}\log(1/\epsilon)$ and the fact that $1 - e^{-\frac{\beta s}{N}x}$ is a concave function. According to Lemma 2 in [41],

$$\mathbb{E}[f_{(i+1)_s}(\mathcal{S})|\mathcal{S}] \geq \frac{\Pr\{\Phi \neq \emptyset\}}{|\mathcal{O}\backslash\mathcal{S}|}\sum_{j\in\mathcal{O}\backslash\mathcal{S}} f_o(\mathcal{S}). \tag{28}$$

Hence,

$$\mathbb{E}\left[f_{(i+1)_s}(\mathcal{S})|\mathcal{S}\right] \geq \frac{1-\epsilon^\beta}{m}\sum_{j\in\mathcal{O}\backslash\mathcal{S}} f_j(\mathcal{S}). \tag{29}$$

On the other hand, employing Lemma 1 with $\mathcal{T} = \mathcal{O}\cup\mathcal{S}$ and invoking monotonicity of f yields

$$\begin{aligned}\frac{f(\mathcal{O}) - f(\mathcal{S})}{C(r)} &\leq \frac{f(\mathcal{O}\cup\mathcal{S}) - f(\mathcal{S})}{C(r)} \leq \sum_{j\in\mathcal{O}\backslash\mathcal{S}} f_j(\mathcal{S})\\ &\leq \frac{m}{1-\epsilon^\beta}\mathbb{E}\left[f_{(i+1)_s}(\mathcal{S})|\mathcal{S}\right], \end{aligned} \tag{30}$$

where $|\mathcal{O}\backslash\mathcal{S}| = r$. Let $c = \max\{\mathcal{C}_f, 1\}$. Applying the law of total expectation and the fact that $C(r) \leq c$ yields

$$\mathbb{E}[f(\mathcal{S}\cup\{(i+1)_s\}) - f(\mathcal{S})] \geq \frac{1-\epsilon^\beta}{mc}(f(\mathcal{O}) - \mathbb{E}[f(\mathcal{S})]). \tag{31}$$

With the established result, the proof simplifies to that of the classical greedy algorithm [52]. Therefore, by using a simple inductive argument,

$$\begin{aligned}\mathbb{E}[f(\mathcal{S}_{rg})] &\geq \left(1 - \left(1 - \frac{1-\epsilon^\beta}{mc}\right)^m\right)f(\mathcal{O})\\ &\geq \left(1 - e^{-\frac{1}{c}} - \frac{\epsilon^\beta}{c}\right)f(\mathcal{O}) = \alpha f(\mathcal{O}), \end{aligned} \tag{32}$$

where the last inequality is due to the facts that $(1+x)^y \leq e^{xy}$ for $y > 0$ and $e^{ax} \leq 1 + axe^a$ for $0 < x < 1$. Finally, the stated result follows by using the definition of $f(\mathcal{S})$. This completes the proof.

Appendix E. Proof of Theorem 3

Consider the ith iteration of Algorithm 2. Let $\mathcal{S}$ denote the current sampling set and let $(i+1)_g$ and $(i+1)_{rg}$ denote indices of the nodes selected at the $(i+1)$st iteration of the greedy sampling algorithm [15,19,20] and Algorithm 2, respectively. Similar to the proof of Theorem (2), we start by reducing the proof to that of the classical greedy algorithm. To this end, we employ Lemma 1 with $\mathcal{T} = \mathcal{O}\cup\mathcal{S}$ and use monotonicity of f to obtain

$$f(\mathcal{O}) - f(\mathcal{S}) \leq f(\mathcal{O}\cup\mathcal{S}) - f(\mathcal{S}) \leq c\sum_{j\in\mathcal{O}\backslash\mathcal{S}} f_j(\mathcal{S}). \tag{33}$$

Note that given the current sampling set $\mathcal{S}$, from the selection criteria of greedy and randomized-greedy algorithms for all j it follows that

$$f(\mathcal{O}) - f(\mathcal{S}) \leq cmf_{(i+1)_g}(\mathcal{S}), \tag{34}$$

where we used the fact that $|\mathcal{O}\backslash\mathcal{S}| \leq m$. On the other hand,

$$\begin{aligned}f(\mathcal{S}\cup\{(i+1)_{rg}\}) - f(\mathcal{S}) &= f_{(i+1)_{rg}}(\mathcal{S})\\ &= \eta_{i+1}f_{(i+1)_g}(\mathcal{S}). \end{aligned} \tag{35}$$

Combining (34) and (35) yields

$$f(\mathcal{S}\cup\{(i+1)_{rg}\}) - f(\mathcal{S}) \geq \frac{\eta_{i+1}}{mc}(f(\mathcal{O}) - f(\mathcal{S})). \tag{36}$$

Using a similar inductive argument as we did in the proof of Theorem 2 and due to the fact that $(1+x)^y \leq e^{xy}$ for $y > 0$, it follows that

$$\begin{aligned}f(\mathcal{S}_{rg}) &\geq \left(1 - \left(1 - \sum_{i=1}^{m}\frac{\eta_i}{mc}\right)\right)f(\mathcal{O})\\ &\geq \left(1 - e^{-\sum_{i=1}^{m}\frac{\eta_i}{mc}}\right)f(\mathcal{O}). \end{aligned} \tag{37}$$

Note that if we assume $\{\eta_i\}$ are independent, the term $\sum_{i=1}^{m}\eta_i$ is a sum of independent bounded random variables. Since $\{\eta_i\}$ are bounded random variables, by Popoviciu's inequality [54] for all $i \in [m]$ it holds that $\mathrm{Var}[\eta_i] \leq \frac{1}{4}$. Therefore, using Bernstein's inequality[54] it holds that for all $0 < q < 1$

$$\Pr\{\sum_{i=1}^{m}\eta_i < (1-q)m\mu_\epsilon\} \leq e^{-\frac{m(1-q)^2\mu_\epsilon^2}{\frac{1-q}{3}\mu_\epsilon + \frac{1}{4}}} = e^{-C(\epsilon,q)m}. \tag{38}$$

Employing this results in (37) yields

$$f(\mathcal{S}_{rg}) \geq \left(1 - e^{-\frac{(1-q)\mu_\epsilon}{c}}\right)f(\mathcal{O}), \tag{39}$$

with probability at least $1 - e^{C(\epsilon,q)m}$. Recalling the definition of $f(\mathcal{S})$ leads to the stated bound which in turn completes the proof.

Appendix F. Proof of Theorem 4

To prove the stated result, we begin by exploiting the recursive formulation of the marginal gain derived in Proposition 2 to establish a sufficient condition for weak submodularity of $f(\mathcal{S})$. More specifically, from the definition of the maximum element-wise curvature and (13), for all $(\mathcal{S},\mathcal{T},j)\in\mathcal{X}_l$ we have

$$\frac{f_j(\mathcal{T})}{f_j(\mathcal{S})}=\frac{(\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{T}}^2\mathbf{u}_j)(\sigma_j^2+\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}\mathbf{u}_j)}{(\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{S}}^2\mathbf{u}_j)(\sigma_j^2+\mathbf{u}_j^\top\bar{\mathbf{\Sigma}}_{\mathcal{T}}\mathbf{u}_j)}. \tag{40}$$

Next, we employ Courant-Fischer min-max theorem [42] to obtain

$$\begin{aligned}\frac{f_j(\mathcal{T})}{f_j(\mathcal{S})} &\le \frac{\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{T}}^2)(\sigma_j^2+\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{S}})\|\mathbf{u}_j\|_2^2)}{\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{S}}^2)(\sigma_j^2+\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{T}})\|\mathbf{u}_j\|_2^2)}\\ &\overset{(a)}{\le}\frac{\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{T}}^2)(\sigma_j^2+\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{S}}))}{\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{S}}^2)(\sigma_j^2+\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{T}}))},\end{aligned} \tag{41}$$

where (a) holds since

$$g(x)=\frac{\sigma_j^2+\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{S}})x}{\sigma_j^2+\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{T}})x} \tag{42}$$

is a monotonically increasing function for $x>0$ and $\|\mathbf{u}\|_2^2\le 1$. Given the fact that $\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{S}})=\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{S}}^{-1})^{-1}$, (41) simplifies to

$$\frac{f_j(\mathcal{T})}{f_j(\mathcal{S})}\le\frac{\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{T}}^{-1})^{-2}(\sigma_j^2+\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{S}}^{-1})^{-1})}{\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{S}}^{-1})^{-2}(\sigma_j^2+\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{T}}^{-1})^{-1})}. \tag{43}$$

By Weyl's inequality [42], for all $(\mathcal{S},\mathcal{T},j)\in\mathcal{X}_l$ it holds that $\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{N}}^{-1})\ge\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{T}}^{-1})\ge\lambda_{\min}(\bar{\mathbf{\Sigma}}_{\mathcal{S}}^{-1})\ge\lambda_{\min}(\mathbf{W}^{-1})$ and $\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{N}}^{-1})\ge\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{T}}^{-1})\ge\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{S}}^{-1})\ge\lambda_{\max}(\mathbf{W}^{-1})$. Hence, by definition of maximum element-wise curvature we have

$$\begin{aligned}\mathcal{C}_{\max} &\le \max_{j\in\mathcal{N}}\frac{\lambda_{\max}(\mathbf{W})^2(\sigma_j^2+\lambda_{\max}(\mathbf{W}))}{\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{N}}^{-1})^{-2}(\sigma_j^2+\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{N}}^{-1})^{-1})}\\ &\overset{(a)}{\le}\max_{j\in\mathcal{N}}\frac{(\sigma_j^2+\lambda_{\max}(\mathbf{W}))(\lambda_{\min}(\mathbf{W})^{-1}+\sigma_j^{-2})^2}{\lambda_{\max}(\mathbf{W})^{-2}(\sigma_j^2+(\lambda_{\min}(\mathbf{W})^{-1}+\sigma_j^{-2})^{-1})},\end{aligned} \tag{44}$$

where (a) follows since $\lambda_{\max}(\bar{\mathbf{\Sigma}}_{\mathcal{N}}^{-1})\le\lambda_{\max}(\mathbf{W}^{-1}+\sigma_j^{-2}\mathbf{I}_N)$ and because the maximum eigenvalue of a positive definite matrix satisfies the triangle inequality. Note that the denominator of the last inequality is always strictly larger than σ_j^2, and that $\lambda_{\max}(\mathbf{W})\ge\lambda_{\min}(\mathbf{W})$. Following some straight-forward algebra, we obtain

$$\mathcal{C}_{\max}\le\max_{j\in\mathcal{N}}\frac{\lambda_{\max}^2(\mathbf{W})}{\lambda_{\min}^2(\mathbf{W})}\left(1+\frac{\lambda_{\max}(\mathbf{W})}{\sigma_j^2}\right)^3 \tag{45}$$

which is the stated result. This completes the proof.

References

[1] A. Hashemi, R. Shafipour, H. Vikalo, G. Mateos, Sampling and reconstruction of graph signals via weak submodularity and semidefinite relaxation, in: Int. Conf. Acoust., Speech, Signal Process. (ICASSP), IEEE, 2018, pp. 4179–4183.

[2] A. Hashemi, R. Shafipour, H. Vikalo, G. Mateos, A novel scheme for support identification and iterative sampling of bandlimited graph signals, in: Global Conf. Signal and Inf. Process. (GlobalSIP), IEEE, 2018. (submitted)

[3] W. Huang, L. Goldsberry, N.F. Wymbs, S.T. Grafton, D.S. Bassett, A. Ribeiro, Graph frequency analysis of brain signals, IEEE J. Sel. Top. Signal Process. 10 (7) (2016) 1189–1203.

[4] J.A. Deri, J.M. Moura, New York City taxi analysis with graph signal processing, in: Global Conf. Signal and Inf. Process. (GlobalSIP), IEEE, 2016, pp. 1275–1279.

[5] D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, P. Vandergheynst, The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains, IEEE Signal Proces. Mag. 30 (3) (2013) 83–98.

[6] A. Sandryhaila, J.M.F. Moura, Discrete signal processing on graphs, IEEE Trans. Signal Process. 61 (7) (2013) 1644–1656.

[7] H. Shomorony, A.S. Avestimehr, Sampling large data on graphs, in: Global Conf. Signal and Inf. Process. (GlobalSIP), IEEE, 2014, pp. 933–936.

[8] M. Tsitsvero, S. Barbarossa, P. Di Lorenzo, Signals on graphs: uncertainty principle and sampling, IEEE Trans. Signal Process. 64 (18) (2016) 4845–4860.

[9] A. Anis, A. Gadde, A. Ortega, Efficient sampling set selection for bandlimited graph signals using graph spectral proxies, IEEE Trans. Signal Process. 64 (14) (2016) 3775–3789.

[10] S. Chen, R. Varma, A. Sandryhaila, J. Kovacevic, Discrete signal processing on graphs: sampling theory, IEEE Trans. Signal Process. 24 (63) (2015) 6510–6523.

[11] S.P. Chepuri, G. Leus, Subsampling for graph power spectrum estimation, in: Sensor Array and Multichannel Signal Process. Workshop (SAM), IEEE, 2016, pp. 1–5.

[12] A.G. Marques, S. Segarra, G. Leus, A. Ribeiro, Sampling of graph signals with successive local aggregations, IEEE Trans. Signal Process. 64 (7) (2016) 1832–1843.

[13] F. Gama, A.G. Marques, G. Mateos, A. Ribeiro, Rethinking sketching as sampling: linear transforms of graph signals, in: Asilomar Conf. Signals, Syst. and Computers, IEEE, 2016, pp. 522–526.

[14] A. Jayawant, A. Ortega, A distance-based formulation for sampling signals on graphs, in: Int. Conf. Acoust., Speech, Signal Process. (ICASSP), IEEE, 2018, pp. 6318–6322.

[15] L.F. Chamon, A. Ribeiro, Greedy sampling of graph signals, IEEE Trans. Signal Process. 66 (1) (2018) 34–47.

[16] S. Chen, R. Varma, A. Singh, J. Kovacevic, Signal recovery on graphs: fundamental limits of sampling strategies, IEEE Trans. Signal Inf. Process. Netw. 2 (4) (2016) 539–554.

[17] G. Puy, N. Tremblay, R. Gribonval, P. Vandergheynst, Random sampling of bandlimited signals on graphs, Appl. Comput. Harmonic Anal. (2018).

[18] N. Tremblay, P.-O. Amblard, S. Barthelmé, Graph sampling with determinantal processes, in: European Signal Process. Conf. (EUSIPCO), EURASIP, 2017, pp. 1674–1678.

[19] M. Shamaiah, S. Banerjee, H. Vikalo, Greedy sensor selection: leveraging submodularity, in: Conf. Decision and Control (CDC), IEEE, 2010, pp. 2572–2577.

[20] M. Shamaiah, S. Banerjee, H. Vikalo, Greedy sensor selection under channel uncertainty, IEEE Wirel. Commun. Lett. 1 (4) (2012) 376–379.

[21] A.G. Marques, S. Segarra, G. Leus, A. Ribeiro, Stationary graph processes and spectral estimation, IEEE Trans. Signal Process. 65 (22) (2017) 5911–5926.

[22] N. Perraudin, P. Vandergheynst, Stationary signal processing on graphs, IEEE Trans. Signal Process. 65 (13) (2017) 3462–3477.

[23] B. Girault, Stationary graph signals using an isometric graph translation, in: European Signal Process. Conf. (EUSIPCO), EURASIP, 2015, pp. 1516–1520.

[24] I. Pesenson, Sampling in Paley-Wiener spaces on combinatorial graphs, Trans. Am. Math. Soc. 360 (10) (2008) 5603–5627.

[25] S.K. Narang, A. Gadde, A. Ortega, Signal processing techniques for interpolation in graph structured data, in: Int. Conf. Acoust., Speech, Signal Process. (ICASSP), IEEE, 2013, pp. 5445–5449.

[26] A. Anis, A. Gadde, A. Ortega, Towards a sampling theorem for signals on arbitrary graphs, in: Int. Conf. Acoust., Speech, Signal Process. (ICASSP), IEEE, 2014, pp. 3864–3868.

[27] X. Wang, P. Liu, Y. Gu, Local-set-based graph signal reconstruction, IEEE Trans. Signal Process. 63 (9) (2015) 2432–2444.

[28] P. Di Lorenzo, S. Barbarossa, P. Banelli, S. Sardellitti, Adaptive least mean squares estimation of graph signals, IEEE Trans. Signal Inf. Process. Netw. 2 (4) (2016) 555–568.

[29] D. Romero, M. Ma, G.B. Giannakis, Kernel-based reconstruction of graph signals, IEEE Trans. Signal Process. 65 (3) (2017) 764–778.

[30] Y.C. Pati, R. Rezaiifar, P. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: Asilomar Conf. Signals, Syst. and Computers, IEEE, 1993, pp. 40–44.

[31] S. Joshi, S. Boyd, Sensor selection via convex optimization, IEEE Trans. Signal Process. 57 (2) (2009) 451–462.

[32] A. Sandryhaila, J.M.F. Moura, Discrete signal processing on graphs: frequency analysis, IEEE Trans. Signal Process. 62 (12) (2014) 3042–3054, doi:10.1109/TSP.2014.2321121.

[33] J.A. Deri, J.M. Moura, Spectral projector-based graph fourier transforms, IEEE J. Sel. Top. Signal Process. 11 (6) (2017) 785–795.

[34] R. Shafipour, A. Khodabakhsh, G. Mateos, E. Nikolova, A directed graph fourier transform with spread frequency components, IEEE Trans. Signal Process. (2018). (submitted; see also arXiv:1804.03000 [eess.SP])

[35] S. Sardellitti, S. Barbarossa, P. Di Lorenzo, On the graph Fourier transform for directed graphs, IEEE J. Sel. Top. Signal Process. 11 (6) (2017) 796–811.

[36] A.K. Farahat, A. Elgohary, A. Ghodsi, M.S. Kamel, Greedy column subset selection for large-scale data sets, Knowl. Inf. Syst. 45 (1) (2015) 1–34.

[37] S.M. Kay, Fundamentals of Statistical Signal Processing, Prentice Hall PTR, 1993.

[38] A. Hashemi, M. Ghasemi, H. Vikalo, U. Topcu, A randomized greedy algorithm for near-optimal sensor scheduling in large-scale sensor networks, in: American Control Conf. (ACC), IEEE, 2018.

[39] L. Chamon, A. Ribeiro, Approximate supermodularity bounds for experimental design, in: Advances in Neural Inf. Process. Syst. (NIPS), 2017, pp. 5409–5418.

[40] L.F. Chamon, G.J. Pappas, A. Ribeiro, The mean square error in Kalman filtering sensor selection is approximately supermodular, in: Conf. Decision and Control (CDC), IEEE, 2017, pp. 343–350.

[41] B. Mirzasoleiman, A. Badanidiyuru, A. Karbasi, J. Vondrak, A. Krause, Lazier than lazy greedy, in: AAAI Conf. Artificial Intelligence, AAAI, 2015.

[42] R. Bellman, Introduction to Matrix Analysis, SIAM, 1997.

[43] L.G. Valiant, A theory of the learnable, Commun ACM 27 (11) (1984) 1134–1142.
[44] L. Valiant, Probably approximately correct: nature's algorithms for learning and prospering in a complex world (2013).
[45] J.A. Tropp, An introduction to matrix concentration inequalities, Foundations Trends® in Mach.Learn. 8 (1–2) (2015) 1–230.
[46] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.
[47] C. You, D. Robinson, R. Vidal, Scalable sparse subspace clustering by orthogonal matching pursuit, in: Proceedings of the IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE, 2016, pp. 3918–3927.
[48] A. Hashemi, H. Vikalo, Evolutionary subspace clustering: discovering structure in self-expressive time-series data, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 3707–3711.
[49] A. Hashemi, H. Vikalo, Evolutionary self-expressive models for subspace clustering, IEEE J. Sel. Top. Signal Process. 12 (6) (2018) 1534–1546.
[50] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[51] A. Hashemi, M. Ghasemi, H. Vikalo, U. Topcu, Randomized greedy sensor selection: leveraging weak submodularity, IEEE Trans. Autom. Control (2021) 1–14.
[52] G.L. Nemhauser, L.A. Wolsey, M.L. Fisher, An analysis of approximations for maximizing submodular set functions I, Math. Program. 14 (1) (1978) 265–294.
[53] S. Lang, Algebraic Number Theory, vol. 110, Springer Science & Business Media, 2013.
[54] R.V. Hogg, A.T. Craig, Introduction to Mathematical Statistics, Prentice Hall, Upper Saddle River, New Jersey, 1995.