ROBUST GRAPH SIGNAL SAMPLING

*Başak Güler** *Ajinkya Jayawant** *A. Salman Avestimehr* *Antonio Ortega*

University of Southern California
Ming Hsieh Department of Electrical and Computer Engineering
Los Angeles, CA

ABSTRACT

This paper considers the graph signal sampling problem when some of the selected samples are lost or unavailable due to sensor failures or adversarial erasures. We formulate a robust graph signal sampling problem where only a subset of selected samples are received, and the goal is to maximize the worst-case performance. We propose a novel greedy robust sample selection algorithm and study its performance guarantees. Our numerical results demonstrate the performance improvement of the proposed algorithm over the existing schemes.

***Index Terms*—** Graph signal sampling, robust optimization, approximate submodular function maximization.

1. INTRODUCTION

Graphs are important tools for representing relationships, e.g., similarities, between entities in large datasets. Graph signals are signals defined over the vertices of a graph, which have been shown to be useful in analyzing real-world systems, such as sensor networks, biological data, or machine learning systems, using tools from graph signal processing [1–3].

Due to the size of most real-world graphs, it is often unfeasible to observe all of the data points on the graph. In such scenarios, one needs to select a small set of samples to observe, and make inferences about the remaining nodes in the network by leveraging the data obtained from the selected samples. Such setups are also related to the problem of active semi-supervised learning, where one chooses a small set of data points to label, and learns the missing labels by utilizing the labeled data along with the graph topology. The question is then to choose the best data points to sample in order to reconstruct the underlying data structure as accurately as possible. This is known as the graph signal sampling problem [4].

The conditions under which one can perfectly reconstruct the unknown graph signals from the sampled data points have been explored in various studies [4–8]. When the received samples are noisy, however, perfect reconstruction is not possible, and one has to look for the best approximate reconstruction in terms of the worst-case or average-case performance. Greedy algorithms have been proved to be effective in finding good sample sets under such scenarios [7,9,10], due to their approximate submodularity properties [11]. In particular, greedy set selection algorithms are known to provide a constant factor guarantee for approximating the optimal solution of a submodular function [12]. Approximate submodularity quantifies how submodular a function is and also provides a constant factor approximation guarantee for greedy algorithms [13]. It has been shown recently that the optimality criteria used in graph sampling [14] often exhibit properties similar to submodular functions, with approximate submodularity in noisy environments [11].

Most graph signal sampling methodologies assume that all samples from the selected sampling set can actually be observed. In environments where sensors can fail or samples can deleted by an adversary, however, one may receive only an unknown subset of the selected samples. This phenomenon can also be observed in social sensing scenarios such as crowdsourcing [15], where one may query a large group of users only to receive answers from a small subset of them. Similarly, in the case of sample set selection on graphs some samples may be lost. To address the issue, this paper deals with the problem of finding a sampling strategy for maximizing the reconstruction mean-squared error (MSE) of graph signals in the face of worst case sample deletions.

Motivated by such scenarios, in this work, we formulate a robust graph sampling problem, where one selects a set of k samples, but only receives $k-\tau$ of them, with the remaining τ samples being lost. Furthermore, it is not known beforehand which τ samples will be lost. The goal is to select the best sampling set of size k that has the best worst-case performance over *any* subset of size $k-\tau$ samples. We propose a robust greedy sample set selection algorithm buffers against the possible loss of samples and study its performance guarantees. As such, our work extends the greedy sample set selection setup from [11] to sampling scenarios in which one receives only a subset of the selected samples. Our robust graph signal sampling problem is also related to the existing robust optimization frameworks [16–18], which mainly focus on the optimization of a monotone function with a given approximate submodularity parameter. Our focus is on the functions used as an optimality criterion for graph signal sampling, and through numerical evaluations, we demonstrate how in this setup the proposed approach outperforms state-of-the-art robust maximization techniques.

In the remainder of the paper, x stands for a scalar variable, whereas $\mathbf{x}$ is a vector. $\mathcal{X}$ represents a set with cardinality $|\mathcal{X}|$. $\mathbf{A}$ is a matrix with A_{ij} denoting its $(i,j)^{th}$ element. We use $\mathrm{tr}(\mathbf{A})$ to denote the trace of $\mathbf{A}$. $\mathrm{diag}(\cdot)$ represents a diagonal matrix.

2. SYSTEM MODEL

We consider an undirected graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ with $|\mathcal{V}|=N$ nodes. A graph signal $\mathbf{x}$ is defined over the vertices of $\mathcal{G}$, where x_i represents the graph signal associated with node $i \in V$. We define an adjacency matrix $\mathbf{A}$ associated with $\mathcal{G}$, where $A_{ij}=1$ if and only if $(i,j)\in\mathcal{E}$ and $A_{ij}=0$ otherwise. The graph Laplacian of $\mathcal{G}$ is then given by $\mathbf{L}=\mathbf{D}-\mathbf{A}$, where $\mathbf{D}$ is a diagonal matrix such that $D_{ii}=\sum_{j:(i,j)\in\mathcal{E}} A_{ij}$. Assuming that $\mathbf{L}$ has eigenvalues $0=\lambda_1 \leq \ldots \leq \lambda_N$ with corresponding eigenvectors $\mathbf{V}=[\mathbf{v}_1,\ldots,\mathbf{v}_N]$, we can write $\mathbf{L}=\mathbf{V}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{\Sigma}=\mathrm{diag}(\lambda_1,\ldots,\lambda_N)$. The graph

* Equal contribution.

Fourier transform of $\mathbf{x}$ is defined as $\tilde{\mathbf{x}} = \mathbf{V}^T\mathbf{x}$ [1]. A graph signal is called $\mathcal{K}$-bandlimited if $\tilde{\mathbf{x}} = [\tilde{\mathbf{x}}_{\mathcal{K}}^T \ \tilde{\mathbf{x}}_{V\setminus\mathcal{K}}^T]^T$ such that $\tilde{\mathbf{x}}_{V\setminus\mathcal{K}} = \mathbf{0}$. One can represent $\mathbf{x}$ as, $\mathbf{x} = \mathbf{V}_{\mathcal{K}}\tilde{\mathbf{x}}_{\mathcal{K}}$ where $\mathbf{V}_{\mathcal{K}} = [\mathbf{v}_1 \dots \mathbf{v}_{|\mathcal{K}|}]$.

We consider a signal model along the lines of [11]. In particular, we study the class of graph signals for which $\tilde{\mathbf{x}}_{\mathcal{K}}$ has mean $\mathbb{E}[\tilde{\mathbf{x}}_{\mathcal{K}}] = \mathbf{0}$ and covariance $\mathbb{E}[\tilde{\mathbf{x}}_{\mathcal{K}}\tilde{\mathbf{x}}_{\mathcal{K}}^T] = \mathbf{\Lambda} = \text{diag}(\delta_1, \dots, \delta_{\mathcal{K}})$. Without loss of generality, we assume $\mathbf{\Lambda}$ has full-rank, as one can always adjust $\mathcal{K}$ to remove the elements for which $\delta_i = 0$ otherwise. The sampling process has access to a noisy version of the graph signal,

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{1}$$

where the noise vector $\mathbf{n}$ has mean $\mathbb{E}[\mathbf{n}] = \mathbf{0}$ and covariance $\mathbb{E}[\mathbf{n}\mathbf{n}^T] = \mathbf{\Omega} = \text{diag}(\mu_1, \dots, \mu_N)$.

The sampling process chooses a set of samples, denoted by $\mathcal{S} \subseteq \mathcal{V}$, from the signal in (1). This process is represented by a sampling matrix $\mathbf{S} \in \{0,1\}^{|\mathcal{S}|\times N}$ in which the i^{th} row corresponds to the i^{th} row of the $N \times N$ identity matrix $\mathbf{I}$. The samples are given by:

$$\mathbf{y}_{\mathcal{S}} = \mathbf{S}\mathbf{y}. \tag{2}$$

For the noiseless case $\mathbf{n}$, one can obtain an exact reconstruction of $\mathbf{x}$ from the samples $\mathbf{y}_{\mathcal{S}}$ under the conditions identified in [5,6]. For the noisy setup, however, one has to approximate $\mathbf{x}$ from the noisy samples $\mathbf{y}_{\mathcal{S}}$, as perfect reconstruction is no longer feasible. Accordingly, we represent the reconstructed signal by $\hat{\mathbf{x}} = \mathbf{W}\mathbf{y}_{\mathcal{S}}$ where $\mathbf{W}$ is an $N \times |\mathcal{S}|$ reconstruction matrix. The corresponding error covariance matrix is given by:

$$\mathbf{K} = \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T]. \tag{3}$$

The optimal reconstruction matrix $\mathbf{W}$ is chosen as to minimize the scalar cost function $J(\mathbf{W}) = \mathbf{z}^T\mathbf{K}\mathbf{z}$ over the error covariance matrix for all $\mathbf{z} \in \mathbb{R}^N$, as detailed in [11], leading to the following error covariance matrix:

$$\mathbf{K} = \mathbf{V}_{\mathcal{K}}(\mathbf{\Lambda}^{-1} + \mathbf{V}_{\mathcal{K}}^T\mathbf{S}^T\mathbf{S}\mathbf{\Omega}^{-1}\mathbf{S}^T\mathbf{S}\mathbf{V}_{\mathcal{K}})^{-1}\mathbf{V}_{\mathcal{K}}^T. \tag{4}$$

As can be observed from (4), the chosen sampling set $\mathcal{S}$ affects the error covariance matrix through the following term:

$$\mathbf{K}(\mathcal{S}) = (\mathbf{\Lambda}^{-1} + \mathbf{V}_{\mathcal{K}}^T\mathbf{S}^T\mathbf{S}\mathbf{\Omega}^{-1}\mathbf{S}^T\mathbf{S}\mathbf{V}_{\mathcal{K}})^{-1} \tag{5}$$

$$= (\mathbf{\Lambda}^{-1} + \sum_{i\in\mathcal{S}} \mu_i^{-1}\mathbf{r}_i\mathbf{r}_i^T)^{-1}. \tag{6}$$

where $\mathbf{r}_i^T$ is the i^{th} row of $\mathbf{V}_{\mathcal{K}}$. Then, the sampling set selection problem is to choose the best sampling set $\mathcal{S}^*$ of a given size k with respect to an objective function defined over the matrix from (6),

$$\mathcal{S}^* = \arg\max_{\mathcal{S}:|\mathcal{S}|=k} f((\mathbf{\Lambda}^{-1} + \sum_{i\in\mathcal{S}} \mu_i^{-1}\mathbf{r}_i\mathbf{r}_i^T)^{-1}). \tag{7}$$

It is useful to note that the formulation from (7) depends on the choice of function $f(.)$ In this work, we let $f(.)$ be a monotonically non-decreasing, non-negative function with $f(\emptyset) = 0$.

Finding $\mathcal{S}^*$ is in general an NP-hard problem [19]. When the objective function (7) is submodular, greedy algorithms can approximate it in polynomial time. However, most functions used as optimality criteria in graph signal sampling are not submodular. In such cases, performance of greedy algorithms has been studied using the notion of *approximate submodularity*, which measures how close a function is to being submodular [11].

3. PROBLEM FORMULATION

The problem from (7) identifies the best sampling set assuming one receives all the samples from the selected set $\mathcal{S}^*$. If there are sensor failures or in adversarial environments, not all selected sensors will provide samples. Accordingly, we define the robust graph sampling problem as the selection of a set of k samples under the condition that one receives only $k - \tau$ of them, where we do not have have prior knowledge of which samples will be lost. Formally, we wish to select a set of samples $\mathcal{S}$ to maximize the worst-case performance:

$$\mathcal{S}^* = \arg\max_{\mathcal{S}:|\mathcal{S}|=k} \min_{\substack{\mathcal{A}:\mathcal{A}\subseteq\mathcal{S} \\ |\mathcal{A}|=k-\tau}} f((\mathbf{\Lambda}^{-1} + \sum_{i\in\mathcal{A}} \mu_i^{-1}\mathbf{r}_i\mathbf{r}_i^T)^{-1}). \tag{8}$$

To solve (8) we propose a greedy algorithm consisting of two stages (see Algorithm 1). In the first stage, the algorithm selects τ nodes in an oblivious manner, similar to the first stage of existing two-stage algorithms [17, 18]; at each step, the best node is selected from the set of available nodes, by assuming all the previous ones may be lost. In the second stage, we know that at least j nodes will be received from the first stage. The algorithm then selects the next node to maximize the worst case performance when combined with any j nodes from the samples selected so far. This is unlike the selection criterion used in [18], which discards the nodes selected in the first stage. For $\tau = 0$, Algorithm 1 reduces to the conventional greedy algorithm [11].

Algorithm 1 Robust Graph Sampling

1: Initialize graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and sampling set $\mathcal{S} = \emptyset$.
2: **Stage 1:**
3: **for** $i = 1, \dots, \tau$
4: Choose node u^* such that:

$$u^* \leftarrow \arg\max_{u\in\mathcal{V}\setminus\mathcal{S}} f(\{u\}) \tag{9}$$

▷ $\mathcal{S}$ is the set of already selected nodes at the beginning of iteration i.
5: $\mathcal{S} = \mathcal{S} \cup \{u^*\}$
6: Set $j := 0$
7: **Stage 2:**
8: **for** $i = \tau + 1, \dots, k$
9: Choose node u^* such that:

$$u^* \leftarrow \arg\max_{u\in\mathcal{V}\setminus\mathcal{S}} \min_{\mathcal{A}\in\mathcal{S}^j} f\big(\mathcal{A} \cup \{u\}\big) \tag{10}$$

▷ $\mathcal{S}^j$ denotes the set of j-element subsets of $\mathcal{S}$.
10: $\mathcal{S} = \mathcal{S} \cup \{u^*\}$
11: $j = j + 1$

In order to study the performance of Algorithm 1, we utilize the following definitions. The first one is related to the approximate supermodularity notion from [11]. In the sequel, we use the shorthand notation $f(\mathcal{A}) = f((\mathbf{\Lambda}^{-1} + \sum_{i\in\mathcal{A}} \mu_i^{-1}\mathbf{r}_i\mathbf{r}_i^T)^{-1})$.

Definition 1. *(Approximate submodularity) A function f is α-approximately submodular if,*

$$f(\mathcal{A} \cup \{i\}) - f(\mathcal{A}) \geq \alpha(f(\mathcal{B} \cup \{i\}) - f(\mathcal{B})) \tag{11}$$

for all $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ and $i \in \mathcal{V}\setminus\mathcal{B}$. where $\alpha \in [0, 1]$ is chosen as the largest scalar satisfying (11)*.*

We note that $\alpha = 0$ always holds since f is monotonically non-decreasing, and f becomes submodular when $\alpha = 1$.

Definition 2. *(Bipartite subadditivity ratio) [18] The bipartite subadditivity ratio of f is the largest $\theta \in [0, 1]$ such that*

$$\frac{f(\mathcal{A}) + f(\mathcal{B})}{f(\mathcal{A} \cup \mathcal{B})} \geq \theta, \quad \forall \mathcal{A}, \mathcal{B} \subseteq \mathcal{V} \text{ such that } \mathcal{A} \cap \mathcal{B} = \emptyset. \tag{12}$$

Next, we provide a lower bound on the performance of the robust greedy sample set selection algorithm.

Theorem 1. *Let $\mathcal{S}$ be the sampling set selected by Algorithm 1, with $\mathcal{S}_0$ and $\mathcal{S}_1$ representing the samples selected in the first and second stages, respectively. Denote the worst-case performance of $\mathcal{S}$ by,*

$$\phi(\mathcal{S}) = \min_{\substack{\mathcal{A}:\mathcal{A}\subseteq\mathcal{S} \\ |\mathcal{A}|=k-\tau}} f(\mathcal{A}) \tag{13}$$

and let $\phi(\mathcal{S}^*)$ denote the optimal solution of (8)*. Then,*

$$\phi(\mathcal{S}) \geq \left(1-e^{-\bar{\alpha}}\right)\left(\theta\,\phi(\mathcal{S}^*) - \tau\frac{\beta}{\alpha}\right) \tag{14}$$

where $\bar{\alpha}$ is the approximate submodularity ratio from (11) *of the function,*

$$g_{\mathcal{S}_0}(\mathcal{S}_1) = \min_{\substack{\mathcal{A}:\mathcal{A}\subseteq\mathcal{S}_0\cup\mathcal{S}_1 \\ |\mathcal{A}|=|\mathcal{S}_1|}} f(\mathcal{A}). \tag{15}$$

where $\mathcal{S}_1 \subseteq \mathcal{V}\backslash\mathcal{S}_0$, and $\beta = \max_{i\in\mathcal{V}} f(\{i\})$.

Proof. Consider the optimal sampling set $\mathcal{S}^*$ from (8), and let $\mathcal{W} \subseteq \mathcal{S}^*$ be its worst-case subset of size $k-\tau$, i.e., $f(\mathcal{W}) \leq f(\mathcal{W}')$ for all $\mathcal{W}' \subseteq \mathcal{S}^*$ with $|\mathcal{W}'| = k-\tau$. Then,

$$\phi(\mathcal{S}^*) = f(\mathcal{W}) \leq f(\mathcal{W}\cup\mathcal{S}_0\cup\mathcal{R}) \tag{16}$$

where $\mathcal{R} \subseteq \mathcal{V}\backslash(\mathcal{W}\cup\mathcal{S}_0)$ is an arbitrary set that satisfies the condition $|\mathcal{W}\cup\mathcal{S}_0\cup\mathcal{R}| = k$, and (16) follows from the fact that f is monotonically non-decreasing. Let $\mathcal{M} \subseteq \mathcal{W}\cup\mathcal{S}_0\cup\mathcal{R}$ be a set of size $|\mathcal{M}| = k-\tau$ such that $f(\mathcal{M}) \leq f(\mathcal{M}')$ for all $\mathcal{M}' \subseteq \mathcal{W}\cup\mathcal{S}_0\cup\mathcal{R}$ with $|\mathcal{M}'| = k-\tau$. Then, we have from (16),

$$\begin{aligned} f(\mathcal{W}\cup\mathcal{S}_0\cup\mathcal{R}) &= f(\mathcal{M}\cup(\mathcal{W}\cup\mathcal{S}_0\cup\mathcal{R})\backslash\mathcal{M}) && (17)\\ &\leq \frac{1}{\theta}(f(\mathcal{M}) + f((\mathcal{W}\cup\mathcal{S}_0\cup\mathcal{R})\backslash\mathcal{M})) && (18)\\ &\leq \frac{1}{\theta}(g_{\mathcal{S}_0}(\mathcal{S}_1^*) + f((\mathcal{W}\cup\mathcal{S}_0\cup\mathcal{R})\backslash\mathcal{M})) && (19)\end{aligned}$$

where (18) follows from (12), and $g_{\mathcal{S}_0}(\mathcal{S}_1^*)$ in (19) is defined as,

$$g_{\mathcal{S}_0}(\mathcal{S}_1^*) = \max_{\substack{\mathcal{S}_1:\mathcal{S}_1\subseteq\mathcal{V}\backslash\mathcal{S}_0 \\ |\mathcal{S}_1|=k-\tau}} \min_{\substack{\mathcal{A}:\mathcal{A}\subseteq\mathcal{S}_0\cup\mathcal{S}_1 \\ |\mathcal{A}|=|\mathcal{S}_1|}} f(\mathcal{A}), \tag{20}$$

hence,

$$\begin{aligned} g_{\mathcal{S}_0}(\mathcal{S}_1^*) &\geq \min_{\substack{\mathcal{A}:\mathcal{A}\subseteq\mathcal{S}_0\cup\mathcal{T} \\ |\mathcal{A}|=k-\tau}} f(\mathcal{A}),\ \forall\,\mathcal{T}\subseteq\mathcal{V}\backslash\mathcal{S}_0 \text{ s.t. } |\mathcal{T}|=k-\tau && (21)\\ &\geq f(\mathcal{M}) && (22)\end{aligned}$$

from which (19) follows. By denoting $(\mathcal{W}\cup\mathcal{S}_0\cup\mathcal{R})\backslash\mathcal{M} \triangleq \{e_1,\ldots,e_\tau\}$, we find from (19) that

$$\begin{aligned} &\frac{1}{\theta}(g_{\mathcal{S}_0}(\mathcal{S}_1^*) + f((\mathcal{W}\cup\mathcal{S}_0\cup\mathcal{R})\backslash\mathcal{M}))\\ &= \frac{1}{\theta}(g_{\mathcal{S}_0}(\mathcal{S}_1^*) + f(\{e_1,\ldots,e_\tau\}) && (23)\\ &= \frac{1}{\theta}(g_{\mathcal{S}_0}(\mathcal{S}_1^*) + \sum_{i=1}^{\tau}(f(\{e_i\}\cup\{e_{i+1},\ldots,e_\tau\}) - f(\{e_{i+1},\ldots,e_\tau\})) && (24)\\ &\leq \frac{1}{\theta}(g_{\mathcal{S}_0}(\mathcal{S}_1^*) + \sum_{i=1}^{\tau}\frac{1}{\alpha}f(\{e_i\}) && (25)\\ &\leq \frac{1}{\theta}\left(g_{\mathcal{S}_0}(\mathcal{S}_1^*) + \tau\frac{\beta}{\alpha}\right) && (26)\end{aligned}$$

where (24) is from telescopic sum, (25) is from (11) and $f(\emptyset) = 0$.

Next, note that $\mathcal{S}_0$ is the set of nodes selected in the first stage of Algorithm 1. In the second stage, the algorithm aims to solve the following set selection problem:

$$g_{\mathcal{S}_0}(\mathcal{S}_1^*) = \max_{\substack{\mathcal{S}_1:\mathcal{S}_1\subseteq\mathcal{V}\backslash\mathcal{S}_0 \\ |\mathcal{S}_1|=k-\tau}} \min_{\substack{\mathcal{A}:\mathcal{A}\subseteq\mathcal{S}_0\cup\mathcal{S}_1 \\ |\mathcal{A}|=|\mathcal{S}_1|}} f(\mathcal{A}). \tag{27}$$

in a greedy manner. That is, (10) constructs a set $\mathcal{S}_1$ of size $k-\tau$ iteratively, at each iteration by selecting the node u that essentially maximizes the function,

$$\begin{aligned} u^* &= \arg\max_u g_{\mathcal{S}_0}(\mathcal{U}_i\cup\{u\}) && (28)\\ &= \arg\max_u \min_{\substack{\mathcal{A}:\mathcal{A}\subseteq\mathcal{S}_0\cup\mathcal{U}_i\cup\{u\} \\ |\mathcal{A}|=|\mathcal{U}_i|+1}} f(\mathcal{A}) && (29)\\ &= \arg\max_u \min_{\substack{\mathcal{A}:\mathcal{A}\subseteq\mathcal{S}_0\cup\mathcal{U}_i \\ |\mathcal{A}|=|\mathcal{U}_i|}} f(\mathcal{A}\cup\{u\}) && (30)\end{aligned}$$

where $\mathcal{U}_i$ denotes the already selected nodes at iteration i, accordingly, $\mathcal{S}_1 = \mathcal{U}_{k-\tau}$. Function $g_{\mathcal{S}_0}(\cdot)$ in (28) is monotonically non-decreasing, which can be proved by contradiction. Then, by letting $\bar{\alpha}$ denote the approximate submodularity ratio of $g_{\mathcal{S}_0}(\cdot)$, one can show through similar steps from [11, 20] for bounding the performance of greedy algorithms that,

$$\phi(\mathcal{S}) = g_{\mathcal{S}_0}(\mathcal{S}_1) \geq (1-e^{\bar{\alpha}})g_{\mathcal{S}_0}(\mathcal{S}_1^*) \tag{31}$$

which, combined with (26), leads to (14). □

An important optimality criterion based on the error covariance $\mathbf{K}(\mathcal{S})$ from (6) is minimizing the mean-squared error (MSE) of the reconstructed signal. Also known as A-optimality, the MSE criterion is quantified by $\mathrm{tr}(\mathbf{K}(\mathcal{S}))$. This problem can be equivalently represented in the following maximization form,

$$\begin{aligned} f(\mathcal{S}) &= \mathrm{tr}(\mathbf{\Lambda}) - \mathrm{tr}(\mathbf{K}(\mathcal{S})) && (32)\\ &= \mathrm{tr}(\mathbf{\Lambda}) - \mathrm{tr}((\mathbf{\Lambda}^{-1} + \sum_{i\in\mathcal{S}}\mu_i^{-1}\mathbf{r}_i\mathbf{r}_i^T)^{-1}) && (33)\end{aligned}$$

which is a non-negative, monotonically non-decreasing set function, with $f(\emptyset) = 0$. In the following, we study the approximate submodularity characteristics of the MSE function.

For tractability of our further analysis, we let $\mathbf{\Lambda} = \sigma_x^2\mathbf{I}$ and $\mathbf{\Omega} = \sigma_n^2\mathbf{I}$, from which (33) can be written as

$$f(\mathcal{S}) = \sigma_x^2\left(|\mathcal{K}| - \mathrm{tr}((\mathbf{I} + \gamma\sum_{i\in\mathcal{S}}\mathbf{r}_i\mathbf{r}_i^T)^{-1})\right) \tag{34}$$

where $\gamma = \sigma_x^2/\sigma_n^2$ is the SNR of the graph signals.

The submodularity ratio α for function f in (34) can be bounded following the same steps in [11][Theorem 3]. The submodularity ratio $\bar{\alpha}$, however, is based on function (15) instead, which may be different f in general. As such, the next result provides a lower bound on the approximate submodularity ratio $\bar{\alpha}$.

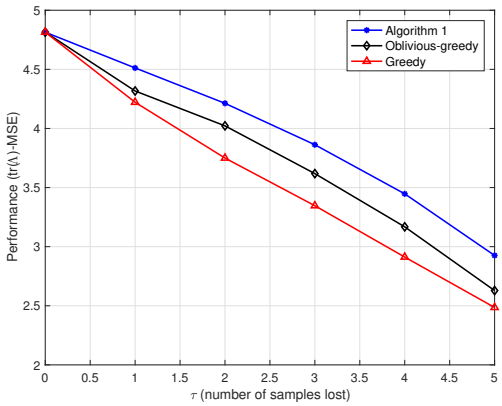
Proposition 1. *The approximate submodularity ratio $\bar{\alpha}$ in* (15)*, where f is the MSE criterion from* (34)*, can be bounded below by,*

$$\bar{\alpha} \geq \frac{\gamma^{-1} + (1+\gamma)^{-1}}{\gamma^{-1}+\rho}\rho(1+\gamma)^{-2} \tag{35}$$

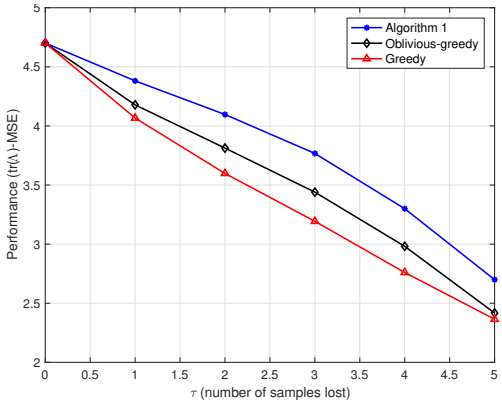
where $\rho = \min_{i\in\mathcal{V}}||\mathbf{r}_i||^2$.

Proof. The proof follows the lines of [11] but over the worst-case solutions in (15), and is omitted due to space considerations. □

We observe that $\bar{\alpha}$ increases as SNR decreases, hence the function becomes more submodular. As a result, the greedy algorithm

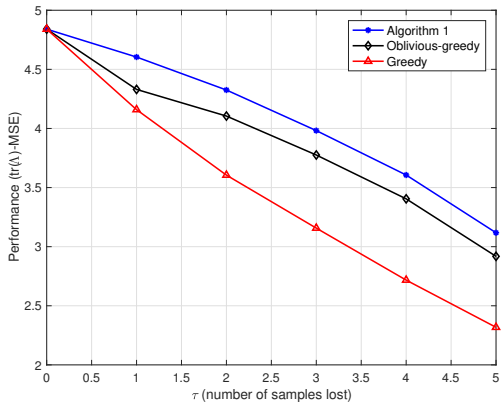


(a)

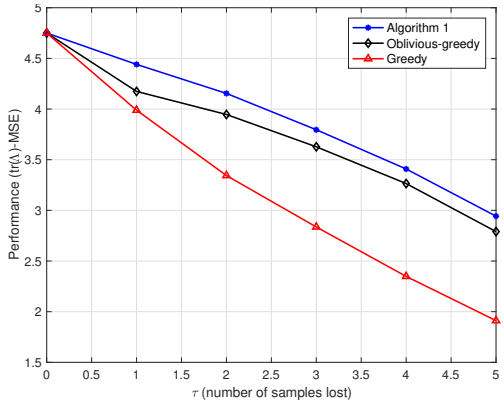


(b)

Fig. 1. Performance comparisons for Barabasi-Albert graph with (a) $N = 100$, (b) $N = 200$.



(a)



(b)

Fig. 2. Performance comparisons for Erdös-Rényi graph with (a) $N = 100$, (b) $N = 200$.

provides a good approximation of the optimal solution in low SNR environments. In contrast, for the noiseless case almost every subset of $|\mathcal{K}|$ samples provides perfect reconstruction, hence in high SNR scenarios the specific choice of sampling set is less relevant [11].

4. PERFORMANCE EVALUATIONS

In our simulations, we let $\mathbf{\Lambda} = \mathbf{I}$ and $\mathbf{\Omega} = \sigma_n^2\mathbf{I}$ with $\sigma_n^2 = 10^{-2}$, and $|\mathcal{K}| = 5$. We compare three algorithms, Algorithm 1, the conventional greedy set selection algorithm, and the robust greedy optimization algorithm (termed as oblivious-greedy algorithm) from [18, Algorithm 1]. The oblivious-greedy algorithm is a 2 stage algorithm as Algorithm 1. In stage 1, nodes are selected obliviously as in Algorithm 1. In stage 2, however, the two algorithms differ. The oblivious-greedy algorithm discards the τ nodes selected in the first stage, and starting from an empty set, applies the conventional greedy algorithm for selecting $k - \tau$ nodes in the second stage. On the other hand, Algorithm 1 combines the information from samples selected in the first stage and selects the next node to maximize the performance of the worst case subset. The performance of a selected set $\mathcal{S}$ is measured by the worst-case performance $\min_{\substack{\mathcal{A}:\mathcal{A}\subseteq\mathcal{S} \\ |\mathcal{A}|=k-\tau}} (\mathrm{tr}(\mathbf{\Lambda}) - \mathrm{tr}(\mathbf{K}(\mathcal{A}))) \triangleq \mathrm{tr}(\mathbf{\Lambda}) - MSE$.

We first consider a Barabasi-Albert graph created from 4 seed nodes, which has a scale-free degree structure like many real-world topologies such as the WWW. The results, given in Fig. 1, show that Algorithm 1 can provide performance gains of up to 20% improvement over the greedy algorithm and up to 12% improvement over the oblivious-greedy algorithm. Next, we consider an Erdös-Rényi graph with each edge drawn with probability $p = 0.2$. The results are illustrated in Fig. 2. We observe that in this setup Algorithm 1 can provide performance gains of up to 54% over the conventional greedy algorithm, and up to 7% improvement over the oblivious-greedy algorithm. In our performances $\mathrm{tr}(\mathbf{\Lambda}) = 5$. So, in Fig. 1 and Fig. 2 at a performance level of 4.5 a drop in the performance by 1 for a different method indicates a difference in the SNR of 4.7dB between those two methods whereas at the performance level of 3 a drop in the performance by 1 unit indicates a drop in the SNR by 1.76dB. As expected, the performance of all three algorithms is the same when $\tau = 0$, since in this case both Algorithm 1 and oblivious-greedy reduce to the greedy algorithm. The performance improvement of Algorithm 1 over the oblivious-greedy becomes more significant for the scale-free network topology.

5. CONCLUSION

We considered a graph sampling problem in which one receives only a subset of the selected samples. For this problem, we proposed a greedy robust sample selection algorithm and investigated its performance guarantees. Numerical experiments show that the proposed setup can significantly improve the performance over conventional greedy sample selection algorithms, as well as state-of-the-art robust set selection algorithms.

6. REFERENCES

[1] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[2] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[3] Fernando Gama, Antonio G Marques, Geert Leus, and Alejandro Ribeiro, "Convolutional neural network architectures for signals supported on graphs," *IEEE Transactions on Signal Processing*, vol. 67, no. 4, pp. 1034–1049, 2019.

[4] Isaac Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Transactions of the American Mathematical Society*, vol. 360, no. 10, pp. 5603–5627, 2008.

[5] Aamir Anis, Akshay Gadde, and Antonio Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.

[6] I. Shomorony and A. S. Avestimehr, "Sampling large data on graphs," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP'14)*, Dec 2014, pp. 933–936.

[7] Siheng Chen, Rohan Varma, Aliaksei Sandryhaila, and Jelena Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, 2015.

[8] Sundeep Prabhakar Chepuri and Geert Leus, "Graph sampling for covariance estimation," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 451–466, 2017.

[9] Mikhail Tsitsvero, Sergio Barbarossa, and Paolo Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, 2016.

[10] Ajinkya Jayawant and Antonio Ortega, "A distance-based formulation for sampling signals on graphs," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6318–6322.

[11] Luiz F. O. Chamon and Alejandro Ribeiro, "Greedy sampling of graph signals," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 34–47, 2018.

[12] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher, "An analysis of approximations for maximizing submodular set functions-I," *Mathematical programming*, vol. 14, no. 1, pp. 265–294, 1978.

[13] Andrew An Bian, Joachim M Buhmann, Andreas Krause, and Sebastian Tschiatschek, "Guarantees for greedy maximization of non-submodular functions with applications," *arXiv preprint arXiv:1703.02100*, 2017.

[14] Friedrich Pukelsheim, *Optimal design of experiments*, vol. 50, SIAM, 1993.

[15] Javier Maroto and Antonio Ortega, "Efficient worker assignment in crowdsourced data labeling using graph signal processing," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2271–2275.

[16] Andreas Krause, H Brendan McMahan, Carlos Guestrin, and Anupam Gupta, "Robust submodular observation selection," *Journal of Machine Learning Research*, vol. 9, no. Dec, pp. 2761–2801, 2008.

[17] James B Orlin, Andreas S Schulz, and Rajan Udwani, "Robust monotone submodular function maximization," in *International Conference on Integer Programming and Combinatorial Optimization*. Springer, 2016, pp. 312–324.

[18] Ilija Bogunovic, Junyao Zhao, and Volkan Cevher, "Robust maximization of non-submodular objectives," *International Conference on Artificial Intelligence and Statistics (AISTATS'18)*, vol. 84, 2018.

[19] Andreas Krause and Daniel Golovin, *Submodular function maximization*, Cambridge University Press, 2014.

[20] Abhimanyu Das and David Kempe, "Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection," *arXiv preprint arXiv:1102.3975*, 2011.