

Fast Graph Sampling Set Selection Using Gershgorin Disc Alignment

Yuanchao Bai, *Student Member, IEEE*, Fen Wang, Gene Cheung, *Senior Member, IEEE*,
Yuji Nakatsukasa, Wen Gao, *Fellow, IEEE*

Abstract—Graph sampling set selection, where a subset of nodes are chosen to collect samples to reconstruct a smooth graph signal, is a fundamental problem in graph signal processing (GSP). Previous works employ an unbiased least-squares (LS) signal reconstruction scheme and select samples via expensive extreme eigenvector computation. Instead, we assume a biased graph Laplacian regularization (GLR) based scheme that solves a system of linear equations for reconstruction. We then choose samples to minimize the condition number of the coefficient matrix—specifically, maximize the smallest eigenvalue λ_{\min} . Circumventing explicit eigenvalue computation, we maximize instead the lower bound of λ_{\min} , designated by the smallest left-end of all Gershgorin discs of the matrix. To achieve this efficiently, we first convert the optimization to a dual problem, where we minimize the number of samples needed to align all Gershgorin disc left-ends at a chosen lower-bound target T . Algebraically, the dual problem amounts to optimizing two disc operations: i) shifting of disc centers due to sampling, and ii) scaling of disc radii due to a similarity transformation of the matrix. We further reinterpret the dual as an intuitive disc coverage problem bearing strong resemblance to the famous NP-hard set cover (SC) problem. The reinterpretation enables us to derive a fast approximation scheme from a known SC error-bounded approximation algorithm. We find an appropriate target T efficiently via binary search. Extensive simulation experiments show that our disc-based sampling algorithm runs substantially faster than existing sampling schemes and outperforms other eigen-decomposition-free sampling schemes in reconstruction error.

Index Terms—Graph signal processing, graph sampling, Gershgorin circle theorem, combinatorial optimization

I. INTRODUCTION

Sampling of smooth signals—and subsequent signal reconstruction from collected samples—are fundamental problems in signal processing. Graph sampling extends the well-known studies of Nyquist sampling on regular data kernels to irregular data kernels described by graphs. Graph sampling can be studied under different contexts: aggregation sampling [1], [2], local measurement [3] and subset sampling [4]. In this paper, we focus on *graph sampling set selection*: a subset of nodes are chosen to collect samples to reconstruct a graph signal with a well-defined notion of smoothness.

Yuanchao Bai and Wen Gao are with the School of Electronics Engineering and Computer Science, Peking University, Beijing, 100871, China. (e-mail: {yuanchao.bai, wgao}@pku.edu.cn).

Gene Cheung is the *corresponding author* and is with the Department of Electrical Engineering & Computer Science, York University, Toronto, Canada. (e-mail: genec@yorku.ca).

Fen Wang is with State Key Laboratory of ISN, Xidian University, Xi'an 710071, Shaanxi, China (e-mail: fenwang@stu.xidian.edu.cn)

Yuji Nakatsukasa is with the University of Oxford, UK, and the National Institute of Informatics, Japan. (e-mail: nakatsukasa@maths.ox.ac.uk)

For the noiseless case, graph sampling set selection is well investigated [5]–[7]; in particular, [7] shows that random selection of K samples on a graph for a target K -bandlimited signal can enable perfect reconstruction with high probability. The noisy case where acquired samples are corrupted by additive noise is more challenging, and there exists many proposals. Assuming that the target signal is bandlimited, a class of schemes compute one or more extreme eigenvectors of a graph Laplacian matrix (or sub-matrix) and examine the per-node magnitudes to greedily determine a sample set. While these methods show good sampling performance using a least-squares (LS) signal reconstruction scheme, their complexity is high. On the other hand, there are faster sampling methods based on the notion of local operators [8], [9], but these methods are ad-hoc in nature, and there are no global notions of optimality associated with the selected sampling sets.

In practice, real-world graph signals, such as temperature in sensor network [10], category labels in graph-based classifier [11] and ratings in recommendation system [12], are typically not strictly bandlimited with known bandwidths. Instead, generally “smooth” signals with respect to known graph structures are relatively more common. Different from standard least-squares reconstruction only considering spectrum energy under a given cutoff frequency, we assume a different signal smoothness notion defined using *graph Laplacian regularization* (GLR), and then propose a corresponding sampling scheme with roughly linear complexity. Further, our scheme can be proven to minimize the global worst case mean squared error (MSE); thus our sampling scheme can gracefully scale to very large graphs and has excellent signal reconstruction performance. Specifically, we first assume a biased GLR based scheme that solves a system of linear equations for reconstruction given collected samples [13]. We then choose samples to minimize the condition number of the coefficient matrix; *i.e.*, maximize the smallest eigenvalue λ_{\min} . Circumventing computation of extreme eigenvalues, we maximize instead the lower bound of λ_{\min} , designated by the smallest left-end of all Gershgorin discs of the matrix [14].

To achieve this efficiently, we first convert the optimization to a dual problem, where we minimize the number of samples needed to align all Gershgorin disc left-ends at a pre-selected lower-bound target T . Algebraically, the dual problem amounts to two disc operations: i) shifting of disc centers due to sampling, and ii) scaling of disc radii due to a similarity transformation of the matrix. We further reinterpret the dual as an intuitive *disc coverage* (DC) problem, which bears strong resemblance to the famous *set cover* (SC)

problem [15]. Though DC remains NP-hard (we provide a proof by reduction), the reinterpretation enables us to derive a fast approximation scheme from a known SC error-bounded approximation algorithm [15]. An appropriate target T can be located quickly via simple binary search. Extensive simulation experiments show that our disc-based sampling algorithm runs substantially faster than existing eigen-decomposition-based schemes (hundreds to thousands times faster) for medium-sized graphs and outperforms other eigen-decomposition-free schemes in reconstruction error for large-sized graphs.

The outline of the paper is as follows. We first overview related works in Section II. We introduce notations and a biased reconstruction method in Section III. In Section IV, we first propose our sampling formulation to minimize the upper bound of MSE and then derive its dual problem based on Gershgorin circle theorem. We develop a fast sampling set selection algorithm to tackle our sampling formulation with roughly linear complexity in Section V. Experimental results are presented in Section VI. We conclude in Section VII.

II. RELATED WORKS

The problem of graph sampling can be defined from different contexts. From an *aggregation sampling* context, the authors in [1], [2] proposed to aggregate the sampled signal at one node (or a subset) but with the graph signal shifted successively by a known shift operator. From a *local weighted sampling* context, [3] proposed to collect samples via a weighted sum of signals from different local sets. In contrast, we focus on graph sampling in the context of set selection in this paper: how to select a subset of nodes to collect samples to reconstruct a smooth signal in high fidelity. We next focus on detailed reviews of related works under the umbrella of set selection-based graph sampling.

A. Deterministic Graph Sampling Set Selection

Sampling theory for noise-free graph signals was first studied in [5], where a *uniqueness set* is defined for perfect recovery of *bandlimited* graph signals. Starting from this definition, [16]–[18] proved a sufficient and necessary condition for a uniqueness set, and proposed a lightweight sampling set selection strategy based on the notion of graph spectral proxies.

In the presence of noise, assuming an unbiased *least-squares* (LS) signal reconstruction with independent and identically distributed (i.i.d) noise model, the minimum mean square error (MMSE) function for sampling is actually the known A-optimal criterion [19]–[21]. In response, the authors in [22], [23] proposed to directly optimize the A-optimal function with greedy scheme. Alternatively, [24] used the E-optimality criterion for graph sampling, which can be interpreted as minimizing the worst case of MSE. However, the criteria in [22]–[24] all required extreme eigen-pair computation of the graph variation operator to choose samples, which is not scalable for large graphs. Though [20] avoided eigen-pair computation by developing a sampling method based on truncated Neumann series, it must compute many matrix-series multiplications for good approximation, which is expensive.

To tackle the complexity issue in large scale graphs, very recently, [8], [9] proposed an eigen-decomposition-free graph sampling method by successively maximizing the coverage of the localization operators, which was implemented via Chebyshev polynomial approximation [25]. Nonetheless, this sampling idea utilizes local message heuristically and has no global performance guarantee. Orthogonally, our early work [26] proposed a fast graph sampling algorithm via Gershgorin disc alignment without any eigen-pair computation, but its performance was sub-optimal due to the node sampling strategy simply based on breadth first search (BFS). It is known that random sampling [27], [28] can lead to very low computational complexity, but it typically requires more samples for the same signal reconstruction quality compared to its deterministic counterparts.

In this paper, we propose a roughly linear-time deterministic graph sampling method without any eigen-pair computation, while its reconstruction performance remains superior. Unlike previous fast sampling works [9], [16], [26], our method can be proven to optimize the robustness of the linear reconstruction system resulting from GLR and reduce the upper-bound of the global reconstruction MSE, while each sampling step is performed locally for speed.

B. Sampling of GMRF Graph Signals

Besides the bandlimitedness assumption, graph signal (label information) in classical semi-supervised learning is interpreted as a Gaussian Markov random field (GMRF), where the signal is smooth with respect to the graph Laplacian operator [29]. With this property, active semi-supervised learning can be interpreted as graph sampling on a nearest neighbor graph constructed by node features [11]. An adaptive graph sampling method was proposed [30] to collect samples based on both successively modified graph and selected samples. To improve efficiency, authors in [31] proposed a hybrid adaptive and non-adaptive sampling method to combine the graph sampling and active learning methods. Under the GMRF model, [32] showed that the MSE function for graph sampling is the trace of the posterior covariance matrix, based on which they developed a greedy sampling method. However, this method is very inefficient in large graphs because of the matrix inverse computation in each greedy step. In this paper, we derive a fast sampling strategy for GMRF graph signals with smoothness prior, but it is also applicable for bandlimited graph signals.

III. PRELIMINARY

A. Background and Notations

A graph \mathcal{G} is defined by a triplet $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$, where \mathcal{V} and \mathcal{E} represent sets of N nodes and M edges in the graph, respectively. Associated with each edge $(i, j) \in \mathcal{E}$ is a weight w_{ij} , which reflects the correlation or similarity between two connected nodes i and j . In this paper, we assume a connected, undirected graph; i.e., $w_{ij} = w_{ji}, \forall i, j \in \mathcal{V}$. \mathbf{W} is an *adjacency* matrix with w_{ij} as the (i, j) -th entry of the matrix. We assume a non-negative weight matrix with elements $w_{ij} > 0$ if $(i, j) \in \mathcal{E}$, and $w_{ij} = 0$ otherwise.

Given \mathbf{W} , the *combinatorial graph Laplacian* matrix \mathbf{L} is computed as [33], [34]:

$$\mathbf{L} \triangleq \mathbf{D} - \mathbf{W} \quad (1)$$

where $\mathbf{D} = \text{diag}(\mathbf{W}\mathbf{1})$ is a diagonal *degree* matrix. $\mathbf{1}$ is a vector of all 1's and $\text{diag}(\cdot)$ is an operator that returns a diagonal matrix with the elements of an input vector on the main diagonal. As \mathbf{L} is real and symmetric, it can be eigen-decomposed via the *Spectral Theorem* [35],

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top \quad (2)$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing real eigenvalues θ_k along its diagonal, and \mathbf{U} is an orthonormal matrix composed of real eigenvectors \mathbf{u}_k 's as columns. Since $w_{ij} \geq 0$ for $\forall (i, j) \in \mathcal{E}$, $\mathbf{x}^\top \mathbf{L} \mathbf{x} = \sum_i \sum_j w_{ij} (x_i - x_j)^2 \geq 0$ for $\forall \mathbf{x} \in \mathbb{R}^N$. Thus, \mathbf{L} is a *positive semi-definite* (PSD) matrix. The non-negative eigenvalues θ_k 's are interpreted as *graph frequencies*. By sorting θ_k 's in a non-decreasing order, i.e., $0 = \theta_1 \leq \theta_2 \leq \dots \leq \theta_N$, θ_k 's are ordered from lowest to highest frequencies. θ_1 represents the lowest DC graph frequency and θ_N represents the highest AC graph frequency. The corresponding eigenvectors in \mathbf{U} are interpreted as Fourier-like graph frequency basis; together they form the *graph spectrum* for a graph \mathcal{G} .

Given a graph frequency basis \mathbf{U} , a graph signal \mathbf{x} can be transformed to its graph frequency domain representation $\boldsymbol{\alpha}$ via *graph Fourier transform* (GFT) [33], [34]:

$$\boldsymbol{\alpha} = \mathbf{U}^\top \mathbf{x} = [\mathbf{u}_1, \dots, \mathbf{u}_N]^\top \mathbf{x} \quad (3)$$

A graph signal \mathbf{x} is called *ω -bandlimited* if $\alpha_i = 0$, for $\forall i \in \{i \mid \theta_i > \omega\}$.

A graph signal $\mathbf{x} \in \mathbb{R}^N$ is defined on the nodes of a graph, i.e., one scalar value x_i is assigned to the i -th node. Given a graph signal \mathbf{x} and subset $\mathcal{S} \subset \{1, \dots, N\}$ with cardinality $|\mathcal{S}| = K < N$, we define the sampling process as [5], [7], [9], [16], [20], [22]–[24], [27], [36]:

$$\mathbf{y} = \mathbf{x}_{\mathcal{S}} + \mathbf{n} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (4)$$

where $\mathbf{y} \in \mathbb{R}^K$ is a sampled observation of length K . $\mathbf{x}_{\mathcal{S}} \in \mathbb{R}^K$ denotes a sub-vector of \mathbf{x} consisting of elements indexed by \mathcal{S} and \mathbf{n} is an additive noise term. The *sampling matrix* $\mathbf{H} \in \{0, 1\}^{K \times N}$ is defined as [20]:

$$\mathbf{H}_{ij} = \begin{cases} 1, & j = S_i; \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

where S_i denotes the i -th element of set \mathcal{S} . The complement set of \mathcal{S} on \mathcal{V} is denoted by $\mathcal{S}^c = \mathcal{V} \setminus \mathcal{S}$. We define $L_2(\mathcal{S})$ as the space of all graph signals that are zero except on the subset of nodes indexed by \mathcal{S} [16]:

$$L_2(\mathcal{S}) = \{\mathbf{x} \in \mathbb{R}^N \mid \mathbf{x}_{\mathcal{S}^c} = \mathbf{0}\} \quad (6)$$

B. Signal Reconstruction with Graph Laplacian Regularizer

Instead of assuming an unbiased *least-squares* (LS) signal reconstruction scheme from sparse samples [20], [22]–[24], we adopt a popular biased graph signal reconstruction scheme based on *graph Laplacian regularization* (GLR) [13], which

has demonstrated its effectiveness in numerous practical applications, such as image processing [13], [37]–[41], computer graphics [42] and semi-supervised learning [43], [44]. Biased signal reconstruction using GLR leads to an unconstrained l_2 - l_2 -norm minimization, whose solution can be obtained by solving a system of linear equations, efficiently computed using state-of-the-art numerical linear algebra algorithms such as *conjugate gradient* (CG) [45]. For large noise variance, it is shown that the biased estimation has smaller reconstructed errors on average than the unbiased counterpart [20], [46].

Specifically, given a noise-corrupted sampled observation $\mathbf{y} \in \mathbb{R}^K$ on a graph \mathcal{G} , one can formulate an optimization for the target signal $\hat{\mathbf{x}} \in \mathbb{R}^N$ using GLR as follows:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \mu \mathbf{x}^\top \mathbf{L} \mathbf{x} \quad (7)$$

where \mathbf{H} is a sampling matrix defined in (5). $\mu > 0$ is a tradeoff parameter to balance GLR against the l_2 -norm data fidelity term. The law of selecting an optimal μ with respect to both noise variance and graph spectrum is studied in [46]. The focus of this paper is on graph sampling rather than graph signal reconstruction, and interested readers can refer to [46] for a discussion on the appropriate selection of μ .

Since objective (7) is quadratic, the optimal solution $\hat{\mathbf{x}}$ can be obtained by solving a system of linear equations:

$$(\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L}) \hat{\mathbf{x}} = \mathbf{H}^\top \mathbf{y}. \quad (8)$$

For \mathbf{L} constructed on a connected graph \mathcal{G} , DC graph frequency $\mathbf{u}_1 = \mathbf{1}/\sqrt{N}$ leads to $\mathbf{u}_1^\top \mathbf{L} \mathbf{u}_1 = 0$ and $\|\mathbf{H}\mathbf{u}_1\|_2 > 0$ for $\forall \mathcal{S}$ with $K \geq 1$; for any nonzero graph signal $\mathbf{x} \in \mathbb{R}^N \neq \mathbf{u}_1$, $\mathbf{x}^\top \mathbf{L} \mathbf{x} > 0$ and $\|\mathbf{H}\mathbf{x}\|_2 \geq 0$. Therefore, $\forall \mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^\top (\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L}) \mathbf{x} = \|\mathbf{H}\mathbf{x}\|_2^2 + \mu \mathbf{x}^\top \mathbf{L} \mathbf{x} > 0$. Hence, coefficient matrix $\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L}$ must be *positive definite* (PD), and (8) has a unique solution $\hat{\mathbf{x}} = (\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L})^{-1} \mathbf{H}^\top \mathbf{y}$.

IV. PROBLEM FORMULATION

A. Reconstruction Stability and Sampling Set

Although $\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L}$ is a PD matrix, (8) can potentially be poorly conditioned, because the minimum eigenvalue λ_{\min} of $\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L}$ can be very close to zero. For instance, if \mathbf{H} is chosen such that there exists a low frequency graph signal $\mathbf{x} \in L_2(\mathcal{S}^c)$ and its bandlimit ω is very small, then

$$\begin{aligned} \lambda_{\min} &\leq \frac{\mathbf{x}^\top (\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L}) \mathbf{x}}{\|\mathbf{x}\|_2^2} \stackrel{(1)}{=} \mu \frac{\mathbf{x}^\top \mathbf{L} \mathbf{x}}{\|\mathbf{x}\|_2^2} \\ &= \mu \frac{\alpha_1^2 \theta_1 + \dots + \alpha_k^2 \theta_k}{\|\boldsymbol{\alpha}\|_2^2} \leq \mu \omega. \end{aligned} \quad (9)$$

where $\stackrel{(1)}{=}$ holds because of the definition in (6). For small μ , λ_{\min} is then upper-bounded by a small number.

This means that condition number $\rho = \lambda_{\max}/\lambda_{\min}$ for coefficient matrix $\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L}$ can be very large, resulting in instability when $\hat{\mathbf{x}}$ is numerically computed in (8), and more importantly, $\hat{\mathbf{x}}$ is sensitive to the noise in the observation \mathbf{y} . Our goal then is to maximize λ_{\min} by appropriately choosing sampling matrix \mathbf{H} , so that ρ is minimized¹.

¹Largest eigenvalue λ_{\max} is upper-bounded by $1 + 2\mu d_{\max} \ll \infty$, where d_{\max} is the maximum node degree of the graph \mathcal{G} . Thus to minimize condition number $\rho = \lambda_{\max}/\lambda_{\min}$, it is sufficient to maximize λ_{\min} .

Denote a diagonal matrix by $\mathbf{A} = \mathbf{H}^\top \mathbf{H} = \text{diag}(\mathbf{a}) = \text{diag}(a_1, \dots, a_N) \in \mathbb{R}^{N \times N}$, where *sampling vector* \mathbf{a} satisfies

$$a_i = \begin{cases} 1, & i \in \mathcal{S}, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

In other words, \mathbf{a} is 0-1 bit array encoded sampling set \mathcal{S} . Denote by $\mathbf{B} = \mathbf{A} + \mu \mathbf{L}$.

Given matrix \mathbf{L} , parameter μ and a sampling budget K where $0 < K < N$, our objective is to find a sampling set \mathcal{S} with $|\mathcal{S}| \leq K$ (or, equivalently, a sampling vector \mathbf{a} with $\sum_{i=1}^N a_i \leq K$), to maximize λ_{\min} of \mathbf{B} , i.e.,

$$\begin{aligned} \max_{\mathbf{a}} \quad & \lambda_{\min}(\mathbf{B}) \\ \text{s.t.} \quad & \mathbf{B} = \mathbf{A} + \mu \mathbf{L}, \end{aligned} \quad (11)$$

$$\mathbf{A} = \text{diag}(\mathbf{a}), \quad a_i \in \{0, 1\}, \quad \sum_{i=1}^N a_i \leq K.$$

We next show that maximizing $\lambda_{\min}(\mathbf{B})$ in (11) is equivalent to minimizing the worst case of MMSE.

Proposition 1. *Maximizing $\lambda_{\min}(\mathbf{B})$ in (11) minimizes the upper bound of MSE between original signal \mathbf{x} and reconstructed signal $\hat{\mathbf{x}}$.*

Proof. Reconstructed signal $\hat{\mathbf{x}}$ from a noisy sampled signal \mathbf{y} using (8) is

$$\begin{aligned} \hat{\mathbf{x}} &= (\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L})^{-1} \mathbf{H}^\top \mathbf{y} \\ &= (\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L})^{-1} \mathbf{H}^\top (\mathbf{H} \mathbf{x} + \mathbf{n}) \\ &= (\mathbf{H}^\top \mathbf{H} + \mu \mathbf{L})^{-1} \mathbf{H}^\top \mathbf{H} (\mathbf{x} + \tilde{\mathbf{n}}) \end{aligned} \quad (12)$$

where $\tilde{\mathbf{n}} \in \mathbb{R}^N$ is an N -dimensional additive noise. Denote by $\mathbf{B} = \mathbf{A} + \mu \mathbf{L} = \mathbf{H}^\top \mathbf{H} + \mu \mathbf{L}$. We have

$$\hat{\mathbf{x}} = \mathbf{B}^{-1}(\mathbf{B} - \mu \mathbf{L})(\mathbf{x} + \tilde{\mathbf{n}}) = (\mathbf{I} - \mu \mathbf{B}^{-1} \mathbf{L})(\mathbf{x} + \tilde{\mathbf{n}}) \quad (13)$$

To minimize MSE between original signal \mathbf{x} and reconstructed signal $\hat{\mathbf{x}}$ is equivalent to minimizing $\|\hat{\mathbf{x}} - \mathbf{x}\|_2$.

$$\begin{aligned} \|\hat{\mathbf{x}} - \mathbf{x}\|_2 &= \|(\mathbf{I} - \mu \mathbf{B}^{-1} \mathbf{L})(\mathbf{x} + \tilde{\mathbf{n}}) - \mathbf{x}\|_2 \\ &= \|-\mu \mathbf{B}^{-1} \mathbf{L}(\mathbf{x} + \tilde{\mathbf{n}}) + \tilde{\mathbf{n}}\|_2 \\ &\leq \|-\mu \mathbf{B}^{-1} \mathbf{L}(\mathbf{x} + \tilde{\mathbf{n}})\|_2 + \|\tilde{\mathbf{n}}\|_2 \\ &\leq \mu \|\mathbf{B}^{-1}\|_2 \|\mathbf{L}(\mathbf{x} + \tilde{\mathbf{n}})\|_2 + \|\tilde{\mathbf{n}}\|_2 \end{aligned} \quad (14)$$

Since \mathbf{B} is a symmetric PD matrix, $\|\mathbf{B}^{-1}\|_2 = \frac{1}{\lambda_{\min}(\mathbf{B})}$. Thus, maximizing $\lambda_{\min}(\mathbf{B})$ minimizes the upper bound of MSE, given fixed μ , \mathbf{L} and \mathbf{x} . \square

B. Maximizing the lower bound of $\lambda_{\min}(\mathbf{B})$

Solving (11) directly is a challenging combinatorial optimization problem. To begin with, for large graphs, computing $\lambda_{\min}(\mathbf{B})$ given \mathbf{a} using a state-of-the-art numerical linear algebra algorithm like the Lanczos algorithm [47] or the LOBPCG algorithm [48] is still computation-expensive. Moreover, we need to compute $\lambda_{\min}(\mathbf{B})$ for exponential number of candidate sampling vectors \mathbf{a} where $\sum_{i=1}^N a_i = K$.

In order to study $\lambda_{\min}(\mathbf{B})$ *without* performing expensive eigen-pair computation, we design an efficient algorithm based

on *Gershgorin Circle Theorem* (GCT) [35]. We first describe GCT below:

Theorem 1 (Gershgorin Circle Theorem). *Let \mathbf{X} be a complex $n \times n$ matrix with entries x_{ij} . For $i \in \{1, \dots, n\}$, let $R_i = \sum_{j \neq i} |x_{ij}|$ be the sum of the absolute values of the off-diagonal entries in the i -th row. Consider the n Gershgorin discs*

$$\Psi_i(x_{ii}, R_i) = \{z \in \mathbb{C} \mid |z - x_{ii}| \leq R_i\}, \quad i \in \{1, \dots, n\}. \quad (15)$$

where x_{ii} and R_i are the center and radius of disc i , respectively. Each eigenvalue of \mathbf{X} lies within at least one Gershgorin disc $\Psi_i(x_{ii}, R_i)$.

For our specific real-valued coefficient matrix \mathbf{B} , each real eigenvalue λ lies within at least one *Gershgorin disc* $\Psi_i(b_{ii}, R_i)$ with disc center b_{ii} and radius R_i , i.e.,

$$b_{ii} - R_i \leq \lambda \leq b_{ii} + R_i, \quad (16)$$

where $R_i = \sum_{j \neq i} |b_{ij}| = \mu \sum_j w_{ij} = \mu d_i$, and d_i is the degree of node i . The third equality is true since there are no self-loops in \mathcal{G} . Center of disc i is $b_{ii} = \mu d_i + a_i$.

Using GCT, we can compute the lower bound of $\lambda_{\min}(\mathbf{B})$ by examining left-ends of Gershgorin discs. Specifically, the lower bound of $\lambda_{\min}(\mathbf{B})$ is computed as the smallest left-end of all Gershgorin discs, i.e.,

$$\min_i \{b_{ii} - R_i\} = \min_i a_i \leq \lambda_{\min}(\mathbf{B}). \quad (17)$$

For intuition, we examine the two extreme sampling cases. One extreme case is when no nodes are sampled, i.e., $\mathcal{S} = \emptyset$, and both the lower bound and λ_{\min} of $\mathbf{B} = \mu \mathbf{L}$ are 0. The other extreme case is when all nodes are sampled, i.e., $\mathcal{S} = \mathcal{V}$, and both the lower bound and λ_{\min} of $\mathbf{B} = \mathbf{I} + \mu \mathbf{L}$ are 1. Note that in both extreme cases, the smallest Gershgorin disc left-end (lower bound) is also exactly λ_{\min} . The observation inspires us to maximize the lower bound of $\lambda_{\min}(\mathbf{B})$ instead of maximizing $\lambda_{\min}(\mathbf{B})$ directly.

In general, the tightness of the Gershgorin lower bound of λ_{\min} does not hold for any sampling sets. In order to tighten and maximize the lower bound of $\lambda_{\min}(\mathbf{B})$, we introduce two basic operations to manipulate Gershgorin discs:

Disc Shifting: The first operation is *disc shifting* via sampling. In (16), the left-end of the i -th Gershgorin disc Ψ_i in matrix \mathbf{B} is $b_{ii} - R_i = a_i$. When node i is sampled, we have

$$a_i : 0 \rightarrow 1, \quad i \in \mathcal{S} \quad (18)$$

The corresponding left-end of disc Ψ_i shifts from 0 to 1.

Disc Scaling: The second operation is *disc scaling* via similarity transformation. For $0 < K < N$, there is at least one $a_i = 0$, $i \in \mathcal{S}^c$, and the lower bound of $\lambda_{\min}(\mathbf{B})$ in (17) is always 0. We introduce disc scaling to pull left-ends of unsampled discs away from 0, resulting in a tighter lower bound for $\lambda_{\min}(\mathbf{B})$. Specifically, disc scaling is defined as

$$\mathbf{C} = \mathbf{S} \mathbf{B} \mathbf{S}^{-1}, \quad (19)$$

where \mathbf{S} is a chosen invertible square matrix, named *scaling matrix*. Matrix \mathbf{C} is the similarity transformation of the square

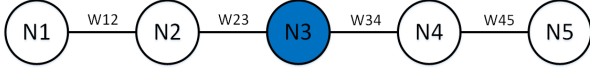


Fig. 1. An illustrative example of a 5-node path graph.

matrix \mathbf{B} and shares the same eigenvalues with \mathbf{B} (but with different eigenvectors) [14]. Thus, we can examine Gershgorin disc left-ends of transformed matrix \mathbf{C} instead of original \mathbf{B} .

To simplify the design of \mathbf{S} , we employ a PD diagonal matrix $\mathbf{S} = \text{diag}(\mathbf{s})$, $s_i > 0$, $\forall i \in \{1, \dots, N\}$, to scale Gershgorin discs of \mathbf{B} , so that left-ends of Gershgorin discs of resulting transformed \mathbf{C} can be moved right. Specifically, the i -th diagonal term s_i of \mathbf{S} in (19) and its corresponding element s_i^{-1} in \mathbf{S}^{-1} are used to scale the radius R_i of Ψ_i and the radii of its neighbors' discs Ψ_j respectively, where $j \in \mathcal{N}_i = \{j \mid w_{ij} > 0\}$. For example, if a scalar $s_i > 1$ is used to *expand* R_i , then its neighbors' discs are *shrunk* with $s_i^{-1} < 1$. Since s_i is always offset by s_i^{-1} on the main diagonal, the center b_{ii} of disc Ψ_i is unchanged.

5-node Path Graph Example: To illustrate the two basic operations (shifting and scaling) on Gershgorin discs, as an example we consider a 5-node path graph and start by sampling node 3, as shown in Fig. 1. Assuming $\mu = 1$, the graph's coefficient matrix \mathbf{B} with (3,3)-th entry updated is shown in Fig. 2a. The left-end of disc Ψ_3 is shifted from 0 to 1, as shown in Fig. 2d. After disc Ψ_3 is shifted, we apply a scalar $s_3 > 1$ on the third row of \mathbf{B} to expand radius and center of disc Ψ_3 in Fig. 2b and 2e. Afterwards, its neighbors' discs Ψ_2 and Ψ_4 are shrunk when applying $s_3^{-1} < 1$ on the third column of \mathbf{B} , as shown in Fig. 2c and 2f. Note that s_3 is offset by s_3^{-1} on the main diagonal of disc Ψ_3 , so its center $b_{33} = 1 + d_3$ is unchanged.

With the above two operations, we can select a sampling set and *shift* left-ends of the corresponding Gershgorin discs from 0 to 1. At the same time, we *scale* radii of all Gershgorin discs to align their left-ends, resulting in a tighter lower bound of $\lambda_{\min}(\mathbf{B})$. Given a graph \mathcal{G} and a sampling budget K , maximizing the lower bound using the two disc operations (encoded in vectors \mathbf{a} and \mathbf{s}) can be formulated as:

$$\begin{aligned} \max_{\mathbf{a}, \mathbf{s}} \quad & \min_{i \in \{1, \dots, N\}} c_{ii} - \sum_{j \neq i} |c_{ij}| \\ \text{s.t.} \quad & \mathbf{C} = \mathbf{S}(\mathbf{A} + \mu \mathbf{L}) \mathbf{S}^{-1} \\ & \mathbf{A} = \text{diag}(\mathbf{a}), \quad a_i \in \{0, 1\}, \quad \sum_{i=1}^N a_i \leq K, \\ & \mathbf{S} = \text{diag}(\mathbf{s}), \quad s_i > 0. \end{aligned} \quad (20)$$

where the objective is to maximize the smallest left-end $\min_{i \in \{1, \dots, N\}} c_{ii} - \sum_{j \neq i} |c_{ij}|$ of all Gershgorin discs of similarity-transformed matrix \mathbf{C} of original $\mathbf{A} + \mu \mathbf{L}$. We call this formulated optimization (20) the *primal problem*.

C. Formulating the Dual Problem

The primal problem (20) is difficult to solve because of the combinatorial nature of a_i 's and the max-min objective.

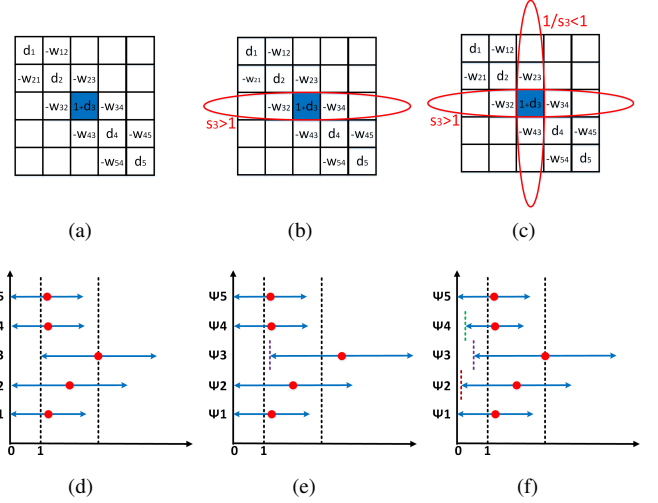


Fig. 2. An illustration of disc operations. (a) & (d) Shifting disc Ψ_3 . (b) & (e) Shifting and expanding disc Ψ_3 with $s_3 > 1$. (c) & (f) Shifting and expanding disc Ψ_3 with $s_3 > 1$ and $s_3^{-1} < 1$. The center of Ψ_3 is unchanged. The radii of disc Ψ_2 and Ψ_4 are shrunk because $s_3^{-1} < 1$. In (a), (b) and (c), we show disc operations on coefficient matrices \mathbf{B} . In (d), (e) and (f), red dots and blue arrows represent disc centers and radii respectively.

Instead, we formulate a *dual problem*² (21) that is essentially a rewriting of (20) with the objective and the constraint on $\sum_{i=1}^N a_i$ swapping roles. Specifically, given a graph \mathcal{G} and a lower bound target T , (21) minimizes the number of sampled nodes subject to a constraint to lower-bound all left-ends:

$$\begin{aligned} \min_{\mathbf{a}, \mathbf{s}} \quad & \sum_{i=1}^N a_i \\ \text{s.t.} \quad & \mathbf{C} = \mathbf{S}(\mathbf{A} + \mu \mathbf{L}) \mathbf{S}^{-1}, \quad c_{ii} - \sum_{j \neq i} |c_{ij}| \geq T, \quad \forall i \\ & \mathbf{A} = \text{diag}(\mathbf{a}), \quad a_i \in \{0, 1\}, \\ & \mathbf{S} = \text{diag}(\mathbf{s}), \quad s_i > 0. \end{aligned} \quad (21)$$

Instead of a max-min criterion, the objective in (21) is now linear in a_i 's.

Optimal solutions to the primal problem (20) and the dual problem (21) are related via the following proposition:

Proposition 2. *If T is chosen such that there exists at least one optimal solution $(\hat{\mathbf{a}}, \hat{\mathbf{s}})$ to the dual problem (21) satisfying $\sum_i \hat{a}_i = K$, then there exists one optimal solution to the dual problem (21) at lower bound target T that is also an optimal solution to the primal problem (20).*

Proof. For feasible solutions (\mathbf{a}, \mathbf{s}) to the dual problem (21) with $\sum_i a_i > K$, they are not feasible to the primal problem (20), and thus can be ignored. For each solution (\mathbf{a}, \mathbf{s}) to the dual problem (21) with $\sum_i a_i < K$ satisfying all constraints except $\min_i c_{ii} - \sum_{j \neq i} |c_{ij}| \geq T$, they must have $\min_i c_{ii} - \sum_{j \neq i} |c_{ij}| < T$, otherwise $(\hat{\mathbf{a}}, \hat{\mathbf{s}})$ would not be the optimal solution with $\sum_i \hat{a}_i = K$. Thus this solution (\mathbf{a}, \mathbf{s}) , while

²Though the dual problem is not exactly the same definition as classical optimization setups in linear programming [49] and convex programming [50], the essence of exchanging the roles of objective function and constraints to obtain a related optimization problem is identical.

feasible in the primal, has a worse objective than $(\hat{\mathbf{a}}, \hat{\mathbf{s}})$ in the primal problem. So the optimal solution to the primal problem (20) must be a solution to the dual problem (21) with $\sum_i a_i = K$ (though not necessarily the one $(\hat{\mathbf{a}}, \hat{\mathbf{s}})$ if multiple optimal solutions exist). \square

Denote by $(\hat{\mathbf{a}}, \hat{\mathbf{s}})$ an optimal solution to the dual problem (21) given lower bound target T , and \hat{K}_T the corresponding objective function value. We can further prove that \hat{K}_T is non-decreasing with respect to T .

Proposition 3. *Let \hat{K}_T be the optimal objective function value to (21) given lower bound T . Then, we have $\hat{K}_{T_1} \geq \hat{K}_{T_2}$, $\forall T_1 \geq T_2$.*

Proof. Since $T_1 \geq T_2$, \hat{K}_{T_1} is a feasible objective function value to (21) given lower bound T_2 . Hence, $\hat{K}_{T_1} \geq \hat{K}_{T_2}$ must be satisfied, since \hat{K}_{T_2} is the optimal (minimum) objective function value to (21) given T_2 . \square

Proposition 2 and Proposition 3 provide us with an important insight:

Instead of solving the primal problem (20) directly, we can solve the dual problem (21) iteratively for different $T \in (0, 1)$, where we employ a binary search to seek the largest \hat{T} such that the objective value of (21) is K exactly.

When this largest \hat{T} is found, we achieve an optimal solution $(\hat{\mathbf{a}}, \hat{\mathbf{s}})$ to (20), and \hat{T} is the corresponding optimal objective value. When solving the dual problem (21), we select the smallest sampling set and scale radii of all Gershgorin discs, so that the discs' left-ends (the lower bound of $\lambda_{\min}(\mathbf{B})$) is no smaller than target T . Hence, we refer to the dual problem (21) also as a *disc alignment* problem.

When left-ends of all discs are aligned to T as closely as possible, not only is the smallest left-end of Gershgorin discs a lower bound for $\lambda_{\min}(\mathbf{B})$, as stated in GCT, but we can also prove that the largest left-end gives an upper bound:

Proposition 4. *Let Γ_i be the Gershgorin disc from the i -th row of matrix $\mathbf{C} = \mathbf{SBS}^{-1}$, where \mathbf{S} is a non-singular diagonal matrix satisfying $\mathbf{S} = \text{diag}(\mathbf{s})$ where $\mathbf{s} > \mathbf{0}$. Denote by ℓ_i the left-end of Γ_i , we have*

$$\min_i \ell_i \leq \lambda_{\min}(\mathbf{B}) \leq \max_i \ell_i \quad (22)$$

In words, while the smallest left-end ℓ_i gives a lower bound for $\lambda_{\min}(\mathbf{B})$, the largest left-end ℓ_i gives an upper bound.

The proof of Proposition 4 is in the Appendix A. Note that $\lambda_{\min}(\mathbf{B})$ can be exactly the lower bound T , if the left-ends of all Gershgorin discs are aligned at T satisfying $\min_i \ell_i = \max_i \ell_i$.

V. SAMPLING SET SELECTION ALGORITHM DEVELOPMENT

Although, unlike the primal (20), the dual (21) has a linear objective, it is nonetheless difficult to solve optimally. Instead of manipulating all discs simultaneously, we first define a coverage subset Ω_i for each node i —the maximum subset of discs whose left-ends can be aligned at or beyond T by

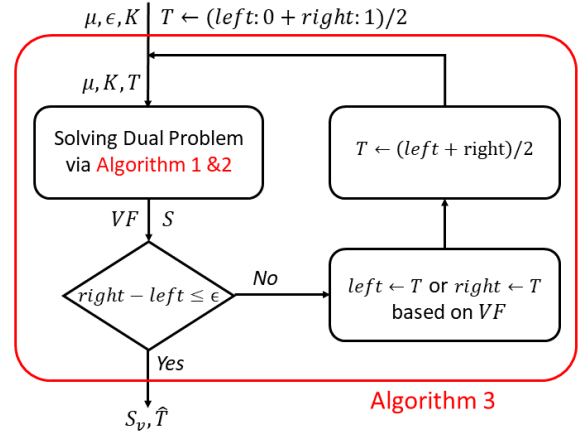


Fig. 3. The workflow of the complete sampling algorithm (Algorithm 3). Algorithm 1 and 2 solve the dual (21). Binary search seeks the largest \hat{T} such that the objective of the dual (21) is K exactly.

sampling only node i . We introduce an approximation algorithm to pre-compute each coverage subset Ω_i (Algorithm 1) in Section V-A.

Pre-computed coverage subsets lead to a coarse but intuitive reinterpretation of the dual (21) as the disc coverage problem, which we prove to be NP-hard via reduction from the famous *set cover* (SC) problem in Section V-B. The reinterpretation enables us to derive a fast algorithm for the dual (21) based on a known SC approximation algorithm (Algorithm 2) in the combinatorial optimization literature in Section V-C.

We further employ a binary search to seek the largest \hat{T} such that the objective of the dual (21) is K exactly. The complete sampling algorithm (Algorithm 3) is proposed in Section V-D, which is summarized in Fig. 3. Finally, we discuss the computational complexities of different graph sampling approaches in Section V-E.

A. Coverage Subset of Each Sampled Node

Before we approximately solve (21)—align all Gershgorin discs' left-ends at or beyond target $T \in (0, 1)$ using as few sampled nodes as possible—we first define a *coverage subset* Ω_i for each node i .

Definition 1 (Coverage subset). *A coverage subset Ω_i for each node i is the optimal subset \mathcal{X} satisfying*

$$\begin{aligned} \max_{\mathcal{X}} \quad & |\mathcal{X}| \\ \text{s.t.} \quad & \mathbf{C} = \mathbf{S}(\mathbf{A} + \mu\mathbf{L})\mathbf{S}^{-1} \\ & \mathbf{A} = \text{diag}(\mathbf{a}), \quad a_i = 1, \quad a_{j \neq i} = 0 \\ & \mathbf{S} = \text{diag}(\mathbf{s}), \quad s_j > 0 \\ & \mathcal{X} = \{j \mid c_{jj} - \sum_{k \neq j} |c_{jk}| \geq T\}, \quad \forall j, k \in \{1, \dots, N\} \end{aligned} \quad (23)$$

In words, a coverage subset Ω_i is the maximum subset of discs whose left-ends can be aligned at or beyond T by sampling only node i .

We propose an approximation algorithm to estimate each coverage subset Ω_i . As described in Section IV-B, disc scaling

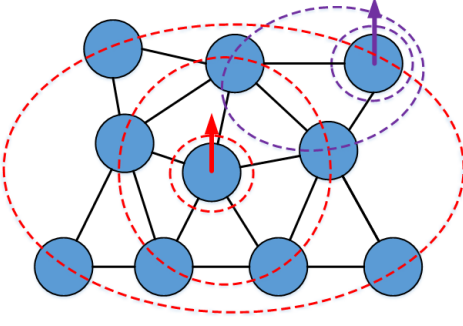


Fig. 4. Coverage subset estimation. Starting from each sampled node i (marked by vertical arrows), we progressively employ disc scaling operations to align left-ends of discs at T with increasing hops from node i . The algorithm stops when farther nodes cannot be aligned.

operation can balance left-ends of discs in the local neighbourhood of a sampled node. Starting from the sampled node i , we progressively employ disc scaling operations to align left-ends of discs at T with increasing hops from the node i , as illustrated in Fig. 4. The algorithm stops when farther nodes cannot be aligned.

Specifically, we first sample node i (thus moving the corresponding disc Ψ_i 's center b_{ii} from μd_i to $1 + \mu d_i$, and Ψ_i 's left-end a_i from 0 to 1). We next apply scalar s_i to expand Ψ_i 's radius R_i and align its left-end at exactly T . Scalar s_i must hence satisfy

$$a_i + \mu \left(d_i - s_i \cdot \sum_{j \in \mathcal{N}_i} \frac{w_{ij}}{s_j} \right) = T, \quad (24)$$

where initially $s_j = 1, \forall j$. \mathcal{N}_i denotes the neighbors of node i satisfying $w_{ij} > 0$. Solving for s_i in (24), we obtain

$$s_i = \frac{a_i + \mu d_i - T}{\mu \sum_{j \in \mathcal{N}_i} \frac{w_{ij}}{s_j}}. \quad (25)$$

Computing s_i for sample node i using (25) implies that $s_i > 1$ (i.e., expansion of Ψ_i 's radius), which means that node i 's neighbors' discs Ψ_j 's radii will shrink due to s_i^{-1} . Specifically, disc left-end $b_{jj} - R_j$ of an unsampled neighbor j 's disc (thus $a_j = 0$) is now:

$$b_{jj} - R_j = a_j + \mu \left(d_j - s_j \cdot \sum_{k \in \mathcal{N}_j \setminus \{i\}} \frac{w_{jk}}{s_k} - s_j \cdot \frac{w_{ji}}{s_i} \right). \quad (26)$$

If a neighboring disc Ψ_j 's left-end remains smaller than T , then nothing further needs to be done. However, if Ψ_j 's left-end is larger than T , then j belongs to Ω_i . We then subsequently expand its radius to align its left-end at T using (25). This shrinks the disc radii of node j 's neighbors and so on; The sequence of nodes explored starting from sample i is determined using *Breadth First Search* (BFS) [15]. Expansion factor s_j decreases with hops away from node i . Since we always expand a current disc ($s_j \geq 1$) leading to shrinking of neighboring discs ($s_j^{-1} \leq 1$) in each step, the left-end of each already scaled node remains larger than or equal to T .

Note that $1 \leq |\Omega_i| \leq N$ for any value of $T \in (0, 1)$. When $T \rightarrow 0$, $|\Omega_i| \rightarrow N$, and we need to visit all N nodes

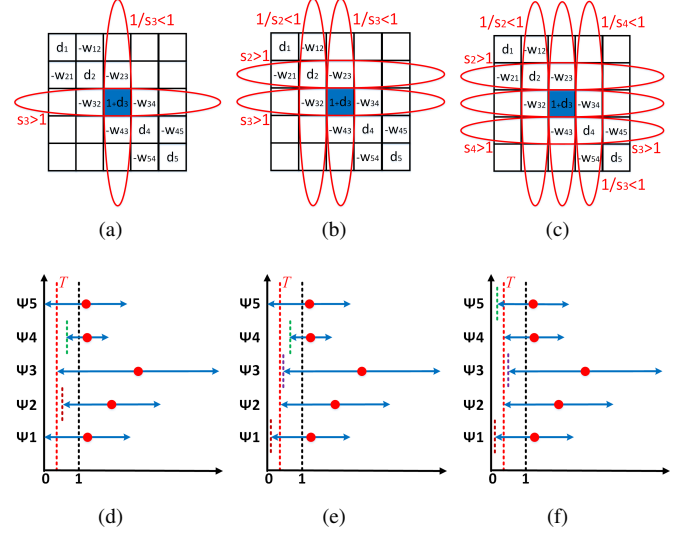


Fig. 5. An illustration of estimating Ω_3 on a 5-node path graph. (a) & (d) aligning left-end of disc Ψ_3 to T . (b) & (e) aligning left-end of disc Ψ_2 to T . (c) & (f) aligning left-end of disc Ψ_4 to T . In (a), (b) and (c), we show disc operations on coefficient matrices \mathbf{B} . In (d), (e) and (f), red dots and blue arrows represent disc centers and radii respectively.

using BFS. The worst-case time complexity for computing a single Ω_i is $\mathcal{O}(N + M)$, where $M = |\mathcal{E}|$. Typically, sampling budget K is much larger than 1, and it is not necessary to compute Ω_i corresponding to a very small T . To enable faster computation, we set a parameter p to estimate Ω_i only within a p -hop neighborhood from node i . Thus, the time complexity of estimating Ω_i within p hops is $\mathcal{O}(P + Q)$, where P is the number of nodes within p hops from node i , and Q is the number of edges within p hops from node i . Assuming a maximum degree per node d_{\max} , we have $P \leq Q \leq d_{\max}^p$. d_{\max}^p is achieved only if graph \mathcal{G} is a tree. Since p is typically set much smaller than the number of hops required to traverse the entire graph with BFS, $P \ll N$ and $Q \ll M$. The procedure of estimating coverage subset Ω_i is sketched in Algorithm 1.

Illustrative Example on 5-node Path Graph: We use an example to illustrate the operations of the above algorithm, shown in Fig. 5. We assume the same 5-node path graph in Fig. 1 and choose to sample node 3. Assuming $\mu = 1$, the graph's coefficient matrix \mathbf{B} with entry $(3, 3)$ updated is shown in Fig. 2a. The left-end of node 3's Gershgorin disc shifts from 0 to 1, as shown in Fig. 2d.

We next perform disc scaling to align the left-end of disc Ψ_3 to T . As shown in Fig. 5a, scalar s_3 is applied to the third row of \mathbf{B} , and thus the radius of disc Ψ_3 is expanded by s_3 , where $s_3 > 1$. Simultaneously, scalar s_3^{-1} is applied to the third column, and thus the radii of discs Ψ_2 and Ψ_4 are shrunk due to the scaling of w_{23} and w_{43} respectively by s_3^{-1} . Note that entry $(3, 3)$ of \mathbf{B} (and Ψ_3 's disc center) is unchanged, since scalar s_3 is offset by s_3^{-1} . We see that due to the expansion of disc Ψ_3 radius, the disc left-ends of its neighboring nodes (nodes 2 and 4) move beyond threshold T , as shown in Fig. 5d.

We next apply scalar s_2 to disc Ψ_2 to expand its radius by

Algorithm 1 Estimating Coverage Subset Ω_i

Input: Graph \mathcal{G} , lower bound T , node i , hop constraint p and μ .

```

1: Initialize  $\mathbf{w} = \mathbf{W}\mathbf{1}$ ,  $\mathbf{s} = \mathbf{1}$  and  $\mathbf{a} = [0, \dots, 0, a_i = 1, 0, \dots, 0]$ .
2: Initialize  $\mathbf{h} = \mathbf{0}$  for hop number.
3: Initialize  $\Omega_i = \emptyset$ .
4: Initialize  $\mathcal{Q} = \emptyset$  for enqueued nodes.
5: Initialize an empty queue.
6: Enqueue(queue, i) and  $\mathcal{Q} \leftarrow \mathcal{Q} \cup \{i\}$ .
7: while queue is not empty do
8:    $k \leftarrow \text{Dequeue}(\text{queue})$ .
9:   Update  $s_k$  using (25).
10:  if  $s_k \geq 1$  and  $h_k \leq p$  do
11:     $\Omega_i \leftarrow \Omega_i \cup \{k\}$ .
12:    for  $t$  in  $k$ 's neighbours  $\mathcal{N}_k$  do
13:      if  $t \notin \mathcal{Q}$  do
14:        Enqueue(queue, t) and  $\mathcal{Q} \leftarrow \mathcal{Q} \cup \{t\}$ .
15:         $h_t \leftarrow h_k + 1$ .
16:      endif
17:    endfor
18:  endif
19: endwhile
Output: Subset  $\Omega_i$ .

```

s_2 , where $s_3 > s_2 > 1$, and the radii of discs Ψ_1 and Ψ_3 are shrunk due to the scaling of w_{12} and w_{32} by s_2^{-1} , as shown in Fig. 5b. s_2 must be smaller than s_3 for the left-end of Ψ_2 not to move past 0. The discs are shown in Fig. 5e. Scaling Ψ_2 does not move the left-end of disc Ψ_1 beyond threshold T . Thus, we stop aligning the further-away disc Ψ_1 .

Subsequently, similar disc operations can be performed on Ψ_4 as shown in Fig. 5c and 5f. Scaling Ψ_4 does not move the left-end of disc Ψ_5 beyond T . Thus, we stop aligning the further-away disc Ψ_5 . Finally, the left-ends of discs Ψ_2 , Ψ_3 and Ψ_4 are moved beyond threshold T , i.e., $\Omega_3 = \{2, 3, 4\}$.

Analogy of Throwing Pebbles in a Pond: To impart intuition, we consider the following real-world analogy of computing subset Ω_i for sample i . Suppose a planar graph \mathcal{G} models the water surface of a pond, where nodes denote water surface locations. Sampling a node i is analogous to throwing a pebble at a water surface location i in the pond. The thrown pebble causes a ripple to adjacent locations $j \in \mathcal{N}_i$, where the magnitude of the ripple at location j is smaller than at i . Analogously, when aligning left-ends of Gershgorin discs at T , scalar s_j of neighbor $j \in \mathcal{N}_i$ is necessarily smaller than s_i . The ripple in the pond dissipates away from the thrown location i . Analogously, scalars s_j becomes smaller further away in hop count from sample i . The ripple stops increasing in size when the dissipating fringes become unnoticeable. Analogously, subset Ω_i stops growing when left-ends of neighboring discs fail to move beyond target T .

B. NP-Hardness of Disc Coverage Problem

After pre-computing coverage subsets, we can reinterpret the dual (21) more coarsely but in a more intuitive manner: given pre-computed coverage subsets $\Omega_i, \forall i \in \{1, \dots, N\}$, how to select minimum-cost sampling vector \mathbf{a} so that every node j belongs to at least one sample's coverage subset Ω_i ,

i.e.,

$$\begin{aligned} \min_{\mathbf{a}} \quad & \sum_{i=1}^N a_i \\ \text{s.t.} \quad & j \in \bigcup_{i \mid a_i=1} \Omega_i, \quad \forall j \in \{1, \dots, N\} \\ & a_i \in \{0, 1\} \end{aligned} \quad (27)$$

We term this the *disc coverage* (DC) problem. DC (27) is related to the dual (21) in the sense that: i) they have the same optimization objective, and ii) every feasible solution \mathbf{a} to (27) is also a feasible solution to (21), where the scalars \mathbf{s} corresponding to \mathbf{a} are the ones calculated when coverage subsets Ω_i 's are pre-computed. Thus solving DC provides an approximate feasible solution to the dual.

Unfortunately, even this DC problem (27) is NP-hard. A simple NP-hardness proof via a reduction from the famous NP-hard set cover (SC) problem [15] is provided in Appendix B. However, the resemblance of the DC problem to SC enables us to derive an efficient approximation algorithm to the dual (21), which we describe next.

C. Disc Alignment via Greedy Set Cover

Given nodes $\mathcal{U} = \{1, 2, \dots, N\}$ and a collection of pre-computed N coverage subsets $\Omega = \{\Omega_1, \dots, \Omega_N\}$, we derive a fast algorithm to DC (27) based on an error-bounded approximation algorithm to SC. We first describe SC formally.

Definition 2 (Set cover problem). *An instance $(\mathcal{X}, \mathcal{F})$ of the set cover problem consists of a finite element set \mathcal{X} and a family \mathcal{F} of subsets of \mathcal{X} , such that every element $x \in \mathcal{X}$ belongs to at least one subset $S \in \mathcal{F}$:*

$$x \in \bigcup_{S \in \mathcal{F}} S, \quad \forall x \in \mathcal{X} \quad (28)$$

We say that a subset $S \in \mathcal{F}$ covers its elements. The problem is to find a minimum-size subset collection $\mathcal{C} \subseteq \mathcal{F}$ whose members cover all of \mathcal{X} :

$$x \in \bigcup_{S \in \mathcal{C}} S, \quad \forall x \in \mathcal{X} \quad (29)$$

Given the resemblance between DC and SC, we borrow a known greedy approximation algorithm for SC with error-bounded performance [15]: at each iteration, we select a yet-to-be-chosen subset S that covers the largest number of uncovered elements in \mathcal{X} . Similarly, for our algorithm for DC, we select sample i and coverage subset Ω_i with the largest uncovered nodes at each iteration. When a sampling vector \mathbf{a} satisfying the condition that each node belongs to at least one sample coverage subset Ω_i is computed, the greedy algorithm can achieve a $H(\max_i |\Omega_i|)$ -approximate solution to DC [15], i.e.,

$$|\mathbf{a}| \leq H(\max_i |\Omega_i|) \cdot |\hat{\mathbf{a}}|, \quad (30)$$

where $H(n) = \sum_{i=1}^n 1/i$ is the harmonic series and $\hat{\mathbf{a}}$ is the optimal solution to DC. $|\mathbf{a}|$ denotes $\sum_i a_i$.

Since we are given a sampling budget K , we terminate the algorithm when the number of selected samples reaches K for practical implementation. The resulting sampling vector \mathbf{a} is valid only if $|\mathbf{a}| \leq K$ and each node belongs to at least one sample coverage subset Ω_i . The procedure of disc alignment via greedy set-covering is outlined in Algorithm 2.

Algorithm 2 Disc Alignment via Greedy Set Cover

Input: Graph \mathcal{G} , lower bound T , sampling budget K and μ .

```

1: Initialize  $\mathcal{U} = \{1, \dots, N\}$ .
2: Initialize  $\Omega = \emptyset$  and  $\mathcal{S} = \emptyset$ .
3: Initialize validity flag  $VF = \text{true}$  and  $n = 1$ .
4: for  $i = 1 \rightarrow N$  do
5:   Estimate  $\Omega_i$  using Algorithm 1.
6:    $\Omega \leftarrow \Omega \cup \{\Omega_i\}$ .
7: endfor
8: while  $|\mathcal{U}| \neq \emptyset$  and  $n \leq K$  do
9:   Select  $\Omega_i \in \Omega$  that maximizes  $|\Omega_i \cap \mathcal{U}|$ .
10:   $\mathcal{U} \leftarrow \mathcal{U} \setminus \{\Omega_i \cap \mathcal{U}\}$ .
11:   $\mathcal{S} \leftarrow \mathcal{S} \cup \{i\}$ .
12:   $n \leftarrow n + 1$ .
13: endwhile
14: if  $|\mathcal{U}| \neq \emptyset$  do
15:    $VF \leftarrow \text{false}$ .
16: endif

```

Output: validity flag VF , sampling set \mathcal{S} .

The time complexity of N subsets preparation within p hops is $\mathcal{O}(PN + QN)$. We represent uncovered elements in \mathcal{U} using 0-1 bit array. The uncovered elements in Ω_i to \mathcal{U} can be computed through at most $\mathcal{O}(P)$ times fast bit operations, since the number of “1” in the bit array of Ω_i is at most P . We compute uncovered elements of N subsets at each subset selection stage. We employ no more than K times subset selections, thus the time complexity of the greedy subset selection algorithm is $\mathcal{O}(NKP)$. The total time complexity of Algorithm 2 comprises $\mathcal{O}(PN + QN)$ node search plus $\mathcal{O}(NKP)$ bit operations.

D. Binary Searching the Maximum Lower Bound

According to Proposition 3, the number of sampled nodes K is non-decreasing with respect to lower bound T in the dual problem (21). Hence, we can perform binary search to find the maximum lower bound $\hat{T} \in (0, 1)$ corresponding to K sampled nodes. We name the proposed algorithm *Binary Search with Gershgorin Disc Alignment* (BS-GDA), as outlined in Algorithm 3.

At each iteration, we estimate validity flag VF and sampling set \mathcal{S} given a certain lower bound T using Algorithm 2. If $VF = \text{false}$, then the lower bound T is set too large, and we update $right$ to reduce T . Otherwise, $VF = \text{true}$, then T may be still too small, leading to $|\mathcal{S}| < K$ or not the maximum T corresponding to $|\mathcal{S}| = K$. We update $left$ to increase T . When $right - left \leq \epsilon$, BS-GDA converges and we find the maximum lower bound \hat{T} with numerical error lower than ϵ .

In order to achieve numerical precision ϵ in BS-GDA, we need to employ $\mathcal{O}(\log_2 \frac{1}{\epsilon})$ times binary search. For instance, to achieve numerical precision $\epsilon = 10^{-3}$, we employ 10 times binary search. The total time complexity of Algorithm 3 comprises $\mathcal{O}(N(P + Q) \log_2 \frac{1}{\epsilon})$ node search plus $\mathcal{O}(NKP \log_2 \frac{1}{\epsilon})$ bit operations.

Algorithm 3 Binary Search with Gershgorin Disc Alignment

Input: Graph \mathcal{G} , sampling budget K , numerical precision ϵ and μ .

```

1: Initialize  $left = 0$ ,  $right = 1$ .
2: Initialize valid sampling set  $\mathcal{S}_v = \emptyset$ .
3: while  $right - left > \epsilon$  do
4:    $T \leftarrow (left + right)/2$ .
5:   Estimate  $VF, \mathcal{S}$  given  $T$  using Algorithm 2.
6:   if  $VF$  is false do
7:      $right \leftarrow T$ 
8:   else
9:      $left \leftarrow T$ 
10:     $\mathcal{S}_v \leftarrow \mathcal{S}$ 
11:   endif
12: endwhile
13:  $\hat{T} \leftarrow left$ .

```

Output: Sampling set \mathcal{S}_v , maximum lower bound \hat{T} .

E. Complexity Analysis

We here summarize the theoretical computational complexities of different graph sampling approaches, and list them in Table I, assuming that the signal bandwidth ω is the K -th eigenvalue θ_K of \mathbf{L} and the graph is sparse. We separate the complexity into two parts: *preparation* (for eigendecomposition or computing initial preparation information) and *sampling* (during greedy sampling step). The competing algorithms are referred to as E-optimal [24], Spectral proxies (SP) [16], MFN [22], MIA [20] and Eigendecomposition-free (Ed-free) [9], respectively. Complexities of methods E-optimal, SP, MFN and MIA are directly borrowed from paper [20], with modified notations. For E-optimal and MFN, T_1 is the convergence steps for computing the first K eigen-pairs of Laplacian \mathbf{L} and R is a constant for combining complexities. In strategy SP, k is the degree of spectral proxies and T_2 is the convergence steps for obtaining the first eigen-pair of \mathbf{L} via LOBPCG [16]. The parameters q and F in MIA are the degree of Chebyshev polynomials for a low-pass filter approximation and the truncated parameter of a Neumann series, respectively. As for Ed-free method, paper [9] claimed that its complexity was dominated by the non-zero entries in matrix $\mathbf{L}^{\bar{q}}$, i.e., J in paper [9], where \bar{q} is degree of Chebyshev polynomial approximation. Similarly, the non-zero entries in $\mathbf{L}^{\bar{q}}$ is at most $Nd_{\max}^{\bar{q}}$ in a degree-bounded graph, so we use \bar{Q} to denote the largest number of edges in \bar{q} -hop neighborhood, which implies $J = \mathcal{O}(N\bar{Q})$. From Table I, we can see that the complexities of Ed-free and our proposed BS-GDA method are both roughly linear in terms of graph size N , which is substantially lower than other eigendecomposition-based sampling methods and implies those two methods are applicable in fairly large graphs. The explicit execution time comparison will be demonstrated in the coming experimental section.

VI. EXPERIMENTS

A. Experimental Setup

We demonstrate the efficacy of our proposed fast graph subset sampling algorithm (BS-GDA) via extensive simulations. Our experimental platform is Windows 10 desktop computer with Intel i5-4670K CPU and 24GB memory. All algorithms are performed on MATLAB R2016b.

TABLE I
THEORETICAL COMPLEXITY COMPARISON OF DIFFERENT GRAPH SAMPLING STRATEGIES.

	E-optimal	SP	MFN	MIA	Ed-free	proposed BS-GDA
Preparation	$\mathcal{O}((MK + RK^3)T_1)$	$\mathcal{O}(kMKT_2)$	$\mathcal{O}((MK + RK^3)T_1)$	$\mathcal{O}(qNM)$	$\mathcal{O}(M\bar{q} + N\bar{Q})$	$\mathcal{O}(N(P + Q)\log_2 \frac{1}{\epsilon})$
Sampling	$\mathcal{O}(NK^4)$	$\mathcal{O}(NK)$	$\mathcal{O}(NK^4)$	$\mathcal{O}(NFK^{3.373})$	$\mathcal{O}(NK\bar{Q})$	$\mathcal{O}(NKP\log_2 \frac{1}{\epsilon})$

We use four types of graphs in GSPBOX [51] for testing:

- 1) Random sensor graph with N nodes.
- 2) Community graph with N nodes and $\lfloor \sqrt{N}/2 \rfloor$ random communities.
- 3) Barabási-Albert graph with N nodes. Each graph is constructed by adding new nodes, each with $m = 1$ edge, that are preferentially attached to existing nodes with high degrees.
- 4) Minnesota road graph with fixed $N = 2642$ nodes.

Random sensor graphs are weighted sparse graphs generated by GSPBOX, where each node connects to its six nearest neighbours. Edge weights are computed using:

$$w_{ij} = \exp \left\{ -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma_x^2} \right\} \quad (31)$$

where \mathbf{x}_i is the 2D coordinate of node i , and σ_x is set automatically by the toolbox. Community graphs, Barabási-Albert graphs and Minnesota road graph are unweighted sparse graphs generated by GSPBOX, where we manually compute weight for each edge. For community graphs and Minnesota road graph, we compute edge weights using (31). We set parameter $\sigma_x = 1$ for community graphs and set parameter $\sigma_x = 0.1$ for Minnesota road graphs. Since Barabási-Albert graphs are scale-free graphs without 2D coordinates, we randomly generate edge weights using a uniform probability distribution with interval $(0, 1)$. Illustrations of four types of graphs are shown in Fig. 6.

For each graph type, we consider two types of graph signals described as follows:

- 1) **GS1**: The true signals are exactly ω -bandlimited, where $\omega = \theta_{\lfloor N/10 \rfloor}$ is the $\lfloor N/10 \rfloor$ -th eigenvalue of \mathbf{L} . For example, $\omega = \theta_{50}$ for $N = 500$. The non-zero GFT coefficients are randomly generated from $\mathcal{N}(0, 10)$. We then add i.i.d. Gaussian noise generated from $\mathcal{N}(0, 0.1^2)$ such that the SNR equals to 20 dB.
- 2) **GS2**: The true signals are generated from multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, (\mathbf{L} + \delta \mathbf{I})^{-1})$, where $\delta = 10^{-5}$. Because the power of the generated graph signals is inconsistent, we normalize the signals using $\mathbf{x}' = \frac{\mathbf{x} - \text{mean}(\mathbf{x})}{\text{std}(\mathbf{x})}$, where $\text{mean}(\mathbf{x}) = \sum_i x_i / N$ and $\text{std}(\mathbf{x}) = \sqrt{\frac{\sum_i (x_i - \text{mean}(\mathbf{x}))^2}{N}}$, respectively. We then add i.i.d. Gaussian noise generated from $\mathcal{N}(0, 0.1^2)$ such that the SNR equals to 20 dB.

We compare the performance of our proposed algorithm with six recent graph sampling set selection methods, of which implementations are available:

- 1) Random graph sampling with non-uniform probability distribution (referred to as Random) [27].

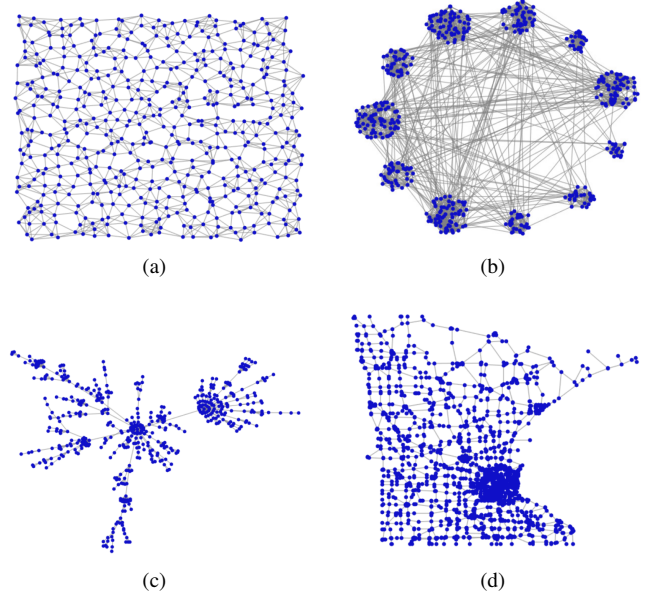


Fig. 6. Illustrations of four types of graphs. (a) Random sensor graph. (b) Community graph. (c) Barabási-Albert graph. (d) Minnesota road graph.

- 2) Deterministic sampling set selection methods, including E-optimal [24], Spectral proxies (SP) [16], MFN [22], MIA [20] and Eigendecomposition-free (Ed-free) [9], respectively.

Our proposed algorithm BS-GDA avoids expensive eigendecomposition and matrix computations, thus we build MEX functions for the iterative BFS and the greedy set covering, leading to a MATLAB+MEX implementation. For all experiments, we set the target numerical precision $\epsilon = 10^{-5}$ for binary search and $\mu = 0.01$. We empirically set hop constraint $p = 12$ to consider no more than 12 hops for each subset Ω_i estimation. It is analogous to the neighborhood considered by Chebyshev polynomial approximation with order 12 in [9].

B. Running Time Comparisons

We compare the running time of our proposed BS-GDA with other competing algorithms on random sensor graphs and community graphs with increasing graph size N . We select sampling budget $K = N/10$ nodes for both graphs. Fig. 7 shows the curves of running time versus graph size N . Specifically, for $N = 3000$, we further compute the speedup factor (SF) [9] of our algorithm with respect to the competing algorithms using

$$SF = \frac{\text{Running time of competing method}}{\text{Running time of our method}}, \quad (32)$$

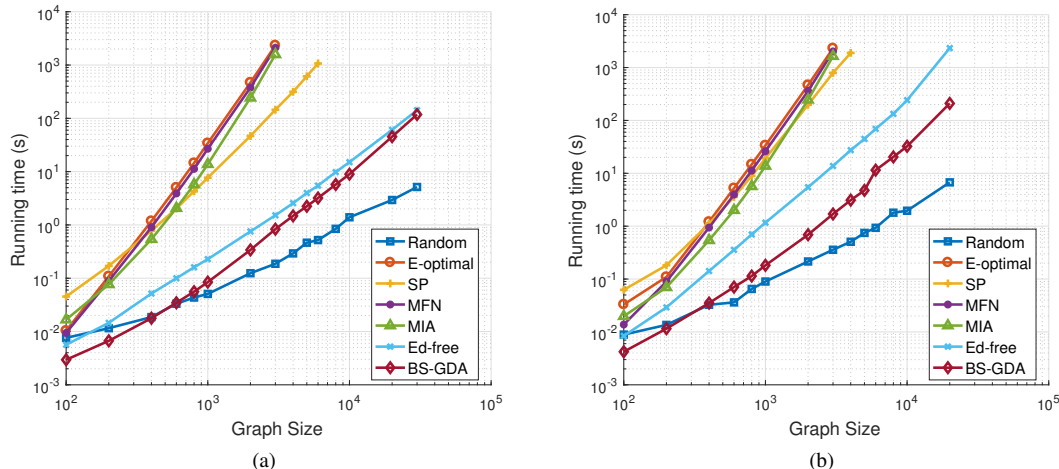


Fig. 7. Running time comparisons on two different graphs. (a) Random sensor graph. (b) Community graph. Both axes are represented on logarithmic scales.

TABLE II
SPEEDUP FACTORS OF OUR ALGORITHM WITH RESPECT TO OTHER
SAMPLING ALGORITHMS FOR $N = 3000$

Sampling Algorithms	Sensor	Community
Random [27]	0.22	0.21
E-optimal [24]	2812.77	1360.76
SP [16]	174.09	466.18
MFN [22]	2532.91	1184.23
MIA [20]	1896.19	964.65
Ed-free [9]	1.82	8.11

as summarized in Table. II.

We observe that BS-GDA is the fastest deterministic sampling set selection algorithm for both graphs. As mentioned, E-optimal, SP and MFN all require computation of extreme eigen-pairs of the graph Laplacian matrix (or submatrix), which is computationally demanding. Though there is no explicit computation of eigen-pairs in MIA, the large matrix series computation also results in large computation cost. Thus, BS-GDA is hundreds to thousands times faster than E-optimal, SP, MFN and MIA.

The recent eigen-decomposition-free method, Ed-free [9], utilizes the localized operator to design an intuitive sampling strategy, whose complexity is $\mathcal{O}(NK\bar{Q})$ in Table I. Our BS-GDA is faster than Ed-free for both graphs, in particular on the community graph. The reason is that edges within each community are relatively dense, and thus \bar{Q} for community graph is much larger than \bar{Q} for random sensor graph, leading to a slower speed for Ed-free. In contrast, the computation efficiency of BS-GDA is similar on both graphs.

Random [27] is faster than our proposed BS-GDA, but its reconstruction error performance is noticeably worse than BS-GDA, as shown in the following section.

C. Reconstruction Error Comparisons

We also compare the resulting reconstruction MSEs between our proposed algorithm and the competing ones. We choose $N = 500$ for random sensor graphs, community graphs and Barabási-Albert graphs. Minnesota road graph is of fixed size

$N = 2642$. All the graph sampling set selection algorithms are compared on these four graphs with increasing sampling budgets.

For each graph, we consider two types of graph signals, as described in Sec. VI-A. We randomly generate 50 graph signals for each type, and for each signal we add 50 different i.i.d. Gaussian noise, resulting in 2500 noisy signals for each experiment. For signal type **GS2**, because the generated signals are not strictly bandlimited, we approximately set the bandwidth to $\omega = \theta_{\lfloor N/10 \rfloor}$ for sampling algorithms that require this parameter.

We use our proposed BS-GDA and the competing algorithms to sample the noisy signals. All the competing algorithms are run with the default or the recommended settings in the referred publications. We then reconstruct original graph signals by solving GLR-based reconstruction (8) with $\mu = 0.01$.

The average MSEs in terms of sampling budgets are illustrated in Fig. 8. As shown, our proposed BS-GDA achieves comparable or smaller MSE values to the competing deterministic schemes on all the four graphs and achieves much better performance than the random sampling scheme Random.

For eigen-decomposition based algorithms, the performances of E-optimal, MFN and SP are unsatisfactory when the sampling budget is much smaller than the bandwidth. Further, these eigen-decomposition based algorithms are computationally too expensive for large graphs, such as the Minnesota road graph. For eigen-decomposition-free algorithms, MIA and Ed-free work well on the random sensor graphs, the community graphs and the Barabási-Albert graphs, but their performance is poor on the large Minnesota road graph. To avoid explicit eigen-decomposition, symmetrically normalized \mathbf{L}_n is designated for MIA and Ed-free, whose eigenvalues are theoretically bounded in $[0, 2]$, to utilize Chebyshev polynomials of \mathbf{L}_n to approximate the low-pass graph filter operator. These two schemes are not robust enough on the graph signals generated on the Minnesota road graph. It has been observed that, in general, the performance of Random

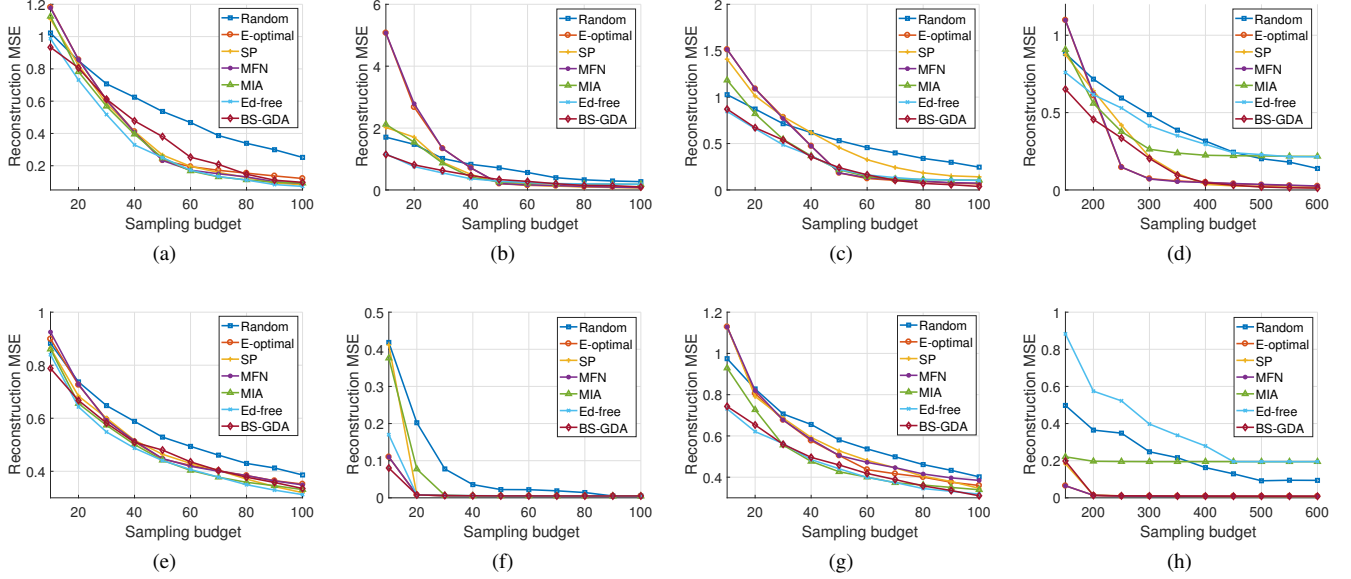


Fig. 8. Reconstruction MSEs for different graphs and sampling algorithms using GLR-based reconstruction. (a) Random sensor graph of size $N = 500$ and signal type **GS1** with bandwidth $\omega = \theta_{50}$. (b) Community graph of size $N = 500$ and 11 random communities and signal type **GS1** with bandwidth $\omega = \theta_{50}$. (c) Barabási-Albert graph of size $N = 500$ and signal type **GS1** with bandwidth $\omega = \theta_{50}$. (d) Minnesota road graph of size $N = 2642$ and signal type **GS1** with bandwidth $\omega = \theta_{264}$. (e) Random sensor graph of size $N = 500$ and signal type **GS2**. (f) Community graph of size $N = 500$ and 11 random communities and signal type **GS2**. (g) Barabási-Albert graph of size $N = 500$ and signal type **GS2**. (h) Minnesota road graph of size $N = 2642$ and signal type **GS2**.

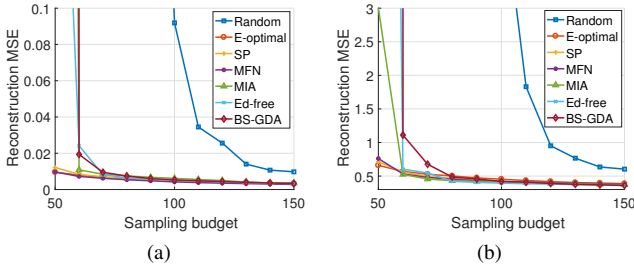


Fig. 9. Reconstruction MSEs for different sampling algorithms using least-squares reconstruction on random sensor graph. (a) Random sensor graph of size $N = 500$ and signal type **GS1** with bandwidth $\omega = \theta_{50}$. (b) Random sensor graph of size $N = 500$ and signal type **GS2** assuming bandwidth $\omega = \theta_{50}$.

is not comparable to deterministic graph sampling algorithms.

Since the previous deterministic graph sampling algorithms such as E-optimal and MFN are designed assuming a standard least-squares reconstruction, we further test all the sampling algorithms with least-squares reconstruction on random sensor graphs, as shown in Fig. 9. We set the sampling budget to be no smaller than the bandwidth for **GS1** or **GS2**. As observed in Fig. 9, E-optimal, MFN and SP achieve robust performance. For MIA, Ed-free and the proposed BS-GDA, their reconstruction MSEs are somewhat larger when the sampling budget is close to the bandwidth. As the sampling budget increases, all deterministic sampling algorithms have similar reconstruction MSEs. In contrast, the performance of Random is noticeably worse than the deterministic sampling algorithms.

The proposed BS-GDA solves the dual (21) via greedy set-

covering and performs a binary search to find the optimal lower bound of the primal (20), which reduces the global upper bound of the MSE value. Hence, BS-GDA guarantees robust performance for different graphs in the experiments. We further show an illustrative example of sampling $K = 11$ nodes on the community graph of size $N = 500$ with 11 communities in Fig. 10. This illustration clearly shows that BS-GDA captures the graph structure very well and evenly samples one node in each community. This matches our sampling analogy to the set cover (SC) problem, where each sampled node covers many uncovered nodes.

VII. CONCLUSION

To address the “graph sampling with noise” problem, in this paper, we first assume a biased graph Laplacian regularization (GLR) based signal reconstruction scheme from samples, and then propose a fast graph sampling set selection algorithm to maximize the stability of the resulting linear system. In particular, our proposed BS-GDA maximizes the lower-bound of λ_{\min} of the coefficient matrix via Gershgorin disc alignment, which is proven to mean also minimizing the upper-bound of reconstruction MSE. Unlike most previous graph sampling schemes, our proposed BS-GDA does not require computation of any eigen-pairs. Extensive simulation experiments show that our disc-based sampling algorithm runs substantially faster than existing eigen-decomposition-based schemes (hundreds to thousands times faster) for medium-sized graphs and outperforms other eigen-decomposition-free schemes in reconstruction error for large-sized graphs.

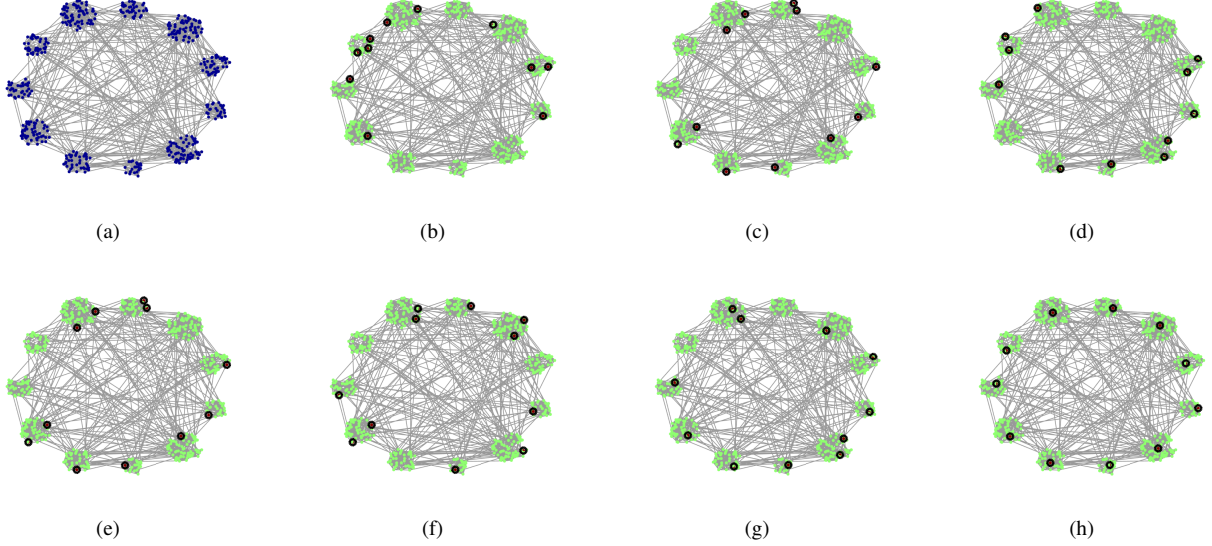


Fig. 10. Visualization of selected nodes on the community graph ($N = 500, K = 11$). Black circles denote sampled nodes. (a) Original graph. (b) Random [27]. (c) E-optimal [24]. (d) SP [16]. (e) MFN [22]. (f) MIA [20]. (g) Ed-free [9]. (h) The proposed BS-GDA.

APPENDIX A PROOF OF PROPOSITION 4

Proof. Since matrix \mathbf{C} is similar to \mathbf{B} , we have $\lambda_{\min}(\mathbf{B}) = \lambda_{\min}(\mathbf{C})$. Hence (22) is equivalent to

$$\min_i \ell_i \leq \lambda_{\min}(\mathbf{C}) \leq \max_i \ell_i \quad (33)$$

From GCT, we see that $\lambda_{\min}(\mathbf{C})$ is bounded from below by the smallest left-end ℓ_i , i.e.

$$\lambda_{\min}(\mathbf{C}) \geq \min_i \ell_i = \min_i \left(c_{ii} - \sum_{j \neq i} |c_{ij}| \right). \quad (34)$$

We next prove that the largest left-end ℓ_i is the upper bound of $\lambda_{\min}(\mathbf{C})$, i.e.,

$$\lambda_{\min}(\mathbf{C}) \leq \max_i \ell_i = \max_i \left(c_{ii} - \sum_{j \neq i} |c_{ij}| \right). \quad (35)$$

Because $\mathbf{B} = \mathbf{A} + \mu\mathbf{L}$ and $\mathbf{S} = \text{diag}(\mathbf{s})$, $\mathbf{s} > \mathbf{0}$, the diagonal elements of \mathbf{C} are non-negative $c_{ii} \geq 0$ and the off-diagonal elements of \mathbf{C} are non-positive $c_{ij} \leq 0$ for $i \neq j$. We consider the matrix $\tilde{\mathbf{C}} := -\mathbf{C} + t\mathbf{I}$, where t is chosen large enough so that the diagonal elements of $\tilde{\mathbf{C}}$ are all non-negative, and hence the whole matrix $\tilde{\mathbf{C}}$ is non-negative. Then by the *Perron-Frobenius Theorem* [35], $\tilde{\mathbf{C}}$ has a positive eigenvalue ϕ_{\max} (with a non-negative eigenvector), which is also the eigenvalue of largest absolute value. Since the eigenvalues λ 's of \mathbf{C} and those ϕ 's of $\tilde{\mathbf{C}}$ are related simply by $\phi = -\lambda + t$, it follows that

$$\lambda_{\min}(\mathbf{C}) = -\phi_{\max} + t \quad (36)$$

is the minimum eigenvalue of \mathbf{C} .

Let $\mathbf{v} = [1, 1, \dots, 1]^\top$ and consider the vector (recalling that $\tilde{c}_{ij} = -c_{ij} = |c_{ij}|$ for $i \neq j$)

$$\tilde{\mathbf{C}}\mathbf{v} = \begin{bmatrix} \tilde{c}_{11} + \sum_{j \neq 1} \tilde{c}_{1j} \\ \tilde{c}_{22} + \sum_{j \neq 2} \tilde{c}_{2j} \\ \vdots \\ \tilde{c}_{nn} + \sum_{j \neq n} \tilde{c}_{nj} \end{bmatrix} = \begin{bmatrix} -c_{11} + t + \sum_{j \neq 1} |c_{1j}| \\ -c_{22} + t + \sum_{j \neq 2} |c_{2j}| \\ \vdots \\ -c_{nn} + t + \sum_{j \neq n} |c_{nj}| \end{bmatrix}. \quad (37)$$

Now by *Collatz-Wielandt' bound* [35], we have

$$\phi_{\max} \geq \min_i (\tilde{\mathbf{C}}\mathbf{v})_i = \min_i \left(-c_{ii} + t + \sum_{j \neq i} |c_{ij}| \right). \quad (38)$$

Together with (36), we have

$$t - \lambda_{\min}(\mathbf{C}) \geq \min_i (\tilde{\mathbf{C}}\mathbf{v})_i = \min_i \left(-c_{ii} + t + \sum_{j \neq i} |c_{ij}| \right). \quad (39)$$

Thus, we conclude that

$$\begin{aligned} \lambda_{\min}(\mathbf{C}) &\leq t - \min_i \left(-c_{ii} + t + \sum_{j \neq i} |c_{ij}| \right) \\ &= \max_i \left(c_{ii} - \sum_{j \neq i} |c_{ij}| \right) = \max_i \ell_i, \end{aligned} \quad (40)$$

as required. \square

APPENDIX B PROOF OF NP-COMPLETENESS

We prove that the decision version of the *disc coverage* (DC) problem (27) is NP-complete by transforming the decision version of the *set cover* (SC) problem [15] to a special case of DC. The SC decision problem asks the following binary decision question: given a family of subsets \mathcal{F} of a finite

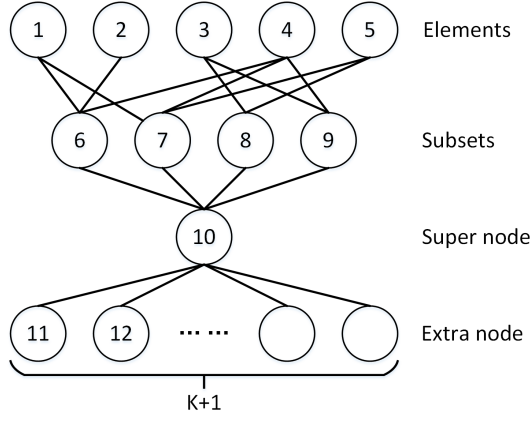


Fig. 11. Example of a mapped dual graph sampling problem from a SC problem.

element set \mathcal{X} and positive integer $K \leq |\mathcal{F}|$, does there exist a cover \mathcal{C} of size K or less, i.e., $\mathcal{C} \subseteq \mathcal{F}$ and $|\mathcal{C}| \leq K$, where each element $x \in \mathcal{X}$ is in at least one member of \mathcal{C} ?

For each instance of SC, we perform the following transformation. First, for each element j in set, $x_j \in \mathcal{X}$, we construct one *element node* j in a graph \mathcal{G} . Next, for each subset in collection $c_i \in \mathcal{F}$, we construct one *subset node* $|\mathcal{X}| + i$ and connect it to element nodes j that correspond to elements $x_j \in c_i$. We construct one *super node* $|\mathcal{X}| + |\mathcal{F}| + 1$ and connect it to $|\mathcal{F}|$ subset nodes. Finally, we construct $K + 1$ *extra nodes* and connect them to the super node. Coverage subset Ω_i for each node i is set to be all 1-hop neighbors of node i in the constructed graph.

As an example, consider the following SC instance:

$$\begin{aligned} \mathcal{X} &= \{1, 2, 3, 4, 5\}, \quad \mathcal{F} = \{c_1, c_2, c_3, c_4\} \\ c_1 &= \{1, 2, 4\}, \quad c_2 = \{1, 4, 5\} \\ c_3 &= \{3, 5\}, \quad c_4 = \{3, 4\} \end{aligned} \quad (41)$$

Fig. 11 illustrates the corresponding constructed graph. The corresponding DC decision problem is: does there exist a sampling vector \mathbf{a} with objective of (27) $K + 1$ or less, while having each graph node i included in at least one sample i 's coverage subset Ω_i ?

We prove that the SC decision problem is true iff the corresponding DC decision problem is true. Suppose first that there exists a cover \mathcal{C} of size $|\mathcal{C}| \leq K$. We observe that a corresponding sample selection of super node $|\mathcal{X}| + |\mathcal{F}| + 1$ and subset nodes representing members $c_i \in \mathcal{C}$ will mean that each graph node is included in at least one sample coverage subset Ω_i , i.e., all nodes are within one hop of selected samples.

Now suppose that there exists a sample selection of size $K + 1$ such that all nodes are within one hop of selected samples. We first observe that the super node must be selected. Otherwise, just covering the $K + 1$ extra nodes would require $K + 1$ node selections, leaving no leftover budget to select and cover element and subset nodes.

Given that the super node is selected, there remain K sample budget to cover $|\mathcal{X}|$ element nodes. We can identify a cover of K' subsets in SC, $K' \leq K$, from the chosen K node samples as follows. Suppose there are K_1 , K_2 and K_3

selected element, subset and extra nodes respectively, where $K_1 + K_2 + K_3 = K$. For each one of K_2 chosen subset node, we select the corresponding subset in SC. For each one of K_1 chosen element node, we randomly select one of its connected subset node and choose the corresponding subset in SC (if not already chosen in previous step). We ignore all K_3 chosen extra nodes. We see that the number of chosen subsets in SC is no larger than $K_1 + K_2 \leq K$, but covers all elements in set \mathcal{X} . Thus a feasible sample set for the DC decision problem maps to a feasible cover \mathcal{C} for the SC decision problem. We have thus proven that an instance of the SC problem maps to our construction of the DC problem. Therefore, we can conclude that the DC problem is at least as hard as the SC problem. Since the SC decision problem is known to be NP-complete, the DC decision problem is also NP-complete.

Given the DC decision problem is NP-complete, the corresponding DC optimization problem in (27) is NP-hard. \square

REFERENCES

- [1] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1832–1843, 2016.
- [2] D. Valsesia, G. Fracastoro, and E. Magli, "Sampling of graph signals via randomized local aggregations," *IEEE Transactions on Signal and Information Processing over Networks*, 2018.
- [3] X. Wang, J. Chen, and Y. Gu, "Local measurement and reconstruction for noisy bandlimited graph signals," *Signal Processing*, vol. 129, pp. 119–129, 2016.
- [4] S. Chen, A. Sandryhaila, and J. Kovačević, "Sampling theory for graph signals," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 3392–3396.
- [5] I. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Transactions of the American Mathematical Society*, vol. 360, no. 10, pp. 5603–5627, 2008.
- [6] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal recovery on graphs: Random versus experimentally designed sampling," in *2015 International Conference on Sampling Theory and Applications (SampTA)*. IEEE, 2015, pp. 337–341.
- [7] H. Shomorony and A. S. Avestimehr, "Sampling large data on graphs," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2014, pp. 933–936.
- [8] A. Sakiyama, Y. Tanaka, T. Tanaka, and A. Ortega, "Accelerated sensor position selection using graph localization operator," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 5890–5894.
- [9] —, "Eigendecomposition-free sampling set selection for graph signals," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2679–2692, May 2019.
- [10] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal recovery on graphs: Fundamental limits of sampling strategies," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 539–554, Dec 2016.
- [11] A. Gadde, A. Anis, and A. Ortega, "Active semi-supervised learning using sampling theory for graph signals," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '14. New York, NY, USA: ACM, 2014, pp. 492–501.
- [12] G. Ortiz-Jiménez, M. Coutino, S. P. Chepuri, and G. Leus, "Sparse sampling for inverse problems with tensors," *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3272–3286, June 2019.
- [13] J. Pang and G. Cheung, "Graph Laplacian regularization for image denoising: Analysis in the continuous domain," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1770–1785, April 2017.
- [14] R. S. Varga, *Geršgorin and His Circles*. Springer, 2004.
- [15] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2009.
- [16] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.

- [17] A. Anis, A. El Gamal, A. S. Avestimehr, and A. Ortega, "A sampling theory perspective of graph-based semi-supervised learning," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2322–2342, 2018.
- [18] A. Gadde and A. Ortega, "A probabilistic interpretation of sampling theory of graph signals," in *2015 IEEE international conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 3257–3261.
- [19] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [20] F. Wang, Y. Wang, and G. Cheung, "A-optimal sampling and robust reconstruction for graph signals via truncated neumann series," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 680–684, 2018.
- [21] F. Pukelsheim, *Optimal design of experiments*. SIAM, 1993, vol. 50.
- [22] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.
- [23] L. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 34–47, 2018.
- [24] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, 2015.
- [25] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.
- [26] Y. Bai, G. Cheung, F. Wang, X. Liu, and W. Gao, "Reconstruction-cognizant graph sampling using gershgorin disc alignment," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 5396–5400.
- [27] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Applied and Computational Harmonic Analysis*, vol. 44, no. 2, pp. 446–475, 2018.
- [28] G. Puy and P. Pérez, "Structured sampling and fast reconstruction of smooth graph signals," *Information and Inference: A Journal of the IMA*, vol. 7, no. 4, pp. 657–688, 2018.
- [29] C. Zhang, D. Florêncio, and P. A. Chou, "Graph signal processing-a probabilistic framework," *Microsoft Res., Redmond, WA, USA, Tech. Rep. MSR-TR-2015-31*, 2015.
- [30] G. Dasarthy, R. Nowak, and X. Zhu, "S2: An efficient graph based active learning algorithm with application to nonparametric classification," in *Conference on Learning Theory*, 2015, pp. 503–522.
- [31] E. E. Gad, A. Gadde, A. S. Avestimehr, and A. Ortega, "Active learning on weighted graphs using adaptive and non-adaptive approaches," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 6175–6179.
- [32] P.-Y. Chen and D. Wei, "On the supermodularity of active graph-based semi-supervised learning with stieltjes matrix regularization," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2801–2805.
- [33] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," in *IEEE Signal Processing Magazine*, vol. 30, no.3, May 2013, pp. 83–98.
- [34] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.
- [35] R. A. Horn, R. A. Horn, and C. R. Johnson, *Matrix analysis*. Cambridge university press, 1990.
- [36] A. Anis, A. Gadde, and A. Ortega, "Towards a sampling theorem for signals on arbitrary graphs," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, Conference Proceedings, pp. 3864–3868.
- [37] X. Liu, G. Cheung, X. Wu, and D. Zhao, "Random walk graph laplacian-based smoothness prior for soft decoding of jpeg images," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 509–524, Feb 2017.
- [38] G. Cheung, E. Magli, Y. Tanaka, and M. K. Ng, "Graph spectral image processing," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 907–930, 2018.
- [39] Y. Bai, G. Cheung, X. Liu, and W. Gao, "Graph-based blind image deblurring from a single photograph," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1404–1418, March 2019.
- [40] X. Liu, G. Cheung, X. Ji, D. Zhao, and W. Gao, "Graph-based joint dequantization and contrast enhancement of poorly lit jpeg images," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1205–1219, 2018.
- [41] C. Dinesh, G. Cheung, and I. V. Bajić, "3D point cloud denoising via bipartite graph approximation and reweighted graph laplacian," *arXiv preprint arXiv:1812.07711*, 2018.
- [42] O. Sorkine, "Laplacian mesh processing," in *Eurographics 2005 - State of the Art Reports*, Y. Chrysanthou and M. Magnor, Eds. The Eurographics Association, 2005.
- [43] M. Belkin, I. Matveeva, and P. Niyogi, "Regularization and semi-supervised learning on large graphs," in *Learning Theory*, J. Shawe-Taylor and Y. Singer, Eds. Springer Berlin Heidelberg, 2004, Conference Proceedings, pp. 624–638.
- [44] G. Cheung, S. W.-T., M. Y., and L. C.-W., "Robust semisupervised graph classifier learning with negative edge weights," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 4, pp. 712–726, 2018.
- [45] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*. Journal of Research of the National Bureau of Standards, 1952, vol. 49, no. 1.
- [46] P. Chen and S. Liu, "Bias-variance tradeoff of graph laplacian regularizer," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1118–1122, Aug 2017.
- [47] B. N. Parlett, *The symmetric eigenvalue problem*. SIAM, 1998, vol. 20.
- [48] A. V. Knyazev, "Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method," *SIAM journal on scientific computing*, vol. 23, no. 2, pp. 517–541, 2001.
- [49] K. G. Murty, *Linear programming*. Chichester, 1983.
- [50] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.
- [51] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," *ArXiv e-prints*, Aug. 2014.