Signals on Graphs: Uncertainty Principle and Sampling

Mikhail Tsitsvero, Sergio Barbarossa, *Fellow, IEEE,* and Paolo Di Lorenzo, *Member, IEEE*

***Abstract*—In many applications of current interest, the observations are represented as a signal defined over a graph. The analysis of such signals requires the extension of standard signal processing tools. Building on the recently introduced Graph Fourier Transform, the first contribution of this paper is to provide an uncertainty principle for signals on graph. As a by-product of this theory, we show how to build a dictionary of maximally concentrated signals on vertex/frequency domains. Then, we establish a direct relation between uncertainty principle and sampling, which forms the basis for a sampling theorem for graph signals. Since samples' location plays a key role in the performance of signal recovery algorithms, we suggest and compare a few alternative sampling strategies. Finally, we provide the conditions for perfect recovery of a useful signal corrupted by sparse noise, showing that this problem is also intrinsically related to vertex-frequency localization properties.**

***Index Terms*—signals on graphs, Graph Fourier Transform, uncertainty principle, sampling theory**

I. Introduction

In many applications, from sensor to social networks, transportation systems, gene regulatory networks or big data, the signals of interest are defined over the vertices of a graph [1], [2]. Over the last few years, a series of papers produced a significant advancement in the development of processing tools for the analysis of signals defined over a graph, or graph signals for short [1], [3]. One of the unique features in graph signal processing is that the analysis tools come to depend on the graph topology. This paves the way to a plethora of methods, each emphasizing different aspects of the problem. A central role is of course played by spectral analysis of graph signals, which passes through the introduction of the so called Graph Fourier Transform (GFT). Alternative definitions of GFT exist, depending on the different perspectives used to extend classical tools. Two basic approaches are available, proposing the projection of the graph signal onto the eigenvectors of either the graph Laplacian, see, e.g., [1], [4] or of the adjacency matrix, see, e.g. [3], [5]. The first approach builds on the spectral clustering properties of the Laplacian eigenvectors and the minimization of the ℓ_2 norm graph total variation; the second approach is based on the interpretation of the adjacency operator as the graph shift operator, which lies at the heart of all linear shift-invariant filtering methods for graph signals [6], [7].

M. Tsitsvero, S. Barbarossa and P. Di Lorenzo are with the Department of Information Eng., Electronics and Telecommunications, Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy, e-mail: tsitsvero@gmail.com, sergio.barbarossa@uniroma1.it, dilorenzo@infocom.uniroma1.it

Given the definition of the GFT, in [8] and very recently in [9], [10], [11] it was derived a graph uncertainty principle aimed at expressing the fundamental relation between the spread of a signal over the vertex and spectral domains. The approach used in [8] is basically a transposition of classical Heisenberg's method to graph signals. However, although conceptually interesting, this transposition presents a series of shortcomings, essentially related to the fact that while time and frequency domains are inherently metric spaces, the vertex domain is not. This requires a careful definition of spread in vertex and GFT domain, which should not make any assumption about ordering and metrics over the vertex domain.

The next fundamental tool in signal processing is sampling theory. A basic contribution to the extension of sampling theory to graph signals was given in [12]. The theory developed in [12] aimed to show that, given a subset of samples, there exists a cutoff frequency ω such that, if the spectral support of the signal is in $[0, \omega]$, the overall signal can be reconstructed with no errors. Later, [13] extended the results of [12] providing a method to compute the cut-off frequency and to interpolate signals which are not exactly band-limited. Two very recent works address also the problem of sampling on graphs providing conditions for perfect recovery [5], based on the adjacency matrix formulation of the GFT, and local-set graph signal reconstructions [14].

The goal of this paper is to contribute to graph signal processing in three main aspects:

A) Uncertainty principle: We derive an uncertainty principle for graph signals, based on the generalization of classical Slepian-Landau-Pollack seminal works [15], [16]; extending those works to graph signals, we derive the boundaries of the uncertainty region in closed form and show that they are attainable; differently from [8], the proposed methodology is perfectly suitable to work on signals defined over a non-metric space, as it does not involve any measure of distance neither in the signal nor in the transformed domain. We derive the conditions for perfect localization in *both* vertex and frequency domains and build a class of maximally concentrated signals in the joint vertex-frequency domain; this extends the choice of suitable dictionaries for the joint vertex-frequency analysis proposed, e.g., in [17];

B) Sampling: We establish a link between uncertainty principle and sampling theory; based on this link, we derive the necessary and sufficient conditions for the recovery of band-limited signals from a subset of observations and suggest alternative sampling strategies aimed at improving the performance of the recovery algorithms; differently from [12] and [18], in our

band-limited model, the signal support may be the union of disjoint sets;

C) Signal recovery from noisy observations: We provide algorithms for recovering band-limited signals from noisy observations; as a particular case, we show that, under suitable conditions, if noise is confined to a small subset of vertices, perfect recovery is possible, irrespective of the signal-to-noise ratio.

Notation: We consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of a set of N nodes $\mathcal{V} = \{1, 2, ..., N\}$, along with a set of weighted edges $\mathcal{E} = \{a_{ij}\}_{i,j\in\mathcal{V}}$, such that $a_{ij} > 0$, if there is a link from node j to node i, or $a_{ij} = 0$, otherwise. The symbol $|\mathcal{S}|$ denotes the cardinality of set $\mathcal{S}$, i.e., the number of elements of $\mathcal{S}$. A signal $\boldsymbol{x}$ over a graph $\mathcal{G}$ is defined as a mapping from the vertex set to complex vectors of size N, i.e. $\boldsymbol{x} : \mathcal{V} \to \mathbb{C}^{|\mathcal{V}|}$. The adjacency matrix $\mathbf{A}$ of a graph is the collection of all the weights $a_{ij}, i, j = 1, \ldots, N$. The degree of node i is $k_i := \sum_{j=1}^{N} a_{ij}$. The degree matrix is a diagonal matrix having the node degrees on its diagonal: $\mathbf{K} = \mathrm{diag}\{k_1, k_2, ..., k_N\}$. The combinatorial Laplacian matrix is defined as $\mathbf{L} = \mathbf{K} - \mathbf{A}$. In the literature, it is also common to use the normalized graph Laplacian matrix $\mathcal{L} = \mathbf{K}^{-1/2}\mathbf{L}\mathbf{K}^{-1/2}$.

The Graph Fourier Transform (GFT) has been defined in alternative ways, depending on different ways to identify the Fourier basis, either from the eigenvectors of the Laplacian, adjacency matrix or, more generally, as some orthonormal set of complex-valued vectors specifying some analysing features. If the graph is undirected, the Laplacian matrix is symmetric and positive semi-definite. Hence, it may be diagonalized as

$$\mathbf{L} = \mathbf{U}\boldsymbol{\Xi}\mathbf{U}^* = \sum_{i=1}^{N} \xi_i \boldsymbol{u}_i \boldsymbol{u}_i^*, \tag{1}$$

where $\boldsymbol{\Xi}$ is a diagonal matrix with non-negative real eigenvalues $\{\xi_i\}$ on its diagonal, $\{\boldsymbol{u}_i\}$ is the set of real-valued orthonormal eigenvectors and $(\cdot)^*$ denotes conjugate transpose. We denote the set of all indices enumerating the GFT basis vectors as $\mathcal{V}^* = \{1, 2, ..., N\}$. From spectral graph theory, see e.g., [19], the Laplacian eigenvectors are well known to possess useful clustering properties. Furthermore, defining the ℓ_2-norm total variation of a signal $\boldsymbol{x}$ as

$$TV_{\mathcal{G}}^{(2)}(\boldsymbol{x}) := \left[\frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(x_i - x_j)^2\right]^{1/2} = \left(\boldsymbol{x}^*\mathbf{L}\boldsymbol{x}\right)^{1/2}, \tag{2}$$

the eigenvectors of $\mathbf{L}$ constitute a set of orthonormal vectors that minimize the graph ℓ_2-norm total variation [4]. The previous properties supported the definition of the Graph Fourier Transform of a vector $\boldsymbol{x}$ as the projection onto the space spanned by the Laplacian eigenvectors [1], [4], [8], [13]:

$$\hat{\boldsymbol{x}} = \mathbf{U}^*\boldsymbol{x} \tag{3}$$

with inverse Fourier transform

$$\boldsymbol{x} = \mathbf{U}\hat{\boldsymbol{x}}. \tag{4}$$

Alternatively, the authors of [20] started from the Jordan decomposition of the adjacency matrix as

$$\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}$$

where $\mathbf{J}$ is a block diagonal matrix representing the Jordan normal form of $\mathbf{A}$, and defined the Graph Fourier Transform of a vector $\boldsymbol{x}$ as

$$\hat{\boldsymbol{x}} = \mathbf{V}^{-1}\boldsymbol{x}.$$

This definition is rooted on the so called *algebraic signal processing theory* [6], [7], where the shift operator z^{-1} of classical discrete time signal processing is identified as the multiplication by the adjacency matrix $\mathbf{A}$ in the graph domain. In case of undirected *regular* graphs, i.e. graphs where all nodes have the same degree k, we have $\mathbf{L} = k\mathbf{I} - \mathbf{A}$. In such a case, the eigenvectors of $\mathbf{A}$ and $\mathbf{L}$ coincide and then the two alternative definitions of GFT boil down to the same operator.

In this paper, in all our numerical examples we use the definition of GFT based on the Laplacian matrix. Nevertheless, all our theoretical derivations can be applied to a more general case of complex-valued orthonormal basis.

Even though graph signals are typically defined as a mapping from the vertices of the graph to the complex numbers, we can envisage a broader sense definition, where the signal represents a mapping from, in principle, any subset of $\mathcal{V}$ to the complex valued space of dimension equal to the number of subsets. For example, we can define a signal over the edges of a graph as a mapping from the edge set to complex numbers, i.e., $\boldsymbol{x} : \mathcal{E} \to \mathbb{C}^{|\mathcal{E}|}$. Nevertheless, in this paper, we will be primarily interested in signals defined over the vertices of a graph.

Given a subset of vertices $\mathcal{S} \subseteq \mathcal{V}$, we define a vertex-limiting operator as a diagonal matrix $\mathbf{D}_{\mathcal{S}}$ such that

$$\mathbf{D}_{\mathcal{S}} = \mathrm{Diag}\{\mathbf{1}_{\mathcal{S}}\} \tag{5}$$

where $\mathbf{1}_{\mathcal{S}}$ is the set indicator vector, whose i-th entry is equal to one, if $i \in \mathcal{S}$, or zero otherwise. Similarly, given a subset of frequency indices $\mathcal{F} \subseteq \mathcal{V}^*$, we introduce the filtering operator

$$\mathbf{B}_{\mathcal{F}} = \mathbf{U}\boldsymbol{\Sigma}_{\mathcal{F}}\mathbf{U}^*, \tag{6}$$

where $\boldsymbol{\Sigma}_{\mathcal{F}}$ is a diagonal matrix defined as $\boldsymbol{\Sigma}_{\mathcal{F}} = \mathrm{Diag}\{\mathbf{1}_{\mathcal{F}}\}$. It is immediate to check that both matrices $\mathbf{D}_{\mathcal{S}}$ and $\mathbf{B}_{\mathcal{F}}$ are self-adjoint and idempotent, and then they represent orthogonal projectors. As in [12], we refer to the space of all signals whose GFT is exactly supported on the set $\mathcal{F}$, as the *Paley-Wiener space* for the set $\mathcal{F}$. We denote by $\mathcal{B}_{\mathcal{F}} \subseteq L_2(\mathcal{G})$ the set of all finite ℓ_2-norm signals belonging to the Paley-Wiener space associated to $\mathcal{F}$. Similarly, we denote by $\mathcal{D}_{\mathcal{S}} \subseteq L_2(\mathcal{G})$ the set of all finite ℓ_2-norm signals with support on the vertex subset $\mathcal{S}$. In the rest of the paper, whenever there will be no ambiguities in the specification of the sets, we will drop the subscripts referring to the sets. Finally, given a set $\mathcal{S}$, we denote its complement set as $\overline{\mathcal{S}}$, such that $\mathcal{V} = \mathcal{S} \cup \overline{\mathcal{S}}$ and $\mathcal{S} \cap \overline{\mathcal{S}} = \emptyset$. Correspondingly, we define the vertex-projector onto $\overline{\mathcal{S}}$ as $\overline{\mathbf{D}}$ and, similarly, the frequency projector onto the frequency domain $\overline{\mathcal{F}}$ as $\overline{\mathbf{B}}$.

The rest of the paper is organized as follows. In Section II we derive the localization properties of graph signals, illustrating as a particular case the conditions enabling perfect localization in *both* vertex and frequency domains. Building on these tools, in Section III we derive an uncertainty principle for graph signals and, in Section IV, we derive the necessary and

sufficient conditions for recovering band-limited graph signal from its samples and propose alternative recovery algorithms. In Section V we analyze the effect of observation noise on signal recovery and, finally, in Section VI we propose and compare several sampling strategies.

II. Localization Properties

Scope of this section is to derive the class of signals maximally concentrated over given subsets $\mathcal{S}$ and $\mathcal{F}$ in vertex and frequency domains. We say that a vector $\boldsymbol{x}$ is perfectly localized over the subset $\mathcal{S} \subseteq \mathcal{V}$ if

$$\mathbf{D}\boldsymbol{x} = \boldsymbol{x}, \tag{7}$$

with $\mathbf{D}$ defined as in (5). Similarly, a vector $\boldsymbol{x}$ is perfectly localized over the frequency set $\mathcal{F} \subseteq \mathcal{V}^*$ if

$$\mathbf{B}\boldsymbol{x} = \boldsymbol{x}, \tag{8}$$

with $\mathbf{B}$ given in (6). Differently from continuous-time signals, a graph signal can be perfectly localized in *both* vertex and frequency domains. This is stated in the following

Theorem 2.1: A vector $\boldsymbol{x} \in L_2(\mathcal{G})$ is perfectly localized over both vertex set $\mathcal{S}$ and frequency set $\mathcal{F}$ if and only if the operator $\mathbf{BDB}$ has an eigenvalue equal to one; in such a case, $\boldsymbol{x}$ is an eigenvector associated to the unit eigenvalue.

Proof: Let us start proving that, if a vector $\boldsymbol{x}$ is perfectly localized in both vertex and frequency domains, then it must be an eigenvector of $\mathbf{BDB}$ associated to a unit eigenvalue. Indeed, by repeated applications of (7) and (8), it follows

$$\mathbf{BDB}\boldsymbol{x} = \mathbf{BD}\boldsymbol{x} = \mathbf{B}\boldsymbol{x} = \boldsymbol{x}. \tag{9}$$

This proves the first part. Now, let us prove that, if $\boldsymbol{x}$ is an eigenvector of $\mathbf{BDB}$ associated to a unit eigenvalue, then $\boldsymbol{x}$ must satisfy (7) and (8). Indeed, starting from

$$\mathbf{BDB}\boldsymbol{x} = \boldsymbol{x} \tag{10}$$

and multiplying from the left side by $\mathbf{B}$, taking into account that $\mathbf{B}^2 = \mathbf{B}$, we get

$$\mathbf{BDB}\boldsymbol{x} = \mathbf{B}\boldsymbol{x} \tag{11}$$

Equating (10) to (11), we get

$$\mathbf{B}\boldsymbol{x} = \boldsymbol{x}, \tag{12}$$

which implies that $\boldsymbol{x}$ is perfectly localized in the frequency domain. Now, using (12) and the Rayleigh-Ritz theorem, we can also write

$$1 = \max_{\boldsymbol{x}} \frac{\boldsymbol{x}^*\mathbf{BDB}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}} = \max_{\boldsymbol{x}} \frac{\boldsymbol{x}^*\mathbf{D}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}. \tag{13}$$

This shows that $\boldsymbol{x}$ satisfies also (7), i.e., $\boldsymbol{x}$ is also perfectly localized in the vertex domain. ■ ■

Equivalently, the perfect localization properties can be expressed in terms of the operators $\mathbf{BD}$ and $\mathbf{DB}$. First of all, we prove the following

Lemma 2.2: The operators $\mathbf{BD}$ and $\mathbf{DB}$ have the same singular values, i.e. $\sigma_i(\mathbf{BD}) = \sigma_i(\mathbf{DB}),\ i = 1, \dots, N$.

Proof: Since both matrices $\mathbf{B}$ and $\mathbf{D}$ are Hermitian, $(\mathbf{BD})^* = \mathbf{DB}$. But the singular values of a matrix coincide with the singular values of its Hermitian conjugate. ■ ■

Hence, combining Lemma 2.2 and (9), perfect localization onto the sets $\mathcal{S}$ and $\mathcal{F}$ can be achieved if the following properties hold true:

$$\|\mathbf{BD}\|_2 = 1; \quad \|\mathbf{DB}\|_2 = 1. \tag{14}$$

From Theorem 2.1, perfect localization is possible if and only if $\mathbf{BDB}$ admits a unit eigenvalue. If this happens, (7) and (8) hold true. As a consequence,

$$\mathbf{BD}\boldsymbol{x} = \boldsymbol{x}. \tag{15}$$

Since $\mathbf{B} + \overline{\mathbf{B}} = \mathbf{I}$, by construction, we then have

$$\mathbf{D}\boldsymbol{x} = \boldsymbol{x} = (\mathbf{B} + \overline{\mathbf{B}})\,\mathbf{D}\boldsymbol{x} = \mathbf{BD}\boldsymbol{x} + \overline{\mathbf{B}}\mathbf{D}\boldsymbol{x}.$$

This means that

$$\overline{\mathbf{B}}\mathbf{D}\boldsymbol{x} = \mathbf{0}. \tag{16}$$

From (16), using (6) applied to $\overline{\mathcal{F}}$, since $\mathbf{U}$ is a unitary matrix, perfect localization is feasible only if the following system of linear equations admits a non-trivial solution

$$\overline{\mathbf{\Sigma}}\mathbf{U}^*\mathbf{D}\boldsymbol{x} = \mathbf{0}. \tag{17}$$

This equation can be rewritten in compact form by retaining only the nonzero rows corresponding to the indices i belonging to the complement set $\overline{\mathcal{F}}$, and the columns corresponding to the indices belonging to $\mathcal{S}$. More specifically, denoting by $i_1, \dots, i_{|\overline{\mathcal{F}}|}$ the indices in $\overline{\mathcal{F}}$ and by $j_1, \dots, j_{|\mathcal{S}|}$ the indices in $\mathcal{S}$, (17) can be rewritten in compact form as

$$\mathbf{W}\boldsymbol{\phi} = \mathbf{0}, \tag{18}$$

where $(\mathbf{W})_{k\ell} = u_{i_k}(j_\ell)$, for $k = 1, \dots, N - |\mathcal{F}|$ and $\ell = 1, \dots, |\mathcal{S}|$, is the j_ℓ-th entry of the i_k-th column of $\mathbf{U}$. Matrix $\mathbf{W}$ has dimensionality $(N - |\mathcal{F}|) \times |\mathcal{S}|$. Hence, a sufficient condition for the existence of perfectly localized vectors in both vertex and frequency domains is

$$|\mathcal{S}| \geq N - |\mathcal{F}| + 1. \tag{19}$$

This property marks a fundamental difference with respect to continuous-time signals that cannot be perfectly localized in both time and frequency domains. Conversely, if $|\mathcal{S}| < (N - |\mathcal{F}|)$, system (18) can still admit a non-trivial solution, but only if the matrix $\mathbf{W}$ is not full column rank.

Typically, given two generic domains $\mathcal{S}$ and $\mathcal{F}$, we might not have perfectly concentrated signals in both domains. In such a case, it is worth finding the class of signals with limited support in one domain and maximally concentrated on the other. For example, we may search for the class of perfectly band-limited signals, i.e. satisfying $\mathbf{B}\boldsymbol{x} = \boldsymbol{x}$, which are maximally concentrated in a vertex domain $\mathcal{S}$ or, viceversa, the class of signals with support on a subset of vertices, i.e. satisfying $\mathbf{D}\boldsymbol{x} = \boldsymbol{x}$, which are maximally concentrated in a frequency domain $\mathcal{F}$. Let us start with the band-limited scenario.

Theorem 2.3: The class of orthonormal band-limited vectors $\boldsymbol{\psi}_i$, $i = 1, \dots, N$, with $\mathbf{B}\boldsymbol{\psi}_i = \boldsymbol{\psi}_i$, maximally concentrated over a vertex set $\mathcal{S}$, is given by the eigenvectors of the $\mathbf{BDB}$, i.e.

$$\mathbf{BDB}\boldsymbol{\psi}_i = \lambda_i\boldsymbol{\psi}_i, \tag{20}$$

with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$. Furthermore, these vectors are orthogonal over the set $\mathcal{S}$, i.e.

$$\langle \boldsymbol{\psi}_i, \mathbf{D}\boldsymbol{\psi}_j \rangle = \lambda_j \delta_{ij}, \tag{21}$$

where δ_{ij} is the Kronecker symbol.

Proof: The vectors $\boldsymbol{\psi}_i$ can be built as a solution of the following optimization problem:

$$\begin{aligned} \boldsymbol{\psi}_i \quad &= \arg\max_{\boldsymbol{\psi}_i} \|\mathbf{D}\boldsymbol{\psi}_i\|_2 \\ &\quad \text{s. t. } \|\boldsymbol{\psi}_i\|_2 = 1, \\ &\quad \mathbf{B}\boldsymbol{\psi}_i = \boldsymbol{\psi}_i, \\ &\quad \langle \boldsymbol{\psi}_i, \boldsymbol{\psi}_j \rangle = 0, j \neq i. \end{aligned}$$

Substituting the band-limiting constraint within the objective functions, we get

$$\begin{aligned} \boldsymbol{\psi}_i \quad &= \arg\max_{\boldsymbol{\psi}_i} \|\mathbf{DB}\boldsymbol{\psi}_i\|_2 \\ &\quad \text{s. t. } \|\boldsymbol{\psi}_i\|_2 = 1, \langle \boldsymbol{\psi}_i, \boldsymbol{\psi}_j \rangle = 0, j \neq i. \end{aligned} \tag{22}$$

Using Rayleigh-Ritz theorem, the solutions of (22) are the eigenvectors of $(\mathbf{DB})^* \mathbf{DB} = \mathbf{BDB}$, i.e. the solutions of (20). This proves the first part of the theorem. The second part is proven by noting that, using $\mathbf{B}\boldsymbol{\psi}_i = \boldsymbol{\psi}_i$ and $\mathbf{B}^* = \mathbf{B}$, we obtain $\langle \boldsymbol{\psi}_i, \mathbf{BDB}\boldsymbol{\psi}_j \rangle = \langle \boldsymbol{\psi}_i, \mathbf{D}\boldsymbol{\psi}_j \rangle = \lambda_j \delta_{ij}$. ■ ■

The vectors $\boldsymbol{\psi}_i$ are the counterpart of the prolate spheroidal wave functions introduced by Slepian and Pollack for continuous-time signals [15]. In particular, the eigenvector $\boldsymbol{\psi}_1$ is a band-limited signal with the highest energy concentration on $\mathcal{S}$; the eigenvector $\boldsymbol{\psi}_2$ is a band-limited signal orthogonal to $\boldsymbol{\psi}_1$ and which is maximally concentrated on $\mathcal{S}$, and so on.

III. Uncertainty principle

A cornerstone property of continuous-time signals is the Heisenberg's principle, stating that a signal cannot be perfectly localized in both time and frequency domains (see, e.g., [21] for a survey on the uncertainty principle). More specifically, given a continuous-time signal $x(t)$ and its Fourier Transform $X(f)$, introducing the time spread

$$\Delta_t^2 = \frac{\int_{-\infty}^{\infty} (t - t_0)^2 |x(t)|^2 \, \mathrm{d}t}{\int_{-\infty}^{\infty} |x(t)|^2 \, \mathrm{d}t} \tag{23}$$

with

$$t_0 = \frac{\int_{-\infty}^{\infty} t \, |x(t)|^2 \, \mathrm{d}t}{\int_{-\infty}^{\infty} |x(t)|^2 \, \mathrm{d}t}$$

and the frequency spread

$$\Delta_f^2 = \frac{\int_{-\infty}^{\infty} (f - f_0)^2 |X(f)|^2 \, \mathrm{d}f}{\int_{-\infty}^{\infty} |X(f)|^2 \, \mathrm{d}f}, \tag{24}$$

with

$$f_0 = \frac{\int_{-\infty}^{\infty} f \, |X(f)|^2 \, \mathrm{d}f}{\int_{-\infty}^{\infty} |X(f)|^2 \, \mathrm{d}f},$$

the uncertainty principle states that

$$\Delta_t^2 \Delta_f^2 \geq \frac{1}{(4\pi)^2}.$$

Quite recently, the uncertainty principle was extended to signals on graphs in [8] by following an approach based on the transposition of the previous definitions of time and frequency spreads to graph signals. However, although interesting, this transposition hides a number of subtleties, which can limit the status of the result as a "fundamental" result, i.e. a result not constrained to any specific, thus arbitrary, choice. More specifically, what happens is that the second order moments used in (23) and (24) contain a measure of square distances in the time and frequency domains, represented by the terms $(t - t_0)^2$ and $(f - f_0)^2$, respectively. When transposing these formulas to graph signals, it is necessary to define a distance between vertices of a graph. This is done in [8] by using the common measure of graph distance, defined as the sum of weights along the shortest path between two vertices. But, then, the question is: Is it correct to exchange vertex or frequency distances with a graph distance defined in such a way ? In an attempt to overcome the above problem, in this paper we resort to an alternative definition of spread in vertex and frequency domain, generalizing the works of Slepian, Landau and Pollack [15], [16]. In those works, the effective duration T of a signal centered around a time instant t_0 is the value such that the percentage of energy falling in the interval $[t_0 - T/2, t_0 + T/2]$ assumes a specified value α^2, i.e.

$$\frac{\int_{t_0 - T/2}^{t_0 + T/2} |x(t)|^2 \, \mathrm{d}t}{\int_{-\infty}^{\infty} |x(t)|^2 \, \mathrm{d}t} = \alpha^2.$$

Similarly, the effective bandwidth is the value W such that

$$\frac{\int_{f_0 - W/2}^{f_0 + W/2} |X(f)|^2 \, \mathrm{d}f}{\int_{-\infty}^{\infty} |X(f)|^2 \, \mathrm{d}f} = \beta^2.$$

Equivalently, in the graph domain, given a vertex set $\mathcal{S}$ and a frequency set $\mathcal{F}$, we denote by α^2 and β^2 the percentage of energy falling within the sets $\mathcal{S}$ and $\mathcal{F}$, respectively, as

$$\frac{\|\mathbf{D}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} = \alpha^2; \ \frac{\|\mathbf{B}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} = \beta^2. \tag{25}$$

Generalizing the approach of [16], our goal is to find out the region of all admissible pairs (α, β) and to illustrate which are the signals able to attain all the points in such a region. The uncertainty principle is stated in the following

Theorem 3.1: There exists a vector $\boldsymbol{f} \in L_2(\mathcal{G})$ such that $\|\boldsymbol{f}\|_2 = 1$, $\|\mathbf{D}\boldsymbol{f}\|_2 = \alpha$, $\|\mathbf{B}\boldsymbol{f}\|_2 = \beta$ if and only if $(\alpha, \beta) \in \Gamma$, where

$$\begin{aligned} \Gamma = \{(\alpha, \beta) : \\ &\cos^{-1}\alpha + \cos^{-1}\beta \geq \cos^{-1}\sigma_{max}\left(\mathbf{BD}\right), \\ &\cos^{-1}\sqrt{1-\alpha^2} + \cos^{-1}\beta \geq \cos^{-1}\sigma_{max}\left(\mathbf{B}\overline{\mathbf{D}}\right), \\ &\cos^{-1}\alpha + \cos^{-1}\sqrt{1-\beta^2} \geq \cos^{-1}\sigma_{max}\left(\overline{\mathbf{B}}\mathbf{D}\right), \\ &\cos^{-1}\sqrt{1-\alpha^2} + \cos^{-1}\sqrt{1-\beta^2} \geq \cos^{-1}\sigma_{max}\left(\overline{\mathbf{B}}\overline{\mathbf{D}}\right)\Big\}. \end{aligned} \tag{26}$$

Proof: The proof is reported in Appendix A. ■

An illustrative example of admissible region Γ is reported in Fig. 1. A few remarks about the border of the region Γ are of interest. The curve in the upper right corner of Γ specifies

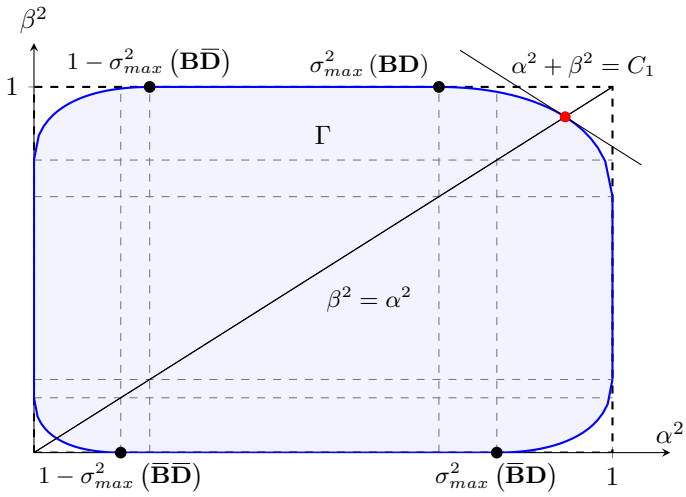

Fig. 1: Admissible region Γ of unit norm signals $\boldsymbol{f} \in L^2(\mathcal{G})$ with $\|\mathbf{D}\boldsymbol{f}\|_2 = \alpha$ and $\|\mathbf{B}\boldsymbol{f}\|_2 = \beta$.

the pairs (α, β) that yield maximum concentration. This curve has equation

$$\cos^{-1}\alpha + \cos^{-1}\beta = \cos^{-1}\sigma_{max}(\mathbf{BD}). \tag{27}$$

Solving with respect to β, and setting $\sigma^2_{max} := \sigma^2_{max}(\mathbf{BD})$, we get

$$\beta = \alpha\,\sigma_{max} + \sqrt{(1-\alpha^2)(1-\sigma^2_{max})}. \tag{28}$$

For any given subset of nodes $\mathcal{S}$, as the cardinality of $\mathcal{F}$ increases, this upper curve gets closer and closer to the upper right corner. Typically, the matrix $\mathbf{W}$ in (18) does not loose rank and hence, as soon as $|\mathcal{F}|$ exceeds the value $N - |\mathcal{S}|$, $\sigma_{max} = 1$ and then the curve collapses onto the upper right corner $(1, 1)$, indicating perfect localization in both vertex and frequency domains. In general, any of the four curves at the corners of region Γ in Fig. 1 may collapse onto the corresponding corner, whenever the conditions for perfect localization of the corresponding operator hold true.

Furthermore, if we are interested in the allocation of energy within the sets $\mathcal{S}$ and $\mathcal{F}$ that maximizes, for example, the sum of the (relative) energies $\alpha^2 + \beta^2$ falling in the vertex and frequency domains, the result is given by the intersection of the upper right curve, i.e. (28), with the line $\alpha^2 + \beta^2 = \text{const}$. Given the symmetry of the curve (27), the result is achieved by setting $\alpha = \beta$, which yields

$$\alpha^2 = \frac{1}{2}(1 + \sigma_{max}). \tag{29}$$

The corresponding function $\boldsymbol{f}'$ may be written in close form

$$\boldsymbol{f}' = \frac{\boldsymbol{\psi}_1 - \mathbf{D}\boldsymbol{\psi}_1}{\sqrt{2\,(1+\sigma_{max})}} + \sqrt{\frac{1+\sigma_{max}}{2\sigma^2_{max}}}\mathbf{D}\boldsymbol{\psi}_1, \tag{30}$$

where $\boldsymbol{\psi}_1$ is the eigenvector of $\mathbf{BDB}$ corresponding to σ^2_{max} and the result (98) from the appendix was used (please check App. A for further details on the structure of signals attaining each point in the admissible region).

A numerical example is useful to grasp the advantages of tolerating some energy spill-over in representing a graph signal. The example is built as follows. We consider a random geometric graph composed of 100 vertices, located over a unit square, with coverage radius r_0. The graph is then made periodic to avoid problems related to border nodes and put every node in the same statistical conditions. Then, we picked a vertex i_0 at random and identify the set $\mathcal{S}$ as the ensemble of nodes falling within a distance R_0 from i_0. Then we let R_0 to increase and, for each value of R_0, we evaluate the cardinality of $\mathcal{S}$ and we build $\mathcal{F}$ as the set of indices $\{1, 2, \ldots, k\}$ enumerating the first k eigenvectors of the Laplacian matrix $\mathbf{L}$, with k as the maximum number such that the (relative) spill-over energy $1-\alpha^2 = 1-\sigma^2_{max}$ is less than a prescribed value ε^2. In Fig. 2 we plot $|\mathcal{F}| = k$ as a function of $|\mathcal{S}|$, for different values of ε^2. The dashed line represents the case $\varepsilon^2 = 0$. This is the curve $N = |\mathcal{S}| + |\mathcal{F}|$. The interesting result is that, as we allow for some spill-over energy, we can get a substantial reduction of the "bandwidth" $|\mathcal{F}|$ necessary to contain a signal defined on a given vertex set $\mathcal{S}$.

IV. Sampling

Given a signal $\boldsymbol{f} \in \mathcal{B}$ defined on the vertices of a graph, let us denote by $\boldsymbol{f}_{\mathcal{S}} \in \mathcal{D}$ the vector equal to $\boldsymbol{f}$ on a subset $\mathcal{S} \subseteq \mathcal{V}$ and zero outside:

$$\boldsymbol{f}_{\mathcal{S}} := \mathbf{D}\boldsymbol{f}. \tag{31}$$

We wish to find out the conditions and the means for perfect recovery of $\boldsymbol{f}$ from $\boldsymbol{f}_{\mathcal{S}}$. The necessary and sufficient conditions are stated in the following

Theorem 4.1 (Sampling Theorem): Given a band-limited vector $\boldsymbol{f} \in \mathcal{B}$, it is possible to recover $\boldsymbol{f}$ from its samples taken from the set $\mathcal{S}$, if and only if

$$\|\mathbf{B}\overline{\mathbf{D}}\|_2 < 1, \tag{32}$$

i.e. if the matrix $\mathbf{B}\overline{\mathbf{D}}\mathbf{B}$ does not have any eigenvector that is perfectly localized on $\overline{\mathcal{S}}$ and bandlimited on $\mathcal{F}$.

Proof: We prove first that condition (32) is sufficient for perfect recovery. Let us denote by $\mathbf{Q}$ a matrix enabling the reconstruction of $\boldsymbol{f}$ from $\boldsymbol{f}_{\mathcal{S}}$ as $\mathbf{Q}\boldsymbol{f}_{\mathcal{S}}$. If such a matrix exists, the corresponding reconstruction error is

$$\boldsymbol{f} - \mathbf{Q}\boldsymbol{f}_{\mathcal{S}} = \boldsymbol{f} - \mathbf{Q}\left(\mathbf{I} - \overline{\mathbf{D}}\right)\boldsymbol{f} = \boldsymbol{f} - \mathbf{Q}\left(\mathbf{I} - \overline{\mathbf{D}}\mathbf{B}\right)\boldsymbol{f}, \tag{33}$$

where, in the second equality, we exploited the band-limited nature of $\boldsymbol{f}$. This error can be made equal to zero by taking $\mathbf{Q} = \left(\mathbf{I} - \overline{\mathbf{D}}\mathbf{B}\right)^{-1}$. Hence, checking for the existence of $\mathbf{Q}$ is equivalent to check if $\left(\mathbf{I} - \overline{\mathbf{D}}\mathbf{B}\right)$ is invertible. This happens if (32) holds true. Conversely, if $\|\mathbf{B}\overline{\mathbf{D}}\|_2 = 1$ and, equivalently, $\|\overline{\mathbf{D}}\mathbf{B}\|_2 = 1$, from (14) we know that there exist band-limited signals that are perfectly localized over $\overline{\mathcal{S}}$. This implies that, if we sample one of such signals over the set $\mathcal{S}$, we get only zero values and then it would be impossible to recover $\boldsymbol{f}$ from those samples. This proves that condition (32) is also necessary. ■ ■

Equivalently, we note that, if $\boldsymbol{f} \in \mathcal{B}$ then

$$\left(\mathbf{I} - \overline{\mathbf{D}}\mathbf{B}\right)\boldsymbol{f} = \mathbf{D}\mathbf{B}\boldsymbol{f}. \tag{34}$$

The operator $\mathbf{DB}$ is invertible, for any $\boldsymbol{f} \in \mathcal{B}$, if the dimensionality of the image of $\mathbf{DB}$ is equal to the $\operatorname{rank}\mathbf{B}$, i.e.

$$\operatorname{rank}\mathbf{DB} = \operatorname{rank}\mathbf{B}. \tag{35}$$

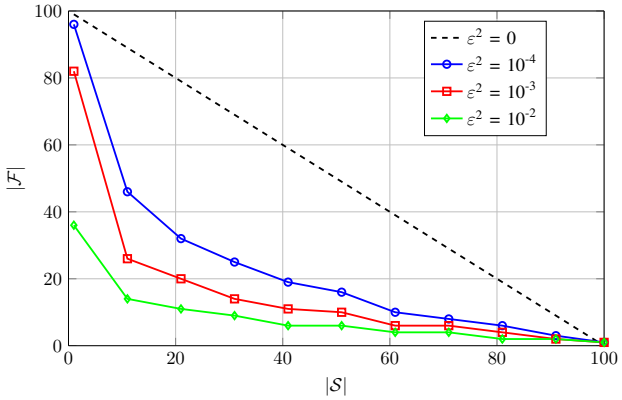

Fig. 2: Relation between the dimensions of the support over the vertex set $\mathcal{S}$ and the frequency domain $\mathcal{F}$ guaranteeing a spill-over energy ε^2.

This condition is then equivalent to the condition of Theorem 4.1. In this case the singular vectors of $\mathbf{DB}$ corresponding to non-zero singular values constitute a basis for $\mathcal{B}$.

Theorem 4.1 suggests also a way to recover the original signal from its samples as $\left(\mathbf{I}-\overline{\mathbf{D}}\mathbf{B}\right)^{-1}\boldsymbol{f}_{\mathcal{S}}$. However, this requires the inversion of a matrix, which in many practical cases may be problematic. To avoid the inversion, we provide next a reconstruction algorithm that leverages on the properties of maximally concentrated signals described in Section II.

Theorem 4.2: If condition (32) of the sampling theorem holds true, then any bandlimited signal $\boldsymbol{f}\in\mathcal{B}$ can be reconstructed from its sampled version $\boldsymbol{f}_{\mathcal{S}}\in\mathcal{D}$ by the following formula

$$\boldsymbol{f}=\sum_{i=1}^{|\mathcal{F}|}\frac{1}{\sigma_i^2}\langle \boldsymbol{f}_{\boldsymbol{\mathcal{S}}},\boldsymbol{\psi}_i\rangle\boldsymbol{\psi}_i, \tag{36}$$

where $\{\boldsymbol{\psi}_i\}_{i=1..N}$ and $\left\{\sigma_i^2\right\}_{i=1..N}$ are the eigenvectors and eigenvalues of $\mathbf{BDB}$.

Proof: For any $\boldsymbol{g}\in L_2(\mathcal{G})$, we can write

$$\boldsymbol{g}=\sum_{i=1}^{N}\langle \boldsymbol{g},\boldsymbol{\psi}_i\rangle\boldsymbol{\psi}_i, \tag{37}$$

and, for its bandlimited projection

$$\mathbf{B}\boldsymbol{g}=\sum_{i=1}^{K}\langle \mathbf{B}\boldsymbol{g},\boldsymbol{\psi}_i\rangle\boldsymbol{\psi}_i, \tag{38}$$

where $|\mathcal{F}|\leq K\leq N$. Because of (32), there is no bandlimited vector in $\mathcal{B}$ perfectly localized on $\overline{\mathcal{S}}$. Hence, all the eigenvectors from $\ker(\mathbf{BDB})$ are out of band $\mathcal{F}$ or, equivalently, belong to $\overline{\mathcal{B}}$, so that $K=|\mathcal{F}|$. Setting $\boldsymbol{f}=\mathbf{B}\boldsymbol{g}$, since $\mathbf{BDB}\boldsymbol{\psi}_i=\sigma_i^2\boldsymbol{\psi}_i$ with $\sigma_i\neq 0$ for $i\in\mathcal{F}$, we can then write

$$\boldsymbol{f}=\sum_{i=1}^{|\mathcal{F}|}\langle \boldsymbol{f},\frac{1}{\sigma_i^2}\mathbf{BDB}\boldsymbol{\psi}_i\rangle\boldsymbol{\psi}_i=\sum_{i=1}^{|\mathcal{F}|}\frac{1}{\sigma_i^2}\langle \mathbf{D}\boldsymbol{f},\boldsymbol{\psi}_i\rangle\boldsymbol{\psi}_i, \tag{39}$$

where we have used the property that the operators $\mathbf{D}$ and $\mathbf{B}$ are self-adjoint and the eigenvectors $\{\boldsymbol{\psi}_i\}_{i=1,\dots,|\mathcal{F}|}$ are bandlimited. ■ ■

Let us study now the implications of condition (32) of Theorem 4.1 on the sampling strategy. To fulfill (32), we need to guarantee that there exist no band-limited signals, i.e. $\mathbf{B}\boldsymbol{x}=\boldsymbol{x}$, such that $\mathbf{B}\overline{\mathbf{D}}\boldsymbol{x}=\boldsymbol{x}$. To make (32) hold true, we must then ensure that $\mathbf{B}\overline{\mathbf{D}}\boldsymbol{x}\neq\boldsymbol{x}$ or, equivalently, recalling Lemma 2.2, $\overline{\mathbf{D}}\mathbf{B}\boldsymbol{x}\neq\boldsymbol{x}$. Since

$$\mathbf{B}\boldsymbol{x}=\boldsymbol{x}=\mathbf{DB}\boldsymbol{x}+\overline{\mathbf{D}}\mathbf{B}\boldsymbol{x}, \tag{40}$$

we need to guarantee that $\mathbf{DB}\boldsymbol{x}\neq\mathbf{0}$. Proceeding as in Section II, let us define now the $|\mathcal{S}|\times|\mathcal{F}|$ matrix $\mathbf{G}$ as

$$\mathbf{G}=\begin{pmatrix} u_{i_1}(j_1) & u_{i_2}(j_1) & \cdots & u_{i_{|\mathcal{F}|}}(j_1)\\ \vdots & \vdots & \vdots & \vdots\\ u_{i_1}(j_{|\mathcal{S}|}) & u_{i_2}(j_{|\mathcal{S}|}) & \cdots & u_{i_{|\mathcal{F}|}}(j_{|\mathcal{S}|})\end{pmatrix}$$

whose ℓ-th column is the eigenvector of index i_ℓ of the Laplacian matrix (or any orthonormal set of basis vectors), sampled at the positions indicated by the indices $j_1,\dots,j_{|\mathcal{S}|}$. Condition (32) is equivalent to require $\mathbf{G}$ to be full column rank.

Of course, a necessary condition for the existence of a non trivial vector $\boldsymbol{x}$ satisfying $\mathbf{DB}\boldsymbol{x}\neq\mathbf{0}$, and then enabling sampling theorem, is that $|\mathcal{S}|\geq|\mathcal{F}|$. However, this condition is not sufficient, because $\mathbf{G}$ may loose rank, depending on graph topology and samples' location. As an extreme case, if the graph is not connected, the vertices can be labeled so that the Laplacian (adjacency) matrix can be written as a block diagonal matrix, with a number of blocks equal to the number of connected components. Correspondingly, each eigenvector of $\mathbf{L}$ (or $\mathbf{A}$) can be expressed as a vector having all zero elements, except the entries corresponding to the connected component, which that eigenvector is associated to. This implies that, if there are no samples over the vertices corresponding to the non-null entries of the eigenvectors with index included in $\mathcal{F}$, $\mathbf{G}$ looses rank. In principle, a signal defined over a disconnected graph can still be reconstructed from its samples, but only provided that the number of samples belonging to each connected component is at least equal to the number of eigenvectors with indices in $\mathcal{F}$ associated to that component. More generally, even if the graph is connected, there may easily occur situations where matrix $\mathbf{G}$ is not rank-deficient, but it is ill-conditioned, depending on graph topology and samples' location. This suggests that the location of samples plays a key role in the performance of the reconstruction algorithm. For this reason, In Section VI we will suggest and compare a few alternative sampling strategies satisfying different optimization criteria.

Frame-based reconstruction: The problem of sampling on graphs using Dirac frames was initially studied by [12], [22], where the conditions for the existence of such frames were derived. Here we approach the problem using the above developed theory of maximally vertex-frequency concentrated signals on graph.

Definition [23] A set of elements $\{\boldsymbol{g}_i\}_{i\in\mathcal{I}}$, is a frame for the Hilbert space $\mathcal{H}$, if for all $\boldsymbol{f}\in\mathcal{H}$ there exist constants $0<A\leq B<\infty$ such that

$$A\|\boldsymbol{f}\|_2^2\leq\sum_{i\in\mathcal{I}}|\langle \boldsymbol{f},\boldsymbol{g}_i\rangle|^2\leq B\|\boldsymbol{f}\|_2^2. \tag{41}$$

Constants A and B are called frame bounds, while the largest A and the smallest B are called the tightest frame bounds. It is useful to note that condition (41) guarantees the frame operator $\mathbf{T}$ to be bounded and invertible.

Definition Given a frame $\{\boldsymbol{g}_i\}_{i\in\mathcal{I}}$, the linear operator $\mathbf{T}: \mathcal{H}\to\mathcal{H}$ defined as

$$\mathbf{T}\boldsymbol{f} = \sum_{i\in\mathcal{I}} \langle \boldsymbol{f}, \boldsymbol{g}_i\rangle \boldsymbol{g}_i \tag{42}$$

is called the frame operator.

Now, introducing the canonical basis vector $\boldsymbol{\delta}_u$, with $u\in\mathcal{V}$, i.e. having all zero entries except the u-th entry equal to 1, we investigate under what conditions a set of vectors $\{\mathbf{B}\boldsymbol{\delta}_u\}_{u\in\mathcal{S}}$ constitutes a frame for $\mathcal{B}$. The frame operator in this case is

$$\mathbf{T}_\delta \boldsymbol{f} = \sum_{u\in\mathcal{S}} \langle \boldsymbol{f}, \mathbf{B}\boldsymbol{\delta}_u\rangle \mathbf{B}\boldsymbol{\delta}_u = \sum_{u\in\mathcal{S}} f(u)\mathbf{B}\boldsymbol{\delta}_u. \tag{43}$$

First, we observe that the frame operator $\mathbf{T}_\delta$ may be also expressed as

$$\mathbf{T}_\delta = \mathbf{BDDB} = \mathbf{BDB}. \tag{44}$$

Operator $\mathbf{T}_\delta$ has a spectral norm $\|\mathbf{T}_\delta\|_2$ equal to $\sigma^2_{max}(\mathbf{BD})$. Hence, to guarantee that $\{\mathbf{B}\boldsymbol{\delta}_u\}_{u\in\mathcal{S}}$ is a frame, it is sufficient to check when $\mathbf{T}_\delta$ is invertible, for any $\boldsymbol{f}\in\mathcal{B}$. The operator $\mathbf{BD}$, on its turn, is invertible for any $\boldsymbol{f}\in\mathcal{B}$ if its singular vectors, not belonging to its kernel, constitute a basis for the $|\mathcal{F}|$-dimensional space $\mathcal{B}$, or, formally, if

$$\operatorname{rank}\mathbf{BDB} = \operatorname{rank}\mathbf{B}. \tag{45}$$

Taking into account (35) and Lemma 2.2, we conclude that the condition for a frame-based reconstruction based on a canonical-vector frames coincides with the condition of Theorem 4.1.

In general, however, the reconstruction based on the canonical-vector frame may be non robust in the presence of observation noise. For this reason, we generalize now the sampling frame operator $\mathbf{T}_\delta$ by introducing the operator $\mathbf{T}_Y$ as

$$\mathbf{T}_Y \boldsymbol{f} = \mathbf{BYDB}\boldsymbol{f} = \sum_{u\in\mathcal{S}} f(u)\boldsymbol{y}_u, \tag{46}$$

where $\mathbf{Y}$ is a bounded matrix whose columns $\boldsymbol{y}_i$, without loss of generality, can be taken as band-limited in $\mathcal{B}$, i.e. $\mathbf{B}\boldsymbol{y}_i = \boldsymbol{y}_i$, so that the image of $\mathbf{Y}$ is also bandlimited. Let us consider now the reconstruction of $\boldsymbol{f}\in\mathcal{B}$ from its samples on $\mathcal{S}$, based on $\mathbf{T}_Y$. This requires checking under what conditions the operator $\mathbf{T}_Y$ is bounded and invertible. Since the columns of $\mathbf{YD}$ corresponding to indices that do not belong to the set $\mathcal{S}$ are null, we can limit our attention to matrices $\mathbf{Y}$ that are invariant to the right-side multiplication by $\mathbf{D}$, i.e. $\mathbf{YD}=\mathbf{Y}$. Finally, we arrive at the following sampling theorem.

Theorem 4.3: Let $\mathcal{F}\subseteq\mathcal{V}^*$ be the set of frequencies and $\mathcal{S}\subseteq\mathcal{V}$ be the sampling set of vertices and let $\mathbf{Y}\in\mathbb{C}^{N\times N}: \mathcal{B}\to L_2(\mathcal{G})$ be an arbitrary bounded operator. If

$$\operatorname{rank}\mathbf{BYDB} = \operatorname{rank}\mathbf{B} \tag{47}$$

then $\{\mathbf{B}\boldsymbol{y}_i\}_{i\in\mathcal{S}}$ is a frame for $\mathcal{B}$.

Proof: The proof follows directly from the invertibility conditions for the operator $\mathbf{BYDB}$. ■ ■

The tightest frame bounds, according to the Rayleigh-Ritz theorem, are defined by the minimum and maximum singular values of $\mathbf{BYDB}$

$$\sigma_{min}\|\boldsymbol{f}\|_2^2 \le \sum_{u\in S} |\langle \boldsymbol{f}, \boldsymbol{y}_u\rangle|^2 \le \sigma_{max}\|\boldsymbol{f}\|_2^2, \tag{48}$$

which is valid for every $\boldsymbol{f}\in\mathcal{B}$. As an example of matrix $\mathbf{Y}$, encompassing the approaches proposed in [14] and [24], we have the following frame operator

$$\mathbf{T}_1\boldsymbol{f} = \mathbf{BY}_1\mathbf{DB}\boldsymbol{f} = \sum_{u\in\mathcal{S}} f(u)\mathbf{B}\boldsymbol{\delta}_{\mathcal{N}(u)}, \tag{49}$$

where $\boldsymbol{\delta}_{\mathcal{N}(u)}$ is the characteristic function of set $\mathcal{N}(u)$, defined as $\delta_{\mathcal{N}(u)}(v)=1$, if $v\in\mathcal{N}(u)$, and zero otherwise. In this case, the graph signal is supposed to be sampled sparsely in such a way that around each sampled vertex there is a non-empty neighborhood of vertices $\mathcal{N}(u)$ that altogether could cover the whole graph. However, this choice is not necessarily the best one. In Section VI we will provide numerical results showing how the generalized frame-based approach can yield better performance results in the presence of observation noise.

V. Reconstruction from noisy observations

Let us consider now the reconstruction of band-limited signals from noisy samples, where the observation model is

$$\boldsymbol{r} = \mathbf{D}(\boldsymbol{s}+\boldsymbol{n}), \tag{50}$$

where $\boldsymbol{n}$ is a noise vector. Applying (36) to $\boldsymbol{r}$, the reconstructed signal $\tilde{\boldsymbol{s}}$ is

$$\tilde{\boldsymbol{s}} = \sum_{i=1}^{|\mathcal{F}|} \frac{1}{\sigma_i^2}\langle \mathbf{D}\boldsymbol{s}, \boldsymbol{\psi}_i\rangle\boldsymbol{\psi}_i + \sum_{i=1}^{|\mathcal{F}|} \frac{1}{\sigma_i^2}\langle \mathbf{D}\boldsymbol{n}, \boldsymbol{\psi}_i\rangle\boldsymbol{\psi}_i. \tag{51}$$

Exploiting the orthonormality of $\boldsymbol{\psi}_i$, the mean square error is

$$\begin{aligned} MSE &= \mathbb{E}\left\{\|\tilde{\boldsymbol{s}}-\boldsymbol{s}\|_2^2\right\} = \mathbb{E}\left\{\sum_{i=1}^{|\mathcal{F}|}\frac{1}{\sigma_i^4}|\langle \mathbf{D}\boldsymbol{n}, \boldsymbol{\psi}_i\rangle|^2\right\} \\ &= \sum_{i=1}^{|\mathcal{F}|}\frac{1}{\sigma_i^4}\boldsymbol{\psi}_i^*\mathbf{D}\mathbb{E}\{\boldsymbol{n}\boldsymbol{n}^*\}\mathbf{D}\boldsymbol{\psi}_i. \end{aligned} \tag{52}$$

In case of identically distributed uncorrelated noise, i.e. $\mathbb{E}\{\boldsymbol{n}\boldsymbol{n}^*\} = \beta_n^2\mathbf{I}$, using (21), we get

$$\begin{aligned} MSE_G &= \sum_{i=1}^{|\mathcal{F}|}\frac{\beta_n^2}{\sigma_i^4}\operatorname{tr}(\mathbf{D}\boldsymbol{\psi}_i\boldsymbol{\psi}_i^*\mathbf{D}) \\ &= \sum_{i=1}^{|\mathcal{F}|}\frac{\beta_n^2}{\sigma_i^4}\operatorname{tr}(\boldsymbol{\psi}_i^*\mathbf{D}\boldsymbol{\psi}_i) = \beta_n^2\sum_{i=1}^{|\mathcal{F}|}\frac{1}{\sigma_i^2}. \end{aligned} \tag{53}$$

Since the non-null singular values of the Moore-Penrose left pseudo-inverse $(\mathbf{BD})^+$ are the inverses of singular values of $\mathbf{BD}$, i.e. $\lambda_i\left((\mathbf{BDB})^+\right) = \lambda_i^{-1}(\mathbf{BDB})$, (53) can be rewritten as

$$MSE_G = \beta_n^2\,\|(\mathbf{BDB})^+\|_F. \tag{54}$$

Proceeding exactly in the same way, the mean square error for the frame-based sampling scheme (46) is

$$MSE_F = \beta_n^2 \left\| \left(\mathbf{BYDB}\right)^+ \right\|_F. \tag{55}$$

Based on previous formulas, a possible optimal sampling strategy consists in selecting the vertices that minimize (54) or (55). This aspect will be analyzed in Section VI.

A. ℓ_1-norm reconstruction

Let us consider now a different observation model, where the signal is observed everywhere, but a subset of nodes $\mathcal{S}$ is strongly corrupted by noise, i.e.

$$\boldsymbol{r} = \boldsymbol{s} + \mathbf{D}\boldsymbol{n}, \tag{56}$$

where the noise is arbitrary but bounded, i.e., $\|\boldsymbol{n}\|_1 < \infty$. This model was considered in [25] and it is relevant, for example, in sensor networks, where a subset of sensors can be damaged or highly interfered. The problem in this case is whether it is possible to recover the signal $\boldsymbol{s}$ exactly, i.e. irrespective of noise. Even though this is not a sampling problem, the solution is still related to sampling theory. Clearly, if the signal $\boldsymbol{s}$ is bandlimited and if the indices of the noisy observations are known, the answer is simple: $\boldsymbol{s} \in \mathcal{B}$ can be perfectly recovered from the noisy-free observations, i.e. by completely discarding the noisy observations, if the sampling theorem condition (32) holds true. But of course, the challenging situation occurs when the location of the noisy observations is not known. In such a case, we may resort to an ℓ_1-norm minimization, by formulating the problem as follows

$$\tilde{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}' \in \mathcal{B}} \|\boldsymbol{r} - \boldsymbol{s}'\|_1. \tag{57}$$

We will show next under what assumptions it is still possible to recover a bandlimited signal *perfectly*, even without knowing exactly the position of the corrupted observations.

To start with, the following lemma provides a sufficient condition for the convergence of (57).

Lemma 5.1: For any $\boldsymbol{g} \in \mathcal{B}$, given the observation model (56), if

$$\|\mathbf{D}\boldsymbol{g}\|_1 < \|\overline{\mathbf{D}}\boldsymbol{g}\|_1, \tag{58}$$

then the ℓ_1-reconstruction algorithm (57) is able to recover $\boldsymbol{g}$ perfectly.

Proof: To prove this, we show first that for a signal consisting of noise only, i.e. $\boldsymbol{r} = \mathbf{D}\boldsymbol{n}$, the best bandlimited ℓ_1-norm approximation $\boldsymbol{g}$ is the zero vector. In fact,

$$\begin{aligned} \|\mathbf{D}\boldsymbol{n} - \boldsymbol{g}\|_1 &= \|\mathbf{D}\left(\boldsymbol{n} - \boldsymbol{g}\right)\|_1 + \|\overline{\mathbf{D}}\boldsymbol{g}\|_1 \\ &\geq \|\mathbf{D}\boldsymbol{n}\|_1 - \|\mathbf{D}\boldsymbol{g}\|_1 + \|\overline{\mathbf{D}}\boldsymbol{g}\|_1 \\ &> \|\mathbf{D}\boldsymbol{n}\|_1. \end{aligned} \tag{59}$$

Now, suppose instead that $\boldsymbol{s} \neq \mathbf{0}$. We can observe that

$$\|\boldsymbol{r} - \boldsymbol{g}\|_1 = \|\boldsymbol{s} + \mathbf{D}\boldsymbol{n} - \boldsymbol{g}\|_1 = \|\mathbf{D}\boldsymbol{n} + \left(\boldsymbol{s} - \boldsymbol{g}\right)\|_1. \tag{60}$$

Since we proved before that, under (58), the best bandlimited approximation of $\mathbf{D}\boldsymbol{n}$ is the null vector, (60) is minimized by the vector $\boldsymbol{g} = \boldsymbol{s}$. ■ ■

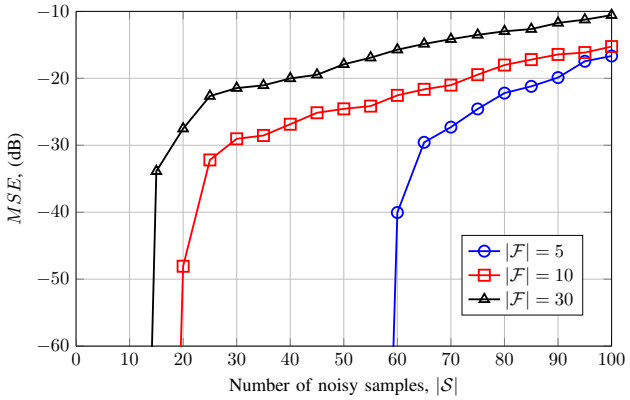

Fig. 3: Behavior of Mean Squared Error versus number of noisy samples, for different signal bandwidths.

From the previous lemma it is hard to say if for a given $\mathcal{S}$ and $\mathcal{F}$ condition (58) will hold or not. Next theorem provides such a condition.

Theorem 5.2: For any $\boldsymbol{g} \in \mathcal{B}$, given the observation model (56), if

$$\max_{j \in \mathcal{F}} \sum_i \left| \left(\mathbf{DB}\right)_{ij} \right| < \min_{j \in \mathcal{F}} \sum_i \left| \left(\overline{\mathbf{D}}\mathbf{B}\right)_{ij} \right|, \tag{61}$$

then the ℓ_1-reconstruction method (57) recovers the signal perfectly, i.e. $\tilde{\boldsymbol{s}} = \boldsymbol{s}$.

Proof: Since

$$\sup_{\substack{\boldsymbol{g} \in \mathcal{B} \\ \|\boldsymbol{g}\|_1 = 1}} \|\mathbf{DB}\boldsymbol{g}\|_1 = \max_{j \in \mathcal{F}} \sum_i \left| \left(\mathbf{DB}\right)_{ij} \right| \tag{62}$$

and

$$\inf_{\substack{\boldsymbol{g} \in \mathcal{B} \\ \|\boldsymbol{g}\|_1 = 1}} \|\overline{\mathbf{D}}\mathbf{B}\boldsymbol{g}\|_1 = \min_{j \in \mathcal{F}} \sum_i \left| \left(\overline{\mathbf{D}}\mathbf{B}\right)_{ij} \right|, \tag{63}$$

if (61) holds true, then

$$\sup_{\boldsymbol{g} \in \mathcal{B}} \frac{\|\mathbf{DB}\boldsymbol{g}\|_1}{\|\boldsymbol{g}\|_1} < \inf_{\boldsymbol{g} \in \mathcal{B}} \frac{\|\overline{\mathbf{D}}\mathbf{B}\boldsymbol{g}\|_1}{\|\boldsymbol{g}\|_1}. \tag{64}$$

As a consequence, for every bandlimited signal $\boldsymbol{g}$, (61) implies (58) and then, by Lemma 5.1, it guarantees perfect recovery. ■ ■

Besides establishing perfect recovery conditions, Theorem 5.2 provides hints on how to select the vertices to be discarded still enabling perfect reconstruction of a bandlimited signal through an ℓ_1-norm reconstruction.

An example of ℓ_1 reconstruction based on (57) is useful to grasp some interesting features. We consider a graph composed of 100 nodes connected by a scale-free topology. The signal is assumed to be band limited, with a spectral content limited to the first $|\mathcal{F}|$ eigenvectors of the Laplacian matrix. In Fig. 3, we report the behavior of the MSE associated to the ℓ_1-norm estimate in (57), versus the number of noisy samples, considering different values of bandwidth $|\mathcal{F}|$. As we can notice from Fig. 3, for any value of $|\mathcal{F}|$, there exists a threshold value such that, if the number of noisy samples is lower than the threshold, the reconstruction of the signal is error free. As expected, a smaller signal bandwidth allows perfect reconstruction with a larger number of noisy samples.

We provide next some theoretical bounds on the cardinality of $\mathcal{S}$ and $\mathcal{F}$ enabling ℓ_1-norm recovery. To this purpose, we start proving the following lemma.

Lemma 5.3: It holds true that

$$\sup_{\boldsymbol{f}\in\mathcal{B}} \frac{\|\mathbf{D}\boldsymbol{f}\|_1}{\|\boldsymbol{f}\|_1} \leq \mu^2 \left|\mathcal{S}\right| \left|\mathcal{F}\right|, \tag{65}$$

where μ is defined as

$$\mu := \max_{\substack{j\in\mathcal{V}^*\\ i\in\mathcal{V}}} \left|u_j(i)\right|. \tag{66}$$

Proof: Let us consider the expansion formula for $\boldsymbol{f}\in\mathcal{B}$

$$f(k) = \sum_{j\in\mathcal{F}} u_j(k) \sum_{i\in\mathcal{V}} f(i) u_j^*(i) = \sum_{i\in\mathcal{V}} f(i) \sum_{j\in\mathcal{F}} u_j(k) u_j^*(i) \tag{67}$$

which yields

$$|f(k)| \leq \|\boldsymbol{f}\|_\infty \leq \sum_{i\in\mathcal{V}} |f(i)| \sum_{j\in\mathcal{F}} \mu^2 = \mu^2 \left|\mathcal{F}\right| \|\boldsymbol{f}\|_1. \tag{68}$$

or

$$\|\boldsymbol{f}\|_1 \geq \frac{\|\boldsymbol{f}\|_\infty}{\mu^2 \left|\mathcal{F}\right|}. \tag{69}$$

By combining

$$\|\mathbf{D}\boldsymbol{f}\|_1 \leq \|\boldsymbol{f}\|_\infty \left|\mathcal{S}\right| \tag{70}$$

with (69), we come to

$$\frac{\|\mathbf{D}\boldsymbol{f}\|_1}{\|\boldsymbol{f}\|_1} \leq \mu^2 \left|\mathcal{S}\right| \left|\mathcal{F}\right|. \tag{71}$$

■ ■

Theorem 5.4 (ℓ_1-uncertainty): Let $\boldsymbol{f}$ be a unit norm signal α_1-concentrated to the set of vertices $\mathcal{S}$, i.e. $\|\mathbf{D}\boldsymbol{f}\|_1 \geq \alpha_1$ and β_1-bandlimited to the set of frequencies $\mathcal{F}$, i.e. $\|\mathbf{B}\boldsymbol{f}\|_1 \geq \beta_1$, then

$$\left|\mathcal{S}\right| \left|\mathcal{F}\right| \geq \frac{(\alpha_1+\beta_1-1)}{\mu^2\,(2-\beta_1)}. \tag{72}$$

Proof: If $\|\mathbf{B}\boldsymbol{f}\|_1 \geq \beta_1$, then by definition there exists a $\boldsymbol{g}\in\mathcal{B}$ such that $\|\boldsymbol{g}-\boldsymbol{f}\|_1 \leq 1-\beta_1$ and, for this $\boldsymbol{g}$, we can write

$$\|\mathbf{D}\boldsymbol{g}\|_1 \geq \|\mathbf{D}\boldsymbol{f}\|_1 - \|\mathbf{D}\left(\boldsymbol{g}-\boldsymbol{f}\right)\|_1 \geq \|\mathbf{D}\boldsymbol{f}\|_1 - 1 + \beta_1 \tag{73}$$

and

$$\|\boldsymbol{g}\|_1 \leq \|\boldsymbol{f}\|_1 + 1 - \beta_1. \tag{74}$$

Therefore

$$\frac{\|\mathbf{D}\boldsymbol{g}\|_1}{\|\boldsymbol{g}\|_1} \geq \frac{\|\mathbf{D}\boldsymbol{f}\|_1 - 1 + \beta_1}{\|\boldsymbol{f}\|_1 + 1 - \beta_1} \geq \frac{\alpha_1+\beta_1-1}{2-\beta_1}. \tag{75}$$

Combining this result with the results of Lemma 5.3, we finally get (72). ■ ■

It is worth noting that an ℓ_2-uncertainty principle analogous to Theorem 5.4 may also be easily derived. Finally, we provide the condition for perfect reconstruction using (57) in the case when $\mathcal{S}$ and $\mathcal{F}$ are not known.

Theorem 5.5: Defining

$$\mu := \max_{\substack{j\in\mathcal{V}^*\\ i\in\mathcal{V}}} \left|u_j(i)\right|, \tag{76}$$

if for some *unknown* $\mathcal{F}$ and $\mathcal{S}$, we have

$$\left|\mathcal{S}\right| \left|\mathcal{F}\right| < \frac{1}{2\mu^2}, \tag{77}$$

then the ℓ_1-norm reconstruction method (57) recovers $\boldsymbol{s}\in\mathcal{B}$ perfectly, i.e. $\tilde{\boldsymbol{s}} = \boldsymbol{s}$, for any arbitrary noise $\boldsymbol{n}$ present on at most $\left|\mathcal{S}\right|$ vertices.

Proof: For a band-limited signal $\boldsymbol{g}\in\mathcal{B}$ satisfying (58), we can also write

$$\frac{\|\mathbf{D}\boldsymbol{g}\|_1}{\|\boldsymbol{g}\|_1} < \frac{1}{2}. \tag{78}$$

On the other hand, from Lemma 5.3 we know that the supremum of the previous ratio among all $\boldsymbol{g}\in\mathcal{B}$ is upper bounded by $\mu^2\left|\mathcal{S}\right|\left|\mathcal{F}\right|$. Hence, by Lemma 5.1, all band-limited signals satisfying (77) satisfy also condition (78) or, equivalently (58), for perfect ℓ_1-norm recovery. ■ ■

VI. Sampling strategies

When sampling graph signals, besides choosing the right number of samples, whenever possible it is also fundamental to have a strategy indicating *where* to sample, as the samples' location plays a key role in the performance of reconstruction algorithms. Building on the analysis of signal reconstruction algorithms in the presence of noise carried out in Section V, a possible strategy is to select the location in order to minimize the MSE. From (54), taking into account that

$$\lambda_i\left(\mathbf{BDB}\right) = \sigma_i^2\left(\mathbf{BD}\right) = \sigma_i^2\left(\mathbf{\Sigma U}^*\mathbf{D}\right), \tag{79}$$

the problem is equivalent to selecting the right columns of the matrix $\mathbf{\Sigma U}^*$ in order to minimize the Frobenius norm of the pseudo-inverse $\left(\mathbf{\Sigma U}^*\mathbf{D}\right)^+$. This problem is combinatorial and NP-hard. Nevertheless, the problem of choosing columns from a matrix so as to minimize the Frobenius norm of its pseudo-inverse was studied for example in [26], so that we can take advantage of these methods for our purposes. Next we provide a few alternative strategies for selecting the samples' locations.

1) Greedy Selection - Minimization of Frobenius norm of $\left(\mathbf{\Sigma U}^*\mathbf{D}\right)^+$: This strategy aims at minimizing the MSE in (53). The method selects the columns of the matrix $\mathbf{\Sigma U}^*$ so that the Frobenius norm of the pseudo-inverse of the resulting matrix is minimized. In case of uncorrelated noise, this is equivalent to minimizing $\sum_{i=1}^{|\mathcal{F}|} 1/\sigma_i^2$. We propose a greedy approach to tackle this selection problem. The resulting sampling strategy is summarized in Algorithm 1. Note that $\mathcal{S}$ is the sampling set, indicating which columns to select, $\tilde{\mathbf{U}}$ denotes the first $|\mathcal{F}|$ rows of $\mathbf{U}^*$, and the symbol $\tilde{\mathbf{U}}_{\mathcal{A}}$ denotes the matrix formed with the columns of $\tilde{\mathbf{U}}$ belonging to set $\mathcal{A}$.

2) Maximization of the Frobenius norm of $\mathbf{\Sigma U}^*\mathbf{D}$: The second strategy aims at selecting the columns of the matrix $\tilde{\mathbf{U}}$ in order to maximize its Frobenius norm. Even if this strategy is not directly related to the optimization of the MSE in (53), it leads to a very easy implementation that shows

Algorithm 1 : Greedy selection based on minimum Frobenius norm of $(\mathbf{\Sigma U}^*\mathbf{D})^+$

Input Data : $\tilde{\mathbf{U}}$, the first $|\mathcal{F}|$ rows of $\mathbf{U}^*$;
M, the number of samples.
Output Data : $\mathcal{S}$, the sampling set.
Function : initialize $\mathcal{S} \equiv \emptyset$
while $|\mathcal{S}| < M$
$s = \arg\min_j \sum_{i=1}^{|\mathcal{F}|} \frac{1}{\sigma_i^2(\tilde{\mathbf{U}}_{\mathcal{S}\cup\{j\}})}$;
$\mathcal{S} \leftarrow \mathcal{S} \cup \{s\}$;
end

good performance in practice, as we will see in the sequel. In particular, since we have

$$\max_{\mathcal{S}} \|\tilde{\mathbf{U}}\mathbf{D}\|_2^2 = \max_{\mathcal{S}} \sum_{i\in\mathcal{S}} \|(\tilde{\mathbf{U}})_i\|_2^2, \tag{80}$$

the optimal selection strategy simply consists in selecting the M columns from $\tilde{\mathbf{U}}$ with largest ℓ_2-norm.

3) Greedy Selection - Maximization of the vector products norm: The previous strategy is very easy to implement, but it considers only the norms of the columns of $\tilde{\mathbf{U}}$. Intuitively, if we take into account also the angle between the selected vectors, we should get some performance improvement. To incorporate this further feature in the selection, we propose a greedy approach, whose main steps are listed in Algorithm 2. The algorithm starts including the column with the largest norm in $\tilde{\mathbf{U}}$, and then adds iteratively the columns having the largest norms and, at the same time, are as orthogonal as possible to the vectors already contained in $\mathcal{S}$.

Algorithm 2 : Greedy selection based on maximum norm of vector products

Input Data : $\tilde{\mathbf{U}}$, the first $|\mathcal{F}|$ rows of $\mathbf{U}^*$;
M, the number of samples;
$\mathcal{S} = \{s_{\max}\}$, where $s_{\max}$ is the index of the column with largest norm in $\tilde{\mathbf{U}}$;
Output Data : $\mathcal{S}$, the sampling set.
Function : initialize $\mathcal{S} \equiv \emptyset$
while $|\mathcal{S}| < M - 1$
for all $j \notin \mathcal{S}$
Compute $v_j(s) = \tilde{\mathbf{U}}_{\{j\}}^* \tilde{\mathbf{U}}_{\{s\}}$, $\forall s \in \mathcal{S}$;
Set $V_j = \|\tilde{\mathbf{U}}_{\{j\}}\|_2 \cdot \prod_{s=1}^{|S|} \sqrt{1 - v_j(s)^2}$;
end
$\hat{s} = \arg\min_j V_j$;
$S \leftarrow \mathcal{S} \cup \{\hat{s}\}$;
end

Comparison of sampling strategies: We compare now the performance obtained with the proposed sampling strategies, with random sampling and with the strategy proposed in [5] maximizing the minimum singular value of $\mathbf{\Sigma U}^*\mathbf{D}$. We consider two random graph models: scale-free and random geometric graphs. The corresponding results are shown in Fig. 4 and 5, respectively, reporting the normalized MSE, defined as the mean square error per node, divided by the noise variance, versus the number of samples. In each case, we consider band-limited signals with two different bandwidth $|\mathcal{F}| = 5$ and $|\mathcal{F}| = 10$. The number of nodes is 20. The observation model is (50) where the additive noise is generated as a unit variance, uncorrelated, zero mean Gaussian random vector. The results shown in the figures have been obtained by averaging over 100 independent realizations of graph topology and noise. We compare five different sampling strategies, namely: (i) the random strategy, which picks nodes randomly; (ii) the greedy selection method of Algorithm 1, minimizing the Frobenius norm of $(\mathbf{\Sigma U}^*\mathbf{D})^+$ (MinPinv); (iii) the Max Frobenius norm (MaxFro) strategy; (iv) the greedy selection method of Algorithm 2, maximizing the norm of the vector products (MaxVecProd); and (v) the greedy algorithm (MaxSigMin) maximizing the minimum singular value of $\mathbf{\Sigma U}^*\mathbf{D}$, recently proposed in [5]. The behavior of the globally optimal strategy obtained through an exhaustive search over all possible selections is also reported as a benchmark. From Figs. 4 and 5 we can observe that, as expected, as the number of samples increases, the mean squared error decreases. As a general remark, we can notice how random sampling can perform quite poorly. This shows that, when sampling a graph signal, what matters is not only the number of samples, but also (and most important) *where* the samples are taken. Our greedy MinPinv strategy shows the best performance among all other methods, also because it is matched to the performance metric (MSE). Even if sub-optimal, MinPinv shows also performance very close to the optimal combinatorial benchmark, at least for sufficiently large number of samples. Interestingly, the MaxFro method, in spite of its simplicity, shows quite good performance, at least for sufficient number of samples. It is also interesting to observe how, increasing the signal bandwidth, see, e.g. Fig. 4 (b), the (MaxVecProd) shows a behavior very close to the globally optimal strategy. In this last case, also the MaxFro strategy performs very well. Comparing Figs. 4 and 5, we observe qualitatively similar results with perhaps the only noticeable difference that in the largest bandwidth case, i.e. $|\mathcal{F}| = 10$, for the random geometric case it is better to take a number of samples slightly larger than the minimum value to achieve good performance. The intuitive explanation of this behavior is the worse connectivity features of the RGG model as opposed to the scale-free graph, for the given parameter settings.

The next example shows how to improve robustness to noise by using the frame-based reconstruction method. In (49) we provided a possible choice of frame operator to be used for sampling. In the following, we show how the mean square error MSE_F in (55) behaves for different choices of graph coverings by the sets $\mathcal{N}(v)$. For this example, we consider a random geometric graph having 100 nodes with connectivity radius r_0. The graph is sampled randomly and, around each sample, taken at vertex v, the local set $\mathcal{N}(v)$ is composed of the nodes falling inside a ball of radius r_1 centered on v. The local sets associated to each sample can intersect each other and their union does not necessarily cover the whole graph. In Fig. 6, we show the MSE as a function of the covering

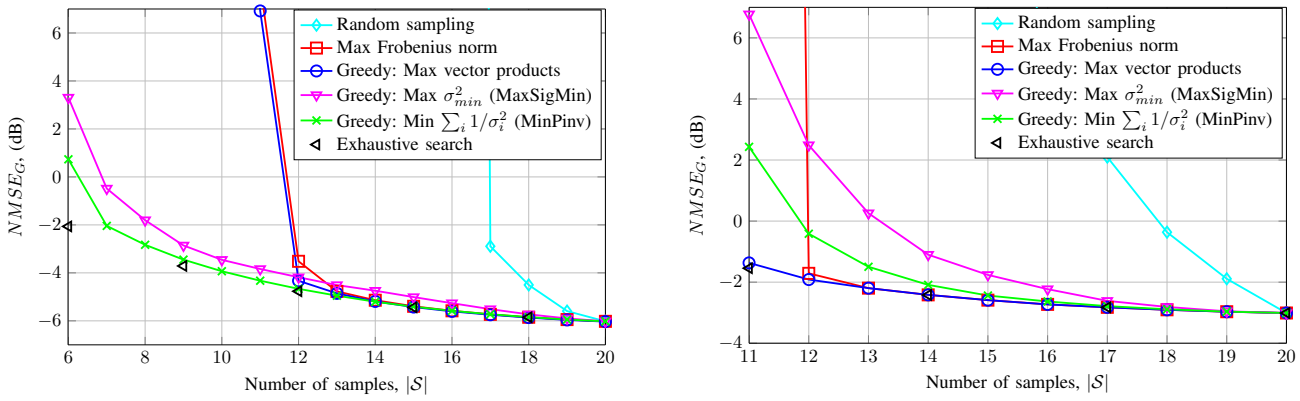

Fig. 4: Behavior of Mean Squared Error versus number of samples, for different sampling strategies for a scale-free topology. (a) for the case of $|\mathcal{F}| = 5$, whereas (b) for the case $|\mathcal{F}| = 10$.

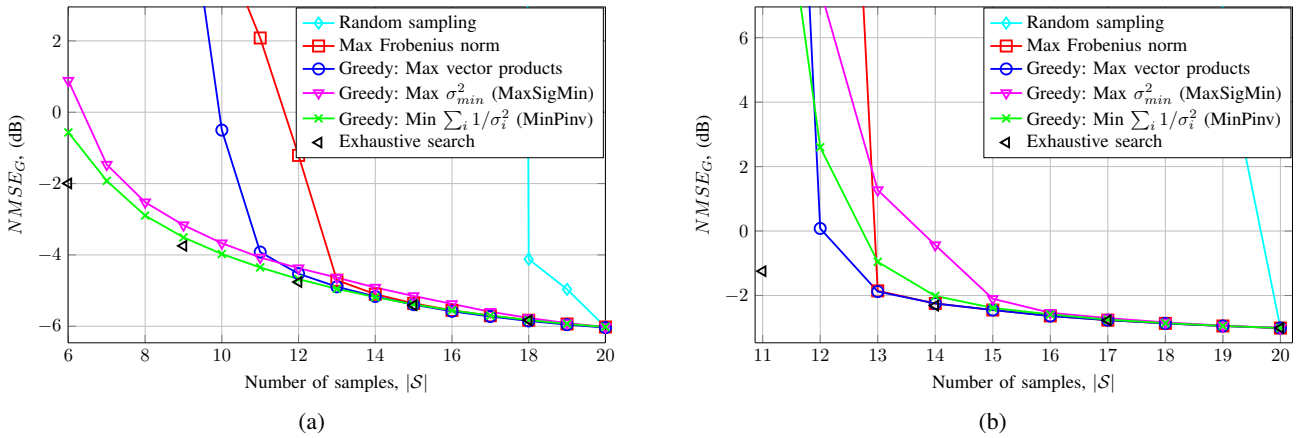

Fig. 5: Behavior of Mean Squared Error versus number of samples, for different sampling strategies on an RGG. (a) considers $|\mathcal{F}| = 5$, whereas (b) considers $|\mathcal{F}| = 10$.

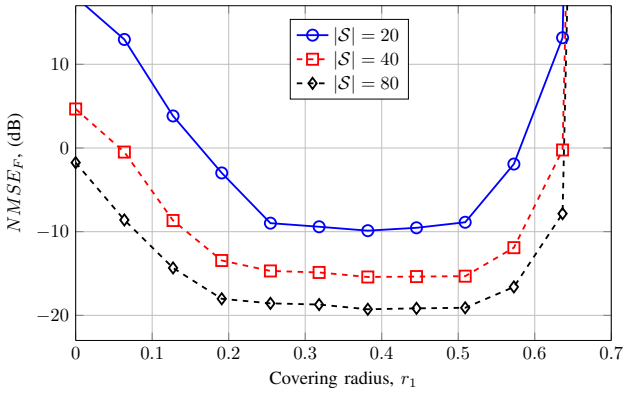

Fig. 6: Dependence of mean square error MSE_F on the covering radius.

radius r_1. We can see from Fig. 6 that there exists an optimal size of covering local-sets which minimizes the mean square error. Furthermore, we can see how, increasing the number of samples, for a given bandwidth, the MSE decreases.

VII. Conclusion

In this paper we have derived an uncertainty principle suitable for graph signals, showing all the admissible pairs of energy concentration in the vertex and frequency domain. Even if in the paper we referred to the recently defined Graph Fourier Transform based on the projection of the signal onto the subspace spanned by the eigenvectors of the Laplacian (or adjacency) matrix, the theory has more general validity as the only assumption is the choice of a basis of orthonormal vectors forming the matrix $\mathbf{U}$ in (3). From the graph perspective, an appealing basis is the one that enables a parsimonious, or sparse, representation of the observed signal. Furthermore, as suggested in the Introduction, even if the signal was defined over the vertices of the graph, the theory can be directly extended to signals defined over subsets of vertices, e.g., edges, triplets and so on. In building the fundamental blocks leading to the uncertainty principle, we derived the class of maximally concentrated graph signals, which may form the basis for a joint vertex-frequency analysis over graphs. One of the interesting results of the paper is the establishment of a relation between uncertainty principle and sampling theorem. This link provided also suggestions on how to identify sampling strategy robust against additive observation noise. It has been shown in fact that, in sampling graph signals, the location of the samples plays a key role in the final performance. Interesting further developments include the extension to hypergraphs, the robustness analysis in the case of non perfectly band-limited signals and the identification of further robust recovery

algorithms.

APPENDIX A
PROOF OF THEOREM 3.1

Before proceeding to the proof, we introduce some useful notation and provide several results that will be used for proving Theorem 3.1. Using the usual definition of the scalar product $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}^* \boldsymbol{b}$, we can define the angle between two vectors $\theta(\boldsymbol{a}, \boldsymbol{b})$ as

$$\theta(\boldsymbol{a}, \boldsymbol{b}) = \cos^{-1} \frac{\Re\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\|\boldsymbol{a}\|_2 \|\boldsymbol{b}\|_2}. \tag{81}$$

By Schwartz inequality $\langle \boldsymbol{a}, \boldsymbol{b} \rangle \leq \|\boldsymbol{a}\|_2 \|\boldsymbol{b}\|_2$ and the fact that $|\Re\langle \boldsymbol{a}, \boldsymbol{b} \rangle| \leq |\langle \boldsymbol{a}, \boldsymbol{b} \rangle|$ it is clear that

$$-1 \leq \frac{\Re\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\|\boldsymbol{a}\|_2 \|\boldsymbol{b}\|_2} \leq 1$$

and $\theta(\boldsymbol{a}, \boldsymbol{b}) = 0$ only if $\boldsymbol{b} = const \cdot \boldsymbol{a}$, i.e. when two vectors are colinear. Now, let us consider two vectors $\boldsymbol{f} \in \mathcal{B}$ and $\boldsymbol{g} \in \mathcal{D}$. For the beginning let us consider a fixed function $\boldsymbol{f} \in \mathcal{B}$ and an arbitrary $\boldsymbol{g} \in \mathcal{D}$. In this case the following lemma gives us an achievable lower bound of $\theta(\boldsymbol{f}, \boldsymbol{g})$.

Lemma A.1: For a given $\boldsymbol{f} \in \mathcal{B}$ there exists

$$\inf_{\boldsymbol{g} \in \mathcal{D}} \theta(\boldsymbol{f}, \boldsymbol{g}) = \cos^{-1} \frac{\|\mathbf{D}\boldsymbol{f}\|_2}{\|\boldsymbol{f}\|_2}, \tag{82}$$

which is achieved by $\boldsymbol{g} = k\mathbf{D}\boldsymbol{f}$ for any $k > 0$.

Proof: For any $\boldsymbol{g} \in \mathcal{D}$ it holds

$$\Re\langle \boldsymbol{f}, \boldsymbol{g} \rangle \leq |\langle \boldsymbol{f}, \boldsymbol{g} \rangle| = |\langle \mathbf{D}\boldsymbol{f}, \boldsymbol{g} \rangle|$$

and

$$|\langle \mathbf{D}\boldsymbol{f}, \boldsymbol{g} \rangle| \leq \|\mathbf{D}\boldsymbol{f}\|_2 \cdot \|\boldsymbol{g}\|_2.$$

So we can write

$$\frac{\Re\langle \boldsymbol{f}, \boldsymbol{g} \rangle}{\|\boldsymbol{f}\|_2 \cdot \|\boldsymbol{g}\|_2} \leq \frac{\|\mathbf{D}\boldsymbol{f}\|_2}{\|\boldsymbol{f}\|_2} = \frac{\Re\langle \boldsymbol{f}, \mathbf{D}\boldsymbol{f} \rangle}{\|\mathbf{D}\boldsymbol{f}\|_2 \cdot \|\boldsymbol{f}\|_2}$$

Taking into account that $\cos\theta$ decreases monotonically in $[0, \pi]$, it follows that for any $\boldsymbol{g} \in \mathcal{D}$

$$\theta(\boldsymbol{f}, \boldsymbol{g}) \geq \theta(\boldsymbol{f}, \mathbf{D}\boldsymbol{f}),$$

with equality when $\boldsymbol{g}$ and $\mathbf{D}\boldsymbol{f}$ are proportional. ■ ■

If the quantity

$$\theta_{min} = \inf_{\substack{\boldsymbol{f} \in \mathcal{B} \\ \boldsymbol{g} \in \mathcal{D}}} \theta(\boldsymbol{f}, \boldsymbol{g}) \tag{83}$$

is assumed by some specific $\boldsymbol{f} \in \mathcal{B}$ and $\boldsymbol{g} \in \mathcal{D}$ then we will say that $\mathcal{B}$ and $\mathcal{D}$ form the minimum angle θ_{min}. We then provide the following

Theorem A.2: The minimum angle θ_{min} between $\mathcal{B}$ and $\mathcal{D}$ exists and equals to

$$\theta_{min} = \cos^{-1} \sigma_{max}(\mathbf{BD}), \tag{84}$$

and is achieved by $\boldsymbol{f} = \boldsymbol{\psi}_1$ and $\boldsymbol{g} = \mathbf{D}\boldsymbol{\psi}_1$, where $\boldsymbol{\psi}_1$ is an eigenvector of $\mathbf{BDB}$ corresponding to the eigenvalue $\sigma^2_{max}(\mathbf{BD})$.

Proof: Using the result of Lemma A.1 we can write

$$\inf_{\substack{\boldsymbol{f} \in \mathcal{B} \\ \boldsymbol{g} \in \mathcal{D}}} \theta(\boldsymbol{f}, \boldsymbol{g}) = \inf_{\boldsymbol{f} \in \mathcal{B}} \cos^{-1} \frac{\|\mathbf{D}\boldsymbol{f}\|_2}{\|\boldsymbol{f}\|_2} = \inf_{\boldsymbol{f} \in \mathcal{B}} \cos^{-1} \frac{|\langle \boldsymbol{f}, \mathbf{D}\boldsymbol{f} \rangle|}{\|\boldsymbol{f}\|_2},$$

where infimum on the left side is achieved if the infimum on the right side is achieved. Since $\cos\theta$ decreases monotonically in $[0, \pi]$, we can apply the result of Theorem 2.3), from which it follows that infumum is achieved by the eigenvector $\boldsymbol{\psi}_1$ of $\mathbf{BDB}$ corresponding to the maximum eigenvalue $\sigma^2_{max}(\mathbf{BD})$. Therefore we conclude that

$$\theta_{min} = \inf_{\boldsymbol{f} \in \mathcal{B}} \cos^{-1} \frac{|\langle \boldsymbol{f}, \mathbf{D}\boldsymbol{f} \rangle|}{\|\boldsymbol{f}\|_2} = \cos^{-1} \sigma_{max}(\mathbf{BD}).$$

■ ■

Notice that, under perfect localization conditions, i.e. Theorem 2.1, $\sigma_{max}(\mathbf{BD}) = 1$ and the minimum angle is 0, thus implying that there are some vectors which lie in both subspaces $\mathcal{B}$ and $\mathcal{D}$. Next, we derive, without loss of generality, which values of β are attainable for every choice of α, assuming unit norm vectors $\boldsymbol{f} \in L_2(\mathcal{G})$.

The case $\alpha = 1$ means that all the energy of signal is supported only on $\mathcal{S}$. According to (22) and Lemma 2.2 the minimally concentrated on $\mathcal{F}$ vector from $\mathcal{D}$ is the eigenvector of $\mathbf{D}\overline{\mathbf{B}}\mathbf{D}$, corresponding to the eigenvalue $\sigma^2_{max}(\mathbf{D}\overline{\mathbf{B}})$, while the maximally concentrated on $\mathcal{F}$ vector from $\mathcal{D}$ is the eigenvector of $\mathbf{DBD}$, corresponding to the eigenvalue $\sigma^2_{max}(\mathbf{DB})$. Therefore

$$\inf_{\substack{\boldsymbol{f} \in \mathcal{D} \\ \|\boldsymbol{f}\|_2 = 1}} \beta^2 = 1 - \sigma^2_{max}\left(\mathbf{D}\overline{\mathbf{B}}\right) \tag{85}$$

and

$$\sup_{\substack{\boldsymbol{f} \in \mathcal{D} \\ \|\boldsymbol{f}\|_2 = 1}} \beta^2 = \sigma^2_{max}(\mathbf{DB}) \tag{86}$$

for the case $\alpha = 1$. All the values in between are attainable by the function $\boldsymbol{f} = \sum_{i=1}^{K} a_i \boldsymbol{\psi}_i$ with $\sum_{i=1}^{K} a_i^2 = 1$, where $\{\boldsymbol{\psi}_i\}_{i=1..K}$ are the eigenvectors of $\mathbf{DBD}$ belonging to $\mathcal{D}$ and corresponding to the eigenvalues from the interval $[1 - \sigma^2_{max}\left(\mathbf{D}\overline{\mathbf{B}}\right), \sigma^2_{max}(\mathbf{DB})]$.

Next let us consider the behavior of β for α belonging to $\alpha \in (0,\, 1)$. First, we will show that

$$\cos^{-1}\alpha + \cos^{-1}\beta \geq \cos^{-1} \sigma_{max}(\mathbf{BD}). \tag{87}$$

We can decompose any vector $\boldsymbol{f}$ as

$$\boldsymbol{f} = \lambda \mathbf{D}\boldsymbol{f} + \gamma \mathbf{B}\boldsymbol{f} + \boldsymbol{g}, \tag{88}$$

where $\boldsymbol{g}$ is a vector orthogonal to both $\mathcal{B}$ and $\mathcal{D}$ and again we consider a unit norm $\boldsymbol{f}$ with $\|\mathbf{D}\boldsymbol{f}\|_2 = \alpha$. Our goal is to find the nearest vector to $\boldsymbol{f}$ in the space spanned by $\mathbf{D}\boldsymbol{f}$ and $\mathbf{B}\boldsymbol{f}$.

First, we calculate the inner products of (88) successively with $\boldsymbol{f}, \mathbf{D}\boldsymbol{f}, \mathbf{B}\boldsymbol{f}$ and $\boldsymbol{g}$ and arrive to the system of equations

$$\begin{cases} 1 & = \lambda\alpha^2 + \gamma\beta^2 + \langle \boldsymbol{g}, \boldsymbol{f} \rangle, \\ \alpha^2 & = \lambda\alpha^2 + \gamma\langle \mathbf{B}\boldsymbol{f}, \mathbf{D}\boldsymbol{f} \rangle, \\ \beta^2 & = \lambda\langle \mathbf{D}\boldsymbol{f}, \mathbf{B}\boldsymbol{f} \rangle + \gamma\beta^2, \\ \langle \boldsymbol{f}, \boldsymbol{g} \rangle & = \langle \boldsymbol{g}, \boldsymbol{g} \rangle. \end{cases} \tag{89}$$

After eliminating $\langle \boldsymbol{g}, \boldsymbol{f} \rangle, \lambda$ and γ from the above system we arrive to

$$\beta^2 - 2\Re\langle \mathbf{D}\boldsymbol{f}, \mathbf{B}\boldsymbol{f} \rangle = -\alpha^2 + \left(1 - \frac{|\langle \mathbf{D}\boldsymbol{f}, \mathbf{B}\boldsymbol{f} \rangle|^2}{\alpha^2\beta^2}\right) - \|\boldsymbol{g}\|_2^2 \left(1 - \frac{|\langle \mathbf{D}\boldsymbol{f}, \mathbf{B}\boldsymbol{f} \rangle|^2}{\alpha^2\beta^2}\right). \quad (90)$$

According to (81) we define

$$\cos\theta = \Re \frac{\langle \mathbf{D}\boldsymbol{f}, \mathbf{B}\boldsymbol{f} \rangle}{\|\mathbf{D}\boldsymbol{f}\|_2 \|\mathbf{B}\boldsymbol{f}\|_2}. \quad (91)$$

Because we measure the angle θ between $\mathbf{D}\boldsymbol{f} \in \mathcal{D}$ and $\mathbf{B}\boldsymbol{f} \in \mathcal{B}$, according to Theorem A.2,

$$\theta \geq \cos^{-1} \sigma_{max}(\mathbf{BD}). \quad (92)$$

Due to the fact that

$$\alpha\beta\cos\theta = \Re\langle \mathbf{D}\boldsymbol{f}, \mathbf{B}\boldsymbol{f} \rangle \leq |\langle \mathbf{D}\boldsymbol{f}, \mathbf{B}\boldsymbol{f} \rangle| \leq \alpha\beta, \quad (93)$$

we can write

$$0 \leq 1 - \frac{|\langle \mathbf{D}\boldsymbol{f}, \mathbf{B}\boldsymbol{f} \rangle|^2}{\alpha^2\beta^2} \leq 1 - \cos^2\theta. \quad (94)$$

In (90), after introduction of θ, completion of the square on the left-hand side and use of (94), we finally arrive to

$$(\beta - \alpha\cos\theta)^2 \leq \left(1 - \alpha^2\right)\sin^2\theta, \quad (95)$$

where inequality can be achieved if and only if $\boldsymbol{g} = \mathbf{0}$ and $\langle \mathbf{D}\boldsymbol{f}, \mathbf{B}\boldsymbol{f} \rangle$ is real. Next, from (95) we can write

$$\beta \leq \cos\left(\theta - \cos^{-1}\alpha\right), \quad (96)$$

from which it follows, using bound (92), that

$$\beta \leq \cos\left(\cos^{-1}\sigma_{max}(\mathbf{BD}) - \cos^{-1}\alpha\right), \quad (97)$$

and we immediately come to (87).

Equality in (97) is achieved by

$$\boldsymbol{f}' = p\boldsymbol{\psi}_0 + q\mathbf{D}\boldsymbol{\psi}_0, \quad (98)$$

with

$$p = \sqrt{\frac{1-\alpha^2}{1-\sigma_{max}^2(\mathbf{BD})}}, \quad (99)$$

$$q = \frac{\alpha}{\sigma_{max}(\mathbf{BD})} - \sqrt{\frac{1-\alpha^2}{1-\sigma_{max}^2(\mathbf{BD})}} \quad (100)$$

and where $\boldsymbol{\psi}_1$ is an eigenvector of $\mathbf{BDB}$ corresponding to the eigenvalue $\sigma_{max}^2(\mathbf{BD})$. In (99) and (100) it was supposed that $\sigma_{max}^2(\mathbf{BD}) < 1$, because in the case $\sigma_{max}^2(\mathbf{BD}) = 1$ there exists at least one vector belonging to both $\mathcal{B}$ and $\mathcal{D}$, therefore point with $\alpha = 0$ and $\beta = 1$ belongs to Γ.

Applying the same steps between (87) and (97) to the operators $\mathbf{B}\overline{\mathbf{D}}$, $\overline{\mathbf{B}}\mathbf{D}$ and $\overline{\mathbf{B}}\overline{\mathbf{D}}$, we obtain the three remaining inequalities in (26). For $\beta = 1$ and $\alpha \in \left[1 - \sigma_{max}^2(\mathbf{BD}), \sigma_{max}^2(\mathbf{BD})\right]$ the concentrations are achievable by the eigenvectors of $\mathbf{BDB}$ which belong to $\mathcal{B}$ and their linear combinations. Continuing by analogy one can show that all the values α and β belonging to the border of Γ (see Fig. 1) are achievable. All the points inside Γ are achievable by the functions build up from different combinations of left and right singular vectors of $\mathbf{BD}$, $\mathbf{B}\overline{\mathbf{D}}$, $\overline{\mathbf{B}}\mathbf{D}$ and $\overline{\mathbf{B}}\overline{\mathbf{D}}$.

REFERENCES

[1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Proc. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

[2] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Proc. Mag.*, vol. 31, no. 5, pp. 80–90, 2014.

[3] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. on Signal Proc.*, vol. 61, pp. 1644–1656, 2013.

[4] X. Zhu and M. Rabbat, "Approximating signals supported on graphs," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 3921–3924.

[5] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *arXiv preprint arXiv:1503.05432*, 2015.

[6] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: Foundation and 1-d time," *IEEE Trans. Signal Process.*, vol. 56, pp. 3572–3585, 2008.

[7] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: 1-d space," *IEEE Trans. on Signal Processing*, pp. 3586–3599, 2008.

[8] A. Agaskar and Y. M. Lu, "A spectral graph uncertainty principle," *IEEE Trans. on Inform. Theory*, vol. 59, no. 7, pp. 4338–4356, 2013.

[9] B. Pasdeloup, R. Alami, V. Gripon, and M. Rabbat, "Toward an uncertainty principle for weighted graphs," *arXiv preprint arXiv:1503.03291*, 2015.

[10] J. J. Benedetto and P. J. Koprowski, "Graph theoretic uncertainty principles," *http://www.math.umd.edu/ jjb/graph_theoretic_UP_April_14.pdf*, 2015.

[11] P. J. Koprowski, *Finite Frames and Graph Theoretic Uncertainty Principles*, Ph.D. thesis, 2015.

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

[13] S.K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 5445–5449.

[14] X. Wang, P. Liu, and Y. Gu, "Local-set-based graph signal reconstruction," *IEEE Trans. on Signal Processing*, vol. 63, no. 9, pp. 2432–2444, 2015.

[15] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty. I," *The Bell System Techn. Journal*, vol. 40, no. 1, pp. 43–63, Jan. 1961.

[16] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty, II," *Bell System Technical Journal*, vol. 40, no. 1, pp. 65–84, 1961.

[17] D. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," Apr. 2010, vol. 30, pp. 129–150.

[18] A. Anis, A. Gadde, and A. Ortega, "Towards a sampling theorem for signals on arbitrary graphs," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3864–3868.

[19] F. R. K. Chung, *Spectral Graph Theory*, American Mathematical Society, 1997.

[20] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. on Signal Proc.*, vol. 62, pp. 3042–3054, 2014.

[21] G. B. Folland and A. Sitaram, "The uncertainty principle: A mathematical survey," 1997, pp. 207–238.

[22] I. Z. Pesenson and M. Z. Pesenson, "Sampling, filtering and sparse approximations on combinatorial graphs," *Journal of Fourier Analysis and Applications*, vol. 16, no. 6, pp. 921–942, 2010.

[23] R. J. Duffin and A. C. Schaeffer, "A class of nonharmonic fourier series," *Trans. of the American Mathematical Society*, pp. 341–366, 1952.

[24] I. Z. Pesenson, "Sampling, splines and frames on compact manifolds," *GEM-International Journal on Geomathematics*, vol. 6, no. 1, pp. 43–81, 2015.

[25] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[26] H. Avron and C. Boutsidis, "Faster subset selection for matrices and applications," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 4, pp. 1464–1499, 2013.