Sampling of Graph Signals with Successive Local Aggregations

Antonio G. Marques, *Senior Member, IEEE*, Santiago Segarra, *Student Member, IEEE*, Geert Leus, *Fellow, IEEE*, and Alejandro Ribeiro, *Member, IEEE*

***Abstract*—A new scheme to sample signals defined on the nodes of a graph is proposed. The underlying assumption is that such signals admit a sparse representation in a frequency domain related to the structure of the graph, which is captured by the so-called graph-shift operator. Most of the works that have looked at this problem have focused on using the value of the signal observed at a subset of nodes to recover the signal in the entire graph. Differently, the sampling scheme proposed here uses as input observations taken at a single node. The observations correspond to sequential applications of the graph-shift operator, which are linear combinations of the information gathered by the neighbors of the node. When the graph corresponds to a directed cycle (which is the support of time-varying signals), our method is equivalent to the classical sampling in the time domain. When the graph is more general, we show that the Vandermonde structure of the sampling matrix, which is critical to guarantee recovery when sampling time-varying signals, is preserved. Sampling and interpolation are analyzed first in the absence of noise and then noise is considered. We then study the recovery of the sampled signal when the specific set of frequencies that is active is not known. Moreover, we present a more general sampling scheme, under which, either our aggregation approach or the alternative approach of sampling a graph signal by observing the value of the signal at a subset of nodes can be both viewed as particular cases. The last part of the paper presents numerical experiments that illustrate the results developed through both synthetic graph signals and a real-world graph of the economy of the United States.**

***Index Terms*—Graph signal processing, Sampling, Interpolation, Graph signals, Error covariance, Support selection.**

I. Introduction

Sampling (and subsequent interpolation) is a cornerstone problem in classical signal processing [1]. The emergence of new fields of knowledge such as network science and big data is generating a pressing need to extend the results existing for classical time-varying signals to signals defined on graphs [2]–[5]. This not only entails modifying the algorithms currently available for time-varying signals, but also gaining intuition on what concepts are preserved (and lost) when a signal is defined, not in the classical time grid, but in a more general graph domain.

Work in this paper is supported by Spanish MINECO grant No TEC2013-41604-R and USA NSF CCF-1217963. A. G. Marques is with the Dept. of Signal Theory and Comms., King Juan Carlos Univ. S. Segarra and A. Ribeiro are with the Dept. of Electrical and Systems Eng., Univ. of Pennsylvania. G. Leus is with the Dept. of Electrical Eng., Mathematics and Computer Science, Delft Univ. of Technology. Emails: antonio.garcia.marques@urjc.es, ssegarra@seas.upenn.edu, g.j.t.leus@tudelft.nl, and aribeiro@seas.upenn.edu.

This paper investigates the sampling and posterior recovery of signals that are defined on the nodes of a graph. The underlying assumption is that such signals admit a sparse representation in a (frequency) domain which is related to the structure of the graph where these signals reside. Most of the current efforts in this field have been focused on using the value of the signal observed at a subset of nodes to recover the signal in the entire graph [6]–[10]. Our proposal in this paper is different. We begin by presenting a new sampling method that accounts for the graph structure, can be run at a single node and only requires access to information of neighboring nodes. Moreover, we also show that the proposed method shares similarities with the classical sampling and interpolation of time-varying signals. When the graph corresponds to a directed cycle, which is the support of classical time-varying signals, our method is equivalent to classical sampling. When the graph is more general, the Vandermonde structure of the sampling matrix, which is critical to guarantee recovery in classical sampling [1], is preserved. Such a structure not only facilitates the interpolation process, but also helps to draw some connections between the proposed method and the sampling of time-varying signals. Sampling and interpolation are analyzed first in the absence of noise, where the conditions that guarantee recovery are identified. The conditions depend both on the structure of the graph and the particular node taking the observations. We then analyze the sampling and reconstruction process when noise is present and when the specific frequencies where the signal is sparse are not known. For the noisy case, an interpolator based on the Best Linear Unbiased Estimator (BLUE) is designed and the effect on the corresponding error covariance matrix of different noise models is discussed. For the case of unknown frequency support, we provide conditions under which the signal can be identified. This second problem falls into the category of sparse signal reconstruction [11]–[14] where the main idea is to leverage the structure of the observation matrix to facilitate recovery. The last contribution is the definition of a more general space-shift sampling method that considers a subset of nodes, each of them taking multiple observations. Within that generalization, the approach of sampling a graph signal by observing the value of the signal at a subset of nodes can be viewed as a particular case. Space-shift sampling is relevant not only because it subsumes existing schemes, but also because it can be used to compare and establish relationships between existing approaches to sample signals in graphs and our proposed method.

The paper is organized as follows. Section II introduces

the new aggregation sampling method, compares it to the existing selection sampling method and shows that for classical time-varying signals both methods are equivalent. Section III analyzes our sampling method in more detail and applies it to sample bandlimited graph signals. The analysis includes conditions for recovery, which are formally stated in Section III-B. Section IV investigates the effect of noise in aggregation sampling. It also discusses how to select sampling nodes and observation schemes that lead to a good recovery performance. Corresponding modifications in the interpolation in order to recover the signal when the support is not known are discussed in Section V. Section VI proposes a generalization under which the existing selection sampling and the proposed aggregation sampling can be viewed as particular cases. Finally, several illustrative numerical results are presented in Section VII.

Notation: The entries of a matrix $\mathbf{X}$ and a vector $\mathbf{x}$ are denoted as X_{ij} and x_i; however, when contributing to avoid confusion, the alternative notation $[\mathbf{X}]_{ij}$ and $[\mathbf{x}]_i$ will be used. The notations ^{T} and ^{H} stand for transpose and transpose conjugate, respectively; $\otimes$ is the Kronecker product; $\text{vec}(\mathbf{X})$ is the column-wise vectorized version of matrix $\mathbf{X}$; $\mathbf{e}_i$ is the i-th $N \times 1$ canonical basis vector (all entries of $\mathbf{e}_i$ are zero except the i-th one, which is one); $\mathbf{E}_K := [\mathbf{e}_1, ..., \mathbf{e}_K]$ is a tall matrix collecting the K first canonical basis vectors; and $\mathbf{0}$ and $\mathbf{1}$ are, respectively, the all-zeros and all-ones matrices (if the dimensions are not clear from the context, a subscript will be used). The modulus (remainder) obtained after dividing x by N is denoted as $\text{mod}_N(x)$.

II. Sampling of Graph Signals

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ denote a directed graph. The set of nodes or vertices $\mathcal{N}$ has cardinality N, and the set of links $\mathcal{E}$ is such that $(i, j) \in \mathcal{E}$ if and only if node i is connected to node j. The set $\mathcal{N}_i : \{ j \,|(j, i) \in \mathcal{E}\}$ contains all nodes with an incoming connection to i and is termed the incoming neighborhood of i. For any given graph we define the adjacency matrix $\mathbf{A}$ as a sparse $N \times N$ matrix with non-zero elements A_{ij} if and only if $(j, i) \in \mathcal{E}$. The value of A_{ij} captures the strength of the connection between i and j. The focus of this paper is not on analyzing $\mathcal{G}$, but a graph signal defined on the set of nodes $\mathcal{N}$. Such a signal can be represented as a vector $\mathbf{x} = [x_1, \ldots, x_N]^T \in \mathbb{R}^N$ where the i-th component represents the value of the signal at node i, or, equivalently, as a function $f : \mathcal{N} \to \mathbb{R}$, defined on the vertices of the graph.

The graph $\mathcal{G}$ is endowed with a *graph-shift operator* $\mathbf{S}$ defined as an $N \times N$ matrix whose entry (i, j), denoted as S_{ij}, can be non-zero only if $i = j$ or $(j, i) \in \mathcal{E}$. The sparsity pattern of the matrix $\mathbf{S}$ captures the local structure of $\mathcal{G}$ but we make no specific assumptions on the values of the non-zero entries of $\mathbf{S}$. Common choices for $\mathbf{S}$ are the adjacency matrix of the graph [3], [15], the Laplacian [2], and its generalizations [16]. The intuitive interpretation of $\mathbf{S}$ is that it represents a linear transformation that can be computed locally at the nodes of the graph. Specifically, if $\mathbf{y} = [y_1, \ldots, y_N]^T$ is defined as $\mathbf{y} = \mathbf{S}\mathbf{x}$, then node i can compute y_i provided that it has access to the values of x_j at its incoming neighbors $j \in \mathcal{N}_i$. We assume henceforth that $\mathbf{S}$ is diagonalizable, so that there exists a $N \times N$ matrix $\mathbf{V}$ and a $N \times N$ diagonal matrix $\mathbf{\Lambda}$ that can be used to decompose $\mathbf{S}$ as $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$. When $\mathbf{S}$ is normal, i.e., when $\mathbf{S}\mathbf{S}^H = \mathbf{S}^H\mathbf{S}$, not only $\mathbf{S}$ is diagonalizable but $\mathbf{V}$ is unitary, which implies $\mathbf{V}^{-1} = \mathbf{V}^H$, and leads to the decomposition $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$.

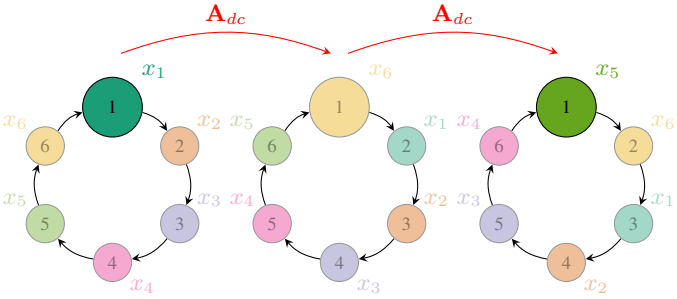


Fig. 1: Conventional sampling in the time domain as aggregation sampling in $\mathcal{G}_{dc}$. We apply the shift operator $\mathbf{S}$ successively and sample the resulting signal observed at a given node (here, node 1).

Let $\mathbf{C}$ denote a fat $K \times N$ *selection* matrix whose elements satisfy: $C_{ij} \in \{0, 1\}$, $\sum_j C_{ij} = 1$ for all i, and $\sum_i C_{ij} \leq 1$ for all j. A natural definition of sampling for a graph signal is to obtain the sampled signal as [9]

$$\bar{\mathbf{x}} := \mathbf{C}\mathbf{x}. \tag{1}$$

Given the structure of $\mathbf{C}$, $\bar{\mathbf{x}}$ is a selection of K out of the N elements of $\mathbf{x}$ and K/N is the sampling rate. Uniform sampling amounts to choosing $\mathbf{C} = [\mathbf{e}_1, \mathbf{e}_{N/K+1}, \ldots, \mathbf{e}_{N-N/K+1}]^T$ and the selection of the first K elements of $\mathbf{x}$ is accomplished by setting $\mathbf{C} = \mathbf{E}_K^T := [\mathbf{e}_1, \ldots, \mathbf{e}_K]^T$. In general, it is not clear how to choose good selection matrices $\mathbf{C}$. This is in contrast to conventional sampling of signals in the time domain where uniform sampling is advantageous [1].

An equally valid, yet less intuitive, definition is to fix a node, say i, and consider the sampling of the signal seen by this node as the shift operator $\mathbf{S}$ is applied recursively. To describe this sampling methodology more clearly, define the l-th shifted signal $\mathbf{y}^{(l)} := \mathbf{S}^l\mathbf{x}$ and further define the $N \times N$ matrix

$$\mathbf{Y} := [\mathbf{y}^{(0)}, \mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N-1)}] = [\mathbf{x}, \mathbf{S}\mathbf{x}, \ldots, \mathbf{S}^{N-1}\mathbf{x}], \tag{2}$$

that groups the signal $\mathbf{x}$ and the result of the first $N - 1$ applications of the shift operator. Associating the i-th row of $\mathbf{Y}$ with node i, we define the successively aggregated signal at i as $\mathbf{y}_i := (\mathbf{e}_i^T\mathbf{Y})^T = \mathbf{Y}^T\mathbf{e}_i$. Sampling is now reduced to the selection of K out of the N elements (rows) of $\mathbf{y}_i$, which is accomplished with a selection matrix $\mathbf{C}$ [cf. (1)]

$$\bar{\mathbf{y}}_i := \mathbf{C}\mathbf{y}_i = \mathbf{C}\left(\mathbf{Y}^T\mathbf{e}_i\right). \tag{3}$$

We say that the signal $\bar{\mathbf{y}}_i$ samples $\mathbf{x}$ with successive local aggregations. This nomenclature follows from the fact that $\mathbf{y}^{(l)}$ can be computed recursively as $\mathbf{y}^{(l)} := \mathbf{S}\mathbf{y}^{(l-1)}$ and that the i-th element of this vector can be computed using signals associated with itself and its incoming neighbors,

$$y_i^{(l)} = \sum_{j \in \mathcal{N}_i} S_{ij} y_j^{(l-1)}. \tag{4}$$

We can then think of the signal $\mathbf{y}_i$ as being computed locally at node i using successive variable exchanges with neighboring

nodes. In fact, $y_i^{(l)}$ can be expressed as a linear combination of the values of x_j at nodes j whose distance (number of hops) from node i is less than or equal to l. This implies that the sampled signal $\bar{\mathbf{y}}_i$ in (3) is a selection of values that node i can determine locally. Indeed, an underlying idea behind the sampling in (3) is to incorporate the structure of the shift into the sampling procedure.

To understand the difference between selection sampling [cf. (1)] and aggregation sampling [cf. (3)], it is instructive to consider their application to a signal defined in the time domain. Classical time-domain signals can be represented as graph signals defined on top of a directed cycle graph [2], [9], as illustrated in Fig. 1. Let $\mathcal{G}_{dc}$ and $\mathbf{A}_{dc}$ denote the directed cycle graph and its adjacency matrix, respectively. In $\mathcal{G}_{dc}$, node i is connected only to node $j = \text{mod}_N(i) + 1$, so that the elements of $\mathbf{A}_{dc}$ are zero except for the ones in the first cyclic subdiagonal, which are one. For a signal $\mathbf{x}$ defined on $\mathcal{G}_{dc}$, we consider selection and aggregation sampling when using $\mathbf{S} = \mathbf{A}_{dc}$ and the uniform selection matrix $\mathbf{C} = [\mathbf{e}_1, \mathbf{e}_{N/K+1}, \ldots, \mathbf{e}_{N-N/K+1}]^T$. In selection sampling, the sampled signal is obtained using (1) as $\bar{\mathbf{x}} = \mathbf{C}\mathbf{x}$. In aggregation sampling, subsequent applications of $\mathbf{S} = \mathbf{A}_{dc}$ are considered. Each of these shifts amounts to rotating the signal clockwise. It follows that the aggregated signal $\mathbf{y}_1$ in (2) is given by $\mathbf{y}_1 = [x_1, x_N, x_{N-1}, \ldots, x_2]$, which upon multiplication by $\mathbf{C}$ [cf. (3)] results in a vector $\bar{\mathbf{y}}_1 = \mathbf{C}\mathbf{y}_1$ that contains the same elements that $\bar{\mathbf{x}}$ contains. Hence, when $\mathbf{S} = \mathbf{A}_{dc}$ both methods can be viewed as generalizations of conventional sampling. However, for more general topologies, selection and aggregation sampling produce different outcomes. In selection sampling we move through nodes to collect samples at points uniquely identified by $\mathbf{C}$, whereas in aggregation sampling we move the signal through the graph while collecting samples at a fixed node.

III. Sampling of bandlimited graph signals

Recovery of $\mathbf{x}$ from its sampled version is possible under the assumption that $\mathbf{x}$ admits a sparse representation. The common practice when addressing the problem of sampling signals in graphs is to suppose that $\mathbf{S}$ plays a key *role in explaining* the signals of interest $\mathbf{x}$. More specifically, that $\mathbf{x}$ can be expressed as a linear combination of a *subset* of the columns of $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_N]$, or, equivalently, that the vector $\widehat{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$ is sparse. In this context, vectors $\mathbf{v}_k$ are interpreted as the graph frequency basis and $\widehat{x}_k$ as the corresponding signal frequency coefficients. To simplify exposition, it will be assumed throughout that the active frequencies are the first K ones, which are associated with the largest eigenvalues [6], [17], so that $\widehat{\mathbf{x}} = [\widehat{x}_1, ..., \widehat{x}_K, 0, ..., 0]^T$. However, the results in the paper can be applied to any set of active frequencies $\mathcal{K}$ of size K provided that $\mathcal{K}$ is known. For convenience, we define $\mathbf{V}_K := [\mathbf{v}_1, ..., \mathbf{v}_K]$ and $\widehat{\mathbf{x}}_K := [\widehat{x}_1, ..., \widehat{x}_K]^T$ so that we may write $\widehat{\mathbf{x}} = [\widehat{\mathbf{x}}_K^T \mid \mathbf{0}_{1\times N-K}]^T$. For $\widehat{\mathbf{x}}$ to be sparse, it is reasonable to assume that $\mathbf{S}$ is involved in the generation of $\mathbf{x}$.

When $\mathcal{G} = \mathcal{G}_{dc}$, setting the shift operator either to $\mathbf{S} = \mathbf{A}_{dc}$ or to $\mathbf{S} = \mathbf{L}_{dc} := \mathbf{I} - \mathbf{A}_{dc}$ gives rise to the Fourier basis $\mathbf{F}$. More formally, since $\mathbf{S}$ is circulant, its right eigenvectors are $\mathbf{V} = \mathbf{F}$, with $F_{ij} := e^{+\mathrm{j}2\pi(i-1)(j-1)/N}/\sqrt{N}$ and $\mathrm{j} := \sqrt{-1}$. Selecting $\mathbf{S} = \mathbf{A}_{dc}$ has the additional advantage of satisfying $\Lambda_{ii} = e^{-\mathrm{j}2\pi(i-1)/N}$, i.e., the eigenvalues of the shift operator correspond to the classical discrete frequencies. Interpretations for the eigenvalues of the Laplacian matrix $\mathbf{L}_{dc}$ also exist [2].

A. Selection sampling of bandlimited graph signals

Under the selection sampling approach [6]–[10], sampling a graph signal amounts to setting $\bar{\mathbf{x}} = \mathbf{C}\mathbf{x}$ [cf. (1)]. Since the $K \times N$ binary matrix $\mathbf{C}$ selects the observed nodes, the issue then is how to design $\mathbf{C}$, i.e., which nodes to select, and how to recover the original signal $\mathbf{x}$ from its samples $\bar{\mathbf{x}}$.

To answer this, it is assumed that $\mathbf{x}$ is bandlimited, so that it can be expressed as a linear combination of the K principal eigenvectors in $\mathbf{V}$. The sampled signal $\bar{\mathbf{x}}$ is then $\bar{\mathbf{x}} = \mathbf{C}\mathbf{x} = \mathbf{C}\mathbf{V}_K\widehat{\mathbf{x}}_K$. Hence, if matrix $\mathbf{C}\mathbf{V}_K$ is invertible, $\widehat{\mathbf{x}}_K$ can be recovered from $\bar{\mathbf{x}}$ and, thus, the original signal $\mathbf{x}$ is obtained as

$$\mathbf{x} = \mathbf{V}_K\widehat{\mathbf{x}}_K = \mathbf{V}_K(\mathbf{C}\mathbf{V}_K)^{-1}\bar{\mathbf{x}}. \tag{5}$$

Perfect signal reconstruction can be guaranteed by selecting a subset of K nodes such that the corresponding rows in $\mathbf{V}_K$ are linearly independent. In the classical domain of time-varying signals, $\mathbf{V}_K = \mathbf{F}\mathbf{E}_K$ has a row-wise Vandermonde structure, which implies that any subset of K rows is invertible. However, for an arbitrary graph this is not guaranteed and algorithms to select a specific subset that guarantees recovery are required [7].

B. Aggregation sampling of bandlimited graph signals

As explained in (3), under the aggregation approach the sampled signal is formed by observations of the shifted signals $\mathbf{y}^{(l)} = \mathbf{S}^l\mathbf{x}$ taken at a given node i. Under this approach, $\mathbf{S}$ plays a *role* not only in *explaining* and *recovering* $\mathbf{x}$, but also in *sampling* $\mathbf{x}$. Another reason to consider this scheme is that the entries of $\mathbf{y}^{(l)}$ can be found by sequentially exchanging information among neighbors. This implies that: a) for setups where graph vertices correspond to nodes of an actual network, the procedure can be implemented distributedly; and b) if recovery is feasible, the observations at a single node can recover the signal in the entire graph.

As done before, we first analyze how the bandlimitedness of $\mathbf{x}$ is manifested on the sampled signal. Then, we identify under which conditions recovery is feasible and describe the corresponding interpolation algorithm. For ease of exposition, the dependence of $\mathbf{y}_i$ on $\widehat{\mathbf{x}}$ is given in the form of a lemma.

Lemma 1: *Define the $N \times 1$ vector $\boldsymbol{\upsilon}_i := \mathbf{V}^T\mathbf{e}_i$, which collects the values of the frequency basis $\{\mathbf{v}_k\}_{k=1}^K$ at node i, and the $N \times N$ (column-wise) Vandermonde matrix*

$$\boldsymbol{\Psi} := \begin{pmatrix} 1 & \ldots & 1 \\ \lambda_1 & \ldots & \lambda_N \\ \vdots & & \vdots \\ \lambda_1^{N-1} & \ldots & \lambda_N^{N-1} \end{pmatrix}. \tag{6}$$

Then, the shifted signal $\mathbf{y}_i$ can be expressed as

$$\mathbf{y}_i = \boldsymbol{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\widehat{\mathbf{x}}. \tag{7}$$

Proof: Since $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, signal $\mathbf{y}^{(l)}$ can be written as

$$\mathbf{y}^{(l)} = \mathbf{S}^l\mathbf{x} = (\mathbf{V}\mathbf{\Lambda}^l\mathbf{V}^{-1})\mathbf{x} = (\mathbf{V}\mathbf{\Lambda}^l)\widehat{\mathbf{x}}. \quad (8)$$

Based on the definitions of $\mathbf{y}_i$ and $\boldsymbol{\upsilon}_i$, it follows that

$$\begin{aligned}\mathbf{y}_i &= \mathbf{Y}^T\mathbf{e}_i = (\mathbf{V}\mathbf{V}^{-1}\mathbf{Y})^T\mathbf{e}_i \\ &= (\mathbf{V}^{-1}\mathbf{Y})^T\mathbf{V}^T\mathbf{e}_i = (\mathbf{V}^{-1}\mathbf{Y})^T\boldsymbol{\upsilon}_i. \end{aligned} \quad (9)$$

Because the l-th column of matrix $\mathbf{Y}$ is $\mathbf{y}^{(l-1)}$, it can be written as $(\mathbf{V}\mathbf{\Lambda}^{l-1})\widehat{\mathbf{x}}$ [cf. (8)]. Hence, the l-th column of matrix $(\mathbf{V}^{-1}\mathbf{Y})$ can be written as $\mathbf{\Lambda}^{l-1}\widehat{\mathbf{x}}$ or, equivalently, as $\text{diag}(\widehat{\mathbf{x}})[\lambda_1^{l-1}, ..., \lambda_N^{l-1}]^T$. Leveraging the fact that the vector containing the l-th power of the eigenvalues corresponds to the row $l+1$ of matrix $\mathbf{\Psi}$, the shifted signal $\mathbf{y}_i$ can be expressed as

$$\begin{aligned}\mathbf{y}_i &= (\mathbf{V}^{-1}\mathbf{Y})^T\boldsymbol{\upsilon}_i = (\text{diag}(\widehat{\mathbf{x}})\mathbf{\Psi}^T)^T\boldsymbol{\upsilon}_i \\ &= \mathbf{\Psi}\text{diag}(\widehat{\mathbf{x}})\boldsymbol{\upsilon}_i = \mathbf{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\widehat{\mathbf{x}}, \end{aligned} \quad (10)$$

which is the claim in the lemma. ■

Note that while in Section III-A the relation between the sparse frequency coefficients $\widehat{\mathbf{x}}$ and the signal to be sampled was simply given by $\mathbf{x} = \mathbf{V}\widehat{\mathbf{x}}$, here it is given by $\mathbf{y}_i = \mathbf{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\widehat{\mathbf{x}}$.

Next, we use Lemma 1 to identify under which conditions recovery is feasible. To do this, let us define the $N \times K$ matrix $\mathbf{\Psi}_i = \mathbf{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{E}_K$. Then, the sampled signal $\bar{\mathbf{y}}_i$ is

$$\bar{\mathbf{y}}_i = \mathbf{C}\mathbf{y}_i = \mathbf{C}\mathbf{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\widehat{\mathbf{x}} = \mathbf{C}\mathbf{\Psi}_i\widehat{\mathbf{x}}_K, \quad (11)$$

where $\mathbf{C}$ is the binary $K \times N$ selection matrix, and $\widehat{\mathbf{x}}_K$ the vector collecting the non-zero elements of $\widehat{\mathbf{x}}$. To simplify exposition, for the time being we will assume that $\mathbf{C} = \mathbf{E}_K^T$, i.e., that the observations correspond to the original signal and the first $K-1$ shifts. The assumption can be relaxed as discussed in Remark 1.

If matrix $\mathbf{C}\mathbf{\Psi}_i$ is invertible, then $\widehat{\mathbf{x}}_K$ can be recovered from $\bar{\mathbf{y}}_i$ [cf. (11)] and, hence, $\mathbf{x}$ can be found as [cf. (5)]

$$\mathbf{x} = \mathbf{V}_K\widehat{\mathbf{x}}_K = \mathbf{V}_K(\mathbf{C}\mathbf{\Psi}_i)^{-1}\bar{\mathbf{y}}_i. \quad (12)$$

Thus, the expression in (12) shows how $\mathbf{x}$ can be interpolated from $\bar{\mathbf{y}}_i$. The interpolator $\mathbf{V}_K(\mathbf{C}\mathbf{\Psi}_i)^{-1}$ may be decomposed into three factors $(\mathbf{V}_K)(\mathbf{E}_K^T\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{E}_K)^{-1}(\mathbf{C}\mathbf{\Psi}\mathbf{E}_K)^{-1}$ to reveal its dependence on the support where the signal is bandlimited, the node taking the samples, and the spectrum of the graph.

Equation (12) requires $\mathbf{C}\mathbf{\Psi}_i$ being invertible. Hence, perfect reconstruction can be guaranteed by selecting samples such that the corresponding rows in $\mathbf{\Psi}_i$ are linearly independent. While for the selection sampling described in Section III-A there is no straightforward way to check the invertibility of $\mathbf{C}\mathbf{V}_K$ (existing algorithms typically do that by inspection [7]), for the aggregation sampling described in (7)-(12) the invertibility of $\mathbf{C}\mathbf{\Psi}_i$ can be guaranteed if the conditions presented in the following proposition hold.

Proposition 1: *Let $\mathbf{x}$ and $\bar{\mathbf{y}}_i$ be, respectively, a bandlimited graph signal with at most K non-zero frequency components and the output of the sampling process defined in* (11)*. Then, the N entries of signal $\mathbf{x}$ can be recovered from the K samples in $\bar{\mathbf{y}}_i$ if the two following conditions hold*
i) The first K eigenvalues of the graph-shift operator $\mathbf{S}$ are distinct; i.e., $\lambda_i \neq \lambda_j$ for all $i \neq j$, $i \leq K$ and $j \leq K$.
ii) The K first entries of $\boldsymbol{\upsilon}_i$ are non-zero.

Proof: To prove the proposition it suffices to show that under *i)* and *ii)*, $\mathbf{C}\mathbf{\Psi}_i$ is invertible [cf. (12)]. Matrix $\mathbf{C}\mathbf{\Psi}_i$ can be understood as the multiplication of two matrices: matrix $(\mathbf{C}\mathbf{\Psi}\mathbf{E}_K)$ and matrix $(\mathbf{E}_K^T\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{E}_K)$. Condition *ii)* guarantees that the second matrix is invertible. Moreover, condition *i)* guarantees invertibility of the first matrix. To see this, note that $(\mathbf{\Psi}\mathbf{E}_K)$ is a $N \times K$ (column-wise) Vandermonde matrix. Hence $\mathbf{C}(\mathbf{\Psi}\mathbf{E}_K)$ is a selection of the first K rows of $(\mathbf{\Psi}\mathbf{E}_K)$, which is also Vandermonde. Any square Vandermonde matrix has full rank provided that the basis (i.e., the eigenvalues of $\mathbf{S}$) are distinct, as required in condition *i)*. ■

One of the implications of Proposition 1 is that there is no need to compute or observe the entire vector $\mathbf{y}_i$, since its first K entries suffice to guarantee recovery. Hence, linear combinations of signals at nodes that are in a neighborhood of radius $K-1$ suffice to reconstruct the entire graph signal. The conditions in the proposition are easy to check, providing additional insights on aggregation sampling. Condition *i)* states that if a graph has two identical frequencies and the signal of interest is a linear combination of both of them, the reconstruction will fail. Condition *ii)* refers to the specific node where the samples are taken. It states that any node in the network can be used to sample the signal provided that $(\mathbf{e}_k^T\boldsymbol{\upsilon}_i) \neq 0$ for $k \leq K$; i.e., that the chosen node participates in the specific frequencies on which signal $\mathbf{x}$ is expressed. It also points to the fact that if $|\mathbf{e}_k^T\boldsymbol{\upsilon}_i|$ is small, the interpolation matrix associated with i may be poorly conditioned. For the particular case of $\mathbf{S} = \mathbf{A}_{dc}$, conditions *i)* and *ii)* are always satisfied.

From the above discussion, one can also understand bandlimited graph signals as signals that can be identified *locally* by relying on observations within a given number of hops. This does not necessarily imply that the variation of the signal among close-by nodes is small, but that the pattern of variation can be inferred just by looking at close-by nodes. For the *recovery* to be implemented locally too, the nodes need to know $\mathbf{V}_K$ and $\{\lambda_k\}_{k=1}^K$, i.e. the structure of the graph where the signal resides. Alternative schemes to reconstruct a bandlimited graph signal using only local interactions have been reported in [18].

Remark 1: The structure of the selection matrix $\mathbf{C}$ and, in particular, the fact that $\mathbf{C}\mathbf{\Psi}\mathbf{E}_K$ is a Vandermonde matrix are instrumental to guarantee the recovery of $\mathbf{x}$. Note that $\mathbf{C}\mathbf{\Psi}\mathbf{E}_K$ is Vandermonde not only when $\mathbf{C} = \mathbf{E}_K^T$, but also when $\mathbf{C} = [\mathbf{e}_1, \mathbf{e}_{1+N_0}, \ldots, \mathbf{e}_{1+(K-1)N_0}]^T$, provided that $1 \leq N_0 \leq N/K$ and $\lambda_{k_1}^{N_0} \neq \lambda_{k_2}^{N_0}$ for all $k_1 \neq k_2$, where $k_1 \leq K$ and $k_2 \leq K$. By setting $N_0 = N/K$, the counterpart of the classical time sampling theorem (which considers uniformly spaced samples) is recovered. Moreover, if none of the frequencies of interest $\{\lambda_k\}_{k=1}^K$ is zero, selection patterns of the form $\mathbf{C} = [\mathbf{e}_{n_0}, \mathbf{e}_{n_0+N_0}, \ldots, \mathbf{e}_{n_0+(K-1)N_0}]^T$ are also guaranteed to lead to invertible matrices. In this case, the resultant matrix is a product of a Vandermonde and

a non-zero diagonal matrix. For reference in the following sections, we define here the $K \times N$ matrix $\mathbf{C}_K(n_0, N_0) := [\mathbf{e}_{n_0}, \mathbf{e}_{n_0+N_0}, \ldots, \mathbf{e}_{n_0+(K-1)N_0}]^T$ and the *set* of admissible $K \times N$ selection matrices $\mathcal{C}_K := \{\mathbf{C}_K(n_0, N_0) \quad | \; N_0 = 1, \ldots, N/K \text{ and } n_0 = 1, \ldots, N - N_0(K-1)\}$.

IV. Sampling and interpolation in the presence of noise

If the samples are noisy, perfect reconstruction is, in general, unfeasible and new issues arise. In Section IV-A, we estimate $\mathbf{x}$ for a general noise model using a Best Linear Unbiased Estimator (BLUE). We then specify noise models that are likely to arise in graph domains. In Section IV-B, we discuss the effect on the interpolation error of selecting the sampling node and the selection matrix.

A. BLUE interpolation

Consider now that the shifted signal $\mathbf{y}_i$ is corrupted by additive noise, so that the *observed* signal $\mathbf{z}_i$ is given by $\mathbf{z}_i = \mathbf{y}_i + \mathbf{w}_i$. The noise $\mathbf{w}_i$ is assumed to be zero-mean, independent of the graph signal, and colored with a covariance matrix $\mathbf{R}_w^{(i)} := \mathbb{E}[\mathbf{w}_i \mathbf{w}_i^H]$. For notational convenience, we define also $\bar{\mathbf{w}}_i = \mathbf{C}\mathbf{w}_i$ and $\bar{\mathbf{R}}_w^{(i)} = \mathbf{C}\mathbf{R}_w^{(i)}\mathbf{C}^H$.

To design the interpolator in the presence of noise, we leverage that the relation between $\bar{\mathbf{z}}_i$ and $\mathbf{x}$ is given by

$$\bar{\mathbf{z}}_i = \mathbf{C}\mathbf{\Psi}_i \widehat{\mathbf{x}}_K + \bar{\mathbf{w}}_i, \tag{13}$$

$$\mathbf{x} = \mathbf{V}_K \widehat{\mathbf{x}}_K. \tag{14}$$

The BLUE estimator of $\widehat{\mathbf{x}}_K$, which minimizes the least squares error, is then given by [19]

$$\hat{\widehat{\mathbf{x}}}_K^{(i)} = \left(\mathbf{\Psi}_i^H \mathbf{C}^H (\bar{\mathbf{R}}_w^{(i)})^{-1} \mathbf{C}\mathbf{\Psi}_i\right)^{-1} \mathbf{\Psi}_i^H \mathbf{C}^H (\bar{\mathbf{R}}_w^{(i)})^{-1} \bar{\mathbf{z}}_i, \tag{15}$$

provided that the inverse in (15) exists. Additionally, for the particular case of Gaussian noise in (13), the estimator in (15) coincides with the Minimum Variance Unbiased (MVU) estimator which attains the Cramér-Rao lower bound. Clearly, the larger the number of rows in (13), the better the estimation is. When the selection matrix $\mathbf{C}$ selects exactly K rows (and not more), (15) reduces to

$$\hat{\widehat{\mathbf{x}}}_K^{(i)} = \left(\mathbf{C}\mathbf{\Psi}_i\right)^{-1} \bar{\mathbf{z}}_i. \tag{16}$$

After obtaining $\hat{\widehat{\mathbf{x}}}_K^{(i)}$ – either via (15) or (16) –, the time signal recovered at the i-th node $\hat{\mathbf{x}}^{(i)}$ can be found as

$$\hat{\mathbf{x}}^{(i)} = \mathbf{V}_K \hat{\widehat{\mathbf{x}}}_K^{(i)}. \tag{17}$$

Finally, the error covariance matrices for the frequency and time estimators $\widehat{\mathbf{R}}_e^{(i)} := \mathbb{E}[(\widehat{\mathbf{x}}_K - \hat{\widehat{\mathbf{x}}}_K^{(i)})(\widehat{\mathbf{x}}_K - \hat{\widehat{\mathbf{x}}}_K^{(i)})^H]$ and $\mathbf{R}_e^{(i)} := \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}}^{(i)})(\mathbf{x} - \hat{\mathbf{x}}^{(i)})^H]$ are [19]

$$\widehat{\mathbf{R}}_e^{(i)} = \left(\mathbf{\Psi}_i^H \mathbf{C}^H (\bar{\mathbf{R}}_w^{(i)})^{-1} \mathbf{C}\mathbf{\Psi}_i\right)^{-1}, \tag{18}$$

$$\mathbf{R}_e^{(i)} = \mathbf{V}_K \widehat{\mathbf{R}}_e^{(i)} \mathbf{V}_K^H. \tag{19}$$

Note that $\mathbf{R}_e^{(i)}$ depends on the noise model, the frequencies of the graph, the node taking the observations, and the sample-selection scheme adopted (cf. Remark 1).

The expressions in (18)-(19) can be used to assess the performance of the estimation. Multiple alternatives to quantify the estimation error exist, as analyzed by the theory of optimal design of experiments [20]. The most common approach is to find an estimator that minimizes the trace of the error covariance

$$e_1 := \mathrm{trace}(\mathbf{R}_e^{(i)}), \tag{20}$$

which corresponds to the minimization of the Mean Square Error (MSE). Other common error metrics based on the error covariance matrix are the largest eigenvalue

$$e_2 := \lambda_{\max}(\mathbf{R}_e^{(i)}), \tag{21}$$

the log determinant

$$e_3 := \log\det(\widehat{\mathbf{R}}_e^{(i)}), \tag{22}$$

and the inverse of the trace of its inverse

$$e_4 := \left[\mathrm{trace}\left(\widehat{\mathbf{R}}_e^{(i)^{-1}}\right)\right]^{-1}. \tag{23}$$

Notice that the error metrics e_3 and e_4 are computed based on the error covariance matrix for the frequency estimator $\widehat{\mathbf{R}}_e^{(i)}$ instead of the time estimator, since $\mathbf{R}_e^{(i)}$ is a singular matrix [cf. (19)].

The results presented so far consider a generic $\mathbf{R}_w^{(i)}$, so that they can be used regardless of the color of the noise. Three particular examples of interest are presented next.

- White noise in the observed signal $\mathbf{z}_i$. This implies that $\mathbf{w}_i$ is white and $\mathbf{R}_w^{(i)} = \sigma^2\mathbf{I}$, with σ^2 denoting the noise power. In this case, the $K \times K$ matrix $\bar{\mathbf{R}}_w^{(i)}$ is given by
$$\bar{\mathbf{R}}_w^{(i)} = \sigma^2\mathbf{I}. \tag{24}$$
- White noise in the original signal $\mathbf{x}$. With $\mathbf{w}$ denoting the white additive noise present in $\mathbf{x}$, we can use the linear observation model to write $\mathbf{w}_i = \mathbf{\Psi}\mathrm{diag}(\boldsymbol{\upsilon}_i)\mathbf{V}^{-1}\mathbf{w}$. Then, the $N \times N$ error covariance matrix is simply given by $\mathbf{R}_w^{(i)} = \sigma^2\mathbf{\Psi}\mathrm{diag}(\boldsymbol{\upsilon}_i)\mathbf{V}^{-1}(\mathbf{V}^{-1})^H\mathrm{diag}(\boldsymbol{\upsilon}_i)^H\mathbf{\Psi}^H$. When the shift is a normal matrix, $\mathbf{V}$ is unitary and the previous expression reduces to $\mathbf{R}_w^{(i)} = \sigma^2\mathbf{\Psi}|\mathrm{diag}(\boldsymbol{\upsilon}_i)|^2\mathbf{\Psi}^H$. The $K \times K$ error covariance matrix is then
$$\bar{\mathbf{R}}_w^{(i)} = \sigma^2\mathbf{C}\mathbf{\Psi}|\mathrm{diag}(\boldsymbol{\upsilon}_i)|^2\mathbf{\Psi}^H\mathbf{C}^H. \tag{25}$$
Equation (25) shows not only that the noise is correlated, but also that the correlation depends on the spectrum of $\mathbf{S}$, the node collecting the observations, and the specific selection of observations.
- White noise in the active frequency coefficients $\widehat{\mathbf{x}}_K$. With $\widehat{\mathbf{w}}_K$ denoting the white additive noise present in $\widehat{\mathbf{x}}_K$, we can use the linear observation model to write $\mathbf{w}_i = \mathbf{\Psi}\mathrm{diag}(\boldsymbol{\upsilon}_i)\mathbf{E}_K\widehat{\mathbf{w}}_K = \mathbf{\Psi}_i\widehat{\mathbf{w}}_K$. It follows that the $N \times N$ and $K \times K$ error covariance matrices are $\mathbf{R}_w^{(i)} = \sigma^2\mathbf{\Psi}_i\mathbf{\Psi}_i^H$ and
$$\bar{\mathbf{R}}_w^{(i)} = \sigma^2\mathbf{C}\mathbf{\Psi}_i\mathbf{\Psi}_i^H\mathbf{C}^H. \tag{26}$$
This model can be appropriate for scenarios where the signal of interest is the output of a given "graph process" – e.g., a diffusion process – and the noise is present in the

input of that process [5], [18]. This noise model can also arise when the signal to be sampled has been previously processed with a low-pass graph filter [15], [18].

There are many other noise models that can be of interest in graph setups. Matrix $\bar{\mathbf{R}}_w^{(i)}$ can be a non-negative weighted sum of (24)-(26), e.g., if noise is present in both the original signal and the observation process. Alternatively, the noise at a specific node can be rendered dependent on the number of neighbors, which is reasonable, e.g., in distributed setups where the information of neighboring nodes is exchanged via noisy channels.

B. Selection of the sampling set

The two elements that specify the set of samples to be interpolated are: *1)* the node i that aggregates the information and *2)* the entries of $\mathbf{y}_i$ selected by $\mathbf{C}$. The design of these two elements is discussed next.

1) Selection of the sampling node: The recovery results in Section III-B show that any node i can be used to sample and recover the entire graph signal, provided that the entries of $\boldsymbol{\upsilon}_i$ corresponding to the active frequencies in $\widehat{\mathbf{x}}$ are non-zero. However, when noise is present, $\mathbf{R}_e^{(i)}$ is different for each i. In this context, it is reasonable to select as a sampling node one leading to a small error. Note that selecting the best node requires the computation of N closed-form expressions, which involve matrix inversions. In scenarios where computational complexity is a limiting factor, the structure of the noise-covariance and the interpolation matrices can be exploited to reduce the burden. E.g., when white noise is present in $\widehat{\mathbf{x}}_K$, after substituting (26) into (18) and (19), it follows that

$$\widehat{\mathbf{R}}_e^{(i)} = \sigma^2\mathbf{I}, \quad \mathbf{R}_e^{(i)} = \sigma^2\mathbf{V}_K\mathbf{V}_K^H. \tag{27}$$

Consequently, for this particular noise model, the estimator performance is independent of the node choice. This is true for every error metric [cf. (20)-(23)]. The result is intuitive: given that the noise and the signal are present in the same frequencies, it is irrelevant if a node amplifies or attenuates a particular frequency. Differently, if the white noise is present in $\mathbf{z}_i$, we can substitute (24) into (18) to obtain

$$\widehat{\mathbf{R}}_e^{(i)} = \sigma^2\big(\mathbf{E}_K^H\text{diag}(\boldsymbol{\upsilon}_i)^H\boldsymbol{\Psi}^H\mathbf{C}^H\mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{E}_K\big)^{-1}. \tag{28}$$

Thus, if we are interested in minimizing, e.g., the error metric e_4 [cf. (23)], our objective may be reformulated as finding the optimal node i^* such that

$$i^* = \arg\max_i \ \text{trace}\left(\mathbf{E}_K^H\text{diag}(\boldsymbol{\upsilon}_i)^H\boldsymbol{\Psi}^H\mathbf{C}^H\mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{E}_K\right). \tag{29}$$

For a selection matrix of the form $\mathbf{C} = \mathbf{C}_K(n_0, N_0)$ (cf. Remark 1), the k-th diagonal element of the matrix in (29) can be written as $|[\boldsymbol{\upsilon}_i]_k|^2\sum_{m=0}^{K-1}|\lambda_k|^{2\,(n_0+mN_0)}$. The trace is simply the sum of those elements, so that the closed form of a geometric series can be used to rewrite (29) as

$$i^* = \arg\max_i \ \sum_{k=1}^{K}|[\boldsymbol{\upsilon}_i]_k|^2\frac{|\lambda_k|^{2n_0}-|\lambda_k|^{2(n_0+N_0K)}}{1-|\lambda_k|^{2N_0}}. \tag{30}$$

Thus, the optimal sampling node i^* will be one with large values of $|[\boldsymbol{\upsilon}_i]_k|$ for the active frequencies $k \leq K$. The relative importance of frequency k is given by the fraction in (30), which depends on the modulus of the associated eigenvalue and the structure of $\mathbf{C}$ (values of n_0 and N_0).

2) Design of the sample selection: For general sampling schemes, designing $\mathbf{C}$ is an inherently combinatorial problem, where the set of candidate matrices $\mathcal{C}$ has cardinality N choose K. The approach when using aggregation sampling is to leverage the Vandermonde structure – together with the error metrics in (20)-(23) and the noise models in (24)-(26) – to render the optimization of $\mathbf{C}$ tractable.

We start by recalling that any matrix in the set of admissible selection matrices $\mathcal{C}_K$ defined in Remark 1 is guaranteed to lead to a feasible recovery. The cardinality of $\mathcal{C}_K$ is much smaller than that of $\mathcal{C}$: N_0 can take at most N/K values, and n_0 at most $(N - N_0(K-1))$. This leads to a significant reduction of the computational burden. Moreover, in some cases the noise structure can be exploited to readily determine the optimal observation strategy. E.g., for the case of white noise in $\widehat{\mathbf{x}}_K$, it is immediate to see that the performance is independent of the sample-selection scheme [cf. (27)]. For the case where the white noise is present in $\mathbf{z}_i$, let us assume that $\mathbf{C} = \mathbf{C}_K(n_0, N_0)$, where N_0 is fixed and we want to design n_0.

If we adopt e_3 in (22) as our error metric, the goal is to find the value n_0^* that minimizes $\det\big(\widehat{\mathbf{R}}_e^{(i)}\big)$. To achieve this, consider two different selection matrices $\mathbf{C}_A = \mathbf{C}_K(n_0, N_0)$ and $\mathbf{C}_B = \mathbf{C}_K(n_0+1, N_0)$. Using (28) and assuming without loss of generality that $\sigma^2 = 1$, the error covariance for $\mathbf{C}_B$ is given by $\widehat{\mathbf{R}}_{e,B}^{(i)} = (\mathbf{E}_K^H\text{diag}(\boldsymbol{\upsilon}_i)^H\boldsymbol{\Psi}^H\mathbf{C}_B^H\mathbf{C}_B\boldsymbol{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{E}_K)^{-1}$. A similar expression can be written for $\widehat{\mathbf{R}}_{e,A}^{(i)}$. Since $\boldsymbol{\Psi}$ is Vandermonde, it is not difficult to show that $\boldsymbol{\Psi}^H\mathbf{C}_B^H\mathbf{C}_B\boldsymbol{\Psi}$ can be written as $\boldsymbol{\Lambda}^H\boldsymbol{\Psi}^H\mathbf{C}_A^H\mathbf{C}_A\boldsymbol{\Psi}\boldsymbol{\Lambda}$. This implies that

$$\begin{aligned}\widehat{\mathbf{R}}_{e,B}^{(i)^{-1}} &= \mathbf{E}_K^H\boldsymbol{\Lambda}^H\text{diag}(\boldsymbol{\upsilon}_i)^H\boldsymbol{\Psi}^H\mathbf{C}_A^H\mathbf{C}_A\boldsymbol{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\boldsymbol{\Lambda}\mathbf{E}_K \\ &= (\mathbf{E}_K^H\boldsymbol{\Lambda}^H\mathbf{E}_K)\widehat{\mathbf{R}}_{e,A}^{(i)^{-1}}(\mathbf{E}_K^H\boldsymbol{\Lambda}\mathbf{E}_K).\end{aligned} \tag{31}$$

For the first equality we have used that the product of diagonal matrices is commutative and for the second one that right and left multiplying by the canonical matrix amounts to selecting the columns and rows of the multiplied matrix. Using (31), we have that

$$\det\big(\widehat{\mathbf{R}}_{e,A}^{(i)}\big) = \det\big(\widehat{\mathbf{R}}_{e,B}^{(i)}\big)\prod_{k=1}^{K}|\lambda_k|^2, \tag{32}$$

which results in the following optimal strategy for the solution of e_3: if $\prod_{k=1}^{K}|\lambda_k|^2 \leq 1$ then $n_0^* = 1$, otherwise n_0^* should be as large as possible; see Remark 2. Equivalently, the optimal strategy states that if one application of $\mathbf{S}$ has an overall effect of amplification in the active frequencies, then we should aim to apply it as many times as possible, whereas if the opposite is true, we should avoid its application. As expected, the optimal design of $\mathbf{C}$ given by (32) depends on the topology of the graph (spectrum of $\mathbf{S}$) and the properties of $\mathbf{x}$ (set of active frequencies).

One can also look at selection matrices that are not in $\mathcal{C}_K$. In that case, the Vandermonde structure cannot be leveraged and the problem has to be formulated as a binary optimization over

$\mathbf{C} \in \mathcal{C}$. Although of interest, developing approximate solutions for $\mathbf{C}$ that exploit the structure of aggregation sampling is out of the scope of this paper and is left as future work.

It is worth stressing that the selection matrix $\mathbf{C}$ that minimizes the error does not have to be the same for all nodes. Hence, both the selection of the sampling node and the sampling shifts can be combined to obtain the best local reconstruction across all nodes in the graph.

Remark 2: Designing $\mathbf{C}$ entails selecting K out of the N entries in $\mathbf{y}_i$. However, $\mathbf{y}_i$ has only N entries because $\mathbf{Y}$ has only N columns [cf. (2)]. Strictly speaking, this restriction is not required and more columns could be added to $\mathbf{Y}$. As a matter of fact, if for a given *noisy* graph signal applying $\mathbf{S}$ attenuates the noise while amplifying the signal, the sampling procedure will benefit from further applications of $\mathbf{S}$, even beyond the size of the graph N. In practice, the maximum number of applications will be limited by the computational and signaling costs associated with the application of the shift.

V. Identifying the support of the graph signal

In the previous sections, it has been assumed that the frequency support of $\mathbf{x}$ corresponded to the K principal eigenvectors, which are the ones associated with the largest eigenvalues. However, the results presented also hold true as long as the basis support, i.e., the frequencies that are present in $\mathbf{x}$, are known. To be specific, let $\mathcal{K} := \{k_1, \ldots, k_K\}$ denote the set of indices where the signal $\mathbf{x}$ is sparse and, based on it, define the $N \times K$ matrices $\mathbf{V}_{\mathcal{K}} := [\mathbf{v}_{k_1}, \ldots, \mathbf{v}_{k_K}]$ and $\mathbf{E}_{\mathcal{K}} := [\mathbf{e}_{k_1}, \ldots, \mathbf{e}_{k_K}]$. Then, all the results presented so far hold true if $\mathbf{V}_K$ is replaced with $\mathbf{V}_{\mathcal{K}}$, and $\mathbf{E}_K$, when used to select the active frequencies, is replaced with $\mathbf{E}_{\mathcal{K}}$.

A related but more challenging problem is to design the sampling and interpolation procedures when the frequency support $\mathcal{K}$ is not known. Generically, this problem falls into the class of sparse signal reconstruction [11], [13], [14]. However, the particularities of our setup can be leveraged to achieve stronger results.

A. Noiseless joint recovery and support identification

Consider the noiseless aggregation sampling of Section III-B, where we know that $\widehat{\mathbf{x}}$ is K-sparse but we do not know $\mathcal{K}$ [cf. (11)]

$$\bar{\mathbf{y}}_i = \mathbf{C}\mathbf{y}_i = \mathbf{C}\mathbf{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\widehat{\mathbf{x}}. \tag{33}$$

When the support is known, it was shown that a matrix $\mathbf{C}$ that selects the first K rows of $\mathbf{\Psi}$ is enough for perfect reconstruction (cf. Proposition 1).

If we reformulate the recovery problem as

$$\begin{aligned} \widehat{\mathbf{x}}^* := \arg\min_{\widehat{\mathbf{x}}} \quad & ||\widehat{\mathbf{x}}||_0 \\ \text{s.t.} \quad & \bar{\mathbf{y}}_i = \mathbf{C}\mathbf{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\widehat{\mathbf{x}}, \end{aligned} \tag{34}$$

for the unknown support case, selecting K samples of $\mathbf{y}_i$ does not imply that the solution to (34) is unique. Indeed, guaranteeing identifiability in this case requires selecting a higher number of rows (samples) [11]. The following proposition, whose proof leverages the Vandermonde structure of $\mathbf{\Psi}$, states this result formally. To simplify notation, we assume that $K \leq N/2$, but the result holds true for any K.

Proposition 2: *Let $\mathbf{x}$ and $\mathbf{C}$ be, respectively, a bandlimited graph signal with at most K non-zero frequency components and a selection matrix with $2K$ rows of the form $\mathbf{C} = \mathbf{C}_{2K}(n_0, N_0)$ (cf. Remark 1). Then, if all the entries in $\boldsymbol{\upsilon}_i$ are non-zero and all the eigenvalues of $\mathbf{S}$ are non-zero while satisfying that $\lambda_k^{N_0} \neq \lambda_{k'}^{N_0}$ for all $k \neq k'$, it holds that*
i) the solution to (34) *is unique; and*
ii) the original graph signal can be recovered as $\mathbf{x} = \mathbf{V}\widehat{\mathbf{x}}^*$.

Proof: The proof proceeds in two steps. The first step is to show that the $2K \times N$ matrix $\mathbf{M} := \mathbf{C}\mathbf{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)$ has full spark [11], i.e., that any selection of $2K$ of its columns has rank $2K$ and, hence, it leads to an invertible $2K \times 2K$ matrix. To prove this, let $\mathcal{F} = \{f_1, \ldots, f_{2K}\}$ be a set containing the indices of the selected columns and define the $N \times 2K$ canonical matrix $\mathbf{E}_{\mathcal{F}} = [\mathbf{e}_{f_1}, \ldots, \mathbf{e}_{f_{2K}}]$. Using this notation, the matrix containing the columns of $\mathbf{M}$ indexed by $\mathcal{F}$ is $\mathbf{M}\mathbf{E}_{\mathcal{F}}$, which can be alternatively written as

$$\mathbf{M}\mathbf{E}_{\mathcal{F}} = \mathbf{C}\mathbf{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{E}_{\mathcal{F}} = (\mathbf{C}\mathbf{\Psi}\mathbf{E}_{\mathcal{F}})(\mathbf{E}_{\mathcal{F}}^T\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{E}_{\mathcal{F}}). \tag{35}$$

This shows that $\mathbf{M}\mathbf{E}_{\mathcal{F}}$ is invertible because it can be written as the product of two invertible matrices. The latter is true because: a) conditions $\mathbf{C} = \mathbf{C}_{2K}(n_0, N_0)$, $\lambda_k^{N_0} \neq \lambda_{k'}^{N_0}$ for all $k \neq k'$, and $\lambda_k \neq 0$ for all k guarantee that $(\mathbf{C}\mathbf{\Psi}\mathbf{E}_{\mathcal{F}})$ is invertible because it is a product of a diagonal and a Vandermonde full-rank matrices (cf. Remark 1); and b) condition $[\boldsymbol{\upsilon}_i]_k \neq 0$ for all k guarantees that $(\mathbf{E}_{\mathcal{F}}^T\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{E}_{\mathcal{F}})$ is an invertible diagonal matrix. This is true for any $\mathcal{F}$. The second step is to show that $2K$ observations guarantee identifiability. To see why this is the case, assume that two different feasible solutions $\widehat{\mathbf{x}}_A$ and $\widehat{\mathbf{x}}_B$ exist. This would imply that $\mathbf{M}(\widehat{\mathbf{x}}_A - \widehat{\mathbf{x}}_B) = \mathbf{0}$. Nevertheless, the vector $(\widehat{\mathbf{x}}_A - \widehat{\mathbf{x}}_B)$ has, at most, $2K$ non-zero components and any choice of $2K$ columns of $\mathbf{M}$ generates a full rank square matrix, which forces $\widehat{\mathbf{x}}_A = \widehat{\mathbf{x}}_B$ and contradicts the assumption of multiple solutions. ■

This result reinforces the intuition that bandlimited graph signals can be recovered by observing a local neighborhood. Suppose that $\mathbf{x}$ is a bandlimited signal with $K = 1$, which implies that $\mathbf{x}$ can be written as $\mathbf{x} = \alpha\mathbf{v}_k$. If the value of k is known, then node i can interpolate the entire signal as $\mathbf{x} = (x_i/[\mathbf{v}_k]_i)\mathbf{v}_k$. If the support is not known, one sample is not enough, but $2K = 2$ samples suffice. Note that the first two shifts, which are linear combinations of the signal values within the one-hop neighborhood of i, yield $[\mathbf{y}_i]_1 = x_i$ and $[\mathbf{y}_i]_2 = S_{ii}x_i + \sum_{j \in \mathcal{N}_i} S_{ij}x_j = \lambda_{\hat{k}}x_i$. Then, node i can identify the active frequency by finding the frequency index $\hat{k}$ satisfying $\lambda_{\hat{k}} = [\mathbf{y}_i]_2/[\mathbf{y}_i]_1$. Once $\hat{k}$ is known, the corresponding frequency coefficient can be estimated as before and the entire graph signal is given by $\mathbf{x} = (x_i/[\mathbf{v}_{\hat{k}}]_i)\mathbf{v}_{\hat{k}}$.

From a computational perspective, the presence of the ℓ_0-norm in (34) renders the optimization non-convex, thus challenging to solve. A straightforward way to convexify it is to replace the ℓ_0-norm with a ℓ_1-norm. Conditions under which this process is guaranteed to identify the frequency support can be found by analyzing the coherence and the restricted

isometry property (RIP) of matrix $\mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)$ [11], [12]. Unfortunately, determining the conditioning of all submatrices of a deterministic matrix (and, hence, the RIP) is challenging [21]. For aggregation sampling, the coherence of the matrix $\mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)$ is not hard to find and it depends on the most similar pair of eigenvalues in $\boldsymbol{\Lambda}$.

B. Noisy joint recovery and support identification

If noise is present and the frequency support of the signal is unknown, the (K-sparse) least squares estimate of $\widehat{\mathbf{x}}$ can be found as the solution to the following optimization problem

$$\widehat{\mathbf{x}}^* := \arg\min_{\widehat{\mathbf{x}}} \|(\bar{\mathbf{R}}_w^{(i)})^{-1/2}\big(\mathbf{C}\mathbf{y}_i - \mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\widehat{\mathbf{x}}\big)\|_2^2 \quad (36)$$
$$\text{s.t.} \qquad ||\widehat{\mathbf{x}}||_0 \leq K,$$

where the matrix multiplication $(\bar{\mathbf{R}}_w^{(i)})^{-1/2}$ in the objective accounts for the fact of the noise being colored. As in the noiseless case, a straightforward approach to convexify the problem is to replace the ℓ_0-norm with the ℓ_1-norm and solve the problem $\widehat{\mathbf{x}}_1^* := \arg\min_{\widehat{\mathbf{x}}} \|(\bar{\mathbf{R}}_w^{(i)})^{-1/2}\big(\mathbf{C}\mathbf{y}_i - \mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{\upsilon}_i)\widehat{\mathbf{x}}\big)\|_2^2 + \gamma||\widehat{\mathbf{x}}||_1$ for different values of the parameter γ.

VI. Space-shift sampling of graph signals

This section presents an alternative – more general – sampling setup that combines the selection sampling presented in Section III-A with the aggregation sampling proposed in Section III-B, termed *space-shift sampling*.

Let us start by defining $\mathbf{Z} = \mathbf{Y} + \mathbf{W}$ as the noisy counterpart of $\mathbf{Y}$ [cf. (2)]. Note that the i-th row of $\mathbf{Z}$ corresponds to $\mathbf{z}_i$, the observed shifted signal at node i. We are now interested in collecting samples at different nodes and shifts, i.e., we want to sample matrix $\mathbf{Z}$. To do so, we first define the vectorized version of $\mathbf{Z}$ as $\underline{\mathbf{z}} = \text{vec}(\mathbf{Z}^T)$. Recall that signal $\mathbf{z}_i$ can be related to $\widehat{\mathbf{x}}_K$ via $\mathbf{z}_i = \boldsymbol{\Psi}\mathbf{E}_K\text{diag}(\bar{\boldsymbol{\upsilon}}_i)\widehat{\mathbf{x}}_K + \mathbf{w}_i$ [cf. (13)], where $\bar{\boldsymbol{\upsilon}}_i = \mathbf{E}_K^T\boldsymbol{\upsilon}_i$. To write a similar equation relating $\underline{\mathbf{z}}$ to $\widehat{\mathbf{x}}_K$, we need to define the $N^2 \times N$ matrix $\boldsymbol{\Upsilon} := [\text{diag}(\boldsymbol{\upsilon}_1), \ldots, \text{diag}(\boldsymbol{\upsilon}_N)]^T$ and its corresponding reduced $NK \times K$ matrix $\bar{\boldsymbol{\Upsilon}} := [\text{diag}(\bar{\boldsymbol{\upsilon}}_1), \ldots, \text{diag}(\bar{\boldsymbol{\upsilon}}_N)]^T$. Based on this, $\underline{\mathbf{z}}$ can be written as

$$\underline{\mathbf{z}} = \Big(\mathbf{I} \otimes (\boldsymbol{\Psi}\mathbf{E}_K)\Big)\bar{\boldsymbol{\Upsilon}}\widehat{\mathbf{x}}_K + \underline{\mathbf{w}}, \quad (37)$$

where $\underline{\mathbf{w}}$ is a vector of length N^2 obtained by concatenating the noise vectors $\mathbf{w}_i$ for all nodes i. This implies that (37) is a system of N^2 linear equations with $K < N$ variables. Thus, our objective is to select K of these equations in order to estimate $\widehat{\mathbf{x}}_K$ – and, hence, $\mathbf{x}$ through (17) – while minimizing the error introduced by the noise $\underline{\mathbf{w}}$. Suppose that for a given node index i we consider the problem of selecting K equations out of the N equations in positions $\{(i-1)N + n\}_{n=1}^N$, then space-shift sampling reduces to local aggregation sampling at node i. Similarly, if we restrict ourselves to select K equations out of the N equations in positions $\{1 + (n-1)N\}_{n=1}^N$, the problem reduces to selection sampling. In this sense, the formulation in (37) is more general. To implement the selection of the K equations out of the N^2 options in (37), we use a binary selection matrix $\mathbf{C}$ as done in previous sections but, in this case, the size of $\mathbf{C}$ is $K \times N^2$. The reduced square system of linear equations can then be written as [cf. (13)]

$$\underline{\bar{\mathbf{z}}} = \mathbf{C}\Big(\mathbf{I} \otimes (\boldsymbol{\Psi}\mathbf{E}_K)\Big)\bar{\boldsymbol{\Upsilon}}\widehat{\mathbf{x}}_K + \mathbf{C}\underline{\mathbf{w}}. \quad (38)$$

The error covariance matrices $\widehat{\mathbf{R}}_{\underline{e}}$ and $\mathbf{R}_{\underline{e}}$ computed based on the solution of (38) are [cf. (18) and (19)]

$$\widehat{\mathbf{R}}_{\underline{e}} = \Big(\bar{\boldsymbol{\Upsilon}}^H \left(\mathbf{I} \otimes (\boldsymbol{\Psi}\mathbf{E}_K)\right)^H \mathbf{C}^H$$
$$\times \left(\mathbf{C}\mathbf{R}_{\underline{w}}\mathbf{C}^H\right)^{-1} \times \mathbf{C}\left(\mathbf{I} \otimes (\boldsymbol{\Psi}\mathbf{E}_K)\right)\bar{\boldsymbol{\Upsilon}}\Big)^{-1}, \quad (39)$$

$$\mathbf{R}_{\underline{e}} = \mathbf{V}_K\widehat{\mathbf{R}}_{\underline{e}}\mathbf{V}_K^H, \quad (40)$$

where $\mathbf{R}_{\underline{w}}$ is the covariance matrix of $\underline{\mathbf{w}}$. In this case, the noise models introduced in Section IV-A are also relevant. For white noise in the observations, we have that $\mathbf{R}_{\underline{w}} = \sigma^2\mathbf{I}$; for white noise in the original signal, we have that $\mathbf{R}_{\underline{w}} = \sigma^2\left(\mathbf{I} \otimes \boldsymbol{\Psi}\right)\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H\left(\mathbf{I} \otimes \boldsymbol{\Psi}\right)^H$; and for white noise in the active frequency coefficients, we have that $\mathbf{R}_{\underline{w}} = \sigma^2\left(\mathbf{I} \otimes (\boldsymbol{\Psi}\mathbf{E}_K)\right)\bar{\boldsymbol{\Upsilon}}\bar{\boldsymbol{\Upsilon}}^H\left(\mathbf{I} \otimes (\boldsymbol{\Psi}\mathbf{E}_K)\right)^H$.

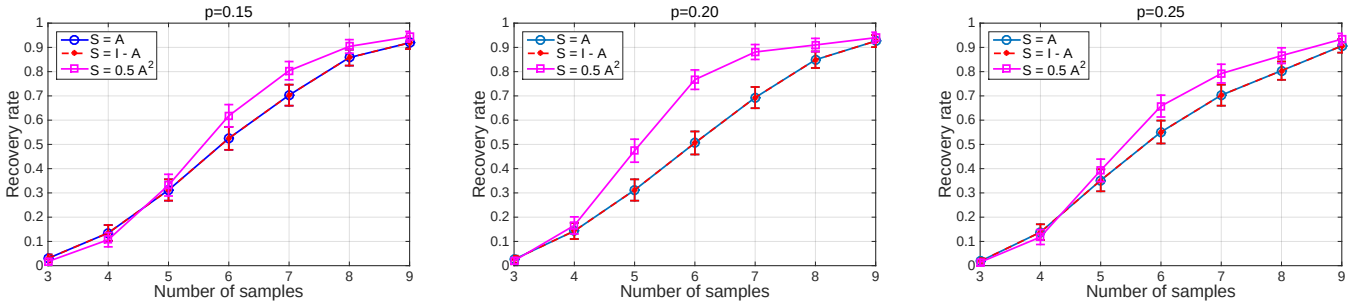
A. Structured observability pattern

In the previous discussion, no structure was assumed in the selection matrix $\mathbf{C}$. A case of particular interest is when the sampling schemes are implemented in a distributed manner using message passing. Suppose that the sampling is performed at node i. To compute $y_i^{(l)}$, the node i needs to have access to $y_j^{(l')}$ for all $j \in \mathcal{N}_i$ and $l' < l$. To simplify notation, and without loss of generality, we will assume that the sampling node is $i = 1$ and that the neighbors of $i = 1$ are $i = 2, \ldots, N_1 + 1$. Suppose also that node $i = 1$ computes L_1 shifts, from $y_1^{(0)}$ up to $y_1^{(L_1)}$. This implies that node $i = 1$ has access to $L_1 + 1$ of its own samples and to L_1 samples of each of its N_1 neighbors. The selection matrix $\mathbf{C}$ can then be written as

$$\mathbf{C} = \begin{bmatrix} \mathbf{E}_{L_1+1}^T & \mathbf{0}_{L_1+1 \times (N^2-N))} \\ \mathbf{0}_{N_1L_1 \times N} & \mathbf{I}_{N_1} \otimes \mathbf{E}_{L_1}^T \quad \mathbf{0}_{N_1L_1 \times (N^2-NN_1-N)} \end{bmatrix}. \quad (41)$$

Matrix $\mathbf{C}$ has $1 + L_1(1 + N_1)$ rows, one per observation. The first $1 + L_1$ rows correspond to the samples at node $i = 1$ and the remaining L_1N_1 to the samples at its neighbors. Note also that matrix $\mathbf{C}\left(\mathbf{I} \otimes (\boldsymbol{\Psi}\mathbf{E}_K)\right)\bar{\boldsymbol{\Upsilon}}$ is not full (row) rank. The reason is that all the samples obtained at node $i = 1$, except for the first one, are linear combinations of the samples at its neighbors. This implies that the number of frequencies that can be recovered using (41) is, at most, $1 + L_1N_1$.

Structured observation models different from the one in (41) can be also of interest. For example, one can consider setups where nodes from different parts of the graph take a few samples each and forward those samples to a central fusion center. In such a case, since the nodes gathering data need not be neighbors, the problem of some of the samples being a linear combination of the others will not necessarily be present.



(a) (b) (c)

Fig. 2: Recovery rate of bandlimited signals in random graphs with unknown frequency support. Error bars represent the standard error (three standard deviations) around the mean recovery rates. Signals are recovered via the ℓ_1-norm relaxation of problem (34) for different numbers of observations K and three different graph-shift operators $\mathbf{S}_1 = \mathbf{A}$ (blue circle), $\mathbf{S}_2 = \mathbf{I} - \mathbf{A}$ (red cross), and $\mathbf{S}_3 = 0.5\mathbf{A}^2$ (magenta square). Random graphs with different edge probabilities were considered: (a) 0.15 , (b) 0.20, and (c) 0.25.

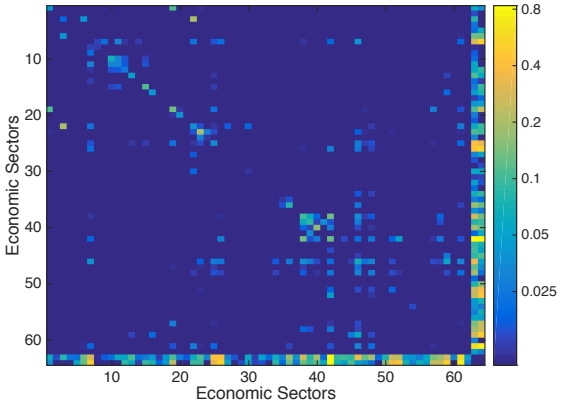


Fig. 3: Heat map of the graph-shift operator $\mathbf{S}$ of the economic network. It is sparse across the real economic sectors (from sector 1 to 62) while the synthetic sectors AV and FU are highly connected.

VII. Numerical Experiments

The purpose of this section is to illustrate and gain intuition about some of the theoretical results presented. We start by illustrating perfect recovery of synthetic noiseless graph signals when the frequency support is not known (Section VII-A). We then present results for real-world graph signals corresponding to the exchange among the different sectors of the economy of the United States. These are used to test recovery under the presence of noise (Section VII-B) as well as to illustrate the space-shift sampling method (Section VII-C).

A. Noiseless recovery and support selection

In this set of experiments we consider realizations of a symmetric Erdős-Rényi random graph with $N = 20$ nodes and edge probability p [22]. With $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}_A\mathbf{V}^H$ denoting the adjacency matrix of a specific realization, three different graph-shift operators are considered: $\mathbf{S}_1 = \mathbf{A}$, $\mathbf{S}_2 = \mathbf{I} - \mathbf{A}$, and $\mathbf{S}_3 = 0.5\mathbf{A}^2$. Notice that, even though the support of $\mathbf{S}_3$ differs from that of $\mathbf{S}_1$ and $\mathbf{S}_2$, the graph-shift operator $\mathbf{S}_3$ still preserves the notion of locality as defined by a two-hop neighborhood. Note also that the three shift operators share the same set of eigenvectors $\mathbf{V}$, so that the bandwidth K of a given signal is the same for all of them. The experiments focus on recovering a signal of bandwidth $K = 3$ whose frequency support is unknown using the ℓ_1-norm relaxation. To assess recovery, Fig. 2 plots the success rate – fraction of realizations for which the actual signal was recovered – for graph-shifts $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$, and different numbers of observations. Ten random graph realizations and five signal realizations per graph were considered. For each of these realizations, every node tries to reconstruct the signal. Points in the plots represent the average success rate, while error bars stand for 3 standard deviations. Each of the three panels corresponds to realizations generated using different edge probabilities: $p = 0.15$, $p = 0.20$, and $p = 0.25$. The recovery rate for $\mathbf{S}_3 = 0.5\mathbf{A}^2$ is consistently higher than the one for the other shift operators considered. This is not surprising: when squaring the adjacency matrix to generate $\mathbf{S}_3$, the dissimilarity between any pair of eigenvalues is increased, which reduces the matrix coherence associated with $\mathbf{S}_3 = 0.5\mathbf{A}^2$ and facilitates sparse recovery (cf. last paragraph in Section V-A).

B. Recovery in the presence of noise

The Bureau of Economic Analysis of the U.S. Department of Commerce publishes a yearly table of input and outputs organized by economic sectors [23]. More precisely, we have a set $\mathcal{N}$ of 62 industrial sectors as defined by the North American Industry Classification System and a similarity function $U : \mathcal{N} \times \mathcal{N} \to \mathbb{R}_+$ where $U(i, i')$ represents how much of the production of sector i, expressed in trillions of dollars per year, was used as an input of sector i' on average during years 2008, 2009, and 2010. Moreover, for each sector we are given two economic markers: the added value (AV) generated and the level of production destined to the market of final users (FU). Thus, we define a graph on the set of $N = 64$ nodes comprising the original 62 sectors plus the two synthetic ones (AV and FU) and an associated symmetric graph-shift operator $\bar{\mathbf{S}}$ defined as $\bar{\mathbf{S}}_{ij} = (U(i, j) + U(j, i))/2$. We then threshold $\bar{\mathbf{S}}$ in order to increase its sparsity by setting to 0 all the values lower than 0.01, giving rise to the shift $\mathbf{S}$ in Fig. 3. The shift $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ is normal given that it is symmetric. Associated with this graph, we consider the signal $\mathbf{x} \in \mathbb{R}^{64}$ that collects the total production – in trillion of dollars – of each sector (including AV and FU) during year 2011. Signal $\mathbf{x}$ is approximately bandlimited in $\mathbf{S}$ since most of the elements of $\widehat{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$ are close to zero; see Fig. 4a (top). In particular,

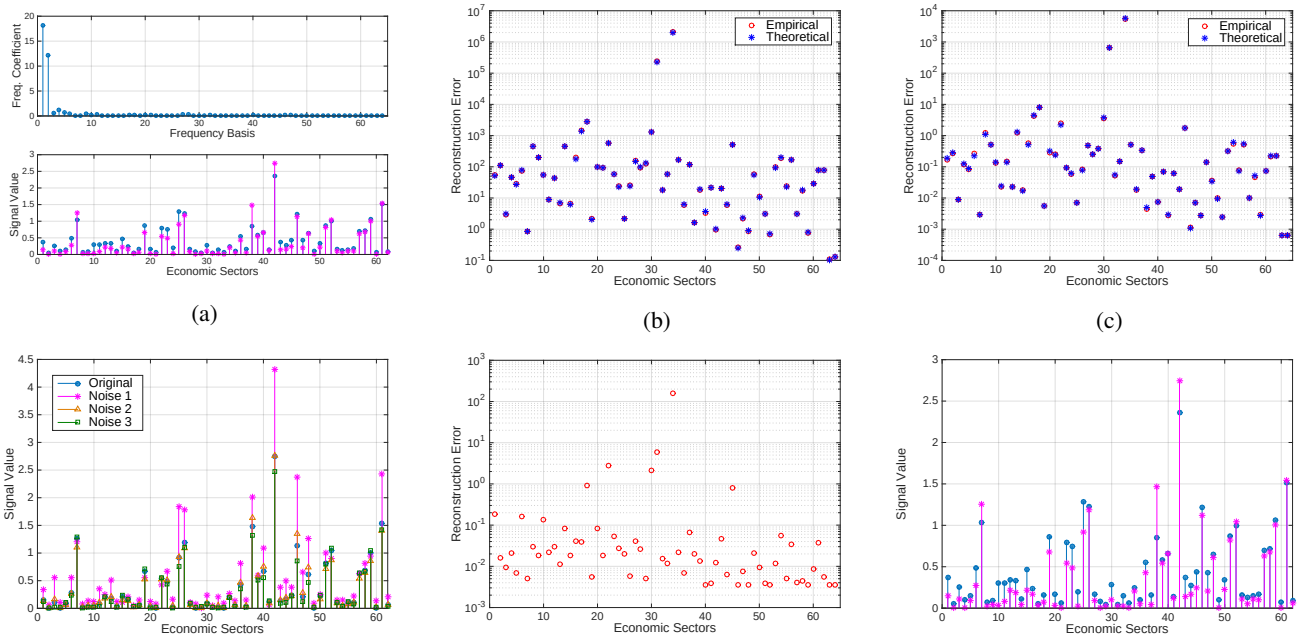


Fig. 4: (a) Top: Frequency representation of the graph signal $\mathbf{x}$ in the basis of eigenvectors of the graph-shift $\mathbf{S}$. The signal is approximately bandlimited. Bottom: Signal $\mathbf{x}$ (blue) and its reconstruction $\mathbf{x}_4$ (magenta) when keeping only the first 4 frequency components. (b) Empirical (red circle) and theoretical (blue star) reconstruction errors for different sampling nodes when white noise is added to the observed signal. (c) Empirical (red circle) and theoretical (blue star) reconstruction errors when white noise is added directly to the signal $\mathbf{x}_4$. (d) Signal $\mathbf{x}_4$ (blue) and the best reconstruction achieved when sampling an economic sector for the three types of noise considered: white noise in the observations (magenta), white noise in the signal (orange) and white noise in the active frequency components (green). (e) Reconstruction errors for different sampling nodes when interpolating signal $\mathbf{x}$ based on four observations. (f) Signal $\mathbf{x}$ (blue) and the best reconstruction (magenta) achieved when performing local aggregation sampling of economic sectors.

the reconstructed signal $\mathbf{x}_4 = \mathbf{V}_4\widehat{\mathbf{x}}_4$ obtained by just keeping the first $K=4$ frequency coefficients attains a reconstruction error of 3.5×10^{-3} computed as the ratio between the energy of the error and the energy of the original signal. This small reconstruction error is nonetheless noticeable when plotting the original signal $\mathbf{x}$ and the reconstructed one $\mathbf{x}_4$; see Fig. 4a (bottom). To present a reasonable scale for illustration, sectors AV and FU are not included in Fig. 4, since $\mathbf{x}_4$ takes out-of-scale values for these sectors.

In Sections VII-B1 to VII-B3 we consider the bandlimited signal $\mathbf{x}_4$ as noiseless and add different types of Gaussian noise to analyze the interpolation performance at different nodes. Differently, in Section VII-B4 we interpret $\mathbf{x}$ as a noisy version of $\mathbf{x}_4$ and analyze the reconstruction error when interpolating $\mathbf{x}$ from just 4 samples.

1) White noise in the observed signal: We perform aggregation sampling of multiple noisy versions of $\mathbf{x}_4$ via successive applications of $\mathbf{S}$ at different economic sectors (nodes). The noisy versions of $\mathbf{x}_4$ are generated by adding noise to the observed signal as described in (24). The noise power σ^2 is the same for all nodes and its value is set so that, when averaged across nodes, the linear signal to noise ratio (SNR) for the first, second, third and fourth observations in each node is 2, 10, 50, and 250, respectively. The increase in the SNR is attributable to the fact that successive applications of $\mathbf{S}$ increase the signal magnitude while σ^2 remains constant. In Fig. 4b we plot both the empirical reconstruction error at different nodes (averaged across 1,000 noisy realizations of $\mathbf{x}_4$) and the theoretical average error given by the trace of $\mathbf{R}_e^{(i)}$ [cf. (19)-(20)]. We first observe that for all 64 nodes the theoretical and empirical errors coincide. A more surprising observation is that the quality of the reconstruction depends heavily on the node collecting the samples. The error is minimized when the samples are taken at the synthetic sectors AV and FU. This is reasonable since these two nodes – unlike other sectors – are closely related to every other sector of the economy (cf. Fig. 3). Furthermore, the sectors achieving the worst reconstruction errors are 'Publishing Industries' and 'Ground Passenger Transportation' corresponding to nodes 34 and 31. The heat map in Fig. 3 shows that these two nodes (especially 31) are poorly connected to the rest of the network. A more rigorous explanation can be provided by analyzing vectors $\bar{\boldsymbol{\upsilon}}_{34} = \mathbf{E}_4^T\boldsymbol{\upsilon}_{34}$ and $\bar{\boldsymbol{\upsilon}}_{31} = \mathbf{E}_4^T\boldsymbol{\upsilon}_{31}$ (cf. Lemma 1). Even though both vectors have all four components different from zero, which guarantees perfect reconstruction in the noiseless case (cf. Proposition 1), they possess an element whose absolute value is in the order of 10^{-4}, increasing the sensitivity of the reconstruction in the presence of noise. For all other nodes the smallest element of $\bar{\boldsymbol{\upsilon}}_i$ is at least one order of magnitude larger, hence sensitivity to noise is much smaller. Fig. 4d presents the reconstruction obtained by aggregation sampling in node 46. This node, which is well connected and corresponds to 'Professional Services', is the best among real economic sectors – i.e., excluding AV and FU –, leading to an error of 0.26. This suggests that node connectivity (centrality)

TABLE I: Minimum and median reconstruction errors for different sampling strategies.

Sampling strategy				Min. error	Median error
$[\mathbf{x}]_i$	$[\mathbf{Sx}]_i$	$[\mathbf{S^2x}]_i$	$[\mathbf{S^3x}]_i$	.0035	.019
$[\mathbf{x}]_i$	$[\mathbf{x}]_j$	$[\mathbf{x}]_k$	$[\mathbf{x}]_l$	.0039	4.2
$[\mathbf{Sx}]_i$	$[\mathbf{Sx}]_j$	$[\mathbf{Sx}]_k$	$[\mathbf{Sx}]_l$	.0035	.030
$[\mathbf{S^2x}]_i$	$[\mathbf{S^2x}]_j$	$[\mathbf{S^2x}]_k$	$[\mathbf{S^2x}]_l$	.0035	.0055
$[\mathbf{S^3x}]_i$	$[\mathbf{S^3x}]_j$	$[\mathbf{S^3x}]_k$	$[\mathbf{S^3x}]_l$	.0035	.0040
$[\mathbf{x}]_i$	$[\mathbf{Sx}]_i$	$[\mathbf{x}]_j$	$[\mathbf{Sx}]_j$	.0035	.039

The first row lists the results for the aggregation sampling scheme. The second row corresponds to selection sampling, i.e., observing the value of the signal $\mathbf{x}$ at 4 different nodes i, j, k, l. The remaining strategies correspond to instances of the more general space-shift sampling presented in Section VI.

can be leveraged for prediction purposes or used as input for heuristics to select a good sampling node.

2) White noise in the original signal: Here we consider that noise is added to $\mathbf{x}_4$ [cf. (25)] and quantify the reconstruction error when aggregation sampling is performed at each of the 64 nodes. The noise power σ^2 is set to induce a linear SNR of 10^2. As was the case in the previous section, the average empirical error across 1,000 realizations matches closely our theoretical estimates; see Fig. 4c. Moreover, the specific nodes that lead to a good (bad) interpolation performance are very similar to those in the previous noise model. Indeed, sectors 34 and 31 have the highest reconstruction error whereas AV and FU attain the best reconstructions. Fig. 4d shows the best reconstruction – excluding AV and FU – which amounts to an error of 0.001 and corresponds to the sector 'Professional Services' at node 46.

3) White noise in the active frequencies: Here white noise is added to $\widehat{\mathbf{x}}_4$, as described in (26). As before, the noise power σ^2 is set to induce a linear SNR of 10^2. The average empirical reconstruction error associated with each node (across 1,000 noisy realizations of $\mathbf{x}_4$) is the same regardless of the node. This validates the analysis in (27), which stated that for this noise model the quality of the reconstruction is node independent. In Fig. 4d we present an example of such a reconstruction, achieving an error of 0.01.

4) Real-world noisy signal: We interpret the graph signal $\mathbf{x}$ as a noisy realization of a signal of bandwidth 4. Hence, our goal is to obtain the best reconstruction of $\mathbf{x}$ based on 4 observations. As described in (19) and shown before, interpolation performance is highly node dependent. Indeed, the reconstruction error when keeping the first 4 observations at each node spans 5 orders of magnitude depending on the sampling node, although for most nodes it is contained between 10^{-3} and 10^{-1}; see Fig. 4e. The best reconstruction among the real sectors is achieved by 'Insurance Carriers' (node 40). The best and the median reconstructions are acceptable, attaining errors of 0.0035 and 0.019, respectively. Fig. 4f depicts the best reconstruction.

C. Space-shift sampling

In Section VII-B4 we analyzed the accuracy of reconstructing the U.S. economic activity using as input the observations generated after running aggregation sampling at different nodes (economic sectors). The minimum and median reconstruction errors are presented in the first row of Table I, where the reconstruction error is quantified as the ratio between the energy of the error and that of the original signal. An alternative approach is to implement selection sampling, i.e. to sample the signal $\mathbf{x}$ in 4 different sectors – excluding the synthetic sectors AV and FU – and interpolate the whole signal from these 4 observations, as explained in Section III-A. Recall that reconstruction is not guaranteed for every subset of 4 nodes since we must have invertibility of $(\mathbf{CV}_K)$ [cf. (5)]. By analyzing the minimum and median reconstruction errors – see the two first rows in Table I – it is clear that the node aggregation sampling outperforms the node selection sampling. This is intuitive since most of the energy of the signal is contained in the two first frequencies [cf. Fig. 4a(top)], which are associated with the largest eigenvalues. Hence, after successive applications of the graph-shift, the error in estimating these frequencies is reduced, resulting in a smaller error in the interpolation of the whole signal.

As developed in Section VI, more general sampling strategies can be implemented. For example, we can sample the value of the signal at 4 nodes after the application of one, two or three graph-shifts. The results – listed in rows 3, 4 and 5 of Table I – reveal a significant reduction in the median error after each graph-shift application, especially when going from no applications – median error of 4.2 – to one application – median error of 0.03. This can be due to the fact that the application of $\mathbf{S}$ amplifies the frequencies associated with large eigenvalues, which are the ones present in $\mathbf{x}_4$. A different alternative is a sampling strategy that selects the original signal and the signal after one shift in two different sectors. The results, listed in the last row of Table I, show that this configuration leads to a very good reconstruction performance: 0.0035 minimum error and 0.039 median error. Note that with this sampling configuration, the two sectors are only required to compute the aggregated activity of their one-hop neighbors.

The performance attained by a specific sampling scheme depends on factors like the operating conditions of the network, the structure of the graph, the noise model, and the properties of the signal. As a general rule, when sampling an approximately bandlimited signal whose active frequencies are associated with large eigenvalues of $\mathbf{S}$, aggregation sampling is expected to give rise to a better interpolation. Successive applications of $\mathbf{S}$ amplify the active frequencies, entailing a better estimation of these frequencies and reducing the interpolation error. By contrast, when the active frequencies are associated with small eigenvalues of $\mathbf{S}$, selection sampling is preferred. Space-shift sampling strategies are useful whenever some active frequencies are related to large eigenvalues and others are related to small eigenvalues. Moreover, space-shift sampling is also a suitable alternative when the magnitudes of the eigenvalues associated with the active frequencies are unknown.

VIII. Conclusions

A novel scheme for sampling bandlimited graph signals – that admit a sparse representation in the frequency domain – was proposed. The scheme was based on the aggregation of local information at a single node after successive applications of the graph-shift operator. This contrasted with

most existing works, which focus on sampling the value of the signal observed at a subset of nodes. Our scheme was shown to be equivalent to classical sampling for directed cycle graphs whereas, for more general graphs, the Vandermonde structure of the sampling matrix was leveraged to determine the conditions for perfect reconstruction in the absence of noise. Reconstruction under correlated noise was analyzed, and design criteria to select the sampling node and shifts leading to optimal noisy reconstruction were discussed. Scenarios where the specific set of frequencies present in the bandlimited signal is not known were also investigated and connections with sparse signal reconstruction were drawn. Finally, a more general sampling scheme was presented which contained, as particular cases, the selection sampling as well as our local aggregation approach. The various sampling and interpolation scenarios were illustrated through numerical experiments in both synthetic and real-world graph signals.

REFERENCES

[1] M. Unser, "Sampling-50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[2] 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 Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

[3] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[4] X. Zhu and M. Rabbat, "Approximating signals supported on graphs," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2012, pp. 3921–3924.

[5] S. Segarra, A. G. Marques, and A. Ribeiro, "Distributed linear network operators using graph filters," *arXiv preprint arXiv:1510.03947*, 2015.

[6] A. Anis, A. Gadde, and A. Ortega, "Towards a sampling theorem for signals on arbitrary graphs," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, May 2014, pp. 3864–3868.

[7] I. Shomorony and A. S. Avestimehr, "Sampling large data on graphs," *arXiv preprint arXiv:1411.3017*, 2014.

[8] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovačević, "Signal recovery on graphs," *arXiv preprint arXiv:1411.7414*, 2014.

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

[10] X. Wang, P. Liu, and Y. Gu, "Local-set-based graph signal reconstruction," *arXiv preprint arXiv:1410.3944*, 2014.

[11] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[12] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[13] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.

[14] X. Zhu and M. Rabbat, "Graph spectral compressed sensing for sensor networks," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2012, pp. 2865–2868.

[15] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, June 2014.

[16] C. Godsil and G. Royle, *Algebraic Graph Theory*. Springer-Verlag, Graduate Texts in Mathematics, 2001, vol. 207.

[17] M. Rabbat and V. Gripon, "Towards a spectral characterization of signals supported on small-world networks," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, May 2014, pp. 4793–4797.

[18] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *arXiv preprint arXiv:1507.08364*, 2015.

[19] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[20] F. Pukelsheim, *Optimal Design of Experiments*. SIAM, 1993, vol. 50.

[21] B. Alexeev, J. Cahill, and D. G. Mixon, "Full spark frames," *J. of Fourier Analysis and Appl.*, vol. 18, no. 6, pp. 1167–1194, 2012.

[22] B. Bollobás, *Random Graphs*. Springer, 1998.

[23] Bureau of Economic Analysis, "Input-output accounts: the use of commodities by industries before redefinitions," *United States Department of Commerce*, 2011, http://www.bea.gov/industry/io_annual.htm.

Antonio G. Marques (SM'13) received the telecommunications engineering degree and the Doctorate degree, both with highest honors, from the Carlos III University of Madrid, Spain, in 2002 and 2007, respectively. In 2007, he became a faculty of the Department of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain, where he currently develops his research and teaching activities as an Associate Professor. From 2005 to 2012 and in 2015, he held different visiting positions at the University of Minnesota, Minneapolis. In 2015 he was a visitor scholar at the University of Pennsylvania.

His research interests lie in the areas of signal processing, networking and communications. His current research focuses on stochastic resource allocation for green and cognitive wireless networks, smart grids, nonlinear network optimization, and signal processing for graphs. Dr. Marques has served the IEEE in a number of posts (currently, he is an Associate Editor of the Signal Process. Letters) and his work has been awarded in several conferences and workshops.

Santiago Segarra (S'12) received the B.Sc. degree in industrial engineering with highest honors (Valedictorian) from the Instituto Tecnológico de Buenos Aires (ITBA), Argentina, in 2011 and the M.Sc. degree in electrical engineering from the University of Pennsylvania, Philadelphia, in 2014. Since 2011, he has been working towards the Ph.D. degree in the Department of Electrical and Systems Engineering at the University of Pennsylvania.

His research interests include network theory, data analysis, machine learning, and graph signal processing. Mr. Segarra received the ITBA's 2011 Best Undergraduate Thesis Award in industrial engineering, the 2011 Outstanding Graduate Award granted by the National Academy of Engineering of Argentina, and the Best Student Paper Award at the 2015 Asilomar Conference.

Geert Leus (F'13) received the MSc and PhD degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in June 1996 and May 2000, respectively. Currently, Geert Leus is an "Antoni van Leeuwenhoek" Full Professor at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, The Netherlands.

His research interests are in the area of signal processing for communications. Geert Leus received a 2002 IEEE Signal Processing Society Young Author Best Paper Award and a 2005 IEEE Signal Processing Society Best Paper Award. He is a Fellow of the IEEE. Geert Leus was the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, and an Associate Editor for the IEEE Transactions on Signal Processing, the IEEE Transactions on Wireless Communications, the IEEE Signal Processing Letters, and the EURASIP Journal on Advances in Signal Processing. Currently, he is a Member-at-Large to the Board of Governors of the IEEE Signal Processing Society and a member of the IEEE Sensor Array and Multichannel Technical Committee. He finally serves as the Editor in Chief of the EURASIP Journal on Advances in Signal Processing.

Alejandro Ribeiro (M'07) received the B.Sc. degree in electrical engineering from the Universidad de la Republica Oriental del Uruguay, Montevideo, in 1998 and the M.Sc. and Ph.D. degree in electrical engineering from the Department of Electrical and Computer Engineering, the University of Minnesota, Minneapolis in 2005 and 2007. From 1998 to 2003, he was a member of the technical staff at Bellsouth Montevideo. After his M.Sc. and Ph.D studies, in 2008 he joined the University of Pennsylvania (Penn), Philadelphia, where he is currently the Rosenbluth Associate Professor at the Department of Electrical and Systems Engineering.

His research interests are in the applications of statistical signal processing to the study of networks and networked phenomena. His current research focuses on wireless networks, network optimization, learning in networks, networked control, robot teams, and structured representations of networked data structures. Dr. Ribeiro received the 2014 O. Hugo Schuck best paper award, the 2012 S. Reid Warren, Jr. Award presented by Penn's undergraduate student body for outstanding teaching, the NSF CAREER Award in 2010, and student paper awards at the 2013 American Control Conference (as adviser), as well as the 2005 and 2006 International Conferences on Acoustics, Speech and Signal Processing. Dr. Ribeiro is a Fulbright scholar and a Penn Fellow.