

Sampling of graph signals via randomized local aggregations

*Original*

Sampling of graph signals via randomized local aggregations / Valsesia, Diego; Fracastoro, Giulia; Magli, Enrico. - In: IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS. - ISSN 2373-776X. - 5:2(2019), pp. 348-359. [10.1109/TSIPN.2018.2869354]

*Availability:*

This version is available at: 11583/2713424 since: 2021-03-16T13:17:08Z

*Publisher:*

Institute of Electrical and Electronics Engineers Inc.

*Published*

DOI:10.1109/TSIPN.2018.2869354

*Terms of use:*

openAccess

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

IEEE postprint/Author's Accepted Manuscript

©2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

# Sampling of Graph Signals via Randomized Local Aggregations

Diego Valsesia, *Member, IEEE*, Giulia Fracastoro, *Member, IEEE*, and Enrico Magli, *Fellow, IEEE*

**Abstract**—Sampling of signals defined over the nodes of a graph is one of the crucial problems in graph signal processing. While in classical signal processing sampling is a well defined operation, when we consider a graph signal many new challenges arise and defining an efficient sampling strategy is not straightforward. Recently, several works have addressed this problem. The most common techniques select a subset of nodes to reconstruct the entire signal. However, such methods often require the knowledge of the signal support and the computation of the sparsity basis before sampling. Instead, in this paper we propose a new approach to this issue. We introduce a novel technique that combines localized sampling with compressed sensing. We first choose a subset of nodes and then, for each node of the subset, we compute random linear combinations of signal coefficients localized at the node itself and its neighborhood. The proposed method provides theoretical guarantees in terms of reconstruction and stability to noise for any graph and any orthonormal basis, even when the support is not known.

**Index Terms**—Graph signal processing, Sampling, Random projections, Compressed sensing

## I. INTRODUCTION

The analysis of signals defined over graphs is a natural extension of classical processing of time-varying signals that has emerged due to the wide array of applications where the signals of interest lie on irregular domains [1]. Fields as varied as sensor and social networks [2], [3], genetics [4], neuroscience [5], computer graphics [6] and image processing [7]–[9] may benefit from graph signal processing (GSP) techniques. The pervasiveness of big data [10] makes sampling and subsequent recovery crucial to creating efficient representations. While sampling is well understood in classical signal processing for time-varying signals, not least thanks to the recent developments in compressed sensing (CS) [11], signals whose coefficients lie on the nodes of a graph pose a unique set of challenges. Most current efforts in the field have been focused on establishing a concept of a frequency domain for graph signals as a linear transform dependent on the graph structure [12], [13], and uncovering the fundamental limits of sampling for signals having a low-dimensional structure in this transformed domain [14]. Efficient sampling of graph signals is concerned with a notion of locality of the sampling procedure where only a few nodes can be directly observed and such nodes can only return the value of the coefficient on the node itself [15]–[19] or an aggregate value determined from the coefficients on its neighbors [20], [21]. Universal sampling strategies such as CS with dense random projections

would be optimal for graph signals as well as classical signals because they are independent of the particular transform enforcing the low-dimensional prior [22]. However, dense random projections violate locality as each projection requires to combine all the coefficients across the graph. This paper shows that it is possible to combine locality with CS to sample graph signals with a low-complexity scheme that provides theoretical guarantees for perfect reconstruction and stability to noise. In particular, the proposed method of randomized local aggregations fits the line of works on aggregation sampling and consists of computing random linear combinations of the signal coefficients restricted to the neighborhood of a subset of nodes in the graph. The theoretical analysis rigorously derives conditions on the number of measurements required to guarantee perfect reconstruction and stability to noise depending on the structure of the graph and of the signal. Finally, the proposed sampling method does not need a priori knowledge of the sparsity support of the signal and can be therefore used even if it is unknown at sampling time. This is common to many problems, as discussed in the literature on CS [11], as well as on GSP [20], [23].

This paper is organized as follows. Section II presents some background on GSP and CS, and discusses relevant works in the literature. Section III introduces the proposed method performing randomized local aggregations. Section IV analyzes the method and provides theoretical guarantees. Section V reports numerical experiments comparing the proposed method with relevant methods in the literature. Finally, Section VI draws some conclusions.

### Notation

Lower-case (upper-case) bold symbols denote real-valued vectors (matrices). Lower-case letters indicate scalars. Upper-case letters denote random variables. Symbols  $\mathbb{P}$  and  $\mathbb{E}$  denote the probability and expectation operators, respectively. Calligraphic letters denote sets. Matrix  $\mathbf{A}_{\mathcal{R}}$  indicates the restriction of  $\mathbf{A}$  to rows indexed by set  $\mathcal{R}$ , while  $\mathbf{A}_{|\mathcal{C}}$  indicates restriction of  $\mathbf{A}$  to columns indexed by set  $\mathcal{C}$ .

## II. BACKGROUND

### A. Graph signal processing

We first introduce a few definitions related to graph signal processing that will be used in the remainder of the paper. Let  $\mathcal{G}(\mathcal{V}, \mathcal{E})$  denote a directed graph where  $\mathcal{V}$  is the set of nodes with  $|\mathcal{V}| = n$  and  $\mathcal{E}$  is the set of edges, being  $(i, j)$  an edge from node  $i$  to node  $j$ .

The authors are with Politecnico di Torino – Department of Electronics and Telecommunications, Italy. Email: {name.surname}@polito.it. This work is supported by the DISLO-MAN project funded by Regione Piemonte.

**Definition 1.** Given  $\mathcal{G}$ , for each  $i \in \mathcal{V}$  we define the in-neighborhood of the node  $i$  as

$$\mathcal{N}_i = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}.$$

Throughout the paper, whenever the terms neighborhood or neighbor are used, they are meant to be an in-neighborhood or an in-neighbor, respectively. For an undirected graph, the in- and out- neighborhoods coincide.

**Definition 2.** Given  $\mathcal{G}$ , a dominating set of the graph  $\mathcal{G}$  is a subset  $\mathcal{D} \subseteq \mathcal{V}$  with the following property

$$\forall i \in \mathcal{V} \exists j \in \mathcal{D} \text{ s.t. } i \in \tilde{\mathcal{N}}_j,$$

where  $\tilde{\mathcal{N}}_j = \mathcal{N}_j \cup \{j\}$ .

Given  $\mathcal{G}$ , we also introduce the adjacency matrix  $\mathbf{A} \in \{0, 1\}^{n \times n}$  where  $a_{ij} = 1$  if  $(i, j) \in \mathcal{E}$  and  $a_{ij} = 0$  otherwise. We define  $\bar{\mathbf{A}}$  as the adjacency matrix of the same graph  $\mathcal{G}$  including self-loops, i.e.  $\bar{a}_{ii} = 1 \forall i \in \mathcal{V}$ . For undirected graphs, the graph Laplacian is defined as  $\mathbf{L} = \mathbf{D} - \mathbf{A}$  where  $\mathbf{D}$  is a diagonal matrix whose  $i$ -th diagonal element  $\mathbf{D}_{ii}$  is the sum of all the edges incident to node  $i$ .  $\mathbf{L}$  is diagonalizable by an orthogonal matrix

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H,$$

where  $\mathbf{U} \in \mathbb{C}^{n \times n}$  is the eigenvector matrix of  $\mathbf{L}$  that contains the eigenvectors as columns and  $\mathbf{\Lambda} \in \mathbb{R}^{n \times n}$  is the diagonal eigenvalue matrix where the eigenvalues are sorted in increasing order. A graph signal  $\mathbf{x} \in \mathbb{R}^n$  in the node domain is a real-valued function defined on the nodes of the graph  $\mathcal{G}$ , such that  $x_i$  with  $i = 1, \dots, n$  is the value of the signal at node  $i \in \mathcal{V}$  [1]. For an undirected graph, the eigenvectors of the Laplacian are used to define the graph Fourier transform (GFT) [1] of the signal  $\mathbf{x}$  as follows

$$\hat{\mathbf{x}} = \mathbf{U}^H \mathbf{x}.$$

If we consider a directed graph, we can define the GFT matrix using the eigenvector matrix of the adjacency matrix  $\mathbf{A}$ , in this case we have  $\mathbf{U} \in \mathbb{C}^{n \times n}$  [24].

### B. Compressed sensing

Compressed sensing (CS) [11], [22] is an established theory for signal sampling relying on linear sampling operators satisfying certain norm-preservation properties and employing nonlinear reconstruction techniques that leverage signal models such as sparsity. The typical setup involves a signal  $\mathbf{x} \in \mathbb{R}^n$ , having a sparse representation under an orthonormal basis  $\mathbf{U}$ ,  $\mathbf{x} = \mathbf{U}\hat{\mathbf{x}}$ , with  $\|\hat{\mathbf{x}}\|_0 \leq k$ , and a vector of measurements  $\mathbf{y} \in \mathbb{R}^m$ ,  $m \ll n$ , computed as

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \mathbf{U} \hat{\mathbf{x}} = \Psi \hat{\mathbf{x}}.$$

The most well-known reconstruction algorithm to recover  $\hat{\mathbf{x}}$  from  $\mathbf{y}$  is  $\ell_1$  norm minimization:

$$\hat{\mathbf{x}}^* = \arg \min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_1 \quad \text{s.t.} \quad \Psi \hat{\mathbf{x}} = \mathbf{y}, \quad (1)$$

with  $\Psi = \Phi \mathbf{U}$ .

It is known that a sufficient condition for perfect reconstruction and stability in presence of additive noise is that matrix

$\Psi$ , the product between the sampling matrix and the sparsity basis, provides a stable, distance-preserving embedding of the set of sparse signals  $\Sigma_k$ . This notion is captured by the Restricted Isometry Property (RIP) [25]–[27].

**Definition 3 (RIP).** A matrix  $\Psi$  satisfies the Restricted Isometry Property of order  $k$  if there exists a  $\delta_k \in (0, 1)$  such that

$$(1 - \delta_k) \|\hat{\mathbf{x}}\|_2^2 \leq \|\Psi \hat{\mathbf{x}}\|_2^2 \leq (1 + \delta_k) \|\hat{\mathbf{x}}\|_2^2$$

holds for all  $\hat{\mathbf{x}} \in \Sigma_k$

Several random constructions for the sampling matrix  $\Phi$  have been studied. Each specific construction interacts in a different way with the sparsity basis  $\mathbf{U}$ . More structured constructions typically need a larger number of measurements for  $\Psi$  to satisfy the RIP with constant probability. It is known that a sampling matrix  $\Phi$  made of independent and identically distributed (i.i.d.) sub-Gaussian random variables is universal, meaning that the RIP holds with  $m = O(k \log n)$  regardless of the basis  $\mathbf{U}$ . However, dense random matrices are not practical in many scenarios due to excessive memory requirements or architectural constraints in the sampling procedure (e.g., in the graph setting, they would violate locality). Toeplitz [28], circulant [29] and block-diagonal [30], [31] matrices are more structured random matrices suitable for common acquisition systems [32], [33] and they are also known to satisfy the RIP, albeit requiring more measurements and not being uniformly good for all bases.

### C. Related work

Several works have addressed the problem of sampling signals defined on graphs, under the assumption of sparsity in a transform domain, i.e., only up to  $k$  nonzero coefficients in the positions defined by the support. The most common model assumes a bandlimited signal, i.e., a signal where the sparsity basis is the GFT and where the support coincides with the first  $k$  frequency coefficients. The sampling problem in the graph signal setting is typically concerned with trading off locality and reconstruction accuracy. Locality implies that the signal is observed only at a small number of nodes or through short-range interactions of a node with its neighbors. We will use the general term *measurement* to refer to either a pointwise selection of a signal coefficient or the result of a more complex operation such as a linear combination. The result of the sampling operation is a vector of measurements such that the original signal can be reconstructed via linear or nonlinear interpolation methods.

Puy et al. [15] show that random selection of nodes can be an effective technique. Indeed, Chen et al. [16] proved that uniform random sampling is optimal for the Erdős-Rényi graph, thus requiring only  $k$  measurements. However, uniform random sampling is in general suboptimal because the required number of measurements is heavily affected by the structure of the graph. Puy et al. [15] improve over uniform sampling by proposing a weighted sampling scheme that uses information on the signal support and the transform basis to bias the sampling procedure towards nodes where the signal is

highly localized. Albeit improving over uniform sampling, this method requires knowledge of the support and computation of the sparsity basis before sampling. It is also suboptimal as it generally requires more than  $k$  measurements.

Tsivero et al. [14] study node-frequency uncertainty principles to derive necessary conditions for signal recovery after pointwise sampling. They also discuss greedy node selection strategies that approximate the optimal sampling objective, aimed at minimizing the MSE of reconstruction. Such greedy strategies require the knowledge of the signal support and are typically computationally expensive, e.g., requiring multiple singular value decompositions.

Sakiyama et al. [34], [35] study node selection methods in the context of sensor position selection. In [34] an assumption of a bandlimited signal is made and nodes are selected so that the cutoff frequency of the signal restricted to the sampling set is maximized. This is done using three possible techniques requiring access to the eigenvectors of the graph Laplacian. Thus this method only targets bandlimited signals (a special case of known support) and requires knowledge of the basis at sampling time. In [35] a different method is proposed where the localization operator is used to choose nodes so that the operators localized at the nodes of choice cover well the whole graph with minimal overlap. This is a rather efficient method since the localization operator can be implemented with polynomial filters, thus not requiring the full basis. While the method does not strictly need support information and can in principle be applied even if the signal support is unknown, there is an implicit assumption on the support in the choice of the kernel for the localization operator. In fact, the exponential kernel used by the authors implicitly assumes a bandlimited model, and it is not obvious how to choose a kernel for an arbitrary support.

Marques et al. [20], motivated by an analogy with classical sampling, propose a sampling strategy computing measurements as the result of iterated applications of the graph shift operator (e.g., the adjacency matrix) observed at one or more nodes. This strategy can guarantee perfect reconstruction with the optimal number of measurements and does not require knowledge of the support of the signal, so it can also be used when the support is unknown. The technique is equivalent to linear sampling where the sampling matrix has a Vandermonde structure. Unfortunately, Vandermonde matrices are known to be poorly conditioned [36] resulting in severe instability even for moderately-sized graphs. The analogy drawn by the authors with classical sampling does not consider that the Vandermonde structure arising in classical sampling is a unique special case that is well conditioned, as it results in the DFT [36]. Moreover, the method is highly dependent on the specific nodes chosen as observation points and it is not clear how to perform the selection.

In this paper, we present a sampling method based on random linear combinations of signal coefficients localized at a subset of nodes and their neighbors. The method is not computationally expensive (e.g., it does not require the sparsity basis to be known) and non-adaptive, i.e., it works without knowledge of the signal support. We provide a sufficient condition to select the sampling set of nodes, which must be

a superset of a dominating set of the graph. Thanks to such sampling nodes and the use of random coefficients in the linear combinations, we are able to prove the RIP for the resulting sampling matrix. The proposed method can achieve the optimal number of measurements when the support is known. When the support is unknown, the RIP guarantees stability provided a sufficient number of measurements is computed. Such number depends on the structure of the graph and the sparsity basis, as shown by the theoretical analysis in Section IV. Compared to the previously discussed approaches in the literature, the main advantage of the proposed method is its generality, i.e., the ability to work both when the signal support is known and when it is unknown, having at the same time theoretical guarantees for perfect and stable reconstruction for any kind of signal support. Such wide-reaching guarantees on performance are only possible since this is a non-adaptive aggregation sampling scheme. Schemes based on selection sampling of nodes may not require cooperation among nodes to exchange information but are adaptive, in the sense that they require more stringent conditions on the signals such as prior knowledge of the signal support or sparsity basis at sampling time or more specific signal models beyond sparsity, such as bandlimitedness. In a nutshell, techniques based on aggregation sampling, and especially the randomized scheme proposed in this paper, trade communications for generality.

### III. RANDOMIZED LOCAL AGGREGATIONS

This section describes the proposed technique to sample graph signals via random linear combinations of signal coefficients localized to neighborhoods of a subset of the nodes of the graph, which we refer to as random local aggregations. The main idea is to select a subset of nodes and at each of them aggregate the signal coefficients of the node and its neighbors with a random linear combination in order to obtain a measurement. The node selection procedure and the use of random linear combinations allows a sampling procedure that is simultaneously *localized*, i.e., it only requires interactions of the chosen node with its neighbors, and can guarantee perfect reconstruction and stability to noise. In particular, the proposed technique is equivalent to randomized linear sampling, similar to the random projections used in CS, with a sampling operator that satisfies the RIP, albeit not being dense, in general.

The proposed method does not rely on the knowledge of the sparsity support of the signal in a transformed domain (e.g., the knowledge of the active set of frequencies in the GFT) and can be used both when that is known or unknown. We want to compute a number of measurements  $m$  greater than or equal to the sparsity  $k$ . When the support is known, choosing  $m \geq k$  improves robustness to noise. Otherwise, this condition is needed to ensure correct recovery (the optimal  $\ell_0$  minimization requires at least  $2k$  measurements). The theoretical analysis in section IV shows that the proposed technique can obtain perfect reconstruction in absence of noise and with known support with the optimal number of measurements (i.e.,  $m = k$ ), for any graph signal with a dense sparsity basis, contrary to other techniques requiring a larger number of measurements unless the graph has a particular structure (e.g., Erdős-Rényi).

As a practical application example, let us think of the problem of deploying a sensor network having many low-power sensors that are equipped with short-range transmitters and a small number of (more expensive) sensors with long-range transmitters. The sensors acquire a signal defined over the nodes of a graph that may represent their communication links, and the goal is to compute a compact representation of the signal to be transmitted to a remote fusion center by the long range transmitters. The proposed algorithm allows that and requires a smaller number of long-range transmitters (or equivalently, a smaller data payload for the same number of transmitters). At the same time, no prior knowledge on the signal to be sampled is needed in the design phase, thus being suitable even for signals with non-stationary supports.

### A. Measurements

We consider a signal  $\mathbf{x} \in \mathbb{R}^n$ , defined over a directed graph  $\mathcal{G}(\mathcal{V}, \mathcal{E})$  with  $|\mathcal{V}| = n$  and  $\mathcal{N}_i$  the in-neighborhood of node  $i$  including a self-loop (from here onwards, the term neighborhood always refers to the in-neighborhood). The signal has a sparse representation under an orthonormal basis  $\mathbf{U}$ ,  $\mathbf{x} = \mathbf{U}\hat{\mathbf{x}}$  with  $\|\hat{\mathbf{x}}\|_0 \leq k$  and support  $\mathcal{S} = \{i : \hat{x}_i \neq 0\}$ . We highlight that the method is highly general and we can choose any orthonormal basis  $\mathbf{U}$ . One example is the GFT matrix defined in Section II-A. Let  $m$  be the desired number of measurements and  $\mathcal{R} \subseteq \mathcal{V}$  be the sampling set, i.e., the set of nodes at which measurements are computed. We also define a function  $\pi : \mathcal{R} \rightarrow [1, m]$  mapping a node  $i \in \mathcal{R}$  to its corresponding measurement index  $\pi(i)$ . Then, a measurement is computed for each  $i \in \mathcal{R}$  as:

$$y_{\pi(i)} = \sum_{j \in \mathcal{N}_i} \varphi_{\pi(i)j} x_j,$$

being each  $\varphi_{\pi(i)j}$  a zero-mean independent Gaussian random variable drawn as

$$\varphi_{\pi(i)j} \sim N\left(0, \frac{1}{g_j}\right)$$

with

$$g_j = \sum_{i \in \mathcal{R}} |\mathcal{N}_i \cap \{j\}| \quad (2)$$

being the node multiplicity, i.e., the number of times node  $j$  appears in the linear combinations. Notice that the node multiplicity only depends on the graph topology and the sampling set, which are known and precomputed before observing the signal.

In matrix form, the measurements are computed as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}$$

where the sampling matrix  $\mathbf{\Phi}$  is defined as follows

$$\varphi_{\pi(i)j} \sim \begin{cases} N\left(0, \frac{1}{g_j}\right), & \text{if } j \in \mathcal{N}_i \\ \delta_0, & \text{if } j \notin \mathcal{N}_i \end{cases}, \quad (3)$$

where  $\delta_0$  is a Dirac's delta distribution centered in 0. For an undirected graph  $\mathbf{\Phi}$  can be seen as the elementwise product of the adjacency matrix of the graph augmented to include self-loops and restricted to the rows of the sampling set  $\hat{\mathbf{A}}_{\mathcal{R}}$  and a random matrix  $\mathbf{\Xi}$  with independent entries,  $\mathbf{\Phi} = \hat{\mathbf{A}}_{\mathcal{R}} \odot \mathbf{\Xi}$ .

### B. Selection of the sampling set

The sampling set must be chosen to ensure that the norm of the signal is preserved in expectation, i.e.  $\mathbb{E}[\mathbf{\Phi}^H \mathbf{\Phi}] = \mathbf{I}$ . A sufficient condition to ensure this is that the sampling set  $\mathcal{R}$  is a superset of a dominating set  $\mathcal{D}$  of the graph, i.e.,  $\mathcal{R} \supseteq \mathcal{D}$ . Choosing the sampling set  $\mathcal{R}$  to be a dominating set ensures that the sampling matrix  $\mathbf{\Phi}$  has no empty columns, since all the nodes (including those in  $\mathcal{R}$ , thanks to self-loops) have at least a neighbor in  $\mathcal{R}$ . Moreover, all the rows have random coefficients from a continuous distribution to ensure that they are linearly independent with probability 1, as proved in Theorem 2.

Given a graph  $\mathcal{G}$ , finding a dominating set of minimal size is known to be NP-hard [37]. However, good approximations can be found efficiently with greedy algorithms. For instance, an approximation within  $O(\log \Delta)$  of the minimal size can be found with complexity  $O(\log |\mathcal{V}| \log \Delta)$  being  $\Delta$  the maximum vertex degree [38].

Depending on the desired number of measurements  $m$  and the size of the dominating set found by the particular algorithm in use, three situations may occur:

- 1)  $|\mathcal{D}| = m$ : in this case  $\mathcal{R} = \mathcal{D}$ ;
- 2)  $|\mathcal{D}| > m$ : we use  $p$ -hop random aggregations, i.e., define an equivalent graph having as adjacency matrix  $\mathbf{B}$  the union of all  $\mathbf{A}^l$  powers with  $l = 1, \dots, p$ , i.e.,  $b_{ij} = 1$  if and only if  $\exists l \in [1, p]$  s.t.  $a_{ij}^l = 1$ ;  $p$  is the minimum positive integer such that  $\mathcal{D}^p$ , the dominating set computed from  $\mathbf{B}$ , has cardinality  $|\mathcal{D}^p| \leq m$ .
- 3)  $|\mathcal{D}^p| < m$ : we add rows to  $\mathbf{\Phi}$  according to one of two strategies

- repetition of dominating nodes;
- insertion of new nodes.

The two strategies essentially differ by which nodes are considered eligible for insertion to increase the number of measurements. Let us call  $\mathcal{I}$  the set of nodes eligible for insertion.

In the former strategy, a dominating node is repeated<sup>1</sup>, i.e.,  $\mathcal{I} = \mathcal{D}^p$ . If node  $i \in \mathcal{D}^p$  is to be repeated, a new random combination of the signal coefficients in the neighborhood  $\mathcal{N}_i$  is computed. However, drawing new random coefficients does not guarantee that the new measurement is linearly independent as there may be a subgraph for which we are computing more measurements than signal coefficients. To avoid this, a solution is to have a greedy algorithm that before inserting the node in  $\mathcal{R}$  checks if the rank of the resulting sampling matrix is full or not. This strategy is advantageous in terms of localization since it does not require increasing the number of transmissions within the nodes of graph, nor does it require observing more than  $|\mathcal{D}^p|$  nodes. Notice, however, that checking if the new measurement is linearly independent can be expensive (up to  $O(n^3)$  with Gaussian elimination) and it is not practical for large graphs.

On the other hand, the latter strategy considers eligible for insertion only nodes that are not already in the sampling set, i.e.,  $\mathcal{I} = \mathcal{V} \setminus \mathcal{R}$ . Insertion of new nodes always guarantees that

<sup>1</sup>With slight abuse of notation we suppose that  $\mathcal{R}$  contains two or more copies of the same node.

the new measurement is linearly independent (the limit case being cliques forming square submatrices which however are still full rank with probability 1 due to the independent random coefficients). The complexity of this strategy is therefore limited to solving (4), which is  $O(n)$ . However, this strategy reduces locality, up to  $m = n$  when all nodes are in the sampling set.

The choice of the order to use for node insertion under both strategies is determined by the theoretical analysis in Section IV. In particular, we choose node  $i^*$  as the eligible node having most neighbors with lowest node multiplicity  $g$ :

$$i^* = \arg \max_{i \in \mathcal{I}} \left| \left\{ j : j \in \mathcal{N}_i, g_j = g_{\min}^{\mathcal{I}} \right\} \right|, \quad (4)$$

$$g_{\min}^{\mathcal{I}} = \min_{\ell \in \bigcup_{i \in \mathcal{I}} \mathcal{N}_i} g_\ell.$$

Informally, this is the node which has most neighbors that have been least seen by the various local aggregations, thus ensuring that all the signal coefficients have been observed enough times. This is particularly important when the signal is highly localized to a set of nodes that are not very well connected with the rest of the graph. Notice that the proposed criterion to select the node bears some similarity with the entropy-based criterion in [39]. However, it differs in the signal model adopted, being sparse in a transform basis rather than a Gaussian process.

Notice that we have proposed  $p$ -hop random aggregations when  $|\mathcal{D}| > m$ . They represent a tradeoff between number of measurements and locality. For instance, in a sensor network the sampling nodes might be the only ones equipped for long-range communication. Under this example, since the proposed method computes linear combinations of coefficients in some neighborhood, such coefficients must be transmitted from the nodes to the nodes in the sampling set. This is also common to other methods performing aggregations, such as [20], [21]. Defining one transmission as a real-valued scalar crossing an edge in the graph, the total number of transmissions for the proposed method depends on the strategy that is used to achieve  $|\mathcal{R}| = m$ :

- Repetition of dominating nodes: in this case, the number of transmissions is only determined by the size of the  $p$ -hop dominating set. Extra measurements will not cost extra transmissions because they are just different random linear combinations of the same signal coefficients already transmitted. In particular, the number of transmissions can be upper bounded<sup>2</sup> by

$$n_{\text{TX}} \leq \sum_{j \in \mathcal{D}^p} \sum_{i=1}^p i \cdot |\mathcal{N}_j^i|$$

being  $\mathcal{N}_j^i$  the set of nodes whose shortest path to node  $j$  is exactly  $i$  edges long.

- Insertion of new nodes: this strategy requires extra transmissions every time a node is inserted:

$$n_{\text{TX}} \leq \sum_{j \in \mathcal{R}} \sum_{i=1}^p i \cdot |\mathcal{N}_j^i|$$

<sup>2</sup>This is an upper bound because some paths may be counted multiple times.

## IV. THEORETICAL ANALYSIS

In this section, we prove the main theoretical results, showing that the proposed method guarantees reconstruction and stability to noise. Then, guided by the results of the theoretical analysis we discuss the effects of system characteristics and free parameters.

### A. Restricted Isometry Property

The proposed sampling method can be seen as a generalization of the method presented in [31], where the authors use block diagonal matrices as sampling matrices. In our case,  $\Phi$  is not a block-diagonal matrix, but we can expand the mathematical tools developed in [31] to our case. For this reason, the technique we use to prove that  $\Psi$  satisfies the RIP is similar to the one proposed in [31].

Before presenting the main results of the section, we first introduce some preliminary definitions, referring to the construction of the sampling matrix  $\Phi$  and sampling set  $\mathcal{R}$  presented in Sec. III.

**Definition 4.** Given a graph  $\mathcal{G}$ , the graph-basis coherence  $\mu(\mathbf{U})$  of an orthobasis  $\mathbf{U} \in \mathbb{C}^{n \times n}$  is defined as follows

$$\mu(\mathbf{U}) = \min \left( \sqrt{|\mathcal{N}^*|} \|\mathbf{U}\|_{\text{MAX}}, 1 \right), \quad (5)$$

where  $|\mathcal{N}^*| = \max_{i \in \mathcal{R}} |\mathcal{N}_i|$  and  $\|\mathbf{U}\|_{\text{MAX}} = \max_{i,j} |\mathbf{U}_{ij}|$ .

**Definition 5.** The minimum node multiplicity  $g_{\min}$  of  $\Phi$  is defined as follows

$$g_{\min} = \min_{k \in \mathcal{V}} g_k$$

with  $g_k$  defined as in (2).

We can now introduce our main theoretical result.

**Theorem 1.** Suppose  $\Phi \in \mathbb{R}^{m \times n}$  is defined as in (3) and suppose  $\mathbf{U} \in \mathbb{C}^{n \times n}$  is an orthobasis for  $\mathbb{C}^n$ . Set  $\Sigma_k = \{\hat{\mathbf{x}} \in \mathbb{C}^n \text{ s.t. } \|\hat{\mathbf{x}}\|_0 \leq k, \|\hat{\mathbf{x}}\|_2 = 1\}$ . If  $k \geq 1$  and

$$g_{\min} \geq c\delta^{-2}\mu^2k \log^2 k \log^2 n,$$

where  $c$  is a positive constant,  $\mu$  is the graph-basis coherence defined in (5) and  $0 < \delta < 1$ , then there exists  $\delta_k \leq \delta < 1$  such that, for all  $\mathbf{x} \in \mathbb{R}^n$  with  $\hat{\mathbf{x}} \in \Sigma_k$ , it holds that

$$(1 - \delta_k) \leq \|\Psi \hat{\mathbf{x}}\|_2^2 \leq (1 + \delta_k)$$

except with a probability of at most  $O(n^{-\log n \log^2 k})$ .

Before presenting the proof of Theorem 1, we first recall an important lemma that will be used during the proof.

**Lemma 1** (Theorem 3.1 in [40]). Let  $\mathcal{T} \subset \mathbb{C}^{m \times n}$  be a set of matrices, and let  $\varepsilon$  be a random vector whose entries are i.i.d., zero-mean, unit-variance random variables with sub-Gaussian norm  $\tau$ . Set

$$d_F(\mathcal{T}) = \sup_{\mathbf{T} \in \mathcal{T}} \|\mathbf{T}\|_F,$$

$$d_2(\mathcal{T}) = \sup_{\mathbf{T} \in \mathcal{T}} \|\mathbf{T}\|_2,$$

and

$$\begin{aligned} E_1 &= \gamma_2(\mathcal{T}, \|\cdot\|_2)(\gamma_2(\mathcal{T}, \|\cdot\|_2) + d_F(\mathcal{T})) + d_F(\mathcal{T})d_2(\mathcal{T}), \\ E_2 &= d_2(\mathcal{T})(\gamma_2(\mathcal{T}, \|\cdot\|_2) + d_F(\mathcal{T})), \\ E_3 &= d_2^2(\mathcal{T}) \end{aligned}$$

where the  $\gamma_2$ -functional of  $\mathcal{T}$  is a geometrical property of  $\mathcal{T}$ , as defined in [40]. Then, for  $t > 0$ ,

$$\log \mathbb{P} \left[ \sup_{\mathbf{T} \in \mathcal{T}} \left| \|\mathbf{T}\boldsymbol{\varepsilon}\|_2^2 - \mathbb{E}[\|\mathbf{T}\boldsymbol{\varepsilon}\|_2^2] \right| \geq E_1 + t \right] \leq -c(\tau) \min \left( \frac{t^2}{E_2^2}, \frac{t}{E_3} \right),$$

where  $c(\tau)$  is a positive constant that depends on  $\tau$ .

After having introduced this result, we can now present the proof of Theorem 1.

*Proof.* First of all, we note that

$$\begin{aligned} \mathbb{E}[\|\Phi\mathbf{x}\|_2^2] &= \mathbb{E}[\|\Psi\hat{\mathbf{x}}\|_2^2] = \hat{\mathbf{x}}^T \mathbb{E}[\mathbf{U}^H \Phi^H \Phi \mathbf{U}] \hat{\mathbf{x}} \\ &= \hat{\mathbf{x}}^T \mathbf{U}^H \mathbb{E}[\Phi^H \Phi] \mathbf{U} \hat{\mathbf{x}} = \hat{\mathbf{x}}^T \mathbf{U}^H \mathbf{U} \hat{\mathbf{x}} = \hat{\mathbf{x}}^T \hat{\mathbf{x}} = 1, \end{aligned}$$

where the fourth equality above follows because the sampling set  $\mathcal{R}$  is a superset of the dominating set  $\mathcal{D}^p$  and  $\varphi_{\pi(i)j} \sim N\left(0, \frac{1}{g_j}\right)$  if  $j \in \tilde{\mathcal{N}}_i$ . In this way, we can define the restricted isometry constant  $\delta_k$  as follows

$$\delta_k = \sup_{\hat{\mathbf{x}} \in \Sigma_k} \left| \|\Psi\hat{\mathbf{x}}\|_2^2 - 1 \right|.$$

Given  $\delta < 1$ , our objective is to show that, under the conditions described in the statement of the theorem,  $\delta_k \leq \delta$ .

We define the diagonal matrix  $\mathbf{G} \in \mathbb{R}^{m \times n}$  as

$$\mathbf{G} = \begin{bmatrix} \frac{1}{\sqrt{g_1}} & & & \\ & \frac{1}{\sqrt{g_2}} & & \\ & & \ddots & \\ & & & \frac{1}{\sqrt{g_n}} \end{bmatrix},$$

in addition, given  $\mathcal{R}$ , we define the vector  $\mathbf{x}_{r_i} \in \mathbb{R}^n$ , where  $r_i$  with  $1 \leq i \leq m$  is the  $i$ -th element of  $\mathcal{R}$ , as follows

$$x_{r_i, j} = \begin{cases} x_j & \text{if } j \in \tilde{\mathcal{N}}_{r_i} \\ 0 & \text{if } j \notin \tilde{\mathcal{N}}_{r_i} \end{cases}$$

and the block diagonal matrix  $\mathbf{T}_{\mathbf{x}} \in \mathbb{R}^{m \times nm}$  as

$$\mathbf{T}_{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_{r_1}^T \mathbf{G} & & & \\ & \mathbf{x}_{r_2}^T \mathbf{G} & & \\ & & \ddots & \\ & & & \mathbf{x}_{r_m}^T \mathbf{G} \end{bmatrix}.$$

Then, we can observe that

$$\begin{aligned} \|\Phi\mathbf{x}\| &= \sum_{i=1}^m |\varphi_i^T \mathbf{x}|^2 = \sum_{i=1}^m |\varphi_i^T \mathbf{x}_{r_i}|^2 = \sum_{i=1}^m |\mathbf{x}_{r_i}^T \varphi_i|^2 \\ &= \sum_{i=1}^m |\mathbf{x}_{r_i}^T \mathbf{G} \boldsymbol{\varepsilon}_i|^2 = \|\mathbf{T}_{\mathbf{x}} \boldsymbol{\varepsilon}\|_2^2, \end{aligned}$$

where  $\varphi_i \in \mathbb{R}^n$  is the  $i$ -th row of  $\Phi$ ,  $\boldsymbol{\varepsilon}_i \in \mathbb{R}^n$  is a rescaled version of  $\varphi_i$  having standard normal distribution

and  $\boldsymbol{\varepsilon} \in \mathbb{R}^{mn}$  is made by stacking all the vectors  $\boldsymbol{\varepsilon}_i$ . Thus, if we define the matrix set  $\mathcal{T}_{\mathbf{x}} = \{\mathbf{T}_{\mathbf{x}} : \hat{\mathbf{x}} \in \Sigma_k\}$ , we have completely expressed our problem in the setting of Lemma 1. To apply this lemma we have first to compute  $d_2(\mathcal{T}_{\mathbf{x}})$ ,  $d_F(\mathcal{T}_{\mathbf{x}})$  and  $\gamma_2(\mathcal{T}_{\mathbf{x}}, \|\cdot\|_2)$ . In order to compute these quantities, we first observe that

$$\begin{aligned} \|\mathbf{T}_{\mathbf{x}}\|_2 &= \|\mathbf{T}_{\mathbf{x}} \mathbf{T}_{\mathbf{x}}^T\|_2^{\frac{1}{2}} \leq \frac{1}{\sqrt{g_{\min}}} \max_{1 \leq i \leq m} \|\mathbf{x}_{r_i}\|_2 \\ &= \frac{1}{\sqrt{g_{\min}}} \max_{1 \leq i \leq m} \|\mathbf{U}_{\tilde{\mathcal{N}}_{r_i}} \hat{\mathbf{x}}\|_2 \\ &\leq \sqrt{\frac{|\tilde{\mathcal{N}}^*|}{g_{\min}}} \max_{1 \leq i \leq n} \langle \mathbf{u}_i, \hat{\mathbf{x}} \rangle \\ &\leq \sqrt{\frac{|\tilde{\mathcal{N}}^*|}{g_{\min}}} \max_{1 \leq i \leq n} \|\mathbf{u}_i\|_{\infty} \|\hat{\mathbf{x}}\|_1 = \sqrt{\frac{|\tilde{\mathcal{N}}^*|}{g_{\min}}} \|\mathbf{U}\|_{\text{MAX}} \|\hat{\mathbf{x}}\|_1, \end{aligned} \quad (6)$$

where  $\mathbf{U}_{\tilde{\mathcal{N}}_{r_i}} \in \mathbb{C}^{|\tilde{\mathcal{N}}_i| \times n}$  is the matrix  $\mathbf{U}$  restricted to the  $\tilde{\mathcal{N}}_{r_i}$  rows and  $\mathbf{u}_i \in \mathbb{C}^n$  is the  $i$ -th row of  $\mathbf{U}$ . On the other hand, we can also write that

$$\begin{aligned} \|\mathbf{T}_{\mathbf{x}}\|_2 &\leq \frac{1}{\sqrt{g_{\min}}} \max_{1 \leq i \leq m} \|\mathbf{x}_{r_i}\|_2 \leq \frac{1}{\sqrt{g_{\min}}} \|\mathbf{x}\|_2 \\ &= \frac{1}{\sqrt{g_{\min}}} \|\hat{\mathbf{x}}\|_2 \leq \frac{1}{\sqrt{g_{\min}}} \|\hat{\mathbf{x}}\|_1. \end{aligned} \quad (7)$$

From (6) and (7) we arrive at

$$\begin{aligned} \|\mathbf{T}_{\mathbf{x}}\|_2 &\leq \frac{1}{\sqrt{g_{\min}}} \min \left( \sqrt{|\tilde{\mathcal{N}}^*|} \|\mathbf{U}\|_{\text{MAX}}, 1 \right) \|\hat{\mathbf{x}}\|_1 \\ &= \frac{\mu}{\sqrt{g_{\min}}} \|\hat{\mathbf{x}}\|_1 \end{aligned}$$

Now, we can compute the quantities used in Lemma 1. First, we can compute  $d_2(\mathcal{T}_{\mathbf{x}})$  as follows

$$d_2(\mathcal{T}_{\mathbf{x}}) = \sup_{\mathbf{T}_{\mathbf{x}} \in \mathcal{T}_{\mathbf{x}}} \|\mathbf{T}_{\mathbf{x}}\|_2 \leq \frac{\mu}{\sqrt{g_{\min}}} \sup_{\hat{\mathbf{x}} \in \Sigma_k} \|\hat{\mathbf{x}}\|_1 \leq \mu \sqrt{\frac{k}{g_{\min}}},$$

Second, we have that

$$\begin{aligned} d_F(\mathcal{T}_{\mathbf{x}}) &= \sup_{\mathbf{T}_{\mathbf{x}} \in \mathcal{T}_{\mathbf{x}}} \sqrt{\text{tr}(\mathbf{T}_{\mathbf{x}} \mathbf{T}_{\mathbf{x}}^T)} = \sup_{\mathbf{T}_{\mathbf{x}} \in \mathcal{T}_{\mathbf{x}}} \sqrt{\sum_{i=1}^m \sum_{j=1}^n \frac{1}{g_j} x_{r_i, j}^2} \\ &= \sup_{\mathbf{T}_{\mathbf{x}} \in \mathcal{T}_{\mathbf{x}}} \|\mathbf{x}\|_2 = \sup_{\mathbf{T}_{\mathbf{x}} \in \mathcal{T}_{\mathbf{x}}} \|\hat{\mathbf{x}}\|_2 = 1. \end{aligned}$$

In order to compute  $\gamma_2(\mathcal{T}_{\mathbf{x}}, \|\cdot\|_2)$ , we apply the same procedure used in [31] and we obtain that

$$\gamma_2(\mathcal{T}_{\mathbf{x}}, \|\cdot\|_2) \leq c\mu \sqrt{\frac{k}{g_{\min}}} \log k \log n.$$

Finally, we can now apply Lemma 1. Given  $\delta$  such that  $0 < \delta < 1$  and assuming that  $g_{\min} \geq c\delta^{-2} \mu^2 k \log^2 k \log^2 n$ , we obtain that

$$\begin{aligned} E_1 &= \gamma_2(\mathcal{T}, \|\cdot\|_2)(\gamma_2(\mathcal{T}, \|\cdot\|_2) + d_F(\mathcal{T})) + d_F(\mathcal{T})d_2(\mathcal{T}) \\ &\leq c\mu \sqrt{\frac{k}{g_{\min}}} \log k \log n \left( c\mu \sqrt{\frac{k}{g_{\min}}} \log k \log n + 1 \right) \\ &\quad + \mu \sqrt{\frac{k}{g_{\min}}} \leq \delta(\delta + 1) + \frac{\delta}{c \log k \log n} \leq c_1 \delta, \end{aligned}$$

$$\begin{aligned}
E_2 &= d_2(\mathcal{T})(\gamma_2(\mathcal{T}, \|\cdot\|_2) + d_F(\mathcal{T})) \\
&\leq \mu \sqrt{\frac{k}{g_{\min}}} \left( c\mu \sqrt{\frac{k}{g_{\min}}} \log k \log n + 1 \right) \\
&\leq \frac{\delta}{c \log k \log n} (\delta + 1) \leq c_2 \frac{\delta}{\log k \log n}, \\
E_3 &= d_2^2(\mathcal{T}) \leq \frac{k\mu^2}{g_{\min}} \leq c_3 \frac{\delta^2}{\log^2 k \log^2 n}.
\end{aligned}$$

Then, the tail bound in Lemma 1 can be written as

$$\begin{aligned}
&\log \left( \mathbb{P} \left[ \sup_{\tilde{\mathbf{x}} \in \Sigma_k} \left| \|\Phi \tilde{\mathbf{x}}\|_2^2 - 1 \right| > c_1 \delta + t \right] \right) \\
&\leq -c(\tau) \min(t^2 \delta^{-2} \log^2 k \log^2 n, t \delta^{-2} \log^2 k \log^2 n),
\end{aligned}$$

If we set  $t = \delta$ , we obtain

$$\log \left( \mathbb{P} \left[ \sup_{\tilde{\mathbf{x}} \in \Sigma_k} \left| \|\Phi \tilde{\mathbf{x}}\|_2^2 - 1 \right| > \delta \right] \right) \leq -c \log^2 k \log^2 n,$$

which completes the proof.  $\square$

We point out that Theorem 1 does not require  $\Phi$  to be full row-rank. However, adding measurements that are linear combination of the previous ones is not efficient because these new measurements do not add new information. If we consider a sampling strategy without repeated nodes like the second one presented in Section III-B, we have the theoretical guarantee that  $\Phi$  is full row-rank, as shown in the following theorem.

**Theorem 2.** *Given a sampling set  $\mathcal{R}$ , where there are no repeated copies of the same node, set  $\Phi = \bar{\mathbf{A}}_{\mathcal{R}} \odot \Xi$ , where  $\Xi$  is a random matrix with independent entries drawn from a continuous distribution. Then,  $\Phi$  is a full row-rank matrix with probability 1.*

*Proof.* Let  $\varphi_{\pi(i)}$  be the row of  $\Phi$  corresponding to node  $i$ . Given a subset  $\mathcal{W}_i$  of the set  $\mathcal{R} - \{i\}$ , we define the union of the supports of  $\{\varphi_{\pi(j)}\}_{j \in \mathcal{W}_i}$  as the set  $\mathcal{C}_i = \bigcup_{j \in \mathcal{W}_i} \bar{\mathcal{N}}_j$ . Since the nonzero entries of  $\Phi$  are drawn independently from a continuous distribution, we have that  $\varphi_{\pi(i)}$  is linearly dependent on  $\{\varphi_{\pi(j)}\}_{j \in \mathcal{W}_i}$  with probability 1 if and only if there exists a subset  $\mathcal{W}_i \subseteq \mathcal{R} - \{i\}$  such that  $\bar{\mathcal{N}}_i \subseteq \mathcal{C}_i$  and  $|\mathcal{W}_i| \geq |\mathcal{C}_i|$ , i.e.,  $\text{span}(\{\varphi_{\pi(j)}\}_{j \in \mathcal{W}_i}) = \mathbb{R}^{|\mathcal{C}_i|}$ . Otherwise,  $\varphi_{\pi(i)}$  is linearly dependent on  $\{\varphi_{\pi(j)}\}_{j \in \mathcal{W}_i}$  with probability 0 if for any  $\mathcal{W}_i \subseteq \mathcal{R} - \{i\}$ ,  $\bar{\mathcal{N}}_i \not\subseteq \mathcal{C}_i$  or  $|\mathcal{W}_i| < |\mathcal{C}_i|$ . Then, every  $j \in \mathcal{W}_i$  also implies  $j \in \mathcal{C}_i$  because of self-loops and the following holds:  $|\mathcal{C}_i| \geq |\mathcal{W}_i|$ . When  $i \in \mathcal{C}_i$  because of edges from a node in  $\mathcal{W}_i$ , it makes  $|\mathcal{C}_i| \geq |\mathcal{W}_i| + 1$ , thus implying that  $\varphi_{\pi(i)}$  is linearly dependent on  $\{\varphi_{\pi(j)}\}_{j \in \mathcal{W}_i}$  with probability 0. Otherwise,  $i \notin \mathcal{C}_i$  implies  $\bar{\mathcal{N}}_i \not\subseteq \mathcal{C}_i$  thanks to the self-loop on  $i$ , thus also implying that  $\varphi_{\pi(i)}$  is linearly dependent on  $\{\varphi_{\pi(j)}\}_{j \in \mathcal{W}_i}$  with probability 0 and concluding the proof. As a remark, we notice that this proof would not be true with repeated nodes because the self-loop on  $i$  could already be included in the supports of  $\{\varphi_{\pi(j)}\}_{j \in \mathcal{W}_i}$  thus making the statement  $|\mathcal{C}_i| \geq |\mathcal{W}_i| + 1$  for the  $i \in \mathcal{C}_i$  case false.  $\square$

### B. Uniqueness of solution with known support

We now provide the conditions enabling the proposed method to achieve perfect reconstruction when the signal support is known a priori.

**Theorem 3.** *Given a graph  $\mathcal{G}$  and an orthobasis  $\mathbf{U} \in \mathbb{C}^{n \times n}$ , we consider a graph signal  $\mathbf{x} \in \mathbb{R}^n$  having a  $k$ -sparse representation under the basis  $\mathbf{U}$  with known support  $\mathcal{S}$ , i.e.,  $\mathbf{x} = \mathbf{U}_{|\mathcal{S}} \tilde{\mathbf{x}}_{\mathcal{S}}$  and measurements  $\mathbf{y} = \Phi \mathbf{x} = \Psi \tilde{\mathbf{x}} = \Psi_{|\mathcal{S}} \tilde{\mathbf{x}}_{\mathcal{S}}$ . Then, perfect reconstruction is achieved with probability 1 when  $g_{\min} \geq k$ .*

*Proof.* In order to show that we can achieve perfect reconstruction, we have to prove that, under the conditions described in the statement of the theorem, the matrix  $\Psi_{|\mathcal{S}}$  is full rank  $\forall \mathcal{S}$  such that  $|\mathcal{S}| = k$ . Therefore, the  $k$  columns of  $\Psi_{|\mathcal{S}}$  should be linearly independent. Let  $\psi_i$  be the  $i$ -th column of  $\Psi_{|\mathcal{S}}$ , given any  $\psi_j$  with  $j \neq i$ , we point out that

$$\mathbb{E}[\psi_j^T \psi_i] = \mathbb{E}[\mathbf{u}_j^T \Phi^T \Phi \mathbf{u}_i] = \mathbf{u}_j^T \mathbb{E}[\Phi^T \Phi] \mathbf{u}_i = \mathbf{u}_j^T \mathbf{u}_i = 0,$$

where  $\mathbf{u}_i$  is the  $i$ -th column of  $\mathbf{U}$ . Thus, we can say that  $\psi_j$  and  $\psi_i$ , with  $i \neq j$ , are uncorrelated. We also recall that the entries of  $\psi_i$ , with  $1 \leq i \leq n$ , are linear combinations of Gaussian random variables, then  $\psi_i$  is a random vector with Gaussian entries. Moreover, it is easy to see that any linear combinations of  $\psi_i$  and  $\psi_j$  is still a random vector with Gaussian entries. Therefore, we can say that the entries of  $\psi_i$  and  $\psi_j$  are jointly Gaussian distributed [41]. Then, since  $\psi_i$  and  $\psi_j$  are uncorrelated and their entries are jointly Gaussian distributed, from Theorem 4.5 in [42] we can say that the entries of  $\psi_i$  and  $\psi_j$  are independent.

Let  $\mathcal{S}_i$  be the support of  $\psi_i$ . Then, since we have shown that  $\forall j \in \mathcal{S}$  with  $j \neq i$   $\psi_i$  and  $\psi_j$  are pairwise independent random vectors whose entries are drawn from a continuous distribution,  $\psi_i$  is linearly dependent with probability 1 if and only if there exists a subset  $\mathcal{C}_i \subseteq \mathcal{S} - \{i\}$  with  $\mathcal{S}_{\mathcal{C}_i} = \bigcup_{j \in \mathcal{C}_i} \mathcal{S}_j$  such that  $\mathcal{S}_i \subseteq \mathcal{S}_{\mathcal{C}_i}$  and  $|\mathcal{C}_i| \geq |\mathcal{S}_{\mathcal{C}_i}|$ , i.e.,  $\text{span}(\{\psi_j\}_{j \in \mathcal{C}_i}) = \mathbb{R}^{|\mathcal{S}_{\mathcal{C}_i}|}$ . Otherwise,  $\psi_i$  is linearly dependent with probability 0 if for any subset  $\mathcal{C}_i \subseteq \mathcal{S} - \{i\}$  with  $\mathcal{S}_{\mathcal{C}_i} = \bigcup_{j \in \mathcal{C}_i} \mathcal{S}_j$ ,  $\mathcal{S}_i \not\subseteq \mathcal{S}_{\mathcal{C}_i}$  or  $|\mathcal{C}_i| < |\mathcal{S}_{\mathcal{C}_i}|$ . Then, we can observe that

$$k - 1 \geq |\mathcal{C}_i| \geq |\mathcal{S}_{\mathcal{C}_i}| \geq |\mathcal{S}_i|.$$

In order to compute  $|\mathcal{S}_i|$ , we recall that  $\psi_i = \Phi \mathbf{u}_i$ . Since  $\mathbf{u}_i$  is an orthonormal vector with respect to  $\mathbf{u}_j$ , we can say that there exists at least one nonzero entry  $l$  of the vector  $\mathbf{u}_i$  and the corresponding  $l$ -th column of  $\Phi$  has  $g_l$  nonzero entries by construction. Therefore, we obtain that  $|\mathcal{S}_i| \geq g_l \geq g_{\min}$ , which completes the proof.  $\square$

### C. Discussion

As we have shown in Section IV-A, if we consider the second sampling strategy presented in Section III-B, we have the theoretical guarantee that new measurements are linearly independent from the previous ones and thus always informative. Instead, the strategy presented in Section III-B that repeats dominating nodes can only have this guarantee if at every insertion the rank of the resulting matrix is checked



to avoid the presence of subgraphs with linearly dependent measurements.

It is also important to underline the fundamental role of the minimum node multiplicity  $g_{\min}$  in Theorem 1. As we can see from Definition 5, the parameter  $g_{\min}$  represents the minimum number of times a node of the graph appears in the measurements and it is subject to the lower bounds in Theorem 1 and 3 in order to satisfy the RIP and perfect reconstruction conditions. For these reasons, the proposed sampling strategies aim to find the sampling node that provides the fastest increase of  $g_{\min}$ , as shown in Eq. (4). We also point out that our method is not affected if the graph is disconnected. In fact, the sampling strategy based on  $g_{\min}$  guarantees that all the connected components are fairly sampled.

Furthermore, we point out that Theorem 3 provides a sufficient condition to guarantee perfect reconstruction. However, in some specific cases the bound might be not very tight and it could be impossible to satisfy the condition  $g_{\min} \geq k$ . In such cases, Theorem 3 becomes uninformative. For example, if we consider a disconnected graph, it may happen that  $g_{\min}$  is always lower than  $k$ , even if we increase the number of measurements. In this specific case, we can easily obtain a more informative lower bound by considering each connected component separately. Therefore, even if the signal is localized in just a few components, we can still reconstruct the signal given a sufficiently large number of measurements.

It is also interesting to highlight the connection between the theoretical results presented in this section and some well-known results of compressed sensing. In particular if we consider a fully connected graph, we have that  $\Phi$  is a dense matrix and  $g_{\min}$  is equal to  $m$ ; then the condition in Theorem 3 becomes  $m \geq k$ , which is a known result for compressed sensing. Instead, if we consider a disconnected graph where each component has the same number of nodes and is fully connected, it can be shown that  $\Phi$  is a random block-diagonal matrix and the RIP property presented in Theorem 1 becomes equal to the result presented in [31].

Moreover, it is worth noting that, if we consider a dense sparsity basis  $\mathbf{U}$ , the condition of Theorem 3 on  $g_{\min}$  can be replaced by requiring  $m \geq k$ , because in this case we have the guarantee that  $\Psi$  is dense.

Another important parameter that appears in Theorem 1 is the graph-basis coherence  $\mu(\mathbf{U})$ . It can be easily verified that

$$\sqrt{\frac{\mathcal{N}^*}{n}} \leq \mu(\mathbf{U}) \leq 1.$$

If we consider an undirected cycle graph and take as orthonormal basis  $\mathbf{U}$  its GFT basis (which in this case corresponds to the Fourier basis), it can be shown that  $\mu(\mathbf{U}) = \sqrt{\frac{2}{n}}$  [31].

## V. EXPERIMENTS

This section compares the performance of the proposed method with some other sampling methods for signals defined over graphs present in the literature.

### A. Experimental setup

We consider a signal that is  $k$ -sparse in the GFT domain, defined as the eigenvectors of the normalized graph Laplacian

Table I  
CONDITION NUMBER OF SAMPLING MATRIX  
ERDŐS-RÉNYI GRAPH,  $n = 100$ ,  $k = 10$

$p_e$	Method	Measurements					
		10	20	30	40	50	100
0.2	[20]	$10^8$	$10^{17}$	$10^{19}$	$10^{20}$	$10^{22}$	$10^{28}$
	Proposed	197.2	3.8	2.5	2.2	2.1	1.6
0.5	[20]	$10^{10}$	$10^{17}$	$10^{18}$	$10^{19}$	$10^{19}$	$10^{25}$
	Proposed	36.7	4.5	3.1	2.3	2.0	1.9
0.8	[20]	$10^{12}$	$10^{17}$	$10^{17}$	$10^{17}$	$10^{18}$	$10^{21}$
	Proposed	42.6	3.6	3.0	2.5	2.3	1.5

L. Depending on the particular experiment, the frequency support can either be the first  $k$  coefficients (bandlimited model) or  $k$  randomly chosen frequencies (random model). Tests are performed using the following graphs as implemented in the Matlab Graph Signal Processing Toolbox [43]: Minnesota, Sensor, Community. The dominating set is computed using a greedy algorithm that at every step adds the node with the largest degree without any neighbors already in the dominating set.

We perform sampling and reconstruction experiments with both known and unknown signal support. When the support is known, we compare the proposed method with uniform random sampling [15], weighted random sampling [15], the MinPinv [14] greedy approximation to the optimal sampling and reconstruction problem and the method based on the localization operator [35] (using a Chebyshev polynomial approximation of order 10 for the kernel). Notice that weighted random sampling and MinPinv require knowledge of the signal support, so we only test uniform random sampling and the localization operator when the support is unknown. Also notice that with respect to the other techniques, the proposed method with the repeated dominating nodes does not reduce locality when increasing the number of measurements: only  $|\mathcal{D}^p| \leq n$  nodes are aggregation points. Also, the hop parameter  $p$  decreases as  $m$  increases. All the results are averaged over 1000 runs. We also provide results about the running time of the sampling procedure (without signal reconstruction) for each method.

Finally, we do not show any reconstruction results for the successive local aggregations method of [20]. This method suffers from numerical instability even for moderately-sized graphs so that the reconstruction error quickly explodes for the sizes used in our experiments. Table I highlights this fact by reporting the high value of the condition number of the matrix  $\Psi$  obtained with such method, for the Erdős-Rényi graph with edge probability  $p_e$ . As mentioned in Section II-C this is due to the poor conditioning of the Vandermonde matrix constructed by the sampling procedure.

### B. System behavior

We show the size of the  $p$ -hop dominating set as returned by the greedy algorithm for the Minnesota and Sensor graphs. From Fig. 1 it can be noticed that the Minnesota graph has a particularly large dominating set due to the low degree of many nodes. However, the dominating set quickly shrinks once multiple hops are considered.

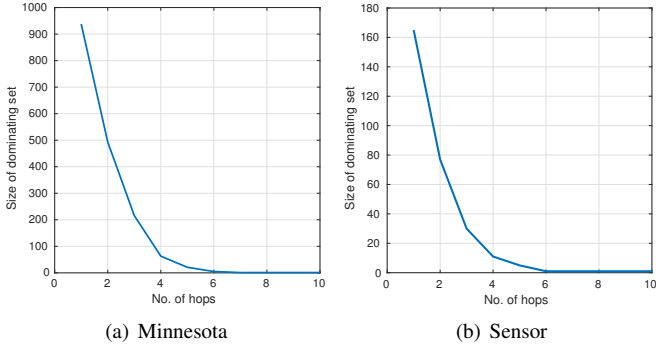


Figure 1. Size of dominating set as function of number of hops used to consider two nodes a neighbors ( $n = 2642$  for Minnesota,  $n = 1000$  for Sensor).

### C. Reconstruction: known support

In this experiment, the signal support in the frequency domain is known. However, the true sparse signal is corrupted by additive white Gaussian noise with standard deviation  $\sigma$ , i.e., measurements are computed as  $\mathbf{y} = \Phi(\mathbf{x} + \mathbf{n})$ . Least-squares reconstruction is performed by computing the pseudoinverse of the product between the sampling operator and the basis, restricted to the the columns in the support:

$$\begin{aligned}\hat{\mathbf{x}}_S^* &= (\Psi|_S)^+ \mathbf{y} \\ \mathbf{x}^* &= \mathbf{U}\hat{\mathbf{x}}^*\end{aligned}$$

Finally, we estimate the mean squared error (MSE) between the reconstructed signal  $\mathbf{x}^*$  and the original noiseless signal  $\mathbf{x}$  as:

$$\text{MSE} = 10 \log_{10} (n^{-1} \|\mathbf{x}^* - \mathbf{x}\|^2).$$

Fig. 2 shows the reconstruction MSE for the Sensor and Community graphs, as function of the number of measurements for a fixed noise standard deviation. We notice that the MinPinv algorithm [14] performs best, as expected since it approximates the optimal solution to the sampling problem which would minimize the reconstruction MSE. Its drawback is that it is computationally expensive and not very flexible, as it can only be applied if the support is known. The proposed method is typically close to MinPinv, especially for a lower number of measurements where it outperforms methods based on uniform or weighted random sampling and it is on par with or slightly better than the approach using the localization operator. We notice that the proposed method has a different error floor with respect to other methods for a large number of measurements due to the construction of the sampling matrix, having a condition number slightly larger than 1. However, it is not particularly interesting as this floor occurs for low values of MSE. We also notice that the proposed method works well even with the random model, when uniform random sampling and the localization operator exhibit poor performance due to their implicit assumptions on bandlimited signals.

Fig. 3 reports the results for the Minnesota graph. It is important to notice that the proposed method is stable in presence of noise, as confirmed by Fig. 3(b) where the slope matches the ones of the quasi-optimal MinPinv. We also observe that the two node selection strategies are fairly equivalent. Finally,

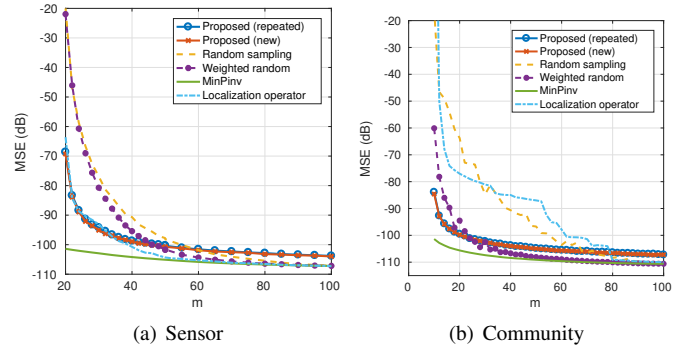


Figure 2. Reconstruction MSE as function of number of measurements. Sensor graph,  $n = 100$ ,  $k = 20$ , bandlimited model. Community graph,  $n = 100$ ,  $k = 10$ , random model. Noise standard deviation  $\sigma = 10^{-5}$ .

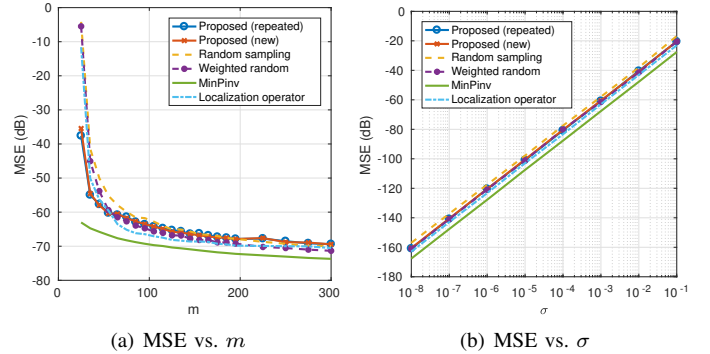


Figure 3. Reconstruction MSE as function of number of measurements  $m$  and noise standard deviation  $\sigma$ . Minnesota graph,  $n = 2642$ ,  $k = 25$ , bandlimited model. (a)  $\sigma = 10^{-3}$ , (b)  $m = 65$ .

we note that all the tests achieved perfect reconstruction with  $m = k$  when no noise was added.

### D. Reconstruction: unknown support

This experiment reconstructs signals with unknown frequency support by using  $\ell_1$  minimization, as defined in Eq. (1). No noise is added in this experiment and we measure the probability of perfect reconstruction, defined as a MSE lower than  $-40$  dB, as a function of the number of measurements.

Fig. 4 shows the probability of perfect reconstruction for the Minnesota and Sensor graphs. It can be noticed that the phase transition of the proposed method occurs at a lower number of measurements with respect to uniform random sampling. Fig. 5 reports the same experiment on the community graph. Notice how random sampling performs well for a bandlimited signal but degrades with respect to the proposed method when the support is chosen at random. The method based on the localization operator performs worse on this graph both under the bandlimited and random models.

Finally, we test the effect of the graph-basis coherence on the reconstruction performance. Theorem 1 predicts that a higher value of  $\mu$  requires a higher  $g_{\min}$  which means a higher number of measurements for a given graph. We constructed an experiment with two graphs with the same number of nodes ( $n = 100$ ) and same size of the dominating set ( $|\mathcal{D}| = 50$ ), but with drastically different values of the graph-basis coherence with the GFT basis. The first graph is a 2D grid, showing a low

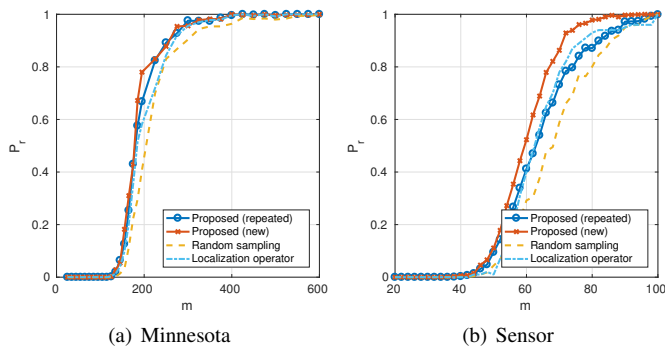


Figure 4. Support unknown: probability of perfect recovery as function of number of measurements  $m$ . Minnesota graph,  $n = 2642$ ,  $k = 25$ , bandlimited model. Sensor graph,  $n = 100$ ,  $k = 20$ , random model.

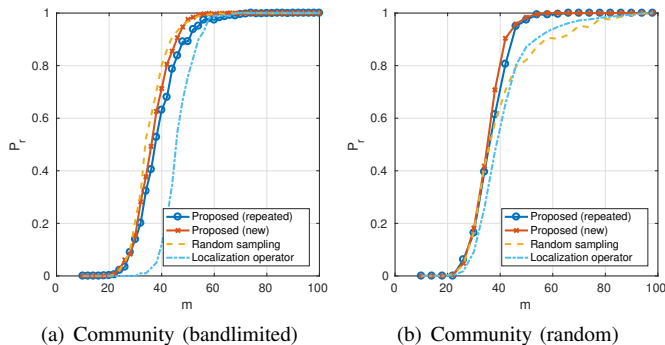


Figure 5. Support unknown: probability of perfect recovery as function of number of measurements  $m$ . Community graph,  $n = 100$ ,  $k = 10$ , (a) bandlimited model, (b) random model.

coherence, while the second is a small-world graph with high coherence. Fig. 6 shows that the phase transition occurs for a larger number of measurements when the coherence is high. We also test the impact of a weighted matrix used to compute the GFT basis. In particular, we draw a random geometric graph with a fixed number of nodes ( $n = 100$ ) on the unit 2D square and an edge is placed if the Euclidean distance between the nodes is below a certain threshold (equal to 0.2). We then repeat the same experiment as before of reconstruction with unknown support, comparing the probability of perfect reconstruction with the GFT as sparsity basis, computed from the normalized Laplacian of the binary adjacency matrix or from the weighted adjacency matrix where the weights of the nonzero entries are  $w_{ij} = e^{-d_E(i,j)}$ , being  $d_E$  the Euclidean distance between two nodes in the 2D space. Fig. 7 shows that there is no substantial difference in performance between the two bases.

### E. Computational complexity

Table II reports the running time for all the tested methods (except uniform random sampling, whose complexity is trivial) on the community graph with  $n = 2000$  nodes, a signal with sparsity  $k = 20$  on a random support and for different number of measurements. The times include all the operations needed to produce the measurements, thus including the computation of the eigenvectors of the Laplacian for the methods requiring them, or the dominating set for the proposed method. Notice

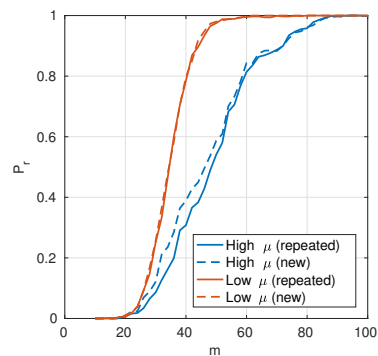


Figure 6. Support unknown: probability of perfect recovery as function of number of measurements  $m$ .  $n = 100$ ,  $k = 10$ , bandlimited model. Low  $\mu$ : 2D grid graph ( $\mu = 0.68$  at  $m = 60$ ), High  $\mu$ : small-world graph ( $\mu = 1$  at  $m = 60$ )

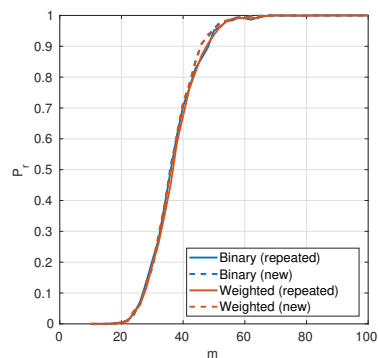


Figure 7. Support unknown: probability of perfect recovery as function of number of measurements  $m$ .  $n = 100$ ,  $k = 10$ , bandlimited model. Random geometric graph in the 2D square with Euclidean distance threshold 0.2. Basis is GFT from binary weights or exponential weights.

that for the proposed method, the 1-hop dominating set has size  $|\mathcal{D}| = 163$ , the 2-hop dominating set has size  $|\mathcal{D}^2| = 48$  and the 3-hop dominating set has size  $|\mathcal{D}^3| = 5$ . The methods were tested in MATLAB on a PC with an Intel Xeon E5620 CPU and 24GB of RAM. It can be noticed that the proposed method is faster, except when the number for measurements is very large and the repetition strategy is used. As discussed in Sec. III-B this is almost entirely due to the check on linear independence of the new measurement. On the other hand, the runtime of the weighted random method is dominated by the computation of the eigenvectors of the Laplacian, and in the case of MinPinv by the partial SVDs. The method based on the localization operator scales well thanks to the use of polynomial filters and is close to the proposed method.

Table II  
SAMPLING RUNTIME (SECONDS)

$m$	Weighted random	MinPinv	Localization operator	Proposed (repeated)	Proposed (new)
20	5.88	8.25	2.06	<b>1.29</b>	1.71
100	5.89	31.56	3.98	<b>1.25</b>	1.87
500	5.94	203.17	14.40	16.49	<b>4.99</b>

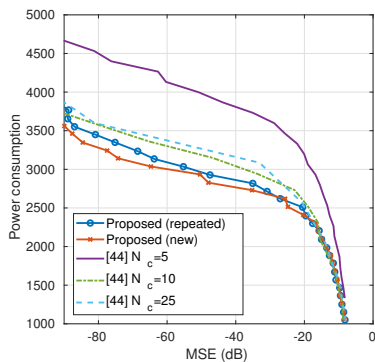


Figure 8. Tradeoff between power consumption and reconstruction MSE for data gathering in a wireless sensor network.  $N_c$  is the number of cluster heads in [44].  $n = 250$ ,  $k = 50$ .

### F. Application: data gathering in sensor networks

Wireless sensor networks are made of multiple, typically inexpensive sensors, tasked with acquiring a signal in a certain area. Such sensors usually have very limited computational capabilities and batteries, so power management is critical. We will consider the same scenario of data gathering presented in [44], in which a signal is sampled by means of a few measurements that are transmitted only by small portion of nodes to a remote base station. [44] also uses an approach based on CS but it is less flexible than the one presented in this paper as it results in a block-diagonal sensing matrix. The setup of the experiments considers  $n$  sensors randomly positioned inside the unit square and a remote base station at a distance equal to 5 times the side of the square. Sensors incur a transmission cost due to the consumed power being equal to the square of the transmission distance:  $P_{ij} = d_{ij}^2$ . The total power consumption is determined by the communications among the sensors and with the base station:  $P_{\text{tot}} = P_{\text{intra}} + P_{\text{BS}}$ . The method in [44] chooses a number of cluster heads, i.e., the sensors aggregating the measurements and transmitting them to the base station, and selects them at random among the sensors. Each sensor which is not a cluster head multiplies its signal coefficient by one or more random Gaussians and sends the value to the closest cluster head. The signal is supposed to be sparse in the DCT domain. For the proposed method, the graph is a nearest-neighbor graph with a radius equal to 0.2. In order to present a fair experiment, we compared the proposed method with [44] in terms of consumed power for the same reconstruction MSE or, equivalently, in terms of reconstruction MSE for a fixed power consumption. This can be observed in Figure 8 which shows that the proposed method achieves a better power-quality tradeoff, i.e., it is advantageous regardless of the number of clusters for the method in [44], as it allows to save power for the same quality of reconstruction or, equivalently, to achieve better quality for the same power.

## VI. CONCLUSIONS

We proposed a sampling algorithm for signals defined over graphs that draws concepts from the theories of compressed sensing and graph signal processing. It allows representing a graph signal by a small number of measurements obtained

through localized operations, i.e., only involving exchanges of information restricted to local neighborhoods. The method has theoretical guarantees in terms of reconstruction and stability to noise for any graph topology and any orthonormal basis where the signal has a sparse representation. The algorithm is agnostic to information on the signal support and therefore can be used when the signal support is both known or unknown.

## 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 Processing Magazine*, vol. 30, no. 3, pp. 83–98, May 2013.
- [2] I. Jablonski, "Graph signal processing in applications to sensor networks, smart grids and smart cities," *IEEE Sensors Journal*, vol. PP, no. 99, pp. 1–1, Dec. 2017.
- [3] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, "Computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, Feb. 2009.
- [4] E. H. Davidson, J. P. Rast, P. Oliveri, A. Ransick, C. Caletani, C.-H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C. T. Brown, C. B. Livi, P. Y. Lee, R. Revilla, A. G. Rust, Z. j. Pan, M. J. Schilstra, P. J. C. Clarke, M. I. Arnone, L. Rowen, R. A. Cameron, D. R. McClay, L. Hood, and H. Bolouri, "A genomic regulatory network for development," *Science*, vol. 295, no. 5560, pp. 1669–1678, Mar. 2002.
- [5] W. Huang, L. Goldsberry, N. F. Wymbs, S. T. Grafton, D. S. Bassett, and A. Ribeiro, "Graph frequency analysis of brain signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 7, pp. 1189–1203, Oct. 2016.
- [6] D. Thanou, P. A. Chou, and P. Frossard, "Graph-based compression of dynamic 3D point cloud sequences," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1765–1778, April 2016.
- [7] G. Fracastoro and E. Magli, "Predictive graph construction for image compression," in *2015 IEEE International Conference on Image Processing (ICIP)*, Sept. 2015, pp. 2204–2208.
- [8] G. Fracastoro, D. Thanou, and P. Frossard, "Graph transform learning for image compression," in *2016 Picture Coding Symposium (PCS)*, Dec. 2016, pp. 1–5.
- [9] W. Hu, G. Cheung, A. Ortega, and O. C. Au, "Multiresolution graph fourier transform for compression of piecewise smooth images," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 419–433, Jan. 2015.
- [10] 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 Processing Magazine*, vol. 31, no. 5, pp. 80–90, Sept. 2014.
- [11] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [12] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, Mar. 2011.
- [13] G. Leus, S. Segarra, A. Ribeiro, and A. G. Marques, "The dual graph shift operator: Identifying the support of the frequency domain," *arXiv preprint arXiv:1705.08987*, 2017.
- [14] M. Tsitsvero, S. Barbarossa, and P. D. Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, Sept. 2016.
- [15] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Applied and Computational Harmonic Analysis*, Mar. 2016.
- [16] S. Chen, R. Varma, A. Sandryhaila, and J. Kovacevic, "Discrete signal processing on graphs: Sampling theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, Dec 2015.
- [17] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, July 2016.
- [18] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 764–778, Feb. 2017.
- [19] L. F. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *arXiv preprint arXiv:1704.01223*, 2017.

- [20] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1832–1843, April 2016.
- [21] X. Wang, J. Chen, and Y. Gu, "Local measurement and reconstruction for noisy bandlimited graph signals," *Signal Processing*, vol. 129, pp. 119–129, Dec. 2016.
- [22] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [23] R. Varma, S. Chen, and J. Kovacevic, "Spectrum-blind signal recovery on graphs," in *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec. 2015, pp. 81–84.
- [24] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis." *IEEE Transaction on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, June 2014.
- [25] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [26] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, Dec. 2008.
- [27] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Uniform uncertainty principle for bernoulli and subgaussian ensembles," *Constructive Approximation*, vol. 28, no. 3, pp. 277–289, Dec. 2008.
- [28] J. Haupt, W. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5862–5875, Nov. 2010.
- [29] J. Vyříral, "A variant of the Johnson-Lindenstrauss lemma for circulant matrices," *Journal of Functional Analysis*, vol. 26, no. 4, pp. 1096–1105, Feb. 2011.
- [30] H. L. Yap, A. Eftekhari, M. B. Wakin, and C. J. Rozell, "The restricted isometry property for block diagonal matrices," in *2011 45th Annual Conference on Information Sciences and Systems*, March 2011, pp. 1–6.
- [31] A. Eftekhari, H. L. Yap, C. J. Rozell, and M. B. Wakin, "The restricted isometry property for random block diagonal matrices," *Applied and Computational Harmonic Analysis*, vol. 38, no. 1, pp. 1–31, Jan. 2015.
- [32] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, Sept. 2011.
- [33] D. Valsesia and E. Magli, "Compressive signal processing with circulant sensing matrices," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 1015–1019.
- [34] A. Sakiyama, Y. Tanaka, T. Tanaka, and A. Ortega, "Efficient sensor position selection using graph signal sampling theory," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 6225–6229.
- [35] —, "Accelerated sensor position selection using graph localization operator," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 5890–5894.
- [36] V. Y. Pan, "How bad are Vandermonde matrices?" *SIAM Journal on Matrix Analysis and Applications*, vol. 37, no. 2, pp. 676–694, 2016.
- [37] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co., 1990.
- [38] L. Jia, R. Rajaraman, and T. Suel, "An efficient distributed algorithm for constructing small dominating sets," *Distributed Computing*, vol. 15, no. 4, pp. 193–205, Dec. 2002.
- [39] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, Jun. 2008.
- [40] F. Kraemer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property," *Communications on Pure and Applied Mathematics*, vol. 67, no. 11, pp. 1877–1904, Nov. 2014.
- [41] A. Gut, *An Intermediate Course in Probability*. Springer, 2009.
- [42] R. V. Hogg, E. Tanis, and D. Zimmerman, *Probability and statistical inference*. Pearson Higher Ed, 2014.
- [43] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," *arXiv preprint arXiv:1408.5781*, 2014.
- [44] M. T. Nguyen and K. A. Teague, "Compressive sensing based data gathering in clustered wireless sensor networks," in *2014 IEEE International Conference on Distributed Computing in Sensor Systems*, May 2014, pp. 187–192.



sensing images, compressed sensing, and deep learning.

**Diego Valsesia** (S'13-M'17) received the Ph.D. degree in Electronic and Communication Engineering from the Politecnico di Torino, Turin, Italy, in 2016 and the M.Sc. degrees in Telecommunications Engineering from Politecnico di Torino and Electrical and Computer Engineering from the University of Illinois at Chicago, Chicago, IL, in 2012 and 2013 respectively. He is currently a Postdoctoral Associate at the Department of Electronics and Telecommunications (DET), Politecnico di Torino. His main research interests include compression of remote



processing, and deep learning.

**Giulia Fracastoro** (S'15-M'18) received the B.Sc. and M.Sc. degrees in Applied Mathematics from Politecnico di Torino, Italy, in 2011 and 2013, respectively, and the Ph.D. degree in Telecommunications Engineering from Politecnico di Torino in 2017. She is currently a Postdoctoral Associate at the Department of Electronics and Telecommunications (DET), Politecnico di Torino. During 2016, she was a visiting student at the Signal Processing Laboratory (LTS4), EPFL, Switzerland. Her research interests include graph signal processing, image processing, and deep learning.



**Enrico Magli** (S'97-M'01-SM'07-F'17) received the M.Sc. and Ph.D. degrees from the Politecnico di Torino, Torino, Italy, in 1997 and 2001, respectively. He is currently a Full Professor with Politecnico di Torino, Torino, Italy. His research interests include compressive sensing, image and video coding, and vision. He is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and the EURASIP Journal on Image and Video Processing, and a former Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA. He is a Fellow of the IEEE, and has been an IEEE Distinguished Lecturer from 2015 to 2016. He was the recipient of the IEEE Geoscience and Remote Sensing Society 2011 Transactions Prize Paper Award, the IEEE ICIP 2015 Best Student Paper Award (as senior author), and the 2010 and 2014 Best Associate Editor Award of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.