# Reconstruction of graph signals: percolation from a single seeding node

Santiago Segarra, Antonio G. Marques, Geert Leus, and Alejandro Ribeiro

***Abstract*—Schemes to reconstruct signals defined in the nodes of a graph are proposed. Our focus is on reconstructing bandlimited graph signals, which are signals that admit a sparse representation in a frequency domain related to the structure of the graph. The schemes, which are designed within the framework of linear shift-invariant graph filters, consider that the signal is injected at a single seeding node. After several sequential applications of the graph-shift operator – which computes linear combinations of the information available at neighboring nodes – the seeding signal percolates across the graph. We show that if the node is allowed to change the seeding signal with each application of the shift operator, the original bandlimited signal can be recovered. Conditions under which such a recovery is feasible are identified for two different reconstruction schemes. We illustrate both reconstruction schemes in synthetic graph signals and we analyze their performance in noisy real-world scenarios.**

***Index Terms*—Graph signal processing, Signal reconstruction, Interpolation, Percolation, Graph shift operator, Bandlimited graph signals**

## I. Introduction

Coping with the needs posed by fields such as network science and big data requires extending the results existing for classical time-varying signals to signals defined on graphs [1], [2]. 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. Two problems that have recently received substantial attention are sampling [3]–[6] and filtering [2] signals that are defined on the nodes of a graph.

This paper investigates the reconstruction of graph signals. In our approach, the reconstructed signal is obtained through percolation of a seeding signal, i.e., through the application of a graph filter to an initial signal. Graph filters are the generalization of the classical time-invariant systems when the signals are defined in a general graph structure. Seeding signals are graph signals defined on a subset of the nodes in the graph. In [7] it was shown that a graph signal of bandwidth $K$ can be reconstructed by using $K$ *seeding nodes* followed by the application of a low-pass graph filter. Different from both what occurs in the classical time domain and the usual approach followed for interpolating graph signals [8]–[11], if we want to reconstruct a signal using graph filters, the values of the seeding signals at the seeding nodes do not correspond with those of the signal to reconstruct [7].

The main contribution of this paper is the design of graph-signal reconstruction schemes via percolation *using only one seeding node*. Inspired by the local aggregation *sampling* procedure in [5], we propose two reconstruction – *interpolation* – schemes where the signal is injected at a single node and percolates throughout the graph via successive applications of the graph shift operator. In Sec. II we introduce the concepts needed for the definition of the reconstruction schemes. In Sec. III we present the two schemes prosed in this paper as well as their conditions for perfect reconstruction. The reconstruction scheme in Sec. III-A consists of successive injections of seeding values each of them followed by an application of the graph-shift operator. Differently, in the scheme considered in Sec. III-B, the seeding node only injects a number of seeding values equal to the bandwidth of the signal to reconstruct. The injection is then followed by a local implementation of a low-pass filter. In Sec. IV we first illustrate the perfect reconstruction of both schemes in synthetic noiseless signals. Then, we analyze the reconstruction performance when noisy signals in real-world networks are considered.[1]

Work in this paper is supported by USA NSF CCF-1217963 and Spanish MINECO TEC2013-41604-R. S. Segarra and A. Ribeiro are with the Dept. of Electrical and Systems Eng., Univ. of Pennsylvania. A. G. Marques is with the Dept. of Signal Theory and Comms., King Juan Carlos Univ. G. Leus is with the Dept. of Electrical Eng., Math. and Computer Sci., Delft Univ. of Tech. Emails: {ssegarra,aribeiro}@seas.upenn.edu, antonio.garcia.marques@urjc.es, g.j.t.leus@tudelft.nl.

## II. Bandlimited graph signals and graph filters

Sec. II-A describes the modeling considerations and presents the concept of the graph shift operator [2], [12]. Sec. II-B reviews the concept of a bandlimited graph signal and establishes some connections with the classical time domain. Sec. II-C introduces the concept of a graph filter and details the notion of a low-pass filter, which is central for the reconstruction scheme in Sec. III-B.

### A. General modeling considerations

Let $\mathcal{G}$ denote a directed graph with a set of nodes or vertices $\mathcal{N}$ (with cardinality $N$) and a set of links $\mathcal{E}$, such that if node $i$ is connected to $j$, then $(i,j) \in \mathcal{E}$. Since $\mathcal{G}$ is directed, the set $\mathcal{N}_i : \{j \ | (j,i) \in \mathcal{E}\}$ stands for 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 $A_{ji}$ if and only if $(i,j) \in \mathcal{E}$. The value of $A_{ji}$ captures the strength of the connection from $i$ to $j$. The focus of the paper is not on analyzing $\mathcal{G}$, but a graph signal defined on the set of nodes $\mathcal{N}$. Formally, such a signal can be represented as a vector $\mathbf{x} = [x_1, ..., x_N]^T \in \mathbb{R}^N$ where the $i$-th component represents the value of the signal at node $i$ or, alternatively, as a function $f : \mathcal{N} \to \mathbb{R}$, defined on the vertices of the graph.

The graph $\mathcal{G}$ can be endowed with the so-called *graph-shift operator* $\mathbf{S}$ [2], [12]. The shift $\mathbf{S}$ is a $N \times N$ matrix whose entry $S_{ji}$ can be non-zero only if $i = j$ or if $(i,j) \in \mathcal{E}$. 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 nonzero entries of $\mathbf{S}$. Widely-used choices for $\mathbf{S}$ are the adjacency matrix of the graph [2], [12] and the Laplacian [1]. The intuition behind $\mathbf{S}$ is to represent a linear transformation that can be computed locally at the nodes of the graph. More rigorously, if $\mathbf{y}$ is defined as $\mathbf{y} = \mathbf{S}\mathbf{x}$, then node $i$ can compute $y_i$ provided that it has access to the value of $x_j$ at $j \in \mathcal{N}_i$. 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 $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$.

[1] **Notation:** Generically, the entries of a matrix $\mathbf{X}$ and a (column) vector $\mathbf{x}$ will be 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 notation $^T$ and $^H$ stands for transpose and transpose conjugate, respectively; $\text{diag}(\mathbf{x})$ is a diagonal matrix satisfying $[\text{diag}(\mathbf{x})]_{ii} = [\mathbf{x}]_i$; $\mathbf{e}_i$ is the $i$-th $N \times 1$ canonical 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 vectors; and $\mathbf{0}$ is the all-zero vector (when not clear from the context, a subscript indicating the dimensions will be used). The modulus (remainder) obtained after dividing $x$ by $N$ will be denoted as $\text{mod}_N(x)$.

The next section presents the concept of a bandlimited graph signal. To establish connections with the notion of bandlimitedness for classical time-varying signals, we define the directed cycle graph $\mathcal{G}_{dc}$, with edge set $\mathcal{E}_{dc}=\{(i,\mathrm{mod}_N(i+1))\}_{i=1}^N$. Its adjacency and Laplacian matrices are denoted, respectively, as $\mathbf{A}_{dc}$ and $\mathbf{L}_{dc}:=\mathbf{I}-\mathbf{A}_{dc}$.

### B. Bandlimited graph signals

A common assumption when studying graph signals is that the graph-shift operator $\mathbf{S}$ 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 [3]. In this context, vectors $\mathbf{v}_i$ are interpreted as the graph frequency basis, $\widehat{x}_i$ as the corresponding signal frequency coefficients, and $\mathbf{x}$ as a $K$ bandlimited graph signal. The superscript $\widehat{\ }$ will be used to emphasize that the corresponding signal pertains to the frequency domain. We will assume that the set of active frequencies are known and, without loss of generality, that those are the first $K$ ones. Then, with $\widehat{\mathbf{x}}_K:=[\widehat{x}_1,\dots,\widehat{x}_K]^T$ being a $K\times 1$ vector collecting the coefficients associated with those frequencies, it holds that $\mathbf{x}$ is a $K$ bandlimited signal if

$$\widehat{\mathbf{x}}=[\widehat{\mathbf{x}}_K,0,\dots,0]^T,\qquad \mathbf{x}=\mathbf{V}\widehat{\mathbf{x}}=\mathbf{V}_K\widehat{\mathbf{x}}_K, \tag{1}$$

where $\mathbf{V}_K:=\mathbf{V}\mathbf{E}_K=[\mathbf{v}_1,...,\mathbf{v}_K]$.

**Remark 1** *Classical time-domain discrete signals can be thought as graph signals on the directed cycle $\mathcal{G}_{dc}$. Setting the shift operator either to $\mathbf{S}=\mathbf{A}_{dc}$ or $\mathbf{S}=\mathbf{L}_{dc}$ gives rise to the Fourier basis $\mathbf{F}$. More formally, the right eigenvectors of $\mathbf{S}$ satisfy $\mathbf{V}=\mathbf{F}$, with $F_{ij}:=\frac{1}{\sqrt{N}}e^{+\sqrt{-1}\frac{2\pi}{N}(i-1)(j-1)}$.*

### C. Graph filters

Let $\mathbf{H}:\ \mathbb{R}^N\to\mathbb{R}^N$ be a graph signal operator, i.e. a map between signals. We are interested in operators of the form $\mathbf{H}:=\sum_{l=0}^{L-1}h_l\mathbf{S}^l$, i.e., in linear transformations that can be expressed as a polynomial (of degree $L-1$) of the graph-shift operator. This type of transformations are called graph filters [2]. Graph filters are of particular interest because they can be implemented locally, e.g., with $L-1$ exchanges of information among neighbors. This is true since the graph-shift operator $\mathbf{S}$ can be computed locally (cf. Sec. II-A).

Note that the graph filter $\mathbf{H}$ can be written as $\mathbf{H}:=\mathbf{V}\big(\sum_{l=0}^{L-1}h_l\mathbf{\Lambda}^l\big)\mathbf{V}^{-1}$. The diagonal matrix $\widehat{\mathbf{H}}:=\sum_{l=0}^{L-1}h_l\mathbf{\Lambda}^l$ can be viewed as the frequency response of $\mathbf{H}$ and it can be alternatively written as $\widehat{\mathbf{H}}=\mathrm{diag}(\widehat{\mathbf{h}})$, where vector $\widehat{\mathbf{h}}$ is a vector that contains the $N$ frequency coefficients of the filter. By defining the $N\times L$ Vandermonde matrix

$$\mathbf{\Psi}:=\begin{pmatrix}1&\lambda_1&\dots&\lambda_1^{L-1}\\ \vdots&\vdots&&\vdots\\ 1&\lambda_N&\dots&\lambda_N^{L-1}\end{pmatrix}, \tag{2}$$

where $\lambda_i$ are the eigenvalues of $\mathbf{S}$, and the vector containing the coefficients of the filter as $\mathbf{h}:=[h_0,\dots,h_{L-1}]^T$, it holds that $\widehat{\mathbf{h}}=\mathbf{\Psi}\mathbf{h}$ and therefore

$$\mathbf{H}=\textstyle\sum_{l=0}^{L-1}h_l\mathbf{S}^l=\mathbf{V}\mathrm{diag}\big(\mathbf{\Psi}\mathbf{h}\big)\mathbf{V}^{-1}=\mathbf{V}\mathrm{diag}(\widehat{\mathbf{h}})\mathbf{V}^{-1}. \tag{3}$$

This implies that if $\mathbf{y}$ is defined as $\mathbf{y}=\mathbf{H}\mathbf{x}$, its frequency representation $\widehat{\mathbf{y}}$ satisfies

$$\widehat{\mathbf{y}}=\mathrm{diag}\big(\mathbf{\Psi}\mathbf{h}\big)\widehat{\mathbf{x}}. \tag{4}$$

Within this context, a *low-pass* graph filter of bandwidth $K$ is one where the frequency response $\widehat{\mathbf{h}}$ is given by

$$\widehat{\mathbf{h}}=[\widehat{\mathbf{h}}_K,0,\dots,0]^T, \tag{5}$$

where $\widehat{\mathbf{h}}_K$ contains the frequency response for the first $K$ frequencies. Notice that when the low-pass filter in (5) is applied to an arbitrary signal $\mathbf{x}$, the output signal is $K$ bandlimited as described in (1). The following proposition states the minimum degree needed for the design of a low-pass filter (cf. Proposition 1 in [7]).

**Proposition 1** Let $D$ denote the number of distinct eigenvalues of $\mathbf{S}$ in $\{\lambda_k\}_{k=K+1}^N$. Then, it holds that if $L>D$, there exist infinitely many nonzero $L\times 1$ vectors of low-pass filter coefficients $\mathbf{h}^*$, i.e. that $\mathbf{\Psi}\mathbf{h}^*$ is of the form (5).

The idea behind the proof is to show that if $L>D$, the nullspace of the $(N-K)\times L$ matrix $\mathbf{\Psi}_{\underline{K},L}:=[\mathbf{e}_{K+1},\dots,\mathbf{e}_N]^T\mathbf{\Psi}$ is non-empty (so that $\mathbf{h}^*$ can be selected from that space). This requires setting $L$ larger than the rank of $\mathbf{\Psi}_{\underline{K},L}$. Since $\mathbf{\Psi}_{\underline{K},L}$ has a Vandermonde structure, its rank is $D$ (the number of distinct $\lambda_i$) and, therefore, the proposition follows. Low-pass filters will be instrumental in the design of the reconstruction scheme in Sec. III-B.

## III. Reconstruction using a single seeding node

We present reconstruction – interpolation – schemes for graph signals where the signal is injected at a single node and percolates through the graph via successive applications of the graph shift operator $\mathbf{S}$. Assume without loss of generality that the first node is the one injecting the signal, and let the $N\times 1$ vector $\bar{\mathbf{x}}^{(t)}=[\bar{x}^{(t)},0,\dots,0]^T$ denote the signal injected at time $t$. The signals are propagated across nodes through successive applications of the graph shift $\mathbf{S}$, so that the obtained signal $\mathbf{y}^{(t)}$ after $t$ shifts is given by

$$\mathbf{y}^{(t)}=\mathbf{S}\mathbf{y}^{(t-1)}+\bar{\mathbf{x}}^{(t)},\qquad \mathbf{y}^{(-1)}=\mathbf{0}. \tag{6}$$

With $\mathbf{y}:=\mathbf{y}^{(\tau-1)}$ denoting the reconstructed signal after $\tau-1$ applications of the shift for some integer $\tau$, we rewrite (6) to obtain

$$\mathbf{y}=\textstyle\sum_{l=0}^{\tau-1}\mathbf{S}^l\bar{\mathbf{x}}^{(\tau-1-l)}=\sum_{l=0}^{\tau-1}\mathbf{S}^l\bar{x}^{(\tau-1-l)}\mathbf{e}_1, \tag{7}$$

which relates the reconstructed signal $\mathbf{y}$ with the successive inputs $\bar{x}^{(\cdot)}$ of the seeding node. Notice that (7) can be interpreted as the application of the graph filter

$$\mathbf{H}=\textstyle\sum_{l=0}^{\tau-1}\bar{x}^{(\tau-1-l)}\mathbf{S}^l \tag{8}$$

of degree $\tau-1$ to the canonical vector $\mathbf{e}_1$. The coefficients of the filter are given by the seeding values $\{\bar{x}^{(t)}\}_{t=0}^{\tau-1}$, which can be alternatively written as $\bar{\mathbf{x}}:=[\bar{x}^{(\tau-1)},\dots,\bar{x}^{(0)}]^T$. Our objective is then to design the amount of seeding signals $\tau$ and each of the seeding values in $\bar{\mathbf{x}}$ so that the reconstructed graph signal $\mathbf{y}$ equals a desired signal $\mathbf{x}$. In Sec. III-A we solve this design problem and state the conditions for guaranteed perfect reconstruction. In Sec. III-B we exploit the fact of $\mathbf{x}$ being bandlimited to propose an alternative reconstruction scheme which first injects a (shorter) seeding signal and then processes it using a low-pass graph filter.

### A. Reconstruction via successive shift applications

Since the reconstructed signal can be understood as the output of a graph filter [cf. (8)], we will address the reconstruction problem in the frequency domain. We begin by defining the frequency representations of the input and the output of the filter as $\widehat{\mathbf{e}}_1:=\mathbf{V}^{-1}\mathbf{e}_1$ and $\widehat{\mathbf{y}}:=\mathbf{V}^{-1}\mathbf{y}$. The first collects the frequency coefficients associated with the first canonical vector (input of the filter) and the second one those of the reconstructed signal (output of the filter). Note that the elements in $\widehat{\mathbf{e}}_1$ represent how strongly the seeding node expresses each of the graph frequencies.

The output-input relationship of filter (8) in the frequency domain can be written as [cf. (4)]

$$\widehat{\mathbf{y}}=\mathrm{diag}(\mathbf{\Psi}\bar{\mathbf{x}})\widehat{\mathbf{e}}_1=\mathrm{diag}(\widehat{\mathbf{e}}_1)\mathbf{\Psi}\bar{\mathbf{x}}, \tag{9}$$

where we have set $L = \tau$ in the definition of $\mathbf{\Psi}$ [cf. (2)] and, for the second equality, we have used the fact that for generic vectors $\mathbf{a}$ and $\mathbf{b}$ it holds that $\text{diag}(\mathbf{a})\mathbf{b} = \text{diag}(\mathbf{b})\mathbf{a}$.

Since we assume $\mathbf{x}$ to be $K$-bandlimited with active frequencies $\widehat{\mathbf{x}}_K$ and our design goal is to have $\widehat{\mathbf{x}} = \widehat{\mathbf{y}}$, (9) implies that the filter output must satisfy

$$[\widehat{\mathbf{x}}_K,\ \mathbf{0}]^T = \text{diag}(\widehat{\mathbf{e}}_1)\mathbf{\Psi}\bar{\mathbf{x}}. \tag{10}$$

The following proposition states the conditions under which (10) can be solved with respect to $\bar{\mathbf{x}}$. Its proof can be found in Appendix A.

**Proposition 2** Let $U_1$ and $U_2$ be, respectively, the number of values in $\{[\widehat{\mathbf{e}}_1]_k\}_{k=1}^K$ and $\{[\widehat{\mathbf{e}}_1]_k\}_{k=K+1}^N$ that are zero; let $D_1$ be the number of repeated values in $\{\lambda_k\}_{k=1}^K$; and let $D_2$ be the number of repeated values in $\{\lambda_k\}_{k\in\mathcal{K}_U}$, where $\mathcal{K}_U := \{k \mid K < k \leq N \text{ and } [\widehat{\mathbf{e}}_1]_k \neq 0\}$. Then, the system of $N$ equations in (10) is guaranteed to have a solution w.r.t. $\bar{\mathbf{x}}$ if the two following conditions hold:
i) $U_1 = 0$ and $D_1 = 0$; and
ii) $\tau \geq N - U_2 - D_2$.

Condition *i)* basically states that the seeding node needs to be able to act on every active frequency ($U_1 = 0$) and this action must be different for each of these frequencies ($D_1 = 0$). Condition *ii)* states the minimum number of seeding values needed for perfect reconstruction. Notice that, if condition *i)* holds, Proposition 2 allows us to set $\tau = N - U_2 - D_2$ and solve (10) to find the values $\{\bar{x}^{(t)}\}_{t=0}^{\tau-1}$ that the seeding node must inject into the network.

As a particular case, consider the connection with the classical time domain where the shift operator is given by the adjacency matrix of the directed cycle, i.e. $\mathbf{S} = \mathbf{A}_{dc}$. In this case, applying the $l$th power of shift-operator $\mathbf{A}_{dc}^l$ to a signal amounts to shifting the signal $l$ time instants. Particularizing (7) for the problem at hand yields that the reconstructed signal $\mathbf{y}$ is $\mathbf{y} = \sum_{l=0}^{\tau-1} \mathbf{A}_{dc}^l \bar{\mathbf{x}}^{(\tau-1-l)}$. Since in this case the seeding values $\bar{\mathbf{x}}^{(t)} = [\bar{x}^{(t)}, 0, \ldots, 0]^T$ correspond to Dirac deltas, $\mathbf{A}_{dc}^l \bar{\mathbf{x}}^{(t)}$ is a shifted Dirac with the non-zero entry at the $(l+1)$th position. Therefore, the goal of designing $\{\bar{x}^{(t)}\}_{t=0}^{\tau-1}$ such that $\mathbf{y} = \mathbf{x}$ can be trivially guaranteed by setting $\tau = N$ and $\bar{x}^{(t)} = x_{N-t}$, with $t = 0, \ldots, N-1$. Clearly, for the entire signal to be propagated across the graph (circular time grid), the percolation procedure needs $N-1$ shifts. This is consistent with Prop. 2 since, for this particular case, $U_2 = D_2 = 0$ [cf. condition *ii)*].

The reconstruction scheme presented can be used to recover graph signals that are not bandlimited. This is not surprising since the way in which the signal percolates through the network – via successive applications of $\mathbf{S}$ – does not take into account the bandwidth of $\mathbf{x}$. A different reconstruction scheme based on local interactions which further exploits the bandlimitedness of $\mathbf{x}$ is presented next.

### *B. Reconstruction via successive shifts and low-pass filtering*

Given a $K$ bandlimited signal $\mathbf{x}$, the reconstruction scheme considered here can be divided into two phases:

- *Seeding phase*. The seeding node injects $K$ signals, corresponding to the percolation dynamics (6) for times $t = 0, \ldots, K-1$.
- *Filtering phase*. The seeding signal is processed using a low-pass graph filter, which can be implemented locally by exchanging information only among neighboring nodes.

The goal of the seeding phase is to inject into the graph the information needed to recover $\mathbf{x}$. Since $\mathbf{x}$ has bandwidth $K$, we need at least $K$ seeding values for perfect reconstruction. Thus, under this scheme, the $K$ seeding values are grouped into vector $\bar{\mathbf{x}}_K := [\bar{x}^{(K-1)}, \ldots, \bar{x}^{(0)}]^T$. In the filtering phase, we further propagate the information available from the seeding phase while annihilating the components present in the frequencies $k > K$ not present in $\mathbf{x}$.

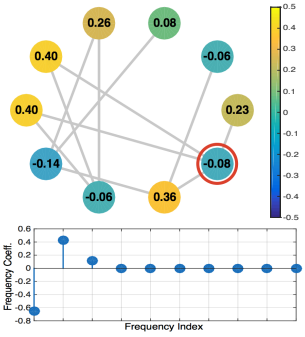



Fig. 1: (top) Graph signal $\mathbf{x}$ to reconstruct. The seeding node is circled in red. (bottom) Frequency components $\widehat{\mathbf{x}}$ of graph signal $\mathbf{x}$.

| t | seed 1 | seed 2 |
|---|---|---|
| 0 | -0.017 | 0.059 |
| 1 | -0.036 | -0.019 |
| 2 | 0.161 | -0.341 |
| 3 | 0.296 | 0 |
| 4 | -0.262 | 0 |
| 5 | -0.465 | 0 |
| 6 | 0.054 | 0 |
| 7 | 0.095 | 0 |

TABLE I: Seeding values injected by the seeding node at each time $t$ for both reconstruction schemes.

Denote by $\mathbf{y}$ the signal obtained from the seeding phase, $\mathbf{z} := \mathbf{H}\mathbf{y}$ the signal obtained from the filtering phase, and $\widehat{\mathbf{y}}$ and $\widehat{\mathbf{z}}$ their corresponding frequency representations. Let $\widehat{\mathbf{z}}_K := \mathbf{E}_K^T\widehat{\mathbf{z}}$ be a vector containing the first $K$ entries of $\widehat{\mathbf{z}}$. Since $\widehat{\mathbf{z}}$ is the output of the reconstruction, we want $\widehat{\mathbf{z}}_K = \widehat{\mathbf{x}}_K$, while the entries of $\widehat{\mathbf{z}}$ for $k > K$ must be zero.

Let $D$ denote the number of distinct eigenvalues in $\{\lambda_k\}_{k=K+1}^N$ and denote by $\mathbf{h}^*$ the coefficients of a low-pass filter (cf. Sec. II-C) of degree $D$. Defining $\widehat{\mathbf{h}}^* := \mathbf{\Psi}\mathbf{h}^*$ we may analyze the application of the low-pass filter in the frequency domain as

$$\widehat{\mathbf{z}}_K = \mathbf{E}_K^T\text{diag}(\widehat{\mathbf{h}}^*)\widehat{\mathbf{y}} = \mathbf{E}_K^T\text{diag}(\widehat{\mathbf{h}}^*)\mathbf{E}_K\mathbf{E}_K^T\widehat{\mathbf{y}}. \tag{11}$$

Further defining $\widehat{\mathbf{h}}_K^* := \mathbf{E}_K^T\mathbf{\Psi}\mathbf{h}^*$ and $\widehat{\mathbf{y}}_K := \mathbf{E}_K^T\widehat{\mathbf{y}}$, (11) can be rewritten as

$$\widehat{\mathbf{z}}_K = \text{diag}(\widehat{\mathbf{h}}_K^*)\widehat{\mathbf{y}}_K. \tag{12}$$

Using (9) to obtain the output of the seeding phase and recalling that we want $\widehat{\mathbf{z}}_K = \widehat{\mathbf{x}}_K$, we rewrite (12) to obtain

$$\widehat{\mathbf{x}}_K = \text{diag}(\widehat{\mathbf{h}}_K^*)\mathbf{E}_K^T\text{diag}(\widehat{\mathbf{e}}_1)\mathbf{\Psi}\bar{\mathbf{x}}_K, \tag{13}$$

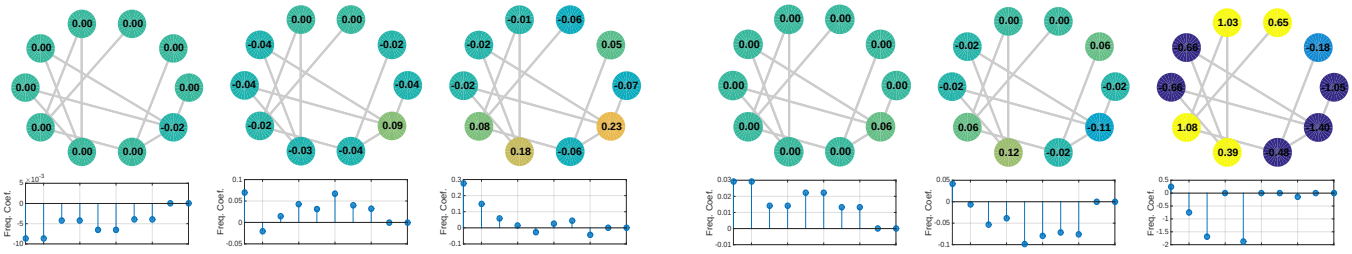
where $\mathbf{\Psi}$ has $K$ columns, i.e. $K = L$ in (2). Expression (13) relates the active frequencies $\widehat{\mathbf{x}}_K$ of the signal to reconstruct and the seeding values $\bar{\mathbf{x}}_K$. The following proposition states the conditions under which (13) can be solved with respect to $\bar{\mathbf{x}}_K$. Its proof can be found in Appendix B.

**Proposition 3** Let $U_1$ and $D_1$ be defined as in Proposition 2. Then, the system of $K$ equations in (13) is guaranteed to have a solution w.r.t. $\bar{\mathbf{x}}_K$ if the following conditions hold:
i) $U_1 = 0$ and $D_1 = 0$; and
ii) $\lambda_{k_1} \neq \lambda_{k_2}$ for all $(\lambda_{k_1}, \lambda_{k_2})$ such that $k_1 \leq K$ and $k_2 > K$.

Condition *i)* is equivalent to that in Prop. 2 whereas condition *ii)* ensures that the low-pass filter with coefficients $\mathbf{h}^*$ does not cancel any of the present frequencies in the signal to reconstruct $\mathbf{x}$.

Note that while in Sec. III-A the seeding node had to inject $\tau = N - U_2 - D_2$ seeding signals, now that number has been reduced to $K$. However, the number of times that the graph shift operator has to be applied is basically the same for both reconstruction schemes. For the scheme in Sec. III-A, we need to apply the shift operator $N - U_2 - D_2 - 1$ times. For the scheme in this section, we need to apply the shift operator $K - 1$ times for the seeding phase and $D$ times for the filtering phase. Whenever no eigenvalues are repeated and $U_2 = 0$, both reconstruction schemes require exactly $N - 1$ graph-shift applications.

In the following section we illustrate the application of both reconstruction schemes in synthetic and real-world graph signals.




(a) (b)

Fig. 2: Intermediate signals obtained when reconstructing the graph signal in Fig. 1 via percolation from a single seeding node, and their frequency representation. (a) Reconstruction via successive shift applications (cf. Sec. III-A) for times $t = 0$ (left), $t = 2$ (mid), and $t = 6$ (right). (b) Reconstruction via successive shifts and low-pass filtering (cf. Sec. III-B) for times $t = 0$ (left), $t = 2$ (mid), and $t = 6$ (right).

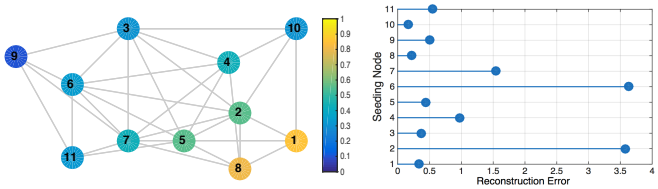
## IV. Numerical experiments

**Test Case 1.- Perfect reconstruction of noiseless graph signals.** Consider the 10-node graph in Fig. 1 generated as an Erdõs-Rényi graph with edge probability 0.3, and a signal defined on the graph taking the values stated at each node and encoded by the nodes' colors. Considering the adjacency matrix of the graph as its shift operator, i.e. $\mathbf{S} = \mathbf{A}$, the graph signal has bandwidth 3 since only three frequency coefficients are different from zero; see Fig. 1 (bottom). Our objective is to reconstruct this signal by injecting seeding values at the seeding node, circled in red in Fig. 1.

We first illustrate the reconstruction scheme presented in Sec. III-A. After checking that condition *i)* in Prop. 2 is satisfied, we fix $\tau = 8$ to satisfy condition *ii)* since $N = 10$, $U_2 = 2$, and $D_2 = 0$. This guarantees that we can perfectly reconstruct the desired signal with $\tau = 8$ seeding values $\bar{\mathbf{x}}$, which can be computed using (10). The column 'seed 1' in Table I presents these seeding values, i.e., the seeding node must inject the value $-0.017$ at time $t = 0$, $-0.036$ at $t = 1$, and so on. In Fig. 2a, we present the intermediate graph signals obtained at times $t = 0$, $t = 2$ and $t = 6$. For $t = 7$, we recover exactly the signal in Fig. 1. Naturally, the signal at $t = 0$ attains the value zero for all nodes except for the seeding node where the injected value is attained. The frequency representation of this signal is simply $-0.017\,\widehat{\mathbf{e}}_1$ and, given that the last two frequency components are exactly zero, it becomes clear that $U_2 = 2$ as stated previously (cf. Prop. 2). Notice that in the signal at time $t = 6$, which corresponds to the intermediate signal before perfect reconstruction, every frequency component is present – except for the last two, which are always zero. This illustrates the fact that intermediate signals are not bandlimited when following this reconstruction scheme.

For the second reconstruction scheme (cf. Sec. III-B), only three seeding values are needed – since the signal to recover has bandwidth three – injected at times $t = 0$ to $t = 2$; see Table I. These values are obtained from (13) and, after checking fulfillment of conditions *i)* and *ii)* in Prop. 3, we are guaranteed perfect reconstruction. From time $t = 3$ onwards, the filtering phase starts and the seeding node does not inject any further values. However, a low-pass filter is implemented through local interactions and, at each time step, one frequency is annihilated. Consequently, at time $t = 6$ (cf. Fig. 2b) we see that some of the frequencies which are not present in $\mathbf{x}$, the signal to recover, are already zero. Indeed, for perfect reconstruction, frequencies 5 and 8 have to be set to zero. This can be achieved through two local interaction steps, so that for time $t = 8$ we perfectly recover the desired signal $\mathbf{x}$.

**Test Case 2.- Reconstruction performance for noisy graph signals.** Consider the social network in Fig. 3a where every node represents a member of the student government at the University of Ljubljana in Slovenia [13]. An edge between two students indicates that they have informally discussed university affairs. Assume that we want to reconstruct a 3-bandlimited signal in this graph, indicated by the node colors in Fig. 3a. We consider the graph-shift operator to be the




(a) (b)

Fig. 3: (a) A social network and a bandlimited graph signal to recover, where the signal values are indicated by the node colors. (b) Average reconstruction error as a function of the seeding node for the noisy graph-shift operator $\tilde{\mathbf{S}}$.

Laplacian of the graph, i.e. $\mathbf{S} = \mathbf{L}$. In this context, each value of the graph signal can be interpreted as a one-dimensional opinion of the corresponding individual regarding some topic, and each successive application of $\mathbf{S}$ can be seen as an opinion update influenced by the opinion of neighboring individuals. Furthermore, reconstruction using a single seeding node amounts to designing the opinion that an individual has to transmit to its neighbors at each time point to induce the desired global opinion in the graph. Within this framework, the first reconstruction scheme (cf. Sec. III-A) is more natural since the collaborative implementation of a low-pass filter needed for the second scheme (cf. III-B) would be an unrealistic assumption. Perfect reconstruction can be achieved as illustrated in Test Case 1 as long as the conditions stated in Prop. 2 are satisfied.

In a real scenario, however, opinion transmission is noisy since it is difficult for individuals to sense exactly the opinion of others. We model this by considering the graph-shift operator $\tilde{\mathbf{S}} = \mathbf{L}+\mathbf{W}$, where $\mathbf{W}$ has the same sparsity pattern as $\mathbf{L}$ and every non-zero element is independently drawn from a zero-mean normal distribution with standard deviation $10^{-3}$. Unlike the noiseless case, the reconstruction performance in noisy scenarios depends on the seeding node. Denoting by $\tilde{\mathbf{x}}_i$ the recovered signal when node $i$ acts as seeding node, we quantify the reconstruction error of node $i$ as $e(i) = \|\tilde{\mathbf{x}}_i - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$. The empirical error across 1000 noisy reconstructions is minimized by node 10 in Fig. 3a with an error of 0.17 and maximized by node 6 with an error of 3.6; see Fig. 3b. The worse performances – seeding nodes 6, 2, and 7 – are related to nodes that have limited action on the active frequencies, i.e. the absolute values of the elements $\{[\widehat{\mathbf{e}}_i]_k\}_{k=1}^3$ are small for $i \in \{6, 2, 7\}$. The opposite is true for the nodes achieving the best performances, i.e., nodes 10, 8, and 1. These issues will be further investigated in the journal version of this paper.

## V. Conclusions

Two schemes for the reconstruction of graph signals through percolation of data injected at a single seeding node were designed. In the first one, the seeding node injected successive values followed by applications of the graph-shift operator. In the second scheme, a shorter seeding phase was followed by the implementation of a low-pass filter. Conditions for perfect reconstruction were identified and illustrated through synthetic and real-world network experiments.

## REFERENCES

[1] D. Shuman, S. 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, Mar. 2013.

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

[3] 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)*, 2014, pp. 3864–3868.

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

[5] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *arXiv preprint arXiv:1504.04687*, 2015.

[6] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Sampling of graph signals: successive local aggregations at a single node," in *Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 8-11 2015.

[7] ——, "Interpolation of graph signals using shift-invariant graph filters," in *European Signal Process. Conf. (EUSIPCO)*, Nice, France, Aug. 31 - Sept. 4 2015.

[8] S. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2013, pp. 5445–5449.

[9] I. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *ArXiv e-prints arXiv:1111.5896*, 2011.

[10] X. Wang, P. Liu, and Y. Gu, "Local-set-based graph signal reconstruction," *CoRR*, vol. abs/1410.3944, 2014.

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

[12] 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.

[13] A. Ferligoj and A. Kramberger, "Recall versus recognition: comparison of the two alternative procedures for collecting social network data," in *Develop. in Stat. and Methodology: Proc. of the Intl. Conf. on Methodology and Stat., Bled, Slovenia, Sept. 24-26, 1992*, vol. 9. FDV, 1993, p. 121.


## APPENDIX A
## PROOF OF PROPOSITION 2

Condition *i)* is necessary for the top $K$ equations in (10)

$$\widehat{\mathbf{x}}_K = \mathbf{E}_K^T \text{diag}(\widehat{\mathbf{e}}_1)\mathbf{\Psi}\bar{\mathbf{x}} \tag{14}$$

to have a solution for an arbitrary $\widehat{\mathbf{x}}_K$. To see why this is the case, note that if $U_1 > 0$ then at least one of the first $K$ elements of $\widehat{\mathbf{e}}_1$, say element $k$, is zero. This implies that, for (14) to hold true, we must have $[\widehat{\mathbf{x}}_K]_k = 0$ which violates the assumption of $\widehat{\mathbf{x}}_K$ being arbitrary. Similarly, if $D_1 > 0$, denote as $k_1$ and $k_2$ a pair of indexes for which $\lambda_{k_1} = \lambda_{k_2}$ and as $\Psi_{k_1}^T$ and $\Psi_{k_2}^T$ the corresponding rows in matrix $\mathbf{\Psi}$. Since $\Psi_{k_1}^T\bar{\mathbf{x}} = \Psi_{k_2}^T\bar{\mathbf{x}}$ regardless of the value of $\bar{\mathbf{x}}$, (14) can only hold true if $[\widehat{\mathbf{x}}_K]_{k_1}/[\widehat{\mathbf{e}}_1]_{k_1} = [\widehat{\mathbf{x}}_K]_{k_2}/[\widehat{\mathbf{e}}_1]_{k_2}$. As before, this violates the assumption of $\widehat{\mathbf{x}}_K$ being arbitrary.

Condition *ii)* establishes the minimum number of variables required to solve the system of linear equations in (10) when condition *i)* is satisfied. Define as $\mathcal{R}$ the set of indices corresponding to zero elements in $\widehat{\mathbf{e}}_1$ or repeated rows in $\mathbf{\Psi}$. Notice that the cardinality of $\mathcal{R}$ is $U_2 + D_2$ and every index in $\mathcal{R}$ must be greater than $K$. Based on $\mathcal{R}$, define the selection matrices $\mathbf{E}_{\mathcal{R}} = [\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \ldots, \mathbf{e}_{i_{U_2+D_2}}]$ for all $i \in \mathcal{R}$ and $\mathbf{E}_{\bar{\mathcal{R}}} = [\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \ldots, \mathbf{e}_{i_{N-U_2+D_2}}]$ for all $i \notin \mathcal{R}$. Hence, we may separate the system of linear equations in (10) into two:

$$\mathbf{E}_{\mathcal{R}}^T[\widehat{\mathbf{x}}_K,\ \mathbf{0}]^T = \mathbf{E}_{\mathcal{R}}^T \text{diag}(\widehat{\mathbf{e}}_1)\mathbf{\Psi}\bar{\mathbf{x}}, \tag{15}$$

$$\mathbf{E}_{\mathcal{R}}^T[\widehat{\mathbf{x}}_K,\ \mathbf{0}]^T = \mathbf{0} = \mathbf{E}_{\mathcal{R}}^T \text{diag}(\widehat{\mathbf{e}}_1)\mathbf{\Psi}\bar{\mathbf{x}}. \tag{16}$$

If we fix $\tau = N - U_2 - D_2$, then $\mathbf{E}_{\bar{\mathcal{R}}}^T \text{diag}(\widehat{\mathbf{e}}_1)\mathbf{\Psi}$ is an invertible square matrix since it can be written as the multiplication of a diagonal matrix with no zeros in the diagonal $\mathbf{E}_{\bar{\mathcal{R}}}^T \text{diag}(\widehat{\mathbf{e}}_1)\mathbf{E}_{\bar{\mathcal{R}}}$ times a full-rank Vandermonde matrix $\mathbf{E}_{\bar{\mathcal{R}}}^T\mathbf{\Psi}$. This ensures that (15) can be solved. Moreover, if $\bar{\mathbf{x}}^*$ is a solution of (15), then (16) is satisfied automatically, i.e. $\mathbf{0} = \mathbf{E}_{\mathcal{R}}^T \text{diag}(\widehat{\mathbf{e}}_1)\mathbf{\Psi}\bar{\mathbf{x}}^*$. To see why this is the case, notice that $U_2$ rows of $\mathbf{E}_{\mathcal{R}}^T \text{diag}(\widehat{\mathbf{e}}_1)\mathbf{\Psi}$ are exactly zero, trivially satisfying (16) for any $\bar{\mathbf{x}}^*$. Also, each of the remaining $D_2$ equations in (16) can be written as a scalar times one of the $N - K - U_2 - D_2$ homogenous equations in (15), guaranteeing that $\bar{\mathbf{x}}^*$ also solves these $D_2$ equations. Since (15) and (16), which are equivalent to (10), can be solved when $\tau = N - U_2 - D_2$ and condition *i)* is satisfied, the proof concludes.

## APPENDIX B
## PROOF OF PROPOSITION 3

If we rewrite (13) as

$$\widehat{\mathbf{x}}_K = \left(\text{diag}(\widehat{\mathbf{h}}_K^*)\right)\left(\mathbf{E}_K^T \text{diag}(\widehat{\mathbf{e}}_1)\mathbf{E}_K\right)\left(\mathbf{E}_K^T\mathbf{\Psi}\right)\bar{\mathbf{x}}_K, \tag{17}$$

then it becomes clear that conditions *i)* and *ii)* ensure invertibility of the three square matrices multiplying $\bar{\mathbf{x}}_K$ in (17). $U_1 = 0$ ensures that $\mathbf{E}_K^T \text{diag}(\widehat{\mathbf{e}}_1)\mathbf{E}_K$ is invertible since it is a diagonal matrix with no zero elements in its diagonal. $D_1 = 0$ ensures that $\mathbf{E}_K^T\mathbf{\Psi}$ is invertible since it is a square Vandermonde matrix with no repeated rows. Finally, condition *ii)* ensures that $\text{diag}(\widehat{\mathbf{h}}_K^*)$ is invertible. To see why this is the case, recall that

$$\widehat{\mathbf{h}}^* = [\widehat{\mathbf{h}}_K^*, \mathbf{0}]^T = \mathbf{\Psi}\mathbf{h}^*. \tag{18}$$

Thus, if condition *ii)* is not satisfied, two rows of $\mathbf{\Psi}$ – one corresponding to a homogenous and one to a nonhomogenous equation in (18) – would be identical, forcing an element of $\widehat{\mathbf{h}}_K^*$ to be zero independently of the low-pass filter coefficients $\mathbf{h}^*$ chosen. This would imply that $\text{diag}(\widehat{\mathbf{h}}_K^*)$ is not invertible. If this is not the case, the independence of the rows of $\mathbf{\Psi}$ guarantees that we can find filter coefficients $\mathbf{h}^*$ such that every element in $\widehat{\mathbf{h}}_K^*$ is different from zero, showing invertibility of $\text{diag}(\widehat{\mathbf{h}}_K^*)$ and concluding the proof.