

HARMONIC ANALYSIS ON DIRECTED GRAPHS AND APPLICATIONS: FROM FOURIER ANALYSIS TO WAVELETS.

HARRY SEVI^{1,2}, GABRIEL RILLING¹, PIERRE BORGNAT²

ABSTRACT. We introduce a novel harmonic analysis for functions defined on the vertices of a strongly connected directed graph of which the random walk operator is the cornerstone. As a first step, we consider the set of eigenvectors of the random walk operator as a non-orthogonal Fourier-type basis for functions over directed graphs. We found a frequency interpretation by linking the variation of the eigenvectors of the random walk operator obtained from their Dirichlet energy to the real part of their associated eigenvalues. From this Fourier basis, we can proceed further and build multi-scale analyses on directed graphs. We propose both a redundant wavelet transform and a decimated wavelet transform by extending the diffusion wavelets framework by Coifman and Maggioni for directed graphs. The development of our harmonic analysis on directed graphs thus leads us to consider both semi-supervised learning problems and signal modeling problems on graphs applied to directed graphs highlighting the efficiency of our framework.

1. INTRODUCTION

In a world where data available for scientific or social purposes accumulates at an exponential pace, managing, exploiting and analyzing this torrent of data has become one of the challenges of our era. Some of them take form as graphs, structures that arise in various fields such as neurosciences, the Internet, genomic data, road transportations or social networks to name a few [78, 79]. Thus there is a need to develop efficient mathematical and computational approaches to process and analyze such graphs and data on graphs.

Among the existing methods, a large number involve the (graph) Laplacian [40, 39, 67, 42, 82]:

- As a fundamental subject in mathematics and physics, the Laplacian is known and used, through its spectral study to extract relevant geometric properties from a manifold [80] or a graph [34] and plays a role in machine learning applications such as spectral clustering [82].
- In continuous space, the eigenfunctions of the Laplace-Beltrami operator constitute the generalization of the Fourier basis on manifolds [80, 81].

¹ CEA, LIST, LABORATOIRE D'ANALYSE DE DONNÉES ET INTELLIGENCE DES SYSTÈMES, 91400 GIF-SUR-YVETTE, FRANCE

² UNIV LYON, ENS DE LYON, UCB LYON 1, CNRS, LABORATOIRE DE PHYSIQUE, F-69342 LYON, FRANCE

Date: February 18, 2019.

Key words and phrases. harmonic analysis, graph signal processing, Fourier analysis, wavelets, directed graphs, random walks, semi-supervised learning.

This work was supported by the ANR-14-CE27-0001 GRAPHISIP grant and the ACADEMICS Grant given by the IDEXLYON project of the Université de Lyon, as part of the "Programme Investissements d'Avenir" ANR-16-IDEX-0005.

- In discrete space, the eigenvectors of the graph Laplacian form an orthonormal Fourier-type basis for functions on undirected graphs [8, 7], when undirected graphs are considered as a discretization or sampling of a manifold [38, 37, 67, 42]. The eigenvalue associated with each eigenvector is related to the notion of frequency [8].

Thanks to these properties, the graph Laplacian bridges the gap between spectral graph theory and signal processing through the mathematical framework called "signal processing on graphs" [8, 7]. This framework aims at extending the concepts of classical signal processing such as filtering or sampling, for functions defined on the vertices of a graph.

First developed in the context of undirected graphs, signal processing on graphs considers the graph Laplacian (combinatorial or normalized) [34] as its core element. The graph Laplacian is a symmetric positive semidefinite operator. By the spectral theorem, the graph Laplacian admits an orthonormal basis of eigenvectors which can be considered as Fourier modes, the corresponding eigenvalues being associated with a notion of frequency.

The purpose of this paper is to provide some answers to the extension of the framework of signal processing on graphs to the case of directed graphs. The extension of signal processing on directed graphs is of interest because most data circulating on graphs found in the scientific (e.g. neurosciences) or social (e.g. social networks) fields are directed. Therefore, the analysis of information on directed graphs also requires the development of mathematical approaches and therefore an extension of the framework of signal processing on graphs to the case of directed graphs.

Formally, the direct use of the core element of graph signal processing, i.e. the graph Laplacian, is no longer appropriate. A directed graph is naturally represented by a non-symmetric adjacency matrix. It is always possible to naively define a graph Laplacian in the directed case, but the spectral properties associated with this undirected graph Laplacian (e.g. by obtaining an orthonormal basis and non-negative real eigenvalues) are no longer verified in the case of directed graphs. There is no simple consensus on a definition of a Laplacian for directed graphs for which the variation of eigenvectors is linked to a notion of frequency. This is the main challenge of signal processing on directed graphs and of this paper: which reference operator(s) should a Fourier analysis be built upon that would also generalize Laplacian-based Fourier analysis on undirected graphs ?

Hereafter, the random walk operator on graphs [12, 14, 11] is proposed as being a suitable reference operator for extending the framework of signal processing on directed graphs. Like the graph Laplacian, the random walk operator is associated to the notion of diffusion. It is as well defined on directed graphs as on undirected ones. Assuming that the random walk operator is diagonalizable, it potentially admits conjugated complex eigenvectors as well as associated conjugated complex eigenvalues. Our framework is based on the fact that the variational analysis of the eigenvectors of the random walk operator through their associated Dirichlet energy [14] is related to the real part of the eigenvalue [87]. Therefore, the Dirichlet energy of a given eigenvector of the random walk operator on a directed graph can be considered as its frequency, using the same analogy as in [8], now for directed graphs. Provided with this notion of frequency, the eigenvectors of the random walk

operator form a suitable Fourier basis on directed graphs. It is furthermore consistent with the undirected setting when the Fourier basis is based on the random-walk Laplacian operator (see section 6.2).

By now considering the random walk operator as a reference operator and its associated eigenvectors as a Fourier basis for functions defined on directed graphs, the last part of our harmonic analysis on directed graphs consists in building a multi-resolution analysis for functions defined on the vertices of an directed graph. Complex graphs, both directed and undirected, have structures at different scales. Motivated by the efficiency of multi-resolution wavelet analysis in traditional signal processing [36] and the multi-scale dimension of graph data, a number of multi-resolution graph constructions have emerged in recent years [21, 19, 20, 55, 52]. Here, we propose a new harmonic analysis built around the random walk operator and whose multi-resolution constructions generalize graph spectral wavelets [20, 19] and diffusion wavelets [21, 24] on directed graphs.

1.1. Literature review and related work. Wavelets [36, 50, 49, 46, 47, 48] have been thoroughly investigated for over two decades. Their efficiency and success in analyzing functions defined on the real line have led to their generalization for functions defined on higher dimensional spaces such as the sphere [43, 44, 45] or other manifolds [19, 21, 22, 5, 51, 52, 74]. The development of wavelets on manifolds and the multi-scale feature of graphs and data on graphs led to the extension of wavelets construction to the discrete space setting in recent years. These wavelet-type constructions include wavelets on unweighted graphs [53], lifting wavelets [54], diffusion wavelets [21, 22, 24], diffusion polynomial frames [19], spectral graph wavelets [20, 60, 62], Haar-like wavelets [56, 57, 66, 61], average interpolating wavelets [58], graph wavelets via deep learning [55], multi-scale pyramid transform [63], tight wavelet frames on graphs [52, 4] and intertwining wavelets on graphs via random forests [64]. The wavelet constructions mentioned above were all developed on undirected graphs. In the present work, we intend to construct a harmonic analysis on directed graphs. The first stage of the proposed harmonic analysis is the development of a Fourier analysis.

In recent years, Sandryhaila and Moura generalized some fundamental concepts of traditional signal processing such as filtering to directed graphs using the adjacency matrix as the central component of their framework [9]. They also proposed the generalized eigenvectors of the adjacency matrix obtained by Jordan decomposition as a Fourier-type basis on directed graphs [26]. As far as we know, no wavelet design was developed in this framework, with the exception of the development of the filterbank approach generalized in [6] which can be applied to bipartite directed graphs and uses the adjacency matrix for a polyphase representation of the filters.

More recently, new Fourier-type bases for functions on directed graphs built as the solutions of non-convex optimization problems. On the one hand, Sardellitti et al. have proposed the construction of a Fourier basis on directed graphs as a set of orthonormal vectors obtained by minimizing the Lovász extension of the graph cut size [30]. On the other hand, Shafipour et al. also have proposed the construction of a set of orthonormal vectors minimizing the spectral dispersion function [31]. Consequently, these bases proposed by [30, 31] are based on a specific notion of variation. Contrary to the approaches [30, 31] which are not based on a specific operator, our Fourier basis is non-orthogonal and obtained without solving non-convex optimization problems. In addition, the study of the Dirichlet energy of

the eigenvectors of the random walk operator enables the definition of a notion of frequency associated with the basis. Furthermore, the basis is well-defined mathematically, unlike the previous two since they are based on non-necessarily optimal solutions of non-convex optimization problems.

Mhaskar has recently proposed the first wavelet-type frame construction for functions defined on directed graphs [65] as an extension of the diffusion polynomial frames [19]. Although our redundant wavelet construction on directed graphs is inspired in part by that of polynomial diffusion frames, it differs mainly in the choice of the Fourier basis and the considered operator. Mhaskar proposes in [65] the set of left-singular vectors of the weighted adjacency matrix of a directed graph as a Fourier basis while we propose the set of eigenvectors of the random walk operator.

1.2. Contributions. The contributions of this article are the following:

- (1) **Construction of Fourier bases on directed graphs.** We propose the random walk operator associated with a random walk on a directed graph as a reference operator. We propose the eigenvectors of the random walk operator as a Fourier basis on directed graphs and determine a variational characterization of the eigenvectors of the random walk operator – see Proposition 6.1.
- (2) **Construction of multi-resolution analyses on directed graphs.** We propose an overcomplete spectral wavelet transform on directed graphs in Section 8.1. This construction extends the framework of spectral wavelets on undirected graphs [20] and diffusion polynomial frames [19]. We also propose a critically sampled wavelet transform in Section 8.2 generalizing the framework of diffusion wavelets [21, 24] to the directed case.
- (3) **Efficiency of the theoretical framework through applications.** The development of our harmonic analysis on directed graphs leads us to consider semi-supervised learning problems with regularization of type ℓ_2 in Section 7.1 and ℓ_1 regularization in section 8.3.3 and signal modeling by filtering on directed graphs in Section 7.2 highlighting the effectiveness of our theoretical framework.

1.3. Outline of the paper. This paper is structured as follows. We introduce the essential aspects of graph theory and review the foundations of graph signal processing in Section 2. We present operators defined on directed graphs built from the random walk operator and their properties in Section 4 and propose a new class of operators on directed graphs expressed as the convex combination between the random walk operator and its time version. Section 5 is dedicated to the presentation of a Fourier transform on directed graphs as the set of eigenvectors of the random walk operator. We propose in the Section 6 a Fourier-type analysis for functions defined on directed graphs by studying the variation of the random walk’s eigenvectors. To illustrate our Fourier-type analysis, we study the machine learning problems in Section 7 such as semi-supervised learning and signal modeling on directed graphs and show through numerical experiments the efficiency of our framework with respect to the existing approaches. Section 8 is dedicated to multiresolution analyses on directed graphs. There, we present both a redundant wavelets construction, as well as a critically sampled wavelets construction with the random walk operator as a reference operator and we illustrate these multiresolution analyses through some examples. We conclude in Section 9.

2. FUNDAMENTALS OF SIGNAL PROCESSING ON GRAPHS

Signal processing on graphs [8, 9, 7] is a mathematical framework in which the core concepts of harmonic analysis are generalized to functions defined on the vertices of a given arbitrary graph. In this section, we introduce the essential aspects and give some additional remarks to the signal processing on graphs framework.

2.1. Graph theory setup. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$ be a weighted directed graph where $\mathcal{V} = \{v_1, \dots, v_N\}$ is a finite set of vertices, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of directed edges. Each edge $e_{ij} = (v_i, v_j)$ is directed and represents a link from vertex v_i to vertex v_j . The weight function $w : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}_+$ satisfies the following conditions :

- $w(v_i, v_j) > 0$ if $e_{ij} \in \mathcal{E}$.
- $w(v_i, v_j) = 0$ if $e_{ij} \notin \mathcal{E}$.

We denote by $|\mathcal{V}| = N$, the cardinality of the vertex set \mathcal{V} , that is the total number of vertices in \mathcal{V} . A weighted directed graph \mathcal{G} can be entirely represented by its weighted adjacency matrix $\mathbf{W} = \{w_{ij}\}_{1 \leq i, j \leq N} \in \mathbb{R}_+^{N \times N}$ where $w_{ij} = w(v_i, v_j)$ are the weights associated to the respective edges e_{ij} . We define respectively the out-degree and the in-degree of a vertex $v_i \in \mathcal{V}$ by $d_i^+ = \sum_{j=1}^N w_{ij}$ and $d_i^- = \sum_{j=1}^N w_{ji}$. For the sake of simplicity, we refer to the out-degree d_i^+ of a vertex $v_i \in \mathcal{V}$ as its degree that we denote by d_i .

A limitation of our framework is that it only applies to the case of weighted graphs with nonnegative weights, that is $\mathbf{W} = \{w_{ij}\}_{1 \leq i, j \leq N} \in \mathbb{R}_+^{N \times N}$. Other approaches exist for the case of directed graphs with positive and negative weights but they are not compatible with the one presented here [32, 9, 26].

We assume through the theoretical sections of this paper that the directed graph \mathcal{G} is **strongly connected** and the random walk operators are **diagonalizable**. These two conditions are essential for our theoretical framework. If the directed graph is not strongly connected, the method of building the Google matrix [29] from the original adjacency matrix can be applied to ensure strong connectivity of the resulting directed graph. This will be used in Section 7.2.

2.2. Graph signals. Let $f : \mathcal{V} \mapsto \mathbb{C}$ be a function defined on the vertex set \mathcal{V} of a given directed graph \mathcal{G} . We define a graph signal \mathbf{f} as a column vector representation of the function f applied at each node $v_i \in \mathcal{V}$, that is

$$\mathbf{f} = [f(v_1), \dots, f(v_N)]^\top \in \mathbb{C}^N.$$

We now define the space of functions defined over the vertices of a directed graph.

Definition 2.1 ([33]). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph and $\mu : \mathcal{V} \mapsto [0, \infty)$ be a function on \mathcal{V} considered as a measure on \mathcal{V} by setting $\mu(U) = \sum_{x \in U} \mu(x)$, $U \subset \mathcal{V}$. For $q \in [1, \infty)$, we denote $\ell^q(\mu, \mathcal{V})$, the space of functions $f : \mathcal{V} \rightarrow \mathbb{C}$ such that*

$$\|\mathbf{f}\|_{\ell^q(\mu, \mathcal{V})} = \begin{cases} \left(\sum_{x \in \mathcal{V}} |f(x)|^q \mu(x) \right)^{1/q} < \infty, & q \in [1, \infty). \\ \max_{x \in \mathcal{V}} |f(x)| \mu(x) < \infty, & q = \infty. \end{cases}$$

We assume throughout this paper that the graph signals are defined in $\ell^2(\mathcal{V}, \mu)$ which is the Hilbert space of functions defined over the vertex set \mathcal{V} of \mathcal{G} endowed

with the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle_\mu = \sum_{x \in \mathcal{V}} \overline{f(x)} g(x) \mu(x),$$

for all $\mathbf{f}, \mathbf{g} \in \ell^2(\mathcal{V}, \mu)$ and where $\overline{f(x)}$ denotes the complex conjugate of $f(x)$.

3. LINEAR FILTERS ON GRAPHS

A linear operator on graph is represented by a matrix $\mathbf{H} \in \mathbb{C}^{N \times N}$ that acts on a graph signal input $\mathbf{s} \in \mathbb{C}^N$ and produces a graph signal output $\tilde{\mathbf{s}} \in \mathbb{C}^N$ according to the matrix vector product

$$\tilde{\mathbf{s}} = \mathbf{H}\mathbf{s}.$$

Within the framework of graph signal processing and inspired by conventional signal processing, we are mainly interested in a type of graph filters that commutes with a reference operator \mathbf{R} , i.e. :

$$\mathbf{H}\mathbf{R} = \mathbf{R}\mathbf{H}.$$

We define a graph filter \mathbf{H} as a finite polynomial sum of a reference operator \mathbf{R} , that is

$$\mathbf{H} = \sum_{t=0}^T h_t \mathbf{R}^t, \quad h_t \in \mathbb{C}, \quad \forall t = 0, \dots, T.$$

A graph filter \mathbf{H} commutes with \mathbf{R} . The following theorem establishes, under a particular condition, that a graph filter that commutes with a given reference operator can be expressed as a polynomial of the latter.

Theorem 3.1. *Let \mathbf{R} be a reference operator on a directed graph \mathcal{G} . We assume that \mathbf{R} is diagonalizable. We also assume that the characteristic and minimum polynomials of \mathbf{R} are equal, i.e. $p_{\mathbf{R}}(x) = m_{\mathbf{R}}(x)$. Then a graph filter \mathbf{H} that commutes with \mathbf{R} is expressed as a polynomial finite sum of \mathbf{R}*

$$\mathbf{H} = \sum_{t=0}^T h_t \mathbf{R}^t, \quad h_t \in \mathbb{C}, \quad \forall t = 0, \dots, T.$$

Theorem 3.1 indicates that if the filter \mathbf{H} verifies the following three conditions:

- (1) It commutes with a reference operator \mathbf{R} .
- (2) It is diagonalizable.
- (3) Each of the eigenspaces is of dimension one.

then the filter can be expressed as a polynomial finite sum of \mathbf{R} . We call a graph filter linear "shift" invariant [9] (LSI) all operator polynomial of \mathbf{R} . This notion of invariance with respect to a reference operator on graphs seems rather restrictive insofar as many reference operators on graphs do not systematically admit eigenspaces of dimension equal to 1 but eventually eigenspaces of dimension greater than 1. Thus, it will always be possible to construct a filter on graph which is a polynomial finite sum of a reference operator and therefore always invariant to its reference operator.

A graph filter can also be characterized by its eigenvalues on each eigenspace of \mathbf{R} . Let $\{\mathfrak{E}_j\}_{j=1}^m$ denote the eigenspaces of \mathbf{R} and $\{\mathbf{E}_j\}_{j=1}^m$ the corresponding

spectral projectors, characterized by $\mathbf{E}_i \mathfrak{E}_j = \delta_{ij} \mathfrak{E}_j$. A graph filter with respect to a diagonalizable operator \mathbf{R} is a linear combination of the spectral projectors

$$(1) \quad \mathbf{H} = \sum_{j=1}^m \gamma_j \mathbf{E}_j, \quad \gamma_j \in \mathbb{C}, \quad \mathbf{E}_j \in \mathbb{C}^{N \times N}, \quad \forall j = 1, \dots, m,$$

where γ_j is the eigenvalue of \mathbf{H} associated to eigenspace \mathfrak{E}_j .

4. CANONICAL OPERATORS ON DIRECTED GRAPHS

In this section, we introduce the fundamental linear operators and their properties in order to develop a harmonic analysis on directed graphs.

4.1. Random walk operators on graphs. We define a random walk on a directed graph as follows.

Definition 4.1 ([11, 12]). *A random walk on a strongly connected directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$ is a homogeneous Markov chain $\mathcal{X} = (X_n)_{n \geq 0}$ not necessarily reversible with a finite state space \mathcal{V} and whose transition probabilities are proportional to the edges weights. In particular, the entries of \mathbf{P} , written as $p(x, y)$, $\forall x, y \in \mathcal{V}$, give the transition matrix associated to \mathcal{X} on \mathcal{V} , and they are :*

$$p(x, y) = \mathbb{P}(X_{n+1} = y | X_n = x).$$

Since its entries are probabilities, it follows that

$$p(x, y) \geq 0 \quad \text{and} \quad \sum_{y \in \mathcal{V}} p(x, y) = 1, \quad \forall x, y \in \mathcal{V}.$$

From a graph theory point of view, the transition matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$ is equal to

$$\mathbf{P} = \mathbf{D}^{-1} \mathbf{W},$$

where $\mathbf{D} = \text{diag}\{d_1, \dots, d_n\}$, the diagonal matrix of the out-degrees of the vertices and \mathbf{W} the weighted adjacency matrix. For any graph signal \mathbf{f} , measure $\mu : \mathcal{V} \mapsto [0, \infty)$ and states $x, y \in \mathcal{V}$, we adopt the following conventions [10]:

$$\begin{aligned} \mathbf{P} \mathbf{f}(x) &= \sum_{y \in \mathcal{V}} p(x, y) f(y), \\ \mu \mathbf{P}(y) &= \sum_{x \in \mathcal{V}} \mu(x) p(x, y). \end{aligned}$$

where \mathbf{f} and μ are the vector representations of f and μ respectively. Let us consider the transition matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$ and a delta Kronecker function at vertex $k \in \mathcal{V}$, δ_k as a graph signal. If \mathbf{P} acts on δ_k , that is $\mathbf{P} \delta_k$, the mass at the vertex k propagates on the parent node to the initial node, i.e. the mass propagates from the node k to the node $k-1$. More generally, the random walk operator \mathbf{P} behaves as a local averaging on the children's nodes when it acts on a function on its right hand side. Given the importance of local averaging in signal processing, this suggests that the random walk operator \mathbf{P} may be natural as a core element of graph signal processing. We define the irreducibility of a random walk on a graph as follows.

Definition 4.2. *A random walk \mathcal{X} is said to be irreducible if for any $x, y \in \mathcal{V}$, the probability from x to reach y is strictly positive, in other words:*

$$\forall x, y \in \mathcal{V}, \exists m < \infty : \mathbb{P}(X_{n+m} = y | X_n = x) > 0.$$

Irreducibility is equivalent to the strong connectivity of \mathcal{G} , that is there is a (directed) path from any vertex $v_i \in \mathcal{V}$ to any vertex $v_j \in \mathcal{V}$.

Remark 4.1. *In the case of undirected connected graphs, the associated random walk is irreducible.*

We define the periodicity of a random walk on a graph as follows.

Definition 4.3. *Let \mathcal{X} be a random walk on \mathcal{G} . The period of a vertex $x \in \mathcal{V}$ is:*

$$\varrho(x) = \gcd\{n \in \mathbb{N}^+ : p^{(n)}(x, x) > 0\}.$$

Thus starting in x , the Markov chain can return to x only at multiples of the period ϱ . The state x is aperiodic if $\varrho(x) = 1$ and periodic if $\varrho(x) > 1$. Having defined the notions of irreducibility and periodicity, we are able to define the notion of ergodicity of a random walk.

Definition 4.4. *Let \mathcal{X} be a random walk on \mathcal{G} . The random walk \mathcal{X} is ergodic if it is irreducible and aperiodic.*

We set out the proposition for the stationary distribution of an ergodic random walk on \mathcal{G} .

Proposition 4.1 ([77]). *Let \mathcal{G} be a directed graph with finite state space \mathcal{V} . If a random walk \mathcal{X} with its transition matrix \mathbf{P} is ergodic, i.e. irreducible and aperiodic, the measures $\mathbf{P}^n(x, \cdot)$ converge towards the row vector $\boldsymbol{\pi} = [\pi(v_1), \dots, \pi(v_N)] \in \mathbb{R}_+^N$ as $n \rightarrow \infty$, i.e. the unique stationary distribution. In particular, $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, with :*

$$\sum_{i=1}^N \pi(v_i) = 1, \quad \pi(v_i) \geq 0.$$

4.1.1. *Reversibility.* Given the discrete-time Markov chain $\mathcal{X} = (X_n)_{n \geq 0}$ and $M > 0$ the finite time horizon, we define the time reversed Markov chain as $\mathcal{X}^* = (X_n^*) = (X_{M-n})$ for $n = 0, \dots, M$. We denote by $\mathbf{P}^* = \{p^*(x, y)\}_{x, y \in \mathcal{V}}$ the transition matrix associated with the Markov chain \mathcal{X}^* . It verifies

$$p^*(x, y) = \mathbb{P}(X_{n+1}^* = y | X_n^* = x) = \frac{\mathbb{P}(X_{M-n-1} = y)}{\mathbb{P}(X_{M-n} = x)} p(y, x), \quad \forall x, y \in \mathcal{V}.$$

If we assume that \mathcal{X} is ergodic with stationary distribution π , the time reversed Markov chain \mathcal{X}^* is also ergodic with stationary distribution π and the entries $p^*(x, y) \in \mathbf{P}^*$ are

$$p^*(x, y) = \frac{\pi(y)}{\pi(x)} p(y, x), \quad \forall x, y \in \mathcal{V}$$

or in its transition matrix version

$$\mathbf{P}^* = \boldsymbol{\Pi}^{-1} \mathbf{P}^\top \boldsymbol{\Pi},$$

where $\boldsymbol{\Pi} = \text{diag}\{\pi(v_1), \dots, \pi(v_N)\}$ is the diagonal matrix of the stationary distribution.

Let us introduce the function space $\ell^2(\mathcal{V}, \pi)$ endowed with its inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle_\pi = \sum_{x \in \mathcal{V}} \overline{f(x)} g(x) \pi(x).$$

In this space, \mathbf{P}^* is the adjoint of \mathbf{P} , that is $\langle \mathbf{f}, \mathbf{P}\mathbf{g} \rangle_\pi = \langle \mathbf{P}^*\mathbf{f}, \mathbf{g} \rangle_\pi$, for all $\mathbf{f}, \mathbf{g} \in \ell^2(\mathcal{V}, \pi)$.

The ergodic random walk \mathcal{X} with its transition matrix \mathbf{P} is reversible if and only if we have the following relation

$$\mathbf{P} = \mathbf{P}^*.$$

Remark 4.2. *Ergodic random walks on finite undirected graphs are reversible. That means that the transition matrix associated to the time reversed random walk \mathcal{X}^* , \mathbf{P}^* is equal to the transition matrix \mathbf{P} associated to the original ergodic random walk \mathcal{X} namely $\mathbf{P}^* = \mathbf{P}$.*

Remark 4.3. *In the undirected setting, the stationary distribution π admits a closed form expression. Indeed, on a given weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$ represented by its symmetric adjacency matrix $\mathbf{W} = \{w_{xy}\}_{x,y \in \mathcal{V}} \in \mathbb{R}_+^{N \times N}$ and the degree of a vertex $x \in \mathcal{V}$ is $c(x) = \sum_{y \in \mathcal{V}} w_{xy}$. As a result, the associated random walk is reversible with stationary distribution π defined by $\pi(x) = c(x)/c_{\mathcal{G}}$ where $c_{\mathcal{G}} = \sum_{x \in \mathcal{V}} c(x)$. However, in the directed setting, the stationary distribution π does not admit an analytical form. In order to calculate it, we use iterative methods such as the power iteration method [89] or Markov Chain Monte Carlo (MCMC) methods [90].*

Remark 4.4. *Ergodic random walks defined on directed graphs are typically non-reversible. Nevertheless, reversible ergodic random walks may be constructed on directed graphs by modifying the original non-reversible ergodic random walk. Such a modification is discussed in section 4.2.2.*

4.1.2. Eigenvalue distribution. By the Perron-Frobenius theorem [10], if the random walk \mathcal{X} is ergodic with a diagonalizable transition matrix \mathbf{P} , the diagonalization of \mathbf{P} admits a simple dominant eigenvalue $\lambda_{\max} = 1$. The other eigenvalues $\{\lambda \neq 1\}$ satisfy $|\lambda| < 1$, which implies that all eigenvalues different of λ_{\max} lie within the unit circle.

4.2. Random walk generalizations. Given a random walk \mathcal{X} on a directed graph \mathcal{G} with transition matrix \mathbf{P} , we are able to build new types of random walks based on \mathbf{P} with various purposes.

4.2.1. Lazy random walks. The periodicity of a random walk \mathcal{X} can be overcome by considering the lazy random walk version of \mathcal{X} . The transition matrix $\tilde{\mathbf{P}}$ associated to the lazy random walk $\tilde{\mathcal{X}}$, based upon \mathbf{P} is expressed as

$$\tilde{\mathbf{P}} = \frac{\mathbf{I} + \mathbf{P}}{2}.$$

The lazy random walk can be seen as the random walk on a modified graph $\tilde{\mathcal{G}}$ where an edge from each vertex of \mathcal{G} to itself is added with a weight equal to the vertex's degree in \mathcal{G} . The lazy random walk $\tilde{\mathcal{X}}$ is always aperiodic, hence when \mathcal{G} is strongly connected, $\tilde{\mathcal{X}}$ is also ergodic with a stationary measure π verifying $\pi = \pi \tilde{\mathbf{P}} = \pi \mathbf{P}$.

More generally, let us define $\tilde{\mathcal{P}}$, the class of transition matrices associated with generalized lazy random walks by

$$\tilde{\mathcal{P}} = \left\{ \tilde{\mathbf{P}}_{\gamma} : \tilde{\mathbf{P}}_{\gamma} = (1 - \gamma)\mathbf{P} + \gamma\mathbf{I} \middle| \gamma \in [0, 1] \right\}.$$

We note that the elements $\tilde{\mathbf{P}}_{\gamma} \in \tilde{\mathcal{P}}$ share the same eigenspaces as \mathbf{P} , hence the graph filters with respect to $\tilde{\mathbf{P}}$ are also graph filters with respect to \mathbf{P} .

4.2.2. *Additive reversibilization.* From a non-reversible ergodic random walk \mathcal{X} with transition matrix \mathbf{P} and unique stationary measure π , an additive reversibilization of \mathcal{X} can be constructed [13], denoted by $\bar{\mathcal{X}}$, whose transition matrix is the average between \mathbf{P} and its time reversed \mathbf{P}^* :

$$(2) \quad \bar{\mathbf{P}} = \frac{\mathbf{P} + \mathbf{P}^*}{2}.$$

$\bar{\mathcal{X}}$ is a reversible random walk with the same unique stationary distribution π .

More generally, let us define $\bar{\mathcal{P}}$ the class of convex combinations between the random walk matrix \mathbf{P} and its time reversed version \mathbf{P}^* as

$$(3) \quad \bar{\mathcal{P}} = \left\{ \bar{\mathbf{P}}_\alpha : \bar{\mathbf{P}}_\alpha = (1 - \alpha)\mathbf{P} + \alpha\mathbf{P}^* \mid \alpha \in [0, 1] \right\}.$$

We note that the elements $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$ share the same stationary distribution π but do not have the same eigenspaces. We also note that all the random walks associated to transition matrices $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$ are non-reversible except for $\alpha = 1/2$, namely $\bar{\mathbf{P}}_{1/2} = \bar{\mathbf{P}}$.

Remark 4.5. *Given an ergodic random walk $(\mathcal{X}, \mathbf{P}, \pi)$, we note that all lazy and reversibilized versions, respectively $\tilde{\mathbf{P}}_\gamma \in \tilde{\mathcal{P}}$ and $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$ share the same unique stationary distribution π .*

Remark 4.6. *As a generalization of random walks, we may also consider the multiplicative reversibilization of the ergodic non-reversible random walk \mathbf{P} , that is $\mathbf{P}\mathbf{P}^*$ which is also a reversible Markov chain with stationary distribution π . It is useful e.g. for defining the convergence bounds of non-reversible Markov chains [13].*

4.3. **Laplacians on directed graphs.** In this section, we introduce several definitions of Laplacians on directed graphs as there exists in the undirected case [76, 75].

4.3.1. *Normalized directed graph Laplacian.* The Laplacian on a directed graph \mathcal{G} is expressed in terms of the transition matrix \mathbf{P} of an ergodic random walk \mathcal{X} on \mathcal{G} and is defined as follows.

Definition 4.5 ([2, 1]). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph with $|\mathcal{V}| = N$. Let \mathcal{X} be an ergodic random walk on \mathcal{G} with transition matrix \mathbf{P} and unique stationary distribution π . The normalized Laplacian on \mathcal{G} is defined by*

$$(4) \quad \mathcal{L} = \mathbf{I} - \frac{\mathbf{\Pi}^{1/2}\mathbf{P}\mathbf{\Pi}^{-1/2} + \mathbf{\Pi}^{-1/2}\mathbf{P}^\top\mathbf{\Pi}^{1/2}}{2},$$

where \mathbf{I} is the identity matrix and $\mathbf{\Pi} = \text{diag}\{\pi(v_1), \dots, \pi(v_N)\}$ is the diagonal matrix of the stationary distribution.

4.3.2. *Random walk Laplacian.* Another definition is the random walk Laplacian \mathbf{L}_{RW} , defined as

$$(5) \quad \mathbf{L}_{\text{RW}} = \mathbf{I} - \bar{\mathbf{P}}.$$

The normalized Laplacian on directed graphs \mathcal{L} is connected to the random walk Laplacian through the following relation

$$\mathcal{L} = \mathbf{\Pi}^{1/2}\mathbf{L}_{\text{RW}}\mathbf{\Pi}^{-1/2}.$$

More generally, we define the random walk Laplacian on directed graphs associated to a transition matrix $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$ by

$$\mathbf{L}_{\text{RW},\alpha} = \mathbf{I} - \frac{\bar{\mathbf{P}}_\alpha + \bar{\mathbf{P}}_\alpha^*}{2}.$$

In the following proposition, we show that the random walk Laplacian $\mathbf{L}_{\text{RW},\alpha}$ is equal to \mathbf{L}_{RW} .

Proposition 4.2. *For any $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$, we have*

$$(6) \quad \mathbf{L}_{\text{RW},\alpha} = \mathbf{L}_{\text{RW}}, \quad \forall \alpha \in [0, 1].$$

Proof.

$$\begin{aligned} \mathbf{L}_{\text{RW},\alpha} &= \mathbf{I} - \frac{\bar{\mathbf{P}}_\alpha + \bar{\mathbf{P}}_\alpha^*}{2} \\ &= \mathbf{I} - \frac{\alpha \mathbf{P} + (1-\alpha) \mathbf{P}^* + \alpha \mathbf{P}^* + (1-\alpha) \mathbf{P}}{2} \\ &= \mathbf{I} - \frac{\alpha(\mathbf{P} + \mathbf{P}^*) + (1-\alpha)(\mathbf{P} + \mathbf{P}^*)}{2} \\ &= \mathbf{I} - \frac{\mathbf{P} + \mathbf{P}^*}{2} \\ &= \mathbf{L}_{\text{RW}}. \end{aligned} \quad \square$$

4.3.3. *Combinatorial Laplacian.* Finally, we also use the combinatorial Laplacian \mathbf{L} , defined as

$$(7) \quad \mathbf{L} = \mathbf{\Pi} - \frac{\mathbf{\Pi} \mathbf{P} + \mathbf{P}^\top \mathbf{\Pi}}{2}.$$

The latter is related to the random walk Laplacian through

$$\mathbf{L} = \mathbf{\Pi} \mathbf{L}_{\text{RW}}.$$

Remark 4.7. *As mentioned in remark 4.2, the transition matrix of a random walk on an undirected graph is equal to the transition matrix of its time reversed Markov chain. Furthermore, the stationary distribution π admits a closed form as we mention in remark 4.3. As a result, the definitions eqs. (4), (5) and (7), generalize the usual definitions for undirected graphs [34].*

4.4. **Canonical operators on directed graphs and Hilbert spaces.** Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph, $\ell^2(\mathcal{V})$ and $\ell^2(\mathcal{V}, \pi)$ be the Hilbert spaces associated respectively with the counting measure and with the stationary measure π of the ergodic random walk \mathcal{X} on \mathcal{G} . Let $\varphi : \ell^2(\mathcal{V}) \rightarrow \ell^2(\mathcal{V}, \pi)$ be the linear mapping defined as follows

$$(8) \quad \varphi : \mathbf{f} \mapsto \mathbf{\Pi}^{-1/2} \mathbf{f}, \quad \forall \mathbf{f} \in \ell^2(\mathcal{V}).$$

Definition 4.6 ([88]). *Let \mathbf{H}, \mathbf{K} be two Hilbert spaces. A linear transformation $\mathbf{V} : \mathbf{H} \mapsto \mathbf{K}$ is an isometry if and only if*

$$\|\mathbf{V}(\mathbf{x} - \mathbf{y})\|_{\mathbf{H}} = \|\mathbf{x} - \mathbf{y}\|_{\mathbf{K}}, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{H}.$$

Proposition 4.3. *The linear transformation $\varphi : \ell^2(\mathcal{V}) \mapsto \ell^2(\mathcal{V}, \pi)$ is an isometry.*

Proof. Given $\mathbf{g}, \mathbf{h} \in \ell^2(\mathcal{V})$, and the linear transformation φ , we have

$$\langle \varphi(\mathbf{g}), \varphi(\mathbf{h}) \rangle_\pi = \langle \mathbf{\Pi}^{-1/2} \mathbf{g}, \mathbf{\Pi}^{-1/2} \mathbf{h} \rangle_\pi = \langle \mathbf{g}, \mathbf{h} \rangle. \quad \square$$

Let us introduce the operator $\mathbf{T} \in \ell^2(\mathcal{V})$ defined as

$$(9) \quad \mathbf{T} = \mathbf{\Pi}^{1/2} \mathbf{P} \mathbf{\Pi}^{-1/2}.$$

Definition 4.7 ([88]). *Let \mathbf{H}, \mathbf{K} be two Hilbert spaces. An invertible bounded linear transformation $\mathbf{V} : \mathbf{H} \mapsto \mathbf{K}$ intertwines an operator $\mathbf{M} \in \mathbf{H}$ to an operator $\mathbf{S} \in \mathbf{K}$ if*

$$\mathbf{V} \mathbf{M} = \mathbf{S} \mathbf{V}.$$

The operators \mathbf{M} and \mathbf{S} are called similar.

Proposition 4.4. *The linear operator φ intertwines the operator $\mathbf{T} \in \ell^2(\mathcal{V})$ to the operator $\mathbf{P} \in \ell^2(\mathcal{V}, \pi)$.*

Proof.

$$\varphi(\mathbf{T}) = \mathbf{\Pi}^{-1/2} \mathbf{T} = \mathbf{\Pi}^{-1/2} \mathbf{\Pi}^{1/2} \mathbf{P} \mathbf{\Pi}^{-1/2} = \mathbf{P} \mathbf{\Pi}^{-1/2}. \quad \square$$

Hence the operators \mathbf{T} and \mathbf{P} are similar with respect to φ . Identically, the random walk Laplacian \mathbf{L}_{RW} and the normalized Laplacian \mathcal{L} are also similar with respect to φ .

5. DIRECTED GRAPH FOURIER TRANSFORM

Let \mathbf{P} be the transition matrix of the ergodic random walk \mathcal{X} . We assume \mathbf{P} diagonalizable, that is \mathbf{P} admits an eigenvalue decomposition

$$\mathbf{P} = \mathbf{\Xi} \mathbf{\Theta} \mathbf{\Xi}^{-1},$$

where $\mathbf{\Xi} = [\xi_1, \dots, \xi_N]$ is an eigenbasis with elements $\xi_j, j = 1, \dots, N$ and $\mathbf{\Theta} = \text{diag}\{\vartheta_1, \dots, \vartheta_N\}$ is the corresponding diagonal eigenvalue matrix, non necessarily distinct. Given a graph signal \mathbf{s} , its directed graph Fourier transform denoted by $\hat{\mathbf{s}} = [\hat{s}_1, \dots, \hat{s}_N]$ is

$$\hat{\mathbf{s}} = \mathbf{\Xi}^{-1} \mathbf{s}.$$

The values $\{\hat{s}_j\}_{j=1}^N$ characterize the content of the graph signal \mathbf{s} in the graph Fourier domain.

The graph Fourier transform is not uniquely defined as it depends on the choice of eigenbasis $\mathbf{\Xi}$. First all eigenvectors ξ_j are defined up to a multiplicative scalar. This can be easily resolved by normalizing the vectors and e.g. making the first non-zero coefficient of each eigenvector real and positive. Second, whenever there are eigenspaces with dimension greater than 1, there are infinite choices for the orientations of the basis vectors within such eigenspaces. There is generally no solution to this second issue. This implies that the values of individual Fourier coefficients \hat{s}_j within such eigenspaces are rather meaningless. However, it only impacts cases with eigenspaces of dimension greater than 1, which are not that common in applications.

The inverse directed graph Fourier transform corresponding to an eigenbasis $\mathbf{\Xi}$ is given by

$$\mathbf{s} = \mathbf{\Xi} \hat{\mathbf{s}}.$$

It reconstructs the original signal from its frequency contents by forming a linear combination of eigenvectors weighted by their Fourier coefficients.

More generally, any $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$ admits an eigenvalue decomposition

$$\bar{\mathbf{P}}_\alpha = \mathbf{\Xi}_\alpha \mathbf{\Theta}_\alpha \mathbf{\Xi}_\alpha^{-1}.$$

Consequently, one can build an infinity of Fourier-type bases $\{\Xi_\alpha\}_{\alpha \in [0,1]}$ for functions defined on directed graphs. Among these, the case $\alpha = 1/2$ is particularly interesting insofar as $\bar{\mathbf{P}}_{1/2} = \bar{\mathbf{P}}$ is self-adjoint in $\ell^2(\mathcal{V}, \pi)$. This implies that there is an orthonormal eigenbasis $\Xi_{1/2}$ which yields the following theorem.

Theorem 5.1 (Generalized Parseval's Theorem). *Given an eigenbasis $\Xi_{1/2}$ of $\bar{\mathbf{P}}$ that is orthonormal in $\ell^2(\mathcal{V}, \pi)$, the corresponding Fourier transform is an isometric operator from $\ell^2(\mathcal{V}, \pi)$ to $\ell^2(\{1, \dots, N\})$*

$$(10) \quad \langle \Xi_{1/2} \mathbf{x}, \Xi_{1/2} \mathbf{y} \rangle_\pi = \langle \mathbf{x}, \mathbf{y} \rangle, \quad \forall \mathbf{x}, \mathbf{y} \in \ell^2(\{1, \dots, N\}).$$

Proof. $\Xi_{1/2}^\top \Pi \Xi_{1/2} = \mathbf{I} \implies \langle \Xi_{1/2} \mathbf{x}, \Xi_{1/2} \mathbf{y} \rangle_\pi = \mathbf{x}^\top \Xi_{1/2}^\top \Pi \Xi_{1/2} \mathbf{y} = \mathbf{x}^\top \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle \quad \square$

Another interesting property of $\bar{\mathbf{P}}$ is that its eigenspaces are the same as those of the random walk Laplacian \mathbf{L}_{RW} (5).

Remark 5.1. *In the undirected setting, a suitable Fourier-type basis is usually characterized by an orthonormal eigenbasis of the combinatorial Laplacian \mathbf{L} or its normalized counterpart \mathcal{L} [34]. These Fourier-type bases are both orthonormal for $\ell^2(\mathcal{V})$. Another less common choice is a Fourier-type basis based upon the random walk Laplacian $\mathbf{L}_{\text{RW}} = \mathbf{I} - \mathbf{P}$. As in the directed case with $\bar{\mathbf{P}}$, it leads to a Fourier basis that is orthonormal in $\ell^2(\mathcal{V}, \pi)$, where π is proportional to the degrees of the vertices (see remark 4.3). Some properties of the different definitions are investigated in [75].*

6. FOURIER ANALYSIS ON DIRECTED GRAPHS

In this section, we propose a new analysis on directed graphs that is different from the existing approaches on directed graphs [26, 30, 31, 65]. It is based on the study of variations in the eigenvectors of the random walk operator. Before discussing it in greater detail, we introduce the elements that allow us to study the variation of signals on directed graphs.

6.1. Regularity of signals on graphs. The behavior of a graph signal over a directed graph (or undirected) can be analyzed by measuring its regularity or smoothness. We first define the length of the graph gradient at a given vertex as follows.

Definition 6.1 ([14]). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph and $\nu : \mathcal{E} \rightarrow [0, \infty)$ be a positive measure defined on the edge set \mathcal{E} . The length of the graph gradient of a graph signal \mathbf{f} at vertex $x \in \mathcal{V}$ on an arbitrary graph \mathcal{G} under the measure ν is*

$$|\nabla \mathbf{f}(x)| = \left(\frac{1}{2} \sum_{y \in \mathcal{V}, (x,y) \in \mathcal{E}} |f(x) - f(y)|^2 \nu(x,y) \right)^{1/2}.$$

Intuitively, the length of the graph gradient measures the smoothness of a graph signal around a given vertex. We now introduce the Dirichlet energy as a measure of the smoothness of a signal over a strongly connected graph.

Definition 6.2 ([77]). *The Dirichlet energy of a graph signal \mathbf{f} associated to the ergodic random walk \mathcal{X} with transition matrix \mathbf{P} and stationary distribution π on a strongly connected graph \mathcal{G} is*

$$\begin{aligned}
(11) \quad \mathcal{D}_{\pi, \mathbf{P}}^2(\mathbf{f}) &= \frac{1}{2} \sum_{(x,y) \in \mathcal{E}} \pi(x)p(x,y)|f(x) - f(y)|^2, \\
&= \langle \mathbf{f}, \mathbf{L}_{\text{RW}} \mathbf{f} \rangle_{\pi}.
\end{aligned}$$

As we can appreciate, $\mathcal{D}_{\pi, \mathbf{P}}^2(\mathbf{f}) = \|\nabla \mathbf{f}\|_{2, \nu}^2$ where $\nu(x, y) = \pi(x)p(x, y)$ where $p(x, y)$ is the (x, y) entry of the transition matrix \mathbf{P} and $\pi(x)$ is the stationary distribution at vertex x . For all $\bar{\mathbf{P}}_{\alpha} \in \bar{\mathcal{P}}$, the associated Dirichlet energy is the same:

$$\mathcal{D}_{\pi, \bar{\mathbf{P}}_{\alpha}}^2(\mathbf{f}) = \mathcal{D}_{\pi, \mathbf{P}}^2(\mathbf{f}) = \langle \mathbf{f}, \mathbf{L}_{\text{RW}} \mathbf{f} \rangle_{\pi}.$$

We also introduce the Rayleigh quotient of a graph signal \mathbf{f} associated to the Dirichlet energy $\mathcal{D}_{\pi, \mathbf{P}}^2(\mathbf{f})$ as

$$\mathcal{R}_{\pi, \mathbf{P}}(\mathbf{f}) = \frac{\mathcal{D}_{\pi, \mathbf{P}}^2(\mathbf{f})}{\|\mathbf{f}\|_{\pi}^2}.$$

For any $\bar{\mathbf{P}}_{\alpha} \in \bar{\mathcal{P}}$, we have

$$\mathcal{R}_{\pi, \mathbf{P}}(\mathbf{f}) = \mathcal{R}_{\pi, \bar{\mathbf{P}}_{\alpha}}(\mathbf{f}).$$

6.2. Frequency analysis on directed graphs. The following proposition settles a key connection between the regularity of the eigenvectors of the transition matrix \mathbf{P} of the ergodic random walk \mathcal{X} and their associated eigenvalues.

Proposition 6.1. *Let $\xi \in \mathbb{C}^N$ be an eigenvector of a transition matrix \mathbf{P} of an ergodic random walk \mathcal{X} , with stationary distribution π , associated to the eigenvalue $\vartheta \in \mathbb{C}$. The Rayleigh quotient of ξ is given by*

$$\mathcal{R}_{\pi, \mathbf{P}}(\xi) = 1 - \Re(\vartheta),$$

where $\Re(\vartheta)$ denotes the real part of $\vartheta \in \mathbb{C}$.

Proof.

$$\begin{aligned}
\mathcal{R}_{\pi, \mathbf{P}}(\xi) &= \frac{1}{\|\xi\|_{\pi}^2} \left(\langle \xi, \xi \rangle_{\pi} - \frac{1}{2} \langle \xi, \mathbf{P} \xi \rangle_{\pi} - \frac{1}{2} \langle \xi, \mathbf{P}^* \xi \rangle_{\pi} \right) \\
&= \frac{1}{\|\xi\|_{\pi}^2} \left(\|\xi\|_{\pi}^2 - \frac{\vartheta + \bar{\vartheta}}{2} \|\xi\|_{\pi}^2 \right) \\
\mathcal{R}_{\pi, \mathbf{P}}(\xi) &= [1 - \Re(\vartheta)]. \quad \square
\end{aligned}$$

The latter proposition thus indicates that the smoothness of any eigenvector ξ of \mathbf{P} , as described by its Rayleigh quotient, is associated to the real part of its respective eigenvalue ϑ . More generally, the smoothness of any eigenvector ξ_{α} of $\bar{\mathbf{P}}_{\alpha} \in \bar{\mathcal{P}}$ is characterized exactly in the same manner as in Proposition 6.1, i.e.

$$\mathcal{R}_{\pi, \mathbf{P}}(\xi_{\alpha}) = 1 - \Re(\vartheta_{\alpha}).$$

Therefore, we are now able to associate to each eigenvector ξ of \mathbf{P} and generally to each eigenvector ξ_{α} of a given $\bar{\mathbf{P}}_{\alpha} \in \bar{\mathcal{P}}$, a value ω characterizing its variation that we call *frequency* expressed intuitively as

$$(12) \quad \omega = 1 - \Re(\vartheta), \quad \omega \in [0, 2].$$

Remark 6.1. *In the undirected setting, the random walk is reversible such that the Dirichlet energy associated to the random walk operator \mathbf{P} is*

$$\mathcal{D}_{\pi, \mathbf{P}}(\mathbf{f}) = \langle \mathbf{f}, \mathbf{L}_{\text{RW}} \mathbf{f} \rangle_{\pi},$$

where $\mathbf{L}_{\text{RW}} = \mathbf{I} - \mathbf{P}$. The random walk operator \mathbf{P} is self-adjoint in $\ell^2(\mathcal{V}, \pi)$. As a result, given an eigenvector $\boldsymbol{\xi}$ of \mathbf{P} associated to eigenvalue ϑ , the associated Rayleigh quotient is

$$\mathcal{R}_{\pi, \mathbf{P}}(\boldsymbol{\xi}) = 1 - \vartheta.$$

Therefore, the variation of the eigenvectors of \mathbf{P} is directly related to their respective eigenvalues. A similar result holds when the Fourier basis is based on the combinatorial Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$ thanks to the following identity

$$\mathcal{D}_{\pi, \mathbf{P}}(\mathbf{f}) = \langle \mathbf{f}, \mathbf{L}_{\text{RW}} \mathbf{f} \rangle_{\pi} = \langle \mathbf{f}, \mathbf{L} \mathbf{f} \rangle.$$

As \mathbf{L} is a symmetric semi-definite operator, its eigenvalues are nonnegative and real. Hence the Rayleigh quotient of an eigenvector $\boldsymbol{\phi}$ of \mathbf{L} associated to an eigenvalue λ is $\mathcal{R}_{\pi, \mathbf{P}}(\boldsymbol{\phi}) = \lambda$, which is the definition of frequency considered in [8].

6.2.1. On the content of subspaces associated to the random walk. The eigenvalues of an ergodic transition matrix \mathbf{P} associated to a random walk on a strongly connected directed graph \mathcal{G} are either real or come as complex-conjugate pairs $(\vartheta, \bar{\vartheta}) \in \mathbb{C}^2, \vartheta = \alpha + i\beta$. In the general case, a mono-frequency subspace can be composed of eigenspaces corresponding to any number of pairs of complex-conjugate eigenvalues and possibly one real eigenvalue. In the following, we examine how these individual eigenspaces can be understood with respect to each other.

Let us first consider pairs of eigenspaces related to complex-conjugate eigenvalues $\alpha \pm i\beta$. Assuming that the eigenspaces have dimension equal to 1, the eigenvectors can be chosen as complex-conjugate $\boldsymbol{\xi}, \bar{\boldsymbol{\xi}} \in \mathbb{C}^N$. These eigenvectors share the same frequency $1 - \alpha$ and can be seen as a generalization of the Fourier modes at frequencies ω and $-\omega$ in one-dimensional classical Fourier analysis. One can also define the corresponding real-valued Fourier modes

$$\boldsymbol{\xi}^{\cos} = \frac{\boldsymbol{\xi} + \bar{\boldsymbol{\xi}}}{2}, \quad \boldsymbol{\xi}^{\sin} = \frac{\boldsymbol{\xi} - \bar{\boldsymbol{\xi}}}{2i},$$

which generalize the cosine and sine functions to graph signals. A graph mono-frequency subspace spanned by only two complex-conjugate eigenvectors is thus analogous to the subspace spanned by the cosine and sine functions at a given frequency in classical Fourier analysis. As with the discrete Fourier transform, there are also frequencies for which the frequency subspace has dimension equal to one (the zero frequency and possibly the 1/2 frequency). The case of real eigenvalues in graph signal processing can be seen as a generalization of these.

When the eigenspaces have dimension greater than 1, the situation is similar to the multidimensional classical Fourier analysis setting where the frequency is a vector $\boldsymbol{\omega} = [\omega_1, \dots, \omega_n]^{\top}$.

Let us now consider two eigenvalues with the same real part but different non-conjugate imaginary parts. Continuing the analogy with classical Fourier analysis, in two dimensions Fourier modes with a given frequency ω can be seen as linear combinations of, say, vertical and horizontal Fourier modes with the same frequency, which are two *different orientations*. This is illustrated in section 6.4.2 on the directed toroidal graph.

6.3. Random walk graph filters. In our theoretical framework, our reference operator is the random walk operator \mathbf{P} . Assuming that \mathbf{P} can be diagonalized, we can represent a graph filter \mathbf{H} as a linear combination of the spectral projectors $\{\mathbf{E}_{\vartheta_k}\}_{k=1}^m$ associated with \mathbf{P} , as defined at the eq. (1), i.e. if $\mathbf{R} = \mathbf{P}$,

$$\mathbf{H} = \sum_{k=1}^m \gamma_k \mathbf{E}_{\vartheta_k}, \quad \gamma_k \in \mathbb{C}, \quad k = 1, \dots, m.$$

In the previous section, we have defined the notion of mono-frequency random walk subspace, that is the subspace spanned by the eigenvectors of \mathbf{P} associated to a given frequency obtained by studying the variation of these eigenvectors. Consequently, we first define a random walk graph filter \mathbf{H}_ω as follows

$$\mathbf{H}_\omega = \sum_{\omega \in \boldsymbol{\omega}} \tau_\omega \mathbf{S}_\omega, \quad \tau_\omega \in \mathbb{C}.$$

where each τ_ω is the response depending only on the frequency $\omega \in \boldsymbol{\omega}$ and \mathbf{S}_ω is the projector associated to the graph frequency $\omega \in \boldsymbol{\omega}$ defined as

$$\mathbf{S}_\omega = \sum_{\vartheta: 1 - \Re(\vartheta) = \omega} \mathbf{E}_\vartheta$$

Let $h : \boldsymbol{\omega} \mapsto \mathbb{R}$ (or \mathbb{C}) be the frequency response associated to the frequency $\omega \in \boldsymbol{\omega}$. We also are able to construct a random walk graph filter with frequency response \mathbf{H}_ω as follows

$$(13) \quad \mathbf{H}_\omega = \sum_{\omega \in \boldsymbol{\omega}} h(\omega) \mathbf{S}_\omega, \quad \omega \in [0, 2].$$

This latter definition allows one to play with the frequency response by e.g. shifting, contracting or dilating it along the frequency axis.

6.4. Fourier analysis on finite groups: a graph signal perspective. In order to show the consistency of our Fourier analysis with respect to traditional signal processing, we depict the directed counterparts of two well-known objects: the cycle graph associated with the cyclic group $(\mathbb{Z}/n\mathbb{Z})$ and the toroidal graph being is the direct product of cyclic groups, i.e. $\mathbb{T} = (\mathbb{Z}/n_1\mathbb{Z}) \oplus \dots \oplus (\mathbb{Z}/n_r\mathbb{Z})$ [15].

6.4.1. Fourier analysis on the directed cycle graph. A directed cycle graph \mathcal{C}_N is a graph with N vertices containing a single cycle through all the vertices and where all the edges are directed in the same direction. The directed cycle graph \mathcal{C}_N is represented by its adjacency matrix \mathbf{C}_N which is circulant

$$\mathbf{C}_N = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}$$

As the out-degree matrix of \mathcal{C}_N is $\mathbf{D} = \mathbf{I}_N$, we define a random walk \mathcal{X} on \mathcal{C}_N whose transition matrix is characterized by its adjacency matrix, namely $\mathbf{P} = \mathbf{C}_N$. \mathbf{P} is diagonalizable, i.e. $\mathbf{P} = \mathbf{\Xi} \mathbf{\Theta} \mathbf{\Xi}^{-1}$ where $\mathbf{\Theta} = \text{diag}\{\vartheta_1, \dots, \vartheta_N\} \in \mathbb{C}^{N \times N}$ is the eigenvalue matrix where

$$\vartheta_k = e^{2\pi i(k-1)/N}, \quad k = 1, \dots, N.$$

and $\Xi = [\xi_1, \dots, \xi_N] \in \mathbb{C}^{N \times N}$ is the discrete Fourier basis where

$$\xi_k = \frac{1}{\sqrt{N}} [1, \vartheta_k^1, \dots, \vartheta_k^{N-1}]^\top \quad k = 1, \dots, N.$$

The random walk \mathcal{X} is irreducible and periodic. As our frequency analysis only deals with ergodic random walks, we need \mathcal{X} to be aperiodic. In order to overcome the periodicity issue, we consider a γ -lazy random walk $\tilde{\mathcal{X}}$ characterized by its transition matrix $\tilde{\mathbf{P}}_\gamma \in \tilde{\mathcal{P}}$ defined at section 4.2.1. As a consequence, the Rayleigh quotient associated to a given eigenvector ξ_k is given by

$$\mathcal{R}_{\pi, \mathbf{P}}(\xi_k) = (1 - \gamma) \left[1 - \cos \left(\frac{2\pi(k-1)}{N} \right) \right] = \omega_k, \quad k = 1, \dots, N.$$

Thus, the ordering of the frequencies $\{\omega_k\}_{k=1}^N$ associated with the eigenvectors $\{\xi_k\}_{k=1}^N$ by the Rayleigh quotient coincides exactly with the classical signal processing approach.

By taking the limit $\gamma \rightarrow 0$, we could define frequencies for no-lazy random walks. More generally, our Fourier analysis works perfectly with irreducible random walks by taking the same kind of limit.

6.4.2. Fourier analysis on the directed toroidal graph. A directed toroidal graph $\mathcal{T}_{m,n}$ is the Cartesian product of the directed cycle graphs \mathcal{C}_m and \mathcal{C}_n , namely $\mathcal{T}_{m,n} = \mathcal{C}_m \square \mathcal{C}_n$. We introduce necessary definitions for the understanding of the section.

Definition 6.3 ([18]). Let $\mathcal{G} = (\mathcal{U}, \mathcal{E})$ and $\mathcal{H} = (\mathcal{V}, \mathcal{F})$ be two graphs with respective vertex sets $\mathcal{U} = (u_1, \dots, u_n)$ and $\mathcal{V} = (v_1, \dots, v_m)$. The Cartesian product of \mathcal{G} and \mathcal{H} is the graph $\mathcal{G} \square \mathcal{H}$ with vertex set $\mathcal{U} \times \mathcal{V}$ in which two vertices $x = (u_i, v_j)$ and $y = (u_p, v_q)$ are adjacent if and only if either $u_i = u_p$ and $(v_j, v_q) \in \mathcal{F}$ or $v_j = v_q$ and $(u_i, u_p) \in \mathcal{E}$.

Definition 6.4 ([18]). Let $\mathcal{G} = (\mathcal{U}, \mathcal{E})$ and $\mathcal{H} = (\mathcal{V}, \mathcal{F})$ be two graphs. The adjacency matrix of the Cartesian product $\mathcal{G} \square \mathcal{H}$, denoted $\mathbf{A}_{\mathcal{G} \square \mathcal{H}}$ is

$$\mathbf{A}_{\mathcal{G} \square \mathcal{H}} = \mathbf{A}_{\mathcal{G}} \otimes \mathbf{I}_{|\mathcal{V}|} + \mathbf{I}_{|\mathcal{U}|} \otimes \mathbf{A}_{\mathcal{H}}.$$

where \otimes is the Kronecker product symbol.

Lemma 6.1. Suppose $\lambda_1, \dots, \lambda_n$ are eigenvalues of $\mathbf{A}_{\mathcal{G}}$ and μ_1, \dots, μ_m are eigenvalues of $\mathbf{A}_{\mathcal{H}}$. Then the eigenvalues of $\mathbf{A}_{\mathcal{G} \square \mathcal{H}}$ are all $\lambda_i + \mu_j$ for $1 \leq i \leq n$ and $1 \leq j \leq m$. Moreover, if u and v are eigenvectors for $\mathbf{A}_{\mathcal{G}}$ and $\mathbf{A}_{\mathcal{H}}$ with eigenvalues λ and μ respectively, then the vector $\mathbf{w} = u \otimes v$ is an eigenvector of $\mathbf{A}_{\mathcal{G} \square \mathcal{H}}$ with eigenvalue $\lambda + \mu$.

Directed toroidal graph. Let $\mathcal{T}_{m,n} = \mathcal{C}_m \square \mathcal{C}_n$ be the directed toroidal graph characterized by its adjacency matrix $\mathbf{A}_{\mathcal{T}_{m,n}} = \mathbf{C}_m \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{C}_n$. The eigenvalues of \mathbf{C}_m are $\lambda_k = e^{2i\pi(k-1)/m}$ for $k \in \llbracket 1, m \rrbracket$ and the eigenvectors are denoted by $[\phi_1, \dots, \phi_m]$. Identically, the eigenvalues of \mathbf{C}_n are $\mu_\ell = e^{2i\pi(\ell-1)/n}$ for $\ell \in \llbracket 1, n \rrbracket$ and the eigenvectors are denoted by $[\psi_1, \dots, \psi_n]$. The directed toroidal graph $\mathcal{T}_{m,n}$ is a directed 2-regular graph. Consequently, the out-degree matrix of $\mathcal{T}_{m,n}$ is $\mathbf{D}_{\mathcal{T}_{m,n}} = 2\mathbf{I}_{m \times n}$. We thus define a random walk \mathcal{X} on $\mathcal{T}_{m,n}$ with a diagonalizable transition matrix $\mathbf{P} = \mathbf{D}_{\mathcal{T}_{m,n}}^{-1} \mathbf{A}_{\mathcal{T}_{m,n}}$. We specify its spectral properties in the following lemma.

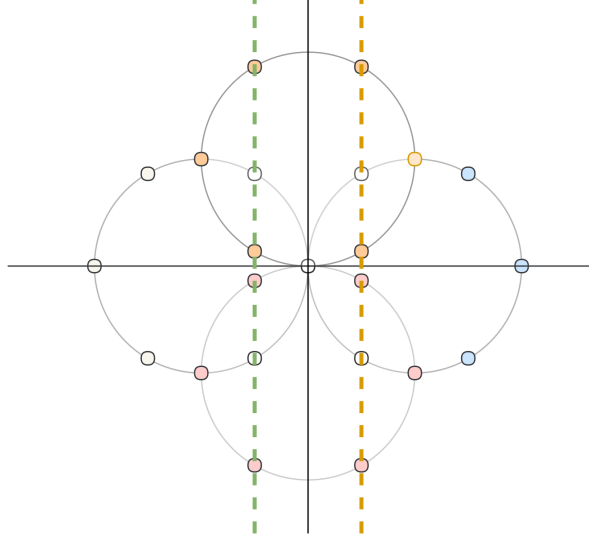


FIGURE 1. Eigenvalue distribution of the random walk matrix of a directed toroidal graph $\mathcal{T}_{6,4}$. Eigenvalues on the dotted green line: $\Re(\lambda^g) = -0.25, \forall \lambda^g \in \Lambda_g$. Eigenvalues on the dotted orange line: $\Re(\lambda^o) = 0.25, \forall \lambda^o \in \Lambda_o$.

Lemma 6.2. *Let $\mathcal{T}_{m,n}$ be a directed toroidal graph and \mathcal{X} be a random walk associated with the diagonalizable transition matrix \mathbf{P} . The directed toroidal graph $\mathcal{T}_{m,n}$ is a directed 2-regular graph. Consequently, the spectrum of \mathbf{P} forms the set*

$$\sigma(\mathbf{P}) = \left\{ \zeta_{i,j} = \frac{\lambda_i + \mu_j}{2}, i \in \llbracket 1, m \rrbracket, j \in \llbracket 1, n \rrbracket \right\},$$

with the associated eigenvectors

$$\left\{ \phi_i \otimes \psi_j, i \in \llbracket 1, m \rrbracket, j \in \llbracket 1, n \rrbracket \right\}.$$

As at the section 6.4.1, the associated random walk \mathcal{X} is irreducible and periodic. In order to overcome this periodicity issue, we consider a γ -lazy random walk $\tilde{\mathcal{X}}$ with transition matrix $\tilde{\mathbf{P}}_\gamma \in \tilde{\mathcal{P}}$.

Figure X 1 illustrates the eigenvalue distribution of the transition matrix of a directed toroidal graph. As we can appreciate, the distribution of eigenvalues is remarkable. We denote by Λ_g , the set of eigenvalues located on the green dotted line and Λ_o the set of eigenvalues located on the orange dotted line. Sets Λ_g , and Λ_o , include respectively eigenvalues having the same real part and different imaginary parts i.e. $\Re(\Lambda_g) = -0.25$ and $\Re(\Lambda_o) = 0.25$. In order to illustrate the insights discussed at the section 6.2.1, we exhibit on the figure 2 the two-dimensional representation of some eigenvectors from the transition matrix of a directed toroidal graph $\mathcal{T}_{54,36}$. Selected eigenvectors $\xi_{28,49}$ and $\xi_{28,50}$ have the same real part, that is $\Re(\lambda(\xi_{28,49})) = \Re(\lambda(\xi_{28,50}))$ but different non conjugated imaginary parts, that is $\Im(\lambda(\xi_{28,49})) \neq \pm \Im(\lambda(\xi_{28,50}))$. Analytically, the Rayleigh quotients of the selected eigenvectors are equal, as explained in section 6.2.1. As we can notice on

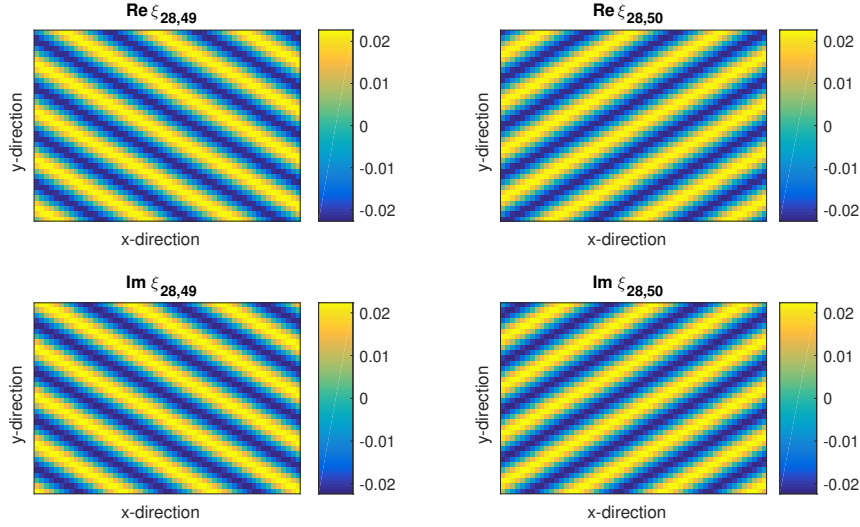


FIGURE 2. Two-dimensional representation of the eigenvectors $\xi_{28,49}$ and $\xi_{28,50}$ from the transition matrix of the directed toroidal $\mathcal{T}_{54,36}$. The axis is the label of a vertex in \mathbf{C}_m (resp. \mathbf{C}_m) in the x -direction (resp. y -direction).

figure 2, the selected eigenvectors have the orientation difference. This orientation difference is associated to the influence of the imaginary parts of the eigenvalues of the selected eigenvectors.

7. APPLICATIONS

This section is devoted to illustrating our theoretical framework through examples of semi-supervised learning and signal modeling on directed graphs. It turns out that in the following applications, we are able to work either in the Hilbert space $\ell^2(\mathcal{V})$ or $\ell^2(\mathcal{V}, \pi)$ where π is the unique stationary distribution of an ergodic random walk. Furthermore, by the results obtained in the Section 4.4, we define the isometry φ going from $\ell^2(\mathcal{V})$ to $\ell^2(\mathcal{V}, \pi)$. For the sake of simplicity, we will always map signals living in $\ell^2(\mathcal{V}, \pi)$ to $\ell^2(\mathcal{V})$, so that we can work in $\ell^2(\mathcal{V})$ where the scalar product coincides with the one between vectors in \mathbb{C}^N . For example, we have measured a graph signal \mathbf{f} in the real world. We decide in which space the signal lives. We therefore make an arbitrary choice between the fact that the signal lives in a space associated with the counting measure or in a space associated with the stationary measure π of an ergodic random walk. If we assume that the signal lives in $\ell^2(\mathcal{V}, \pi)$, we convert the signal into $\ell^2(\mathcal{V})$ using the application φ^{-1} defined as

$$(14) \quad \varphi^{-1} : \mathbf{f} \mapsto \mathbf{\Pi}^{1/2} \mathbf{f}, \quad \forall \mathbf{f} \in \ell^2(\mathcal{V}, \pi).$$

We perform all the necessary operations in $\ell^2(\mathcal{V})$. We then go back to the $\ell^2(\mathcal{V}, \pi)$ space in which the signal originally lived by using φ . In $\ell^2(\mathcal{V})$, the operator \mathbf{T} defined in eq. (9) corresponds to the random walk operator and \mathcal{L} is the operator such that $\langle \mathbf{f}, \mathcal{L} \mathbf{f} \rangle$ is the Dirichlet energy.

7.1. Semi-supervised learning on directed graph via ℓ_2 -regularization. We discuss the semi-supervised learning problem on directed graphs with a regularization term of type ℓ_2 . The following problem aims to show the efficiency and relevance of the Dirichlet energy defined at eq. (11) as a regularization term for signals defined on directed graphs.

7.1.1. Definition of the problem. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a strongly connected directed graph where $\mathcal{V} = (v_1, \dots, v_N)$ is the vertex set, \mathcal{E} is the edge set and with cardinality $|\mathcal{V}| = N$. The directed graph \mathcal{G} is represented by its adjacency matrix $\mathbf{W} \in \mathbb{R}_+^{N \times N}$. Let \mathcal{X} be the associated random walk on \mathcal{G} . The random walk \mathcal{X} is represented by its random walk operator $\mathbf{P} \in \mathbb{R}_+^{N \times N}$ with unique stationary distribution π .

We also introduce \mathcal{G}_{sym} the symmetrized version of the directed graph \mathcal{G} . The undirected graph \mathcal{G}_{sym} is represented by its adjacency matrix $\mathbf{W}_{\text{sym}} = (\mathbf{W} + \mathbf{W}^\top)/2 \in \mathbb{R}_+^{N \times N}$. Let \mathcal{X}_{sym} be the associated random walk on \mathcal{G}_{sym} . The random walk \mathcal{X}_{sym} is represented by its random walk operator $\mathbf{P}_{\text{sym}} \in \mathbb{R}_+^{N \times N}$ with unique stationary distribution π_{sym} .

Let $y : \mathcal{V} \rightarrow \{-1, 1\}$ be the function defined on the vertex set \mathcal{V} . From the function y , we obtain the graph signal \mathbf{y} defined as follows

$$\mathbf{y} = [y(v_1), \dots, y(v_N)]^\top.$$

Given the values $y(v_k)_{\{1 \leq k \leq N\}}$ on only a subset of labeled vertices $\mathcal{U} \subset \mathcal{V}$, the aim is to estimate the labels of the remaining unlabeled vertices. This is the semi-supervised learning problem on graphs [27, 39, 91]. We introduce the formulation of the problem in the next section.

7.1.2. Resolution of the problem. The problem of semi-supervised learning on graphs is formulated as follows

$$(15) \quad \underset{\mathbf{f}}{\operatorname{argmin}} \{c \|\mathbf{M}_l(\mathbf{f} - \mathbf{y})\|^2 + \varrho_1 \|\mathbf{M}_u \mathbf{f}\|^2 + \varrho_2 \mathcal{S}(\mathbf{f})\}, \quad c, \varrho_1, \varrho_2 > 0.$$

The term $\|\mathbf{M}_l(\mathbf{f} - \mathbf{y})\|^2$ is the labeled data fidelity term with $\mathbf{M}_l = \operatorname{diag}\{m_1, \dots, m_N\}$ a diagonal matrix where $m_i = \mathbb{1}_{v_i \in \mathcal{U}}$, for all $i = 1 \dots, N$ (hence 0 on vertices with unknown labels). In the second term, $\|\mathbf{M}_u \mathbf{f}\|^2$ appears \mathbf{M}_u defined as $\mathbf{M}_u = \mathbf{I}_N - \mathbf{M}_l$ and the third term, $\mathcal{S}(\mathbf{f})$ is the variational regularization term defined as

$$\mathcal{S}(\mathbf{f}) = \langle \mathbf{f}, \mathbf{X} \mathbf{f} \rangle, \quad \mathbf{X} \in \mathbb{R}^{N \times N}.$$

The semi-supervised learning on graphs problem is formulated in the same manner for the directed graph \mathcal{G} or its symmetrized version \mathcal{G}_{sym} , except that the variational regularization term $\mathcal{S}(\mathbf{f})$ differs. We propose to compare three methods for the directed case and its symmetrized one.

Semi-supervised learning on \mathcal{G} .

Method 1: For this method, we use a modification of (15), previously investigated by [27]. If the graph signal $\mathbf{y} \in \ell^2(\mathcal{V})$, the optimization problem is

$$(16) \quad \underset{\mathbf{f}}{\operatorname{argmin}} \{c \|\mathbf{M}_l(\mathbf{f} - \mathbf{y})\|^2 + c \|\mathbf{M}_u \mathbf{f}\|^2 + \varrho_2 \langle \mathbf{f}, \mathcal{L} \mathbf{f} \rangle\}, \quad c, \varrho_2 > 0.$$

with \mathcal{L} the directed normalized Laplacian defined at eq. (4). The term $\langle \mathbf{f}, \mathcal{L} \mathbf{f} \rangle$ is the Dirichlet energy in $\ell^2(\mathcal{V})$.

Method 2: For this method, we also use the formulation (16), except that the graph signal $\mathbf{y} \in \ell^2(\mathcal{V}, \pi)$. Hence, we define $\tilde{\mathbf{y}} \in \ell^2(\mathcal{V})$ as follows

$$\tilde{\mathbf{y}} = \varphi^{-1}(\mathbf{y}) = \mathbf{\Pi}^{1/2} \mathbf{y}, \quad \mathbf{y} \in \ell^2(\mathcal{V}, \pi).$$

The optimization problem is therefore

$$(17) \quad \underset{\tilde{\mathbf{f}}}{\operatorname{argmin}} \{c \|\mathbf{M}_l(\tilde{\mathbf{f}} - \tilde{\mathbf{y}})\|^2 + c \|\mathbf{M}_u \tilde{\mathbf{f}}\|^2 + \varrho_2 \langle \tilde{\mathbf{f}}, \mathcal{L} \tilde{\mathbf{f}} \rangle\}, \quad c, \varrho_2 > 0.$$

Equation (17) is the same as eq. (16) except for \mathbf{y} which is now $\tilde{\mathbf{y}}$. We can rewrite the optimization problem (17) as follows

$$(18) \quad \underset{\mathbf{f}}{\operatorname{argmin}} \{c \|\mathbf{M}_l(\mathbf{f} - \mathbf{y})\|_\pi^2 + c \|\mathbf{M}_u \mathbf{f}\|_\pi^2 + \varrho_2 \langle \mathbf{f}, \mathbf{L}_{\text{RW}} \mathbf{f} \rangle_\pi\}, \quad c, \varrho_2 > 0.$$

with \mathbf{L}_{RW} the random walk Laplacian on directed graphs defined at Eq. (5). The term $\langle \mathbf{f}, \mathbf{L}_{\text{RW}} \mathbf{f} \rangle_\pi$ is the Dirichlet energy defined at Eq. (11).

Method 3: In the framework of "Discrete Signal Processing on Graphs" [9, 26], the semi-supervised learning problem for directed graphs is formulated as follows

$$(19) \quad \underset{\mathbf{f}}{\operatorname{argmin}} \{c \|\mathbf{M}_l(\mathbf{f} - \mathbf{y})\|^2 + \varrho_2 \|\mathbf{f} - \mathbf{W}^{\text{norm}} \mathbf{f}\|_2^2\}, \quad c, \varrho_2 > 0,$$

with \mathbf{W}^{norm} , a normalized version of the the adjacency matrix \mathbf{W} so that its spectral norm is equal to one.

Semi-supervised learning on \mathcal{G}_{sym} . We use the same methods for the semi supervised learning on \mathcal{G} .

Method 1: If the graph signal $\mathbf{y} \in \ell^2(\mathcal{V})$, the optimization problem is

$$(20) \quad \underset{\mathbf{f}}{\operatorname{argmin}} \{c \|\mathbf{M}_l(\mathbf{f} - \mathbf{y})\|^2 + c \|\mathbf{M}_u \mathbf{f}\|^2 + \varrho_2 \langle \mathbf{f}, \mathcal{L}_{\text{sym}} \mathbf{f} \rangle\}, \quad c, \varrho_2 > 0.$$

with \mathcal{L}_{sym} the normalized Laplacian defined as

$$\mathcal{L}_{\text{sym}} = \mathbf{I} - \mathbf{\Pi}_{\text{sym}}^{1/2} \mathbf{P}_{\text{sym}} \mathbf{\Pi}_{\text{sym}}^{-1/2}, \quad \mathbf{\Pi}_{\text{sym}} = \operatorname{diag}\{\pi_{\text{sym}}(v_1), \dots, \pi_{\text{sym}}(v_N)\}.$$

Method 2: The optimization problem is

$$(21) \quad \underset{\tilde{\mathbf{f}}}{\operatorname{argmin}} \{c \|\mathbf{M}_l(\tilde{\mathbf{f}} - \tilde{\mathbf{y}})\|^2 + c \|\mathbf{M}_u \tilde{\mathbf{f}}\|^2 + \varrho_2 \langle \tilde{\mathbf{f}}, \mathcal{L}_{\text{sym}} \tilde{\mathbf{f}} \rangle\}, \quad c, \varrho_2 > 0.$$

We can rewrite the optimization problem (21) as follows

$$(22) \quad \underset{\mathbf{f}}{\operatorname{argmin}} \{c \|\mathbf{M}_l(\mathbf{f} - \mathbf{y})\|_\pi^2 + c \|\mathbf{M}_u \mathbf{f}\|_\pi^2 + \varrho_2 \langle \mathbf{f}, \mathbf{L}_{\text{RW}, \text{sym}} \mathbf{f} \rangle_\pi\}, \quad c, \varrho_2 > 0.$$

with $\mathbf{L}_{\text{RW}, \text{sym}}$ the random walk Laplacian defined as

$$\mathbf{L}_{\text{RW}, \text{sym}} = \mathbf{I} - \mathbf{P}_{\text{sym}}.$$

7.1.3. *Solution of the problem.* The problem (15) is quadratic and convex and therefore admits a closed form solution.

Semi-supervised learning on \mathcal{G} .

Method 1: The solution of the optimization problem (16) is

$$(23) \quad \mathbf{f}^* = \text{sign}[(\mathbf{I} + \gamma \mathcal{L})^{-1} \mathbf{M}_l \mathbf{y}], \quad \gamma = \varrho_2/c, \quad \mathbf{f}^* \in \ell^2(\mathcal{V}).$$

Method 2: The solution of the optimization problem (17) is

$$(24) \quad \mathbf{f}^* = \text{sign}[(\mathbf{I} + \gamma \mathbf{L}_{\text{RW}})^{-1} \mathbf{M}_l \mathbf{y}], \quad \gamma = \varrho_2/c, \quad \mathbf{f}^* \in \ell^2(\mathcal{V}, \pi).$$

Method 3: The solution of the optimization problem (19) is

$$\mathbf{f}^* = \text{sign}[(\mathbf{M}_l + \gamma \mathbf{R}_M^{-1}) \mathbf{y}], \quad \gamma = \varrho_2/c, \quad \mathbf{f}^* \in \ell^2(\mathcal{V}).$$

$$\text{with } \mathbf{R}_M = (\mathbf{I} - \mathbf{W}^{\text{norm}})^\top (\mathbf{I} - \mathbf{W}^{\text{norm}}).$$

Semi-supervised learning on \mathcal{G}_{sym} .

Method 1: The solution of the optimization problem (20) is

$$(25) \quad \mathbf{f}^* = \text{sign}[(\mathbf{I} + \gamma \mathcal{L}_{\text{sym}})^{-1} \mathbf{M}_l \mathbf{y}], \quad \gamma = \varrho_2/c, \quad \mathbf{f}^* \in \ell^2(\mathcal{V}).$$

Method 2: The solution of the optimization problem (21) is

$$(26) \quad \mathbf{f}^* = \text{sign}[(\mathbf{I} + \gamma \mathbf{L}_{\text{RW}, \text{sym}})^{-1} \mathbf{M}_l \mathbf{y}], \quad \gamma = \varrho_2/c, \quad \mathbf{f}^* \in \ell^2(\mathcal{V}, \pi_{\text{sym}}).$$

We summarize the different methods in Table 1.

Method	\mathcal{G}	\mathcal{G}_{sym}
Method 1	\mathcal{L}	\mathcal{L}_{sym}
Method 2	\mathbf{L}_{RW}	$\mathbf{L}_{\text{RW}, \text{sym}}$
Method 3	\mathbf{R}_M [9]	

TABLE 1. Table of the different operators associated with the semi-supervised learning methods according to \mathcal{G} and \mathcal{G}_{sym} .

7.1.4. Experiments. Let us consider the dataset of the political blogs of the 2004 US presidential campaign [28]. The dataset consists of 1224 political blogs where each political blog is associated with a political orientation, either republican or democrat. This dataset is modeled by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where each vertex $v \in \mathcal{V}$ is associated to a blog and an edge between two vertices $\{v_i, v_j\}$ indicates the presence of hyperlinks from the blog associated to the vertex v_i to the vertex v_j . The political orientations of blogs are modeled by a graph signal $\mathbf{f}_0 = \{f_0(v_1), \dots, f_0(v_N)\}$ where $f_0(v_i) \in \mathcal{O}$ where each vertex is associated with a label belongs to the set $\mathcal{O} = \{-1, 1\}$ characterizing its political orientation, the political orientation associated with Democrats being “+1” and the political orientation associated with Republicans being “-1”. The directed graph \mathcal{G} associated to the political blogs is not strongly connected. Our theoretical framework leads us to only consider strongly connected graphs. As a result, we consider the largest strongly connected subgraph of \mathcal{G} , denoted by $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ which is made up of $|\mathcal{V}'| = N' = 793$ vertices (hence roughly 65% of the vertex set \mathcal{V}) and its associated graph signal \mathbf{f}'_0 . The respective performances of the approaches eqs. (16), (18) and (19) and their symmetrized counterparts eqs. (20) and (22) are compared in fig. 3.

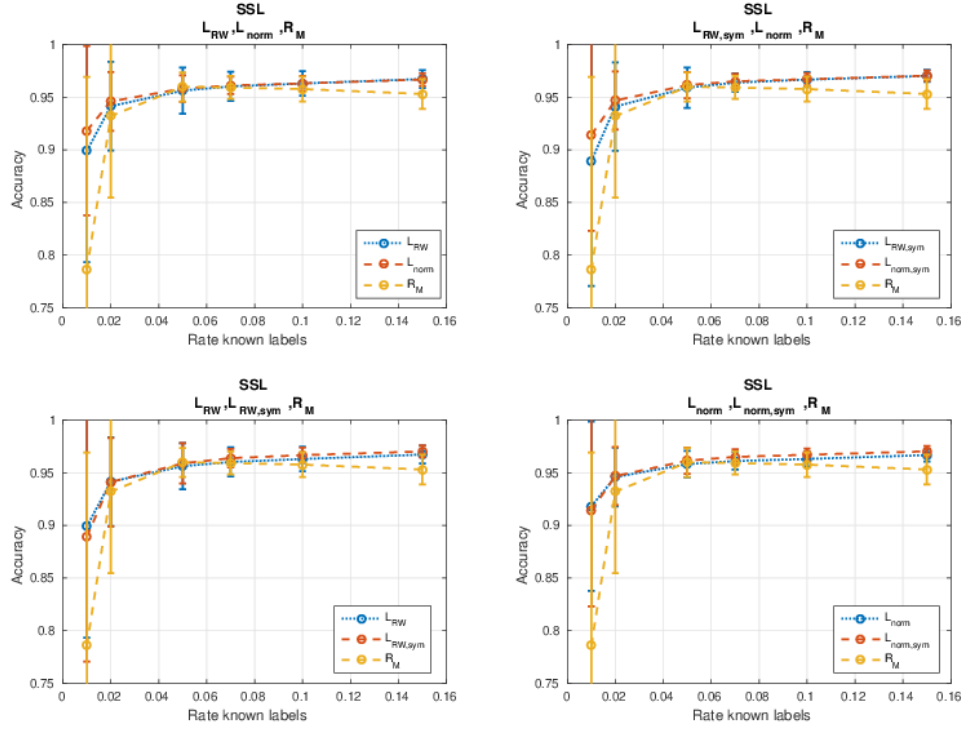


FIGURE 3. Semi-supervised learning on \mathcal{G}' of the political blogs of the 2004 US presidential campaign. The notations \mathbf{L}_{norm} and $\mathbf{L}_{\text{norm,sym}}$ correspond respectively to \mathcal{L} and \mathcal{L}_{sym} .

Numerical simulations. Figure 3 displays the performance of the semi-supervised approaches eqs. (16), (18) to (20) and (22). The performance rate is an average of 500 realizations calculated by determining the parameter $\gamma \in [0, 10]$ on the training set giving the best performance rate.

Firstly, we remark that for all proportions of known labels, the performance rates of semi supervised approaches based on \mathcal{L} , \mathbf{L}_{RW} and its symmetrized counterparts \mathcal{L}_{sym} , $\mathbf{L}_{\text{RW,sym}}$ are better than the performance rates of the semi-supervised approach based on \mathbf{R}_M . On the top left of fig. 3, we compare the semi-supervised approaches involving the operators \mathcal{L} , \mathbf{L}_{RW} and \mathbf{R}_M . The performance rate of the semi-supervised approach associated with \mathcal{L} is the best compared to the approaches involving \mathbf{L}_{RW} , \mathbf{R}_M for low rates of known labels. Furthermore, the performance rate approach associated with \mathcal{L} is slightly better than the approach involving \mathbf{L}_{RW} . On the top right of fig. 3, we compare the semi-supervised approaches involving the operators \mathcal{L}_{sym} , $\mathbf{L}_{\text{RW,sym}}$ and \mathbf{R}_M . Similarly the top left figure, the performance rate of the semi-supervised approach associated with \mathcal{L}_{sym} is slightly better than the approach involving \mathbf{L}_{RW} . On the bottom left of fig. 3, we compare the semi-supervised approaches involving the operators \mathbf{L}_{RW} , $\mathbf{L}_{\text{RW,sym}}$ and \mathbf{R}_M . The performance rates between the approach involving \mathbf{L}_{RW} and $\mathbf{L}_{\text{RW,sym}}$ are quite identical for any proportion of known labels as well as to the bottom right of fig. 3 comparing the semi-supervised approaches involving \mathbf{L}_{RW} , $\mathbf{L}_{\text{RW,sym}}$ and \mathbf{R}_M . The

comparison of performance rates of the semi-supervised approaches involving the different operators listed in the table 1 associated with this example leads us to the following conclusions :

- The semi supervised learning approaches involving Dirichlet energies based upon \mathcal{L} and \mathbf{L}_{RW} (and their symmetrized counterparts) yields better performance rates compared to the approach proposed by Sandryhaila and Moura [9] based upon \mathbf{R}_M . Consequently, semi-supervised learning approaches on directed graphs with Dirichlet energies based on the random walk operator seem more appropriate.
- The Hilbert space where we consider the graph signal seems to have an influence on the performance rates of the semi-supervised problem. Here, considering the graph signal \mathbf{y} belongs to $\ell^2(\mathcal{V})$, or equivalently considering the operator \mathcal{L} in the regularization term yields to better performance than considering the graph signal \mathbf{y} belonging to $\ell^2(\mathcal{V}, \pi)$ or equivalently considering the operator \mathcal{L}_{RW} in the regularization term.

7.2. Signal modeling on directed graphs via filtering. In this section, we consider a way to model the relationships between the values of a graph signal using a graph filter. The model is expressed as a graph filter that takes some values from the graph signal and reconstructs the other values. A possible application of such a model could be the lossy compression of the signal, where knowing only the graph, the coefficients of the graph filter and a few signal values enable the reconstruction of the whole signal. The major difference with the previous application is that we assume the whole graph signal known to learn the filter.

7.2.1. Problem formulation. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph with cardinality $|\mathcal{V}| = N$ and $\mu : \mathcal{V} \rightarrow \mathbb{R}_+$ a positive measure. The directed graph \mathcal{G} is characterized by its adjacency matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$. Given a random graph signal \mathbf{y} , we wish to determine the missing values of the signal by using a graph filter \mathbf{H} expressed as a polynomial finite sum of a reference operator \mathbf{R} , that is

$$(27) \quad \mathbf{H} = \sum_{k=0}^K \theta_k \mathbf{R}^k, \quad \theta_k \in \mathbb{R}, \quad k = 0, \dots, K.$$

The whole graph signal $f_0 \in \ell^2(\mathcal{V}, \mu)$ is defined as

$$\mathbf{f}_0 = [f_0(v_1), \dots, f_0(v_N)]^\top.$$

The random signal $\mathbf{y} = [y_1, \dots, y_N]^\top$ is a multivariate random variable indexed on the vertex set \mathcal{V} . The random variables $y_j, j = 1, \dots, N$ can be expressed as $y_j = \varepsilon_j f_0(v_j)$, where the ε_j are assumed independent and Bernoulli distributed with parameters δ_j : $\varepsilon_j \sim \text{Ber}(\delta_j)$. The random variable $\bar{\varepsilon}$ defined as

$$(28) \quad p = \frac{1}{N} \sum_{i=1}^N \varepsilon_i,$$

is the proportion of known values of \mathbf{f}_0 , all collected in \mathbf{y} . This signal \mathbf{y} can be viewed as a random sampling of the graph signal \mathbf{f}_0 . The optimization problem is the following

$$(29) \quad \underset{\boldsymbol{\theta} = \{\theta_k\}_{k=0}^K \in \mathbb{R}^{K+1}}{\text{argmin}} \quad \mathbb{E} \left[\left\| \mathbf{f}_0 - \sum_{k=0}^K \theta_k \mathbf{R}^k \mathbf{y} \right\|_\mu^2 \right],$$

7.2.2. *Solution of the problem.* The solution of the problem (29) is the following

$$(30) \quad \hat{\boldsymbol{\theta}} = \mathbf{Z}^{-1} \mathbf{M} \mathbf{Q}^\top \mathbf{f}_0,$$

where $\mathbf{M} = \text{diag}\{\mu(v_1), \dots, \mu(v_N)\}$ is the diagonal matrix of the associated measure μ , \mathbf{Z} the matrix where each entry $Z_{k\ell} \in \mathbf{Z}$ is expressed as

$$Z_{k\ell} = \text{Tr}([\mathbf{R}^k]^\top \mathbf{M} \mathbf{R}^\ell \mathbb{E}(\mathbf{y} \mathbf{y}^\top)), \quad \forall \{k, \ell\} \in \llbracket 0, K \rrbracket^2,$$

and $\mathbf{Q} = [\mathbf{q}_0, \dots, \mathbf{q}_K] \in \mathbb{R}^{N \times (K+1)}$ where each vector \mathbf{q}_j is

$$\mathbf{q}_j = \mathbf{R}^j \mathbb{E}(\mathbf{y}), \quad \forall j \in \llbracket 0, K \rrbracket.$$

Consequently, the resulting graph filter $\hat{\mathbf{H}}$ is

$$\hat{\mathbf{H}} = \sum_{k=0}^K \hat{\theta}_k \mathbf{R}^k, \quad \hat{\theta}_k \in \mathbb{R}, \quad k = 0, \dots, K.$$

7.2.3. *An exploration of modeling accuracy with different graph operators.* The reconstruction quality using the previous approach depends on several factors: the choice of reference operator \mathbf{R} , the order of the filter K and the random sampling strategy defined by the Bernoulli parameters δ_k . Here, we evaluate primarily the influence of \mathbf{R} and compare two possible sampling strategies. As an example, we also consider the dataset of the political blogs of the 2004 US presidential campaign [28], described in section 7.1.4.

We consider the directed graph \mathcal{G}' . The directed graph \mathcal{G}' is represented by its adjacency matrix \mathbf{W} . On \mathcal{G}' , we consider the following operators

- \mathbf{W}_{norm} the normalized version of the adjacency matrix \mathbf{W} whose norm is equal to one.
- The random walk operator \mathbf{P} associated with an ergodic random walk \mathcal{X} with unique stationary distribution π .
- $\bar{\mathbf{P}}$ the additive reversibilization of \mathbf{P} .
- $\bar{\mathbf{T}} = \mathbf{\Pi}^{1/2} \bar{\mathbf{P}} \mathbf{\Pi}^{-1/2}$ the operator similar to $\bar{\mathbf{P}}$.
- $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$ with $\bar{\mathcal{P}} = \{\bar{\mathbf{P}}_\alpha : \bar{\mathbf{P}}_\alpha = (1 - \alpha)\mathbf{P} + \alpha\mathbf{P}^* | \alpha \in [0, 1]\}$
- $\bar{\mathbf{T}}_\alpha \in \bar{\mathcal{T}}$ with $\bar{\mathcal{T}} = \{\bar{\mathbf{T}}_\alpha : \bar{\mathbf{T}}_\alpha = (1 - \alpha)\mathbf{\Pi}^{1/2} \mathbf{P} \mathbf{\Pi}^{-1/2} + \alpha\mathbf{\Pi}^{1/2} \mathbf{P}^* \mathbf{\Pi}^{-1/2} | \alpha \in [0, 1]\}$

We also consider \mathcal{G}'_{sym} , the symmetrized version of \mathcal{G}' . The undirected graph \mathcal{G}'_{sym} is represented by its adjacency matrix $\mathbf{W}_{sym} = (\mathbf{W} + \mathbf{W}^\top)/2$. On \mathcal{G}'_{sym} , we define the following operators

- \mathbf{W}_{sym}^{norm} the normalized version of the adjacency matrix \mathbf{W}_{sym} whose the norm is equal to one.
- The random walk operator \mathbf{P}_{sym} corresponding to the ergodic random walk \mathcal{X}_{sym} with stationary distribution π_{sym} .
- $\mathbf{T}_{sym} = \mathbf{\Pi}_{sym}^{1/2} \mathbf{P}_{sym} \mathbf{\Pi}_{sym}^{-1/2}$ the operator similar to \mathbf{P}_{sym} .

Finally, we consider \mathcal{G} , the full graph. The directed graph \mathcal{G} is represented by its adjacency matrix \mathbf{W} . The graph \mathcal{G} is not strongly connected. Consequently, we can not build directly a random walk operator on \mathcal{G} . We need to deal with its property of not being strongly connected, as required by our framework.

Indeed, some vertices in \mathcal{G} have an out-degree equal to zero and we cannot create a transition matrix directly from the adjacency matrix. To overcome this problem, we propose two approaches that make the graph strongly connected and the associated random walk ergodic:

Approach 1: Rank-one perturbation: from the original adjacency matrix \mathbf{W} , we construct a new adjacency matrix \mathbf{W}_ϵ as follows

$$\mathbf{W}_\epsilon = \mathbf{W} + \epsilon \mathbf{J},$$

where $\mathbf{J} = \mathbf{1}\mathbf{1}^\top/N$ is a rank-one matrix and ϵ is small. The weak perturbation of the adjacency matrix \mathbf{W} by the matrix \mathbf{J} ensures that the random walk on \mathcal{G} is ergodic with stationary measure π_ϵ and its associated transition matrix \mathbf{P}_ϵ is well-defined. For our experiments, we choose $\epsilon = 10^{-4}$.

Approach 2: Construction of the Google matrix of \mathcal{G} [29]. This is achieved in two steps. Firstly, we construct an adjacency matrix $\tilde{\mathbf{W}}$ from \mathbf{W} by adding a weight one from all dangling nodes, that is nodes with no out-edges towards all the nodes in the graph. From $\tilde{\mathbf{W}}$ we can construct the transition matrix \mathbf{S} . Secondly, we define the Google matrix \mathbf{P}_G as

$$\mathbf{P}_G = (1 - \gamma)\mathbf{S} + \gamma\mathbf{J}.$$

where $\gamma = 0.85$ [29].

We will compare the reconstruction accuracy using a proportion of correctly reconstructed labels obtained via

$$\hat{\mathbf{f}} = \text{sign}(\hat{\mathbf{H}}\mathbf{y}).$$

Subproblem 1. We solve the problem (29) by learning a polynomial graph filter with $K = 10$ on \mathcal{G}' and $\mathcal{G}'_{\text{sym}}$. We consider as reference operators the ones listed on \mathcal{G}' and $\mathcal{G}'_{\text{sym}}$. In this subproblem, the random variables $y_j, j = 1, \dots, N'$ are distributed according to one of the two following cases

- Random variables $y_j = \varepsilon_j f'_0(v_j)$ where $\varepsilon_j \sim \text{Ber}(p)$ with p the proportion of known labels.
- Random variables $y_j = \varepsilon_j f'_0(v_j)$ where $\varepsilon_j \sim \text{Ber}(\alpha\pi_j)$, such that $\sum_j \mathbb{E}(\varepsilon_j) = pN'$ with p the proportion of known labels.

The proportions of correctly reconstructed labels are measured for various p values using 500 realizations of \mathbf{y} to estimate its mean and covariance used in the solution (30).

We summarize the different cases in table 2.

Case	Distribution \mathbf{y}	Reference operators
Case 1	$y_j = \varepsilon_j f'_0(v_j), \varepsilon_j \sim \text{Ber}(p)$	$\mathbf{W}^{norm}, \mathbf{P}, \bar{\mathbf{P}}, \mathbf{T}, \bar{\mathbf{T}}, \mathbf{W}_{sym}^{norm}, \mathbf{P}_{sym}, \mathbf{T}_{sym}$
Case 2	$y_j = \varepsilon_j f'_0(v_j), \varepsilon_j \sim \text{Ber}(\alpha\pi_j), \sum_j \mathbb{E}(\varepsilon_j) = pN'$	$\mathbf{W}^{norm}, \mathbf{P}, \bar{\mathbf{P}}, \mathbf{T}, \bar{\mathbf{T}}, \mathbf{W}_{sym}^{norm}, \mathbf{P}_{sym}, \mathbf{T}_{sym}$

TABLE 2. Summary table of the different cases of the subproblem 1.

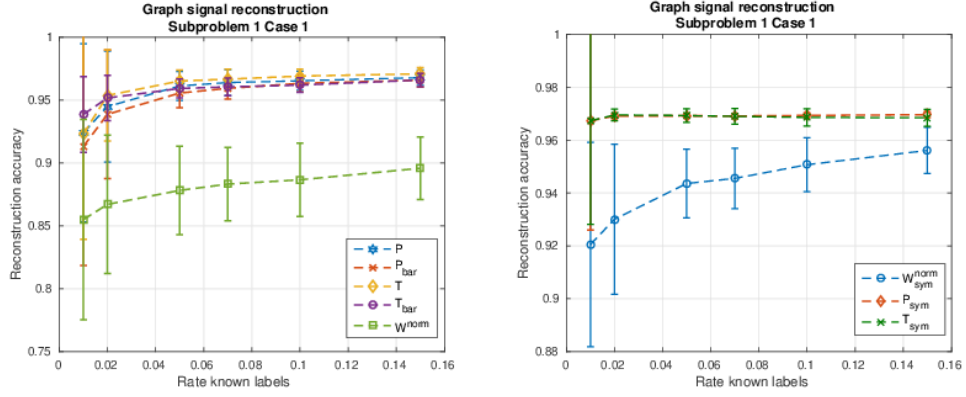


FIGURE 4. (Subproblem 1, Case 1) Left: Reconstruction of the graph signal \mathbf{f}'_0 on \mathcal{G}' . Right: Reconstruction of the graph signal \mathbf{f}'_0 on \mathcal{G}'_{sym} . The notation \mathbf{P}_{bar} (resp. \mathbf{T}_{bar}) correspond to $\bar{\mathbf{P}}$ (resp. $\bar{\mathbf{T}}$).

Numerical simulations: Case 1. We evaluate the reconstruction performance of the graph signal \mathbf{f}'_0 in figure 4. For all proportions p of known labels, the average rate of correctly reconstructed labels in the graph signal \mathbf{f}'_0 from a learned polynomial graph filter of \mathbf{P} or $\bar{\mathbf{P}}$ is significantly better than the one based on a learned polynomial graph filter of \mathbf{W}_{norm} in the case where we consider either the subgraph \mathcal{G}' or its symmetrized version \mathcal{G}'_{sym} . Furthermore, the reconstruction performance using a filter based on \mathbf{P} is slightly better than using a filter based on $\bar{\mathbf{P}}$. The reconstruction performance using a filter based on \mathbf{T} is the best. For small proportions p of known labels, the average rate of correctly reconstructed labels from a learned polynomial graph filter of $\bar{\mathbf{T}}$ is slightly better. We also notice that the reconstruction performance using a filter based on \mathbf{T}_{sym} is identical to using a filter based on \mathbf{P}_{sym} for all proportions p of known labels.

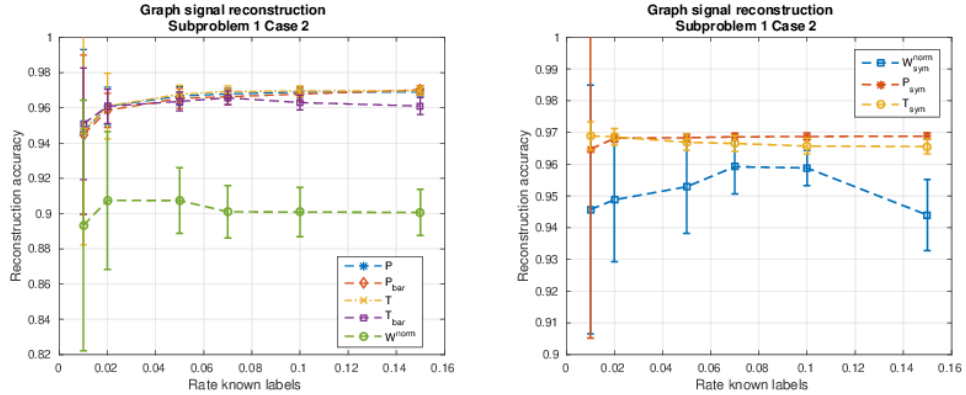


FIGURE 5. (Subproblem 1, Case 2) Left: Reconstruction of the graph signal \mathbf{f}'_0 on the subgraph \mathcal{G}' . Right: Reconstruction of the graph signal \mathbf{f}'_0 on \mathcal{G}'_{sym} . The notation \mathbf{P}_{bar} (resp. \mathbf{T}_{bar}) correspond to $\bar{\mathbf{P}}$ (resp. $\bar{\mathbf{T}}$).

Numerical simulations: Case 2. The reconstruction performance is shown in figure 5. We observe a significantly better performance using a filter based on \mathbf{P} or $\bar{\mathbf{P}}$ than based on \mathbf{W}_{norm} in the case where we consider either the subgraph \mathcal{G}' or its symmetrized version \mathcal{G}'_{sym} and a slightly better performance using \mathbf{P} rather than $\bar{\mathbf{P}}$. For all proportions p of known labels, the average rates of correctly reconstructed labels are identical using either a filter based on \mathbf{T} or $\bar{\mathbf{T}}$ are better than using a filter based on \mathbf{P} or $\bar{\mathbf{P}}$. We notice the reconstruction performance based on \mathbf{W}_{norm} has a large variability and even decreases for larger p .

Case 1 vs. Case 2. The reconstruction performance using a filter based on \mathbf{P} or $\bar{\mathbf{P}}$ is slightly better when the samples in \mathbf{y} are selected according to a distribution proportional to the stationary distribution π (case 2) than when the distribution is uniform (case 1). It is also the case by using a filter based on \mathbf{T} or $\bar{\mathbf{T}}$. This is expected as vertices with larger π_k values correspond to better-connected blogs, which are likely to have an influence on a larger number of other blogs. Although the reconstruction performance using a filter based on \mathbf{W}_{norm} is better in case 2 than in case 1, the large variability and the poorer performance at large p in case 2 suggest that these filters provide generally poorer models of the signal. The better reconstruction performance using a filter based on \mathbf{T} or $\bar{\mathbf{T}}$ with respect to using a filter based on \mathbf{P} or $\bar{\mathbf{P}}$ suggest that the graph signal seems more suitable to live in $\ell^2(\mathcal{V})$ than $\ell^2(\mathcal{V}, \pi)$.

Subproblem 2. Here we consider the same problem (29) as in the previous section but now compare the performance using reference operators $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$ and their equivalents $\bar{\mathbf{T}}_\alpha = \mathbf{\Pi}^{1/2} \bar{\mathbf{P}}_\alpha \mathbf{\Pi}^{-1/2}$. The random variables $y_j, j = 1, \dots, N'$ are distributed according to one of the two following cases:

- Random variables $y_j = \varepsilon_j f'_0(v_j)$ where $\varepsilon_j \sim \text{Ber}(p)$ with p the proportion of known labels.
- Random variables $y_j = \varepsilon_j f'_0(v_j)$ where $\varepsilon_j \sim \text{Ber}(\alpha\pi_j)$, such that $\sum_j \mathbb{E}(\varepsilon_j) = pN'$ with p the proportion of known labels.

We summarize the different cases in table 3.

Case	Distribution \mathbf{y}	Reference operators
Case 1	$y_j = \varepsilon_j f'_0(v_j), \varepsilon_j \sim \text{Ber}(p)$	$\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}, \bar{\mathbf{T}}_\alpha \in \bar{\mathcal{T}}$
Case 2	$y_j = \varepsilon_j f'_0(v_j), \varepsilon_j \sim \text{Ber}(\alpha\pi_j), \sum_j \mathbb{E}(\varepsilon_j) = pN'$	$\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}, \bar{\mathbf{T}}_\alpha \in \bar{\mathcal{T}}$

TABLE 3. Summary table of the different cases of the subproblem 2.

Numerical simulations: Case 1. We evaluate the reconstruction performance of f'_0 using filters either based on $\bar{\mathbf{P}}_\alpha, \alpha \in [0, 1]$ at the top of fig. 6 or either based on $\bar{\mathbf{T}}_\alpha, \alpha \in [0, 1]$ at the bottom of fig. 6. We first consider the top of fig. 6. For small proportions of p of known labels, we notice that the reconstruction performance using a filter based on $\bar{\mathbf{P}}_\alpha, \alpha \simeq 0.2$ is the best while for higher proportions p , the

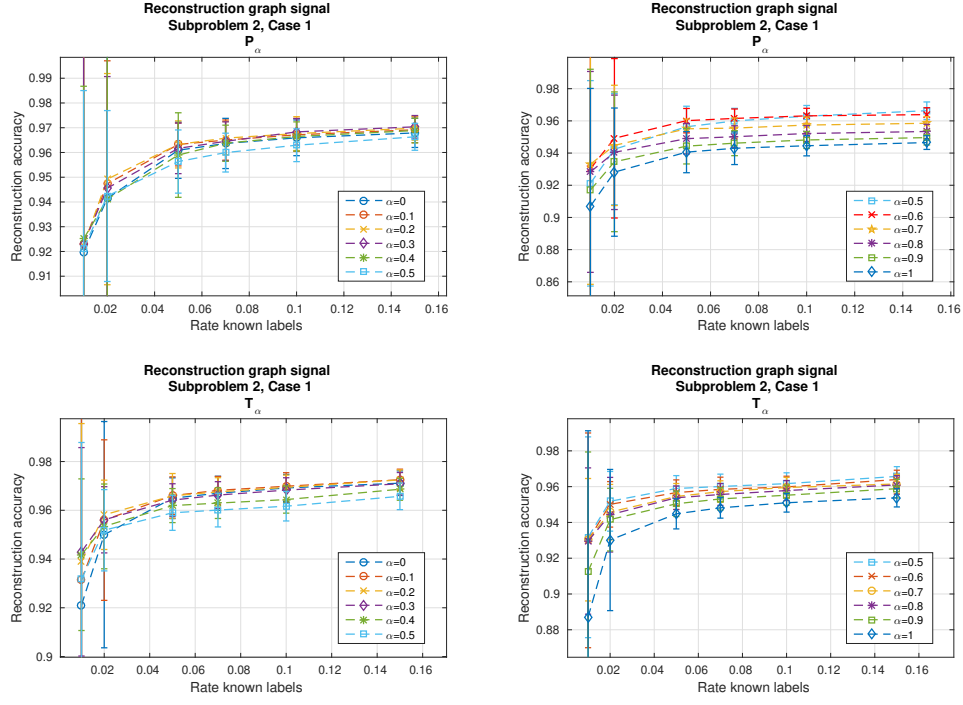


FIGURE 6. (Subproblem 2, Case 1) Up: Reconstruction of the graph signal \mathbf{f}'_0 on \mathcal{G}' , $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$. Down: Reconstruction of the graph signal \mathbf{f}'_0 on \mathcal{G}' , $\bar{\mathbf{T}}_\alpha \in \bar{\mathcal{T}}$.

reconstruction performance using a filter based on $\bar{\mathbf{P}}_\alpha$, $\alpha \simeq 0.3$ is the best. We notice poorer performance overall when α increases from $\alpha = 0.5$ to $\alpha = 1$. At the bottom of fig. 6, the reconstruction performances using a filter based on $\bar{\mathbf{T}}_\alpha$, $\alpha \simeq 0.2$ are the best except for the rate $p = 0.01$. We notice poorer performance overall when α increases from $\alpha = 0.5$ to $\alpha = 1$.

For all proportions p of known labels, the reconstruction performance are slightly better overall if we consider filters based on $\bar{\mathbf{T}}_\alpha$, $\alpha \in [0, 1]$ than filters based on $\bar{\mathbf{P}}_\alpha$, $\alpha \in [0, 1]$. That means the problem is more adapted when the graph signal \mathbf{f}'_0 is in $\ell^2(\mathcal{V})$ than \mathbf{f}'_0 is in $\ell^2(\mathcal{V}, \pi)$. Furthermore, the study of this case suggests that, for this signal, graph filters $\bar{\mathbf{P}}_\alpha$ perform better when the convex combination of \mathbf{P} and \mathbf{P}^* involves more the random walk \mathbf{P} , i.e. $\alpha \in (0, 0.5)$, than the random \mathbf{P}^* , i.e. $\alpha \in (0.5, 1)$.

Numerical simulations: Case 2. We evaluate the reconstruction performance of \mathbf{f}'_0 using filters either based on $\bar{\mathbf{P}}_\alpha$, $\alpha \in [0, 1]$ at the top of fig. 7 or either based on $\bar{\mathbf{T}}_\alpha$, $\alpha \in [0, 1]$ at the bottom of fig. 7. Let us consider the top of fig. 7. We notice that the reconstruction performances using a filter based on $\bar{\mathbf{P}}_\alpha$, $\alpha \simeq 0.3$ are the best except for the rates $p = \{0.01, 0.02\}$. We notice poorer performance overall when α increases from $\alpha = 0.5$ to $\alpha = 1$. This suggests that, for this signal, graph filters based on $\bar{\mathbf{P}}_\alpha$ perform better when the convex combination of \mathbf{P} and \mathbf{P}^* involves more the forward random walk \mathbf{P} , i.e. $\alpha \in (0, 0.5)$, than the backward

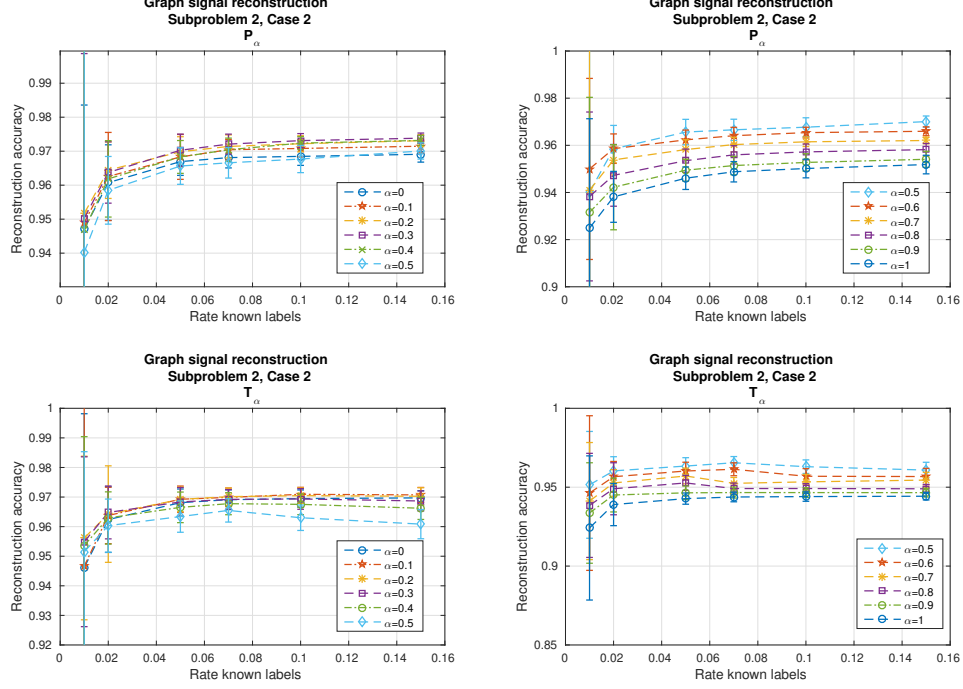


FIGURE 7. (Subproblem 2, case 2) Up: Reconstruction of the graph signal \mathbf{f}'_0 on \mathcal{G}' , $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$. Down: Reconstruction of the graph signal \mathbf{f}'_0 on \mathcal{G}' , $\bar{\mathbf{T}}_\alpha \in \bar{\mathcal{T}}$.

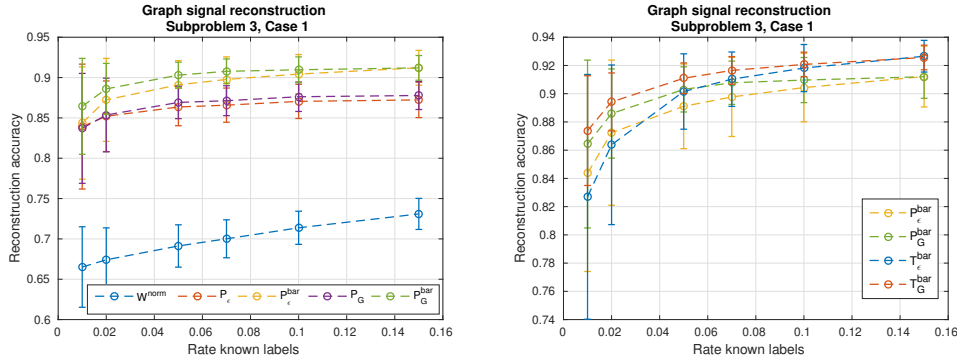
random walk \mathbf{P}^* , i.e. $\alpha \in (0.5, 1)$. At the bottom of fig. 7, the reconstruction performances using a filter based on $\bar{\mathbf{T}}_\alpha$, $\alpha \simeq 0.3$ are the best for all proportions p of known labels. Similarly to the previous cases, we notice poorer performance overall when α increases from $\alpha = 0.5$ to $\alpha = 1$.

Firstly, the reconstruction performance is better using filters based on $\bar{\mathbf{P}}_\alpha$ than $\bar{\mathbf{T}}_\alpha$. That means the problem is more suitable for graph signal $\mathbf{f}'_0 \in \ell^2(\mathcal{V}, \pi)$ is slightly better than performance with respect to the graph signal belonging to $\ell^2(\mathcal{V})$ when the samples in \mathbf{y} are selected from a distribution proportional to the stationary distribution π than when the distribution is uniform. Conversely, the problem is more suitable for graph signal \mathbf{f}'_0 is $\ell^2(\mathcal{V})$ than $\ell^2(\mathcal{V}, \pi)$ when the distribution in \mathbf{y} is uniform while the samples in \mathbf{y} are selected according to a distribution proportional to the stationary distribution π . As a result, the distribution of the graph signal has an influence in the reconstruction performance through the use of learned filters based either on $\bar{\mathbf{P}}_\alpha$ or $\bar{\mathbf{T}}_\alpha$.

Subproblem 3. We now consider the same problem (29) on the whole graph \mathcal{G} . In this subproblem, the random variables $y_j, j = 1 \dots, N$ are distributed as $y_j = \varepsilon_j f_0(v_j), \varepsilon_j \sim \text{Ber}(p)$. For each proportion of known labels p , the good reconstruction rates are derived by averaging 500 graph signals realizations of \mathbf{y} . We summarize the following case in the table 4.

Case	Distribution \mathbf{y}	Reference operators
Case 1	$y_j = \varepsilon_j f'_0(v_j), \varepsilon_j \sim \text{Ber}(p)$	$\mathbf{W}^{norm}, \mathbf{P}_\epsilon, \mathbf{P}_G, \bar{\mathbf{P}}_\epsilon, \bar{\mathbf{P}}_G, \mathbf{T}_\epsilon, \mathbf{T}_G, \bar{\mathbf{T}}_\epsilon, \bar{\mathbf{T}}_G$

TABLE 4. Summary table of the case of the subproblem 3.

FIGURE 8. (Subproblem 3, Case 1) Reconstruction of the graph signal \mathbf{f}_0 on \mathcal{G} . The notation $\mathbf{P}_\epsilon^{\text{bar}}$ (resp. $\mathbf{P}_G^{\text{bar}}, \mathbf{T}_G^{\text{bar}}, \mathbf{T}_\epsilon^{\text{bar}}$) corresponds to $\bar{\mathbf{P}}_\epsilon$ (resp. $\bar{\mathbf{P}}_G, \bar{\mathbf{T}}_G, \bar{\mathbf{T}}_\epsilon$).

Numerical simulations. The reconstruction performance is shown in fig. 8. For all p values, the reconstruction performance using filters based on \mathbf{W}_{norm} is significantly worse than with all other reference operators. Compared to the other reference operators, \mathbf{W}_{norm} is the only one that can be used without modifying the graph to make it strongly connected. Still, its performance is always worse.

Among the other reference operators, we notice a clearly better performance generally when using the reversibilizations $\bar{\mathbf{P}}_\epsilon$ and $\bar{\mathbf{P}}_G$ compared to the non-reversible random walks \mathbf{P}_ϵ and \mathbf{P}_G . Furthermore, the best reconstruction performance is obtained using a filter based on $\bar{\mathbf{T}}_G$ and the reconstruction performance using a filter based on $\bar{\mathbf{T}}_\epsilon$ is identical to $\bar{\mathbf{T}}_G$ for higher rate of known labels. This differs from the results on \mathcal{G}' where both would perform similarly.

Finally, the reconstruction performance does not depend as much on the approach we use to make the graph strongly connected. It seems though that the Google approach slightly outperforms the rank-one approach, both for the non-reversible random walks and their reversibilizations.

Subproblem 4. Here we consider the same problem (29) as in the previous section but now compare the performance using reference operators $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$ and their respective equivalents $\bar{\mathbf{T}}_\alpha = \mathbf{\Pi}^{1/2} \bar{\mathbf{P}}_\alpha \mathbf{\Pi}^{-1/2}$. The random walk operator is built from the Google matrix approach (Approach 1). The random variables $y_j, j = 1 \dots, N$ are distributed as $y_j = \varepsilon_j f_0(v_j), \varepsilon_j \sim \text{Ber}(p)$. We summarize the case in the table 5.

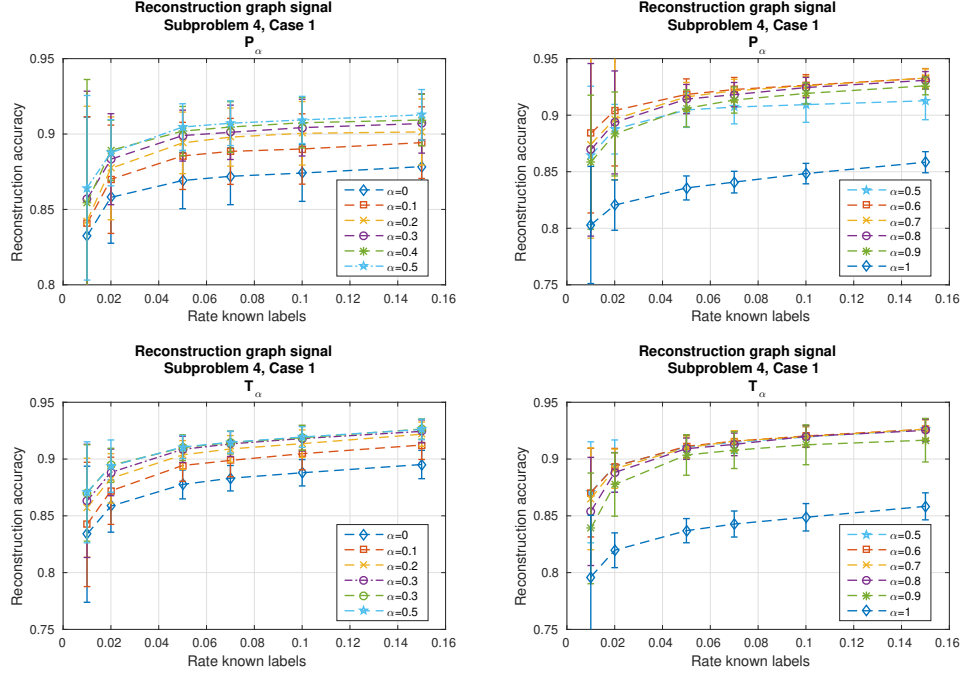


FIGURE 9. (Subproblem 4, case 1) Up: Reconstruction of the graph signal \mathbf{f}_0 on \mathcal{G} , $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$. Down: Reconstruction of the graph signal \mathbf{f}_0 on \mathcal{G} , $\bar{\mathbf{T}}_\alpha \in \bar{\mathcal{T}}$.

Case	Distribution \mathbf{y}	Reference operators
Case 1	$y_j = \varepsilon_j f_0(v_j), \varepsilon_j \sim \text{Ber}(p)$	$\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}, \bar{\mathbf{T}}_\alpha \in \bar{\mathcal{T}}$

TABLE 5. Table of the case of the subproblem 4.

Numerical simulations: Case 1. We evaluate the reconstruction performance of \mathbf{f}'_0 using filters either based on $\bar{\mathbf{P}}_\alpha, \alpha \in [0, 1]$ at the top of the fig. 9 or either based on $\bar{\mathbf{T}}_\alpha, \alpha \in [0, 1]$ at the bottom of the fig. 9. For all proportions p of known labels, the reconstruction performances using a filter based on $\bar{\mathbf{P}}_\alpha, \alpha \simeq 0.6$ are the best. We also notice that the reconstruction performance using a filter with $\alpha \in \{0.7, 0.8\}$ are better than $\alpha \simeq 0.5$. This suggests that, for this signal, graph filters based on $\bar{\mathbf{P}}_\alpha$ perform better when the convex combination of \mathbf{P} and \mathbf{P}^* involves more the random walk \mathbf{P}^* , i.e. $\alpha \in [0.6, 0.8]$, than the random walk \mathbf{P} , i.e. $\alpha \in [0.1, 0.5]$. At the bottom of fig. 9, the reconstruction performances using a filter based on $\bar{\mathbf{T}}_\alpha, \alpha \simeq 0.5$ are the best, for all proportions p of known labels. We notice poorer performance overall when α increases from $\alpha = 0.5$ to $\alpha = 1$. For all proportions p of known labels, the reconstruction performance are better overall if the graph signal \mathbf{f}_0 is in $\ell^2(\mathcal{V})$ than \mathbf{f}_0 is in $\ell^2(\mathcal{V}, \pi)$. In other words, the reconstruction performance using filters based on $\bar{\mathbf{T}}_\alpha$ is better than using filters based on $\bar{\mathbf{P}}_\alpha$.

To summarize :

- The Hilbert space where we consider the graph signal or the distribution of the sampled graph signal has an influence on the performance rates of the reconstruction of the graph signal. In the example discussed in the current section, we obtain better performances when we consider the Hilbert space of the graph signals associated to the measure corresponding to the sampling probability.
- The construction of graph filters based on the random walk operator yields better performance rates than filters based on the adjacency matrix.
- The choice of the parameter α in the convex combination between \mathbf{P} and \mathbf{P}^* has an influence on the performance rates. In the example discussed in the current section, the optimal convex combination \mathbf{P} and \mathbf{P}^* (resp. \mathbf{T} and \mathbf{T}^*) is rather forward than backward with one exception.

8. MULTI-RESOLUTION ANALYSES ON DIRECTED GRAPHS

In the previous sections, we proposed a Fourier like basis on directed graphs, as the set of random walk's eigenvectors and determine a frequency analysis by studying the smoothness of the random walk's eigenvectors. We are thus able to construct multiresolution analyses of functions over directed graphs. In a first instance, we propose wavelet frames made of analysis and synthesis graph filter banks. This multi-scale construction is closely related to the diffusion polynomial frames construction [19] and spectral graph wavelets [20]. In a second instance, we propose a critically sampled wavelet construction on directed graphs generalizing the diffusion wavelets framework [21]. In our theoretical constructions, we propose to work in the $\ell^2(\mathcal{V})$ space. We use the linear transformation φ defined in eq. (8) to go into the $\ell^2(\mathcal{V}, \pi)$ space.

8.1. Redundant wavelet transform on directed graphs. In this section, we propose a redundant wavelet transform on directed graphs. We follow the construction of spectral wavelets on undirected graphs [20] and polynomial diffusion frames [19]. The novelty is the construction of filters designed in the frequency domain via a frequency response function as a linear combination of projectors onto the associated mono-frequency random walk subspaces. We introduce the necessary elements for the construction of redundant wavelets on directed graphs.

8.1.1. Theoretical framework. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a strongly connected directed graph with cardinality $|\mathcal{V}| = N$. The directed graph \mathcal{G} is characterized by its adjacency matrix $\mathbf{W} \in \mathbb{R}_+^{N \times N}$. On \mathcal{G} , we defined a random walk \mathcal{X} characterized by its transition matrix $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W} \in \mathbb{R}_+^{N \times N}$. The \mathcal{X} is ergodic with stationary distribution π . We introduce $\mathbf{T} = \mathbf{\Pi}^{1/2}\mathbf{P}\mathbf{\Pi}^{-1/2} \in \ell^2(\mathcal{V})$, the operator similar to \mathbf{P} . We assume \mathbf{T} diagonalizable. We define a low-pass operator as follows.

Definition 8.1. A low pass operator at dilation t , $\mathbf{H}_t^{\mathcal{G}}$ on a strongly connected directed graph \mathcal{G} is defined by

$$\mathbf{H}_t^{\mathcal{G}} = \sum_{\ell=1}^p h(t\omega_{\ell})\mathbf{S}_{\ell}, \quad t \in \mathbb{N}.$$

with $h : [0, 2] \rightarrow \mathbb{R}$, a function giving a low-pass frequency response and $\{\mathbf{S}_{\ell}\}_{\ell=1}^p$, the set of mono-frequency random walk projectors associated to the mono-frequency

random walk subspaces $\{\mathbf{S}_\ell\}_{\ell=1}^p$ of \mathbf{T} . The scaling function at dilation t and translation y is denoted by

$$\mathbf{h}_{t,y}^\mathcal{G} = \mathbf{H}_t^\mathcal{G} \boldsymbol{\delta}_y$$

where $\boldsymbol{\delta}_y$ is the Kronecker delta function at the vertex $y \in \mathcal{V}$.

We define a bandpass operator as follows.

Definition 8.2. A band pass operator at dilation t , $\mathbf{G}_t^\mathcal{G}$ on a strongly connected directed graph \mathcal{G} is defined as

$$\mathbf{G}_t^\mathcal{G} = \sum_{\ell=1}^p g(t\omega_\ell) \mathbf{S}_\ell, \quad t \in \mathbb{N}.$$

with $g : [0, 2] \mapsto \mathbb{R}$, a function giving a band-pass frequency response and $\{\mathbf{S}_\ell\}_{\ell=1}^p$, the set of mono-frequency random walk projectors associated to the mono-frequency random walk subspaces $\{\mathbf{S}_\ell\}_{\ell=1}^p$ of \mathbf{T} . The wavelet function at dilation t and translation y is denoted by

$$\mathbf{g}_{t,y}^\mathcal{G} = \mathbf{G}_t^\mathcal{G} \boldsymbol{\delta}_y,$$

where $\boldsymbol{\delta}_y$ is the Kronecker delta function at the vertex $y \in \mathcal{V}$.

Having defined the low-pass and band-pass operators on directed graphs, we are able to build analysis and synthesis filter banks on directed graphs.

Definition 8.3. We define a bank of synthesis filters \mathcal{K} as the set of a low-pass filter at dilation t_J , $\mathbf{H}_{t_J}^\mathcal{G}$ and a series of band-pass filters at increasing dilations $\{t_j\}_{j=1}^J$, $\{\mathbf{G}_{t_j}\}_{j=1}^J$:

$$\mathcal{K} = \{\mathbf{H}_{t_J}^\mathcal{G}, \mathbf{G}_{t_1}^\mathcal{G}, \dots, \mathbf{G}_{t_J}^\mathcal{G}\}.$$

Definition 8.4. We define a bank of analysis filters $\tilde{\mathcal{K}}$ as the set of a filter at dilation t_J , $\tilde{\mathbf{H}}_{t_J}^\mathcal{G}$ and a series of filters at increasing dilations $\{t_j\}_{j=1}^J$, $\{\tilde{\mathbf{G}}_{t_j}\}_{j=1}^J$:

$$\tilde{\mathcal{K}} = \{\tilde{\mathbf{H}}_{t_J}^\mathcal{G}, \tilde{\mathbf{G}}_{t_1}^\mathcal{G}, \dots, \tilde{\mathbf{G}}_{t_J}^\mathcal{G}\},$$

where

$$\tilde{\mathbf{H}}_{t_J}^\mathcal{G} = \sum_{\ell=1}^p \tilde{h}(t\omega_\ell) \mathbf{S}_\ell \in \mathbb{R}^{N \times N},$$

and

$$\tilde{\mathbf{G}}_{t_j}^\mathcal{G} = \sum_{\ell=1}^p \tilde{g}(t\omega_\ell) \mathbf{S}_\ell \in \mathbb{R}^{N \times N}, \quad \forall j = 1, \dots, J.$$

We also define $\tilde{\mathbf{h}}_{t,k}$ and $\tilde{\mathbf{g}}_{t,k}$ as row vectors as follows:

$$\tilde{\mathbf{h}}_{t,k} = \boldsymbol{\delta}_k^\top \tilde{\mathbf{H}}_t^\mathcal{G}, \quad \tilde{\mathbf{g}}_{t,k} = \boldsymbol{\delta}_k^\top \tilde{\mathbf{G}}_t^\mathcal{G}.$$

Proposition 8.1. Given a fixed set of increasing dilations $\{t_j\}_{j=1}^J$, the synthesis and analysis filters sets respectively \mathcal{K} and $\tilde{\mathcal{K}}$, the perfect reconstruction condition

$$(31) \quad \mathbf{H}_{t_J}^\mathcal{G} \tilde{\mathbf{H}}_{t_J}^\mathcal{G} + \sum_{j=1}^J \mathbf{G}_{t_j}^\mathcal{G} \tilde{\mathbf{G}}_{t_j}^\mathcal{G} = \mathbf{I}$$

is guaranteed if and only if

$$h(t_J\omega_\ell) \tilde{h}(t_J\omega_\ell) + \sum_{j=1}^J g(t_j\omega_\ell) \tilde{g}(t_j\omega_\ell) = 1, \quad \forall \omega_\ell \in [0, 2].$$

Proof.

$$\begin{aligned}
\mathbf{H}_{t_J}^{\mathcal{G}} \tilde{\mathbf{H}}_{t_J}^{\mathcal{G}} + \sum_{j=1}^J \mathbf{G}_{t_j}^{\mathcal{G}} \tilde{\mathbf{G}}_{t_j}^{\mathcal{G}} &= \sum_{\ell, \ell'=1}^p h(t_J \omega_{\ell}) \tilde{h}(t_J \omega_{\ell'}) \mathbf{S}_{\ell} \mathbf{S}_{\ell'} + \sum_{j=1}^J \sum_{\ell, \ell'=1}^p g(t_j \omega_{\ell}) \tilde{g}(t_j \omega_{\ell'}) \mathbf{S}_{\ell} \mathbf{S}_{\ell'} \\
&= \sum_{\ell=1}^p h(t_J \omega_{\ell}) \tilde{h}(t_J \omega_{\ell}) \mathbf{S}_{\ell} + \sum_{j=1}^J \sum_{\ell=1}^p g(t_j \omega_{\ell}) \tilde{g}(t_j \omega_{\ell}) \mathbf{S}_{\ell} \\
&= \sum_{\ell=1}^p [h(t_J \omega_{\ell}) \tilde{h}(t_J \omega_{\ell}) + \sum_{j=1}^J g(t_j \omega_{\ell}) \tilde{g}(t_j \omega_{\ell})] \mathbf{S}_{\ell}
\end{aligned}$$

Consequently, (31) is guaranteed if and only if

$$h(t_J \omega_{\ell}) \tilde{h}(t_J \omega_{\ell}) + \sum_{j=1}^J g(t_j \omega_{\ell}) \tilde{g}(t_j \omega_{\ell}) = 1 \quad \forall \ell = 1 \dots p. \quad \square$$

We now define a frame as follows.

Definition 8.5. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph and μ a measure on \mathcal{V} . A countable family of elements $\{\mathbf{f}_k\}_{k=1}^n \in \ell^2(\mathcal{V})$ is said to be a frame if for any graph signal $\mathbf{f} \in \ell^2(\mathcal{V})$ we have

$$A \|\mathbf{f}\|^2 \leq \sum_{k=1}^n |\langle \mathbf{f}, \mathbf{f}_k \rangle|^2 \leq B \|\mathbf{f}\|^2,$$

for some constants $0 < A \leq B < \infty$ which are called frame bounds.

We introduce the rectangular matrix $\mathbf{K} = (\mathbf{H}_{t_J}^{\mathcal{G}}, \mathbf{G}_{t_1}^{\mathcal{G}}, \dots, \mathbf{G}_{t_J}^{\mathcal{G}}) \in \mathbb{R}^{N \times N(J+1)}$ where :

$$\mathbf{K} = (\mathbf{h}_{t_J,1}^{\mathcal{G}}, \dots, \mathbf{h}_{t_J,N}^{\mathcal{G}}, \mathbf{g}_{t_1,1}^{\mathcal{G}}, \dots, \mathbf{g}_{t_J,1}^{\mathcal{G}}, \dots, \mathbf{g}_{t_J,N}^{\mathcal{G}}) \in \mathbb{R}^{N \times N(J+1)}.$$

We also introduce the following rectangular matrix $\tilde{\mathbf{K}} = (\tilde{\mathbf{H}}_{t_J}^{\mathcal{G}}, \tilde{\mathbf{G}}_{t_1}^{\mathcal{G}}, \dots, \tilde{\mathbf{G}}_{t_J}^{\mathcal{G}}) \in \mathbb{R}^{N(J+1) \times N}$

$$(32) \quad \tilde{\mathbf{K}} = \begin{pmatrix} \tilde{\mathbf{h}}_{t_J,1}^{\mathcal{G}} \\ \vdots \\ \tilde{\mathbf{h}}_{t_J,N}^{\mathcal{G}} \\ \tilde{\mathbf{g}}_{t_1,1}^{\mathcal{G}} \\ \vdots \\ \tilde{\mathbf{g}}_{t_J,1}^{\mathcal{G}} \\ \vdots \\ \tilde{\mathbf{g}}_{t_J,N}^{\mathcal{G}} \end{pmatrix} \in \mathbb{R}^{N(J+1) \times N}.$$

Proposition 8.2. Let \mathbf{K} and $\tilde{\mathbf{K}}$ be respectively the synthesis and analysis filter banks. We assume that the filter banks \mathbf{K} and $\tilde{\mathbf{K}}$ verifies the perfect reconstruction condition enunciated at proposition 8.1. Consequently, $\tilde{\mathbf{K}}$ is a frame for $\ell^2(\mathcal{V})$ with lower frame bound $1/\|\mathbf{K}\|^2$ and upper frame bound $\|\tilde{\mathbf{K}}\|^2$.

Proof. We have

$$\|\tilde{\mathbf{K}} \mathbf{f}\|^2 = \sum_{k=1}^N |\langle \tilde{\mathbf{h}}_{t_J,k}^{\mathcal{G}}, \mathbf{f} \rangle|^2 + \sum_{j=1}^J \sum_{k=1}^N |\langle \tilde{\mathbf{g}}_{t_j,k}^{\mathcal{G}}, \mathbf{f} \rangle|^2.$$

Consequently, we can write

$$\begin{aligned} A\|\mathbf{f}\|^2 &\leq \|\tilde{\mathbf{K}}\mathbf{f}\|^2 \leq B\|\mathbf{f}\|^2, \\ A\|\mathbf{f}\|^2 &\leq \sum_{k=1}^N |\langle \tilde{\mathbf{h}}_{t_J,k}^{\mathcal{G}}, \mathbf{f} \rangle|^2 + \sum_{j=1}^J \sum_{k=1}^N |\langle \tilde{\mathbf{g}}_{t_j,k}^{\mathcal{G}}, \mathbf{f} \rangle|^2 \leq B\|\mathbf{f}\|^2, \end{aligned}$$

where $A = 1/\|\mathbf{K}\|^2$ et $B = \|\tilde{\mathbf{K}}\|^2$.

The right side is obtained by Cauchy-Schwarz inequality applied on $\|\tilde{\mathbf{K}}\mathbf{f}\|^2$. The left side is obtained by the fact that $\|\mathbf{f}\|^2 = \|\mathbf{K}\tilde{\mathbf{K}}\mathbf{f}\|^2 \leq \|\mathbf{K}\|^2 \|\tilde{\mathbf{K}}\mathbf{f}\|^2$ and application of the Cauchy-Schwarz inequality. As a result, $\tilde{\mathbf{K}}$ is a frame. \square

Definition 8.6 (Wavelet decomposition of a signal on a directed graph). *Any graph signal $\mathbf{f} \in \ell^2(\mathcal{V})$ can be expressed as follows :*

$$\mathbf{f} = \sum_{k=1}^N \langle \mathbf{f}, \tilde{\mathbf{h}}_{t_J,k} \rangle \mathbf{h}_{t_J,k} + \sum_{j=1}^J \sum_{k=1}^N \langle \mathbf{f}, \tilde{\mathbf{g}}_{t_j,k} \rangle \mathbf{g}_{t_j,k}.$$

8.2. Critically sampled wavelet transform on directed graphs.

8.2.1. Notations. We first recall relevant and useful notations for a clear understanding of the construction of diffusion wavelets [21, 22]. Two notations are introduced: one for the representation of linear transformations as matrices and a second for the representation of sets of vectors as matrices where columns correspond to a vector in the set. These notations were used in [21, 22] except that we adopt a column vector convention whereas a row vector convention was used in [21, 22].

Let $\mathbf{V}_0 = \ell^2(\mathcal{V})$ be the space of functions defined over the vertices of a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. If \mathbf{L} is a linear transformation of \mathbf{V}_0 into \mathbf{V}_0 , $[\mathbf{L}]_{B_1}^{B_2}$ indicates the matrix representing the linear transformation \mathbf{L} with respect to the basis B_1 in the domain and B_2 in the range. A set of vectors X represented in a given basis B will be written in the matrix form $[X]_B$ where the columns of $[X]_B$ are the coordinates of the vectors X in the coordinate system defined by B . Generally, $[B_1]_{B_2} = [\mathbf{I}]_{B_1}^{B_2}$ represent the basis vectors B_1 in terms of the basis B_2 . We will note that, for a given basis B if the input and output bases are the same, $[\mathbf{I}]_B^B = \mathbf{I}$ the identity matrix. We will also abuse this notation if B_2 spans a subspace of the space spanned by B_1 . If B_1 and B_2 are two bases, the representations of X in B_1 and B_2 are related as follows: $[X]_{B_2} = [\mathbf{I}]_{B_1}^{B_2} [X]_{B_1}$. If B_1, B_2, B_3, B_4 are arbitrary bases, the matrix representation of \mathbf{L} with respect to the basis B_1 in the domain and B_2 in the range in the different bases is expressed as

$$[\mathbf{L}]_{B_1}^{B_2} = [\mathbf{I}]_{B_4}^{B_2} [\mathbf{L}]_{B_3}^{B_4} [\mathbf{I}]_{B_2}^{B_1}.$$

Furthermore, if B_1 and B_2 are linearly independent sets of vectors that do not span the whole space \mathbf{V}_0 then we will still use the notation $[\mathbf{L}]_{B_1}^{B_2}$, but in that case it will represent the *restriction* of the linear transformation \mathbf{L} to the domain and range subspaces spanned by B_1 and B_2 .

8.2.2. Diffusion wavelets. The construction of the diffusion wavelets [21, 22] enables a multi-resolution analysis of functions on graphs generalizing the concept of classical multiresolution analysis [36]. The starting point of this construction is a diffusion operator \mathbf{T} . Similarly to the classical multiresolution analysis, diffusion wavelets is characterized by a family of nested subspaces $\mathbf{V}_0 \supseteq \mathbf{V}_1, \dots, \mathbf{V}_j \supseteq \dots$,

where each subspace V_j is spanned by a basis of scaling functions Φ_j . The complement of V_{j+1} into V_j is called W_j and is spanned by a set of diffusion wavelets Ψ_j .

8.2.3. Construction. The construction of diffusion wavelets proceeds with a diffusion operator \mathbf{T} defined on a directed strongly connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. In the original diffusion wavelets framework by Coifman and Maggioni [21], the graph is assumed undirected and they suggest using the reversible random walk operator \mathbf{P} .

More generally, any operator \mathbf{T} can be used as long as it is a low-pass filter when it is applied to localized graph signals. Given a directed graph \mathcal{G} , an appropriate choice is to use the operator $\mathbf{T} = \mathbf{\Pi}^{1/2} \mathbf{P} \mathbf{\Pi}^{-1/2}$ where \mathbf{P} is the random walk operator with low-pass frequency response with respect to the definition proposed in the section 6.

Assuming an ordering of vertices in \mathcal{V} , \mathbf{T} is originally represented on the canonical basis $\Phi_0 = \{\delta_k\}_{k \in \mathcal{V}}$ of V_0 where δ_k is the Kronecker delta function corresponding to the vertex $k \in \mathcal{V}$. Using the notations from the section 8.2.1, $[\mathbf{T}]_{\Phi_0}^{\Phi_0}$ is the matrix representation of the linear operator \mathbf{T} with respect to the basis Φ_0 in the domain and Φ_0 in the range.

The columns of the matrix $[\mathbf{T}]_{\Phi_0}^{\Phi_0}$ can be described as a set of functions $\tilde{\Phi}_1 = \{\tilde{\phi}_{1,k}\}_{k \in \mathcal{V}}$ represented in the basis Φ_0 . Each element $\tilde{\phi}_{1,k}, k \in \mathcal{V}$ of $\tilde{\Phi}_1$ corresponds to the action of \mathbf{T} on the Kronecker δ -function at vertex k and is represented as $[\tilde{\phi}_{1,k}]_{\Phi_0} = [\mathbf{T}]_{\Phi_0}^{\Phi_0} \delta_k$.

The low-pass property of \mathbf{T} leads to the fact that each element $\tilde{\phi}_{1,k}$ of $\tilde{\Phi}_1$ is a function localized around its vertex k whose support extends to its close neighbors. In terms of the diffusion wavelets construction, the elements $\tilde{\phi}_{1,k}, k \in \mathcal{V}$ are therefore interpreted as *scaling* functions. Due to the fact that the support of each element $\tilde{\phi}_{1,k}$ covers a small neighborhood around their respective vertices, these elements $\tilde{\phi}_{1,k}$ can generally be well approximated by a linear combination of the other functions $\tilde{\phi}_{1,l}$ with $l \neq k$. The next stage of the construction is a column subset selection stage [25]. The aim is to find a small number of columns of $\tilde{\Phi}_1$, forming a set \mathcal{C} such that the residual $\|[\tilde{\Phi}_1]_{\Phi_0} - \Pi_{\mathcal{C}}[\tilde{\Phi}_1]_{\Phi_0}\|_{\beta}$ is as minimal as possible where $\Pi_{\mathcal{C}}$ is the projection matrix onto the space spanned by the columns of \mathcal{C} and β denotes the spectral norm or the Frobenius norm.

More precisely, this step involves the selection of a subset $\{\tilde{\phi}_{1,k}, k \in \mathcal{I}_1\}$ from $\tilde{\Phi}_1$, $|\mathcal{I}_1| \leq |\mathcal{V}|$. We select a subset $\{\tilde{\phi}_{1,k}, k \in \mathcal{I}_1\}$ such that all $\tilde{\phi}_{1,k}$ are well-enough approximated by linear combinations of the functions in the subset. In classical signal processing terms, this step is analogous to a subsampling of a set of scaling functions in the classical discrete wavelet transform. Coifman and Maggioni used a greedy approach to build the set \mathcal{I}_1 iteratively by using a modified Gram-Schmidt orthogonalization procedure [21].

The subset $\{\tilde{\phi}_{1,k}, k \in \mathcal{I}_1\}$ spans a subspace V_1 which corresponds to the first approximation subspace of the multi-resolution analysis. We will thus denote $\{\tilde{\phi}_{1,k}, k \in \mathcal{I}_1\}$ as Φ_1 , where Φ_1 is by definition a basis of V_1 that is generally not orthogonal. The vectors in Φ_1 are the *scaling functions* at scale 1.

In order to define the next scales of scaling functions, we consider the compression step of the diffusion operator \mathbf{T} on the subspace V_1 , that is its restriction on V_1 . The latter can be represented in Φ_1 as

$$[\mathbf{T}]_{\Phi_1}^{\Phi_1} = [\mathbf{I}]_{\Phi_0}^{\Phi_1} [\mathbf{T}]_{\Phi_0}^{\Phi_0} [\mathbf{I}]_{\Phi_1}^{\Phi_0}.$$

where $[\mathbf{I}]_{\Phi_0}^{\Phi_1}$ represents the restriction of the signal space V_0 to V_1 (with respective bases Φ_0 and Φ_1) and $[\mathbf{I}]_{\Phi_1}^{\Phi_0}$ represents the embedding of V_1 in V_0 .

The next approximation space $V_2 \subset V_1$ and its associated basis Φ_2 can be obtained in the same way as the definition of V_1 and its basis Φ_1 except that we now consider the operator \mathbf{T}^2 restricted to V_1 instead of \mathbf{T} in V_0 .

The columns of $[\mathbf{T}^2]_{\Phi_1}^{\Phi_1} = ([\mathbf{T}]_{\Phi_1}^{\Phi_1})^2$ can be interpreted as scaling functions at scale 2, $\tilde{\Phi}_2 = \{\tilde{\phi}_{2,k}\}$, represented in the basis Φ_1 . From these functions we extract a subset $\Phi_2 = \{\tilde{\phi}_{2,k}, k \in \mathcal{I}_2\}$ such that any function in $\tilde{\Phi}_2$ is well-approximated by a linear combination of functions of Φ_2 .

After j iterations of this procedure we have defined j approximation subspaces $V_j \subset V_{j-1} \subset \dots \subset V_1$ with corresponding bases $\Phi_j, \Phi_{j-1}, \dots, \Phi_1$. At each step the basis Φ_j is defined by its representation in the basis Φ_{j-1} based on the restriction of the operator \mathbf{T}^{2^j} to V_{j-1} . In order to represent these functions in the original basis Φ_0 of V_0 we can use

$$[\Phi_j]_{\Phi_0} = [\mathbf{I}]_{\Phi_j}^{\Phi_0} = [\mathbf{I}]_{\Phi_0}^{\Phi_0} [\mathbf{I}]_{\Phi_1}^{\Phi_0} \dots [\mathbf{I}]_{\Phi_{j-1}}^{\Phi_{j-2}} [\mathbf{I}]_{\Phi_j}^{\Phi_{j-1}}.$$

Since every function in Φ_0 is defined on V_0 , so is every function on Φ_j . Hence any function in the approximation space V_j can be extended naturally to the whole space V_0 .

Regarding the construction of the wavelets, we propose to construct the wavelet bases Ψ_j for the subspaces W_j by selecting a subset of the columns of the band pass operator $[\mathbf{I}_{V_j} - \Phi_{j+1} \Phi_{j+1}^\dagger]_{\Phi_j}^{\Phi_j}$ which is the orthogonal projector on the complement of V_{j+1} into V_j . The wavelets capture the detail lost from going from V_j to V_{j+1} . As our framework falls into bi-orthogonal scope, we need to build the dual wavelet bases. For a scale j , we have a wavelet base Ψ_j and we need to construct the associated dual base $\hat{\Psi}_j$ obtained such that we have the following relation

$$\hat{\Psi}_j^\dagger \Psi_j = \mathbf{I}.$$

where $\hat{\Psi}_j^\dagger$ is the pseudo-inverse of Ψ_j . We mention that in [22], biorthogonal wavelet transform is proposed but only the scaling functions are actually defined. Our construction is identical but we also propose a definition for wavelets.

8.2.4. A generalization using more arbitrary scaling operators. We propose a generalization of the diffusion wavelet framework that enables it to be combined for instance with the wavelet construction presented in Section 8.2. The idea is merely to replace the powers of \mathbf{T} as multiresolution scaling operators by more arbitrary low-pass filters. More precisely, where the operator \mathbf{T}^{2^j} is used to define the approximation space V_{j+1} in section 8.2.3, we propose to use instead a low-pass graph filter \mathbf{H}_j . If the graph filters \mathbf{H}_j correspond to the scaling operators defined in section 8.1 with appropriately increasing scales, this approach provides a way to reduce the redundancy of the sets of scaling functions. Similarly, the same approach could be used to reduce the redundancy of the wavelet functions defined in Section 8.2.

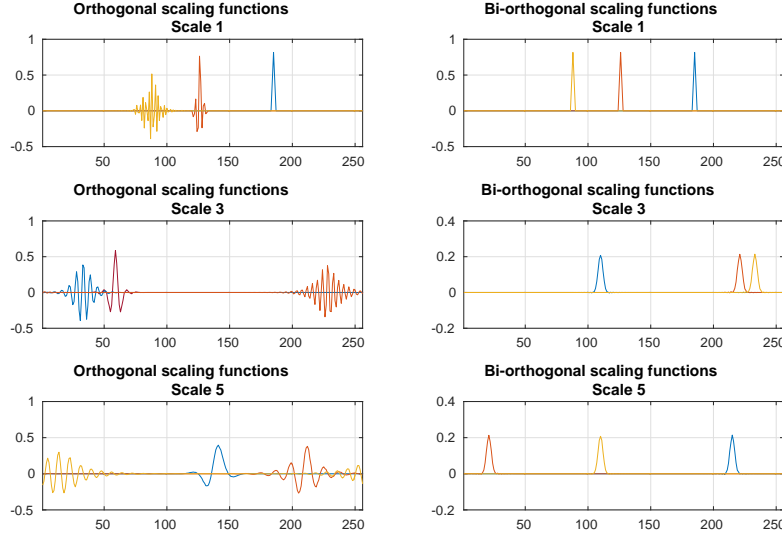


FIGURE 10. Orthogonal and biorthogonal scaling functions on the directed cycle graph \mathcal{C}_{256} .

8.3. Applications. This section is devoted to illustrate our multiresolution analysis framework through the visualization of wavelets and scaling functions on regular-type structures and semi-supervised learning on directed graphs. The application framework will be the same as in section 7.

8.3.1. Multiresolution analysis on the directed cycle graph. We show an example of multi-resolution analysis on the directed cycle graph \mathcal{C}_N with $N = 256$. We use the same assumptions made in section 6.4.1. We construct both orthogonal and biorthogonal multi-resolution analyses on \mathcal{C}_N through the framework of the diffusion wavelets applied on the dyadic powers of \mathbf{T} , i.e. $\{\mathbf{T}^{2^j}\}_{j=1}^J$. We set the number of scales at $J = 6$.

Figure 10 shows the orthogonal and biorthogonal scaling functions at scales 1, 3 and 5. At each scale, we represent 3 or 4 scaling functions.

Figure 11 shows the orthogonal and biorthogonal wavelet functions also at scales 1, 3 and 5.

We note at each scale that the support of the orthogonal scaling functions is larger than the support of the biorthogonal scaling functions. We also note at scale 3 and 5 that biorthogonal wavelet functions have support slightly smaller than the orthogonal wavelets functions. Furthermore, we note that orthogonal scaling functions as well orthogonal wavelet functions have more oscillations than their biorthogonal counterparts. That means that orthogonal scaling and wavelet functions have a poorer frequency localization. The conclusions about the spatial localization of the scaling functions are identical to the conclusions in [24]. The novelty here is we work on the directed circle graph and we show the orthogonal and biorthogonal wavelets functions in addition to the orthogonal and biorthogonal scaling functions. Finally, the scaling and wavelets functions are not centered

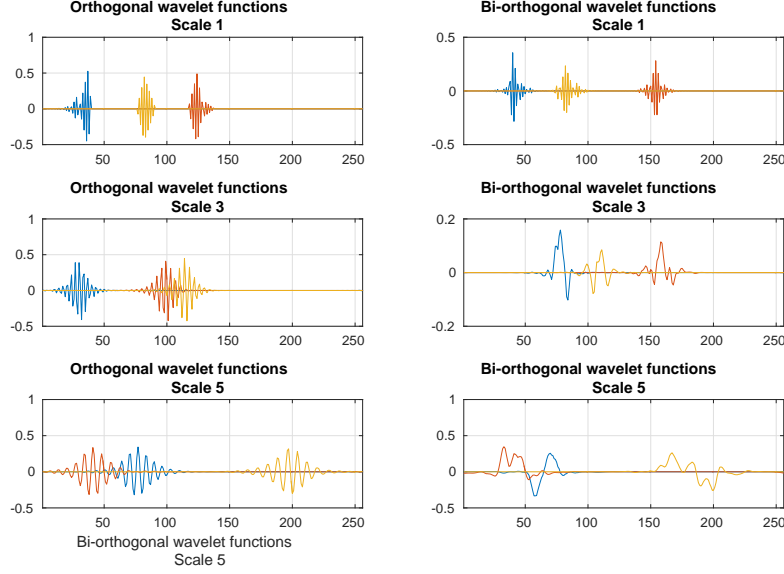


FIGURE 11. Orthogonal and biorthogonal wavelet functions on the directed cycle graph \mathcal{C}_{256} .

around the vertex where they have been selected. They propagate towards the parents' node of the selected vertex. We can also note the orthogonal wavelet transform is more robust with respect to perturbations than the biorthogonal wavelet transform. Let $\mathcal{W}_{ortho} = [\Psi_1, \dots, \Psi_5, \Phi_5]$ be the inverse wavelet transform and \mathcal{W}_{biorth} its biorthogonal counterpart. We have that the condition numbers $\kappa(\mathcal{W}_{ortho}) \approx 33$ and $\kappa(\mathcal{W}_{biorth}) \approx 2 \times 10^4$. Consequently, the biorthogonal wavelet transform is much more sensitive to perturbations than its orthogonal counterpart.

8.3.2. Multiresolution analysis on the directed Watts-Strogatz graph. In this section, we show examples of multi-resolution analyses on the directed version of a graph from the Watts-Strogatz model. The Watts-Strogatz model [85] is a undirected random graph model exhibiting small world properties including short average path lengths and high clustering [86]. The construction of a directed graph from the Watts-Strogatz model starts with a directed cycle graph with N vertices. Each node is connected to its k next nodes following the direction of the directed cycle graph. For the sake of simplicity, we consider the k next nodes connected to a given node i following the direction of the directed cycle graph as its "closest" neighbors. Starting from an arbitrary vertex, we apply the following procedure to each vertex in a clockwise manner. At vertex i , the edge that connects i to each of its next nodes is randomly rewired with probability p or remains untouched with probability $1 - p$. This procedure is repeated cyclically for each successive vertex until the vertex i is selected again. We denote by $\mathcal{G} \sim \text{DWS}(N, K, \beta)$ a graph constructed following a directed Watts-Strogatz model with N vertices, K nearest neighbors and rewiring probability β .

In a first instance, we analyze the scaling functions built from low-pass filters based on $\bar{\mathbf{T}}_\alpha$ at a given scale of $\mathcal{G}_1 \sim \text{DWS}(64, 2, 0)$. That is a special case of the

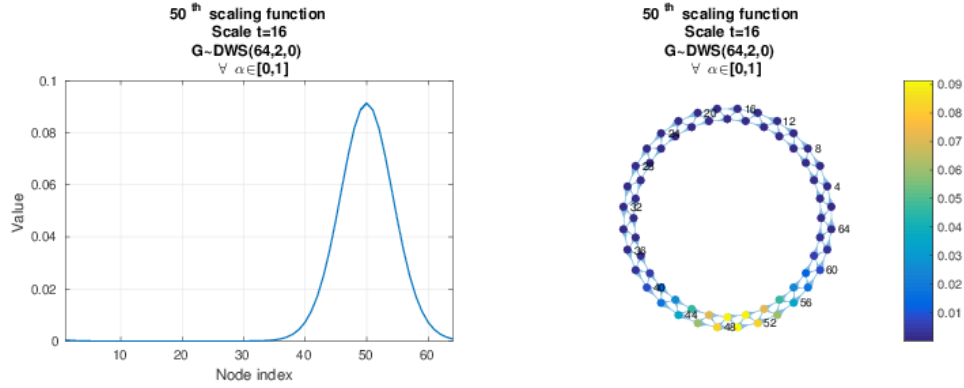


FIGURE 12. (Case 1) 50th scaling function at scale $t=16$ on the graph $\mathcal{G} \sim \text{DWS}(64, 2, 0)$, for all $\alpha \in [0, 1]$.

directed Watts-Strogatz model with no rewiring. In a second instance, we analyze some scaling functions built from low-pass filters based on $\bar{\mathbf{T}}_\alpha$ at a given scale of $\mathcal{G}_2 \sim \text{DWS}(64, 2, 0.02)$. Lastly, we show both orthogonal and biorthogonal scaling and wavelets functions built following the construction Section 8.2. We summarize the following cases in table 6.

Case	Graph	Reference operators
Case 1 : Scaling functions built wrt. sect. 8.1	$\mathcal{G}_1 \sim \text{DWS}(64, 2, 0)$	$\bar{\mathbf{T}}_\alpha \in \bar{\mathcal{T}}$
Case 2: Scaling functions built wrt. sect. 8.1	$\mathcal{G}_2 \sim \text{DWS}(64, 2, 0.02)$	$\bar{\mathbf{T}}_\alpha \in \bar{\mathcal{T}}$
Case 3: Scaling functions built wrt. sect. 8.2	$\mathcal{G}_2 \sim \text{DWS}(64, 2, 0.02)$	$\bar{\mathbf{T}}$

TABLE 6. Table of the different cases for the directed Watts Strogatz graph.

Case 1. In this case, we analyze some scaling functions built from low-pass filters based on $\bar{\mathbf{P}}_\alpha$ at scale $t = 2^4$ of a directed Watts-Strogatz graph $\mathcal{G}_2 \sim \text{DWS}(64, 2, 0)$. We consider the following low-pass filters

$$(33) \quad \bar{\mathbf{T}}_\alpha = \mathbf{\Pi}^{1/2} \bar{\mathbf{P}}_\alpha \mathbf{\Pi}^{-1/2}, \quad \forall \alpha \in [0, 1].$$

and we construct a collection of low-pass graph filters at a given scale following the construction presented in section 8.1. More precisely, we build the filters \mathbf{H}_α as follows

$$\mathbf{H}_\alpha = \sum_{\omega \in \omega} h(t\omega) \mathbf{S}_{\omega, \alpha},$$

with $t = 2^4$ and $h(x) = \exp(-x)$.

Figure 12 represents the 50th scaling function $\mathbf{h}_{\alpha, 50} = \mathbf{H}_\alpha \delta_{50}$ at scale $t = 2^4$ built from low-pass filters \mathbf{T}_α , for all $\alpha \in [0, 1]$ according the construction section 8.1 on the directed Watts-Strogatz graph $\mathcal{G} \sim \text{DWS}(64, 2, 0)$. The particularity of this case is that the following graph has no rewiring. Consequently, the associated adjacency

matrix is circulant. All the low-pass filters \mathbf{T}_α admit the same discrete Fourier basis and the associated scaling functions $\mathbf{h}_{\alpha,50}$ are exactly the same. We observe that the scaling function is centered around the node 50 and has a symmetric shape. In this case, for non symmetric graph filters \mathbf{H}_α , the scaling functions have symmetric shape and are centered around its given node.

Case 2. In this case, we analyze some scaling functions built from low-pass filters based on $\bar{\mathbf{P}}_\alpha$ at scale 10 of $\mathcal{G}_2 \sim \text{DWS}(64, 2, 0.02)$. We also consider the following low-pass filters

$$(34) \quad \bar{\mathbf{T}}_\alpha = \mathbf{\Pi}^{1/2} \bar{\mathbf{P}}_\alpha \mathbf{\Pi}^{-1/2}, \quad \forall \alpha \in \{0, 0.5, 1\}.$$

and we construct a collection of low-pass graph filters at a given scale following the construction presented in section 8.1. More precisely, we build the filters \mathbf{H}_α as follows

$$\mathbf{H}_\alpha = \sum_{\omega \in \omega} h(t\omega) \mathbf{S}_{\omega, \alpha}$$

with $t = 2^4$ and $h(x) = \exp(-x)$. We observe the 50^{th} scaling functions at the scale $t = 2^4$, that is $\mathbf{h}_{\alpha,50} = \mathbf{H}_\alpha \delta_{50}$.

Figure 13 represent the 50^{th} scaling function at the scale $t = 2^4$, for different α , $\mathbf{h}_{\alpha,50}$. The graph \mathcal{G}_2 admits a directed edge from the node 50 to the node 11 and a directed edge from the node 30 to the node 50. As α increases towards 1, we observe that the scaling function $\mathbf{h}_{\alpha,50}$ propagates around the child node 11. This means the more α increases, the more the influence of the backward random walk \mathbf{P}^* is important. We note for $\alpha = 0.5$, the scaling function $\mathbf{h}_{\alpha,50}$ is centered around the child node 11 and the parent node 30. This case seems to be the equilibrium. Finally, as α goes to 0, we observe that $\mathbf{h}_{\alpha,50}$ propagates around the parent node 30.

Case 3. Here, we consider orthogonal and biorthogonal scaling functions on a graph $\mathcal{G} \sim \text{DWS}(64, 2, 0.02)$ used on the Case 2. These orthogonal and biorthogonal scaling functions are built from the diffusion wavelets procedure. We start the procedure with the low-pass filter $\mathbf{H} = \mathbf{T}^2$ and we look at the dyadic powers of \mathbf{H} i.e. $\{\mathbf{H}^{2^j}\}_{j=1}^J$. We set the number of scales at $J = 5$. We observe the orthogonal and biorthogonal scaling functions at node 49 at scale 3 obtained by the diffusion wavelets procedure and we compare to the scaling function at node 49 built from a graph filter based on $\bar{\mathbf{T}}$ at the scale $t = 16$.

Figure 14 shows the orthogonal and bi-orthogonal scaling functions at node 49 built w.r.t. the diffusion wavelets framework and the scaling function at node 49 built w.r.t. the spectral graph wavelets framework. We note that the biorthogonal scaling function and the scaling function built by the spectral graph wavelet framework are similar, i.e. well localized around nodes 11 and 30 and localized around 49 with a large size support. By contrast, the orthogonal scaling function has a poor spatial localization around the nodes 11 and 30.

8.3.3. Semi-supervised learning on directed graphs with ℓ_1 -regularization. We discuss the semi-supervised learning approach in the case of functions over directed graphs with ℓ_1 regularization on the wavelet coefficients. Our aim is to show that the performance of the semi-supervised learning problem with ℓ_1 regularization is competitive to the existing approaches, e.g. the semi-supervised learning problem studied in section 7. We discuss the following semi-supervised learning problem,

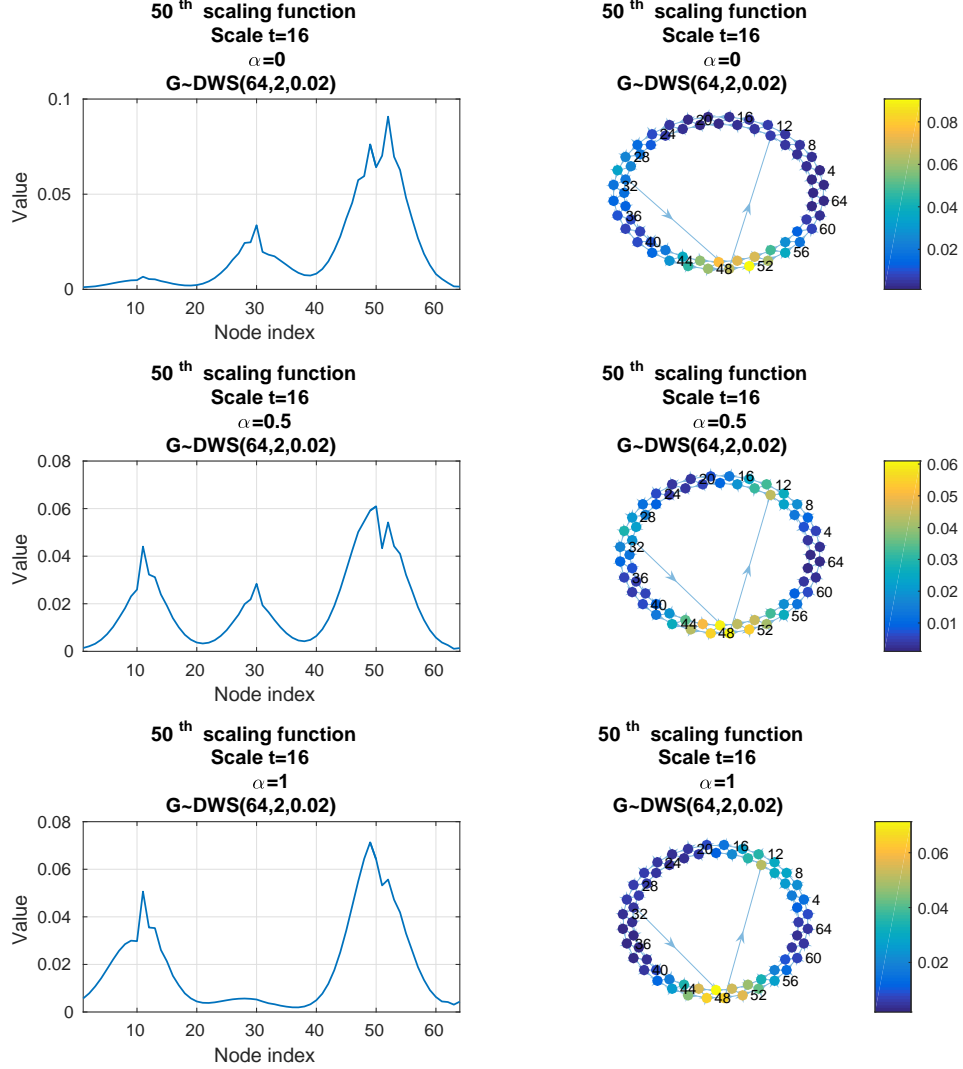


FIGURE 13. 50th scaling function at scale 4 on a graph $\mathcal{G} \sim \text{DWS}(64, 2, 0.02)$, $\alpha \in \{0, 0.5, 1\}$, eq. (33).

again on the example of the political blogs dataset [28]. We use the same notations described in section 7 and work on the subgraph \mathcal{G}' . The ℓ_1 -regularized semi-supervised learning problem is

$$(35) \quad \mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \|\tilde{\mathbf{y}} - \mathbf{M}\mathbf{K}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1, \quad \lambda \in \mathbb{R}^+.$$

The graph signal $\tilde{\mathbf{y}} = \mathbf{M}\mathbf{y}$ is the partially labeled graph signal with $\mathbf{M} = \{m_{ij}\}_{\{1 \leq i, j \leq N\}}$ is the mask operator, i.e. the diagonal matrix where $m_{ii} = \mathbb{1}_{v_i \in \mathcal{O}}$ where $\mathcal{O} \subset \mathcal{V}$ is the subset of known labels.

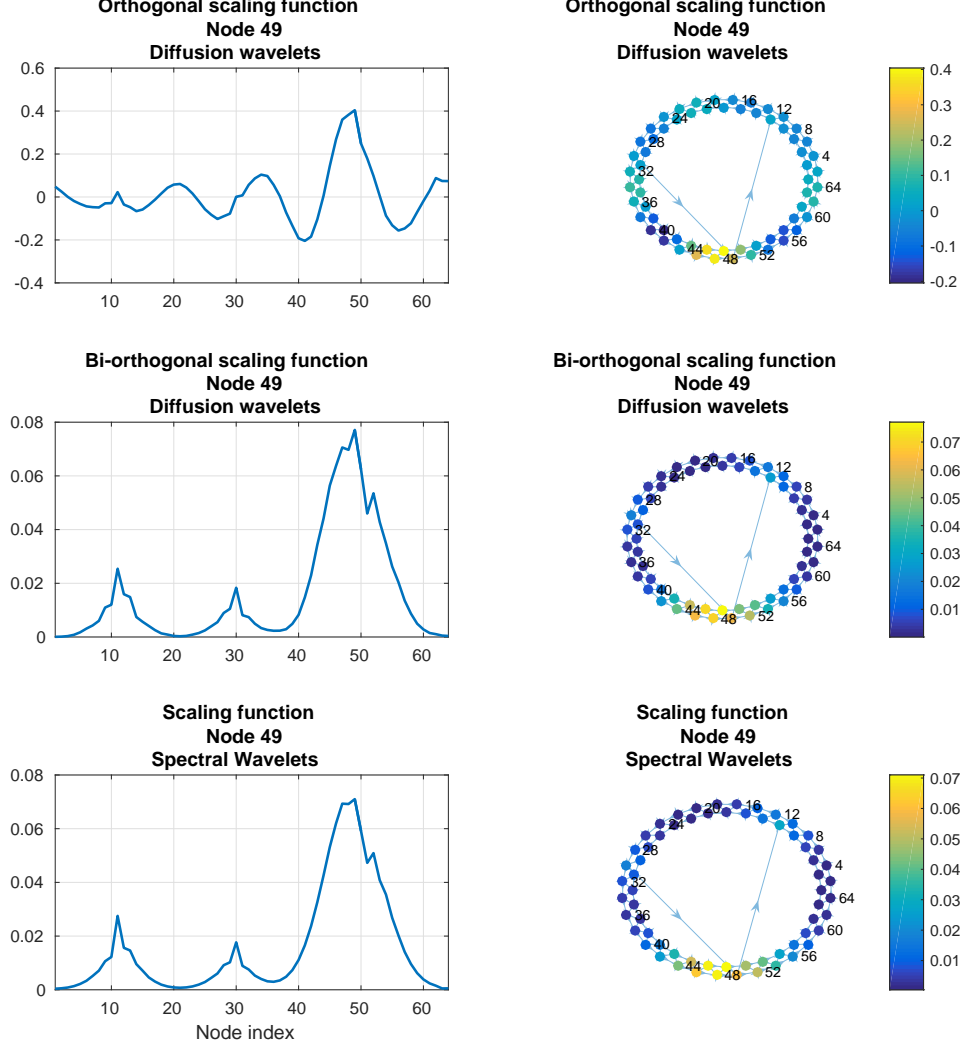


FIGURE 14. Orthogonal and bi-orthogonal scaling functions built w.r.t the diffusion wavelet framework versus scaling function built w.r.t spectral wavelets framework.

The matrix $\mathbf{K} = (\mathbf{H}_J, \mathbf{G}_1, \dots, \mathbf{G}_J)$ is the synthesis filter bank. If we set $\mathbf{X} = \mathbf{MK}$, the equation (35) can be rewritten as

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \|\tilde{\mathbf{y}} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1, \quad \lambda \in \mathbb{R}.$$

The formulation (35) is identical to the formulation of the problem of signal restoration with redundant wavelet transforms in [68], except that \mathbf{K} is the synthesis wavelet transform for functions defined over directed graphs. Furthermore, previous approaches of semi-supervised learning on undirected graphs have been investigated using overcomplete graph wavelets [69] or critically sampled spline graph

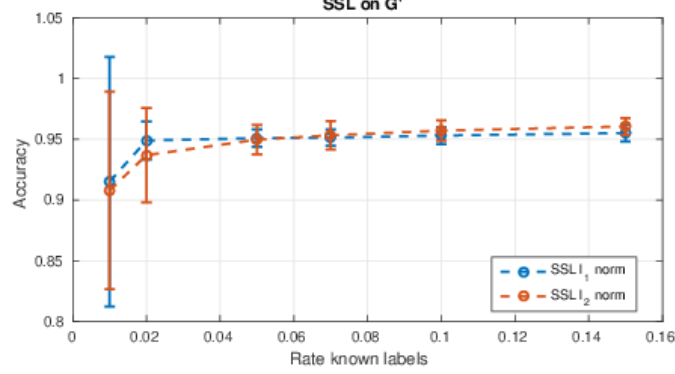


FIGURE 15. SSL on the largest connected graph of US political blogs.

wavelets [70]. The ℓ_1 -regularized synthesis semi-supervised learning problem is convex and can be solved efficiently using e.g. proximal splitting methods [71, 72, 73]. From the solution \mathbf{w}^* of eq. (35), we define the restored signal \mathbf{f}^* as

$$\mathbf{f}^* = \text{sign}(\mathbf{K}\mathbf{w}^*).$$

where \mathcal{L} is the directed normalized Laplacian defined at (4) and the graph bandpass filters $\mathbf{G}_1, \mathbf{G}_2$ are

We compare the performances of formulations (35) and (16). The formulation (35) requires a synthesis filter bank \mathbf{K} . We choose a two scales synthesis graph filter \mathbf{K} . We choose a two scales synthesis graph filter bank $\mathbf{K} = (\mathbf{H}_2, \mathbf{G}_1, \mathbf{G}_2)$ based on a heat kernel construction [83]

$$\mathbf{H}_j = e^{-\mathcal{L}t_j}, \quad t_j = 2^j, \quad j = 1, 2.$$

with \mathcal{L} is the directed Laplacian defined at (4) and the graph bandpass filters $\mathbf{G}_1, \mathbf{G}_2$ as

$$\mathbf{G}_2 = \mathbf{H}_1 - \mathbf{H}_2, \quad \mathbf{G}_1 = \mathbf{I} - \mathbf{H}_1.$$

Figure 15 shows the performance of semi-supervised learning between the formulation (35) and (16) on the largest strongly connected graph obtained from US political blogs. The performance is obtained by averaging 200 realizations and determining the value of the parameters λ and γ associated respectively to the approaches (35) and (16) giving the best performance rates. On fig. 15, the performance based on the ℓ_1 penalization of redundant graph wavelets coefficients on directed graphs is competitive against the problem with ℓ_2 Dirichlet regularization term for all percentage of known labels. Normally, one should expect better performances in the ℓ_1 case but it seems not the case because our graph wavelets are not good in terms of spatial localization. Still, in order to improve the performance of the method (35), it could be wise to build new wavelets with a better space-frequency characterization, and choose the best suitable number of scales with respect to the characteristics of the data set.

9. CONCLUSION

We introduced a novel harmonic analysis on directed graphs. First, we proposed a frequency analysis for functions defined on directed graphs based on the eigenvectors of the random walk operator on a directed graph. From this Fourier-type frequency interpretation, we showed how to construct redundant wavelets on directed graphs as well as critically sampled wavelets by generalizing the diffusion wavelets framework. Finally, we illustrated our harmonic analysis through examples of semi-supervised learning and graph signal modeling on directed graphs and showed the relevance of our framework to existing approaches.

10. ACKNOWLEDGEMENTS

The authors would like to thank Romain Cosentino, Randall Balestriero, Yanis Bahroun, Benjamin Girault and Antonio Ortega for the useful discussions.

REFERENCES

- [1] Steve Butler. Interlacing for weighted graphs using the normalized laplacian. *Electronic Journal of Linear Algebra*, 16(1):8, 2007.
- [2] Fan Chung. Laplacians and the cheeger inequality for directed graphs. *Annals of Combinatorics*, 9(1):1–19, 2005.
- [3] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- [4] Franziska Göbel, Gilles Blanchard, and Ulrike von Luxburg. Construction of tight frames on graphs and application to denoising. In *Handbook of Big Data Analytics*, pages 503–522. Springer, 2018.
- [5] Thierry Coulhon, Gerard Kerkycharian, and Pencho Petrushev. Heat kernel generated frames in the setting of dirichlet spaces. *Journal of Fourier Analysis and Applications*, 18(5):995–1066, 2012.
- [6] D. B. H. Tay and A. Ortega. Bipartite graph filter banks: Polyphase analysis and generalization. *IEEE Transactions on Signal Processing*, 65(18):4833–4846, Sept 2017.
- [7] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst. Graph signal processing. *arXiv preprint arXiv:1712.00468*, 2017.
- [8] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.
- [9] Aliaksei Sandryhaila and José MF Moura. Discrete signal processing on graphs. *IEEE transactions on signal processing*, 61(7):1644–1656, 2013.
- [10] Pierre Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, volume 31. Springer Science & Business Media, 2013.
- [11] David Aldous and Jim Fill. Reversible markov chains and random walks on graphs, 2002.
- [12] László Lovász. Random walks on graphs. *Combinatorics, Paul erdos is eighty*, 2(1-46):4, 1993.
- [13] James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process. *The annals of applied probability*, pages 62–87, 1991.
- [14] Thierry Coulhon and Alexander Grigoryan. Random walks on graphs with regular volume growth. *Geometric and Functional Analysis*, 8(4):656–701, 1998.
- [15] Audrey Terras. *Fourier analysis on finite groups and applications*, volume 43. Cambridge University Press, 1999.
- [16] MI Ostrovskii. Sobolev spaces on graphs. *Quaestiones Mathematicae*, 28(4):501–523, 2005.
- [17] A Kaveh and H Rahami. A unified method for eigendecomposition of graph products. *International Journal for Numerical Methods in Biomedical Engineering*, 21(7):377–388, 2005.
- [18] Ali Kaveh. *Optimal analysis of structures by concepts of symmetry and regularity*. Springer, 2013.

- [19] M Maggioni and HN Mhaskar. Diffusion polynomial frames on metric measure spaces. *Applied and Computational Harmonic Analysis*, 24(3):329–353, 2008.
- [20] David K Hammond, Pierre Vandergheynst, and Rémi Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.
- [21] Ronald R Coifman and Mauro Maggioni. Diffusion wavelets. *Applied and Computational Harmonic Analysis*, 21(1):53–94, 2006.
- [22] Mauro Maggioni, James C Bremer, Ronald R Coifman, and Arthur D Szlam. Biorthogonal diffusion wavelets for multiscale representations on manifolds and graphs. In *Wavelets XI*, volume 5914, page 59141M. International Society for Optics and Photonics, 2005.
- [23] Ole Christensen et al. *An introduction to frames and Riesz bases*. Springer, 2016.
- [24] James C Bremer, Ronald R Coifman, Mauro Maggioni, and Arthur D Szlam. Diffusion wavelet packets. *Applied and Computational Harmonic Analysis*, 21(1):95–112, 2006.
- [25] Ali Civril. *Column Subset Selection for approximating data matrices*. PhD thesis, Rensselaer Polytechnic Institute, 2009.
- [26] Aliaksei Sandryhaila and Jose MF Moura. Discrete signal processing on graphs: Frequency analysis. *IEEE Trans. Signal Processing*, 62(12):3042–3054, 2014.
- [27] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning from labeled and unlabeled data on a directed graph. In *Proceedings of the 22nd international conference on Machine learning*, pages 1036–1043. ACM, 2005.
- [28] Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM, 2005.
- [29] Amy N Langville and Carl D Meyer. *Google’s PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2011.
- [30] Stefania Sardellitti, Sergio Barbarossa, and Paolo Di Lorenzo. On the graph fourier transform for directed graphs. *IEEE Journal of Selected Topics in Signal Processing*, 11(6):796–811, 2017.
- [31] Rasoul Shafipour, Ali Khodabakhsh, Gonzalo Mateos, and Evdokia Nikolova. A digraph fourier transform with spread frequency components. *arXiv preprint arXiv:1705.10821*, 2017.
- [32] Frank Bauer. Normalized graph laplacians for directed graphs. *Linear Algebra and its Applications*, 436(11):4193–4222, 2012.
- [33] Delio Mugnolo. *Semigroup methods for evolution equations on networks*. Springer.
- [34] Fan RK Chung. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.
- [35] David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.
- [36] Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.
- [37] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. Graph laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8(Jun):1325–1368, 2007.
- [38] Mikhail Belkin and Partha Niyogi. Convergence of laplacian eigenmaps. In *Advances in Neural Information Processing Systems*, pages 129–136, 2007.
- [39] Mikhail Belkin, Irina Matveeva, and Partha Niyogi. Regularization and semi-supervised learning on large graphs. In *International Conference on Computational Learning Theory*, pages 624–638. Springer, 2004.
- [40] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.
- [41] Yonathan Aflalo and Ron Kimmel. Spectral multidimensional scaling. *Proceedings of the National Academy of Sciences*, 110(45):18052–18057, 2013.
- [42] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.
- [43] Peter Schröder and Wim Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 161–172. ACM, 1995.
- [44] J-P Antoine and Pierre Vandergheynst. Wavelets on the 2-sphere: A group-theoretical approach. *Applied and Computational Harmonic Analysis*, 7(3):262–291, 1999.

- [45] Y Wiaux, JD McEwen, P Vandergheynst, and O Blanc. Exact reconstruction with directional wavelets on the sphere. *Monthly Notices of the Royal Astronomical Society*, 388(2):770–788, 2008.
- [46] Nick Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Applied and computational harmonic analysis*, 10(3):234–253, 2001.
- [47] Emmanuel J Candès and David L Donoho. New tight frames of curvelets and optimal representations of objects with piecewise c^2 singularities. *Communications on pure and applied mathematics*, 57(2):219–266, 2004.
- [48] Gabriel Peyré and Stéphane Mallat. Orthogonal bandelet bases for geometric images approximation. *Communications on Pure and Applied Mathematics*, 61(9):1173–1212, 2008.
- [49] Wim Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and computational harmonic analysis*, 3(2):186–200, 1996.
- [50] Albert Cohen, Ingrid Daubechies, and J-C Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 45(5):485–560, 1992.
- [51] Daryl Geller and Azita Mayeli. Continuous wavelets on compact manifolds. *Mathematische Zeitschrift*, 262(4):895, 2009.
- [52] Bin Dong. Sparse representation on graphs by tight wavelet frames and applications. *Applied and Computational Harmonic Analysis*, 42(3):452–479, 2017.
- [53] Mark Crovella and Eric Kolaczyk. Graph wavelets for spatial traffic analysis. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 3, pages 1848–1857. IEEE, 2003.
- [54] Maarten Jansen, Guy P Nason, and Bernard W Silverman. Multiscale methods for data on graphs and irregular multidimensional situations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1):97–125, 2009.
- [55] Raif Rustamov and Leonidas J Guibas. Wavelets on graphs via deep learning. In *Advances in neural information processing systems*, pages 998–1006, 2013.
- [56] Arthur D Szlam, Mauro Maggioni, Ronald R Coifman, and James C Bremer. Diffusion-driven multiscale analysis on manifolds and graphs: top-down and bottom-up constructions. In *Wavelets XI*, volume 5914, page 59141D. International Society for Optics and Photonics, 2005.
- [57] Matan Gavish, Boaz Nadler, and Ronald R Coifman. Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning. In *ICML*, pages 367–374, 2010.
- [58] Raif M Rustamov. Average interpolating wavelets on point clouds and graphs. *arXiv preprint arXiv:1110.2227*, 2011.
- [59] Franziska Göbel, Gilles Blanchard, and Ulrike von Luxburg. Construction of tight frames on graphs and application to denoising. *arXiv preprint arXiv:1408.4012*, 2014.
- [60] Nicolas Tremblay and Pierre Borgnat. Graph wavelets for multiscale community mining. *IEEE Transactions on Signal Processing*, 62(20):5227–5239, 2014.
- [61] Nicolas Tremblay and Pierre Borgnat. Subgraph-based filterbanks for graph signals. *IEEE Transactions on Signal Processing*, 64(15):3827–3840, 2016.
- [62] Sunil K Narang and Antonio Ortega. Perfect reconstruction two-channel wavelet filter banks for graph structured data. *IEEE Transactions on Signal Processing*, 60(6):2786–2799, 2012.
- [63] David I Shuman, Mohammad Javad Faraji, and Pierre Vandergheynst. A multiscale pyramid transform for graph signals. *IEEE Transactions on Signal Processing*, 64(8):2119–2134, 2016.
- [64] Luca Avena, Fabienne Castell, Alexandre Gaudillière, and Clothilde Mélot. Intertwining wavelets or multiresolution analysis on graphs through random forests. *arXiv preprint arXiv:1707.04616*, 2017.
- [65] Hrushikesh N Mhaskar. A unified framework for harmonic analysis of functions on directed graphs and changing data. *Applied and Computational Harmonic Analysis*, 2016.
- [66] Jeff Irion and Naoki Saito. Hierarchical graph laplacian eigen transforms. *JSIAM Letters*, 6:21–24, 2014.
- [67] Evarist Giné, Vladimir Koltchinskii, et al. Empirical graph laplacian approximation of laplace–beltrami operators: Large sample results. In *High dimensional probability*, pages 238–259. Institute of Mathematical Statistics, 2006.
- [68] Ivan W Selesnick and Mario AT Figueiredo. Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors. In *Wavelets XIII*, volume 7446, page 74460D. International Society for Optics and Photonics, 2009.

- [69] David I Shuman, Mohammadjavad Faraji, and Pierre Vandergheynst. Semi-supervised learning with spectral graph wavelets. In *Proceedings of the International Conference on Sampling Theory and Applications (SampTA)*, number EPFL-CONF-164765, 2011.
- [70] Venkatesan N Ekambaram, Giulia Fanti, Babak Ayazifar, and Kannan Ramchandran. Wavelet-regularized graph semi-supervised learning. In *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, pages 423–426. IEEE, 2013.
- [71] Patrick L Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.
- [72] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.
- [73] Stephen Becker, Jérôme Bobin, and Emmanuel J Candès. NESTA: A fast and accurate first-order method for sparse recovery. *SIAM Journal on Imaging Sciences*, 4(1):1–39, 2011.
- [74] Inam Ur Rahman, Iddo Drori, Victoria C Stodden, David L Donoho, and Peter Schröder. Multiscale representations for manifold-valued data. *Multiscale Modeling & Simulation*, 4(4):1201–1232, 2005.
- [75] Benjamin Girault, Antonio Ortega, and Shrikanth Narayanan. Irregularity-aware graph fourier transforms. *arXiv preprint arXiv:1802.10220*, 2018.
- [76] Aamir Anis, Akshay Gadde, and Antonio Ortega. Efficient sampling set selection for bandlimited graph signals using graph spectral proxies. *IEEE Transactions on Signal Processing*, 64(14):3775–3789, 2016.
- [77] Ravi Montenegro, Prasad Tetali, et al. Mathematical aspects of mixing times in markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006.
- [78] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1-3):1–101, 2011.
- [79] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge university press, 2008.
- [80] Steven Rosenberg. *The Laplacian on a Riemannian manifold: an introduction to analysis on manifolds*, volume 31. Cambridge University Press, 1997.
- [81] Olivier Lablée. *Spectral Theory in Riemannian Geometry*. 2015.
- [82] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.
- [83] Martin T Barlow. *Random walks and heat kernels on graphs*, volume 438. Cambridge University Press, 2017.
- [84] H Francis Song and Xiao-Jing Wang. Simple, distance-dependent formulation of the watts-strogatz model for directed and undirected small-world networks. *Physical Review E*, 90(6):062801, 2014.
- [85] Duncan J Watts and Steven H Strogatz. Collective dynamics of "small-world" networks. *nature*, 393(6684):440, 1998.
- [86] MEJ Newman. MeJ newman, *siam rev.* 45, 167 (2003). *Siam Rev.*, 45:167, 2003.
- [87] Harry Sevi, Gabriel Rilling, and Pierre Borgnat. Multiresolution analysis of functions on directed networks. In *Wavelets and Sparsity XVII*, volume 10394, page 103941Q. International Society for Optics and Photonics, 2017.
- [88] Carlos S Kubrusly. Hilbert space operators. In *Hilbert Space Operators*, pages 13–22. Springer, 2003.
- [89] Gene H Golub and Charles F Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.
- [90] Yves Atchadé, Gersende Fort, Eric Moulines, and Pierre Priouret. Adaptive markov chain monte carlo: theory and methods. *Preprint*, 2009.
- [91] Xiaojin Zhu, John Lafferty, and Ronald Rosenfeld. *Semi-supervised learning with graphs*. PhD thesis, Carnegie Mellon University, language technologies institute, school of computer science Pittsburgh, PA, 2005.