



HAL
open science

Subgraph-based filterbanks for graph signals

Nicolas Tremblay, Pierre Borgnat

► **To cite this version:**

Nicolas Tremblay, Pierre Borgnat. Subgraph-based filterbanks for graph signals. *IEEE Transactions on Signal Processing*, 2016, 64, pp.3827-3840. 10.1109/tsp.2016.2544747 . hal-01243889

HAL Id: hal-01243889

<https://hal.science/hal-01243889>

Submitted on 15 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Subgraph-based filterbanks for graph signals

Nicolas Tremblay, Pierre Borgnat, *Member, IEEE*

Abstract—We design a critically-sampled compact-support biorthogonal transform for graph signals, via graph filterbanks. Instead of partitioning the nodes in two sets so as to remove one every two nodes in the filterbank downsampling operations, the design is based on a partition of the graph in connected subgraphs. Coarsening is achieved by defining one “supernode” for each subgraph and the edges for this coarsened graph derives from the connectivity between the subgraphs. Unlike the “one every two nodes” downsampling on bipartite graphs, this coarsening operation does not have an exact formulation in the graph Fourier domain. Instead, we rely on the local Fourier bases of each subgraph to define filtering operations. We apply successfully this method to decompose graph signals, and show promising performance on compression and denoising.

Index Terms—Graph signal processing, Laplacian pyramid, filterbanks, community detection, wavelet.

I. INTRODUCTION

Graphs are a modeling tool suitable to many applications involving networks, may they be social, neuronal, or driven from computer science, molecular biology [2]... Data on these graphs may be defined as a scalar (or vector) on each of its nodes, forming a so-called *graph signal* [3]. In a sense, a graph signal is the extension of the 1-D discrete classical signal (where the signal is defined on the circular graph, each node having exactly two neighbors) to any arbitrary discrete topology where each node may have an arbitrary number of neighbors. Temperature measured by a sensor network, age of the individuals in a social network, Internet traffic in a router network, etc. are all examples of such graph signals.

Adapting classical signal processing tools to signals defined on graphs has raised significant interests in the last few years [3], [4]. For instance, the graph Fourier transform, the fundamental building block of signal processing, has several possible definitions, either based on the diagonalisation of one of the Laplacian matrices [5], or based on Jordan’s decomposition of the adjacency matrix [6], [7]. Building upon this graph Fourier transform, authors have defined different sampling and interpolation procedures [8]–[11], windowed Fourier transform [12], graph empirical mode decomposition [13], different wavelet transforms, including spectral graph wavelets [5], [14], [15], diffusion wavelets [16], and wavelets defined via filterbanks [17]–[20]. Among the applications of graph signal processing, one may cite works on fMRI data [21], on multi-scale community detection [22], image compression [23], etc... In fact, such graph signal processing tools are general enough to deal with many types of irregular data [4].

Authors are at the Physics Laboratory, CNRS, ENS Lyon, University of Lyon 1, Lyon, France. Email: firstname.lastname@ens-lyon.fr. Work supported by the ANR-14-CE27-0001 GRAPHSIP grant. A preliminary approach of this work was presented at Asilomar Conference on Signals, Systems, and Computers 2015 [1]. Manuscript received XXX revised XXXX.

Graph filterbanks using downsampling have been initially defined for bipartite graphs [17] because: i) Bipartite graphs, by definition, contain two sets of nodes that are natural candidates for the sampling operations; ii) Downsampling followed by upsampling (which forces to zero the signal on one of the two sets of nodes) can be exactly written as a filter in the graph Fourier space. This enables to write exact anti-aliasing equations for the low-pass and high-pass filters to control the spectral folding phenomenon due to sampling [17]. However, for arbitrary graphs, one needs to decompose the graph in a (non-unique) sum of bipartite graphs [18], [19], [24], and analyze each of them separately. Another solution for arbitrary graph is based on downsampling according to the polarity of the graph Fourier mode of highest frequency [25].

We propose a significantly different way of defining filterbanks. Instead of trying to find an exact equivalent of both the decimation operator, hereafter (\downarrow), and a filtering operator C , we directly define a decimated filtering operator $L = (\downarrow)C$; following here the notations of [26], where L is not to be confused with the Laplacian operator of the graph, noted \mathcal{L} . Consider the 1-D straight-line graph where each node has two neighbors, and a partition of this graph in subgraphs of pairs of adjacent nodes. The classical Haar low-pass (resp. high-pass) channel samples one node per subgraph and defines on it the local average (resp. difference) of the signal. By analogy, we consider a partition of the graph in connected subgraphs, not necessarily of same size. Creating one “supernode” per subgraph, the low-pass channel (resp. high-pass channels) defines on it the local, i.e., over the subgraph, average (resp. differences) of the signal. The coarsened graphs on which those downsampled signals are defined, are then derived from the connectivity between the subgraphs: two supernodes are linked if there are edges between the associated subgraphs.

With this approach, we design a critically-sampled, compact-support biorthogonal filterbank that is valid for any partition in connected subgraphs. Depending on the application at hand, one has the choice on how to detect such partitions. For image approximation and denoising, an adequate way is to use a partition in communities, i.e. groups of nodes more connected with themselves than with the rest of the network [27]. This community structure is indeed linked to the low frequencies of graph signals [22], [28].

Section II recalls the definition of the graph Fourier transform we use. In Section III, after detailing the difficulties to extend classical filterbanks to graph signals, we discuss the state-of-the-art of graph filterbanks. The main contribution is in Section IV, first discussed as an analogy to the Haar filterbank, before presenting fully the proposed graph filterbank design. Section V proposes how to obtain a relevant partition in connected subgraph. Section VI shows applications, in compression and denoising. We conclude in Section VII.

II. THE GRAPH FOURIER TRANSFORM

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ be an undirected weighted graph with \mathcal{V} the set of nodes, \mathcal{E} the set of edges, and \mathbf{A} the weighted adjacency matrix such that $\mathbf{A}_{ij} = \mathbf{A}_{ji} \geq 0$ is the weight of the edge between nodes i and j . Let N be the total number of nodes. Let us define the graph's Laplacian matrix $\mathcal{L} = \mathbf{D} - \mathbf{A}$ where \mathbf{D} is a diagonal matrix with $\mathbf{D}_{ii} = \mathbf{d}_i = \sum_{j \neq i} \mathbf{A}_{ij}$ the strength of node i . \mathcal{L} is real symmetric, therefore diagonalizable: its spectrum is composed of $(\lambda_l)_{l=1 \dots N}$ its set of eigenvalues that we sort: $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_N$; and of \mathbf{Q} the matrix of its normalized eigenvectors: $\mathbf{Q} = (\mathbf{q}_1 | \mathbf{q}_2 | \dots | \mathbf{q}_N)$. Considering only connected graphs, the multiplicity of eigenvalue $\lambda_1 = 0$ is 1 [29]. By analogy to the continuous Laplacian operator whose eigenfunctions are the continuous Fourier modes and eigenvalues their squared frequencies, \mathbf{Q} is considered as the matrix of the graph's Fourier modes, and $(\sqrt{\lambda_l})_{l=1 \dots N}$ its set of associated "frequencies" [3]. For instance, the graph Fourier transform $\hat{\mathbf{x}}$ of a signal \mathbf{x} defined on the nodes of the graph reads: $\hat{\mathbf{x}} = \mathbf{Q}^\top \mathbf{x}$.

III. STATE OF THE ART

A. Classical 1-D filterbank and the Haar filterbanks

In classical setting, the decimation ($\downarrow 2$) operator by 2 is paramount. It keeps one every two nodes and follows what we call the "one every two nodes paradigm", as seen on Fig. 1a). The classical design of a filterbank is to find a set of operators, e.g. a low-pass filter \mathbf{C} and a high-pass filter \mathbf{D} , that combine well with decimation such that perfect recovery is possible from the decimated low- and high-pass filtered signals [26].

The usual Haar filterbank will be used as a leading example to expose the main issues encountered when attempting to extend filterbanks to graph signals. Consider the 1-D discrete signal \mathbf{x} of size N . Let us recall that, at the first level of the classical Haar filterbank, \mathbf{x} is decomposed into [26]:

- its approximation \mathbf{x}_1 of size $N/2$: $\mathbf{x}_1 = (\downarrow 2)\mathbf{C}\mathbf{x}$, where \mathbf{C} is the sliding average operator, here in matrix form:

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & \dots \\ 0 & 1 & 1 & 0 & \dots \\ 0 & 0 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (1)$$

- its detail \mathbf{x}_2 of size $N/2$: $\mathbf{x}_2 = (\downarrow 2)\mathbf{D}\mathbf{x}$, where \mathbf{D} is the sliding difference operator:

$$\mathbf{D} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 & 0 & 0 & \dots \\ 0 & -1 & 1 & 0 & \dots \\ 0 & 0 & -1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (2)$$

This Haar filterbank is orthogonal and critically sampled. Our objective is to generalize it to signals on arbitrary graphs.

B. Adapting filterbanks to graph signals

For graph signals, a key difficulty is the design of a suitable decimation operator, and it comes in two separate problems:

i) how to choose the nodes to keep?

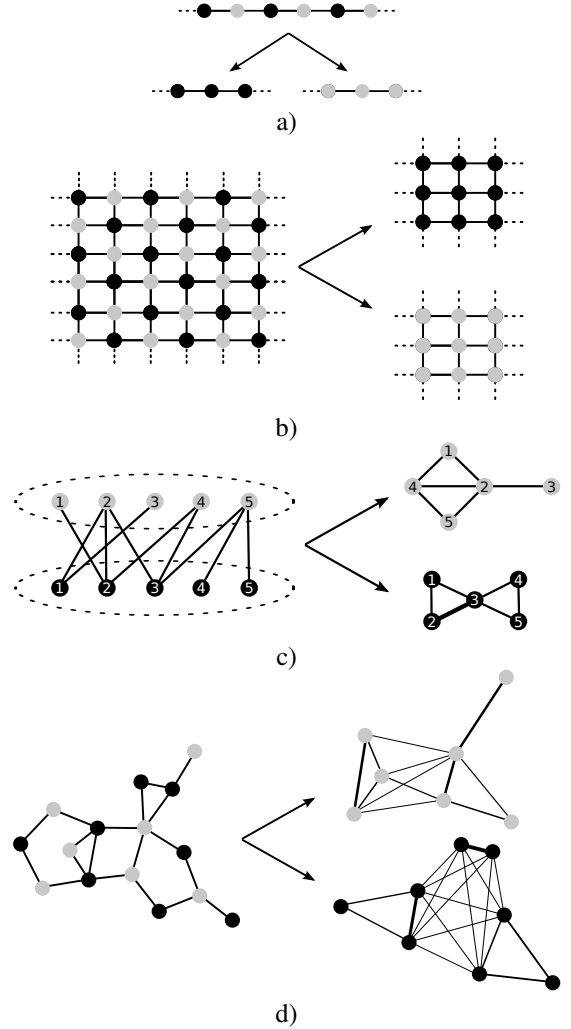


Fig. 1. State-of-the-art graph downsampling procedure following the "one every two nodes" paradigm (one set in black, the other in gray): a) (resp. b) the classical 1D (resp. 2D) downsampling; c) bipartite graph downsampling [17]; d) polarity downsampling from the highest frequency graph Fourier mode [25].

ii) how to wire together the nodes that are kept, so as to create the downsampled graph?

On a straight line or a regular grid, issue i) is solved by the one every two nodes paradigm and issue ii) does not exist: the structure after downsampling is exactly the same as the original (straight line or 2D grid); see Figs. 1 a) and b).

To tackle these issues, the following works propose a way to adapt the one every two nodes paradigm to arbitrary graphs.

1) *A design for signals defined on bipartite graphs:* Narang and Ortega [17], [24] consider first signals defined on bipartite graphs (i.e. two-colourable graphs). In this particular case, one may still downsample the graph by naturally keeping one every two nodes, as shown in Fig. 1c). For non-bipartite graphs, they develop a pre-processing step of the structure where the graph is decomposed in a sum of bipartite subgraphs, on which the filterbank is successively applied. Other methods to create bipartite graphs have been proposed, by oversampling [18], or with maximum spanning tree [19]. For issue ii), the downsampled structure has edges between two nodes if they have at least a common neighbor in the initial graph.

As seen in Fig. 1c), a downsampled bipartite graph is not necessarily bipartite and the pre-processing of the structure is mandatory to iterate the filterbank design. Still, an interesting property of this design is the specific behavior of bipartite graphs Laplacian's eigenvalues, that enable the authors to write specific anti-aliasing equations for the design of the filters \mathbf{C} and \mathbf{D} (see Eq. (13) of [17]) on bipartite graphs.

2) *A downsampling based on the Laplacian's last eigenvector*: Shuman et al. [25] focus on the eigenvector associated to the Laplacian's largest eigenvalue. They create two sets of nodes depending on this eigenvector's sign, as illustrated in Fig. 1d). According to the Fourier interpretation of the Laplacian's eigenbasis, the last eigenvector corresponds to the "highest frequency" of a graph signal. This idea, inspired by graph coloring studies [30], generalizes the fact that, for structured grids and bipartite graphs, the sign of this eigenvector does alternate every two nodes. To tackle issue ii), the authors in [25] rely on the Kron reduction [31] of the initial graph to obtain new graphs, and post-process them to remove links from otherwise very dense downsampled graphs, implying some degree of arbitrary choices.

In both cases, one sees that there is a part of arbitrariness in the pre- or post-processing steps to obtain suitable graphs and decimated graphs on which the filterbank can be cascaded.

IV. FILTERBANKS ON CONNECTED SUBGRAPHS

The idea we explore lets go of the "one every two nodes" sampling paradigm, and concentrate on graph coarsening: given a partition in connected subgraphs of the initial graph, the approximation and detail(s) will be obtained on each "supernode" that represent each connected subgraph. Hence we will not attempt to define separately analogies of down-sampling ($\downarrow 2$) and filtering \mathbf{C} and \mathbf{D} and, instead, we directly define analogies to graph signals of the decimated sliding average operator $\mathbf{L} = (\downarrow 2)\mathbf{C}$ and of the decimated sliding difference operator $\mathbf{B} = (\downarrow 2)\mathbf{D}$. They read for the Haar filterbanks:

$$\mathbf{L} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \in \mathbb{R}^{N/2 \times N} \quad (3)$$

and:

$$\mathbf{B} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & -1 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & -1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \in \mathbb{R}^{N/2 \times N}. \quad (4)$$

In Section IV-A, we first take a close look at the effect of these two Haar operators \mathbf{L} and \mathbf{B} on the input signal \mathbf{x} , to give insight in the fundamental analogy that is further formalized in Section IV-B to IV-D. In Section IV-E, we detail the analysis atoms created by the proposed filterbanks.

A. Introducing the design by analogy to the Haar filterbank

Let us rephrase the Haar filterbank from the proposed new point of view of operators on connected subgraphs.

1) *Replace decimation by partition*: Consider the 1-D classical signal \mathbf{x} of even size N defined on the straight line graph \mathcal{G} , of size N , where each node has two neighbors. We consider the partition \mathbf{c} of this graph in $K = N/2$ connected subgraphs $\{\mathcal{G}^{(k)}\}_{k \in \{1, K\}}$ connecting neighbors two-by-two: we call it the Haar partition, and it reads (when coded as a vector):

$$\mathbf{c} = (1, 1, 2, 2, 3, 3, \dots, K, K)^\top, \quad (5)$$

where $\mathbf{c}(i)$ is the label of node i 's subgraph.

2) *Interpret operators \mathbf{L} and \mathbf{B} in terms of local Fourier modes*: Consider subgraph $\mathcal{G}^{(k)}$ and $\mathbf{x}^{(k)}$ the restriction of \mathbf{x} to this subgraph. Define $\mathcal{G}^{(k)}$'s local adjacency matrix $\mathbf{A}^{(k)}$:

$$\forall k \in \{1, N/2\} \quad \mathbf{A}^{(k)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (6)$$

Its Laplacian matrix is diagonalisable with two local Fourier modes: $\mathbf{q}_1^{(k)\top} = \frac{1}{\sqrt{2}}(1, 1)$ of associated eigenvalue $\lambda_1^{(k)} = 0$ and $\mathbf{q}_2^{(k)\top} = \frac{1}{\sqrt{2}}(-1, 1)$ of associated eigenvalue $\lambda_2^{(k)} = 2$.

The actual effect of the operation $\mathbf{x}_1 = \mathbf{L}\mathbf{x}$ in Haar filterbank is to assign to each subgraph $\mathcal{G}^{(k)}$ the first local Fourier component of $\mathbf{x}^{(k)}$:

$$\forall k \in \{1, \dots, N/2\} \quad \mathbf{x}_1(k) = \mathbf{q}_1^{(k)\top} \mathbf{x}^{(k)}. \quad (7)$$

Similarly, the actual effect of the operation $\mathbf{x}_2 = \mathbf{B}\mathbf{x}$ may be rewritten as:

$$\forall k \in \{1, \dots, N/2\} \quad \mathbf{x}_2(k) = \mathbf{q}_2^{(k)\top} \mathbf{x}^{(k)}. \quad (8)$$

In other words $[\mathbf{x}_1(k) \ \mathbf{x}_2(k)]^\top$ is the local (reduced to $\mathcal{G}^{(k)}$) Fourier transform of $\mathbf{x}^{(k)}$.

3) *Analogy for graph signals*: Consider a graph \mathcal{G} and a partition \mathbf{c} of this graph in K connected subgraphs $\{\mathcal{G}^{(k)}\}_{k \in \{1, K\}}$. Consider one of these subgraphs $\mathcal{G}^{(k)}$ of size N_k . We propose the following fundamental analogy: consider the k -th component of the approximation and all detail signals $[\mathbf{x}_1(k) \ \mathbf{x}_2(k) \ \dots \ \mathbf{x}_{N_k}(k)]^\top$ as the local graph Fourier transform of $\mathbf{x}^{(k)}$, the graph signal reduced to $\mathcal{G}^{(k)}$. To this end, we diagonalize $\mathcal{G}^{(k)}$'s local Laplacian matrix to find its N_k eigenvectors (a.k.a. local Fourier modes) sorted w.r.t. their eigenvalues and compute the successive inner products.

4) *Graph support of the decimated components*: For Haar filterbank of 1-D signals, the two components are defined on straight-line graphs of size $N/2$. For arbitrary graphs, all subgraphs have not necessarily the same size. This implies that only subgraphs of size at least l contribute to \mathbf{x}_l . For instance, a subgraph $\mathcal{G}^{(k)}$ of three nodes will have three eigenvectors for its Laplacian and its first (resp. second, third) eigenvector will contribute to $\mathbf{x}_1(k)$ (resp. $\mathbf{x}_2(k)$, $\mathbf{x}_3(k)$).

The proposed analogy naturally defines the graphs on which are defined the downsampled signals \mathbf{x}_l . Let us introduce a supernode k standing for each subgraph $\mathcal{G}^{(k)}$. For $l = 1$, the approximation signal \mathbf{x}_1 lies naturally on a graph of adjacency matrix \mathbf{A}_1 where $\mathbf{A}_1(k, k')$ is the sum of the weights of the edges connecting subgraph k to subgraph k' in the original graph. Then, for the subsequent graph of adjacency matrix \mathbf{A}_l on which is defined the detail signal \mathbf{x}_l , only supernodes standing for subgraphs of size at least l are needed and \mathbf{A}_l is defined the same way by summing the edges between involved subgraphs. We formalize this analogy in the following.

B. Formalization of the operators necessary to the design

1) *Subgraph sampling operators*: Consider an arbitrary graph \mathcal{G} and an arbitrary partition \mathcal{c} of this graph in K connected subgraphs $\{\mathcal{G}^{(k)}\}_{k \in \{1, \dots, K\}}$. Write N_k the number of nodes in subgraph $\mathcal{G}^{(k)}$ and $\Gamma^{(k)} \subset \mathcal{V}$ the list of nodes in $\mathcal{G}^{(k)}$. For each subgraph $\mathcal{G}^{(k)}$, let us define the sampling operator $\mathbf{C}^{(k)} \in \mathbb{R}^{N \times N_k}$ such that:

$$\begin{aligned} \mathbf{C}^{(k)}(i, j) &= 1 \text{ if } \Gamma^{(k)}(j) = i, \\ &= 0 \text{ if not.} \end{aligned} \quad (9)$$

Applying $\mathbf{C}^{(k)\top}$ to a graph signal \mathbf{x} , one obtains the graph signal reduced to subgraph $\mathcal{G}^{(k)}$, i.e. $\mathbf{x}^{(k)}$. Conversely, applying $\mathbf{C}^{(k)}$ expands a signal defined on $\mathcal{G}^{(k)}$ to a graph signal on \mathcal{G} with zero-padding. It then turns out that the intra-subgraph adjacency matrix of $\mathcal{G}^{(k)}$, keeping only the links within $\mathcal{G}^{(k)}$ and noted $\mathbf{A}_{int}^{(k)} \in \mathbb{R}^{N_k^2}$, satisfies:

$$\forall k \in \{1, \dots, K\} \quad \mathbf{A}_{int}^{(k)} = \mathbf{C}^{(k)\top} \mathbf{A}_{int} \mathbf{C}^{(k)}. \quad (10)$$

As a consequence, the adjacency matrix \mathbf{A} may be written as the sum of two adjacency matrices:

- the intra-subgraph adjacency matrix, grouping the $\mathbf{A}_{int}^{(k)}$ in a block matrix keeping the edges within subgraphs:

$$\mathbf{A}_{int} = \sum_{k=1}^K \mathbf{C}^{(k)} \mathbf{C}^{(k)\top} \mathbf{A} \mathbf{C}^{(k)} \mathbf{C}^{(k)\top}. \quad (11)$$

- the inter-subgraph adjacency matrix $\mathbf{A}_{ext} = \mathbf{A} - \mathbf{A}_{int}$ that keeps only the links connecting subgraphs together.

2) *Subgraph Laplacian operators*: On each $\mathcal{G}^{(k)}$, let us define $\mathcal{L}_{int}^{(k)}$ the local Laplacian, computed from $\mathbf{A}_{int}^{(k)}$. It is diagonalisable:

$$\forall k \in \{1, \dots, K\} \quad \mathcal{L}_{int}^{(k)} = \mathbf{Q}^{(k)} \mathbf{\Lambda}^{(k)} \mathbf{P}^{(k)\top}, \quad (12)$$

with $\mathbf{\Lambda}^{(k)}$ the diagonal matrix of sorted eigenvalues ($\lambda_1^{(k)}$ is the smallest):

$$\mathbf{\Lambda}^{(k)} = \text{diag} \left(\lambda_1^{(k)}, \lambda_2^{(k)}, \dots, \lambda_{N_k}^{(k)} \right), \quad (13)$$

and $\mathbf{Q}^{(k)}$ the basis of local Fourier modes:

$$\mathbf{Q}^{(k)} = \left(\mathbf{q}_1^{(k)} | \mathbf{q}_2^{(k)} | \dots | \mathbf{q}_{N_k}^{(k)} \right). \quad (14)$$

and $\mathbf{P}^{(k)\top} = \left(\mathbf{Q}^{(k)} \right)^{-1}$ with:

$$\mathbf{P}^{(k)\top} = \left(\mathbf{p}_1^{(k)} | \mathbf{p}_2^{(k)} | \dots | \mathbf{p}_{N_k}^{(k)} \right)^\top. \quad (15)$$

We normalize the $\mathbf{q}_i^{(k)}$ with the L_p norm:

$$\forall i \in [1, N_k] \quad \|\mathbf{q}_i^{(k)}\|_p = \left(\sum_j |\mathbf{q}_i^{(k)}(j)|^p \right)^{1/p} = 1. \quad (16)$$

We discuss the choice of p (usually 1 or 2) in Section IV-E.

For each $\mathbf{q}_i^{(k)}$ of size N_k defined on the local subgraph $\mathcal{G}^{(k)}$, let $\bar{\mathbf{q}}_i^{(k)}$ be its zero-padded extension to the whole global graph:

$$\forall k \in \{1, K\} \quad \forall i \in \{1, N_k\} \quad \bar{\mathbf{q}}_i^{(k)} = \mathbf{C}^{(k)} \mathbf{q}_i^{(k)}. \quad (17)$$

Similarly $\bar{\mathbf{p}}_i^{(k)}$ stands for the zero-padded extension of $\mathbf{p}_i^{(k)}$.

Nota bene: we could have of course obtained the $\bar{\mathbf{q}}_i^{(k)}$ and $\bar{\mathbf{p}}_i^{(k)}$ by direct diagonalization of the Laplacian matrix of \mathbf{A}_{int} , but the introduced notations are used for the next definitions.

3) *Analysis, Synthesis and Group operators*: Considering the connected subgraph partition \mathcal{c} , define \tilde{N}_1 the maximum number of nodes in any subgraph:

$$\tilde{N}_1 = \max_k N_k. \quad (18)$$

For any $l \in \{1, \dots, \tilde{N}_1\}$, we note \mathcal{I}_l the list of subgraph labels containing at least l nodes:

$$\forall l \in \{1, \dots, \tilde{N}_1\} \quad \mathcal{I}_l = \{k \in \{1, K\} \text{ s.t } N_k \geq l\}. \quad (19)$$

For instance, as all subgraphs contain at least 1 node, \mathcal{I}_1 contains all the K subgraph labels. Also, necessarily: $|\mathcal{I}_1| \geq |\mathcal{I}_2| \geq \dots \geq |\mathcal{I}_K|$ (where $|\cdot|$ denotes the cardinality). By construction, we have that $\sum_l |\mathcal{I}_l| = N$.

The family of Analysis operators $\{\Theta_l \in \mathbb{R}^{N \times |\mathcal{I}_l|}\}$ are the \tilde{N}_1 operators that generalize \mathbf{L} and \mathbf{B} of the Haar filterbank:

$$\forall l \in \{1, \dots, \tilde{N}_1\} \quad \Theta_l = \left(\bar{\mathbf{q}}_l^{\mathcal{I}_l(1)} | \bar{\mathbf{q}}_l^{\mathcal{I}_l(2)} | \dots | \bar{\mathbf{q}}_l^{\mathcal{I}_l(|\mathcal{I}_l|)} \right). \quad (20)$$

This means that operator Θ_l groups together all local Fourier modes associated to the l -th eigenvalue of all subgraphs containing at least l nodes.

The family of \tilde{N}_1 Synthesis operators $\{\Pi_l \in \mathbb{R}^{N \times |\mathcal{I}_l|}\}$ read:

$$\forall l \in \{1, \dots, \tilde{N}_1\} \quad \Pi_l = \left(\bar{\mathbf{p}}_l^{\mathcal{I}_l(1)} | \bar{\mathbf{p}}_l^{\mathcal{I}_l(2)} | \dots | \bar{\mathbf{p}}_l^{\mathcal{I}_l(|\mathcal{I}_l|)} \right). \quad (21)$$

The family of Group operators $\{\Omega_l \in \mathbb{R}^{N \times |\mathcal{I}_l|}\}$ also contains \tilde{N}_1 operators defined as:

$$\begin{aligned} \forall l \in \{1, \dots, \tilde{N}_1\} \quad \Omega_l(i, j) &= 1 \text{ if } i \in \Gamma^{(\mathcal{I}_l(j))}, \\ &= 0 \text{ if not.} \end{aligned} \quad (22)$$

This means that Ω_l groups together indicator functions of subgraphs containing at least l nodes.

4) *On the operators' uniqueness*: Operators as we have defined them are not unique if no further rules are enforced. In fact, for each eigenvector $\mathbf{q}_i^{(k)}$, its opposite $-\mathbf{q}_i^{(k)}$ is also an eigenvector. Moreover, in the case of eigenvalue multiplicity, associated eigenvectors are not unique. To enforce uniqueness, any set of deterministic rules to extract eigenvectors will work. To solve the orientation issue, one may for instance decide to set the first non-zero coefficient of all vectors to be positive. For eigenvalues with multiplicity, we discuss a possible set of rules in Appendix A that guarantees uniqueness.

C. The filterbank design

1) *Analysis block*: Given a signal \mathbf{x} defined on the graph whose adjacency matrix is \mathbf{A} , one first needs to find a partition \mathcal{c} in connected subgraphs. The maximum number of nodes in any subgraph of \mathcal{c} is \tilde{N}_1 (see eq. (18)) and it determines the number of channels through which \mathbf{x} will be analyzed. Then, from \mathcal{c} , one constructs all operators Θ_l and Ω_l as described in Section IV-B. Finally, the signal \mathbf{x} defined on the graph is decomposed, through the \tilde{N}_1 channels, in \tilde{N}_1 signals:

$$\forall l \in \{1, \dots, \tilde{N}_1\} \quad \mathbf{x}_l = \Theta_l^\top \mathbf{x}, \quad (23)$$

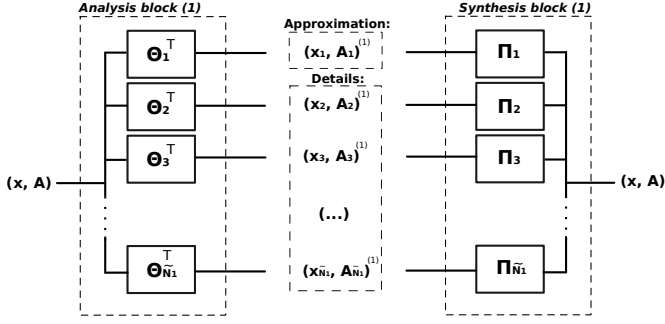


Fig. 2. Schematic representation of the analysis and synthesis blocks.

each of them defined on a graph whose adjacency matrix reads:

$$\mathbf{A}_l = \mathbf{\Omega}_l^\top \mathbf{A}_{ext} \mathbf{\Omega}_l. \quad (24)$$

By this formula, the convention is that there is no self-loop, i.e., for $k = k'$, $\mathbf{A}_l(k, k) = 0$. Also, adjacency matrices \mathbf{A}_l for $l \geq 2$ are only needed if one decides to cascade the filterbank on detail channels; here, the cascade will only be done on \mathbf{A}_1 (see Fig.3 and Section IV-D).

2) *A remark on the storage of structural information:* An important question is whether the total amount of stored information pre- and post-analysis equal or not? In terms of signal information only (i.e., discarding the structural information), the total amount of stored information is equal on both sides of the analysis block as each of the downsampled signals \mathbf{x}_l is of size $|\mathcal{I}_l|$ and $\sum_l |\mathcal{I}_l| = N$.

On the other hand, in terms of structural information (i.e. the information of the adjacency matrices), one needs to keep both the structural information pre- and post-analysis. Indeed, the post-analysis structural information is not enough to reconstruct the original graph (unlike the signal \mathbf{x} who can be perfectly reconstructed from its approximation and details, as we will see in Section IV-C3). The amount of stored structural information therefore increases after analysis and this is –at least for now– an irreducible storage price to pay. This is a common downfall of all graph filterbanks yet proposed, e.g. [17]–[19], [24], [25]. Finding ways to critically sample both the structure and the signal defined on it is part of our ongoing research and is not in the scope of this article.

In the following, all graph structures are stored in the form $\mathbf{A} = \mathbf{A}_{int} + \mathbf{A}_{ext}$, as this does not increase the amount of information (the number of links) but still subtly encodes the connected subgraph structure: indeed the partition \mathbf{c} can be exactly recovered by detecting the connected components of \mathbf{A}_{int} . In Narang et al. [17], [24], authors also need to keep an information equivalent to \mathbf{c} : the bipartite graph decomposition of \mathbf{A} and, for each bipartite graph, the information of the two sets of nodes. It is also the case in Shuman et al. [25]’s work, where authors need to keep the downsampling vector \mathbf{m} .

3) *Synthesis block:* One starts the synthesis with the list of signals $\{\mathbf{x}_l\}_{l \in \{1, \dots, \tilde{N}_1\}}$ and the structure information $\mathbf{A} = \mathbf{A}_{int} + \mathbf{A}_{ext}$. We show here that \mathbf{x} may be exactly recovered from that information. First of all, one extracts \mathbf{c} from \mathbf{A}_{int} , which in turn enables to compute all Synthesis operators $\{\mathbf{\Pi}_l\}$ following Sections IV-B2 and IV-B3. Moreover:

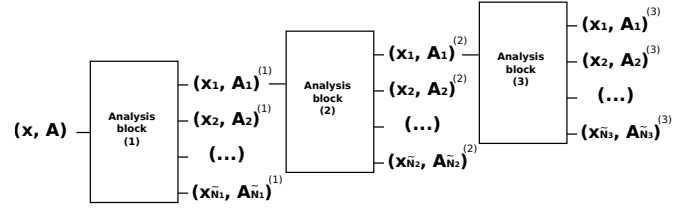


Fig. 3. The first three levels of the analysis cascade.

Lemma 1. *[Perfect reconstruction] \mathbf{x} is perfectly reconstructed from $\{\mathbf{x}_l\}_{l \in \{1, \dots, \tilde{N}_1\}}$ by applying successively each Synthesis operator to its corresponding channel:*

$$\sum_{l=1}^{\tilde{N}_1} \mathbf{\Pi}_l \mathbf{x}_l = \mathbf{x}. \quad (25)$$

Proof. Combining Eqs. (23) and (25), one has:

$$\sum_{l=1}^{\tilde{N}_1} \mathbf{\Pi}_l \mathbf{x}_l = \left(\sum_{l=1}^{\tilde{N}_1} \mathbf{\Pi}_l \mathbf{\Theta}_l^\top \right) \mathbf{x} = \left(\sum_{l=1}^{\tilde{N}_1} \sum_{j=1}^{|\mathcal{I}_l|} \mathbf{p}_l^{\mathcal{I}_l(j)} \mathbf{q}_l^{\mathcal{I}_l(j)\top} \right) \mathbf{x}. \quad (26)$$

$\mathbf{p}_l^{\mathcal{I}_l(j)} \mathbf{q}_l^{\mathcal{I}_l(j)\top}$ is a matrix of size $N \times N$, with non zero coefficients only for indices in subgraph $\mathcal{I}_l(j)$. In this non-zero block, it equals $\mathbf{p}_l^{\mathcal{I}_l(j)} \mathbf{q}_l^{\mathcal{I}_l(j)\top}$. One finally obtains:

$$\sum_{l=1}^{\tilde{N}_1} \mathbf{\Pi}_l \mathbf{\Theta}_l^\top = \begin{bmatrix} \mathbf{P}^{(1)} \mathbf{Q}^{(1)\top} & 0 & \dots & 0 \\ 0 & \mathbf{P}^{(2)} \mathbf{Q}^{(2)\top} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{P}^{(K)} \mathbf{Q}^{(K)\top} \end{bmatrix} = \mathbf{I}_N \quad (27)$$

the Identity, as $\mathbf{P}^{(k)\top} = (\mathbf{Q}^{(k)})^{-1}$, which ends the proof. \square

We show in Fig. 2 a schematic representation of the analysis and synthesis blocks of the proposed graph filterbanks.

4) *Critical sampling and biorthogonality:* Starting with data \mathbf{x} of size N , the analysis block provides vectors \mathbf{x}_l of size $|\mathcal{I}_l| \forall l \in \{1, \dots, \tilde{N}_1\}$. Storing all the \mathbf{x}_l accounts for $\sum_l |\mathcal{I}_l| = N$ values: the filterbank is critically sampled.

Furthermore, this filterbank is biorthogonal. Indeed, the analysis filter combining the \tilde{N}_1 channels may be written as $[\mathbf{\Theta}_1 \ \mathbf{\Theta}_2 \ \dots \ \mathbf{\Theta}_{\tilde{N}_1}]^\top$. Moreover, the overall synthesis filter reads $[\mathbf{\Pi}_1 \ \mathbf{\Pi}_2 \ \dots \ \mathbf{\Pi}_{\tilde{N}_1}]$. This filterbank is biorthogonal as we have:

$$[\mathbf{\Pi}_1 \ \mathbf{\Pi}_2 \ \dots \ \mathbf{\Pi}_{\tilde{N}_1}] [\mathbf{\Theta}_1 \ \mathbf{\Theta}_2 \ \dots \ \mathbf{\Theta}_{\tilde{N}_1}]^\top = \mathbf{I}_N,$$

as shown by Lemma 1. In the specific case where we choose $p = 2$ for the normalization of Eq. (16), we have $\mathbf{P}^{(k)} = \mathbf{Q}^{(k)}$, therefore $\mathbf{\Pi}_l = \mathbf{\Theta}_l$: in this case, the filterbank simply codes for an *orthogonal* transform.

D. The analysis cascade

Filterbanks are easily organised in cascade, each level of which contains the analysis and synthesis operators defined in Section IV-C. Given an original graph of adjacency matrix \mathbf{A} , a signal \mathbf{x} defined on it, and a partition \mathbf{c} of the graph, one

may obtain the \tilde{N}_1 Analysis operators $\{\Theta_l^{(1)}\}$ where the index in parenthesis stands for the level of the analysis cascade. The cascade's first level comprises of:

$$\forall l \in [1, \tilde{N}_1] \quad \mathbf{x}_l^{(1)} = \Theta_l^{(1)\top} \mathbf{x} \text{ and } \mathbf{A}_l^{(1)} = \Omega_l^{(1)\top} \mathbf{A}_{ext} \Omega_l^{(1)}.$$

For each of these channels, one may iterate the same analysis scheme, thereby obtaining successive approximations and details of the original signal at different scales of analysis. Classically, and we will follow this approach in the following, one iterates the analysis scheme only on the approximation signal at each level of the cascade, as shown on Fig. 3. Considering the approximation signal $\mathbf{x}_1^{(j)}$ at level (j) defined on the approximation graph coded by $\mathbf{A}_1^{(j)}$, and a partition $\mathbf{c}^{(j)}$ of this graph, one may obtain the \tilde{N}_{j+1} Analysis operators $\{\Theta_l^{(j+1)}\}$, that define the next scale of analysis. Similarly, we note $\{\Pi_l^{(j+1)}\}$ the associated Synthesis operators.

Note that the number of channels is not necessarily constant down the cascade. It is in fact adaptative: the number of channels \tilde{N}_j of the analysis block (j) depends on the maximal number of nodes contained in the subgraphs given by $\mathbf{c}^{(j)}$.

E. Atoms of analysis and the choice of normalization

To study the effect of the filterbank, one may look at the dictionary of analysis atoms, and of recovery atoms. With additional assumptions, it could be wavelets. However, we will avoid using the term wavelet and prefer the more general term of “atom”, as they do not necessarily have wavelet's properties: they are not related to translation on the graph, nor are they band-pass in the global graph Fourier domain (see Fig 8).

To each output of the analysis cascade (approximations and details at all levels) is associated an analysis atom. Approximation analysis atoms (“scaling function-like”) are associated to the first channel of each level: at level (j) , they are the columns of $\Theta_1^{(j)}$, interpolated back to the original graph's size:

$$\Phi^{(j)} = \Theta_1^{(1)} \times \dots \times \Theta_1^{(j-1)} \times \Theta_1^{(j)}. \quad (28)$$

Detail analysis atoms (“wavelet-like”) are obtained from all but the first channel at each level. More precisely at level (j) , the detail analysis atoms associated to channel $l \neq 1$ are the columns of $\Theta_l^{(j)}$, interpolated back to the original graph's size:

$$\Psi_l^{(j)} = \Theta_1^{(1)} \times \dots \times \Theta_1^{(j-1)} \times \Theta_l^{(j)}. \quad (29)$$

The choice of the L_p norm in Eq. (16) is now dictated by the wished properties of the atoms. A first possible choice is $p = 1$, i.e. normalization in L_1 , as it is the only normalization that ensures that the detail analysis atoms $\Psi_l^{(j)}$ have zero mean – a desirable feature to have atoms as close as possible to a wavelet interpretation. Another possibility would be to normalize in L_2 (as for the Haar filterbank). In this case, detail atoms do not have zero mean in general, however the energy of the modes is constant. In Section VI, the normalization will be application-dependent. The default normalization is with L_1 .

A property of the atoms is that their support is always compact: each is defined and non-zero only on one subgraph. On the other hand, in the global Fourier domain, the atoms

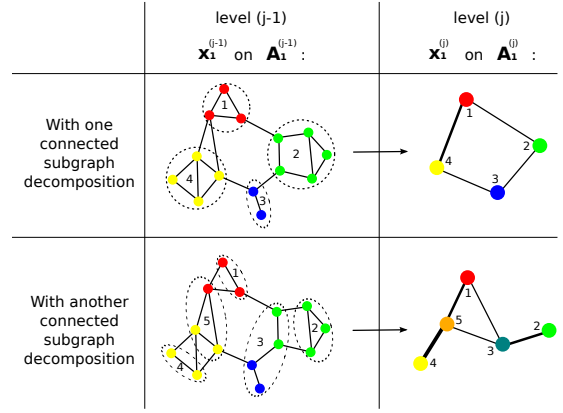


Fig. 4. Approximation $(\mathbf{x}_1^{(j)}, \mathbf{A}_1^{(j)})$ at level (j) of an analysis cascade of the signal-structure couple $(\mathbf{x}_1^{(j-1)}, \mathbf{A}_1^{(j-1)})$, given two different partitions in connected subgraphs (represented by the dotted lines). The signal is represented by colors on the nodes. Subgraph k in the original graph is represented by supernode k in the coarsened graph. As seen in Section IV, the signal on supernode k is the signal's average on the original subgraph k . Note the strong impact of the partition on the approximation signal.

are exactly localized only if the decomposition in subgraphs corresponds exactly to different connected components of the whole graph (in this case, the global Fourier matrix is the concatenation of all local Fourier matrices). If not, the further away is the graph from this disconnected model, the less localized are the atoms in the global Fourier domain.

V. DETECTING A PARTITION OF CONNECTED SUBGRAPHS

The proposed filterbank explicitly integrates the graph structure in connected subgraphs. A central question arises: how does one choose a particular partition \mathbf{c} of the graph in connected subgraphs? The partition choice has a strong influence on what the filterbank achieves, as shown in Fig. 4 where we compare the effect of downsampling for two different partitions on a toy graph. The practitioner has the choice among a wide variety of options to find such a partition: he could follow graph partitioning techniques of [32] or [33], or use graph nodal domains [34] – either very high frequency ones as in [25] or others — or any other solution... While the proposed filterbanks is well-defined for any of these partitions, the final decision regarding the partitioning algorithm will depend on what the user wants the filterbanks to achieve.

In the following, we show applications for compression and denoising. We seek to typically transform the original signal in a sparser one after analysis. For that, we look for partitions that separate the graph into groups of nodes more connected to themselves than with the rest of the graph: they are known as communities. Indeed, as in image or video compression, we suppose that low-frequencies contain the useful information of the signal. Approximating a community of nodes, each one with its signal value, by a supernode on which is the average over the community is a way to keep such low-frequencies.

A. Community detection procedure

Literature is abundant on community detection (see the survey [27]). To detect non-overlapping communities, we use the

greedy Louvain method [35]. It maximizes (approximately) over all the possible partitions \mathbf{c} , the so-called modularity (see [27]), a well-known objective function that measures the quality of a partition in communities \mathbf{c} , defined as:

$$Q(\mathbf{c}) = \frac{1}{2m} \sum_{ij} \left(\mathbf{A}_{ij} - \frac{d_i d_j}{2m} \right) \delta(\mathbf{c}_i, \mathbf{c}_j) \quad (30)$$

where $d_i = \sum_j \mathbf{A}_{ij}$ and $2m = \sum_i d_i$. The Louvain method iteratively repeats two main phases, starting from an initial situation where each node is in its own community: 1) Select a node and group it with its adjacent node that causes the largest increase of modularity; do this sequentially with all other nodes, until no individual move can improve the modularity; 2) Aggregate each community in a “supernode” and build a new adjacency matrix of this “supernode” graph. Phase 1 is then applied to this new graph, and so on and so forth. The algorithm stops when phase 1 is not able to increase the modularity anymore. We modify this algorithm and have two different implementations:

The SC (Small Communities) implementation. It consists in performing phase 1 only once: this implementation ensures that the partition separates the graph in small communities (typically smaller than 10 nodes).

The LC (Large Communities) implementation. When performing the usual algorithm, a stopping criterion is added: the algorithm is stopped (if not already stopped thanks to the first criterion) before the size of the largest community becomes larger than a given threshold τ . In fact, iterating both phases, communities become gradually larger; and recall that our proposal relies on the diagonalisation of the local Laplacian matrices, which has a cubic computation cost. In order to control computation time, we do not allow communities larger than the threshold, hereafter $\tau = 1000$ nodes.

For comparison, in Section VI-B2, we will show some results obtained with another famous multiscale community detection algorithm, called Infomap [36]. With this algorithm also, one may define analog SC and LC implementations.

Note on stochasticity: The Louvain and the Infomap algorithms are stochastic: they do not necessarily output the same partition at every run on the same data. This implies that the output of the analysis cascade of our filterbank may differ from a realisation to another (this is also the case of other methods such as the filterbanks based on bipartite graphs). Stochasticity is not an issue as synthesis operators are built according to the solutions found during the analysis: reconstruction is always perfect. For the results in Table I and Figures 10, 11 and 14, we show the median computed after 10 realisations.

B. Choice of adjacency matrix

When performing community detection, one may choose to use only the original adjacency matrix \mathbf{A} as it is, or incorporate some information about the graph signal \mathbf{x} to follow more closely its evolution. We explore two choices:

CoSub, short for Connected Subgraphs, is based on simply applying the Louvain algorithm on the adjacency matrix \mathbf{A} ;

EdAwCoSub, short for Edge Aware¹ Connected Subgraphs,

¹We use the term edge aware in relation to usage in the Signal processing community; the reader can think of it as “signal-adapted” if preferred.

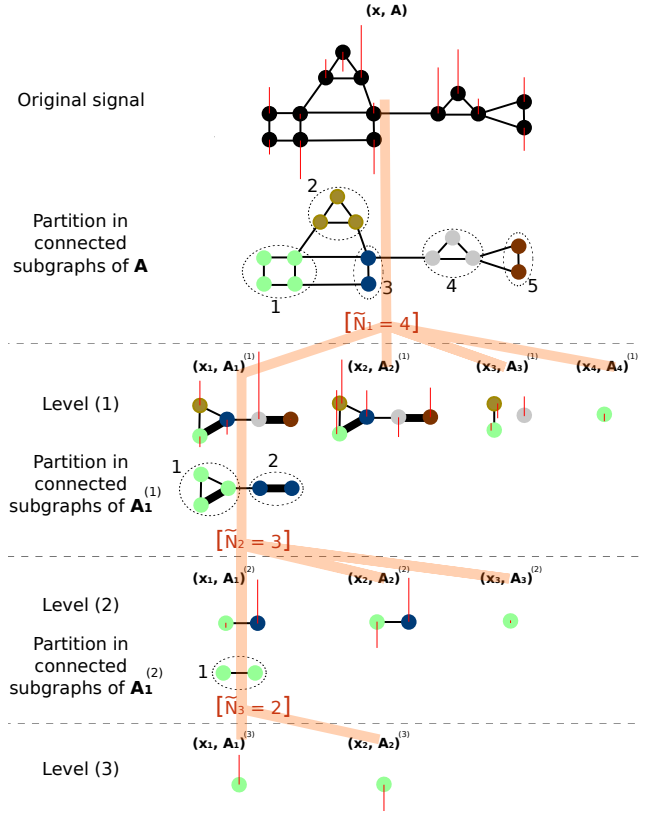


Fig. 5. An analysis cascade of a signal \mathbf{x} defined on a toy graph \mathbf{A} (top figure). On each node, the red vertical bar is proportional to the value of this node’s signal. Each edge’s width is proportional to its weight. Under the top figure is represented a partition in connected subgraph (see Section V); for clarity’s sake, one color is assigned to each subgraph. The largest subgraph contains $\tilde{N}_1 = 4$ nodes: the first level of the cascade therefore contains $\tilde{N}_1 = 4$ channels. At level (1), we represent the approximation $(\mathbf{x}_1, \mathbf{A}_1)^{(1)}$ and the three details $\{(\mathbf{x}_l, \mathbf{A}_l)^{(1)}\}_{l=2,3,4}$. For each of the coarsened graphs, the color of each supernode corresponds to the color of \mathbf{A} ’s subgraph it represents. We then iterate the analysis on the successive approximation signals, until level (3), where the approximation signal $(\mathbf{x}_1, \mathbf{A}_1)^{(3)}$ is reduced to a single node. The underlying orange tree is a guide to the eye down the analysis cascade. From the 6 detail and 1 approximation signals of each of its leaves, one may perfectly recover the original signal \mathbf{x} .

takes the signal \mathbf{x} into account for subgraph partitioning and modify the adjacency matrix into

$$\begin{aligned} \mathbf{A}_{\mathbf{x}}(i, j) &= e^{-\frac{(\mathbf{x}(i) - \mathbf{x}(j))^2}{2\sigma_x^2}} & \text{if } \mathbf{A}(i, j) \neq 0 \\ &= 0 & \text{if } \mathbf{A}(i, j) = 0 \end{aligned} \quad (31)$$

where $\sigma_x = \text{std}(\{|\mathbf{x}(i) - \mathbf{x}(j)|\}_{i \sim j})$ ($i \sim j$ means i neighbor to j in \mathbf{A}). This choice of σ_x is classical in the clustering literature [28]. The Louvain algorithm is then applied on $\mathbf{A}_{\mathbf{x}}$.

In both cases, the obtained partition enables us to decompose the original adjacency matrix in $\mathbf{A} = \mathbf{A}_{int} + \mathbf{A}_{ext}$. Edge-awareness could also be implemented, as in [23], by adapting classical image segmentation methods to graph signals; such an advanced comparison between edge-awareness methods is left for future work. In Section VI, we compare the 4 different implementations of the proposed filterbank: CoSub SC and LC, EdAwCoSub SC and LC; to methods from the literature.

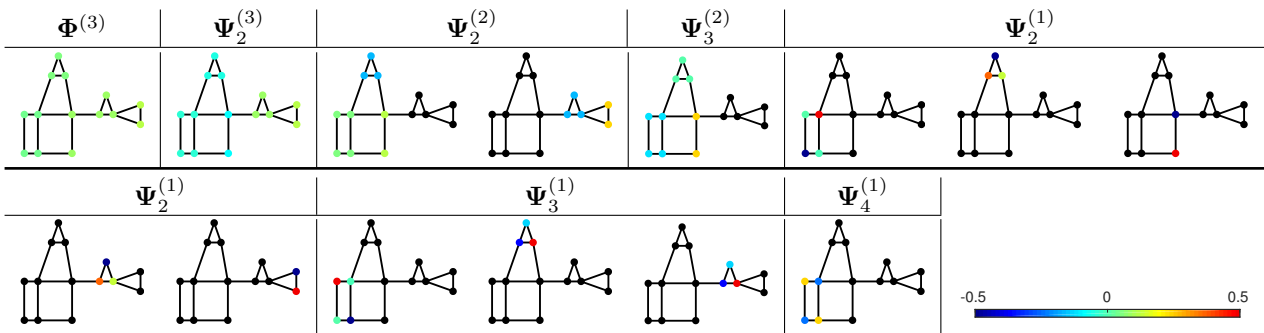


Fig. 6. The 14 analysis atoms of the proposed filterbank for the graph represented at the top of Fig. 5. The top left atom is the approximation atom (first channel at level (3) of the analysis cascade). Then, from left to right, and top to bottom, are listed the 13 detail atoms, from large to small scales. A node’s color represents the atom’s value on that node, as indicated by the colorbar. A black node corresponds to a *strictly* null value: indeed, atoms are compact-support.

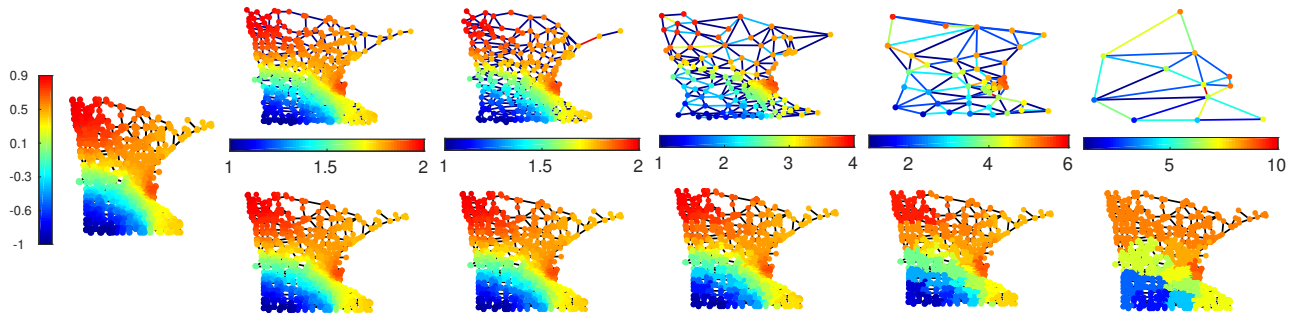


Fig. 7. Left: original smooth graph signal (sum of the first five global graph Fourier modes normalized by its maximum absolute value) defined on the Minnesota traffic graph. The vertical colorbar of this figure is valid for all graph signals represented on this figure. Top row: signal approximation $\mathbf{x}_1^{(l)}$ on the graph approximation $\mathbf{A}_1^{(l)}$ after each level (l) of the analysis cascade (with the Cosub SC implementation). The horizontal colorbar on the bottom of each figure corresponds to the weights of the links of the corresponding coarsened graph. Figures who do not have a bottom horizontal colorbar represent binary graphs. Lower row: for each of the successive approximations, the reconstructed graph signal is represented.

VI. APPLICATIONS

All the reported examples are computed using a developed Matlab toolbox that is available for download². The comparisons with methods from the literature use the implementations from the original authors, when they are available.

A. Three illustrative examples

1) *A signal defined on a simple toy graph:* Fig. 5 shows the analysis cascade on a toy signal defined on a very simple graph of size 14. Let us take a close look at the subgraph number 5 of the original graph: it contains two nodes and the signal on each of its node is of same absolute value but of opposite signs. As expected, $\mathbf{x}_1^{(1)}(5)$, the approximation signal on the corresponding supernode at level (1) is zero (average of the two original values); and $\mathbf{x}_2^{(1)}(5)$, the first detail signal is large: it is the difference between the two original values. Moreover, as this subgraph contains only 2 nodes, its Laplacian does not have a third eigenvector: this subgraph does not participate to the second and third detail signals and its associated supernode does not appear in $(\mathbf{x}_3, \mathbf{A}_3)^{(1)}$ nor $(\mathbf{x}_4, \mathbf{A}_4)^{(1)}$.

The analysis cascade decomposes the original signal in 3 detail signals (of sizes 5, 3 and 1) at level (1), 2 detail signals (of sizes 2 and 1) at level (2), 1 detail signal and one approximation signal (both of size 1) at level (3). From these 7 downsampled signals of total size 14, one may perfectly

reconstruct the original signal, using the synthesis operators defined in Section IV-C3.

In Fig. 6, we show the 14 analysis atoms corresponding to this analysis cascade: one approximation atom (from the approximation channel at the last level of the cascade); and 13 detail atoms that represent the other channels.

2) *An example of approximated graph signal:* Fig. 7 shows successive approximated signals $\{(\mathbf{x}_1, \mathbf{A}_1)^{(l)}\}_{l=1:5}$ of a smooth signal defined on the Minnesota traffic graph [24], using the CoSub SC implementation. Notice how the last level’s approximated signal, even if small in size (12 nodes) still captures the original signal’s information remarkably well.

3) *A small image:* Images can be studied as graph signals defined on the two-dimensional regular grid (each pixel is a node, and each node has four neighbors), and may therefore be analyzed by the proposed graph-based filterbank. Consider the small 32×32 image of Fig. 8: it is a graph signal defined on a regular graph of size 1024. Its graph Fourier transform is represented on the same figure. We analyze this image with the EdAwCoSub SC algorithm and the rest of Fig. 8 represents a selection of atoms of the filterbank, shown both in the node and the global graph Fourier domain, and partially reconstructed images from the projection of the original image on these atoms. Note that the support of the subgraphs are clearly impacted by edge-awareness. We see that each atom is compactly supported (in the node domain) and only (very) approximately localized in the global graph

²URL: http://perso.ens-lyon.fr/nicolas.tremblay/files/CoSubFB_toolbox.zip

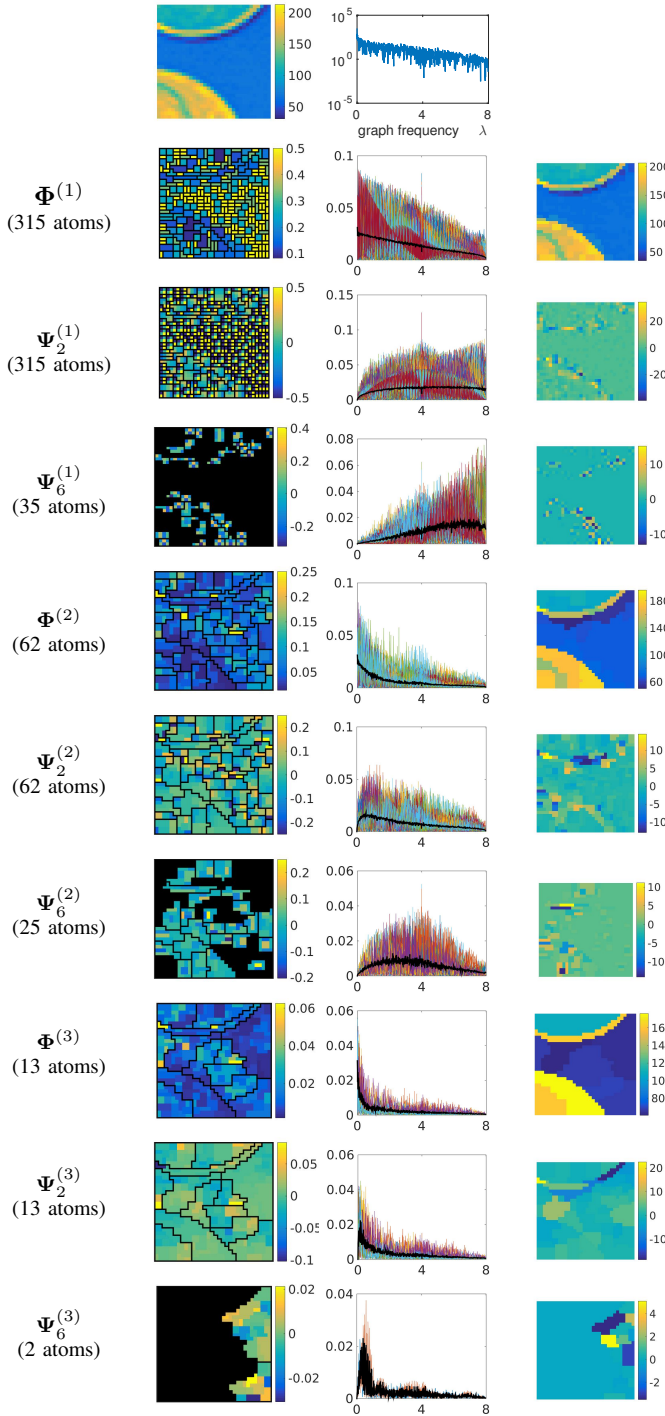


Fig. 8. Selected analysis atoms of the filterbank applied to a 32×32 regular 2 dimensional grid. Top line of figures: the top middle figure represents the graph Fourier transform of the image (top left) seen as a graph signal on the 2D grid. Other lines of figures: each line represents a chosen set of analysis atoms, corresponding to a given channel of the filterbank. For instance, the second line represents the atoms $\Phi^{(1)}$, i.e., the atoms corresponding to the first channel of the first level of the analysis cascade. The black lines on the first column's image represents the partition of the graph in 315 subgraphs, which is, in this case, the output of the Edge-Aware SC implementation of our filterbank. To each subgraph is associated a local atom, by construction, such that we can represent all of them on the same global image. The second column's figure represents all global Fourier transforms of all 315 atoms, and, superimposed with a thick black line, is their average. The third column represents the reconstruction of the original image if one keeps only the information that went through this analysis channel, and discards the rest.



Fig. 9. Benchmark images. From left to right: *cameraman*, *coins*, *synthetic*.

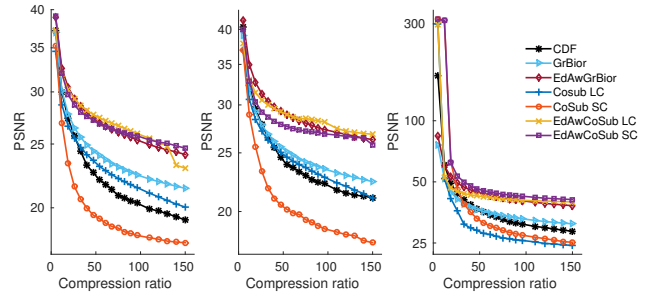


Fig. 10. Comparison of compression performance on the three benchmark images of Fig. 9 (from left to right: *cameraman*, *coins*, *synthetic*).

Fourier domain. Moreover, we see how, within a given level (j), the mean frequency of $\Psi_l^{(j)}$ increases as l increases. The cause of the frequency delocalisation is that regular grids are not decomposable in a sum of almost disconnected subgraphs: the local Fourier modes on which we base our design are therefore far from localized in the global Fourier domain. In fact, regular grids are a typical structure for which our method (and the graph partition in communities) is not very appropriate. Nevertheless, we still show results on images for pedagogical purposes and in order to compare performance with other methods from the literature.

B. Graph signal reconstruction via non-linear approximation

One of the use of classical filterbanks is compression. The main idea relies on the fact that natural signals are approximately smooth at different scales of analysis, and have therefore a sparse representation on filterbanks' atoms. One may thus transform the signal with the filterbanks, keep the low-pass coefficients and a fraction of high-pass coefficients while setting the others to zero, and still obtain a decent reconstruction of the original signal. In the following, we apply this non-linear approximation (NLA) scheme on images and on the Minnesota traffic graph.

Typical filterbank comparisons using NLA look at reconstruction results after three levels of analysis. In our case, as we do not know beforehand in how many subgraphs the partitioning algorithm will cut the graph, we cannot predict how many low-pass coefficients will be left after a given number of levels of the analysis cascade. Thus, for a comparison with other methods, and for a given compression ratio, we will compute all non-linear approximations corresponding to all different levels of the cascade, and keep the level for which the reconstruction result is the best.

Compr. ratio	5				26				54				96			
	lev	# LP	# HP	PSNR	lev	# LP	# HP	PSNR	lev	# LP	# HP	PSNR	lev	# LP	# HP	PSNR
CDF 9/7	3	1024	12083	37.1	3	1024	1497	25.7	4	256	958	22.0	4	256	427	20.3
GrBior	6	16	13091	36.9	5	64	2457	26.5	6	16	1198	24.0	6	16	667	22.4
CoSub LC	1	282	12825	34.6	1	274	2247	25.5	1	273	941	23.1	1	279	404	21.5
CoSub SC	3	2017	11090	35.3	4	282	2239	21.5	4	281	933	19.3	4	282	401	18.2
EdAwGrBior	5	64	13043	39.0	6	16	2505	29.3	6	16	1198	27.0	6	16	667	25.4
EdAwCoSub LC	1	478	12629	37.3	1	486	2035	29.3	1	486	728	27.4	1	486	197	25.9
EdAwCoSub SC	2	2584	10523	39.1	3	430	2091	28.7	3	422	792	26.8	3	439	244	25.7

TABLE I
COMPARISON OF THE COMPRESSION PERFORMANCE WITH OTHER METHODS ON THE BENCHMARK IMAGE *cameraman* SHOWN IN FIG. 9.

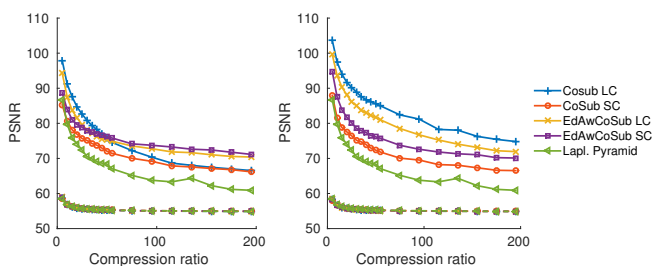


Fig. 11. Left: (full lines) compression results of the graph signal of Fig. 7, and (dotted lines) a Gaussian random signal of same energy defined on the same graph. Right: same comparison using the Infomap clustering algorithm rather than the Louvain algorithm to find partitions in connected subgraphs (see Section V-B). Results with the random signal are averaged over 10 realisations.

1) *Reconstruction of images*: Consider for instance the benchmark image *cameraman* shown in Fig. 9 (left). Its size is 256×256 ($N = 65536$ is the size of the associated graph signal). Table I compares the reconstruction details after NLA of CoSub LC and SC, EdAwCoSub LC and SC, to the classical image filterbank CDF 9/7, the Graph Bior filterbank [24] with Zero DC, *graphBior*(6,6) filters and Gain Compensation (GrBior); and the same Graph Bior filterbank but including edge-awareness [23] (EdAwGrBior). Also, Fig. 10 recaps the PSNR of reconstruction for each of the three benchmark images of Fig. 9. We see that our method, without edge-awareness, does not perform as well as GrBior. This is due to the fact that our method is not well suitable to 2D grids as they do not have a natural structure in communities and, on the contrary, GrBior is best suitable to bipartite graphs, of which 2D grids are an example. On the other hand, when adding edge-awareness, the graph becomes more structured and we obtain results similar to EdAwGrBior.

Note on the LC implementation. We see here that the best reconstruction for the LC implementation is always obtained after only one level of analysis. In this case, the filterbanks may hardly be seen as a multiscale analysis, but rather as a graph windowed Fourier transform. To observe a true multiscale analysis with the LC implementation, one needs to either decrease the threshold τ or increase the graph's size.

2) *Reconstruction performance for graph signals*: The graph signal model undermining the NLA scheme is that graph signals should be smooth with respect to the topology on which they are defined. For instance, let us consider the graph signal of Fig. 7 (left): it is by construction smooth with respect to the underlying graph as it is the sum of the first five eigenvectors of its Laplacian matrix. We compare

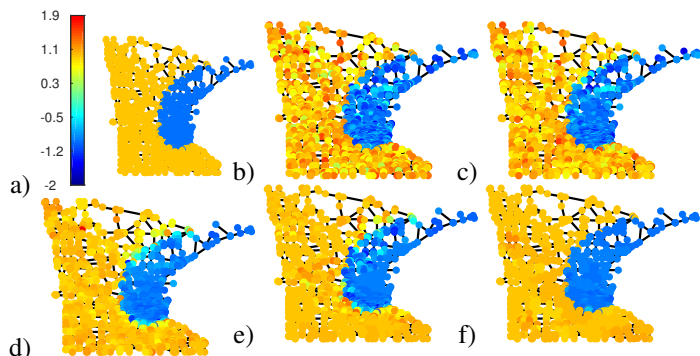


Fig. 12. a) Piece-wise constant signal on the Minnesota traffic graph (only two values: ± 1), and b) its corrupted version with an additive Gaussian noise of standard deviation $\sigma = 1/4$. The four other figures are denoised signals after a one-level analysis and hard-thresholding all high-pass coefficients with $T = 3\sigma$. c) GrBior; d) Laplacian pyramid; e) CoSub LC; f) EdAwCoSub LC.

in Fig. 11 (left, full lines) the reconstruction performance for our filterbank implementations, to Shuman's Laplacian pyramid filterbanks [25], with respect to the compression ratio. The dotted lines on the same Figure show the performances for a random Gaussian signal of zero mean and variance 1, normalized to have the same energy as the smooth signal. As expected, random signals are dense on the proposed analysis atoms, and reconstruction is comparatively poor. Moreover, we see that our proposed filterbanks really have an edge for signals who are smooth compared to the community structure of the underlying graph. On the right of Fig. 11 are represented the performances obtained with the Infomap algorithm rather than the Louvain algorithm (see Section V-B). In this particular case, performances with Infomap are better. Empirically, we find that using the Louvain algorithm or the Infomap algorithm yields in general similar results.

Note on the typical size of communities: In this Minnesota example (resp. *cameraman* example), the typical community size of the first level of the cascade is 2 (resp. 2) for CoSub SC, 5 (resp. 5) for EdAwCoSub SC, 80 (resp. 200) for CoSub LC, and 40 (resp. 100) for EdAwCoSub LC.

C. Application in denoising, on the Minnesota traffic graph

Another application of filterbanks is denoising. We consider first a piece-wise constant graph signal (that has only two possible values: +1 and -1) defined on the Minnesota traffic graph, as shown in Fig. 12a; so as to compare our proposition with previously published methods. We corrupt this signal with an additive Gaussian noise of standard deviation σ . Fig. 12b

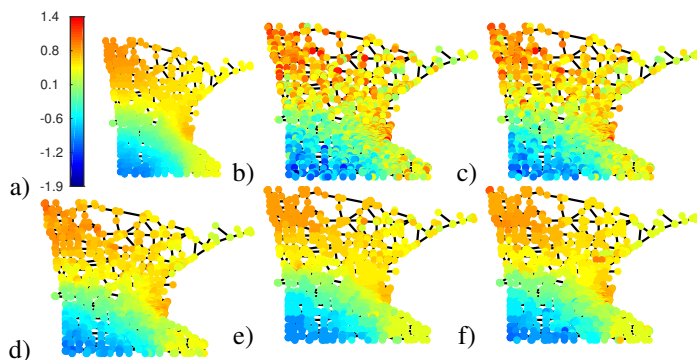


Fig. 13. a) Smooth signal on the Minnesota traffic graph (the same as in Fig. 7), and b) its corrupted version with an additive Gaussian noise of standard deviation $\sigma = 1/4$. The four other figures are denoised signals after a one-level analysis and hard-thresholding all high-pass coefficients with $T = 3\sigma$. c) GrBior; d) Laplacian Pyramid; e) CoSub LC; f) EdAwCoSub LC.

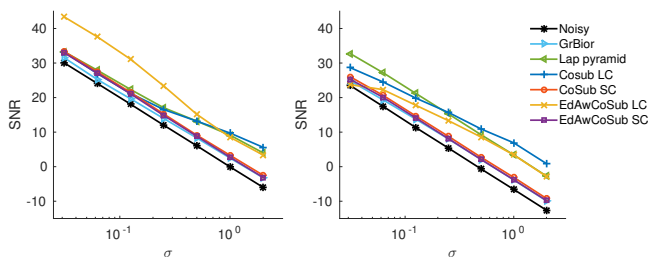


Fig. 14. Comparison of the denoising performance for (left) the piecewise constant signal of Fig. 12a and (right) the smooth signal of Fig. 13a; vs. the standard deviation of a Gaussian corrupting noise. Both signals are normalized such that their maximum absolute value is one. Results are averaged over ten realisations of the corrupting noise.

shows such a corrupted signal with $\sigma = 1/4$. We then attempt to recover the original image by computing the first level of the analysis cascade, and reconstructing the signal from all low-pass coefficients and thresholded high-pass coefficients having absolute value higher than $T = 3\sigma$. In order for such a thresholding scheme to be justified for denoising, the energy of coefficients associated to noise should have a constant variance in all the details after analysis. For that, we use here a L_2 normalization of the local Fourier modes (rather than L_1) for this denoising experiment (see the discussion in Section IV-E).

Fig. 12 compares results obtained with Graph Bior filterbanks (to our knowledge, the Edge-Aware Graph Bior algorithm is not implemented for graphs different than the 2D grid) and with the Laplacian Pyramid filterbanks to our proposition of filterbanks, for $\sigma = 1/4$. Fig. 14 (left) summarizes SNR results for different values of σ . These results may be compared to those obtained by Sakiyama et al. and summarized in Table 5 of [18]. We study also the denoising on the smooth signal of Fig. 7. Results are shown in Fig. 13 for $\sigma = 1/4$ and summarized in Fig. 14 (right) for different values of σ .

All four of our implementations outperform GrBior. Moreover, CoSub LC and the Laplacian pyramid obtain similar results; our method performing slightly better at high compression rates. Edge-awareness helps in the case of the piecewise constant signal and not so much for the smooth signal.

VII. CONCLUSION

While previous methods are based on global filters defined in the global Fourier space of the graph, we defined local filters based on the local Fourier spaces of each connected subgraph. Thanks to this paradigm, a simple form of filterbanks is designed, that one could call Haar graph filterbank: the k -th value of the downsampled approximation (resp. details) is the average (resp. differences) of the signal on subgraph k .

We first illustrated this for compression on images, mainly for pedagogical and state-of-the-art comparison purposes. In fact, without edge-awareness, our proposition is not really good for such regular structures. Edge-awareness, on the other hand, by giving structure to the network raises performance to the state-of-the-art. The improvement over existing methods becomes truly apparent for irregular graphs for which a community structure exists. Existence of communities is a very common, if not universal, property of real-world graphs; and our filterbanks rely on this particular organization of complex networks. For such graphs, our proposition outperforms existing ones on non-linear approximation experiments and equals state-of-the-art on denoising experiments.

Within this framework, future work will concentrate on extending the local filters to more sophisticated filters, and on finding ways to critically sample jointly the graph structure and the graph signal defined on it.

APPENDIX

A. A proposition for uniqueness of the operators

To enforce uniqueness of the graph Fourier basis in the case of eigenvalue λ having multiplicity $m > 1$, a possible rule can be set as follows. Consider the first vector of this eigenspace. All we know is its orthogonality with vectors of all other eigenspaces, i.e. $N - m$ vectors. We decide to arbitrarily force its last m coefficients to zero, and then find the unique $N - m$ coefficients that respects orthogonality with other known vectors and proper normalization. Note that, if at least one of the vectors of the other eigenspaces have non-zero values only on these last m coefficients, we then look for the set of m coefficients closest possible to the last one such that uniqueness is guaranteed. For the second vector, it has to be orthogonal to the already decided $N - m + 1$ vectors: we arbitrarily force its $m - 1$ last coefficients to zero and find the unique set of its set of coefficients thanks to orthogonality. And so on and so forth up to the multiplicity m .

Note that, for practical implementations, classical functions for eigenvector computation (for instance `eig` or `svd` in Matlab) empirically output the same choice of eigenvectors when run on two *exactly* identical inputs, even when there are eigenvalues with multiplicity.

REFERENCES

- [1] N. Tremblay and P. Borgnat, "Joint filtering of graph and graph-signals," in *proceedings of the Asilomar Conference on Signals, Systems, and Computers*, Nov 2015.
- [2] M. Newman, *Networks: an introduction*. Oxford University Press, 2010.
- [3] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 83–98, May 2013.

- [4] A. Sandryhaila and J. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *Signal Processing Magazine, IEEE*, vol. 31, no. 5, pp. 80–90, Sept 2014.
- [5] D. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.
- [6] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs," *Signal Processing, IEEE Transactions on*, vol. 61, no. 7, pp. 1644–1656, April 2013.
- [7] —, "Discrete signal processing on graphs: Graph fourier transform," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 6167–6170.
- [8] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *CoRR*, vol. abs/1510.00297, 2015. [Online]. Available: <http://arxiv.org/abs/1510.00297>
- [9] S. Chen, R. Varma, A. Sandryhaila, and J. Kovacevic, "Discrete signal processing on graphs: Sampling theory," *Signal Processing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.
- [10] X. Wang, P. Liu, and Y. Gu, "Local-set-based graph signal reconstruction," *Signal Processing, IEEE Transactions on*, vol. 63, no. 9, pp. 2432–2444, May 2015.
- [11] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *CoRR*, vol. abs/1511.05118, 2015. [Online]. Available: <http://arxiv.org/abs/1511.05118>
- [12] D. Shuman, B. Ricaud, and P. Vandergheynst, "A windowed graph fourier transform," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, Aug 2012, pp. 133–136.
- [13] N. Tremblay, P. Borgnat, and P. Flandrin, "Graph empirical mode decomposition," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, Sept 2014, pp. 2350–2354.
- [14] D. Shuman, C. Wiesmeyr, N. Holighaus, and P. Vandergheynst, "Spectrum-adapted tight graph wavelet and vertex-frequency frames," *Signal Processing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.
- [15] N. Leonardi and D. Van De Ville, "Tight wavelet frames on multislice graphs," *Signal Processing, IEEE Transactions on*, vol. 61, no. 13, pp. 3357–3367, July 2013.
- [16] R. Coifman and M. Maggioni, "Diffusion wavelets," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 53–94, 2006.
- [17] S. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *Signal Processing, IEEE Transactions on*, vol. 60, no. 6, pp. 2786–2799, June 2012.
- [18] A. Sakiyama and Y. Tanaka, "Oversampled graph Laplacian matrix for graph filter banks," *Signal Processing, IEEE Transactions on*, vol. 62, no. 24, pp. 6425–6437, Dec 2014.
- [19] H. Nguyen and M. Do, "Downsampling of signals on graphs via maximum spanning trees," *Signal Processing, IEEE Transactions on*, vol. 63, no. 1, pp. 182–191, Jan 2015.
- [20] V. Ekambaram, G. Fanti, B. Ayazifar, and K. Ramchandran, "Critically-sampled perfect-reconstruction spline-wavelet filterbanks for graph signals," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, Dec 2013, pp. 475–478.
- [21] H. Behjat, N. Leonardi, L. Sornmo, and D. Van De Ville, "Canonical cerebellar graph wavelets and their application to fmri activation mapping," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, Aug 2014, pp. 1039–1042.
- [22] N. Tremblay and P. Borgnat, "Graph wavelets for multiscale community mining," *Signal Processing, IEEE Transactions on*, vol. 62, no. 20, pp. 5227–5239, Oct 2014.
- [23] S. Narang, Y. H. Chao, and A. Ortega, "Graph-wavelet filterbanks for edge-aware image processing," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, Aug 2012, pp. 141–144.
- [24] S. Narang and A. Ortega, "Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs," *Signal Processing, IEEE Transactions on*, vol. 61, no. 19, pp. 4673–4685, Oct 2013.
- [25] D. I. Shuman, M. J. Faraji, and P. Vandergheynst, "A framework for multiscale transforms on graphs," *CoRR*, vol. abs/1308.4942, 2013. [Online]. Available: <http://arxiv.org/abs/1308.4942>
- [26] G. Strang and T. Nguyen, *Wavelets and filter banks*. SIAM, 1996.
- [27] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3-5, pp. 75–174, 2010.
- [28] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [29] F. Chung, *Spectral graph theory*. Amer Mathematical Society, 1997, no. 92.
- [30] B. Aspvall and J. R. Gilbert, "Graph coloring using eigenvalue decomposition," *SIAM Journal on Algebraic Discrete Methods*, vol. 5, no. 4, pp. 526–538, 1984. [Online]. Available: <http://dx.doi.org/10.1137/0605051>
- [31] F. Dorfler and F. Bullo, "Kron reduction of graphs with applications to electrical networks," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 60, no. 1, pp. 150–163, Jan 2013.
- [32] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 359–392, 1998. [Online]. Available: <http://dx.doi.org/10.1137/S1064827595287997>
- [33] S.-H. Teng, "Coarsening, sampling, and smoothing: Elements of the multilevel method," in *Algorithms for Parallel Processing*, ser. The IMA Volumes in Mathematics and its Applications. Springer New York, 1999, vol. 105, pp. 247–276.
- [34] R. Band, I. Oren, and U. Smilansky, "Nodal domains on graphs - how to count them and why?" *Analysis on Graphs and its applications Proc. Symp. Pure Math.*, 2008.
- [35] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.
- [36] M. Rosvall and C. T. Bergstrom, "Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems," *PLoS ONE*, vol. 6, no. 4, p. e18209, 04 2011.