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, Institute of Electrical and Electronics Engineers, 2016, 64, pp.3827-3840. 10.1109/tsp.2016.2544747. hal-01243889

HAL Id: hal-01243889
https://hal.archives-ouvertes.fr/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 multiscale 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].

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 (↓), and a filtering operator $\mathcal{C}$, we directly define a decimated filtering operator $L = (\downarrow) \mathcal{C}$; 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.

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.
II. THE GRAPH FOURIER TRANSFORM

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ be a 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 $A_{ij} = A_{ji} \geq 0$ 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 $\mathbf{L} = \mathbf{D} - \mathbf{A}$ where $\mathbf{D}$ is a diagonal matrix with $D_{ii} = d_i = \sum_{j \neq i} A_{ij}$ the strength of node $i$. $\mathbf{L}$ is real symmetric, therefore diagonalizable: its spectrum is composed of $(\lambda_l)_{l=1..N}$ its set of eigenvalues that we sort: $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_N$; and of $\mathbf{Q}$ the matrix of its normalized eigenvectors: $\mathbf{Q} = (q_1|q_2|\ldots|q_N)$. 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 $(\sqrt{\lambda_l})_{l=1..N}$ its set of associated “frequencies” [3]. For instance, the graph Fourier transform $\hat{x}$ of a signal $x$ defined on the nodes of the graph reads: $\hat{x} = \mathbf{Q}^\top x$.

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 $C$ and a high-pass filter $D$, 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 $x$ of size $N$. Let us recall that, at the first level of the classical Haar filterbank, $x$ is decomposed into [26]:

- its approximation $x_1$ of size $N/2$: $x_1 = (\downarrow 2)Cx$, where $C$ is the sliding average operator, here in matrix form:

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

- its detail $x_2$ of size $N/2$: $x_2 = (\downarrow 2)Dx$, where $D$ is the sliding difference operator:

$$D = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 & 0 & 0 & \ldots \\ 0 & -1 & 1 & 0 & \ldots \\ 0 & 0 & -1 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$  \hfill (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?

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 downsampling 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 \( C \) and \( D \) (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, by graph coloring studies [30], generalizes the fact that, “highest frequency” of a graph signal. This idea, inspired by graph coloring studies [30], generalizes the fact that, “highest frequency” of a graph signal. This idea, inspired by graph coloring studies [30], generalizes the fact that, “highest frequency” of a graph signal.

In Section IV-A, we first take a close look at the effect of these downsampled graphs on which the filterbank can be cascaded.

The idea we explore lets go of the “one every two nodes” property of this design is the specific behavior of bipartite graphs 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, “highest frequency” of a graph signal. This idea, inspired by graph coloring studies [30], generalizes the fact that, “highest frequency” of a graph signal. This idea, inspired by graph coloring studies [30], generalizes the fact that, “highest frequency” of a graph signal. This idea, inspired by graph coloring studies [30], generalizes the fact that, “highest frequency” of a graph signal.

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

\[
e = (1, 1, 2, 2, 3, 3, \ldots, K, K) \top,
\]

where \( c(i) \) is the label of node \( i \)’s subgraph.

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

\[
\forall k \in \{1, N/2\} \quad A(k) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.
\]

Its Laplacian matrix is diagonalisable with two local Fourier modes: \( q_{1}^{(k)} ) = \frac{1}{\sqrt{2}} (1, -1) \) of associated eigenvalue \( \lambda_{1}^{(k)} = 0 \) and \( q_{2}^{(k)\top} = \frac{1}{\sqrt{2}} (-1, 1) \) of associated eigenvalue \( \lambda_{2}^{(k)} = 2 \).

The actual effect of the operation \( x_1 = Lx \) in Haar filterbank is to assign to each subgraph \( G'(k) \) the first local Fourier component of \( x(k) \) :

\[
\forall k \in \{1, \ldots, N/2\} \quad x_1(k) = q_{1}^{(k)\top} x(k).
\]

Similarly, the actual effect of the operation \( x_2 = Bx \) may be rewritten as :

\[
\forall k \in \{1, \ldots, N/2\} \quad x_2(k) = q_{2}^{(k)\top} x(k).
\]

In other words \( x_1(k) \) \( x_2(k) \) is the low (reduced to \( G'(k) \)) Fourier transform of \( x(k) \).

3) Analogy for graph signals: Consider a graph \( G \) and a partition \( e \) of this graph in \( K \) connected subgraphs \( \{ G'(k) \} _ {k \in \{1,K\}} \) Consider one of these subgraphs \( G'(k) \) of size \( N_k \). We propose the following fundamental analogy: consider the \( k \)-th component of the approximation and all detail signals \( x_1(k) \) \( x_2(k) \) \( x_{N_k}(k) \) as the local graph Fourier transform of \( x(k) \), the graph signal reduced to \( G'(k) \). To this end, we diagonalize \( G'(k) \)'s local Laplacian matrix to find its \( N_k \) eigenvectors (u.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 \( x_l \). For instance, a subgraph \( G'(k) \) of three nodes will have have three eigenvectors for its Laplacian and its first (resp. second, third) eigenvector will contribute to \( x_1(k) \) (resp. \( x_2(k) \), \( x_3(k) \)).

The proposed analogy naturally defines the graphs on which are defined the downsampled signals \( x_1 \). Let us introduce a supernode \( k \) standing for each subgraph \( G'(k) \). For \( l = 1 \), the approximation signal \( x_1 \) lies naturally on a graph of adjacency matrix \( A_1 \) where \( A_1(k, k') \) 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 \( A_l \) on which is defined the detail signal \( x_l \), only supernodes standing for subgraphs of size at least \( l \) are needed and \( A_l \) 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 $c$ of this graph in $K$ connected subgraphs $\{\mathcal{G}(k)\}_{k \in \{1, \ldots, K\}}$. 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 $C(k) \in \mathbb{R}^{N \times N_k}$ such that:

$$C(k)(i, j) = 1 \text{ if } \Gamma(k)(j) = i,$$

$$= 0 \text{ if not.} \quad (9)$$

Applying $C(k)^T$ to a graph signal $x$, one obtains the graph signal reduced to subgraph $\mathcal{G}(k)$, i.e. $x(k)$. Conversely, applying $C(k)$ 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 $A_{\text{int}}(k) \in \mathbb{R}^{N_k \times N_k}$, satisfies:

$$\forall k \in \{1, \ldots, K\} \quad A_{\text{int}}(k) = C(k)^T A_{\text{int}} C(k). \quad (10)$$

As a consequence, the adjacency matrix $A$ may be written as the sum of two adjacency matrices:

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

$$A_{\text{int}} = \sum_{k=1}^{K} C(k){C(k)}^T A C(k){C(k)}^T. \quad (11)$$

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

2) Subgraph Laplacian operators: On each $\mathcal{G}(k)$, let us define $L_{\text{int}}(k)$ the local Laplacian, computed from $A_{\text{int}}(k)$. It is diagonalisable:

$$\forall k \in \{1, \ldots, K\} \quad L_{\text{int}}(k) = Q(k) \Lambda(k) P(k)^T, \quad (12)$$

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

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

and $Q(k)$ the basis of local Fourier modes:

$$Q(k) = \begin{bmatrix} q_{1}(k) & q_{2}(k) & \cdots & q_{N_k}(k) \end{bmatrix}. \quad (14)$$

and $P(k)^T = (Q(k))^{-1}$ with:

$$P(k)^T = \begin{bmatrix} p_{1}(k) & p_{2}(k) & \cdots & p_{N_k}(k) \end{bmatrix}^T. \quad (15)$$

We normalize the $q_i(k)$ with the $L_p$ norm:

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

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

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

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

Similarly $\tilde{p}_i(k)$ stands for the zero-padded extension of $p_i(k)$.

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

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

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

For any $l \in \{1, \ldots, \hat{N}_1\}$, we note $I_l$ the list of subgraph labels containing at least $l$ nodes:

$$\forall l \in \{1, \ldots, \hat{N}_1\} \quad I_l = \{ k \in \{1, K\} \text{ s.t } N_k \geq l \}. \quad (19)$$

For instance, as all subgraphs contain at least 1 node, $I_1$ contains all the $K$ subgraph labels. Also, necessarily: $|I_l| \geq |I_2| \geq \cdots \geq |I_K|$ (where $|.|$ denotes the cardinality). By construction, we have that $\sum_l |I_l| = N$.

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

$$\forall l \in \{1, \ldots, \hat{N}_1\} \quad \Theta_l = \begin{bmatrix} \tilde{q}_1^T(l) \tilde{q}_2^T(l) \cdots \tilde{q}_{|I_l|}^T(l) \end{bmatrix}. \quad (20)$$

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

The family of $\hat{N}_1$ Synthesis operators $\{\Pi_l \in \mathbb{R}^{N \times |I_l|}\}$ read:

$$\forall l \in \{1, \ldots, \hat{N}_1\} \quad \Pi_l = \begin{bmatrix} \tilde{p}_1^T(l) \tilde{p}_2^T(l) \cdots \tilde{p}_{|I_l|}^T(l) \end{bmatrix}. \quad (21)$$

The family of Group operators $\{\Omega_l \in \mathbb{R}^{N \times |I_l|}\}$ also contains $\hat{N}_1$ operators defined as:

$$\forall l \in \{1, \ldots, \hat{N}_1\} \quad \Omega_l(i, j) = 1 \text{ if } i \in \Gamma(I_l(j)), \quad (22)$$

$$= 0 \text{ if not.}$$

This means that $\Omega_l$ 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 $\tilde{q}_i(k)$, its opposite $-\tilde{q}_i(k)$ 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 $x$ defined on the graph whose adjacency matrix is $A$, one first needs to find a partition $c$ in connected subgraphs. The maximum number of nodes in any subgraph of $c$ is $\hat{N}_1$ (see eq. (18)) and it determines the number of channels through which $x$ will be analyzed. Then, from $c$, one constructs all operators $\Theta_l$ and $\Omega_l$ as described in Section IV-B. Finally, the signal $x$ defined on the graph is decomposed, through the $\hat{N}_1$ channels, in $\hat{N}_1$ signals:

$$\forall l \in \{1, \ldots, \hat{N}_1\} \quad x_l = \Theta_l^T x, \quad (23)$$

Similarly $\tilde{x}_l$ stands for the zero-padded extension of $x_l$. 

For any $l \in \{1, \ldots, \hat{N}_1\}$, $x_l$ is the zero-padded extension of $x_l$.
each of them defined on a graph whose adjacency matrix reads:

\[ A_l = \Omega_l^T A_{ext} \Omega_l. \]  

By this formula, the convention is that there is no self-loop, i.e., for \( k = k' \), \( A_l(k, k) = 0 \). Also, adjacency matrices \( A_l \) 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 \( A_1 \) (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 \( x_l \) is of size \( |I_l| \) and \( \sum_l |I_l| = N \).

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 \( x \) 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 and this is –at least for now– an irreducible storage price to pay.

In the following, all graph structures are stored in the form \( A = A_{int} + A_{ext} \), as this does not amount the number of links but still subsumes the connected subgraph structure: indeed the partition \( c \) can be exactly recovered by detecting the connected components of \( A_{int} \). In Narang et al. [17], [24], authors also need to keep an equivalent information to \( c \): the bipartite graph decomposition of \( A \) 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 \( m \).

3) Synthesis block: One starts the synthesis with the list of signals \( \{x_l\}_{l \in \{1, N_1\}} \) and the structure information \( A = A_{int} + A_{ext} \). We show here that \( x \) may be exactly recovered from that information. First of all, one extracts \( c \) from \( A_{int} \), which in turn enables to compute all Synthesis operators \( \{\Pi_l\} \) following Sections IV-B2 and IV-B3. Moreover:

**Lemma 1.** Perfect reconstruction \( x \) is perfectly reconstructed from \( \{x_l\}_{l \in \{1, \ldots, N_1\}} \) by applying successively each Synthesis operator to its corresponding channel:

\[ \sum_{l=1}^{N_1} \Pi_l x_l = x. \]

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

\[ \sum_{l=1}^{N_1} \Pi_l x_l = (\sum_{l=1}^{N_1} \Pi_l \Theta_l^T) x = (\sum_{l=1}^{N_1} \Pi_l \Theta_l^T) (\sum_{j=1}^{N_1} p_l^{T(j)} q_l^{T(j)^T}) x. \]

where \( \Theta_l^T \) is a matrix of size \( N \times N \), with non zero coefficients only for indices in subgraph \( I_l(j) \). In this non-zero block, it equals \( p_l^{T(j)} q_l^{T(j)^T} \). One finally obtains:

\[ \sum_{l=1}^{N_1} \Pi_l \Theta_l^T = \begin{bmatrix} P^{(1)} Q^{(1)^T} & 0 & \cdots & 0 \\ 0 & P^{(2)} Q^{(2)^T} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P^{(K)} Q^{(K)^T} \end{bmatrix} = I_N \]

the identity, as \( P^{(k)^T} = (Q^{(k)})^{-1} \), which ends the proof.

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

![Fig. 2. Schematic representation of the analysis and synthesis blocks.](image)

4) Critical sampling and biorthogonality: Starting with data \( x \) of size \( N \), the analysis block provides vectors \( x_l \) of size \( |I_l| \) \( \forall l \in \{1, \ldots, N_1\} \). Storing all the \( x_l \) accounts for \( \sum_l |I_l| = N \) values: the filterbank is critically sampled.

Furthermore, this filterbank is biorthogonal. Indeed, the analysis filter combining the \( N_1 \) channels may be written as \( [\Theta_1 \ \Theta_2 \ \cdots \ \Theta_{N_1}]^T \). Moreover, the overall synthesis filter reads \( [\Pi_1 \ \Pi_2 \ \cdots \ \Pi_{N_1}]^T \). This filterbank is biorthogonal as we have:

\[ [\Pi_1 \ \Pi_2 \ \cdots \ \Pi_{N_1}] [\Theta_1 \ \Theta_2 \ \cdots \ \Theta_{N_1}]^T = I_N, \]

as shown by Lemma 1. In the specific case where we choose \( p = 2 \) for the normalization of Eq. (16), we have \( P^{(k)^T} = Q^{(k)} \), therefore \( \Pi_l = \Theta_l \); 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 \( A \), a signal \( x \) defined on it, and a partition \( c \) of the graph, one
may obtain the $\tilde{N}_1$ Analysis operators $\{\Theta_j^{(1)}\}$ where the index in parenthesis stands for the level of the analysis cascade. The cascade’s first level comprises of:

$$\forall l \in [1, \tilde{N}_1] \quad x_l^{(1)} = \Theta_1^{(1)\top} x \quad \text{and} \quad A_l^{(1)} = \Omega_k^{(1)\top} A_{ext} \Omega_k^{(1)}.$$  

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 $x_1^{(j)}$ at level $(j)$ defined on the approximation graph coded by $A_1^{(j)}$, and a partition $c^{(j)}$ of this graph, one may obtain the $\tilde{N}_{j+1}$ Analysis operators $\{\Theta^{(j+1)}\}$, that define the next scale of analysis. Similarly, we note $\{\Pi^{(j+1)}\}$ the associated Synthesis operators.

Note that the number of channels is not necessarily constant down the cascade. It is in fact adaptive: 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 $c^{(j)}$.

### 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:

$$\Psi^{(j)} = \Theta_1^{(1)} \times \ldots \times \Theta_1^{(j-1)} \times \Theta_1^{(j)}.$$  

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:

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

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 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 $c$ 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 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 \( c \), the so-called modularity (see [27]), a well-known objective function that measures the quality of a partition in communities \( c \), defined as:

\[
Q(c) = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{d_i d_j}{2m} \right) \delta(c_i, c_j)
\]

where \( d_i = \sum_j A_{ij} \) 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 \( \tau \). 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 \( A \) as it is, or incorporate some information about the graph signal \( x \) 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 \( A \):

EdAwCoSub, short for Edge Aware Connected Subgraphs,

\footnote{We 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.}


takes the signal \( x \) into account for subgraph partitioning and modify the adjacency matrix into:

\[
A_x(i,j) = \begin{cases} 
\frac{(x_i-x_j)^2}{2\sigma_x^2} & \text{if } A(i,j) \neq 0 \\
0 & \text{if } A(i,j) = 0
\end{cases}
\]

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

In both cases, the obtained partition enables us to decompose the original adjacency matrix in \( A = \text{A}_{\text{int}} + \text{A}_{\text{ext}}. \) 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.
VI. APPLICATIONS

All the reported examples are computed using a developed Matlab toolbox that is available for download\(^2\). 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, \(x_1^{(1)}(5)\), the approximation signal on the corresponding supernode at level (1) is zero (average of the two original values); and \(x_2^{(1)}(5)\), 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 \((x_3, A_3)^{(1)}\) nor \((x_4, A_4)^{(1)}\).

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 \(\{x_1, A_1^{(l)}\}_{l=1:5}\) 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 \(32 \times 32\) 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

Fig. 8. Selected analysis atoms of the filterbank applied to a $32 \times 32$ 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 represents the reconstruction of the original image if one keeps only the information that went through this analysis channel, and discards the rest.

Fourier domain. Moreover, we see how, within a given level ($j$), the mean frequency of $\Psi^{(j)}_l$ 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.
1) Reconstruction of images: Consider for instance the benchmark image cameraman shown in Fig. 9 (left). Its size is $256 \times 256$ ($N = 65536$ 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 $\tau$ or increase the graph’s size.

2) Reconstruction performance for graph signals: The graph signal model underlying 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 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 provides a much better approach.

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 $\sigma$. Fig. 12b

![Graph Image](image-url)

**Table I**

Comparison of the compression performance with other methods on the benchmark image cameraman shown in Fig. 9.
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. 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 $\sigma$. 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 $\sigma$.

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 piece-wise 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 $\lambda$ 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


