Towards Signal Processing on Graphs

December 19, 2013

1 Summary

Many modern problems involve solving computational problems on large scale datasets. In recent years, it has become common to model the relationships between data points – the geometry of a data set – via very large graphs. A graph is a fundamental, and therefore well-studied, mathematical structure to represent collections of abstract objects connected by links. The strength of graphs is that objects can be anything and their links, edges in the graphs, can represent virtually any notion of affinity or distance. Recently, the prevalence of graphs in data driven science has been fuelled up by algorithmic and software engineering progresses, such as the availability of efficient graph databases and graph-based computational abstractions. Manipulating data on such structures in an efficient way and with principled approaches offering performance guarantees is therefore key to reaping the benefits of Big Data.

In this project, we propose to lay the foundations for a generalisation of signal processing to data defined on large graphs. Our approach is based on using spectral graph theory, the spectral theory of the graph Laplacian, to revisit computational harmonic analysis on graphs. In particular, we will revisit the fundamental aspects of signal processing via a new definition of localised filtering on graphs. We will study the fundamental limits of the new theory and seek to revitalise the link between signal processing and learning theory via new interpolation and regularisation frameworks based on generalised notions of sparsity and smoothness for data on graphs. We will explore the asymptotics of these frameworks towards manifold-supported distributions. Finally, we will leverage recent computational abstractions and frameworks to design efficient algorithms that scale up to very large graphs.

The corpus of our results will pave the way towards efficiently analysing and processing datasets, in very much the same way we process signals and images today, thereby opening new computational avenues for data driven science and technology.

2 State of the Art

2.1 Spectral Graph Theory

A weighted graph $\mathcal{G} = \{E, V, w\}$ consists of a set of vertices V, a set of edges E, and a weight function $w : E \to \mathbb{R}^+$ which assigns a positive weight to each edge. We consider here only finite graphs where $|V| = N < \infty$. The adjacency matrix A for a weighted graph $\mathcal{G}$ is the $N \times N$ matrix with entries $a_{m,n}$ where

$$a_{m,n} = \begin{cases} w(e) \text{ if } e \in E \text{ connects vertices } m \text{ and } n \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

In this proposal, we first consider undirected graphs, which correspond to symmetric adjacency matrices. We do not consider the possibility of negative weights. The degree of each vertex is the sum of the weights of all the edges incident to it. We define the degree matrix D to have diagonal elements equal to the degrees, and zeros elsewhere. We will also write $f \in \mathbb{R}^N$, a function on the vertices of the graph, and $f(m)$ for the value on the mth vertex.

The non-normalized Laplacian is defined as $\mathcal{L} = D - A$. It can be verified that for any $f \in \mathbb{R}^N$, $\mathcal{L}$ satisfies

$$(\mathcal{L}f)(m) = \sum_{m \sim n} a_{m,n} \cdot (f(m) - f(n)), \tag{2}$$

where the sum over $m \sim n$ indicates summation over all vertices n that are connected to the vertex m, and $a_{m,n}$ denotes the weight of the edge connecting m and n. For a graph arising from a regular mesh, the graph Laplacian corresponds to the standard stencil approximation of the continuous Laplacian, intuitively justifying this terminology. Note that there exist several slightly different definitions, but most of them will still fit the framework defined here.

As the graph Laplacian $\mathcal{L}$ is a real symmetric matrix, it has a complete set of orthonormal eigenvectors. We denote these by χ_ℓ for $\ell = 0, \ldots, N-1$, with associated eigenvalues λ_ℓ

$$\mathcal{L}\chi_\ell = \lambda_\ell \chi_\ell. \tag{3}$$

As $\mathcal{L}$ is symmetric, each of the λ_ℓ are real. For the graph Laplacian, it can be shown that the eigenvalues are all non-negative, and that 0 appears as an eigenvalue with multiplicity equal to the number of connected components of the graph [1]. Henceforth, we assume the graph $\mathcal{G}$ to be connected, we may thus order the eigenvalues such that

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 ... \leq \lambda_{N-1} \tag{4}$$

For any function $f \in \mathbb{R}^N$ defined on the vertices of $\mathcal{G}$, its graph Fourier transform $\hat{f}$ is defined by

$$\hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{n=1}^{N} \chi_\ell^*(n) f(n). \tag{5}$$

2

where we adopt the convention that the inner product be conjugate-linear in the first argument. The inverse transform reads as

$$f(n) = \sum_{\ell=0}^{N-1} \hat{f}(\ell)\chi_\ell(n). \tag{6}$$

2.2 Spectral Graph Wavelets

The spectral graph wavelet transform, introduced recently in [**?**], is generated by wavelet operators that are operator-valued functions of the Laplacian. Let us summarize the main steps of this construction. The transform is determined by the choice of a kernel function $g : \mathbb{R}^+ \to \mathbb{R}^+$, which is analogous to the Fourier transform of a wavelet in the classical setting. This kernel g should behave as a band-pass filter, i.e. it satisfies $g(0) = 0$ and $\lim_{x\to\infty} g(x) = 0$. The wavelet operator $T_g = g(\mathcal{L})$ is defined through its action on a given function f as :

$$\widehat{T_g f}(\ell) = g(\lambda_\ell)\hat{f}(\ell). \tag{7}$$

The wavelet operators at scale t is then defined by $T_g^t = g(t\mathcal{L})$. It should be emphasized that even though the "spatial domain" for the graph is discrete, the domain of the kernel g is continuous and thus the scaling may be defined for any positive real number t.

The spectral graph wavelets are then realized through localizing these operators by applying them to the impulse on a single vertex, i.e.

$$\psi_{t,n} = T_g^t \delta_n. \tag{8}$$

Expanding this explicitly in the graph domain shows

$$\psi_{t,n}(m) = \sum_{\ell=0}^{N-1} g(t\lambda_\ell)\chi_\ell^*(n)\chi_\ell(m). \tag{9}$$

Formally, the wavelet coefficients of a given function f are produced by taking the inner product with these wavelets, as

$$\begin{aligned} W_f(t,n) &= \langle \psi_{t,n}, f \rangle \\ &= \left(T_g^t f\right)(n) \\ &= \sum_{\ell=0}^{N-1} g(t\lambda_\ell)\hat{f}(\ell)\chi_\ell(n). \end{aligned}$$

It is convenient to introduce a second class of waveforms, analogous to the lowpass residual scaling functions from classical wavelet analysis. They are determined by a single real valued function $h : \mathbb{R}^+ \to \mathbb{R}$, which acts as a lowpass filter, and satisfies $h(0) > 0$ and $h(x) \to 0$ as $x \to \infty$. By properly discretizing

the scale variable, one obtains frame of spectral graph wavelets. More precisely, the lower and upper frame bounds, A and B are given by :

$$A = \min_{\lambda \in [0, \lambda_{N-1}]} G(\lambda), \ B = \max_{\lambda \in [0, \lambda_{N-1}]} G(\lambda)$$

where $G(\lambda) = h^2(\lambda) + \sum_j g(t_j \lambda)^2$.

An important property of spectral graph wavelets is that their localization at small scales is guaranteed by simple constrains on the kernel g. This property ensures that graph wavelets behave like discrete multiscale differential operators on graphs and thus can sparsely approximate piecewise smooth signals. A precise mathematical formulation of the links between smoothness on graphs and sparsity in graph wavelet frames will be one of the topics covered by this proposal.

2.3   Towards Signal Processing on Graphs

For signals $f, g \in L^2(\mathbb{R})$, the classical convolution product $h = f * g$ is defined as

$$h(t) = (f * g)(t) := \int_{\mathbb{R}} f(\tau) g(t - \tau) d\tau. \tag{10}$$

Since the simple translation $g(t - \tau)$ cannot be directly extended to the graph setting, we cannot directly generalize (10). However, the classical convolution product also satisfies

$$h(t) = (f * g)(t) = \int_{\mathbb{R}} \hat{h}(k) \psi_k(t) dk = \int_{\mathbb{R}} \hat{f}(k) \hat{g}(k) \psi_k(t) dk, \tag{11}$$

where $\psi_k(t) = e^{2\pi i k t}$. This important property that convolution in the time domain is equivalent to multiplication in the Fourier domain is the notion we generalize instead. Specifically, by replacing the complex exponentials in (11) with the graph Laplacian eigenvectors, we define a *generalized convolution* of signals $f, g \in \mathbb{R}^N$ on a graph by

$$(f * g)(n) := \sum_{\ell=0}^{N-1} \hat{f}(\ell) \hat{g}(\ell) \chi_\ell(n). \tag{12}$$

Proposition 1 *The generalized convolution product defined in* (12) *satisfies the following properties:*

1. *Generalized convolution in the vertex domain is multiplication in the graph spectral domain:*

$$\widehat{f * g} = \hat{f} \hat{g}. \tag{13}$$

2. *Let $\alpha \in \mathbb{R}$ be arbitrary. Then*

$$\alpha(f * g) = (\alpha f) * g = f * (\alpha g). \tag{14}$$

4

3. *Commutativity:*

$$f * g = g * f. \tag{15}$$

4. *Distributivity:*

$$f * (g + h) = f * g + f * h. \tag{16}$$

5. *Associativity:*

$$(f * g) * h = f * (g * h). \tag{17}$$

6. *Define a function $g_0 \in \mathbb{R}^N$ by $g_0(n) := \sum_{\ell=0}^{N-1} \chi_\ell(n)$. Then g_0 is an identity for the generalized convolution product:*

$$f * g_0 = f. \tag{18}$$

7. *An invariance property with respect to the graph Laplacian (a difference operator):*

$$\mathcal{L}(f * g) = (\mathcal{L}f) * g = f * (\mathcal{L}g). \tag{19}$$

8. *The sum of the generalized convolution product of two signals is a constant times the product of the sums of the two signals:*

$$\sum_{n=1}^{N}(f * g)(n) = \frac{1}{\sqrt{N}}\left[\sum_{n=1}^{N} f(n)\right]\left[\sum_{n=1}^{N} g(n)\right] = \sqrt{N}\hat{f}(0)\hat{g}(0). \tag{20}$$

Now the application of the classical translation operator T_u to a function $f \in L^2(\mathbb{R})$ can be seen as a convolution with δ_u:

$$(T_u f)(t) := f(t - u) = (f * \delta_u)(t) \overset{(11)}{=} \int_{\mathbb{R}} \hat{f}(k)\widehat{\delta_u}(k)\psi_k(t)dk = \int_{\mathbb{R}} \hat{f}(k)\psi_k^*(u)\psi_k(t)dk,$$

where the equalities are in the weak sense. Thus, for any signal $f \in \mathbb{R}^N$ defined on the the graph $\mathcal{G}$ and any $i \in \{1, 2, \ldots, N\}$, we also define a *generalized translation operator* $T_i : \mathbb{R}^N \to \mathbb{R}^N$ via generalized convolution with a delta centered at vertex i:

$$(T_i f)(n) := \sqrt{N}(f * \delta_i)(n) \overset{(12)}{=} \sqrt{N} \sum_{\ell=0}^{N-1} \hat{f}(\ell)\chi_\ell^*(i)\chi_\ell(n). \tag{21}$$

The translation (21) is a kernelized operator. The window to be shifted around the graph is defined in the graph spectral domain via the kernel or filter $\hat{f}(\cdot)$. To translate this window to vertex i, the ℓth component of the kernel is multiplied by $\chi_\ell^*(i)$, and then an inverse graph Fourier transform is applied. As an example, in Figure 1, we apply generalized translation operators to the normalized heat kernel. We can see that doing so has the desired effect of shifting a window around the graph, centering it at any given vertex i.

Some expected properties of the generalized translation operator follow immediately from the generalized convolution properties of Proposition 1.
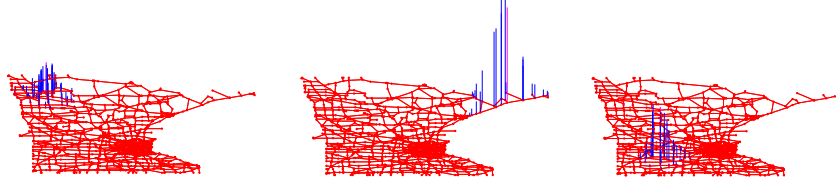
5

Figure 1: Translation operator applied to the heat kernel of the Minnessota graph. The resulting window is sharply localised around the target vertex.

Corollary 1 *For any $f, g \in \mathbb{R}^N$ and $i, j \in \{1, 2, \ldots, N\}$,*

1. $T_i(f * g) = (T_i f) * g = f * (T_i g)$.

2. $T_i T_j f = T_j T_i f$.

3. $\sum_n (T_i f)(n) = \sqrt{N} \hat{f}(0) = \sum_n f(n)$.

However, the niceties end there, and we should also point out some properties that are true for the classical translation operator, but not for the generalized translation operator for signals on graphs. First, unlike the classical case, the set of translation operators $\{T_i\}_{i \in \{1,2,\ldots,N\}}$ do not form a mathematical group; i.e., $T_i T_j \neq T_{i+j}$. In the very special case of the unweighted ring graph, we have

$$T_i T_j = T_{\left[\left((i-1)+(j-1)\right) \bmod N\right]+1}, \ \forall i, j \in \{1, 2, \ldots, N\}. \qquad (22)$$

However, (22) is not true in general for arbitrary graphs. Moreover, while the idea of successive translations $T_i T_j$ carries a clear meaning in the classical case, it is not a particularly meaningful concept in the graph setting due to our definition of generalized translation as a kernelized operator.

Second, unlike the classical translation operator, the generalized translation operator is not an isometric operator; i.e., $\|T_i f\|_2 \neq \|f\|_2$ for all indices i and signals f. Rather, we have

Lemma 1 *For any $f \in \mathbb{R}^N$,*

$$|\hat{f}(0)| \leq \|T_i f\|_2 \leq \sqrt{N} \tilde{\mu}_i \|f\|_2 \leq \sqrt{N} \mu \|f\|_2. \qquad (23)$$

$|\chi_\ell^*(i)|$ may be small or even zero for some i and ℓ, and thus $\|T_i f\|_2$ may be significantly smaller than $\|f\|_2$, and in fact may even be zero. Meanwhile, if μ is close to 1, then we may also have the case that $\|T_i f\|_2 \approx \sqrt{N} \|f\|_2$.

These preliminary results hint at important similitudes and differences with classical Digital Signal Processing. The cornerstone of this emerging corpus is the graph convolution which, with appropriate filters, outlines an important class of linear operators on graphs. These structures are fully compatible with the spectral decomposition of the Laplacian. On the other hand the lack of any

6

specific structure on $\mathcal{G}$ results in the loss of translation invariance. However, properties of $\mathcal{G}$ clearly influence the shift operator as shown in Lemma 1. A full understanding of this proposed theory of signal processing on graphs, at the mathematical and computational levels, is the main goal of this project.

2.4 Related approaches

Even though setting up a generalised theory of signal processing on graphs is a fairly new idea, there has been several approaches that implement multiresolution, or wavelet-like transform, for graph data. For example, Crovella and Kolaczyk [2] defined wavelets on unweighted graphs for analyzing computer network traffic. Their construction was based on the n-hop distance, such that the value of a wavelet centered at a vertex n on vertex m depended only on the shortest-path distance between m and n. However, they did not study the invertibility or frame properties of their transform. Coifman and Gavish have proposed a tree-like construction of wavelets on graphs that easily yields an orthonormal basis. However their scheme involved partitioning the graph into a balanced tree and the authors did not provide a universal algorithm to do so. The closest construction is the "diffusion wavelet" transform of Coifman an Maggioni [3]. Their construction interacts with the underlying graph through repeated applications of a diffusion operator T, analogously to how our construction is parametrized by the choice of the graph Laplacian $\mathcal{L}$. The largest difference between their work and ours is that the diffusion wavelets are designed to be orthonormal. This is achieved by running a localized orthogonalization procedure after applying dyadic powers of T at each scale to yield nested approximation spaces, wavelets are then produced by locally orthogonalizing vectors spanning the difference of these approximation spaces. While an orthogonal transform is desirable for many applications, notably operator and signal compression, the use of the orthogonalization procedure complicates the construction of the transform, and somewhat obscures the relation between the diffusion operator T and the resulting wavelets.

Finally there exist a wide range of interesting techniques for graph coarsening that can be used for implementing multi resolution on graphs. To mention just a few, Lafon and Lee [4] downsample based on *diffusion distances* and form new edge weights based on random walk transition probabilities; the greedy *seed selection* algorithm of Ron et al. [5] leverages an *algebraic distance* measure to downsample the vertices; recursive spectral bisection [6] repeatedly divides the graph into parts according to the polarity (signs) of the Fiedler vectors $\mathbf{u}_1$ of successive subgraphs; Narang and Ortega [7] minimize the number of edges connecting two vertices in the same downsampled subset; and another generally-applicable method which yields the natural downsampling on bipartite graphs ([8, Chapter 3.6]) is to partition V into two subsets according to the polarity of the components of the graph Laplacian eigenvector $\mathbf{u}_{N-1}$ associated with the largest eigenvalue $\lambda_{\max}$. We refer readers to [5, 9] and references therein for more thorough reviews of the graph coarsening literature.

7

3 Earlier contributions to the state-of-the-art by the project investigator

The Principal Investigator, Prof. Pierre Vandergheynst, introduced Spectral Graph Wavelets in [**?**] (see also the upcoming review paper [10]). Among other applications, Prof. Vandergheynst discussed transductive learning with graph wavelets in [11]. He has also further studied distributed computational aspects of graph filtering in [12].

3.1 Research expertise of the research applicant

Pierre Vandergheynst received the M.S. degree in physics and the Ph.D. degree in mathematical physics from the Universit Catholique de Louvain, Louvain-la-Neuve, Belgium, in 1995 and 1998, respectively. From 1998 to 2001, he was a Postdoctoral Researcher with the Signal Processing Laboratory, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. He was Assistant Professor at EPFL (2002 to 2007), where he is now an Associate Professor. His research focuses on harmonic analysis, sparse approximations and mathematical data processing in general with applications covering signal, image and high dimensional data processing, sensor networks, computer vision. He was co-Editor-in-Chief of Signal Processing (2002 to 2006) and Associate Editor of the IEEE Transactions on Signal Processing (2007 to 2011), the flagship journal of the signal processing community. He has been on the Technical Committee of various conferences, serves on the steering committee of the SPARS workshop and was co- General Chairman of the EUSIPCO 2008 conference.

Pierre Vandergheynst is the author or co-author of more than 60 journal papers, one monograph and several book chapters. He has received two IEEE best paper awards. Professor Vandergheynst is a senior member of the IEEE, a laureate of the Apple 2007 ARTS award and of the 2009–2010 De Boelpaepe prize of the Royal Academy of Sciences of Belgium for fundamental contributions in Imaging Sciences.

Professor Pierre Vandergheynst pioneered the use of spectral graph theory as a novel platform for processing signals on graphs. He has a strong background in Computational Harmonic Analysis, having contributed to seminal papers on wavelets on manifolds [13–16]. More generally, Prof. Pierre Vandergheynst has co-authored several recent papers applying graph based modelling for applications in data mining [17–19] or computer vision [20]. He also has extensive experience in sparsity based modelling and compressive sensing theory, which have strong natural connections to learning.

4 Detailed Research Plan

4.1 General Strategy

The project is organised as follows. The management, dissemination and networking tasks will be strictly overlooked by the Principal Investigator (WP0). The research is broken into three main thrusts, each corresponding to a Workpackage. These WPs have been chosen so they each represent the effort of a PhD student over 36 months. There will be cross-fertilisation among WPs, but the topics have been carefully chosen to avoid counter-productive overlaps. For each WP, a list of milestones is provided as a way to assess the success of the corresponding work.

4.2 WP0: Management of the project, dissemination of results and networking

Task T0.1: Management
The project management will be provided by Prof Vandergheynst, who will also commit resources for administrative assistance. No funding required.
Task T0.2: Dissemination
Dissemination activities will be essentially based on publications in international journals with high impact factors and in selected international conferences.
Task T0.3: Networking
Although the topic of this proposal is emerging, we have identified a few research groups with similar agendas. Some of these groups, for instance Prof. Ortega's group at USC, have already expressed their willingness to collaborate. Professor Vandergheynst will seek ways to establish visits and exchange of ideas. No funding required.

4.3 WP1: Foundations of Signal Processing on Graphs

The main goal of this workpackage will be to lead a thorough study of the mathematical properties of graph signal processing. Some early results illustrated below are reproduced from our recent preprint [21].
Task T1.1: Convolution and Localization
Our focus will first be on the translation operator and the associated convolution via filters that was introduced above. First, it is easy to see that a polynomial kernel with degree K that is translated to a given center vertex is strictly localized in a ball of radius K around the center vertex, where the distance $d_{\mathcal{G}}(\cdot, \cdot)$ used to define the ball is the *geodesic* or *shortest path* distance (i.e., the distance between two vertices is the minimum number of edges in any path connecting them). Note that this choice of distance measure ignores the weights of the edges and only depends on the unweighted adjacency matrix of the graph.

Lemma 2 *Let $\widehat{p_K}$ be a polynomial kernel with degree K; i.e.,*

$$\widehat{p_K}(\ell) = \sum_{k=0}^{K} a_k \lambda_\ell^k \tag{24}$$

for some coefficients $\{a_k\}_{k=0,1,\ldots,K}$. If $d_{\mathcal{G}}(i,n) > K$, then $(T_i p_K)(n) = 0$.

More generally, as seen in Figure 1, if we translate a smooth kernel to a given center vertex i, the magnitude of the translated kernel at another vertex n decays as the distance between i and n increases. In the following theorem, we provide one estimate of this localization by combining the strict localization of polynomial kernels with a classical result on the minimax polynomial approximation error.

Theorem 1 *Let $\hat{g} : [0, \lambda_{\max}] \to \mathbb{R}$ be a kernel, define $d_{in} := d_{\mathcal{G}}(i,n)$, and define the minimax polynomial approximation error*

$$B_{\hat{g}}(K) := \inf_{\widehat{p_K}} \left\{ \sup_{\lambda \in [0, \lambda_{\max}]} |\hat{g}(\lambda) - \widehat{p_K}(\lambda)| \right\},$$

where the infimum is taken over all polynomials of degree K, as defined in (24). Then

$$|(T_i g)(n)| \leq \sqrt{N} B_{\hat{g}}(d_{in} - 1). \tag{25}$$

Moreover, if $\hat{g}(\cdot)$ is d_{in}-times continuously differentiable on $[0, \lambda_{\max}]$, then

$$|(T_i g)(n)| \leq \left[\frac{2\sqrt{N}}{d_{in}!} \left(\frac{\lambda_{\max}}{4} \right)^{d_{in}} \right] \sup_{\lambda \in [0, \lambda_{\max}]} \left| \hat{g}^{(d_{in})}(\lambda) \right|. \tag{26}$$

Our main endeavour in this task will therefore be to extend this result into a comprehensive and constructive theory of localisation on graphs, with explicit ties to graph properties and an emphasis on important graph classes (scale-free networks, etc ...) as well as generalising these results to weighted graphs.

Task T1.2: Uncertainty in Graph Representations

To formulate a useful theory of signal processing on graphs, we must also understand the fundamental limits of its main operations. The Heisenberg Uncertainty Principle is one such limit in the classical setting, dictating how that signals and filters cannot be arbitrarily well jointly localised in the impulse and frequency domains. We will therefore investigate limits on the joint vertex and spectral localisation properties of graph filters. One such phenomenon has been recently discussed in [22], but their arguments only apply to low pass filters, whereas we already know that band pass filters such as wavelets can be localised arbitrarily well. When manipulating the spectral domain of the Laplacian matrix, which is finite, one should not expect classical arguments to hold. In this project, we will focus on uncertainty relations measured via indirect ways, in particular joint vertex/spectral decomposition. Indeed, Lieb [23] has

shown that the concentration of the ambiguity function in the classical case is also subject to an uncertainty principle. In this task we will construct a graph based ambiguity function via an appropriate definition of modulations on the vertex set. Motivated by the fact that the classical modulation is a multiplication by a Laplacian eigenfunction, we define, for any $k \in \{0, 1, \ldots, N-1\}$, a *generalized modulation operator* $M_k : \mathbb{R}^N \to \mathbb{R}^N$ by

$$(M_k f)(n) := \sqrt{N} f(n) \chi_k(n).$$

First, note that M_0 is the identity operator, as $\chi_0(n) = \frac{1}{\sqrt{N}}$ for all n for connected graphs. In the classical case, the modulation operator represents a translation in the Fourier domain:

$$\widehat{M_\xi f}(\omega) = \hat{f}(\omega - \xi), \forall \omega \in \mathbb{R}.$$

This property is not true in general for our modulation operator on graphs due to the discrete nature of the graph. However, we do have the nice property that if $\hat{g}(\ell) = \delta_0(\lambda_\ell)$, then

$$\widehat{M_k g}(\ell) = \sum_{n=1}^{N} \chi_\ell^*(n)(M_k g)(n)$$

$$= \sum_{n=1}^{N} \chi_\ell^*(n)\sqrt{N}\chi_k(n)\frac{1}{\sqrt{N}} = \delta_0(\lambda_\ell - \lambda_k),$$

so M_k maps the DC component of any signal $f \in \mathbb{R}^N$ to $\hat{f}(0)\chi_k$. Moreover, if we start with a function f that is localized around the eigenvalue 0 in the graph spectral domain then $M_k f$ will be localized around the eigenvalue λ_k in the graph spectral domain. We have recently quantified this statement in the following result [24]

Theorem 2 *Given a weighted graph $\mathcal{G}$ with N vertices, if for some $\kappa > 0$, a kernel $\hat{f}$ satisfies*

$$\sqrt{N} \sum_{\ell=1}^{N-1} \mu_\ell |\hat{f}(\ell)| \leq \frac{|\hat{f}(0)|}{1+\kappa}, \tag{27}$$

then

$$|\widehat{M_k f}(k)| \geq \kappa |\widehat{M_k f}(\ell)| \quad \text{for all } \ell \neq k. \tag{28}$$

Based on this localisation result, and inspired by Lieb, we will therefore study the localisation of the ambiguity function associated to a graph filter g :

$$\mathcal{A}_g(i, k) = \langle T_i M_k g, g \rangle.$$

Our starting point will be to look for estimates of $\|\mathcal{A}_g\|_{1,1}$ as a proxy for the concentration of the ambiguity.

Task T1.3: Extension to edge-centric signals and directed graphs
This task is motivated by applications where the connectivity of the graph is fixed (the set of edges), but the edge weights represent the signal to be processed, i.e the information is not on the vertices but on the edges of the graph. An important application scenario is the processing of large Origin-Destination matrices or flows: these matrices represent graphs whose edge weights measure the amount of "traffic" between two regions represented by vertices. The vertices and their connections are fixed, but the amount of traffic can vary. The Principal Investigator has exclusive access to Origin/Destination matrices that represent large scale pedestrian flows in Swiss railway stations. However, processing this data requires to generalise the above constructed theory of signal processing on graphs to signals defined at the edges of a directed graph. We propose to carry this generalisation via the construction of the *line graph* $L(\mathcal{G})$ of $\mathcal{G}$ [25]. The line graph essentially exchanges the role of vertices and edges in the original graph and therefore constructing filters on $L(\mathcal{G})$ will result in an appropriate setting for processing edge signals. Finally, we will consider a direct application of Chung's directed graph Laplacian [26] to generalise the above mentioned ideas to directed graphs. The advantage of Chung's construction is that the corresponding Laplacian operator remains symmetric and hence has real eigenvalues.

Milestones: A comprehensive theory of localisation on graphs, with sharp results for important classes of graphs. A new breed of uncertainty principles based on joint vertex/spectral analysis. A constructive extension of signal processing on graphs for edge-based signals and directed graphs, with a simple software toolbox for experimentation and validation.

4.4   WP2: Connections with Learning Theory

In this workpackage we would like to set the foundations of a new learning theory using graphical structures. To this end we will leverage and generalise the idea of sparse representations of functions to signals on graphs, focusing on characterising precisely what signal properties lead to sparse wavelet expansions. This will lead to transductive learning results. In a second step, we will develop inductive learning results by studying a mechanism for out-of-sample extension of graph wavelet regularisation. Leveraging recent results on the convergence of the graph laplacian to the Laplace-Beltrami operator on manifolds, we will set up new connections between classes of smooth functions on graphs and their asymptotic convergence to smooth functions on manifolds. The out-of-sample extension mechanism will therefore allow us to quantify precisely what type of distributions can be inferred by partial observations on graphs.

Task 2.1: Beyond global smoothness
The success of sparse representations of signals and images can be explained, in large parts, by the existence of sound mathematical results linking sparsity to signal structures. In particular, it is well known that sparsity is strongly linked to various notions of smoothness. For example, piecewise-smooth signals are optimally approximated by sparse representations in a wavelet basis

because the wavelet coefficients of such signals decay optimally fast. That notion of *local* smoothness is best captured by Besov or Hölder norms. Intuitively, wavelets sparsify piecewise smooth signals because they behave like multiresolution differential operators : at sufficiently small scales, most wavelet coefficients $\langle \psi_{t,b}, f \rangle$ will be small because the signal will appear locally polynomial. There are also many interesting signal models for which the underlying assumption is a *global* smoothness hypothesis. These models are best captured by Sobolev regularity, which also translates directly into sparsity (both for wavelets or Fourier coefficients).

Since smoothness is the signal structure that induces sparsity for signals and images, it is vital to precisely understand what is the equivalent property for graph data : what structure induces sparsity of the spectral graph wavelet representation. This seemingly simple question is key to a coherent mathematical framework linking learning with graphs and harmonic analysis on graphs. However, due to the fundamentally discrete nature of graphs, there is no straightforward generalization.

Our starting point is be a notion of global regularity of graph data captured by an extension of Sobolev norms. Recall that the Sobolev (semi-)norm of a function f on the circle can be defined by controlling the p-norm of the signal derivatives :

$$\|f\|_{W^{k,p}} = \Big(\sum_{\alpha \leq k} \|f^{(\alpha)}\|_p^p \Big)^{1/p}, \quad 1 \leq p < +\infty$$

and the extension to $p = \infty$ is done by using the corresponding norm. The case $p = 2$ is particularly interesting because it translates immediately into a control of the decay of the Fourier coefficients of f. Basically, $f \in W^{k,2}$ means it is square integrable and

$$\sum_{n \in \mathbb{Z}} \big(1 + |n|^2\big)^k |\hat{f}(n)|^2.$$

This characterization is key to extending to graphs. Indeed, the boundedness of the 2-Dirichlet form $S_2(f)$ is precisely equivalent to requiring that f belongs to a natural generalization of $W^{1,2}$, where the graph Fourier transform coefficients are used.

Let us now outline why this particular notion of regularity is connected to sparsity of the spectral graph wavelet coefficients. Remark that :

$$\sum_i |\langle \psi_{t_n,i}, f \rangle|^2 = \sum_\ell |g(t_n \lambda_\ell)|^2 |\hat{f}(\ell)|^2,$$

and so

$$\sum_n t_n^2 \sum_i |\langle \psi_{t_n,i}, f \rangle|^2 \approx \|f\|_{W^{1,2}},$$

which in turns imposes that the wavelet transform coefficients decay sufficiently fast with scale. Note though that this reasoning implicitly assumes a large spectral radius and that the effects of finite "frequency range", or finite resolution, should be investigated. The conclusion, though, is that a function that is

smooth on the graph in the sense of having a small 2-Dirichlet form has sparse wavelet coefficients in the spectral graph wavelet domain.

As explained above, it is also important to connect sparsity in the spectral graph wavelet representation to more local notions of smoothness. This could be done by either extending the equivalence between Sobolev norms and norms on graph wavelet coefficients to the case $p = \infty$ (Lipschitz smoothness) or by directly studying graph based extension of Hölder regularity. A possible definition would be that f is α-Hölder regular if it satisfies the following

$$|f(i) - f(j)| \leq C \cdot d_G(i,j)^\alpha,$$

where $d_G(.,.)$ is the graph geodesic distance between vertices. In the classical euclidean wavelet transform, this condition also translates into the decay at small scales of the wavelet coefficients. Two key ingredients in the proof are the zero-mean condition of wavelets and their localization at small scales. These two requirements are also valid for spectral graph wavelets and we will thus try to use them to extend the equivalence between local Hölder regularity and sparsity as future work.

This connection between smoothness and sparsity allows us to set up a straightforward algorithm for transductive learning by leveraging the smoothness assumption. Assume we want to regress an unknown piecewise smooth f function on a connected graph from observations limited to few randomly selected samples. We can cast this task as a sparsity-constrained inverse problem of the form :

$$f = \arg\min_u \|\mathbf{M}Wu - y\|_2^2 + \lambda\|u\|_1 \tag{29}$$

where the minimization is performed over wavelet coefficients u, the observed data is y, W is the wavelet synthesis operator (adjoint of the forward transform) and sparsity is enforced by the use of the one norm $\|u\|_1$. The restriction operator $\mathbf{M}$ simply sets unobserved vertices values to zero. This convex optimization problem can be solved efficiently by iterative soft-thresholding of the wavelet coefficients using the fast spectral graph wavelet transform algorithm.

One of our goals in the forthcoming months will be to leverage the extensive body of results on harmonic analysis, sparsity and inverse problems to semi-supervised learning using spectral graph wavelets, in particular using notions of structured sparsity.

Task 2.2: Sobolev regular interpolation

When performing regression from scattered values, an essential tool in traditional signal processing is spline interpolation [27]. One of the most interesting properties of splines is that they allow to precisely control the regularity of the interpolated function. Recently, Pezenson [28] has proposed a construction of interpolating splines on graphs. We will basically seek to adapt this construction to our framework by explicitly using our localisation operator and the proper smoothness class defined via the spectral decomposition of the Laplacian. Starting with a regularized version $\mathcal{L}_\epsilon^t$ of the t-th power of the Laplacian, the Green's function localized at w, which we will note $E_{2t,\epsilon}^w$, satisfies:

$$(\epsilon I + \mathcal{L})^t E_{2t,\epsilon}^w = \delta_w.$$

14

This relationship is the building block of variational splines. In the Graph Fourier domain, this reads:

$$(\epsilon + \lambda_\ell)^t g_{2t,\epsilon}^w(\lambda_\ell) = \phi_\ell(w).$$

Let us define the kernel $g_{2t,\epsilon}^w$ of the localized Green's function via the Fourier shift or Localization operator acting on the basic Green kernel:

$$g_{2t,\epsilon}^w(\lambda_\ell) \equiv g_{2t,\epsilon}(\lambda_\ell)\phi_\ell(w),$$

which leads directly to :

$$(\epsilon + \lambda_\ell)^t g_{2t,\epsilon}(\lambda_\ell) = 1.$$

We will be using a slightly modified definition of the Sobolev norm introduced above :

$$\|f\|_{H_{t,\epsilon}}^2 = \sum_\ell (\epsilon + \lambda_\ell)^{2t} |\hat{f}(\ell)|^2.$$

It is interesting to notice that the Sobolev norm of localized Green's functions is independent of the localization, as it should :

$$\|g_{2t,\epsilon}^w\|_{H_{t,\epsilon}}^2 = \sum_\ell (\epsilon + \lambda_\ell)^{2t} |g_{2t,\epsilon}^w(\lambda_\ell)|^2 = 1.$$

These preliminaries allow us to construct a Parseval frame of localised Green's functions on $\mathcal{G}$:

Theorem 3 *The set of localized Green's functions $E_{2t,\epsilon}^w$, $w \in V$ is a Parseval frame for the Sobolev space $H_{t,\epsilon}$, meaning that for every $\alpha \in \ell_2(V)$*

$$\|\sum_{u \in V} \alpha[u] E_{2t,\epsilon}^u\|_{H_{t,\epsilon}}^2 = \|\alpha\|_2^2.$$

This construction will constitute our starting point for using interpolating splines for learning smooth functions over $\mathcal{G}$. It will be interesting to study more precisely the localisation properties of those splines and see wether the interpolated function can be computed with dedicated algorithms using the particular form of the frame. Finally, since Green's functions are naturally related to the pseudo-inverse of the Laplacian[1], it might also be advantageous to study the properties of the frame in the resistive metric rather than in the traditional geodesic metric over $\mathcal{G}$.

Task 2.3: convergence to manifolds

Suppose we are given a graph $\mathcal{G}$ that is sampled from a manifold $\mathcal{M}$ with Laplace-Beltrami operator $\Delta_\mathcal{M}$. The connections on $\mathcal{G}$ are computed by associating a function of the euclidean distance in embedding space, typically an

---

[1]In fact it is easy to check that

$$\langle E_{2t,\epsilon}^u, E_{2t,\epsilon}^v \rangle_{H_{t,\epsilon}} = \langle \delta_u, (\mathcal{L}_\epsilon^t)^\dagger \delta_v \rangle.$$

exponential, to the k-nearest neighbour graph of samples. Let $\mathcal{L}$ be the Laplacian matrix of that graph. A recent series of fascinating papers [29, 30] have focused on studying the convergence of $\mathcal{L}$ to $\Delta_{\mathcal{M}}$ as well as the convergence of the associated eigenvectors. Interestingly, this convergence allows us to estimate the action of $\Delta_{\mathcal{M}}$ on a given function f at a point x from sampled values $f(x_i)$:

$$\Delta_{\mathcal{M}}f(x) \approx -\frac{1}{t}(4\pi t)^{\frac{n}{2}}\left(f(x) - \sum_{x_i} f(x_i)e^{-\frac{\|x-x_i\|^2}{4t}}\right). \tag{30}$$

Naturally, this expression paves the way to extending the action of linear filters to out-of-sample values by simply inserting the expression for the estimated Laplacian action into the filter definition. One particular application we will investigate is as follows. Suppose we have solved the LASSO problem (29) in a transductive way (i.e for sampled values only). We can now use (30) to extend the wavelet synthesis operator to any new data point x and synthesise a solution based only the pre-computed solution coefficients u. We will study this mechanism, in particular trying to understand the connection with the smoothness of the asymptotic solution. One guideline will be the potential connection to wavelets on manifolds constructed directly from the eigendecomposition of $\Delta_{\mathcal{M}}$ by Geyer and Mayeli [31].

Milestones: Results on the sparsifying properties of graph wavelets for locally smooth signals on graphs. Principled approaches based on linear filters for regression of piece-wise smooth functions on graphs. Application of these results to transductive and semi-supervised learning. Out of sample extension of spectral constructions based on convergence of the graph Laplacian to the Laplace-Beltrami operator.

4.5   WP3: Computational Aspects

One of the main application of graph based algorithms will be in the processing of very large datasets. The need for efficient algorithms and implementations is therefore stringent. Obviously, one cannot compute the eigenvectors of very large laplacian matrices and directly work in the spectral domain, but other efficient methods can be used to implement graph filtering. First [?] has introduced a very fast algorithm to compute the spectral graph wavelet transform and its adjoint. The algorithm is based on the approximation of the kernel g by an expansion in Chebychef polynomials:

$$g(t_n x) = \frac{1}{2}c_{n,0} + \sum_{k=1}^{\infty} c_{n,k}\overline{T}_k(x). \tag{31}$$

When the Laplacian matrix is sparse, the corresponding approximate wavelet transform is computed efficiently by sparse matrix-vector multiplications and by making use of recurrence relations for the polynomials $\overline{T}_k$. This algorithm is of particular interest for distributed applications, for instance in sensor networks, where the whole method is implemented via passing scalar messages. It

can of course be used with any filter. In this workpackage we will be focusing on efficient implementations for centralised, eventually parallel, architectures by explicitly leveraging recent advances in software engineering of graph based applications. The vertex centric model at the heart of the proposed theory of signal processing on graphs lends itself remarkably well to some of the latest and most promising computational frameworks for large scale processing with graph data structures. Furthermore, we will exploit the multi resolution nature of wavelets to derive scalable coarse-to-fine algorithms for graph signal processing.

Task 3.1: Large scale processing with GraphLab and GraphChi

GraphLab [32] is a recent framework for distributed calculations involving graphs. It is based on a vertex centric framework that implements an abstraction level very similar to the well-known Map-Reduce framework, but built from scratch for dependencies embodied in a graph structure. GraphChi [33] is a similar abstraction framework but specifically built for large scale applications on a single machine, with optimal use of memory and disk accesses. GraphChi does not necessitate special computational resources (cluster) and can therefore be used more easily than GraphLab. It is particularly well-suited for localised filtering on graphs with sparse laplacian matrices. Indeed, localised filters can be thought of as windows of a fixed and controlled size (geodesic radius), sliding over the graph and in which simple recursive calculation are done to implement the action of the filter. These calculations only invoke edge weights and the spectral signature of the filter or, equivalently, its Chebyshev expansion coefficients. This sliding window metaphor is precisely at the heart of GraphChi and we therefore expect strong benefits in implementing all basic signal processing on graphs operations via this framework. One of the main outcomes of this work package will be a toolkit (in the GraphLab terminology) and a set of easy-to-use python wrappers.

Task 3.2: Principled coarse-to-fine computations on graphs

This second task will typically build on Task 3.1 and go one step further via the idea of coarse-to-fine calculations. Indeed, in most wavelet based problems, multiresolution allows to approximately solve the problem at a coarse scale with low computational burden and later refine the solution locally at finer scales if necessary. We believe this is one of the main strengths of spectral graph wavelets, but they will only live up to their expectations if this can be turned into a computational statement. To deploy coarse-to-fine computations, we will first study various methods to downsample a large graph, if possible in a way that is compatible with the spectral graph wavelets structure. One such technique involves recursively partitioning the graph using a colouring scheme based on the polarity of the largest eigenvector of the Laplacian. We will start by exploring this idea in the framework of a generalised graph Laplacian Pyramid where each level involves only linear graph filtering implemented with GraphChi.

Task 3.3: Large scale demonstrations

The final task in this workpackage will consist in validating the software approach on various datasets and building connections with WP1 and WP2 by providing novel and efficient computational approaches. We will be looking

to apply our approach to large scale problems, in particular social network datasets available on the internet and for which reference approaches have been published. But we will also apply our code to original, even exclusive large scale datasets. We have direct access to several challenging datasets: large asteroid meshes with Albedo data on the vertices, high resolution Planck Cosmic Microwave Data (non-uniformly sampled over a sphere) as well a large graph of similarities between songs from the Montreux Jazz Festival Digital Archive.
Milestones: GraphChi implementation of linear filtering on graphs, spectral graph wavelet transform and associated optimisation problems. Laplacian pyramid on graph with efficient GraphChi implementation for one or more choices of graph partitioning. Evaluation on large scale datasets.

5 Timeline

The detailed research timeline of the project is depicted below. The project lead is carried out by Prof. P. Vandergheynst. An experienced post-doc researcher, Dr Benjamin Ricaud, will help supervise the PhD students.
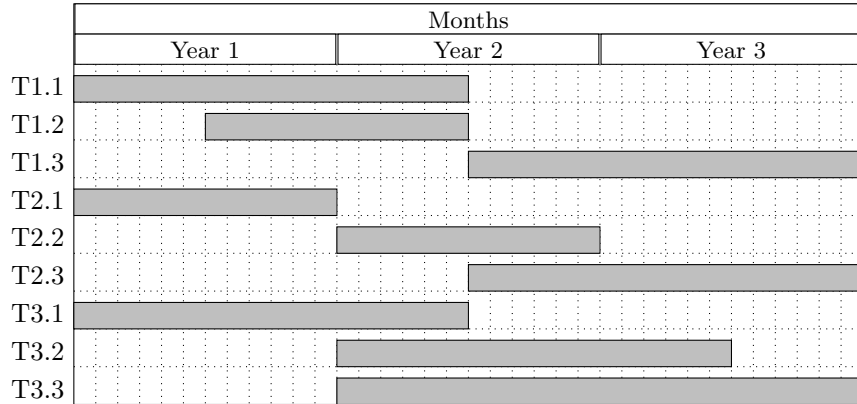

Figure 2: Project timeline

6 Significance of Planned Research

An era of fast paced scientific and technological innovations ("Data Driven Science") based on the availability of large scale datasets ("Big Data") is definitely within our reach. The span of problems involving big data is staggering, from brain understanding to cosmology. However, exploiting these vast amounts of data necessitates to solve several challenges. This project aims at providing the necessary mathematical framework, algorithms and computational tools to overcome some of the deadlocks of big data processing.

We expect that the results of this project will make it as easy to analyse and process large graph based data as it is currently to process signals or images, thereby opening the door to making big data analysis part of the standard toolbox of any engineer. Given the ubiquity of these problems, this would be a very significant achievement. Moreover, since the industry is also strongly embracing the big data paradigm, we expect our results to fuel technological innovations in the ICT sphere, strengthening the position of Switzerland in that domain.

7 Risk Management

Although the three research axes proposed in this project are fairly adventurous, they each contain elements to mitigate risks. In WP1, some amount of work has already been carried out as preparation by Nathanal Perraudin during his MSc thesis dedicated to uncertainty in graph representations. Moreover, the basic elements allowing to study localisation are already in place. For WP2, Prof. Pierre Vandergheynst has carried extensive experiments in transductive learning with spectral graph wavelets. In particular, all the necessary software is already available. Finally in WP3, the PhD student we would hire on this project (Kirell Benzi) has already joined EPFL and has been trained in the necessary software engineering tools.

8 Ethical Aspects of the Research

Not applicable.

References

[1] F. K. Chung, *Spectral Graph Theory.* Vol. 92 of the CBMS Regional Conference Series in Mathematics, AMS Bokstore, 1997.

[2] M. Crovella and E. Kolaczyk, "Graph Wavelets for Spatial Traffic Analysis," in *Proc. IEEE INFOCOM*, Mar. 2003, pp. 1848–1857 vol.3.

[3] R. R. Coifman and M. Maggioni, "Diffusion Wavelets," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 53–94, 2006.

[4] S. Lafon and A. B. Lee, "Diffusion Maps and Coarse-Graining: A Unified Framework for Dimensionality Reduction, Graph Partitioning, and Data Set Parameterization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1393–1403, Sep. 2006.

[5] D. Ron, I. Safro, and A. Brandt, "Relaxation-Based Coarsening and Multiscale Graph Organization," *Multiscale Model. Simul.*, vol. 9, no. 1, pp. 407–423, Sep. 2011.

[6] S. T. Barnard and H. D. Simon, "Fast Multilevel Implementation of Recursive Spectral Bisection for Partitioning Unstructured Problems," *Concurrency: Practice and Experience*, vol. 6, no. 2, pp. 101–117, Apr. 1994.

[7] S. K. Narang and A. Ortega, "Lifting based wavelet transforms on graphs," in *Proc. APSIPA ASC*, Sapporo, Japan, Oct. 2009, pp. 441–444.

[8] T. Biyikouglu, J. Leydold, and P. F. Stadler, *Laplacian Eigenvectors of Graphs*. Lecture Notes in Mathematics, vol. 1915, Springer, 2007.

[9] C. Walshaw, "The Graph Partitioning Archive."

[10] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains," *arXiv.org*, Nov. 2012.

[11] D. I. Shuman, M. J. Faraji, and P. Vandergheynst, "Semi-Supervised Learning with Spectral Graph Wavelets," 2011.

[12] D. I. Shuman, P. Vandergheynst, and P. Frossard, "Chebyshev Polynomial Approximation for Distributed Signal Processing," *arXiv.org*, May 2011.

[13] J.-P. Antoine and P. Vandergheynst, "Wavelets on the n-sphere and related manifolds," *Journal of Mathematical Physics*, vol. 39, no. 8, pp. 3987–4008, 1998.

[14] ——, "Wavelets on the 2-Sphere : A Group Theoretical Approach," *Applied and Computational Harmonic Analysis*, vol. 7, no. 3, pp. 262–291, Nov. 1999.

[15] I. Bogdanova, P. Vandergheynst, and J. P. Gazeau, "Continuous wavelet transform on the hyperboloid," *Applied and Computational . . .*, 2007.

[16] J.-P. Antoine and P. Vandergheynst, "Wavelets on the two-sphere and other conic sections," *Journal of Fourier Analysis and . . .*, 2007.

[17] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov, "Clustering With Multi-Layer Graphs: A Spectral Perspective," *Signal Processing, IEEE Transactions on*, vol. 60, no. 11, pp. 5820–5831, 2012.

[18] X. Dong, P. Frossard, and P. Vandergheynst, "Clustering on Multi-Layer Graphs via Subspace Analysis on Grassmann Manifolds," *arXiv.org*, 2013.

[19] X. Dong, A. Ortega, P. Frossard, and P. Vandergheynst, "Inference of Mobility Patterns via Spectral Graph Wavelets," in *International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, 2013.

[20] L. Bagnato, P. Frossard, and P. Vandergheynst, "OPTICAL FLOW AND DEPTH FROM MOTION FOR OMNIDIRECTIONAL IMAGES USING A TV-L1 VARIATIONAL FRAMEWORK ON GRAPHS," in *Proceedings of ICIP*. Cairo, Egypt: EPFL, 2009, pp. 1469–1472.

[21] D. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-Frequency Analysis on Graphs," *Applied and Computational Harmonic Analysis.*

[22] A. Agaskar and Y. M. Lu, "A Spectral Graph Uncertainty Principle," *arXiv.org*, Jun. 2012.

[23] E. Lieb, "Integral bounds for radar ambiguity functions and Wigner distributions," *Journal of Mathematical Physics*, vol. 31, p. 594, 1990.

[24] D. Shuman, B. Ricaud, and P. Vandergheynst, "A Windowed Graph Fourier Transform," in *2012 Ieee Statistical Signal Processing Workshop (Ssp)*, EPFL. New York: Ieee, 2012, pp. 133–136.

[25] M. Aigner, "On the linegraph of a directed graph," *Mathematische Zeitschrift*, vol. 102, no. 1, pp. 56–61, 1967.

[26] F. Chung, "Laplacians and the Cheeger Inequality for Directed Graphs," *Annals of Combinatorics*, vol. 9, no. 1, pp. 1–19, 2005.

[27] M. Unser, "Sampling-50 Years after Shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[28] I. Pesenson, "Variational Splines and Paley–Wiener Spaces on Combinatorial Graphs," *arXiv.org*, Nov. 2011.

[29] A. Singer, "From Graph to Manifold Laplacian: The Convergence Rate," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 128–134, Jul. 2006.

[30] M. Belkin and P. Niyogi, "Towards a Theoretical Foundation for Laplacian-Based Manifold Methods," *J. Comp. Syst. Sciences*, vol. 74, pp. 1289–1308, 2008.

[31] D. Geller and A. Mayeli, "Continuous wavelets on compact manifolds," *Mathematische Zeitschrift*, vol. 262, no. 4, pp. 895–927, 2009.

[32] J. Gonzalez, Y. Low, H. Gu, D. Bickson, and C. Guestrin, "PowerGraph: Distributed graph-parallel computation on natural graphs," in *Proc. of the 10th USENIX conference on Operating systems design and implementation*, 2012.

[33] A. Kyrola, G. Blelloch, and C. Guestrin, "GraphChi: Large-scale graph computation on just a PC," in *Proceedings of the 10th conference on Symposium on Opearting Systems Design & Implementation*, 2012.