SBrT 2019 1570554422

On the Use of Vertex-Frequency Analysis for Anomaly Detection in Graph Signals

Gabriela Lewenfus, Wallace A. Martins, Symeon Chatzinotas, and Björn Ottersten

***Abstract*— Graph signals (GS) are widespread in many areas of data analysis, such as in social, genetics, and biomolecular networks as well as in several engineering applications. Detecting localized properties of GS using spectral tools while taking into account the underlying graph topology is still an active research topic called vertex-frequency analysis (VFA). This paper provides a brief and up-to-date overview on state-of-the-art VFA tools, namely windowed graph Fourier transform and spectral graph wavelet transform. In addition, the paper shows how VFA can be applied to detect and localize anomalies in GS. In the particular example of localizing a malfunctioning weather station, the average area under ROC curve achieved by the local factor outlier technique can be improved from 72% to 87% when fed with VFA-extracted features to detect small drifts in temperature measurements, ranging from 0.5°C to 4°C.**

***Keywords*— GSP, vertex-frequency analysis, Fourier transform, wavelets, anomaly detection**

I. Introduction

Many current practical problems can be modeled via data signals defined on the nodes of a weighted graph. Social media [1], traffic [2], wireless networks [3], genetic networks [4], and functional relationship across brain regions [5] are instances of this kind of data. Graphs could also describe similarities among high dimensional data points in statistical learning, being useful to machine learning solutions [6]. *Graph signal processing* (GSP) extends classical discrete signal processing (DSP) techniques in order to reveal relevant information about these unstructured data by exploring the underlying topology.

In DSP, both wavelet transform and windowed Fourier transform (also termed short-time Fourier transform) are well-known frequency-analysis tools that also provide localized information in time/space. *Graph signals* (GS) can also have different localized spectral properties across nodes and *vertex-frequency analysis* (VFA) approaches have been developed in the past decade to deal with GS [7].

This paper provides an up-to-date overview on graph spectral methods for VFA and proposes a general methodology for anomaly detection and localization in GS. Section II defines notations and provides a brief review on GSP. Section III describes the *windowed graph Fourier transform* (WGFT) [8], followed by the *spectral graph wavelet transform* (SGWT) [9] in Section IV. In Section V, we propose a framework for applying VFA concepts to the problem of anomaly detection in graphs and, in Section VI, we provide experimental results with real data. Section VII contains some concluding remarks and future research directions.

Ms. Lewenfus is with the Federal University of Rio de Janeiro (UFRJ), e-mail: gabriela.lewenfus@gmail.com; Prof. Martins is with UFRJ (on leave) and with the University of Luxembourg (Research Associate), e-mail: wallace.martins@smt.ufrj.br; Prof. Chatzinotas is with the University of Luxembourg (SnT), e-mail: symeon.chatzinotas@uni.lu; and Prof. Ottersten is with the University of Luxembourg (SnT), e-mail: bjorn.ottersten@uni.lu.
This work was financed in part by the ERC project AGNOSTIC and by the Brazilian research agency FAPERJ.

II. GSP: Notation and Background

Let $\mathcal{G} \triangleq (\mathcal{V}, \mathcal{E}, \mathbf{A})$ be a weighted graph, where $\mathcal{V} \triangleq \{v_1, \ldots, v_N\}$ is the set of N nodes, $\mathcal{E}$ is the set of edges, and $\mathbf{A}$ is an $N \times N$ matrix containing edge weights A_{mn}. In this paper, we only consider undirected weighted graphs (i.e., $\mathbf{A}$ is symmetric) with no self loops. The degree of node n, d_n, is the sum of weights of all edges connected to node n. The degree matrix $\mathbf{D}$ is a diagonal matrix containing each node degree d_n. Filters described in this text are based on the Laplacian matrix $\mathbf{L} \triangleq \mathbf{D} - \mathbf{A}$, which resembles a discrete version of the Laplace-Betrami operator [10]. The Laplacian matrix is symmetric semidefinite having a zero eigenvalue. The constant function is thus always the smoothest eigenvector, providing an interpretation of DC component to the magnitude of the frequency content corresponding to the zero eigenvalue. Other approaches use matrix $\mathbf{A}$ as building block for graph filters [11] and, to this date, there is no consensus regarding which approach brings more advantages.

In this paper, a GS is a real-valued scalar function $x : \mathcal{V} \to \mathbb{R}$ taking values on the graph nodes, and it will be represented by the N-dimensional vector $\mathbf{x}$. The order of arrangement of the nodes in the vector is arbitrary and does not have any impact on the frequency representation of the GS; we shall associate $\mathcal{V}$ with the set $\{1, \ldots, N\}$ so that the signal value on the n^{th} vertex, $x(v_n)$, can be denoted as $x[n]$.

Let $0 = \lambda_1 \leq \ldots \leq \lambda_N \triangleq \lambda_{\max}$ and $\mathbf{u}_1, \ldots, \mathbf{u}_N$ be the eigenvalues and corresponding orthonormal eigenvectors of the Laplacian matrix $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$. The rows of $\mathbf{U}$ will be denoted as $\bar{\mathbf{u}}_n^{\mathrm{T}}$ for $n \in \{1, \ldots, N\}$. The Laplacian eigenvectors are the Fourier basis and, therefore, the *graph Fourier transform* (GFT) of a GS $\mathbf{x}$ is $\hat{\mathbf{x}} \triangleq \mathbf{U}^{\mathrm{T}}\mathbf{x}$. Notice that the Laplacian eigenvalues satisfy $\lambda_k = \mathbf{u}_k^{\mathrm{T}}\mathbf{L}\mathbf{u}_k = \sum_{m,n} A_{mn}(u_k[n] - u_k[m])^2$ and, therefore, a low λ_k is associated with an eigenvector with small variations between nodes, assuming $A_{mn} \geq 0$.

The GS convolution between $\mathbf{x}$ and $\mathbf{y}$ is inspired by the classic convolution theorem, being defined in the frequency domain as the element-wise multiplication $\hat{\mathbf{x}} \odot \hat{\mathbf{y}}$ [9]. Filtering is also defined in the frequency domain as follows: consider $h : [0, \lambda_{\max}] \to \mathbb{R}$ and let $\tilde{\mathbf{H}} = h(\mathbf{\Lambda})$ be a diagonal matrix with diagonal entries $h(\lambda_n)$; the output signal $\mathbf{y}$ is

$$\mathbf{y} = \mathbf{U}\hat{\mathbf{y}} = \mathbf{U}\tilde{\mathbf{H}}\hat{\mathbf{x}} = \mathbf{U}\tilde{\mathbf{H}}\mathbf{U}^{\mathrm{T}}\mathbf{x} = \mathbf{H}\mathbf{x}, \tag{1}$$

in which $\mathbf{H} \triangleq \mathbf{U}\tilde{\mathbf{H}}\mathbf{U}^{\mathrm{T}}$ is the graph-domain representation of filter h. If $\mathbf{H} = h(\mathbf{L})$ is a polynomial on $\mathbf{L}$, then it is a linear shift-invariant filter when the shift of $\mathbf{x}$ on $\mathcal{G}$ is $\mathbf{L}\mathbf{x}$ [11].

III. Windowed Graph Fourier Transform

The classical windowed Fourier transform divides the signal into segments, via a sliding window, before applying the discrete Fourier transform (DFT) to each piece. In order to describe the WGFT [8], it is necessary to first generalize the concept of translation. In DSP, centering a window $w[m]$ into

sample n, for example, is the same as convolving $w[m]$ with the Kronecker delta function $\delta_n[m]$. In the GSP framework, the GFT of $\boldsymbol{\delta}_n$ (zero vector with 1 in the n^{th} entry) is given by $\mathbf{U}^{\mathrm{T}}\boldsymbol{\delta}_n = \bar{\mathbf{u}}_n$. The translation of a window function $w : [0, \lambda_{\max}] \to \mathbb{R}$ to node n on $\mathcal{G}$ is

$$\mathbf{t}_n^w \triangleq \mathbf{U} w(\boldsymbol{\Lambda}) \bar{\mathbf{u}}_n = \mathbf{U} \left(\mathbf{w} \odot \bar{\mathbf{u}}_n \right), \tag{2}$$

where $\mathbf{w} \triangleq [w(\lambda_1), \ldots, w(\lambda_N)]^{\mathrm{T}}$. Note that this definition of translation on graphs does not preserve signal energy. In fact, if $w(0) \neq 0$, the energy of the translated window $\mathbf{t}_n^w$ is limited by $\frac{1}{\sqrt{N}}|w(0)|$ and $\max_k\{\bar{\mathbf{u}}_n[k]\}\|\mathbf{w}\|_2$ [8].

In order to achieve localization in the frequency domain, the convolution theorem is used in its dual form: to concentrate the GS $\mathbf{x}$ on a graph frequency k, one has to multiply $\mathbf{x}$ element-wise by the Laplacian eigenvector $\mathbf{u}_k$. This operation is called modulation and is defined as

$$\mathbf{m}_k(\mathbf{x}) \triangleq \mathbf{x} \odot \mathbf{u}_k. \tag{3}$$

Putting together (2) and (3), WGFT atoms and coefficients corresponding to a GS $\mathbf{x}$ are respectively defined as

$$\mathbf{w}_{n,k} \triangleq \mathbf{m}_k\left(\mathbf{t}_n^w\right) \quad \text{and} \quad x_{n,k}^w \triangleq \langle \mathbf{x}, \mathbf{w}_{n,k} \rangle . \tag{4}$$

The matrix of WGFT coefficients $\mathbf{X}^w$ can be computed by first filtering the GS $\mathbf{x}$ by $\mathbf{T} = \mathbf{U} w(\boldsymbol{\Lambda}) \mathbf{U}^{\mathrm{T}}$ and then multiplying each entry of $\mathbf{T}\mathbf{x}$ element-wise by each column of $\mathbf{U}$. The computational cost of this process is N^2 for matrix vector multiplication plus N^2 for element-wise multiplication, yielding an overcomplete representation composed by N^2 coefficients.

If the window function w has nonzero mean, the original GS $\mathbf{x}$ is perfectly recovered from its WGFT coefficients by [8]

$$x[n] = \frac{1}{\|\mathbf{t}_n^w\|_2^2} \sum_{m=1}^{N} \sum_{k=1}^{N} x_{m,k}^w w_{m,k}[n]. \tag{5}$$

One can notice that the WGFT has some disadvantages: (i) it increases the number of coefficients to represent a GS from N to N^2; (ii) it is computationally expensive and has poor scalability; indeed, WGFT atoms depend on the computation of Laplacian eigendecomposition, being prohibitive for large graphs. Actually, translation can be approximated by a Chebyshev polynomial on the Laplacian matrix (see Section IV), however, to the best of our knowledge, there is no method for approximating the modulation operation in (3); (iii) it does not provide a tight frame[1] for irregular graphs [8]; tight frames allow the interpretation of spectograms as an energy density function, improving stability in recovering signals from noisy measurements, while also enabling faster computations [12]; (iv) the resolution in the frequency domain is uniform, just like in the classical DSP.

Wavelets, on the other hand, can: (i) have a more compact representation with $N \cdot S$ coefficients, with the number of scales satisfying $S < N$; (ii) be fully implemented without resorting to Laplacian eigendecomposition; (iii) provide a tight frame, depending on the choice of the mother/scale functions; (iv) have adaptive resolution in accordance with Heisenberg's theorem [12]. Next section introduces VFA using wavelets.

[1]When $A\|\mathbf{x}\|_2^2 \leq \sum_{n,k} |\langle \mathbf{x}, \mathbf{w}_{n,k} \rangle|^2 \leq B\|\mathbf{x}\|_2^2$ for some constant $A, B > 0$, then the dictionary of atoms $\{\mathbf{w}_{n,k}\}$ is a *frame*. If $A = B$ the frame is said to be *tight*.

IV. Spectral Graph Wavelet Transform

As in the WGFT definition, the SGWT is defined by a basic function, called wavelet mother/kernel, $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ in the frequency domain [9]. This mother function must be a band-pass signal satisfying $\psi(0) = 0$ and $\lim_{z\to\infty} \psi(z) = 0$. Wavelet atoms $\boldsymbol{\psi}_{n,s}$, with $(n, s) \in \{1, \ldots, N\} \times \{1, \ldots, S\}$, are GS constructed by dilating the mother function ψ by a factor $\alpha_s \in \mathbb{R}^+$, and then translating it to vertex n, as follows:

$$\boldsymbol{\psi}_{n,s} \triangleq \mathbf{U}\psi(\alpha_s \boldsymbol{\Lambda}) \bar{\mathbf{u}}_n = \left(\mathbf{U}\psi(\alpha_s \boldsymbol{\Lambda}) \mathbf{U}^{\mathrm{T}} \right) \boldsymbol{\delta}_n. \tag{6}$$

Given a GS $\mathbf{x}$, SGWT coefficients are defined as

$$x_{n,s}^{\psi} \triangleq \langle \mathbf{x}, \boldsymbol{\psi}_{n,s} \rangle , \tag{7}$$

and collecting all coefficients in a vector $\mathbf{x}^{\psi}$, one can write $\mathbf{x}^{\psi} = \boldsymbol{\Psi}^{\mathrm{T}}\mathbf{x}$, where $\boldsymbol{\Psi} \in \mathbb{R}^{N \times SN}$ has columns $\boldsymbol{\psi}_{n,s}$.

Since wavelet atoms are orthogonal to the eigenvector $\mathbf{u}_1$ associated with $\lambda_1 = 0$, it is necessary to introduce a scaling function, as in classical DSP, in order to represent low frequencies. The scaling function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ must be a low-pass filter, satisfying $\phi(0) > 0$ and $\lim_{z\to\infty} \phi(z) = 0$.

A signal $\tilde{\mathbf{x}}$ is recovered from a set of SGWT coefficients $\mathbf{x}^{\psi}$ by solving the least-squares problem $\tilde{\mathbf{x}} \triangleq \operatorname{argmin} \|\mathbf{x}^{\psi} - \boldsymbol{\Psi}^{\mathrm{T}}\mathbf{y}\|_2$ over $\mathbf{y}$.

The following scaling functions ϕ and wavelet mothers ψ will be used to construct SGWTs:

(a) cubic spline [9];
(b) Meyer [13], with parameters $(a, M) = (\frac{2}{3}, 2)$;
(c) Hann [14], with parameters $(K, a_0, a_1, R) = (1, \frac{1}{2}, \frac{1}{2}, 3)$;
(d) ideal filter [15].

Designing the wavelet atoms in (6) depends on computing the Laplacian eigendecomposition, which may be prohibitive for large graphs. Filters can actually be approximated by Chebyshev polynomials on the graph Laplacian and the filtering procedure can be performed directly in the vertex domain. Chebyshev polynomials are defined on the interval $[-1, 1]$, thus requiring the mapping of $[0, \lambda_{\max}]$ into $[-1, 1]$ in order to encompass the entire Laplacian spectrum. This eventually means that only the largest eigenvalue must be computed. The Chebyshev polynomials are recursively defined as

$$C_q(\mathbf{L}) = 2\overline{\mathbf{L}} C_{q-1}(\mathbf{L}) - C_{q-2}(\mathbf{L}) \tag{8}$$
$$\text{where } \overline{\mathbf{L}} \triangleq \frac{2}{\lambda_{\max}}\mathbf{L} - \mathbf{I},\ C_0(\mathbf{L}) \triangleq \mathbf{I},\ \text{and } C_1(\mathbf{L}) \triangleq \overline{\mathbf{L}},$$

and the coefficients for approximating $\psi_s(\mathbf{L}) \triangleq \mathbf{U}\psi(\alpha_s\boldsymbol{\Lambda})\mathbf{U}^{\mathrm{T}}$ in (6) via $\widetilde{\psi}_s(\mathbf{L}) \triangleq \frac{1}{2}c_{s,0}\mathbf{I} + \sum_{q=1}^{Q} c_{s,q} C_q(\mathbf{L})$ are

$$c_{s,q} \triangleq \frac{2}{\pi} \int_0^{\pi} \cos(q\theta) \psi\left(\alpha_s \frac{\lambda_{\max}}{2} (\cos\theta + 1) \right) \mathrm{d}\theta . \tag{9}$$

The Chebyshev polynomials in (8) allows faster computations and each $C_q(\mathbf{L})$ needs to be computed once for all filters. Only the coefficients in (9) need to be updated. Fig. 1 depicts the approximation of cubic splines, Meyer kernels, Hann kernels, and ideal filters by a 50^{th}-order Chebyshev polynomial. Ideal filters are approximated by the Jackson-Chebyshev polynomial [16], which is able to cancel the ripples at the cost of expanding the transition band. This approximation is useful to enhance the stop-band attenuation. As can be seen, the cubic spline provides higher localization in low frequencies, however, it does not generate a tight frame, since $\Psi(\lambda) \triangleq |\phi(\lambda)|^2 + \sum_{s=1}^{S} |\psi(\alpha_s\lambda)|^2$, shown in

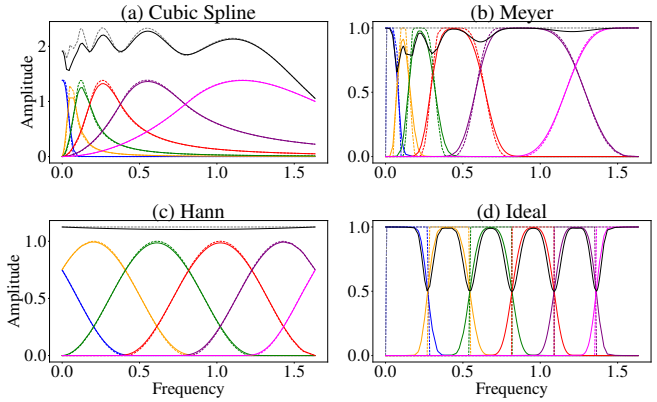

Fig. 1. Wavelets and their approximations in frequency domain for $S = 5$. One low-pass filter (in blue) and five band-pass filters (in yellow, green, red, purple, and magenta) are shown. Dashed lines correspond to the true wavelet kernel in the range $[0, \lambda_{\max}]$. Solid lines show the corresponding 50^{th}-order polynomial approximation of the wavelet kernels: (a) cubic spline; (b) Meyer kernel; (c) Hann kernel; and (d) ideal filter. Solid black lines show $\Psi(\lambda)$ for the approximations and dashed gray lines, for the actual wavelet kernels.

gray dashed line, is not constant within the interval $[0, \lambda_{\max}]$, whereas Meyer, Hann, and ideal kernels do form tight frames. The Meyer wavelet [13] needs a high-degree polynomial to achieve good approximations. The Laplacian matrix is usually sparse but the higher the degree of approximation the denser is the approximated filters. If the graph is extremely large, a high degree approximation may be infeasible not only for matrix multiplication but also for storage. In order to have a better approximation by low-order Chebyshev polynomials, one may construct wavelet kernels based on sinusoidal waves [14], [17]. The graph wavelet based on Hann kernel shown in Fig. 1(c) has uniform band-pass filters; the authors in [14] suggest to warp these Hann wavelets with a function $\omega(\lambda) = \log(\lambda)$ to achieve finest resolutions at low frequencies.

Another important concern on spectral filter design is the graph spectrum distribution. The aforementioned wavelet kernels are adapted to the interval $[0, \lambda_{\max}]$, however, as generic graph Laplacians may have a highly irregular eigenvalue distribution, it is possible to design a graph filter having no spectrum elements in the pass-band. Authors in [14] construct tight wavelet frames adapted to spectrum wavelets by warping kernel filters with a smooth approximation of the cumulative spectral function. This function can also be approximated by Chebyshev polynomials and then spectrum-adapted filters can be designed without Laplacian eigendecomposition. Note that, if nodes n and m are L-hops apart (i.e., the shortest path length between nodes m and n is L), then $(\mathbf{L}^{\ell})_{mn} = 0$ for any $\ell < L$ and therefore the polynomial degree controls the vertex localization of approximated filters.

V. Anomaly Detection in Graph Signals

Anomaly detection is the identification of instances of data differing significantly from the majority of the dataset and is applied in a variety of problems, such as fraud, intrusion, medical imaging, and event detection in sensor networks [18].

In [19], the authors applied GFT and high-pass filtering for detecting anomalies in the *whole network*, but without providing any clue about where is the corrupted measurement. In [20] a graph-filtering-based method is developed to identify anomalies in wireless sensor networks, and in [21] the SGWT is used to identify patterns in dynamic networks. Centrality measures of graphs, such as node degree, shortest path distance, and entries of $\mathbf{u}_N$, have also been used as extracted features to feed anomaly detectors [22]. Aligned with these works, we propose the following methodology for VFA-based classification to detect and localize point anomalies in GS:

(i) define the adjacency matrix $\mathbf{A}$ of the graph based on a similarity measure across nodes. Section VI provides an example of a possible choice for $\mathbf{A}$ in sensor graphs. An important assumption to this approach is that connected nodes have similar measurements.

(ii) compute a VFA representation of the raw input data. For instance, SGWT provides SN coefficients, which can all be used as new features, or just a subset of scales (e.g., coefficients of a single specific scale).

(iii) feed a machine-learning-based or statistics-based classifier with the new transformed features.

VI. Experiments

In this section, we describe some numerical examples that illustrate many of the concepts previously described in the paper. We start with a toy example to better understand the WGFT localization capabilities and limitations, then we move to another example that points to the good SGWT properties as a VFA tool, and we finish the section by showcasing the proposed methodology for anomaly detection based on VFA through a particular example of detecting a malfunctioning sensor in a Brazilian weather station network.

A. Graph description

We employed data from [23], which provides monthly average temperatures recorded by 296 Brazilian weather stations during the period of 1961-1990. The adjacency matrix $\mathbf{A}$ is constructed as in [19]: the node n is in the neighborhood $\mathcal{N}_m$ of m if n represents one of the L nearest weather stations to the station represented by node m or if m represents one of the L nearest weather stations to n. The resulting matrix is symmetric and its nonzero coefficients are

$$A_{mn} \triangleq \frac{\gamma_{mn}}{\sqrt{\left(\sum_{i\in\mathcal{N}_n} \gamma_{in}\right)\left(\sum_{j\in\mathcal{N}_m} \gamma_{jm}\right)}}, \tag{10}$$

where $\gamma_{mn} = \mathrm{e}^{-(d_{mn}^2 + h_{mn}^2)/\sigma^2}$, with d_{mn} and h_{mn} respectively denoting the geodesic distance and the height difference between stations m and n, and $\sigma > 0$ is a free parameter. Unlike [19], the difference in altitude h_{mn} is included because it is strongly correlated with temperature. In addition, all entries of $\mathbf{A}$ are divided by the largest eingenvalue in order to provide more stable operations.

B. Representation using WGFT

Consider Fig. 2, which uses the aforementioned graph structure to convey an artificially generated continuous by parts GS $\mathbf{x}$ defined as

$$x[n] = \begin{cases} 2u_{10}[n] + \epsilon, & \text{if } n \in \mathcal{N}, \\ u_{100}[n] + \epsilon, & \text{if } n \in \mathcal{S}, \end{cases} \tag{11}$$

where $\mathcal{N}$ (north) and $\mathcal{S}$ (south) are subsets of vertexes highlighted in different colors in Fig. 2(a), and ϵ is drawn from a zero-mean Gaussian distribution with variance 0.01. The Gaussian kernel $w(\lambda) = \mathrm{e}^{-\lambda^2/10}$ is used as the spectral window for computing WGFT coefficients. Fig. 2(b) depicts the resulting GS $\mathbf{x}$, whereas Figs. 2(c)-(d) show the absolute values of WGFT coefficients corresponding to the two frequencies

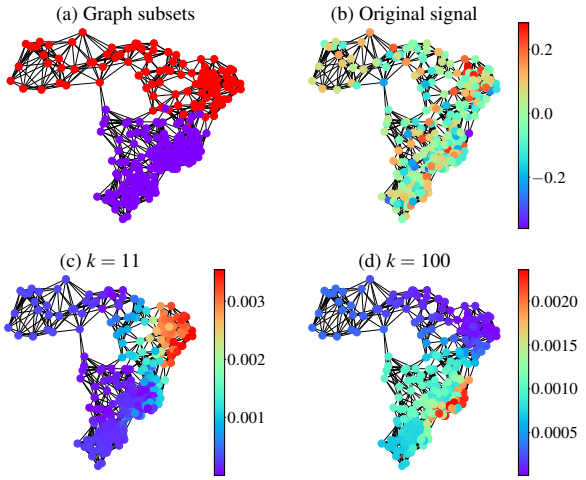

Fig. 2. Brazilian weather station graph: (a) graph divided into two sets of nodes ($\mathcal{N}$ and $\mathcal{S}$); (b) original GS $\mathbf{x}$ in (11); and (c)-(d) WGFT coefficients $x^w_{n,k}$ in (4) at 11^{th} and 100^{th} frequencies, respectively.

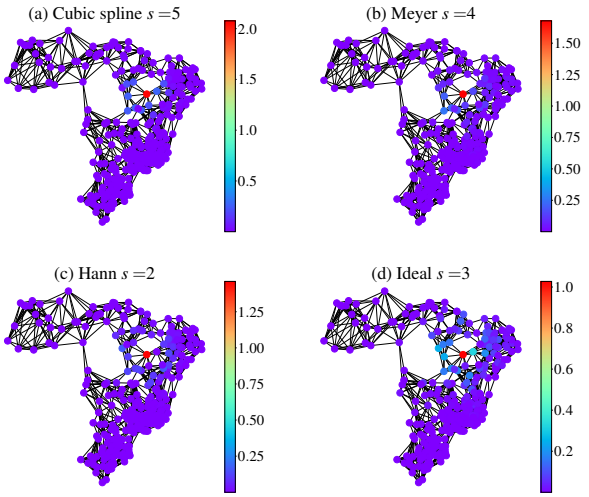

Fig. 3. Magnitude difference between SGWT coefficients of non-corrupted and corrupted December temperatures for each wavelet kernel. Each s corresponds to the scale with the highest magnitude difference.

with largest coefficients' energies. As expected, nodes in the north region present larger WGFT coefficient magnitudes for frequencies around the 10^{th} Laplacian eigenvalue, whereas nodes in the south region present larger WGFT coefficient magnitudes for frequencies at the 100^{th} eigenvalue.

Many nodes in the Brazilian north region in Fig. 2(c) achieve very small values because the eigenvector $\mathbf{u}_{11}$ is concentrated on the northeast of the graph. WGFT atoms $\mathbf{w}_{n,k}$ are not always jointly well-localized around n in vertex domain and λ_k in frequency domain, for some eigenvectors can be too much concentrated on certain vertexes. Indeed, it is true that $\mathbf{t}^w_n$ in (2) is localized around vertex n, but if $u_k[n]$ is close to zero, then $w_{n,k}[n] = t^w_n[n]u_k[n]$ in (4) will be small. Similarly, $\mathbf{m}_k(\cdot)$ in (3) is localized around λ_k in frequency domain, but if $\bar{u}_n[k]$ is close to zero, $w_{n,k}[k]$ is also going to be small. If an eigenvector $\mathbf{u}_k$ is concentrated around vertex n, it will have small values on the other vertexes, giving rise to the aforementioned problem. Therefore, simultaneous localization in both vertex and frequency domains is limited by the graph coherence $\mu \triangleq \max_{n,k} |u_k[n]| \leq 1$. The sensor graph in Fig. 2 has $\mu = 0.94$ (a large value) meaning that some of its WGFT atoms will not be well-localized, which is a limitation of the WGFT representation.

C. Representation using SGWT

We now use the actual temperature measurements of December for the GS $\mathbf{x}$. We also generated a disturbed GS by introducing an anomaly to a single vertex in the northeast region by adding 2^{o}C to its value—a mild drift from the actual value. Each graph in Fig. 3 depicts the absolute difference between SGWT coefficients $x^{\psi}_{n,s}$ of the original signal and the corrupted one, for a given scale s and with n varying in $\{1, \ldots, N\}$. Wavelet atoms were constructed from the wavelet approximation in Fig. 1. Each wavelet scale s is chosen such that the magnitude difference between coefficients of the corrupted node is the largest one. In fact, each SGWT detected a very localized variation in the transformed coefficients, but the largest magnitude differences appear in different scales of detail. For instance, the corruption of the vertex measurement was better detected by the scale of detail $s = 5$ when an SGWT with cubic spline kernel is applied, and by the scale $s = 2$ when an SGWT with Hann kernel is used. As expected, the ideal filter approximation yields the worst localization.

D. SGWT-based anomaly detector

Now, we showcase the methodology described in Section V. We started from the original dataset in [23] to create a new dataset containing some malfunctioning weather stations. Then the usual division into training and test sets is considered to design and assess the performance of a VFA-based classifier.

To generate training anomaly/corrupted data, a Gaussian noise with variance 4^{o}C^2, truncated for absolute values smaller than 0.5, was added to the temperature in June in a quarter of the nodes, which were randomly selected. The set of non-corrupted training data was generated using another quarter of nodes, but now without modifying its corresponding temperature measurements. The test set was composed by the remaining nodes, and 50% of these nodes had their measurements corrupted, in December, by the same (truncated) Gaussian distribution as in the training set.

Following the general methodology in Section V, the SGWT was applied to the six months before the current month (first semester for training and second semester for test) for each node in traning and test sets in order to feed a Gaussian process classifier (GPC) with radial basis function (RBF) kernel [24]. For example, suppose node n is in the training dataset, then SGWT was applied to the signals of temperature in January, February, March, April, May and, June. The transformed coefficients $x^{\psi}_{n,s}$ for a predefined scale s were the extracted feature that fed GPC. The underlying hyperparameters are: kernel function, number of neighbors L in the adjacency matrix, number of wavelet scales S, wavelet coefficients used in feature extraction for a given scale s (we are using only one SGWT scale as features), and the order Q of the Chebyshev approximation. Table I contains the chosen hyperparameters for each kernel function and the respective $f1$-score.[2]

Feeding GPC with raw vertex-domain data provided an average $f1$-score of 51%, much worse than using SGWT; the four VFA-based classifiers achieved an $f1$-score greater than 85% even with low-oder approximations.

Unlike the scenario described in the previous experiment,

[2] $f1$-score is a measure of accuracy that combines the precision p, which is the number of actual positive samples classified as positives divided by the total number of samples classified as positives, and recall r, which is the number of actual positive samples classified as positive divided by the number of actual positive samples. $f1$-score is then given by $2\frac{p \cdot r}{p+r}$.

TABLE I
Hiperparameters of the VFA-based classifiers and corresponding $f1$-score

Kernel	L	S	s	Q	$f1$-score
Cubic spline	10	10	3	10	88%
Meyer	20	5	4	10	87%
Hann	20	5	3	3	86%
Ideal filter	10	5	5	3	85%

where a balanced dataset was used, practical problems of point-anomaly detection have to deal with a small percentage of corrupted data. In order to evaluate the performance of the proposed VFA-based methodology from Section V in unbalanced scenario, a new dataset was generated for an unsupervised learning setup. This time, only 10% of the nodes from the entire dataset were corrupted by a Gaussian noise with variance 4°C^2 in the month of December. The SGWT was applied to the new dataset as in the supervised experiment and the transformed features fed the local outlier factor (LOF) [25] from *scikit-learn*, an unsupervised outlier detector. In order to report an average trend of the algorithm behavior without being biased by the particular choice of the corrupted nodes, the entire process (dataset definition and outlier detection) was repeated 50 times. Fig. 4 depicts the *boxplot* of the overall performance achieved by each LOF combined with the SGWT using the receiver operating characteristic (ROC) curve that measures true positive rate against false positive rate. The boxplot shows the area under the curve (AUC). We also evaluated the score of the LOF fed with raw data in vertex domain. The four VFA-based LOF classifiers outperformed the LOF classifier fed with raw input, especially the Hann kernel.

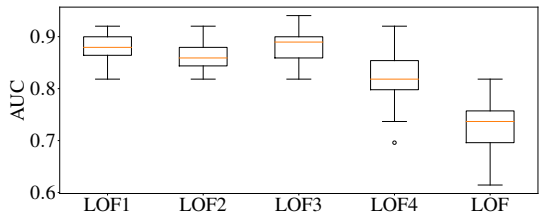

Fig. 4. *Boxplot* of AUC of ROC for each VFA-based classifier and LOF in 50 runs. LOF1, LOF2, LOF3, and LOF4 represent the LOF combined with cubic spline, Meyer, Hann and ideal SGWT, respectively. Orange lines represent the median of distributions, boxes represent the quartiles, and vertical lines extend to the most extreme non-outlier data point. Parameters in LOF are default except for contamination (set as 0.2).

VII. Conclusions

VFA is a useful tool to analyze GS simultaneously in frequency and vertex domains, being a potential tool for current big data applications. In the anomaly detection experiment, SGWT coefficients extracted good features from data, so that very small drifts in temperature measurements due to malfunctioning sensors could be detected with high accuracy by GPC and LOF, having a clear advantage over classifiers that consider only the time-series structure of temperature measurements (like feeding GPC and LOF with raw data). While WGFT is a natural extension of the standard short-time Fourier transform, it has many disadvantages, which are overcome by SGWT. Indeed, SGWT scales well for large graphs, as filtering can be performed without computing the Laplacian eigendecomposition and, depending on the application, a low-order polynomial approximation may be highly satisfactory. As future works, it would be interesting to consider wavelet transforms that are able to mix time and vertex domains and extending VFA methods for directed graphs that consider directional flow of data in networks.

References

[1] P. N. Akella, T. J. Connors, J. Kelly, C. J. Roon, A. M. Rosenbaum, B. T. Smith IV, and C. S. Tolles, "Creation and maintenance of social relationship network graphs," *Google Patents*, May, 2009.
[2] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," *IEEE Int. Conf. on Comp. Comm.*, vol. 3, pp. 1848–1857, Mar., 2003.
[3] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE Journal on Selected Areas in Comm.*, vol. 27, pp. 1029–1046, Sept., 2009.
[4] C. Chaouiya, E. Remy, P. Ruet, and D. Thieffry, "Qualitative modelling of genetic networks: From logical regulatory graphs to standard Petri nets," in *Int. Conf. on Application and Theory of Petri Nets*, pp. 137–156, 2004.
[5] L. Goldsberry, W. Huang, N. F. Wymbs, S. T. Grafton, D. S. Bassett, and A. Ribeiro, "Brain signal analytics from graph signal processing perspective," *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, pp. 851–855, May, 2017.
[6] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, pp. 808–828, Apr., 2018.
[7] L. Stanković and E. Sejdić, *Vertex-Frequency Analysis of Graph Signals*. Eds. New York, NY, USA: Springer, 2019.
[8] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "A windowed graph Fourier transform," in *IEEE Statistical Signal Process. Workshop*, pp. 133–136, 2012.
[9] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Comp. Harmonic Analysis*, vol. 30, pp. 129–150, Mar., 2011.
[10] M. Wardetzky, S. Mathur, F. Kälberer, and E. Grinspun, "Discrete Laplace operators: no free lunch," in *Symposium on Geometry Process.*, pp. 33–37, July, 2007.
[11] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE Trans. on Signal Process.*, vol. 61, pp. 1644–1656, Jan., 2013.
[12] S. Mallat, *A wavelet tour of signal processing*. London,U.K.: Academic Press, 3rd ed., 2009.
[13] N. Leonardi and D. Van De Ville, "Tight wavelet frames on multislice graphs," *IEEE Trans. on Signal Process.*, vol. 61, pp. 3357–3367, Apr., 2013.
[14] D. I. Shuman, C. Wiesmeyr, N. Holighaus, and P. Vandergheynst, "Spectrum-adapted tight graph wavelet and vertex-frequency frames," *IEEE Trans. on Signal Process.*, vol. 63, pp. 4223–4235, Apr., 2015.
[15] Y. Jin and D. I. Shuman, "An M-channel critically sampled filter bank for graph signals," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, pp. 3909–3913, June, 2017.
[16] S. Li, Y. Jin, and D. I. Shuman, "Scalable m-channel critically sampled filter banks for graph signals," *arXiv preprint arXiv:1608.03171*, 2016.
[17] A. Sakiyama, K. Watanabe, and Y. Tanaka, "Spectral graph wavelets and filter banks with low approximation error," *IEEE Trans. on Signal and Inf. Process. over Networks*, vol. 2, pp. 230–245, June, 2016.
[18] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, p. 15, July, 2009.
[19] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. on Signal Process.*, vol. 62, pp. 3042–3054, Apr., 2014.
[20] H. E. Egilmez and A. Ortega, "Spectral anomaly detection using graph-based filtering for wireless sensor networks," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, pp. 1085–1089, July, 2014.
[21] A. Dal Col, P. Valdivia, F. Petronetto, F. Dias, C. T. Silva, and L. G. Nonato, "Wavelet-based visual analysis of dynamic networks," *IEEE Trans. on Visual. and Comp. Graphics*, vol. 24, pp. 2456–2469, Aug., 2017.
[22] P.-Y. Chen, S. Choudhury, and A. O. Hero, "Multi-centrality graph spectral decompositions and their application to cyber intrusion detection," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, pp. 4553–4557, May, 2016.
[23] I. N. de Meteorologia (IMET), "Normais climatológicas do brasil." `http://www.inmet.gov.br/portal/index.php?r=clima/normaisClimatologicas`. [Online; accessed 2018].
[24] M. D. Buhmann, "Radial basis functions," *Acta Numerica*, vol. 9, pp. 1–38, Jan., 2000.
[25] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," in *ACM Sigmod Record*, vol. 29, pp. 93–104, May, 2000.