On the Transferability of Spectral Graph Filters

Levie, Ron; Isufi, Elvin; Kutyniok, Gitta

On the Transferability of Spectral Graph Filters

Ron Levie*, Elvin Isufi†, and Gitta Kutyniok*

*Technische Universität Berlin, Institut für Mathematik, Berlin, Germany

†Delft University of Technology, Department of Microelectronics, Delft, Netherlands.

Emails: levie@math.tu-berlin.de, E.Isufi-1@tudelft.nl, kutyniok@math.tu-berlin.de

Abstract—This paper focuses on spectral filters on graphs, namely filters defined as elementwise multiplication in the frequency domain of a graph. In many graph signal processing settings, it is important to transfer a filter from one graph to another. One example is in graph convolutional neural networks (ConvNets), where the dataset consists of signals defined on many different graphs, and the learned filters should generalize to signals on new graphs, not present in the training set. A necessary condition for transferability (the ability to transfer filters) is stability. Namely, given a graph filter, if we add a small perturbation to the graph, then the filter on the perturbed graph is a small perturbation of the original filter. It is a common misconception that spectral filters are not stable, and this paper aims at debunking this mistake. We introduce a space of filters, called the Cayley smoothness space, that contains the filters of state-of-the-art spectral filtering methods, and whose filters can approximate any generic spectral filter. For filters in this space, the perturbation in the filter is bounded by a constant times the perturbation in the graph, and filters in the Cayley smoothness space are thus termed linearly stable. By combining stability with the known property of equivariance, we prove that graph spectral filters are transferable.

I. INTRODUCTION

The success of convolutional neural networks (ConvNets) on Euclidean domains ignited an interest in recent years in extending these methods to graph structured data. A graph ConvNet is a mapping, that receives a signal defined over the vertices of a graph, and returns a value in some output space. A graph ConvNet consists of many layers of computations, where each layer computes a set of filters of the output of the previous layer, followed by a pointwise nonlinearity, and optionally a pooling step and non-convolution layers. In a machine learning setting, the general architecture of the ConvNet is fixed, but the specific filters to use in each layer are free parameters. In training, the filter coefficients are optimized to minimize some loss function. In some situations, both the graph and the signal defined on the graph are variables in the input space of the ConvNet. In these situations, if two graphs represent the same underlying phenomenon, and the two signals given on the two graphs are similar in some sense, the output of the ConvNet on both signals should be similar as well. This property is typically termed transferability, and is an essential requirement if we wish the ConvNet to generalize well on the test set. Analyzing and proving transferability is the subject of this paper.

A necessary condition of any reasonable definition of transferability is stability. Namely, given a filter, if the topology of a graph is perturbed, then the filter on the perturbed graph is close to the filter on the un-perturbed graph. Without stability it is not even possible to transfer a filter from a graph to another very close graph, and thus stability is necessary for transferability. Previous work studied the behavior of graph filters with respect variations in the graph. [1] provided numerical results on the robustness of polynomial graph filters to additive Gaussian perturbations of the eigenvectors of the graph Laplacian. Since the eigendecomposition is not stable to perturbations in the topology of the graph, this result does not prove robustness to such perturbations. [2] showed that the expected graph filter under random edge losses is equal to the accurate output. However, [2] did not bound the error in the output in terms of the error in the graph topology. In this paper we show the linear stability of graph filters to general perturbations in the topology.

There are generally two approaches to defining convolution on graphs, both generalizing the standard convolution on Euclidean domains [3], [4]. Spatial approaches generalize the idea of a sliding window to graphs. Here, the main challenge is to define a way to translate a filter kernel along the vertices of the graph. Some popular examples of spatial methods are [5], [6], [7]. Spectral methods are inspired by the convolution theorem in Euclidean domains, that states that convolution in the spatial domain is equivalent to pointwise multiplication in the frequency domain. The challenge here is to define the frequency domain and the Fourier transform of graphs. The basic idea is to define the graph Laplacian, or some other graph operator that we interpreted as a shift operator, and to use its eigenvalues as frequencies and its eigenvectors as the corresponding pure harmonics [8]. Decomposing a graph signal to its pure harmonic coefficients is by definition the graph Fourier transform, and filters are defined by multiplying the different frequency components by different values. For some examples of spectral methods see, e.g., [9], [10], [11], [12]. Additional references for both methods can be found in [4].

The great majority of researchers from the graph ConvNet community currently focus on developing spatial methods. One typical motivation for favoring spatial methods is the claim that spectral methods are not transferable, and thus do not generalize well on graphs unseen in the training set. The goal in this paper is to debunk this misconception, and to show that state-of-the-art spectral graph filtering methods are transferable. This paper does not argue against spatial methods, but shows the potential of spectral approaches to cope with datasets having varying graphs. We would like to

encourage researches to reconsider spectral methods in such situations. Interestingly, [13] obtained state-of-the-art results using spectral graph filters on variable graphs, without any modification to compensate for the "non-transferability".

II. PRELIMINARIES

A. Spectral graph filters

Consider an undirected weighted graph $\mathcal{G} = \{E, V, \mathbf{W}\}$, with vertices $V = \{1, \ldots, N\}$, edges E, and adjacency matrix $\mathbf{W}$. The adjacency matrix $\mathbf{W}$ is symmetric and represents the weights of the edges, where $w_{n,m}$ is nonzero only if vertex n is connected to vertex m. Consider the degree matrix $\mathbf{D}$, defined as the diagonal matrix with entries $d_{n,n} = \sum_{m=1}^{N} w_{n,m}$.

The frequency domain of a graph is determined by choosing a shift operator, namely a self-adjoint operator $\mathbf{\Delta}$ that respects the connectivity of the graph. As a prototypical example, we consider the unnormalized Laplacian $\mathbf{\Delta} = \mathbf{D} - \mathbf{W}$, which depends (affine) linearly on $\mathbf{W}$. Other examples of common shift operators are the normalized Laplacian $\mathbf{\Delta}_{\mathrm{n}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, and the adjacency matrix itself. In this paper we call a generic self-adjoint shift operator *Laplacian*, and denote it by $\mathbf{\Delta}$. Denote the eigenvalues of $\mathbf{\Delta}$ by $\{\lambda_n\}_{n=1}^{N}$, and the eigenvectors by $\{\phi_n : V \to \mathbb{C}\}_{n=1}^{N}$. The Fourier transform of a graph signal $f : V \to \mathbb{C}$ is given by the vector of frequency intensities

$$\mathcal{F}f = (\langle f, \phi_n \rangle)_{n=1}^{N},$$

where $\langle u, v \rangle$ is an inner product in $\mathbb{C}^N$, e.g., the standard dot product. The inverse Fourier transform of the vector $(v_n)_{n=1}^{N}$ is given by

$$\mathcal{F}^*(v_n)_{n=1}^{N} = \sum_{n=1}^{N} v_n \phi_n.$$

Since $\{\phi_n\}_{n=1}^{N}$ is an orthonormal basis, $\mathcal{F}^*$ is the inverse of $\mathcal{F}$. A spectral graph filter $\mathbf{G}$ based on the coefficients $(g_n)_{n=1}^{N}$ is defined by

$$\mathbf{G}f = \sum_{n=1}^{N} g_n \langle f, \phi_n \rangle \phi_n. \tag{1}$$

Any spectral filter defined by (1) is permutation *equivariant*, namely, does not depend on the indexing of the vertices. Re-indexing the vertices in the input, results in the same re-indexing of vertices in the output.

Spectral filters defined by (1) have two disadvantages. First, as shown in Subsection II-B, they are not transferable. Second, they entail high computational complexity. Formula (1) requires the computation of the eigendecomposition of the Laplacian $\mathbf{\Delta}$, which is computationally demanding and can be unstable when the number of vertices N is large. Moreover, there is no general "graph FFT" algorithm for computing the Fourier transform of a signal $f \in L^2(V)$, and (1) requires computing the frequency components $\langle f, \phi_n \rangle$ and their summation directly.

To overcome these two limitations, state-of-the-art methods, like [10], [14], [11], [12], are based on *functional calculus*. Functional calculus is the theory of applying functions $g :$

$\mathbb{C} \to \mathbb{C}$ on normal operators in Hilbert spaces. In the special case of a self-adjoint or unitary operator $\mathbf{T}$ with a discrete spectrum, $g(\mathbf{T})$ is defined by

$$g(\mathbf{T})f = \sum_n g(\lambda_n) \langle f, \phi_n \rangle \phi_n, \tag{2}$$

for any vector f in the Hilbert space, where $\{\lambda_n, \phi_n\}$ is the eigendecomposition of the operator $\mathbf{T}$. The operator $g(\mathbf{T})$ is normal for general $g : \mathbb{C} \to \mathbb{C}$, self-adjoint for $g : \mathbb{C} \to \mathbb{R}$, and unitary for $g : \mathbb{C} \to e^{i\mathbb{R}}$ (where $e^{i\mathbb{R}}$ is the unit complex circle).

Definition (2) is canonical in the following sense. In the special case where

$$g(\lambda) = \frac{\sum_{l=0}^{L} c_l \lambda^l}{\sum_{l=0}^{L} d_l \lambda^l}$$

is a rational function, $g(\mathbf{T})$ can be defined in two ways. First, by (2), and second by compositions, linear combinations, and inversions, as

$$g(\mathbf{T}) = \Big(\sum_{l=0}^{L} c_l \mathbf{T}^l \Big) \Big(\sum_{l=0}^{L} d_l \mathbf{T}^l \Big)^{-1} \tag{3}$$

It can be shown that (2) and (3) are equivalent. Moreover, definition (2) is also canonical in regard to non-rational functions. Loosly speaking, if a rational function q approximates the function g, then the operator $q(\mathbf{T})$ approximates the operator $g(\mathbf{T})$.

Implementation (3) overcomes the limitation of definition (1), where now filters are defined via (2) with polynomial or rational function g. By relying on the spatial operations of compositions, linear combinations, and inversions, the computation of a spectral filter is carried out entirely in the spatial domain, without ever resorting to spectral computations. Thus, no eigendecomposition and Fourier transforms are ever computed. The inversions in $g(\mathbf{T})f$ involve solving systems of linear equations, which can be computed directly if N is small, or by some iterative approximation method for large N. Methods like [10], [15], [8], [12] use polynomial filters, and [14], [11], [13] use rational function filters. We term spectral methods based on functional calculus *functional calculus filters*.

B. The misconception of non-transferability of spectral graph filters

The non-transferability claim is formulated based on the sensitivity of the Laplacian eigendecomposition to small perturbation in $\mathbf{W}$, or equivalently in $\mathbf{\Delta}$. Namely, a small perturbation of $\mathbf{\Delta}$ can result in a large perturbation of the eigendecomposition $\{\lambda_n, \phi_n\}_{n=1}^{N}$, which results in a large change in the filter defined via (1). This claim was stated in [3] only for spectral filters defined via (1), for which it is true. However, later papers misinterpreted this claim and applied it to functional calculus filters. The instability argument does not prove non-transferability, since state-of-the-art spectral methods do not explicitly use the eigenvectors, and do not parametrize the filter coefficients g_n via the index

n of the eigenvalues. Instead, state-of-the-art methods are based on functional calculus, and define the filter coefficients using a function $g : \mathbb{R} \to \mathbb{C}$, as $g(\lambda_n)$. The parametrization of the filter coefficients by g is indifferent to the specifics of how the spectrum is indexed, and instead represents an overall response in the frequency domain, where the *value* of each frequency determines its response, and not its index. In functional calculus filters defined by (2), a small perturbation of $\boldsymbol{\Delta}$ that results in a perturbation of λ_n, also results in a perturbation of the coefficients $g(\lambda_n)$. It turns out, as we prove in this paper, that the perturbation in $g(\lambda_n)$ implicitly compensates for the instability of the eigendecomposition, and functional calculus spectral filters are stable.

III. Main results

A. Transferability of functional calculus filters

In this paper, we define transferability as the linear robustness of the filter to re-indexing of the vertices and perturbation of the topology of the graph. Thus, to formulate transferability, we combine equivariance with stability. Since spectral filters are known to be equivariant, transferability is equivalent to stability. Thus, our goal is to prove stability.

B. Linear stability of spectral filters

Stability is proven on a dense subspace of filters in $L^p(\mathbb{R})$, which we term the Cayley smoothness space. The definition of the Cayley smoothness space is based on the Cayley transform $\mathcal{C} : \mathbb{R} \to e^{i\mathbb{R}}$, defined by $\mathcal{C}(x) = \frac{x-i}{x+i}$.

Definition 1. *The* Cayley smoothness space $Cay^1(\mathbb{R})$ *is the subspace of functions* $g \in L^2(\mathbb{R})$ *of the form* $g(\lambda) = q(\mathcal{C}(\lambda))$, *where* $q : e^{i\mathbb{R}} \to \mathbb{C}$ *is in* $L^2(e^{i\mathbb{R}})$, *and has classical Fourier coefficients* $\{c_l\}_{l=-\infty}^{\infty}$ *satisfying* $\|g\|_{\mathcal{C}} := \sum_{l=-\infty}^{\infty} l \, |c_l| < \infty$.

The mapping $g \mapsto \|g\|_{\mathcal{C}}$ is a seminorm. It is not difficult to show that $Cay^1(\mathbb{R})$ is dense in each $L^p(\mathbb{R})$ space with $1 \leq p < \infty$. Intuitively, Cayley smoothness implies decay of the filter kernel in the spatial domain, since it models smoothness in a frequency domain. This can be formulated rigorously for graph filters based on Cayley polynomials $(g(\lambda) = q(\mathcal{C}(\lambda))$ with finite expansion $\{c_l\}_{l=-L}^{L}$ [11, Theorem 4].

For filters in the Cayley smoothness space we have linear stability, which is our main contribution.

Theorem 2. *Let* $\boldsymbol{\Delta} \in \mathbb{C}^{N \times N}$ *be a self-adjoint matrix that we call Laplacian. Let* $\boldsymbol{\Delta}' = \boldsymbol{\Delta} + \mathbf{E}$ *be self-adjoint, such that* $\|\mathbf{E}\| < 1$. *Let* $g \in Cay^1(\mathbb{R})$. *Then*

$$\left\| g(\boldsymbol{\Delta}) - g(\boldsymbol{\Delta}') \right\| \leq \|g\|_{\mathcal{C}} \left((\|\boldsymbol{\Delta}\| + 1)\frac{\|\mathbf{E}\|}{1 - \|\mathbf{E}\|} + \|\mathbf{E}\| \right)$$
$$= O(\|\boldsymbol{\Delta} - \boldsymbol{\Delta}'\|).$$
(4)

IV. Examples

ChebNets. Consider the normalized translated Laplacian $\boldsymbol{\Delta}_n - \mathbf{I}$. In ChebNets [10], g is a polynomial, and since the spectrum of $\boldsymbol{\Delta}_n - \mathbf{I}$ is in $[-1, 1]$, the values of g outside $[-1, 1]$ do not affect the filter $g(\boldsymbol{\Delta}_n - \mathbf{I})$. Thus, we may assume that
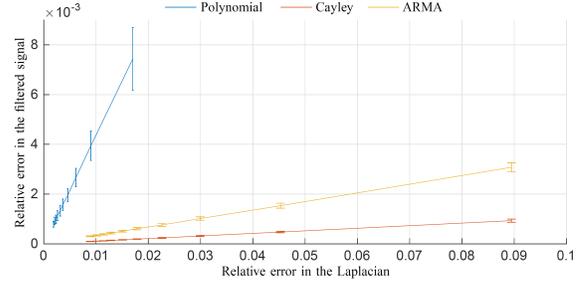

Fig. 1. Average relative error in the output of the filter as a function of the average relative error in the Laplacian, for three fixed filters: low-pass polynomial and Cayley filters of order 3, and all-pass ARMA filter of order 3.

g is a polynomial in $[-1, 1]$, and padded outside $[-1, 1]$ to obtain a smooth compactly supported function. It is easy to see that such a g is in $Cay^1(\mathbb{R})$. Thus, for two translated normalized Laplacians $\boldsymbol{\Delta}_n - \mathbf{I}$ and $\boldsymbol{\Delta}'_n - \mathbf{I}$ of two graphs, $\|g(\boldsymbol{\Delta}_n - \mathbf{I}) - g(\boldsymbol{\Delta}'_n - \mathbf{I})\| = O(\|\boldsymbol{\Delta}_n - \boldsymbol{\Delta}'_n\|)$.
General rational functions. The above claim is also true for general rational functions, if we assume that the spectrum of $\boldsymbol{\Delta}, \boldsymbol{\Delta}'$ is contained in some pre-defined band $[0, M]$. Thus, the polynomial filters of [15], [12] and the ARMA rational function filters of [14], [13] are also transferable, under the assumption of uniformly bounded Laplacians.
CayleyNets. CayleyNets [11] are always transferable, since a Cayley filter is by definition in $Cay^1(\mathbb{R})$, with finite expansion.

To corroborate the proposed theoretical result, in Figure 1 we test the above three examples in the Molene weather dataset[1]. The graph comprises $N = 32$ weather stations, with weights given as the Gaussian of the physical distances between stations. Each of the 744 graph signals is a temperature recording. For the polynomial filter we consider the normalized Laplacian, while for the Cayley and ARMA filters we consider the unnormalized Laplacian. The results are averaged over 100 different perturbations in the topology and the 744 graph signals. The experimental results concord with the theoretical linear stability property.

V. Proof of Theorem 2

We start with a useful lemma.

Lemma 3. *Suppose* $\mathbf{B}, \mathbf{D}, \mathbf{E} \in \mathbb{C}^{N \times N}$ *are matrices satisfying* $\mathbf{B} = \mathbf{D} + \mathbf{E}$, *and* $\|\mathbf{B}\|, \|\mathbf{D}\| \leq C$ *for some* $C > 0$. *Then for every* $l \geq 0$

$$\left\| \mathbf{B}^l - \mathbf{D}^l \right\| \leq lC^{l-1} \|\mathbf{E}\|.$$
(5)

Proof. Let $l \in \mathbb{N}$.

$$\mathbf{D}^l - \mathbf{B}^l = \mathbf{D}^{l-1}(\mathbf{D} - \mathbf{B}) + (\mathbf{D}^{l-1} - \mathbf{B}^{l-1})\mathbf{B}$$
(6)

so

$$\left\| \mathbf{D}^l - \mathbf{B}^l \right\| \leq C^{l-1} \|\mathbf{E}\| + \left\| \mathbf{D}^{l-1} - \mathbf{B}^{l-1} \right\| C.$$
(7)

Now, (5) follows by repeatedly using (7) with decreasing powers $l - j$, $j = 1, \ldots, l - 1$. □

Next, we cite a general property from spectral theory.

Lemma 4. *Let T be a bounded normal operator in a Hilbert space. Let σ be the spectrum of T. Define the infinity norm on the space of bounded continuous functions $f : \sigma \to \mathbb{C}$ by*

$$\|f - g\|_\infty^\sigma = \sup_{x \in \sigma} |f(x) - g(x)|.$$

Then

$$\|f(T) - g(T)\| = \|f - g\|_\infty^\sigma$$

where the norm in the left-hand-side is the operator norm.

To prove Theorem 2, we start with a version of the theorem restricted to $g = q \circ \mathcal{C} \in Cay^1(\mathbb{R})$ where q has a finite expansion with coefficients $(c_l)_{l=-L}^L$.

Proof of Theorem 2 for finite Cayley expansions. We prove for q with with coefficients $(c_l)_{l=1}^L$. The part with coefficients $(c_l)_{l=-L}^0$ is treated similarly. Note that

$$\mathcal{C}(\boldsymbol{\Delta}) - \mathcal{C}(\boldsymbol{\Delta}') = (\boldsymbol{\Delta} - i)(\boldsymbol{\Delta} + i)^{-1} - (\boldsymbol{\Delta} - i)(\boldsymbol{\Delta}' + i)^{-1}$$
$$+ (\boldsymbol{\Delta} - i)(\boldsymbol{\Delta}' + i)^{-1} - (\boldsymbol{\Delta}' - i)(\boldsymbol{\Delta}' + i)^{-1}$$

so

$$\left\| \mathcal{C}(\boldsymbol{\Delta}) - \mathcal{C}(\boldsymbol{\Delta}') \right\|$$
$$\leq \left\| (\boldsymbol{\Delta} - i) \right\| \left\| (\boldsymbol{\Delta} + i)^{-1} - (\boldsymbol{\Delta}' + i)^{-1} \right\| + \left\| (\boldsymbol{\Delta}' + i)^{-1} \right\| \|\mathbf{E}\|.$$

By the fact that the spectrum of $\boldsymbol{\Delta}'$ is real, $\left\| (\boldsymbol{\Delta}' + i)^{-1} \right\| \leq 1$, and we have

$$\left\| \mathcal{C}(\boldsymbol{\Delta}) - \mathcal{C}(\boldsymbol{\Delta}') \right\|$$
$$\leq (\|\boldsymbol{\Delta}\| + 1) \left\| (\boldsymbol{\Delta} + i)^{-1} - (\boldsymbol{\Delta}' + i)^{-1} \right\| + \|\mathbf{E}\|. \qquad (8)$$

Let us bound $\left\| (\boldsymbol{\Delta} + i)^{-1} - (\boldsymbol{\Delta}' + i)^{-1} \right\|$ in terms of $\|\mathbf{E}\|$. Since $\|\mathbf{E}\| < 1$ we may expand

$$(\boldsymbol{\Delta} + i + \mathbf{E})^{-1} = (\boldsymbol{\Delta} + i)^{-1} \left(\mathbf{I} + \mathbf{E}(\boldsymbol{\Delta} + i)^{-1} \right)^{-1}$$
$$= (\boldsymbol{\Delta} + i)^{-1} \left(\sum_{k=0}^\infty (-1)^k (\mathbf{E}(\boldsymbol{\Delta} + i)^{-1})^k \right)$$

so, by $\left\| (\boldsymbol{\Delta} + i)^{-1} \right\| \leq 1$,

$$\left\| (\boldsymbol{\Delta} + i)^{-1} - (\boldsymbol{\Delta}' + i)^{-1} \right\| \leq \frac{\|\mathbf{E}\|}{1 - \|\mathbf{E}\|}. \qquad (9)$$

Now, by (8) and (9),

$$\left\| \mathcal{C}(\boldsymbol{\Delta}) - \mathcal{C}(\boldsymbol{\Delta}') \right\| \leq (\|\boldsymbol{\Delta}\| + 1) \frac{\|\mathbf{E}\|}{1 - \|\mathbf{E}\|} + \|\mathbf{E}\|. \qquad (10)$$

Observe that $\mathcal{C}(\boldsymbol{\Delta})$ and $\mathcal{C}(\boldsymbol{\Delta}')$ are unitary, so their spectrum is bounded by $C = 1$. Thus, by Lemma 3 and the triangle inequality on the polynomial expansions of $q(\mathcal{C}(\boldsymbol{\Delta})) - q(\mathcal{C}(\boldsymbol{\Delta}'))$,

$$\left\| g(\boldsymbol{\Delta}) - g(\boldsymbol{\Delta}') \right\| = \left\| q(\mathcal{C}(\boldsymbol{\Delta})) - q(\mathcal{C}(\boldsymbol{\Delta}')) \right\|$$
$$\leq \sum_{l=1}^L l c_l \left\| \mathcal{C}(\boldsymbol{\Delta}) - \mathcal{C}(\boldsymbol{\Delta}') \right\|$$

which gives (4). □

Proof of Theorem 2. Theorem 2 follows the above result by a simple density argument. Given $g = q \circ \mathcal{C} \in Cay^1(\mathbb{R})$, we consider the truncations $g_L = q_L \circ \mathcal{C}$, where q_L is restricted to the coefficients $(c_l)_{l=-L}^L$. We base a three-epsilon argument on the expansion, for any $L \in \mathbb{N}$,

$$\left\| g(\boldsymbol{\Delta}) - g(\boldsymbol{\Delta}') \right\| \leq \left\| g(\boldsymbol{\Delta}) - g_L(\boldsymbol{\Delta}) \right\|$$
$$+ \left\| g_L(\boldsymbol{\Delta}) - g_L(\boldsymbol{\Delta}') \right\| \qquad (11)$$
$$+ \left\| g_L(\boldsymbol{\Delta}') - g(\boldsymbol{\Delta}') \right\|.$$

For any $\epsilon > 0$, by Lemma 4, the first and the last terms of the right-hand side of 11 can be made smaller than $\epsilon/2$ by choosing L large enough. Moreover, for any $L \in \mathbb{N}$,

$$\|g_L\|_{\mathcal{C}} = \sum_{l=-L}^L l |c_l| \leq \sum_{l=-\infty}^\infty l |c_l| = \|g\|_{\mathcal{C}},$$

so, by Theorem 2 for finite Cayley expansions, for every $\epsilon > 0$

$$\left\| g(\boldsymbol{\Delta}) - g(\boldsymbol{\Delta}') \right\| \leq \|g\|_{\mathcal{C}} \left((\|\boldsymbol{\Delta}\| + 1) \frac{\|\mathbf{E}\|}{1 - \|\mathbf{E}\|} + \|\mathbf{E}\| \right) + \epsilon. \qquad (12)$$

Since (12) is true for every $\epsilon > 0$, we must have (4). □

REFERENCES

[1] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 4117–4131, 2017.

[2] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Filtering random graph processes over random time-varying graphs," *IEEE Transactions on Signal Processing*, vol. 65, no. EPFL-ARTICLE-230521, pp. 4406–4421, 2017.

[3] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, July 2017.

[4] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *arXiv preprint arXiv:1901.00596*, 2019.

[5] M. Gori, G. Monfardini, and F. Scarselli, "A new model for learning in graph domains," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 2, July 2005, pp. 729–734 vol. 2.

[6] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, Jan 2009.

[7] F. Monti, D. Boscaini, J. Masci, E. Rodolà, J. Svoboda, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model cnns," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5425–5434, 2017.

[8] 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, no. 5, pp. 808–828, 2018.

[9] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *Proc. ICLR*, 2013.

[10] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. NIPS*, 2016.

[11] R. Levie, F. Monti, X. Bresson, and M. M. Bronstein, "Cayleynets: Graph convolutional neural networks with complex rational spectral filters," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 97–109, Jan 2019.

[12] F. Gama, A. G. Marques, G. Leus, and A. Ribeiro, "Convolutional neural networks architectures for signals supported on graphs," *arXiv preprint arXiv:1805.00165*, 2018.

[13] F. M. Bianchi, D. Grattarola, L. Livi, and C. Alippi, "Graph neural networks with convolutional arma filters," *arXiv preprint arXiv:1901.01343*, 2019.

[14] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive moving average graph filtering," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 274–288, Jan 2017.

[15] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *In Proceedings of International Conference on Learning Representations (ICLR)*, 2017.