

# ON THE STABILITY OF POLYNOMIAL SPECTRAL GRAPH FILTERS

Henry Kenlay<sup>†</sup>      Dorina Thanou<sup>‡</sup>      Xiaowen Dong<sup>†</sup>

<sup>†</sup>University of Oxford      <sup>‡</sup>Swiss Data Science Center

## ABSTRACT

Spectral graph filters are a key component in state-of-the-art machine learning models used for graph-based learning, such as graph neural networks. For certain tasks stability of the spectral graph filters is important for learning suitable representations. Understanding the type of structural perturbation to which spectral graph filters are robust lets us reason as to when we may expect them to be well suited to a learning task. In this work, we first prove that polynomial graph filters are stable with respect to the change in the normalised graph Laplacian matrix. We then show empirically that properties of a structural perturbation, specifically the relative locality of the edges removed in a binary graph, effect the change in the normalised graph Laplacian. Together, our results have implications on designing robust graph filters and representations under structural perturbation.

**Index Terms**— Graph signal processing, spectral graph filters, stability analysis

## 1. INTRODUCTION

In a multitude of different fields, one may be tasked with the analysis of data which naturally resides in a graph-structured domain. Modern machine learning techniques are not well suited to handle this raw data as input. As a practical example, consider a social network where users are represented by nodes in a graph and edges between users represent a social connection. Given access to information about the users, and the knowledge of how a small number of the users plan to vote in an election, can we learn to predict how other users will vote? A natural framework for this learning task is graph-based semi-supervised learning [1]. An explicit assumption in this example is the homophily principle which states that individuals tend to associate themselves with others who are similar [2], hence voting tends to be similar between connected individuals. Recently, graph signal processing has emerged as a field which extends high dimensional data analysis to graph-structured data [3]. One of the challenges in graph signal processing is to adapt ideas from the signal processing literature and to generalise them to signals which reside in a graph-structured domain.

A fundamental operation in the analysis and transformation of signals is filtering. Filtering amplifies or attenuates frequencies of the input signal by performing pointwise multiplication of the signal in the frequency domain with a so-called transfer function evaluated at the frequencies. This idea can be generalised to filtering of signals in a graph-structured domain. The challenge then is to define a suitable frequency domain for signals defined on graphs. Recent advances in spectral graph theory and graph signal processing provide us with a notion of a frequency domain on graphs via the graph Fourier transform. The graph Fourier transform helps define the concept of a filtering operation as well as convolution for signals defined

on graphs [3]. This enables the design of efficient graph representation learning models, such as graph neural networks [4].

When utilising spectral graph filters for learning representations a necessary condition for transferability in certain tasks is stability. Stability can be loosely defined to be the property such that if we add a small perturbation to the input graph, the output of the filter is also perturbed by a small amount. In the given semi-supervised learning example, small changes to people’s social circles is unlikely to drastically change the way they vote. Based on this reasoning, if we use graph filters as part of an end-to-end graph-based classification system we would desire the property of stability. However, bounding the change in filter output by the absolute change in a graph matrix, e.g., the normalised graph Laplacian, may not be desirable as the latter is not always a natural or interpretable metric of change. In the social network example, a more natural unit of change is the addition or deletion of an edge (a friendship). Given a structural perturbation to a graph under a fixed budget, e.g., number of edges allowed to be changed, it is possible to observe different magnitudes of change in the normalised Laplacian matrix of the graph.

Our first contribution in this work is to show that polynomial spectral graph filters are stable, by proving that the change in the output of the filters is linearly bounded by that in the normalised graph Laplacian matrix. This result is similar to a recent work of Levie et al. [5] which proves that spectral graph filters in the Cayley smoothness space (which includes polynomial filters) are linearly stable, but does not require the perturbations to be sufficiently small as in their case. Gama et al. [6] study as well the stability of convolutional graph neural networks using spectral graph filters for convolutional layers. However, compared to our work, they consider a different metric to measure the distance between two graphs. Our work constitutes another approach to stability analysis of spectral graph filters, and contributes more generally to the existing literature on robustness of graph signal processing and graph analysis tools [7], [8], [9], [10]. Our second contribution is to make the first step towards demonstrating the importance of understanding the structure of perturbation in the graph, by showing empirically that the change in the normalised graph Laplacian, given a fixed budget of edge removals, is affected by the relative locality of these edges in the graph. We believe that the combination of these two lines of investigation will pave the way to designing more stable graph filters in light of the nature of the structural perturbation in the graph, with implications in scenarios such as adversarial attacks and defence [11].

## 2. PRELIMINARIES

We consider an undirected graph without self loops  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$  with  $n = |\mathcal{V}|$  vertices and  $m = |\mathcal{E}|$  edges and an adjacency matrix  $\mathbf{W}$ . The non-zero entry of the adjacency matrix  $\mathbf{W}_{ij}$  denotes the weight of an edge  $\{i, j\} \in \mathcal{E}$ . For binary weighted graphs the degree  $d_u$  of the node  $u$  is the number of nodes adjacent to  $u$ . A node with degree 0 is called an isolated node. The  $k$ -hop neighbourhood

This work was supported by the EPSRC Centre for Doctoral Training in Autonomous Intelligent Machines and Systems EP/L015897/1.

of a node  $u$  in a graph is the set of all nodes which can be reached from  $u$  via a path of at most  $k$  edges. The normalised Laplacian matrix of a graph is defined to be  $\mathcal{L}(\mathcal{G}) = \mathbf{I}_n - \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}$  where  $\mathbf{I}_n \in \mathbb{R}^{n \times n}$  is the  $n \times n$  identity matrix and  $\mathbf{D}$  is a diagonal matrix with  $\mathbf{D}_{ii} = d_i$ . As the normalised Laplacian is a real symmetric matrix, it admits an eigendecomposition  $\mathcal{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$ . Here  $\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{n-1})$  is the diagonal matrix of real eigenvalues of  $\mathcal{L}$  in increasing order and  $\mathbf{U} = [\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{n-1}]$  is the orthonormal matrix of corresponding eigenvectors as columns. It is a well known result that the eigenvalues are contained in the interval  $[0, 2]$  with  $\lambda_0 = 0$ , and  $\lambda_{n-1} = 2$  if and only if  $\mathcal{G}$  is bipartite [12].

A signal on a graph is a function  $x : \mathcal{V} \rightarrow \mathbb{R}$  that can be compactly represented as a vector  $\in \mathbb{R}^n$  with  $i$  representing the signal value of node  $i$ . The graph Fourier transform of a graph signal is defined as  $\hat{x} = \mathbf{U}^T x$  and the inverse graph Fourier transform is given by  $\mathbf{U} \hat{x}$ . Filtering in graph signal processing can be defined as operating in the frequency domain using a smooth transfer function  $g(\lambda_i)$  which amplifies or attenuates each of the frequency components  $\mathbf{u}_i$  of the graph signal. Computing the eigendecomposition explicitly is prohibitively expensive for large graphs; however, we can instead operate directly on the normalised Laplacian matrix:

$$\mathbf{U} \text{diag}(g(\lambda_0), \dots, g(\lambda_{n-1})) \mathbf{U}^T = g(\mathcal{L}),$$

to obtain an equivalent result. For diagonalisable matrices this is one of the multiple equivalent definitions of matrix functions [13].

An order  $K$  polynomial graph filter applied to the normalised graph Laplacian matrix  $\mathcal{L}$  and graph signal is computed as so:

$$g_\theta(\mathcal{L}) = \sum_{k=0}^K \theta_k \mathcal{L}^k,$$

where  $\theta = (\theta_0, \dots, \theta_K) \in \mathbb{R}^{K+1}$  are the parameters of the filter. The graph filter  $g_\theta : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$  is a matrix-valued function. A scaling of the normalised Laplacian matrix can be used to ensure that the eigenvalues lie in the range  $[-1, 1]$  to improve numerical stability of the filtering operation. In [14] the scaled normalised Laplacian is defined to be  $2\mathcal{L}/\lambda_{n-1} - \mathbf{I}_n$ . In [15] the simpler scaling of  $\tilde{\mathcal{L}} = \mathcal{L} - \mathbf{I}_n$  was used which does not require calculating the largest eigenvalue of the normalised Laplacian matrix. We will adopt the latter of these two scalings and refer to this matrix as the scaled normalised Laplacian matrix.

### 3. STABILITY OF POLYNOMIAL FILTERS

In this section, we prove that polynomial graph filters are stable. In particular, we state that a spectral graph filter  $g$  is (linearly) stable if

$$\|g(\mathcal{L}) - g(\mathcal{L}_p)\|_2 \in \mathcal{O}(\|\mathcal{L} - \mathcal{L}_p\|_2), \quad (1)$$

where  $\mathcal{L}_p$  represents the normalised Laplacian of the perturbed input graph and  $\|\mathbf{A}\|_2$  is the operator norm of a matrix  $\mathbf{A}$ . We call the left hand side of Eq. (1) the filter distance, and  $\|\mathcal{L} - \mathcal{L}_p\|_2$  the Laplacian distance.

We begin by motivating why we wish to bound the filter distance. Assume the signal is non-zero. Consider the relative output distance of a polynomial graph filter  $g_\theta$  relative to some perturbation

$$\frac{\|g_\theta(\mathcal{L}) - g_\theta(\mathcal{L}_p)\|_2}{\|\mathcal{L}\|_2},$$

where  $\mathcal{L}$  is the input normalised graph Laplacian and  $\mathcal{L}_p$  is the normalised graph Laplacian of the perturbed graph. In our setup, we assume that the signal parameters and the input signal are fixed.

We can see that by definition the relative output distance is bounded by the filter distance

$$\frac{\|g_\theta(\mathcal{L}) - g_\theta(\mathcal{L}_p)\|_2}{\|\mathcal{L}\|_2} \leq \sup_{\neq 0} \frac{\|g_\theta(\mathcal{L}) - g_\theta(\mathcal{L}_p)\|_2}{\|\mathcal{L}\|_2},$$

$$\stackrel{\text{def}}{=} \|g_\theta(\mathcal{L}) - g_\theta(\mathcal{L}_p)\|_2.$$

The looseness of this bound depends on the signal in relation to where the perturbation is taking place in the graph.

The simplest case to consider is the first order polynomial graph filter (i.e.,  $K = 1$ ) where  $\|g_\theta(\mathcal{L}) - g_\theta(\mathcal{L}_p)\|_2 = |\theta_1| \|\mathcal{L} - \mathcal{L}_p\|_2$ . Thus, the output distance is proportional to the Laplacian distance.

#### 3.1. Higher-order filters

We now consider higher-order polynomial filters. We prove that the filter distance is bounded above by some constant times the Laplacian distance, where the constant depends on the filter parameters. To do so, we will use Taylor's Theorem for matrix-valued functions where the magnitude of the remainder term under the 2-norm gives us exactly the filter distance. We will then bound the remainder term to show our main result. We begin by stating Taylor's theorem for matrix-valued functions.

**Theorem 1** (Theorem 2.2, [16]). *Let  $f$  have a power series expansion about the origin with radius of convergence  $r$  and let  $\mathcal{D} \subset \mathbb{C}$  be a simply connected open set within the circle of radius  $r$  centred at 0. Let  $\mathbf{A}, \mathbf{E} \in \mathbb{C}^{n \times n}$  be such that  $\Lambda(\mathbf{A}), \Lambda(\mathbf{A} + \mathbf{E}) \subset \mathcal{D}$ . Then for any  $k \in \mathbb{N}$*

$$f(\mathbf{A} + \mathbf{E}) = T_k(\mathbf{A}, \mathbf{E}) + R_k(\mathbf{A}, \mathbf{E}), \quad (2)$$

where

$$T_k(\mathbf{A}, \mathbf{E}) = \sum_{j=0}^k \frac{1}{j!} D_f^{[j]}(\mathbf{A}, \mathbf{E}),$$

$$R_k(\mathbf{A}, \mathbf{E}) =$$

$$\frac{1}{2\pi i} \int_{\Gamma} f(z)(z\mathbf{I}_n - \mathbf{A} - \mathbf{E})^{-1} [\mathbf{E}(z\mathbf{I}_n - \mathbf{A})^{-1}]^{k+1} dz,$$

and  $\Gamma$  is a closed contour in  $\mathcal{D}$  enclosing  $\Lambda(\mathbf{A})$  and  $\Lambda(\mathbf{A} + \mathbf{E})$ .

In the theorem, the terms

$$D_f^{[j]}(\mathbf{A}, \mathbf{E}) = \left. \frac{d^j}{dt^j} \right|_{t=0} f(\mathbf{A} + t\mathbf{E})$$

denote the  $j$ th order Fréchet derivatives [17]. The term  $R_k(\mathbf{A}, \mathbf{E})$  in Eq. (2) is the remainder term in the Taylor expansion which the following Lemma provides an analytic bound for.

**Lemma 1** (Lemma 3.1, [16]). *Let  $f$  and  $\mathcal{D}$  satisfy the criteria of Theorem 1. Furthermore, let  $\epsilon > 0$  be such that  $\Lambda_\epsilon(\mathbf{A}) \subset \mathcal{D}$  and  $\Lambda_\epsilon(\mathbf{A} + \mathbf{E}) \subset \mathcal{D}$ , and take  $\tilde{\Gamma}_\epsilon \subset \mathcal{D}$  to be a closed contour that encloses both  $\Lambda_\epsilon(\mathbf{A})$  and  $\Lambda_\epsilon(\mathbf{A} + \mathbf{E})$ . Then the remainder term  $R_k(\mathbf{A}, \mathbf{E})$  is bounded by*

$$\|R_k(\mathbf{A}, \mathbf{E})\| \leq \frac{\|\mathbf{E}\|^{k+1} \tilde{L}_\epsilon}{2\pi \epsilon^{k+2}} \max_{z \in \tilde{\Gamma}_\epsilon} |f(z)|,$$

where  $\tilde{L}_\epsilon$  is the length of  $\tilde{\Gamma}_\epsilon$ . In particular, when a circular contour centred at 0 is used,

$$\|R_k(\mathbf{A}, \mathbf{E})\| \leq \frac{\|\mathbf{E}\|^{k+1} \tilde{\rho}_\epsilon}{\epsilon^{k+2}} \max_{\phi \in [0, 2\pi]} \left| f(\tilde{\rho}_\epsilon e^{i\phi}) \right|, \quad (3)$$

where  $\tilde{\rho}_\epsilon = \max \{|z| : z \in \Lambda_\epsilon(\mathbf{A} + \mathbf{E}) \cap \Lambda_\epsilon(\mathbf{A})\}$  is the radius of the circle.

In the above lemma,  $\Lambda_\epsilon(\mathbf{X}) = \{z \in \mathbb{C} : \|(z\mathbf{I}_n - \mathbf{A})^{-1}\| \geq \epsilon^{-1}\}$  is the  $\epsilon$ -pseudospectrum of a matrix  $\mathbf{X} \in \mathbb{C}^{n \times n}$  [18].

The following lemmas will be used to further bound the remainder term given in Eq. (3). The proofs are given in Appendix A.

**Lemma 2.** Let  $f(z)$  be a complex polynomial of degree  $K$ , if  $|z| \geq 1$  then

$$|f(z)| \leq |z|^K \max_{\phi \in [0, 2\pi]} |f(e^{i\phi})|.$$

**Lemma 3.** Let  $K \geq 2$  then

$$\min_{\epsilon > 0} \frac{(1 + \epsilon)^{K+1}}{\epsilon^2} = \frac{1}{4}(K^2 - 1) \left( \frac{K+1}{K-1} \right)^K.$$

We now use the above developments to state and prove the main result of this section.

**Theorem 2.** Consider a polynomial graph filter of order  $K$ :

$$g_\theta(\mathcal{L}) = \sum_{k=0}^K \theta_k \tilde{\mathcal{L}}^k,$$

where  $\theta \in \mathbb{R}^{K+1}$  are the polynomial coefficients and  $\tilde{\mathcal{L}}$  is the scaled normalised Laplacian of an input graph. Consider perturbing the graph and let the Laplacian of the perturbed graph be  $\mathcal{L}_p$ . Then the following holds:

$$\|g_\theta(\mathcal{L}) - g_\theta(\mathcal{L}_p)\|_2 \leq \frac{1}{4} \|\theta_{-0}\|_1 (K^2 - 1) \left( \frac{K+1}{K-1} \right)^K \|\mathcal{L} - \mathcal{L}_p\|_2,$$

where  $\theta_{-0} = (\theta_1, \dots, \theta_K)$  is the vector of polynomial coefficients for all terms apart from the constant term, with the 1-norm of a vector is defined as  $\|\cdot\|_1 = \sum_i |\cdot|$ .

*Proof.* We proceed by applying Theorem 1 to our polynomial graph filter. Polynomial functions have an infinite radius of convergence so we may take  $\mathcal{D} = \mathbb{C}$ . According to Eq. (2) with  $k = 0$  we get

$$g_\theta(\mathcal{L}_p) - g_\theta(\mathcal{L}) = R_0(\mathcal{L}, \mathcal{L}_p - \mathcal{L}).$$

Under the 2-norm, the  $\epsilon$ -pseudospectrum  $\Lambda_\epsilon(A)$  of a normal matrix  $A$  is the union of open balls of radius  $\epsilon$  around the eigenvalues of  $A$  [18, Theorem 2.2]. In particular, if a point lies in the  $\epsilon$ -pseudospectrum of a scaled normalised Laplacian matrix, it is at most distance  $\epsilon$  from some point in  $[-1, 1]$ . Thus, by the triangle inequality all points in the  $\epsilon$ -pseudospectrum are at most distance  $1 + \epsilon$  from the origin. Using this, we can bound the remainder term by Lemma 1 using Eq. (3) using a circle contour of radius  $1 + \epsilon$  centred at 0. This gives us that

$$\|g_\theta(\mathcal{L}) - g_\theta(\mathcal{L}_p)\|_2 \leq \frac{\|\mathcal{L} - \mathcal{L}_p\|_2 (1 + \epsilon)}{\epsilon^2} \max_{\phi \in [0, 2\pi]} |g_\theta((1 + \epsilon)e^{i\phi})|.$$

By taking  $z = (1 + \epsilon)e^{i\theta}$  in Lemma 2 we get

$$\|g_\theta(\mathcal{L}) - g_\theta(\mathcal{L}_p)\|_2 \leq \frac{\|\mathcal{L} - \mathcal{L}_p\|_2 (1 + \epsilon)^{K+1}}{\epsilon^2} \max_{\phi \in [0, 2\pi]} |g_\theta(e^{i\phi})|.$$

Since  $\epsilon$  is an arbitrary non-negative scalar value, we may minimise the right hand side with respect to  $\epsilon$ . By Lemma 3 we get

$$\|g_\theta(\mathcal{L}) - g_\theta(\mathcal{L}_p)\|_2 \leq \frac{1}{4}(K^2 - 1) \left( \frac{K+1}{K-1} \right)^K \|\mathcal{L} - \mathcal{L}_p\|_2 \max_{\phi \in [0, 2\pi]} |g_\theta(e^{i\phi})|.$$

We give a naive upper bound to the magnitude of the polynomial over the unit circle by noting that for all  $\phi \in [0, 2\pi]$  we have that

$$|g_\theta(e^{i\phi})| = \left| \sum_{k=0}^K \theta_k e^{i\phi k} \right| \leq \sum_{k=0}^K |\theta_k| |e^{i\phi k}| = \|\theta\|_1. \quad (4)$$

Finally, note that the  $\theta_0 \mathbf{I}_n$  terms cancel out in the calculation of  $g_\theta(\mathcal{L}) - g_\theta(\mathcal{L}_p)$  so we may, without loss of generality, assume  $\theta_0$  is zero. Using this observation, combined with the bound given in Eq. (4), we obtain the result.  $\square$

Theorem 2 shows that the difference between the outputs of two polynomial filters is bounded by a quantity which scales linearly with the Laplacian distance, hence proves the stability of polynomial spectral filters. Empirically, however, we found this bound to be quite loose and the development of practical bounds is an open research direction.

#### 4. IMPORTANCE OF STRUCTURAL PERTURBATION

Theorem 2 bounds the filter distance by the Laplacian distance. However, absolute change in the graph Laplacian is not a quantity that is easy to interpret, in the sense that it does not provide any insight into the actual structural perturbation of the graph. Furthermore, different structural perturbation under a fixed budget can lead to different magnitudes of change in the Laplacian. For example, considering perturbing undirected binary graphs by removing a fixed number of edges, the locality of these edges in the graph would lead to different Laplacian distance.

Our intuition for this is guided by the way the entries of the Laplacian matrix change under perturbation of the graph topology. Note that if the edges are sufficiently far apart, the changes in the matrix incurred by each edge removal is likely to be independent. However, if two edges are close then an entry may change due to both edge removals. Based on the same intuition of considering the change of entries of the Laplacian matrix, we expect the degree of the endpoints of the edges we remove to also play a role. We leave further investigation of this as future work.

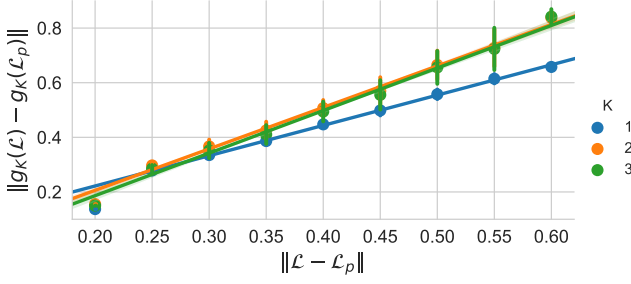
To quantify the relative locality of a set of edges, we propose the following definition.

**Definition 1.** A set of edges  $\mathcal{R} \subseteq \mathcal{E}$  is  $k$ -hop localised in a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with respect to a node  $v \in \mathcal{V}$  if the edge set of the graph induced by the  $k$ -hop neighbourhood of  $v$  contains all edges in  $\mathcal{R}$ .

In this paper, we empirically test the effect of relative locality of edge removals in affecting the change in the normalised graph Laplacian, under different types of random graph models. We present these results in the following section.

#### 5. EXPERIMENTS

We begin this section by showing that the filter distance empirically scales linearly with the Laplacian distance in line with the result of Theorem 2. We then test our hypothesis that the locality of edges



**Fig. 1:** A plot of the Laplacian distance and the filter distance for different order polynomial filters. The bars indicate the standard deviation of the filter distance.

being removed affects the magnitude of the Laplacian distance as discussed in Section 4. Code for reproducing the results in this section is available online<sup>1</sup>.

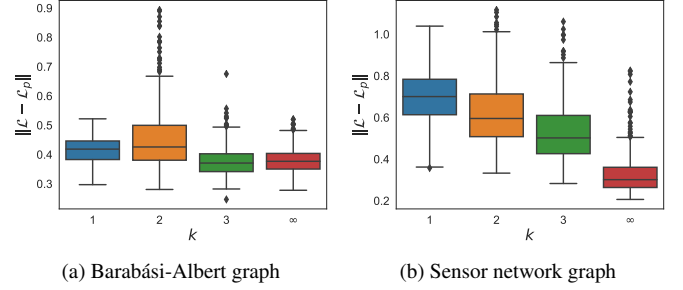
To demonstrate the  $\mathcal{O}(\|L - L_p\|_2)$  property of the result of Theorem 2 we consider the following experimental setup. We generate Barabási-Albert graphs with  $n = 200$  nodes and randomly remove each edge with independent probability of 0.5 to give a perturbed graph [19]. We only consider perturbations which do not disconnect the graph. We then look at the filter distance for low pass polynomial filter of order  $K \in \{1, 2, 3\}$ . The results are shown in Fig 1. The filter distance can be seen to scale linearly with the Laplacian distance consistent with Theorem 2 which states that polynomial filters are linearly stable.

To test the effect that the locality of edge removals has on the Laplacian distance we conducted repeats of the following experiment. We perturbed a graph by randomly selecting a node and remove a fixed number of edges in the  $k$ -hop subgraph around that node. We also considered removing edges uniformly at random (we say in this case the edges are  $\infty$ -hop localised). We only consider an experiment valid if the graph remains connected. We consider two random graph models. The first random graph model we consider is the Barabási-Albert random graph model with  $n = 500$  where we connect a node with 3 edges at each step of the network generation. The second random graph model we consider is a sensor network model. Synthetic sensor networks are generated by uniformly sampling  $n = 500$  points in the unit square and connecting nodes which are under some threshold distance. The settings of these experiments are summarised in Table 1; obtained results are illustrated in Fig. 2.

In the experiments, we see that as we increase the size of the neighbourhood, the median Laplacian distance is reduced. The effect is more pronounced for the sensor network graph. We believe this is due to the sensor networks greater diameter, leading the perturbations under different  $k$ -hop neighbourhoods to be more distinct.

**Table 1:** Summary statistics of the graphs used in experiments

	Barabási-Albert	Sensor network
Number of edges $m$	1491	2724-3130
Diameter	5-6	18-24
Edges removed	74 (5%)	136-156 (5%)



**Fig. 2:** Varying effects of removing a fixed percentage of edges with different localisation.

## 6. CONCLUSION

In this work, we proved that polynomial graph filters are linearly stable with respect to the change in the normalised Laplacian. Furthermore, we have demonstrated empirically how the distribution of the observed Laplacian distance can change under a fixed structural perturbation budget. Further research is required to develop a theoretical understanding of how the normalised Laplacian changes under structural perturbation. In future work, we hope to understand how stability of graph filters with respect to the underlying topology can effect the robustness of graph-based representations and models, e.g., the ones that utilise graph filters in a learning setting such as graph neural networks [4].

### A. PROOFS OF LEMMAS

#### A.1. Proof of Lemma 2

*Proof.* The function  $z^K f(1/z)$  is holomorphic in the disk  $|z| \leq 1$  and thus reaches its maximum on the boundary  $|z| = 1$  by the maximum modulus principle. For  $|z| \leq 1$ ,

$$|z^K f(1/z)| = |z|^K |f(1/z)| \leq \max_{|z|=1} f(1/z) = \max_{|z|=1} f(z). \quad (5)$$

Now, let  $w$  be so that  $|w| \geq 1$  then  $z = w^{-1}$  is inside the unit disk. Using Eq. (5) we get

$$|w|^{-K} |f(w)| = |z|^K |f(1/z)| \leq \max_{|z|=1} f(z). \quad (6)$$

Multiplying the inequality by  $|w|^K$  yields the result.  $\square$

#### A.2. Proof of Lemma 3

*Proof.* Taking the derivative of the expression with respect to  $\epsilon$  we get

$$\frac{\partial}{\partial \epsilon} \frac{(1 + \epsilon)^{K+1}}{\epsilon^2} = \frac{(1 + \epsilon)^K ((K - 1)\epsilon - 2)}{\epsilon^3}$$

which has a unique zero in the positive reals given by  $\epsilon^* = 2/(K - 1)$ . The second derivative is calculated to be

$$\frac{\partial^2}{\partial \epsilon^2} \frac{(1 + \epsilon)^{K+1}}{\epsilon^2} = \frac{(K^2 - 3K + 2)\epsilon^2 - 4(K - 2)\epsilon + 6}{\epsilon^4 (\epsilon + 1)^{1-K}}.$$

It can be verified that this is strictly positive when evaluated at  $\epsilon^*$ . By plugging  $\epsilon^*$  into the original expression we obtain the result.  $\square$

<sup>1</sup><https://github.com/henrykenlay/spgf>.

## 7. REFERENCES

- [1] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*, MIT Press, 2006.
- [2] M. McPherson, L. Smith-Lovin, and J. M. Cook, “Birds of a feather: Homophily in social networks,” *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, Aug. 2001.
- [3] 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,” *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, May 2013.
- [4] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A comprehensive survey on graph neural networks,” *arXiv:1901.00596*, 2019.
- [5] R. Levie, E. Isufi, and G. Kutyniok, “On the transferability of spectral graph filters,” *arXiv:1901.10524*, 2019.
- [6] F. Gama, J. Bruna, and A. Ribeiro, “Stability properties of graph neural networks,” *arXiv:1905.04497*, 2019.
- [7] 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. 16, pp. 4406–4421, Aug. 2017.
- [8] S. Segarra and A. Ribeiro, “Stability and continuity of centrality measures in weighted graphs,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 3387–3391.
- [9] E. Ceci and S. Barbarossa, “Small perturbation analysis of network topologies,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 4194–4198.
- [10] E. Ceci and S. Barbarossa, “Robust graph signal processing in the presence of uncertainties on graph topology,” in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2018, pp. 1–5.
- [11] D. Zügner, Akbarnejad A, and S. Günnemann, “Adversarial attacks on neural networks for graph data,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, July 2018, KDD ’18, pp. 2847–2856.
- [12] F. Chung, *Spectral graph theory*, American Mathematical Society, 1997.
- [13] N. J. Higham, *Functions of Matrices: Theory and Computation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.
- [14] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,” in *Advances in Neural Information Processing Systems* 29, 2016, pp. 3844–3852.
- [15] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *International Conference on Learning Representations (ICLR)*, 2017.
- [16] E. Deadman and S. D. Relton, “Taylor’s theorem for matrix functions with applications to condition number estimation,” *Linear Algebra and its Applications*, vol. 504, pp. 354–371, Sept. 2016.
- [17] N. J. Higham and S. D. Relton, “Higher order fréchet derivatives of matrix functions and the level-2 condition number,” *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 3, pp. 1019–1037, July 2014.
- [18] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, 2005.
- [19] A. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.