Appl. Comput. Harmon. Anal. 63 (2023) 48–83

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

journal homepage: www.elsevier.com/locate/acha

Transferability of graph neural networks: An extended graphon approach

Sohir Maskey [a,*], Ron Levie [b], Gitta Kutyniok [a,c,d]

[a] Department of Mathematics, LMU Munich, 80333 Munich, Germany
[b] Faculty of Mathematics, Technion - Israel Institute of Technology, Israel
[c] Department of Physics and Technology, University of Tromsø, 9019 Tromsø, Norway
[d] Munich Center for Machine Learning (MCML), Germany

ARTICLE INFO

Article history:
Received 18 October 2021
Received in revised form 25 June 2022
Accepted 24 November 2022
Available online 28 November 2022
Communicated by Radu Balan

MSC:
68T07
68R10
47A60

Keywords:
Graph neural network
Spectral methods
Graphon
Stability
Generalization
Transferability

ABSTRACT

We study spectral graph convolutional neural networks (GCNNs), where filters are defined as continuous functions of the graph shift operator (GSO) through functional calculus. A spectral GCNN is not tailored to one specific graph and can be transferred between different graphs. It is hence important to study the *GCNN transferability*: the capacity of the network to have approximately the same repercussion on different graphs that represent the same phenomenon. Transferability ensures that GCNNs trained on certain graphs generalize if the graphs in the test set represent the same phenomena as the graphs in the training set.

In this paper, we consider a model of transferability based on graphon analysis. Graphons are limit objects of graphs, and, in the graph paradigm, two graphs represent the same phenomenon if both approximate the same graphon. Our main contributions can be summarized as follows: 1) we prove that any fixed GCNN with continuous filters is transferable under graphs that approximate the same graphon, 2) we prove transferability for graphs that approximate unbounded graphon shift operators, which are defined in this paper, and 3) we obtain non-asymptotic approximation results, proving linear stability of GCNNs. This extends current state-of-the-art results which show asymptotic transferability for polynomial filters under graphs that approximate bounded graphons.

1. Introduction

CNNs [22] have shown great performances in a variety of tasks on data defined on Euclidean domains. In recent years, there has been a growing interest in machine learning on non-Euclidean data, and especially, data represented by graphs, leading to graph convolutional neural networks (GCNNs). Typically, GCNNs fall under two categories: spatial and spectral methods. Spatial-based approaches define graph convolutions by

* Corresponding author.
E-mail addresses: maskey@math.lmu.de (S. Maskey), levieron@technion.ac.il (R. Levie), kutyniok@math.lmu.de (G. Kutyniok).

https://doi.org/10.1016/j.acha.2022.11.008
1063-5203/

information propagation in the spatial domain. Spectral-based methods define convolutions in the frequency domain, taking inspiration from the classical Euclidean convolution theorem as follows. First, a frequency domain is defined for graphs by generalizing the Fourier transform on them. Then, graph convolutions are defined as element-wise multiplication in the graph frequency domain. In this paper, we study the *functional calculus approach* [9,20,25,3] for filtering, which bridges the gap between the spatial and spectral methods. Even though the functional calculus GCNNs are defined as spectral methods, they can be implemented directly in the spatial domain by applying filters on the graph shift operator (GSO). For a survey on spectral and spatial graph neural networks, we refer to [8] and [35].

In the functional calculus approach, the filters of the GCNN are defined via scalar functions $h : \mathbb{R} \to \mathbb{R}$ applied on the GSO $\mathbf{\Delta}$ via functional calculus, namely, as $h(\mathbf{\Delta})$ (see Subsection 2.2 for more details). The functions h are parameterized by some trainable parameters. For example, when h are polynomials or rational functions, the trainable parameters are the coefficients of h [9,25]. Any GCNN is defined by its underlying filters h, which are not linked to a specific graph. This means that an individual GCNN can be used to process any signal on any graph by plugging different GSOs $\mathbf{\Delta}$, corresponding to different graphs, in the filters h of the GCNN.

In *multi-graph settings*, the datasets consist of a set of signals defined on different graphs. Since the test set may consist of different graphs than the training set, the GCNN should generalize to graphs unseen during training. To guarantee a good generalization capability, it is desirable to show that the GCNN is *transferable* between graphs. In [23], the notion of GCNN transferability was described as follows: a GCNN is transferable, if, whenever two graphs represent the same phenomenon, the GCNN has approximately the same repercussion on both graphs. Transferability promotes generalization, since we may encounter a graph that realizes some phenomenon in the training set, and then encounter a different graph representing the same phenomenon in the test set. Having learned how to cope with the first realization of the phenomenon, the network handles its other versions well. For example, in a dataset of mesh graphs, surfaces are the underlying phenomena. If a filter is learned on one mesh-graph, discretizing a surface, it is desirable to show that the same filter has approximately the same effect on a different mesh-graph, discretizing the same surface.

Transferability is analyzed formally using a modeling paradigm that was introduced in [23]. First, we require a mathematical model of "two graphs representing the same phenomenon." Second, a notion that a fixed GCNN "has approximately the same repercussion on two graphs" is needed. Equipped with two such mathematical definitions, one can formulate rigorously and prove transferability. Some examples of transferability frameworks are the stability approach [24,17,18,16,10], the sampling approach [23,19], and the graphon approach [30]. Each framework differs in the way it formalizes the transferability paradigm. In Subsection 1.2 we survey transferability frameworks from the literature. In this work, we extend and generalize the graphon transferability approach.

1.1. Comparison with other works

The approach we consider in this paper for modeling graphs that represent the same phenomenon is called the *graphon approach*. *Graphons* [7] are limit objects of sequences of weighted graphs. In some sense, a graphon can be thought of as an adjacency matrix, or GSO, of a graph with a continuous node index set. The continuous node index set is typically taken to be $[0, 1]$, and a graphon is formally a function $W : [0, 1]^2 \to [0, 1]$. The value $W(x, y)$ is interpreted as the probability of having an edge between the nodes x and y, or as the weight of the edge (x, y). To model graphs that represent the same phenomenon, the graphon approach relies on the notion of homomorphism density. Given a simple graph F, informally, the *homomorphism density* $t(F, G)$ is the probability that F occurs as a substructure of the graph G (see Definition 3.1 for a precise formulation). Similarly, the homomorphism density $t(F, W)$ is also defined for graphons (see Definition 3.3 and [7]). In the graphon approach, two graphs G_1 and G_2 represent the same

phenomenon if for every simple graph F we have $t(F, G_1) \approx t(F, G_2)$. The sequence of graphs $(G_n)_n$ converges to the graphon W if for every simple graph F, $t(F, G_n) \to t(F, W)$.

In this subsection, we compare our transferability results with other related transferability results [19,28, 31,30,32], which also use the graphon approach. The analysis of transferability of GCNNs in the graphon approach in the papers [19,28,31,30,32] consists of two steps: 1) given a sequence of graphs G_n and a graphon W, first prove that G_n converges to W, and 2) given a filter h and a sequence of graphs G_n converging to a graphon W, prove that the graph filter $h(G_n)$ converges to the graphon filter $h(W)$ (see Definition 2.2 for more detail). The first step is treated already in other works, e.g., [6,7,21] show that sampling random graphs from a graphon with increasing number of nodes leads to convergence, and [32] shows that graphs that are sampled in a regular grid from the graphon converge to the graphon. Hence, in this paper we only focus on the second step of the analysis.

We have already seen that analyzing transferability in the graphon approach consists of performing an approximation analysis. For this, one can consider various types of assumptions on the object of study. We call *topological assumptions* any construction that depends on the topology or metric of the underlying space. For example, assuming that a function $f : [0, 1]^2 \to \mathbb{R}$ is Lipschitz means that we treat the space $[0, 1]$ as a metric space. A *measure theoretic assumption* treats the underlying space as a measure space. For example, assuming that a function $f : [0, 1]^2 \to \mathbb{R}$ is only measurable or square-integrable implies that we treat the space $[0, 1]$ only as a measure space, not as a topological. Since there is a natural measure space associated with topological spaces, namely its Borel space, assuming only measure theoretic assumption gives more general results than topological assumptions.

To analyze the transferability of GCNN filters in the graphon approach it is important to not assume any topological constraints on the graphon. By this, the graphon approach relies purely on measure-theoretical objects defined on the atom-free standard probability space $[0, 1]$ equipped with the Lebesgue measure. A well-known fact is that every atom-free standard space is isomorphic up to a measure-preserving isomorphism. Standard probability spaces cover a large class of possible spaces, e.g., $\mathbb{R}^n, \mathbb{C}^n$, every Lie group with countable components, every Polish space and standard Borel space with atom-free probability measure. In particular, since there is no notion of dimension in measure theory, the space $[0, 1]$ is equivalent as a probability space to a rich class of probability spaces over topological spaces of any dimension. Hence, if no topological constraints are assumed for the graphon and the graphon signal, the transferability results hold for all atom-free standard probability space.

The related works [28,31,30] consider only measure-theoretical assumptions on the graphon but show transferability without giving rates of converges. The papers [19,32] show transferability with rates of converges, but impose strong topological constraints on the graphon. Our work is the first to derive transferability bounds with rate of convergence, while treating graphon purely in a measure theoretic setting, which makes our results very general.

In the following, we recall in more detail the related works [28,31,30,19,33,32] and compare them to our results. The authors in [28,31,30] consider simple graphs that are represented by their adjacency matrix. It is shown that polynomial filters are asymptotically transferable for a sequence of graphs that converges to a graphon with weights in $[0, 1]$. The assumptions on the considered graphons are purely measure-theoretical, and thus generalize to any atom-free standard probability space. However, the approach in [28,31,30] does not give any rates of convergence for the transferability results. They are moreover restricted only to polynomial filters, while filters in practice are often non-polynomial [25,3]. We generalize these results by allowing graphs with arbitrary GSOs, having possibly negative weights, which includes graph Laplacians. We also allow general unbounded graphons, which we define in this paper in Definition 4.8, and any continuous filter (not just polynomials). Furthermore, we obtain convergence rates while only considering a pure measure theoretic construction (no topological constraints).

To derive a rate of convergence, [19,33,32] consider a framework in which a sequence of graphs is sampled from a template graphon. The authors show that highly regular filters (polynomials or analytic functions)

applied to such a sequence of graphs converge to the filter applied to template graphon. This result holds only under strong topological constraints on the limit graphon and the graphon signal (e.g., Lipschitz continuity). Thus, the transferability results (that include rates of convergence) in [19,33,32] do not generalize to all atom-free standard probability spaces and to general measurable graphon (that can for example be discontinuous everywhere).

In contrast, we prove transferability with rates of convergence while assuming no topological constraints on the graphon and on the graphon signal. In particular, our theory then applies to underlying spaces of any dimension and irregular graphons that need only be measurable and can even be discontinuous everywhere. While other papers use basic linear algebra to derive their results, we generalize and improve the theory by using tools from graphon theory and functional analysis, in particular, functional calculus and operator Lipschitz functions [2].

1.2. Survey of GCNN transferability approaches

In this subsection, we survey different approaches for defining "two graphs represent the same phenomenon" and a fixed filter having "approximately the same repercussion" on two graphs.

Stability under small graph perturbations

This framework allows studying the transferability of graphs sharing the same node set. Here, two graphs represent the same phenomenon if their GSOs are small perturbations of each other. The notion of "small perturbation" is modeled differently in different works. Moreover, the change in repercussion of the filter due to the perturbation of the GSO is also defined in various ways. The first work that considered this framework was by Levie et al. in [24]. There, the perturbation of the GSO is defined in the induced l_2-norm sense, between the GSO $\mathbf{\Delta}_1$ and its perturbed version $\mathbf{\Delta}_2$. The repercussion error of the filter was also defined by the induced l_2-norm. Given a function h from the so-called Cayley smoothness class, and assuming that $\|\mathbf{\Delta}_1 - \mathbf{\Delta}_2\|_2 \leq 1$, the following stability result was proven

$$\|h(\mathbf{\Delta}_1) - h(\mathbf{\Delta}_2)\|_2 \leq C\|\mathbf{\Delta}_1 - \mathbf{\Delta}_2\|_2, \tag{1}$$

where the constant $C > 0$ depends on h and $\|\mathbf{\Delta}_1\|_2$.

Similar results were also proven in [18], where graphs are assumed to be simple. The authors show that polynomial and low-pass filters are stable to arbitrary perturbations, in the sense that $\|\mathbf{\Delta}_1 - \mathbf{\Delta}_2\|_2 \leq 1$ does not need to be assumed in (1). Furthermore, if the structural change of the graph consists of adding and deleting edges, the authors constrain the GSO error $\|\mathbf{\Delta}_1 - \mathbf{\Delta}_2\|_2$ by interpretable properties directly related to this structural change. In [10], Gama et al. consider weighted graphs with GSOs. The authors showed stability of Lipschitz continuous polynomial filters in a similar setting as (1). The difference with respect to the previous papers is that the measure of graph perturbation is the norm distance, minimized over all permutations of the node set of $\mathbf{\Delta}_2$.

Closely related to stability of GCNNs, [4,5] also studied the robustness of some GCNNs on simple graphs. The authors provided certificates that guarantee that GCNNs are stable in a semi-supervised graph or node classification task. The authors guarantee that the classification of nodes and graphs remains correct while structurally perturbing the input graph or perturbing the nodes' input features. The structural perturbations are the deletion or addition of edges. The repercussion error of a fixed GCNN is defined either pointwise on a single node of interest, or the classification outcome for the whole graph. The definition of the GCNN in these papers was restricted to a subclass of spectral methods.

Transferability under sampling from latent spaces

A major drawback of analyzing transferability of GCNNs under GSO perturbations is the assumption that the graphs share the same node set. In contrast, in the sampling approach to transferability, graphs

are assumed to represent the same phenomenon if their nodes are sampled from the same underlying latent space, which is typically a metric, topological, or measure space $\mathcal{M}$. This allows comparing graphs of arbitrary sizes and topologies. In [23], Levie et al. introduce a framework where graph nodes and GSOs are sampled from a "continuous" underlying model. The model consists of a measure space $\mathcal{M}$ and a continuous Laplacian $\mathcal{L}$, which is a self-adjoint operator with discrete spectrum. The repercussion of a filter is treated in the "sampling–interpolation" approach. It is assumed that two latent sampling operators S_1 and S_2 exist for the two graphs G_1 and G_2, which describe how the graph data is generated from the latent space. For example, the nodes of the graph can be assumed to belong to a latent space, and sampling evaluates signals defined over the latent space on these sample nodes, to generate a graph signal. Conversely, two interpolation operators R_1 and R_2 take graph signals and return signals over the topological space. For a fixed filter h, the transferability error of the filter on graph signals sampled from a latent continuous signal $\psi \in L^2(\mathcal{M})$ is defined as $\|R_1 h(\boldsymbol{\Delta}_1) S_1 \psi - R_2 h(\boldsymbol{\Delta}_2) S_2 \psi\|_{L^2(\mathcal{M})}$. It is shown that the transferability error of the filter is bounded by the transferability of the Laplacian and the consistency error, i.e.,

$$\begin{aligned} \|R_1 h(\boldsymbol{\Delta}_1) S_1 \psi - R_2 h(\boldsymbol{\Delta}_2) S_2 \psi\|_{L^2(\mathcal{M})} \leq C_1 \sum_{j=1,2} \|\mathcal{L}\psi - R_j \boldsymbol{\Delta}_j S_j \psi\|_{L^2(\mathcal{M})} \\ + C_2 \sum_{j=1,2} \|\psi - R_j S_j \psi\|_{L^2(\mathcal{M})}. \end{aligned} \tag{2}$$

Based on this inequality, the authors develop a transferability inequality for end-to-end GCNNs. Moreover, it was shown that the right-hand-side of (2) is small when the graphs are sampled randomly from the latent space, hence proving transferability for randomly generated graphs.

A related approach was proposed in [19], where the authors consider random graph models. A random graph model is a tuple of a compact metric space equipped with a probability measure and a "continuous" shift operator, defined as a graphon. The probability measure is used to sample nodes, and the graphon is used to sample edges, to generate a graph from the continuous model. The stability of analytic filters and GCNNs to small deformations of the underlying random graph model was proved.

Transferability between graphs converging to the same graphon

Ruiz et al. first used the graphon framework in [30] to define a transferability analysis of graph filters. In that work, the repercussion of a polynomial filter h is measured in an asymptotic sense, using the notion of signal induction. For a signal $\mathbf{x} \in \mathbb{R}^n$, its induced graphon signal $\psi_{\mathbf{x}} \in L^2[0,1]$ is a piecewise constant representation of $\mathbf{x}$ over the continuous index set $[0,1]$ (see Definition 3.11). Furthermore, a graphon W induces the integral operator $T_W f(u) := \int_0^1 W(u,v) f(v) dv$. The result is given in [30, Theorem 4] and can be stated informally as follows.

Theorem 1.1 *(informal). Let h be a polynomial Lipschitz continuous filter and $(G_n)_n$ be a sequence of graphs with adjacency matrices $(\boldsymbol{A}_n)_n$. Suppose that G_n converges in the homomorphism density sense to the graphon W. Let $\mathbf{x}_n \in \mathbb{R}^n$ be a sequence of graph signals converging to the graphon signal $y \in L^2[0,1]$ in the sense that the induced signals $\psi_{\mathbf{x}_n}$ converge to y, i.e., $\|\psi_{\mathbf{x}_n} - y\|_{L^2[0,1]} \xrightarrow{n\to\infty} 0$. Denote $\mathbf{z}_n = h(\boldsymbol{A}_n/n)\mathbf{x}_n$ for $n \in \mathbb{N}$. Then, the output of the graph filters converges in the induction sense to the output of the graphon filter, namely,*

$$\|\psi_{\mathbf{z}_n} - h(T_W) y\|_{L^2[0,1]} \xrightarrow{n\to\infty} 0.$$

[28,19,32] consider transferability of GCNNs under sampling from a graphon W. The authors consider a sampling procedure leading to a graph sequence G_n that converges to W in the homomorphism density sense (see Definition 3.4). The main results then show, under topological restrictions on the graphon, the

graphon signal and the filter, that the output of the graph filters converge in the induction sense to the output of the graphon filter, similarly to Theorem 1.1.

As described in Subsection 1.1, related transferability results in the graphon approach are either purely asymptotic while generalizing to other standard probability spaces, or a rate of convergence is calculated, but the results do not generalize to other standard probability spaces since the graphon is assumed to be a Lipschitz continuous function.

1.3. Summary of our main contribution

We consider in this paper dense graphs, namely, graphs in which the number of edges is of order of the number of nodes squared. Graphons are the natural limit object of such graphs.

Definition 1.2. Given a graph G with n nodes and GSO $\mathbf{\Delta}$, we define the *graph weight matrix (GWM)* $\boldsymbol{A}$ as the matrix

$$\boldsymbol{A} = n\mathbf{\Delta}. \tag{3}$$

The relation (3) is interpreted as follows. GSOs $\mathbf{\Delta}$ define how graphs operate on graph signals. To accommodate the transferability between dense graphs of different sizes, GSOs are assumed to be roughly normalized in the induced l^2-norm. On the other hand, GWMs describe the probability of each edge in the graph to exist, in case the entries of the GWM are in $[0, 1]$. Accordingly, GWMs are roughly normalized in the entry-wise infinity norm. This motivates Definition 1.2.

The way to compare how graphs of different sizes operate on signals in the graphon approach is by *inducing* their GWMs to graphons. Given a GWM $\boldsymbol{A}$, its *induced graphon* $W_{\boldsymbol{A}}$ is a piecewise continuous representation of the matrix $\boldsymbol{A}$ over the continuous index set [0,1] [7]. The *graphon shift operator (GRSO)* corresponding to the graphon W is defined as the integral operator $T_W f(u) := \int_0^1 W(u, v)f(v)dv$. GRSOs are the limit objects of GSOs. Since filters $h(\mathbf{\Delta})$ are GSOs themselves, and since GSOs and GWMs are related by (3), the notation $W_{nh(\mathbf{\Delta})}$ means the graphon induced by the GWM corresponding to the GSO $h(\mathbf{\Delta})$. Hence, given two graphs of size n_1 and n_2, the way to compare $h(\mathbf{\Delta}_1)$ to $h(\mathbf{\Delta}_2)$ is by

$$\|T_{W_{n_1 h(\mathbf{\Delta}_1)}} - T_{W_{n_2 h(\mathbf{\Delta}_2)}}\|_{L^2[0,1]\to L^2[0,1]}.$$

In this work we generalize Theorem 1.1 in [30] by allowing graphs with arbitrary GSOs, general unbounded graphon shift operators with values in $\mathbb{R}$, and general continuous filters. Another important contribution is that we derive non-asymptotic transferability bounds without any topological assumptions. We moreover consider end-to-end graph convolutional networks. We summarize our contributions as follows.

Asymptotic transferability

A common approach in data science for studying and comparing graphs is counting motifs, which are substructures of the graphs, namely, simple graphs (see, e.g., [12, Section 2.1.1]). Since homomorphism densities can be interpreted as the density of motifs in a graph (see Subsection 3.1), we choose the notion of homomorphism densities to model graphs that represent the same phenomenon. On the other hand, when doing signal processing, graph filters act on graph signals as operators. Thus, when comparing the repercussion of a filter realized on different graphs, we consider the L^2-induced operator norm of the induced GRSOs and the L^2-norm of the induced graphon signals. Since our transferability framework is built on these two definitions, namely, homomorphism densities and L^2-induced operator norms, we note that they are closely related: convergence in homomorphism density is equivalent (up to relabeling of the graphs' nodes) to convergence in the L^2-induced operator norm (see Theorem 3.10).

We prove asymptotic transferability for continuous filters in the following sense, that we formulate here informally, and write rigorously in Theorem 5.2.

Theorem 1.3 *(informal). Let $(G_n)_n$ be a sequence of graphs with GSOs $(\mathbf{\Delta}_n)_n$. Suppose that G_n converges to a graphon $W : [0,1] \times [0,1] \to \mathbb{R}$ in the homomorphism density sense (see Definition 3.4). Let $h : \mathbb{R} \to \mathbb{R}$ be a continuous filter with $h(0) = 0$. Then the, by $h(\mathbf{\Delta}_n)$ induced, integral operators $(T_{W_{h(\mathbf{\Delta}_n)}})_n$ converge to the graphon filter $h(T_W)$ in operator norm, namely,*

$$\|T_{W_{nh(\mathbf{\Delta}_n)}} - h(T_W)\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0.$$

In fact, Theorem 1.3 with $h(x) = x$ shows that convergence of graphs in homomorphism density implies convergence of the induced GRSOs in operator norm. Theorem 1.3 can be interpreted as a tranferability result between graphs as follows. Let $\varepsilon > 0$ be a desired error tolerance for the repercussion error of the filter h. There exists a large enough N such that for every $m, n > N$, by the triangle inequality,

$$\|T_{W_{nh(\mathbf{\Delta}_n)}} - T_{W_{mh(\mathbf{\Delta}_m)}}\|_{L^2[0,1]\to L^2[0,1]} < \varepsilon.$$

Here, we interpret $\mathbf{\Delta}_n$ and $\mathbf{\Delta}_m$ as a pair of arbitrary GSOs that represent the same phenomenon in the homomorphism density sense.

Approximation rate

Theorem 1.3 shows asymptotic transferability in a very general setting, but with no convergence rate. In Theorem 1.4 we present non-asymptotic transferability bounds under some assumption on the filter. We give the following informal result, which we formulate rigorously in Proposition 5.6.

Theorem 1.4 *(informal). Let $h \in C^1$ be such that h' is Lipschitz and $h(0) = 0$. Suppose that G_1 and G_2 are graphs with n_1 and n_2 nodes respectively, represented by the GSOs $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$. Let $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ be the GWMs corresponding to $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$. Then*

$$\|T_{W_{n_1h(\mathbf{\Delta}_1)}} - T_{W_{n_2h(\mathbf{\Delta}_2)}}\|_{L^2[0,1]\to L^2[0,1]} \leq C\|T_{W_{\boldsymbol{A}_1}} - T_{W_{\boldsymbol{A}_2}}\|_{L^2[0,1]\to L^2[0,1]},$$

where $C \in \mathbb{N}$ is a constant that depends on the filter h.

We mention that in our setting the graphs do not need to be of the same size and can be represented by arbitrary symmetric matrices. Furthermore, we allow more general filters than polynomials.

Graphs approximating unbounded graphon shift operators

GRSOs cannot represent all important limit objects of graph shift operators, e.g., Laplace-Beltrami operators. To allow such examples, we define *unbounded graphon shift operators (unbounded GRSOs)* – self-adjoint unbounded operators that become GRSOs when restricted to their Paley-Wiener spaces. As opposed to the similar approach in [23], we allow an accumulation point at 0 in the spectrum of the unbounded GRSO, which allows treating a richer family of operators. The following theorem is an informal version of Theorem 5.11.

Theorem 1.5 *(informal). Let $h : \mathbb{R} \to \mathbb{R}$ be a function with additional assumptions given in Theorem 5.11. Let $(G_n)_n$ be a sequence of graphs with GSOs $(\mathbf{\Delta}_n)_n$, converging to an unbounded GRSO $\mathcal{L}$. Then, for every $\lambda > 0$*

$$\|P_{\mathcal{L}}(\lambda)T_{W_{nh(\mathbf{\Delta}_n)}}P_{\mathcal{L}}(\lambda) - h(\mathcal{L})\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0,$$

where $P_{\mathcal{L}}(\lambda)$ is the projection onto the space of band limited signals on the band $[-\lambda, \lambda]$, also called the Paley-Wiener space associated with $\mathcal{L}$ (see Definition 4.7).

1.4. Overview

In Section 2 we give an introduction to graph shift operators and graph filters. Section 3 starts by introducing convergence for sequences of graphs. Then, graphons and graphon shift operators, limit objects of convergent sequences of graphs and GSOs, are discussed. The notation and most results cited there are from [27]. We introduce our transferability framework and the framework to analyze unbounded GRSOs in Section 4. In Section 5, we present the main results about transferability of graph filters and GCNNs. In the Appendix, we provide proofs and theoretical backgrounds. The foundations of graphons are given in Appendix A.2.

2. Graph convolutional neural networks (GCNN)

In this section, we recall basic definitions of graphs with GSOs and the definition of functional calculus filters.

2.1. Graph shift operators

There are various ways to represent the structure of a graph G. In this work, we consider representing G by symmetric matrices $\mathbf{\Delta} \in \mathbb{R}^{n\times n}$, which are called *graph shift operators (GSOs)* or *graph Laplacians*. Accordingly, we use the following definition of a graph.

Definition 2.1. A graph G with GSO $\mathbf{\Delta}$ is a triplet $G = (V, E, \mathbf{\Delta})$ where $V = \{1, \dots, n\}$ is the node set, $E \subset V \times V$ is the edge set, and $\mathbf{\Delta} \in \mathbb{R}^{n\times n}$ is a symmetric matrix called the GSO.

In practice, graphs in datasets are given by an adjacency matrix $\mathbf{W}$, with entries in $[0, 1]$ if the graph is weighted, and in $\{0, 1\}$ if it is unweighted. In practical deep learning, one typically modifies the "raw data" adjacency matrix $\mathbf{W}$ and constructs a GSO $\mathbf{\Delta}$ with entries in $\mathbb{R}$. Some common choices for $\mathbf{\Delta}$ are the combinatorial and the symmetric graph Laplacians or the graphs' adjacency matrix itself. The filters are then defined with respect to $\mathbf{\Delta}$. In our analysis, we assume that the graph is already given with the GSO $\mathbf{\Delta}$, no matter how it was constructed from the "raw data" $\mathbf{W}$. The graph weight matrix GWM, not to be confused with the raw GSO, is defined directly from $\mathbf{\Delta}$ as $\boldsymbol{A} = n\mathbf{\Delta}$ (see Definition 1.2).

2.2. Graph filters

In this subsection, we recall the definition of *functional calculus filters*, as described in [23]. For a graph G with GSO $\mathbf{\Delta} \in \mathbb{R}^{n\times n}$, a *graph signal* $\mathbf{x} = (x_i)_{i=1}^n \in \mathbb{C}^n$ is a vector where x_i is called the feature of the i'th node. Given $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, the inner product between $\mathbf{x}$ and $\mathbf{y}$ is defined to be

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{C}^n} = \sum_{i=1}^{n} x_i \bar{y}_i.$$

Definition 2.2. Let G be a graph with a GSO $\mathbf{\Delta}$, and let $\{\lambda_j, \phi_j\}_{j=1}^n$ be the eigenvalues and eigenvectors of $\mathbf{\Delta}$.

(1) A *filter* is a continuous function $h : \mathbb{R} \to \mathbb{R}$.

(2) The *realization of the filter* $h(\mathbf{\Delta})$ is defined to be the operator

$$h(\mathbf{\Delta})\mathbf{x} = \sum_{j=1}^{n} h(\lambda_j)\langle \mathbf{x}, \phi_j \rangle \phi_j, \tag{4}$$

applied on any graph signal $\mathbf{x} \in \mathbb{C}^n$.

For brevity, we often call $h(\mathbf{\Delta})$ simply a filter. As an example, for a rational function

$$h(\lambda) = \frac{\sum_{n=0}^{N} h_n \lambda^n}{\sum_{m=0}^{M} k_m \lambda^m},$$

it is easy to see that

$$h(\mathbf{\Delta}) = \left(\sum_{n=0}^{N} h_n \mathbf{\Delta}^n \right) \cdot \left(\sum_{m=0}^{M} k_m \mathbf{\Delta}^m \right)^{-1}. \tag{5}$$

Thus, the polynomial filters in [20,9] and rational filters in [25,3] are examples of functional calculus filters (4).

2.3. Spectral graph convolutional neural networks

Similarly to CNNs, a GCNN is defined as a deep convolutional architecture with three main components per layer: a set of graph filters, a non-linear activation function and optionally pooling. In this work, we study architectures without pooling. Example applications are semi-supervised node classification and node regression.

GCNNs are defined by realizing a *spectral convolutional neural network (SCNN)* on a graph. SCNNs are defined independently of a particular graph topology and can be even applied on non-graph data, namely, graphons, as we explain in Subsection 3.4.

Definition 2.3. Let $L \in \mathbb{N}$ denote the number of layers. Let $F_l \in \mathbb{N}$ denote the number of features in the layer $l \in \{1, \dots, L\}$. For each $l \in \{1, \dots, L\}$, let $H_l := (h_l^{jk})_{j \le F_l, k \le F_{l-1}} \in (C(\mathbb{R}))^{F_l \times F_{l-1}}$ be an array of filter functions and $\mathbf{M}_l := (m_l^{jk})_{j \le F_l, k \le F_{l-1}} \in \mathbb{R}^{F_l \times F_{l-1}}$ a matrix. Let $\rho : \mathbb{R} \to \mathbb{R}$ be a continuous function that we call the *activation function*. Then, the corresponding *spectral convolutional neural network (SCNN)* ϕ is defined to be

$$\phi := (H, \mathbf{M}, \rho) := ((H_l)_{l=1}^{L}, (\mathbf{M}_l)_{l=1}^{L}, \rho). \tag{6}$$

In comparison to the single filter case in Section 2.2, a graph signal is not restricted to be an element in $\mathbb{C}^n$. We consider *feature maps* given by a tuple of graph signals, i.e., by a $\mathbf{x} := (\mathbf{x}^k)_{k=1}^{F_l} \in \mathbb{C}^{n \times F_l}$, where $F_l \in \mathbb{N}$ is the feature dimension.

Definition 2.4. Let G be a graph with GSO $\mathbf{\Delta}$ and ϕ be an SCNN, as defined in Definition 2.3. For each $l \in \{1, \dots, L\}$, we define $\phi_{\mathbf{\Delta}}^l$ as the mapping that maps input feature maps to the feature maps in the l-th layer, i.e.,

$$\phi_{\mathbf{\Delta}}^l : \mathbb{C}^{n \times F_0} \to \mathbb{C}^{n \times F_l}, \ \mathbf{x} \mapsto \mathbf{x}_l = (\mathbf{x}_l^j)_{j=1}^{F_l} \tag{7}$$

with

$$\mathbf{x}_l^j = \rho(\sum_{k=1}^{F_{l-1}} m_l^{jk} h_l^{jk}(\boldsymbol{\Delta})(\mathbf{x}_{l-1}^k)), \tag{8}$$

for $j \in \{1, \dots, F_l\}$ and $\mathbf{x}_0 = \mathbf{x}$. We write $\phi_{\boldsymbol{\Delta}} := \phi_{\boldsymbol{\Delta}}^L$ and call it the realization of ϕ on the graph G with GSO $\boldsymbol{\Delta}$. Furthermore, we define $\tilde{\phi}_{\boldsymbol{\Delta}}^l$ as the mapping that maps input features to the feature maps in the l-th layer before applying the activation function in the l-th layer, i.e.,

$$\tilde{\phi}_{\boldsymbol{\Delta}}^l : \mathbb{C}^{n \times F_0} \to \mathbb{C}^{n \times F_l},\ \mathbf{x} \mapsto \tilde{\mathbf{x}}_l = (\tilde{\mathbf{x}}_l^j)_{j=1}^{F_l}$$

with

$$\tilde{\mathbf{x}}_l^j = \sum_{k=1}^{F_{l-1}} m_l^{jk} h_l^{jk}(\boldsymbol{\Delta})(\phi_{\boldsymbol{\Delta}}^{l-1}(\mathbf{x})^k),$$

for $j \in \{1, \dots, F_l\}$.

For brevity, we typically do not distinguish between an SCNN and its realization on a graph.

3. Graphon analysis

In this section, we recall the notion of a *graphon*, which we describe informally as limit objects of sequences of graphs in Subsection 1.2. The definition of convergence in the graphon setting is based on homomorphism numbers, which we recall in Subsection 3.1. We give the definition of graphs converging to graphons in Subsection 3.2. In Subsection 3.3 we discuss shift operators acting on the space of graphons.

3.1. Homomorphism numbers

Lovász and Szegedy introduce in [26] the following definition.

Definition 3.1. Let $F = (V(F), E(F))$ be a simple graph and G a graph with GSO $\boldsymbol{\Delta}$ and node set $V(G)$. Let $\boldsymbol{A} = n\boldsymbol{\Delta}$ be the GWM of G.

(1) Given a map $\psi : V(F) \to V(G)$, we define the homomorphism number

$$hom_\psi(F, G) = \prod_{(u,v) \in E(F)} \boldsymbol{A}_{(\psi(u), \psi(v))}.$$

(2) We define the homomorphism number

$$hom(F, G) = \sum_{\psi : V(F) \to V(G)} hom_\psi(F, G), \tag{9}$$

where the sum goes over all maps $\psi : V(F) \to V(G)$.

(3) The *homomorphism density* with respect to the graph G is a function $t(\cdot, G)$ that assigns to each simple graph F the following value

$$t(F, G) := \frac{|hom(F, G)|}{|V(G)|^{|V(F)|}}. \tag{10}$$

Let us give an interpretation for the homomorphism numbers. Suppose that G is a simple graph with adjacency matrix $\boldsymbol{W} \in \{0,1\}^{n\times n}$ and GWM $\boldsymbol{A} = \boldsymbol{W}$. We recall that a homomorphism between two graphs is an adjacency-preserving mapping between their nodes. Then, $hom_\psi(F,G)$ is one if ψ is a homomorphism and zero otherwise. $hom(F,G)$ counts the number of homomorphisms of F into G. Since $|V(G)|^{|V(F)|}$ is the total number of maps between the nodes of F and G, the homomorphism density $t(F,G)$ gives the probability that a random map from F to G is a homomorphism. We interpret F as a simple substructure, called a *motif*, so that $hom(F,G)$ counts how many times the motif F appears in G. If G is a weighted graph with weights in $[0,1]$, interpreted as the probability of edges to exist, then $hom(F,G)$ is the probability of the motif F to appear in G. For further interpretations, see [27, Chapter 5.3].

3.2. Graphons

The seminal work [26] studied sequences of graphs G_n such that for every motif F the parameter $t(F,G_n)$ converges. Graphons were shown to be the natural limit object of such sequences.

Definition 3.2. A *graphon* is a bounded, symmetric and measurable function $W : [0,1]\times[0,1] \to \mathbb{R}$.[1] We denote the space of all graphons by $\mathcal{W}$. Let $\Gamma > 0$. The subspace $\mathcal{W}_\Gamma \subset \mathcal{W}$ is defined as the set of all graphons $W \in \mathcal{W}$ such that $|W(u,v)| \leq \Gamma$ for every $(u,v) \in [0,1]\times[0,1]$.

Graphons are interpreted as the spaces from which graphs are sampled, and to which graphs are embedded in order to compare them. Since in applications we only observe the graphs, we call the graphon corresponding to a graph its *latent space*. We do not distinguish between functions that are almost everywhere equal. The way graphons are seen as limit objects of graphs is via the homomorphism density of graphons.

Definition 3.3. For a graphon W and a simple graph $F = (V,E)$, the homomorphism density of F in W is defined as

$$t(F,W) := \int_{[0,1]^n} \prod_{(i,j)\in E} W(u_i,u_j) \prod_{i\in V} du_i, \tag{11}$$

where n is the number of nodes in F.

Since the integral can be interpreted as a continuous analogue to the sum, (11) is a continuous analogue to the homomorphism density for graphs with GSO, and interpreted as the density of the motif F in W. We next recall a notion of convergence of sequences of graphs [26, Section 2.2].

Definition 3.4. Let $(G_n)_n$ be a sequence of graphs with GSOs and GWMs $(\boldsymbol{A}_n)_n$.

(1) We say that $(G_n)_n$ is a bounded sequence if there exists a constant $C > 0$ such that $\|\boldsymbol{A}_n\| \leq C$ for every $n \in \mathbb{N}$.
(2) We say that $(G_n)_n$ is *convergent in homomorphism density*, if for all simple graphs F the sequence $(t(F,G_n))_n$ is convergent.

The following theorem, taken from [6, Theorem 3.8 and Corollary 3.9], shows that graphons are natural limit objects of convergent sequences of graphs.

[1] Graphons can be defined over any atom-free standard probability space (see Subsection 1.1).

Theorem 3.5 *([6, Theorem 3.8 and Corollary 3.9]). For every bounded sequence of graphs $(G_n)_n$ with GSO that converges in homomorphism density, there exists a graphon $W \in \mathcal{W}$ such that for every simple graph F*

$$t(F, G_n) \xrightarrow{n\to\infty} t(F, W).$$

Remark 3.6. If $(G_n)_n$ is convergent according to Definition 3.4, then given the graphon W from Theorem 3.5, we say that $(G_n)_n$ converges to W in homomorphism density and write $G_n \xrightarrow[n\to\infty]{H} W$. Conversely, given a graphon $W \in \mathcal{W}$, there exists a sequence of graphs $(G_n)_n$ such that $G_n \xrightarrow[n\to\infty]{H} W$. The sequence $(G_n)_n$ can be constructed by first sampling i.i.d. uniform random points and then using the values of W to define the GWM (see [6, Theorem 4.5]).

Remark 3.7. In the following, we only consider a sequence of graphs $(G_n)_n$ with $v(G_n) \xrightarrow{n\to\infty} \infty$, where $v(G_n)$ denotes the number of nodes of G_n. For simplicity of notation, we assume that $v(G_n) = n$.

In the following, we describe a method for defining a graphon corresponding to the graph G using the notion of induced graphons. For this, we define for every $n \in \mathbb{N}$ the standard partition $\mathcal{P}_n = \{P_1, \dots, P_n\}$ of $[0, 1]$ by the sub-intervals $P_k = [k - 1/n, k/n)$ for $k = 1, \dots, n$.

Let χ_A denote the characteristic function of a subset $A \subset [0, 1]$, i.e.,

$$\chi_A(u) = \begin{cases} 1, & \text{for } u \in A \\ 0 & \text{else.} \end{cases}$$

We use characteristic functions to define the induction of a graph to a graphon.

Definition 3.8. Let G be a graph with GSO $\boldsymbol{\Delta}$ and GWM $\boldsymbol{A}$ and n nodes. The graphon $W_{\boldsymbol{A}}$ *induced* by G is defined as[2]

$$W_{\boldsymbol{A}}(u, v) := \sum_{i,j \le n} \boldsymbol{A}_{(i,j)} \chi_{P_i}(u) \chi_{P_j}(v). \tag{12}$$

With Definition 3.8, we clearly have

$$t(F, G) = t(F, W_{\boldsymbol{A}}) \tag{13}$$

for every graph G with GWM $\boldsymbol{A}$ and every simple graph F. We recall that GSOs $\boldsymbol{\Delta}$ and filters $h(\boldsymbol{\Delta})$, which are GSOs themselves, are related to GWMs by (3). Hence, we use the notation $W_{nh(\boldsymbol{\Delta})}$ to describe the graphon induced by the GWM corresponding to the GSO $h(\boldsymbol{\Delta})$.

Next we discuss a setting for performing signal processing with graphons, using L^2-norms. The space of *graphon signals* is defined as $L^2[0, 1]$. In the context of signal processing, we define convergence of graphons by interpreting them as kernel operators. For this, we recall that in Subsection 1.3 we defined the graphon shift operator (GRSO) corresponding to the graphon $W \in \mathcal{W}$ as the operator

$$T_W \psi(v) := \int_0^1 W(v, u) \psi(u) d\mu(u) \tag{14}$$

[2] We note that we could use any partition $\tilde{\mathcal{P}} = \{\tilde{P}_1, \dots, \tilde{P}_n\}$ such that $\mu(\tilde{P}_i) = 1/n$ for every $i = 1, \dots, n$ and $[0, 1]$ is the disjoint union of all element in $\tilde{\mathcal{P}}$.

for signals $\psi \in L^2[0,1]$. The operator norm of T_W is closely related to convergence in homomorphism density. To understand this relation, we define the *relabeling* of a graph by a permutation π.

Definition 3.9. Let G be a graph with GSO $\mathbf{\Delta}$ and edge set E. We define Π_G as the set of all permutations of the node set of G. Let $\pi \in \Pi_G$. The graph $\pi(G)$ is defined as the graph $\pi(G) = (\pi(V), \pi(E), \pi(\mathbf{\Delta}))$ with GSO where $(i,j) \in \pi(E)$ if there exists an edge $(k,l) \in E$ such that $(\pi(k), \pi(l)) = (i,j)$. The GSO is defined by $\pi(\mathbf{\Delta}) = (\delta_{\pi(i),\pi(j)})_{i,j}$, where $\mathbf{\Delta} = (\delta_{i,j})_{i,j}$. We call $\pi(G)$ the *relabeled graph.*

The following result relates convergence in homomorphism density to the operator norm.

Theorem 3.10. *Let $(G_n)_n$ a sequence of graphs with GSOs $(\mathbf{\Delta}_n)_n$ and $(\boldsymbol{A}_n)_n$ be the corresponding sequence of GWMs. Let $W \in \mathcal{W}$ satisfy $G_n \xrightarrow[n\to\infty]{H} W$. Let $(W_{\boldsymbol{A}_n})_n$ be the sequence of graphons induced by $(G_n)_n$. Then, there exists a sequence of permutations $(\pi_n)_n$ such that $\|T_{W_{\pi_n(\boldsymbol{A}_n)}} - T_W\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0$.*

Theorem 3.10 follows from a well-known result in graphon theory (see Appendix A.2). In fact, the converse direction in Theorem 3.10 is also true, showing the equivalence between convergence in homomorphism density and convergence of the induced GRSOs in the L^2-induced operator norm up to relabeling of the nodes.

We now follow a similar route as Ruiz et al. in [30] to define convergence of graph signals to graphon signals. Given a sequence of graph signals $(\mathbf{x}_n)_n$ on graphs $(G_n)_n$ with a growing number of nodes, we can interpret graphon signals as the limit of the graph signal sequence. We recall the definition of induced graphon signals from graph signals [30].

Definition 3.11. Let $\mathbf{x} \in \mathbb{C}^n$ be a signal on a graph G. The graphon signal $\psi_{\mathbf{x}} \in L^2(\mathcal{J})$ *induced* by $\mathbf{x}$ is defined as

$$\psi_{\mathbf{x}}(u) := \sum_{j \le n} x_j \chi_{P_j}(u). \tag{15}$$

We then define the following set of admissible permutations.

Definition 3.12. Let $(G_n)_n$ be a sequence of graphs with GSOs and corresponding GWMs $(\boldsymbol{A}_n)_n$. Let $W \in \mathcal{W}$ satisfy $G_n \xrightarrow[n\to\infty]{H} W$. The set of *admissible permutations* is defined as

$$\mathbf{P} = \left\{ (\pi_n)_n \in (\Pi_{G_n})_n \mid \|T_{W_{\pi_n(\boldsymbol{A}_n)}} - T_W\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0 \right\}.$$

Theorem 3.10 ensures that the set of admissible permutations is not empty. We use the set of admissible permutations to define convergence of graph signals to graphon signals.

Definition 3.13. Let $(G_n)_n$ be a sequence of graphs with GSOs and $(\mathbf{x}_n)_n$ a sequence of graph signals. We say that $(\mathbf{x}_n)_n$ converges to a graphon signal $\psi \in L^2[0,1]$ if there exists a graphon W such that $G_n \xrightarrow[n\to\infty]{H} W$ and for some admissible permutation $(\pi_n)_n \in \mathbf{P}$

$$\|\psi_{\pi_n(\mathbf{x}_n)} - \psi\|_{L^2[0,1]} \xrightarrow{n\to\infty} 0. \tag{16}$$

We write $\mathbf{x}_n \xrightarrow[n\to\infty]{I} \psi$.

3.3. Graphon shift operators and graphon filters

We define realizations of filters acting on graphon signals using GRSOs. Since W is symmetric, T_W is a self-adjoint Hilbert-Schmidt operator, and hence has a discrete eigendecomposition.

Definition 3.14. Let $h : \mathbb{R} \to \mathbb{R}$ be a filter. Let $W \in \mathcal{W}$, and let T_W be the associated GRSO. Let $\{\lambda_j, \phi_j\}_j$ be the eigendecomposition (with possibly repeating eigenvalues and orthonormal eigenvectors) of T_W. The *realization of the filter* $h(T_W)$ is defined as

$$h(T_W)\psi = \sum_{j=1}^{\infty} h(\lambda_j)\langle \psi, \phi_j \rangle \phi_j$$

for $\psi \in L^2[0,1]$.

Similarly to graph filters, we often call $h(T_W)$ simply a filter.

Example 3.15. If $h(\lambda) = \sum_{j=0}^{n} p_j \lambda^j / \sum_{i=0}^{m} q_i \lambda^i$ is a rational function, the filter $h(T_W)$ is given by

$$h(T_W) = \left(\sum_{j=0}^{n} p_j T_W^j \right) \left(\sum_{i=0}^{m} q_i T_W^i \right)^{-1}.$$

3.4. Graphon convolutional neural networks (GCNNon)

In this subsection, we define graphon convolutional neural networks (GCNNons). The GCNNon architecture can be defined equivalently to the one of GCNNs, given in Subsection 2.3. GCNNons are defined equivalently to GCNNs, given in Subsection 2.3, by applying an SCNN on a GRSO and a graphon signal. The construction is akin to WNNs in [32] and continuous GCNs in [19]. Nevertheless, we have fewer constraints on the filters and the graphon.

Let ϕ be SCNN. Let $W \in \mathcal{W}$ be a graphon and T_W the GRSO from (14). Similarly to GCNNs, we define the mapping ϕ_W^l by only replacing the discrete GSO by the GRSO T_W and propagating square-integrable functions instead of discrete graph signals in Definition 2.4. This means that the GCNNon is based on vector valued signals $\psi := (\psi^g)_{g=1}^{F_l} \in (L^2[0,1])^{F_l}$ that we call *graphon features maps*, as inputs and outputs.

4. The transferability framework

Induced signals and induced graphons provide a framework to compare the repercussion of filters acting on graphs that approximate the same graphon. Given two graphs G_1, G_2 with GSOs $\mathbf{\Delta}_1 \in \mathbb{R}^{n_1 \times n_1}, \mathbf{\Delta}_2 \in \mathbb{R}^{n_2 \times n_2}$, we compare the repercussions of the filters $h(\mathbf{\Delta}_1)$ and $h(\mathbf{\Delta}_2)$ as follows. We map the filters into the latent space by defining the induced graphons $W_{n_1 h(\mathbf{\Delta}_1)}$ and $W_{n_2 h(\mathbf{\Delta}_2)}$. Then, the repercussion error of the filters $h(\mathbf{\Delta}_1)$ and $h(\mathbf{\Delta}_2)$ in the latent space is defined by

$$\|T_{W_{n_1 h(\mathbf{\Delta}_1)}} - T_{W_{n_2 h(\mathbf{\Delta}_2)}}\|_{L^2[0,1] \to L^2[0,1]}.$$

For graph signals $\mathbf{x}_1 \in \mathbb{C}^{n_1}$ and $\mathbf{x}_2 \in \mathbb{C}^{n_2}$ with $\|\psi_{\mathbf{x}_1} - \psi_{\mathbf{x}_2}\|_{L^2[0,1]}$ sufficiently small, the repercussion error of the filter outputs of $h(\mathbf{\Delta}_1)$ and $h(\mathbf{\Delta}_2)$ is defined by

$$\|\psi_{h(\mathbf{\Delta}_1)\mathbf{x}_1} - \psi_{h(\mathbf{\Delta}_2)\mathbf{x}_2}\|_{L^2[0,1]}.$$

In this section, we develop tools for analyzing the above repercussion errors and extend the analysis to accommodate unbounded operators on graphon signals.

4.1. The spaces of induced graphons and induced signals

We study in this subsection the relation between the space of graph signals $\mathbb{C}^n$ and its image under the induction operation of Definition 3.11. The image space of the induction operator is a space of step functions.

Definition 4.1. For $n \in \mathbb{N}$, we define the space of step functions $[0,1]/\mathcal{P}_n \subset L^2[0,1]$ as $[0,1]/\mathcal{P}_n = \text{span}\{\chi_{P_1}, \dots, \chi_{P_n}\}$.

The following lemma can be easily verified.

Lemma 4.2. *For $n \in \mathbb{N}$, the correspondence $\mathbf{x} \mapsto \psi_{\mathbf{x}}$, defined in Definition 3.11, between the space $\mathbb{C}^n$ and $[0,1]/\mathcal{P}_n$ is an isomorphism. Further, it holds*

$$\frac{1}{\sqrt{n}}\|\mathbf{x}\|_{\mathbb{C}^n} = \|\psi_{\mathbf{x}}\|_{L^2[0,1]} \text{ and } \frac{1}{n}\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbb{C}^n} = \langle \psi_{\mathbf{x}}, \psi_{\mathbf{y}}\rangle_{L^2[0,1]}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.

Lemma 4.3. *Let G be a graph with GSO $\boldsymbol{\Delta}$ and associated GWM $\boldsymbol{A}$. Denote the eigendecomposition of $\boldsymbol{\Delta}$ by $\{\lambda^i, \mathbf{x}^i\}_{i=1}^n$. Then, $T_{W_{\boldsymbol{A}}}$ admits the eigendecomposition $\{\lambda^i, \sqrt{n}\psi_{\mathbf{x}^i}\}_{i=1}^n \cup \{0, \psi_j\}_{j=n+1}^{\infty}$, where the eigenvectors are an orthonormal basis for $L^2[0,1]$.*

Proof. Let $\{\lambda^i, \mathbf{x}^i\}_{i=1}^n$ denote the eigendecomposition of the GSO $\boldsymbol{\Delta}$ with $|\lambda_1| \leq |\lambda_2| \leq \dots \leq |\lambda_n|$. Let $N \in \{1, \dots, n\}$. Then, for almost every $u \in [0,1]$,

$$\begin{aligned}
T_{W_{\boldsymbol{A}}}(\sqrt{n}\psi_{\mathbf{x}^N})(u) &= \int_0^1 (\sqrt{n} \sum_{k,l \leq n} \chi_{P_k}(u)\chi_{P_l}(v)\boldsymbol{A}_{(k,l)})(\sum_{j \leq n} \chi_{P_j}(v)x_j^N)dv \\
&= \sum_{k \leq n} \chi_{P_k}(u) \int_0^1 (\sqrt{n} \sum_{l,j \leq n} \chi_{P_j}(v)\chi_{P_l}(v)\boldsymbol{A}_{(k,l)}x_j^N)dv \\
&= \sum_{k \leq n} \chi_{P_k}(u) \int_0^1 (\sqrt{n} \sum_{j \leq n} \chi_{P_j}(v)\boldsymbol{A}_{(k,j)}x_j^N)dv \\
&= \sum_{k \leq n} \chi_{P_k}(u) \frac{1}{n}\sqrt{n} \sum_{j \leq n} \boldsymbol{A}_{(k,j)}x_j^N \\
&= \sum_{k \leq n} \chi_{P_k}(u) \sqrt{n} \sum_{j \leq n} \boldsymbol{\Delta}_{(k,j)}x_j^N \\
&= \sqrt{n} \sum_{k \leq n} \chi_{P_k}(u)\lambda^N x_k^N = \lambda^N \sqrt{n}\psi_{\mathbf{x}^N}(u).
\end{aligned}$$

Hence, for every eigenvalue-eigenvector pair $(\lambda^N, \mathbf{x}^N)$ of $\boldsymbol{\Delta}$, $(\lambda^N, \sqrt{n}\psi_{\mathbf{x}^N})$ is one of $T_{W_{\boldsymbol{A}}}$. Since the image of $T_{W_{\boldsymbol{A}}}$ is $[0,1]/\mathcal{P}_n$ and the induced signals $\{\sqrt{n}\psi_{\mathbf{x}^1}, \dots, \sqrt{n}\psi_{\mathbf{x}^n}\}$ form an orthonormal basis for $[0,1]/\mathcal{P}_n$, there can not exist any other non-zero eigenvalue. □

The following lemma shows that given a graph G with GSO $\mathbf{\Delta}$ and corresponding GWM $\boldsymbol{A}$, we can either first induce it to the graphon space $W_{\boldsymbol{A}}$, then build the GRSO $T_{W_{\boldsymbol{A}}}$ and the filter $h(T_{W_{\boldsymbol{A}}})$, or first apply the discrete filter on the GSO $h(\mathbf{\Delta})$ and, only then, induce it to the graphon space.

Proposition 4.4. *Let G be a graph with GSO $\mathbf{\Delta}$ and corresponding GWM $\boldsymbol{A}$. Let $h : \mathbb{R} \to \mathbb{R}$ be continuous with $h(0) = 0$. Then, $h(T_{\boldsymbol{A}}) = T_{W_{nh(\mathbf{\Delta})}}$.*

Proof. Let $\{\lambda^i, \mathbf{x}^i\}_{i=1}^n$ denote the eigendecomposition of $\mathbf{\Delta}$. Then, $h(\mathbf{\Delta})$ admits the eigendecomposition $\{h(\lambda^i), \mathbf{x}^i\}_{i=1}^n$. By Lemma 4.3, the eigendecomposition of $T_{W_{nh(\mathbf{\Delta})}}$ is $\{h(\lambda^i), \sqrt{n}\psi_{\mathbf{x}^i}\}_{i=1}^n$ with infinitely many eigenvectors corresponding to the kernel $\{0, \psi_j\}_{j=n+1}^{\infty}$.

We compare this to the eigendecomposition of $h(T_{W_{\boldsymbol{A}}})$. By Lemma 4.3, $T_{W_{\boldsymbol{A}}}$ has eigendecomposition $\{\lambda^i, \sqrt{n}\psi_{\mathbf{x}^i}\}_{i=1}^n$ with infinitely many eigenvectors corresponding to the kernel $\{0, \psi_j\}_{j=n+1}^{\infty}$. Then, the eigendecomposition of $h(T_{W_{\boldsymbol{A}}})$ is given by $\{h(\lambda^i), \sqrt{n}\psi_{\mathbf{x}^i}\}_{i=1}^n \cup \{h(0), \psi_j\}_{j=n+1}^{\infty}$. The condition $h(0) = 0$, leads to $h(T_{W_{\boldsymbol{A}}}) = T_{W_{nh(\mathbf{\Delta})}}$. □

The condition $h(0) = 0$ in Proposition 4.4 can be removed if we restrict the input signals to induced graphon signals $\psi_{\mathbf{x}}$, instead of studying $T_{W_{nh(\mathbf{\Delta})}}$ and $h(T_{W_{\boldsymbol{A}}})$ on the whole space $L^2(\mathcal{J})$.

Proposition 4.5. *Let G be a graph with GSO $\mathbf{\Delta}$ and corresponding GWM $\boldsymbol{A}$. Let $h : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then, for every signal $\mathbf{x} \in \mathbb{C}^n$*

$$\psi_{h(\mathbf{\Delta})\mathbf{x}} = h(T_{W_{\boldsymbol{A}}})\psi_{\mathbf{x}}. \tag{17}$$

Proof. Similarly to the calculations and notations from the proof of Proposition 4.4, we can write

$$h(\mathbf{\Delta})\mathbf{x} = \sum_{i=1}^{n} h(\lambda^i)\langle \mathbf{x}, \mathbf{x}^i\rangle_{\mathbb{C}^n}\mathbf{x}^i$$

and

$$h(T_{W_{\boldsymbol{A}}})\psi_{\mathbf{x}} = \sum_{i=1}^{n} h(\lambda^i)\langle \psi_{\mathbf{x}}, \sqrt{n}\psi_{\mathbf{x}^i}\rangle_{L^2([0,1])}\sqrt{n}\psi_{\mathbf{x}^i}$$

for $\mathbf{x} \in \mathbb{C}^n$. By Definition 3.11 of induced signals, we have

$$\begin{aligned}
\psi_{h(\mathbf{\Delta})\mathbf{x}}(\cdot) &= \sum_{j=1}^{n} \chi_{I_j}(\cdot) \sum_{i=1}^{n} h(\lambda^i)\langle \mathbf{x}, \mathbf{x}^i\rangle_{\mathbb{C}^n} x_j^i = \sum_{i=1}^{n} h(\lambda_i)\langle \mathbf{x}, \mathbf{x}^i\rangle_{\mathbb{C}^n} \sum_{j=1}^{n} \chi_{I_j}(\cdot) x_j^i \\
&= \sum_{i=1}^{n} h(\lambda_i)\langle \mathbf{x}, \mathbf{x}^i\rangle_{\mathbb{C}^n}\psi_{\mathbf{x}^i} = \sum_{i=1}^{n} h(\lambda_i)\langle \psi_{\mathbf{x}}, \sqrt{n}\psi_{\mathbf{x}^i}\rangle_{L^2([0,1])}\sqrt{n}\psi_{\mathbf{x}^i} \\
&= h(T_{W_{\boldsymbol{A}}})\psi_{\mathbf{x}}.
\end{aligned}$$

The second to last equation holds by Remark 4.2. □

4.2. General unbounded shift operator

Graphon shift operators cannot represent all important limit objects of GSOs. An important example are filters in Euclidean domains, which are the backbone of standard convolutional neural networks.

Example 4.6. Let D denote the space of 1-periodic functions on $\mathbb{R}$ with absolutely continuous first derivative, equipped with the $L^2[0,1]$ inner product. Let Δ be the Laplace operator,

$$\Delta : D \to D, \quad f \mapsto -f''.$$

Let h be a function and $(h_n)_n = (h(4\pi^2 n^2))_n$. Define the Euclidean circular convolution operator[3] C_h by

$$C_h f = f * (\sum h_n e^{2\pi i n(\cdot)})$$

for any input signal f. Operators of this kind exhaust the space of all circular convolution operators with even kernels. By the convolution theorem, this can be written in the frequency domain as

$$\mathcal{F} C_h \mathcal{F}^*(\hat{f}_n)_n = (h_n \hat{f}_n)_n.$$

Applying the inverse Fourier transform leads to

$$C_h f = \sum_n \hat{f}_n h_n e^{2\pi i n(\cdot)} = h(\Delta) f.$$

This means that Euclidean circular convolution operators can be written as functional calculus filters of the Laplace operator.

We consider a discretization of $[0,1]$ by a grid. The grid is a graph, defined by connecting each grid point to its neighbors by edges. This means that $[0,1]$, with a grid, are a special case of a graphon latent space and a graph. In this subsection, we show algorithmic alignment of GCNNons with CNNs on Euclidean data, where filters are defined by functional calculus of the Laplace operator as described in Example 4.6. The Laplace operator is not bounded, so there exists no graphon W with $\Delta = T_W$. To be able to treat discrete grid convolutions as graph convolutions approximating Δ, our analysis should allow *unbounded GRSOs*. We first define the Paley-Wiener spaces of self-adjoint operators.

Definition 4.7. Let $\mathcal{L}$ be a self-adjoint operator with spectrum consisting only of eigenvalues. Denote the eigenvalues, eigenspaces, and projections upon the eigenspaces of $\mathcal{L}$ by $\{\lambda_j, W_j, P_j\}_{j\in\mathbb{N}}$. For each $\lambda > 0$, we define the λ'th *Paley-Wiener* space of $\mathcal{L}$ as

$$PW_{\mathcal{L}}(\lambda) = \bigoplus_{j\in\mathbb{N}} \{W_j \,|\, |\lambda_j| \leq \lambda\}.$$

The *spectral projection* $P_{\mathcal{L}}(\lambda)$ upon $PW_{\mathcal{L}}(\lambda)$ is defined by

$$P_{\mathcal{L}}(\lambda) = \sum_{|\lambda_j|\leq\lambda} P_j.$$

Definition 4.8. Let $\mathcal{L}$ be a self-adjoint operator defined on a dense subset of $L^2[0,1]$, with spectrum consisting only of eigenvalues. We say that $\mathcal{L}$ is an *unbounded graphon shift operator (unbounded GRSO)* if it is self-adjoint and for every $\lambda > 0$, there exists a graphon $W^\lambda \in \mathcal{W}$ such that

$$\mathcal{L} P_{\mathcal{L}}(\lambda) = T_{W^\lambda}. \tag{18}$$

[3] A circular convolution in $L^2[0,1]$ between f and q is defined by $f * q(x) = \int_0^1 f(y) q(x-y) dy$, where values of q outside $[0,1]$ are taken from the periodic extension of q.

The following proposition can be easily verified.

Proposition 4.9. *A self-adjoint operator $\mathcal{L}$ with spectrum $\sigma(\mathcal{L}) = \{\lambda_1, \lambda_2, \ldots\}$ consisting only of eigenvalues is an unbounded GRSO if and only if the eigenvalues in each compact interval are square summable, i.e., for every $a < b \in \mathbb{R}$*

$$\sum_{\lambda \in \sigma(\mathcal{L}) \cap [a,b]} \lambda^2 < \infty.$$

Next, we define convergence of GSOs to unbounded graphon shift operators.

Definition 4.10. Let $(G_n)_n$ be a sequence of graphs with GSOs $(\boldsymbol{\Delta}_n)_n$ and $(\boldsymbol{A}_n)_n$ be the associated sequence of GWMs. We say that $(G_n)_n$ converges to the unbounded GRSO $\mathcal{L}$ if there exists a sequence of permutations $(\pi_n)_n$ such that for every $\lambda \in \mathbb{R}$

$$\|P_{\mathcal{L}}(\lambda) T_{W_{\pi_n(\boldsymbol{A}_n)}} P_{\mathcal{L}}(\lambda) - \mathcal{L} P_{\mathcal{L}}(\lambda)\|_{L^2[0,1] \to L^2[0,1]} \xrightarrow{n \to \infty} 0. \tag{19}$$

We write $G_n \xrightarrow[n\to\infty]{U} \mathcal{L}$.

In the following example, we show that the classical Laplace operator is an unbounded GRSO.

Example 4.11. Let Δ be the Laplace operator from Example 4.6. The spectrum of Δ consists only of eigenvalues, given by

$$\sigma(\Delta) = \{0, 4\pi^2, 4\pi^2 2^2, \ldots, 4\pi^2 k^2, \ldots\}.$$

The corresponding eigenvectors $e^{2i\pi kx}$ form the classical Fourier basis. For every $a < b \in \mathbb{R}$, the intersection $\sigma(\Delta) \cap [a, b]$ is a finite set. Hence, by Proposition 4.9, the Laplace operator Δ is an unbounded GRSO.

There exists a sequence of graphs approximating the Laplace operator in the sense of Definition 4.10. The construction of the approximating sequence of graphs can be taken from the proof of Proposition 4.12, showing a more general result.

Proposition 4.12. *Let $\mathcal{L}$ be an unbounded GRSO. Then, there exists a sequence of weighted graphs $(G_n)_n$ with GSOs $(\boldsymbol{\Delta}_n)_n$ such that $G_n \xrightarrow[n\to\infty]{U} \mathcal{L}$.*

The proof of Proposition 4.12 is given in Appendix A.1.

5. Main results on transferability

First, we give results on asymptotic transferability of continuous graph filters. Secondly, by restricting the analysis to filters with higher regularity, we achieve linear approximation speed. Then, we prove transferability results for graphs that approximate unbounded graphon shift operators. Last, we present transferability results for end-to-end graph convolutional neural networks.

5.1. Asymptotic transferability of graph filters

Given a sequence of graphs $(G_n)_n$ with GSOs $(\boldsymbol{\Delta}_n)_n$ and a graphon W such that $G_n \xrightarrow[n\to\infty]{H} W$, we show that the associated sequence of graph filters $(h(\boldsymbol{\Delta}_n))_n$ is convergent in operator norm. The limit object is the filtered induced GRSO $h(T_W)$.

Lemma 5.1 *(Convergence of graph filters). Let $(G_n)_n$ be a sequence of graphs with GSOs $(\mathbf{\Delta}_n)_n$ and its associated sequence of GWMs $(\boldsymbol{A}_n)_n$. Let $W \in \mathcal{W}$ satisfy $G_n \xrightarrow[n\to\infty]{H} W$. Let h be a filter. Then, there exists a sequence of permutations $(\pi_n)_n$ such that*

$$\|h(T_{W_{\pi_n(\boldsymbol{A}_n)}}) - h(T_W)\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0.$$

Consequently, consider a sequence of graph signals $(\mathbf{x}_n)_n$ on the relabeled sequence of graphs. Assume that the induced signals, defined in Definition 3.11, converge to some $\psi \in L^2[0,1]$, i.e., $\|\psi_{\mathbf{x}_n} - \psi\|_{L^2[0,1]} \xrightarrow{n\to\infty} 0$. Then,

$$\|h(T_{W_{\pi_n(\boldsymbol{A}_n)}})\psi_{\mathbf{x}_n} - h(T_W)\psi\|_{L^2[0,1]} \xrightarrow{n\to\infty} 0. \tag{20}$$

The proof for Lemma 5.1 is given in Appendix A.2. In practice, filters are applied on graph shift operators. Hence, we want to show that the outputs of graph filters applied on GSOs, associated to graphs that approximate the same graphon, are approximately the same. The following corollaries formulate the transferability error by first applying the filter on GSOs and then taking the induced graphons.

Theorem 5.2. *Let $(G_n)_n$ be a sequence of graphs with GSOs $(\mathbf{\Delta}_n)_n$ and its associated sequence of GWMs $(\boldsymbol{A}_n)_n$. Let $W \in \mathcal{W}$ satisfy $G_n \xrightarrow[n\to\infty]{H} W$. Let h be a filter with $h(0) = 0$. Then,*

$$\|T_{W_{nh(\pi_n(\mathbf{\Delta}_n))}} - T_{W_{mh(\pi_m(\mathbf{\Delta}_m))}}\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n,m\to\infty} 0.$$

Proof. By Proposition 4.4 and Lemma 5.1, we have

$$\|T_{W_{nh(\pi_n(\mathbf{\Delta}_n))}} - h(T_W)\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0.$$

The proof now follows from the triangle inequality. □

In the following result we get rid of the assumption of $h(0) = 0$ from Theorem 5.2 by considering the convergence of induced graphon signals instead of the convergence of induced GRSOs.

Theorem 5.3. *Let $(G_n)_n$ be a sequence of graphs with GSOs $(\mathbf{\Delta}_n)_n$ and its associated sequence of GWMs $(\boldsymbol{A}_n)_n$. Let $W \in \mathcal{W}$ satisfy $G_n \xrightarrow[n\to\infty]{H} W$. Let h be a filter. Consider a sequence of graph signals $(\mathbf{x}_n)_n$ on $(G_n)_n$ and a graphon signal $\psi \in L^2[0,1]$ such that $\mathbf{x}_n \xrightarrow[n\to\infty]{I} \psi$. Then,*

$$\|\psi_{h(\mathbf{\Delta}_n)\mathbf{x}_n} - \psi_{h(\mathbf{\Delta}_m)\mathbf{x}_m}\|_{L^2[0,1]} \xrightarrow{n,m\to\infty} 0.$$

Proof. By Proposition 4.5 and the Lemma 5.1, we have $h(\mathbf{\Delta}_n)\mathbf{x}_n \xrightarrow[n\to\infty]{I} h(T_W)\psi$. The proof now follows from the triangle inequality. □

5.2. Non-asymptotic transferability of graph filters

In the following, we derive filters $h : \mathbb{R} \to \mathbb{R}$ that allow improved rates of convergence. For $U \subset \mathbb{R}$ and $p \in \mathbb{N}$, we write $f \in C^p(U)$ if f is p times continuously differentiable for all $x \in U$.

We begin by calculating the convergence rate for the transferability in the *Schatten p-norm*. For $1 \le p \le \infty$, the Schatten p-norm $\|\cdot\|_{S_p}$ of a compact self-adjoint operator is defined as the l_p-norm of the sequence of its eigenvalues.

Theorem 5.4. *Let h be Lipschitz continuous with Lipschitz constant $\|h\|_{Lip}$ and $h(0) = 0$. Let $(G_n)_n$ be a sequence of graphs with GSOs $(\boldsymbol{\Delta}_n)_n$ and associated GWMs $(\boldsymbol{A}_n)_n$ such that there exists a $W \in \mathcal{W}$ with $G_n \xrightarrow[n\to\infty]{H} W$. There exists a sequence of permutations $(\pi_n)_n$ such that for every $2 < p < \infty$*

$$\begin{aligned}&\|T_{W_{nh(\pi_n(\boldsymbol{\Delta}_n))}} - T_{W_{mh(\pi_m(\boldsymbol{\Delta}_m))}}\|_{S_p}\\ &\leq 2\|h\|_{Lip}K_p\|T_{W_{\pi_n(\boldsymbol{\Delta}_n)}} - T_{W_{\pi_m(\boldsymbol{\Delta}_m)}}\|_{S_p} \xrightarrow{n,m\to\infty} 0,\end{aligned}$$

where K_p is a universal constant described in [29, Theorem 1].

Remark 5.5. If we restrict ourselves in Theorem 5.4 to graphons $W \in \mathcal{W}_\Gamma$ for a $\Gamma > 0$, the convergence rate in Theorem 5.4 holds for any filter h that is Lipschitz continuous in a neighborhood of $[-\Gamma, \Gamma]$. Thus, the convergence rate in Theorem 5.4 holds also for polynomials.

The proof for Theorem 5.4 is given in Appendix A.3. We give another formulation of transferability based on Fourier coefficients of the filter h.

Theorem 5.6 *(Linear convergence). Let $\Gamma \in \mathbb{N}$ and $h \in C^1([-\Gamma, \Gamma])$ such that h' is Lipschitz and $h(0) = 0$. Suppose that G_1 and G_2 are graphs with GSOs $\boldsymbol{\Delta}_1 \in \mathbb{R}^{n_1\times n_1}$ and $\boldsymbol{\Delta}_2 \in \mathbb{R}^{n_2\times n_2}$. Let $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ be the associated GWMs and suppose that there exists a $\Gamma > 0$ with $\|\boldsymbol{A}_1\|_{\mathbb{C}^{n_1}\to\mathbb{C}^{n_1}}, \|\boldsymbol{A}_2\|_{\mathbb{C}^{n_2}\to\mathbb{C}^{n_2}} \leq \Gamma$. We have for all permutations π_1 and π_2 of the graph's nodes*

$$\begin{aligned}&\|T_{W_{n_1h(\pi_1(\boldsymbol{\Delta}_1))}} - T_{W_{n_2h(\pi_2(\boldsymbol{\Delta}_2))}}\|_{L^2[0,1]\to L^2[0,1]}\\ &\qquad\leq C\|T_{W_{\pi_1(\boldsymbol{A}_1)}} - T_{W_{\pi_2(\boldsymbol{A}_2)}}\|_{L^2[0,1]\to L^2[0,1]},\end{aligned} \tag{21}$$

where C is given by (A.10).

5.3. Transferability between graphs and unbounded GRSOs

In this subsection, we generalize the asymptotic transferability results of Subsection 5.1 to unbounded GRSOs. The following lemma is a direct consequence of Lemma Appendix A.6.

Lemma 5.7. *Let $h : \mathbb{R} \to \mathbb{R}$ be a continuous function. Let $(G_n)_n$ be a sequence of graphs with GSOs $(\boldsymbol{\Delta}_n)_n$ and associated GWMs $(\boldsymbol{A}_n)_n$. Suppose that $\mathcal{L}$ is an unbounded GRSO such that $G_n \xrightarrow[n\to\infty]{U} \mathcal{L}$. Then, there exists a sequence of permutations $(\pi_n)_n$ such that for every $\lambda > 0$*

$$\|h(P_{\mathcal{L}}(\lambda)T_{W_{\pi_n(\boldsymbol{A}_n)}}P_{\mathcal{L}}(\lambda)) - h(\mathcal{L}P_{\mathcal{L}}(\lambda))\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0.$$

In Lemma 5.7, the filter is applied to the induced graphon shift operator, projected to the Paley-Wiener space. However, in practical graph deep learning, the filters are applied to the GSO. To bridge the gap between Lemma 5.7 and practical graph deep learning, we introduce the following definition.

Definition 5.8. Let $(G_n)_n$ be a sequence of graphs with GWMs $(\boldsymbol{A}_n)_n$, Θ be a family of filters, and $\mathcal{L}$ be an unbounded GSO. We say that Θ *approximately commutes* with $P_{\mathcal{L}}$ on $(G_n)_n$ if for every $h \in \Theta$ and every $\lambda > 0$

$$\|P_{\mathcal{L}}(\lambda)h(T_{\boldsymbol{A}_n})P_{\mathcal{L}}(\lambda) - h\big(P_{\mathcal{L}}(\lambda)T_{\boldsymbol{A}_n}P_{\mathcal{L}}(\lambda)\big)\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0.$$

In the following claim, we show that the family of monomial filters approximately commutes with P_Δ in our motivating example – the classical Laplace operator Δ.

Claim 5.9. *Let Δ be the Laplace operator from Example 4.11. Suppose that $(\boldsymbol{\Delta}_n)_n$ is the finite difference GSO, i.e.,*

$$\boldsymbol{\Delta}_n = \begin{bmatrix} -2 & 1 & 0 & \dots & 0 & 1 \\ 1 & -2 & 1 & 0 & \dots & 0 \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & 1 & -2 & 1 \\ 1 & 0 & \dots & 0 & 1 & -2 \end{bmatrix}.$$

Let $(G_n)_n$ be the corresponding sequence of graphs, and $(\boldsymbol{A}_n)_n$ the GWMs. Let $\Theta = \{z^k \,|\, k \in \mathbb{N}\}$ be the family of monomial filters. Then, Θ approximately commutes with P_Δ on $(G_n)_n$, i.e., for every $\lambda > 0$ and $k \in \mathbb{N}$,

$$\|P_\Delta(\lambda) T^k_{\boldsymbol{A}_n} P_\Delta(\lambda) - \Big(P_\Delta(\lambda) T_{\boldsymbol{A}_n} P_\Delta(\lambda)\Big)^k \|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0.$$

The proof of Claim 5.9 is in Appendix A.5. The following lemma extends approximate commutations of Θ with $P_\mathcal{L}$ to linear combinations of Θ.

Lemma 5.10. *Let $(G_n)_n$ be a sequence of graphs with GWMs $(\boldsymbol{A}_n)_n$, $\mathcal{F}$ be a family of filters and $\mathcal{L}$ an unbounded GSO. Suppose that Θ approximately commutes with $P_\mathcal{L}$ on $(G_n)_n$. Then,* $\mathrm{span}\Theta$ *approximately commutes with $P_\mathcal{L}$ on $(G_n)_n$.*

Proof. Let $h = \sum_{k=0}^n \alpha_k h_k \in \mathrm{span}\Theta$. Then, for $\lambda > 0$

$$\begin{aligned}
&\|P_\mathcal{L}(\lambda) h(T_{W_{\boldsymbol{A}_n}}) P_\mathcal{L}(\lambda) - h\Big(P_\mathcal{L}(\lambda) T_{W_{\boldsymbol{A}_n}} P_\mathcal{L}(\lambda)\Big)\|_{L^2[0,1]\to L^2[0,1]} \\
&= \|P_\mathcal{L}(\lambda)\Big(\sum_{k=0}^n \alpha_k h_k(T_{W_{\boldsymbol{A}_n}})\Big) P_\mathcal{L}(\lambda) - \sum_{k=0}^n \alpha_k h_k\Big(P_\mathcal{L}(\lambda) T_{W_{\boldsymbol{A}_n}} P_\mathcal{L}(\lambda)\Big)\|_{L^2[0,1]\to L^2[0,1]} \\
&= \|\sum_{k=0}^n \alpha_k P_\mathcal{L}(\lambda) h_k(T_{W_{\boldsymbol{A}_n}}) P_\mathcal{L}(\lambda) - \sum_{k=0}^n \alpha_k h_k\Big(P_\mathcal{L}(\lambda) T_{W_{\boldsymbol{A}_n}} P_\mathcal{L}(\lambda)\Big)\|_{L^2[0,1]\to L^2[0,1]} \\
&\le \sum_{k=0}^n |\alpha_k| \|P_\mathcal{L}(\lambda) h_k(T_{W_{\boldsymbol{A}_n}}) P_\mathcal{L}(\lambda) - h_k\Big(P_\mathcal{L}(\lambda) T_{W_{\boldsymbol{A}_n}} P_\mathcal{L}(\lambda)\Big)\|_{L^2[0,1]\to L^2[0,1]} \\
&\xrightarrow{n\to\infty} 0. \quad \square
\end{aligned}$$

The following proposition shows that filters that approximately commute with spectral projections of an unbounded GRSOs are transferable between graphs.

Theorem 5.11. *Let $(G_n)_n$ be a sequence of graphs with GSOs $(\boldsymbol{\Delta}_n)_n$ and associated GWMs $(\boldsymbol{A}_n)_n$. Suppose that $\mathcal{L}$ is an unbounded GRSO such that $G_n \xrightarrow[n\to\infty]{U} \mathcal{L}$. Let Θ be a family of filters with $h(0) = 0$ for every $h \in \Theta$. Suppose that Θ approximately commutes with $P_\mathcal{L}$ on $(G_n)_n$. Then, there exists a sequence of permutations $(\pi_n)_n$ such that for every $\lambda > 0$ and every $h \in \mathrm{span}\Theta$*

$$\|P_\mathcal{L}(\lambda) T_{W_{nh(\pi_n(\boldsymbol{\Delta}_n))}} P_\mathcal{L}(\lambda) - P_\mathcal{L}(\lambda) T_{W_{mh(\pi_m(\boldsymbol{\Delta}_m))}} P_\mathcal{L}(\lambda)\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n,m\to\infty} 0.$$

Proof. Let $h \in \mathrm{span}\Theta$ and $(\pi_n)_n$ be the sequence that we get from Lemma 5.7. It holds $h(0) = 0$. We use the triangle inequality to achieve

$$\begin{aligned}
&\|P_{\mathcal{L}}(\lambda)T_{W_{nh(\pi_n(\boldsymbol{\Delta}_n))}}P_{\mathcal{L}}(\lambda) - h(\mathcal{L})P_{\mathcal{L}}(\lambda)\|_{L^2[0,1]} \\
&\leq \|P_{\mathcal{L}}(\lambda)T_{W_{nh(\pi_n(\boldsymbol{\Delta}_n))}}P_{\mathcal{L}}(\lambda) - h(P_{\mathcal{L}}(\lambda)T_{W_{\pi_n(\boldsymbol{A}_n)}}P_{\mathcal{L}}(\lambda))\|_{L^2[0,1]} \\
&+ \|h(P_{\mathcal{L}}(\lambda)T_{W_{\pi_n(\boldsymbol{A}_n)}}P_{\mathcal{L}}(\lambda)) - h(\mathcal{L})P_{\mathcal{L}}(\lambda)\|_{L^2[0,1]}.
\end{aligned}$$

Since span$\mathcal{F}$ approximately commutes with $P_{\mathcal{L}}$ on G_n, by Lemma 5.7 and Lemma 4.4, the last term goes to zero for $n \to \infty$. The proof follows by application of the triangle inequality. □

As an example application of Theorem 5.11 by Claim 5.9 and Lemma 5.10, polynomial filters are transferable between graphs with different finite difference GSO.

5.4. Asymptotic transferability of graph neural networks

In this subsection, we prove transferability of SCNNs between graphs approximating the same graphon. GCNNs and GCNNons are nonlinear operators, so we do not measure their repercussion error in operator norm. Instead, we analyze the repercussion error of GCNNs and GCNNons point-wise on feature maps. For this, we give the following definitions.

Definition 5.12. Let $d \in \mathbb{N}$ and $y \in L^2[0,1]^d$ be a graphon feature map and $\mathbf{x} \in \mathbb{C}^{n\times d}$ a graph feature map.

(1) We define

$$\|y\|_{L^2[0,1]^d} = \max_{k=1,\dots,d} \|y^k\|_{L^2[0,1]}. \tag{22}$$

$$\|\mathbf{x}\|_{\mathbb{C}^{n\times d}} = \max_{k=1,\dots,d} \|\mathbf{x}^k\|_{\mathbb{C}^n}. \tag{23}$$

(2) We say that y is *normalized* if $\max_{k=1,\dots,d} \|y^k\|_{L^2[0,1]} \leq 1$. We say that $\mathbf{x}$ is *normalized* if $\max_{k=1,\dots,d} \|\mathbf{x}^k\|_{\mathbb{C}^n} \leq 1$.

(3) Let $(G_n)_n$ be a sequence of graphs with GSOs and $(\mathbf{x}_n)_n$ a sequence of graph feature maps with $\mathbf{x}_n \in \mathbb{C}^{n\times d}$ for every $n \in \mathbb{N}$. Let $\mathbf{x}_n^k$ and y^k denote the k-th feature of $\mathbf{x}_n$ and y. We say that $(\mathbf{x}_n)_n$ converges to y if for every $k = 1, \dots, n$ it holds $\mathbf{x}_n^k \xrightarrow[n\to\infty]{I} y^k$, (see Definition 3.13). We write in this case $\mathbf{x}_n \xrightarrow[n\to\infty]{I} y$.

The following lemma shows the transferability of SCNN between graphons and induced graphons from a convergent sequence of graphs.

Lemma 5.13. *Let $\phi := (H, \mathbf{M}, \rho)$ be an SCNN with L layers and Lipschitz continuous activation function $\rho : \mathbb{R} \to \mathbb{R}$. Let F_l be the number of features in the l-th layer, for $l = 1, \dots, L$. Let $(G_n)_n$ be a sequence of graphs with GSOs $(\boldsymbol{\Delta}_n)_n$ and associated GWMs $(\boldsymbol{A}_n)_n$. Let $W \in \mathcal{W}$ satisfy $G_n \xrightarrow[n\to\infty]{H} W$. Suppose that $(y_n)_n$ is a sequence in $(L^2[0,1])^{F_0}$ and $y \in (L^2[0,1])^{F_0}$ such that*

$$\|y_n - y\|_{L^2[0,1]^{F_0}} \xrightarrow{n\to\infty} 0. \tag{24}$$

Then there exists a sequence of permutations $(\pi_n)_n$ such that,

$$\|\phi_{W_{\pi_n(\boldsymbol{A}_n)}}(y_n) - \phi_W(y)\|_{L^2[0,1]^{F_l}} \xrightarrow{n\to\infty} 0. \tag{25}$$

The proof of Lemma 5.13 is given in Appendix A.6. The following Lemma is a generalization of Proposition 4.5 to SCNNs.

Lemma 5.14. *Let $\phi := \phi(H, \mathbf{M}, \rho)$ be an SCNN. Let G be a graph with GSO $\boldsymbol{\Delta}$ and associated GWM $\boldsymbol{A}$. Let $\mathbf{x} \in \mathbb{C}^{n\times d}$. Then, $\psi_{\phi_{\boldsymbol{\Delta}}(\mathbf{x})} = \phi_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})$.*

Proof. We prove by induction over the layers $l = 1, \dots, L$. Let $l = 1$, and $j = 1, \dots, F_1$. It holds

$$\phi^1_{\Delta}(\mathbf{x})^j = \rho\Big(\sum_{k=1}^{F_0} m_1^{jk} h_1^{jk}(\Delta)\mathbf{x}^k\Big) = \rho(\tilde{\phi}^1_{\Delta}(\mathbf{x})^j).$$

We induce the resulting signal to the graphon space, and obtain

$$\begin{aligned}\psi^1_{\rho(\tilde{\phi}_{\Delta}(\mathbf{x})^j)} &= \sum_{i=1}^{n} \chi_{P_i}(\cdot)\rho(\tilde{\phi}^1_{\Delta}(\mathbf{x})^j_i) = \rho\Big(\sum_{i=1}^{n} \chi_{P_i}(\cdot)(\tilde{\phi}^1_{\Delta}(\mathbf{x})^j_i)\Big) \\ &= \rho(\psi^1_{\tilde{\phi}_{\Delta}(\mathbf{x})^j}) = \rho\Big(\sum_{k=1}^{F_{L-1}} \psi_{m_1^{jk} h_1^{jk}(\Delta)\mathbf{x}^k}\Big) = \rho\Big(\sum_{k=1}^{F_0} m_1^{jk} h_1^{jk}(T_{W_{\boldsymbol{A}}})\psi_{\mathbf{x}^k}\Big) \\ &= \phi^1_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^j.\end{aligned} \tag{26}$$

The third-to-last equality holds since the induction operator commutes with the addition of filters. The second-to-last equality holds by Proposition 4.5. The proof follows by a similar computation for the other layers. □

The following result is a generalization of Theorem 5.3 to GCNNs.

Theorem 5.15. *Consider the same conditions as in Lemma 5.13. Let $(\mathbf{x}_n)_n$ be a sequence of graph feature maps, and $y \in (L^2[0,1])^{F_0}$, such that $\mathbf{x}_n \xrightarrow[n\to\infty]{I} y$. Then, there exists a sequence of permutations $(\pi_n)_n$ such that*

$$\|\psi_{\phi_{\pi_n(\boldsymbol{\Delta}_n)}(\mathbf{x}_n)} - \psi_{\phi_{\pi_m(\boldsymbol{\Delta}_m)}(\mathbf{x}_m)}\|_{L^2[0,1]^{F_L}} \xrightarrow{n,m\to\infty} 0.$$

Proof. By Lemma 5.13 and Lemma 5.14, there exists a sequence of permutations $(\pi_n)_n$ such that

$$\|\psi_{\phi_{\pi_n(\boldsymbol{\Delta}_n)}(\mathbf{x}_n)} - \phi_W(y)\|_{L^2[0,1]^{F_L}} \xrightarrow{n\to\infty} 0.$$

The proof now follows by the triangle inequality. □

5.5. Non-asymptotic transferability of graph neural networks

In this section, we derive stability bounds for GCNNs. We restrict ourselves to *contractive* activation functions.

Definition 5.16. An activation function $\rho : \mathbb{R} \to \mathbb{R}$ is called *contractive* if $|\rho(x) - \rho(y)| \leq |x - y|$ for every $x, y \in \mathbb{R}$.

The following lemma shows that GCNNons, based on filters with higher regularity, achieve linear approximation rate.

Lemma 5.17. *Let $\phi := \phi(H, \mathbf{M}, \rho)$ be an SCNN with contractive activation function ρ. Let G be a graph with GSO $\boldsymbol{\Delta}$ and associated GWM $\boldsymbol{A}$. Let $W \in \mathcal{W}$ and suppose that*

$$\|\boldsymbol{A}\|_{\mathbb{C}^n\to\mathbb{C}^n} \leq \Gamma, \quad \|T_W\|_{L^2[0,1]\to L^2[0,1]} \leq \Gamma$$

for $\Gamma > 0$. *Let* $\varepsilon > 0$, *and suppose there exists a node permutation* π *of* G *satisfying*

$$\|T_{W_{\pi(\boldsymbol{A})}} - T_W\|_{L^2[0,1]\to L^2[0,1]} \leq \varepsilon.$$

Let $\mathbf{x}$ *be a normalized graph feature map and* y *a graphon feature map satisfying*

$$\|\psi_{\mathbf{x}} - y\|_{L^2[0,1]^{F_0}} \leq \varepsilon.$$

Suppose that for every filter h *in the SCNN* ϕ, *we have* $h \in C^1[-\Gamma,\Gamma]$ *with Lipschitz continuous derivative and* $\|h\|_{L^\infty[-\Gamma,\Gamma]} \leq 1$. *Then, there exists a* $C_L \in \mathbb{N}$ *such that*

$$\|\phi_{W_{\pi(\boldsymbol{A})}}(\psi_{\mathbf{x}}) - \phi_W(y)\|_{L^2[0,1]^{F_L}} \leq C_L\varepsilon, \tag{27}$$

where C_L *is a constant that depends on* $\mathbf{M}$ *and* H, *and is described in Remark 5.18.*

The proof of Lemma 5.17 is in Appendix A.7.

Remark 5.18. Consider the setting of Lemma 5.17. For every $l = 1,\dots,L$, let C_l^{jk} denote the constant C of (21) corresponding to the filter $h = h_l^{jk}$, for $j = 1,\dots,F_l$, $k = 1,\dots,F_{l-1}$. Let $C = \max_{l,j,k} C_l^{jk}$, where the maximum is over $l = 1,\dots,L$, $j = 1,\dots,F_l$, and $k = 1,\dots,F_{l-1}$. Further, let $M = \max_{l=1,\dots,L} \|\mathbf{M}_l\|_\infty$. Then, the constant C_L in (27) is given by

$$C_L = M^L(1 + LC). \tag{28}$$

Remark 5.19. In Lemma 5.17, the constant C_L can increase exponentially in L. We can control C_L by demanding $\|\mathbf{M}_l\|_\infty \leq 1$ for every $l = 1,\dots,L$, which gives

$$\|\phi_{W_{\pi(\boldsymbol{A})}}(\psi_{\mathbf{x}}) - \phi_W(y)\|_{L^2[0,1]^{F_L}} \leq (1 + LC)\varepsilon.$$

The condition $\|\mathbf{M}_l\|_\infty \leq 1$ is relaxed to weight decay regularization in practical deep learning. Hence (28) indicates that weight decay might be an important regularization for promoting transferability in practical deep learning. A similar observation was given in [23].

The following theorem shows GCNNs achieve linear approximation rate.

Theorem 5.20. *Let* $\phi := \phi(H, \mathbf{M}, \rho)$ *be an SCNN with contractive activation function* ρ. *Let* G_1, G_2 *be graphs with GSOs* $\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2$ *and associated GWM* $\boldsymbol{A}_1, \boldsymbol{A}_2$. *Suppose that*

$$\|\boldsymbol{A}_1\|_{\mathbb{C}^n\to\mathbb{C}^n} \leq \Gamma, \quad \|\boldsymbol{A}_2\|_{\mathbb{C}^m\to\mathbb{C}^m} \leq \Gamma$$

for $\Gamma > 0$. *Let* $\varepsilon > 0$, *and suppose there exist node permutations* π_1 *of* G_1 *and* π_2 *of* G_2 *satisfying*

$$\|T_{W_{\pi_1(\boldsymbol{A}_1)}} - T_{W_{\pi_2(\boldsymbol{A}_2)}}\|_{L^2[0,1]\to L^2[0,1]} \leq \varepsilon.$$

Let $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *be normalized graph feature maps satisfying*

$$\|\psi_{\mathbf{x}_1} - \psi_{\mathbf{x}_2}\|_{L^2[0,1]^{F_0}} \leq \varepsilon.$$

Then,

$$\|\psi_{\phi_{\pi(\mathbf{\Delta_1})}(\mathbf{x_1})} - \psi_{\phi_{\pi(\mathbf{\Delta_2})}(\mathbf{x_2})}\|_{L^2[0,1]^{F_L}} \leq C_L \varepsilon,$$

where C_L is given in (27).

The proof of Theorem 5.20 follows directly from Lemma 5.17 and Lemma 5.14.

6. Conclusions

In this paper, we proved that continuous graph filters and end-to-end graph neural networks with continuous filters are transferable in the graphon sense. Generally, a filter is said to be transferable if it has approximately the same repercussion on graphs that represent the same phenomenon. To specify the definition of transferability in the graphon sense, we used the space of graphons as a latent space from which graphs are sampled, and to which graphs are embedded, via induction, in order to compare them. In our setting, graphs represent the same phenomenon if they approximate the same graphon in homomorphism density.

In Theorem 5.2 we showed that continuous filters are transferable, were the repercussion error is measured in the induced graphon sense. In Theorem 5.15 we showed that any GCNN with continuous filters is transferable, where the repercussion error is measured, for a given graph signals, in the induced signal sense. We provide the first work to consider graphon-based transferability for GCNNs with non-polynomial filters, which are used in GCNN architectures like [25,3].

In Theorem 5.6 and Theorem 5.20 we introduced approximation rates for the repercussion error of filters and GCNNs with filters that have Lipschitz continuous derivative. We showed that the repercussion errors are linearly stable with respect to $\|T_{W_{n_1(\mathbf{\Delta_1})}} - T_{W_{n_2(\mathbf{\Delta_2})}}\|_{L^2[0,1]\to L^2[0,1]}$. The bound from Theorem 5.20 may hint to the importance of weight decay regularization in graph deep learning.

Furthermore, we introduced a framework for analyzing graphs that approximate unbounded GRSOs. In Subsection 4.2 we show that SCNNs on Euclidean spaces with the Laplace operator are classical CNNs. Discretizations of different resolutions of the same Euclidean space by grids are graphs that represent the same phenomenon. To show transferability between classical discrete Euclidean signals on grids of different resolutions, we must be able to work with unbounded GRSOs, since the classical Laplace operator is unbounded. Our analysis allows this, showing that our framework of graph transferability aligns with the stability of classical CNNs between resolutions.

Acknowledgments

S.M acknowledges partial support by the NSF–Simons Research Collaboration on the Mathematical and Scientific Foundations of Deep Learning (MoDL) (NSF DMS 2031985) and DFG SPP 1798, KU 1446/27-2.

R.L. acknowledges support by the DFG SPP 1798, KU 1446/21-2 “Compressed Sensing in Information Processing” through Project Massive MIMO-II.

G.K. acknowledges partial support by the NSF–Simons Research Collaboration on the Mathematical and Scientific Foundations of Deep Learning (MoDL) (NSF DMS 2031985), the ONE Munich Strategy Forum (LMU Munich, TU Munich, and the Bavarian Ministry for Science and Art), the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research, as well as the German Research Foundation under under grants DFG SPP 2298, KU 1446/31-1 and KU 1446/32-1 and under Grant DFG SFB/TR 109, Project C09 and the Federal Ministry of Education and Research under Grant MaGriDo.

Appendix A. Proofs

A.1. Proof of Proposition 4.12

Let $(\lambda_m)_m$ be a sequence of bands such that $\lambda_m \xrightarrow{m\to\infty} \infty$. By Definition 4.8 there exists a graphon $W^{\lambda_m} \in \mathcal{W}$ such that $\mathcal{L}P_{\mathcal{L}}(\lambda_m) = T_{W^{\lambda_m}}$ for every $m \in \mathbb{N}$. Then, for every $m \in \mathbb{N}$, by [6, Theorem 4.5] there exists a sequence of graphs with GSO $(G_n^m)_n$, where n denotes the number of nodes, such that $G_n^m \xrightarrow[n\to\infty]{H} W^{\lambda_m}$. For every $m \in \mathbb{N}$, let $(\boldsymbol{A}_n^m)_n$ be the sequence of GWMs associated to $(G_n^m)_n$. Theorem Appendix A.3 and Theorem Appendix A.4 imply that, after adequate relabeling of the graph nodes, $\|T_{W_{\boldsymbol{A}_n^m}} - T_{W^{\lambda_m}}\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0$. Then, for every $m \in \mathbb{N}$

$$\|T_{W_{\boldsymbol{A}_n^m}} P_{\mathcal{L}}(\lambda_m) - \mathcal{L}P_{\mathcal{L}}(\lambda_m)\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0.$$

We choose for every $m \in \mathbb{N}$ an $n_m \in \mathbb{N}$ such that

$$\|T_{W_{\boldsymbol{A}_{n_m}^m}} P_{\mathcal{L}}(\lambda_m) - \mathcal{L}P_{\mathcal{L}}(\lambda_m)\|_{L^2[0,1]\to L^2[0,1]} \leq 1/m.$$

We denote by abuse of notation $G_m := G_{n_m}^m$ and consider the sequence $(G_m)_m$ with GWMs $(\boldsymbol{A}_m)_m$. It is now easy to see that $(G_m)_m$ converges to $\mathcal{L}$ in the sense of Definition 4.10.

A.2. Proof of Lemma 5.1

We start this subsection by defining the *cut distance* – a pseudo-metric in the space of graphons. As shown in Theorem Appendix A.2 below, convergence in cut distance is equivalent to convergence in homomorphism density. Borgs et al. defined the cut distance, based on the cut norm, as a permutation invariant distance (see [6]). For this, denote by $S_{[0,1]}$ the set of all bijective measure preserving maps between the unit interval and itself.

Definition Appendix A.1. Let $W, U \in \mathcal{W}$ be two graphons. The *cut distance* between W and U is defined by

$$\delta_\square(W,U) := \inf_{f\in S_{[0,1]}} \|W - U^f\|_\square, \tag{A.1}$$

where $U^f(x,y) = U(f(x), f(y))$.

The following theorem from [6, Theorem 3.7] shows that convergence in homomorphism density is equivalent to convergence in cut distance.

Theorem Appendix A.2 *([6, Theorem 3.7]). Let $U, W \in \mathcal{W}$ and $C = \max\{1, \|W\|_\infty, \|U\|_\infty\}$.*

(1) *For a simple graph F with m edges*

$$|t(F,U) - t(F,W)| \leq 4mC^{m-1}\delta_\square(U,W).$$

(2) *If $|t(F,U) - t(F,W)| \leq 3^{-k^2}$ for every simple graph F with k nodes, then*

$$\delta_\square(U,W) \leq \frac{22C}{\sqrt{log_2 k}}.$$

Note that $t(F,G) = t(F, W_{\boldsymbol{A}})$ for every graph G with GWM $\boldsymbol{A}$. As a result, whenever we have a sequence of graphs $(G_n)_n$ with GWMs $(\boldsymbol{A}_n)_n$ and $W \in \mathcal{W}$ such that $G_n \xrightarrow[n\to\infty]{H} W$, then $W_{\boldsymbol{A}_n} \xrightarrow[n\to\infty]{H} W$ as well. Theorem Appendix A.2 implies that this is true if and only if

$$\delta_\square(W_{\boldsymbol{A}_n}, W) \xrightarrow{n\to\infty} 0.$$

Lovász and Szegedy introduced in [26] the *cut norm* on the space of graphons as

$$\|W\|_\square := \sup_{S,T\subset[0,1]} \left| \int_{S\times T} W(u,v)d\mu(u)d\mu(v) \right|, \tag{A.2}$$

taking the supremum over all measurable sets S and T. The following result, taken from [6, Lemma 5.3], shows that convergence in homomorphism density is closely related to convergence in cut norm.

Theorem Appendix A.3 *([6, Lemma 5.3]). Let $(G_n)_n$ a sequence of graphs with GSOs $(\boldsymbol{\Delta}_n)_n$ and $(\boldsymbol{A}_n)_n$ be the corresponding sequence of GWMs. Let $W \in \mathcal{W}$ satisfy $G_n \xrightarrow[n\to\infty]{H} W$. Let $(W_{\boldsymbol{A}_n})_n$ be the sequence of graphons induced by $(G_n)_n$. Then, there exists a sequence of permutations $(\pi_n)_n$ such that $\|W_{\pi_n(\boldsymbol{A}_n)} - W\|_\square \xrightarrow{n\to\infty} 0$.*

Theorem Appendix A.3 together with the following lemma show that convergence in homomorphism density is indeed equivalent (up to relabeling) to convergence of the induced GRSOs in operator norm and Schatten p-norms.

Lemma Appendix A.4. *For any $2 < p < \infty$ and $\Gamma > 0$, if $W \in \mathcal{W}_\Gamma$, then*

$$\|W\|_\square \le \|T_W\|_{L^2[0,1]\to L^2[0,1]} \le \|T_W\|_{S_p} \le \sqrt{2}\Gamma^{1/2+1/p}\|W\|_\square^{1/2-1/p}. \tag{A.3}$$

Lemma A.3 follows directly from [14, Lemma E.7(i)], in which W is assumed to be in $\mathcal{W}_1$. To prove Lemma A.3 we simply rescale the graphon W so it is in $\mathcal{W}_1$, and use [14, Lemma E.7(i)].

Let $\mathcal{H}$ be a Hilbert space. We denote by $\mathcal{B}(\mathcal{H})$ the space of bounded linear operators on $\mathcal{H}$. We denote the spectrum of the operator $A \in \mathcal{B}(\mathcal{H})$ by $\sigma(A)$. The following result shows that, for every filter h, the operator norm of a realization of h is equal to the infinity norm of h [34, Satz VII.1.4].

Lemma Appendix A.5 *([34, Satz VII.1.4]). Let $\mathcal{H}$ be a Hilbert space and $A \in \mathcal{B}(\mathcal{H})$ be normal. Let $h : \mathbb{R} \to \mathbb{C}$ be a continuous function, then $\|h(A)\|_{\mathcal{H}\to\mathcal{H}} = \|\chi_{\sigma(A)}h\|_\infty$, where $\chi_{\sigma(A)}$ is the indicator function of $\sigma(A)$.*

The following lemma shows that, given a convergent sequence of bounded operators on a Hilbert space, the continuous functional calculus preserves the convergence.

Lemma Appendix A.6. *Let $\mathcal{H}$ be a Hilbert space and $\mathcal{S}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ be the space of bounded self-adjoint operators with the operator norm topology. Let $h : \mathbb{R} \to \mathbb{C}$ be a continuous function. Then, the mapping*

$$\mathcal{S}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}), \quad A \mapsto h(A), \tag{A.4}$$

is continuous.

Lemma Appendix A.6 can be easily verified, since the continuity of (A.4) is trivial for polynomials h, and follows for continuous functions from the Weierstrass theorem ([1, Section 8]).

Proof of Lemma 5.1. Let $(G_n)_n$ be a sequence of graphs with GSOs $(\mathbf{\Delta}_n)_n$ and associated sequence of GWMs $(\boldsymbol{A}_n)_n$ such that $G_n \xrightarrow[n\to\infty]{H} W$. Then, by Theorem Appendix A.3 and Lemma Appendix A.4, there exists a sequence of permutation $(\pi_n)_n$ such that

$$\|T_{W_{\pi_n(\boldsymbol{A}_n)}} - T_W\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0.$$

Let $h : \mathbb{R} \to \mathbb{R}$ be continuous. Then, by Lemma Appendix A.6,

$$\|h(T_{W_{\pi_n(\boldsymbol{A}_n)}}) - h(T_W)\|_{L^2[0,1]\to L^2[0,1]} \xrightarrow{n\to\infty} 0.$$

Let $(\mathbf{x}_n)_n$ be as in the formulation of Lemma 5.1. Then,

$$\begin{aligned}
&\|h(T_{W_{\pi_n(\boldsymbol{A}_n)}})\psi_{\mathbf{x}_n} - h(T_W)\psi\|_{L^2([0,1])} \\
&\le \|h(T_{W_{\pi_n(\boldsymbol{A}_n)}})\psi_{\mathbf{x}_n} - h(T_W)\psi_{\mathbf{x}_n}\|_{L^2([0,1])} + \|h(T_W)\psi_{\mathbf{x}_n} - h(T_W)\psi\|_{L^2([0,1])} \\
&\le \|h(T_{W_{\pi_n(\boldsymbol{A}_n)}}) - h(T_W)\|_{L^2[0,1]\to L^2[0,1]}\|\psi_{\mathbf{x}_n}\|_{L^2([0,1])} \\
&+ \|h(T_W)\|_{L^2[0,1]\to L^2[0,1]}\|\psi_{\mathbf{x}_n} - \psi\|_{L^2([0,1])} \xrightarrow{n\to\infty} 0. \quad \square
\end{aligned}$$

A.3. Proof of Theorem 5.4

Let $(G_n)_n$ be a sequence of graphs such that $G_n \xrightarrow[n\to\infty]{H} W$. Then, by Theorem Appendix A.3 and Lemma Appendix A.4, there exists a sequence of permutations $(\pi_n)_n$ such that $\|T_{W_{\pi_n(\boldsymbol{A}_n)}} - T_W\|_{S_p} \xrightarrow{n\to\infty} 0$. Every Lipschitz continuous function h with Lipschitz constant $\|h\|_{Lip}$ satisfies

$$\|h(A) - h(B)\|_{S_p} \le \|h\|_{Lip} K_p \|A - B\|_{S_p}$$

for self-adjoint operators A and B such that $\|A - B\|_{S_p} < \infty$, see [29, Theorem 1]. Hence,

$$\|h(T_{W_{\pi_n(\boldsymbol{A}_n)}}) - h(T_W)\|_p \le K_p \|h\|_{Lip} \|T_{W_{\pi_n(\boldsymbol{A}_n)}} - T_W\|_p.$$

Using Proposition 4.4 and the triangle inequality we finish the proof.

A.4. Proof of Theorem 5.6

We first recall basic results from Fourier analysis. Let $\gamma > 0$. The space $L^2(\gamma\mathbb{T})$ is defined as the space of γ-periodic functions $h : \mathbb{R} \to \mathbb{R}$ with $\|h\|_{L^2[0,\gamma]} < \infty$. For $L^2(\gamma\mathbb{T})$, the Fourier coefficients $\hat{h}(n)$ are defined for every $n \in \mathbb{Z}$ by

$$\hat{h}(n) = \frac{1}{\gamma} \int_{-\gamma/2}^{\gamma/2} h(t) e^{-2\pi i n t/\gamma}.$$

The N-th partial sum of the Fourier series of h is defined by

$$(S_N h)(t) := \sum_{n=-N}^{N} \hat{h}(n) e^{2\pi i n t/\gamma}.$$

The following lemma is taken from [2, Example 5].

Lemma Appendix A.7 *([2, Example 5]). Let $a \in \mathbb{R}$ and A and B be a pair of bounded self-adjoint operators on the Hilbert space $\mathcal{H}$. Then,*

$$\|e^{iAa} - e^{iBa}\|_{\mathcal{H}\to\mathcal{H}} \leq |a|\|A - B\|_{\mathcal{H}\to\mathcal{H}}. \tag{A.5}$$

The following lemma is a step in proving Theorem 5.6.

Lemma Appendix A.8. *Let $\gamma > 0$ and $\varepsilon > 0$. Let $h \in L^2[-\gamma/2, \gamma/2]$ such that $(\hat{h}(n)n)_n \in l^1(\mathbb{Z})$. Let A and B be self-adjoint operators on a Hilbert space $\mathcal{H}$, satisfying $\|A - B\|_{\mathcal{H}\to\mathcal{H}} \leq \varepsilon$ for some $\varepsilon > 0$. Then,*

$$\|h(A) - h(B)\|_{\mathcal{H}\to\mathcal{H}} \leq C\varepsilon, \tag{A.6}$$

where $C = 2 + 2\pi/\gamma\|(\hat{h}(n)n)_n\|_{l^1(\mathbb{Z})}$.

Proof. Let $\varepsilon > 0$. By the fact that $(\hat{h}(n)n)_n \in l^1(\mathbb{Z})$, we have $\|h - S_n h\|_\infty \xrightarrow{n\to\infty} 0$. We choose $N \in \mathbb{N}$ large enough such that $\|h - S_N h\|_\infty \leq \varepsilon$. We evaluate $S_N h(A)$ and $S_N h(B)$ and bound their difference in operator norm. By Lemma Appendix A.7,

$$\begin{aligned}
\|(S_N h)(A) - (S_N h)(B)\|_{\mathcal{H}\to\mathcal{H}} &= \|\sum_{n=-N}^{N} \hat{h}(n)e^{2\pi inA/\gamma} - \sum_{n=-N}^{N} \hat{h}(N)e^{2\pi inB/\gamma}\|_{\mathcal{H}\to\mathcal{H}} \\
&\leq \sum_{k=-N}^{N} |\hat{h}(n)||n|2\pi/\gamma\|A - B\|_{\mathcal{H}\to\mathcal{H}} \\
&\leq C2\pi/\gamma\|A - B\|_{\mathcal{H}\to\mathcal{H}},
\end{aligned}$$

where $C := \|(\hat{h}(n)n)_n\|_{l^1(\mathbb{N})}$. We summarize,

$$\begin{aligned}
\|h(A) - h(B)\|_{\mathcal{H}\to\mathcal{H}} &\leq \|h(A) - (S_N h)(A)\|_{\mathcal{H}\to\mathcal{H}} + \|(S_N h)(A) - (S_N h)(B)\|_{\mathcal{H}\to\mathcal{H}} \\
&\quad + \|(S_N h)(B) - h(B)\|_{\mathcal{H}\to\mathcal{H}} \\
&\leq \|h - S_N h\|_\infty + \|(S_N h)(A) - (S_N h)(B)\|_{\mathcal{H}\to\mathcal{H}} + \|h - S_N h\|_\infty \\
&\leq \varepsilon + C2\pi/\gamma\|A - B\|_{\mathcal{H}\to\mathcal{H}} + \varepsilon \leq (2 + 2\pi C/\gamma)\varepsilon,
\end{aligned}$$

where the second inequality holds by Lemma Appendix A.5. □

We continue by formulating sufficient conditions for filter functions $h \in L^2[-\Gamma, \Gamma]$ to fulfil the assumptions of Lemma Appendix A.8.

Definition Appendix A.9. A *modulus of continuity* is a function $w : [0, \infty] \to [0, \infty]$. A function f admits w as a modulus of continuity if

$$|f(x) - f(y)| \leq w(|x - y|),$$

for all x and y in the domain of f.

The following theorem, given in [13, p. 21 ff.], provides us with a sufficient condition for uniform convergence of the Fourier series of a function.

Theorem Appendix A.10 *([13, p. 21 ff.]). Let $h \in C^p(\mathbb{R})$ be such that $h^{(p)}$ has modulus of continuity w. Suppose that h is of period $\gamma > 0$ Then, for the partial sum $S_n h$ of the Fourier series of h,*

$$\|h - S_n h\|_\infty \leq K \frac{ln(n)}{n^p} w(\gamma/n). \tag{A.7}$$

where, $K \in \mathbb{R}$ denotes an universal constant.

Next we recall the *Wiener algebra* [15].

Definition Appendix A.11. The *Wiener algebra*, denoted by $A(\gamma\mathbb{T})$, is defined as the space of all $f \in L^2(\gamma\mathbb{T})$ with absolutely convergent Fourier series.

The following result can be found in [15].

Lemma Appendix A.12 *([15]). Let $f \in L^2(\gamma\mathbb{T})$. Suppose that one of the following conditions is fulfilled.*

(1) *$f \in C^1$.*
(2) *f belongs to a α-Hölder class for $\alpha > 1/2$.*
(3) *f is of bounded variation and belongs to a α-Hölder class for some $\alpha > 0$.*

Then, $f \in A(\gamma\mathbb{T})$. Furthermore,

$$A(\gamma\mathbb{T}) \subset C(\gamma\mathbb{T}). \tag{A.8}$$

We can now prove Theorem 5.6.

Proof of Theorem 5.6. By assumption, the induced graphon $W_{\boldsymbol{A}_i}$ satisfies $\|T_{W_{\boldsymbol{A}_i}}\| \leq \Gamma$, where $i \in \{1, 2\}$. Then, $h(T_{W_{\boldsymbol{A}_i}})$ only depends on the values of h in $[-\Gamma, \Gamma]$. Hence, we can extend h to a function $h_{\text{ext}} : [-\Gamma - 1, \Gamma + 1] \to \mathbb{R}$ such that $h_{\text{ext}}(x) = h(x)$ for all $x \in [-\Gamma, \Gamma]$ and h_{ext} admits a periodical extension on the whole real line with Lipschitz continuous derivative. By abuse of notation, we denote the periodical extension of h_{ext} by h_{ext}.

There exists a constant $L \in \mathbb{R}$ such that the modulus of continuity w for h'_{ext} is given by $w(t) = Lt$. By Theorem Appendix A.10, we have

$$|h_{\text{ext}}(x) - (S_n h_{\text{ext}})(x)| \leq \tilde{K} \frac{ln(n)}{n^2} \tag{A.9}$$

for every $n \in \mathbb{N}$, where $\tilde{K} = KL2(\Gamma + 1)$ only depends on h_{ext}. The Fourier series of h'_{ext} is given by

$$h'_{\text{ext}}(\zeta) = \sum_{n \in \mathbb{Z}} \hat{h}_{\text{ext}}(n) in/(\Gamma + 1) e^{in\zeta/(\Gamma+1)}$$

since h'_{ext} is Lipschitz. Lipschitz continuity implies bounded variation and 1-Hölder continuity. Hence, Lemma Appendix A.12 leads to

$$\sum_{n \in \mathbb{Z}} |\hat{h}_{\text{ext}}(n)||n| = C < \infty. \tag{A.10}$$

Lemma Appendix A.8 leads to

$$\|h_{\text{ext}}(T_{W_{\pi_1(\boldsymbol{A}_1)}}) - h_{\text{ext}}(T_{W_{\pi_2(\boldsymbol{\Delta}_2)}})\|_{L^2[0,1] \to L^2[0,1]} \leq C\varepsilon.$$

By $h_{\text{ext}}(0) = 0$, applying Proposition 4.4 finishes the proof. □

A.5. Proof of Claim 5.9

Proof. Let $\lambda > 0$. The Paley-Wiener space $PW_\Delta(\lambda)$ is finite-dimensional, since Δ has finitely many eigenvalues in $[-\lambda, \lambda]$ (see Example 4.11). Hence, all norms are equivalent and instead of considering the operator norm topology we consider the strong topology. For each $n \in \mathbb{N}$, we denote the equidistant partition of the unit interval into n intervals of the same length by $\mathcal{P}_n = \{I_1, I_2, \dots, I_n\}$ with $I_j = [(j-1)/n, j/n]$. Let g be in the domain of Δ and $f := P_\Delta(\lambda)g \in PW_\Delta(\lambda)$, i.e.,

$$f(x) = \sum_{k \in \mathbb{Z}:\, 4\pi^2 k^2 \leq \lambda} \langle f, e^{2i\pi k(\cdot)} \rangle e^{2i\pi kx}.$$

Hence, f is a smooth function on $[0, 1]$, so

$$|f(x) - n \int_{x-1/2n}^{x+1/2n} f(y)dy| \xrightarrow{n\to\infty} 0. \tag{A.11}$$

Let $k \in \mathbb{N}$. We have

$$\max_{i=1,\dots,n} |(\Delta^k f)(x_i) - (\Delta_n^k f_n)_i| \xrightarrow{n\to\infty} 0, \tag{A.12}$$

since the discretization of the Laplace operator with finite difference approximations on equidistant grid points is consistent (see [11, Chapter 2]). Let $h(z) = z^k \in \mathcal{H}$. Since $h(0) = 0$, Proposition 4.4 leads to $h(T_{\boldsymbol{A}_n}) = T_{nh(\boldsymbol{\Delta}_n)}$, i.e., $T^k_{\boldsymbol{A}_n} = T_{n\boldsymbol{\Delta}_n^k}$. We denote the center of the interval $I_i := [i - 1/n, i/n]$ by x_i, and calculate

$$\begin{aligned} T_{n\boldsymbol{\Delta}_n^k}(x_i) &= n \left(\int_0^1 W_{\boldsymbol{\Delta}_n^k}(x_i, v) f(v) dv \right) \\ &= n \left(\int_0^1 \sum_{g=1}^n \sum_{l=1}^n (\Delta_n^k)_{g,l} \chi_{I_g}(x_i) \chi_{I_l}(v) f(v) dv \right) \\ &= n \left(\int_0^1 \sum_{l=1}^n (\Delta_n^k)_{i,l} \chi_{I_l}(v) f(v) dv \right) = n \sum_{l=1}^n (\Delta_n^k)_{i,l} \int_{I_l} f(v) dv. \end{aligned} \tag{A.13}$$

This leads to

$$\begin{aligned} \max_{i=1,\dots,n} |T^k_{\boldsymbol{A}_n} f(x_i) - (\Delta_n^k f_n)_i| &= \max_{i=1,\dots,n} \left| n \sum_{l=1}^n (\Delta_n^k)_{i,l} \int_{I_l} f(v) dv - (\Delta_n^k)_{i,l} (f_n)_l \right| \\ &= \max_{i=1,\dots,n} \left| \sum_{l=1}^n (\Delta_n^k)_{i,l} \left(n \int_{I_l} f(v) dv - (f_n)_l \right) \right| \\ &= \max_{i=1,\dots,n} \left| \sum_{l=1}^n (\Delta_n^k)_{i,l} \left(n \int_{x_l - 1/2n}^{x_l + 1/2n} f(v) dv - f(x_l) \right) \right|. \end{aligned}$$

For the fixed $k \in \mathbb{N}$, the norms of the matrices Δ_n^k are uniformly bounded with bound independent of n. Hence, by (A.11), we get

$$\max_{i=1,\dots,n} |T_{\boldsymbol{A}_n}^k f(x_i) - (\Delta_n^k f_n)_i| \xrightarrow{n\to\infty} 0. \tag{A.14}$$

The convergence results (A.12) and (A.14) lead to

$$\begin{aligned} &\max_{i=1,\dots,n} |\left(T_{\boldsymbol{A}_n}^k f(x_i) - \Delta^k f(x_i)\right)| \\ &\leq \max_{i=1,\dots,n} |T_{\boldsymbol{A}_n}^k f(x_i) - (\Delta_n^k f_n)_i| + |(\Delta_n^k f_n)_i - \Delta^k f(x_i)| \xrightarrow{n\to\infty} 0. \end{aligned} \tag{A.15}$$

Let $\varepsilon > 0$. Since $\Delta^k f$ is smooth and $\sum_{i=1}^n \chi_{I_i}(\cdot)\Delta^k f(x_i)$ is piecewise constant on the compact set $[0,1]$, there exists $j \in \mathbb{N}$ and $x_n' \in I_j = [x_j - 1/2n, x_j + 1/2n]$ such that

$$\|\sum_{i=1}^n \chi_{I_i}(\cdot)\Delta^k f(x_i) - \Delta^k f\|_{L^\infty([0,1])} = |\Delta^k f(x_j) - \Delta^k f(x_n')|.$$

Since $\Delta^k f$ is Lipschitz continuous, it follows

$$|\Delta^k f(x_j) - \Delta^k f(x_n')| \leq C|x_{j_n} - x_n'| \xrightarrow{n\to\infty} 0.$$

Hence, we can choose $n \in \mathbb{N}$ large enough such that

$$\|\sum_{i=1}^n \chi_{I_i}(\cdot)\Delta^k f(x_i) - \Delta^k f\|_{L^2[0,1]} < \varepsilon, \tag{A.16}$$

and together with (A.15)

$$\max_{i=1,\dots,n} |T_{\boldsymbol{A}_n}^k f(x_i) - \Delta^k f(x_i)| < \varepsilon. \tag{A.17}$$

As a result,

$$\begin{aligned} \|T_{\boldsymbol{A}_n}^k f - \Delta^k f\|_{L^2[0,1]} &= \|\sum_{i=1}^n \chi_{I_i}(\cdot)T_{\boldsymbol{A}_n}^k f(x_i) - \Delta^k f\|_{L^2[0,1]} \\ &\leq \|\sum_{i=1}^n \chi_{I_i}(\cdot)T_{\boldsymbol{A}_n}^k f(x_i) - \sum_{i=1}^n \chi_{I_i}(\cdot)\Delta^k f(x_i)\|_{L^2[0,1]} \\ &+ \|\sum_{i=1}^n \chi_{I_i}(\cdot)\Delta^k f(x_i) - \Delta^k f\|_{L^2[0,1]} \\ &\leq \|\sum_{i=1}^n \chi_{I_i}(\cdot)T_{\boldsymbol{A}_n}^k f(x_i) - \sum_{i=1}^n \chi_{I_i}(\cdot)\Delta^k f(x_i)\|_{L^\infty([0,1])} \\ &+ \|\sum_{i=1}^n \chi_{I_i}(\cdot)\Delta^k f(x_i) - \Delta^k f\|_{L^2[0,1]} \\ &= \max_{i=1,\dots,n} |T_{\boldsymbol{A}_n}^k f(x_i) - \Delta^k f(x_i)| \\ &+ \|\sum_{i=1}^n \chi_{I_i}(\cdot)\Delta^k f(x_i) - \Delta^k P_{\mathcal{L}}(\lambda) f\|_{L^2[0,1]} < 2\varepsilon, \end{aligned}$$

i.e.,

$$\|P_\Delta(\lambda)T^k_{\boldsymbol{A}_n}P_\Delta(\lambda)g-\Delta^k P_\Delta(\lambda)g\|_{L^2[0,1]}\leq\|T^k_{\boldsymbol{A}_n}P_\Delta(\lambda)g-\Delta^k P_\Delta(\lambda)g\|_{L^2[0,1]}\xrightarrow{n\to\infty}0.$$

Since $(G_n)_n\xrightarrow[n\to\infty]{U}\Delta$, Lemma 5.7 leads to

$$\begin{aligned}&\|P_\Delta(\lambda)T^k_{\boldsymbol{A}_n}P_\Delta(\lambda)g-\Big(P_\Delta(\lambda)T_{\boldsymbol{A}_n}P_\Delta(\lambda)\Big)^k g\|\\&\leq\|P_\Delta(\lambda)T^k_{\boldsymbol{A}_n}P_\Delta(\lambda)g-\Delta^k P_\Delta(\lambda)g\|_{L^2[0,1]}\\&+\|\Delta^k P_\Delta(\lambda)g-\Big(P_\Delta(\lambda)T_{\boldsymbol{A}_n}P_\Delta(\lambda)\Big)^k g\|_{L^2[0,1]}\xrightarrow{n\to\infty}0,\end{aligned}$$

finishing the proof. □

A.6. Proof of Lemma 5.13

For $l=1$ and $j=1,\dots,F_1$, we have

$$\begin{aligned}&\|(\tilde{\phi}^1_{W_{\pi_n(\boldsymbol{A}_n)}}(y_n))^j-(\tilde{\phi}^1_W(y))^j\|_{L^2[0,1]}\\&=\|\sum_{k=1}^{F_0}m_1^{jk}h_1^{jk}(T_{W_{\pi_n(\boldsymbol{A}_n)}})(y_n^k)-\sum_{k=1}^{F_0}m_1^{jk}h_1^{jk}(T_W)(y^k)\|_{L^2[0,1]}\\&\leq\sum_{k=1}^{F_0}\|m_1^{jk}h_1^{jk}(T_{W_{\pi_n(\boldsymbol{A}_n)}})(y_n^k)-m_1^{jk}h_1^{jk}(T_W)(y_n^k)\|_{L^2[0,1]}\\&+\sum_{k=1}^{F_0}\|m_1^{jk}h_1^{jk}(T_W)(y_n^k)-m_1^{jk}h_1^{jk}(T_W)(y^k)\|_{L^2[0,1]}\\&\leq\sum_{k=1}^{F_0}|m_1^{jk}|\|h_1^{jk}(T_{W_{\pi_n(\boldsymbol{A}_n)}})(y_n^k)-h_1^{jk}(T_W)(y_n^k)\|_{L^2[0,1]}\\&+\sum_{k=1}^{F_0}|m_1^{jk}|\|h_1^{jk}(T_W)(y_n^k)-h_1^{jk}(T_W)(y^k)\|_{L^2[0,1]}\end{aligned}\tag{A.18}$$

By Lemma 5.1 and by (24), every summand in the-right-hand-side of (A.18) converges to zero as $n\to\infty$. Since $\rho:\mathbb{R}\to\mathbb{R}$ is Lipschitz continuous, we have

$$\begin{aligned}&\int_0^1|\rho((\tilde{\phi}^1_{W_{\pi_n(\boldsymbol{A}_n)}}(y_n))^j(x))-\rho((\tilde{\phi}^1_W(y))^j(x))|^2dx\\&\leq\Lambda^2\int_0^1|(\tilde{\phi}^1_{W_{\pi_n(\boldsymbol{A}_n)}}(y_n))^j-(\tilde{\phi}^1_W(y))^j|^2dx,\end{aligned}$$

where $\Lambda\in\mathbb{R}$ is the Lipschitz constant of ρ. Hence, for every feature $j=1,\dots,F_1$, we have $\|(\phi^1_{W_{\pi_n(\boldsymbol{A}_n)}}(y_n))^j-(\phi^1_W(y))^j\|_{L^2[0,1]}\xrightarrow{n\to\infty}0$. By induction over the layers, we have

$$\|(\phi^L_{W_{\pi_n(\boldsymbol{A}_n)}}(y_n))^j-(\phi^L_W(\tilde{\phi}^{L-1}_W(y)))^j\|_{L^2[0,1]}\xrightarrow{n\to\infty}0$$

for all features $j\in\{1,\dots,F_L\}$.

A.7. Proof of Lemma 5.17

For every $l = 1, \dots, L$, let C_l^{jk} denote the constant C of (21) corresponding to the filter $h = h_l^{jk}$, for $j = 1, \dots, F_l$, $k = 1, \dots, F_{l-1}$. Let $C = \max_{l,j,k} C_l^{jk}$, where the maximum is over $l = 1, \dots, L$, $j = 1, \dots, F_l$, and $k = 1, \dots, F_{l-1}$. Further, let $M = \max_{l=1,\dots,L} \|\mathbf{M}_l\|_\infty$.

For $j = 1, \dots, F_L$, we have

$$\begin{aligned}
&\|(\tilde{\phi}^L_{W_{\pi(\boldsymbol{A})}}(\psi_{\mathbf{x}}))^j - (\tilde{\phi}^L_W(y))^j\|_{L^2[0,1]} \\
&\leq \sum_{k=1}^{F_{L-1}} \|m_L^{jk} h_L^{jk}(T_{W_{\pi(\boldsymbol{A})}})(\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k) - m_L^{jk} h_L^{jk}(T_W)(\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k)\|_{L^2[0,1]} \\
&+ \sum_{k=1}^{F_{L-1}} \|m_L^{jk} h_L^{jk}(T_W)(\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k) - m_L^{jk} h_L^{jk}(T_W)(\phi^{L-1}_{W}(y)^k)\|_{L^2[0,1]}.
\end{aligned} \tag{A.19}$$

For the first summand on the right hand side of (A.19), we have

$$\begin{aligned}
&\sum_{k=1}^{F_{L-1}} \|m_L^{jk} h_L^{jk}(T_{W_{\pi(\boldsymbol{A})}})(\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k) - m_L^{jk} h_L^{jk}(T_W)(\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k)\|_{L^2[0,1]} \\
&\leq \sum_{k=1}^{F_{L-1}} |m_L^{jk}| \|h_L^{jk}(T_{W_{\pi(\boldsymbol{A})}}) - h_L^{jk}(T_W)\| \|\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k\|_{L^2[0,1]} \\
&\leq C\varepsilon \sum_{k=1}^{F_{L-1}} |m_L^{jk}| \|\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k\|_{L^2[0,1]} \leq \varepsilon C M \max_{k=1,\dots,F_{L-1}} \|\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k\|_{L^2[0,1]}.
\end{aligned} \tag{A.20}$$

For the second summand in (A.19), we calculate

$$\begin{aligned}
&\sum_{k=1}^{F_{L-1}} \|m_L^{jk} h_L^{jk}(T_W)(\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k) - m_L^{jk} h_L^{jk}(T_W)(\phi^{L-1}_{W}(y)^k)\|_{L^2[0,1]} \\
&\leq \sum_{k=1}^{F_{L-1}} \|m_L^{jk} h_L^{jk}(T_W)\| \|\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k - \phi^{L-1}_{W}(y)^k\|_{L^2[0,1]} \\
&\leq \sum_{k=1}^{F_{L-1}} |m_L^{jk}| \|h_L^{jk}\|_\infty \|\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k - \phi^{L-1}_{W}(y)^k\|_{L^2[0,1]} \leq M \max_k \mathcal{R}^k_{L-1},
\end{aligned} \tag{A.21}$$

where $\mathcal{R}^k_{L-1} := \|\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k - \phi^{L-1}_{W}(y)^k\|_{L^2[0,1]}$. Together, (A.21) and (A.20) lead to

$$\begin{aligned}
&\|(\tilde{\phi}^L_{W_{\pi(\pi(\boldsymbol{A}))}}(\psi_{\mathbf{x}}))^j - (\tilde{\phi}^L_W(y))^j\|_{L^2[0,1]} \\
&\leq M \max_{k=1,\dots,F_{L-1}} \mathcal{R}^k_{L-1} + C\varepsilon M \max_{k=1,\dots,F_{L-1}} \|\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k\|_{L^2[0,1]}
\end{aligned}$$

for every $j = 1, \dots, F_L$. Then,

$$\begin{aligned}
&\|(\phi^L_{W_{\pi(\pi(\boldsymbol{A}))}})^j(\psi_{\mathbf{x}}) - (\phi^L_W)^j(y)\|_{L^2[0,1]} \\
&\leq M \max_{k=1,\dots,F_{L-1}} \mathcal{R}^k_{L-1} + C\varepsilon M \max_{k=1,\dots,F_{L-1}} \|\phi^{L-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k\|_{L^2[0,1]},
\end{aligned} \tag{A.22}$$

since the activation function is contractive. We solve the recurrence relation in (A.22). For this, we estimate $\max_{k=1,\dots,F_l} \|\phi^l_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k\|_{L^2[0,1]}$ for $l = 1, \dots, L$. Now, for $k = 1, \dots, F_l$,

$$\begin{aligned}\|\phi^l_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k\|_{L^2[0,1]} &= \sum_{k'=1}^{F_{l-1}} \|m_l^{jk'} h_l^{jk'} (T_{W_{\pi(\boldsymbol{A})}})(\phi^{l-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^{k'})\|_{L^2[0,1]} \\ &\leq M \max_{k'=1,\dots,F_{l-1}} \|\phi^{l-1}_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^{k'}\|_{L^2[0,1]}.\end{aligned}$$

Hence,

$$\|\phi^L_{W_{\boldsymbol{A}}}(\psi_{\mathbf{x}})^k\|_{L^2[0,1]} \leq M^L \max_{k'=1,\dots,F_{l-1}} \|\psi^{k'}_{\mathbf{x}}\|_{L^2[0,1]}.$$

This leads to

$$\|(\phi^L_{W_{\pi(\pi(\boldsymbol{A}))}}(\psi_{\mathbf{x}}))^j - (\phi^L_W(y))^j\|_{L^2[0,1]} \leq M^L(1 + LC) \max_{k=1,\dots,F_0} \|\psi^k_{\mathbf{x}} - y^k\|_{L^2[0,1]}.$$

References

[1] A.B. Aleksandrov, V.V. Peller, Operator Hölder–Zygmund functions, Adv. Math. 224 (2010) 910–966.
[2] A.B. Aleksandrov, V.V. Peller, Operator Lipschitz functions, Russ. Math. Surv. 71 (2016) 605–702.
[3] F.M. Bianchi, D. Grattarola, L. Livi, C. Alippi, Graph neural networks with convolutional ARMA filters, IEEE Trans. Pattern Anal. Mach. Intell. (2021) 1.
[4] A. Bojchevski, S. Günnemann, Certifiable robustness to graph perturbations, in: Advances in Neural Information Processing Systems, Curran Associates Inc., 2019.
[5] A. Bojchevski, J. Klicpera, S. Günnemann, Efficient robustness certificates for discrete data: sparsity-aware randomized smoothing for graphs, images and more, in: Proceedings of the 37th International Conference on Machine Learning, PMLR, 2020.
[6] C. Borgs, J.T. Chayes, L. Lovasz, V.T. Sós, K. Vesztergombi, Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing, Adv. Math. 219 (2008) 1801–1851.
[7] C. Borgs, J.T. Chayes, L. Lovasz, V.T. Sós, K. Vesztergombi, Convergent sequences of dense graphs II. Multiway cuts and statistical physics, Ann. Math. 176 (2012) 151–219.
[8] M.M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, P. Vandergheynst, Geometric deep learning: going beyond euclidean data, IEEE Signal Process. Mag. 34 (2017) 18–42.
[9] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, Curran Associates Inc., 2016.
[10] F. Gama, J. Bruna, A. Ribeiro, Stability properties of graph neural networks, IEEE Trans. Signal Process. 68 (2020) 5680–5695.
[11] C. Grossmann, H.-G. Roos, M. Stynes, Numerical Treatment of Partial Differential Equations, Revised translation of the 3rd German edition of 'Numerische Behandlung partieller Differentialgleichungen' by Martin Stynes, Vieweg+Teubner Verlag, 01 2007.
[12] W.L. Hamilton, Graph representation learning, in: Synthesis Lectures on Artifical Intelligence and Machine Learning, vol. 14, 2020, pp. 1–159.
[13] D. Jackson, The Theory of Approximation, American Mathematical Society, New York, 1931.
[14] S. Janson, Graphons, Cut Norm and Distance, Couplings and Rearrangements, 2010, arXiv: Combinatorics.
[15] Y. Katznelson, An Introduction to Harmonic Analysis, Cambridge Mathematical Library, 3 ed., Cambridge University Press, 2004.
[16] H. Kenlay, D. Thano, X. Dong, On the stability of graph convolutional neural networks under edge rewiring, in: International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021.
[17] H. Kenlay, D. Thanou, X. Dong, On the stability of polynomial spectral graph filters, in: International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020.
[18] H. Kenlay, D. Thanou, X. Dong, Interpretable stability bounds for spectral graph filters, in: Proceedings of the 38th International Conference on Machine Learning, PMLR, 2021.
[19] N. Keriven, A. Bietti, S. Vaiter, Convergence and stability of graph convolutional networks on large random graphs, arXiv:2006.01868, 2020.
[20] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: Proceedings of the 5th International Conference on Learning Representations, ICLR, 2017.
[21] O. Klopp, N. Verzelen, Optimal graphon estimation in cut distance, Probab. Theory Relat. Fields 174 (2019) 1033–1090.
[22] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (1998) 2278–2324.

[23] R. Levie, W. Huang, L. Bucci, M.M. Bronstein, G. Kutyniok, Transferability of spectral graph convolutional neural networks, arXiv:1907.12972, 2019.
[24] R. Levie, E. Isufi, G. Kutyniok, On the transferability of spectral graph filters, in: 13th International Conference on Sampling Theory and Applications (SampTA), IEEE, 2019.
[25] R. Levie, F. Monti, X. Bresson, M.M. Bronstein, Cayleynets: graph convolutional neural networks with complex rational spectral filters, IEEE Trans. Signal Process. 67 (2019) 97–109.
[26] L. Lovász, B. Szegedy, Limits of dense graph sequences, J. Comb. Theory, Ser. B 96 (2006) 933–957.
[27] L. Lovász, Large Networks and Graph Limits, Colloquium Publications, Budapest, 2012.
[28] M.W. Morency, G. Leus, Graphon filters: graph signal processing in the limit, IEEE Trans. Signal Process. 69 (2021) 1740–1754.
[29] D. Potapov, F. Sukochev, Operator-Lipschitz functions in Schatten-von Neumann classes, Acta Math. 207 (2011) 375–389.
[30] L. Ruiz, L.F.O. Chamon, A. Ribeiro, Graphon signal processing, arXiv:2003.05030, 2020.
[31] L. Ruiz, L.F.O. Chamon, A. Ribeiro, Graphon filters: signal processing in very large graphs, in: 2020 28th European Signal Processing Conference (EUSIPCO), 2021, pp. 1050–1054.
[32] L. Ruiz, F. Gama, A. Ribeiro, Graph neural networks: architectures, stability, and transferability, Proc. IEEE 109 (2021) 660–682.
[33] L. Ruiz, Z. Wang, A. Ribeiro, Graphon and graph neural network stability, in: International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021.
[34] D. Werner, Funktionalanalysis, Springer-Verlag, Berlin, Heidelberg, 01 2007.
[35] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P.S. Yu, A comprehensive survey on graph neural networks, IEEE Trans. Neural Netw. Learn. Syst. 32 (2021) 4–24.