Graphon Filters:
Signal Processing in Very Large Graphs

Luana Ruiz, Luiz F. O. Chamon and Alejandro Ribeiro

***Abstract*—Graph filters are at the core of network information processing architectures, with applications in machine learning and distributed collaborative systems. Yet, designing filters for very large graphs is challenging because filter design techniques do not always scale with graph size. We overcome this by using graphons, which are infinite-dimensional representations of graphs that are at once random graph models and limit objects of sequences of graphs. Explicitly, we define graphon filters and leverage convergence properties of graph sequences and spectral properties of both graphs and graphons to show that graph filters converge to graphon filters. Filters designed on a graphon can therefore be applied to finite graphs sampled from it with guaranteed convergence properties. We illustrate our findings in two experiments, which corroborate filter response convergence and illustrate transferability of graph filters even in graphs that are not related to one another through a graphon, but that are built from the same type of data.**

***Index Terms*—graphons, convergent graph sequences, graph filters, graph signal processing**

I. Introduction

Graph filters are central components of many network information processing tools arising from graph signal processing (GSP), e.g., distributed collaborative filtering [1], and are key building blocks in machine learning models used for network data, e.g., graph convolutional neural networks [2], [3]. Yet, designing graph filters has become increasingly challenging due to the scale of most real-world networks and their often dynamic nature. This is the case, for instance, in Internet of Things (IoT) applications, in which the number of interconnected devices can reach the order of billions, and where connections are typically intermittent [4].

The first issue with large-scale networks is the difficulty of even acquiring measurements from them. Often, only a subset of nodes can be sampled at a time. However, many applications involving both linear graph filters [5], [6] and nonlinear architectures based on them (e.g., convolutional graph neural networks [2]) depend on the full graph matrix representation. Even if these networks could be efficiently measured, storing them can be challenging, especially if they are not sparse. While more compact representations can be achieved by using low-rank approximations [7], obtaining these representations can be difficult as they involve solving large eigenvalue problems [8], [9]. These issues are exacerbated in situations where the graph dynamically grows or changes, since, unless these changes involve structured node relabeling [3] or small perturbations aligned with the original graph eigenspace [10], these computations need to be performed repeatedly. Otherwise, it is not straightforward to know the effect of the filter designed for the original graph on the modified graph.

We propose to mitigate these issues by studying "graph families" identified by infinite-dimensional objects called *graphons* [11], [12]. Graphons are bounded symmetric kernels that can be interpreted as both random graph models and limit objects of sequences of graphs. The first interpretation is more common and has been extensively used to perform random graph model estimation [13]–[15]. The second, more relevant to our problem, facilitates the study of properties such as node centrality [16], network game equilibria [17], and control of linear systems [18] on very large networks.

Explicitly, we define the concept of a *graphon signal* as a specific limit of a sequence of graph signals in which both the signal and the underlying graph grow. We then propose a definition of *linear graphon filters* (Definition 2) and show that, under mild conditions, both the frequency response and the vertex-domain output of linear graph filters converge to them. This result has a two-sided implication. On the one hand, it shows that filtering a sequence of graph signals converging to a graphon signal is equivalent to filtering the graphon signal itself. On the other hand, it implies that we can design filters that will be used on large or dynamic graph signals directly on the graphon from which they arise, reducing computation costs and increasing the filter robustness to perturbations. This raises an interesting parallel with the design of digital filters in classical signal processing, where it is usually easier to think in terms of their continuous-time counterparts.

Our findings are illustrated in two numerical experiments. In the first, we observe filter response convergence by comparing the outcomes of a graphon signal diffusion and of a graph signal diffusion on a sequence of sensor networks sampled from the graphon model. This toy example can be interpreted as a one-hop communication between sensors in an IoT network. In the second, we use real movie ratings to train a linear filter on a user network built from a small cohort of users, which is then used to predict movie ratings in a larger user network. With this experiment, we aim to illustrate a more practical application of our results—transfer learning—on graphs that are not necessarily related by any common generating graphon, but that are known to be similar in an empirical or statistical sense (e.g., built from data with the same underlying distribution).

II. Preliminary definitions

A. From graphs to graphons

We represent undirected graphs as $\mathbf{G} = (\mathcal{V}, \mathcal{E}, A)$, where $\mathcal{V}$ is a set of n nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges

Supported by NSF CCF 1717120, ARO W911NF1710438, ARL DCIST CRA W911NF-17-2-0181, ISTC-WAS and Intel DevCloud. The authors are with the Dept. of Electrical and Systems Eng., Univ. of Pennsylvania. Email: {rubruiz,luizf,aribeiro}@seas.upenn.edu.

and $A : \mathcal{E} \to \mathcal{I} \subseteq \mathbb{R}$ is a weight function assigning real-valued weights $A(i,j) = A(j,i)$ in the set $\mathcal{I}$ to edges $(i,j) \in \mathcal{E}$. If $\mathcal{I} = \{0,1\}$, the graph is unweighted and we omit $\mathcal{A}$ to write $\mathbf{G} = (\mathcal{V}, \mathcal{E}, A) = (\mathcal{V}, \mathcal{E})$. Graphons are bounded, symmetric, measurable functions $\mathbf{W} : [0,1]^2 \to [0,1]$. If we consider points u and v of the unit line to be nodes, $\mathbf{W}(u,v)$ can be seen as the weight of the edge connecting u and v. Another way to see $\mathbf{W}$ is as the abstraction of a graph with an uncountable number of nodes to which sequences of graphs $\{\mathbf{G}_n\}_{n=1}^{\infty}$ converge in a sense that we formalize next.

To study the convergence of $\{\mathbf{G}_n\}$, consider arbitrary graphs $\mathbf{F} = (\mathcal{V}', \mathcal{E}')$ that are unweighted and undirected, or *simple* for short. We can think of $\mathbf{F}$ as a graph motif. Homomorphisms of $\mathbf{F}$ into $\mathbf{G}$ are adjacency preserving maps $\beta : \mathcal{V}' \to \mathcal{V}$ in which $(i,j) \in \mathcal{E}'$ implies $(i,j) \in \mathcal{E}$. While there are $|\mathcal{V}|^{|\mathcal{V}'|} = n^{n'}$ maps from $\mathcal{V}'$ to $\mathcal{V}$, only some of them are homomorphisms. Denoting by $\hom(\mathbf{F}, \mathbf{G}) = \sum_{\beta} \prod_{(i,j)\in\mathcal{E}'} A(\beta(i), \beta(j))$ the weighted sum of the total number of homomorphisms that map $\mathbf{F}$ into $\mathbf{G}$, we can therefore define the density of homomorphisms as [11]

$$t(\mathbf{F}, \mathbf{G}) := \frac{\hom(\mathbf{F}, \mathbf{G})}{n^{n'}} := \frac{\sum_{\beta} \prod_{(i,j)\in\mathcal{E}'} A(\beta(i), \beta(j))}{n^{n'}}. \quad (1)$$

When $\mathbf{G}$ is unweighted, $t(\mathbf{F}, \mathbf{G})$ has a straightforward interpretation because $\prod_{(i,j)\in\mathcal{E}'} A(\beta(i), \beta(j)) = 1$ counts the number of ways in which the motif graph $\mathbf{F}$ can be mapped into $\mathbf{G}$. The value of $t(\mathbf{F}, \mathbf{G})$ quantifies how often this motif appears in $\mathbf{G}$ relative to the maximum number of times it could appear.

Homomorphism densities can be generalized to graphons by replacing the sum with an integral. We define the homomorphism density of the graph $\mathbf{F}$ into the graphon $\mathbf{W}$ as

$$t(\mathbf{F}, \mathbf{W}) := \int_{[0,1]^{\mathcal{V}'}} \prod_{(i,j)\in\mathcal{E}'} \mathbf{W}(u_i, u_j) \prod_{i\in\mathcal{V}'} du_i. \quad (2)$$

Using 1 and 2, we say that a sequence $\{\mathbf{G}_n\}$ converges to the graphon $\mathbf{W}$ if, for all finite motifs $\mathbf{F}$, it holds that

$$\lim_{n\to\infty} t(\mathbf{F}, \mathbf{G}_n) = t(\mathbf{F}, \mathbf{W}). \quad (3)$$

Interpreting (3) is easiest when we consider graphs $\mathbf{G}_n$ sampled from the graphon $\mathbf{W}$, which are called $\mathbf{W}$-*random graphs*. $\mathbf{W}$-random graphs have labels $u_i \sim U[0,1]$ drawn i.i.d. from $[0,1]$, and their edges are Bernoulli random variables such that $(u_i, u_j) \in \mathcal{E}$ with probability $\mathbf{W}(u_i, u_j)$. It can be shown that (3) holds with probability 1 for sequences of $\mathbf{W}$-random graphs [11]. Naturally, the convergence mode in (3) also allows for other, more general graph sequences than those consisting of $\mathbf{W}$-random graphs. As a matter of fact, every graphon is the limit object of a convergent graph sequence and, conversely, every convergent graph sequence converges to a graphon [11]. Thus, if the limit $\lim_{n\to\infty} t(\mathbf{F}, \mathbf{G}_n)$ exists, it can be written as (2) for some limit graphon $\mathbf{W}$.

B. Graph signals and graph filters

Graph signals are pairs $(\mathbf{G}, \mathbf{x})$ where $\mathbf{G}$ is a graph and $\mathbf{x} \in \mathbb{R}^n$ is a vector mapping signal values to nodes. In graph signal processing, $\mathbf{G}$ is often represented by the graph shift operator (GSO) $\mathbf{S} \in \mathbb{R}^{n\times n}$, which satisfies $[\mathbf{S}]_{ij} \neq 0$ if and only if (i,j) is an edge or if $i = j$. The GSO gets its name from the fact that, when applied to the signal $(\mathbf{G}, \mathbf{x})$, it "shifts" (or diffuses) the information contained in $[\mathbf{x}]_j$ to every 1-hop neighbor of j. To see this, notice that in $[\mathbf{Sx}]_i = \sum_{j=1}^{n} [\mathbf{S}]_{ij} x_j$, $[\mathbf{S}]_{ij}$ is nonzero only if $i = j$ or if j is in the neighborhood of i.

The GSO induces a larger class of operations called linear shift-invariant graph filters (LSI-GFs), which are defined as

$$\mathbf{H}(\mathbf{S})\mathbf{x} = \sum_{k=0}^{K} h_k \mathbf{S}^k \mathbf{x} \ . \quad (4)$$

The weights $h_0, \dots, h_k$ are the graph filter coefficients. For undirected graphs, $\mathbf{S}$ is symmetric and diagonalizable as $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathsf{H}}$, where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues and the columns of $\mathbf{V}$ are the graph eigenvectors. Because the eigenvectors of $\mathbf{S}$ form an orthonormal basis of $\mathbb{R}^n$—the *graph spectral basis*—a change of basis can always be defined as $\hat{\mathbf{x}} = \mathbf{V}^{\mathsf{H}}\mathbf{x}$. In particular, replacing $\mathbf{S}$ by its diagonalization in (4) and calculating $\hat{\mathbf{y}} = \mathbf{V}^{\mathsf{H}}\mathbf{H}(\mathbf{S})\mathbf{x}$, we get

$$\hat{\mathbf{y}} = \sum_{k=0}^{K} h_k \mathbf{\Lambda}^k \hat{\mathbf{x}} = \hat{\mathbf{H}}(\mathbf{\Lambda})\hat{\mathbf{x}} \quad (5)$$

from which we conclude that $\mathbf{H}(\mathbf{S})$ has a spectral representation $\hat{\mathbf{H}}(\mathbf{\Lambda}) = \sum_{k=0}^{K} h_k \mathbf{\Lambda}^k$ that is polynomial on the eigenvalues of the GSO. This is an important result, as polynomials can be used to approximate functions in C^{∞} arbitrarily well through their Taylor series expansion. In particular, any filter with frequency response $\hat{\mathbf{H}}(\lambda) = h(\lambda)$ can be written as a LSI-GF around values of λ where h is infinitely differentiable.

III. Graphon signals and graphon filters

A graphon signal is a pair $(\mathbf{W}, \phi)$ where $\mathbf{W}$ is a graphon and $\phi : [0,1] \to \mathbb{R}$ is a function mapping points of the unit interval (i.e., the "graphon nodes") to signal values. We focus on finite-energy graphon signals, i.e., $\phi \in L_2([0,1])$. Any n-node graph signal $(\mathbf{G}, \mathbf{x})$ induces a graphon signal $(\mathbf{W}_{\mathbf{G}}, \phi_{\mathbf{G}})$

$$\mathbf{W}_{\mathbf{G}}(u,v) = [\mathbf{S}]_{ij} \times \mathbb{I}(u \in I_i)\mathbb{I}(v \in I_j) \quad (6)$$

$$\phi_{\mathbf{G}}(v) = [\mathbf{x}]_j \times \mathbb{I}(v \in I_j) \quad (7)$$

where $\{I_j\}_{j=1}^{n}$ is a partition of the unit interval.

Given that there are sequences of graphs that converge to graphons, we can analogously define sequences of graph signals that converge to graphon signals.

Definition 1 (Convergent sequences of graph signals)**.** A sequence of graph signals $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ is said to converge to the graphon signal $(\mathbf{W}, \phi)$ if there exists a sequence of permutations $\{\pi_n\}$ such that, for all finite, simple $\mathbf{F}$,

$$\lim_{n\to\infty} t(\mathbf{F}, \mathbf{G}_n) \to t(\mathbf{F}, \mathbf{W}) \quad (8)$$

i.e., $\mathbf{G}_n \to \mathbf{W}$ in the homomorphism density sense, and

$$\lim_{n\to\infty} \|\phi_{\pi_n(\mathbf{G}_n)} - \phi\|_{L_2} = 0 \quad (9)$$

where $(\mathbf{W}_{\pi_n(\mathbf{G}_n)}, \phi_{\pi_n(\mathbf{G}_n)})$ is the graphon signal induced by the permuted graph signal $(\pi_n(\mathbf{G}_n), \pi_n(\mathbf{x}_n))$ [cf. 6].

A sequence of graph signals thus converges when the graphs converge and when the step functions induced by $\mathbf{x}_n$ converge to a function in L_2 (up to permutations).

Every graphon $\mathbf{W}$ defines an integral operator $T_{\mathbf{W}} : L_2([0,1]) \to L_2([0,1])$ as

$$(T_{\mathbf{W}}\phi)(v) := \int_0^1 \mathbf{W}(u,v)\phi(u)du \; . \tag{10}$$

Since graphons are bounded and symmetric, $T_{\mathbf{W}}$ is a self-adjoint Hilbert-Schmidt operator. It can thus be decomposed in the operator's basis as

$$\mathbf{W}(u,v) = \sum_{i=0}^{\infty} \lambda_i \varphi_i(u)\varphi_i(v) \tag{11}$$

where $\{\lambda_i\}_{i=1}^{\infty}$, $\lambda_i \in [-1,1]$ are the graphon eigenvalues and $\{\varphi_i\}_{i=1}^{\infty}$, $\varphi_i : [0,1] \to \mathbb{R}$ are its eigenfunctions, which constitute an orthonormal basis of $L_2([0,1])$ called the *graphon spectral basis*. Positive and negative eigenvalues are separated by reordering λ_j with indices $j \in \mathbb{Z} \setminus \{0\}$ according to their sign and in decreasing order of absolute value, that is, $1 \geq \lambda_1 \geq \lambda_2 \geq \ldots \geq 0 \geq \ldots \geq \lambda_{-2} \geq \lambda_{-1} \geq -1$. The eigenvalues and eigenfunctions are also countable, with λ_j converging to 0 as $|j| \to \infty$. Zero is the only point of accumulation and so every $\lambda_j \neq 0$ has finite multiplicity [19].

Using (11), we can write $T_{\mathbf{W}}$ as

$$(T_{\mathbf{W}}\phi)(v) = \sum_{j=0}^{\infty} \lambda_j \varphi_j(v) \int_0^1 \varphi_j(u)\phi(u)du \tag{12}$$

$$= \sum_{j=0}^{\infty} \lambda_j \varphi_j(v) \hat{\phi}(\lambda_j), \tag{13}$$

$$\text{where } \hat{\phi}(\lambda_j) = \int_0^1 \varphi_j(u)\phi(u)du \; . \tag{14}$$

The integrals $\int_0^1 \varphi_j(u)\phi(u)du$ effectively project the signal $(\mathbf{W},\phi)$ onto the graphon spectral basis. We write the jth component of this projection as $\hat{\phi}(\lambda_j)$.

Similarly to how the GSO diffuses graph signals to the 1-hop neighborhood of each node, we can think of $T_{\mathbf{W}}$ as inducing a 1-step "diffusion" of ϕ on the graphon. Building upon this parallel, we refer to $T_{\mathbf{W}}$ as the graphon shift operator (WSO) and use it to define *linear shift-invariant graphon filters* as follows.

Definition 2 (Linear shift-invariant graphon filters)**.** Let $(\mathbf{W},\phi)$ be a graphon signal. A linear shift-invariant graphon filter $T_{\mathbf{H}} : L_2([0,1]) \to L_2([0,1])$ maps $(\mathbf{W},\phi)$ to $(\mathbf{W},\gamma)$,

$$\gamma(v) = (T_{\mathbf{H}}\phi)(v) = \sum_{k=0}^{K} h_k T_{\mathbf{W}}^{(k)}\phi(v) \tag{15}$$

$$(T_{\mathbf{W}}^{(k)}\phi)(v) = \int_0^1 \mathbf{W}(u,v)(T_{\mathbf{W}}^{(k-1)}\phi)(u)du \tag{16}$$

for $T_{\mathbf{W}}^{(1)} := T_{\mathbf{W}}$ and $T_{\mathbf{W}}^{(0)} := \mathbb{I}$, the identity operator. The h_k are the filter coefficients or taps.

The spectral decomposition of $T_{\mathbf{W}}$ in (11) allows us to write $T_{\mathbf{H}}$ as

$$(T_{\mathbf{H}}\phi)(v) = \sum_{j=1}^{\infty}\sum_{k=0}^{K} h_k \lambda_j^k \hat{\phi}(\lambda_j)\varphi_j(v) \tag{17}$$

where we note that, analogously to LSI-GFs, $T_{\mathbf{H}}$ admits a spectral representation on the graphon basis given by

$$\hat{T}_{\mathbf{H}}(\lambda_j) = \sum_{k=0}^{K} h_k \lambda_j^k \text{ for } j \in \mathbb{Z} \setminus \{0\} \; . \tag{18}$$

Comparing expressions (5) and (18), we observe that they are they are finite and infinite counterparts of each other. Because the frequency response of the linear shift-invariant graphon filter is also polynomial in the graphon eigenvalues, they too can approximate any smooth filter response arbitrarily well. Formally, a graphon filter with frequency response $\hat{T}_{\mathbf{H}}(\lambda) = h(\lambda)$ can be written as the filter in Definition 2 with $K \to \infty$, provided that $h(\lambda)$ is infinitely differentiable at $\{\lambda_j\}_{j \in \mathbb{Z}\setminus\{0\}}$.

IV. Convergence of graph filter responses

Herein we show that the spectral and vertex responses of graph filters with filter function $h(\lambda)$ converge to the respective filter responses of the graphon filter with same filter function $h(\lambda)$. We start by showing this convergence in the spectral domain (Theorem 1). This result follows directly from an eigenvalue convergence result proved in [20][Theorem 6.7] and reproduced here as Lemma 1.

Lemma 1 (Eigenvalue convergence [20][Theorem 6.7)**.** Let $\{\mathbf{G}_n\}$ be a sequence of graphs with eigenvalues $\{\lambda_j(\mathbf{S}_n)\}_{j \in \mathbb{Z}\setminus\{0\}}$, and $\mathbf{W}$ a graphon with eigenvalues $\{\lambda_j(T_{\mathbf{W}})\}_{j \in \mathbb{Z}\setminus\{0\}}$. Assume that in both cases the eigenvalues are ordered by decreasing order of absolute value and indexed according to their sign. If $\{\mathbf{G}_n\}$ converges to $\mathbf{W}$, then

$$\lim_{n\to\infty} \frac{\lambda_j(\mathbf{S}_n)}{n} = \lambda_j(T_{\mathbf{W}}) \text{ for all } j \; . \tag{19}$$

Theorem 1 (Convergence of graph filter spectral response)**.** For the graph sequence $\{\mathbf{G}_n\}$, let $\mathbf{H}_n(\mathbf{S}_n)$ be a sequence of filters of the form $\mathbf{H}_n(\mathbf{S}_n) = \mathbf{V}_n h(\mathbf{\Lambda}_n(\mathbf{S}_n)/n)\mathbf{V}_n^{\mathsf{H}}$. For the graphon $\mathbf{W}$, define the filter $T_{\mathbf{H}} : L_2([0,1]) \to L_2([0,1])$ where $(T_{\mathbf{H}}\phi)(v) = \sum_{j=1}^{\infty} h(\lambda_j(T_{\mathbf{W}}))\hat{\phi}(\lambda_j)\varphi_j(v)$. If $\{\mathbf{G}_n\}$ converges to $\mathbf{W}$ and $h : [0,1] \to \mathbb{R}$ is continuous, then

$$\lim_{n\to\infty} \hat{\mathbf{H}}_n(\lambda_j(\mathbf{S}_n)/n) = \hat{T}_{\mathbf{H}}(\lambda_j(T_{\mathbf{W}})) \; . \tag{20}$$

Proof. Since $\mathbf{G}_n \to \mathbf{W}$, from Lemma 1 it holds that $\lambda_j(\mathbf{S}_n)/n \to \lambda_j$. Because h is continuous, this implies $h(\lambda_j(\mathbf{S}_n)/n) \to h(\lambda_j)$, which completes the proof. $\square$

Theorem 1 is an important result because states that we can design filters with a certain spectral behavior on the graphon and expect a similar spectral behavior when the same filter is applied to a large graph sampled from the graphon. However, it gives no filter behavior guarantees on the vertex domain, i.e., for the output of the filter. Vertex domain convergence can be

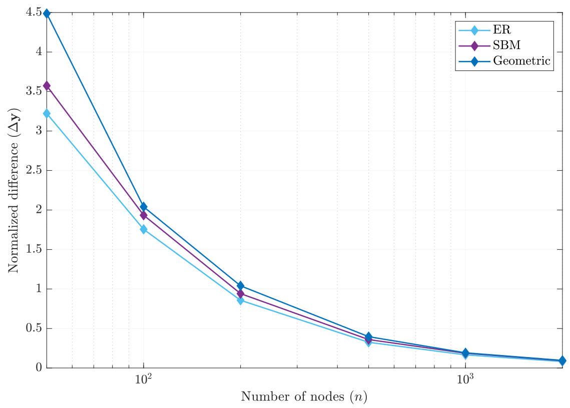

Figure 1: Norm difference between $\mathbf{S}_n\mathbf{x}_n$ and $\mathbf{S}_\mathbf{W}\mathbf{x}_\mathbf{W}$ interpolated at $\{u_i\}_{i=1}^n$. Three graphons were considered: ER ($\mathbf{W}(u,v) = 0.4$), SBM ($\mathbf{W}(u,v) = 0.8$ if $u,v \in [0,0.5]$ or $u,v \in [0.5,1]$, $\mathbf{W}(u,v) = 0.2$ otherwise), and geometric ($\mathbf{W}(u,v) = \exp(-2.3(u-v)^2)$). The diffusion output has been normalized for every n.

established for *Lipschitz graph and graphon filters*, whose filter function $h(\lambda)$ satisfies

$$\|h(\lambda) - h(\lambda')\| \leq L\|\lambda - \lambda'\| \tag{21}$$

and for *non-derogatory* graphons.

Definition 3. A graphon $\mathbf{W}$ is non-derogatory if $\lambda_i \neq \lambda_j$ for all $i \neq j$ and $i,j \in \mathbb{Z} \setminus \{0\}$.

Filter response convergence in the vertex domain is formally stated in Theorem 2, whose proof can be found in [21].

Theorem 2 (Convergence of filter response for Lipschitz continuous graph filters)**.** Let $\{(\mathbf{G}_n, \mathbf{y}_n)\}$ be the sequence of graph signals obtained by applying filters $\mathbf{H}_n(\mathbf{S}_n) = \mathbf{V}_n h(\mathbf{\Lambda}_n(\mathbf{S}_n)/n)\mathbf{V}_n^\mathsf{H}$ to the sequence $\{(\mathbf{G}_n, \mathbf{x}_n)\}$, and let $(\mathbf{W}, \gamma)$ be the graphon signal obtained by applying the graphon filter $(T_\mathbf{H}\phi)(v) = \sum_{j=1}^\infty h(\lambda_j)\hat{\phi}(\lambda_j)\varphi_j(v)$ to the signal $(\mathbf{W}, \phi)$, where $\mathbf{W}$ is non-derogatory. If $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ converges to $(\mathbf{W}, \phi)$ and the function h is Lipschitz continuous, then $\{(\mathbf{G}_n, \mathbf{y}_n)\}$ converges to $(\mathbf{W}, \gamma)$ in the sense of Definition 1.

Theorems 1 and 2 can be leveraged to simplify the design of graph filters on a family of graphs related to the same graphon (e.g., graphs sampled from the graphon). If the graphon is known, we can substitute the design of multiple filters in different graphs for the design of a single graphon filter from which graph filters can be later "sampled" by evaluating $h(\lambda)$ at each graph's eigenvalues. This is especially valuable in the case of large graphs or graphs that grow, for which repeating eigenvalue computations can be costly or unfeasible.

When it is possible to identify two graphs as belonging to the same class (i.e., as having the same generating graphon), these results also suggest the possibility of using a filter designed for one of these graphs on the other to approximately obtain the desired behavior. *Transfer learning*—the ability to transfer a learned architecture from one support to another—is a topic of growing interest in signal processing and machine learning [22]–[25]. Using the intuition provided by Theorem 2, we illustrate it in the experiments of Section V.

K	Number of users 50	100	200	400	600	800	Base
1	9.70%	4.70%	1.90%	0.45%	0.17%	0.04%	**0.77**
2	22.30%	20.47%	14.42%	5.48%	2.22%	0.37%	**0.72**
3	28.17%	13.58%	3.47%	0.32%	0.41%	−0.12%	**0.65**

Table I: Relative RMSE difference for rating prediction based on $K = 1,2,3$ filters obtained from $50, 100, 200, 400, 600$ and 800-user networks, w.r.t. to the base RMSE of filters obtained from the full user network.

V. Numerical experiments

A. GMRF diffusion on sensor networks

In this experiment, we simulate a Gaussian Markov Random Field (GMRF) measured and diffused on different sensor networks to analyze convergence of the filter $\mathbf{H}(\mathbf{S}) = \mathbf{S}$. This could be thought of an example of 1-hop information exchange between sensors in an IoT network. A graph signal $(\mathbf{G}, \mathbf{x})$ is a GMRF on $\mathbf{G}$ if $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{x}, \mathbf{\Sigma}_\mathbf{x})$ and $\mathbf{\Sigma}_\mathbf{x}$ is given by

$$\mathbf{\Sigma}_\mathbf{x} = |a_0|^2(\mathbf{I} - a\mathbf{S})^{-1}[(\mathbf{I} - a\mathbf{S})^{-1}]^\mathsf{H} \tag{22}$$

where the covariance matrix is calculated after sampling $\mathbf{G}$ from a random graph model for the sensor network, from which we obtain $\mathbf{S}$. Three random graph models, or graphons, are considered: Erdös-Rényi (ER), stochastic block model (SBM), and soft random geometric graph.

To simulate a graphon GMRF, we approximate the graphons as matrices $\mathbf{S}_\mathbf{W}$, obtained by evaluating each random graph model on $10^4 \times 10^4$ regularly spaced points of the unit square. The graphon GMRF is then obtained by sampling $\mathbf{x}_\mathbf{W} \in \mathbb{R}^{10^4}$ from the zero-mean multivariate Gaussian with covariance matrix given by (22) for $\mathbf{S} = \mathbf{S}_\mathbf{W}$. In order to observe convergence, we need to compare the outcome of the diffusion of the graphon GMRF with the outcome of the diffusion of a n-node graph signal sampled from it, for increasing n. This is done by uniformly sampling points $\{u_i\}_{i=1}^n$ from the unit line and generating graphs $\mathbf{G}_n$ where nodes i and j are connected (i.e., $[\mathbf{S}_n]_{ij} = 1$) with probability $\mathbf{W}(u_i, u_j)$. The corresponding graph signals $(\mathbf{G}_n, \mathbf{x}_n)$ are then generated by interpolating $\mathbf{x}_\mathbf{W}$ at each u_i.

We calculate the diffused graph signals $\mathbf{S}_n\mathbf{x}_n$ and interpolate the approximation of the diffused graphon signal $\mathbf{S}_\mathbf{W}\mathbf{x}_\mathbf{W}$ at $\{u_i\}_{i=1}^n$. To compare them, we compute their norm difference for increasing values of n. The average norm differences over 100 realizations of the graphon GMRF $\mathbf{x}_\mathbf{W}$ are graphed in Figure 1. We observe that, for all graphon models, the norm differences decrease with n. This indicates that the vertex response of $\mathbf{H}(\mathbf{S}) = \mathbf{S}$ converges as the graphs $\mathbf{G}_n$ grow, corroborating the result of Theorem 2.

B. Movie rating prediction via user-based graph filtering

In this experiment, we use a linear graph filter approach to predict ratings given by $U = 943$ users to $M = 1682$ movies in the MovieLens 100k dataset [26], which consists of 100,000 movie ratings ranging from 1 to 5. This problem can be interpreted as a network problem if we look at the movie ratings as graphs signals on top of a user similarity network. This network is constructed by computing correlations between user ratings to movies that pairs of users have rated, and connecting the top-40 nearest-neighbors. Each movie's rating vector is then a graph signal where user ratings are matched to nodes of the graph and where the value of the signal at nodes corresponding to users who have not rated the movie is 0.

By interpreting a movie's incomplete rating vector as a graph signal, we can think of the problem of predicting ratings as a problem of graph information processing and, indeed, a number of graph-based recommender systems have been proposed in the literature [1], [3], [27]. We focus on the method introduced in [1], which solves an optimization problem to find the optimal filter coefficients of a linear graph filter as in equation (4). Our objective is to find these coefficients using the ratings from a smaller network of users, and observe how the resulting graph filter generalizes when predicting ratings in the complete user network.

On user subnetworks of size ranging between $U = 50$ and $U = 800$ nodes, we obtain the optimal filter coefficients of filters with $K = 1$, 2 and 3 filter taps. Then, we compare the RMSE obtained by predicting ratings using the filters designed on the smaller networks versus the filters designed on the full network. The relative RMSE differences and the base RMSE (obtained from the filters calculated using the full 943-user network) are shown in Table I. For a network with U users, the reported RMSE difference corresponds to that of the average among filters designed on $\lfloor 943/U \rfloor$ different networks. The U users were picked at random.

From Table I, we observe that the RMSE difference decreases as the network size increases for all values of K. For $K = 1$ and $K = 3$ in particular, the relative RMSE difference is less than 1% for filters designed on networks with under half the total number of users in the dataset. These results show that graph filter *transferability*, proved in Theorems 1 and 2 for graphs and graphons, is not restricted to graphs related through some common statistical model; it can also be observed in real-world settings, where graphs are built from data that is empirically or statistically similar in some sense.

VI. Conclusions

In this paper, we have introduced the concept of a graphon filter to facilitate the design of graph filters on very large graphs. Using the convergence properties of graph sequences and the spectral properties of graphs and graphons, we were able to show in Theorems 1 and 2 that the response of graph filters converges to that of graphon filters in both the spectral and vertex domains. These findings were illustrated in experiments where we were able to numerically verify filter response convergence and, additionally, to see that filter transferability can be observed even in graphs that are similar because they are built from the same type of data, and not necessarily because they are related through a graphon.

References

[1] W. Huang, A. G. Marques, and A. Ribeiro, "Rating prediction via graph signal processing," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5066 – 5081, Oct. 2018.

[2] F. Gama, A. G. Marques, G. Leus, and A. Ribeiro, "Convolutional neural network architectures for signals supported on graphs," *arXiv:1805.00165v1 [eess.SP]*, 1 May 2018.

[3] L. Ruiz, F. Gama, G. Marques, and A. Ribeiro, "Invariance-preserving localized activation functions for graph neural networks," *arXiv:1903.12575 [eess.SP]*, 29 March 2019.

[4] Cisco and/or its affiliates, "Cisco annual internet report (2018-2023)," 2020.

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

[6] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Graph filters," in *IEEE Int. Conf. Acoust., Speech and Signal Process.* IEEE, 2013, pp. 6163–6166.

[7] L. Rui, H. Nejati, and N. Cheung, "Dimensionality reduction of brain imaging data using graph signal processing," in *2016 IEEE Int. Conf. on Image Process.*, Phoenix, AZ, 25-28 Sep. 2016, pp. 1329–1333, IEEE.

[8] R. B. Morgan, "Computing interior eigenvalues of large matrices," *Linear Algebra and its Applications*, vol. 154, pp. 289–309, 1991.

[9] C. C. Paige, *The computation of eigenvalues and eigenvectors of very large sparse matrices*, Ph.D. thesis, University of London, 1971.

[10] F. Gama, J. Bruna, and A. Ribeiro, "Stability properties of graph neural networks," *arXiv:1905.04497 [cs.LG]*, 2019.

[11] L. Lovász, *Large networks and graph limits*, vol. 60, American Mathematical Society, 2012.

[12] L. Lovász and B. Szegedy, "Limits of dense graph sequences," *Journal of Combinatorial Theory, Series B*, vol. 96, no. 6, pp. 933–957, 2006.

[13] P. J. Wolfe and S. C. Olhede, "Nonparametric graphon estimation," *arXiv:1309.5936 [math.ST]*, 2013.

[14] E. M. Airoldi, T. B. Costa, and S. H. Chan, "Stochastic blockmodel approximation of a graphon: Theory and consistent estimation," in *Advances in Neural Information Processing Systems*, 2013, pp. 692–700.

[15] Y. De Castro, C. Lacour, and T. M. P. Ngoc, "Adaptive estimation of nonparametric geometric graphs," *arXiv:1708.02107 [math.PR]*, 2017.

[16] M. Avella-Medina, F. Parise, M. Schaub, and S. Segarra, "Centrality measures for graphons: Accounting for uncertainty in networks," *IEEE Transactions on Network Science and Engineering*, 2018.

[17] F. Parise and A. Ozdaglar, "Graphon games," in *Proceedings of the 2019 ACM Conference on Economics and Computation*. ACM, 2019, pp. 457–458.

[18] S. Gao and P. E. Caines, "Graphon control of large-scale networks of linear systems," *arXiv:1807.03412 [math.OC]*, 2018.

[19] P. D. Lax, *Functional Analysis*, Wiley, 2002.

[20] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, "Convergent sequences of dense graphs II. multiway cuts and statistical physics," *Annals of Mathematics*, vol. 176, no. 1, pp. 151–219, 2012.

[21] L. Ruiz, L. F. O. Chamon, and A. Ribeiro, "Graphon filters: Signal processing in very large graphs (extended)," http://www.seas.upenn.edu/~rubruiz/graphon_eusipco_extended.pdf, 2020.

[22] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *International Conference on Artificial Neural Networks*. Springer, 2018, pp. 270–279.

[23] M. Eisen and A. Ribeiro, "Optimal wireless resource allocation with random edge graph neural networks," *arXiv:1909.01865 [eess.SP]*, 2019.

[24] R. Levie, E. Isufi, and G. Leus, Kutyniok, "On the transferability of spectral graph filters," *arXiv:1901.10524 [cs.LG]*, 29 Jan. 2019.

[25] A. Khan, E. Tolstaya, A. Ribeiro, and V. Kumar, "Graph policy gradients for large scale robot control," *arXiv:1907.03822 [cs.RO]*, 2019.

[26] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *ACM Transactions on Interactive Intelligent Systems (TiiS) - Regular Articles and Special issue on New Directions in Eye Gaze for Interactive Intelligent Systems*, vol. 5, pp. 1–19, Jan. 2016.

[27] F. Monti, M. M. Bronstein, and X. Bresson, "Geometric matrix completion with recurrent multi-graph neural networks," in *2017 Annu. Conf. Neural Inform. Process. Syst.*, Long Beach, CA, 4-9 Dec. 2017, NIPS Foundation.