Graphon Signal Processing

Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro, *Member, IEEE*

***Abstract*—Graphons are infinite-dimensional objects that represent the limit of convergent sequences of graphs as their number of nodes goes to infinity. This paper derives a theory of graphon signal processing centered on the notions of *graphon Fourier transform* and linear shift invariant *graphon filters*, the graphon counterparts of the graph Fourier transform and graph filters. It is shown that for convergent sequences of graphs and associated graph signals: (i) the graph Fourier transform converges to the graphon Fourier transform when the graphon signal is bandlimited; (ii) the spectral and vertex responses of graph filters converge to the spectral and vertex responses of graphon filters with the same coefficients. These theorems imply that for graphs that belong to certain families, i.e., that are part of sequences that converge to a certain graphon, graph Fourier analysis and graph filter design have well defined limits. In turn, these facts extend applicability of graph signal processing to graphs with large number of nodes — since signal processing pipelines designed for limit graphons can be applied to finite graphs — and to dynamic graphs — since we can relate the result of SP pipelines designed for different graphs from the same convergent graph sequence.**

***Index Terms*—Graphons, convergent graph sequences, graph filters, graph Fourier transform, graph signal processing.**

I. Introduction

GRAPH signal processing (GSP) provides an array of tools to process signals supported on graphs [1]–[3] but suffers from limitations in the case of graphs with *large* number of nodes or *dynamic* topologies. In these cases, just the acquisition of the graph may be challenging, which hinders the use of GSP tools such as filtering [4], [5] and graph neural network design [6] because they take the graph structure as a given. Other GSP tools like sampling [7]–[9] deal precisely with acquiring compact representations of graph signals. However, the *design* of sampling sets [10] requires not only access to the graph but the computation of an eigendecomposition that can be very costly for large matrices [11], [12, Chapter 1.1]. Challenges are most acute when the graph is both large *and* dynamic. In such cases, costly numerical computations must, in principle, be repeated as the graph changes, because the effect of graph perturbations is understood only in the case of relabelings [13] or small perturbations that induce small changes on the original eigenspace [14].

Yet, large graphs can often be identified as being similar to each other in the sense that they share structural properties. For instance, Figs. 1(b)–(c) show two instances of a random graph with 20 nodes, and Fig. 1(d) a random graph with 50 nodes. These graphs look similar and one can therefore foresee that analyzing signals supported on either of them should yield similar results. If this were the case, it would mitigate the challenge of dynamic variation since we could then design a filter for the graph in Fig. 1(b) and use it in the graph in Fig. 1(c). Similarly, it would mitigate the challenge of large size because we could design a filter for the graph in Fig. 1(b) and use it to process signals supported on the graph in Fig. 1(d). This paper formalizes this intuition ans shows that this graph interchangeability is possible when the graphs belong to the same "family," where each family is identified by a different *graphon*; see Fig. 1(a).

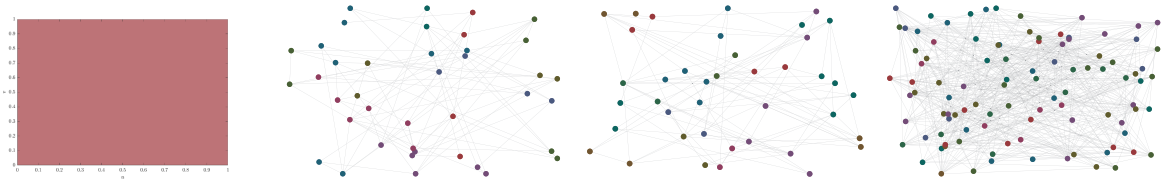
Graphons can be thought of as the infinite-dimensional counterparts of graphs, i.e., as graphs with an uncountable number of nodes. Appearing in many disciplines, they have been used to estimate random graph models in mathematics and statistics [15]–[20]; stabilize large-scale networks of linear systems in controls [21]; and perform graph partitioning [22], [23], node centrality [24] and network game equilibria computations [25] in very large networks. Graphons have two theoretical interpretations. They can be seen as generative models for families of graphs with weighted or stochastic edges [26, Chapter 10], and as the limit objects of convergent sequences of graphs [26, Chapter 7], [27]. In practice, these two interpretations suggest that graphons identify families of networks that are similar in the sense that the density of certain "motifs" is preserved. This motivates the study of signal processing on graphons as a way to enable the analysis of signals supported on large and/or dynamic graphs.

In this work, we thus introduce graphon signal processing (WSP), a framework to synthesize, analyze and process signals on graphons. More specifically, we put forward three novel technical contributions: (i) we define graphon signals and their graphon Fourier transforms (Definition 4), which can be seen as the continuous counterparts of graph signals and of their graph Fourier transforms; (ii) we show, by building upon the results of [28], that the graph Fourier transform converges to the graphon Fourier transform (Theorem 1) when the graphon signal is bandlimited (Definition 5); (iii) we define linear-shift-invariant (LSI) graphon filters (Definition 3), and prove that LSI graph filters converge to LSI graphon filters in both the spectral (Theorem 2) and vertex domains (Theorems 3–4).

Manuscript received March 30, 2021; revised July 1, 2021 and August 10, 2021; accepted August 11, 2021. Date of publication August 24, 2021; date of current version September 10, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. David Ramírez. The work was supported in part by NSF CCF 1717120, ARO10.13039/100000183 W911NF1710438, ARL DCIST CRA W911NF-17-2-0181, ISTC-WAS, and in part by Intel DevCloud. The preliminary results of this work were accepted for publication at the ICASSP20 conference [DOI: 10.1109/ICASSP40776.2020.9053439]. *(Corresponding author: Luana Ruiz.)*

The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104-6243 USA (e-mail: rubruiz@seas.upenn.edu; lfochamon@berkeley.edu; aribeiro@seas.upenn.edu).

Digital Object Identifier 10.1109/TSP.2021.3106857



(a) Graphon (b) $n = 20$ (c) $n = 20$ (d) $n = 50$

Fig. 1. Erdös–Rényi (constant) graphon with probability $p = 0.2$ and three n-node graphs sampled from it, illustrating scenarios where WSP can be useful. We can analyze signals and design systems on a graphon, to apply them on a graph sampled from it ((a) $\rightarrow$ (b)); on a graph, to apply on another graph of same size ((b) $\rightarrow$ (c)); and on a small graph, to apply on a larger graph ((c) $\rightarrow$ (d)).

Theorems 1–4 are especially important because they provide theoretical justification for transferring signal analysis methods and information processing architectures across graphs arising from (or leading to) the same graphon. Indeed, the ability of GSP systems to be transferred between graphs, also known as *transferability*, has been demonstrated empirically in network problems in wireless [29] and robotics [30]. We identify three transferability scenarios for which the results in this work provide a theoretical foundation: **(S1)** graphon to graph (Figs. 1(a)$\rightarrow$1(b)); **(S2)** graph to graph of same size (Figs. 1(b)$\rightarrow$1(c)); and **(S3)** graph to larger graph (Figs. 1(c)$\rightarrow$1(d)).

Attesting to the practical value of the WSP framework, each of these scenarios is illustrated in a numerical experiment in Section V. For instance, to showcase **(S1)**, we compare filter responses *on a graphon and on a graph sampled from this graphon*. The filter is a simple diffusion filter applied to a Gaussian Markov random field (GMRF). Interestingly, this example makes for a parallel with classical signal processing, where even if the application is digital, it is sometimes convenient to design and study filters in continuous time. For **(S2)**, we perform signal analysis in *different graphs of same size drawn from a common graphon*. Two n-node air pollution sensor networks are considered. As n grows, we compare the Fourier transforms of the same air pollution signal on top of them. This illustrates the behavior of SP tools when applied to networks for which we only have access to an approximated or perturbed version of the graph. Finally, we illustrate **(S3)** by *transferring filters designed on small graphs to large graphs* in a movie recommendation example. Using real data from the MovieLens dataset, we calculate the optimal coefficients of a rating prediction filter on networks containing only a subset of all users, and then use it to predict movie ratings on the full user network. The goal of this experiment is to show that graph filters are transferable at scale, which significantly simplifies signal processing on large graphs.

The rest of this paper is organized as follows. Preliminary definitions are recalled in Section II. Section III introduces the WSP framework and convergence results are stated in Section IV. Section V details the numerical experiments described above (**S1**–**S3**). Proofs are deferred to the appendices. Unless otherwise specified, $\|\cdot\|$ refers to the L^2 norm. When referring to the operator norm induced by the L^2 norm (spectral norm), we use the notation $|||\cdot|||$.

II. Graphs and Graphons

Graphs are triplets $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ where $\mathcal{V}$ is a set of n nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of edges and $\mathcal{W} : \mathcal{E} \rightarrow \mathbb{R}$ is a weight function assigning weights $\mathcal{W}(i,j) = w_{ij}$ to edges (i,j) in $\mathcal{E}$. The graph $\mathbf{G}$ can be equivalently represented by a number of matrix representations, which in the context of graph signal processing (GSP) are generically termed graph shift operators (GSOs). In this paper, we fix the GSO to be the *adjacency matrix* $\mathbf{S}$ satisfying $S_{ij} = w_{ij}$ if and only if $(i,j) \in \mathcal{E}$. We restrict our attention to undirected graphs with edge weights in the [0,1] interval, so that $\mathbf{S} = \mathbf{S}^{\mathsf{T}}$ and $\mathbf{S} \in [0,1]^{n \times n}$. We will also use the notations $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ and $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \mathbf{S})$ interchangeably.

A graphon is a bounded symmetric measurable function

$$\mathbf{W} : [0,1]^2 \rightarrow [0,1] \tag{1}$$

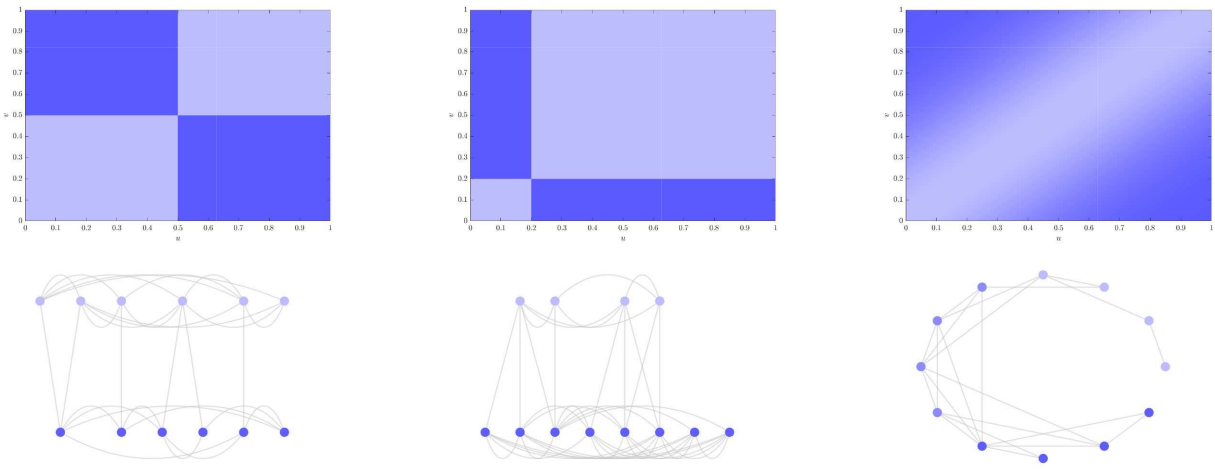
which represents a graph with an uncountable number of nodes [26, Chapter 7], [15]. By construction, graphons can also be interpreted as generative models for the edges of weighted or stochastic graphs. Namely, by associating sample points $u_i \in [0,1]$ to nodes $i \in \{1, 2, \ldots, n\}$, we can construct n-node graphs $\mathbf{G}_n$ where the edges are defined either by assigning edge weight $\mathbf{W}(u_i, u_j)$ to (i,j) or by connecting i and j with probability

$$p_{ij} = \mathbf{W}(u_i, u_j). \tag{2}$$

In the latter case, the $\mathbf{G}_n$ are unweighted. If, additionally, the u_i are sampled independently and uniformly at random, these stochastic graphs are called $\mathbf{W}$-*random graphs*. Three examples of graphons and of $\mathbf{W}$-random graphs sampled from them are shown in Fig. 2. The one in Fig. 2(a) is a stochastic block model (SBM) graphon with two balanced communities where the intra-community edge probability is 0.8 and the inter-community edge probability is 0.2. The one in Fig. 2(b) is also a SBM graphon with the same inter- and intra-community probabilities, but with unbalanced communities. The one in Fig. 2(c) is an exponential graphon, which can be used to generate graphs where nodes are connected if their labels u_i and u_j are close.

A. Convergent Sequences of Graphs

A second and perhaps more interesting interpretation of graphons is as the limit objects of convergent graph sequences. A sequence of graphs $\{\mathbf{G}_n\}$ is said to converge if and only if the density of homomorphisms between any finite, undirected and unweighted graph $\mathbf{F} = (\mathcal{V}', \mathcal{E}')$, which we call a *motif*, and the $\mathbf{G}_n$ converges [27]. Homomorphisms between a motif $\mathbf{F}$ and an arbitrary graph $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \mathbf{S})$ are adjacency preserving maps from $\mathcal{V}'$ to $\mathcal{V}$, i.e., a map $\beta : \mathcal{V}' \rightarrow \mathcal{V}$ is a homomorphism if,



(a) SBM with balanced communities (b) SBM with imbalanced communities (c) Exponential

Fig. 2. Graphons and 12-node $\mathbf{W}$-random graphs. (a) and (b) show SBM graphons and graphs with 2 communities and $p_{c_ic_i} = 0.8, p_{c_ic_j} = 0.2$. (c) shows an exponential graphon $\mathbf{W}(u, v) = \exp(-2.3(u - v)^2)$ and the corresponding graph.

for every $(i, j) \in \mathcal{E}'$, $(\beta(i), \beta(j)) \in \mathcal{E}$. The graph $\mathbf{F}$ can thus be interpreted as a graph pattern that we want to "identify" in $\mathbf{G}$.

A motif $\mathbf{F}$ will typically occur in multiple locations of the graph $\mathbf{G}$. Thus, we can count the number of homomorphisms between $\mathbf{F}$ and $\mathbf{G}$, which we denote $\hom(\mathbf{F}, \mathbf{G})$. Since there are a total of $|\mathcal{V}|^{|\mathcal{V}'|}$ possible maps between the vertices of $\mathbf{F}$ and $\mathbf{G}$ but only a fraction of them are homomorphisms, we further define the *density of homomorphisms* from $\mathbf{F}$ to $\mathbf{G}$ as

$$t(\mathbf{F}, \mathbf{G}) = \frac{\hom(\mathbf{F}, \mathbf{G})}{|\mathcal{V}|^{|\mathcal{V}'|}} = \frac{\sum_\beta \prod_{(i,j)\in\mathcal{E}'} [\mathbf{S}]_{\beta(i)\beta(j)}}{|\mathcal{V}|^{|\mathcal{V}'|}}. \tag{3}$$

This is easiest to understand when $\mathbf{G}$ is unweighted, in which case $t(\mathbf{F}, \mathbf{G})$ is simply the fraction of the total number of ways in which the motif $\mathbf{F}$ can be mapped into $\mathbf{G}$.

The concept of homomorphism densities can also be generalized to graphons. We define the density of homomorphisms between the motif $\mathbf{F}$ and the graphon $\mathbf{W}$ as

$$t(\mathbf{F}, \mathbf{W}) = \int_{[0,1]^{|\mathcal{V}'|}} \prod_{(i,j)\in\mathcal{E}'} \mathbf{W}(u_i, u_j) \prod_{i\in\mathcal{V}'} du_i. \tag{4}$$

This can be interpreted as the probability of sampling the motif $\mathbf{F}$ from the graphon $\mathbf{W}$. With these definitions in hand, a sequence of undirected graphs $\{\mathbf{G}_n\}$ is said to converge to the graphon $\mathbf{W}$ if, for all finite simple graphs $\mathbf{F}$,

$$\lim_{n\to\infty} t(\mathbf{F}, \mathbf{G}_n) = t(\mathbf{F}, \mathbf{W}). \tag{5}$$

In this case, we refer to $\mathbf{W}$ as the *limit graphon* of the sequence. This form of convergence is called "convergence in the homomorphism density sense". An example of convergent graph sequence that is easy to visualize is that of a sequence of $\mathbf{W}$-random graphs. The sequence of graphs $\{\mathbf{G}_n\}$ generated by sampling $\{u_i\}_{i=1}^n$ uniformly at random as $n \to \infty$ can be shown to converge in the homomorphism density sense with probability one [26, Example 11.6, Lemma 11.8].

To conclude, we point out that, while the two interpretations of a graphon — as a generative model for graph families and as the limit object of graph sequences — are theoretical, their practical value lies in that they can be used to identify sets of graphs with large number of nodes and similar structure. This simplifies the study of the properties of large graphs.

B. Convergence in Cut Norm

Similarly to how graphs can be obtained by sampling or evaluating a graphon, graphons can be defined, or *induced*, by graphs. Every undirected graph $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \mathbf{S})$ with $|\mathcal{V}| = n$ and $\mathbf{S} \in [0, 1]^{n\times n}$ admits an induced graphon representation $\mathbf{W}_\mathbf{G}$. This graphon is obtained in two steps. First, we construct a regular partition $I_1 \cup \ldots \cup I_n$ of [0,1], i.e., the partition given by $I_j = [(j - 1)/n, j/n)$ for $1 \leq j \leq n - 1$ and $I_n = [(n - 1)/n, 1]$. Then, the induced graphon $\mathbf{W}_\mathbf{G}$ is defined as [26, Chapter 7.1],[24, Sec. 5]

$$\mathbf{W}_\mathbf{G}(u, v) = \sum_{j=1}^n \sum_{k=1}^n [\mathbf{S}]_{jk} \times \mathbb{I}(u \in I_j)\mathbb{I}(v \in I_k). \tag{6}$$

The concept of induced graphon is useful to define a second mode of convergence for graph sequences — convergence in cut norm. The cut norm of a graphon $\mathbf{W}$ is defined as [26, eq. (8.13)]

$$\|\mathbf{W}\|_\square = \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} \mathbf{W}(u, v) du dv \right| \tag{7}$$

i.e., it is equal to the size of its maximum cut. The following lemma, adapted from [26, Theorem 11.57], states that if a sequence of graphs $\{\mathbf{G}_n\}$ converges to $\mathbf{W}$ in the homomorphism density sense, then it also converges to $\mathbf{W}$ in the cut norm.

Lemma 1 (Cut norm convergence): If $\{\mathbf{G}_n\} \to \mathbf{W}$ in the homomorphism density sense, then there exists a sequence of

permutations $\{\pi_n\}$ such that

$$\|\mathbf{W}_{\pi_n(\mathbf{G}_n)} - \mathbf{W}\|_\square \to 0 \tag{8}$$

where $\mathbf{W}_{\pi_n(\mathbf{G}_n)}$ is the graphon induced by the graph $\pi_n(\mathbf{G}_n)$.

Therefore, for every convergent sequence $\{\mathbf{G}_n\}$ there exists a non-empty set of permutation sequences $\{\pi_n\}$ for which the cut norm of the induced graphons $\mathbf{W}_{\pi_n(\mathbf{G}_n)}$ converges as in (8). This is formalized in Definition 1.

Definition 1 (Set of admissible permutations): Given a sequence $\{\mathbf{G}_n\}$ converging to $\mathbf{W}$ in the homomorphism density sense, the family of convergent permutation sequences $\mathcal{P}$ associated with $\{\mathbf{G}_n\}$ is defined as

$$\mathcal{P} = \left\{ \{\pi_n\} \mid \|\mathbf{W}_{\pi_n(\mathbf{G}_n)} - \mathbf{W}\|_\square \to 0 \right\}.$$

The set $\mathcal{P}$ will be especially important in the convergence analyses of Sections III–IV. In particular, in the definition of convergent sequences of graph signals (Definition 2), we will use permutation sequences $\{\pi_n\} \in \mathcal{P}$ to "organize" the signals on the graphs of a convergent sequence so that the labeling of the signals matches the node labeling for which the graphs converge.

C. Graph Signal Processing

GSP deals with signals defined on top of a graph $\mathbf{G}$. Formally, a graph signal is a map from the vertex set $\mathcal{V}$ onto the real numbers, which we write as the pair $(\mathbf{G}, \mathbf{x})$ and where the ith component x_i is the value of the signal at node i. The three fundamental concepts of GSP are shift operations, LSI filters and graph Fourier transforms (GFTs). We say that $\mathbf{z}$ is the result of shifting $\mathbf{x}$ on the graph $\mathbf{S}$ if $\mathbf{z} = \mathbf{S}\mathbf{x}$. Shifts can be composed to produce k-order shifted signals $\mathbf{S}^k\mathbf{x}$ and, as in the case of time signals, a weighted sum of shifted signals defines the LSI filter $\mathbf{H}(\mathbf{S})$ as the linear map

$$\mathbf{y} = \mathbf{H}(\mathbf{S})\mathbf{x} = \sum_{k=0}^{K} h_k \mathbf{S}^k \mathbf{x}. \tag{9}$$

In (9), the weights h_k are called graph filter taps [4]. The filter $\mathbf{H}(\mathbf{S})$ is said to be shift-invariant because, if $\mathbf{y} = \mathbf{H}(\mathbf{S})\mathbf{x}$ and we shift the input to $\mathbf{x}' = \mathbf{S}\mathbf{x}$, the output $\mathbf{y}' = \mathbf{H}(\mathbf{S})\mathbf{x}'$ is simply the shifted version of $\mathbf{y}$, $\mathbf{y}' = \mathbf{S}\mathbf{y}$. The LSI filter $\mathbf{H}(\mathbf{S})$ is also sometimes referred to as a graph convolutional filter [31].

Since $\mathbf{S}$ is symmetric, it is diagonalizable as $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathsf{H}}$. The matrix $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues and the columns of $\mathbf{V}$ are the graph eigenvectors. Herein, we assume that the eigenvalues are ordered according to their sign and in decreasing order of absolute value, i.e., $\lambda_1(\mathbf{S}) \geq \lambda_2(\mathbf{S}) \geq \ldots \geq 0 \geq \ldots \geq \lambda_{-2}(\mathbf{S}) \geq \lambda_{-1}(\mathbf{S})$ (e.g., Fig. 3). The graph Fourier transform (GFT) of the graph signal $(\mathbf{G}, \mathbf{x})$ is then defined as the projection of $\mathbf{x}$ onto the eigenvector basis $\mathbf{V}$, i.e.,

$$\hat{\mathbf{x}} = \text{GFT}\{(\mathbf{G}, \mathbf{x})\} = \mathbf{V}^{\mathsf{H}}\mathbf{x}. \tag{10}$$

In particular, $[\hat{\mathbf{x}}]_j = \mathbf{v}_j^{\mathsf{H}}\mathbf{x}$ for each index j, where $\mathbf{v}_j$ is the eigenvector associated with λ_j. This operation has the effect of decomposing $(\mathbf{G}, \mathbf{x})$ in the eigenbasis of the graph, which makes sense if we interpret the eigenvalues as frequencies. Similarly,

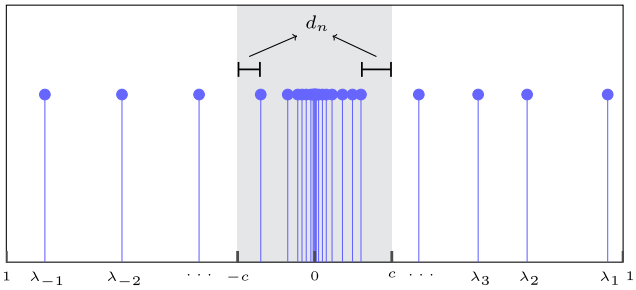


Fig. 3. Graphon eigenvalues. A graphon has an infinite number of eigenvalues λ_i but for any fixed constant c the number of eigenvalues $|\lambda_i| \geq c$ is finite. Thus, eigenvalues accumulate at 0 and this is the only accumulation point for graphon eigenvalues. The quantity d_n is approximately equal to the minimum distance between c (or $-c$) and the eigenvalues in the set $\{\lambda_i \mid \lambda_i < c\}$.

the inverse graph Fourier transform (iGFT) is defined as

$$\text{iGFT}\{\hat{\mathbf{x}}\} = \mathbf{V}\hat{\mathbf{x}} = \mathbf{x}.$$

Since $\mathbf{V}^{\mathsf{H}}\mathbf{V} = \mathbf{I}$, the iGFT is a proper inverse and can recover $\mathbf{x}$ from $\hat{\mathbf{x}}$ without loss of information.

The LSI filters defined in (9) also admit a spectral representation $\hat{\mathbf{H}}(\mathbf{\Lambda})$. This spectral representation is given by

$$\hat{\mathbf{H}}(\mathbf{\Lambda}) = \sum_{k=0}^{K} h_k \mathbf{\Lambda}^k. \tag{11}$$

Therefore, if we consider the action of the filter $\mathbf{H}(\mathbf{S})$ in the frequency domain, we see that $\hat{\mathbf{y}} = \hat{\mathbf{H}}(\mathbf{\Lambda})\hat{\mathbf{x}}$, i.e., graph filters are pointwise operators in the GFT domain.

Another interesting observation regarding the spectral response of graph filters is that, for any set of filter taps, we can define the frequency response [32]

$$h(\lambda) = \sum_{k=0}^{K} h_k \lambda^k. \tag{12}$$

Comparing (11) and (12), we see that using the same set of coefficients on different graphs induces different responses depending on the eigenvalues of $\mathbf{S}$. Indeed, if we let λ_i denote the ith eigenvalue of $\mathbf{S}$ and $\hat{x}_i$ and $\hat{y}_i$ the ith components of the GFTs $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, we have that $\hat{y}_i = h(\lambda_i)\hat{x}_i$.

The goal of this paper is to generalize the definitions of graph signals, GFTs, and convolutional graph filters to graphon signals, graphon Fourier transforms, and convolutional graphon filters (Section III). In introducing this graphon signal processing framework, we intend to do for large-scale graph signal processing what graphons do for very large and dynamic networks. Namely, we leverage the fact that working with limits is easier and focus on graph signal limits — graphon signals — to facilitate GSP on large-scale and dynamic networks. Importantly, we also show that, for sequences of graphs that converge to a graphon in the sense of (5), corresponding sequences of GFTs and graph filters converge to the respective graphon Fourier transforms and graphon filters (Section IV).

III. Graphon Signal Processing

The central concept in graphon signal processing is the graphon signal. Graphon signals are defined as pairs $(\mathbf{W}, X)$ where the function $X : [0,1] \to \mathbb{R}$ maps points of the unit interval to the real numbers, i.e., the signal values. The graphon signals that we consider have finite energy, i.e., X is a function in $L^2([0,1])$. As in the case of graphons, graphon signals can be *induced by graph signals*. Given a n-node graph signal $(\mathbf{G}, \mathbf{x})$, the induced graphon signal $(\mathbf{W}_\mathbf{G}, X_\mathbf{G})$ is defined as

$$X_\mathbf{G}(v) = \sum_{j=1}^{n} [\mathbf{x}]_j \times \mathbb{I}(v \in I_j) \tag{13}$$

where $\mathbf{W}_\mathbf{G}$ is the graphon induced by $\mathbf{G}$ [cf. (6)] and $I_1 \cup \ldots \cup I_n$ is the regular partition of the unit interval, i.e., $I_j = [(j-1)/n, j/n)$ for $1 \leq j \leq n-1$ and $I_n = [(n-1)/n, 1]$.

A. Convergent Sequences of Graph Signals

We define convergent sequences of graph signals as follows.

Definition 2 (Convergent sequences of graph signals): A sequence of graph signals $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ is said to converge to the graphon signal $(\mathbf{W}, X)$ if, for every graph motif $\mathbf{F}$,

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

and if there exists a sequence of permutations $\{\pi_n\} \in \mathcal{P}$ such that

$$\lim_{n\to\infty} \|X_{\pi_n(\mathbf{G}_n)} - X\| = 0$$

where $\mathcal{P}$ is the set of admissible permutations for the sequence $\{\mathbf{G}_n\}$ [cf. Definition 1] and $(\mathbf{W}_{\pi_n(\mathbf{G}_n)}, X_{\pi_n(\mathbf{G}_n)})$ is the graphon signal induced by the permuted graph signal $(\pi_n(\mathbf{G}_n), \pi_n(\mathbf{x}_n))$ [cf. (13)].

A sequence of graph signals is thus convergent if (i) the underlying graphs converge and (ii) the graphon signals induced by some permutation sequence $\{\pi_n\} \in \mathcal{P}$ of the graph signals converge in L^2. The role of the permutation sequence is to match the labels of the signals $\mathbf{x}_n$ to those of the sequence of graphs $\mathbf{G}_n$ that converges in cut norm [cf. Lemma 1]. This is a similar requirement to the isometric embeddings necessary to define Gromov-Hausdorff convergence in metric spaces [33].

Importantly, the graph signal limit $(\mathbf{W}, X)$ is unique for each $\{\pi_n\} \in \mathcal{P}$. Indeed, suppose that it is not, i.e., that $\|X_{\pi_n(\mathbf{G}_n)} - X\| \to 0$ and $\|X_{\pi_n(\mathbf{G}_n)} - Y\| \to 0$ with $\|X - Y\| \geq \epsilon > 0$. Using the triangle inequality, we get

$$\begin{aligned}\|X - Y\| &= \|X - X_{\pi_n(\mathbf{G}_n)} + X_{\pi_n(\mathbf{G}_n)} - Y\| \\ &\leq \|X - X_{\pi_n(\mathbf{G}_n)}\| + \|X_{\pi_n(\mathbf{G}_n)} - Y\| \to 0,\end{aligned}$$

which contradicts the hypothesis since there must then exist n_0 such that $\|X - X_{\pi_n(\mathbf{G}_n)}\| + \|X_{\pi_n(\mathbf{G}_n)} - Y\| < \epsilon$ for $n > n_0$. See also Remark 1.

B. The Graphon Operator and Graphon Filters

Every graphon $\mathbf{W}$ induces an integral operator $T_\mathbf{W} : L^2([0,1]) \to L^2([0,1])$, which maps a signal $(\mathbf{W}, X)$ to the signal $(\mathbf{W}, Y)$ given by

$$Y(v) = (T_\mathbf{W} X)(v) = \int_0^1 \mathbf{W}(u,v) X(u) du. \tag{14}$$

We refer to $T_\mathbf{W}$ as the *graphon shift operator* (WSO) because it induces a diffusion of $(\mathbf{W}, X)$ on the graphon analogous to the diffusion induced by the adjacency matrix $\mathbf{S}$ on a graph. Building upon this parallel, LSI graphon filters are defined as follows.

Definition 3 (LSI graphon filters): Let $(\mathbf{W}, X)$ be a graphon signal. A LSI graphon filter $T_\mathbf{H} : L^2([0,1]) \to L^2([0,1])$ maps $(\mathbf{W}, X) \mapsto (\mathbf{W}, Y)$ with $(\mathbf{W}, Y)$ given by

$$\begin{aligned} Y(v) &= (T_\mathbf{H} X)(v) = \sum_{k=0}^{K} h_k (T_\mathbf{W}^{(k)} X)(v) \\ \text{where } (T_\mathbf{W}^{(k)} X)(v) &= \int_0^1 \mathbf{W}(u,v) (T_\mathbf{W}^{(k-1)} X)(u) du, \; k \geq 1 \end{aligned} \tag{15}$$

and $T_\mathbf{W}^{(0)} = \mathbb{I}$, the identity operator. The h_k are known as the *graphon filter taps*.

Similarly to LSI graph filters, the graphon filters in Definition 3 are shift-invariant because, given an input graphon signal X and denoting the graphon filter ouput $Y = T_\mathbf{H} X$, applying $T_\mathbf{H}$ to a shifted version of the input $X' = T_\mathbf{W} X$ yields $Y' = T_\mathbf{H} X' = T_\mathbf{W} Y$, i.e., Y' is a shifted version of Y.

C. Graphon Spectra and the Graphon Fourier Transform

Because $\mathbf{W}$ is a bounded symmetric function, $T_\mathbf{W}$ is a self-adjoint Hilbert-Schmidt operator (see Lemma 5, Appendix C in the supplementary material). As such, it can be decomposed in the operator's basis as

$$\mathbf{W}(u,v) = \sum_{i \in \mathbb{Z}\setminus\{0\}} \lambda_i \varphi_i(u) \varphi_i(v) \tag{16}$$

with eigenvalues $\lambda_i \in [-1,1]$, and eigenfunctions $\varphi_i : [0,1] \to \mathbb{R}$. As before, we separate positive and negative eigenvalues by ordering them according to their sign and in decreasing order of absolute value. Therefore, we have $1 \geq \lambda_1 \geq \lambda_2 \geq \ldots \geq 0 \geq \ldots \geq \lambda_{-2} \geq \lambda_{-1} \geq -1$. The eigenfunctions form an orthonormal basis of $L^2([0,1])$. Note that the eigenvalues, and hence the eigenfunctions, are countable. What is more, since $\|\mathbf{W}\|^2 \leq 1$ the trace of $T_\mathbf{W}^* T_\mathbf{W}$ is bounded by one and so the λ_i converge to 0 for $|i| \to \infty$ as depicted in Fig. 3. Zero is the only point of accumulation, which in turn implies that all $\lambda_i \neq 0$ have finite multiplicity [34].

(16) allows writing $T_\mathbf{W}$ as

$$(T_\mathbf{W} X)(v) = \sum_{i \in \mathbb{Z}\setminus\{0\}} \lambda_i \varphi_i(v) \int_0^1 \varphi_i(u) X(u) du. \tag{17}$$

The integral terms $\int_0^1 \varphi_i(u) X(u) du$ are the L^2 inner products $\langle X, \varphi_i \rangle$ between the signal X and the eigenfunctions φ_i. Since the φ_i form a complete orthonormal basis of $L^2([0,1])$, the inner products $\langle X, \varphi_i \rangle$ provide a complete representation of $(\mathbf{W}, X)$ on the graphon basis. Although there is an infinite number of

eigenfunctions, they are countable and so the change of basis can always be defined. This change of basis operation is called the *graphon Fourier transform* (WFT).

Definition 4 (Graphon Fourier Transform): Consider the graphon signal $(\mathbf{W}, X)$, and let $\{\lambda_i\}_{i\in\mathbb{Z}\setminus\{0\}}$ and $\{\varphi_i\}_{i\in\mathbb{Z}\setminus\{0\}}$ be the eigenvalues and eigenfunctions of $T_{\mathbf{W}}$. The graphon Fourier transform (WFT) of $(\mathbf{W}, X)$ is defined as

$$\text{WFT}\left[(\mathbf{W}, X)\right] = \hat{X} \quad \text{with}$$

$$[\hat{X}]_i = \hat{X}(\lambda_i) = \int_0^1 X(u)\varphi_i(u)du.$$

The inverse Graphon Fourier Transform (iWFT) of $\hat{X}$ is defined as

$$\text{iWFT}\left[\hat{X}\right] = \sum_{i\in\mathbb{Z}\setminus\{0\}} \hat{X}(\lambda_i)\varphi_i = X.$$

Since the $\{\varphi_i\}_{j\in\mathbb{Z}\setminus\{0\}}$ are orthonormal, the iWFT is the proper inverse transformation of the WFT. Definition 4 further allows defining graphon signals that are *bandlimited*.

Definition 5 (Bandlimited graphon signals): A graphon signal $(\mathbf{W}, X)$ is *c-bandlimited* with bandwith $c \in [0, 1]$ if $\hat{X}(\lambda_i) = 0$ for all i such that $|\lambda_i| < c$.

Because all nonzero eigenvalues have finite multiplicity, the WFT of a bandlimited graphon signal is finite-dimensional. Bandlimited graphon signals have two noteworthy properties. The first is that they extend the notion of graph bandlimited signals, which are the most common type of graph signal in practical GSP applications [35]. The second is that, since they only depend on a finite number of graphon eigenfunctions, their WFT can be computed analytically. Although countability of the φ_i allows us to write the definition of the WFT (Definition 4) for any graphon signal, calculating all inner products $\langle X, \varphi_i\rangle$ is infeasible because the graphon basis is infinite-dimensional.

D. Graphon Filter Frequency Response

The WFT also allows obtaining the frequency response of graphon filters. Using the spectral decomposition of $T_{\mathbf{W}}$ (16), we can rewrite the LSI graphon filter (15) as

$$Y(v) = (T_{\mathbf{H}}X)(v) = \sum_{i\in\mathbb{Z}\setminus\{0\}} \sum_{k=0}^{K} h_k \lambda_i^k \hat{X}(\lambda_i)\varphi_i(v).$$

Hence, the frequency response of $T_{\mathbf{H}}$ is given by

$$\hat{T}_{\mathbf{H}}(\lambda) = h(\lambda) = \sum_{k=0}^{K} h_k \lambda^k. \tag{18}$$

Equation (18) is the infinite counterpart of the frequency response of a LSI graph filter (12). Note that, to understand the behavior of this filter on a specific graphon $\mathbf{W}$, we need to evaluate $h(\lambda)$ at each graphon eigenvalue λ_i. But (18) is otherwise independent of the graphon. In other words, the frequency response of a graphon filter always has the same shape, irrespective of the graphon. A third important remark pertaining to (18) is that LSI graphon filters can approximate any filter with analytic frequency response $h(\lambda)$ arbitrarily well as $K \to \infty$. This is because the frequency response of a LSI graphon filter (18) is a polynomial of the eigenvalues of the graphon. Put formally, a graphon filter with frequency response $\hat{T}_{\mathbf{H}}(\lambda) = h(\lambda)$ can be written as a LSI graphon filter (Definition 3) provided that $h(\lambda)$ is analytic, i.e., that it is infinitely differentiable at $\{\lambda_i\}_{i\in\mathbb{Z}\setminus\{0\}}$ and its Taylor series converges pointwise.

We conclude this section by stressing that the goal of Defs. 3 and 4, as well as of the definition of a graphon signal, is to *generalize* GSP concepts to graphons. These concepts are not realizable in the way that graph signals, graph filters, and the GFT are because, unlike graphs, graphons are intangible theoretical objects. Nonetheless, their value lies in that they help understand the behavior of graph signals in the limit of large-scale networks. This provides the theoretical foundations to enable the practical scenarios (**S1**–**S3**) discussed in the introduction (see also Section V). Indeed, as we show next, the WFT and the LSI graphon filter are mathematical limits of the GFT and of the LSI graph filter on convergent sequences of graph signals.

Remark 1 (Uniqueness of limit graphon signal on $\mathbf{W}$*):* A sequence of graphs $\{\mathbf{G}_n\}$ converges to a graphon $\mathbf{W}$ if and only if the homomorphism densities $t(\mathbf{F}, \mathbf{G}_n)$ converge to $t(\mathbf{F}, \mathbf{W})$ for every motif $\mathbf{F}$. Naturally, there may be other graphons $\mathbf{W}'$ with same homomorphism densities $t(\mathbf{F}, \mathbf{W}') = t(\mathbf{F}, \mathbf{W})$ for all graphs $\mathbf{F}$ and so the limit graphon $\mathbf{W}$ is not necessarily unique, but in this paper we select one of these limits—the graphon $\mathbf{W}$— without loss of generality[1] and use its implicit node labeling to define sequences of graph signals $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ that converge to graphon signals $(\mathbf{W}, X)$ [cf. Definition 2]. Since $\mathbf{W}$ is fixed, for each sequence $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ the limit signal $(\mathbf{W}, X)$ is unique in L^2.

IV. GSP Converges to WSP

In this section, we leverage the properties of convergent graph sequences to prove a series of convergence results which show that GSP converges to WSP. Our first result describes the limit behavior of the GFT on convergent sequences of graph signals (Section IV-A). We show that, when the limit graphon signal is bandlimited, the GFT converges to the WFT (Theorem 1). Our second result shows that the spectral responses of graph filters converges to the spectral response of a well-defined graphon filter (Theorem 2, Section IV-B). Theorem 2 is further extended to show that convergence of the graph signal and graph filter also implies convergence of the filter response in the *vertex domain*. This is our third convergence result. It is stated with increasing levels of generality in Corollary 1, which follows directly from convergence of the GFT and of the filter spectral responses; and in Theorems 3 and 4, which do not require the graphon signal to be bandlimited. These findings, particularly the more general Theorem 4, are the main technical contributions of this

[1]Graphons $\mathbf{W}'$ and $\mathbf{W}$ with same homomorphism densities $t(\mathbf{F}, \mathbf{W}') = t(\mathbf{F}, \mathbf{W})$ for all $\mathbf{F}$ are called weakly isomorphic. We can select $\mathbf{W}$ without loss of generality because two graphons $\mathbf{W}'$ and $\mathbf{W}$ are weakly isomorphic if and only if $\delta_{\square}(\mathbf{W}, \mathbf{W}') = 0$, i.e., if their cut distance is zero [26, Chapters 7.3, 8.2.2]. Therefore, the limit graphon is unique w.r.t. the metric induced by the cut distance in the space of *unlabeled* graphons.

TABLE I
TABLE SUMMARY OF GSP AND WSP

	Graph	Graphon	Convergence result
Signal	$(\mathbf{G}, \mathbf{x})$	$(\mathbf{W}, X)$	Def. 2
FT	$\hat{\mathbf{x}}$ (eq. (10))	$\hat{X}$ (Def. 4)	Thm. 1
Filter	$\mathbf{H}(\mathbf{S})$ (eq. (9))	$T_{\mathbf{H}}$ (Def. 3)	Thm. 4

work. At the end of the section, a table summary of the GSP and WSP definitions of a signal, of the Fourier Transform and of linear shift-invariant filters can be found in Table I. This table also highlights the relationships between these concepts as established by Theorems 1–4.

A. Convergence of the GFT

When a sequence of graph signals converges to a bandlimited graphon signal, we can show that the GFT converges to the WFT as long as the limit graphon is *non-derogatory* (Definition 6). This is the main result of this section, presented in Theorem 1.

Definition 6: A graphon $\mathbf{W}$ is non-derogatory if $\lambda_i \neq \lambda_j$ for all $i \neq j$ and $i, j \in \mathbb{Z} \setminus \{0\}$.

Theorem 1 (GFT $\to$ WFT for BL graphon signals): Let $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ be a sequence of graph signals and let $\{\pi_n\} \in \mathcal{P}$ be a sequence of permutations such that $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ converges to the c-bandlimited graphon signal $(\mathbf{W}, X)$ in the sense of Definition 2, where $\mathbf{W}$ is non-derogatory. Then,

$$\text{GFT}\left[(\pi_n(\mathbf{G}_n), \pi_n(\mathbf{x}_n))\right] \to \text{WFT}\left[(\mathbf{W}, X)\right].$$

Conversely, if $\{(\mathbf{G}_n, \hat{\mathbf{x}}_n)\}$ is a sequence of GFTs converging to the WFT $(\mathbf{W}, \hat{X})$, then there exists a sequence of permutations $\{\pi_n\} \in \mathcal{P}$ such that

$$\pi_n\left(\text{iGFT}\left[\mathbf{x}_n\right]\right) \to \text{iWFT}\left[\hat{X}\right].$$

Note that the GFT coefficients $[\hat{\mathbf{x}}]_i$ and the WFT coefficients $[\hat{X}]_i$ inherit the ordering of the eigenvalue sets of the graphon $\mathbf{W}$ and of the graphs $\mathbf{G}_n$, which are both ordered with indices $i \in \mathbb{Z} \setminus \{0\}$ according to their sign and in decreasing order of absolute value.

Theorem 1 relates the GFT, a Fourier transform for "discrete" graph signals, to the WFT, a Fourier transform for "continuous" graphon signals. This comparison is only possible because, like graphs, graphons have discrete spectra. This unveils an interesting parallel with the relationship between the discrete Fourier transform (DFT) — a discrete transform for discrete signals — and the Fourier series —- a discrete transform for continuous signals — in Euclidean domains. Another important consequence of Theorem 1 is that it allows inferring the spectral content of graph signals by analyzing the spectral content of the graphon signals giving rise to them (and vice-versa). This is useful, for instance, when these signals and/or the graphs on which they are supported are corrupted or only partially observable, in which case the WFT (or the GFT on another graph of the same family) can be used as an approximation of the GFT on the original graph. We also point out that the requirement that the graphon be non-derogatory is not very restrictive: as stated in the following proposition, the space of non-derogatory graphons is dense in the space of graphons.

Proposition 1 (Density of $\mathfrak{W}$): Let $\mathfrak{W}$ denote the space of all bounded symmetric measurable functions $\mathbf{W} : [0, 1]^2 \to \mathbb{R}$, i.e., the space of graphons. The space of *non-derogatory* graphons is dense in $\mathfrak{W}$.

Proof: Refer to Appendix D in the supplementary material. ■

Prop. 1 tells us that, even if a graphon is derogatory, there exists a non-derogatory graphon arbitrarily close to it for which the GFT convergence result from Theorem 1 holds.

In order to establish Theorem 1 and subsequent results, we will work with the graphon signals induced by the graph signals we consider. We have already described in (13) how their (vertex domain) values are related. In the sequel, we will also need to relate their spectral properties. This relationship is formalized in Lemma 2. Note that, although the adjacencies $\mathbf{S}_n$ of the graphs $\mathbf{G}_n$ have a finite number of eigenvalues $\lambda_i(\mathbf{S}_n)$, we still associate the eigenvalue sign with its index and order the eigenvalues in decreasing order of absolute value. The indices i are now defined on some finite set $\mathcal{L} \subseteq \mathbb{Z} \setminus \{0\}$.

Lemma 2: Let $(\mathbf{W}_{\mathbf{G}}, X_{\mathbf{G}})$ be the graphon signal induced by the graph signal $(\mathbf{G}, \mathbf{x})$ on n nodes. Then, for $i \in \mathcal{L}$ we have

$$\lambda_i(T_{\mathbf{W}_{\mathbf{G}}}) = \frac{\lambda_i(\mathbf{S})}{n}$$

$$\varphi_i(T_{\mathbf{W}_{\mathbf{G}}})(u) = [\mathbf{v}_i]_j \times \sqrt{n}\mathbb{I}\left(u \in I_j\right),\ j = 0, \ldots, n \text{ and}$$

$$[\hat{X}_{\mathbf{G}}]_i = \frac{[\hat{\mathbf{x}}]_i}{\sqrt{n}}$$

where $\lambda_i(\mathbf{S})$ are the eigenvalues of the adjacency matrix. For $i \notin \mathcal{L}$, we let $\lambda_i(T_{\mathbf{W}_{\mathbf{G}}}) = [\hat{X}_{\mathbf{G}}]_i = 0$ and $\varphi_i(T_{\mathbf{W}_{\mathbf{G}}}) = \psi_i$ such that $\{\varphi_i(T_{\mathbf{W}_{\mathbf{G}}})\} \cup \{\psi_i\}$ forms an orthonormal basis of $L^2([0, 1])$.

Proof: Refer to Appendix A.

Proof of Theorem 1: We now prove that, since the finite set $\mathcal{L}$ converges to $\mathbb{Z} \setminus \{0\}$ as n goes to infinity, $\text{WFT}\{(\mathbf{W}_{\pi_n(\mathbf{G}_n)}, \pi_n(X_{\mathbf{G}_n}))\} \to \text{WFT}\{(\mathbf{W}, X)\}$. We leave the dependence of the convergent signal sequence $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ on $\{\pi_n\}$ implicit and write $\mathbf{W}_n = \mathbf{W}_{\pi_n(\mathbf{G}_n)}$ and $X_n = X_{\pi_n(\mathbf{G}_n)}$. Next, we use the eigenvector convergence result from the following lemma. Theorem 1 then follows from the fact that inner products are continuous in the product topology that they induce.

Lemma 3: Let $\mathcal{C} = \{i \in \mathbb{Z} \setminus \{0\} \mid |\lambda_i(T_{\mathbf{W}})| \geq c\}$ be the set of indices of the non-vanishing eigenvalues and denote $\mathcal{S}$ the subspace spanned by the eigenfunctions $\{\varphi_i(T_{\mathbf{W}})\}_{i \notin \mathcal{C}}$. Then,

(i) for $i \in \mathcal{C}$, $\varphi_i(T_{\mathbf{W}_n}) \to \varphi_i(T_{\mathbf{W}})$ in L_2, and
(ii) for $i \notin \mathcal{C}$, $\varphi_i(T_{\mathbf{W}_n}) \to \Psi_i \in \mathcal{S}$ weakly.

Proof: Refer to Appendix B. ■

Starting with the eigenvectors with indices in $\mathcal{C}$, for any $\epsilon > 0$ it holds from Lemma 3 and from the convergence of X_n in L^2 that there exist n_1 and n_2 such that

$$\|\varphi_i(T_{\mathbf{W}_n}) - \varphi_i(T_{\mathbf{W}})\| \leq \frac{\epsilon}{2\|X\|}, \text{ for all } n > n_1$$

$$\text{and} \quad \|X_n - X\| \leq \frac{\epsilon}{2}, \text{ for all } n > n_2.$$

Recall that $\|\varphi_i(T_{\mathbf{W}_n})\| \leq 1$ for all n and $i \in \mathcal{C}$ because the graphon spectral basis is orthonormal. Since the sequence $\{X_n\}$ is convergent, it is bounded and $\|X\| < \infty$. Let $m = \max\{n_1, n_2\}$. Then, it holds that

$$\begin{aligned}|[\hat{X}_n]_i - [\hat{X}]_i| &= |\langle X_n, \varphi_i(T_{\mathbf{W}_n})\rangle - \langle X, \varphi_i(T_{\mathbf{W}})\rangle| \\ &= |\langle X_n - X, \varphi_i(T_{\mathbf{W}_n})\rangle + \langle X, \varphi_i(T_{\mathbf{W}_n}) - \varphi_i(T_{\mathbf{W}})\rangle| \\ &\leq \|X_n - X\|\|\varphi_i(T_{\mathbf{W}_n})\| + \|X\|\|\varphi_i(T_{\mathbf{W}_n}) - \varphi_i(T_{\mathbf{W}})\| \\ &\leq \frac{\epsilon}{2}\|\varphi_i(T_{\mathbf{W}_n})\| + \|X\|\frac{\epsilon}{2\|X\|} \leq \epsilon \text{ for all } n > m.\end{aligned}$$

For $i \notin \mathcal{C}$, the eigenfunctions $\varphi_i(T_{\mathbf{W}_n})$ may not converge to $\varphi_i(T_{\mathbf{W}})$, but they do converge to some function $\Psi_i \in \mathcal{S}$. Given that the graphon signal $(\mathbf{W}, X)$ is c-bandlimited, we have $\langle X, \varphi_i(T_{\mathbf{W}})\rangle = 0$ for $i \notin \mathcal{C}$, so that X must be orthogonal to all functions in $\mathcal{S}$. Using the same argument as for $i \in \mathcal{C}$ yields that the remaining GFT coefficients also converge to the WFT. Formally,

$$\langle \varphi_i(T_{\mathbf{W}_n}), X_n\rangle \to \langle \Psi_i, X\rangle = 0 = \langle \varphi_i(T_{\mathbf{W}}), X\rangle.$$

Convergence of the iGFT to the iWFT follows directly from these results and from Lemma 3. Explicitly, use the triangle inequality to write

$$\begin{aligned}&\left\| \sum_{i\in\mathbb{Z}\setminus\{0\}} [\hat{X}]_i \varphi_i(T_{\mathbf{W}}) - \sum_{i\in\mathbb{Z}\setminus\{0\}} [\hat{X}_n]_i \varphi_i(T_{\mathbf{W}_n}) \right\| \\ &\quad \leq \sum_{i\in\mathbb{Z}\setminus\{0\}} \|[\hat{X}]_i \varphi_i(T_{\mathbf{W}}) - [\hat{X}]_i \varphi_i(T_{\mathbf{W}_n})\| \\ &\quad + \sum_{i\in\mathbb{Z}\setminus\{0\}} \|[\hat{X}]_i \varphi_i(T_{\mathbf{W}_n}) - [\hat{X}_n]_i \varphi_i(T_{\mathbf{W}_n})\|.\end{aligned}$$

Applying the Cauchy-Schwarz inequality and splitting the sums between $i \in \mathcal{C}$ and $i \notin \mathcal{C}$, we get

$$\begin{aligned}&\left\| \sum_{i\in\mathbb{Z}\setminus\{0\}} [\hat{X}]_i \varphi_i(T_{\mathbf{W}}) - \sum_{i\in\mathbb{Z}\setminus\{0\}} [\hat{X}_n]_i \varphi_i(T_{\mathbf{W}_n}) \right\| \\ &\quad \leq \sum_{i\in\mathcal{C}} |[\hat{X}]_i| \|\varphi_i(T_{\mathbf{W}}) - \varphi_i(T_{\mathbf{W}_n})\| \\ &\quad + \sum_{i\in\mathcal{C}} |[\hat{X}]_i - [\hat{X}_n]_i| \|\varphi_i(T_{\mathbf{W}_n})\| \\ &\quad + \sum_{i\notin\mathcal{C}} |[\hat{X}_n]_i| \|\varphi_i(T_{\mathbf{W}_n})\| \to 0. \qquad (19)\end{aligned}$$

The first term on the right hand side of (19) vanishes because $\varphi_i(T_{\mathbf{W}_n}) \to \varphi_i(T_{\mathbf{W}})$ for $i \in \mathcal{C}$ by Lemma 3; the second term, because the GFT coefficients $[\hat{X}_n]_i$ converge to $[\hat{X}]_i$ for $i \in \mathcal{C}$; and the third term, because $[\hat{X}_n]_i \to [\hat{X}]_i = 0$ for $i \notin \mathcal{C}$. □

B. Convergence of Graph Filter Responses in the Spectral and Vertex Domains

Our second convergence result involves the frequency response of graph filters. As we have already noted, the frequency responses of LSI graph filters (12) and of their graphon counterparts (18) have the same expression as long as the coefficients h_k (or, equivalently, the function h) are the same. In the following, we show that these frequency responses actually converge to one another as the number of nodes goes to infinity.

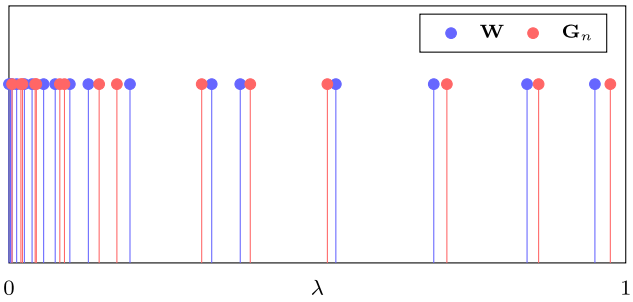


Fig. 4. Comparison of graphon eigenvalues (blue) and eigenvalues of a graph $\mathbf{G}_n$ taken from a convergent graph sequence (red). Only the positive eigenvalues are depicted. For $n \to \infty$, the eigenvalues of $\mathbf{G}_n$ converge to the eigenvalues of $\mathbf{W}$. However, the accumulation of graphon eigenvalues close to $\lambda = 0$ means that the GFT converges to the WFT only for graphon bandlimited signals.

Theorem 2 (Convergence of graph filter frequency response): On the graph sequence $\{\mathbf{G}_n\}$, let $\mathbf{H}_n(\mathbf{S}_n)$ be a sequence of filters of the form $\mathbf{H}_n(\mathbf{S}_n) = \mathbf{V}_n h(\mathbf{\Lambda}(\mathbf{S}_n)/n)\mathbf{V}_n^{\mathsf{H}}$; and, on the graphon $\mathbf{W}$, define the filter

$$(T_{\mathbf{H}}X)(v) = \sum_{i\in\mathbb{Z}\setminus\{0\}} h(\lambda_i(T_{\mathbf{W}}))\hat{X}(\lambda_i)\varphi_i(v).$$

If $\{\mathbf{G}_n\} \to \mathbf{W}$ and $h : [0, 1] \to \mathbb{R}$ is continuous, then

$$\lim_{n\to\infty} \hat{\mathbf{H}}_n(\lambda_i(\mathbf{S}_n)/n) = \hat{T}_{\mathbf{H}}(\lambda_i(T_{\mathbf{W}}))$$

where $\hat{\mathbf{H}}_n$ and $\hat{T}_{\mathbf{H}}$ are the frequency responses of $\mathbf{H}_n$ and $T_{\mathbf{H}}$ respectively.

Proof: This is a direct consequence of a result due to [36, Theorem 6.7] and restated here as Lemma 4.

Lemma 4 (Eigenvalue convergence): Let $\{\mathbf{G}_n\}$ be a sequence of graphs with eigenvalues $\{\lambda_i(\mathbf{S}_n)\}_{i\in\mathbb{Z}\setminus\{0\}}$, and $\mathbf{W}$ a graphon with eigenvalues $\{\lambda_i(T_{\mathbf{W}})\}_{i\in\mathbb{Z}\setminus\{0\}}$. If $\{\mathbf{G}_n\} \to \mathbf{W}$,

$$\lim_{n\to\infty} \frac{\lambda_i(\mathbf{S}_n)}{n} = \lim_{n\to\infty} \lambda_i(T_{\mathbf{W}_{\mathbf{G}_n}}) = \lambda_i(T_{\mathbf{W}}) \text{ for all } i. \qquad (20)$$

Proof: Refer to Appendix B. ■

Lemma 4 tells us that, in any convergent graph sequence, the eigenvalues of the graph converge to the eigenvalues of the limit graphon under a $1/n$ rescaling. This is illustrated in Fig. 4 for $\lambda > 0$. Using Lemma 4, we will show that the transfer functions of arbitrary graph filters $\mathbf{H}_n(\mathbf{S}_n)$ converge to the transfer function of the graphon filter $T_{\mathbf{H}}$ with same filter function $h(\lambda)$.

Consider the graphon signal $(\mathbf{W}, X)$. Applying $T_{\mathbf{H}}$ to $(\mathbf{W}, X)$ as in (15), we get

$$Y(v) = \sum_{i\in\mathbb{Z}\setminus\{0\}} h(\lambda_i)\hat{X}(\lambda_i)\varphi_i(v) \qquad (21)$$

where we have omitted the dependence on $T_{\mathbf{W}}$ by writing $\lambda_i = \lambda_i(T_{\mathbf{W}})$. The WFT of $(\mathbf{W}, Y)$ is given by

$$\begin{aligned}[\hat{Y}]_j &= \int_0^1 Y(v)\varphi_j(v)dv \\ &= \sum_{i\in\mathbb{Z}\setminus\{0\}} h(\lambda_i)\left(\int_0^1 \varphi_i(v)\varphi_j(v)dv\right)\hat{X}(\lambda_i) \\ &= h(\lambda_j)\hat{X}(\lambda_j)\end{aligned}$$

from which we conclude that $\hat{T}_{\mathbf{H}}(\lambda_j) = h(\lambda_j)$.

We now determine the frequency response of $\mathbf{H}_n(\mathbf{S}_n)$. Applying $\mathbf{H}_n(\mathbf{S}_n)$ to the graph signal $(\mathbf{G}_n, \mathbf{x}_n)$, we get

$$\begin{aligned}\mathbf{y}_n = \mathbf{H}_n(\mathbf{S}_n)\mathbf{x}_n &= \mathbf{V}_n h(\mathbf{\Lambda}(\mathbf{S}_n)/n)\mathbf{V}_n^{\mathsf{H}}\mathbf{x}_n \\ &= \mathbf{V}_n h(\mathbf{\Lambda}(\mathbf{S}_n)/n)\hat{\mathbf{x}}_n.\end{aligned}$$

The GFT of $(\mathbf{G}_n, \mathbf{y}_n)$ is given by

$$[\hat{\mathbf{y}}_n]_j = \mathbf{v}_{nj}^{\mathsf{H}}\mathbf{V}_n h(\mathbf{\Lambda}(\mathbf{S}_n)/n)\hat{\mathbf{x}}_n = h(\lambda_j(\mathbf{S}_n)/n)[\hat{\mathbf{x}}_n]_j$$

and therefore $\hat{\mathbf{H}}_n(\lambda_j(\mathbf{S}_n)) = h(\lambda_j(\mathbf{S}_n)/n)$.

Since $\mathbf{G}_n \to \mathbf{W}$, from Lemma 4 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. □

The spectral or *frequency response* of a graph filter thus converges to that of the corresponding graphon filter provided that the frequency response of the filter, h, is continuous. In particular, this is the case for polynomials, so that the frequency response induced by LSI graph filters (12) converges to that of their corresponding graphon filters (18). To understand the importance of this result, suppose that we design a filter with a certain spectral behavior on the graphon; Theorem 2 tells us that the same spectral behavior can be expected from the application of this filter (or, more precisely, of the graph filter with same coefficients) on graphs sampled from the graphon. By a simple triangle inequality argument, it then follows that filters can be transferred between graphs associated with the same graphon to obtain a similar spectral behavior. This is the first evidence of graph filter *transferability*.

Nevertheless, Theorem 2 has a limitation. It only gives account of the limit behavior of the graph filter response in the frequency domain, while graph filters operate in the node domain. To analyze the vertex domain behavior of graph filters, we start with the simple case of bandlimited signals. Putting together Theorems 1 and 2, we can show that, when the limit of the graph signals is bandlimited, the graph filter outputs converge in the vertex domain.

Corollary 1 (Convergence of graph filter response for bandlimited graphon signals): Let $\{\mathbf{G}_n\}$ be a sequence of graphs converging to the graphon $\mathbf{W}$, where $\mathbf{W}$ is non-derogatory. Let $\mathbf{H}_n(\mathbf{S}_n) = \mathbf{V}_n h(\mathbf{\Lambda}(\mathbf{S}_n)/n)\mathbf{V}_n^{\mathsf{H}}$ be a sequence of filters on the graphs $\{\mathbf{G}_n\}$, and let $(T_{\mathbf{H}}X)(v) = \sum_{i\in\mathbb{Z}\setminus\{0\}} h(\lambda_i)\hat{X}(\lambda_i)\varphi_i(v)$ be a filter on the graphon $\mathbf{W}$. Consider the sequence of graph signals $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ and let $\{\pi_n\}$ be a sequence of permutations such that $\{(\mathbf{G}_n, \mathbf{x}_n)\} \to (\mathbf{W}, X)$ in the sense of Definition 2. Then, $\mathbf{y}_n = \mathbf{H}(\pi_n(\mathbf{S}_n))\pi_n(\mathbf{x}_n)$ converges to $Y = T_{\mathbf{H}}X$.

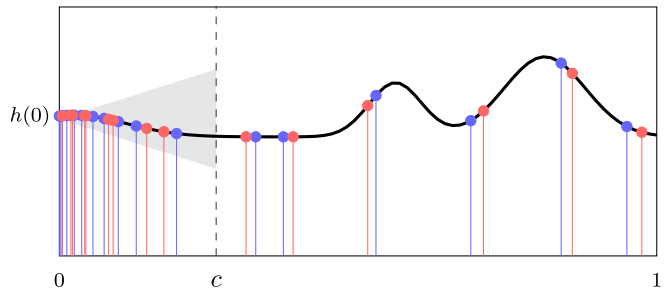


Fig. 5. Lipschitz continuous filter function $h(\lambda)$ with Lipschitz constant L. Only the positive eigenvalue axis is depicted for simplicity. Lipschitz filters eliminate the requirement that the graphon signal be bandlimited because they bound the filter variation for signal components associated with eigenvalues smaller than c.

Proof: Without loss of generality, assume that the permutation sequence $\{\pi_n\}$ is a sequence of identity permutations, i.e., the labeling of the signals $\mathbf{x}_n$ already matches the labeling for which the graphs converge in cut norm. Let the WFT of $(\mathbf{W}, Y)$ be $[\hat{Y}]_i = \hat{T}_{\mathbf{H}}(\lambda_i)[\hat{X}]_i$ and the GFT of $(\mathbf{G}_n, \mathbf{y}_n)$ be $[\hat{\mathbf{y}}_n]_i = \hat{\mathbf{H}}_n(\lambda_i(\mathbf{S}_n)/n)[\hat{\mathbf{x}}_n]_i$. By Theorem 1, $[\hat{\mathbf{x}}_n]_i \to [\hat{X}]_i$, and, by Theorem 2, $\hat{\mathbf{H}}_n(\lambda_i(\mathbf{S}_n)/n) \to \hat{T}_{\mathbf{H}}(\lambda_i)$. Since $\hat{T}_{\mathbf{H}}$ is a linear operator, and hence continuous, $[\hat{\mathbf{y}}_n]_i \to [\hat{Y}]_i$. Applying Theorem 1 once again for the iGFT, we conclude that $\mathbf{y}_n \to Y$.■

Cor. 1 extends upon Theorem 2 by showing that, provided that the sequence of input signals $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ converges to a bandlimited graphon signal, the output signals obtained by applying the filters $\mathbf{H}_n(\mathbf{S})$ converge in the same sense as $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ in the vertex domain. The requirement that the graphon signal be bandlimited arises from the difficulty of matching the GFT and WFT coefficients associated with small eigenvalues, i.e., eigenvalues λ_i for which $|i|$ is large. This is illustrated in Fig. 4. Note that, as the eigenvalues approach 0, it becomes hard to tell which graph eigenvalue converges to which graphon eigenvalue, as the eigenvalue difference $\lambda_i - \lambda_{i+1}$ tends to zero as $i \to \infty$.

This requirement can be eliminated by considering *Lipschitz graph and graphon filters*, i.e., filters with Lipschitz continuous $h(\lambda)$. A function $h : [0, 1] \to \mathbb{R}$ is L-Lipschitz continuous if, for all $\lambda, \lambda' \in [0, 1]$,

$$|h(\lambda) - h(\lambda')| \leq L|\lambda - \lambda'|. \tag{22}$$

For h differentiable, this is equivalent to bounding $dh/d\lambda$ by L in absolute value. An example of Lipschitz continuous filter is shown in Fig. 5. For filter functions h satisfying (22), we can show that the graph filters converge in the vertex domain for *any* graphon signal, not only bandlimited ones, because the variation of Lipschitz filters can be bounded close to zero [cf. Fig. 5].

Theorem 3 (Convergence of filter response for Lipschitz continuous graph filters): Let $\{\mathbf{G}_n\}$ be a sequence of graphs converging to the graphon $\mathbf{W}$, where $\mathbf{W}$ is non-derogatory. Let $\mathbf{H}_n(\mathbf{S}_n) = \mathbf{V}_n h(\mathbf{\Lambda}(\mathbf{S}_n)/n)\mathbf{V}_n^{\mathsf{H}}$ be a sequence of filters on the graphs $\{\mathbf{G}_n\}$, and let $(T_{\mathbf{H}}X)(v) = \sum_{i\in\mathbb{Z}\setminus\{0\}} h(\lambda_i)\hat{X}(\lambda_i)\varphi_i(v)$ be a filter on the graphon $\mathbf{W}$. Consider a sequence of graph signals $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ and let $\{\pi_n\} \in \mathcal{P}$ be a sequence of permutations such that $\{(\mathbf{G}_n, \mathbf{x}_n)\} \to (\mathbf{W}, X)$

in the sense of Definition 2. Then, $\mathbf{y}_n = \mathbf{H}(\pi_n(\mathbf{S}_n))\pi_n(\mathbf{x}_n)$ converges to $Y = T_{\mathbf{H}}X$.

Proof: To prove convergence of the $(\mathbf{G}_n, \mathbf{y}_n)$ to $(\mathbf{W}, Y)$, we consider the graphon signals $(\mathbf{W}_{\pi_n(\mathbf{G}_n)}, X_{\pi_n(\mathbf{G}_n)})$ induced by the graph signals $(\pi_n(\mathbf{G}_n), \pi_n(\mathbf{x}_n))$. The spectral properties of these signals and of the corresponding graph signals are related through Lemma 2. To simplify notation, we once again leave the dependence on $\pi_n(\mathbf{G}_n)$ implicit and write $\mathbf{W}_n = \mathbf{W}_{\pi_n(\mathbf{G}_n)}$ and $X_n = X_{\pi_n(\mathbf{G}_n)}$. We also denote the induced graphon eigenvalues $\lambda_i^n = \lambda_i(T_{\mathbf{W}_n})$. Recall that these are given by $\lambda_i^n = \lambda_i(\mathbf{S}_n)/n$ per Lemma 2. Without loss of generality, consider the normalized filter function $\bar{h}(\lambda) = h(\lambda)/\sup_{\lambda\in[0,1]}|h(\lambda)|$. The signal $(\mathbf{W}, Y)$ obtained by applying $T_{\bar{\mathbf{H}}}$ to $(\mathbf{W}, X)$ can be written as

$$Y(v) = \sum_{i\in\mathbb{Z}\setminus\{0\}} \bar{h}(\lambda_i)\hat{X}(\lambda_i)\varphi_i(v) \tag{23}$$

and $(\mathbf{W}_n, Y_n)$, which is induced by $\mathbf{y}_n = \bar{\mathbf{H}}(\mathbf{S}_n)\mathbf{x}_n$, as

$$Y_n(v) = \sum_{i\in\mathbb{Z}\setminus\{0\}} \bar{h}(\lambda_i^n)\hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n})(v). \tag{24}$$

The dependence of the eigenfunctions $\varphi_i(T_{\mathbf{W}_n})$ on $T_{\mathbf{W}_n}$ is made explicit to distinguish them from φ_i, the eigenvalues of $T_{\mathbf{W}}$.

To show that the $(\mathbf{W}_n, Y_n)$ converge to $(\mathbf{W}, Y)$, we start by writing their norm difference using (23) and (24),

$$\|Y - Y_n\| = \left\| \sum_{i\in\mathbb{Z}\setminus\{0\}} \bar{h}(\lambda_i)\hat{X}(\lambda_i)\varphi_i - \sum_{i\in\mathbb{Z}\setminus\{0\}} \bar{h}(\lambda_i^n)\hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\|. \tag{25}$$

Defining the set $\mathcal{C} = \{i \mid |\lambda_i| \geq c\}$ for $c = (1 - |\bar{h}_0|)/L(2\|X\|\epsilon^{-1} + 1)$ with $\epsilon > 0$ and $\bar{h}_0 = \bar{h}(0)$, these sums can be split up between $i \in \mathcal{C}$ and $i \notin \mathcal{C}$ to yield

$$\begin{aligned} &\left\| \sum_{i\in\mathbb{Z}\setminus\{0\}} \bar{h}(\lambda_i)\hat{X}(\lambda_i)\varphi_i - \sum_{i\in\mathbb{Z}\setminus\{0\}} \bar{h}(\lambda_i^n)\hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\| \\ &\leq \left\| \sum_{i\in\mathcal{C}} \bar{h}(\lambda_i)\hat{X}(\lambda_i)\varphi_i - \sum_{i\in\mathcal{C}} \bar{h}(\lambda_i^n)\hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\| \ \textbf{(i)} \\ &+ \left\| \sum_{i\notin\mathcal{C}} \bar{h}(\lambda_i)\hat{X}(\lambda_i)\varphi_i - \sum_{i\notin\mathcal{C}} \bar{h}(\lambda_i^n)\hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\| \ \textbf{(ii)}. \end{aligned} \tag{26}$$

Note that **(i)** corresponds to the difference between two bandlimited graphon signals. By Cor. 1, there exists n_0 such that, for all $n > n_0$,

$$\left\| \sum_{i\in\mathcal{C}} \bar{h}(\lambda_i)\hat{X}(\lambda_i)\varphi_i - \sum_{i\in\mathcal{C}} \bar{h}(\lambda_i^n)\hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\| < \epsilon. \tag{27}$$

For **(ii)**, we use the filter's Lipschitz property and Cauchy-Schwarz to write

$$\begin{aligned} &\left\| \sum_{i\notin\mathcal{C}} \bar{h}(\lambda_i)\hat{X}(\lambda_i)\varphi_i - \sum_{i\notin\mathcal{C}} \bar{h}(\lambda_i^n)\hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\| \\ &\leq \left\| \sum_{i\notin\mathcal{C}} (\bar{h}_0 + Lc)\hat{X}(\lambda_i)\varphi_i - \sum_{i\notin\mathcal{C}} (\bar{h}_0 - Lc)\hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\| \\ &\leq |\bar{h}_0| \left\| \sum_{i\notin\mathcal{C}} \left[\hat{X}(\lambda_i)\varphi_i - \hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right] \right\| \\ &+ Lc \left\| \sum_{i\notin\mathcal{C}} \hat{X}(\lambda_i)\varphi_i \right\| + Lc \left\| \sum_{i\notin\mathcal{C}} \hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\| \end{aligned} \tag{28}$$

where the last inequality follows from the triangle inequality.

Because $\{\varphi_i\}$ and $\{\varphi_i(T_{\mathbf{W}_n})\}$ form complete bases of L^2, $\sum_{i\notin\mathcal{C}} \hat{X}(\lambda_i)\varphi_i$ and $\sum_{i\notin\mathcal{C}} \hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n})$ can be written as

$$\sum_{i\notin\mathcal{C}} \hat{X}(\lambda_i)\varphi_i = X - \sum_{i\in\mathcal{C}} \hat{X}(\lambda_i)\varphi_i \quad \text{and} \tag{29}$$

$$\sum_{i\notin\mathcal{C}} \hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) = X_n - \sum_{i\in\mathcal{C}} \hat{X}(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \tag{30}$$

i.e., as the difference between the input signal and a bandlimited signal. Using these identities and the triangle inequality, we leverage the fact that $X_n \to X$ in L^2 and apply Theorem 1 to show that there exists n_1 such that, for all $n > n_1$,

$$\begin{aligned} &\left\| \sum_{i\notin\mathcal{C}} \hat{X}(\lambda_i)\varphi_i - \hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\| \\ &\leq \|X - X_n\| + \left\| \sum_{i\in\mathcal{C}} \hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) - \hat{X}(\lambda_i)\varphi_i \right\| < \epsilon. \end{aligned} \tag{31}$$

As for $\|\sum_{i\notin\mathcal{C}} \hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n})\|$, we can use the identities in (29) and (30) together with the triangle inequality to write

$$\begin{aligned} \left\| \sum_{i\notin\mathcal{C}} \hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\| &\leq \|X_n - X\| + \left\| \sum_{i\notin\mathcal{C}} \hat{X}(\lambda_i)\varphi_i \right\| \\ &+ \left\| \sum_{i\in\mathcal{C}} \hat{X}(\lambda_i)\varphi_i - \sum_{i\in\mathcal{C}} \hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\|. \end{aligned}$$

From Theorem 1 and the fact that $X_n \to X$ in L^2,

$$\left\| \sum_{i\notin\mathcal{C}} \hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\| \leq \epsilon + \left\| \sum_{i\notin\mathcal{C}} \hat{X}(\lambda_i)\varphi_i \right\| \text{ for } n > n_1. \tag{32}$$

Applying the Cauchy-Schwarz and triangle inequalities and substituting (31) and (32) in (28), we arrive at a bound for **(ii)**,

$$\left\| \sum_{i\notin\mathcal{C}} \bar{h}(\lambda_i)\hat{X}(\lambda_i)\varphi_i - \sum_{i\notin\mathcal{C}} \bar{h}(\lambda_i^n)\hat{X}_n(\lambda_i^n)\varphi_i(T_{\mathbf{W}_n}) \right\|$$

$$\leq (|\bar{h}_0| + Lc)\epsilon + 2Lc \left\| \sum_{i \notin \mathcal{C}} \hat{X}(\lambda_i)\varphi_i \right\|$$

$$\leq (|\bar{h}_0| + Lc)\epsilon + 2Lc\|X\| = \epsilon. \tag{33}$$

Putting (27) and (33) together, we have thus proved that for all $n > \max\{n_0, n_1\}$, $\|Y - Y_n\| < 2\epsilon$, i.e., the output of $\bar{\mathbf{H}}(\mathbf{S}_n)$ converges to the output of $T_{\bar{\mathbf{H}}}$ in the vertex domain. ■

Theorem 3 broadens the scope of Cor. 1 by extending the filter response convergence result to sequences of graph signals converging to *generic* finite energy graphon signals that are not necessarily bandlimited. The Lipschitz condition on the filter h allows bounding the variability of the filter response for signal components associated with eigenvalues smaller than some threshold $c \in [0, 1]$, which can be made arbitrarily small [cf. Fig. 5].

Theorem 3 can be further generalized to *any graphon* as opposed to only non-derogatory ones. The difference in the case of derogatory graphons is that the WFT cannot be defined, so Theorem 1 cannot be used in the proof of Theorem 4. The proof argument needed in this case is therefore slightly different. However, this is extenuated by Prop. 2. As long as eigengaps between adjacent graphon eigenspaces can be defined, this proposition ensures convergence not only of the eigenvectors, but also of the finite-dimensional eigenspaces associated with the repeated eigenvalues of an arbitrary graphon.

Proposition 2 (Graphon subspace convergence): Let $\{\mathbf{G}_n\}$ be a sequence of graphs with eigenvalues $\lambda_i(\mathbf{S}_n)$ converging to the graphon $\mathbf{W}$ with eigenvalues λ_i. If a given λ_i has multiplicity m_i and $\lambda^n_{i_k} = \lambda_{i_k}(\mathbf{S}_n)/n$ are the eigenvalues of $\mathbf{W}_{\mathbf{G}_n}$ (i.e., of the graphon induced by $\mathbf{G}_n$) converging to λ_i [cf. Lemma 2], then there exists a sequence of permutations $\{\pi_n\} \in \mathcal{P}$ such that

$$\left|\left| E_{T_{\mathbf{W}_{\pi_n(\mathbf{G}_n)}}}(\{\lambda^n_{i_k}\}) - E_{T_{\mathbf{W}}}(\lambda_i) \right|\right| \to 0$$

where $\mathcal{P}$ is the set of admissible permutation sequences for the sequence $\{\mathbf{G}_n\}$ (Definition 1) and $E_T(\Lambda)$ is the projection operator onto the subspace associated with the eigenvalues in the set Λ of the operator T.

Proof: Refer to Appendix E in the supplementary material. ■

With Prop. 2, we are now equipped to state our most general result: vertex domain convergence of Lipschitz continuous graph filters for graph sequences converging to arbitrary graphons. This result is presented in Theorem 4. We defer the proof to the appendices.

Theorem 4 (Convergence of filter response for Lipschitz continuous graph filters): Let $\{\mathbf{G}_n\}$ be a sequence of graphs converging to the graphon $\mathbf{W}$. Let $\mathbf{H}_n(\mathbf{S}_n) = \mathbf{V}_n h(\mathbf{\Lambda}(\mathbf{S}_n)/n)\mathbf{V}_n^{\mathsf{H}}$ be a sequence of filters on the graphs $\{\mathbf{G}_n\}$, and let $(T_{\mathbf{H}}X)(v) = \sum_{i \in \mathbb{Z}\setminus\{0\}} h(\lambda_i)\hat{X}(\lambda_i)\varphi_i(v)$ be a filter on the graphon $\mathbf{W}$. Consider a sequence of graph signals $\{(\mathbf{G}_n, \mathbf{x}_n)\}$ and let $\{\pi_n\} \in \mathcal{P}$ be a sequence of permutations such that $\{(\mathbf{G}_n, \mathbf{x}_n)\} \to (\mathbf{W}, X)$ in the sense of Definition 2. Then, $\mathbf{y}_n = \mathbf{H}(\pi_n(\mathbf{S}_n))\pi_n(\mathbf{x}_n)$ converges to $Y = T_{\mathbf{H}}X$.

Proof: Refer to Appendix F in the supplementary material. ■

The main takeaway from Theorems 3 and 4 is that, if the limit graphon is known, we can trade the design of multiple filters in different graphs by the centralized design of a single graphon filter from which graph filters can then be sampled. In practice, a more relevant implication of these theorems is that graph filters can be *transferred* across graphs associated with the same graphon. The ability to transfer graph filters is especially important when graphs are large or dynamic, as the operations involved in designing filters for these graphs can come out costly. This property is also inherited by graph neural networks (GNNs) based on these graph filters [37]. Transferability of GNNs has been demonstrated empirically in a number of applications [29], [30], and is formally characterized in [31], where transferability bounds are derived for both GNNs and graph filters. Transferability of graph filters will also be illustrated in the numerical experiments of Section V.

Remark 2: Note that, while the results presented in Theorems 1–4 may appear intuitive, their proofs are not. For instance, our Fourier convergence theorem (Theorem 1) requires that the graph and graphon signals be bandlimited for the GFT to converge to the WFT. This is in constrast to classical signal processing, where for any convergent sequence of length-n discrete time signals on [0,1] the discrete Fourier transform (DFT) converges to the Fourier transform (FT) regardless of the underlying spectral properties. This occurs because the regular line graphs underlying these signals have spectra that are evenly distributed on $[-1, 1]$ and therefore never accumulate around zero. Unexpectedly, however, these conditions are not needed to show convergence of graph filter outputs. Indeed, while one would expect that graph filter outputs converge only for bandlimited signals, this is not the case in Theorems 3–4. Instead, these theorems require the filter to be Lipschitz for $|\lambda| < c$ [cf. Fig. 5]. This arises from the fact that, for small λ, the graph eigenspaces can become hard to match to the corresponding graphon eigenspaces since the eigenvalues of the latter accumulate near zero. We can therefore replace bandlimitedness by a filter regularity condition.

V. Numerical Experiments

In this section, we present three numerical experiments to illustrate the results of Theorems 1 through 4. In the first, we sample graph signals from a Gaussian Markov Random Field (GMRF) on ER, SBM and random geometric sensor networks and compare the output of a graph diffusion process as the number of sensors increases. In the second, we compare the WFT of pollutant dispersion signals drawn from the same model on two geometric graphs corresponding to pollution sensor networks in different cities. Finally, in Section V-C a linear graph filter is optimized to predict movie ratings on a small user network and is then applied to a large one.

A. GMRF Diffusion (*S1*)

In this experiment, we simulate a GMRF measured and diffused on different sensor networks to analyze convergence of the filter $\mathbf{H}(\mathbf{S}) = \mathbf{S}$ in networks of growing size. 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

TABLE II
EXPRESSION OF $\mathbf{W}(u_i, u_j)$ FOR THE DIFFERENT GRAPHON MODELS IN SECTION V-A

Model	Expression of $\mathbf{W}(u_i, u_j)$
ER	$= 0.4$ for all u_i, u_j
SBM	$= \begin{cases} 0.8, & \text{if } u_i, u_j \leq 0.5 \text{ or } u_i, u_j > 0.5 \\ 0.2, & \text{otherwise} \end{cases}$
Geom.	$= \exp(-\beta(u_i - u_j)^2)$, $\beta = 2.3$

by [40]

$$\mathbf{\Sigma_x} = |a_0|^2(\mathbf{I} - a\mathbf{S})^{-1}[(\mathbf{I} - a\mathbf{S})^{-1}]^{\mathsf{H}} \tag{34}$$

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 graphons are considered. They are an Erdös-Rényi (ER) [cf. Fig. 1(a)], a stochastic block model (SBM) [cf. Fig. 2(a)], and a soft random geometric graph [cf. Fig. 2(c)]. Their expressions are presented in Table II.

To compare the diffusion outcomes of graph and graphon signals, we first need to define a graphon signal equivalent of the GMRF. We work with its approximation, which is obtained by approximating the graphons as matrices $\mathbf{S_W}$. These matrices are calculated by evaluating $\mathbf{W}(u_i, u_j)$ on $10^4 \times 10^4$ regularly spaced points of the unit square. Then, the graphon GMRF is obtained by sampling $\mathbf{x_W} \in \mathbb{R}^{10^4}$ from the zero-mean multivariate Gaussian with covariance matrix given by (34) for $\mathbf{S} = \mathbf{S_W}$.

In order to observe convergence, we 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 the edges (i, j) are Bernoulli random variables with success probability $\mathbf{W}(u_i, u_j)$, i.e, $[\mathbf{S}_n]_{ij} = [\mathbf{S}_n]_{ji} \sim \text{Bernoulli}(\mathbf{W}(u_i, u_j))$. The graph signals $\mathbf{x}_n$ are obtained by interpolating $\mathbf{x_W}$ at each u_i.

We calculate the diffused graph signals $\mathbf{y}_n = \mathbf{S}_n\mathbf{x}_n$ and interpolate the approximation of the diffused graphon signal $\mathbf{y_W} = \mathbf{S_W}\mathbf{x_W}$ at $\{u_i\}_{i=1}^n$, then compare them by computing their norm difference for increasing values of n. The average normalized norm difference is shown in Fig. 6 for 100 realizations of the graphon GMRF $\mathbf{x_W}$. 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, as expected from Theorem 3.

B. Spectral Analysis of Air Pollution on Sensor Networks (S2)

The objective of this experiment is to compare the spectral representations of air pollution signals collected at the nodes of two distinct sensor networks of same size to illustrate GFT convergence (Theorem 1). This problem can be interpreted as comparing the spectra of graph pollution data in two cities, for instance, New York and Philadelphia. The air pollution sensor networks are modeled as soft random geometric graphs [38] where, given nodes i and j and their coordinates (x_i, y_i) and (x_j, y_j), the probability of connecting i and j is

$$p(i, j) \propto \exp\left(-\beta\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\right). \tag{35}$$

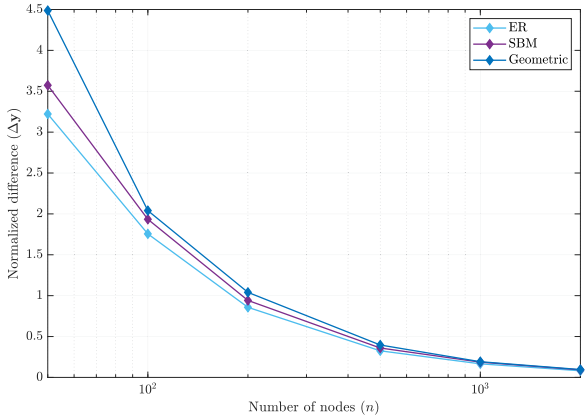


Fig. 6. Norm difference between GMRF graphon signals diffused on ER, SBM and geometric graphons and the corresponding graph signals diffused on sample graphs of increasing size. The diffusion outputs have been normalized by n.

Fixing the x coordinate at $x_i = x_j = x$ and normalizing y as $u = y/y_{\text{max}}$, we can rewrite $p(i, j)$ to fit the expression of the graphon $\mathbf{W}(u_i, u_j) = \exp(-\beta\sqrt{(u_i - u_j)^2})$.

In the cross-wind direction and at fixed altitude, the simplest model for air pollution dispersion is a Gaussian on the distance to the source of pollution in the cross-wind direction. Having fixed x, we assume the cross-wind direction to be y. The air pollution dispersion model is then

$$\mathbf{s}(y) \propto \exp\left(-\frac{(y - y_{\text{source}})^2}{2\sigma^2}\right),$$

where $s(y)$ is the concentration of pollutants at the coordinate y and the variance σ^2 represents the cross-wind mixing [39, Chapter 9]. If we assume $y_{\text{source}} = 0$ and once again normalize y as $u = y/y_{\text{max}}$, this dispersion model can be interpreted as a signal $X(u) \propto \exp(-u^2/2\sigma^2)$ on the graphon associated with the soft random geometric graph model of the sensor networks.

For multiple values of n and using coordinates $\{u_i^{(1)}\}_{i=1}^n$ and $\{u_i^{(2)}\}_{i=1}^n$ sampled uniformly at random from the unit line, we sample two distinct n-node graphs $\mathbf{G}_1$ and $\mathbf{G}_2$ from (35). In each of these graphs, the graph signals are the pollutant concentrations $[\mathbf{s}_1]_i = s(u_i^{(1)})$ and $[\mathbf{s}_2]_i = s(u_i^{(2)})$. We then compute the GFTs $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$, and sort them to find the minimum norm difference $\min \|\hat{\mathbf{s}}_1 - \hat{\mathbf{s}}_2\|$ over different permutations of the labels of these graphs. After repeating the experiment 50 times for each n in $n = 5, 10, 20, 50, 100, 200, 500, 1000$, we graph the 68%, 95% and 99.7% quantile curves of the GFT norm difference (normalized by $\|\hat{\mathbf{s}}_1\|$) in Fig. 7. All confidence intervals shrink consistently around the mean as n increases, indicating that the GFTs of the air pollution signals in $\mathbf{G}_1$ and $\mathbf{G}_2$ indeed converge as expected from Theorem 1.

C. Movie Rating Prediction Via User-Based Graph Filtering (S3)

Given U users and M movies, movie rating prediction consists of completing a $U \times M$ incomplete rating matrix by predicting the ratings users would give to movies that they have not yet

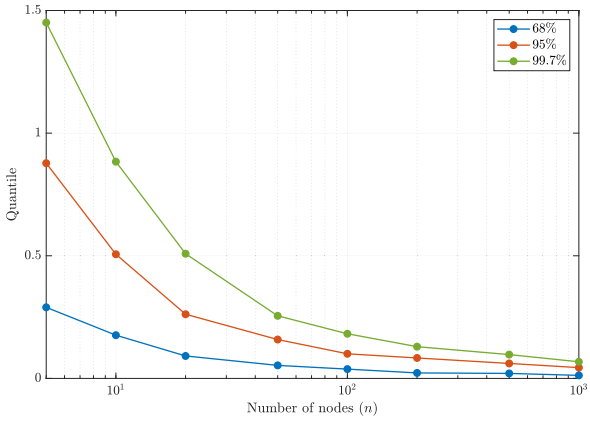


Fig. 7. Quantiles $(68\%, 95\%, 99.7\%)$ of the minimum normalized difference between GFTs of air pollution signals on graphs drawn from the same geometric model ($\mathbf{G}_1$ and $\mathbf{G}_2$) for $n = 5, 10, 20, 50, 100, 200, 500, 1000$, over 50 iterations for each n.

rated. We interpret this problem as a GSP problem by considering movie ratings (i.e., the columns of the rating matrix) to be graph signals on a network connecting similar users. A number of graph-based models for movie rating prediction have been proposed in the literature [13], [41], [42]. We consider one of the methods in [41], which completes the rating matrix by first solving an optimization problem to obtain the optimal coefficients of a linear graph filter, and then applying it to the graph signals corresponding to each movie's rating vector on the user network. Our objective is to calculate this graph filter in subnetworks corresponding to small cohorts of users, and observe how well it generalizes when applied to the full user network.

The dataset we use is the MovieLens 100 k dataset [43], which contains 100,000 ratings by $U = 943$ to $M = 1582$ movies. The user similarity network is built from the data by computing pairwise correlations from ratings given by each pair of users to movies that they both have rated and, then, keeping only the top-40 nearest neighbors to each user. Although these are networks built from real data, i.e, to which we cannot attribute a common generative model or graphon, the goal of this section is to illustrate how our results can be implicitly observed even in graphs that are not related by a common probability model, but that are "similar" in some other empirical or statistical sense. This is illustrated in Fig. 8, where user networks with 100 and 400 users are depicted. Even if the user network on the right has 4 times more users than the one on the left, we can see that the large-scale structure of these networks is similar.

The coefficients of filters with $K = 1$, 2 and 3 filter taps are optimized on networks of size $50, 100, 200, 400, 600, 800$ and 943 nodes. We then compare the RMSE obtained by predicting ratings using the filters calculated on the smaller networks and the filters calculated on the full user network. The relative RMSE differences and the base RMSE (obtained from the filter calculated on the full user network) are shown in Table III. For a network with n users, the reported RMSE difference corresponds to that of the average among filters trained on $\lfloor 943/n \rfloor$ different networks. Users were picked at random. We observe

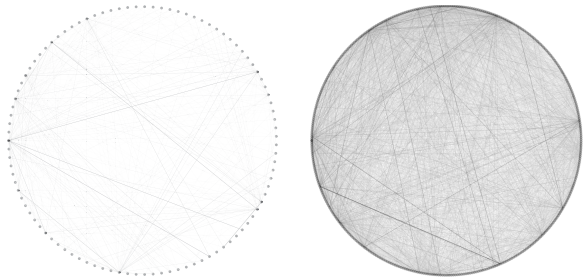


Fig. 8. User networks built from the ratings of 100 (left) ad 400 (right) users in the MovieLens 100 k dataset. The signals on these graphs correspond to the ratings given to the movie "Toy Story". The darker the node, the higher the rating, and the darker the edge, the higher the rating difference between the endnodes.

TABLE III
RELATIVE RMSE DIFFERENCE FOR RATING PREDICTION BASED ON $K = 1, 2, 3$ FILTERS OBTAINED ON $50, 100, 200, 400, 600$ AND 800-USER NETWORKS, WITH RESPECT TO THE BASE RMSE OF THE SAME FILTERS OBTAINED ON THE FULL 943-USER NETWORK

	Number of users						
K	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**

that, for all K, the RMSE difference gets steadily smaller as the network size increases. In particular, for $K = 1$ and $K = 3$ the relative RMSE difference is less than 1% for filters obtained on networks with under half the number of total users in the dataset.

VI. CONCLUSION

We have proposed a novel graphon signal processing framework which simplifies the analysis of signals and the design of filters on very large and dynamic networks. This framework introduces graphon signals, the graphon Fourier transform and LSI graphon filters. We have shown that graphon filters and the WFT are the limit objects of graph filters and of the GFT. These results justify transferring signal analysis methods and information processing systems from graphs to graphons or between graphs associated with the same graphon. GFT and graph filter convergence were demonstrated in two experiments involving graphs drawn from the same graphon, and, in a third experiment, we illustrate how graph filter behavior can be *transferred* even in situations where graphs are built from model-free data and can only empirically or statistically be said to belong to the same "class".

APPENDIX A
PROOF OF LEMMA 2

The proof follows by direct computation. For $j \in \mathcal{L}$,

$$(T_{\mathbf{W_G}}\varphi_j)(u) = \int_0^1 \mathbf{W_G}(u,v)\varphi_j(v)dv$$

$$= \sqrt{n}\mathbb{I}\left(u \in I_k\right) \int_0^1 [\mathbf{S}]_{k\ell}[\mathbf{v}_j]_k \times \mathbb{I}\left(v \in I_\ell\right) dv$$

$$= \sqrt{n}\mathbb{I}\left(u \in I_k\right) \sum_{\ell=1}^{n} [\mathbf{S}]_{k\ell} [\mathbf{v}_j]_k \int_{I_\ell} dv$$

$$= \frac{[\mathbf{S}\mathbf{v}_j]_k}{n} \times \sqrt{n}\mathbb{I}\left(u \in I_k\right)$$

$$= \frac{\lambda_j(\mathbf{S})}{n} \left[[\mathbf{v}_j]_k \times \sqrt{n}\mathbb{I}\left(u \in I_k\right) \right] = \lambda_j(T_{\mathbf{W_G}}) \varphi_j(u).$$

If $j \notin \mathcal{L}$, then $\langle \varphi_j, \varphi_k \rangle = 0$ for all $k \in \mathcal{L}$. In this case, we can trivially write $(T_{\mathbf{W_G}} \varphi_j)(u) = 0 = \lambda_j(T_{\mathbf{W_G}}) \varphi_j(u)$. Note that since the $\mathbf{v}_k$ are orthonormal, so are the $\{\varphi_k(T_{\mathbf{W_G}})\}$ and therefore a basis completion $\{\varphi_j\}$ can always be obtained. To conclude, compute for $j \in \mathcal{L}$

$$[\hat{X}_{\mathbf{G}}]_j = \int_0^1 \varphi_j(v) X_{\mathbf{G}}(v) dv$$

$$= \sqrt{n} \int_0^1 [\mathbf{v}_j]_\ell [\mathbf{x}]_\ell \times \mathbb{I}\left(v \in I_\ell\right) dv$$

$$= \sqrt{n} \sum_{\ell=1}^{n} [\mathbf{v}_j]_\ell [\mathbf{x}]_\ell \int_{I_\ell} dv = \frac{\mathbf{v}_j^{\mathsf{T}} \mathbf{x}}{\sqrt{n}} = \frac{[\hat{\mathbf{x}}]_j}{\sqrt{n}}.$$

If $j \notin \mathcal{L}$, recall that since the $\{\mathbf{v}_j\}$ form a basis of $\mathbb{R}^n$, we can write $\mathbf{x} = \sum_{k \in \mathcal{L}} c_k \mathbf{v}_k$. Hence,

$$[\hat{X}_{\mathbf{G}}]_j = \int_0^1 \varphi_j(v) X_{\mathbf{G}}(v) dv$$

$$= \int_0^1 [\mathbf{x}]_\ell \times \mathbb{I}\left(v \in I_\ell\right) \varphi_j(v) dv$$

$$= \int_0^1 \sum_{k \in \mathcal{L}} c_k [\mathbf{v}_k]_\ell \times \mathbb{I}\left(v \in I_\ell\right) \varphi_j(v) dv$$

$$= \frac{1}{\sqrt{n}} \sum_{k \in \mathcal{L}} c_k \int_0^1 \varphi_k(v) \varphi_j(v) dv = 0. \quad \square$$

APPENDIX B
PROOF OF LEMMA 3 AND LEMMA 4

To prove Lemma 3, we first repeat Lemma 4 below.

Lemma 4 (Eigenvalue convergence): 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, for all j

$$\lim_{n \to \infty} \frac{\lambda_j(\mathbf{S}_n)}{n} = \lim_{n \to \infty} \lambda_j(T_{\mathbf{W_{G}}_n}) = \lambda_j(T_{\mathbf{W}}). \tag{36}$$

Proof: The proof is essentially the one for [36, Theorem 6.7], but we reproduce it here using our notation. Recall that since the sequence $\{\mathbf{G}_n\}$ converges to $\mathbf{W}$, the density of homomorphisms for any motif also converges. The result then follows by choosing a homomorphism connected to the eigenvalues of their induced operators, namely the k-cycle $\mathbf{C}_k$. Indeed, notice that for any graphon $\mathbf{W}'$ and $k \geq 2$, we have, by definition, that $t(\mathbf{C}_k, \mathbf{W}') = \sum_{j \in \mathbb{Z} \setminus \{0\}} \lambda_j(T_{\mathbf{W}'})^k$. Hence,

$$\lim_{n \to \infty} \sum_{j \in \mathbb{Z} \setminus \{0\}} \lambda_j(T_{\mathbf{W}_n})^k = \sum_{j \in \mathbb{Z} \setminus \{0\}} \lambda_j(T_{\mathbf{W}})^k, \text{ for } k \geq 2 \tag{37}$$

where $T_{\mathbf{W}_n} = T_{\mathbf{W_{G}}_n}$. It now suffices to show that (37) implies $\lambda_j(T_{\mathbf{W}_n}) \to \lambda_j(T_{\mathbf{W}})$.

We start by bounding the eigenvalues of any graphon $\mathbf{W}'$ in terms of its density of homomorphisms. In particular, for $k = 4$ we obtain that

$$\sum_{j=1}^{m} \lambda_j(T_{\mathbf{W}'})^4 \leq \sum_{j \in \mathbb{Z} \setminus \{0\}} \lambda_j(T_{\mathbf{W}'})^4 = t(\mathbf{C}_4, \mathbf{W}') \Rightarrow$$

$$\lambda_m(T_{\mathbf{W}'}) \leq \left[\frac{t(\mathbf{C}_4, \mathbf{W}')}{m} \right]^{1/4} \text{ and}$$

$$\sum_{j=-m}^{-1} \lambda_j(T_{\mathbf{W}'})^4 \leq \sum_{j \in \mathbb{Z} \setminus \{0\}} \lambda_j(T_{\mathbf{W}'})^4 = t(\mathbf{C}_4, \mathbf{W}') \Rightarrow$$

$$\lambda_{-m}(T_{\mathbf{W}'}) \geq -\left[\frac{t(\mathbf{C}_4, \mathbf{W}')}{m} \right]^{1/4}.$$

Since $t(\mathbf{C}_4, \mathbf{W}_n)$ is a convergent sequence, it has a bound B [36], which implies that

$$|\lambda_j(T_{\mathbf{W}_n})| \leq \left(\frac{B}{|j|} \right)^{1/4}, \text{ for all } j \in \mathbb{Z} \setminus \{0\}. \tag{38}$$

Note that for $k \geq 5$, we can take the limit in (37) term-by-term since, as $|\lambda_j(T_{\mathbf{W}_n})^k| \leq (B/|j|)^{k/4}$ and the series $\sum_i (B/|j|)^{k/4}$ is convergent for $k > 4$, $\sum_{j \in \mathbb{Z} \setminus \{0\}} |\lambda_j(T_{\mathbf{W}_n})^k|$ also converges. Hence, from (37), we have

$$\lim_{n \to \infty} \sum_{j \in \mathbb{Z} \setminus \{0\}} \lambda_j(T_{\mathbf{W}_n})^k = \sum_{j \in \mathbb{Z} \setminus \{0\}} \zeta_j^k = \sum_{j \in \mathbb{Z} \setminus \{0\}} \lambda_j(T_{\mathbf{W}})^k \tag{39}$$

for $k \geq 5$, where $\zeta_j^k = \lim_{n \to \infty} \lambda_j(T_{\mathbf{W}_n})^k$.

To conclude, we proceed by induction over an ordering of the sequence of eigenvalues $\lambda_j(T_{\mathbf{W}})$, namely over j_ℓ, $\ell = 1, 2, \ldots$, such that $|\lambda_{j_1}(T_{\mathbf{W}})| \geq |\lambda_{j_2}(T_{\mathbf{W}})| \geq \cdots \geq |\lambda_{j_\ell}(T_{\mathbf{W}})|$. Suppose that $\zeta_{j_\ell} = \lambda_{j_\ell}(T_{\mathbf{W}})$ for $\ell < \ell^*$ and let $\lambda_{j_{\ell^*}}(T_{\mathbf{W}})$ be of multiplicity a and appear b times in the sequence $\{\zeta_j\}$ and $-\lambda_{j_{\ell^*}}(T_{\mathbf{W}})$ be of multiplicity a' and appear b' times in $\{\zeta_j\}$. The identity in (39) then reduces to

$$\left[b + (-1)^k b'\right] + \sum_{\ell > \ell^*} \left(\frac{\zeta_{j_\ell}}{\lambda_{j_{\ell^*}}(T_{\mathbf{W}})} \right)^k$$

$$= \left[a + (-1)^k a'\right] + \sum_{\ell > \ell^*} \left(\frac{\lambda_{j_\ell}(T_{\mathbf{W}})}{\lambda_{j_{\ell^*}}(T_{\mathbf{W}})} \right)^k, \text{ for } k \geq 5,$$

where we divided both sides by $\lambda_{j_{\ell^*}}(T_{\mathbf{W}})^k$. Due to the ordering of the λ_{j_ℓ}, for $k \to \infty$ through the even numbers we get $b + b' = a + a'$ and through the odd numbers we get $b - b' = a - a'$. Immediately, we have that $a = a'$ and $b = b'$, so that $\zeta_{j_{\ell^*}} = \lambda_{j_{\ell^*}}$. Although this argument assumes $\zeta_{j_\ell} < \lambda_{j_{\ell^*}}$ for all $\ell > \ell^*$, applying the same procedure to an ordering of the sequence $\{\zeta_j\}$ yields the same conclusion. ■

We will also require the following well known result about the perturbation of self-adjoint operators. For σ a subset of the eigenvalues of a self-adjoint operator T, define the spectral projection $E_T(\sigma)$ as the projection onto the subspace spanned by the eigenfunctions relative to the eigenvalues in σ.

Proposition 3: Let T and T' be two self-adjoint operators on a separable Hilbert space $\mathcal{H}$ whose spectra are partitioned as $\sigma \cup \Sigma$ and $\omega \cup \Omega$ respectively, with $\sigma \cap \Sigma = \emptyset$ and $\omega \cap \Omega = \emptyset$. If there exists $d > 0$ such that $\min_{x\in\sigma,\, y\in\Omega} |x - y| \geq d$ and $\min_{x\in\omega,\, y\in\Sigma} |x - y| \geq d$, then

$$|||E_T(\sigma) - E_{T'}(\omega)||| \leq \frac{\pi}{2} \frac{|||T - T'|||}{d} \tag{40}$$

Proof: See [44]. ∎

Lastly, we need two results related to the graphon norm. The first is Lemma 1, which states that if a sequence of graphs converges to a graphon in the homomorphism density sense, it also converges in the cut norm (7). The second, here presented as Prop. 4, is due to [26, Theorem 11.57] and bounds the L^2-induced norm of the graphon operator by is cut norm.

Proposition 4: Let $T_{\mathbf{W}}$ be the operator induced by the kernel $\mathbf{W}$. Then, $\|\mathbf{W}\|_\square \leq |||T_{\mathbf{W}}||| \leq \sqrt{8\|\mathbf{W}\|_\square}$.

This is a direct consequence of [45, Theorem 3.7(a)] and of the fact that $t(\mathbf{C}_2, \mathbf{W})$ is the Hilbert-Schmidt norm of $T_{\mathbf{W}}$, which dominates the L^2-induced operator norm.

We can now proceed with the proof of our lemma:

Proof of Lemma 3: For $j \in \mathcal{C}$, let $\sigma = \lambda_j(T_{\mathbf{W}})$, $\Sigma = \{\lambda_i(T_{\mathbf{W}})\}_{i\neq j}$, $\omega = \lambda_j(T_{\mathbf{W}_n})$, and $\Omega = \{\lambda_i(T_{\mathbf{W}_n})\}_{i\neq j}$ in Prop. 3 to get

$$|||E_j - E_{j,n}||| \leq \frac{\pi}{2} \frac{|||T_{\mathbf{W}_n} - T_{\mathbf{W}}|||}{d_{j,n}} \tag{41}$$

where E_j and $E_{j,n}$ are the spectral projections of $T_{\mathbf{W}}$ and $T_{\mathbf{W}_n}$ with respect to their j-th eigenvalue and

$$d_{j,n} = \min\left(|\lambda_j - \lambda_{j+1}(T_{\mathbf{W}_n})|, |\lambda_j - \lambda_{j-1}(T_{\mathbf{W}_n})|, \right.$$
$$\left. |\lambda_{j+1} - \lambda_j(T_{\mathbf{W}_n})|, |\lambda_{j-1} - \lambda_j(T_{\mathbf{W}_n})|\right),$$

where we omitted the dependence on $\mathbf{W}$ by writing $\lambda_j = \lambda_j(T_{\mathbf{W}})$.

Fix $\epsilon > 0$. From Lemma 4, we know we can find n_1 such that $|d_{j,n} - \delta_j| \leq \delta_j/2$ for all $n > n_1$, where

$$\delta_j = \min\left(|\lambda_j - \lambda_{j+1}|, |\lambda_j - \lambda_{j-1}|\right).$$

Since $\mathbf{W}$ is non-derogatory, $\delta_j > 0$. Additionally, the cut norm convergence of graphon sequences (Lemma 1) together with Prop. 4 implies there exists n_2 such that $|||T_{\mathbf{W}_n} - T_{\mathbf{W}}||| \leq \epsilon\delta_j/\pi$. Hence, for all $n > \max(n_1, n_2)$ it holds from (41) that

$$|||E_j - E_{j,n}||| \leq \frac{\pi}{2} \frac{\epsilon\delta_j/\pi}{\delta_j/2} = \epsilon. \tag{42}$$

Since ϵ is arbitrary, (42) proves that the projections onto the eigenfunctions of the same eigenvalue converge. I.e., the eigenfunction sequence $\varphi_j(T_{\mathbf{W}_n})$ itself converges weakly. Because the norms of the $\varphi_j(T_{\mathbf{W}_n})$ and $\varphi_j(T_{\mathbf{W}})$ are always equal to one, in this case weak convergence also implies strong convergence. To see this, note that $\|\varphi_j(T_{\mathbf{W}_n}) - \varphi_j(T_{\mathbf{W}})\|^2$ can be written as

$$\begin{aligned}
&\|\varphi_j(T_{\mathbf{W}_n}) - \varphi_j(T_{\mathbf{W}})\|^2 \\
&= \langle \varphi_j(T_{\mathbf{W}_n}) - \varphi_j(T_{\mathbf{W}}), \varphi_j(T_{\mathbf{W}_n}) - \varphi_j(T_{\mathbf{W}})\rangle \\
&= \langle \varphi_j(T_{\mathbf{W}_n}), \varphi_j(T_{\mathbf{W}_n}) - \varphi_j(T_{\mathbf{W}})\rangle \\
&\quad - \langle \varphi_j(T_{\mathbf{W}}), \varphi_j(T_{\mathbf{W}_n}) - \varphi_j(T_{\mathbf{W}})\rangle \\
&= \|\varphi_j(T_{\mathbf{W}_n})\|^2 - 2\langle \varphi_j(T_{\mathbf{W}_n}), \varphi_j(T_{\mathbf{W}})\rangle + \|\varphi_j(T_{\mathbf{W}})\|^2 \\
&\quad \to \|\varphi_j(T_{\mathbf{W}_n})\|^2 - 2\langle \varphi_j(T_{\mathbf{W}}), \varphi_j(T_{\mathbf{W}})\rangle + \|\varphi_j(T_{\mathbf{W}})\|^2 \\
&\quad = \|\varphi_j(T_{\mathbf{W}_n})\|^2 - \|\varphi_j(T_{\mathbf{W}})\|^2 = 1 - 1 = 0
\end{aligned}$$

where the sixth line follows from weak convergence of the $\varphi_j(T_{\mathbf{W}_n})$ to $\varphi_j(T_{\mathbf{W}})$.

To proceed, let us apply Prop. 3 to the subspace spanned by the remaining eigenfunctions with indices not in $\mathcal{C}$. Let $\sigma = \{\lambda_i(T_{\mathbf{W}})\}_{i\notin\mathcal{C}}$, $\Sigma = \{\lambda_i(T_{\mathbf{W}})\}_{i\in\mathcal{C}}$, $\omega = \{\lambda_i(T_{\mathbf{W}_n})\}_{i\notin\mathcal{C}}$, and $\Omega = \{\lambda_i(T_{\mathbf{W}_n})\}_{i\in\mathcal{C}}$ in (40) to get

$$|||E' - E'_n||| \leq \frac{\pi}{2} \frac{|||T_{\mathbf{W}_n} - T_{\mathbf{W}}|||}{d_n}, \tag{43}$$

where E' and E'_n are the projections onto the subspaces given by $\mathcal{S} = \text{span}(\{\varphi_i(T_{\mathbf{W}})\}_{i\notin\mathcal{C}})$ and $\mathcal{S}_n = \text{span}(\{\varphi_i(T_{\mathbf{W}_n})\}_{i\notin\mathcal{C}})$ respectively. From Prop. 3, the denominator d_n must satisfy $d_n \leq \min_{i\notin\mathcal{C}, j\in\mathcal{C}} |\lambda_i(T_{\mathbf{W}_n}) - \lambda_j(T_{\mathbf{W}})| = d^{(1)}$ and $d_n \leq \min_{i\notin\mathcal{C}, j\in\mathcal{C}} |\lambda_i(T_{\mathbf{W}}) - \lambda_j(T_{\mathbf{W}_n})| = d^{(2)}$. For $j \in \mathcal{C}$, we have $|\lambda_j(T_{\mathbf{W}})| \geq c$ and so $d^{(1)} \geq \min_{i\notin\mathcal{C}} c - |\lambda_i(T_{\mathbf{W}_n})|$. As for $d^{(2)}$, there exists n_0 such that $d^{(2)} \geq \min_{i\notin\mathcal{C}} c - |\lambda_i(T_{\mathbf{W}})|$ for $n > n_0$ because $\lambda_j(T_{\mathbf{W}_n}) \to \lambda_j(T_{\mathbf{W}})$ for all j from Lemma 4. Thus, for $n > n_0$ Prop. 3 holds with d_n given by

$$d_n \leq \min\left[\min_{i\notin\mathcal{C}} c - |\lambda_i(T_{\mathbf{W}_n})|,\ \min_{i\notin\mathcal{C}} c - |\lambda_i(T_{\mathbf{W}})|\right]$$

which is satisfied by $d_n = \inf_{i\notin\mathcal{C}} c - |\lambda_i(T_{\mathbf{W}_n})|$. Since the graphon $\mathbf{W}$ is non-derogatory, there exists an n_1 such that $d_n > 0$ for all $n > \max(n_0, n_1)$ and we can use the same argument as above to obtain that $E'_n \to E'$ in operator norm. The quantity d_n is illustrated in Fig. 3.

To see how this implies that for all $i \notin \mathcal{C}$ the function $\varphi_i(T_{\mathbf{W}_n})$ converges weakly to a function in the subspace $\mathcal{S}$—which we denote Ψ_i—, let $\Phi \in \mathcal{S}^\perp$. Then,

$$\begin{aligned}
|\langle \varphi_i(T_{\mathbf{W}_n}), \Phi\rangle| &= |\langle E'_n \varphi_i(T_{\mathbf{W}_n}), \Phi\rangle| \\
&= |\langle E'_n \varphi_i(T_{\mathbf{W}_n}), \Phi\rangle - \langle E' \varphi_i(T_{\mathbf{W}_n}), \Phi\rangle|
\end{aligned}$$

where the last equality holds because $\langle E' \varphi_i(T_{\mathbf{W}_n}), \Phi\rangle = 0$ due to $\Phi \in \mathcal{S}^\perp$. From the linearity of inner products, this can be rewritten as

$$\begin{aligned}
|\langle \varphi_i(T_{\mathbf{W}_n}), \Phi\rangle| &= |\langle E'_n \varphi_i(T_{\mathbf{W}_n}) - E' \varphi_i(T_{\mathbf{W}_n}), \Phi\rangle| \\
&= |\langle (E'_n - E') \varphi_i(T_{\mathbf{W}_n}), \Phi\rangle|
\end{aligned}$$

and, applying Cauchy-Schwarz,

$$|\langle \varphi_i(T_{\mathbf{W}_n}), \Phi\rangle| \leq \|E'_n - E'\| \|\Phi\|.$$

Taking the limit on both sides of the inequality, we get

$$\lim_{n\to\infty} |\langle \varphi_i(T_{\mathbf{W}_n}), \Phi\rangle| \leq \|\Phi\| \lim_{n\to\infty} \|E'_n - E'\| = 0.$$

Hence, $\varphi_i(T_{\mathbf{W}_n})$ converges weakly to a Ψ_i that is perpendicular to elements of $\mathcal{S}^{\perp}$, i.e., $\Psi_i \in \mathcal{S}$. ■

REFERENCES

[1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[2] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[3] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.

[4] 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.

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

[6] F. Gama, A. G. Marques, G. Leus, and A. Ribeiro, "Convolutional neural network architectures for signals supported on graphs," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1034–1049, Feb. 2019.

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

[8] S. Chen, R. Varma, A. Sandryhaila, and J. Kovacevic, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.

[9] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1832–1843, Apr. 2016.

[10] L. F. O. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 34–47, Jan. 2018.

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

[12] C. C. Paige, "The computation of eigenvalues and eigenvectors of very large sparse matrices," Ph.D. dissertation, Inst. Comput. Sci., Univ. London, 1971.

[13] L. Ruiz, F. Gama, G. Marques, and A. Ribeiro, "Invariance-preserving localized activation functions for graph neural networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 127–141, 2020, doi: 10.1109/TSP.2019.2955832.

[14] F. Gama, J. Bruna, and A. Ribeiro, "Stability properties of graph neural networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 5680–5695, 2020, doi: 10.1109/TSP.2020.3026980.

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

[16] E. M. Airoldi, T. B. Costa, and S. H. Chan, "Stochastic blockmodel approximation of a graphon: Theory and consistent estimation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 692–700.

[17] Y. De Castro, C. Lacour, and T. M. P. Ngoc, "Adaptive estimation of nonparametric geometric graphs," *Math. Stat. Learn*, vol. 2, no. 3, pp. 217–274, 2019.

[18] J. Xu, "Rates of convergence of spectral methods for graphon estimation," in *Proc. Mach. Learning Res.*, 2018, pp. 5433–5442.

[19] C. Gao, Y. Lu, and H. H. E. A. Zhou, "Rate-optimal graphon estimation," *Ann. Statist.*, vol. 43, no. 6, pp. 2624–2652, 2015.

[20] J. Xu, L. Massoulié, and M. Lelarge, "Edge label inference in generalized stochastic block models: From spectral theory to impossibility results," in *Proc. Conf. Learn. Theory*, 2014, pp. 903–920.

[21] S. Gao and P. E. Caines, "Graphon control of large-scale networks of linear systems," *IEEE Trans. Automat. Control*, vol. 65, no. 10, pp. 4090–4105, Oct. 2020, doi: 10.1109/TAC.2019.2955976.

[22] K. Rohe, S. Chatterjee, and B. E. A. Yu, "Spectral clustering and the high-dimensional stochastic blockmodel," *Ann. Statist.*, vol. 39, no. 4, pp. 1878–1915, 2011.

[23] P. Diao, D. Guillot, A. Khare, and B. Rajaratnam, "Model-free consistency of graph partitioning," 2016, *arXiv:1608.03860* [*math.CO*].

[24] M. Avella-Medina, F. Parise, M. Schaub, and S. Segarra, "Centrality measures for graphons: Accounting for uncertainty in networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 520–537, Jan.-Mar. 2020.

[25] F. Parise and A. Ozdaglar, "Graphon games," in *Proc. ACM Conf. Econ. Comput.*, 2019, pp. 457–458.

[26] L. Lovász, *Large Networks and Graph Limits*. Providence, RI, USA: American Mathematical Society, 2012.

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

[28] M. W. Morency and G. Leus, "Signal processing on kernel-based random graphs," in *Proc. Eur. Signal Process. Conf.*, 2017, pp. 365–369.

[29] M. Eisen and A. Ribeiro, "Optimal wireless resource allocation with random edge graph neural networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 2977–2991, 2020, doi: 10.1109/TSP.2020.2988255.

[30] E. Tolstaya, F. Gama, J. Paulos, G. Pappas, V. Kumar, and A. Ribeiro, "Learning decentralized controllers for robot swarms with graph neural networks," in *Proc. Conf. Robot Learn.*, 2020, pp. 671–682.

[31] L. Ruiz, F. Gama, and A. Ribeiro, "Graph neural networks: Architectures, stability and transferability," *Proc. IEEE*, vol. 109, no. 5, pp. 660–682, May 2021, doi: 10.1109/JPROC.2021.3055400.

[32] R. Levie, E. Isufi, and G. Leus Kutyniok, "On the transferability of spectral graph filters," in *Proc. 13th Int. conf. Sampling Theory and Appl.*, 2019, pp. 1–5, doi: 10.1109/SampTA45681.2019.9030932.

[33] D. Kerr and H. Li, "On Gromov-Hausdorff convergence for operator metric spaces," *J. Operator Theory*, vol. 62, no. 1, pp. 83–109, 2009.

[34] P. D. Lax, *Functional Analysis*. Ann Arbor, MI, USA: Wiley, 2002.

[35] R. Ramakrishna, H. Wai, and A. Scaglione, "A user guide to low-pass graph signal processing and its applications: Tools and applications," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 74–85, Nov. 2020.

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

[37] L. Ruiz, L. F. O. Chamon, and A. Ribeiro, "Graphon neural networks and the transferability of graph neural networks," in *Proc. Neural Inform. Process. Syst.*, Vancouver, Canada, Dec. 2020, pp. 1702–1712.

[38] M. D. Penrose, "Connectivity of soft random geometric graphs," *Ann. Appl. Probability*, vol. 6, no. 2, pp. 986–1028, 2016.

[39] S. P. E. A. Arya, *Air Pollution Meteorology and Dispersion*. New York, NY, USA:Oxford Univ. Press, 1999.

[40] F. Gama and A. Ribeiro, "Ergodicity in stationary graph processes: A weak law of large numbers," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2761–2774, May 2019.

[41] 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.

[42] F. Monti, M. M. Bronstein, and X. Bresson, "Geometric matrix completion with recurrent multi-graph neural networks," in *Proc. Neural Inform. Process. Syst.*, Long Beach, CA, USA, Dec. 2017, pp. 3700–3710.

[43] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *ACM Trans. Interactive Intell. Syst. -Regular Articles Special Issue New Directions Eye Gaze Interactive Intell. Syst.*, vol. 5, pp. 1–19, Jan. 2016.

[44] A. Seelmann, "Notes on the sin 2Θ theorem," *Integral Equations Operator Theory*, vol. 79, no. 4, pp. 579–597, 2014.

[45] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, "Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing," *Adv. Math.*, vol. 219, no. 6, pp. 1801–1851, 2008.