Graphon Filters

Graph Signal Processing in the Limit

Morency, Matthew; Leus, Geert

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Graphon Filters: Graph Signal Processing in the Limit

Matthew W. Morency , *Student Member, IEEE*, and Geert Leus , *Fellow, IEEE*

Abstract—Graph signal processing is an emerging field which aims to model processes that exist on the nodes of a network and are explained through diffusion over this structure. Graph signal processing works have heretofore assumed knowledge of the graph shift operator. Our approach is to investigate the question of graph filtering on a graph about which we only know a model. To do this we leverage the theory of graphons proposed by L. Lovasz and B. Szegedy. We make three key contributions to the emerging field of graph signal processing. We show first that filters defined over the scaled adjacency matrix of a random graph drawn from a graphon converge to filters defined over the Fredholm integral operator with the graphon as its kernel. Second, leveraging classical findings from the theory of the numerical solution of Fredholm integral equations, we define the Fourier-Galerkin shift operator. Lastly, using the Fourier-Galerkin shift operator, we derive a graph filter design algorithm which only depends on the graphon, and thus depends only on the probabilistic structure of the graph instead of the particular graph itself. The derived graphon filtering algorithm is verified through simulations on a variety of random graph models.

Index Terms—Graph signal processing, graph filter design, graphons, random graphs.

I. Introduction

GRAPH signal processing is an emerging topic which endeavours to explain the evolution of signals supported on an irregular domain, modeled as a graph [1]–[10]. The interconnections between nodes of the graph in some way capture interdependencies of data supported on the nodes. Such considerations have previously been made in the domain of machine learning, specifically in the domain of semi-supervised learning [11]–[13]. For example in [11] the network in particular is a weighted graph, the weights of which are similarities between the nodes. Tying these domains together is the challenge of processing data with an underlying structure, modeled by a graph. Such data can come from many different fields. This information is encoded in graph shift operators, which also capture operations that are performed on the node-supported data. For example, the action of the scaled adjacency matrix on the data supported on the nodes of the graph leads to a weighted sum of the values on the neighbours of every node.

As an irregular domain analogue of classical signal processing, one of the fundamental operations of graph signal processing is that of filtering. The simple graph filter is a polynomial in the graph shift operator, the coefficients of which are chosen to satisfy as closely as possible a desired filter response. The graph frequencies themselves are taken to be the eigenvalues of the graph shift operator. Both finite impulse response (FIR) [3] and infinite impulse response (IIR) [6] filters have been investigated and designed. The simple graph filter imposes on the network process a uniform rule: each delay is to be multiplied by the same coefficient by every node in the network. More agile and general filters have been designed whereby each node is able to choose a different coefficient for all of its neighbours (node-varying filters) [7], or each node is able to scale the signal it receives from each of its neighbours independently (edge-varying filters) [8].

Disregarding for the moment universal graph filter design [4], each of these strategies relies on knowledge of the graph shift operator in order to design the graph filter. In this paper, we investigate the line of questioning raised by Laszlo Lovasz in [18]. Specifically, how shall we design graph filters when the graph is so large that we cannot store or evaluate the entire graph and its diffusion, but only statistically probe the graph and derive models which in some way capture the graph's structure. To this end, [18] and references therein developed a probabilistic and function theoretic framework known as graphons. The word *graphon* itself is a portmanteau of graph and function. These are two-dimensional kernel functions supported on $[0,1] \times [0,1]$. Random graphs can be drawn from a *graphon* by drawing two samples uniformly distributed over $[0,1]$ for each pair of nodes $x_i, x_j, \ \forall i, j$ and interpreting the two samples as coordinates on the *graphon*. The value of the *graphon* at these coordinates is then the parameter p for a Bernoulli trial which decides the connection of the nodes. Random graphs drawn in such a way are structurally "related" to the graphon through graph homomorphism densities [18], and the spectra of the scaled adjacency matrices associated to such random graphs can be shown to converge to integral operators with the *graphon* as a kernel function [22]. In this paper we limit the scope of investigation to dense graphs, whose convergence properties are more easily understood. The case of sparse graphs, which is more applicable to real world graphs has been investigated in, for example, [19]. Other works in the signal processing literature have leveraged the spectral convergence properties of random graphs. For example, in [10] the authors leverage the spectral convergence properties of the scaled adjacency matrix to extend the notion of the graph Fourier transform to graphons, which they dub the *Graphon Fourier Transform*.

Leveraging these properties of *graphons*, we make three key contributions. The first is to show that a graph filter defined with the scaled adjacency matrix of a kernel-based random graph as its shift operator converges to the graphon filter defined by the same kernel. The major difficulty in performing this comparison is bridging the gap between the graph filter, which is a finite dimensional linear operator, and the graphon filter which is an infinite dimensional linear operator. To accomplish this task, we "lift" the graph filter into the infinite dimensional vector space as a step function defined over [0,1]. As a corollary of this finding, we draw a parallel to the numerical solution of Fredholm integral equations. Using projections of the kernel onto different sets of orthonormal bases we can calculate the output of the graph filter with a sparse shift operator, as opposed to a dense one, and thus reduce the cost of computation to $\mathcal{O}(N\log(N))$ where N in this case is the size of the basis chosen to represent the Fredholm integral equation. Finally, using the Fourier-Galerkin shift operator, we derive the optimal order-k graphon filtering algorithm.

To differentiate our approach from other graph filtering approaches, the graphon filtering algorithm is agnostic with respect to the actual graph. Although similar to the universal graph filtering approach, our approach differs from it in a fundamental sense. Graph filtering operates on the graph "frequencies" and the output of the graph filter is thus some projection onto the scaled eigenvectors of the graph. Thus, while the filter design may be universal with respect to graph frequencies, the filter design is not universal in terms of its input-output relationship. In contrast, the graphon filter is designed precisely based on the input-output relationship within a standardized basis for all classes of graphs. Thus if the filter is reachable for two given classes of graphs, their outputs will be similar with the same inputs. This is more analogous to filtering in classical signal processing. Our algorithm allows the user to both design a distributed graph filter, and also make predictions about the signal value at nodes in an arbitrary graph from the class for which the filter was designed. As such, it presents potentially great computational benefits in the domain of extremely large graphs, as the complexity of the algorithm depends on the graph complexity, which we call its "frequency content," and not its size.

The structure of this paper is as follows. Section II introduces graph signal processing preliminaries. In section III the basics of kernel based random graphs are introduced. In section IV we introduce some basic properties of Hilbert-Schmidt operators of which Fredholm integral operators with graphons as their kernels are a subset, and define graphon filters as the counterpart to graph filters. In section V, we introduce the theory of sparse approximations of Fredholm integral equations via expansion methods. In section VI we investigate the convergence of graph filters to their graphon counterparts. In section VII and VIII we define the Fourier-Galerkin shift operator and derive the graphon filtering algorithm. Section IX demonstrates the findings of the paper in simulations with several random graph models. Importantly, in subsection IX-C we investigate how to use the proposed methods contained in this paper on unlabeled graphs from which a graphon must be estimated, which is the situation that practically confronts anyone who would like to use the findings of this paper. This is followed finally by conclusions.

In the following, lower-case letters are scalar variables, upper-case letters are scalar constants, bold lower-case letters denote vectors, bold upper-case letters denote matrices, and caligraphic letters denote special mathematical objects such as distributions, sets, graphs, or functions, and double bold letters correspond to a field, e.g. $\mathbb{C}$ representing the complex numbers, or $\mathbb{K}$ representing a general, unspecified (infinite) field.

II. NETWORKS, ACTION, AND SHIFT OPERATORS

A graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is a double defined on a set of nodes $\mathcal{V} = \{v_1, \ldots, v_N\}$, and the set of connections between the nodes $\mathcal{E}$. The set of neighbours to a given node is denoted as $\mathcal{E}_j$. A graph signal $\mathbf{x} \in \mathbb{R}^N$ is one that is supported on the vertices of an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$. The graph shift operator $\mathbf{S}$ in some way captures the structure of the graph. For example, $\mathbf{S} = \mathbf{A}$, the adjacency matrix where $\mathbf{A}_{i,j} = \mathbf{A}_{j,i} = 1$ if $i, j \in \mathcal{E}$, and 0 otherwise. There are many other possible shift operators including the graph Laplacian, or weighted versions of the adjacency and Laplacian. Note that since the $\mathcal{G}$ under consideration is undirected, the graph shift operator $\mathbf{S}$ is symmetric, and thus has real eigenvalues and a complete set of orthonormal eigenvectors. In graph signal processing, it is assumed that the network structure in some way captures the evolution of the process supported on the graph nodes. However, it is not just the graph connections which influence the evolution of the graph signal, but also the action that the nodes take on the graph signal at each diffusion step. Any action that can be implemented by a linear operator can be represented as

$$[\mathbf{x}]_{j,t} = a[\mathbf{x}]_{j,t-1} + \sum_{i \in \mathcal{E}_j} b_{i,j}[\mathbf{x}]_{i,t-1}$$

where $b_{i,j}$ is the edge-weight connecting the i-th and j-th nodes in the graph, a is whatever weight the node gives to its own observation at "time" $t - 1$, and $[\mathbf{x}]_{i,t}$ is the i-th element of $\mathbf{x}$ at "time" t. Time here meaning discrete ordered diffusions through the graph shift operator: in other words "graph time." The signal $\mathbf{x}_{j,t}$ depends on information existing on the nodes of the graph at time $t - 1$. For example, in the case of the scaled adjacency matrix the operation each node performs on the signal is

$$[\mathbf{x}]_{j,t} = \frac{1}{N} \sum_{i \in \mathcal{E}_j} [\mathbf{x}]_{i,t-1}$$

In this paper, we consider only the scaled adjacency matrix $\mathbf{S} = \frac{1}{N}\mathbf{A}$ as the graph shift operator.

The Graph Fourier transform (GFT) of a graph signal $\mathbf{x}$ supported on a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is defined as

$$\hat{\mathbf{x}} = \mathbf{U}^T \mathbf{x}$$

where $\mathbf{S} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ is the eigenvalue decomposition of the graph shift operator. The eigenvalues of the shift operator are taken to be the "graph frequencies." Similarly, the inverse graph Fourier transform is defined as $\mathbf{x} = \mathbf{U}\hat{\mathbf{x}}$. Though two networks may have the exact same connections, depending on the actions performed at each node, the shift operator that models diffusion over the

graph may have a vastly different eigendecomposition, e.g. a scaled adjacency matrix and a Laplacian, and thus different modes and frequencies. Thus they will also have different frequency responses and different filtering operations altogether for the same filter coefficients.

Having defined the GFT for a fixed network and a given node action, the task of graph filtering can be introduced. The GFT is the expression of the signal supported on the network nodes in the modes of the graph. The filtering operation is then on the eigenvalues of the shift operator. Specifically

$$\mathbf{y} = \mathbf{U}\mathbf{D}'\mathbf{U}^T\mathbf{x} \qquad (1)$$

where $\mathbf{D}'$ is the matrix with the desired frequency coefficients along the diagonal. For example, if we wanted to filter out the i-th graph frequency, we would set $[\mathbf{D}']_{i,i} = 0$. As clear as this definition of a GFT is, it is abstract. It remains unclear how to actually implement the filtering operation on the network through local diffusion. By eliminating certain modes, the operator in (1) will likely develop non-zero entries in positions corresponding to non-existent connections in the graph. This indicates that the "ideal" graph filtering operation may not be realizable with local diffusion.

To filter the graph frequencies, a polynomial approach is thus adopted. The graph signal is diffused over the network by applying the shift operator K times, weighted by a coefficient and summed to produce the filtered output. In view of the definition of the ideal graph filter, the design objective for polynomial graph filters is to achieve a desired filter response in the graph frequencies. Specifically, certain graph frequencies are to be attenuated as much as possible, while holding others constant. With sufficiently many taps, any response can be modeled by

$$\mathbf{H} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k. \qquad (2)$$

In [3] it is proven that any filter that commutes with graph shift (i.e, a H such that for all x we have $\mathbf{S}(\mathbf{H}x) = \mathbf{H}(\mathbf{S}x)$) must be a polynomial: $\mathbf{H} = \mathrm{poly}(\mathbf{S})$. Since the filter is expressed through the diffusion over the graph through the action of the nodes, any such filter is, by definition, reachable. The filter in (2) has the implicit constraint that every node is constrained to use the same filter coefficient as every other node at each delay. We call such graph filters "simple." This constraint can be relaxed to produce node-varying and even edge-varying filters. In this paper, we consider only simple filters.

III. RANDOM GRAPH MODELS AND GRAPHONS

A. Kernel-Based Models

Our motivation is to conduct signal processing tasks on large graphs about whose structure we know little. This could be the case in the real world where the graphs are either too large, or too time-varying to have a single reliable representation. Such networks are typically not well-modeled by the classical Erdös-Renyí random graphs [14], [18]. Real-world networks typically have heavy-tailed degree-distributions, exhibit "small-world" phenomena, they tend to be clustered, and have neighborhood

density higher than the average edge density [14], [16], [18]. This has given rise to new, more general random graph models, one of which are Kernel-based models [20]. These are characterized by the use of a symmetric kernel function to generate probabilities controlling the formation of edges between nodes.

A kernel-based random graph is a triple $\mathcal{G}(N, \mathcal{W}, \mu)$ where N is the number of vertices in the graph, $\mathcal{W} : [0, 1]^2 \to [0, 1]$ is a symmetric measurable function, and μ is a random variable defined on $[0, 1]$. To generate a graph from this triple, for each pair of vertices v_i and v_j a sample is drawn from μ. Then each realization of μ is treated as a coordinate and mapped to a probability by the symmetric function $\mathcal{W}$. This probability is then used to perform a Bernoulli trial to establish if the two vertices are connected, in a manner analogous to the formation of an Erdös-Renyí graph. The symmetric function $\mathcal{W}$ is called a *graphon* by Lovasz and Szegedy [18]. In this paper, we assume that μ is uniform, i.e. μ is evenly distributed over [0,1]. It is important to mention that graphs generated in such a manner are dense [18]. The case of sparse graphs has been considered in, for example [19].

Conversely, the adjacency matrix of a graph $\mathcal{G}$ induces a graphon in the following way. Consider a weighted graph with vertex weights α_n, where $\sum_n \alpha_n = 1$, and edge weights $0 \leq \beta_{i,j} \leq 1$. Then, to construct a symmetric function, for each element of the adjacency matrix place a square of area $\alpha_i \times \alpha_j$ in the corresponding position of $[0, 1]^2$, in which the symmetric function will be the constant edge weight $\beta_{i,j}$. This graphon is denoted $\mathcal{W}_{\mathcal{G}}$. Here we assume $\alpha_i = \frac{1}{N}, \beta_{i,j} = 1$. Such a graphon is referred to as an *empirical graphon*.

B. Graph Sequences and Convergence

Consider a sequence of random graphs $(\mathcal{G}_n)$ with $n \to \infty$ where each $\mathcal{G}_n$ is drawn from a kernel model $\mathcal{G}(n, \mathcal{W}, \mu)$, for $n \to \infty$. A natural question is whether this sequence converges to any particular object. If so, then, while the graph in question is still random, its emergent properties would remain deterministic. The graph properties that we consider in this paper are homomorphism densities. The homomorphism density of a simple graph $\mathcal{F}$ with K nodes in a given graph $\mathcal{G}$ with N nodes is defined as

$$t(\mathcal{F}, \mathcal{G}) = \frac{hom(\mathcal{F}, \mathcal{G})}{hom(\mathcal{F}, \mathcal{K}_N)}$$

$$= \frac{hom(\mathcal{F}, \mathcal{G})}{N^K} \qquad (3)$$

where $hom(\mathcal{F}, \mathcal{G})$ is the number of adjacency preserving maps (homomorphisms) $\mathcal{V}(\mathcal{G}) \to \mathcal{V}(\mathcal{F})$, and $\mathcal{K}_N$ is the complete graph on N nodes, $\mathcal{K}_N$. Roughly speaking $t(\mathcal{F}, \mathcal{G})$ is a ratio between the number of copies of $\mathcal{F}$ in $\mathcal{G}$ and the number of copies of $\mathcal{F}$ in the complete graph on N nodes. The graph homomorphism densities of different graphs $\mathcal{F}$ in $\mathcal{G}$ provide structural information about the graph. For example, if $\mathcal{F} = \mathcal{K}_3$ then $t(\mathcal{F}, \mathcal{G})$ is the triangle density of $\mathcal{G}$. A sequence of graphs $(\mathcal{G}_n)$ is said to be *convergent* if $t(\mathcal{F}, \mathcal{G}_n)$ converges for any simple graph $\mathcal{F}$ [15].

We can further use the symmetric function $\mathcal{W}$ to define a limit object of the graph sequence $(\mathcal{G}_n)$. Let $\mathcal{W}$ be a bounded

symmetric function $\mathcal{W} : [0,1]^2 \to [0,1]$, and $\mathcal{F}$ be a simple graph with K nodes. Then the graph homomorphism limit of $\mathcal{F}$ in $\mathcal{W}$ is defined as

$$t(\mathcal{F}, \mathcal{W}) = \int_{[0,1]^K} \prod_{(i,j)\in\mathcal{E}(\mathcal{F})} \mathcal{W}(x_i, x_j)\mathrm{d}\mathbf{x} \qquad (4)$$

The following two theorems, proven in [17], form the fundamental link between convergent graph sequences and graphons, and hence the link between observed graphs and explanatory graphon models.

Theorem 1: For every convergent graph sequence $(\mathcal{G}_n)$ there exists a bounded measurable symmetric function $\mathcal{W} : [0,1]^2 \to [0,1]$ such that $\lim_{n\to\infty} t(\mathcal{F}, \mathcal{G}_n) = t(F, \mathcal{W})$

Theorem 2: The graph sequence, $(\mathcal{G}_n)$, where $\mathcal{G}_n$ is drawn from $\mathcal{G}(n, \mathcal{W}, \mu)$, is convergent with probability 1, and its limit object is the function $\mathcal{W}$.

Thus the graphon itself can be used to study the properties of the graph sequence. As will be observed in the next section, these are not limited to structural properties of the graph, as we can also use the graphon to calculate the spectra of the graphs in $(\mathcal{G}_n)$ as $n \to \infty$.

IV. FREDHOLM INTEGRAL EQUATIONS AND GRAPHON FILTERS

We begin by defining a Hilbert-Schmidt kernel function.

Definition IV.1: Let X and Y be intervals in $\mathbb{R}$ and $\mathcal{W} : X \times Y \to \mathbb{R}$. If $\int_X \int_Y |\mathcal{W}(x,y)|^2 \mathrm{d}x\,\mathrm{d}y < \infty$ then $\mathcal{W}$ is a Hilbert-Schmidt kernel function.

Graphons, as defined in the previous section are clearly Hilbert-Schmidt kernel functions, being both bounded and measurable. Hilbert-Schmidt kernel functions induce bounded integral operators on the space of square integrable functions $L^2(Y) \to L^2(X)$. In the case of graphons, they induce an integral operator $T : L^2[0,1] \to L^2[0,1]$

$$Tf(x_1) := \int_0^1 \mathcal{W}(x_1, x_2)f(x_2)\mathrm{d}x_2. \qquad (5)$$

Equations of the form

$$g(x) = Tf(x) \qquad (6)$$

are known as Fredholm integral equations of the first kind [21].

Since the graphons are required to be symmetric, the operators they induce are self-adjoint. This, combined with the fact that Hilbert-Schmidt operators are compact [21] implies that the spectrum of the operator (5) consists of a countable number of real-valued eigenvalues with $\lambda_n \to 0$. The eigenvalues of (5) and corresponding eigenfunctions can be found by solving the resolvent equation

$$\lambda f(x_1) = \int_0^1 \mathcal{W}(x_1, x_2)f(x_2)\mathrm{d}x_2 \qquad (7)$$

for $f(x)$ and λ, given the kernel $\mathcal{W}$. Since we know from Theorems 1–2 that graph sequences converge to their limit object, and that every graph sequence has a limit object, knowledge of the spectrum of (5) provides knowledge of the spectrum of the graphs in $(\mathcal{G}_n)$, particularly when the number of vertices becomes large. Indeed, it is not difficult to see from the empirical

graphons $\mathcal{W}_{\mathcal{G}_n}$ that the spectrum of $\mathcal{W}$ is the spectrum of the scaled adjacency matrix corresponding to $\mathcal{G}_n$ as $n \to \infty$. Theorem 11.53 in [18] gives a formal statement and proof of this fact. The *Graphon Fourier Transform* has also be defined via the resolvent equation in [10]. Similarly, the degree matrix is also determined by the graphon [28], and thus, the spectrum of the graph Laplacian can also be investigated through the resolvent equation. We do not pursue this here, though.

As was shown in section II a distributed simple graph filter $\mathbf{H}$ can be implemented as a polynomial in the graph shift operator $\mathbf{S}$

$$\mathbf{H} = h_0\mathbf{I} + \sum_{k=1}^{K} h_k\mathbf{S}^k, \qquad (8)$$

also known as a k-th order FIR graph filter. The graph filter of order 1 with $h_0 = 0$ is simply the matrix vector product $\mathbf{y} = \mathbf{S}\mathbf{x}$ which as we've mentioned converges spectrally to the equation $g(x) = Tf(x)$. The connection between $g(x), f(x), T$ and $\mathbf{S}$ is the subject of section VI. By defining the powers of T through operator composition

$$T^2 f(x) = TTf(x) = \int_0^1 \mathcal{W}(x, z) \int_0^1 \mathcal{W}(z, y)f(y)dy\,dz. \qquad (9)$$

we arrive at the definition of a graphon filter. The higher powers of T are derived inductively. We draw attention to the fact that there must be an intermediate dimension added for each power of T. A graphon filter is then defined as

$$\mathcal{H}f(x) = h_0 f(x) + \sum_{k=1}^{K} h_k T^k f(x). \qquad (10)$$

V. EXPANSION METHODS FOR FREDHOLM INTEGRAL EQUATIONS

Comparing (6) and $\mathbf{y} = \mathbf{S}\mathbf{x}$ there is one fundamental difference which must be reconciled. The Fredholm integral operator in (6) operates on functions with a continuous domain [0,1], whereas the equation $\mathbf{y} = \mathbf{S}\mathbf{x}$ is over a finite dimensional vector space. We relate the two via the expansion method. Let $\mathcal{B} = \{b_1, b_2, \cdots\}$ be a complete orthonormal basis for $L^2[0,1]$. The expansion method seeks to approximate the (forward) solution of the Fredholm integral equation, as $g(x) = \sum_{i=1}^{\infty} g_i b_i(x)$. We seek to find the coefficients g_i given the graphon $\mathcal{W}(x,y)$, and the input function $f(x)$. To do this, we use the inner product over $L^2[0,1]$ to project both sides of (6) onto the basis $\mathcal{B}$, as

$$\int_0^1 g(x)b_i(x)\mathrm{d}x = \int_0^1 b_i(x) \int_0^1 \mathcal{W}(x,y)f(y)\mathrm{d}y\,\mathrm{d}x$$

$$\text{for } i \in 1, 2, \ldots \quad (11)$$

Substituting the expansions of $g(x)$ and $f(y)$ allows us to write

$$g_i = \int_0^1 b_i(x) \int_0^1 \mathcal{W}(x,y) \sum_{j=1}^{\infty} f_j b_j(y)\mathrm{d}y\,\mathrm{d}x \qquad (12)$$

where the left hand side follows from the orthonormality of $\mathcal{B}$. Inspecting the right hand side, we see that for each j, the integral with respect to y can be factored out. The integral with respect

to y thus becomes the projection of $\mathcal{W}(x, y)$ onto $b_j(y)$, leading to $\mathcal{W}_j(x) = \int_0^1 \mathcal{W}(x, y)b_j(y)\mathrm{d}y$.

$$g_i = \sum_{j=1}^{\infty} f_j \int_0^1 b_i(x)\mathcal{W}_j(x)\mathrm{d}x \qquad (13)$$

Since the sum is invariant with respect to x, the order of summation and integration can be swapped resulting in (13). The projections of $\mathcal{W}_j(x)$ onto $\mathcal{B}$ are now no longer orthogonal to the basis functions $b_i(x)$. By inspection, for i and j ranging over $1, 2, \ldots$, (13) can be written as the following matrix vector equation.

$$\mathbf{g} = \mathbf{Wf} \qquad (14)$$

where $[\mathbf{W}]_{i,j} = \int_0^1 b_i(x) \int_0^1 \mathcal{W}(x, y)b_j(y)\mathrm{d}x \ \mathrm{d}y$ and $[\mathbf{f}]_j = \int_0^1 f(y)b_j(y)\mathrm{d}y$. If $\mathcal{B}$ is complete and orthonormal, $g(x)$ can be approximated without error by (14). We call $\mathbf{W}$ the operator matrix, and (14) the operator equation. The operator equation is a linear algebraic equation of countably infinite dimension. Thus, by truncating (14) we arrive at a finite linear approximation of an integral equation over an infinite domain. This truncated finite approximation we denote as

$$\mathbf{g}_a = \mathbf{W}_a\mathbf{f}_a. \qquad (15)$$

As $\mathbf{g}_a$ is a representation of the solution of the Fredholm integral equation in the basis $\mathcal{B}$, the approximate solution to the integral equation can then be resampled with as many points as desired as

$$\mathbf{y}' = \sum_{i=1}^{N}[\mathbf{g}_a]_i\mathbf{b}_i \qquad (16)$$

where N is the number of approximating functions in the expansion, and $\mathbf{b}_i$ is the vector of samples drawn uniformly from the basis function $b_i(x)$ from the interval $[0, 1]$.

To make an analogy to classical signal processing, (15) is like the representation of a "time" domain signal (6) in "frequency" domain.

Remark 1: In the above derivation, we have assumed the orthonormality of the set $\mathcal{B}$. Practically, basis functions which are not even orthogonal can be used by introducing a weighting function $w(x)$, as we do in section VII. Normality can be enforced by renormalization.

Remark 2: In most cases, $\mathbf{W}_a$ will be a principal sub-matrix of $\mathbf{W}$. However, this is not necessarily the case. In the next section, the size of the basis determines the basis functions themselves. To remind the reader of this distinction, we mark such bases as $b_i'(x)$.

Remark 3: The choice of basis affects the matrix $\mathbf{W}$. Specifically, the size of the matrix and its sparsity depend both on the choice of the basis and the properties of the graphon itself. Smooth graphons will be easily approximated by a few non-zero coefficients with respect to smooth basis functions, while graphons with abrupt changes will require more basis functions and more non-zero coefficients.

VI. EMPIRICAL GRAPHONS AND THE FREDHOLM EQUATION

The purpose of this section is to make explicit the connection between graph and graphon filters. We will do this in a series of steps illustrated in the below commutative diagram.

The link ϕ bridges the gap between the finite dimensional shift operator and the infinite dimensional graphon Fredholm operator defined by the empirical graphon corresponding to the graph in question, operating on the piecewise step-functions defined on $[0, 1]$. The first right arrow then shows that, since the piecewise step-functions are dense in the continuous functions, that there is a stepfunction arbitrarily close to any element of the function $L^2[0, 1]$, and therefore the composition of the empirical operator defined on a piece-wise approximation of a function in $L^2[0, 1]$ converges to the empirical operator composed on that function. And finally, the last right arrow concerns the probabilistic convergence of the empirical operator to the true graphon operator. The logic is that the shift operator is injectively related to an object which deterministically converges to something which probabilistically converges to the desired result.

This chain concerns the convergence of the scaled adjacency matrix operating on approximations of the function space $L^2[0, 1]$ to the true graphon operator defined over the whole space. Graph and graphon filters, however, are comprised of powers of these operators. The convergence of the powers of $\mathbf{S}$ to the powers of T is demonstrated in the final subsection.

As a general note, while the bases used in section V were arbitrary complete orthonormal bases, in this section specific bases are used to illustrate key points. Thus, for example, $\mathbf{W}_a$ in this section is a specific instance of $\mathbf{W}_a$ from the previous section.

A. From Finite to Infinite Dimensions

To compare a finite vector to a function in $L^2[0, 1]$ we must define the map ϕ. Given a basis $\mathcal{B}$ and a finite dimensional vector $\mathbf{x}$ we define

$$\phi(\mathbf{x}) := f_a(x) = \sum_{i=1}^{N}[\mathbf{x}]_ib_i(x), \ b_i(x) \in \mathcal{B}. \qquad (17)$$

The map ϕ^{-1} then returns the coefficients $[\mathbf{x}]_i$. The map ϕ could also be similarly defined (though in the opposite sense) given a function $f(x)$, however, our concern is analyzing the convergence of the shift operator $\mathbf{S}$ to a continuous operator T. In order to avoid confusing terms in the following derivation we remind the reader that $[\mathbf{x}]_i$ is the value of the graph signal at node i, $f_a(x)$ is a step-function in $L^2[0, 1]$, and $[\mathbf{f}]_i$ is the projection of a function $f(x)$ onto the i-th member of an orthonormal basis for $L^2[0, 1]$.

Proposition 1: There exists a basis $\mathcal{B}$ such that ϕ is bijective.

Proof: Let $\mathcal{G}$ be a kernel-based random graph of N nodes drawn from a kernel $\mathcal{W}(x, y)$. We now show that the evaluation of $T_e f_a(x)$ is equivalent to $\mathbf{Sx}$. That is the finite vector $\phi^{-1}(g_a(x))$ where $g_a(x) = T_e f_a(x)$ will equal $\mathbf{y}$. Assume that all the nodes are evenly weighted with node-weight $\alpha = 1/N$, and connections between nodes i and j are weighted by $\beta_{i,j} = 1$. Then, with x_i being the i-th strip of width $1/N$ of the interval $[0,$

1] centered at $i \cdot 1/N$, the empirical graphon $\mathcal{W}_e(x_i, y_j) = 1$ if nodes i and j are connected, and $\mathcal{W}_e(x_i, y_j) = 0$ otherwise.

This empirical graphon can be approximated exactly using the procedure described in the previous section, using a basis of N orthonormal functions. Take as the basis $\{b_1(x), \ldots, b_N(x)\}$ the functions

$$b_i'(x) = \begin{cases} N & \text{if } x \in [i \cdot \frac{1}{N}, i \cdot \frac{1}{N} + \frac{1}{N}) \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

The approximation $\mathbf{W}$ of the empirical graphon $\mathcal{W}_e(x, y)$ is then the adjacency matrix of the graph with respect to this basis, for any size N. To see this, we use (14) to calculate $\mathbf{W}_a$,

$$[\mathbf{W}_a]_{i,j} = \int_0^1 \int_0^1 \mathcal{W}_e(x, y) b_i'(x) b_j'(y) \, \mathrm{d}x \mathrm{d}y$$

$$= 1 \cdot \frac{1}{N^2} \cdot N^2 = 1 \text{ if } (i,j) \in \mathcal{E}, \ 0 \text{ o.w.} \quad (19)$$

which is the adjacency matrix of $\mathcal{G}$, by definition.

Thus the truncated Fredholm integral operator (15) can be written as

$$g_a(x) = \int_0^1 \mathcal{W}_e(x, y) f_a(y) \mathrm{d}y \quad (20)$$

$$g_a(x) = \sum_{i=1}^{N} \sum_{j \in \mathcal{E}_i} \frac{1}{N} [\mathbf{x}]_j b_i'(x)$$

$$\Rightarrow \phi^{-1}(g_a(x)) = \frac{1}{N} \mathbf{W}_a \phi^{-1}(f_a(x)) = \mathbf{S}\mathbf{x} = \mathbf{y}. \quad (21)$$

This simple derivation leverages the fact that $\mathbf{W}_a$ is exactly $\mathbf{A}$ with respect to the basis $\mathcal{B}$. Having noted this, the integral in (20) is equivalently represented by the averaging operation in (21), and thus $\mathbf{g}_a$ is exactly $\mathbf{y}$. Thus, there is no error introduced by the replacement of the integral in (14) by the summation (implemented by the matrix product) and division operations in (21). Equivalence of higher powers of T_e^k follow inductively from the observation that T_e maps stepfunctions to stepfunctions. ∎

B. Right Arrow 1

Let T_e be a Fredholm integral operator with the empirical graphon of $\mathcal{G}$ as its kernel, and T_a be the Fredholm operator in (15) with the approximation of $\mathcal{W}_e$ as its kernel, using the basis defined in the previous section. Then, in order to show convergence of $T_a f_a(x) = g_a(x)$ to $T_e f(x) = g_e(x)$ we denote the error function as $e_a = g_a(x) - g_e(x)$. Then, it can be seen that

$$T_e f(x) - T_a f_a(x) = g_e(x) - g_a(x)$$

$$\Rightarrow T_e(f(x) - f_a(x)) = g_e(x) - g_a(x)$$

$$\Rightarrow \|e_a\| \le \|T_e\| \cdot \|f(x) - f_a(x)\|. \quad (22)$$

where $\|T_e\| = \sup_{\|v(x)\|=1} \|T_e v(x)\|, v(x) \in L^2[0,1]$. The first step is implied by the fact that

$$\mathcal{W}_e(x, y) = \sum_{i=1}^{N} \sum_{j=1}^{N} [\mathbf{W}_a]_{i,j} b_i'(x) b_j'(y) \quad (23)$$

where $\mathbf{W}_a$ is as defined in the previous subsection, and the second step uses the Cauchy-Schwarz inequality. It is not, in general, true that T_e and T_a as defined at the beginning of this subsection will be equal, it is only through the specific choice of basis that this holds. The norm $\| \cdot \|$ is with respect to the space $L^2[0,1]$.

In this case, the approximation error is entirely due to the quadrature error of the input $f(x)$, and is scaled by the spectral radius (T is self-adjoint) of T. We know that this is less than 1 from [18]. Thus, to guarantee convergence of the matrix approximation of the empirical graphon Fredholm equation by (14) as $N \to \infty$, we only need to assume continuity of the input $f(x)$. As a note, the continuous functions are dense in $L^2[0,1]$, and as $N \to \infty$, the set of basis functions (18) are a basis for this space.

With the previous result in mind, we return to the comparison between (8) and (10) with the goal of bounding the total error of the approximation of (10) defined by the empirical graphon by (8) where $\mathbf{S}$ is given by the scaled adjacency matrix.

Define the output of a filter acting on $L^2[0,1]$ as $z_e(x) = (Ih_0 + \sum_{k=1}^{K} h_k T_e^k) f(x)$ and the output of a graph filter using the scaled adjacency matrix lifted into $L^2[0,1]$ via the equivalence relationship between (15) and (21) as $z_a(x) = (h_0 I + \sum_{k=1}^{K} h_k T_a^k) f_a(x)$. We emphasize that $z_a(x)$ is a function in $L^2[0,1]$, but it is equivalently represented as a vector in $\mathbb{R}^N$ as the output of a graph filter defined by the scaled adjacency matrix. Then, taking the norm

$$\|z_e(x) - z_a(x)\| = \|h_0 I(f_e(x) - f_a(x)) - h_1 T_e(f_e(x)$$
$$- f_a(x)) + \cdots + T_e^k(f_e(x) - f_a(x))\|$$

$$\le \|\sum_{k=0}^{K} h_k T_e^k\| \cdot \|f(x) - f_a(x)\|$$

$$(24)$$

Since the spectral radius of T_e is strictly less than 1, $\|T_e^k\| < \|T_e\|$ yielding the result

$$\|z_e(x) - z_a(x)\| < \left\|\sum_{k=0}^{K} h_k\right\| \cdot \|T_e\| \cdot \|f(x) - f_a(x)\| \quad (25)$$

As the dimension $N \to \infty$, the approximation $\|f(x) - f_a(x)\|$ tends to zero assuming $f(x)$ is continuous. The quantities $\|\sum_{k=0}^{K} h_k\|$ and $\|T_e\|$ are constants with respect to the limit as $N \to \infty$ and thus the graph filter defined over the scaled adjacency matrix converges to the graphon filter defined over the empirical graphon.

C. Right Arrow 2

The previous sections define the sense in which the diffusion of a signal on the nodes of a graph converge to the solution of a Fredholm integral equation. However, the Fredholm integral operator in the previous section has as its kernel the empirical graphon. What remains to be shown is how the output of the empirical Fredholm equation relates to the Fredholm equation associated with the "true" graphon.

Let $g_e(x) = \int_0^1 \mathcal{W}_e(x,y)f(y)dy$, then the error we wish to quantify is $\|g(x) - g_e(x)\|$.

$$\|g(x) - g_e(x)\| = \left\| \int_0^1 \mathcal{W}(x,y)f(y)dy \right.$$
$$\left. - \int_0^1 \mathcal{W}_e(x,y)f(y)dy \right\|$$
$$= \left\| \int_0^1 (\mathcal{W}(x,y) - \mathcal{W}_e(x,y))f(y)dy \right\|$$
$$\leq \|\mathcal{W}(x,y) - \mathcal{W}_e(x,y)\| \cdot \|f(y)\| \quad (26)$$

where the second step follows from Cauchy-Schwarz.

In [18], it is shown that, in general $\|\mathcal{W}(x,y) - \mathcal{W}_e(x,y)\|_\square$ where $\|\cdot\|_\square = \sup_{S,T \subset [0,1]} |\int_{S \times T} \mathcal{W}(x,y) \, \mathrm{d}x \, \mathrm{d}y|$, does not converge to 0 as $N \to \infty$, as the cut norm varies widely with the node labeling of the graph to which the empirical graphon corresponds. The cut metric defined as

$$\delta_\square(\mathcal{W}, \mathcal{W}_e) \triangleq \inf_\psi \|\mathcal{W} - \mathcal{W}_e^\psi\|_\square \quad (27)$$

where ψ is an invertible, measure preserving map on [0,1], and $\mathcal{W}_e^\psi = \mathcal{W}_e(\psi(x), \psi(y))$ can be shown to converge. Since $\|\mathcal{W}\|_\square \leq \|\mathcal{W}\|$ it is certainly true that $\|\mathcal{W} - \mathcal{W}_e\|$ does not converge, in general. However, it was shown in Corollary 1.1 of [22] that if $\|\mathcal{W}_e\|$ converges to $\|\mathcal{W}\|$, and $\|\mathcal{W} - \mathcal{W}_e\|_\square \to 0$, then $\mathcal{W}_e$ also converges in the L_2 topology. Therefore, the conditions under which $\|\mathcal{W}(x,y) - \mathcal{W}_e(x,y)\| \to 0$ should be investigated.

In [28], several rates for $\mathcal{W}_e(x,y)$ to converge to $\mathcal{W}$ are given assuming that the sample μ_i are ordered, i.e. $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_N$. In the first case, the graphons are assumed to be piece-wise Lipschitz continuous. In the appendix, however, they give a convergence rate for a general graphon, based on Lemma 10.16 of [18]. For the purposes of completing the commutative diagram, it is sufficient to note one such rate.

Proposition 2: If μ_i are ordered, then, with probability at least $1 - \exp(-\frac{N}{2\log N})$, $\|\mathcal{W}_e(x,y) - \mathcal{W}(x,y)\| \leq \sqrt{176}/\log(N)^{1/4}$.

In particular, $\lim_{N \to \infty} \|\mathcal{W}_e(x,y) - \mathcal{W}(x,y)\| \to 0$.

Informally, the sorting of μ_i introduces a labeling, thus eliminating the need for the measure preserving map ψ. Visually, the effect of sorting on μ_i conveys the point of proposition 2 quite clearly. However, in practice, we encounter unlabeled graphs, and not graphons or the latent variables μ_i. What is important is that the labeling of the graph is consistent with the map ψ, not that we know the "true" map ψ. Following the findings of [28], centrality measures such as degree or page rank are both coherent concepts for graphons, and the centrality measures of the graphs drawn from graphons also converge to their graphon counterparts. Such centrality measures can be used to enforce a labeling on an unlabeled graph, and so long as this labeling remains consistent with the graphon estimate based on the observed sample, the output will also be consistent. We demonstrate this procedure using the graphon estimation technique of [29] and degree sorting based on [28] in Section IX-C.
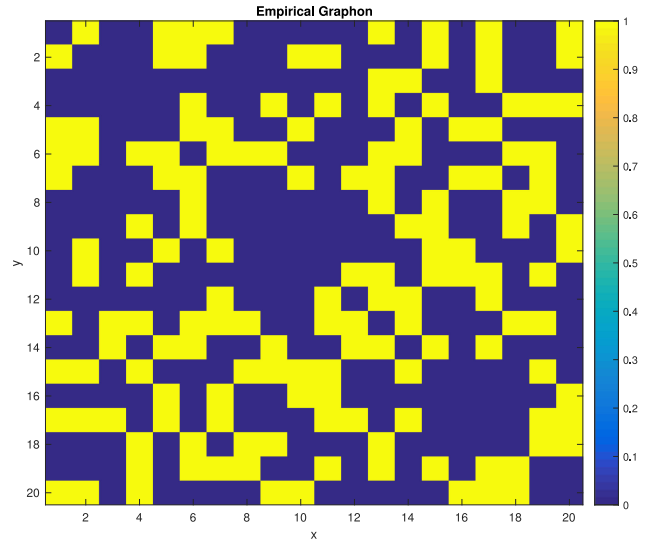

Fig. 1. Empirical graphon of a random graph of 20 nodes generated from the graphon $\mathcal{W}(x,y) = e^{-1/2(x+y)}$.

Fig. 2. Empirical graphon corresponding to unsorted samples of $\mathcal{W}(x,y) = 1/3 + 1/3 \cdot \sin(3 \cdot \pi x \cdot y)$.
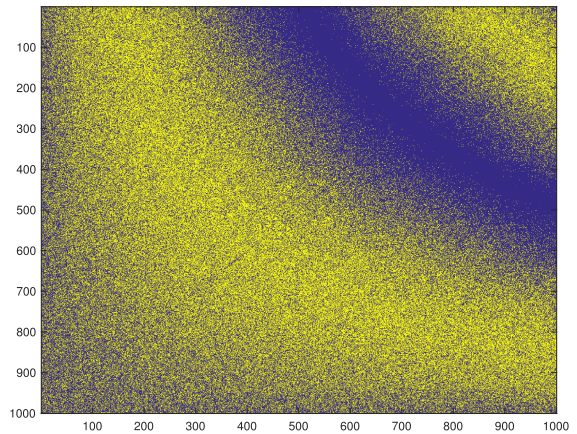

Fig. 3. Empirical graphon corresponding to sorted samples of $\mathcal{W}(x,y) = 1/3 + 1/3 \cdot \sin(3 \cdot \pi x \cdot y)$.
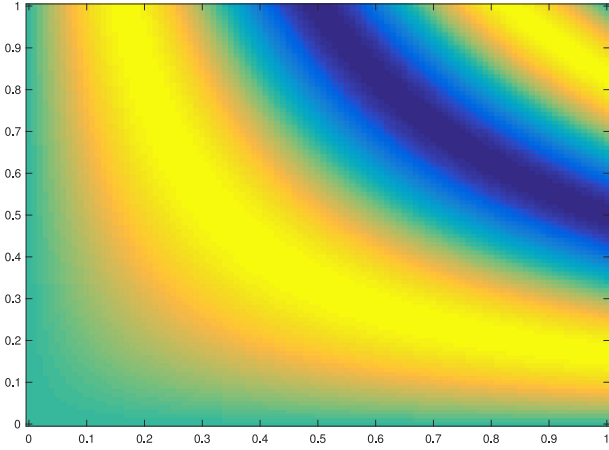
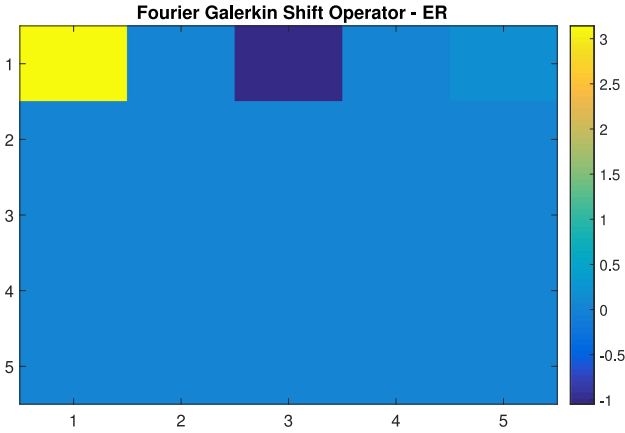Fig. 4.   The graphon $\mathcal{W}(x, y) = 1/3 + 1/3 \cdot \sin(3 \cdot \pi x \cdot y)$.

Fig. 5.   Fourier-Galerkin shift operator corresponding to an Erdös-Renyí model with $p = 0.5$. There is one non-zero row, the first, indicating that the range space of this operator is the constant functions.

In Figs. 3 and 4 we see the effect of the sorting on μ_i in it's convergence in norm to the graphon displayed in Fig. 5. Alternatively, we can assume that the samples μ_i might not be ordered with respect to the reals, but consistently ordered from sample to sample. In that case, the graphon to which the empirical graphon converges in norm is $\mathcal{W}_e^{\psi^*}$ where ψ^* is the measure preserving map which minimizes $\|\mathcal{W} - \mathcal{W}_e^\psi\|_\square$. Specifically, it is shown in [23] that the empirical graphon associated with the adjacency matrix not only converges to $\mathcal{W}(x, y)$ but is rate optimal in the minimax sense. To wit, for k-step graphons, it can be shown that

$$\delta_\square(\mathcal{W}, \mathcal{W}_e) \leq C\sqrt{\frac{k}{N\log(k)}} \qquad (28)$$

where C is a numerical constant [23]. In general, however, for an arbitrary graphon, the most that can be shown is a convergence rate proportional to $\log(N)^{-1/2}$ [18], [23].

What remains to be shown is that the powers T_e^k converge to the power of the true graphon operator T^k. We proceed by proof by induction. In the following, for ease of notation we make use of the equivalence of norms to write the equations in terms of the

operator 2 norm, in this case equivalent to the spectral radius.

$$\|T^k - T_e^k\| = \|TT^{k-1} - T_eT_e^{k-1} + T_eT^{k-1} - T_eT^{k-1}\|$$

$$= \|(T - T_e)T^{k-1} + T_e(T^{k-1} - T_e^{k-1})\|$$

$$\leq \|T^{k-1}\| \cdot \|T - T_e\| + \|T_e\| \cdot \|T^{k-1} - T_e^{k-1}\|, \qquad (29)$$

where the final implication is a combination of Minkowski's inequality and Cauchy-Schwarz. We then proceed by induction to write

$$\|T^{k-1} - T_e^{k-1}\| \leq \|T^{k-2}\| \cdot \|T - T_e\|$$
$$+ \|T_e\| \cdot \|T^{k-2} - T_e^{k-2}\| \qquad (30)$$

Repeating this process $k - 1$ times yields the following expansion.

$$\|T^k - T_e^k\| <= \|T - T_e\| \sum_{i=1}^{k} \|T^{i-1}\|\|T_e\|^{k-i} \qquad (31)$$

Since $\|T^k\| \leq \|T\|^k$ for a finite k in a normed space, we can simplify.

$$\|T^k - T_e^k\| \leq \|T - T_e\| \sum_{i=1}^{k} \|T^{i-1}\|\|T_e\|^{k-i}$$

$$\leq \|T - T_e\| \sum_{i=1}^{k} \|T\|^{i-1}\|T_e\|^{k-i}$$

$$= \|T - T_e\| \sum_{i=1}^{k} \lambda_{max}^{i-1}\tilde{\lambda}_{max}^{k-i}$$

$$\leq k \cdot \|T - T_e\| \qquad (32)$$

where λ_{max} and $\tilde{\lambda}_{max}$ are the largest eigenvalues of T and T_e respectively, each being not greater than 1. This allows us to state the following convergence result for the

$$\|z(x) - z_e(x)\| = \left\|\left(\sum_{k=0}^{K} h_0 T^k - \sum_{k=0}^{K} h_0 T_e^k\right) f(x)\right\|$$

$$\leq \sum_{k=0}^{K} |h_k| \cdot \|T^k - T_e^k\| \cdot \|f(x)\|$$

$$\leq \sum_{k=0}^{K} k \cdot |h_k| \cdot \|T - T_e\| \cdot \|f(x)\|$$

$$= \sum_{k=0}^{K} k \cdot |h_k| \cdot \|T - T_e\| \text{ w.l.o.g.}$$

$$\text{as } N \to \infty \qquad (33)$$

The result of Theorem 1 can be easily combined with Proposition 2 to give a rate of convergence.

Corollary 1: With probability at least $1 - \exp(-\frac{N}{2\log N})$, $\|\mathcal{H} - \mathcal{H}_e\| \leq \sqrt{176}/\log(N)^{1/4} \sum_{k=0}^{K} k \cdot |h_k|$.

In summary, ϕ has been shown to be bijective with $T_e f_a(x)$, which has been shown to converge to $Tf(x)$. Formally, we write

$$\|\mathcal{H}_e f_a(x) - \mathcal{H}f(x)\| = \|\mathcal{H}_e f_a(x) + \mathcal{H}f_a(x)$$
$$- \mathcal{H}f_a(x) - \mathcal{H}f(x)\|$$
$$\leq \|\mathcal{H}_e - \mathcal{H}\| \cdot \|f_a(x)\|$$
$$+ \|\mathcal{H}\| \cdot \|f_a(x) - f(x)\|$$
$$\leq \|\mathcal{H}_e - \mathcal{H}\| + \sum_{k=0}^{K} |h_k| \cdot \delta(N) \quad (34)$$

where $\delta(N) = \|f_a(x) - f(x)\|$. In the above derivation we have used Minkowski's inequality as well as Cauchy-Schwarz.

Theorem 3: $1 - \exp(-\frac{N}{2\log N})$, $\|\mathcal{H}_e f_a(x) - \mathcal{H}f(x)\| \leq \sum_{k=0}^{K} k \cdot |h_k|(\sqrt{176}/\log(N)^{1/4} + \sum_{k=0}^{K} |h_k| \delta(N) \square$.

VII. SHIFT OPERATOR DEFINITION

In the previous sections we have demonstrated the convergence of a graph filter defined on a random graph to a graphon filter defined on the kernel-model from which the random graph was generated. In order to make the comparison between a filter defined on a finite dimensional vector space, and a filter defined on an infinite vector space, we had to introduce the concept of an operator matrix. To do this, the input and output functions and the kernel were projected onto an orthonormal basis for the space $L^2[0, 1]$. Thus we arrive at the question of how to appropriately choose a basis in order to arrive at an operator which captures the behaviour of the operator T as concisely as possible.

We know of a way to choose a matrix on the basis of a graphon that converges to the Fredholm integral equation: a random scaled adjacency matrix drawn from the graphon by the procedure mentioned in Section III has this property. However, the convergence rate is extremely slow as we saw in the previous sections. One could choose the step-function basis used in Section VI however this would also require many step-functions to converge, especially for smooth graphons. This would defy our purpose stated in the introduction of how to perform graph filtering on graphs that are too large to study as a whole. Clearly, the choice of basis affects the resulting operator matrix: both in its size and complexity. Thus, the basis should be chosen such that the resulting operator matrix is as small as possible, as sparse as possible, and can be calculated as efficiently as possible.

The primary practical concern is that one should be able to solve the integrals in (15), and they should be sufficiently general to be able to approximate as many different types of graphons as possible. To this end, we choose the Tchebyshev basis functions. Firstly, these functions are the best approximators of continuous functions in $L^2[0, 1]$. Secondly, the Tchebyshev functions allow us to use Gauss-Tchebyshev quadrature, thus transforming the analytical integrals in (15) into weighted sums which can be numerically evaluated.

A. Fourier-Galerkin Shift Operator

In the Galerkin approach [24] to the solution of Fredholm integral equations (6), a basis of orthogonal polynomials are chosen, and the function $f(x)$ and kernel $\mathcal{W}(x, y)$ are taken to be the projections onto this basis. Following [24], [25] the orthogonal basis is chosen to be the Chebyshev polynomials of the first kind, $c_i(x)$. Working under the frame work of the operator matrix vector equation (14), we now obtain

$$[\mathbf{f}]_i = \int_{-1}^{1} \frac{f(u)c_i(u)}{(1 - u^2)^{1/2}} du \quad (35)$$

and

$$[\mathbf{W}]_{i,j} = \int_{-1}^{1} \int_{-1}^{1} \frac{\mathcal{W}(u, v)c_i(u)c_j(v)}{(1 - u^2)^{1/2}} du \, dv \quad (36)$$

Notably, the Tchebyshev functions of the first kind are only orthogonal with respect to the weight function $1/(1 - x^2)^{1/2}$, hence its inclusion in equation (36). However, by mapping the interval of integration from $[0, 1]$ to $[-1, 1]$ with the substitution rule $u = 2x - 1$ we can use the Gauss-Tchebyshev quadrature rule to convert the equations (36) into tractable weighted sums. Specifically, equations (35) and (36) can be evaluated with the rule

$$\int_{-1}^{1} \frac{f(u)}{\sqrt{1 - u^2}} du \approx \sum_{i=1}^{N} w_i f(u_i) \quad (37)$$

where $w_i = \frac{\pi}{P}$ and $u_i = cos(\frac{2i-1}{P}\pi)$, where P is the number of quadrature points, and N is the number of Tchebyshev functions. Using (37) and using the identity

$$c_i(u) = cos(icos^{-1}(u)) \quad (38)$$

we can rewrite equations (35) and (36) as

$$[\mathbf{f}]_i = \frac{\pi}{P} \sum_{k=0}^{\tilde{P}} f\left(cos\left(\frac{\pi i}{N}\right)\right) cos(\pi(k - 1)) \quad (39)$$

$$[\tilde{\mathbf{W}}]_{i,j} = \frac{\pi^2}{P^2} \sum_{m=0}^{\tilde{P}} \sum_{n=0}^{\tilde{P}} \mathcal{W}\left(cos\left(\frac{\pi m}{P}\right), cos\left(\frac{\pi n}{P}\right)\right)$$
$$cos\left(\frac{\pi m(i - 1)}{P}\right) cos\left(\frac{\pi n(j - 1)}{P}\right) \quad (40)$$

where the operator $\tilde{\sum}$ denotes that the first and last terms are multiplied by $\frac{1}{2}$. The operator $\tilde{\mathbf{W}}_{i,j}$ is, in fact, the numerical solution to the integrals

$$\int_{-1}^{1} \frac{c_i(u)}{\sqrt{1 - u^2}} \int_{-1}^{1} \frac{\mathcal{W}(u, v)c_j(v)}{\sqrt{1 - v^2}} du \, dv \quad (41)$$

which has the extra weight function $\frac{1}{\sqrt{1-v^2}}$ compared to the defining equations (36). Thus, approximating T by (41) will incur large errors. However, we can derive the operator $\mathbf{W}$ from the operator $\tilde{\mathbf{W}}$ by multiplying (41) by the Tchebyshev expansion of $\sqrt{1 - v^2}$

$$\sqrt{1 - v^2} = \frac{2}{\pi} - \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{c_{2k}(v)}{4k^2 - 1}. \quad (42)$$

Algorithm 1: Fourier-Galerkin Shift Operator $(\mathcal{W}(x,y), P, N)$ Complexity $\mathcal{O}(P^2)$.

---

calculate $\tilde{\mathbf{W}}$ with (40)

calculate $\mathbf{W}$ with (44)

return $N \times N$ principal submatrix of $\mathbf{W}$

By multiplying the series (40) by the series (42), and using the identity

$$c_p(u)c_q(u) = \frac{1}{2}\left(c_{p+q}(u) + c_{|p-q|}(u)\right) \tag{43}$$

the operator $\mathbf{W}$ can be approximately calculated as

$$\frac{\pi}{2}[\mathbf{W}]_{i,j} \approx [\tilde{\mathbf{W}}]_{i,j} - \sum_{l=1}^{P} \frac{1}{4\,l^2-1}\left([\tilde{\mathbf{W}}]_{i,j+2\,l} + [\tilde{\mathbf{W}}]_{i,|j-2l|}\right) \tag{44}$$

where the operator $\tilde{\mathbf{W}}$ in the above equation is equal to the operator $\tilde{\mathbf{W}}$ in (40) but padded with extra zeros depending on the length of the series (42). The calculation of the Fourier-Galerkin shift operator is summarized in Algorithm 1.

Illustrative Example: The simplest example, structurally speaking, is that of the Erdös-Renyí (ER) model. Applying (36) to the function $\mathcal{W}(x,y) = p$ readily gives the result that $[\mathbf{W}]_{1,1} = \pi^2 p$, with $[\mathbf{W}]_{i,j} = 0$, $\forall i \neq j$ and $i,j \geq 1$, since $\sum_{m=1}^{M} \cos(\frac{\pi(m-1)(i-1)}{P}) = 0$, with the exception of $i=1$. Therefore, diffusion over an Erdös-Renyí graph, as captured by multiplication by the scaled adjacency matrix, can be encoded with a single coefficient. Mathematically, the point becomes clear by investigating the resolvent equation corresponding to an ER kernel

$$\lambda f(x) = \int_0^1 p\mathbf{1}(x,y)f(y)\,\mathrm{d}y \tag{45}$$

the solutions to this equation are clearly $\lambda = p$, which corresponds to the eigenfunction $\mathbf{1}(x)$, and $\lambda = 0$ with infinite multiplicity, which corresponds to any function which integrates to 0.

Fig. 5 shows the Fourier-Galerkin shift operator corresponding to an ER random graph model with parameter $p = 0.5$. A total of 5 Tchebyshev functions and 10 quadrature points were used in the evaluation of (41). It can be observed that the matrix $\mathbf{W}$ has only one non-zero row, specifically the first row, indicating that the range space of the operator is the set of constant functions. To demonstrate that the matrix $\mathbf{W}$ is the correct linear map corresponding to the kernel function $p = 0.5$ we first must assume a function $f(y)$. We thus assume the function $f(y) = y$ and evaluate the integral $\int_0^1 p\mathbf{1}(x,y)y\mathrm{d}y$ having the obvious solution $\frac{1}{4}\mathbf{1}(x)$. More details on the solution of resolvent equations and their relationship to the graphon can be found in [9], [10]. A close correspondence between the analytical solution, the Fourier-Galerkin solution, and the evaluation via the scaled adjacency matrix is observed in Fig. 6.

Figs. 7 and 8 show the Fourier-Galerkin shift operator corresponding to an exponential random graph model with kernel
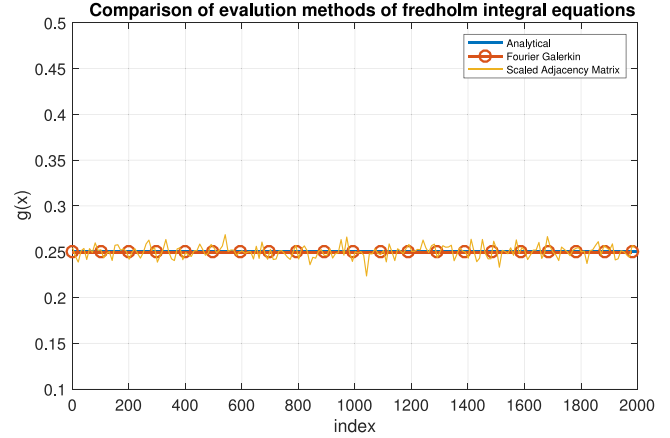

Fig. 6. Evaluation of the integral equation $\int_0^1 p\mathbf{1}(x,y)y\mathrm{d}y$ for $p = 0.5$ via analytical solution, Fourier-Galerkin method, and diffusion through the scaled adjacency matrix of an ER graph with parameter $p = 0.5$ and $N = 2000$ nodes.
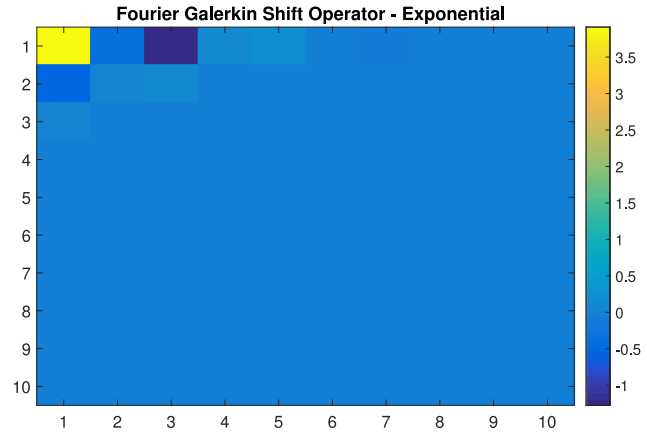

Fig. 7. Fourier-Galerkin shift operator corresponding to the graphon $e^{(-1/2(x+y))}$. There are non-zero coefficients in more than the first row.

function $e^{-\frac{1}{2}(x+y)}$ and the evaluation of the integral equation $\int_0^1 e^{-\frac{1}{2}(x+y)}y\mathrm{d}y$ which has the solution $(4 - \frac{6}{e^{1/2}})e^{-1/2x}$, respectively. A close correspondence between the evaluation of the integral equation by the scaled adjacency matrix, Fourier-Galerkin method, and the analytical solution is again observed.

Remark 4: The convergence of the scaled adjacency matrix to its limit object occurs as the number of nodes and connections in the network grows. The convergence of the solution of the integral equation by quadrature to the analytical solution only depends on the complexity of the kernel. Thus, given an arbitrarily large graph, given the Kernel function explaining the graph, the Fourier-Galerkin shift operator achieves a compression of constant order (with respect to the graph size) of the random graph. Additionally, given the properties of approximation by Tchebyshev functions, the approximation error of the solution of the Fredholm integral equation can be made arbitrarily small. The error in approximation of the graph filter by the graphon analog is thus dominated by the convergence of the scaled adjacency matrix to the limit object, the Fredholm integral equation.
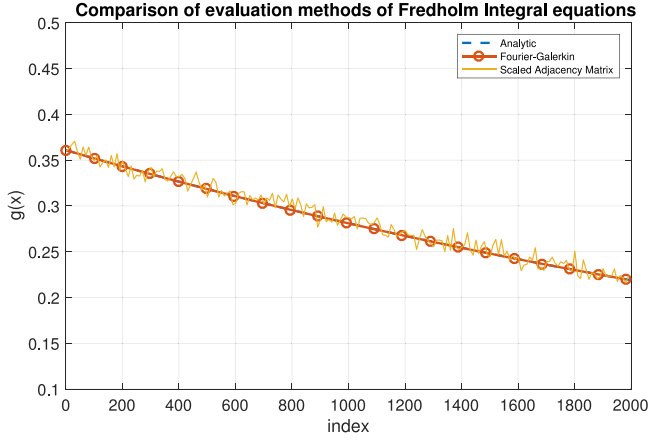
Fig. 8. Evaluation of the integral equation $\int_0^1 e^{(-1/2(x+y))} y \mathrm{d}y$ via analytical solution, Fourier-Galerkin method, and diffusion through the scaled adjacency matrix with $N = 2000$ nodes.

VIII. GRAPHON FILTER DESIGN

In this section, we address the design of simple graphon filters given the graphon. As was shown in the previous section, the Fourier-Galerkin shift operator is a linear map between the Tchebyshev representation of the input function to a Tchebyshev representation of the output function (the solution of the integral equation (6)). In essence, filtering in this context is thus controlling what Tchebyshev representations are possible as an output of the evaluation of the filter operator. As such, the goal of designing a graphon filter is the attenuation of one or more of the Tchebyshev coefficients of the output.

The first step of the graphon filtering algorithm is to create a representation of the input graph signal $\mathbf{x}$. To do this, we set $f(x) = \sum_{i=1}^{N} [\mathbf{x}]_i b_i'(x)$, where $b_i'(x)$ is the step basis functions used in Section VI supported on $[-1, 1]$. Then, to produce $\mathbf{f}$ we use equation (35). The ordering of $\mathbf{x}$ must reflect the ordering of the graphon $\mathcal{W}$ in order to avoid the problems described in Section 26.

The graph "frequencies" correspond to the various Tchebyshev functions respectively. Thus, an "all pass" graphon filter is simply the identity matrix. A "frequency" band is analogous to selecting a certain number of rows of the identity matrix. Thus, the ideal graphon filter response is contained in the diagonal of a matrix $\mathbf{D}$. Using the graphon filter definition proposed in section IV in combination with the convergence of the scaled adjacency matrix to a Fredholm integral equation investigated in the previous section, a simple graphon filter $\mathbf{H}$ is a polynomial in the Fourier-Galerkin shift operator.

$$\mathbf{H} = \sum_{k=0}^{K-1} h_k \mathbf{W}^k \qquad (46)$$

The design of the simple graph filter can be stated as a simple least squares problem

$$\min_{\{h_k\}}, \left\| \sum_{k=0}^{K-1} h_k \mathbf{W}^k - \mathbf{D} \right\|_F^2 \qquad (47)$$

Algorithm 2: Graphon Filter ($\mathbf{W}$, order k, $\mathbf{D}$) Complexity $\mathcal{O}(N^6)$.

> Produce $\mathbf{f}$ with (35)
> $\mathbf{A} = [\mathrm{vec}(\mathbf{W}^0)\mathrm{vec}(\mathbf{W}) \cdots \mathrm{vec}(\mathbf{W}^{k-1})]$, $\mathbf{b} = \mathrm{vec}(\mathbf{D})$
> $\min_{\mathbf{h}} \|\mathbf{A}\mathbf{h} - \mathbf{b}\|$
> return $\mathbf{h} = \mathbf{A}^\dagger \mathbf{b}$
> $\mathbf{g} = \mathbf{H}\mathbf{f}$
> $\mathbf{y}' = \sum_{n=1}^{N} [\mathbf{g}]_n \mathbf{t}_n$

---

which can be restated in matrix vector form via the vectorization operator as

$$\min_{\mathbf{h}}, \|\mathbf{A}\mathbf{h} - \mathbf{b}\|_2^2$$

where the k-th column of $\mathbf{A}$ is $\mathrm{vec}(\mathbf{W}^{k-1})$ and $\mathbf{b} = \mathrm{vec}(\mathbf{D})$, which is solved optimally via least squares. Due to the polynomial nature of the columns of $\mathbf{A}$, the matrix $\mathbf{A}$ will be highly ill-conditioned resulting in inaccurate coefficients of extreme magnitude. Thus, for reasons of numerical stability, the graphon filter should be designed via truncated SVD.

We remind the reader that here N is the number of Tchebyshev functions used to construct $\mathbf{W}$, not the number of nodes on a graph. Thus N^6 is in the order of millions at most. To move back from the graphon "frequency" domain to the graph node domain, we use resampling. That is, given the Tchebyshev representation of the input f, and the implementation of the filter $\mathbf{g} = \mathbf{H}f$, the node domain representation would be given by the T dimensional vector

$$\mathbf{y}' = \sum_{n=1}^{N} [\mathbf{g}]_n \mathbf{t}_n \qquad (48)$$

where $\mathbf{t}_n$ is n-th first-kind Tchebyshev function sampled at T points, where T is the size of the graph, uniformly over $[-1, 1]$, and $\mathbf{y}'$ denotes the resampled node domain graphon filtered output, as opposed to the graph filtered output $\mathbf{y} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k \mathbf{x}$. Crucially, we point out that the coefficients designed by the graphon filtering algorithm are the coefficients for the graph filtering operation. These steps are summarized in Algorithm 2.

A. Filter Response

One interpretative benefit of the graphon filter algorithm is that filtering is accomplished in terms of a standard basis. Specifically, while different graphons (and therefore, the random graphs drawn from the graphons) may have different representations in this standard basis, but the output will always be in the standard basis. In this sense, a "low pass" or "high pass" filter has the same meaning regardless of the graph structure. The output of the filter will only have non-zero coefficients corresponding to the desired basis functions. This does not have a parallel to graph filtering using the eigenbasis of a graph shift operator, wherein the graph filtering operation and output are graph dependent. In contrast, the graphon filtering algorithm is not only agnostic with respect to the particular graph, but also with respect to the graphon. Only the reachability of $\mathbf{D}$ depends on the graphon.
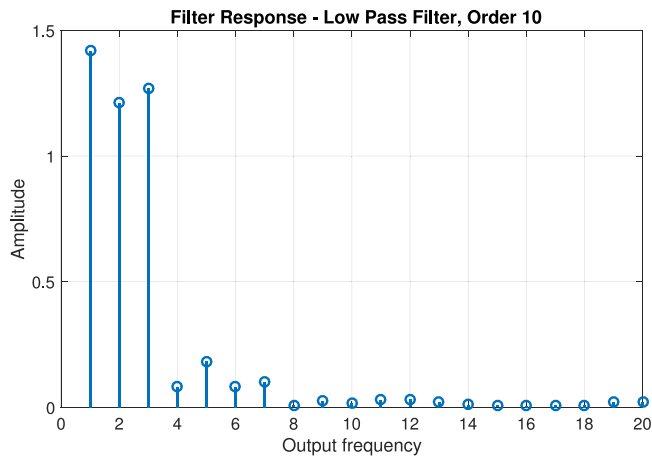
Fig. 9. Low-pass filter response of a 10th order graphon filter implemented over a kernel of $e^{-5|x-y|}$. The filter was designed to preserve the first three graphon frequencies while nullifying any higher frequencies.
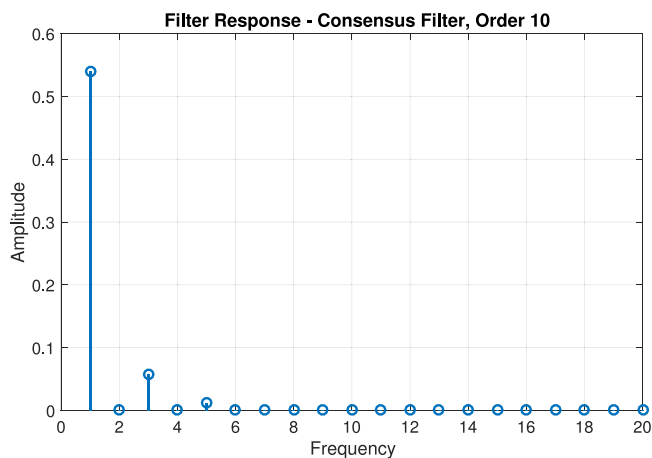

Fig. 10. Consensus filter response of a 10th order graphon filter implemented over a kernel of $e^{-5|x-y|}$. The filter was designed to have as small of an output as possible for all except the constant frequency.

Additionally, the matrix operator provides a clear interpretation of the frequency response of the graphon filter. Since the amplitudes of the input frequencies are the j-th coefficient of the input vector $\mathbf{f}$, the frequency response of the graphon filter $\mathbf{H}$ is the output of $\mathbf{H1}$. This can easily be seen by considering the response to each frequency individually. The response of $\mathbf{H}$ to the j-th frequency is obviously $\mathbf{He}_j$, and the response to all frequencies is just the sum over all frequencies j, which is exactly $\mathbf{H1}$. In contrast to classical signal processing, the frequencies are discrete. There are no graph frequencies "between" the Tchebyshev frequencies. Thus the frequency response is not a continuous function, but a vector.

Fig. 9 shows the frequency response of a "low-pass" graphon-filter designed for the graphon $e^{-5|x-y|}$. Fig. 10 shows the frequency response corresponding to a "consensus" filter designed for the same graphon. Comparing the two figures we can clearly see that the graphon filtering algorithm is able to both preserve and nullify the selected frequencies and thus implement a variety of filtering tasks.
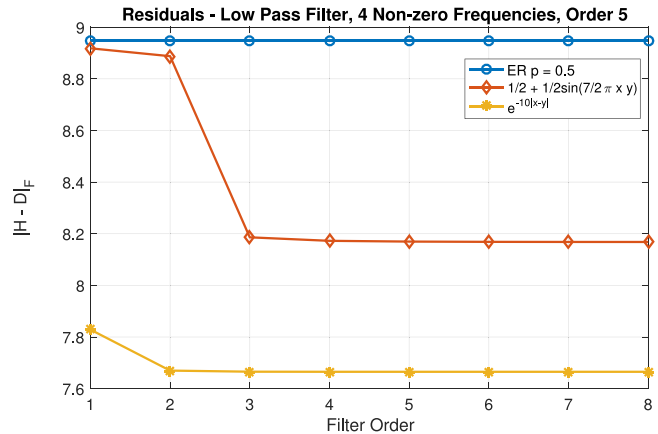

Fig. 11. Residuals of the least squares filter design algorithm for a Low Pass filter. A large difference in the ability of different random graph models to implement the "low pass" filtering task is observed.

IX. SIMULATIONS

In order to demonstrate the graphon filtering algorithm, the effects of the graphon on the design and reachability of certain filters, as well as the convergence results presented in this paper, we examine three graphons: and ER-graphon with $p = 0.5$, $\frac{1}{2} + \frac{1}{2}\sin(\frac{7}{2}\pi xy)$, and $e^{-10|x-y|}$. Two separate filtering tasks are implemented: "low-pass" filtering, wherein the aim is to preserve several low frequencies, and consensus filtering where the aim is to preserve only the constant frequency present in the graph signal. All graph filters are implemented with the coefficients designed by Algorithm 2, without knowledge of the specific graph sample. The input function is $f(x) = x + sin(x)$. Empirical samples are drawn with $N = 2000$ nodes.

A. Low Pass Filter

We implement the graphon filtering algorithm described in Algorithm 2 for the aforementioned three graphons, for orders 1 through 8, in order to implement a low-pass graphon filter. The matrix $\mathbf{D}$ in the least squares formulation in Algorithm 2 was here specified to be $\mathrm{diag}(\mathbf{b})$ where $\mathbf{b} = [1, 5, 5, 10, 0, \ldots, 0]$. Fig. 11 shows the residuals from the least squares filter design algorithm.

Inspecting Fig. 11 it becomes clear that the underlying graphon has an enormous effect on the ability to match the desired graphon filter operator. With an ER graphon, Algorithm 2 is completely unable to match the desired graphon filter operator $\mathbf{D}$. This is because the related Fourier-Galerkin shift operator has no higher order frequency content. Thus all higher order frequencies are filtered out by default, leaving only the constant frequency. Other graphon models allow Algorithm 2 to match the ideal graphon filter operator, specifically because these models have higher frequency content. The graphon with the highest frequency content here considered, $e^{-10|x-y|}$ is the best able to match the ideal graphon filter. Moreover, the graphon with the higher frequency content is able to implement its filter design with a lower order filter.

Fig. 12 shows the output of low pass filters designed over the aforementioned three graphons, with order 5. As Theorem
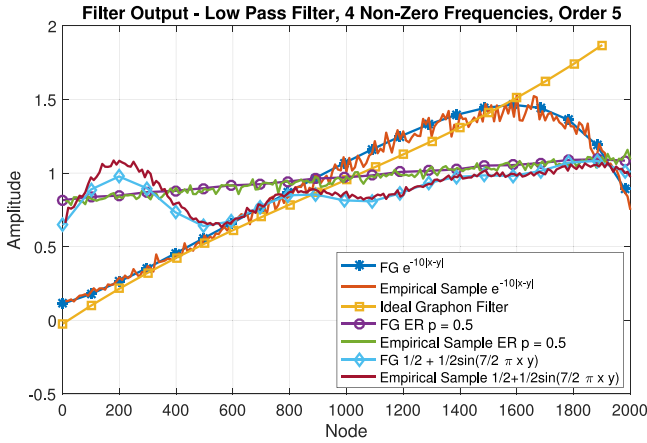
Fig. 12. The output of 3 order 5 graphon filters. The first 4 graphon frequencies were to be preserved while the rest were to be nullified. The input function was $y + sin(y)$ over the domain $[0, 1]$.
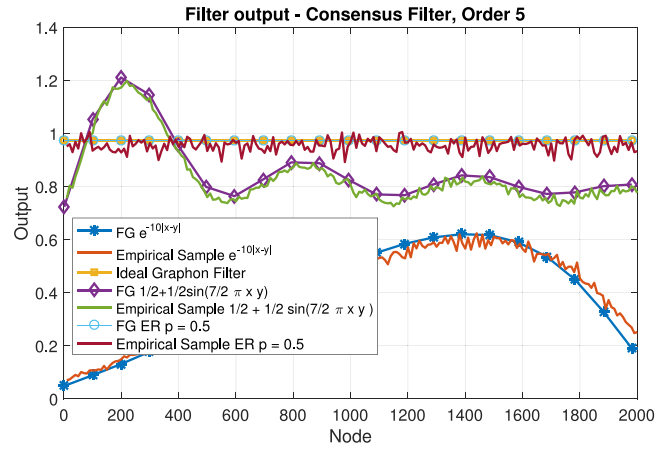

Fig. 14. The output of three order 5 graphon consensus filters. The input function was $y + sin(y)$ over the domain $[0, 1]$.
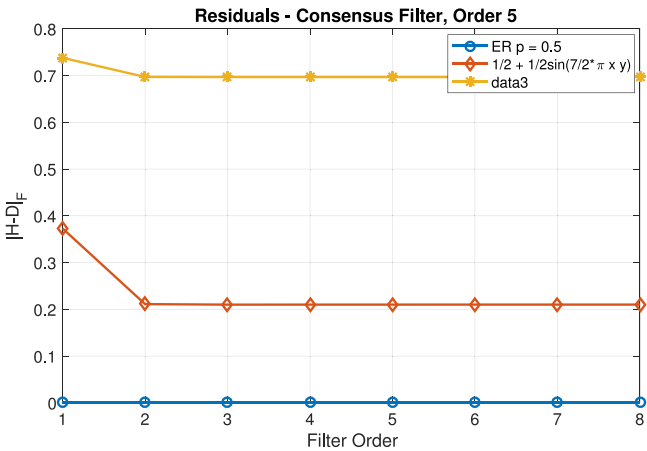

Fig. 13. Residuals of the least squares filter design algorithm for a consensus filter. A large difference in the ability of different random graph models to implement the "consensus" filtering task is observed.

?? predicts, the output of all graphon filters closely track the empirical samples. The output of the ideal filter is plotted for comparison purposes. It can be observed that the output of the graphon filter designed over $e^{-10|x-y|}$ fits closest to the output of the ideal filter. Whereas the other two filters have much lower frequency content they have much more difficulty in matching the ideal filter output.

B. Consensus Filtering

We again implement the graphon filtering algorithm described in Algorithm 2 but this time in order to implement a consensus filtering task. The Algorithm is again implemented over the same 3 graphons. The ideal graphon filter for this operation is $\mathbf{D} = \mathrm{diag}(\mathbf{e}_1)$. Fig. 13 shows the residuals from the least squares consensus filter design algorithm.

Contrary to what was observed in the low pass example, the graphon filtering algorithm is able to almost-perfectly implement the consensus filter. In direct contrast to the low pass filtering task, higher frequency content in the Fourier-Galerkin shift operator is undesirable as these higher order frequencies

will need to be nullified. In this case the higher order frequencies cannot be effectively nullified using a simple graphon filter.

Fig. 14 shows the output of graphon consensus filters designed over the aforementioned graphons compared with the output of the ideal graphon filter. Again we see a close correspondence between the empirical graph filter and the prediction made by the graphon filter. Contrary to the previous example, random graphs drawn from the graphon with the highest frequency content, $e^{-10|x-y|}$, have the most trouble performing the consensus task, while the graphon with the least frequency content, the Erdös-Renyí, performs the consensus task with a high degree of accuracy. The excess frequency content presents a challenge for the simple graphon filtering procedure. In order to overcome the structural challenges presented by the various graphon filtering tasks, the action of the nodes must be altered in some way: either via a node or edge varying filter, or a different shift action, or both.

C. Unknown Graphon and Unlabeled Graphs

We now address how to apply the method to a graph drawn from an unknown graphon. The only assumption used in this section is that the graph is drawn from a graphon, the graphon here being $e^{-4(x+y)}$. There are two key differences in the exposition in figures 15–17 from those of Section 8. First, the "true" graphon is not used to produce the estimate of the FG shift operator. One must first be estimated. Secondly, the graph is unlabeled, that is we don't know the samples μ_i. As was shown in section 7, even if we had the "true" graphon, without knowing the order of the samples μ_i, the empirical sample of the graph does not converge to the true graphon, and thus, neither does its diffusion. In order to resolve these issues, we first sort the graph sample by degree. As was shown in [28], centrality measures such as degree also converge between the graph sample estimates. Then, the USVT method of [29] is used to provide an estimate of the graphon. Fig. 16 shows the estimate of the underlying graphon via the method [29] using the single largest eigenvalue and its associated eigenvector, specifically $\mathbf{W}_{svd} = \mathbf{u}_1 \mathbf{u}_1^T \lambda_1$. Then $\mathbf{W}_{svd}$ becomes a lookup table of the
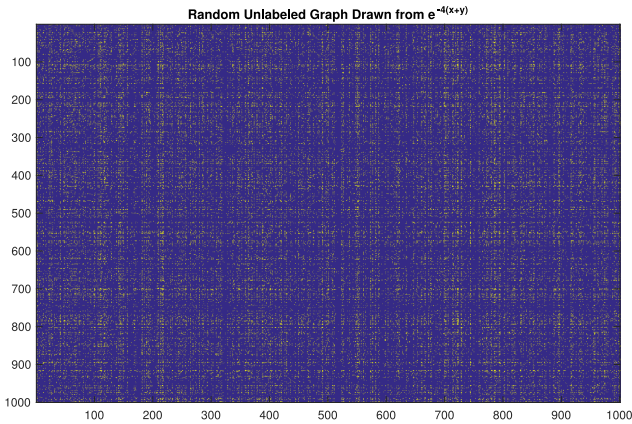
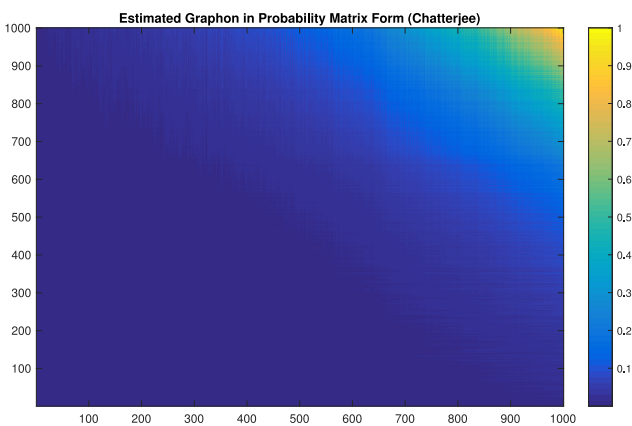Fig. 15.   Random unlabeled graph drawn from $e^{-4(x+y)}$.

Fig. 16.   Estimate of underlying graphon by first degree sorting the observed graph and then using the USVT technique of [29]. The underlying graphon is $e^{-4(x+y)}$.
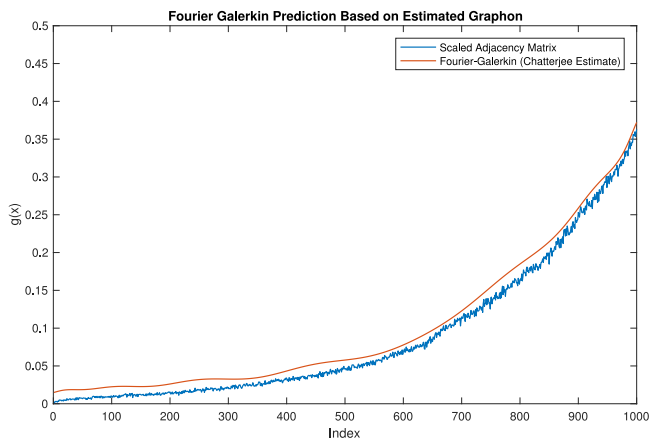

Fig. 17.   Comparison between the FG diffusion prediction based on the estimated graphon and the output of the degree-sorted scaled-adjacency matrix.

underlying graphon $\mathcal{W}$ for the quadrature procedure outlined in Section 8. Specifically, the value of $[\mathbf{W}_{svd}]_{i,j}$ is assumed to be the sample of $\mathcal{W}$ at $\frac{i}{N}, \frac{j}{N}$ where NxN is the dimension of $\mathbf{W}_{svd}$. Quadrature points lying between values of $\frac{i}{N}$ and $\frac{i+1}{N}$ are interpolated from the values of $\mathbf{W}_{svd}$. Fig. 17 demonstrates the close correspondence between the output of the FG shift

operator based on the estimate $\mathbf{W}_{svd}$ and the degree sorted scaled adjacency matrix. We note that the eigenvectors of the graph shift operator can be deduced from the graph signal itself using, for example, the method developed in [30].

X. CONCLUSION

In this report we defined several concepts. Firstly, we defined the concept of a graphon filter, both in discrete and continuous time. Secondly, the sparse approximation of Fredholm integral equations allowed us to define the relationship between diffusion of a signal on a graph, represented by the graph shift operator, and a Fredholm integral equation with the empirical graphon as its kernel. The approximation error was shown to be a function of an integral quadrature error, and the spectral radius of the empirical graphon. Finally, it was shown that the solution of the Fredholm integral equation defined by the empirical graphon converges to the solution of the Fredholm integral equation defined by the continuous graphon. Simulation results verified these theoretical results.

REFERENCES

[1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending highdimensional 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. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7 pp. 1644–1656, Apr. 2013.

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

[4] D. I. Shuman, P. Vandergheynst, D. Kressner, and P. Frossard, "Distributed signal processing via chebyshev polynomial approximation," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 4, pp. 736–751, Dec. 2018.

[5] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 260–291, 2016.

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

[7] S. Segarra, A. 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.

[8] M. Coutino, E. Isufi, and G. Leus, "Advances in distributed graph filtering," *IEEE Trans. Signal Process.* vol. 67, no. 9 pp. 2320–2333, May 2019.

[9] M. W. Morency and G. Leus, "Signal processing on kernel based random graphs," in *Proc. 25th Eur. Signal Process. Conf.*, 2017, pp. 365–369.

[10] L. Ruiz, L. F. O. Chamon, and A. Ribeiro, "The graphon fourier transform," 2019, *arXiv:1910.10195*.

[11] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *Proc. 20th Int. Conf. Mach. Learn.*, 2003, pp. 912–919.

[12] K. Avrachenkov, A. Mishenin, P. Gonalves, and M. Sokol, "Generalized optimization framework for graph-based semi-supervised learning," in *Proc. SIAM Int. Conf. Data Mining*, 2012, pp. 966–974

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

[14] R. Albert and A. L. Barabasi, "Statistical mechanics of complex networks," *Rev. Modern Phys.*, vol. 74, pp. 47–97, 2002.

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

[16] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, B. Szegedy, and K. Vesztergombi, "Graph limits and parameter testing," in *Proc. 38th ACM Symp. Theory Comput.*, Seattle, WA, USA, May 21–23, 2006, pp. 261–270

[17] L. Lovász and B. Szegedy, "Limits of dense graph sequences," *Elsevier J. Combinatorial Theory, Series B*, vol. 96, no. 6, pp. 933–957, Nov. 2006. MSR Tech. Rep. MSR-TR-2004-79. [Online]. Available: ftp://ftp.research.microsoft.com/pub/tr/TR-2004-79.pdf

[18] L. Lovász, *Large Networks and Graph Limits*. vol. 60, American Mathematical Society, p. 475, 2012.

[19] C. Borgs, J. Chayes, H. Cohn, and N. Holden, "Sparse exchangeable graphs and their limits via graphon processes," *J. Mach. Learn. Res.*, vol. 18, pp. 1–71, 2018.

[20] B. Bollobas, S. Janson, and O. Riordan, "The phase transition in inhomogenious random graphs," *Random Struct. Algorithms*, vol. 31, no. 1, pp. 3–122, 2007.

[21] M. Haase, *Elementary Functional Analysis Graduate Studies in Mathematics*, vol. 156, Providence, RI, USA, 2014.

[22] B. Szegedy, "Limits of kernel operators and the spectral regularity lemma," *Elsevier Eur. J. Combinatorics.*, vol. 32, no. 7, pp. 1156–1167, Oct. 2011.

[23] O. Klopp and N. Verzelen, "Optimal estimation of graphons in cut distance," *arXiv:1703.05101v1*.

[24] W. A. Essah and L. M. Delves, "The numerical solution of first kind integral equations," *J. Appl. Comput. Math.*, vol. 27, pp. 363–387, 1989.

[25] L. M. Delves, "A fast method for the solution of fredholm integral equations," *IMA J. Inst. Math. Appl.*, vol. 20, pp. 173–182, 1977.

[26] B. K. Alpert, "A class of bases in L^2 for the sparse representation of integral equations," *SIAM J. Math. Anal.*, vol. 24, pp. 246–262, 1993.

[27] L. Lovasz and B. Szegedy, "Limits of dense graph sequences," *J. Combinatorial Theory B*, vol. 96, pp. 933–957, 2006.

[28] M. A. Medina, F. Parise, M. T. Schaub, and S. Segarra, "Centrality measures for graphons: Accounting for uncertainty in networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 520–37, Nov. 2018.

[29] S. Chatterjee, "Matrix estimation by universal singular value thresholding," *Ann. Statist.*, vol. 43, no. 1, 177–214, 2015.

[30] S. P. Chepuri and G. Leus, "Subsampling for graph power spectrum estimation," in *Proc. 9th IEEE Sensor Array Multichannel Signal Process. Workshop*, Rio de Janeiro, Brazil, Jul. 2016, pp. 1–5.

Geert Leus (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from the Katholieke Universiteit Leuven, Leuven, Belgium, in June 1996 and May 2000, respectively. He is currently an "Antoni van Leeuwenhoek" Full Professor with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands. His research interests include the broad area of signal processing, with a specific focus on wireless communications, array processing, sensor networks, and graph signal processing. He was the recipient of the 2002 IEEE Signal Processing Society Young Author Best Paper Award and the 2005 IEEE Signal Processing Society Best Paper Award. He is a Fellow of EURASIP. He was a Member-at-Large of the Board of Governors of the IEEE Signal Processing Society, the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, and the Editor-in-Chief of the *EURASIP Journal on Advances in Signal Processing*. He was also on the Editorial Boards of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE SIGNAL PROCESSING LETTERS, and the *EURASIP Journal on Advances in Signal Processing*. He is currently a Member of the IEEE Sensor Array and Multichannel Technical Committee, an Associate Editor for the *Foundations and Trends in Signal Processing,* and the Editor-in-Chief of the *EURASIP Signal Processing*.

Matthew W. Morency (Student Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Alberta, Edmonton, AB, Canada and an M.Sc. degree (with distinction) in electrical engineering from Aalto University, Espoo, Finland. From 2016 to 2020, he pursued Ph.D. studies with Delft University of Technology, Delft, The Netherlands. In late 2020, he became an Engineer with Nokia Networks, Espoo, Finland. His research interests include signal processing, with specific focus on array processing, graph signal processing, and sparse sensing. He was the recipient of the 2015 IEEE CAMSAP Best Student Paper Award and the NSERC Postgraduate Scholarship.