GRAPHON AND GRAPH NEURAL NETWORK STABILITY

Luana Ruiz, Zhiyang Wang and Alejandro Ribeiro

Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, USA

ABSTRACT

Graph neural networks (GNNs) are learning architectures that rely on knowledge of the graph structure to generate meaningful representations of large-scale network data. GNN stability is thus important as in real-world scenarios there are typically uncertainties associated with the graph. We analyze GNN stability using kernel objects called graphons. Graphons are both limits of convergent graph sequences and generating models for deterministic and stochastic graphs. Building upon the theory of graphon signal processing, we define graphon neural networks and analyze their stability to graphon perturbations. We then extend this analysis by interpreting the graphon neural network as a generating model for GNNs on deterministic and stochastic graphs instantiated from the original and perturbed graphons. We observe that GNNs are stable to graphon perturbations with a stability bound that decreases asymptotically with the size of the graph. This asymptotic behavior is further demonstrated in an experiment of movie recommendation.

Index Terms— graph neural networks, stability, graphons, graph signal processing

1. INTRODUCTION

Graph neural networks (GNNs) are layered information processing architectures alternating graph convolutional filters and pointwise nonlinearities [1–4]. Graph convolutional filters, or graph convolutions for short, are shift-and-sum operations akin to the time and spatial convolutions of convolutional neural networks (CNNs), but parametrized so as to couple the data with the graph through learnable weights. The coupling between graph and data is achieved by a graph matrix representation called graph shift operator (GSO)—an umbrella term for matrices that, like the adjacency or the Laplacian, encode the sparsity pattern of the graph [5]. Thus, when implementing graph convolutions and GNNs, the underlying assumption is that the graph structure is fully known.

In practice, however, this is not always the case. In networks that need to be measured, uncertainties around the graph structure usually come from measurement noise [6]. In physical networks, they can also be caused by faulty sensors or faded communication channels [7]. These issues are even more pronounced in large and dynamic graphs [8,9]. Two important research questions that arise are then: how is the performance of GNNs affected by changes of the underlying graph? and, what is the effect of the graph size on their stability to graph perturbations?

We answer these questions by drawing on graphons. Graphons are kernel objects that are both limits of convergent graph sequences and generative models for graphs [10–13]. From their graph limit interpretation, we can think of them as objects identifying families of graphs that are structurally similar in the sense that they share certain densities of motifs also shared by the graphon. Graphs associated with different graphon families are structurally different and can be seen as perturbations of one another. Hence, we focus on perturbations of graphons and use their generative model interpretation to instantiate a graph and its perturbation from the original and the perturbed graphons.

To carry out our stability analysis, we take a step-by-step approach relying on graphon signal processing [14, 15]. This theoretical frame-

work defines graphon signals and graphon filters as abstractions of graph data and graph filters in the limit, which we use to define graphon neural networks (WNNs) and study their stability in Section 3. From the WNN stability result (Theorem 1), we prove stability of GNNs supported on deterministic graphs instantiated from the graphon (Theorem 2) using the GNN-WNN approximation result from [16, Theorem 1]. This result is then extended to stochastic graphs in Theorem 3. The main takeaways from these theorems are that: (i) there exists a trade-off between the stability and discriminability of GNNs in the form of a restriction on the passing band of the graph convolutional filters, and (ii) stability increases asymptotically with the size of the graph. To conclude, we illustrate GNN stability in a movie recommendation example in Section 5.

Most of the GNN stability literature considers perturbations acting directly on the graph. Absolute and relative graph perturbations are considered in [17,18], which study the stability of GNNs and graph scattering transforms respectively, while [19] analyzes the stability of graph scattering transforms to permutations and perturbations of the spectra. Other related work includes studies of the transferability of GNNs on graphs instantiated from graphons [16] and generic topological spaces [20]. Most in line with this paper, [21] analyzes the stability of GNNs on standard random graph models, but it does not make the spectral considerations that lead to the key stability-discriminability trade-off we observe.

2. PRELIMINARY DEFINITIONS

In order to analyze stability to graphon perturbations, we start by reviewing GNNs and the graphon signal processing framework.

2.1. Graph neural networks

Graphs are triplets $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, where $\mathcal{V}$, $|\mathcal{V}| = n$, is the set of nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges, and $\mathcal{W} : \mathcal{E} \to \mathbb{R}$ is a function assigning weights to the edges of $\mathbf{G}$. We focus on undirected graphs, so that $\mathcal{W}(i, j) = \mathcal{W}(j, i)$. Data supported on networks is modeled as graph signals $\mathbf{x} \in \mathbb{R}^n$, where $[\mathbf{x}]_i = x_i$ corresponds to the value of the data at node i [5,22]. Graph operations on graph signals are parametrized by the graph shift operator (GSO) $\mathbf{S} \in \mathbb{R}^{n \times n}$. The GSO is a matrix that encodes the sparsity pattern of $\mathbf{G}$ by satisfying $[\mathbf{S}]_{ij} = s_{ij} \neq 0$ if and only if $i = j$ or $(i, j) \in \mathcal{E}$. Examples of GSOs include the graph adjacency matrix $[\mathbf{A}]_{ij} = \mathcal{W}(i, j)$ [23], the graph Laplacian $\mathbf{L} = \text{diag}(\mathbf{A}\mathbf{1}) - \mathbf{A}$ [22], and their normalized counterparts. In this paper, we consider $\mathbf{S} = \mathbf{A}$.

When applied to a graph signal, the GSO effectively shifts or diffuses graph data over the network edges. To see this, note that, at each node i, $[\mathbf{Sx}]_i = \sum_{j|(i,j) \in \mathcal{E}} s_{ij}x_j$. In other words, nodes j shift their data values x_j, weighted by the proximity measure s_{ij}, to neighbors i. Equipped with this notion of shift, we define the graph convolution as a weighted sum of data shifted to neighbors at most $K - 1$ hops away. Explicitly,

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

The vector $\mathbf{h} = [h_0, \ldots h_{K-1}]$ groups the filter coefficients and $*_{\mathbf{S}}$ denotes the convolution operation with GSO $\mathbf{S}$ [24,25].

Because $\mathbf{S}$ is symmetric, it can be diagonalized as $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathsf{H}}$, where $\mathbf{\Lambda}$ contains the graph eigenvalues and $\mathbf{V}$ the graph eigenvectors.

The eigenvector basis $\mathbf{V}$, which we call the graph spectral basis, is orthonormal. Substituting $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{H}}$ in (1) and calculating the change of basis $\mathbf{V}^{\mathsf{H}}\mathbf{H}(\mathbf{S})\mathbf{x}$, we get

$$\mathbf{V}^{\mathsf{H}}\mathbf{H}(\mathbf{S})\mathbf{x} = \sum_{k=0}^{K-1} h_k \boldsymbol{\Lambda}^k \mathbf{V}^{\mathsf{H}}\mathbf{x} = h(\boldsymbol{\Lambda})\mathbf{V}^{\mathsf{H}}\mathbf{x}. \qquad (2)$$

Thus, the graph convolution $\mathbf{H}(\mathbf{S})$ has spectral representation $h(\lambda) = \sum_{k=0}^{K-1} h_k \lambda^k$, which only depends on $\mathbf{h}$ and on the eigenvalues of $\mathbf{G}$.

A GNN is made up of layers consisting of a bank of filters like the one in (1) followed by a pointwise nonlinear activation function. Denoting the activation function σ, the ℓth layer of a GNN can be written as

$$\mathbf{x}_\ell^f = \sigma\left(\sum_{g=1}^{F_{\ell-1}} \mathbf{h}_\ell^{fg} *_{\mathbf{S}} \mathbf{x}_{\ell-1}^g\right) \qquad (3)$$

where $\mathbf{x}_\ell^f$ denotes the fth feature of layer ℓ for $1 \le f \le F_\ell$ and $1 \le \ell \le L$. At each layer ℓ, $F_{\ell-1}$ and F_ℓ are the numbers of input and output features respectively. The input features of the first layer, denoted $\mathbf{x}_0^g$, are the input data $\mathbf{x}^g$ for $1 \le g \le F_0$. The GNN output is $\mathbf{y}^f = \mathbf{x}_L^f$. For simplicity, this GNN can be succinctly represented as the map $\mathbf{y} = \boldsymbol{\Phi}(\mathcal{H}; \mathbf{S}; \mathbf{x})$. The set $\mathcal{H}$ groups the learnable parameters of all layers and is defined as $\mathcal{H} = \{\mathbf{h}_\ell^{fg}\}_{\ell, f, g}$.

2.2. Graphon signal processing

A graphon $\mathbf{W} : [0, 1]^2 \to [0, 1]$ is a bounded, symmetric, measurable kernel representation of an undirected graph with an uncountable number of nodes. To see this, associate points $u_i, u_j \in [0, 1]$ with graph nodes i, j, and assign weights $\mathbf{W}(u_i, u_j)$ to edges (i, j). Graphons are both limit objects of convergent graph sequences and generative models for deterministic and random graphs. The first interpretation is important because it allows grouping graphs in *graph families* associated with different graphons. The second provides a way of instantiating or sampling graphs belonging to a graphon family.

Consider arbitrary unweighted and undirected graphs $\mathbf{F} = (\mathcal{V}', \mathcal{E}')$. We refer to a given $\mathbf{F}$ as a "graph motif". Homomorphisms of $\mathbf{F}$ into $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ are defined as adjacency preserving maps in which $(i, j) \in \mathcal{E}'$ implies $(i, j) \in \mathcal{E}$. Note that these maps are only a fraction of the total number of ways in which the nodes of $\mathbf{F}$ can be mapped into $\mathbf{G}$. Therefore, we can define a density of homomorphisms $t(\mathbf{F}, \mathbf{G})$, which quantifies the frequency of the motif $\mathbf{F}$ in the graph $\mathbf{G}$. Homomorphisms of graphs into graphons are defined analogously, so we can also define the density of homomorphisms of the graph $\mathbf{F}$ into the graphon $\mathbf{W}$—denoted $t(\mathbf{F}, \mathbf{W})$. A sequence $\{\mathbf{G}_n\}$ is said to converge to the graphon $\mathbf{W}$ if, for all finite, unweighted and undirected graphs $\mathbf{F}$,

$$\lim_{n \to \infty} t(\mathbf{F}, \mathbf{G}_n) = t(\mathbf{F}, \mathbf{W}). \qquad (4)$$

It can be shown that every graphon is the limit object of a convergent graph sequence, and every convergent graph sequence converges to a graphon [26, Chapter 11]. Indeed, convergent graph sequences can be obtained by drawing on the generative model interpretation of graphons. The first type of graphs that can be generated from graphons are *deterministic graphs* $\bar{\mathbf{G}}_n$. A deterministic graph on n nodes is constructed by building a regular partition $u_i = (i - 1)/n$ of $[0, 1]$ for $1 \le i \le n$ and attaching each point u_i to a node i. The GSO $\bar{\mathbf{S}}_n$ is then defined as

$$[\bar{\mathbf{S}}_n]_{ij} = \bar{s}_{ij} = \mathbf{W}(u_i, u_j) \qquad (5)$$

i.e., edges (i, j) have weight $\mathbf{W}(u_i, u_j)$.

The second type of graphs that can be generated from a graphon are *stochastic graphs* $\mathbf{G}_n$. These are unweighted graphs (i.e., whose edge weights are either 0 or 1) obtained by sampling edges (i, j) as Bernoulli random variables with probability $[\bar{\mathbf{S}}_n]_{ij} = \bar{s}_{ij}$. Explicitly,

$$[\mathbf{S}_n]_{ij} \sim \text{Bernoulli}([\bar{\mathbf{S}}_n]_{ij}). \qquad (6)$$

Also known as $\mathbf{W}$-random graphs, these graphs converge to the graphon $\mathbf{W}$ with probability 1 [26, Chapter 10.1].

In the same way that we can instantiate graphs from graphons, the *graphon induced by a graph* $\mathbf{G}_n$ can always be defined. For instance, consider the deterministic graph $\bar{\mathbf{G}}_n$ and let $I_i = [(i - 1)/n, i/n]$, $1 \le i \le n$, be a regular partition of the unit interval. The graphon induced by $\bar{\mathbf{G}}_n$ is denoted $\mathbf{W}_{\bar{\mathbf{G}}_n} = \bar{\mathbf{W}}_n$ and defined as

$$\bar{\mathbf{W}}_n(u, v) = [\bar{\mathbf{S}}_n]_{ij}\, \mathbb{I}(u \in I_i) \times \mathbb{I}(v \in I_j) \qquad (7)$$

where $\mathbb{I}$ is the indicator function and $\times$ the Cartesian product. The graphon $\mathbf{W}_{\mathbf{G}_n} = \mathbf{W}_n$ induced by the stochastic graph $\mathbf{G}_n$ can be defined analogously.

2.2.1. Graphon signals and graphon convolutions

Graphon signals are defined as functions $X \in L_2([0, 1])$. They can be seen as limits of graph signals supported on graphs converging to a graphon [27], or as generating models for graph signals $\mathbf{x}_n$ supported on the deterministic graph $\bar{\mathbf{G}}_n$ (5) or on the stochastic graph $\mathbf{G}_n$ (6) and given by

$$[\mathbf{x}_n]_i = X(u_i) \qquad (8)$$

where $u_i = (i - 1)/n$ for $1 \le i \le n$. Graphon signals can also be induced by graph signals. The graphon signal induced by a graph signal $\mathbf{x}_n \in \mathbb{R}^n$ is given by [27]

$$X_n(u) = [\mathbf{x}_n]_i\, \mathbb{I}(u \in I_i) \qquad (9)$$

where $I_i = [(i - 1)/n, i/n]$, $1 \le i \le n$, is the regular partition of $[0, 1]$.

The integral operator with kernel $\mathbf{W}$ defines a basic operation for graphon signals. Explicitly, this operator is given by

$$(T_{\mathbf{W}}X)(v) := \int_0^1 \mathbf{W}(u, v)X(u)du \qquad (10)$$

and we call it graphon shift operator (WSO) in analogy with the GSO. Since $\mathbf{W}$ is bounded and symmetric, $T_{\mathbf{W}}$ is a self-adjoint Hilbert-Schmidt (HS) operator. Hence, the graphon can be expressed in the operator's spectral basis as $\mathbf{W}(u, v) = \sum_{i \in \mathbb{Z} \setminus \{0\}} \lambda_i \varphi_i(u)\varphi_i(v)$ and we can rewrite $T_{\mathbf{W}}$ as

$$(T_{\mathbf{W}}X)(v) = \sum_{i \in \mathbb{Z} \setminus \{0\}} \lambda_i \varphi_i(v) \int_0^1 \varphi_i(u)X(u)du. \qquad (11)$$

The eigenvalues λ_i, $i \in \mathbb{Z} \setminus \{0\}$, are ordered as $1 \ge \lambda_1 \ge \lambda_2 \ge \ldots \ge \ldots \ge \lambda_{-2} \ge \lambda_{-1} \ge -1$ following their sign and in decreasing order of absolute value. As $|i| \to \infty$, they accumulate around 0 due to the spectral theorem for self-adjoint HS operators [28, Theorem 3, Chapter 28].

Analogously to graph convolutions, graphon convolutions are defined as shift-and-sum operations in which each shift corresponds to an application of the WSO. The graphon convolution $\mathbf{h} *_{\mathbf{W}} X$ is given by

$$\mathbf{h} *_{\mathbf{W}} X = \sum_{k=0}^{K-1} h_k (T_{\mathbf{W}}^{(k)} X)(v) = (T_{\mathbf{H}}X)(v) \quad \text{with}$$
$$(T_{\mathbf{W}}^{(k)} X)(v) = \int_0^1 \mathbf{W}(u, v)(T_{\mathbf{W}}^{(k-1)}X)(u)du \qquad (12)$$

where $T_{\mathbf{W}}^{(0)} = \mathbf{I}$ is the identity operator and $\mathbf{h}$ is the vector of filter coefficients $\mathbf{h} = [h_0, \ldots, h_{K-1}]$ [14]. Alternatively, we can use the graphon spectral decomposition to write $T_{\mathbf{H}}$ as

$$(T_{\mathbf{H}}X)(v) = \sum_{i \in \mathbb{Z} \setminus \{0\}} \sum_{k=0}^{K-1} h_k \lambda_i^k \varphi_i(v) \int_0^1 \varphi_i(u)X(u)du$$
$$= \sum_{i \in \mathbb{Z} \setminus \{0\}} h(\lambda_i)\varphi_i(v) \int_0^1 \varphi_i(u)X(u)du. \qquad (13)$$

Observe that $T_{\mathbf{H}}$ has spectral representation $h(\lambda) = \sum_{k=0}^{K-1} h_k \lambda^k$, which only depends on the graphon eigenvalues and the coefficients h_k.

3. GRAPHON NEURAL NETWORKS

Leveraging the definition of the convolution operation for graphon signals, we can analogously define graphon neural networks as layered architectures consisting of banks of graphon convolutional filters and nonlinear activation functions. Denoting the nonlinear activation function σ, the ℓth layer of a graphon neural network can be written as [16]

$$X_\ell^f = \sigma \left(\sum_{g=1}^{F_{\ell-1}} \mathbf{h}_\ell^{fg} *_{\mathbf{w}} X_{\ell-1}^g \right) \qquad (14)$$

where $1 \leq f \leq F_\ell$ and F_ℓ is the number of features at the output of layer ℓ. For a WNN with L layers, the WNN output is given by $Y^f = X_L^f$. The input features at the first layer, X_0^g, are the input data X^g for $1 \leq g \leq F_0$. Like the GNN, this WNN can be written more succinctly as the map $Y = \mathbf{\Phi}(\mathcal{H}; \mathbf{W}; X)$, where $\mathcal{H} = \{\mathbf{h}_\ell^{fg}\}_{\ell,f,g}$.

In the map $\mathbf{\Phi}(\mathcal{H}; \mathbf{W}; X)$, $\mathcal{H}$ is independent of the graphon $\mathbf{W}$. More importantly, in the GNN $\mathbf{\Phi}(\mathcal{H}; \mathbf{S}_n; \mathbf{x}_n)$ and in the WNN $\mathbf{\Phi}(\mathcal{H}; \mathbf{W}; X)$ the sets $\mathcal{H}$ can be the same. This allows building GNNs from WNNs by instantiating graphs $\mathbf{G}_n$ and signals $\mathbf{x}_n$ from graphons $\mathbf{W}$ and signals X. Similarly, we can define WNNs induced by GNNs. The WNN induced by a GNN $\mathbf{\Phi}(\mathcal{H}; \mathbf{S}_n; \mathbf{x}_n)$ is given by $\mathbf{\Phi}(\mathcal{H}; \mathbf{W}_n; X_n)$, where $\mathbf{W}_n$ is the graphon induced by $\mathbf{G}_n$ (7) and X_n is the graphon signal induced by $\mathbf{x}_n$ (9). The definition of WNNs induced by GNNs is useful because it allows comparing GNNs with WNNs. This will be leveraged to break the stability analysis of GNNs down into simpler steps, the first of which is the stability analysis of WNNs.

3.1. Stability of WNNs

Let $Y = \mathbf{\Phi}(\mathcal{H}; \mathbf{W}; X)$ be a WNN on the graphon $\mathbf{W}$ and suppose that $\mathbf{W}$ is perturbed by a symmetric kernel $\mathbf{A}$ to yield the graphon $\mathbf{W}' = \mathbf{W} + \mathbf{A}$ and the WNN $Y' = \mathbf{\Phi}(\mathcal{H}; \mathbf{W}'; X)$. The stability of this WNN can be quantified by bounding $\|Y - Y'\|$. Let us start by introducing the constants $n_c^{(p)}$ and $\delta_c^{(pq)}$ which will play an important role in the upcoming analysis.

Definition 1. *Let* $\mathbf{W}^{(p)}$ *and* $\mathbf{W}^{(q)}$ *be two graphons with labels* (p) *and* (q) *assigned arbitrarily. Let* $c \in [0,1]$. *We define* $n_c^{(p)}$ *and* $\delta_c^{(pq)}$ *as*

$$n_c^{(p)} = |\mathcal{C}^{(p)}| \quad \text{with} \quad \mathcal{C}^{(p)} = \{i \mid |\lambda_i^{(p)}| \geq c\}$$

$$\delta_c^{(pq)} = \min_{i \in \mathcal{C}^{(q)}} \{|\lambda_i^{(p)} - \lambda_{i+sgn(i)}^{(q)}|, |\lambda_i^{(q)} - \lambda_{i+sgn(i)}^{(p)}|, \qquad (15)$$
$$|\lambda_1^{(p)} - \lambda_{-1}^{(q)}|, |\lambda_1^{(q)} - \lambda_{-1}^{(p)}|\}$$

where $\lambda_i^{(p)}$ *and* $\lambda_i^{(q)}$ *are the eigenvalues of* $\mathbf{W}^{(p)}$ *and* $\mathbf{W}^{(q)}$ *respectively.*

For a graphon labeled $\mathbf{W}^{(p)}$, $n_c^{(p)}$ corresponds to the number of eigenvalues with indices in the set $\mathcal{C}^{(p)}$, i.e., to the number of eigenvalues satisfying $\lambda_i^{(p)} \geq c$. For any two graphons $\mathbf{W}^{(p)}$ and $\mathbf{W}^{(q)}$, $\delta_c^{(pq)}$ corresponds to the minimum eigenvalue difference (or eigengap) between eigenvalues $\lambda_i^{(q)}$ with $i \in \mathcal{C}^{(q)}$ and $\lambda_j^{(p)}$ with j consecutive to i.

The following assumptions are needed to prove stability of WNNs in Theorem 1, whose proof we defer to the appendices [29].

AS1. *The convolutional filters* $h(\lambda)$ *are* A_2*-Lipschitz and non-amplifying, i.e.,* $|h(\lambda) - h(\mu)| \leq A_2|\lambda - \mu|$ *and* $|h(\lambda)| < 1$.

AS2. *The activation functions are normalized Lipschitz, i.e.,* $|\sigma(x_2) - \sigma(x_1)| \leq |x_2 - x_1|$, *and* $\sigma(0) = 0$.

Theorem 1 (WNN stability). *Let* $\mathbf{W} : [0,1]^2 \to [0,1]$ *be a graphon and let* $\mathbf{W}' : [0,1]^2 \to [0,1]$ *be a perturbation of this graphon such that* $\mathbf{W}' - \mathbf{W} = \mathbf{A}$ *where* $\mathbf{A}$ *is symmetric and* $\|\mathbf{A}\| \leq \varepsilon$. *Consider the* L*-layer WNNs given by* $Y = \mathbf{\Phi}(\mathcal{H}; \mathbf{W}; X)$ *and* $Y' = \mathbf{\Phi}(\mathcal{H}; \mathbf{W}'; X)$

where $F_0 = F_L = 1$ *and* $F_\ell = F$ *for* $1 \leq \ell \leq L - 1$. *Let the graphon convolutions* (12) *in all layers be such that* $h(\lambda)$ *is constant for* $|\lambda| < c$. *Then, under Assumptions 1–2, it holds*

$$\|Y' - Y\| \leq LF^{L-1} \left(A_2 + \frac{\pi n_c}{\delta_c} \right) \varepsilon \|X\|$$

where we have labeled the graphons $\mathbf{W}^{(1)} := \mathbf{W}$ *and* $\mathbf{W}^{(2)} := \mathbf{W}'$ *and defined* $n_c := n_c^{(2)}$ *and* $\delta_c := \delta_c^{(12)}$ *[cf. Definition 1].*

WNNs are thus Lipschitz stable to absolute graphon perturbations with a stability constant that depends on the hyperparameters of the WNN and on the passing band of the graphon convolutions. The dependence on the architecture is exerted by the number of layers L, the number of features F, and the convolutional filters $h(\lambda)$ through the Lipschitz constants A_2, n_c and δ_c. The constant A_2 measures the variability of the filter. The values of n_c and δ_c depend on the spectra of both $\mathbf{W}$ and $\mathbf{W}'$, but are primarily related to the length of the filter passing band $1 - c$. Since graphon eigenvalues accumulate around zero for $|i| \to \infty$, small values of c translate into large n_c, i.e., large number of eigenvalues within the passing band. They also produce small δ_c, i.e., small minimum eigengap between these eigenvalues. Hence, the larger the passing band, the worse WNN stability, pointing to a trade-off between stability and discriminability in WNNs.

4. GRAPH NEURAL NETWORK STABILITY

Theorem 1 can be combined with the GNN-WNN approximation theorem from [16, Theorem 1] to prove stability of GNNs supported on deterministic graphs instantiated from $\mathbf{W}$ and $\mathbf{W}'$. This theorem bounds the error incurred when using a GNN $\mathbf{\Phi}(\mathcal{H}; \bar{\mathbf{S}}_n; \mathbf{x}_n)$ to approximate the WNN $\mathbf{\Phi}(\mathcal{H}; \mathbf{W}; X)$, where $\bar{\mathbf{S}}_n$ and $\mathbf{x}_n$ are obtained from $\mathbf{W}$ and X as in (5) and (8). Stability of GNNs $\mathbf{\Phi}(\mathcal{H}; \bar{\mathbf{S}}_n; \mathbf{x}_n)$ to graphon perturbations is stated and proved in Theorem 2, under the same assumptions of Theorem 1 and Assumptions 3 and 4 below.

AS3. *The graphon* $\mathbf{W}$ *is* A_1*-Lipschitz, i.e.,* $|\mathbf{W}(x_2, y_2) - \mathbf{W}(x_1, y_1)| \leq A_1(|x_2 - x_1| + |y_2 - y_1|)$.

AS4. *The perturbation* $\mathbf{A} = \mathbf{W}' - \mathbf{W}$ *is* A_3*-Lipschitz.*

Theorem 2. *Let* $\bar{\mathbf{G}}_n$ *and* $\bar{\mathbf{G}}'_n$ *be* n*-node deterministic graphs* (8) *obtained from the graphons* $\mathbf{W}$ *and* $\mathbf{W}' = \mathbf{W} + \mathbf{A}$, *where* $\mathbf{A}$ *is symmetric and* $\|\mathbf{A}\| \leq \varepsilon$. *Let* $\bar{\mathbf{S}}_n, \bar{\mathbf{S}}'_n \in \mathbb{R}^{n \times n}$ *denote their GSOs. Consider the* L*-layer GNNs given by* $\bar{\mathbf{y}}_n = \mathbf{\Phi}(\mathcal{H}; \bar{\mathbf{S}}_n; \mathbf{x}_n)$ *and* $\bar{\mathbf{y}}'_n = \mathbf{\Phi}(\mathcal{H}; \bar{\mathbf{S}}'_n; \mathbf{x}_n)$ *where* $F_0 = F_L = 1$ *and* $F_\ell = F$ *for* $1 \leq \ell \leq L - 1$. *Let the graph convolutions* (1) *in all layers be such that* $h(\lambda/n)$ *is constant for* $|\lambda/n| < c$. *Then, under Assumptions 1–4, it holds*

$$\|\bar{\mathbf{y}}'_n - \bar{\mathbf{y}}_n\| \leq LF^{L-1} \left(A_2 + \frac{\pi \hat{n}_c}{\hat{\delta}_c} \right) \left(\varepsilon + \frac{B}{\sqrt{n}} \right) \|\mathbf{x}_n\|$$

where $B = \sqrt{A_1} + \sqrt{A_1 + A_3}$ *and* $\hat{n}_c$ *and* $\hat{\delta}_c$ *are defined as* $\hat{n}_c := \max_{p \in \{2,3,4\}} n_c^{(p)}$ *and* $\hat{\delta}_c := \min\{\delta_c^{(12)}, \delta_c^{(13)}, \delta_c^{(24)}\}$ *for* $\mathbf{W}^{(1)} := \mathbf{W}$, $\mathbf{W}^{(2)} := \mathbf{W}'$, $\mathbf{W}^{(3)} := \bar{\mathbf{W}}_n$ *and* $\mathbf{W}^{(4)} := \bar{\mathbf{W}}'_n$ *[cf. Definition 1].*

Proof. Leveraging the identities $\|\bar{\mathbf{y}}'_n - \bar{\mathbf{y}}_n\|_2 = \sqrt{n} \|\bar{Y}'_n - \bar{Y}_n\|_{L^2([0,1])}$ and $\|\mathbf{x}_n\|_2 = \sqrt{n} \|X_n\|_{L^2([0,1])}$, we reason in terms of the induced graphon signals $\bar{Y}'_n$ and $\bar{Y}_n$ (9) to bound $\|\bar{Y}'_n - \bar{Y}_n\|_{L^2([0,1])}$ as

$$\|\bar{Y}'_n - \bar{Y}_n\| = \|\bar{Y}'_n - Y' + Y' - Y + Y - \bar{Y}_n\|$$
$$\leq \|\bar{Y}'_n - Y'\| + \|Y' - Y\| + \|Y - \bar{Y}_n\| \qquad (16)$$

using the triangle inequality. The first and third terms on the RHS of this expression are bounded by [16, Theorem 1]. The second term is bounded by Theorem 1. Taking the maximum among the $n_c^{(p)}$ and the minimum among the $\delta_c^{(pq)}$ completes the proof. $\qquad \square$
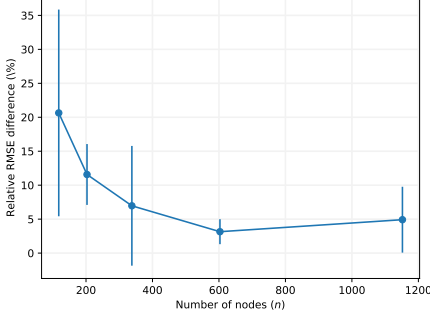
Fig. 1: Relative RMSE difference on the test set between graphs $\mathbf{G}_n$ and $\mathbf{G}'_n$ for the movie "Contact".

Theorem 2 can be further extended to GNNs $\mathbf{\Phi}(\mathcal{H}; \mathbf{S}_n; \mathbf{x}_n)$ and $\mathbf{\Phi}(\mathcal{H}; \mathbf{S}'_n; \mathbf{x}_n)$ defined on the stochastic graphs $\mathbf{G}_n$ and $\mathbf{G}'_n$ (6).

AS5. *For a fixed value of $\xi \in (0,1)$, n is such that $n > 2$,*

$$n - \frac{\log 2n/\xi}{d_{\mathbf{W}}} > 3\frac{A_1}{d_{\mathbf{W}}} \quad and \quad n - \frac{\log 2n/\xi}{d_{\mathbf{W}'}} > 3\frac{A_1 + A_3}{d_{\mathbf{W}'}} \quad (17)$$

where $d_{\mathbf{W}}, d_{\mathbf{W}'}$ denote the maximum degree of the graphons $\mathbf{W}, \mathbf{W}'$, i.e., $d_{\mathbf{W}} = \max_x \int_0^1 \mathbf{W}(x,y)dy$.

Theorem 3. *Let $\mathbf{G}_n$ and $\mathbf{G}'_n$ be n-node stochastic graphs (6) obtained from the graphons $\mathbf{W}$ and $\mathbf{W}' = \mathbf{W} + \mathbf{A}$, where $\mathbf{A}$ is symmetric and $\|\mathbf{A}\| \le \varepsilon$. Let $\mathbf{S}_n, \mathbf{S}'_n \in \mathbb{R}^{n \times n}$ denote their GSOs. Consider the L-layer GNNs given by $\mathbf{y}_n = \mathbf{\Phi}(\mathcal{H}; \mathbf{S}_n; \mathbf{x}_n)$ and $\mathbf{y}'_n = \mathbf{\Phi}(\mathcal{H}; \mathbf{S}'_n; \mathbf{x}_n)$ where $F_0 = F_L = 1$ and $F_\ell = F$ for $1 \le \ell \le L-1$. Let the graph convolutions (1) in all layers be such that $h(\lambda/n)$ is constant for $|\lambda/n| < c$. Then, under Assumptions 1–5, it holds with probability at least $1 - \xi$*

$$\|\mathbf{y}'_n - \mathbf{y}_n\| \le LF^{L-1}\left(A_2 + \frac{\pi \check{n}_c}{\check{\delta}_c}\right)\left(\varepsilon + \frac{B + 4\sqrt{\log 2n/\xi}}{\sqrt{n}}\right)\|\mathbf{x}_n\|$$

where $B = \sqrt{A_1} + \sqrt{A_1 + A_3}$ and $\check{n}_c$ and $\check{\delta}_c$ are defined as $\check{n}_c := \max\{\max_{p \in \{5,6\}} n_c^{(p)}, \hat{n}_c\}$ and $\check{\delta}_c := \min\{\delta_c^{(35)}, \delta_c^{(46)}, \hat{\delta}_c\}$ for $\mathbf{W}^{(3)} := \bar{\mathbf{W}}_n$, $\mathbf{W}^{(4)} := \bar{\mathbf{W}}'_n$, $\mathbf{W}^{(5)} := \mathbf{W}_n$, $\mathbf{W}^{(6)} := \mathbf{W}'_n$ [cf. Definition 1] and $\hat{n}_c, \hat{\delta}_c$ as in Theorem 2.

Proof of Theorem 3. The proof follows trivially from Theorem 2 and the following lemma, whose proof we defer to the appendices [29].

Lemma 1. *Let $\bar{\mathbf{G}}_n$ be an undirected graph with $\bar{\mathbf{S}}_n \in [0,1]^{n \times n}$. Let $\mathbf{G}_n$ be a stochastic graph obtained from $\bar{\mathbf{G}}_n$ as in (6). Consider the L-layer GNNs given by $\bar{\mathbf{y}}_n = \mathbf{\Phi}(\mathcal{H}; \bar{\mathbf{S}}_n; \mathbf{x}_n)$ and $\mathbf{y}_n = \mathbf{\Phi}(\mathcal{H}; \mathbf{S}_n; \mathbf{x}_n)$ where $F_0 = F_L = 1$ and $F_\ell = F$ for $1 \le \ell \le L-1$. Let the graph convolutions (1) in all layers be such that $h(\lambda/n)$ is constant for $|\lambda/n| < c$. Then, under Assumptions 1,2 and 5, it holds*

$$\|\mathbf{y}_n - \bar{\mathbf{y}}_n\| \le LF^{L-1}\left(A_2 + \frac{\pi n_c^{(q)}}{\delta_c^{(pq)}}\right)\left(2\sqrt{\log 2n/\xi}\sqrt{n}\right)\|\mathbf{x}_n\|$$

where $n_c^{(q)}$ and $\delta_c^{(pq)}$ are as in Definition 1 for $\mathbf{W}^{(p)} = \bar{\mathbf{W}}_n$, the graphon induced by $\bar{\mathbf{G}}_n$, and $\mathbf{W}^{(q)} = \mathbf{W}_n$, the graphon induced by $\mathbf{G}_n$ (7).

Use the triangle inequality to write

$$\|\mathbf{y}_n - \check{\mathbf{y}}'_n\| = \|\mathbf{y}_n - \bar{\mathbf{y}}_n + \bar{\mathbf{y}}_n - \bar{\mathbf{y}}'_n + \bar{\mathbf{y}}'_n - \mathbf{y}'_n\|$$
$$\le \|\mathbf{y}_n - \bar{\mathbf{y}}_n\| + \|\bar{\mathbf{y}}_n - \bar{\mathbf{y}}'_n\| + \|\bar{\mathbf{y}}'_n - \mathbf{y}'_n\|. \quad (18)$$

The first and third terms on the RHS of this expression are bounded by Lemma 1. The second term is bounded by Theorem 2. Taking the maximum of the $n_c^{(p)}$ and the minimum of the $\delta_c^{(pq)}$ completes the proof.

□

From Theorems 2 and 3, we conclude that GNNs are asymptotically stable to absolute graphon perturbations on both deterministic and stochastic graphs associated with a graphon. This is an important result, as absolute graphon perturbations can be used to model a variety of graph perturbations. We further observe that the stability bound depends on the graphon variability A_1 and on the perturbation variability A_3. It is also inversely proportional to the size of the graph. In particular, as $n \to \infty$ the GNN stability bound converges to the stability bound of the WNN [cf. Theorem 1]. We demonstrate this asymptotic behavior in the numerical experiment of Section 5.

Like in WNNs, in GNNs stability decreases with the architecture depth L and width F. It also depends on the variability A_2 and on the passing band $1 - c$ of the convolutional filters $h(\lambda)$. The wider the passing band, the greater the number of eigenvalues $\hat{n}_c, \check{n}_c$ larger than c, and the smaller the minimum eigengap $\hat{\delta}_c, \check{\delta}_c$. Therefore, GNNs also exhibit a trade-off between stability and discriminability. In practice, however, this issue is alleviated by the nonlinearities σ, which act as rectifiers that scatter spectral components associated with small eigenvalues to upper parts of the spectrum where they can be discriminated.

5. NUMERICAL EXPERIMENTS

Using the MovieLens 100k dataset [30], we implement a GNN supported on a movie similarity network to predict the ratings that different users would give to the movie "Contact" considering their ratings to other movies in the network and the similarities between these movies. The dataset consists of 100,000 integer ratings between 1 and 5 given by $U = 943$ users to $M = 1682$ movies. To build the movie similarity graph, we first store the data in a $U \times M$ matrix $\mathbf{R}$ such that $[\mathbf{R}]_{um}$ is the rating of user u to movie m if it exists, and zero otherwise; then, we compute the edge weights by calculating pairwise correlations between different movies, i.e., betweeen the columns of $\mathbf{R}$. Denoting the index of the movie "Contact" m_0, the input samples $\mathbf{x}_u$ are generated by zero-ing out the entry $[\mathbf{R}]_{um_0}$ of the user vectors $\mathbf{r}_u = [\mathbf{R}]_{u:}$. The output samples are vectors $\mathbf{y}_u$ where $[\mathbf{y}_u]_m = [\mathbf{R}]_{um_0}$ if $m = m_0$ and zero otherwise.

To analyze the stability of GNNs to graphon perturbations in this setting, we consider the full 1682-movie network to be a SBM graphon $\mathbf{W}$, and perturb it to obtain $\mathbf{W}' = \mathbf{W} + \mathbf{A}$ where $\mathbf{A}(u,v) = (1 - \exp(\mathbf{W}(u,v)^{-1}))/10$. We then instantiate deterministic graphs $\bar{\mathbf{G}}_n$ and $\bar{\mathbf{G}}'_n$ with $n = 118, 203, 338, 603, 1152$, and use a 90-10 data split to train a GNN with $L = 2$, $F = 64$ and $K = 5$ on the $\bar{\mathbf{G}}_n$ by optimizing the MSE loss. The trained GNNs are tested on both the $\bar{\mathbf{G}}_n$ and $\bar{\mathbf{G}}'_n$; the relative RMSE difference obtained on the original and perturbed graphs is shown in Fig. 1. We observe the asymptotic behavior predicted by Theorem 2—the larger the graph, the more stable the GNN.

6. CONCLUSIONS

In this paper, we have defined graphon neural networks (WNNs) and characterized their stability to perturbations of the graphon. By interpreting WNNs as generating models for GNNs, we have further analyzed stability of GNNs defined on deterministic and stochastic graphs instantiated from the original and perturbed graphons. These are key contributions to the study of the theoretical properties of GNNs as graphon perturbations, being perturbations of graph models, make for a more general type of graph perturbation than matrix perturbations. We conclude that GNNs are stable to graphon perturbations with a stability bound that decreases asymptotically with the size of the graph. This result has been demonstrated numerically in a movie recommendation experiment.

A. PROOF OF THEOREM 1

We start by proving stability of graphon convolutions to absolute graphon perturbations, which is stated in Theorem 4.

Theorem 4. *Consider the graphon convolution given by $Y = T_\mathbf{H} X$ as defined in (13), where $h(\lambda)$ is constant for $|\lambda| < c$. Let $Y' = T_{\mathbf{H}'} X$ be the perturbed graphon convolution. Under Assumption 1, it holds that:*

$$\|Y - Y'\| \leq \left(A_2 + \frac{\pi n_c}{\delta_c} \right) \epsilon \|X\|, \tag{19}$$

where we have labeled the graphons $\mathbf{W}^{(1)} := \mathbf{W}$ and $\mathbf{W}^{(2)} := \mathbf{W}'$, and defined $n_c := n_c^{(2)}$ and $\delta_c := \delta_c^{(12)}$ [cf. Definition 1].

To prove this theorem, we need the following propositions.

Proposition 1. *Let T and T' be two self-adjoint operators on a separable Hilbert space $\mathcal{H}$ whose spectra are partitioned as $\gamma \cup \Gamma$ and $\omega \cup \Omega$ respectively, with $\gamma \cap \Gamma = \emptyset$ and $\omega \cap \Omega = \emptyset$. If there exists $d > 0$ such that $\min_{x \in \gamma, y \in \Omega} |x - y| \geq d$ and $\min_{x \in \omega, y \in \Gamma} |x - y| \geq d$, then*

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

Proof. See [31]. □

Proposition 2. *Let $\mathbf{W} : [0,1]^2 \to [0,1]$ and $\mathbf{W}' : [0,1]^2 \to [0,1]$ be two graphons with eigenvalues given by $\{\lambda_i(T_\mathbf{W})\}_{i \in \mathbb{Z} \setminus \{0\}}$ and $\{\lambda_i(T_{\mathbf{W}'})\}_{i \in \mathbb{Z} \setminus \{0\}}$, ordered according to their sign and in decreasing order of absolute value. Then, for all $i \in \mathbb{Z} \setminus \{0\}$, the following inequalities hold*

$$|\lambda_i(T_{\mathbf{W}'}) - \lambda_i(T_\mathbf{W})| \leq \|T_{\mathbf{W}' - \mathbf{W}}\| \leq \|\mathbf{W}' - \mathbf{W}\|.$$

Proof. See [16]. □

We first prove Theorem 4 for filters satisfying $h(\lambda) = 0$ for $|\lambda| < c$. Writing the inner product $\int_0^1 X(u)\varphi_i(u)du$ as $\hat{X}(\lambda_i)$, the filter output norm difference can be expressed as

$$\|T_\mathbf{H} X - T_{\mathbf{H}'} X\| = \left\| \sum_i h(\lambda_i)\hat{X}(\lambda_i)\varphi_i - \sum_i h(\lambda_i')\hat{X}(\lambda_i')\varphi_i' \right\|$$
$$= \left\| \sum_i h(\lambda_i)\hat{X}(\lambda_i)\varphi_i - h(\lambda_i')\hat{X}(\lambda_i')\varphi_i' \right\|.$$

Adding and subtracting $h(\lambda_i')\hat{X}(\lambda_i)\varphi_i$ and applying the triangle inequality, we get

$$\|T_\mathbf{H} X - T_{\mathbf{H}'} X\| = \left\| \sum_i h(\lambda_i)\hat{X}(\lambda_i)\varphi_i - h(\lambda_i')\hat{X}(\lambda_i')\varphi_i' \right\|$$
$$= \left\| \sum_i h(\lambda_i)\hat{X}(\lambda_i)\varphi_i - h(\lambda_i')\hat{X}(\lambda_i)\varphi_i \right.$$
$$\left. + h(\lambda_i')\hat{X}(\lambda_i)\varphi_i - h(\lambda_i')\hat{X}(\lambda_i')\varphi_i' \right\|$$
$$\leq \left\| \sum_i \left(h(\lambda_i) - h(\lambda_i') \right) \hat{X}(\lambda_i)\varphi_i \right\| \quad \textbf{(1.1)}$$
$$+ \left\| \sum_i h(\lambda_i') \left(\hat{X}(\lambda_i)\varphi_i - \hat{X}(\lambda_i')\varphi_i' \right) \right\| \quad \textbf{(1.2)}$$

which we have split between **(1.1)** and **(1.2)**.

We start by bounding **1.1**. To do so, we leverage the Lipschitz continuous assumption, which can be expressed as $|h(\lambda_i) - h(\lambda_i')| \leq A_2 \|\lambda_i - \lambda_i'\|$. Using also Prop. 2 and the Cauchy-Schwarz inequality, we have

$$\left\| \sum_i \left(h(\lambda_i) - h(\lambda_i') \right) \hat{X}(\lambda_i)\varphi_i \right\| \leq A_2 \|\mathbf{W} - \mathbf{W}'\| \left\| \sum_i \hat{X}(\lambda_i)\varphi_i \right\|$$

and, since $\|\mathbf{W} - \mathbf{W}'\| \leq \|\mathbf{A}\| = \epsilon$,

$$\left\| \sum_i \left(h(\lambda_i) - h(\lambda_i') \right) \hat{X}(\lambda_i)\varphi_i \right\| \leq A_2 \epsilon \|X\| \tag{20}$$

For (**1.2**), we use the triangle and Cauchy-Schwarz inequalities to obtain

$$\left\| \sum_i h(\lambda_i') \left(\hat{X}(\lambda_i)\varphi_i - \hat{X}(\lambda_i')\varphi_i' \right) \right\|$$
$$= \left\| \sum_i h(\lambda_i') \left(\hat{X}(\lambda_i)\varphi_i + \hat{X}(\lambda_i)\varphi_i' - \hat{X}(\lambda_i)\varphi_i' - \hat{X}(\lambda_i')\varphi_i' \right) \right\|$$
$$\leq \left\| \sum_i h(\lambda_i')\hat{X}(\lambda_i)(\varphi_i - \varphi_i') \right\| + \left\| \sum_i h(\lambda_i')\varphi_i'\langle X, \varphi_i - \varphi_i' \rangle \right\|$$
$$\leq 2 \sum_i \|h(\lambda_i')\|\|X\|\|\varphi_i - \varphi_i'\|.$$

The norm difference of the eigenfunctions is bounded by Prop. 1. Thus, we can write

$$\left\| \sum_i h(\lambda_i') \left(\hat{X}(\lambda_i)\varphi_i - \hat{X}(\lambda_i')\varphi_i' \right) \right\| \leq \|X\| \sum_i \|h(\lambda_i')\| \frac{\pi\|T_\mathbf{W} - T_{\mathbf{W}'}\|}{d_i}$$

where d_i is the minimum between $\min(|\lambda_i - \lambda_{i+1}'|, |\lambda_i - \lambda_{i-1}'|)$ and $\min(|\lambda_i' - \lambda_{i+1}|, |\lambda_i' - \lambda_{i-1}|)$ for each i. By Definition 1, we have $\delta_c \leq d_i$ for all i and $\|T_\mathbf{W} - T_{\mathbf{W}'}\| \leq \|\mathbf{W} - \mathbf{W}'\|$ (i.e., the Hilbert-Schmidt norm dominates the operator norm). Therefore, we obtain

$$\left\| \sum_i h(\lambda_i') \left(\hat{X}(\lambda_i)\varphi_i - \hat{X}(\lambda_i')\varphi_i' \right) \right\| \leq \frac{\pi\|\mathbf{W} - \mathbf{W}'\|}{\delta_c} \|X\| \sum_i \|h(\lambda_i')\|.$$

Note that, since $|h(\lambda)| < 1$ and $h(\lambda) = 0$ for $|\lambda| < c$, the sum on the right hand side is finite. Denoting the number of eigenvalues $|\lambda'| \geq c$ as n_c, this inequality can be rewritten as

$$\left\| \sum_i h(\lambda_i') \left(\hat{X}(\lambda_i)\varphi_i - \hat{X}(\lambda_i')\varphi_i' \right) \right\| \leq \frac{\pi\epsilon}{\delta_c} \|X\| n_c. \tag{21}$$

The result stated in Theorem 4 follows from (20) and (21).

In Theorem 4, the filter $h(\lambda)$ is constant for $|\lambda| < c$, but not necessarily zero. This type of filter can be constructed by adding a low-pass filter with band $|\lambda| \geq c$ and a high-pass filter $g(\lambda)$ which is constant for $|\lambda| < c$. The high-pass filter is stable because it can be seen as the sum of a constant filter with gain G and a low-pass filter with negative gain $-G$.

We now extend the stability result to the WNN. To compute a bound for $\|Y - Y'\|$, we start by writing it in terms of the features of the final layer, i.e.

$$\|Y - Y'\|^2 = \sum_{f=1}^{F_L} \left\| X_L^f - X_L'^f \right\|^2. \tag{22}$$

At layer ℓ of the WNN $\mathbf{\Phi}(\mathcal{H}; \mathbf{W}; X)$, we have

$$X_\ell^f = \sigma \left(\sum_{g=1}^{F_{\ell-1}} \mathbf{h}_\ell^{fg} *_\mathbf{W} X_{\ell-1}^g \right) = \sigma \left(\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g \right)$$

and, similarly for $\Phi(\mathcal{H}; \mathbf{W}'; X)$,

$$X_\ell^f = \sigma\left(\sum_{g=1}^{F_{\ell-1}} \mathbf{h}_\ell^{'fg} *_{\mathbf{W}} X_{\ell-1}^{'g}\right) = \sigma\left(\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_\ell^{'fg}} X_{\ell-1}^{'g}\right).$$

We can therefore write $\|X_\ell^f - X_\ell^{'f}\|$ as

$$\left\|X_\ell^f - X_\ell^{'f}\right\| = \left\|\sigma\left(\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g\right) - \sigma\left(\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_\ell^{'fg}} X_{\ell-1}^{'g}\right)\right\|$$

and, since σ is normalized Lipschitz,

$$\left\|X_\ell^f - X_\ell^{'f}\right\| \leq \left\|\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g - T_{\mathbf{H}_\ell^{'fg}} X_{\ell-1}^{'g}\right\|$$

$$\leq \sum_{g=1}^{F_{\ell-1}} \left\|T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g - T_{\mathbf{H}_\ell^{'fg}} X_{\ell-1}^{'g}\right\|.$$

where the second inequality follows from the triangle inequality. Looking at each feature g independently, we apply the triangle inequality once again to get

$$\left\|T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g - T_{\mathbf{H}_\ell^{'fg}} X_{\ell-1}^{'g}\right\|$$

$$\leq \left\|T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g - T_{\mathbf{H}_\ell^{'fg}} X_{\ell-1}^g\right\|_{L_2} + \left\|T_{\mathbf{H}_\ell^{'fg}}\left(X_{\ell-1}^g - X_{\ell-1}^{'g}\right)\right\|.$$

The first term on the right hand side of this inequality is bounded by (19) in Theorem 4. The second term can be decomposed by using Cauchy-Schwarz and recalling that $|h(\lambda)| < 1$ for all graphon convolutions in the WNN. We thus obtain a recursion for $\|X_\ell^f - X_\ell^{'f}\|$, which is given by

$$\left\|X_\ell^f - X_\ell^{'f}\right\| \leq \sum_{g=1}^{F_{\ell-1}} \left(A_2 + \frac{\pi n_c}{\delta_c}\right)\epsilon\|X_{\ell-1}^g\| + \sum_{g=1}^{F_{\ell-1}} \left\|X_{\ell-1}^g - X_{\ell-1}^{'g}\right\| \tag{23}$$

and whose first term, $\sum_{g=1}^{F_0}\|X_0^g - X_0^{'g}\|$, is equal to 0. To solve this recursion, we need to compute the norm $\|X_{\ell-1}^g\|$. Since the nonlinearity σ is normalized Lipschitz and $\sigma(0) = 0$ by Assumption 1, this bound can be written as

$$\|X_{\ell-1}^g\| \leq \left\|\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g\right\|$$

and using the triangle and Cauchy Schwarz inequalities,

$$\|X_{\ell-1}^g\| \leq \sum_{g=1}^{F_{\ell-1}} \left\|T_{\mathbf{H}_\ell^{fg}}\right\| \|X_{\ell-1}^g\| \leq \sum_{g=1}^{F_{\ell-1}} \|X_{\ell-1}^g\|$$

where the second inequality follows from $|h(\lambda)| < 1$. Expanding this expression with initial condition $X_0^g = X^g$ yields

$$\|X_{\ell-1}^g\| \leq \prod_{\ell'=1}^{\ell-1} F_{\ell'} \sum_{g=1}^{F_0} \|X^g\|. \tag{24}$$

and substituting it back in (23) to solve the recursion, we get

$$\left\|X_\ell^f - X_\ell^{'f}\right\| \leq L\left(A_2 + \frac{\pi n_c}{\delta_c}\right)\epsilon\left(\prod_{\ell'=1}^{\ell-1} F_{\ell'}\right)\sum_{g=1}^{F_0}\|X^g\|. \tag{25}$$

To arrive at the result of Theorem 1, we evaluate (25) with $\ell = L$ and substitute it into (22) to obtain

$$\|Y - Y'\|^2 = \sum_{f=1}^{F_L} \left\|X_L^f - X_L^{'f}\right\|^2$$

$$\leq \sum_{f=1}^{F_L}\left(L\left(A_2 + \frac{\pi n_c}{\delta_c}\right)\epsilon\left(\prod_{\ell=1}^{L-1} F_\ell\right)\sum_{g=1}^{F_0}\|X^g\|\right)^2. \tag{26}$$

Finally, since $F_0 = F_L = 1$ and $F_\ell = F$ for $1 \leq \ell \leq L - 1$,

$$\|Y - Y'\| \leq LF^{L-1}\left(A_2 + \frac{\pi n_c}{\delta_c}\right)\epsilon\|X\|. \tag{27}$$

B. PROOF OF LEMMA 1

In order to prove Lemma 1, we need the following lemma.

Lemma 2. *Let $\bar{\mathbf{G}}_n$ be an undirected graph with $\bar{\mathbf{S}}_n \in [0,1]^{n\times n}$. Let $\mathbf{G}_n$ be a stochastic graph obtained from $\bar{\mathbf{G}}_n$ as in (6). Consider the graph convolution $\mathbf{H}(\mathbf{S})$ with frequency response h such that $h(\lambda/n)$ is constant for $|\lambda/n| < c$ and let $\bar{\mathbf{y}}_n = \mathbf{H}(\bar{\mathbf{S}}_n)\mathbf{x}_n$ and $\mathbf{y}_n = \mathbf{H}(\mathbf{S}_n)\mathbf{x}_n$ be the outputs of this filter on the graphs $\bar{\mathbf{G}}_n$ and $\mathbf{G}_n$ respectively. Then, under Assumptions 1,2 and 5, it holds*

$$\|\mathbf{y}_n - \bar{\mathbf{y}}_n\| \leq \left(A_2 + \frac{\pi n_c^{(q)}}{\delta_c^{(pq)}}\right)\left(2\sqrt{\log 2n/\xi}\sqrt{n}\right)\|\mathbf{x}_n\|$$

where $n_c^{(q)}$ and $\delta_c^{(pq)}$ are as in Definition 1 for $\mathbf{W}^{(p)} = \bar{\mathbf{W}}_n$, the graphon induced by $\bar{\mathbf{G}}_n$, and $\mathbf{W}^{(q)} = \mathbf{W}_n$, the graphon induced by $\mathbf{G}_n$ (7).

Proof. Following the same reasoning used in the proof of Theorem 4, we analyze low-pass filters $h(\lambda/n) = 0$ for $|\lambda/n| < c$. The result of the lemma then follows from the fact that the filters we consider are the sum of low-pass filters with constant filters, and constant filters are always transferable. Also similarly to what we did in the proof of Theorem 4, we use the triangle inequality to bound $\|\bar{\mathbf{y}}_n - \mathbf{y}_n\|$ as

$$\|\mathbf{H}(\bar{\mathbf{S}}_n)\mathbf{x}_n - \mathbf{H}(\mathbf{S}_n)\mathbf{x}_n\| \leq \left\|\sum_i \left(h(\bar{\lambda}_i/n) - h(\lambda_i/n)\right)[\hat{\bar{\mathbf{x}}}_n]_i\bar{\mathbf{v}}_i\right\| \quad (2.1)$$

$$+ \left\|\sum_i h(\lambda_i)\left([\hat{\bar{\mathbf{x}}}_n]_i\bar{\mathbf{v}}_i - [\hat{\mathbf{x}}_n]_i\mathbf{v}_i\right)\right\| \quad (2.2)$$

where $\bar{\lambda}_i$ and $\bar{\mathbf{v}}_i$ are the eigenvalues and eigenvectors of $\bar{\mathbf{S}}_n$, λ_i and $\mathbf{v}_i$ those of $\mathbf{S}_n$, $[\hat{\bar{\mathbf{x}}}_n]_i = \bar{\mathbf{v}}_i^\mathsf{T}\mathbf{x}_n$ and $[\hat{\mathbf{x}}_n]_i = \mathbf{v}_i^\mathsf{T}\mathbf{x}_n$.

Using the Lipschitz property of the filter in the passing band together with Prop. 2 and the Cauchy-Schwarz inequality, we can rewrite **(2.1)** as

$$\|\mathbf{H}(\bar{\mathbf{S}}_n)\mathbf{x}_n - \mathbf{H}(\mathbf{S}_n)\mathbf{x}_n\| \leq \left\|\sum_i \left(h(\bar{\lambda}_i/n) - h(\lambda_i/n)\right)[\hat{\bar{\mathbf{x}}}_n]_i\bar{\mathbf{v}}_i\right\|$$

$$\leq A_2\frac{\|\bar{\mathbf{S}}_n - \mathbf{S}_n\|}{n}\|\mathbf{x}_n\|.$$

The term $\|\bar{\mathbf{S}}_n - \mathbf{S}_n\|$ can be bounded by [32, Theorem 1] with probability $1 - \xi$ provided that $d_{\mathbf{G}_n}$, the maximum expected degree of the graph $\mathbf{G}_n$, satisfies $d_{\mathbf{G}_n} > 4\log(2n/\xi)$. This is the case for graph $\mathbf{G}_n$ for which n satisfies Assumption 5 [13]. Hence, we can write

$$\|\mathbf{H}(\bar{\mathbf{S}}_n)\mathbf{x}_n - \mathbf{H}(\mathbf{S}_n)\mathbf{x}_n\| \leq A_2\frac{2\sqrt{n\log(2n/\xi)}}{n}\|\mathbf{x}_n\|$$

$$\leq A_2\frac{2\sqrt{\log(2n/\xi)}}{\sqrt{n}}\|\mathbf{x}_n\| \tag{28}$$

.

For (2.2), we use the triangle inequality to decompose the norm difference as:

$$\left\| \sum_i h(\lambda_i/n) \left([\hat{\bar{\mathbf{x}}}_n]_i \bar{\mathbf{v}}_i - [\hat{\mathbf{x}}_n]_i \mathbf{v}_i \right) \right\|$$

$$= \left\| \sum_i h(\lambda_i/n) \left([\hat{\bar{\mathbf{x}}}_n]_i \bar{\mathbf{v}}_i + [\hat{\bar{\mathbf{x}}}_n]_i \mathbf{v}_i - [\hat{\bar{\mathbf{x}}}_n]_i \mathbf{v}_i - [\hat{\mathbf{x}}_n]_i \mathbf{v}_i \right) \right\|$$

$$\leq \left\| \sum_i h(\lambda_i/n)[\hat{\bar{\mathbf{x}}}_n]_i (\bar{\mathbf{v}}_i - \mathbf{v}_i) \right\| + \left\| \sum_i h(\lambda_i/n)\mathbf{v}_i \langle \mathbf{x}_n, \bar{\mathbf{v}}_i - \mathbf{v}_i \rangle \right\|$$

$$\leq 2 \sum_i \| h(\lambda_i/n) \| \| X \| \| \bar{\mathbf{v}}_i - \mathbf{v}_i \|$$

where the last inequality follows from the Cauchy-Schwarz inequality.

Prop. 1 gives an upper bound to the norm difference between the eigenvectors, by which we obtain:

$$\left\| \sum_i h\left(\lambda_i/n\right) \left([\hat{\bar{\mathbf{x}}}_n]_i \bar{\mathbf{v}}_i - [\hat{\mathbf{x}}_n]_i \mathbf{v}_i \right) \right\| \leq \|\mathbf{x}_n\| \sum_i \| h\left(\lambda_i/n\right) \|$$
$$\times \frac{\pi \|\bar{\mathbf{S}}_n - \mathbf{S}_n\|}{d_i}$$

where d_i is the minimum between $\min(|\bar{\lambda}_i - \lambda_{i+1}|, |\bar{\lambda}_i - \{\lambda_{i-1}\})$ and $\min(|\lambda_i - \bar{\lambda}_{i+1}|, |\lambda_i - \bar{\lambda}_{i-1}|)$ for each i. Due to the fact that the eigenvalues of the graphon induced by a graph are equal to the eigenvalues of the graph divided by n [14, Lemma 1], and by Definition 1, we have $n\delta_c^{(pq)} \leq d_i$ for all i. Leveraging this together with [32, Theorem 1], it thus holds with probability $1 - \xi$ that

$$\left\| \sum_i h\left(\lambda_i/n\right) \left([\hat{\bar{\mathbf{x}}}_n]_i \bar{\mathbf{v}}_i - [\hat{\mathbf{x}}_n]_i \mathbf{v}_i \right) \right\| \leq \|\mathbf{x}_n\| n_c^{(q)} \frac{2\pi \sqrt{n \log(2n/\xi)}}{n\delta_c^{(pq)}}$$
$$\leq \|\mathbf{x}_n\| n_c^{(q)} \frac{2\pi \sqrt{\log(2n/\xi)}}{\delta_c^{(pq)}\sqrt{n}} \quad (29)$$

where $n_c^{(q)}$ is the number of eigenvalues of $\mathbf{S}_n$ in the passing band of h. Putting (28) and (29) together completes the proof. $\square$

C. REFERENCES

[1] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.

[2] 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.

[3] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *5th Int. Conf. Learning Representations*, Toulon, France, 24-26 Apr. 2017, pp. 1–14, Assoc. Comput. Linguistics.

[4] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *30th Conf. Neural Inform. Process. Syst.*, Barcelona, Spain, 5-10 Dec. 2016, pp. 3844–3858, Neural Inform. Process. Foundation.

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

[6] D. Moore, J. Leonard, D. Rus, and S. Teller, "Robust distributed network localization with noisy range measurements," in *Proceedings of the 2nd international conference on Embedded networked sensor systems*, 2004, pp. 50–61.

[7] B. Nosrat-Makouei, J. G. Andrews, and R. W. Heath, "MIMO interference alignment over correlated channels with imperfect CSI," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2783–2794, 2011.

[8] E. Adar and C. Re, "Managing uncertainty in social networks.," *IEEE Data Eng. Bull.*, vol. 30, no. 2, pp. 15–22, 2007.

[9] B. Y. Chen, W. H. K. Lam, A. Sumalee, Q. Li, and Z. Li, "Vulnerability analysis for large-scale and congested road networks with demand uncertainty," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 3, pp. 501–516, 2012.

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

[11] 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," *Advances in Mathematics*, vol. 219, no. 6, pp. 1801–1851, 2008.

[12] S. Gao and P. E. Caines, "Graphon control of large-scale networks of linear systems," *To appear: IEEE Transactions on Automatic Control*, 2020.

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

[14] L. Ruiz, L. F. O. Chamon, and A. Ribeiro, "Graphon signal processing," *arXiv:2003.05030 [eess.SP]*, 2020.

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

[16] L. Ruiz, L. F. O. Chamon, and A. Ribeiro, "Graphon neural networks and the transferability of graph neural networks," *arXiv:2006.03548 [cs.LG]*, 2020.

[17] F. Gama, J. Bruna, and A. Ribeiro, "Stability properties of graph neural networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 5680–5695, 25 Sep. 2020.

[18] F. Gama, J. Bruna, and A. Ribeiro, "Stability of graph scattering transforms," in *33rd Conf. Neural Inform. Process. Syst.*, Vancouver, BC, 8-14 Dec. 2019, pp. 8038–8048, Neural Inform. Process. Syst. Foundation.

[19] D. Zou and G. Lerman, "Graph convolutional neural networks via scattering," *Appl. Comput. Harmonic Anal.*, 13 June 2019, accepted for publication (in press).

[20] R. Levie, M. M. Bronstein, and G. Kutyniok, "Transferability of spectral graph convolutional neural networks," *arXiv:1907.12972 [cs.LG]*, 2019.

[21] N. Keriven, A. Bietti, and S. Vaiter, "Convergence and stability of graph convolutional networks on large random graphs," *arXiv:2006.01868 [cs.LG]*, 2020.

[22] 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.

[23] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[24] J. Du, J. Shi, S. Kar, and J. M. F. Moura, "On graph convolution for graph CNNs," in *2018 IEEE Data Sci. Workshop*, Lausanne, Switzerland, 4-6 June 2018, pp. 239–243, IEEE.

[25] 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.

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

[27] L. Ruiz, L. F. O. Chamon, and A. Ribeiro, "The Graphon Fourier Transform," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5660–5664.

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

[29] L. Ruiz, Z. Wang, and A. Ribeiro, "Graphon and graph neural network stability (extended)," *Available at https://www.seas.upenn.edu/ rubruiz/graphons-icassp21*, 2020.

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

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

[32] F. Chung and M. Radcliffe, "On the spectra of general random graphs," *The Electronic Journal of Combinatorics*, 2011.