

On Local Distributions in Graph Signal Processing

T. Mitchell Roddenberry , *Graduate Student Member, IEEE*, Fernando Gama , *Member, IEEE*,
Richard G. Baraniuk , *Fellow, IEEE*, and Santiago Segarra , *Senior Member, IEEE*

Abstract—Graph filtering is the cornerstone operation in graph signal processing (GSP). Thus, understanding it is key in developing potent GSP methods. Graph filters are local and distributed linear operations, whose output depends only on the local neighborhood of each node. Moreover, a graph filter’s output can be computed separately at each node by carrying out repeated exchanges with immediate neighbors. Graph filters can be compactly written as polynomials of a graph shift operator (typically, a sparse matrix description of the graph). This has led to relating the properties of the filters with the spectral properties of the corresponding matrix – which encodes global structure of the graph. In this work, we propose a framework that relies solely on the local distribution of the neighborhoods of a graph. The crux of this approach is to describe graphs and graph signals in terms of a measurable space of rooted balls. Leveraging this, we are able to seamlessly compare graphs of different sizes and coming from different models, yielding results on the convergence of spectral densities, transferability of filters across arbitrary graphs, and continuity of graph signal properties with respect to the distribution of local substructures.

Index Terms—Graph signal processing, graph filtering, transferability, graph Fourier transform, graphing signal processing, weak convergence.

I. INTRODUCTION

GRAPH filters are a fundamental building block in graph signal processing, where cascaded applications of a graph shift operator model diffusion on the nodes of a graph [1], [2]. The analogy between filtering in discrete time and filtering on graphs has led to a fruitful research direction, with applications including robotics [3], neuroscience [4], and recommender systems [5]. Due to their typical implementation as low-degree matrix polynomials, graph filters are *local operators*, where the output of a graph filter at a given node is strictly dependent on the connectivity structure and signal values on the node’s local neighborhood. This highlights an invariance property of graph filters, typically summarized by the property of *permutation equivariance*. However, the equivariance of graph filters is much stronger than not being sensitive to permutations of nodes. If the same filter is applied to two different graphs, and two nodes within each of those graphs have identical neighborhoods,

then the graph filter output at those nodes will be identical as well [6]. Indeed, graph filters in their usual implementation are equivariant to *local substructures*, which have been shown to be of primary importance in real-world networks [7], often leading to useful properties such as scale invariance and robustness [8].

The importance of local substructures is an incipient development in the literature. This view has been considered before in the graph wavelet literature [9], [10], [11], where wavelet atoms are constructed by applying graph filters to impulse functions on each node in the graph, yielding a dictionary of atoms with known spectral properties that also exhibit (approximate) spatial localization. As a representational tool for graphs predating the development of graph signal processing, graph kernel methods have put forth the idea of graphs as bags of motifs [12], such as in the paper on *graphlets* [13]. Works such as [14], [15] have considered the extension of graph kernels for graphs with continuous labels, typically via a neighborhood aggregation or discretization approach. In [16], equivariance to local structures is proposed as a more useful invariant for graph neural networks, as opposed to permutation equivariance. For instance, [17] considers neural networks for graph classification that act on small subgraphs over an entire graph, and [18] considers how to design graph neural networks that can recognize structures more expressive than those of the Weisfeiler-Lehman test. The treatment of graphs as distributions of ego-networks in [19] was used to devise a novel loss function for training of graph neural networks based on the maximization of mutual information. Additionally, local structures at each node need not be deterministic, as [20] considers random walk features and defines the notion of an estimable parameter under such a model.

In this work, we develop machinery for reasoning about basic notions in graph signal processing, with the primacy of local substructures in mind. After reviewing basic definitions for graph signal processing in Section II, we make the following theoretical contributions:

- 1) We introduce rooted graphs and rooted graph filters as vehicles for describing localized graph filters. In doing so, we develop a measure-theoretic view of graph signal processing that considers distributions of rooted balls and signals supported on them (Section III).
- 2) Within the proposed framework, we illustrate how convergence of Fourier spectra of graph signals can be understood in terms of weak convergence of measures (Section IV).
- 3) We apply the proposed framework to yield a principled understanding of the transferability of graph summaries via integral probability metrics (Section V).

Manuscript received 3 March 2022; revised 30 September 2022; accepted 8 November 2022. Date of publication 18 November 2022; date of current version 30 November 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. David I Shuman. The work of T. Mitchell Roddenberry and Santiago Segarra was supported by NSF under Grant CCF-2008555. (*Corresponding author: T. Mitchell Roddenberry.*)

The authors are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (e-mail: mitch@rice.edu; fgama@ieee.org; richb@rice.edu; segarra@rice.edu).

Digital Object Identifier 10.1109/TSP.2022.3223217

- 4) To highlight the flexibility of the proposed approach, we extend the relationship between distributions of local graph structures and the Fourier spectra of graph signals to weighted graphs (Section VI).
- 5) We identify graphings and signals supported on them as the appropriate limiting objects in the proposed framework, and develop basic notions of graphing signal processing. We prove that the proposed framework applies directly to graphing signal processing in a natural way (Section VII), yielding a suitable spectral theory.

II. GRAPH SIGNAL PROCESSING

Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where \mathcal{V} is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. Since the graph is undirected, it holds that if an edge $(u, v) \in \mathcal{E}$ for $u, v \in \mathcal{V}$, then $(v, u) \in \mathcal{E}$. We further assume the graph to have no self-loops, i.e., $(v, v) \notin \mathcal{E}$ for all $v \in \mathcal{V}$, and to be unweighted. An extension to weighted graphs can be found in Section VI.

The neighborhood $\mathcal{N}(\mathcal{V}')$ of a given collection of nodes $\mathcal{V}' \subseteq \mathcal{V}$ is defined as

$$\mathcal{N}(\mathcal{V}') = \{u \in \mathcal{V} : (u, v) \in \mathcal{E} \text{ for some } v \in \mathcal{V}'\} \cup \mathcal{V}'. \quad (1)$$

That is, the neighborhood of a collection of nodes is that set of nodes \mathcal{V}' as well as those nodes immediately connected to it. The k -hop neighborhood $\mathcal{N}^k(\mathcal{V}')$ can then be conveniently defined in a recursive manner as $\mathcal{N}^k(\mathcal{V}') = \mathcal{N}(\mathcal{N}^{k-1}(\mathcal{V}'))$ for integers $k \geq 1$ and with $\mathcal{N}^0(\mathcal{V}') = \mathcal{V}'$. Note that the k -hop neighborhood of a singleton $\mathcal{V}' = \{v\}$ is denoted as $\mathcal{N}^k(v)$, and that its degree can be easily computed as $\deg(v) = |\mathcal{N}(v)| - 1$.

Graph signals can be associated with a graph structure and are defined as a map $\mathbf{x} : \mathcal{V} \rightarrow \mathbb{R}$ between the node set \mathcal{V} and the real numbers \mathbb{R} . That is, a graph signal simply attaches a single real number $[\mathbf{x}]_v \in \mathbb{R}$ to each node of the graph $v \in \mathcal{V}$. For a given graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we can define the space of graph signals as $\mathbb{X}(\mathcal{G}) = \{\mathbf{x} : \mathcal{V} \rightarrow \mathbb{R}\}$.

A graph shift operator (GSO) \mathbf{S} can also be associated with the graph structure \mathcal{G} as a means of relating graph signals explicitly with the underlying graph support. More precisely, the GSO is defined as a linear operator between graph signals $\mathbf{S} : \mathbb{X}(\mathcal{G}) \rightarrow \mathbb{X}(\mathcal{G})$ such that the output graph signal $\mathbf{y} = \mathbf{S}\mathbf{x}$ is computed as

$$[\mathbf{y}]_v = \sum_{u \in \mathcal{N}(v)} [\mathbf{S}]_{vu} [\mathbf{x}]_u. \quad (2)$$

The operation (2) shifts the signal around the graph and is analogous to the time-shift in discrete-time signal processing. Note that the GSO can be completely characterized by specifying the adjacency rule that assigns the values $[\mathbf{S}]_{vu}$ for every $(u, v) \in \mathcal{E}$. Examples of GSOs that have found widespread use in the literature include the adjacency matrix, the Laplacian matrix, the Markov transition matrix, and their normalized counterparts. While it is technically possible to design adjacency rules that lead to arbitrary values of $[\mathbf{S}]_{vu}$, we only consider rules that are determined exclusively by the local, combinatorial structure of the graph.

Remark 1 (More general signals): As indicated by the recent popularity of graph neural networks, it is of interest to consider how to understand multidimensional or categorical data on graphs. This fits into our framework in a seamless fashion by expanding the definition of “graph signal” to include any function from the nodes of a graph to a metric space. For simplicity of exposition, we restrict our view to real-valued functions, keeping in mind that such extensions require very little extra work.

A graph filter can then be built from the GSO. More specifically, given a collection of $(K + 1)$ scalar coefficients $\{h_k\}_{k=0}^K$, a K -tap graph filter $\mathcal{H}(\mathbf{S}) : \mathbb{X}(\mathcal{G}) \rightarrow \mathbb{X}(\mathcal{G})$ is defined as the linear mapping between two graph signals $\mathbf{y} = \mathcal{H}(\mathbf{S})\mathbf{x}$ given by

$$\mathcal{H}(\mathbf{S}) = \sum_{k=0}^K h_k \mathbf{S}^k \quad (3)$$

where \mathbf{S}^k denotes k repeated applications of the GSO (see (2)) to the input signal \mathbf{x} . The operator \mathbf{S}^0 is understood to be the identity map on $\mathbb{X}(\mathcal{G})$.

Graph filters are linear and local operators. They are linear in the input graph signal \mathbf{x} . They are local in that they only require information up to the K -hop neighborhood. More specifically, the output of the graph filter at each node $[\mathbf{y}]_v$ can be computed by K repeated exchanges of information with one-hop neighbors $\mathcal{N}(v)$, thus collecting the signal values contained in up to the K -hop neighborhood $\mathcal{N}^K(v)$. Each node can compute their output separately from the rest. The key observation is that the nodes need not know the global structure of the whole graph \mathcal{G} but only their local neighborhood. Thus, graph filters can be analyzed and understood entirely from looking at this local neighborhood structure.

III. A LOCAL FRAMEWORK FOR GSP

The local nature of graph filters calls for a local framework to analyze their effects. Towards this end, define a rooted K -ball $\bar{\mathcal{G}}_K(r) = (\bar{\mathcal{V}}_K, \bar{\mathcal{E}}_K, r)$ as a graph with a root $r \in \bar{\mathcal{V}}_K$ such that all nodes are at most K -hops away from the root, i.e., $\bar{\mathcal{V}}_K = \mathcal{N}^K(r)$. Note that, since the rooted K -ball $\bar{\mathcal{G}}_K(r)$ is a graph, the notions of graph shift operator and graph signal are immediately well-defined. We denote these as $\bar{\mathbf{S}}_K(r)$ and $\bar{\mathbf{x}}_K(r)$, respectively. For ease of exposition, the specification of the root node will be dropped, unless necessary to avoid confusion.

Of particular interest is the case of rooted K -balls obtained as the induced subgraph of the K -hop neighborhood of a node. More specifically, given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a rooted K -ball $\bar{\mathcal{G}}_K$ can be constructed by selecting a node $r \in \mathcal{V}$, setting $\bar{\mathcal{V}}_K \equiv \mathcal{N}^K(r)$ and $\bar{\mathcal{E}}_K \equiv \mathcal{E} \cap \mathcal{N}^K(r) \times \mathcal{N}^K(r)$. In this context, the graph signal $\bar{\mathbf{x}}_K$ corresponding to the rooted K -ball can be tied to the graph signal \mathbf{x} defined on \mathcal{G} by setting $[\bar{\mathbf{x}}_K]_v = [\mathbf{x}]_v$ for all $v \in \mathcal{N}^K(r)$.

We can now formalize the notion that a graph filter only relies on local information.

Proposition 1: Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with graph signal \mathbf{x} and a GSO \mathbf{S} . For a node $r \in \mathcal{V}$, denote the corresponding rooted

K -ball as $\bar{\mathcal{G}}_K$, with graph signal $\bar{\mathbf{x}}_K$ and GSO $\bar{\mathbf{S}}_K$, respectively. Then, for any K -tap filter \mathcal{H} , the following equality holds

$$[\mathcal{H}(\mathbf{S})\mathbf{x}]_r = [\mathcal{H}(\bar{\mathbf{S}}_K) \bar{\mathbf{x}}_K]_r. \quad (4)$$

We note that for (4) to hold, the adjacency rule used to construct \mathbf{S} and $\bar{\mathbf{S}}_K$ has to be the same, and it is assumed to depend only on the combinatorial structure of the graphs; see Section II.

Proposition 1 formalizes the well-known fact that a K -tap graph filter evaluated at a node only depends on the subgraph of K -hops centered at that node (i.e., the ego-network of the node [21]). In this sense, it is not necessary to understand how a graph filter behaves on the global structure of a graph. Rather, one only needs to understand the behavior of a graph filter locally, in particular on the rooted K -balls of the graph. This naturally leads to a weaker form of the property of permutation equivariance, discussed next.

For two graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$, we say that a map $\phi: \mathcal{V} \rightarrow \mathcal{V}'$ is a K -morphism if it preserves the structure of rooted K -balls for all nodes in \mathcal{V} . The structure of two rooted K -balls $\bar{\mathcal{G}}_K = (\bar{\mathcal{V}}_K, \bar{\mathcal{E}}_K, r)$ and $\bar{\mathcal{G}}'_K = (\bar{\mathcal{V}}'_K, \bar{\mathcal{E}}'_K, r')$ with signals $\bar{\mathbf{x}}_K$ and $\bar{\mathbf{x}}'_K$, respectively, is preserved, if there exists an isomorphism $\psi: \mathcal{V}_K \rightarrow \mathcal{V}'_K$ such that $\psi(r) = r'$, $(\psi(u), \psi(v)) \in \mathcal{E}'_K$ if and only if $(u, v) \in \mathcal{E}_K$, and $[\bar{\mathbf{x}}_K]_v = [\bar{\mathbf{x}}'_K]_{\psi(v)}$ for all $v \in \bar{\mathcal{V}}_K$. Then, a K -morphism $\phi: \mathcal{V} \rightarrow \mathcal{V}'$ between two graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ with associated signals \mathbf{x} and \mathbf{x}' , respectively, satisfies that the rooted K -balls $(\bar{\mathcal{G}}_K(v), \bar{\mathbf{x}}_K(v))$ and $(\bar{\mathcal{G}}'_K(\phi(v)), \bar{\mathbf{x}}'_K(\phi(v)))$ are isomorphic for all nodes $v \in \mathcal{V}$. Note that K -morphisms are not necessarily invertible, nor are they necessarily surjective.

Graph filters are invariant under K -morphisms, as it follows from Proposition 1, where we see that the same filter applied to two different graphs yields the same result if the local structures are the same.

Corollary 1: Consider two graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ with corresponding signals \mathbf{x} and \mathbf{x}' . Denote by $\bar{\mathcal{G}}_K(v)$ the K -rooted ball for root $v \in \mathcal{V}$, and by $\bar{\mathbf{S}}_K(v)$ and $\bar{\mathbf{x}}_K(v)$ the corresponding GSO and graph signal, respectively. If there exists a K -morphism $\phi: \mathcal{V} \rightarrow \mathcal{V}'$, then for any K -hop filter \mathcal{H} it holds that

$$[\mathcal{H}(\bar{\mathbf{S}}_K(v)) \bar{\mathbf{x}}_K(v)]_v = [\mathcal{H}(\bar{\mathbf{S}}_K(\phi(v))) \bar{\mathbf{x}}_K(\phi(v))]_{\phi(v)} \quad (5)$$

for all $v \in \mathcal{V}$.

Note that Corollary 1 is a generalization of the permutation equivariance property that graph filters have [6]. More generally, Corollary 1 opens up ways to compare the performance of a fixed graph filter across two different graphs with different associated graph signals. It would be expected then, that if two graphs have similar distributions of rooted K -balls with associated signal, then a fixed graph filter acting on one should yield similar results on the other. We formalize these notions in the ensuing discussion.

A. Distributions of Rooted Balls

Proposition 1 states that the output of a K -tap graph filter at each node depends only on the rooted K -ball at that node and the corresponding graph signal values. Therefore, to characterize the

effect of graph filtering, it suffices to characterize the *distribution* of rooted K -balls of a graph. Towards this end, we will construct a sample space, σ -field, and a corresponding measure to describe this distribution.

Denote by $\bar{\mathcal{G}}_k$ a rooted k -ball and by $\mathbb{X}(\bar{\mathcal{G}}_k)$ its corresponding signal space. Consider now the space constructed by the disjoint union of all signal spaces on all possible rooted k -balls (up to isomorphism) with $k \leq K$, given by¹

$$\Omega_K = \coprod_{\bar{\mathcal{G}}_k: k \leq K} \mathbb{X}(\bar{\mathcal{G}}_k). \quad (6)$$

The construction of the space Ω_K is illustrated in Fig. 1. Essentially, one considers a different space $\mathbb{X}(\bar{\mathcal{G}})$ (typically, $|\bar{\mathcal{V}}|$ -dimensional vectors $\bar{\mathbf{x}}$) for each possible rooted K -ball $\bar{\mathcal{G}}$, and then puts them all into a common space where each original space is isolated. An element $\omega \in \Omega_K$ is described by both the rooted K -ball and the graph signal space it indexes, i.e., $\omega = (\bar{\mathcal{G}}, \mathbb{X}(\bar{\mathcal{G}}))$. Making the identification $\mathbb{X}(\bar{\mathcal{G}}) \equiv \mathbb{R}^{|\bar{\mathcal{V}}|}$, the point ω can be thought as a $|\bar{\mathcal{V}}|$ -dimensional vector associated with the rooted K -ball $\bar{\mathcal{G}}$. Note that, while it may happen that $\mathbb{X}(\bar{\mathcal{G}}) \cong \mathbb{X}(\bar{\mathcal{G}}')$ because $|\bar{\mathcal{V}}| = |\bar{\mathcal{V}}'|$, these are different component spaces in the disjoint union Ω_K as they are indexed by different rooted balls $\bar{\mathcal{G}}$ and $\bar{\mathcal{G}}'$, respectively.

To define the corresponding σ -field $\sigma(\Omega_K)$ associated with Ω_K , note that the disjoint union of graph signal spaces (i.e., Euclidean spaces) has a natural topology defined over it, namely the disjoint union topology. Thus, the Borel σ -field $\sigma(\Omega_K)$ generated from the topology of Ω_K is a proper σ -field. Intuitively, measurable sets in Ω_K are constructed by selecting some rooted K -ball $\bar{\mathcal{G}}$ and a Borel set in the space of graph signals $\mathbb{X}(\bar{\mathcal{G}})$, then considering the corresponding set in Ω_K .

A measure μ can be properly defined now that the σ -field associated with Ω_K has been constructed. Let $\Sigma_K: \mathcal{V} \rightarrow \Omega_K$ be a map that takes a node $v \in \mathcal{V}$ and returns the corresponding rooted K -ball and signal, i.e.,

$$\Sigma_K(v) = (\bar{\mathcal{G}}_K(v), \bar{\mathbf{x}}_K(v)). \quad (7)$$

Note that $\Sigma_K(v) \in \Omega_K$. Then, for a Borel set $\mathcal{A} \in \sigma(\Omega_K)$, the measure $\mu: \sigma(\Omega_K) \rightarrow [0, 1]$ is defined as

$$\mu(\mathcal{A}) = \frac{1}{|\mathcal{V}|} |\{v \in \mathcal{V} : \Sigma_K(v) \in \mathcal{A}\}|. \quad (8)$$

Recall that each element in Ω_K can be thought of as a rooted K -ball with an associated signal. Formally, μ is the pushforward of the uniform measure on the nodes of \mathcal{G} by the sampling map Σ_K , which can be denoted as $\mu = (\Sigma_K)_*(\mathcal{G}, \mathbf{x})$ [22]. We observe that any measure on the nodes of the graph can be used to replace the RHS of (8).

The space Ω_K serves as a natural domain in which to characterize K -tap graph filters. Indeed, Proposition 1 indicates that the behavior of a K -tap graph filter can be fully characterized by its behavior over Ω_K . That is to say, any given K -tap graph filter \mathcal{H} has a corresponding map $\bar{\mathcal{H}}: \Omega_K \rightarrow \mathbb{R}$ such that for any

¹In this context, each signal space $\mathbb{X}(\bar{\mathcal{G}}_k)$ is considered modulo the action of the automorphism group of $\bar{\mathcal{G}}_k$. This is clarified in Appendix A.

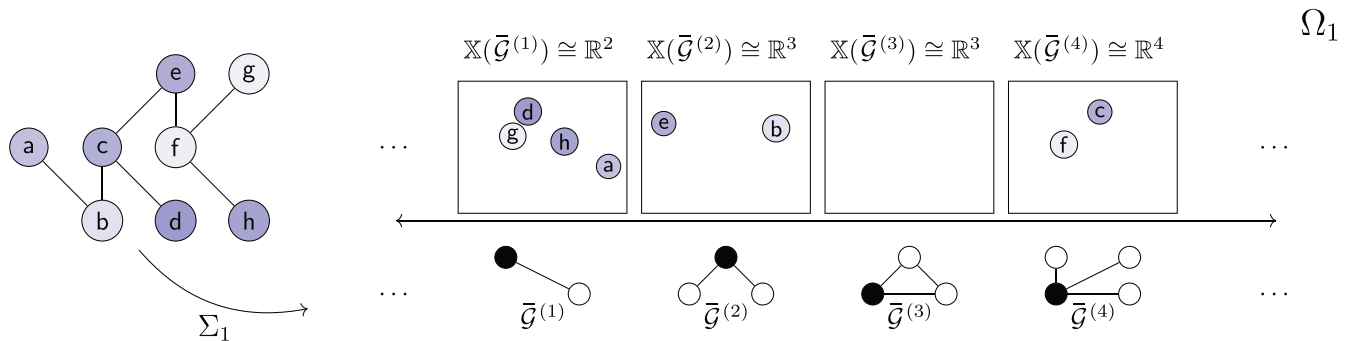


Fig. 1. The space Ω_K of rooted K -balls and signals, constructed via the disjoint union of signal spaces enumerated by rooted K -balls. **(Left)** A graph on nodes $\{a, b, \dots, h\}$, with graph signal denoted by the node coloring. **(Right)** The space Ω_1 , and the distribution μ of points obtained via the pushforward of Σ_1 . Observe that nodes $\{a, d, g, h\}$ all have isomorphic rooted 1-balls, and thus all correspond to points in $\mathbb{X}(\bar{\mathcal{G}}^{(1)})$ (depicted here in random positions within Euclidean space). Similarly, nodes $\{b, e\}$ are mapped to points in $\mathbb{X}(\bar{\mathcal{G}}^{(2)})$, and $\{c, f\}$ are mapped to points in $\mathbb{X}(\bar{\mathcal{G}}^{(4)})$. Since the given graph is triangle-free, no points are mapped to $\mathbb{X}(\bar{\mathcal{G}}^{(3)})$.

graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with GSO \mathbf{S} and signal \mathbf{x} , (4) can be rewritten as:

$$\bar{\mathcal{H}}(\Sigma_K(r)) = [\mathcal{H}(\mathbf{S})\mathbf{x}]_r. \quad (9)$$

To use the terminology of quotient spaces [23], (9) indicates that a K -tap graph filter passes, or *descends*, to a unique map on Ω_K via the map Σ_K . Thus, in the context of Proposition 1 and graph filtering, it is sufficient to characterize local operations over Ω_K , rather than with respect to the global structure of a given graph \mathcal{G} .

Now that the local framework based on rooted balls has been introduced, we proceed to show how it can be leveraged to obtain theoretical results that provide insight into the inner workings of graph signal processing. In Section IV we discuss how to understand power spectral densities, while in Section V we discuss transferability of graph filters across graphs of different size.

Remark 2 (Distributions of edge sets): In [24], the authors consider a pushforward probability measure into the Fourier domain when there is an underlying probability distribution of edges in a fixed graph. This allows for the modeling of graph signals under uncertainty on the edge set. Such a distribution of edges can be incorporated into the distribution μ via the pushforward map $(\Sigma_K)_*$, allowing for a description of the distribution of rooted balls in a graph under a probability distribution on the edges.

Remark 3 (Rooted subtrees): Many authors have considered graph processing methods that are equivariant to local subtree structures [18], such as architectures whose computation resembles the 1-Weisfeiler-Lehman test [12]. In the proposed framework, these subtree structures are strictly “less expressive” than rooted ball structures, in the sense that the rooted subtree of depth K centered at a given node can be determined completely by the rooted K -ball centered at that node. However, the rooted K -ball cannot be determined by the rooted subtree of depth K , as evidenced by the inability of the 1-Weisfeiler-Lehman test to distinguish certain structures [25]. With this in mind, the proposed framework could certainly be applied where the space Ω_K is constructed with rooted subtrees, rather than rooted balls.

However, given that these rooted balls are more expressive than rooted subtrees [18], we state all proceeding results using the rooted ball construction.

IV. POWER SPECTRAL DENSITY

One of the key ideas in graph signal processing is that of the *graph Fourier transform* of a graph signal [2]. For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ on n nodes, consider the GSO \mathbf{S} to be the graph Laplacian, which is a positive semidefinite Hermitian matrix. Thus, it admits the following eigendecomposition

$$\mathbf{S} = \sum_{j=1}^n \lambda_j \mathbf{u}_j \mathbf{u}_j^T, \quad (10)$$

for eigenvalues $0 = \lambda_1 \leq \dots \leq \lambda_n$ with corresponding pairwise orthogonal eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_n$. One can check that for any of these eigenvectors, $\langle \mathbf{u}_j, \mathbf{S}\mathbf{u}_j \rangle = \lambda_j$. The graph Laplacian induces a quadratic form on a graph signal that measures a useful notion of smoothness [2]. For a given graph signal \mathbf{x} on \mathcal{G} , the quadratic form can be shown to be equal to

$$\langle \mathbf{x}, \mathbf{S}\mathbf{x} \rangle = \sum_{v \in \mathcal{V}} \sum_{u \in \mathcal{N}(v)} ([\mathbf{x}]_v - [\mathbf{x}]_u)^2. \quad (11)$$

That is, the quadratic form of the graph Laplacian measures the sum of the squared differences of the graph signal at neighboring pairs of nodes. A signal is called *smooth* if the quadratic form takes a small value relative to its norm. For instance, the eigenvectors of \mathbf{S} with small corresponding eigenvalues are smooth signals.

We note that this notion of smoothness can be extended to normalized versions of the graph Laplacian in a similar fashion. Furthermore, a notion of smoothness can also be defined using the adjacency matrix [26]. In any case, for ease of exposition and conceptual simplicity, we assume that \mathbf{S} is the graph Laplacian throughout this section, but we remark that the results derived herein are also valid for any other GSO built from adjacency rules that rely on the combinatorial structure of the graph.

Since the eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ form an orthonormal basis for \mathbb{R}^n , any graph signal \mathbf{x} admits a unique representation as

a linear combination of the eigenvectors of \mathbf{S} . That is, for any graph signal \mathbf{x} , there exists a set of coefficients $\{\tilde{x}_j\}_{j=1}^n$ such that the following holds:

$$\mathbf{x} = \sum_{j=1}^n \tilde{x}_j \mathbf{u}_j \quad (12)$$

$$\langle \mathbf{x}, \mathbf{S}\mathbf{x} \rangle = \sum_{j=1}^n \lambda_j \tilde{x}_j^2. \quad (13)$$

Thus, the representation of a graph signal as a weighted sum of the eigenvectors can be used to conveniently compute the quadratic form in terms of the spectrum. In an analogy to complex exponential functions being the eigenfunctions of the Laplace operator in discrete-time signal processing, we call the representation of the graph signal \mathbf{x} by the coefficients \tilde{x}_j the *graph Fourier transform* (GFT).

When coupled with their corresponding eigenvalues λ_j , the coefficients \tilde{x}_j can be used to describe the distribution of energy in a graph signal, in a way concordant with the notion of smoothness described by the Laplacian quadratic form (11). To capture this, define the *normalized power spectral distribution* of a graph signal. Given a graph signal \mathbf{x} with Fourier coefficients \tilde{x}_j and corresponding Laplacian eigenvalues $0 = \lambda_1 \leq \dots \leq \lambda_n$, define

$$P_{\mathbf{x}} : \mathbb{R} \rightarrow \mathbb{R}^+ \\ \lambda \mapsto \frac{1}{n} \sum_{j:\lambda_j \leq \lambda} \tilde{x}_j^2. \quad (14)$$

One can easily see that $P_{\mathbf{x}}$ is a monotone, right-continuous function, with $P_{\mathbf{x}}(\lambda) = 0$ for $\lambda < 0$, and at most n discontinuities (one at each eigenvalue λ_j). Moreover, there is a convenient expression for the moments of $P_{\mathbf{x}}$:

$$\mathbf{m}_K(\mathbf{x}) := \int_{\mathbb{R}} \lambda^K dP_{\mathbf{x}}(\lambda) = \frac{1}{n} \langle \mathbf{x}, \mathbf{S}^K \mathbf{x} \rangle. \quad (15)$$

That is to say, the moments of the normalized power spectral distribution are given by scaled quadratic forms of powers of the GSO. Observe that the quadratic form (11) can be expressed as $\langle \mathbf{x}, \mathbf{S}\mathbf{x} \rangle = n \cdot \mathbf{m}_1(\mathbf{x})$. Indeed, the moments of the graph Fourier transform are an essential quantity in the design of low-order graph filters [27]. This points to a very useful idea: the powers of the GSO are the most primitive types of graph filters, so the proposed local framework provides a path for understanding these moments in terms of functions on the space of distributions of rooted balls.

To see this, let us examine (15). Notice that

$$\frac{1}{n} \langle \mathbf{x}, \mathbf{S}^K \mathbf{x} \rangle = \frac{1}{n} \sum_{v \in \mathcal{V}} [\mathbf{x}]_v [\mathbf{S}^K \mathbf{x}]_v. \quad (16)$$

As before, let $\bar{\mathbf{S}}_K(v)$ be the GSO of the K -ball rooted at node v . By Proposition 1, the above equation can be written as

$$\frac{1}{n} \sum_{v \in \mathcal{V}} [\mathbf{x}]_v [\mathbf{S}^K \mathbf{x}]_v = \frac{1}{n} \sum_{v \in \mathcal{V}} [\bar{\mathbf{x}}_K]_v \left[\bar{\mathbf{S}}_K^K(v) \bar{\mathbf{x}}_K(v) \right]_v. \quad (17)$$

That is, the moments of the normalized power spectral distribution of a graph signal can be written as an average of operations that resemble local filtering. In the proposed local framework, this has the following interpretation.

Proposition 2: Let \mathcal{G} be a graph and \mathbf{x} be a graph signal. For a given $K \geq 0$, define the map

$$\bar{\mathbf{m}}_K : \Omega_K \rightarrow \mathbb{R} \quad (18)$$

$$((\mathcal{V}, \mathcal{E}, r), \bar{\mathbf{y}}) \mapsto [\bar{\mathbf{y}}]_r \left[\bar{\mathbf{S}}_K^K \bar{\mathbf{y}} \right]_r, \quad (19)$$

where $\bar{\mathbf{S}}_K$ denotes the Laplacian of the rooted K -ball.

Letting μ be the probability distribution on Ω_K determined by $(\mathcal{G}, \mathbf{x})$, the K^{th} moment of the normalized power spectral distribution $P_{\mathbf{x}}$ is given by the equation

$$\mathbf{m}_K(\mathbf{x}) = \int_{\Omega_K} \bar{\mathbf{m}}_K(\omega) d\mu(\omega) = \mathbb{E}_{\mu}[\bar{\mathbf{m}}_K]. \quad (20)$$

Having reduced the moments of the normalized power spectral distribution to an average over the distribution of rooted balls, we can now reason about the convergence of graph Fourier distributions in terms of weak convergence of measure for graphs with bounded degree D . Although the limit of dense graphs has been considered before via graphon models [28], [29], this cannot capture the behavior of bounded degree graphs, as all infinite sequences of growing graphs of bounded degree converge to the zero graphon [30]. However, in the case of sparse graphs, understanding the descension of maps to the space Ω_K is a feasible approach, with the following compactness property.

Lemma 1: For given integers $K \geq 0, D \geq 1$ and any collection of graphs in \mathbb{G}_D with uniformly bounded signals, there exists a compact subspace $\Gamma \subset \Omega_K$ such that the support of the probability measures corresponding to the graphs/graph signals is contained in Γ .

The proof of Lemma 1 can be found in Appendix C. The constraint of a graph having bounded degree, for instance coming from physical constraints in a real-world system, corresponds to a compact description in the space Ω_K . From this, the following convergence result holds.

Theorem 1: Let integers $K \geq 0, D \geq 1$ be given. Let $\{(\mathcal{G}_j, \mathbf{x}_j)\}_{j=1}^{\infty}$ be a sequence of graphs and graph signals such that each $\mathcal{G}_j \in \mathbb{G}_D$ and the collection of signals \mathbf{x}_j is uniformly bounded. Let μ_j be the probability measure on Ω_K associated with $(\mathcal{G}_j, \mathbf{x}_j)$ for each j , and P_j the corresponding normalized power spectral distribution function.

If the measures μ_j converge weakly, then the K^{th} moments of the normalized power spectral distribution functions converge. Moreover, if weak convergence of measure holds for all $K \geq 0$, then the normalized power spectral distribution functions converge weakly.

The proof of Theorem 1 can be found in Appendix B. Theorem 1 casts the power spectral density of a graph signal in terms of the distribution of local structures in the graph. In particular, we treat the power spectral density as a function from graphs with graph signals to Borel measures on a compact subset of the real line, then prove that this function is continuous with respect to the distribution of rooted balls in the graph, where continuity is understood to be with respect to the weak topology

for both the distribution in Ω_K and the power distribution of the signal. Although the frequency domain representation of a graph signal is typically understood to be a representation in terms of the *global* graph structure, owing to the global support of the graph Laplacian's eigenvectors, this results indicates that it is in essence still subject to fundamentally local phenomena, namely the rooted balls in the graph.

To illustrate this, we consider a toy example of growing sequences of path/cycle graphs with sinusoidal signals on the nodes. For integers $j \geq 1$, let \mathcal{P}_j be the path graph on nodes $[j]$, and \mathcal{C}_j the cycle graph on nodes $[j]$. For either graph, let \mathbf{s}_j be the graph signal mapping each node $v \in [j]$ to $\cos(\pi v/8)$. We now consider the moments of the power spectral distributions for the sequences $\{(\mathcal{P}_j, \mathbf{s}_j)\}_{j=1}^\infty$ and $\{(\mathcal{C}_j, \mathbf{s}_j)\}_{j=1}^\infty$. For any nonnegative integer $K \geq 0$, one can check by explicit calculation that the moments $\mathbf{m}_K(\mathbf{s}_j)$ converge to $(2(1 - \cos(\pi/8)))^K/2$ as $j \rightarrow \infty$ for both the sequence of path graphs and the sequence of cycle graphs. Although the path and cycle graphs with sinusoidal signals are globally distinguishable, they are (approximately) locally indistinguishable, particularly when considered on a large number of nodes.

This characterization of the power spectral distribution in terms of the distribution of rooted balls also reflects the local nature of graph filtering. Indeed, as discussed in [27], the moments of the power spectral distribution are key in the understanding of graph filters. More broadly, when designing a graph filter, we often aim to construct a *matched filter* in the frequency domain, so that the response of the filter aligns with that of the signal, subject to some noise model (additive white Gaussian noise, for instance). Additionally, constraints such as a filter having K taps reduces the space of possible frequency responses to degree K polynomials in the frequency domain, so that the moments become the basic values of interest. This is an immediate consequence of graph filters admitting a distributed implementation [27], so that the expression of a graph filter's performance in terms of the power spectrum of the signal is actually a distillation of the distribution of rooted balls in Ω_K .

We have shown how connecting the moments of the power spectral distribution to the space Ω_K yields insights about how features of graphs and graph signals are dependent only on localized information. In particular, a function $\overline{\mathbf{m}}_K : \Omega_K \rightarrow \mathbb{R}$ was constructed whose expected value under the pushforward measure of a graph computes the moments of the power spectral distribution. This machinery can be immediately extended to other functions on graphs that act locally, allowing us to compare the behavior of a given function on two graphs in terms of their distributions of rooted balls.

V. TRANSFERABILITY

Let J be a function mapping a graph and its corresponding graph signal to a real number. This function is typically known as a *graph summary* and examples include the graph power spectral distribution discussed in the previous section, as well as the average node degree or node centrality values [31], the clustering coefficient [32], or the conductance [33]. We can also consider *cost functions* for a given task as graph summaries, since they ultimately take a graph and its signal as input, and output a real

number [34]. In this context, we want to study how the function J changes across two different graphs with different graph signals $(\mathcal{G}_1, \mathbf{x}_1)$ and $(\mathcal{G}_2, \mathbf{x}_2)$.

To study the transference of the graph summary J across different graphs and graph signals, let us assume that there exists a function $\overline{J} : \Omega_K \rightarrow \mathbb{R}$ such that for any pair $(\mathcal{G}, \mathbf{x})$ whose associated probability distribution on Ω_K is denoted by μ , we have $J(\mathcal{G}, \mathbf{x}) = \mathbb{E}_\mu[\overline{J}]$. This assumption is not too severe, essentially saying that the graph summary is an average over a function on the nodes, with the value at each node determined strictly by the structure of its local neighborhood. Then, consider the *transfer equation*

$$|J(\mathcal{G}_1, \mathbf{x}_1) - J(\mathcal{G}_2, \mathbf{x}_2)| = |\mathbb{E}_\mu[\overline{J}] - \mathbb{E}_\nu[\overline{J}]|. \quad (21)$$

as a way of measuring the transferability of J , and where μ and ν are the distributions of rooted balls on the space Ω_K of the graphs $(\mathcal{G}_1, \mathbf{x}_1)$ and $(\mathcal{G}_2, \mathbf{x}_2)$, respectively.

To understand (21) in a quantitative manner demands a stronger structure on Ω_K than the disjoint union topology. To this end, we then endow Ω_K with the structure of a metric space, in a way that preserves the original topology. For an arbitrary constant $C > 0$, define a metric d_C on Ω_K as follows. For $\omega_1 = (\overline{\mathcal{G}}_1, \overline{\mathbf{x}}_1)$ and $\omega_2 = (\overline{\mathcal{G}}_2, \overline{\mathbf{x}}_2) \in \Omega_K$, put

$$d_C(\omega_1, \omega_2) = \begin{cases} \min\{\|\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\|_2, 2C\} & \overline{\mathcal{G}}_1 = \overline{\mathcal{G}}_2 \\ C & \text{otherwise,} \end{cases} \quad (22)$$

where $\|\cdot\|_2$ is inherited from the identification of $\mathbb{X}(\overline{\mathcal{G}})$ with Euclidean space.² Note that $\|\cdot\|_2$ is well defined since both \mathbf{x}_1 and \mathbf{x}_2 are vectors in the same $|\mathbb{V}|$ -dimensional space when $\overline{\mathcal{G}}_1 = \overline{\mathcal{G}}_2$. One can check that d_C is indeed a metric on Ω_K for any $C > 0$. The metric d_C is constructed in a way that preserves the *local* metric structure of each signal space $\mathbb{X}(\overline{\mathcal{G}})$, while also endowing the whole space Ω_K with a *global* metric structure, at the cost of some local distortion in order to maintain the triangle inequality, as dictated by the constant C . When C is chosen to be very large, the local metric structure on each rooted ball of the space is well-preserved, but these rooted balls are otherwise kept "far apart" from each other. On the other hand, when C is very small, the rooted balls become "close," but the local metric structure on each of them is distorted (i.e. signals belonging to the same rooted ball that are beyond a distance of $2C$ are indistinguishable from signals that belong to different rooted balls). Importantly, the metric topology on Ω_K induced by d_C is equal to the disjoint union topology (6), so that all notions of continuity and weak convergence of measure are preserved.

Given the additional metric structure on Ω_K , we can quantitatively characterize the transfer (21) by comparing distributions of rooted balls on the metric space (Ω_K, d_C) . To do so, we make the following assumption on the function \overline{J} .

Assumption 1: For all rooted K -balls $\overline{\mathcal{G}} \in \overline{\mathcal{G}}_D$, $\overline{J}(\overline{\mathcal{G}}, \overline{\mathbf{x}})$ is L -Lipschitz continuous with respect to the second argument $\overline{\mathbf{x}}$ over the space of bounded signals $\mathbb{X}(\overline{\mathcal{G}}, [-1, 1])$.

This is not a difficult assumption to satisfy. For instance, it is sufficient for \overline{J} to be continuously differentiable on each signal

²This metric is again understood modulo the automorphism group of the rooted ball, discussed in Appendix A.

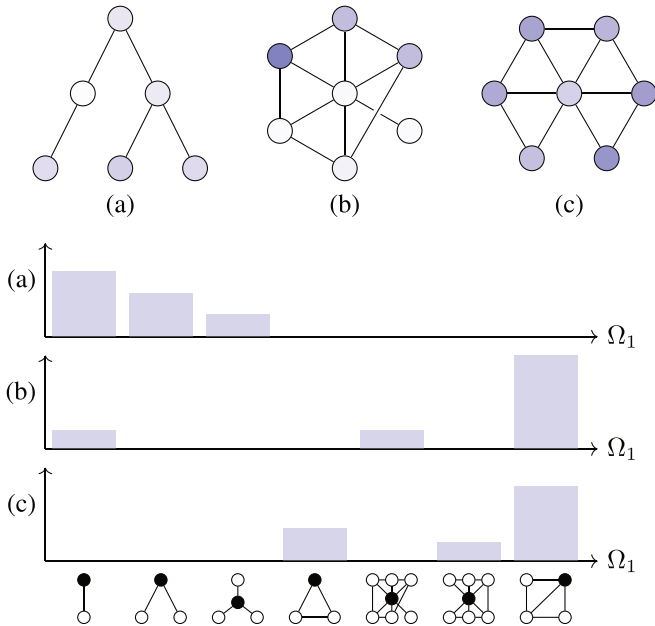


Fig. 2. The 1-Wasserstein distance reflects structural similarity between graphs. **(Top)** Three graphs (a, b, c) with signals indicated by node coloring. **(Bottom)** Distributions of rooted 1-balls in each graph (a, b, c), represented by histograms. The x -axis corresponds to the space Ω_1 , and the y -axis shows the density of each rooted 1-ball in the graph. The 1-Wasserstein distance, then, essentially describes the distance between histograms, subject to a metric structure on Ω_1 . For instance, graph (a) is expected to have a large 1-Wasserstein distance from graphs (b, c), since there is minimal overlap between their histograms. On the other hand, graphs (b, c) have substantial overlap, so their 1-Wasserstein distance will be smaller.

space for this to hold. In fact, the set of Lipschitz continuous functions on a compact space is dense in the set of continuous functions with respect to the uniform norm [35].

Theorem 2: Let graphs $\mathcal{G}_1, \mathcal{G}_2 \in \mathbb{G}_D$ with corresponding bounded signals $\mathbf{x}_1, \mathbf{x}_2$ be given. Denote their corresponding measures in Ω_K by μ and ν , respectively. For a function $\bar{J} : \Omega_K \rightarrow [0, 1]$ that satisfies Assumption 1, it holds that

$$|\mathbb{E}_\mu[\bar{J}] - \mathbb{E}_\nu[\bar{J}]| \leq L \cdot W_1\left(\mu, \nu; \frac{1}{L}\right), \quad (23)$$

where $W_1(\mu, \nu; C)$ denotes the 1-Wasserstein distance between μ and ν under the metric d_C .

The proof of Theorem 2 can be found in Appendix D. Theorem 2 bounds the transfer (21) in terms of the 1-Wasserstein distance between the two distributions of rooted balls. The appearance of the 1-Wasserstein distance in this setting is intuitive: the dual formulation of the 1-Wasserstein distance defines the distance between two distributions via an integral probability metric over 1-Lipschitz functions [36], yielding a transferability bound that holds for all Lipschitz functions. We illustrate the notion of the 1-Wasserstein distance between graphs in Fig. 2.

A. Examples and Discussion

Theorem 2 indicates that smoother functions \bar{J} generally transfer better than ones with large Lipschitz constant. Some examples of graph summaries that fulfill the conditions of Theorem 2 follow.

Example: Power spectral distribution. For any $K \geq 0$, the spectral moment function \mathbf{m}_K as defined in Section IV descends to the expectation of the function $\bar{\mathbf{m}}_K : \Omega_K \rightarrow \mathbb{R}$. For any rooted K -ball $\bar{\mathcal{G}}$, it holds that $\bar{\mathbf{m}}_K(\bar{\mathcal{G}}, \cdot)$ is a polynomial of the graph signal in the second argument, and can thus be shown to satisfy Assumption 1. Thus, the difference in the spectral moments of two graphs can be bounded in terms of the Wasserstein distance of their respective distributions of rooted balls. Moreover, observe that for $K \geq M \geq 0$, the moment function $\bar{\mathbf{m}}_M$ can be defined on Ω_K in a way that preserves the identity (20), so that we can compare the similarity of spectral moments between graph signals by comparing the Lipschitz constants of the functions $\bar{\mathbf{m}}_K$ and $\bar{\mathbf{m}}_M$ on the same space Ω_K . Indeed, one can see that $\bar{\mathbf{m}}_K$ has a larger Lipschitz constant than $\bar{\mathbf{m}}_M$, so that the bound given by Theorem 2 is tighter for the lower moments of the power spectral distribution, contingent upon a relatively small 1-Wasserstein distance between the distributions of rooted balls.

Example: MSE of a graph filter. Let \mathcal{H} be a fixed K -tap graph filter, and let $\sigma^2 > 0$ be given. For any given graph \mathcal{G} on n nodes with associated signal \mathbf{x} and shift operator \mathbf{S} , put the function J as

$$J(\mathcal{G}, \mathbf{x}) = \mathbb{E} \left[\frac{1}{n} \|\mathbf{x} - \mathcal{H}(\mathbf{S})(\mathbf{x} + \boldsymbol{\eta})\|_2^2 \right], \quad (24)$$

where the expectation is taken over the random vector $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. In words, J is the mean squared error when applying \mathcal{H} to $(\mathcal{G}, \mathbf{x})$ under additive white Gaussian noise. Define \bar{J} on rooted K -balls $\bar{\mathcal{G}} = (\mathcal{V}, \mathcal{E}, r)$ with shift operator $\bar{\mathbf{S}}$ as

$$\bar{J}(\bar{\mathcal{G}}, \bar{\mathbf{x}}) = ([\bar{\mathbf{x}}]_r - [\mathcal{H}(\bar{\mathbf{S}}) \bar{\mathbf{x}}]_r)^2 + \sigma^2 \left[(\mathcal{H}(\bar{\mathbf{S}}))^2 \right]_r. \quad (25)$$

If we denote the measure on Ω_K associated with $(\mathcal{G}, \mathbf{x})$ by μ , one can show that

$$J(\mathcal{G}, \mathbf{x}) = \mathbb{E}_\mu[\bar{J}]. \quad (26)$$

Observe that $\bar{J}(\bar{\mathcal{G}}, \bar{\mathbf{x}})$ is a polynomial of the values of $\bar{\mathbf{x}}$ for each rooted K -ball $\bar{\mathcal{G}}$, so that it is Lipschitz continuous on each bounded signal space, thus satisfying Assumption 1. Moreover, the Lipschitz constant of \bar{J} over each signal space $\mathbb{X}(\bar{\mathcal{G}})$ can be shown to be proportional to $\|(\mathbf{I} - \mathcal{H}(\bar{\mathbf{S}}))\boldsymbol{\delta}_r\|_2$, where $\boldsymbol{\delta}_r$ is the signal taking value 1 on the root node and 0 elsewhere. That is to say, the Lipschitz constant is directly influenced by the ‘‘spread’’ of the impulse response of the filter. Intuitively, this indicates that while filters with large coefficients in their higher-order terms may suit a particular signal well, they may transfer across graphs poorly. This aligns with existing results on the stability of graph filters [6], [37].

Estimation of the transferability bound. The computation of the 1-Wasserstein distance in (23) may prove problematic for large graphs. However, given the definition of the 1-Wasserstein distance as an integral probability metric, it is not difficult to compute an empirical estimate of $W_1(\mu, \nu; C)$. By [36, Theorem 6], if $\{\omega_j^\mu\}_{j=1}^m$ and $\{\omega_j^\nu\}_{j=1}^n$ are i.i.d. samples drawn from the respective distributions μ and ν , then the 1-Wasserstein distance between the empirical distributions can be computed via a linear program. To draw i.i.d. samples from these distributions, one can choose a node uniformly at random (with replacement)

from the node set of the underlying graph and then take the rooted K -ball via the map Σ_K . By [36, Corollary 10], the compact support of the distributions μ, ν guarantees almost sure convergence of the empirical estimator to the true Wasserstein distance as $m, n \rightarrow \infty$.

Tighter bounds. In Theorem 2, the difference in expectation between two distributions of rooted balls is characterized in terms of their Wasserstein distance, which is defined in terms of a metric imposed upon Ω_K . Given that the metric d_C is defined for essentially arbitrary C , one can derive a tighter bound on the transfer equation, with less stringent assumptions on the function \bar{J} . First, note that for two graphs $\mathcal{G}_1, \mathcal{G}_2 \in \mathbb{G}_D$ and bounded signals $\mathbf{x}_1, \mathbf{x}_2$, the corresponding measures μ, ν have support in the compact subspace $\Gamma \subset \Omega_K$, where Γ is as described in Lemma 1. Because of this, any continuous function on Γ is bounded. For a function $\bar{J} : \Omega_K \rightarrow \mathbb{R}$ satisfying Assumption 1, define the range of \bar{J} as

$$A = \sup_{\omega \in \Gamma} \bar{J}(\omega) - \inf_{\omega \in \Gamma} \bar{J}(\omega). \quad (27)$$

In the same vein as Theorem 2, it holds that

$$|\mathbb{E}_\mu[\bar{J}] - \mathbb{E}_\nu[\bar{J}]| \leq \inf_{C \in (0,1]} \frac{L}{C} \cdot W_1 \left(\mu, \nu; \frac{AC}{L} \right). \quad (28)$$

This bound illustrates an interplay between the smoothness of \bar{J} , the range of \bar{J} , and the regime of similarity between the two graphs. If two graphs have similar distributions of rooted balls, then the smoothness of \bar{J} plays a stronger role in bounding transferability. If they have highly distinct distributions of rooted balls, the range of \bar{J} , as described by the value A , plays a stronger role in bounding transferability.

Extensions to graph neural networks. Although we have considered maps on graphs with real-valued signals, the analysis in this section is equally applicable when the nodes have other types of features such as categorical labeling. For example, one could consider J to be the cross-entropy loss for a fixed graph neural network with K layers, where the ‘‘graph signal’’ is taken to be a set of discrete input features and a ground-truth label against which the loss is evaluated [38]. In the context of Theorem 2, the Lipschitz constants of the nonlinearity and the operator norm of the weights influence the Lipschitz constant, and the depth of the graph neural network determines the value of K . We aim to address this relationship in greater detail in future work, with a particular focus on understanding how a given graph neural network trained on a single graph might generalize to a distribution of graphs with similar characteristics.

Graph kernels. Wasserstein distances between graphs have been considered in the graph kernel literature [15], where node embeddings in Euclidean space are computed from local structures about each node, followed by a computation of the Wasserstein distance in Euclidean space. Given that the proposed computation for nodes with continuous attributes in this paper is smooth with respect to the node attributes, one can show that the Wasserstein distance between distributions in our case descends to theirs via a continuous map. Also related are the Weisfeiler-Lehman subtree kernels [25] that compare graphs based on their sets of subtree structures, which can be determined by the

rooted ball structures (see Remark 3). The analytical tools we put forth are particularly well-suited for graphs with continuous attributes, unlike the discrete attributes usually considered in the graph kernel literature.

By endowing the space Ω_K with an appropriate metric structure, we have applied the proposed framework to studying transferability of graph summaries that descend to maps on rooted K -balls. In analyzing the relationship between the Lipschitz constant of the graph summary and the distribution of rooted balls, insights have been gleaned as to what drives commonalities between graphs in this regard. Although our focus in this section has been on fixed graph summaries, such as the loss of a fixed graph filter in Section V-A, these ideas can be readily adapted to problems of designing graph summaries. For instance, the presence of the Lipschitz constant in Theorem 2 indicates that a graph filter or graph neural network will transfer better if it is designed to be smooth. This can be cast in the light of stability, where certain conditions on filter banks in graph neural networks can guarantee a response that is stable to structural perturbations [6], [37]. To summarize, in the context of graph filtering, graph kernels, or even graph neural networks, our results formalize the interplay between smoothness of functions and appropriate notions of similarity between graphs to yield a more formal understanding of the phenomenon of transferability in GSP methods. This allows for a precise, yet flexible, way to characterize the application of methods to graphs of different sizes and geometries.

VI. EXTENSION TO WEIGHTED GRAPHS

In this section, we consider how ideas from the unweighted case (Sections IV and V) can be extended to weighted graphs. We define a *weighted graph* as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ coupled with a weight function $w : \mathcal{E} \rightarrow \mathbb{R}_{\geq 0}$. For a given graph \mathcal{G} , the set of all possible weight functions on it is denoted $\mathbb{W}(\mathcal{G}) = \{w : \mathcal{E} \rightarrow \mathbb{R}_{\geq 0}\}$. We use the same notation for the set of weight functions on a rooted graph: $\mathbb{W}(\bar{\mathcal{G}})$.

We now expand our definition of a rooted K -ball with a signal to include weight functions. Define

$$\Omega_K^W = \prod_{\bar{\mathcal{G}}_k : k \leq K} \mathbb{W}(\bar{\mathcal{G}}) \times \mathbb{X}(\bar{\mathcal{G}}). \quad (29)$$

Similar to the definition of Ω_K in (6), the space Ω_K^W consists of all rooted K -balls with both edge weights and graph signals on them. As before, a weighted graph with a signal $(\mathcal{G}, w, \mathbf{x})$ yields elements of Ω_K^W via a suitably defined sampling map Σ_K , defined for nodes $v \in \mathcal{V}$:

$$\Sigma_K(v) = (\bar{\mathcal{G}}_K(v), \bar{w}_K(v), \bar{\mathbf{x}}_K(v)), \quad (30)$$

where $\bar{w}_K(v)$ denotes the restriction of w to the edges of $\bar{\mathcal{G}}_K(v)$. As before, a given weighted graph with a signal defines a probability measure on Ω_K^W via the pushforward of the uniform measure by Σ_K , which we denote by $\mu = (\Sigma_K)_*(\mathcal{G}, w, \mathbf{x})$. For a weighted graph (\mathcal{G}, w) , define the *weighted degree* of a node

$v \in \mathcal{V}$ as follows:

$$\deg(v) = \sum_{u \in \mathcal{N}(v)} w(v, u). \quad (31)$$

The set of weighted graphs whose nodes have weighted degree at most D is denoted by \mathbb{G}_D^W for real number $D > 0$.

We now develop a local description of the graph Fourier transform for weighted graphs, much like in Section IV. For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with weight function w , define the *weighted graph Laplacian* as a shift operator such that, for any $u, v \in \mathcal{V}$,

$$[\mathbf{S}]_{uv} = \begin{cases} \deg(v; w) & u = v \\ -w(u, v) & (u, v) \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases} \quad (32)$$

The weighted Laplacian for a graph on n nodes is positive semidefinite, and admits an eigendecomposition $\mathbf{S} = \sum_{j=1}^n \lambda_j \mathbf{u}_j \mathbf{u}_j^T$, with a Fourier representation for graph signals defined in a way analogous to (12). For a given graph signal $\mathbf{x} \in \mathbb{X}(\mathcal{G})$, there is a power spectral distribution associated with \mathbf{x} via the weighted Laplacian, whose moments are given by (15). We characterize the properties of the power spectral distribution in terms of the distribution of rooted K -balls in the following theorem.

Theorem 3: Let an integer $K \geq 0$ and a real number $D > 0$ be given. Let $\{(\mathcal{G}_j, w_j, \mathbf{x}_j)\}_{j=1}^\infty$ be a sequence of graphs, weight functions, and graph signals such that each graph is contained in \mathbb{G}_D^W and the signals \mathbf{x}_j are uniformly bounded. Let μ_j be the probability measure on Ω_K^W associated with $(\mathcal{G}_j, w_j, \mathbf{x}_j)$ for each j , and P_j the corresponding normalized power spectral distribution function.

If the measures μ_j converge weakly, then the K^{th} moments of the normalized power spectral distribution functions converge. Moreover, if weak convergence of measure holds for all $K \geq 0$, then the normalized power spectral distribution functions converge weakly.

The proof of Theorem 3 can be found in Appendix E. Theorem 3 establishes that the power spectral distribution of a graph signal on a weighted graph is continuous with respect to the weak topology of distributions of rooted K -balls, under boundedness assumptions. Unlike the regime of Theorem 1, the convergence of the power spectral distribution for weighted graphs here is not dependent on the combinatorial constraint of having bounded node degrees. Rather, any node may have an arbitrarily large number of neighbors, as long as the total influence remains bounded. This is a realistic assumption for many systems, where an agent may be allowed to exert influence on a large collection of other agents, but the total influence is bounded by some physical constraint, such as power consumption.

VII. IN THE LIMIT: GRAPHING SIGNAL PROCESSING

In Sections V and VI, the (weak) continuity of the power spectral density with respect to the distribution of rooted balls was established. Specifically, it was shown that weakly convergent sequences of distributions of rooted balls yielded weakly convergent power spectra. Given that these distributions correspond to underlying graphs, it remains to be established precisely

what objects these sequences are converging to. In general, a weakly convergent sequence of finite graphs does not necessarily converge to a finite graph. When studying graph limits using graphon models, the homomorphism density of motifs is of primary concern [39]. This setting is slightly different, depending on the isomorphism density of rooted graphs. In this setting, a more appropriate model is that of a graphing. We show that the basic ideas in the discussion so far can be transferred directly to these limiting objects. Let us define the basic object of study for this section.

Definition 1: A *graphing* of degree D is a triplet $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \lambda)$ such that

- 1) \mathcal{V} is a sample space with a σ -field \mathcal{B} on \mathcal{V}
- 2) λ is a probability measure on \mathcal{B}
- 3) $\mathcal{E} \in \mathcal{B} \times \mathcal{B}$ is such that, for all $A, A' \in \mathcal{B}$, we have

$$\int_A \deg(v, A') d\lambda(v) = \int_{A'} \deg(v, A) d\lambda(v), \quad (33)$$

where $\deg(v, A) = |\{u \in A : (u, v) \in \mathcal{E}\}|$, with the condition that $\deg(v, \mathcal{V}) \leq D$ for all $v \in \mathcal{V}$.

A graphing is a way to describe a graph with a potentially uncountable number of nodes (elements of \mathcal{V}), yet with bounded node degrees. We tacitly assume that all graphings considered are of degree D , for some $D \geq 1$. Unlike a graphon, which describes dense graphs with unbounded node degrees, the structures that arise from a graphing are typically sparse, and notions of graph filtering and graph shifts reduce to finite sums, rather than continuous integrals. Given a graphing $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \lambda)$, a *graphing signal* is a map $\mathbf{x} : \mathcal{V} \rightarrow \mathbb{R}$ between the nodes \mathcal{V} and the real numbers \mathbb{R} such that \mathbf{x} is a measurable function. Endowing this with the usual vector space structure, we define the space of graphing signals as

$$\mathbb{X}(\mathbf{G}) = \{\mathbf{x} : \mathcal{V} \rightarrow \mathbb{R} \mid \mathbf{x} \text{ is measurable}\}. \quad (34)$$

Much like how a graphon describes a model for dense random graphs, a graphing carries with it a distribution of random rooted graphs. As before, let \mathcal{N} be the neighborhood operator, returning all of the one-hop neighbors of a node $v \in \mathcal{V}$, so that \mathcal{N}^K is the K -hop neighborhood operator. For a graphing $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \lambda)$ with associated signal $\mathbf{x} \in \mathbb{X}(\mathbf{G})$, the rooted K -ball at a node $r \in \mathcal{V}$ is defined in the same way as in Section III, denoted by $\bar{\mathcal{G}}_K(r)$, with the same notation used to denote the corresponding graph signal $\bar{\mathbf{x}}_K(r)$. Due to the bounded degree condition on the graphing, $\bar{\mathcal{G}}_K(r)$ is always a finite rooted graph of maximum degree at most D . With this in mind, we define a sampling map $\Sigma_K : \mathcal{V} \rightarrow \Omega_K$ for graphings much like the sampling map (7) for finite graphs:

$$\Sigma_K(v) = (\bar{\mathcal{G}}_K(v), \bar{\mathbf{x}}(v)). \quad (35)$$

The sampling map Σ_K is illustrated in Fig. 3 for a simple graphing. Much like for finite graphs, we can pushforward the measure λ to yield a probability measure μ on Ω_K . That is to say, $\mu = (\Sigma_K)_*(\mathbf{G}, \mathbf{x})$, in the same way as before, where we adopt λ as the probability measure over \mathcal{V} to be pushed forward. This differs slightly from the case of finite graphs, where we implicitly assume that the initial measure was the uniform probability measure on the finite node set.

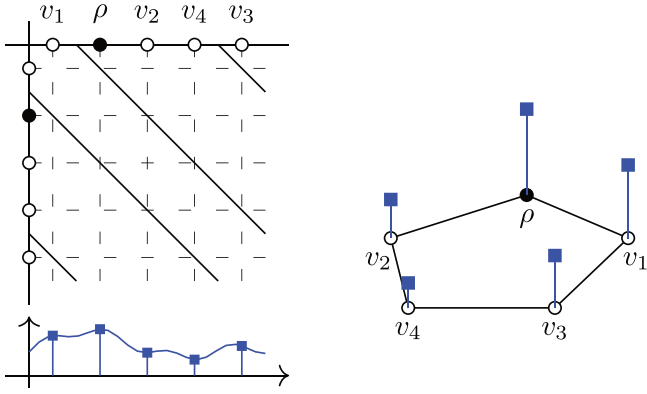


Fig. 3. Sampling of a random rooted graph from a graphing with an associated signal. **(Upper left)** A graphing $\mathbf{G} = ([0, 1], \mathcal{E}, \lambda)$. The diagonal lines indicate adjacency between nodes in $[0, 1]$ via the edge set \mathcal{E} : observe that although there are uncountably many nodes, each node has a finite, bounded number of neighbors. We sample from \mathbf{G} by first selecting a root node ρ at random (colored in black), and then taking the connected component of ρ in the edge set. That is, starting with the node ρ , we see that the nodes v_1, v_2 are neighbors of ρ . Then, the nodes v_3, v_4 are neighbors of v_1, v_2 , with the process stopping there. The intersections of the dashed lines indicate the entries of the adjacency matrix of the subgraph induced by $\{\rho, v_1, v_2, v_3, v_4\}$. **(Lower left)** A graphing signal $\mathbf{x}: [0, 1] \rightarrow \mathbb{R}$. After sampling a node and edge set from the graphing, we sample this function at the corresponding values. **(Right)** The drawn rooted graph and signal sampled from the graphing.

To establish the basic building blocks for signal processing on graphings, we first construct a Laplacian for graphing signals. Unlike graph signals, graphing signals are not necessarily supported on a finite, or even countable space. We specify the graphing Laplacian as a linear operator on the space of graphing signals in the following way. For any $\mathbf{x} \in \mathbb{X}(\mathbf{G})$ and $v \in \mathcal{V}$, put

$$[\mathbf{S}\mathbf{x}]_v = \sum_{u \in \mathcal{N}(v)} [\mathbf{x}]_v - [\mathbf{x}]_u. \quad (36)$$

By the degree boundedness condition from Definition 1, the sum in (36) is well-defined. Due to its singular nature, it is difficult to define a “graphing Fourier transform” directly from the Laplacian: for instance, even fairly tame graphing structures may have a Laplacian whose eigenvalues have uncountable multiplicity. This, for instance, makes the notion of the power spectral distribution of a graphing signal unwieldy.

However, the spectral properties of a graphing signal with respect to the underlying graphing structure can be studied indirectly. Let \mathbf{G} be a graphing with graphing signal $\mathbf{x} \in \mathbb{X}(\mathbf{G})$ and Laplacian \mathbf{S} . Then, for integers $K \geq 0$, we tentatively define

$$\mathbf{m}_K(\mathbf{x}) = \int_{\mathcal{V}} [\mathbf{x}]_v [\mathbf{S}^K \mathbf{x}]_v d\lambda(v), \quad (37)$$

where \mathbf{S}^K again indicates K repeated applications of the Laplacian. See that this definition resembles that of the moments of the power spectral distribution of a graph signal in (15). We define the values $\mathbf{m}_K(\mathbf{x})$ tentatively, as it is not obvious when they are finite, or even well-defined. In the following discussion, we establish sufficient conditions under which the sequence $\{\mathbf{m}_K(\mathbf{x})\}_{K=0}^{\infty}$ exists and corresponds to the moments of a distribution function.

Much like in Sections IV and V, we control the behavior of the graphing by bounding its size. We first do this by determining a suitable notion of boundedness for a graphing signal.

Definition 2: Let $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \lambda)$ be a graphing and $\mathbf{x} \in \mathbb{X}(\mathbf{G})$ be a graphing signal. The signal \mathbf{x} is said to be locally essentially bounded if it is bounded on almost all of the connected components of \mathbf{G} . That is, there exists an $a \geq 0$ such that for all $K \geq 0$,

$$\lambda(\mathcal{N}^K((\mathbf{x}^{-1}[-a, a])^c)) = 0, \quad (38)$$

where $(\mathbf{x}^{-1}[-a, a])^c$ denotes the set of nodes $v \in \mathcal{V}$ such that $[\mathbf{x}]_v \notin [-a, a]$, and \mathcal{N}^K is the K -hop neighborhood of a set. Denote the set of all such graphing signals by $\mathbb{X}_{\text{LB}}(\mathbf{G})$.

In words, local essential boundedness not only controls the size of the set of nodes with large signal value (essential boundedness), but also controls the size of the neighborhoods with which those nodes can interact. Local essential boundedness is an analog of boundedness for finite graph signals adapted to graphing signals. For instance, a graphing signal that is strictly bounded is locally essentially bounded. This condition, however, is stronger than being bounded almost everywhere (see Lemma 4), due to the neighborhood condition in (38). The necessity of this condition stems from the highly singular nature of the graphing Laplacian as an operator on signals.

We will also find it useful to consider graphings that “resemble” finite graphs in some sense.

Definition 3: Let $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \lambda)$ be a graphing with graphing signal $\mathbf{x} \in \mathbb{X}(\mathbf{G})$, and $\{(\mathcal{G}_j, \mathbf{x}_j)\}_{j=1}^{\infty}$ a sequence of graphs with associated graph signals. For all $K \geq 0$, put $\mu_K = (\Sigma_K)_*(\mathbf{G}, \mathbf{x})$, and $\mu_K(j) = (\Sigma_K)_*(\mathcal{G}_j, \mathbf{x}_j)$. Then, we say that the sequence of graphs $\{(\mathcal{G}_j, \mathbf{x}_j)\}_{j=1}^{\infty}$ converges weakly to (\mathbf{G}, \mathbf{x}) if for all $K \geq 0$, $\mu_K(j)$ converges to μ_K as j tends to infinity. We denote weak convergence of graphs to a graphing by $(\mathcal{G}_j, \mathbf{x}_j) \rightharpoonup (\mathbf{G}, \mathbf{x})$.

Much like Theorem 1, we can characterize the properties of a graphing based on the limiting properties of a sequence of graphs that converges to it. We state this formally below.

Theorem 4: Let a graphing \mathbf{G} of degree D with signal $\mathbf{x} \in \mathbb{X}_{\text{LB}}(\mathbf{G})$ be given. Suppose there exists a sequence of graphs and graph signals $\{(\mathcal{G}_j, \mathbf{x}_j)\}_{j=1}^{\infty}$ such that $(\mathcal{G}_j, \mathbf{x}_j) \rightharpoonup (\mathbf{G}, \mathbf{x})$. Then, there exists a unique distribution function $P_{\mathbf{x}}$ supported on $[0, 2D]$ such that the moments of $P_{\mathbf{x}}$ are given by the sequence $\{\mathbf{m}_K(\mathbf{x})\}_{K=0}^{\infty}$.

The proof of Theorem 4 can be found in Appendix F. Theorem 4 defines the power spectral distribution for sufficiently nice graphing signals. In particular, if the graphing and graphing signal are the limit of a sequence of graphs and graph signals, then asking that the power spectral distribution be continuous with respect to the weak topology on distributions forces the distribution to take on a unique value. Namely, the power spectral distribution of the limit of a sequence of graphs and graph signals is merely the weak limit of the power spectral distributions of the graphs as defined in (14).

In Sections IV to VI, we considered how the distribution of rooted balls in graphs yields useful topological or metric structure with which graphs can be related to each other, particularly

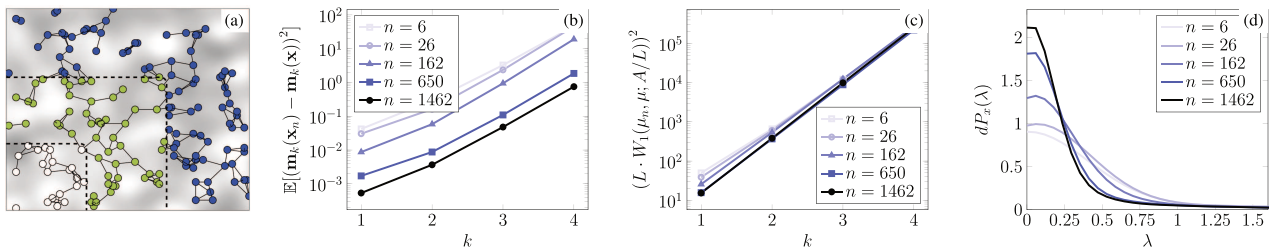


Fig. 4. (a) Pruned random sensor network model, sampling a Gaussian random field over an increasing area. (b) Observed MSE of k^{th} moments compared to the largest ($n = 1462$) random graph. (c) MSE of k^{th} moments relative to ground truth determined by bounds in Theorem 2. (d) Weak convergence of power spectra. For visibility, we apply kernel density estimation to the derivative of the power spectral distribution function, yielding a density plot.

in light of graph summaries such as the power spectral distribution. In all of these cases, the topology of weak convergence was used to show that a convergent sequence of graphs in the sense of the distribution of rooted balls is also convergent in the sense of any appropriate graph summary. This left the question open: what do the graphs converge to? By considering graphings as limiting objects for sparse graphs, we have shown that the extension of graph summaries, such as the power spectral distribution, to graphings maintains continuity in the natural way.

VIII. SIMULATIONS

To illustrate the theoretical results in Sections IV and V, we consider a simulated sensor network example. We assume that a Gaussian random field on \mathbb{R}^2 is being measured, with zero mean and covariance between points $x, y \in \mathbb{R}^2$ given by $C(x, y) = \exp(-\|x - y\|_2^2)$. To measure the properties of the signal, we uniformly place nodes at a density of 6.5 nodes per unit area on squares of edge length 1, 2, 5, 10, 15, yielding random graphs with size $n \in \{6, 26, 162, 650, 1462\}$, as illustrated in Fig. 4(a). Then, a graph is constructed by connecting nodes that are within a distance of 0.5 from each other, and then pruned so that each node has maximum degree 3. The signal on this graph is then sampled from the realization of the Gaussian random field.

By the stationarity of the Gaussian random field, and the locality of the random graph model, we expect the power spectrum of the graph signals to converge as the sampled area increases. We take the k^{th} moments of the power spectral distribution over 25 trials, for $k \in \{1, 2, 3, 4\}$. Treating the largest graph ($n = 1462$) as the ground truth, we plot the mean squared error for each graph size relative to the ground truth in Fig. 4(b), where the $n = 1462$ line plots the variance over the 25 trials. Observe that as the graph grows, the moments converge to that of the ground truth, as expected by Theorem 1.

Additionally, Theorem 2 yields a quantitative understanding of these convergence properties. To illustrate this, we estimate the Lipschitz constant L for each k directly from the data, and then plot the bound on the MSE according to Theorem 2 for each graph size and value of k in Fig. 4(c). Again, the line for $n = 1462$ indicates the variance within the realizations of the graph model of that size. Although there is a difference in the magnitude of the MSE between Fig. 4(b) and 4(c), observe that the relative shapes of the lines are the same. Namely, the larger graphs have tighter bounds with respect to the ground

truth than the smaller ones, as expected. Moreover, since the k^{th} moments of the power spectral distribution have larger Lipschitz constants and more demanding conditions for convergence, the bounds loosen as k increases. This is further demonstrated by the convergence of the power spectral densities of the graph signals shown in Fig. 4(d). By plotting the derivative of the power spectral distribution functions (averaged over the trials) as defined in (14), we roughly see that the “shape” of the signal spectrum converges as the graph grows. Moreover, based on the Lipschitz constants of the moments, the lower-order features of the distribution converge faster than the higher-order ones, e.g., it is clear for all graph sizes that the signal is low-pass, but the degree of (approximate) bandlimitedness is not apparent until larger graphs are considered.

Remark 4 (Complexity of Bound Computation): The calculation of the bounds for Fig. 4(c) is computationally demanding for two reasons: first, the computation of the 1-Wasserstein distance has cubic complexity in the number of nodes, and second, the computation of the distance d_C requires the comparison and identification of all pairs of rooted balls from the graphs by applying the graph isomorphism test to a large collection of small graph structures, which can be done in polynomial time for the bounded-degree case [40]. We would like to note that in the context of this work, we are not proposing the use of Wasserstein distances as a methodological tool for graph signal processing – rather, we are proposing their use as an analytical tool for understanding the phenomenon of transferability.

IX. CONCLUSION

We have developed a local framework for graph signal processing, highlighting the local nature of graph filters. In particular, by mapping a graph and graph signal to a probability distribution on a fixed space, we allow for the application of tools from probability theory to characterize similarities between graphs based on their local structures, even if they do not have a common node set. Using the topology of weak convergence of probability distributions, we define a topology on the space of graphs and graph signals, where notions of continuity can be understood: this was most obviously demonstrated in Section IV, where it was proved that the power spectral distribution of a graph signal is weakly continuous with respect to the distribution of rooted balls in a graph. This approach allows for relatively

painless treatment of limiting objects, such as *graphings*, by simply decomposing them into their constituent distributions.

The analysis carried out in this paper provides proper justification for transferring graph filters across multiple graphs, given the assumption that they share local structural properties. With the particular suitability of this approach to highly sparse graphs, this complements existing work on transferability based on graphon models or assumptions of graphs as discretizing manifold structures. The framework of this paper provides a path for the detailed study of more exotic functions, such as graph neural networks and graph kernels, or the incorporation of properties encoded by distributions of local structures into tasks such as network topology inference [41].

APPENDIX A

METRIC TOPOLOGY OF THE SIGNAL SPACE

Let a rooted ball $\bar{\mathcal{G}} = (\mathcal{V}, \mathcal{E}, r)$ be given. Denote by $\text{Aut}(\bar{\mathcal{G}})$ the automorphism group of maps $\{\phi : \mathcal{V} \rightarrow \mathcal{V}\}$, i.e., the set of all isomorphisms from $\bar{\mathcal{G}}$ to itself, with the group operation given by composition. The group $\text{Aut}(\bar{\mathcal{G}})$ acts on $\mathbb{X}(\bar{\mathcal{G}})$ in a natural way. For any $\phi \in \text{Aut}(\bar{\mathcal{G}})$, $\mathbf{x} \in \mathbb{X}(\bar{\mathcal{G}})$, define the action of ϕ on \mathbf{x} as the graph signal $\phi(\mathbf{x})$ satisfying

$$[\phi(\mathbf{x})]_{\phi(v)} = [\mathbf{x}]_v \quad (39)$$

for all $v \in \mathcal{V}$. Next, define an equivalence relation on $\mathbb{X}(\bar{\mathcal{G}})$ where $\mathbf{x} \sim \phi(\mathbf{x})$ for all $\phi \in \text{Aut}(\bar{\mathcal{G}})$, $\mathbf{x} \in \mathbb{X}(\bar{\mathcal{G}})$, and take the quotient space $\mathbb{X}(\bar{\mathcal{G}})/\sim$, whose elements are the equivalence classes under this relation. Define the metric $\|\cdot\|_2$ on $\mathbb{X}(\bar{\mathcal{G}})/\sim$ so that for any $[\mathbf{x}], [\mathbf{y}] \in \mathbb{X}(\bar{\mathcal{G}})/\sim$,

$$\|[\mathbf{x}] - [\mathbf{y}]\|_2 = \min_{\substack{\mathbf{x}' \in [\mathbf{x}] \\ \mathbf{y}' \in [\mathbf{y}]}} \|\mathbf{x}' - \mathbf{y}'\|_2, \quad (40)$$

where $\|\mathbf{x}' - \mathbf{y}'\|_2$ is the norm on $\mathbb{X}(\bar{\mathcal{G}})$ inherited from identification with $\mathbb{R}^{|\mathcal{V}|}$. As $\text{Aut}(\bar{\mathcal{G}})$ is a finite group, this minimum is well-defined, and satisfies the triangle inequality, thus yielding a valid metric. When defining the space Ω_K in Section III-A we treat $\mathbb{X}(\bar{\mathcal{G}})$ with the structure of $\mathbb{X}(\bar{\mathcal{G}})/\sim$, from which the topology on Ω_K is inherited. Similarly, we use the metric $\|\cdot\|_2$ in Section V as defined on $\mathbb{X}(\bar{\mathcal{G}})/\sim$.

APPENDIX B

PROOF OF THEOREM 1

By Lemma 1, the sequence $\{\mu_j\}_{j=1}^\infty$ has compact support. Since $\bar{\mathbf{m}}_K$ is a continuous function, this implies that $\{\mathbb{E}_{\mu_j}[\bar{\mathbf{m}}_K]\}_{j=1}^\infty$ is a convergent sequence by the definition of weak convergence of probability measures. Applying Proposition 2, the moments $\mathbf{m}_K(\mathbf{x}_j) = \mathbb{E}_{\mu_j}[\bar{\mathbf{m}}_K]$ converge, as desired.

For the case where weak convergence holds for all $K \geq 0$, we complete the proof with the following lemma.

Lemma 2: Let $A = [0, a]$ be a compact subset of the real line, for some $a > 0$. Let $\{\{\mathbf{m}_K(j)\}_{K=0}^\infty\}_{j=1}^\infty$ be an array of numbers such that $\{\mathbf{m}_K(j)\}_{K=0}^\infty$ is the sequence of moments of a finite measure P_j supported on A , and that $\mathbf{m}_K(j) \xrightarrow{j} \mathbf{m}_K$ for each K , where $\{\mathbf{m}_K\}_{K=0}^\infty$ is some sequence of real numbers.

Then, the sequence $\{\mathbf{m}_K\}_{K=0}^\infty$ uniquely corresponds to a distribution P supported on A , such that the sequence of distributions P_j converges weakly to P .

Lemma 2 is a routine derivation following from the Hausdorff moment property [42, Theorem 3.15] and the Weierstrass approximation theorem [43], so we omit the proof. In particular, let $A = [0, 2D]$. The moments $\mathbf{m}_K(\mathbf{x}_j)$ correspond to the power spectral distributions of the graph signals \mathbf{x}_j , which are supported on A by the bounded degree assumption. Since $\mathbf{m}_K(\mathbf{x}_j)$ converges for all $K \geq 0$, we have that the sequence of limits $\{\mathbf{m}_K\}_{K=0}^\infty$ is the sequence of moments of a distribution P supported on A . Moreover, we have that P is the weak limit of the power spectral distributions of the graph signals \mathbf{x}_n , as desired.

APPENDIX C

PROOF OF LEMMA 1

Denote by $\bar{\mathbb{G}}_D$ the set of rooted graphs such that the underlying graph has node degree bounded by D . For any $K \geq 0$ and any graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ contained in $\bar{\mathbb{G}}_D$, note that for all $v \in \mathcal{V}$, it holds that the rooted K -ball centered at v is contained in $\bar{\mathbb{G}}_D$. Moreover, it is straightforward to show that there are only finitely many rooted K -balls contained in $\bar{\mathbb{G}}_D$, up to isomorphism. For any $a \geq 0$ and any rooted graph $\bar{\mathcal{G}}$, denote the set of graph signals taking values in $[-a, a]$ by $\mathbb{X}(\bar{\mathcal{G}}, a)$. It is clear that $\mathbb{X}(\bar{\mathcal{G}}, a)$ is compact as a subset of $\mathbb{X}(\bar{\mathcal{G}})$.

Using these constructions, it immediately follows that for any graph $\mathcal{G} \in \bar{\mathbb{G}}_D$ with graph signal $\mathbf{x} \in \mathbb{X}(\bar{\mathcal{G}}, a)$ for some $a > 0$, the associated distribution $\mu = (\Sigma_K)_*(\mathcal{G}, \mathbf{x})$ satisfies

$$\text{supp}(\mu) \subseteq \coprod_{\bar{\mathcal{G}}_k \in \bar{\mathbb{G}}_D: k \leq K} \mathbb{X}(\bar{\mathcal{G}}, a). \quad (41)$$

Since there are only finitely many rooted K -balls in $\bar{\mathbb{G}}_D$, the above set is a disjoint union of finitely many compact spaces, and is thus a compact space itself, by the properties of the disjoint union of topological spaces. If a family of graphs in $\{\mathcal{G}_j\}_{j \in J} \in \bar{\mathbb{G}}_D$ has uniformly bounded signals $\{\mathbf{x}_j\}_{j \in J}$ for some index set J , this means that there is some $a \geq 0$ such that $\mathbf{x}_j \in \mathbb{X}(\mathcal{G}_j, a)$ for all $j \in J$, so that the condition in (41) holds for all graph/graph signals in the family, as desired.

APPENDIX D

PROOF OF THEOREM 2

We first establish that for a function \bar{J} satisfying Assumption 1, the metric d_C preserves Lipschitz continuity over the subspace $\Gamma \subset \Omega_K$ as defined in Lemma 1.

Lemma 3: Let $\bar{J} : \Omega_K \rightarrow \mathbb{R}$ be a function satisfying Assumption 1. Then, for any $C > 0$, \bar{J} has a Lipschitz constant at most $\max\{L, A/C\}$ on the subspace Γ with the metric d_C , where

$$A = \sup_{\omega \in \Gamma} \bar{J}(\omega) - \inf_{\omega \in \Gamma} \bar{J}(\omega). \quad (42)$$

Lemma 3 can be shown by directly applying the definition of Lipschitz continuity, so we omit the proof. By the dual formulation of the 1-Wasserstein distance [36], Lemma 3 implies

that, for $C > 0$,

$$|\mathbb{E}_\mu[\bar{J}] - \mathbb{E}_\nu[\bar{J}]| \leq \max\left\{L, \frac{A}{C}\right\} \cdot W_1(\mu, \nu; C), \quad (43)$$

owing to the fact that $\text{supp}(\mu), \text{supp}(\nu) \subseteq \Gamma$, by assumption. Observing that $W_1(\mu, \nu; C)$ is monotonically increasing with respect to C , we can restrict our view to $C \geq A/L$. Applying a simple change of variables yields the expression

$$|\mathbb{E}_\mu[\bar{J}] - \mathbb{E}_\nu[\bar{J}]| \leq \inf_{C \in (0, 1]} \frac{L}{C} \cdot W_1\left(\mu, \nu; \frac{AC}{L}\right). \quad (44)$$

Taking $C = 1$ and noting that $A = 1$ by assumption yields the bound (23), as desired.

APPENDIX E PROOF OF THEOREM 3

One can check that $\bar{\mathbf{m}}_K$ is continuous and bounded on $\bigcup_{j=1}^{\infty} \text{supp}(\mu_j)$ due to the weighted degree bound D and the uniform boundedness of the signals, so that $\{\mathbb{E}_{\mu_j}[\bar{\mathbf{m}}_K]\}_{j=1}^{\infty}$ is a convergent sequence by the definition of weak convergence of probability measures. Applying Proposition 2, the moments $\mathbf{m}_K(\mathbf{x}_j) = \mathbb{E}_{\mu_j}[\bar{\mathbf{m}}_K]$ converge, as desired.

When weak convergence holds for all $K \geq 0$, the proof is completed by appealing to Lemma 2 as in the proof of Theorem 1.

APPENDIX F PROOF OF THEOREM 4

We begin by establishing two results regarding locally essentially bounded signals $\mathbb{X}_{\text{LB}}(\mathbf{G})$ and the graphing Laplacian \mathbf{S} .

Lemma 4: For a graphing $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \lambda)$, the set $\mathbb{X}_{\text{LB}}(\mathbf{G})$ is a subset of $\mathcal{L}^\infty(\mathcal{V}, \lambda)$, where $\mathcal{L}^\infty(\mathcal{V}, \lambda)$ denotes the space of essentially bounded functions on \mathcal{V} with respect to the probability measure λ .

We omit the proof of Lemma 4, as it follows directly from Definition 2.

Lemma 5: $\mathbb{X}_{\text{LB}}(\mathbf{G})$ is closed under application of the Laplacian \mathbf{S} , i.e., for any $\mathbf{x} \in \mathbb{X}_{\text{LB}}(\mathbf{G})$, we have that $\mathbf{S}\mathbf{x} \in \mathbb{X}_{\text{LB}}(\mathbf{G})$.

The proof of Lemma 5 can be found in Appendix G.

For any $K \geq 0$, the graphing signal $\mathbf{S}^K \mathbf{x}$ is contained in $\mathbb{X}_{\text{LB}}(\mathbf{G})$ by Lemma 5. The product of two functions in $\mathcal{L}^\infty(\mathcal{V}, \lambda)$ is contained in $\mathcal{L}^\infty(\mathcal{V}, \lambda)$, so that the graphing signal $\mathbf{y} : v \rightarrow [\mathbf{x}]_v \cdot [\mathbf{S}^K \mathbf{x}]_v$ is contained in $\mathcal{L}^\infty(\mathcal{V}, \lambda)$, by Lemma 4. Thus, $\mathbf{m}_K(\mathbf{x}) = \mathbb{E}_\lambda[\mathbf{y}]$ has finite value for all $K \geq 0$.

Let $K \geq 0$ be given, and let $\bar{\mathbf{m}}_K : \Omega_K \rightarrow \mathbb{R}$ be the function as defined in Proposition 2. One can check that $\mathbf{m}_K(\mathbf{x}) = \mathbb{E}_\mu[\bar{\mathbf{m}}_K]$, where $\mu = (\Sigma_K)_*(\mathbf{G}, \mathbf{x})$. By the bounded degree property of the graphing and the assumption that $\mathbf{x} \in \mathcal{L}^\infty(\mathcal{V}, \lambda)$ (as a consequence of Lemma 4), Lemma 1 implies that $\text{supp}(\mu)$ is contained in a compact subset of Ω_K . For each finite graph and graph signal $(\mathcal{G}_n, \mathbf{x}_n)$, denote the K^{th} moment of the signal \mathbf{x}_n by $\mathbf{m}_K(\mathbf{x}_n)$, and the pushforward measure of Σ_K on the space Ω_K by μ_n , so that $\mathbf{m}_K(\mathbf{x}_n) = \mathbb{E}_{\mu_n}[\bar{\mathbf{m}}_K]$, by Proposition 2. The assumption that $(\mathcal{G}_n, \mathbf{x}_n) \rightarrow (\mathbf{G}, \mathbf{x})$ implies that μ_n weakly converges to μ . Since μ has compact support and $\bar{\mathbf{m}}_K$ is a

continuous function, this implies that $\mathbb{E}_{\mu_n}[\bar{\mathbf{m}}_K] \rightarrow \mathbb{E}_\mu[\bar{\mathbf{m}}_K]$. In other words, $\mathbf{m}_K(\mathbf{x}_n) \rightarrow \mathbf{m}_K(\mathbf{x})$.

Having satisfied the conditions of Theorem 1, we have that the power spectral distributions of the finite graphs converge to a unique distribution supported on $[0, 2D]$, and that the moments of this distribution are given by $\{\mathbf{m}_K(\mathbf{x})\}_{K=0}^{\infty}$ by Lemma 2, as desired.

APPENDIX G PROOF OF LEMMA 5

Let a degree D graphing $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \lambda)$ with graphing signal $\mathbf{x} \in \mathbb{X}_{\text{LB}}(\mathbf{G})$ be given. By definition, there exists an $a \geq 0$ such that for all $K \geq 0$, we have

$$\lambda(\mathcal{N}^K((\mathbf{x}^{-1}[-a, a])^c)) = 0. \quad (45)$$

Put $b = 2Da$, and define the graphing signal $\mathbf{y} = \mathbf{S}\mathbf{x}$, where \mathbf{S} denotes the graphing Laplacian. Observe, due to the bounded degree D of the graphing, that

$$(\mathbf{y}^{-1}[-b, b])^c \subseteq \mathcal{N}\left((\mathbf{x}^{-1}[-a, a])^c\right), \quad (46)$$

so that, for all $K \geq 0$,

$$\mathcal{N}^K\left((\mathbf{y}^{-1}[-b, b])^c\right) \subseteq \mathcal{N}^{K+1}\left((\mathbf{x}^{-1}[-a, a])^c\right). \quad (47)$$

By the definition of local essential boundedness, the right hand side of (47) has measure zero (under the probability measure λ), which implies that the left hand side also has measure zero. Since K was arbitrary, this establishes that $\mathbf{y} \in \mathbb{X}_{\text{LB}}(\mathbf{G})$.

REFERENCES

- [1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [2] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.
- [3] F. Gama, Q. Li, E. Tolstaya, A. Prorok, and A. Ribeiro, "Synthesizing decentralized controllers with graph neural networks and imitation learning," *IEEE Trans. Signal Process.*, vol. 70, pp. 1932–1946, 2022.
- [4] W. Huang, T. A. W. Bolton, J. D. Medaglia, D. S. Bassett, A. Ribeiro, and D. Van De Ville, "A graph signal processing perspective on functional brain imaging," *Proc. IEEE*, vol. 106, no. 5, pp. 868–885, May 2018.
- [5] J. Ma, W. Huang, S. Segarra, and A. Ribeiro, "Diffusion filtering of graph signals and its use in recommendation systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016, pp. 4563–4567.
- [6] F. Gama, J. Bruna, and A. Ribeiro, "Stability properties of graph neural networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 5680–5695, 2020.
- [7] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.
- [8] C. N. Zouanas, S. Kim, K. N. Badhiwala, B. W. Avants, and J. T. Robinson, "Hydra vulgaris shows stable responses to thermal stimulation despite large changes in the number of neurons," *iScience*, vol. 24, no. 6, 2021, Art. no. 102490.
- [9] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, 2011.
- [10] D. I. Shuman, C. Wiesmeyer, N. Holighaus, and P. Vandergheynst, "Spectrum-adapted tight graph wavelet and vertex-frequency frames," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4223–4235, Aug. 2015.
- [11] T. M. Roddenberry, F. Frantzen, M. T. Schaumburg, and S. Segarra, "Hodgelets: Localized spectral representations of flows on simplicial complexes," in *Proc. IEEE Int. Conf. Acoust. Speech. Signal Process.*, 2022, pp. 5922–5926.

- [12] N. M. Kriege, F. D. Johansson, and C. Morris, "A survey on graph kernels," *Appl. Netw. Sci.*, vol. 5, no. 1, pp. 1–42, 2020.
- [13] N. Shervashidze, S. Vishwanathan, T. Petri, K. Mehlhorn, and K. Borgwardt, "Efficient graphlet kernels for large graph comparison," in *Proc. Int. Conf. Artif. Intell., Statist.*, 2009, pp. 488–495.
- [14] C. Morris, N. M. Kriege, K. Kersting, and P. Mutzel, "Faster kernels for graphs with continuous attributes via hashing," in *Proc. Int. Conf. Data Mining*, 2016, pp. 1095–1100.
- [15] M. Togninalli, E. Ghisu, F. Llinares-López, B. A. Rieck, and K. Borgwardt, "Wasserstein weisfeiler-lehman graph kernels," in *Proc. Conf. Neural Inf. Process. Syst.*, 2020, pp. 6439–6449.
- [16] P. de Haan, T. S. Cohen, and M. Welling, "Natural graph networks," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 3636–3646, 2020.
- [17] B. Bevilacqua et al., "Equivariant subgraph aggregation networks," in *Proc. Int. Conf. Learn. Representations*, 2022.
- [18] M. Zhang and P. Li, "Nested graph neural networks," in *Proc. Conf. Neural Inf. Process. Syst.*, 2021, pp. 15734–15747.
- [19] Q. Zhu, C. Yang, Y. Xu, H. Wang, C. Zhang, and J. Han, "Transfer learning of graph neural networks with ego-graph information maximization," in *Proc. Conf. Neural Inf. Process. Syst.*, 2021, pp. 1766–1779.
- [20] T. Maehara and H. NT, "Learning on random balls is sufficient for estimating (some) graph parameters," in *Proc. Conf. Neural Inform. Process. Syst.*, 2021, pp. 1126–1141.
- [21] J. Leskovec and J. Mcauley, "Learning to discover social circles in ego networks," in *Proc. Conf. Neural Inf. Process. Syst.*, 2012, pp. 548–556.
- [22] T. Tao, *An Introduction to Measure Theory*, Ann Arbor, MI, USA: Amer. Math. Soc., 2011.
- [23] J. M. Lee, *An Introduction to Smooth Manifolds*, Berlin, Germany: Springer, 2012.
- [24] F. Ji, W. P. Tay, and A. Ortega, "Graph signal processing over a probability space of shift operators," 2021, *arXiv:2108.09192*.
- [25] N. Shervashidze, P. Schweitzer, E. J. Van Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-lehman graph kernels," *J. Mach. Learn. Res.*, vol. 12, no. 9, pp. 2539–2561, 2011.
- [26] F. Gama and A. Ribeiro, "Ergodicity in stationary graph processes: A weak law of large numbers," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2761–2774, May 2019.
- [27] 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.
- [28] M. Avella-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. 520–537, Jan./Mar. 2018.
- [29] L. Ruiz, L. F. O. Chamon, and A. Ribeiro, "Graphon signal processing," *IEEE Trans. Signal Process.*, vol. 69, pp. 4961–4976, 2021.
- [30] C. Borgs, J. T. Chayes, H. Cohn, and N. Holden, "Sparse exchangeable graphs and their limits via graphon processes," *J. Mach. Learn. Res.*, vol. 18, no. 210, pp. 1–71, 2018. [Online]. Available: <http://jmlr.org/papers/v18/16-421.html>
- [31] S. Segarra and A. Ribeiro, "Stability and continuity of centrality measures in weighted graphs," *IEEE Trans. Signal Process.*, vol. 64, no. 3, pp. 543–555, Feb. 2016.
- [32] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [33] B. Bollobás, *Modern Graph Theory*, Berlin, Germany: Springer Sci. Bus. Media, 1998.
- [34] F. Gama, A. G. Marques, G. Mateos, and A. Ribeiro, "Rethinking sketching as sampling: A graph signal processing approach," *Signal Process.*, vol. 169, pp. 1–15, Apr. 2020.
- [35] L. De Branges, "The stone-weierstrass theorem," *Proc. Amer. Math. Soc.*, vol. 10, no. 5, pp. 822–824, 1959.
- [36] B. K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, and G. R. Lanckriet, "On integral probability metrics, ϕ -divergences and binary classification," 2009, *arXiv:0901.2698*.
- [37] L. Ruiz, F. Gama, and A. Ribeiro, "Graph neural networks: Architectures, stability, and transferability," *Proc. IEEE*, vol. 109, no. 5, pp. 660–682, May 2021.
- [38] F. Gama, E. Isufi, G. Leus, and A. Ribeiro, "Graphs, convolutions, and neural networks: From graph filters to graph neural networks," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 128–138, Nov. 2020.
- [39] L. Lovász, *Large Networks and Graph Limits*, Ann Arbor, MI, USA: Amer. Math. Soc., 2012.
- [40] E. M. Luks, "Isomorphism of graphs of bounded valence can be tested in polynomial time," *J. Comput. Syst. Sci.*, vol. 25, no. 1, pp. 42–65, 1982.
- [41] S. Rey, T. M. Roddenberry, S. Segarra, and A. G. Marques, "Enhanced graph-learning schemes driven by similar distributions of motifs," 2022, *arXiv:2207.04747*.
- [42] K. Schmüdgen, *The Moment Problem*, Berlin, Germany: Springer, 2017.
- [43] K. Weierstrass, "Über die analytische darstellbarkeit sogenannter willkürlicher functionen einer reellen veränderlichen," *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, vol. 2, pp. 633–639, 1885.