INVARIANT GRAPHON NETWORKS: APPROXIMATION AND CUT DISTANCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph limit models, like graphons for limits of dense graphs, have recently been used as a tool to study size transferability of graph neural networks (GNNs). While most existing literature focuses on message passing GNNs (MPNNs), we attend to *Invariant Graph Networks* (IGNs), a powerful alternative GNN architecture. In this work, we generalize IGNs to graphons, introducing *Invariant Graphon Networks (IWNs)* which are defined using a subset of the IGN basis corresponding to bounded linear operators. Even with this restricted basis, we show that IWNs of order $k + 1$ are at least as powerful as the k-dimensional Weisfeiler-Leman (WL) test for graphon-signals and we establish universal approximation results for graphon-signals in $\mathcal{L}^p$ distances using *signal-weighted homomorphism densities*. This significantly extends the prior work of Cai & Wang (2022), showing that IWNs—a subset of their *IGN-small*—retain effectively the same expressivity as the full IGN basis in the limit. In contrast to their approach, our blueprint of IWNs also aligns better with the geometry of graphon space, for example facilitating comparability to MPNNs. We also highlight that, unlike other GNN architectures such as MPNNs, IWNs are *discontinuous with respect to cut distance*, which causes their lack of convergence and is inherently tied to the definition of k-WL. Yet, their transferability remains comparable to MPNNs.

1 INTRODUCTION

Graph Neural Networks (GNNs) have emerged as a powerful tool for machine learning on complex graph-structured data, driving advances in fields like social network analysis (Fan et al., 2019), weather prediction (Lam et al., 2023) or materials discovery (Merchant et al., 2023), among others. Message Passing GNNs (MPNNs) (Gilmer et al., 2017; Kipf & Welling, 2017; Veličković et al., 2018; Xu et al., 2019), in which node features are iteratively updated by aggregating messages from neighboring nodes, are a popular architecture paradigm.

The question of *size transferability*—whether an MPNN generalizes to larger graphs than those in the training set—has recently gained attention. Unlike *extrapolation* (Xu et al., 2021; Yehudai et al., 2021; Jegelka, 2022), where generalization to arbitrary graph topologies is considered, size transferability typically assumes structural similarities between the training and evaluation graphs, such as them being sampled from the same random graph model (Keriven et al., 2021), topological space (Levie et al., 2021), or graph limit model (Ruiz et al., 2020; 2023; 2021b; Maskey et al., 2024; Le & Jegelka, 2024). For limits of dense graphs, *graphons* (Lovász & Szegedy, 2006; Lovász, 2012), which extend graphs to node sets on the unit interval and have been used to study extremal graph theory with analytic techniques, have also become a popular choice for studying transferability. In contrast to sparse graph limits (Lovász, 2012; Backhausz & Szegedy, 2022), they offer an established framework with powerful tools, such as embedding the set of all graphs into a compact space and having favorable spectral properties (Ruiz et al., 2021a). In such transferability analyses, a GNN is extended to a function of graphons, and regularity properties of the GNN are then used to bound the difference between outputs of the GNN applied to samples of different sizes from a graphon.

Existing works on transferability have been almost exclusively limited to MPNNs. However, their expressive power has been thoroughly studied and is constrained by the 1-*dimensional Weisfeiler-Leman (1-WL) graph isomorphism test*, also known as the *color refinement algorithm* (Xu et al., 2019; Morris et al., 2019). Hence, standard MPNNs even fail at straightforward tasks such as counting simple patterns like cycles of a specific length. This motivates to extend generalization analyses to more powerful architectures. Most prominent among these are higher-order extensions of MPNNs that are as powerful as the *k-dimensional Weisfeiler-Leman (k-WL) test*, $k > 1$, which iteratively colors k-tuples of nodes (Morris et al., 2019). *Invariant* and *Equivariant Graph Networks* (IGNs/EGNs) (Maron et al., 2018), on which we will mainly focus in this work, are another powerful choice, in which adjacency matrices and node signals are processed through higher-order tensor operations that maintain permutation equivariance. IGNs and EGNs universally approximate any permutation in-/equivariant graph function, and are at least as powerful as k-WL when its tensor orders are restricted to $k + 1$ (Maron et al., 2019; Keriven & Peyré, 2019; Azizian & Lelarge, 2021).

The expressive power of a GNN can also be judged via its *homomorphism expressivity*, i.e., its ability to count the number of homomorphisms from fixed graphs into the input graph. E.g., 1-WL corresponds to counting homomorphisms w.r.t. trees, and its higher-order extensions are related to counting homomorphisms w.r.t. graphs of bounded treewidth (Dvořák, 2010; Dell et al., 2018). In the graphon case, similar results exist for *homomorphism densities* (Böker et al., 2023; Böker, 2023).

In this work, we present an extension of IGNs to graphons, and study their expressivity and continuity properties. The closest related work by Cai & Wang (2022) investigates the convergence of IGNs to a limit graphon using a *partition norm*, which is a vector of norms over all diagonals of a graphon. They observe that convergence of IGNs applied to graphs sampled from a graphon is not always achievable. As a remedy, they propose a reduced model class *IGN-small*, which allows for convergence after estimating edge probabilities under certain regularity conditions. They also demonstrate that IGN-small retains sufficient expressiveness to approximate spectral GNNs. However, considering diagonals of graphons, which correspond to null sets in Lebesgue measure, is somewhat misaligned with the larger body of work in graphon theory, significantly limiting its applicability to their version of IGN limits. Furthermore, their expressivity analysis of IGN-small is rather constrained, especially considering that IGNs are typically universal GNN architectures.

Contributions. We extend IGNs to graphon-signals (Levie, 2023), i.e., node-attributed graphons, introducing *Invariant Graphon Networks (IWNs)*. In contrast to Cai & Wang (2022), we take the view of restricting linear equivariant layers to *bounded operators*, and, thus, our IWNs can be naturally analyzed using $\mathcal{L}^p$ and cut distances, enhancing comparability to the existing graphon literature.

Using only this reduced basis of IWNs, we show that IWNs up to order $k + 1$ are at least as powerful as a natural extension of the k-WL test (Böker, 2023) to graphon-signals. We also establish *universal approximation* results for graphon-signals (Levie, 2023). As IWNs are a subset of IGN-small, this significantly extends the work of Cai & Wang (2022), resolving the open questions posed in their conclusion: We show that the restriction to IGN-small comes at no cost in terms of expressivity, since IGN-small maintains the same expressive power as its discrete counterpart IGN. As a tool for our proofs, we use an extension of homomorphism densities to graphon-signals by signal weighting and show that these signal-weighted homomorphism densities inherit important topological properties from their equivalent in graphon space. We also highlight that IWNs are *discontinuous w.r.t. the cut distance* and only continuous in the finer topologies induced by $\mathcal{L}^p$ distances, which do not accurately represent our intuitive notion of graph similarity (Levie, 2023). This discontinuity is not unique to IWNs, but inherently tied to the way in which k-WL processes edge weights, and results in a large class of higher-order GNNs exhibiting the absence of convergence under sampling simple graphs as observed by Cai & Wang (2022). Yet, despite this discontinuity and the absence of a limit, it is still possible to obtain transferability results for IWNs which are similar to MPNNs.

To the best of our knowledge, this work is the first to extend the framework of IGNs to graphons in a way such that *continuity*, *expressivity*, and *transferability* can be studied and systematically compared to MPNNs, addressing the aforementioned shortcomings of Cai & Wang (2022). In summary, we make the following contributions:

- We define *signal-weighted homomorphism densities*, link them to a natural extension of the k-WL test to graphon-signals, and show how they capture graphon-signal topology.
- We introduce *Invariant Graphon Networks (IWNs)*, restricting linear equivariant layers to bounded operators. We show that $(k + 1)$-order IWNs are at least as powerful as k-WL for graphon-signals, and obtain universal approximation, extending Cai & Wang (2022).
- We point out the cut distance discontinuity of IWNs, its fundamental connection to k-WL, and demonstrate that IWNs are still transferable despite not converging.

2 BACKGROUND

In this section, we provide background on graphon theory, homomorphism expressivity, and the k-WL test for graphons, as well as on how to extend graphons to incorporate node signals. Contents of § 2.1 and § 2.2 are mostly drawn from Lovász (2012); Janson (2013); Zhao (2023), while in § 2.2 we also refer to Böker (2023). In § 2.3 we summarize key results of Levie (2023).

For $n \in \mathbb{N}$, write $[n] := \{1, \ldots, n\}$. Unless stated otherwise, a graph always refers to a *simple graph*, meaning an undirected graph $G = (V, E)$ with a finite node set $V(G) = V$ and edge set $E(G) = E \subseteq \binom{V}{2}$. Define also $v(G) := |V(G)|$, $e(G) := |E(G)|$. We will also consider *multigraphs*, for which the edges are a *multiset*. We consider graphs as a special case of multigraphs. Write λ^k for the k-dimensional Lebesgue measure; $\lambda := \lambda^1$. See also § A for a table of our most important notation used throughout the work.

2.1 GENERAL BACKGROUND ON GRAPHON THEORY

Graphons. Informally, a graphon can be seen as a graph with a continuous node set $[0, 1]$, and the adjacency matrix being represented by a function on the unit square. Intuitively, graphons can be obtained by taking the limit of adjacency matrices of *dense* graph sequences as the number of nodes grows. Formally, we first define a **kernel** as a bounded symmetric measurable function $W : [0, 1]^2 \to \mathbb{R}$. Write $\mathcal{W}$ for the space of all kernels. A **graphon** is a kernel mapping to $[0, 1]$. We define the **cut norm** of a kernel as

$$\|W\|_\square := \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} W \mathrm{d}\lambda^2 \right|, \tag{1}$$

where S, T are tacitly assumed measurable. Let $S_{[0,1]}$ be the set of measure preserving bijections of $[0, 1]$, i.e., $\varphi : [0, 1] \to [0, 1]$ such that $\lambda(\varphi^{-1}(A)) = \lambda(A)$ for all measurable $A \subseteq [0, 1]$. Write $\overline{S}_{[0,1]}$ for the set of *all* measure preserving functions and define $W^\varphi(x, y) := W(\varphi(x), \varphi(y))$ for $\varphi \in \overline{S}_{[0,1]}$. Since the specific ordering of the graphon values does not matter, we work with the **cut distance** (see also Lovász (2012, § 8.2)) between two graphons, defined as

$$\delta_\square(W, V) := \inf_{\varphi\in S_{[0,1]}} \|W - V^\varphi\|_\square = \min_{\varphi,\psi\in\overline{S}_{[0,1]}} \|W^\varphi - V^\psi\|_\square, \tag{2}$$

where the infimum is only guaranteed to be attained in the last expression. Analogously, we can define distances δ_p on graphons based on $\mathcal{L}^p$ norms. Note that $\delta_\square \le \delta_1 \le \delta_p$, where the first inequality follows from moving $|\cdot|$ into the integral in Eq. 2. Among $\{\delta_p\}_p$, the most commonly used is δ_1, which corresponds to the edit distance on graphs. We identify **weakly isomorphic** graphons of distance 0 to obtain the space $\widetilde{\mathcal{W}}_0$ of *unlabeled graphons*. The stricter concept of (strong) isomorphism, namely that the minimum in the second term of Eq. 2 is attained and zero, is less practical. The usefulness of $\delta_\square$ over any δ_p lies in the fact that $(\widetilde{\mathcal{W}}_0, \delta_\square)$ forms a *compact space* (Lovász, 2012, Theorem 9.23).

Discretization and sampling. Any labeled graph G can be identified with its *induced step graphon* $W_G := \sum_{j=1}^n \sum_{k=1}^n A_{jk} \mathbb{1}_{I_j\times I_k}$ for a regular partition $\{I_j\}_{j=1}^n$ of the unit interval, and finite graphs are dense in the graphon space (Zhao, 2023, Theorem 4.2.8). Graphons can also be seen as random graph models: Draw $\mathbf{X} \sim U(0, 1)^n$, and let $W(\boldsymbol{X})$ be a graph with edge weights $W_{ij} = W(X_i, X_j)$. If $e_{ij} \sim \mathrm{Bernoulli}(W_{ij})$ is further sampled, we obtain an unweighted graph $\mathbb{G}(W, \boldsymbol{X})$. Write $\mathbb{H}_n(W)$ and $\mathbb{G}_n(W)$ for the respective distributions. We have almost surely

$$\delta_\square(\mathbb{H}_n(W), W) \le \delta_1(\mathbb{H}_n(W), W) \to 0, \qquad n \to \infty \tag{3}$$

(Zhao, 2023, Lemma 4.9.4). Also

$$\delta_\square(\mathbb{G}_n(W), W) \to 0, \qquad n \to \infty \tag{4}$$

(Lovász, 2012, Proposition 11.32), but this does not hold for δ_1: Take, e.g., $W \equiv 1/2$, then $\delta_1(W, \mathbb{G}_n(W)) = 1/2$ for all $n \in \mathbb{N}$.

Homomorphism densities. Let $\hom(F, G)$ denote the number of homomorphisms from a graph F into a graph G. The corresponding **homomorphism density** is defined as $t(F, G) := \hom(F, G)/v(G)^{v(F)}$, i.e., the proportion of homomorphisms in all maps $V(F) \to V(G)$. We define the homomorphism density of a (multi)graph F with $V(F) = [k]$ to a graphon W by

$$t(F,W) := \int_{[0,1]^k} \prod_{\{i,j\}\in E(F)} W(x_i,x_j)\,\mathrm{d}\lambda^k(\boldsymbol{x}). \tag{5}$$

This generalizes the discrete concept in the sense that $t(F,G) = t(F,W_G)$. We also remark that $t(F,W) = \mathbb{P}(E(\mathbb{G}_k(W)) \supseteq E(F))$, regarding the graphs as labeled. Notably, for a sequence $(W_n)_n$ of graphons, $\delta_\square(W_n,W) \to 0$ if and only if $t(F,W_n) \to t(F,W)$ for all *simple* graphs F, and thus two graphons W,V are weakly isomorphic if and only if $t(F,W) = t(F,V)$ for all simple graphs F. Hence, homomorphism densities can also be seen as a counterpart of moments of a real random variable for W-random graphs, as they fix the distribution of $\mathbb{G}_n(W)$ similarly as the moments would for a sufficiently well-behaved real random variable (Zhao, 2023).

2.2 k-WL and Homomorphism Expressivity

In the discrete setting, the *1-dimensional Weisfeiler-Leman (1-WL) graph isomorphism test*, also known as *color refinement algorithm*, as well as its multidimensional extensions, are widely used to judge the expressive power of a GNN model. Alternatively, the model's **homomorphism expressivity**, i.e., its ability to count the number of homomorphisms from smaller graphs, called *patterns*, into the input graph, can be considered. Via the pattern's homomorphic images (Lovász, 2012, § 6.1), this is also closely related to counting subgraphs of the input graph (Chen et al., 2020; Tahmasebi et al., 2023; Jin et al., 2024). 1-WL expressivity corresponds precisely to distinguishing graphs for which the values of $\hom(F,\cdot)$ differ if F are trees. More generally, k-WL can be precisely characterized as being able to compute $\{\hom(F,\cdot)\}_F$, with F ranging over all simple graphs of treewidth bounded by k (Dvořák, 2010; Dell et al., 2018). See also § B.3 for more information on treewidth and the tree decomposition of a graph. A finer characterization for various MPNN architectural choices was recently shown by Zhang et al. (2024).

For graphons, the color refinement algorithm can be generalized via *distributions of iterated degree measures* (DIDMs) (Grebík & Rocha, 2022; Böker, 2023). Also, the notion of homomorphism expressivity can be naturally extended, simply by considering $t(F,\cdot)$ instead of $\hom(F,\cdot)$. We adopt the definition of k-WL from Böker (2023), who recently generalized the k-WL test to graphons, defining two graphons W,V as k-WL indistinguishable if $t(F,W) = t(F,V)$ for all *multigraphs* F of treewidth at most k.

2.3 Extension to Graphon-Signals

Most common GNNs take a graph-signal $(G,\boldsymbol{f})$ as inputs, i.e., a graph G with node set $[n] := \{1,\dots,n\}$ and a signal $\boldsymbol{f} \in \mathbb{R}^{n\times k}$, with k being the number of features. Levie (2023) extends this definition to graphons. They fix $r > 0$, consider signals in $\mathcal{L}_r^\infty[0,1] := \{f \in \mathcal{L}^\infty[0,1] \mid \|f\|_\infty \le r\}$, and set

$$\|f\|_\square := \sup_{S\subseteq[0,1]} \left|\int_S f\,\mathrm{d}\lambda\right|, \tag{6}$$

with S measurable. They then let $\mathcal{WL}_r := \mathcal{W}_0 \times \mathcal{L}_r^\infty[0,1]$ and define the *cut norm*

$$\|(W,f)\|_\square := \|W\|_\square + \|f\|_\square\,. \tag{7}$$

Define the distances $\delta_\square$ and δ_p, step graphon-signals, and sampling from graphon-signals analogously to the standard case. E.g., write $\mathbb{G}_n(W,f)$ for the distribution of $(\mathbb{G}(W,\boldsymbol{X}), f(\boldsymbol{X}))$, $\boldsymbol{X} \sim U(0,1)^n$. Also, identify weakly isomorphic graphon-signals of cut distance zero to obtain the space $\widetilde{\mathcal{WL}}_r$ of *unlabeled graphon-signals*. Central to their contribution, Levie (2023) proves compactness of the graphon-signal space and provides a bound on its covering number. This is then further used to derive a *sampling lemma*: Namely, for $(W,f) \in \mathcal{WL}_r$ and $n \in \mathbb{N}$, one has

$$\mathbb{E}\left[\delta_\square\big((W,f),\mathbb{G}_n(W,f)\big)\right] \le \frac{15}{\sqrt{\log n}}. \tag{8}$$

3 Signal-Weighted Homomorphism Densities

It is important to note that Böker et al. (2023); Böker (2023) focus exclusively on graphons and do not consider graphon-signals, and Levie (2023) does not introduce a notion of homomorphism densities for graphon-signals either. However, since most GNN architectures in the literature operate on node-featured graphs, we need a concept of homomorphism densities that reflects the properties of the graphon-signal space well. This could then, e.g., be applied to characterize the homomorphism expressivity of GNN models on graphon-signals, similar to the approach for graphons in § 2.2. As in Lovász (2012, § 5.2) for finite graphs, we introduce weighting by signals.

Definition 3.1 (Signal-weighted homomorphism density)**.** *Let F be a multigraph with $V(F) = [k]$, $\boldsymbol{d} \in \mathbb{N}_0^k$, and let $(W, f) \in \mathcal{WL}_r$. We set*

$$t(F, \boldsymbol{d}, (W, f)) := \int_{[0,1]^k} \left(\prod_{i \in V(F)} f(x_i)^{d_i} \right) \left(\prod_{\{i,j\} \in E(F)} W(x_i, x_j) \right) \mathrm{d}\lambda^k(\boldsymbol{x}), \tag{9}$$

calling the functions $t(F, \boldsymbol{d}, \cdot)$ ***signal-weighted homomorphism densities****.*

Note that setting $\boldsymbol{d} = \boldsymbol{0} \in \mathbb{N}_0^k$ recovers the graphon homomorphism densities $t(F, \boldsymbol{0}, (W, f)) = t(F, W)$. $\boldsymbol{d} \neq \boldsymbol{0}$ will allow us to consider moments of the signal, which could alternatively be seen as considering a multiset of *nodes*, similarly to homomorphism densities of multigraphs. This enables us to capture the distribution of the signal, coupled with the graph structure, which will be crucial for Eq. 9 to separate non-weakly isomorphic graphon-signals. In contrast to common approaches in the GNN literature, only considering $\boldsymbol{d} = \boldsymbol{1}$ does not suffice in our case, as this only distinguishes graphs under twin reduction. Restricting the exponents to be the same across all nodes as in Nguyen & Maehara (2020) results in $\{t(F, \boldsymbol{d}, \cdot)\}_{F,\boldsymbol{d}}$ not being closed under multiplication, which would later pose challenges when proving universality.

As a first step, we derive a counting lemma similar to the standard graphon case (Lovász, 2012, Lemma 10.23), which shows that signal-weighted homomorphism densities from *simple* graphs into a graphon-signal are Lipschitz continuous with respect to cut distance.

Lemma 3.2 (Counting lemma for graphon-signals)**.** *Let $(W, f), (V, g) \in \mathcal{WL}_r$ and F be a simple graph, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$. Then, writing $D := \sum_{i \in V(F)} d_i$,*

$$\left| t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g)) \right| \leq 2r^{D-1} \Big(2r \cdot e(F) \left\| W - V \right\|_\square + D \left\| f - g \right\|_\square \Big). \tag{10}$$

We prove the counting lemma in Appendix § C.1. As $t(F, \boldsymbol{d}, \cdot)$ is clearly invariant with respect to measure preserving functions acting on the graphon-signal, the bound of Eq. 10 can be easily extended to $\delta_\square$. The proof of Lemma 3.2 is straightforward, with the only detail requiring a little extra consideration being that, since the signals can take negative values, applying the same method as in the standard graphon case results in a bound using the alternative cut norm $\|\cdot\|_{\square,2}$ (Janson, 2013, Eq. (4.3), (4.4)), which is norm-equivalent to $\|\cdot\|_\square$. In a similar but even simpler way, a statement like Lemma 3.2 can also be shown for all *multigraphs* F using $\|\cdot\|_1$ instead.

However, the main justification for Definition 3.1 is Theorem 3.3 (in the style of Theorem 8.10 from Janson (2013)), as well as the following Corollary 3.4, demonstrating how signal-weighted homomorphism densities capture weak isomorphism and the topological structure of the graphon-signal space in a similar way as do homomorphism densities for graphons:

Theorem 3.3 (Characterizations of weak isomorphism for graphon-signals)**.** *Fix $r > 1$ and let $(W, f), (V, g) \in \mathcal{WL}_r$. Then, the following statements are equivalent:*

(1) $\delta_p((W, f), (V, g)) = 0$ for any $p \in [1, \infty)$;
(2) $\delta_\square((W, f), (V, g)) = 0$;
(3) $t(F, \boldsymbol{d}, (W, f)) = t(F, \boldsymbol{d}, (V, g))$ for all multigraphs F, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$;
(4) $t(F, \boldsymbol{d}, (W, f)) = t(F, \boldsymbol{d}, (V, g))$ for all simple graphs F, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$;
(5) $\mathbb{H}_k(W, f) \overset{\mathcal{D}}{=} \mathbb{H}_k(V, g)$ for all $k \in \mathbb{N}$;
(6) $\mathbb{G}_k(W, f) \overset{\mathcal{D}}{=} \mathbb{G}_k(V, g)$ for all $k \in \mathbb{N}$.

At first, the equivalence (1) $\Leftrightarrow$ (2) in Theorem 3.3, which we show by extending the argument of Lovász (2012, Theorem 8.13), reveals that any δ_p distance (for $p < \infty$) could be alternatively used to define weak isomorphism of two graphon-signals. Thus, any δ_p can also be seen as a metric on $\widetilde{\mathcal{WL}}_r$. The other equivalences show that weak isomorphism of two graphon-signals can be alternatively characterized by them having the same signal-weighted homomorphism densities, and the same random graph distributions. Specifically, $\{t(F, \boldsymbol{d}, \cdot)\}_{F,\boldsymbol{d}}$ fixes the distribution of (W, f)-random graphs similarly as do homomorphism densities for W-random graphs or moments for real-valued random variables. Note that statements (3) and (5) are naturally related to δ_1, while (4) and (6) are to $\delta_\square$. See § C.2 for a full proof of Theorem 3.3. We also remark that the condition $r > 1$ stems from the fact that we use the graphon-signal sampling lemma (Levie, 2023, Theorem 3.7).

The following corollary shows that signal-weighted homomorphism densities of *simple* graphs characterize cut distance convergence (see § C.3 for the proof):

Corollary 3.4 (Convergence in graphon-signal space)**.** *For $(W_n, f_n)_n$, $(W, f) \in \mathcal{WL}_r$ and $r > 1$,*

$$\delta_\square((W_n, f_n), (W, f)) \to 0 \quad \Leftrightarrow \quad t(F, \boldsymbol{d}, (W_n, f_n)) \to t(F, \boldsymbol{d}, (W, f)) \;\; \forall F, \boldsymbol{d} \in \mathbb{N}_0^{v(F)} \tag{11}$$

as $n \to \infty$, with F ranging over all simple graphs.

Finally, we show that signal-weighted homomorphism densities also make sense on the granularity level of the k-WL hierarchy, in the way that their indistinguishability is equivalent to the equality of a natural generalization of DIDMs defined in Böker (2023) to graphon-signals.

Theorem 3.5 (k-WL for graphon-signals, informal)**.** *Two graphon-signals (W, f) and (V, g) are k-WL indistinguishable if and only if $t(F, \boldsymbol{d}, (W, f)) = t(F, \boldsymbol{d}, (V, g))$ for all multigraphs F of treewidth $\leq k$, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$.*

Due to their technical nature, all details are deferred to § C.4. We note that while considering just simple graphs does not make a difference for weak isomorphism (see Theorem 3.3), this is not the case for k-WL, where considering multigraphs is strictly more restrictive (Böker, 2023, § 1.2).

4 Invariant Graphon Networks

In this section, we introduce IWNs. The central building blocks of IWNs will be *linear equivariant layers*, which we generalize from the original definition of Maron et al. (2018) to arbitrary measure spaces and $[0, 1]$ specifically in § 4.1, and derive their dimension as well as a basis. In § 4.2, we define (multilayer) IWNs and draw connections to Cai & Wang (2022).

4.1 Linear Equivariant Layers

We start with generalizing the building blocks of IGNs—namely, the linear equivariant layers. For IGNs, these are linear functions $L : \mathbb{R}^{n^k} \to \mathbb{R}^{n^\ell}$, such that L is equivariant with respect to all *permutations* acting on the n coordinates. We now extend this notion from the set $[n]$ to arbitrary measure spaces. The suitable generalization of permutations will be measure preserving maps:

Definition 4.1 (Linear equivariant layer)**.** *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, simply denoted by $\mathcal{X}$, and let $\overline{S}_\mathcal{X}$ be the set of measure-preserving functions $\varphi : \mathcal{X} \to \mathcal{X}$. Let $k, \ell \in \mathbb{N}_0$. Write $\mathcal{X}^k$ for $(\mathcal{X}^k, \mathcal{A}^{\otimes k}, \mu^{\otimes k})$ and note that $\mathcal{L}^2(\mathcal{X})^{\otimes k} \cong \mathcal{L}^2(\mathcal{X}^k)$. Define the **linear equivariant layers***

$$\mathrm{LE}_{k\to\ell}^{\mathcal{X}} := \left\{ L \in \mathcal{B}(\mathcal{L}^2(\mathcal{X}^k), \mathcal{L}^2(\mathcal{X}^\ell)) \,\middle|\, \forall \varphi \in \overline{S}_\mathcal{X} : L(U^\varphi) = L(U)^\varphi \text{ a.e.} \right\} \tag{12}$$

as the space of all bounded linear operators that are equivariant with respect to all measure preserving functions on $\mathcal{X}$, i.e., all relabelings of $\mathcal{X}$. Here, $U^\varphi(x_1, \ldots, x_k) := U(\varphi(x_1), \ldots, \varphi(x_k))$, and $\mathcal{B}(\cdot, \cdot)$ denotes bounded linear operators.

Note that if we consider $\mathcal{X} := [n]$ with a uniform probability measure (or counting measure), we obtain $\mathcal{L}^2([n]) \cong \mathbb{R}^n$, and $\mathrm{LE}_{k\to\ell}^{[n]}$ can be identified with the space of linear permutation equivariant functions $\mathbb{R}^{n^k} \to \mathbb{R}^{n^\ell}$, as measure preserving functions $[n] \to [n]$ are the permutations S_n. This yields precisely the linear equivariant layers that are building blocks of IGNs, which were studied by Maron et al. (2018). One of their main results is that $\dim \mathrm{LE}_{k\to\ell}^{[n]} = \mathrm{bell}(k+\ell)$, with $\mathrm{bell}(m)$ denoting the number of partitions Γ_m of $[m]$, which does *not* depend on n. They also describe a canonical basis in which every basis element $L_\gamma^{(n)} \in \mathrm{LE}_{k\to\ell}^{[n]}$ corresponds to a partition $\gamma \in \Gamma_{k+\ell}$, with basis elements consisting of simple operations such as extracting diagonals, summing/averaging over axes, and replication.

In our case of graphons, we are interested in $\mathrm{LE}_{k\to\ell} := \mathrm{LE}_{k\to\ell}^{[0,1]}$ as building blocks of IWNs, where $[0, 1]$ is equipped with its Borel σ-algebra and Lebesgue measure. The immediate question is what this space looks like in comparison to $\mathrm{LE}^{[n]}_{k\to\ell}$, i.e., what its dimension is and if there exists a canonical basis we can use to parameterize IWNs later on. It turns out that this space can be seen as just implementing a subset of the possibilities in the discrete setting, which is essentially a consequence of $[0,1]$ being atomless.

Theorem 4.2. *Let $k, \ell \in \mathbb{N}_0$. Then, $\mathrm{LE}_{k\to\ell}$ is a finite-dimensional vector space of dimension*

$$\dim \mathrm{LE}_{k\to\ell} \;=\; \sum_{s=0}^{\min\{k,\ell\}} s!\binom{k}{s}\binom{\ell}{s} \;\leq\; \mathrm{bell}(k+\ell). \tag{13}$$

Central to the argument is the observation that we can consider the action of any $L \in \mathrm{LE}_{k\to\ell}$ on step functions, and apply the characterization from Maron et al. (2018) to a sequence of nested subspaces, which fixes the operator on the entire space $\mathcal{L}^2[0,1]^k$. See § D.1 for the full proof.

A canonical basis of $\mathrm{LE}_{k\to\ell}$. The proof of Theorem 4.2 also provides insight into constructing a canonical basis of $\mathrm{LE}_{k\to\ell}$, which is indexed by the following subset of the partitions Γ_{k+l} of $[k+l]$:

$$\widetilde{\Gamma}_{k,\ell} \;:=\; \Big\{\gamma \in \Gamma_{k+\ell} \;\Big|\; \forall A \in \gamma : |A \cap [k]| \leq 1,\; |A \cap (k+[\ell])| \leq 1\Big\}. \tag{14}$$

For a partition $\gamma \in \widetilde{\Gamma}_{k,\ell}$, suppose that γ contains s sets of size 2 $\{i_1, j_1\}, \dots, \{i_s, j_s\}$ with $i_1, \dots, i_s \in [k]$, $j_1, \dots, j_s \in k + [\ell]$, and let $A = (i_1, \dots, i_s)$, $B = (j_1, \dots, j_s)$. Then, we can write the corresponding basis element $L_\gamma \in \mathrm{LE}_{k\to\ell}$ as

$$L_\gamma(U) \;:=\; \left[[0,1]^\ell \ni \boldsymbol{y} \;\mapsto\; \int_{[0,1]^{k-s}} U(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A})\,\mathrm{d}\lambda^{k-s}(\boldsymbol{x}_{[k]\setminus A})\bigg|_{\boldsymbol{x}_A = \boldsymbol{y}_B}\right]. \tag{15}$$

In comparison to the basis of Maron et al. (2018) (see also § B.1), this corresponds precisely to the basis elements for which no diagonals of the input are selected, and the output is always replicated on the entire space. We also note that the choice $p = 2$ in Definition 4.1 is somewhat arbitrary, and L_γ can indeed be seen as an operator $\mathcal{L}^p \to \mathcal{L}^p$ for any $p \in [1, \infty]$, with $\|L_\gamma\|_{p\to p} = 1$ (see § D.2).

Asymptotics of the dimensions. The dimension of $\mathrm{LE}_{k\to\ell}$ is asymptotically smaller compared to its discrete counterpart. In the case when one of the variables k, ℓ grows more slowly, the difference is tremendous. Assuming $k \to \infty$ but $\ell = \mathcal{O}(1)$, we obtain

$$\dim \mathrm{LE}_{k\to\ell} \;=\; \dim \mathrm{LE}_{\ell\to k} \;=\; \mathcal{O}(k^\ell), \tag{16}$$

while $\dim \mathrm{LE}^{[n]}_{k\to\ell} = \mathrm{bell}(k+\ell)$ grows superexponentially. In the worst case $k \sim \ell$, we have

$$\dim \mathrm{LE}_{k\to k} \geq k! \geq \mathrm{bell}(k), \tag{17}$$

however, the difference is still quadratic in the sense that $\dim \mathrm{LE}^{[n]}_{k\to k} \sim k^{2k}$ but $\dim \mathrm{LE}_{k\to k} \sim k^k$, up to exponential terms. See also § F for the first values for the dimensions and a brief discussion.

4.2 Definition of Invariant Graphon Networks

Using $\mathrm{LE}_{k\to\ell}$ as building blocks, we extend the definitions of IGNs from Maron et al. (2018) to graphons. This also corresponds to the definition given by Cai & Wang (2022), with the restriction that linear equivariant layers are limited to $\mathrm{LE}_{k\to\ell}$.

Definition 4.3 (In- and equivariant graphon networks)**.** *Let $\varrho : \mathbb{R} \to \mathbb{R}$ be Lipschitz continuous and non-polynomial. Let $S \in \mathbb{N}$, and for each $s \in \{0, \dots, S\}$, let $k_s \in \mathbb{N}_0$, $h_s \in \mathbb{N}$. Set $(h_0, k_0) := (2, 2)$. An **Equivariant Graphon Network (EWN)** is a function*

$$\mathcal{N} : \mathcal{WL}_r \ni (W, f) \;\mapsto\; \left(\boldsymbol{T}^{(S)} \circ \varrho \circ \cdots \circ \varrho \circ \boldsymbol{T}^{(1)}\right)(W, f), \tag{18}$$

where for each $s \in [S]$,

$$\boldsymbol{T}^{(s)} : \left(\mathcal{L}^2[0,1]^{k_{s-1}}\right)^{h_{s-1}} \to \left(\mathcal{L}^2[0,1]^{k_s}\right)^{h_s},\; \boldsymbol{U} \mapsto \boldsymbol{L}^{(s)}(\boldsymbol{U}) + \boldsymbol{b}^{(s)}, \tag{19}$$

*with $\boldsymbol{L}^{(s)} \in \left(\mathrm{LE}_{k_{s-1}\to k_s}\right)^{h_s \times h_{s-1}}$, $\boldsymbol{b}^{(s)} \in \mathbb{R}^{h_s}$, and $\boldsymbol{L}^{(s)}(\boldsymbol{U})_i := \sum_{j=1}^{h_{s-1}} L^{(s)}_{ij}(U_j)$ for $i \in [h_s]$. Here, the addition of the bias terms $\boldsymbol{b}^{(s)}$ and application of ϱ are understood elementwise. (W, f) is identified with $\left[(x,y) \mapsto (W(x,y), f(x))\right] \in \mathcal{L}^2\left([0,1]^2, \mathbb{R}^2\right) \cong \left(\mathcal{L}^2[0,1]^2\right)^2$ in the first layer. An **Invariant Graphon Network (IWN)** is an EWN with $(h_S, k_S) = (1, 0)$, i.e., mapping to scalars.*

We call $\max_{s\in[S]} k_s$ the **order** of an EWN $\mathcal{N}$ from Eq. 18, and k_s the orders of the individual layers. We write $\mathcal{IWN}_\varrho$ for the set of all IWNs with nonlinearity ϱ, and $\mathcal{IWN}_\varrho^k$ for the restriction of IWNs to order up to $k \in \mathbb{N}$.

Note that any IWN can be seen as a function $\mathcal{N} : \widetilde{\mathcal{WL}}_r \to \mathbb{R}$, as it is invariant with respect to all $\varphi \in \overline{S}_{[0,1]}$ by the definition of $\mathrm{LE}_{k\to\ell}$. It should also be remarked that in contrast to Maron et al. (2018); Cai & Wang (2022), we only consider real-valued (i.e., constant) bias terms, since $U^\varphi = U$ for all $\varphi \in \overline{S}_{[0,1]}$ only holds if $U \in \mathcal{L}^2[0,1]^\ell$ is constant almost everywhere, i.e., only constant functions are invariant with respect to all measure preserving maps (see Lemma D.1).

We also immediately observe that IWNs yield a parametrization that is closely related to IGN-small proposed by Cai & Wang (2022), a subset of IGNs with more favorable convergence properties under regularity assumptions on the graphon, see Cai & Wang (2022, Theorem 4). In their work, they define IGN-small as continuous IGNs (i.e., defined on graphons with signals) for which grid-sampling commutes with application of the discrete/continuous version of the IGN. For more details on IGN-small, also refer to § B.2.

Proposition 4.4. *Any IWN of the form Eq. 18 is an instance of IGN-small (Cai & Wang, 2022).*

The proof of Proposition 4.4 is an application of invariance under discretization (Lemma D.2) and representation stability of the basis elements. While the condition of IGN-small considers the entire multilayer neural network, we impose our boundedness condition on the individual linear equivariant layers. Therefore, a converse statement, i.e., that individual linear equivariant layers in IGN-small have to be in $\mathrm{LE}_{k\to l}$, does not hold, as there is a lot of ambiguity: For example, a graphon W could be embedded into a higher-dimension diagonal, and in a second step, the integral over this diagonal could be used as output of the continuous IGN. While in this case the *entire network as a whole* does fulfill the consistency requirement, the *individual layers* are not bounded linear operators. Even though IWNs use just a subset of the original IGN basis (see § B.1, § F), we will show in the following section that this suffices to prove strong expressivity results. By Proposition 4.4, these results will also hold for IGN-small.

5 Properties of Invariant Graphon Networks

We start this section by discussing the expressivity of IWNs in § 5.1, relating them to k-WL. In § 5.2, we then investigate the continuity and transferability of IWNs.

5.1 Expressivity of Invariant Graphon Networks

We prove expressivity results for IWNs, namely that IWNs up to order $k+1$ are at least as powerful as k-WL for graphons, and that they are universal approximators in the δ_p distances on any compact *subset* of graphon-signals. This will be done by assessing the signal-weighted homomorphism expressivity of IWNs defined in § 3. Clearly, IWNs are continuous in all δ_p distances:

Lemma 5.1. *Let $\mathcal{N} : \mathcal{WL}_r \to \mathbb{R}$ be an IWN as defined in Eq. 18. Then, $\mathcal{N}$ is Lipschitz continuous w.r.t. δ_p for each $p \in [1,\infty]$.*

See § E.1 for the proof. As a first step towards analyzing expressivity, we show that IWNs can approximate signal-weighted homomorphism densities with respect to graphs of size up to their order. For this, we take inspiration from Keriven & Peyré (2019), modeling the product in the homomorphism densities explicitly while tracking which linear equivariant layers are being used by the IWN. The final result then follows by using a tree decomposition of the graph:

Theorem 5.2 (Approximation of signal-weighted homomorphism densities)**.** *Let $r > 0$, $\varrho : \mathbb{R} \to \mathbb{R}$ Lipschitz continuous and non-polynomial, and F be a multigraph of treewidth $k \in \mathbb{N}$, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$. Fix $\varepsilon > 0$. Then there exists an IWN $\mathcal{N} \in \mathcal{IWN}_\varrho^{k+1}$ of order $k+1$ such that for all $(W,f) \in \mathcal{WL}_r$*

$$|t(F, \boldsymbol{d}, (W,f)) - \mathcal{N}(W,f)| \le \varepsilon. \tag{20}$$

The proof in § E.2 proceeds by induction on the tree decomposition of a graph. As we traverse the tree decomposition, we iteratively introduce IWN layers to incorporate the terms corresponding to newly encountered nodes, while marginalizing over the nodes that have already been processed. Note that as an immediate consequence we can see that IWNs are at least k-WL-expressive:

Corollary 5.3 (k-WL expressivity). *$\mathcal{IWN}_\varrho^{k+1}$ is at least as expressive as the k-WL test (Theorem 3.5) at distinguishing graphon-signals.*

Refer to § E.3 for a proof. As we know from Theorem 3.3 that two graphon-signals are weakly isomorphic if and only if $\{t(F, \boldsymbol{d}, \cdot)\}_{F,\boldsymbol{d}}$ agree for all simple graphs or multigraphs F, Theorem 5.2 gives us an immediate way to prove universal approximation when not restricting the tensor order.

Theorem 5.4 (δ_p-Universality of IWNs). *Let $r > 1$, $p \in [1, \infty)$, $\varrho : \mathbb{R} \to \mathbb{R}$ Lipschitz continuous and non-polynomial. For any compact $K \subset (\widetilde{\mathcal{WL}}_r, \delta_p)$, $\mathcal{IWN}_\varrho$ is dense in the continuous functions $C(K, \mathbb{R})$ w.r.t. $\|\cdot\|_\infty$.*

The proof in § E.4 is a straightforward application of the Stone-Weierstrass theorem: The span of the signal-weighted homomorphism densities forms a subalgebra that, by Theorem 3.3, is point separating. This result also crucially implies that IWNs are capable of distinguishing any two graphon-signals that are not weakly isomorphic. We also want to mention that while IWNs are continuous with respect to δ_∞, our proof of Theorem 3.3 does not extend to this case as $\|\cdot\|_\infty$ is not a smooth norm on $[0, 1]^2$ (see § C.2 for details).

5.2 Cut Distance and Transferability of Invariant Graphon Networks

In this section, we investigate the relation of IWNs to the cut distance and their transferability. As already mentioned, all nontrivial IWNs are discontinuous in the cut distance:

Proposition 5.5. *Let $\varrho : [0, 1] \to \mathbb{R}$. Then, the assignment $\mathcal{W}_0 \ni W \mapsto \varrho(W) \in \mathcal{W}$, where ϱ is applied pointwise, is continuous w.r.t. $\|\cdot\|_\square$ if and only if ϱ is linear.*

See § E.5 for a proof of Proposition 5.5. This discontinuity cannot be observed on finite graphs of any bounded size (as all norms are equivalent there), but, e.g., it appears when sampling finite simple graphs $\mathbb{G}_n(W, f)$ from a graphon-signal. In this case, by the sampling lemma (Eq. 8), for any cut distance continuous function $\mathcal{N}$ we would get $\mathbb{E}\,|\mathcal{N}(\mathbb{G}_n(W, f)) - \mathcal{N}(W, f)| \to 0$ as $n \to \infty$. However, IWNs do *not* display this convergence. This is essentially similar to the result of Cai & Wang (2022) on non-convergence under the "edge probability discrete model".

Crucially, we note that this effect is *not* specific to IWNs, but inherently linked to k-WL. The k-WL test for graphons (Böker, 2023) and graphon-signals (Theorem 3.5) considers *multigraph* homomorphism densities, which are discontinuous in the cut distance. As such, *any* k-WL expressive function defined on graphon-signals would exhibit this discontinuity. The consideration of multigraphs arises from a fundamental difference in how k-WL and 1-WL handle edges. For 1-WL, weighted edges are treated simply as *weights*, i.e., function values of a graphon only act through its shift operator and, thus, carry precisely the meaning of edge probabilities. In contrast, the k-WL test as well as IWNs capture the full *distribution* of these edge weights.

Often, one may analyze the convergence of a graph ML model $\mathcal{N}$ to an underlying limit to study the question of its *transferability*, i.e., if $\mathcal{N}(G_n, \boldsymbol{f}_n) \approx \mathcal{N}(G_m, \boldsymbol{f}_m)$ holds when $(G_n, \boldsymbol{f}_n), (G_m, \boldsymbol{f}_m) \sim \mathbb{G}_n(W, f), \mathbb{G}_m(W, f)$ as $n, m \in \mathbb{N}$ grows. For MPNNs that converge under their respective graph limits, transferability is usually shown simply by invoking the triangle inequality (see, e.g., Ruiz et al. (2023); Le & Jegelka (2024)), which is not possible for IWNs due to Proposition 5.5. Even worse, it is not even guaranteed that random graphs sampled from a graphon-signal become close in the δ_p norms as they grow in size: For example, take $G_n^{(1)}, G_n^{(2)}$ to be independent Erdős–Rényi graphs of size $n \in \mathbb{N}$. By combining Lavrov (2023) and Lovász (2012, Theorem 9.30), we obtain

$$\liminf_{n\to\infty} \mathbb{E}\left[\delta_1(G_n^{(1)}, G_n^{(2)})\right] \geq \tfrac{1}{12} \tag{21}$$

(note that the expectation in Eq. 21 would tend to zero in $\delta_\square$). As IWNs are universal on compact spaces w.r.t. δ_1 by Theorem 5.4, one might expect there to be an adversarial IWN which does not converge to the same value for all such random graphs. However, it turns out that we can "fix" the discontinuity of IWNs and show their transferability with similar worst-case rates as MPNNs.

Theorem 5.6 (Transferability of IWNs). *Let $\varepsilon > 0$ and $r > 1$. Let $\varrho : \mathbb{R} \to \mathbb{R}$ be continuous. Let $\mathcal{N} \in \mathcal{IWN}_\varrho$. Then, there exists a constant $C_{\varepsilon,\mathcal{N}} > 0$ such that for any $(W, f) \in \mathcal{WL}_r$ and $(G_n, \boldsymbol{f}_n), (G_m, \boldsymbol{f}_m) \sim \mathbb{G}_n(W, f), \mathbb{G}_m(W, f)$,*

$$\mathbb{E}\left|\mathcal{N}(G_n, \boldsymbol{f}_n) - \mathcal{N}(G_m, \boldsymbol{f}_m)\right| \leq C_{\varepsilon,\mathcal{N}}\left(\frac{1}{\sqrt{\log n}} + \frac{1}{\sqrt{\log m}}\right) + \varepsilon. \tag{22}$$

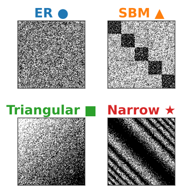
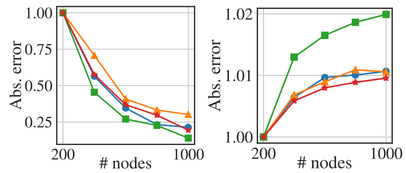
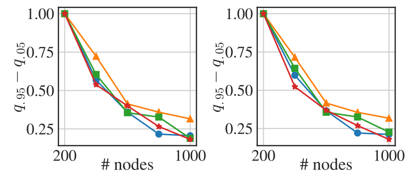


Figure 1: Example graphons W.



(a) MPNN (b) 2-IWN

Figure 2: Absolute output errors $\mathcal{N}(W)$ vs. $\mathcal{N}(\mathbb{G}_n(W))$.



(a) MPNN (b) 2-IWN

Figure 3: .90 pred. interval widths of $\mathcal{N}(\mathbb{G}_n(W))$.

See § E.6 for a proof. The proof relies on the fact that any IWN can be approximated arbitrarily well by a linear combination of signal-weighted homomorphism densities. In any such $t(F, \boldsymbol{d}, \cdot)$, the multigraph F can be replaced by its corresponding simple graph F^{simple}, resulting in a function of $(W, f) \in \mathcal{WL}_r$ which is cut distance continuous (see Lemma 3.2), while still yielding the same function values on 0-1-valued graphons, which includes all graphs $\mathbb{G}_n(W, f)$, $n \in \mathbb{N}$. Applying the standard sampling lemma Eq. 8 then results in the transferability. While the asymptotics of Theorem 5.6 are weak, they agree with the worst case for MPNNs (Levie, 2023), which is expected as we do not impose any additional assumptions on the graphon-signal or model.

We also validate our theoretical findings with a proof-of-concept experiment on the graphons from Figure 1. For *continuity/convergence*, we plot the absolute errors of the model outputs for the sampled simple graphs in comparison to their graphon limits in Figure 2. Due to the $\delta_\square$-continuity of MPNNs, their errors decrease as the graph size grows. For the IWN, however, this does not hold. Yet, the errors for the IWN stabilize with increasing sizes, suggesting that the outputs converge (just *not* to their graphon limit). For *transferability*, we further plot prediction interval widths of the output distributions on simple graphs for each of the sizes in Figure 3. Here, the widths contract for both models and there are only minor differences visible between the MPNN and the IWN. This validates Theorem 5.6 and suggests that IWNs have similar transferability properties as MPNNs, and δ_p-continuity suffices to achieve transferability. For more details, see § G.

6 Conclusion

In this work, we introduce Invariant Graphon Networks (IWNs) as an extension of Invariant Graph Networks (IGNs) to the graphon-signal space (Levie, 2023). By framing IWNs through bounded linear equivariant layers, we conduct a systematic analysis of expressivity, continuity, and transferability properties through $\mathcal{L}^p$ and cut distances on graphons. Significantly extending the results of Cai & Wang (2022), we demonstrate that IWNs, as a subset of their class IGN-small, retain the same expressive power as their discrete counterparts: IWNs up to tensor order $k+1$ are at least as expressive as the k-WL test for graphon-signals, and are universal approximators on graphon-signals with respect to the δ_p distance, $p \in [1, \infty)$. We also introduce signal-weighted homomorphism densities, an extension of the concept of homomorphism densities to graphon-signals, as a key tool.

We highlight that unlike MPNNs, IWNs are discontinuous with respect to cut distance, and therefore, standard size transferability arguments like Ruiz et al. (2023); Levie (2023); Le & Jegelka (2024) do not generalize. This discontinuity is inherently tied to the way edges are handled by k-WL and, as such, *all* k-WL expressive models on graphons will come with the same limitations. We demonstrate that, nevertheless, cut distance discontinuity can be overcome for transferability purposes, and IWNs are provably as transferable as MPNNs in the worst case under the graphon model.

An intriguing avenue for future research could be to investigate parametrizations based on the simple k-WL test (Böker, 2023) which are continuous in the cut distance, and compare expressivity and generalization to IWNs and standard higher-order GNNs (Morris et al., 2019). Another direction could be developing more quantitative bounds of Theorem 5.6 under additional assumptions on the underlying graphon, or analyzing expressive spectral methods (Lim et al., 2022; 2024; Huang et al., 2023) beyond spectral GNNs.

Reproducibility Statement

We provide rigorous proofs of all our statements in § B, § C, § D, and § E of the appendix, along with detailed explanations and the underlying assumptions. A comprehensive overview of our notation is listed in § A. Details for the toy experiments are provided in § G, and an anonymous version of the code is available under https://anonymous.4open.science/r/Higher-Order-WNNs-950F.

References

Waïss Azizian and Marc Lelarge. Expressive Power of Invariant and Equivariant Graph Neural Networks, June 2021. URL http://arxiv.org/abs/2006.15646. arXiv:2006.15646 [cs, stat].

Ágnes Backhausz and Balázs Szegedy. Action convergence of operators and graphs. *Canadian Journal of Mathematics*, 74(1):72–121, February 2022. ISSN 0008-414X, 1496-4279. doi: 10.4153/S0008414X2000070X. URL https://www.cambridge.org/core/journals/canadian-journal-of-mathematics/article/action-convergence-of-operators-and-graphs/F5900CA5BE554C9F4DEAAB518962D2DD.

Jan Böker. Weisfeiler-Leman Indistinguishability of Graphons. *The Electronic Journal of Combinatorics*, 30(4):P4.35, December 2023. ISSN 1077-8926. doi: 10.37236/10973. URL http://arxiv.org/abs/2112.09001. arXiv:2112.09001 [math].

Jan Böker, Ron Levie, Ningyuan Huang, Soledad Villar, and Christopher Morris. Fine-grained Expressivity of Graph Neural Networks. *Advances in Neural Information Processing Systems*, 36:46658–46700, December 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/9200d97ca2bf3a26db7b591844014f00-Abstract-Conference.html.

Chen Cai and Yusu Wang. Convergence of Invariant Graph Networks. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 2457–2484. PMLR, June 2022. URL https://proceedings.mlr.press/v162/cai22b.html. ISSN: 2640-3498.

Zhengdao Chen, Lei Chen, Soledad Villar, and Joan Bruna. Can graph neural networks count substructures? *Advances in neural information processing systems*, 33:10383–10395, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/75877cb75154206c4e65e76b88a12712-Abstract.html.

Holger Dell, Martin Grohe, and Gaurav Rattan. Lov\'asz Meets Weisfeiler and Leman, May 2018. URL http://arxiv.org/abs/1802.08876. arXiv:1802.08876 [cs, math].

Reinhard Diestel. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, Berlin, Heidelberg, 2017. ISBN 978-3-662-53621-6 978-3-662-53622-3. doi: 10.1007/978-3-662-53622-3. URL https://link.springer.com/10.1007/978-3-662-53622-3.

Zdeněk Dvořák. On recognizing graphs by numbers of homomorphisms. *Journal of Graph Theory*, 2010. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jgt.20461.

Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. Graph Neural Networks for Social Recommendation. In *The World Wide Web Conference*, WWW '19, pp. 417–426, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-6674-8. doi: 10.1145/3308558.3313488. URL https://doi.org/10.1145/3308558.3313488.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 1263–1272. PMLR, July 2017. URL https://proceedings.mlr.press/v70/gilmer17a.html. ISSN: 2640-3498.

Jan Grebík and Israel Rocha. Fractional Isomorphism of Graphons. *Combinatorica*, 42(3):365–404, June 2022. ISSN 0209-9683, 1439-6912. doi: 10.1007/s00493-021-4336-9. URL https://link.springer.com/10.1007/s00493-021-4336-9.

Jerzy Grunwald and Grzegorz Serafin. Explicit bounds for Bell numbers and their ratios, August 2024. URL https://arxiv.org/abs/2408.14182v1.

Yinan Huang, William Lu, Joshua Robinson, Yu Yang, Muhan Zhang, Stefanie Jegelka, and Pan Li. On the Stability of Expressive Positional Encodings for Graphs. October 2023. URL https://openreview.net/forum?id=xAqcJ9XoTf.

Svante Janson. Graphons, cut norm and distance, couplings and rearrangements. *NYJM Monographs*, (4), 2013. URL http://arxiv.org/abs/1009.2376. arXiv:1009.2376 [math].

Stefanie Jegelka. Theory of graph neural networks: Representation and learning. In *The International Congress of Mathematicians*, 2022. URL https://ems.press/content/book-chapter-files/33345.

Emily Jin, Michael Bronstein, İsmail İlkan Ceylan, and Matthias Lanzinger. Homomorphism Counts for Graph Neural Networks: All About That Basis, June 2024. URL http://arxiv.org/abs/2402.08595. arXiv:2402.08595 [cs].

Nicolas Keriven and Gabriel Peyré. Universal Invariant and Equivariant Graph Neural Networks. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/ea9268cb43f55d1d12380fb6ea5bf572-Abstract.html.

Nicolas Keriven, Alberto Bietti, and Samuel Vaiter. On the Universality of Graph Neural Networks on Large Random Graphs. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6960–6971. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/38181d991caac98be8fb2ecb8bd0f166-Abstract.html.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks, February 2017. URL http://arxiv.org/abs/1609.02907. arXiv:1609.02907 [cs, stat].

Donald E. Knuth. *The art of computer programming, volume 1 (3rd ed.): fundamental algorithms*. Addison Wesley Longman Publishing Co., Inc., USA, May 1997. ISBN 978-0-201-89683-1.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. GraphCast: Learning skillful medium-range global weather forecasting. *Science*, August 2023. doi: 10.48550/arXiv.2212.12794. URL http://arxiv.org/abs/2212.12794. arXiv:2212.12794 [physics].

Misha Lavrov. Expected graph edit distance between two random graphs, 2023. URL https://math.stackexchange.com/q/4824292(version:2023-12-11).

Thien Le and Stefanie Jegelka. Limits, approximation and size transferability for GNNs on sparse graphs via graphops. *Advances in Neural Information Processing Systems*, 36, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/8154c89c8d3612d39fd1ed6a20f4bab1-Abstract-Conference.html.

Ron Levie. A graphon-signal analysis of graph neural networks. *Advances in Neural Information Processing Systems*, 36:64482–64525, December 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/cb7943be26bb34f036c7e4068c490903-Abstract-Conference.html.

Ron Levie, Wei Huang, Lorenzo Bucci, Michael M. Bronstein, and Gitta Kutyniok. Transferability of Spectral Graph Convolutional Neural Networks, June 2021. URL http://arxiv.org/abs/1907.12972. arXiv:1907.12972 [cs, stat].

Derek Lim, Joshua David Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and Basis Invariant Networks for Spectral Graph Representation Learning. September 2022. URL https://openreview.net/forum?id=Q-UHqMorzil.

Derek Lim, Joshua Robinson, Stefanie Jegelka, and Haggai Maron. Expressive sign equivariant networks for spectral geometric learning. *Advances in Neural Information Processing Systems*, 36, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/3516aa3393f0279e04c099f724664f99-Abstract-Conference.html.

László Lovász. *Large Networks and Graph Limits*. American Mathematical Soc., 2012. ISBN 978-0-8218-9085-1. Google-Books-ID: FsFqHLid8sAC.

László Lovász and Balázs Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, November 2006. ISSN 0095-8956. doi: 10.1016/j.jctb.2006.05.002. URL https://www.sciencedirect.com/science/article/pii/S0095895606000517.

Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and Equivariant Graph Networks. September 2018. URL https://openreview.net/forum?id=Syx72jC9tm.

Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the Universality of Invariant Networks. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 4363–4371. PMLR, May 2019. URL https://proceedings.mlr.press/v97/maron19a.html. ISSN: 2640-3498.

Sohir Maskey, Gitta Kutyniok, and Ron Levie. Generalization Bounds for Message Passing Networks on Mixture of Graphons, April 2024. URL http://arxiv.org/abs/2404.03473. arXiv:2404.03473 [cs].

Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, December 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06735-9. URL https://www.nature.com/articles/s41586-023-06735-9. Publisher: Nature Publishing Group.

Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 4602–4609, 2019. URL https://ojs.aaai.org/index.php/AAAI/article/view/4384. Issue: 01.

Hoang Nguyen and Takanori Maehara. Graph Homomorphism Convolution. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 7306–7316. PMLR, November 2020. URL https://proceedings.mlr.press/v119/nguyen20c.html. ISSN: 2640-3498.

Allan Pinkus. Approximation theory of the MLP model in neural networks. *Acta Numerica*, 8: 143–195, January 1999. ISSN 1474-0508, 0962-4929. doi: 10.1017/S0962492900002919. URL https://www.cambridge.org/core/journals/acta-numerica/article/approximation-theory-of-the-mlp-model-in-neural-networks/18072C558C8410C4F92A82BCC8FC8CF9.

Walter Rudin. *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics. 3 edition, 1976. ISBN 0-07-054235-X.

Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro. Graphon Neural Networks and the Transferability of Graph Neural Networks, October 2020. URL http://arxiv.org/abs/2006.03548. arXiv:2006.03548 [cs, stat].

Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro. Graphon Signal Processing. *IEEE Transactions on Signal Processing*, 69:4961–4976, 2021a. ISSN 1053-587X, 1941-0476. doi: 10.1109/TSP.2021.3106857. URL http://arxiv.org/abs/2003.05030. arXiv:2003.05030 [eess].

Luana Ruiz, Fernando Gama, and Alejandro Ribeiro. Graph Neural Networks: Architectures, Stability and Transferability, January 2021b. URL http://arxiv.org/abs/2008.01767. arXiv:2008.01767 [cs, stat].

Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro. Transferability Properties of Graph Neural Networks. *IEEE Transactions on Signal Processing*, 71:3474–3489, 2023. ISSN 1941-0476. doi: 10.1109/TSP.2023.3297848. URL https://ieeexplore.ieee.org/abstract/document/10190182. Conference Name: IEEE Transactions on Signal Processing.

Konrad Schmüdgen. *The Moment Problem*, volume 277 of *Graduate Texts in Mathematics*. Springer International Publishing, Cham, 2017. ISBN 978-3-319-64545-2 978-3-319-64546-9. doi: 10.1007/978-3-319-64546-9. URL http://link.springer.com/10.1007/978-3-319-64546-9.

Behrooz Tahmasebi, Derek Lim, and Stefanie Jegelka. The Power of Recursion in Graph Neural Networks for Counting Substructures. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, pp. 11023–11042. PMLR, April 2023. URL https://proceedings.mlr.press/v206/tahmasebi23a.html. ISSN: 2640-3498.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks, February 2018. URL http://arxiv.org/abs/1710.10903. arXiv:1710.10903 [cs, stat].

Eric W. Weisstein. Bell Number. URL https://mathworld.wolfram.com/. Publisher: Wolfram Research, Inc.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How Powerful are Graph Neural Networks?, February 2019. URL http://arxiv.org/abs/1810.00826. arXiv:1810.00826 [cs, stat].

Keyulu Xu, Mozhi Zhang, Jingling Li, Simon S. Du, Ken-ichi Kawarabayashi, and Stefanie Jegelka. How Neural Networks Extrapolate: From Feedforward to Graph Neural Networks, March 2021. URL http://arxiv.org/abs/2009.11848. arXiv:2009.11848 [cs, stat].

Gilad Yehudai, Ethan Fetaya, Eli Meirom, Gal Chechik, and Haggai Maron. From Local Structures to Size Generalization in Graph Neural Networks. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 11975–11986. PMLR, July 2021. URL https://proceedings.mlr.press/v139/yehudai21a.html. ISSN: 2640-3498.

Bohang Zhang, Jingchu Gai, Yiheng Du, Qiwei Ye, Di He, and Liwei Wang. Beyond Weisfeiler-Lehman: A Quantitative Framework for GNN Expressiveness, January 2024. URL http://arxiv.org/abs/2401.08514. arXiv:2401.08514 [cs, math].

Yufei Zhao. *Graph Theory and Additive Combinatorics: Exploring Structure and Randomness*. Cambridge University Press, Cambridge, 2023. ISBN 978-1-00-931094-9. doi: 10.1017/9781009310956. URL https://www.cambridge.org/core/books/graph-theory-and-additive-combinatorics/90A4FA3C584FA93E984517D80C7D34CA.

A Notation

Table 1: We list the most important symbols used in this work.

$\mathbb{N}$; $\mathbb{N}_0$; $\mathbb{Q}$; $\mathbb{R}$	Natural, non-negative integer, rational, real numbers.
$[n]$	Set $\{1, \ldots, n\}$ for $n \in \mathbb{N}$.
$\mathbb{1}_A$	Indicator function of a set A.
$V(G)$; $v(G)$	Node set of a graph; number of nodes of a graph G.
$E(G)$; $e(G)$	Edge (multi)set of a (multi)graph; number of edges of a (multi)graph G.
$\mathcal{O}(\cdot)$	"Big-O" notation for asymptotic growth of a function.
$\overline{A}$	Closure of a subset A of a topological space $\mathcal{X}$.
$\mathcal{B}(\mathcal{X})$	Borel σ-algebra of a topological space $\mathcal{X}$.
$\sigma(\cdot)$	Generated σ-algebra.
$\mathbb{P}$	Probability measure.
$\mathbb{E}$	Expected value.
λ; λ^k	1-dimensional Lebesgue measure; k-dimensional Lebesgue measure.
$\mathcal{L}^p(\mathcal{X})$	Space of p-integrable functions on a measure space $\mathcal{X}$, for $p \in [1, \infty]$.
$\mathcal{L}^p_r(\mathcal{X})$	Space of p-integrable functions, with norm bounded by r.
$\\|\cdot\\|_\square$	Cut norm.
$\\|\cdot\\|_p$	$\mathcal{L}^p$ norm of functions on a measure space, for $p \in [1, \infty]$.
$\\|\cdot\\|_{p,\mathcal{X}}$	$\mathcal{L}^p$ norm, with emphasis on the underlying space $\mathcal{X}$.
$\mathcal{B}(V_1, V_2)$	Space of bounded linear operators from normed vector space V_1 to V_2.
$\\|L\\|_{p \to q}$	Operator norm of $L \in \mathcal{B}(\mathcal{L}^p(\mathcal{X}), \mathcal{L}^q(\mathcal{Y}))$.
$C(K, \mathbb{R})$	Space of continuous functions from compact topological space K into $\mathbb{R}$, with uniform norm $\\|\cdot\\|_\infty$.
$\mathcal{W}$	Space of kernels.
$\mathcal{W}_0$	Space of graphons.
$\mathcal{WL}_r$	Space of graphon-signals $\mathcal{W}_0 \times \mathcal{L}^\infty_r[0, 1]$.
$\widetilde{\mathcal{W}_0}$	Space of unlabeled graphons.
$\widetilde{\mathcal{WL}_r}$	Space of unlabeled graphon-signals.
$S_{[0,1]}$	Measure preserving bijections of $[0, 1]$.
$S'_{[0,1]}$	Measure preserving bijections between co-null subsets of $[0, 1]$.
$\overline{S}_{\mathcal{X}}$	Measure preserving functions $\mathcal{X} \to \mathcal{X}$, for a measure space $\mathcal{X}$.
$\delta_\square$	Cut distance.
δ_p	$\mathcal{L}^p$ distance for graphons/kernels.
W_G	Step graphon of a graph G.
$\mathbb{H}_k(W)$; $\mathbb{H}_k(W, f)$	Distribution of weighted graphs/graph-signals of size k sampled from a graphon W/graphon-signal (W, f).
$\mathbb{G}_k(W)$; $\mathbb{G}_k(W, f)$	Distribution of unweighted graphs/graph-signals of size k sampled from a graphon W/graphon-signal (W, f).
$U(0, 1)$	Uniform distribution on the interval $[0, 1]$.
$\hom(F, G)$	Number of homomorphisms from graph F to G.
$t(F, W)$	Homomorphism density from a (multi)graph F into graphon W.
$t(F, \boldsymbol{d}, (W, f))$	Signal-weighted homomorphism density from a (multi)graph F, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$, into graphon-signal (W, f).
$t_{\boldsymbol{x}_A}(F, \boldsymbol{d}, (W, f))$	Signal-weighted homomorphism density from a partially labeled (multi)graph F, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$, $A \subseteq V(F)$, into graphon-signal (W, f).
$\mathrm{LE}^{\mathcal{X}}_{k \to \ell}$	Linear equivariant layers on measure space $\mathcal{X}$.
$\mathrm{LE}_{k \to \ell}$	Linear equivariant layers on $[0, 1]$.
$\mathcal{F}^{(n)}_k$	Regular step functions in $\mathcal{L}^2[0, 1]^k$ at resolution n.
Γ_m	Set of partitions of $[m]$, $m \in \mathbb{N}_0$.
$\mathrm{bell}(m)$	$\lvert\Gamma_m\rvert$, i.e., number of partitions of $[m]$.
$\widetilde{\Gamma}_{k,\ell}$	Set of partitions of $[k + l]$ that index a basis of $\mathrm{LE}_{k \to \ell}$.
$\mathcal{IWN}_\varrho$	Set of IWNs with nonlinearity ϱ.
$\mathcal{IWN}^k_\varrho$	Set of IWNs with nonlinearity ϱ up to order k.

B Extended Background

B.1 Characterization of the IGN Basis

In this section, we restate the characterization of the IGN basis introduced by Cai & Wang (2022). As described in the original IGN paper by Maron et al. (2018), $\dim \mathrm{LE}_{k\to\ell}^{[n]} = \mathrm{bell}(k+\ell)$, i.e., the number of partitions $\Gamma_{k+\ell}$ of the set $[k+\ell]$. In the basis of Cai & Wang (2022), each basis element $L_\gamma^{(n)}$ associated with a partition $\gamma \in \Gamma_{k+\ell}$ can be characterized as a sequence of basic operations.

First, divide γ into 3 subsets $\gamma_1 := \{A \in \gamma \,|\, A \subseteq [k]\}$, $\gamma_2 := \{A \in \gamma \,|\, A \subseteq k + [\ell]\}$, $\gamma_3 := \gamma \setminus (\gamma_1 \cup \gamma_2)$. Here, the numbers $1, \dots, k$ are associated with the input axes and $k+1, \dots, k+\ell$ with the output axes respectively.

① (**Selection**: $U \mapsto U_\gamma$). In a first step, we specify which part of the input tensor $U \in \mathbb{R}^{n^k}$ is under consideration. Take $\gamma\big|_{[k]} := \{A \cap [k] \,|\, A \in \gamma, A \cap [k] \neq 0\}$ and construct a new $|\gamma_1| + |\gamma_2| = |\gamma\big|_{[k]}|$-tensor U_γ by selecting the diagonal of the k-tensor U corresponding with the partition $\gamma\big|_{[k]}$.

② (**Reduction**: $U_\gamma \mapsto U_{\gamma,\mathbf{red}}$). We average U_γ over the axes $\gamma_1 \subseteq \gamma\big|_{[k]}$, resulting in a tensor $U_{\gamma,\mathbf{red}}$ of order $|\gamma_2|$, indexed by $\gamma_2\big|_{[k]}$.

③ (**Alignment**: $U_{\gamma,\mathbf{red}} \mapsto U_{\gamma,\mathbf{align}}$). We align $U_{\gamma,\mathbf{red}}$ with a $|\gamma_2|$-tensor $U_{\gamma,\mathbf{align}}$ indexed by $\gamma_\ell\big|_{k+[\ell]}$, sending for $A \in \gamma_2$ the axis $A \cap [k]$ to $A \cap [\ell]$.

④ (**Replication**: $U_{\gamma,\mathbf{align}} \mapsto U_{\gamma,\mathbf{rep}}$). Replicate the $|\gamma_2|$-tensor $U_{\gamma,\mathbf{align}}$ indexed by $\gamma_2\big|_{k+[\ell]}$ along the axes in γ_3. Note that if $\gamma_2\big|_{k+[\ell]} \cup \gamma_3$ contains non-singleton sets, the output tensor is supported on some diagonal.

The basis element $L_\gamma^{(n)} : \mathbb{R}^{n^k} \to \mathbb{R}^{n^\ell}$ can now be described by the assignment $L_\gamma^{(n)}(U) := U_{\gamma,\mathbf{rep}}$.

B.2 IGN-small (Cai & Wang, 2022)

Cai & Wang (2022) study the convergence of discrete IGNs applied to graphs sampled from a graphon to a continuous version of the IGN defined on graphons. For this, they use the full IGN basis and a *partition norm*, which is for $W \in \mathcal{L}^2[0,1]^k$ a $\mathrm{bell}(k)$-dimensional vector consisting of $\mathcal{L}^2$ norms of W on all possible diagonals. While they show that convergence of a discrete IGN on weighted graphs sampled from a graphon to its continuous counterpart holds, they also demonstrate that this is not the case for unweighted graphs with $\{0,1\}$-valued adjacency matrix (which also follows from our observation on continuity, see Lemma 5.1).

As a remedy, Cai & Wang (2022) constrain the IGN space to *IGN-small*, which consists of IGNs for which applying the discrete version to a grid-sampled step graphon yields the same output as applying the continuous version and grid-sampling afterwards.

In the following, we will formalize this. For $n \in \mathbb{N}$, let $I_j^{(n)} := [\frac{j-1}{n}, \frac{j}{n})$ for $j \in [n-1]$ and $I_n^{(n)} := [\frac{n-1}{n}, 1]$ be a partition of $[0,1]$ into regular intervals. Let $\mathcal{A}_n := \sigma\left(\{I_1^{(n)}, \dots, I_n^{(n)}\}\right)$ denote the σ-algebra generated by this partition and let

$$\mathcal{F}_k^{(n)} := \{W \in \mathcal{L}^2[0,1]^k \,|\, W \text{ is } \mathcal{A}_n^{\otimes k}\text{-measurable}\} \tag{23}$$

be the regular *k-dimensional step functions* on $[0,1]$, $k \in \mathbb{N}_0$.

Definition B.1 (IGN-small). *Let $\mathcal{N}$ be defined as in Definition 4.3, with the only difference that $\mathrm{LE}_{k\to\ell}$ is replaced by the full IGN basis (see § B.1), where averaging steps should be understood as integration. Cai & Wang (2022) call such $\mathcal{N}$ a* continuous IGN. *For any basis element L_γ, $\gamma \in \Gamma_{k+\ell}$, denote its discrete version at resolution $n \in \mathbb{N}$ by $L_\gamma^{(n)}$ and the network obtained by discretizing all equivariant linear layers by $\mathcal{N}^{(n)}$. Let*

$$S^{(n)} : \mathbb{R}^{[0,1]^k \times h} \to \mathbb{R}^{n^k \times h},\ \boldsymbol{U} \mapsto (U(i/n, j/n))_{i,j=1}^n \tag{24}$$

*be the grid-sampling operator. Then, $\mathcal{N}$ is contained in **IGN-small** if*

$$(S^{(n)} \circ \mathcal{N})(W, f) \;=\; (\mathcal{N}^{(n)} \circ S^{(n)})(W, f) \tag{25}$$

for any $(W, f) \in \mathcal{WL}_r$ such that $W \in \mathcal{F}_2^{(n)}$, $f \in \mathcal{F}_1^{(n)}$.

In this case, for a graphon-signal $(W, f) \in \mathcal{WL}_r$, the input to such an IGN is $(x_1, x_2) \mapsto (W(x_1, x_2), \mathbb{1}\{x_1 = x_2\}f(x_1))$.

Cai & Wang (2022) demonstrate that the convergence of IGN-small in a model, where a $\{0, 1\}$-valued adjacency matrix is sampled from the graphon, is achievable under certain restrictive assumptions on the graphon and signal, such as Lipschitz continuity, along with the prior estimation of an edge probability (refer to Theorem 4). Regarding the expressivity of IGN-small, they establish that this model class can approximate spectral GNNs with arbitrary precision (see Theorem 5).

B.3 Tree Decomposition and Treewidth

In this section, we will recall the tree decomposition of a graph and the related notion of *treewidth*, which essentially captures how "far" a graph is from being a tree. See for example Diestel (2017, § 12.3) for a more in-depth discussion of this fundamental graph theoretic concept. We use the specific notation of Böker (2023).

Definition B.2 (Tree Decomposition of a Graph)**.** *Let G be a graph. A **tree decomposition** of G is a pair (T, β), where T is a tree and $\beta : V(T) \to 2^{V(G)}$ such that*

(1) for every $v \in V(G)$, the set $\{t \mid v \in \beta(t)\}$ is nonempty and connected in T,
(2) for every $e \in E(G)$, there is a node $t \in V(T)$ such that $e \subseteq \beta(t)$.

For $t \in V(T)$, the sets $\beta(t) \subseteq V(G)$ are commonly referred to as *bags* of the tree decomposition. Note that every graph G has a trivial tree decomposition, given by a tree consisting of one node, with the bag being the entire node set $V(G)$. However, we are generally interested in finding tree decompositions with smaller bags. This leads us to the concept of *treewidth*:

Definition B.3 (Treewidth of a Graph)**.** *Let G be a graph. For any tree decomposition (T, β) of G, define its **width** as*

$$\max\{|\beta(t)| \mid t \in V(T)\} - 1. \tag{26}$$

*The **treewidth** of a graph G is then the minimum width of all tree decompositions of G.*

Note that, the edge graph of a tree G can be seen as a tree decomposition of G, with each edge being a bag. Hence, the treewidth of a tree is 2. It can also be shown that, e.g., the treewidth of a circle of size at least 3 is 2.

As remarked by Böker (2023), while treewidth is commonly defined just for (simple) graphs, we extend the definition to multigraphs by simply ignoring the edge multiplicities, i.e., considering the *set* of edges instead of the multiset.

C Signal-Weighted Homomorphism Densities

In this appendix, we collect proofs of Lemma 3.2,Theorem 3.3, and Corollary 3.4, and provide further background and explanations regarding our extension of homomorphism densities to graphon-signal space from § 3.

C.1 Proof of Lemma 3.2

Lemma 3.2 (Counting lemma for graphon-signals)**.** *Let $(W, f), (V, g) \in \mathcal{WL}_r$ and F be a simple graph, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$. Then, writing $D := \sum_{i \in V(F)} d_i$,*

$$\left|t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g))\right| \;\le\; 2r^{D-1}\Big(2r \cdot e(F) \left\|W - V\right\|_\square + D \left\|f - g\right\|_\square\Big). \tag{10}$$

Proof. We split the l.h.s. into two parts, bounding the difference of the graphons and the signals seperately:

$$\left|t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g))\right| \le \tag{27}$$

$$\underbrace{\left|\int_{[0,1]^k} \left(\prod_{i \in V(F)} f(x_i)^{d_i}\right)\left(\prod_{\{i,j\} \in E(F)} W(x_i, x_j) - \prod_{\{i,j\} \in E(F)} V(x_i, x_j)\right) \mathrm{d}\lambda^k(\boldsymbol{x})\right|}_{①} \tag{28}$$

$$+ \underbrace{\left|\int_{[0,1]^k} \left(\prod_{\{i,j\} \in E(F)} V(x_i, x_j)\right)\left(\prod_{i \in V(F)} f(x_i)^{d_i} - \prod_{i \in V(F)} g(x_i)^{d_i}\right) \mathrm{d}\lambda^k(\boldsymbol{x})\right|}_{②}. \tag{29}$$

For term ①, we set $D := \sum_i d_i$ and observe that for all $\boldsymbol{x} \in [0, 1]^k$

$$\frac{1}{r^D} \prod_{i \in V(F)} f(x_i)^{d_i} \in [-1, 1], \tag{30}$$

and hence similarly to the standard proof of the classical counting lemma (see, e.g., Zhao (2023)) we can bound

$$① \;\le\; r^D e(F) \left\|W - V\right\|_{\square,2} \;\le\; 4r^D e(F) \left\|W - V\right\|_\square. \tag{31}$$

In comparison to the standard proof, the usage of $\left\|\cdot\right\|_{\square,2}$, an alternative definition of the cut norm, stems from the fact that function values appearing in the integral in ① (renormalizing by r^D) are not necessarily in $[0, 1]$, but $[-1, 1]$. See also Equations (4.3), (4.4) in Janson (2013). For ②, we bound the $\mathcal{L}^1$ difference of the terms involving f and g:

$$\left|\int_{[0,1]^k} \left(\prod_{\{i,j\} \in E(F)} V(x_i, x_j)\right)\left(\prod_{i \in V(F)} f(x_i)^{d_i} - \prod_{i \in V(F)} g(x_i)^{d_i}\right) \mathrm{d}\lambda^k(\boldsymbol{x})\right| \tag{32}$$

$$\le \int_{[0,1]^k} \left|\prod_{\{i,j\} \in E(F)} V(x_i, x_j)\right| \left|\prod_{i \in V(F)} f(x_i)^{d_i} - \prod_{i \in V(F)} g(x_i)^{d_i}\right| \mathrm{d}\lambda^k(\boldsymbol{x}) \tag{33}$$

$$\le \sum_{i \in V(F)} \int_{[0,1]^k} \left|f(x_i)^{d_i} - g(x_i)^{d_i}\right| \left|\prod_{j<i} f(x_i)^{d_i} \prod_{j>i} g(x_i)^{d_i}\right| \mathrm{d}\lambda^k(\boldsymbol{x}) \tag{34}$$

$$\le \sum_{i \in V(F)} r^{\sum_{j \neq i} d_i} \int_{[0,1]} \left|f(x_i)^{d_i} - g(x_i)^{d_i}\right| \mathrm{d}\lambda(x) \tag{35}$$

$$\overset{(*)}{\le} \sum_{i \in V(F)} r^{\sum_{j \neq i} d_i} \cdot d_i r^{d_i - 1} \left\|f - g\right\|_1 \;=\; Dr^{D-1} \left\|f - g\right\|_1 \;\le\; 2Dr^{D-1} \left\|f - g\right\|_\square, \tag{36}$$

where $(*)$ uses $\|f\|_\infty, \|g\|_\infty \leq r$ and hence the Lipschitz constant of $x \mapsto x^{d_i}$ is bounded by the maximum of its derivative $d_i r^{d_i-1}$, and the last inequality uses $\|\cdot\|_1 \leq 2\|\cdot\|_\square$ in one dimension. Combining the two bounds for ① from Eq. 31 and ② from Eq. 36, we obtain

$$\left|t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g))\right| \leq 4r^D e(F) \|W - V\|_\square + 2Dr^{D-1} \|f - g\|_\square, \tag{37}$$

which yields the claim. □

C.2 Proof of Theorem 3.3

Theorem 3.3 (Characterizations of weak isomorphism for graphon-signals). *Fix $r > 1$ and let $(W, f), (V, g) \in \mathcal{WL}_r$. Then, the following statements are equivalent:*

(1) $\delta_p((W, f), (V, g)) = 0$ for any $p \in [1, \infty)$;
(2) $\delta_\square((W, f), (V, g)) = 0$;
(3) $t(F, \boldsymbol{d}, (W, f)) = t(F, \boldsymbol{d}, (V, g))$ for all multigraphs F, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$;
(4) $t(F, \boldsymbol{d}, (W, f)) = t(F, \boldsymbol{d}, (V, g))$ for all simple graphs F, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$;
(5) $\mathbb{H}_k(W, f) \stackrel{\mathcal{D}}{=} \mathbb{H}_k(V, g)$ for all $k \in \mathbb{N}$;
(6) $\mathbb{G}_k(W, f) \stackrel{\mathcal{D}}{=} \mathbb{G}_k(V, g)$ for all $k \in \mathbb{N}$.

To begin, we extend the characterization of $\delta_\square$ from Eq. 2 to graphon-signals, i.e., in the definition of the unlabeled distance a minimum is attained when rearrangements over all measure-preserving maps are taken into account for both involved graphons. We state this in greater generality, particularly also for δ_p, extending Lovász (2012, Theorem 8.13).

Definition C.1 (Smooth and Invariant Norms). *Two norms $N = (N_1, N_2)$, where N_1 is a norm on $\mathcal{L}^\infty[0,1]$ and N_2 on $\mathcal{W}$, are called **smooth** if the two conditions*

(1) $W_n \to W \in \mathcal{W}$, $f_n \to f \in \mathcal{L}^\infty[0,1]$ almost everywhere,
(2) $\sup_{n\in\mathbb{N}} \|W_n\|_\infty \leq \infty$, $\sup_{n\in\mathbb{N}} \|f_n\|_\infty \leq \infty$,

imply that

$$N_1(f_n) \to N_1(f), \qquad N_2(W_n) \to N_2(W). \tag{38}$$

*They are **invariant** if*

$$N_1(f^\varphi) = N_1(f), \quad N_2(W^\varphi) = N_2(W) \qquad \forall \varphi \in S_{[0,1]}, \tag{39}$$

where $(W, f) \in \mathcal{W} \times \mathcal{L}^\infty[0,1]$.

The conditions clearly apply to $\|\cdot\|_\square$ (with the one-dimensional definition from Eq. 6) and $\|\cdot\|_p$ for $p \in [1, \infty)$, but *not* for $p = \infty$ (take for example $W_n = \mathbb{1}_{[0,1/n]^2}$, $f_n = \mathbb{1}_{[0,1/n]}$). We set

$$\delta_N((W, f), (V, g)) := \inf_{\varphi \in S_{[0,1]}} \big(N_2(W - V^\varphi) + N_1(f - g^\varphi)\big). \tag{40}$$

Lemma C.2 (Minima vs. Infima for Smooth Invariant Norms). *Let N be a smooth invariant norm on $\mathcal{W}$ and $\mathcal{L}^\infty[0,1]$. Then, we have the following alternate expressions for δ_N:*

$$\delta_N((W, f), (V, g)) = \inf_{\varphi \in \overline{S}_{[0,1]}} \big(N_2(W - V^\varphi) + N_1(f - g^\varphi)\big) \tag{41}$$

$$= \min_{\varphi, \psi \in \overline{S}_{[0,1]}} \big(N_2(W^\varphi - V^\psi) + N_1(f^\varphi - g^\psi)\big). \tag{42}$$

Proof sketch. We follow the proof of Theorem 8.13 by Lovász (2012), briefly highlighting the necessary adjustments to the argument.

To establish the first equality, approximations by step graphons that converge a.e. are considered, and the crucial point is that any $\varphi \in \overline{S}_{[0,1]}$ can be realized by a suitable $\widetilde{\varphi} \in S_{[0,1]}$ for such step graphons. For graphon-signals, the argument can be transferred if one simply considers partitions respecting each step graphon and step signal simultaneously when constructing the corresponding $\widetilde{\varphi} \in S_{[0,1]}$. For the second equality, which is proven in greater generality with coupling measures over $[0,1]^2$ by Lovász (2012), note that the lower semicontinuity in (8.24) is just shown for kernels (i.e., $\mathcal{L}^\infty[0,1]^2$), but the argument extends verbatim to $\mathcal{L}^\infty[0,1]$, and the sum of two lower semicontinuous functions is still lower semicontinuous. The rest of the argument applies without modification. □

Note that Lemma C.2 justifies our definition in Eq. 40, as $\delta_\square$ on the graphon-signal space was defined slightly differently as the infimum over all measure preserving bijections of co-null sets in $[0,1]$ (with the function being set to zero otherwise) by Levie (2023), i.e.,

$$S'_{[0,1]} := \{\varphi : A \to B \mid A, B \text{ co-null in } [0,1]\}. \tag{43}$$

Since

$$S_{[0,1]} \subseteq S'_{[0,1]} \subseteq \overline{S}_{[0,1]}, \tag{44}$$

this coincides with the definition of $\delta_\square$ in Eq. 40 by the first equality of Lemma C.2.

Additionally, the following lemma will be useful on several occasions when comparing two probability distributions.

Lemma C.3 (Moments of Random Vectors)**.** *Let $m \in \mathbb{N}$ and let $\boldsymbol{X}, \boldsymbol{Y}$ be m-dimensional random vectors which are almost surely bounded, i.e., there is some $R > 0$ such that $\mathbb{P}(\|\boldsymbol{X}\| \leq R) = \mathbb{P}(\|\boldsymbol{Y}\| \leq R) = 1$, for any norm $\|\cdot\|$ on $\mathbb{R}^m$. Suppose that for all $\boldsymbol{d} \in \mathbb{N}_0^m$ the moments of $\boldsymbol{X}, \boldsymbol{Y}$ agree:*

$$\mathbb{E}\left[\prod_{i=1}^m X_i^{d_i}\right] = \mathbb{E}\left[\prod_{i=1}^m Y_i^{d_i}\right]. \tag{45}$$

Then, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are identically distributed, i.e., $\boldsymbol{X} \overset{\mathcal{D}}{=} \boldsymbol{Y}$.

Proof. In the case of more general random variables/vectors, this is known in the literature as the *moment problem* (see, e.g., Schmüdgen (2017)). Under boundedness, however, this is trivial and can for example be proven via the characteristic functions of the random vectors, expanding the exponential function in the definition. □

We are now ready to prove Theorem 3.3. To this end, we will show the following implications:

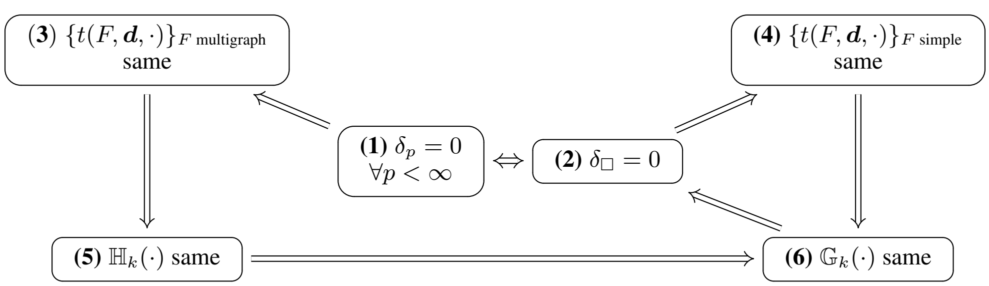


Figure 4: Equivalence chain for the proof of Theorem 3.3.

Proof of Theorem 3.3.

(1) ⇔ (2): Lemma C.2 implies that for any $p \in [1, \infty)$

$$\delta_\square((W,f),(V,g)) = 0 \Leftrightarrow \exists \varphi, \psi \in \overline{S}_{[0,1]} : (W,f)^\varphi = (V,g)^\psi \Leftrightarrow \delta_p((W,f),(V,g)) = 0, \tag{46}$$

where equality in the middle holds in an λ^2-a.e. sense.

(2) ⇒ (4): This follows immediately from Lemma 3.2.

(4) ⇒ (6): Let $(W,f),(V,g) \in \mathcal{WL}_r$ such that $t(F,\boldsymbol{d},(W,f)) = t(F,\boldsymbol{d},(V,g))$ for all simple graphs F, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$. Fix some $k \in \mathbb{N}$. Clearly, the distribution of $\mathbb{G}_k(W,f)$ is uniquely determined by

$$\mathbb{P}_{(G,\boldsymbol{f}) \sim \mathbb{G}_k(W,f)}(G \cong F), \tag{47}$$
$$\mathbb{P}_{(G,\boldsymbol{f}) \sim \mathbb{G}_k(W,f)}(\boldsymbol{f} \in \cdot | G \cong F), \tag{48}$$

i.e., the discrete distribution of the (labeled) random graph G and the conditional distribution of the node features given the graph structure, for every simple graph F of size k. For Eq. 47, we remark that the standard homomorphism densities with respect to just the graphons W, V can be recovered by taking $\boldsymbol{d} = 0$. Thus, the inclusion-exclusion argument from the proof of Theorem 4.9.1 in (Zhao, 2023) can be used verbatim to reconstruct the probabilities from Eq. 47. With a similar inclusion-exclusion argument, we see that for any F

$$\mathbb{1}\{\mathbb{G}_k(W) \cong F\} = \sum_{F' \supseteq F} (-1)^{e(F')-e(F)} \mathbb{1}\{\mathbb{G}_k(W) \supseteq F'\} \tag{49}$$

and therefore

$$\mathbb{E}_{(G,\boldsymbol{f})\sim\mathbb{G}_k(W,f)}\left[\prod_{i\in V(F)} f_i^{d_i} \middle| G \cong F\right] = \frac{\sum_{F'\supseteq F}(-1)^{e(F')-e(F)}\, t(F', \boldsymbol{d}, (W,f))}{\mathbb{P}_{(G,\boldsymbol{f})\sim\mathbb{G}_k(W,f)}(G \cong F)} \tag{50}$$

as long as the denominator is positive (otherwise, the corresponding conditional distribution is arbitrary). Since $(\boldsymbol{f}|G \cong F)$ is a bounded random vector ($\|\boldsymbol{f}\|_\infty \le r$ a.s.), its distribution is uniquely determined by its multidimensional moments, i.e., precisely the expressions from Eq. 50 (see Lemma C.3). Thus, we can conclude $\mathbb{G}_k(W,f) \overset{\mathcal{D}}{=} \mathbb{G}_k(V,g)$ for all $k \in \mathbb{N}$.

(6) ⇒ (2): This implication follows from applying the graphon-signal sampling lemma (Levie, 2023, Theorem 3.7): If **(6)** holds, we can bound

$$\begin{aligned} \delta_\square((W,f),(V,g)) &\le \mathbb{E}\big[\delta_\square((W,f),\mathbb{G}_k(W,f))\big] + \mathbb{E}\big[\delta_\square((V,g),\mathbb{G}_k(W,f))\big] && (51)\\ &= \mathbb{E}\big[\delta_\square((W,f),\mathbb{G}_k(W,f))\big] + \mathbb{E}\big[\delta_\square((V,g),\mathbb{G}_k(V,g))\big] \to 0 && (52) \end{aligned}$$

as $k \to \infty$.

(1) ⇒ (3): With a technique as in Lemma 3.2, bounding the individual graphon terms in a similar way as the signal terms, it is straightforward to show that the signal-weighted homomorphism density $t(F, \boldsymbol{d}, \cdot)$ from any *multigraph* F is also Lipschitz continuous with respect to δ_1. Thus, statement **(3)** follows immediately.

(3) ⇒ (5): Let $(W,f),(V,g) \in \mathcal{WL}_r$ such that $t(F,\boldsymbol{d},(W,f)) = t(F,\boldsymbol{d},(V,g))$ for all multigraphs F, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$. Fix $k \in \mathbb{N}$. Then, $\mathbb{H}_k(W,f)$ and $\mathbb{H}_k(V,g)$ can be seen as (k^2+k)-dimensional random vectors which are clearly bounded, since all graphon entries are in $[0,1]$ and all signal entries in $[-r,r]$. We observe that $\{t(F,\boldsymbol{d},(W,f))\}_{F,\boldsymbol{d}}$ and $\{t(F,\boldsymbol{d},(V,g))\}_{F,\boldsymbol{d}}$, with F ranging over multigraphs of size k, are precisely the multidimensional moments of these random vectors. Lemma C.3 yields statement **(5)**.

(5) ⇒ (6): This is trivial, since $\mathbb{G}_k(\cdot)$ is a function of $\mathbb{H}_k(\cdot)$. □

C.3 Proof of Corollary 3.4

Corollary 3.4 (Convergence in graphon-signal space)**.** *For $(W_n,f_n)_n$, $(W,f) \in \mathcal{WL}_r$ and $r > 1$,*

$$\delta_\square((W_n,f_n),(W,f)) \to 0 \quad \Leftrightarrow \quad t(F,\boldsymbol{d},(W_n,f_n)) \to t(F,\boldsymbol{d},(W,f)) \;\; \forall F, \boldsymbol{d} \in \mathbb{N}_0^{v(F)} \tag{11}$$

as $n \to \infty$, with F ranging over all simple graphs.

Proof. The proof idea is essentially the same as in the classical graphon case (for example, see § 4.9 in Zhao (2023)): An application of Theorem 3.3 that uses compactness of the graphon-signal space. For the sake of completeness, we restate the argument.

"⇒" follows immediately from the counting lemma (Lemma 3.2). For "⇐", let $(W_n,f_n)_n$ be a sequence of graphon-signals that left-converges to $(W,f) \in \mathcal{WL}_r$. By compactness (Levie, 2023, Theorem 3.6), there exists a subsequence $(W_{n_i}, f_{n_i})_i$ converging to some limit (V,g) in cut distance. But then also all signal-weighted homomorphism densities of the subsequence converge, and hence

$$t(F,\boldsymbol{d},(W,f)) = t(F,\boldsymbol{d},(V,g)) \quad \forall F, \boldsymbol{d} \in \mathbb{N}_0^{v(F)}. \tag{53}$$

Theorem 3.3 yields $\delta_\square((W,f),(V,g)) = 0$, i.e., also $(W_n,f_n) \to (W,f)$ in cut distance. □

C.4 Details on Theorem 3.5

Theorem 3.5 (k-WL for graphon-signals, informal)**.** *Two graphon-signals (W, f) and (V, g) are k-WL indistinguishable if and only if $t(F, \boldsymbol{d}, (W, f)) = t(F, \boldsymbol{d}, (V, g))$ for all multigraphs F of treewidth $\leq k$, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$.*

In the following, we will make the statement of Theorem 3.5 precise.

The 1-dimensional Weisfeiler-Leman graph isomorphism test can be generalized to graphons via **Iterated Degree Measures** and their **distributions** (Grebík & Rocha, 2022). Recently, a similar characterization for the k-WL test of graphons has been provided by Böker (2023), and related to *multigraph* homomorphism densities. Inspired by this, we define a natural generalization of the k-WL measure and distribution by Böker (2023) to graphon-signals and point out that their homomorphism expressivity can be described by signal-weighted homomorphism densities, as stated in Theorem 3.5. This provides the final step in establishing the suitability of this extension of homomorphism densities for our analyses. Since the details are very similar to the arguments for the graphon space, we will only highlight which definitions have to be adapted.

Definition C.4 (Weisfeiler-Leman Measures, cf. Böker (2023, Definition 26))**.** *Let $k \geq 1$ and fix $r > 0$. Let $P_0^k := [0,1]^{\binom{[k+1]}{2}} \times [-r, r]^{k+1}$. Inductively define $\mathbb{M}_n^k := \prod_{i \leq n} P_i^k$ and $P_{n+1}^k := \mathcal{P}(\mathbb{M}_n^k)^{k+1}$ for every $n \in \mathbb{N}$, where we consider P_0^k with its standard topology and Borel σ-algebra, and $\mathcal{P}(\cdot)$ denotes the set of all Borel probability measures, equipped with the weak topology. Let $\mathbb{M}^k := \prod_{n \in \mathbb{N}} P_i^k$ and define $p_{m,n} : \mathbb{M}_m^k \to \mathbb{M}_n^k$ to be the natural projection for $n \leq m$. Set*

$$\mathbb{P}^k := \left\{ \alpha \in \mathbb{M}^k \mid (\alpha_{n+1})_j = (p_{n+1,n})_* (\alpha_{n+2})_j \text{ for all } j \in [k+1], n \in \mathbb{N} \right\}, \tag{54}$$

where $(p_{n,n+1})_*$ is the pushforward of measures.

One can think of $\mathbb{P}^k$ as the space of all colors used by the k-WL test, where for $\alpha \in \mathbb{P}^k$ each entry α_n, $n \in \mathbb{N}$, is a *refinement* of the previous one. We can now define the equivalent of the color refinement for a specific graphon-signal.

Definition C.5 (k-WL, cf. Böker (2023, Definition 29))**.** *Let $k \geq 1$, $r > 0$, and let $(W, f) \in \mathcal{WL}_r$. Define $\mathcal{C}_{(W,f),0}^k : [0,1]^{k+1} \to \mathbb{M}_0^k$ by*

$$\mathcal{C}_{(W,f),0}^k(\boldsymbol{x}) := \left(\big(W(x_i, x_j)\big)_{\{i,j\} \in \binom{[k+1]}{2}}, \big(f(x_j)\big)_{j \in [k+1]} \right) \tag{55}$$

for every $\boldsymbol{x} \in [0,1]^{k+1}$. We then inductively define $\mathcal{C}_{(W,f),n+1}^k : [0,1]^{k+1} \to \mathbb{M}_{n+1}^k$ by

$$\mathcal{C}_{(W,f),n+1}^k(\boldsymbol{x}) := \left(\mathcal{C}_{(W,f),n}^k(\boldsymbol{x}), \left(\left(\mathcal{C}_{(W,f),n}^k \circ \boldsymbol{x}[\cdot/j] \right)_* \lambda \right)_{j \in [k+1]} \right) \tag{56}$$

for every $\boldsymbol{x} \in [0,1]^{k+1}$, where λ is the Lebesgue measure on $[0,1]$ and $\boldsymbol{x}[\cdot/j] := (x_1, \ldots, x_{j-1}, \cdot, x_{j+1}, \ldots, x_{k+1}) \in [0,1]^{k+1}$. Let $\mathcal{C}_{(W,f)}^k : [0,1]^{k+1} \to \mathbb{M}^k$ such that $(\mathcal{C}_{(W,f)}^k)_n = (\mathcal{C}_{(W,f),n}^k)_n$ for all $n \in \mathbb{N}$. Call

$$\nu_{(W,f)}^k := \left(\mathcal{C}_{(W,f)}^k \right)_* \lambda^{k+1} \in \mathcal{P}(\mathbb{M}^k) \tag{57}$$

*the k-**dimensional Weisfeiler-Leman distribution** of the graphon-signal (W, f).*

Intuitively, this corresponds with the multiset of colors that is obtained by the k-WL test for graphs (note that, however, for graphons, the partitions do not need to stabilize after a finite number of steps). One can show that $\nu_{(W,f)}^k(\mathbb{P}^k) = 1$. Skipping all the technical details (due to similarity), we get directly to the homomorphism expressivity of this k-WL test, which generalizes Böker (2023, Theorem 5, (1) $\Leftrightarrow$ (2)):

Theorem C.6 (Theorem 3.5, formal)**.** *Let $r > 0$ and $(W, f), (V, g) \in \mathcal{WL}_r$ be two graphon-signals. Then, $\nu_{(W,f)}^k = \nu_{(V,g)}^k$, i.e. the graphon-signals are k-WL indistinguishable, if and only if*

$$t(F, \boldsymbol{d}, (W, f)) = t(F, \boldsymbol{d}, (V, g)) \tag{58}$$

for all multigraphs F of treewidth $\leq k$ and $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$.

Proof sketch. The arguments of Böker (2023) transfer almost verbatim to this setting, as the span of $\{t(F, \boldsymbol{d}, \cdot)\}_{\mathrm{treewidth}(F)\leq k, \boldsymbol{d}}$ defines an algebra. The crux is that in the k-WL test, *both* the distribution of the graphon values W and the signal f are captured already in the first step (Eq. 55) and, therefore, higher-order moments have to be considered. $\square$

D Invariant Graphon Networks

We collect the proofs of our statements from § 4 in this appendix.

D.1 Proof of Theorem 4.2

Theorem 4.2. *Let $k, \ell \in \mathbb{N}_0$. Then, $\mathrm{LE}_{k\to\ell}$ is a finite-dimensional vector space of dimension*

$$\dim\,\mathrm{LE}_{k\to\ell} \;=\; \sum_{s=0}^{\min\{k,\ell\}} s!\binom{k}{s}\binom{\ell}{s} \;\leq\; \mathrm{bell}(k+\ell). \tag{13}$$

We will need some more preparations for the proof, in which we consider any $L \in \mathrm{LE}_{k\to\ell}$ only on step functions using regular intervals first, which will allow us to use the existing results for the discrete case. In order to prove Lemma D.2, we will need another preparation.

Lemma D.1 (Fixed Points of Measure Preserving Functions)**.** *If $k \in \mathbb{N}_0$ and $U \in \mathcal{L}^2[0,1]^k$ such that $U^\varphi = U$ for all $\varphi \in \overline{S}_{[0,1]}$, then U is constant. All involved equalities are meant λ^k-almost everywhere.*

Albeit somewhat tedious, the proof relies on basic measure theory and is rather straightforward. The only aspect requiring additional attention is that φ acts uniformly across all coordinates.

Proof. Let $U \in \mathcal{L}^2[0,1]^k$ such that U is invariant under all measure preserving functions, and suppose that U is not constant λ^k-almost everywhere. Then, there exist $a < b$ such that $A := U^{-1}((-\infty,a])$, $B := U^{-1}([b,\infty))$ have positive Lebesgue measure $\lambda^k(A) = \lambda^k(B) > 0$. Set $\alpha := \lambda^k(A)/\lambda^k(B)$. For $n \in \mathbb{N}$, let $I_j^{(n)} := [\frac{j-1}{n}, \frac{j}{n})$ for $j \in \{1,\dots,n-1\}$ and $I_n^{(n)} := [\frac{n-1}{n}, 1]$ be a partition of $[0,1]$ into regular intervals, and set $\mathcal{P}_k^{(n)} := \{I_{j_1}^{(n)} \times \cdots \times I_{j_k}^{(n)} \,|\, j_1,\dots,j_k \in \{1,\dots,n\}\}$. First, note that we have

$$\lambda^k(Q \cap A) \;\neq\; \alpha\lambda^k(Q \cap B) \tag{59}$$

for some $m \in \mathbb{N}$ and $Q \in \mathcal{P}_k^{(m)}$. Otherwise, equality in Eq. 59 would also hold for all hyperrectangles with rational endpoints, which is a $\cap$-stable generator of $\mathcal{B}([0,1]^k)$. Consequently, equality would hold for all sets in $\mathcal{B}([0,1]^k)$ and thus, $0 < \lambda^k(A) = \alpha\lambda^k(A \cap B) = \alpha\lambda^k(\emptyset) = 0$, which is a contradiction. W.l.o.g. assume $\lambda^k(Q \cap A) > \alpha\lambda^k(Q \cap B)$ in Eq. 59. As

$$\sum_{S\in\mathcal{P}_k^{(m)}} \lambda^k(S \cap A) \;=\; \lambda^k(A) \;=\; \alpha\lambda^k(B) \;=\; \sum_{S\in\mathcal{P}_k^{(m)}} \alpha\lambda^k(S \cap B), \tag{60}$$

there must be another $R \in \mathcal{P}_k^{(m)}$ such that $\lambda^k(R \cap A) < \alpha\lambda^k(R \cap B)$. Set

$$\Delta_k \;:=\; \{\boldsymbol{x} \in [0,1]^k \,|\, |\{x_1,\dots,x_k\}| < k\}, \quad \Delta_k^{(n)} \;:=\; \{Q \in \mathcal{P}_k^{(n)} \,|\, Q \cap \Delta \neq \emptyset\} \tag{61}$$

to be the union of all diagonals on $[0,1]^k$ and the elements of $\mathcal{P}_k^{(n)}$ overlapping with Δ_k respectively for $n \in \mathbb{N}$. As $\lambda^k(\bigcup_{Q\in\Delta_k^{(n)}} Q) \to \lambda^k(\Delta_k) = 0$ as $n \to \infty$, there must exist $m^* \geq m \in \mathbb{N}$ such that there are $Q \supseteq Q^* \in \mathcal{P}_k^{(m^*)} \setminus \Delta_k^{(m^*)}$, $R \supseteq R^* \in \mathcal{P}_k^{(m^*)} \setminus \Delta_k^{(m^*)}$ satisfying

$$\lambda^k(Q^* \cap A) > \alpha\lambda^k(Q^* \cap B), \quad \lambda^k(R^* \cap A) < \alpha\lambda^k(R^* \cap B). \tag{62}$$

Since Q^* and R^* do not overlap with any diagonal, we can now construct $\varphi \in S_{[0,1]}$ such that $\varphi^{\otimes k}$, which clearly defines a measure-preserving bijection from $[0,1]^k$ to itself, sends Q^* to R^*. By invariance of U under all measure-preserving functions, we get

$$\lambda^k(R^* \cap A) \;=\; \lambda^k\left((\varphi^{\otimes k})^{-1}(R^* \cap A)\right) \;=\; \lambda^k(Q^* \cap A), \tag{63}$$

$$\lambda^k(R^* \cap B) \;=\; \lambda^k\left((\varphi^{\otimes k})^{-1}(R^* \cap B)\right) \;=\; \lambda^k(Q^* \cap B), \tag{64}$$

which contradicts Eq. 62. Hence, U must be λ^k-a.e. constant. □

We are now ready to prove Lemma D.2. Let $k, \ell \in \mathbb{N}_0$ and $n \in \mathbb{N}$. Let $I_j^{(n)} := [\frac{j-1}{n}, \frac{j}{n})$ for $j \in \{1, \dots, n-1\}$ and $I_n^{(n)} := [\frac{n-1}{n}, 1]$ be a partition of $[0,1]$ into regular intervals. Let $\mathcal{A}_n := \sigma\left(\{I_1^{(n)}, \dots, I_n^{(n)}\}\right)$ denote the σ-algebra generated by this partition and let $\mathcal{F}_k^{(n)} := \{W \in \mathcal{L}^2[0,1]^k \mid W \text{ is } \mathcal{A}_n^{\otimes k}\text{-measurable}\}$; define $\mathcal{F}_\ell^{(n)}$ similarly.

Lemma D.2 (Invariance under discretization). *Any $L \in \mathrm{LE}_{k\to\ell}$ is invariant under discretization, which means that for any $n \in \mathbb{N}$, $L(\mathcal{F}_k^{(n)}) \subseteq \mathcal{F}_\ell^{(n)}$, where the inclusion should be understood up to sets of measure zero.*

Proof. Let $L \in \mathrm{LE}_{k\to\ell}$ and let $U \in \mathcal{F}_k^{(n)}$. Then, if $\varphi \in \overline{S}_{[0,1]}$ such that

$$\varphi(I_j^{(n)}) \subseteq I_j^{(n)} \tag{65}$$

for any $j \in \{1, \dots, n\}$, we have $U^\varphi = U$ and hence also $L(U)^\varphi = L(U)$ λ^ℓ-almost everywhere. Take any hypercube $Q = I_{j_1}^{(n)} \times \dots \times I_{j_\ell}^{(n)}$ with $j_1, \dots, j_\ell \in \{1, \dots, n\}$ and any measure-preserving function $\varphi : [0, 1/n) \to [0, 1/n)$. We replicate φ on the unit interval as

$$\varphi^*(x) := x \operatorname{div} 1/n \; + \; \varphi\,(x \bmod 1/n)\,, \tag{66}$$

which clearly satisfies Eq. 65, and thus $L(U)^{\varphi^*} = L(U)$ almost everywhere. Since now

$$L(U)\big|_Q \;=\; L(U)^{\varphi^*}\Big|_Q \;=\; \left(L(U)\big|_Q\right)^\varphi, \tag{67}$$

where we identify φ with $\varphi^*\big|_{I_j^{(n)}}$ (which define measure preserving functions on $I_j^{(n)}$), we can use translation invariance and scale equivariance of the Lebesgue measure to conclude by Lemma D.1 that $L(U)\big|_Q$ is constant λ^ℓ-almost everywhere. As Q was chosen arbitrarily, this implies the statement of the lemma. $\square$

Proof of Theorem 4.2. Let $n \in N$ and $L \in \mathrm{LE}_{k\to\ell}$. By Lemma D.2, we know that $L(\mathcal{F}_k^{(n)}) \subseteq \mathcal{F}_\ell^{(n)}$. Since $\mathcal{F}_k^{(n)} \cong (\mathbb{R}^n)^{\otimes k} \cong \mathbb{R}^{n^k}$, we can regard $L\big|_{\mathcal{F}_k^{(n)}} : \mathcal{F}_k^{(n)} \to \mathcal{F}_\ell^{(n)}$ as a linear operator $\mathbb{R}^{n^k} \to \mathbb{R}^{n^\ell}$. Taking for any $\sigma \in S_n$ a measure-preserving transformation $\varphi_\sigma \in S_{[0,1]}$ with $\varphi_\sigma(I_j) = I_{\varphi(j)}$, we can see that $L\big|_{\mathcal{F}_k^{(n)}}$ is also permutation equivariant, and we can use the characterization of the basis elements from § B.1.

Note that for any $n, m \in \mathbb{N}$ we have $\mathcal{F}_k^{(n)} \subseteq \mathcal{F}_k^{(nm)}$ and the canonical basis elements $\{L_\gamma\}_{\gamma\in\Gamma_{k+l}}$ under the identification $\mathcal{F}_k^{(n)} \cong \mathbb{R}^{n^k}, \mathcal{F}_k^{(m)} \cong \mathbb{R}^{m^k}$ are compatible in the sense that

$$L_\gamma^{(nm)}\Big|_{\mathcal{F}_k^{(n)}} = L_\gamma^{(n)}. \tag{68}$$

Hence, the coefficients of $L\big|_{\mathcal{F}_k^{(n)}}$ with respect to the canonical basis $\{L_\gamma^{(n)}\}_{\gamma\in\Gamma_{k+\ell}}$ do not depend on the specific $n \in \mathbb{N}$. W.l.o.g. assume that L restricted to some $\mathcal{F}_k^{(n)}$ is a canonical basis function $L_{\gamma^*}^{(n)}$ (where $\gamma^* \in \Gamma_{k+\ell}$ does not depend on n).

We now take a closer look at the partition γ^* and its induced function $L_{\gamma^*}^{(n)}$ described by the steps **Selection**, **Reduction**, **Alignment**, and **Replication**. Partition γ^* into the 3 subsets $\gamma_1^* := \{A \in \gamma^* \mid A \subseteq [k]\}$, $\gamma_2^* := \{A \in \gamma^* \mid A \subseteq k + [\ell]\}$, $\gamma_3^* := \gamma^* \setminus (\gamma_1^* \cup \gamma_2^*)$. For the constant $U \equiv 1 \in \mathcal{F}_k^{(1)} \subseteq \mathcal{L}^2[0,1]^k$, $L_{\gamma^*}^{(n)}(U) \neq 0$ must also be constant a.e. by compatibility with discretization, so the partition γ^* cannot correspond to a basis function whose images are supported on a diagonal. This is precisely equivalent to $|A \cap (k + [\ell])| \leq 1$ for all $A \in \gamma^*$. Now suppose that the input only depends on a diagonal. Denoting the restriction of the constant $U \equiv 1$ to the diagonal under discretization of $[0,1]$ into n pieces by $U_{\gamma^*}^{(n)}$,

$$L_{\gamma^*}^{(n)}(U_{\gamma^*}^{(n)}) \;=\; L_{\gamma^*}^{(n)}(U) \;=\; L_{\gamma^*}^{(1)}(U) \;\neq\; 0 \tag{69}$$

is constant, but $\|U_{\gamma^*}^{(n)}\|_2 \to 0$ for $n \to \infty$, which contradicts boundedness (i.e., continuity) of the operator $L \in \mathcal{B}(\mathcal{L}^2[0,1]^k, \mathcal{L}^2[0,1]^\ell)$. Hence, γ^* must correspond to a basis function for which the selection step ① is trivial, i.e., $|A \cap [k]| \leq 1$ for all $A \in \gamma^*$.

This leaves us with only the partitions $\gamma \in \Gamma_{k+\ell}$ whose sets $A \in \gamma$ contain at most one element from $[k]$ and $k+[\ell]$ respectively. In the following Lemma D.3 we will check that for all of these partitions, the **Reduction/Alignment/Replication**-procedure (with averaging in the sense of integration over $[0,1]$) indeed yields a valid operator $L_\gamma \in \mathcal{B}(\mathcal{L}^2[0,1]^k, \mathcal{L}^2[0,1]^\ell)$ which agrees with $L_\gamma^{(n)}$ on $\mathcal{F}_k^{(n)}$.

If we can now show that $\bigcup_{n\in\mathbb{N}} \mathcal{F}_k^{(n)} \subseteq \mathcal{L}^2[0,1]^k$ is dense with respect to $\|\cdot\|_2$, we can conclude that $L = L_{\gamma^*}$, as L is continuous and agrees with L_{γ^*} on a dense subset. However, this follows by a simple application of the martingale convergence theorem: Considering $[0,1]^k$ with Lebesgue measure λ^k as a probability space and $U \in L^2[0,1]^k$ as a random variable, we have $\mathbb{E}[U|\mathcal{A}_n^{\otimes k}] \in \mathcal{F}_k^{(n)}$. Also, $\sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{A}_n^{\otimes k}\right) = \mathcal{B}([0,1]^k)$ is the entire Borel σ-Algebra as $\mathcal{A}_n$ contains all intervals with rational endpoints, so $\mathbb{E}[U|\mathcal{A}_n^{\otimes k}] \to \mathbb{E}[U|\mathcal{B}([0,1]^k)] = U$ in $\mathcal{L}^2$.

Define now

$$\widetilde{\Gamma}_{k,\ell} := \{\gamma \in \Gamma_{k+\ell} \,|\, \forall A \in \gamma : |A \cap [k]| \leq 1, |A \cap k+[\ell]| \leq 1\}. \tag{70}$$

It is straightforward to show that

$$\left|\widetilde{\Gamma}_{k,\ell}\right| = \sum_{s=0}^{\min\{k,\ell\}} s! \binom{k}{s}\binom{\ell}{s}, \tag{71}$$

which can be seen as follows: Any partition on the l.h.s. can contain $s \in \{0,\ldots,\min\{k,\ell\}\}$ sets of size 2. Fixing some s, any of these sets can only contain one element from $[k]$ and one from $k+[\ell]$. For the elements occuring in sets of size 2, there are $\binom{k}{s}\binom{\ell}{s}$ options, and there are $s!$ ways to match the s selected elements in $[k]$ with the s elements in $k+[\ell]$, leaving us with the formula on the right. This concludes the proof. □

D.2 Continuity of Linear Equivariant Layers

We first show that all $L \in \mathrm{LE}_{k\to\ell}$ are also continuous with respect to any $\mathcal{L}^p$ norm. Particularly, the special case $p = 2$ is still needed to complete the proof of Theorem 4.2.

Lemma D.3 (Continuity of Linear Equivariant Layers w.r.t. $\|\cdot\|_p$)**.** *Fix $k,\ell \in \mathbb{N}_0$. Let $L \in \mathrm{LE}_{k\to\ell}$ and $p \in [1,\infty]$. Then, L can also be regarded as a bounded linear operator $\mathcal{L}^p[0,1]^k \to \mathcal{L}^p[0,1]^\ell$. Furthermore, all of the canonical basis elements from the proof of Theorem 4.2 have operator norm $\|L\|_{p\to p} = 1$.*

Proof. If suffices to show boundedness of all canonical basis elements. Let $\gamma \in \widetilde{\Gamma}_{k,\ell}$ be a partition corresponding to a basis element L_γ in $\mathrm{LE}_{k\to\ell}$, and suppose that γ contains s sets of size 2 $\{i_1,j_1\},\ldots,\{i_s,j_s\}$ with $i_1,\ldots,i_s \in [k]$, $j_1,\ldots,j_s \in k+[\ell]$, and set $A = (i_1,\ldots,i_s)$, $B = (j_1,\ldots,j_s)$. Then, we can write L_γ as

$$L_\gamma(U) := \left[[0,1]^\ell \ni \boldsymbol{y} \mapsto \int_{[0,1]^{k-s}} U(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A}) \,\mathrm{d}\lambda^{k-s}(\boldsymbol{x}_{[k]\setminus A}) \Big|_{\boldsymbol{x}_A = \boldsymbol{y}_B} \right]. \tag{72}$$

Consider at first $p < \infty$. Clearly, Eq. 72 is also well-defined for $U \in \mathcal{L}^p[0,1]^k$ and

$$\|L_\gamma(U)\|_p^p = \int_{[0,1]^\ell} \left| \int_{[0,1]^{k-s}} U(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A}) \,\mathrm{d}\lambda^{k-s}(\boldsymbol{x}_{[k]\setminus A}) \Big|_{\boldsymbol{x}_A = \boldsymbol{y}_B} \right|^p \mathrm{d}\lambda^\ell(\boldsymbol{y}) \tag{73}$$

$$\leq \int_{[0,1]^\ell} \int_{[0,1]^{k-s}} \left|U(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A})\right|^p \mathrm{d}\lambda^{k-s}(\boldsymbol{x}_{[k]\setminus A}) \Big|_{\boldsymbol{x}_A = \boldsymbol{y}_B} \mathrm{d}\lambda^\ell(\boldsymbol{y}) \tag{74}$$

$$= \int_{[0,1]^s} \int_{[0,1]^{k-s}} \left|U(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A})\right|^p \mathrm{d}\lambda^{k-s}(\boldsymbol{x}_{[k]\setminus A}) \,\mathrm{d}\lambda^s(\boldsymbol{x}_A) = \|U\|_p^p, \tag{75}$$

with Jensen's inequality being applied in the second step. Note that equality holds, e.g., for $U \equiv 1$, so $\|L_\gamma\|_{p\to p} = 1$. For $p = \infty$, we also see

$$\|L_\gamma(U)\|_\infty = \operatorname*{ess\,sup}_{\boldsymbol{y}\in[0,1]^\ell} \left| \int_{[0,1]^{k-s}} U(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A})\, \mathrm{d}\lambda^{k-s}(\boldsymbol{x}_{[k]\setminus A}) \Big|_{\boldsymbol{x}_A=\boldsymbol{y}_B} \right| \tag{76}$$

$$\leq \operatorname*{ess\,sup}_{\boldsymbol{y}\in[0,1]^\ell} \int_{[0,1]^{k-s}} \underbrace{\left|U(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A})\right|}_{\leq \|U\|_\infty \text{ a.e.}} \mathrm{d}\lambda^{k-s}(\boldsymbol{x}_{[k]\setminus A}) \Bigg|_{\boldsymbol{x}_A=\boldsymbol{y}_B} \tag{77}$$

$$\leq \|U\|_\infty, \tag{78}$$

again with equality for $U \equiv 1$. □

D.3 Proof of Proposition 4.4

Proposition 4.4. *Any IWN of the form Eq. 18 is an instance of IGN-small (Cai & Wang, 2022).*

Proof. Let $\mathcal{N} = \boldsymbol{T}^{(S)} \circ \varrho \circ \cdots \circ \varrho \circ \boldsymbol{T}^{(1)}$ be an IWN as in Definition 4.3. We verify that $\mathcal{N}$ fulfills the condition of IGN-small, see Definition B.1.

By invariance under discretization (Lemma D.2), we can see that IWNs fulfill a basis representation stability condition, and under the identification $\mathcal{F}_k^{(n)} \cong \mathbb{R}^{n^k}$, which precisely captures grid-sampling, this directly implies that grid-sampling commutes with application of the IWN and its discrete equivalent. □

E Properties of Invariant Graphon Networks

In this appendix, we will collect the proofs of our statements from § 5.

E.1 Proof of Lemma 5.1

Lemma 5.1. *Let $\mathcal{N} : \mathcal{WL}_r \to \mathbb{R}$ be an IWN as defined in Eq. 18. Then, $\mathcal{N}$ is Lipschitz continuous w.r.t. δ_p for each $p \in [1, \infty]$.*

Proof of Lemma 5.1. Let

$$\mathcal{N} = \boldsymbol{T}^{(S)} \circ \varrho \circ \cdots \circ \varrho \circ \boldsymbol{T}^{(1)} \tag{79}$$

be an IWN as in Definition 4.3. By Lemma D.3, each $L \in \mathrm{LE}_{k\to\ell}$ is Lipschitz continuous with respect to $\|\cdot\|_p$ on the respective input and output space, and this immediately carries over to all $\boldsymbol{T}^{(s)}$, $s \in [S]$. Hence, it suffices to check that the pointwise application of the nonlinearity, i.e., $\mathcal{L}^p[0,1]^k \ni U \mapsto \varrho(U)$, is Lipschitz continuous for every $k \in \mathbb{N}_0$, where ϱ is applied elementwise. If C_ϱ is the Lipschitz constant of ϱ, we have

$$\|\varrho(U) - \varrho(W)\|_p^p = \int_{[0,1]^k} |\varrho(U) - \varrho(W)|^p \,\mathrm{d}\lambda^k \le \int_{[0,1]^k} C_\varrho^p \,|U - W|^p \,\mathrm{d}\lambda^k = C_\varrho^p \|U - W\|_p^p, \tag{80}$$

for $p < \infty$, and a similar argument shows the claim for $\|\cdot\|_\infty$. □

E.2 Proof of Theorem 5.2

E.2.1 General Outline

In this section, we will prove the following theorem:

Theorem 5.4 (δ_p-Universality of IWNs)**.** *Let $r > 1$, $p \in [1, \infty)$, $\varrho : \mathbb{R} \to \mathbb{R}$ Lipschitz continuous and non-polynomial. For any compact $K \subset (\widetilde{\mathcal{WL}_r}, \delta_p)$, $\mathcal{IWN}_\varrho$ is dense in the continuous functions $C(K, \mathbb{R})$ w.r.t. $\|\cdot\|_\infty$.*

As a first step towards this, we show that IWNs can approximate signal-weighted homomorphism densities with respect to graphs of size up to their order. For this, we introduce the notion of homomorphism densities from *partially labeled graphs* (Lovász, 2012, § 7.2), which will later allow us to "glue" together homomorphism densities from parts of a specific graph. Let F be a multigraph with $V(F) = [k]$, $\boldsymbol{d} \in \mathbb{N}_0^k$, and let $A \subseteq [k]$ denote the set of labeled nodes. Then

$$t_{\boldsymbol{x}_A}(F, \boldsymbol{d}, (W, f)) := \int_{[0,1]^{k-|A|}} \left(\prod_{i \in V(F)} f(x_i)^{d_i} \right) \left(\prod_{\{i,j\} \in E(F)} W(x_i, x_j) \right) \mathrm{d}\lambda^{k-|A|}(\boldsymbol{x}_{[k]\setminus A}), \tag{81}$$

i.e., the integral is only formed over the unlabeled nodes, making the expression a function of $\boldsymbol{x}_A$.

Lemma E.1. *Let $r > 0$, $\varrho : \mathbb{R} \to \mathbb{R}$ Lipschitz continuous and non-polynomial, and F be a multigraph with $V(F) = [k]$, $\boldsymbol{d} \in \mathbb{N}_0^k$, and $A \subseteq [k]$. Fix $\varepsilon > 0$. Then there exists an EWN $\mathcal{N}$ of order k such that for all $(W, f) \in \mathcal{WL}_r$*

$$\|t_{\cdot_A}(F, \boldsymbol{d}, (W, f)) - \mathcal{N}(W, f)\|_{\infty, [0,1]^{|A|}} \le \varepsilon. \tag{82}$$

For the proof of Lemma E.1 (see § E.2.3) we take inspiration from Keriven & Peyré (2019), modeling the product in the homomorphism densities explicitly while tracking which linear equivariant layers are being used by the IWN. Note that if we set $A := \emptyset$, this lemma specifically means that we can approximate any signal-weighted homomorphism density from a multigraph of size k with an IWN of order k, up to arbitrary precision. By using a tree decomposition of the graph, we can then get to the statement of Theorem 5.4 (see § E.2.4).

E.2.2 PREPARATIONS

We start with a few preparations for the proofs of Lemma E.1 and Theorem 5.2. Namely, our main goal will be to show that if an EWN of a fixed order can approximate a set of functions, it can also approximate their product (Lemma E.3). This will turn out useful to model signal-weighted homomorphism densities later.

Lemma E.2. *Let $\mathcal{N}$ be an S-layer EWN of order $k \in \mathbb{N}$ as defined in Definition 4.3, such that $k_0 = 2$, $k_s = k$ for $s > 1$, and $h_0 = 2$, $h_S = 1$. Let the nonlinearity ϱ be differentiable at least at one point with nonzero derivative. Let $S^* \in \mathbb{N}$. Fix $\varepsilon > 0$. Then, there exists an $(S + S^*)$-layer EWN $\mathcal{N}^*$ of order k with $k_s = k$ for $s > 1$ such that for all $(W, f) \in \mathcal{WL}_r$*

$$\|\mathcal{N}(W, f) - \mathcal{N}^*(W, f)\|_{\infty,[0,1]^k} \leq \varepsilon. \tag{83}$$

Proof. The argument is an approximation of the identity function, using the nonlinearity ϱ. Let

$$M := \sup_{(W,f)\in\mathcal{WL}_r} \|\mathcal{N}(W, f)\|_{\infty,[0,1]^k}, \tag{84}$$

which is finite since $\mathcal{N}$ is Lipschitz continuous in $\|\cdot\|_\infty$ by Lemma 5.1. Let $x_0 \in \mathbb{R}$ be a point at which ϱ is differentiable with $\varrho'(x_0) \neq 0$. Fix $\delta > 0$. There exists some constant $r > 0$ such that $|x| \leq r$ implies

$$|\varrho(x_0 + x) - \varrho(x_0) - \varrho'(x_0)x| \leq \delta\,|x|. \tag{85}$$

Set

$$\mathrm{id}_\varrho(x) := \frac{M}{r\varrho'(x_0)} \left(\varrho\left(x_0 + \frac{r}{M}x\right) - \varrho(x_0)\right). \tag{86}$$

Then, for any $x \in [-M, M]$, $\left|\frac{r}{M}x\right| \leq r$ and hence

$$\begin{aligned} |\mathrm{id}_\varrho(x) - x| &= \left|\frac{M}{r\varrho'(x_0)}\right| \left|\varrho\left(x_0 + \frac{r}{M}x\right) - \varrho(x_0) - \varrho'(x_0)\frac{r}{M}x\right| && (87)\\ &\leq \left|\frac{M}{r\varrho'(x_0)}\right| \delta \left|\frac{r}{M}x\right| \leq \frac{M}{|\varrho'(x_0)|}\delta. && (88) \end{aligned}$$

If we set

$$\mathcal{N}^* := \mathrm{id}_\varrho^{S^*} \circ \mathcal{N}, \tag{89}$$

where application of id_ϱ is pointwise, we can see that $\mathcal{N}^*$ can be clearly represented by an $(S+S^*)$-layer EWN of order k, and

$$\|\mathcal{N}(W, f) - \mathcal{N}^*(W, f)\|_{\infty,[0,1]^k} \leq \left(\frac{M}{|\varrho'(x_0)|}\delta\right)^{S^*}, \tag{90}$$

which can be made arbitrarily small as $\delta \to 0$. □

Lemma E.3. *Let $\mathcal{N}_1, \ldots, \mathcal{N}_m$ be $m \in \mathbb{N}$ EWNs of order $k \in \mathbb{N}$, each with constant orders after the first layer, and hidden output dimension of 1, i.e., realizing a function $[0, 1]^k \to \mathbb{R}$. Fix $\varepsilon > 0$. Then, there exists an EWN $\mathcal{N}^*$ of order k with constant orders after the first layer, such that for all $(W, f) \in \mathcal{WL}_r$*

$$\left\|\prod_{\ell=1}^m \mathcal{N}_\ell(W, f) - \mathcal{N}^*(W, f)\right\|_{\infty,[0,1]^k} \leq \varepsilon. \tag{91}$$

For the proof, we will draw inspiration from Keriven & Peyré (2019). However, a crucial difference to their proof is that our approach does not rely on modeling multiplication by increasing the tensor orders.

Proof. We will exploit a property of cos that allows us to express products as sums. Namely, it is well-known that for $x_1, \ldots, x_m \in \mathbb{R}$ we have

$$\prod_{j=1}^m \cos(x_j) = \frac{1}{2^m} \sum_{\sigma\in\{\pm1\}^m} \cos\left(\sum_{j=1}^m \sigma_j x_j\right). \tag{92}$$

Fix $\delta > 0$. At first, we will describe how to approximate any $\mathcal{N}_\ell$ using cos as a nonlinearity. By the classical universal approximation theorem (see for example Pinkus (1999, Theorem 3.1)), we can approximate $\varrho : \mathbb{R} \to \mathbb{R}$ by a feedforward neural network $\varrho_{\cos}$ with one hidden layer, using the cosine function as nonlinearity. Fixing compact sets similarly as in Eq. 84 for each layer of the EWNs $\mathcal{N}_\ell$, we can see that replacing each occurence of ϱ by $\varrho_{\cos}$ and absorbing the linear factors into the linear equivariant layers again yields valid EWNs. Call these EWNs $\mathcal{N}_\ell^{\cos}$, $\ell = 1, \dots, m$. The underlying feedforward neural network $\varrho_{\cos}$ can now be chosen such that for all $\ell \in [m]$, $(W, f) \in \mathcal{WL}_r$

$$\|\mathcal{N}_\ell^{\cos}(W, f) - \mathcal{N}_\ell(W, f)\|_{\infty,[0,1]^k} \leq \delta. \tag{93}$$

Note that each $\mathcal{N}_\ell^{\cos}$ might have different numbers of layers. Hence, we invoke Lemma E.2 to equalize the number of layers, and add one more layer of the identity. We obtain EWNs $\widetilde{\mathcal{N}}_\ell^{\cos}$ such that

$$\left\|\mathrm{id}_{\cos}(\widetilde{\mathcal{N}}_\ell^{\cos}(W, f)) - \mathcal{N}_\ell^{\cos}(W, f)\right\|_{\infty,[0,1]^k} \leq \delta \tag{94}$$

for all $(W, f) \in \mathcal{WL}_r$. Using Eq. 92 and setting $\mathrm{id}_{\cos}(x) = c\cdot\cos(ax+b)+d$ for some $a, b, c, d \in \mathbb{R}$, we can now write

$$\prod_{\ell=1}^m \mathrm{id}_{\cos}(\widetilde{\mathcal{N}}_\ell^{\cos}(W, f)) = \prod_{\ell=1}^m \Big(c \cdot \cos(a \cdot \widetilde{\mathcal{N}}_\ell^{\cos}(W, f) + b) + d\Big) \tag{95}$$

$$= \sum_{A \subseteq [m]} c^{|A|} d^{m-|A|} \left(\prod_{\ell \in A} \cos(a \cdot \widetilde{\mathcal{N}}_\ell^{\cos}(W, f) + b)\right) \tag{96}$$

$$= \sum_{A \subseteq [m]} \frac{c^{|A|} d^{m-|A|}}{2^{|A|}} \sum_{\sigma \in \{\pm 1\}^A} \cos \underbrace{\Big(a \cdot \widetilde{\mathcal{N}}_\ell^{\cos}(W, f) + b\Big)}_{(*)}, \tag{97}$$

which can be represented by an EWN of order k and one layer more compared to $\{\widetilde{\mathcal{N}}_\ell^{\cos}\}_\ell$, stacking the expressions from $(*)$, applying the nonlinearity cos and aggregating the outputs as determined by the weighted sum. Let

$$M := \max_{\ell \in [m]} \sup_{(W,f) \in \mathcal{WL}_r} \|\mathcal{N}_\ell(W, f)\|_{\infty,[0,1]^k}. \tag{98}$$

Since for any $(W, f) \in \mathcal{WL}_r$

$$\left\|\prod_{\ell=1}^m \mathrm{id}_{\cos}(\widetilde{\mathcal{N}}_\ell^{\cos}(W, f)) - \prod_{\ell=1}^m \mathcal{N}_\ell(W, f)\right\|_{\infty,[0,1]^k} \tag{99}$$

$$\leq \sum_{\ell=1}^m \left\|\left(\prod_{i>\ell} \mathrm{id}_{\cos}(\widetilde{\mathcal{N}}_\ell^{\cos}(W, f)) \prod_{i<\ell} \mathcal{N}_\ell(W, f)\right) \Big(\mathrm{id}_{\cos}(\widetilde{\mathcal{N}}_\ell^{\cos}(W, f)) - \mathcal{N}_\ell(W, f)\Big)\right\|_{\infty,[0,1]^k} \tag{100}$$

$$\leq \sum_{\ell=1}^m \left\|\prod_{i>\ell} \mathrm{id}_{\cos}(\widetilde{\mathcal{N}}_\ell^{\cos}(W, f)) \prod_{i<\ell} \mathcal{N}_\ell(W, f)\right\|_{\infty,[0,1]^k} \left\|\mathrm{id}_{\cos}(\widetilde{\mathcal{N}}_\ell^{\cos}(W, f)) - \mathcal{N}_\ell(W, f)\right\|_{\infty,[0,1]^k} \tag{101}$$

$$\leq \sum_{\ell=1}^m (M + 2\delta)^{m-\ell} M^{\ell-1} 2\delta, \tag{102}$$

and since Eq. 102 goes to zero as $\delta \to 0$, this shows the claim. □

E.2.3 Proof of Lemma E.1

Lemma E.1. *Let $r > 0$, $\varrho : \mathbb{R} \to \mathbb{R}$ Lipschitz continuous and non-polynomial, and F be a multigraph with $V(F) = [k]$, $\boldsymbol{d} \in \mathbb{N}_0^k$, and $A \subseteq [k]$. Fix $\varepsilon > 0$. Then there exists an EWN $\mathcal{N}$ of order k such that for all $(W, f) \in \mathcal{WL}_r$*

$$\|t_{\cdot A}(F, \boldsymbol{d}, (W, f)) - \mathcal{N}(W, f)\|_{\infty,[0,1]^{|A|}} \leq \varepsilon. \tag{82}$$

The proof boils down to an application of Lemma E.3.

Proof. Fix a multigraph F with $V(F) = [k]$, $\boldsymbol{d} \in \mathbb{N}_0^k$, and $A \subseteq [k]$. Let $(W, f) \in \mathcal{WL}_r$. Define

$$f_i : [0,1]^k \to \mathbb{R},\ \boldsymbol{x} \mapsto f(x_i) \qquad \text{for } i \in V(F), \tag{103}$$

$$W_{\{i,j\}} : [0,1]^k \to \mathbb{R},\ \boldsymbol{x} \mapsto W(x_i, x_j) \qquad \text{for } \{i,j\} \in E(F). \tag{104}$$

Clearly, both $(W, f) \mapsto f_i$ and $(W, f) \mapsto W_{\{i,j\}}$ can be exactly represented by EWNs of order k with one layer (i.e., no application of the nonlinearity at all). By Lemma E.3, the product

$$(W, f) \mapsto \prod_{i \in V(F)} f_i^{d_i} \prod_{\{i,j\} \in E(F)} W_{\{i,j\}} \tag{105}$$

can be approximated by an EWN $\mathcal{N}$ of order k up to ε in $\|\cdot\|_\infty$. However,

$$t_{\cdot_A}(F, \boldsymbol{d}, (W, f)) = \int_{[0,1]^{k-|A|}} \left(\prod_{i \in V(F)} f_i(\boldsymbol{x})^{d_i} \right) \left(\prod_{\{i,j\} \in E(F)} W_{\{i,j\}}(\boldsymbol{x}) \right) \mathrm{d}\lambda^{k-|A|}(\boldsymbol{x}_{[k] \setminus A}) \tag{106}$$

and thus for any $(W, f) \in \mathcal{WL}_r$

$$\left\| t_{\cdot_A}(F, \boldsymbol{d}, (W, f)) - \int_{[0,1]^{k-|A|}} \mathcal{N}(W, f)(\boldsymbol{x}) \, \mathrm{d}\lambda^{k-|A|}(\boldsymbol{x}_{[k] \setminus A}) \right\|_{\infty, [0,1]^{|A|}} \tag{107}$$

$$\leq \left\| \prod_{i \in V(F)} f_i^{d_i} \prod_{\{i,j\} \in E(F)} W_{\{i,j\}} - \mathcal{N}(W, f) \right\|_{\infty, [0,1]^k} \leq \varepsilon, \tag{108}$$

which shows the claim, since the r.h.s. of Eq. 107 can also be implemented by an EWN. $\square$

E.2.4 Assembling the Pieces: Proof of Theorem 5.4

Theorem 5.2 (Approximation of signal-weighted homomorphism densities). *Let $r > 0$, $\varrho : \mathbb{R} \to \mathbb{R}$ Lipschitz continuous and non-polynomial, and F be a multigraph of treewidth $k \in \mathbb{N}$, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$. Fix $\varepsilon > 0$. Then there exists an IWN $\mathcal{N} \in \mathcal{IWN}_\varrho^{k+1}$ of order $k+1$ such that for all $(W, f) \in \mathcal{WL}_r$*

$$|t(F, \boldsymbol{d}, (W, f)) - \mathcal{N}(W, f)| \leq \varepsilon. \tag{20}$$

For the proof, we will proceed by induction over the graph's tree decomposition. The base case is covered by Lemma E.1, after which we can perform induction over the size of the tree in a tree decomposition of graphs.

Proof. Fix $k \in \mathbb{N}$. To show the statement for all graphs of treewidth k, we perform induction over the number of bags in the graphs' tree decomposition, and show that the statement also holds for partially labeled graphs, see Eq. 81, as long as the labeled nodes A are contained in one bag of the tree decomposition.

As a base case, consider a multigraph F whose tree decomposition only consists of one bag, and $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$. This means that $v(F) \leq k+1$, and invoking Lemma E.1 immediately yields that we can approximate the signal-weighted homomorphism density $t_{\cdot_A}(F, \boldsymbol{d}, \cdot)$ for $A \subseteq V(F)$ to arbitrary precision on $\mathcal{WL}_r$ using an IWN of order at most $k+1$.

For the induction step, fix $m \in \mathbb{N}$, $m > 1$. Suppose that we can approximate $t_{\cdot_A}(F, \boldsymbol{d}, \cdot)$ for any F of treewidth at most k such that its tree decomposition consists of at most $m-1$ bags. Now, let F be a multigraph of treewidth k such that F has a tree decomposition (T, β), where T is a tree and

$$\beta : V(T) \to 2^{V(F)}, \tag{109}$$

such that for any $t \in V(T)$, $|\beta(t)| \leq k+1$ and $|V(T)| = m$, i.e., the tree decomposition consists of m bags. Let $t^* \in V(T)$, such that we approximate a signal-weighted homomorphism density from a partially labeled graph with labels $A \subseteq \beta(t^*)$.

Suppose removing t^* from T induces ℓ subtrees $T_1, \ldots, T_\ell$ of T. Let $T_{\ell+1}$ denote the subtree that consists of just the bag t^*. Let

$$\alpha_V : V(F) \to [\ell+1], \quad \alpha_E : E(F) \to [\ell+1] \tag{110}$$

be assignments of nodes and edges to one of the $\ell+1$ subtrees, such that

$$\forall v \in V(F) : v \in \bigcup_{t \in V(T_{\alpha_V(v)})} \beta(t), \quad \forall e \in E(F) : \exists t \in V(T_{\alpha_E(e)}) : e \subseteq \beta(t), \tag{111}$$

i.e., each $v \in V(F)$ is mapped to a subtree that contains v in one of its bags, and each edge in $E(F)$ is mapped to a subtree that contains both ends of the edge in one bag. Then,

$$V_i^* := \{v \in V(F) \,|\, \alpha_V(v) = i\}, \quad E_i^* := \{e \in E(F) \,|\, \alpha_E(e) = i\} \tag{112}$$

for $i \in [\ell+1]$ are partitions of $V(F)$ and $E(F)$ respectively. Further set

$$V_i := \bigcup_{t \in T_i} \beta(t). \tag{113}$$

Now, let $s_1 \in V(T_1), \ldots, s_\ell \in V(T_\ell)$ be the neighbors of t^* in T. As the set of bags that contain a

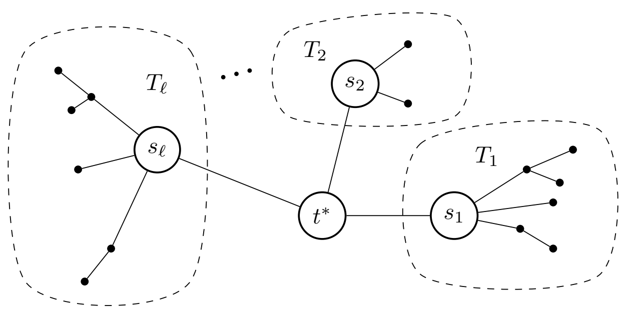


Figure 5: We dissect the tree T into subtrees $T_1, \ldots, T_\ell$, coming from t^*.

certain node from $V(F)$ is connected in T, it must hold that

$$B_i := V_i \cap \beta(t^*) \subseteq \beta(s_i). \tag{114}$$

Further, define the vectors $\boldsymbol{d}_i^* \in \mathbb{N}_0^{|V_i|}$ to coincide with $\boldsymbol{d}$ on V_i^*, and to be zero on $V_i \setminus V_i^*$, for $i \in [\ell+1]$. By the induction hypothesis, for any $i \in [\ell+1]$ and $\delta > 0$ there exists an EWN of order at most $k+1$ that can approximate

$$t_{\cdot_{B_i}}(F[V_i, E_i^*], \boldsymbol{d}_i^*, \cdot) \tag{115}$$

on $\mathcal{WL}_r$ to precision δ in $\|\cdot\|_\infty$, since each T_i defines a valid tree decomposition of the subgraph $F[V_i, E_i^*]$ (obtained from F by selecting the nodes in V_i and edges in E_i^*), consisting of $< m$ bags and with each bag containing at most $k+1$ nodes. Replicating the output of Eq. 115 along the axes in $\beta(t^*) \setminus B_i$ and padding $\boldsymbol{d}_i^*$ with zeros, we can use Eq. 115 to create an EWN $\mathcal{N}_i$ for each $i \in [\ell+1]$ mapping to functions $[0,1]^{|\beta(t^*)|} \to \mathbb{R}$, such that for any $(W, f) \in \mathcal{WL}_r$

$$\left\| t_{\cdot_{\beta(t^*)}}(F[E_i^*], \boldsymbol{d}_i^*, (W,f)) - \mathcal{N}_i(W,f) \right\|_{\infty, [0,1]^{|\beta(t^*)|}} \leq \delta. \tag{116}$$

However,

$$\prod_{i=1}^{\ell+1} t_{\cdot_{\beta(t^*)}}(F[E_i^*], \boldsymbol{d}_i^*, \cdot) = t_{\cdot_{\beta(t^*)}}(F, \boldsymbol{d}, \cdot), \tag{117}$$

and we can invoke Lemma E.3 to conclude that there is an EWN of order $k+1$ that can approximate the product in Eq. 117 to arbitrary precision. Finally, integrating over the axes corresponding to the set of unlabeled nodes $\beta(t^*) \setminus A$ yields the claim. $\square$

E.3 Proof of Corollary 5.3

Having shown Theorem 5.2, k-WL expressivity can be obtained effectively by definition.

Corollary 5.3 (k-WL expressivity)**.** *$\mathcal{IWN}_{\varrho}^{k+1}$ is at least as expressive as the k-WL test (Theorem 3.5) at distinguishing graphon-signals.*

Proof. Let $(W, f), (V, g) \in \mathcal{WL}_r$ be distinguishable by k-WL. By Theorem 3.5, this means that there exists a multigraph F of treewidth at most k and $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$ for which

$$t(F, \boldsymbol{d}, (W, f)) \neq t(F, \boldsymbol{d}, (V, g)). \tag{118}$$

Let $\varepsilon := |t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g))| > 0$. By Theorem 5.2, take an IWN $\mathcal{N}$ such that

$$\sup_{(U,h)\in\mathcal{WL}_r} |\mathcal{N}(U, h) - t(F, \boldsymbol{d}, (U, h))| \leq \varepsilon/3. \tag{119}$$

We obtain

$$\begin{aligned}
&|\mathcal{N}(W, f) - \mathcal{N}(V, g)| && (120)\\
&= \Big|\mathcal{N}(W, f) - t(F, \boldsymbol{d}, (W, f)) + t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g)) + t(F, \boldsymbol{d}, (V, g)) - \mathcal{N}(V, g)\Big| && (121)\\
&\geq |t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g))| && (122)\\
&\quad - |\mathcal{N}(W, f) - t(F, \boldsymbol{d}, (W, f)))| - |t(F, \boldsymbol{d}, (V, g)) - \mathcal{N}(V, g)|\\
&\geq \varepsilon - \varepsilon/3 - \varepsilon/3 = \varepsilon/3 > 0, && (123)
\end{aligned}$$

which yields the claim. □

E.4 Proof of Theorem 5.4

Note that Theorem 5.2 tells us that IWNs of *arbitrary order* can approximate any signal-weighted homomorphism density. As these separate points in $\widetilde{\mathcal{WL}}_r$ by Theorem 3.3, it is straightforward to obtain universality on any set which is compact with respect to a distance on $\widetilde{\mathcal{WL}}_r$ under which the signal-weighted homomorphism densities are continuous.

Theorem 5.4 (δ_p-Universality of IWNs)**.** *Let $r > 1$, $p \in [1, \infty)$, $\varrho : \mathbb{R} \to \mathbb{R}$ Lipschitz continuous and non-polynomial. For any compact $K \subset (\widetilde{\mathcal{WL}}_r, \delta_p)$, $\mathcal{IWN}_{\varrho}$ is dense in the continuous functions $C(K, \mathbb{R})$ w.r.t. $\|\cdot\|_{\infty}$.*

Proof. We show this statement by applying the Stone-Weierstrass theorem (see for example Rudin (1976, Theorem 7.32)). In principle, proving approximation of the signal-weighted homomorphism densities (Theorem 5.2) is the main difficulty. Fix a compact subset $K \subset (\widetilde{\mathcal{WL}}_r, \delta_p)$. Consider the space of all *graphon-signal motif parameters*, i.e.,

$$\mathcal{D} := \operatorname{span}\{t(F, \boldsymbol{d}, \cdot) \,|\, F \text{ simple graph}, \boldsymbol{d} \in \mathbb{N}_0^{v(F)}\} \subseteq C(K, \mathbb{R}). \tag{124}$$

Clearly, $\mathcal{D}$ is a linear subspace, and $\mathcal{D}$ contains a non-zero constant function as we can take a homomorphism density of a graph F with no edges and $\boldsymbol{d} = 0$. Also, it is straightforward to see that $\mathcal{D}$ is a subalgebra, as for any two simple graphs F_1, F_2, $\boldsymbol{d}_1 \in \mathbb{N}_0^{v(F_1)}$, $\boldsymbol{d}_2 \in \mathbb{N}_0^{v(F_2)}$,

$$t(F_1, \boldsymbol{d}_1, \cdot) \cdot t(F_2, \boldsymbol{d}_2, \cdot) = t(F_1 \sqcup F_2, \boldsymbol{d}_1 \| \boldsymbol{d}_2, \cdot) \in \mathcal{D}, \tag{125}$$

i.e., the product of homomorphism densities with respect to two simple graphs can be rewritten as the homomorphism density with respect to their disjoint union. By Theorem 3.3, $\mathcal{D}$ also separates points, and we can apply Stone-Weierstrass (note that K is a metric space and, thus, particularly Hausdorff) to conclude that $\mathcal{D} \subseteq C(K, \mathbb{R})$ is dense. However, by Theorem 5.2, any element of $\mathcal{D}$ can be approximated with arbitrary precision by $\mathcal{IWN}_{\varrho}$, and thus

$$C(K, \mathbb{R}) = \overline{\mathcal{D}} \subseteq \overline{\mathcal{IWN}_{\varrho}} \subseteq C(K, \mathbb{R}). \tag{126}$$

This concludes the proof. □

E.5 PROOF OF PROPOSITION 5.5

Proposition 5.5. *Let $\varrho : [0,1] \to \mathbb{R}$. Then, the assignment $\mathcal{W}_0 \ni W \mapsto \varrho(W) \in \mathcal{W}$, where ϱ is applied pointwise, is continuous w.r.t. $\|\cdot\|_\square$ if and only if ϱ is linear.*

Proof of Proposition 5.5. We will show that the assignment

$$\mathcal{W}_0 \ni W \mapsto \varrho(W) \in \mathcal{W}, \tag{127}$$

where ϱ is applied pointwise, is continuous if and only if ϱ is linear. First, note that $\mathcal{W} \in W \mapsto \int_{[0,1]^2} W \, d\lambda^2$ is linear and continuous with respect to $\|\cdot\|_\square$, since

$$\left| \int_{[0,1]^2} W \, d\lambda^2 \right| \le \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W \, d\lambda^2 \right| = \|W\|_\square. \tag{128}$$

Let $\varrho : [0,1] \to \mathbb{R}$ such that $W \mapsto \varrho(W)$ is continuous and let $p \in [0,1]$. Then, also $W \mapsto \int_{[0,1]^2} \varrho(W) \, d\lambda^2$ is continuous. Let $p \in (0,1)$ and set $W_p := p$ to be a constant graphon. If we sample $G_p^{(n)} \sim \mathbb{G}_n(W_p)$, i.e., from an Erdős–Rényi model with edge probability p, $G_p^{(n)} \to W_p$ in the cut norm almost surely. But

$$\int_{[0,1]^2} \varrho(W_{G_p^{(n)}}) \, d\lambda^2 \to p \cdot \varrho(1) + (1-p) \cdot \varrho(0), \tag{129}$$

while $\int_{[0,1]^2} \varrho(W_p) \, d\lambda^2 = \varrho(p)$. This implies

$$\forall p \in (0,1): \ \varrho(p) = p \cdot \varrho(1) + (1-p) \cdot \varrho(0), \tag{130}$$

i.e., ϱ is a linear function. It is trivial to check that if $\varrho(x) = ax + b$ is a linear function, $W \mapsto \varrho(W)$ is indeed continuous. ☐

E.6 PROOF OF THEOREM 5.6

Theorem 5.6 (Transferability of IWNs). *Let $\varepsilon > 0$ and $r > 1$. Let $\varrho : \mathbb{R} \to \mathbb{R}$ be continuous. Let $\mathcal{N} \in \mathcal{IWN}_\varrho$. Then, there exists a constant $C_{\varepsilon,\mathcal{N}} > 0$ such that for any $(W, f) \in \mathcal{WL}_r$ and $(G_n, \boldsymbol{f}_n), (G_m, \boldsymbol{f}_m) \sim \mathbb{G}_n(W, f), \mathbb{G}_m(W, f)$,*

$$\mathbb{E} \left| \mathcal{N}(G_n, \boldsymbol{f}_n) - \mathcal{N}(G_m, \boldsymbol{f}_m) \right| \le C_{\varepsilon,\mathcal{N}} \left(\frac{1}{\sqrt{\log n}} + \frac{1}{\sqrt{\log m}} \right) + \varepsilon. \tag{22}$$

Proof of Theorem 5.6. Let $p : \mathbb{R} \to \mathbb{R}$ be a polynomial such that the IWN $\mathcal{N}_p$ which is obtained from $\mathcal{N}$ by replacing each occurrence of ϱ with p fulfills

$$\|\mathcal{N}_p - \mathcal{N}\|_{\infty, \mathcal{WL}_r} := \sup_{(W,f) \in \mathcal{WL}_r} |\mathcal{N}_p(W, f) - \mathcal{N}(W, f)| \le \varepsilon/2. \tag{131}$$

Such a p exists: We can approximate $\varrho : \mathbb{R} \to \mathbb{R}$ uniformly arbitrarily well on compact subsets of $\mathbb{R}$ by the standard Weierstrass theorem, and, as the domain of $\mathcal{N}$ only contains bounded functions W and f, for *any* input $(W, f) \in \mathcal{WL}_r$, ϱ is only ever considered on some fixed bounded set (which depends on the model parameters). The argument is the same as switching activation functions, as done on multiple occasions in § E.2.2.

Now, observe that $\mathcal{N}_p$ can be reduced to an integral over $[0,1]^n$ of a polynomial in the variables $W(x_i, x_j)$ and $f(x_k)$, for $i, j, k \in [n]$ and some $n \in \mathbb{N}$. This essentially means that $\mathcal{N}_p$ is a linear combination of signal-weighted homomorphism densities. Thus, there are finite collections of $\{\alpha_i\}_i \in \mathbb{R}$, multigraphs $\{F_i\}_i$, and exponents $\{\boldsymbol{d}_i\}_i$, $\boldsymbol{d}_i \in \mathbb{N}_0^{v(F_i)}$, such that

$$\mathcal{N}_p(W, f) = \sum_i \alpha_i \cdot t(F_i, \boldsymbol{d}_i, (W, f)) \tag{132}$$

for any $(W, f) \in \mathcal{WL}_r$. Set

$$\widetilde{\mathcal{N}}_p(W, f) = \sum_i \alpha_i \cdot t(F_i^{\text{simple}}, \boldsymbol{d}_i, (W, f)), \tag{133}$$

where $F_i \mapsto F_i^{\text{simple}}$ removes parallel edges. $\widetilde{\mathcal{N}}_p$ is Lipschitz continuous in the cut distance by Lemma 3.2 and, crucially, agrees with $\mathcal{N}_p$ on 0-1-valued graphons (since any monomial $x \mapsto x^d$ has 0 and 1 as fixed points). Let $M > 0$ denote the $\delta_\square$-Lipschitz constant of $\widetilde{\mathcal{N}}_p$.

Now, consider $(G_n, \boldsymbol{f}_n), (G_m, \boldsymbol{f}_m) \sim \mathbb{G}_n(W, f), \mathbb{G}_m(W, f)$. By the graphon-signal sampling lemma (Levie (2023); Eq. 8), we can bound

$$
\begin{aligned}
&\mathbb{E}\big[|\mathcal{N}(G_n, \boldsymbol{f}_n) - \mathcal{N}(G_m, \boldsymbol{f}_m)|\big] && (134)\\
&\leq \underbrace{\mathbb{E}\big[|\mathcal{N}(G_n, \boldsymbol{f}_n) - \mathcal{N}_p(G_n, \boldsymbol{f}_n)|\big]}_{\leq \varepsilon/2} + \mathbb{E}\big[|\widetilde{\mathcal{N}}_p(G_n, \boldsymbol{f}_n) - \widetilde{\mathcal{N}}_p(G_m, \boldsymbol{f}_m)|\big] && (135)\\
&\quad + \underbrace{\mathbb{E}\big[|\mathcal{N}_p(G_m, \boldsymbol{f}_m) - \mathcal{N}(G_m, \boldsymbol{f}_m)|\big]}_{\leq \varepsilon/2} && (136)\\
&\leq \varepsilon + M \cdot \mathbb{E}\left[\delta_\square((G_n, \boldsymbol{f}_n), (G_m, \boldsymbol{f}_m))\right] && (137)\\
&\leq \varepsilon + M \cdot \Big(\mathbb{E}\big[\delta_\square((G_n, \boldsymbol{f}_n), (W, f))\big] + \mathbb{E}\big[\delta_\square((W, f), (G_m, \boldsymbol{f}_m))\big]\Big) && (138)\\
&\overset{(*)}{\leq} \varepsilon + M \left(\frac{15}{\sqrt{\log n}} + \frac{15}{\sqrt{\log m}}\right) \leq \varepsilon + \underbrace{15M}_{=:C_{\varepsilon,\mathcal{N}}} \left(\frac{1}{\sqrt{\log n}} + \frac{1}{\sqrt{\log m}}\right), && (139)
\end{aligned}
$$

where the sampling lemma was used in $(*)$. This completes the proof. □

F Asymptotic Dimension Analysis of $\mathrm{LE}_{k\to\ell}^{[0,1]}$

In this section, we briefly analyze the asymptotic differences in dimension between $\mathrm{LE}_{k\to\ell}^{[n]}$, the linear equivariant layer space of discrete IGNs, and $\mathrm{LE}_{k\to\ell} = \mathrm{LE}_{k\to\ell}^{[0,1]}$, of IWNs.

Recall that

$$\dim \mathrm{LE}_{k\to\ell}^{[n]} = \mathrm{bell}(k+\ell), \tag{140}$$

$$\dim \mathrm{LE}_{k\to\ell}^{[0,1]} = \sum_{s=0}^{\min\{k,\ell\}} s!\binom{k}{s}\binom{\ell}{s}. \tag{141}$$

For a comparison of the dimensions for the first few pairs (k, ℓ), see Table 2.

Table 2: Dimensions of $\mathrm{LE}_{k\to\ell}^{[n]}$ and $\mathrm{LE}_{k\to\ell}^{[0,1]}$.

$\dim \mathrm{LE}_{k\to\ell}^{[n]}$	**0**	**1**	**2**	**3**	**4**
0	1	1	2	5	15
1	1	2	5	15	52
2	2	5	15	52	203
3	5	15	52	203	877
4	15	52	203	877	4140

$\dim \mathrm{LE}_{k\to\ell}^{[0,1]}$	**0**	**1**	**2**	**3**	**4**
0	1	1	1	1	1
1	1	2	3	4	5
2	1	3	7	13	21
3	1	4	13	34	73
4	1	5	21	73	209

The case of bounded k or ℓ. Immediately visible from Table 2 is the vastly different behavior of the two expressions as long as one of the variables k, ℓ is bounded: In the discrete case, whenever $k \to \infty$ or $\ell \to \infty$, we have $\mathrm{bell}(k+\ell) \to \infty$ superexponentially. However, for the case of $[0,1]$, suppose w.l.o.g. that only $k \to \infty$ and $\ell = \mathcal{O}(1)$ remains constant. Then, the corresponding dimension growth is bounded by

$$\dim \mathrm{LE}_{k\to\ell}^{[0,1]} = \dim \mathrm{LE}_{\ell\to k}^{[0,1]} = \mathcal{O}(k^\ell), \tag{142}$$

as Eq. 141 is dominated by $\binom{k}{\ell}$ in this case.

The case of $k \sim \ell$. We will now consider the worst case, i.e., when k grows roughly as fast as ℓ. For simplicity, assume $k = \ell$, and thus

$$\dim \mathrm{LE}_{k\to k}^{[n]} = \mathrm{bell}(2k), \quad \dim \mathrm{LE}_{k\to k}^{[0,1]} = \sum_{s=0}^{k} s!\binom{k}{s}^2. \tag{143}$$

The bell numbers grow superexponentially, as can be seen by one of its asymptotic formulas (e.g., refer to Weisstein, Equation 19):

$$\mathrm{bell}(n) \sim \frac{1}{\sqrt{n}}\left(\frac{n}{W(n)}\right)^{n+1/2} \exp\left(\frac{n}{W(n)} - n - 1\right), \tag{144}$$

where W denotes the Lambert W-function, i.e., the inverse of $x \mapsto x\exp(x)$, or a simpler characterization due to Grunwald & Serafin (2024, Proposition 4.7), which is not strictly asymptotically correct but suffices in our case:

$$\left(\frac{1}{e}\frac{n}{\log n}\right)^n \le \mathrm{bell}(n) \le \left(\frac{3}{4}\frac{n}{\log n}\right)^n, \tag{145}$$

as long as $n \ge 2$. Therefore, the dimension of linear equivariant layers in the discrete case can be bounded as

$$\dim \mathrm{LE}_{k\to k}^{[n]} \ge \left(\frac{1}{e}\frac{2k}{\log 2k}\right)^{2k}. \tag{146}$$

We will now provide bounds on the dimension in the continuous case. First note that by only considering the last addend,

$$\dim \mathrm{LE}_{k\to k}^{[0,1]} \ge k! \ge \mathrm{bell}(k) \tag{147}$$

still grows superexponentially. A well-known bound on the factorial (see, e.g., Knuth (1997, § 1.2.5, Ex. 24)) is

$$\frac{n^n}{e^{n-1}} \le n! \le \frac{n^{n+1}}{e^{n-1}}, \tag{148}$$

for $n \in \mathbb{N}$. For a rough upper bound on the dimension, we consider just an even tensor order k:

$$\dim \mathrm{LE}_{k\to k}^{[0,1]} = \sum_{s=0}^{k} s! \binom{k}{s}^2 \tag{149}$$

$$\le (k+1)k! \binom{k}{k/2}^2 = (k+1)\frac{k!^3}{(k/2)!^4} \tag{150}$$

$$\overset{Eq.\ 148}{\le} (k+1)\frac{k^{3k+3}}{e^{3k-3}} \frac{e^{2k-4}}{(k/2)^{2k}} = \frac{1}{e}(k+1)k^3 \left(\frac{4}{e}k\right)^k, \tag{151}$$

which still grows significantly slower than Eq. 146.

G Experiment Details

In this section, we provide details for the toy experiments shown in Figure 1, Figure 2, and Figure 3. The purpose of these experiments is to empirically validate the findings from § 5.2 regarding the cut distance discontinuity and the transferability of IWNs.

Data. We keep the signal fixed at a constant value of 1 and look at the following 4 different graphons:

- **Erdős–Rényi (ER):** $W := 1/2$ constant.
- **Stochastic Block Model (SBM):** We take 5 blocks, each with intra-cluster edge probabilities of $p = 0.8$ and inter-cluster edge probabilites $q = 0.3$.
- **Triangular:** Here, $W(x, y) = (x + y)/2$ (the sets $\{W \geq z\}, z \in [0, 1]$ are triangles).
- **Narrow:** $W(x, y) = \exp\left(\sin\left(\frac{(x-y)^2}{\gamma}\right)^2\right)$, with $\gamma := 0.05$.

From each of these graphons, we sample 100 simple graphs of each of the sizes $\{200, 400, 600, 800, 1000\}$. We use a *weighted* graph of 1000 nodes sampled from each graphon as an approximation of the respective graphon itself.

Models. The following two models are compared:

- **MPNN**: A standard MPNN with mean aggregation over the *entire* node set, as used in the analyses of Levie (2023); Böker et al. (2023).
- **2-IWN**: An IWN of order 2, i.e., with a basis dimension of 7.

For both models, we use a simple setup of 2 layers, a hidden dimension of 16, and the sigmoid function as activation.

Experiment. In Figure 2, we plot the absolute errors of the model outputs for the sampled simple graphs in comparison to their graphon limits. Due to the cut distance continuity of MPNNs and the sampling lemma (Levie, 2023, Theorem 4.3), the MPNN outputs decrease as the graph size grows. While the convergence is slow, it is still significantly faster than the worst-case bound. This is expected as the considered graphons are fairly regular, and the convergence rates can be improved under additional regularity assumptions (Le & Jegelka, 2024; Ruiz et al., 2023). The IWN, however, is discontinuous in the cut distance (Proposition 5.5) and, as such, the errors do not decrease. In Figure 3, we further plot the difference between the 0.95 and 0.05 quantiles of the output distributions on simple graphs for each of the considered sizes. Notably, there are only minor differences visible between the MPNN and the IWN. This validates Theorem 5.6 and suggests that IWNs have similar transferability properties as MPNNs (also beyond the worst-case bound), and δ_p-continuity suffices for transferability.

Code. An anonymous version of the code used for the experiments is provided under `https://anonymous.4open.science/r/Higher-Order-WNNs-950F`.