Invariant Graphon Networks: Approximation and Cut Distance

Editors: List of editors' names

Abstract

Graph limit models, like graphons for limits of dense graphs, have recently been used 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 (Maron et al., 2018). 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 universal approximation for graphon-signals in $\mathcal{L}^p$ distances using *signal-weighted homomorphism densities*. In contrast to the work of Cai and Wang (2022), our results reveal stronger expressivity and better align with graphon space geometry. We also highlight that, unlike other architectures such as MPNNs, IWNs are *discontinuous with respect to cut distance*. Yet, their transferability remains comparable to MPNNs.

Keywords: Graph neural networks, invariant graph networks, universal approximation, graph limits, graphons, transferability, homomorphism densities, machine learning theory.

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 weather prediction (Lam et al., 2023) or materials discovery (Merchant et al., 2023). *Size transferability*—whether a trained GNN generalizes to larger graphs than those in the training set—has recently gained attention for message passing GNNs (MPNNs). Here, graphs are typically assumed to be 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; Le and Jegelka, 2024)—specifically *graphons* for dense graphs, which extend graphs to continuous node sets.

Beyond MPNNs, *Invariant* and *Equivariant Graph Networks* (IGNs/EGNs) (Maron et al., 2018) are another powerful GNN architecture, in which adjacency matrices and node signals are processed through tensor operations that maintain permutation equivariance. IGNs and EGNs have been demonstrated to universally approximate any permutation in- or equivariant graph function (Maron et al., 2019; Keriven and Peyré, 2019).

In this paper, we extend IGNs to graphon-signals (Levie, 2023), introducing *Invariant Graphon Networks (IWNs)*. The closest prior work by Cai and Wang (2022) examines the convergence of IGNs using a *partition norm*, a vector of norms over all diagonals of a graphon. However, these diagonals are null sets, making this approach incompatible with the standard geometry of graphon space. They also introduce a reduced model class *IGN-small* which enables convergence under edge probability estimation. In contrast, we take the different viewpoint of restricting linear equivariant layers to be bounded operators, which yields a subset of IGN-small. While Cai and Wang (2022) only demonstrate that spectral

GNNs can be approximated using their restriction, we establish stronger universal approximation results for graphon-signals (Levie, 2023) in $\mathcal{L}^p$ distances, leveraging an extension of homomorphism densities to graphon-signals which might be of independent interest. We also point out that unlike MPNNs, IWNs are discontinuous w.r.t. cut distance, which renders standard transferability arguments inapplicable. Yet, this discontinuity can be *fixed*, and one can still obtain transferability for IWNs similar to the worst case for MPNNs.

2. Background

This section mainly draws from Lovász (2012); Janson (2013). Also refer to Appendix A.

General Background on Graphons. A *kernel* is 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 $\|W\|_\square := \sup_{S,T\subseteq[0,1]} \left|\int_{S\times T} W \,\mathrm{d}\lambda^2\right|$. Let $S_{[0,1]}$ be the set of measure preserving bijections on $[0,1]$, and $W^\varphi(x,y) := W(\varphi(x),\varphi(y))$ for $\varphi \in S_{[0,1]}$. The *cut distance* between two graphons is $\delta_\square(U,W) := \inf_{\varphi\in S_{[0,1]}} \|U - W^\varphi\|_\square$. Analogously, we can define distances δ_p on graphons based on $\mathcal{L}^p$ norms; $\delta_\square \leq \delta_p$. Among these, 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 usefulness of $\delta_\square$ over any δ_p lies in the fact that $(\widetilde{\mathcal{W}}_0, \delta_\square)$ forms a *compact space*. Any labeled graph G can be identified with its *induced step graphon* W_G, and finite graphs are dense in the graphon space. 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 simple edges are 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 a.s. $\delta_\square(\mathbb{H}_n(W), W) \leq \delta_1(\mathbb{H}_n(W), W) \to 0$. Also $\delta_\square(\mathbb{G}_n(W), W) \to 0$, but this does not hold for δ_1. *Homomorphism densities* are defined as $t(F,W) := \int_{[0,1]^k} \prod_{\{i,j\}\in E(F)} W(x_i,x_j)\,\mathrm{d}\lambda^k(\boldsymbol{x})$ for a simple graph F with $V(F) = [k]$. Notably, $\delta_\square(W_n, W) \to 0$ if and only if $t(F, W_n) \to t(F,W)$ for all simple graphs F.

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 $\boldsymbol{f} \in \mathbb{R}^{n\times k}$. Levie (2023) extends this to graphons. They fix $r > 0$, consider signals in $\mathcal{L}_r^\infty[0,1] := \{f \in \mathcal{L}^\infty[0,1] \mid \|f\|_\infty \leq r\}$, and set $\|f\|_\square := \sup_{S\subseteq[0,1]} \left|\int_S f\,\mathrm{d}\lambda\right|$. They then let $\mathcal{WL}_r := \mathcal{W}_0 \times \mathcal{L}_r^\infty[0,1]$ and define the *cut norm* $\|(W,f)\|_\square := \|W\|_\square + \|f\|_\square$. Define $\delta_\square$ 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 $(\mathbb{G}(W,\boldsymbol{X}), f(\boldsymbol{X}))$, $\boldsymbol{X} \sim U(0,1)^n$. Crucially, Levie (2023) proves compactness of the graphon-signal space, which is then used for a sampling lemma.

3. Signal-Weighted Homomorphism Densities

Homomorphism densities are analogous to moments of a real random variable for W-random graphs in the sense that they fix the distribution of $\mathbb{H}_n(W)$ and $\mathbb{G}_n(W)$ (Zhao, 2023). We introduce an extension of homomorphism densities to graphon-signals: Let F be a multigraph with nodes $V(F)$, edges $E(F)$, $k := |V(F)|$, $\boldsymbol{d} \in \mathbb{N}_0^k$, and $(W,f) \in \mathcal{WL}_r$. We set

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

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

As a first step, we show a counting lemma similar to the standard graphon case (see Appendix B.1). The following theorem in the style of Theorem 8.10 from Janson (2013) and corollary justify our definition, as they show that signal-weighted homomorphism densities fix the distribution of W-random graph-signals and characterize convergence, similar as homomorphism densities do for graphons. Refer to Appendix B for the proofs.

Theorem 1 *Fix $r > 1$ and let $(W, f), (V, g) \in \mathcal{WL}_r$. Then, the following 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 simple graphs F, $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$.
(4) $\mathbb{G}_k(W, f) \stackrel{\mathcal{D}}{=} \mathbb{G}_k(V, g)$ for all $k \in \mathbb{N}$.

Corollary 2 *For $(W_n, f_n)_n, (W, f) \in \mathcal{WL}_r$ and $r > 1$, $\delta_\square((W_n, f_n), (W, f)) \to 0$ as $n \to \infty$ if and only if $t(F, \boldsymbol{d}, (W_n, f_n)) \to t(F, \boldsymbol{d}, (W, f))$ for all simple graphs F and $\boldsymbol{d} \in \mathbb{N}_0^{|V(F)|}$.*

4. Invariant Graphon Networks

Linear Equivariant Layers. We start by defining the building blocks of IWNs, *linear equivariant layers*, generalizing Maron et al. (2018). 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 $k, l \in \mathbb{N}_0$. 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

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

where $W^\varphi(x_1, \ldots, x_k) := W(\varphi(x_1), \ldots, \varphi(x_k))$ and $\mathcal{B}(\cdot, \cdot)$ denote bounded linear operators.

For $\mathcal{X} := [n]$, we obtain $\mathcal{L}^2([n]) \cong \mathbb{R}^n$ and $\mathrm{LE}_{k\to l}^{[n]}$ can be identified with the space of linear permutation equivariant functions $\mathbb{R}^{n^k} \to \mathbb{R}^{n^l}$, which was studied by Maron et al. (2018). One of their main results is that $\dim \mathrm{LE}_{k\to l}^{[n]} = \mathrm{bell}(k+l)$, with $\mathrm{bell}(m)$ denoting the number of partitions of $[m]$, and there is a canonical basis in which every basis element corresponds to one partition. We prove the following about our space of interest $\mathrm{LE}_{k\to l} := \mathrm{LE}_{k\to l}^{[0,1]}$:

Theorem 3 *Let $k, l \in \mathbb{N}_0$. Then, $\mathrm{LE}_{k\to l}$ is a finite-dimensional vector space of dimension* $\dim \mathrm{LE}_{k\to l} = \sum_{s=0}^{\min\{k,l\}} s!\binom{k}{s}\binom{l}{s} \leq \mathrm{bell}(k + l)$.

The proof (see Appendix C.2) builds on the basis characterization from Cai and Wang (2022) as sequences of basic operations like *selection*, *reduction*, *alignment*, and *replication* (see Appendix C.1), nesting subspaces of step kernels and selecting extensible basis elements. An analysis of the dimensions of $\mathrm{LE}_{k\to l}$ and $\mathrm{LE}_{k\to l}^{[n]}$ can be found in Appendix G.

Invariant Graphon Networks. Using $\mathrm{LE}_{k\to l}$ as building blocks, we extend the definitions of IGNs from Keriven and Peyré (2019) to graphons. Let $\varrho : \mathbb{R} \to \mathbb{R}$ be continuous and non-polynomial. An *Invariant Graphon Network (IWN)* is a function

$$\mathcal{N} : \mathcal{WL}_r \to \mathbb{R}, \quad (W, f) \mapsto \sum_{s=1}^{S} L_s^{(2)}\Big(\varrho(L_s^{(1)}(W) + \widetilde{L}_s^{(1)}(f) + b_s^{(1)})\Big) + b^{(2)}, \tag{3}$$

where $S \in \mathbb{N}_0$, $L_s^{(1)} \in \mathrm{LE}_{2\to k_s}$, $\widetilde{L}_s^{(1)} \in \mathrm{LE}_{1\to k_s}$, $L_s^{(2)} \in \mathrm{LE}_{k_s\to 0}$, $b_s^{(1)}, b^{(2)} \in \mathbb{R}$ for $k_s \in \mathbb{N}$ and $s \in \{1, \dots, S\}$. Application of ϱ and addition of the bias terms are understood elementwise. Note that any IWN is invariant w.r.t. all $\varphi \in \overline{S}_{[0,1]}$. We observe that any IWN can be represented as an instance of IGN-small (Cai and Wang, 2022); refer to Appendix E.

Properties of Invariant Graphon Networks. First, any IWN is Lipschitz continuous w.r.t. all $\mathcal{L}^p$ norms, but *not* w.r.t. cut norm (see Appendix D.1 for the proof):

Theorem 4 (Continuity of IWNs)
(1) Let $\mathcal{N} : \mathcal{WL}_r \to \mathbb{R}$ be an IWN as defined in Equation (3). Then, $\mathcal{N}$ is Lipschitz continuous w.r.t. δ_p for each $p \in [1, \infty]$.
(2) 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.

While discontinuity w.r.t. $\|\cdot\|_\square$ is a severe drawback compared to other universal GNN models, one can still use the Stone-Weierstrass theorem to show that IWNs are universal approximators of continuous functions on compact subsets of $(\widetilde{\mathcal{WL}}_r, \delta_p)$:

Theorem 5 (δ_p-Universality of IWNs) *Let $r > 1$, $p \in [1, \infty)$, $\varrho : \mathbb{R} \to \mathbb{R}$ continuous and non-polynomial, and let $\mathcal{IWN}_\varrho$ be the set of IWNs as defined in Equation (3). 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. $\|\cdot\|_\infty$.*

See Appendix D.2 for a proof. Furthermore, it turns out that the $\delta_\square$-discontinuity (cf. Theorem 4), which causes the non-convergence observed by Cai and Wang (2022), can be "fixed" and IWNs *are* indeed transferable:

Theorem 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 $(G_n, \boldsymbol{f}_n), (G_m, \boldsymbol{f}_m) \sim \mathbb{G}_n(W, f), \mathbb{G}_m(W, f)$,*

$$\mathbb{E}\left|\mathcal{N}(G_n, \boldsymbol{f}_n) - \mathcal{N}(G_m, \boldsymbol{f}_m)\right| \;\le\; C_{\varepsilon,\mathcal{N}}\left(\frac{1}{\sqrt{\log n}} + \frac{1}{\sqrt{\log m}}\right) + \varepsilon. \tag{4}$$

See Appendix D.4 for a proof. While the asymptotics are weak, they agree with the worst case for MPNNs (Levie, 2023), which is expected as we do not impose additional assumptions on the graphon-signal or model.

5. Discussion

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 an analysis of expressivity, continuity, and transferability through $\mathcal{L}^p$ and cut distances on graphons. Significantly extending Cai and Wang (2022), we demonstrate that IWNs, as a subset of their class IGN-small, retain the same expressive power as their discrete counterparts. 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 and Jegelka (2024) do not generalize. We demonstrate that, nevertheless, cut distance discontinuity can be overcome for transferability purposes, and IWNs are provably just as transferable as MPNNs in the worst case.

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].

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*, pages 2457–2484. PMLR, June 2022. URL https://proceedings.mlr.press/v162/cai22b.html. ISSN: 2640-3498.

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].

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.

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

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].

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, pages 6960–6971. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/38181d991caac98be8fb2ecb8bd0f166-Abstract.html.

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].

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].

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

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*, pages 4363–4371. PMLR, May 2019. URL https://proceedings.mlr.press/v97/maron19a.html. ISSN: 2640-3498.

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.

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

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. 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.

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

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.

Appendix A. More Background on Graphon Theory

In this section, we will provide more comprehensive background on graphon theory compared to Section 2. Contents of Appendix A.1 are mostly drawn from Lovász (2012) and Janson (2013), and in Appendix A.2 we summarize key results of Levie (2023).

For $n \in \mathbb{N}$, write $[n] := \{1, \dots, n\}$. 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; $\lambda := \lambda^1$.

A.1. General Background on Graphons

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, such graphons can be obtained by taking the limit of adjacency matrices of sequences of *dense* graphs as the number of nodes grows. Formally, we first define a *kernel* to be a bounded symmetric measurable function $W : [0, 1]^2 \to \mathbb{R}$. We write $\mathcal{W}$ for the space of all kernels and $\mathcal{W}_r := \{W \in \mathcal{W} \,|\, \|W\|_\infty \le r\}$, $\mathcal{W}_r^+ := \{W \in \mathcal{W}_r \,|\, W \ge 0\}$. A *graphon* W is defined as an element of $\mathcal{W}_0 := \mathcal{W}_1^+$.

The most important way to measure distance on the space of graphons is the *cut norm/metric*. This generalizes the cut norm for finite graphs which in some sense measures how far a given graph is from being bipartite. For $W \in \mathcal{W}$, we define its *cut norm*

$$\|W\|_\square := \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W \,\mathrm{d}\lambda^2 \right|, \tag{5}$$

considering measurable S, T, and set $d_\square(U, W) := \|U - W\|_\square$ to be the *cut metric*. Just as in the finite case where the node ordering is arbitrary, we want to identify graphons up to reorderings of $[0,1]$, which is formalized by *measure preserving functions*, 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 measure preserving functions, $S_{[0,1]}$ for the set of measure preserving bijections on $[0,1]$, and $W^{\varphi}(x,y) := W(\varphi(x), \varphi(y))$ for $\varphi \in \overline{S}_{[0,1]}$. The *cut distance* between two graphons $U, W \in \mathcal{W}_0$ is defined as

$$\delta_\square(U,W) := \inf_{\varphi \in S_{[0,1]}} d_\square(U, W^{\varphi}) = \min_{\varphi,\psi \in \overline{S}_{[0,1]}} d_\square(U^{\varphi}, W^{\psi}). \tag{6}$$

In Equation (6), the infimum is only guaranteed to be attained in the last expression. Analogously, we can define distances on graphons based on $\mathcal{L}^p$ norms, which we call d_p and δ_p for $p \in [1,\infty]$. Clearly, $\delta_\square \leq \delta_p$. Also, $\delta_\square(U,W) = 0$ if and only if $\delta_p(U,W) = 0$, but the induced topologies differ. δ_1, which is most commonly used, corresponds to the edit distance on graphs. $\delta_\square$ and δ_p only define pseudometrics on $\mathcal{W}_0$, hence we identify *weakly isomorphic* graphons of cut distance 0 to obtain the space $\widetilde{\mathcal{W}}_0$ of *unlabeled graphons*. The important distinguishing property of the cut distance $\delta_\square$ compared to any δ_p is that $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is a *compact space*.

Homomorphism Densities. Another closely related concept are *homomorphism densities*

$$t(F,W) := \int_{[0,1]^k} \prod_{\{i,j\} \in E(F)} W(x_i, x_j) \, \mathrm{d}\lambda^k(\boldsymbol{x}) \tag{7}$$

for $W \in \mathcal{W}_0$ and a simple graph F (i.e., an undirected, unweighted graph without self loops) with $V(F) = [k]$. Notably, homomorphism densities are closely related to cut distance as $\delta_\square(W_n, W) \to 0$ if and only if $t(F, W_n) \to t(F,W)$ for all simple graphs F (the latter being called *left convergence*).

Step Graphons. Any (potentially, but not necessarily weighted) graph G with n vertices labeled $[n]$ and adjacency matrix $\boldsymbol{A} \in [0,1]^{n \times n}$ can be regarded as a graphon by identifying it with its *induced step graphon* W_G: Let $I_j := [\frac{j-1}{n}, \frac{j}{n})$ for $j \in [n-1]$ and $I_n := [\frac{n-1}{n}, 1]$ be a partition of $[0,1]$ into regular intervals, and set

$$W_G := \sum_{j=1}^{n} \sum_{k=1}^{n} A_{jk} \mathbb{1}_{I_j \times I_k} \in \mathcal{W}_0. \tag{8}$$

Finite graphs are dense in the graphon space in the sense that for any $W \in \mathcal{W}_0$, there exists a sequence of labeled graphs $(G_n)_n$ such that $\|W_{G_n} - W\|_\square \to 0$ for $n \to \infty$. Note that, however, graphons are only a suitable limit model for *dense* graphs, as for $(G_n)_n$ with $e(G_n) = o(n^2)$, $\|W_{G_n}\| \to 0$.

***W*-random graphs.** Graphons can also be seen as generative models for random graphs in the sense that we sample a random graph of size n from W by drawing n i.i.d. random variables $\mathbf{X} = (X_1, \ldots, X_n) \sim U(0,1)$ from the uniform distribution on the unit interval, and letting $W(\boldsymbol{X})$ be the *random weighted graph* with edge weights $W_{ij} = W(X_i, X_j)$. We write $\mathbb{H}_n(W)$ for the distribution of $W(\boldsymbol{X})$. Further sample $A_{ij} \sim \text{Bernoulli}(W(X_i, X_j))$ and take $\boldsymbol{A}$ as the adjacency matrix of the resulting unweighted graph $\mathbb{G}(W, \boldsymbol{X})$. Write $\mathbb{G}_n(W)$ for its distribution.

For $H_n \sim \mathbb{H}_n(W)$, we have $\delta_\square(W_{H_n}, W) \le \delta_1(W_{H_n}, W) \to 0$ almost surely. If $G_n \sim \mathbb{G}_n(W)$, then also $\delta_\square(W_{G_n}, W) \to 0$, but this does not hold for δ_1.

A.2. Extension to Graphon-Signals

Most common graph neural networks take a graph-signal $(G, \boldsymbol{f})$ as inputs, i.e., a graph G with node set $[n]$ and a signal $\boldsymbol{f} \in \mathbb{R}^{n\times k}$, meaning that each node is equipped with a feature in $\mathbb{R}^k$. Levie (2023) extends this definition to graphons and introduces a *graphon-signal theory* that shows many parallels to classical graphon theory.

Graphon-Signal Space. Formally, Levie (2023) fixes $r > 0$ and considers signals in $\mathcal{L}_r^\infty[0,1] := \{f \in \mathcal{L}^\infty[0,1] \mid \|f\|_\infty \le r\}$. They then define the *graphon-signal space* as $\mathcal{WL}_r := \mathcal{W}_0 \times \mathcal{L}_r^\infty[0,1]$. The *cut norm* of a signal f is

$$\|f\|_\square := \sup_{S\subseteq[0,1]} \left| \int_S f \,\mathrm{d}\lambda \right|, \tag{9}$$

where S is measurable. The *graphon-signal cut norm* for $(W, f) \in \mathcal{WL}_r$ is $\|(W, f)\|_\square := \|W\|_\square + \|f\|_\square$ and the *cut metric* is $d_\square((W, f), (V, g) := \|(W, f) - (V, g)\|_\square$. As a technical detail for the cut distance, write $S'_{[0,1]} := \{\varphi : A \to B \mid A, B \text{ co-null in } [0,1]\}$, where φ is a measure preserving bijection and $A, B, S \subseteq [0,1]$ measurable. For (W, f), $(V, g) \in \mathcal{WL}_r$ set

$$\delta_\square\left((W, f), (V, g)\right) := \inf_{\varphi\in S'_{[0,1]}} d_\square\left((W, f), (V, g)^\varphi\right), \tag{10}$$

where $(V, g)^\varphi = (V^\varphi, g^\varphi)$. Just as in the standard graphon case, identify graphon-signals with cut distance zero and write $\widetilde{\mathcal{WL}}_r$ for the resulting quotient space.

Define step graphon-signals and sampling from graphon signals analogously to the standard case, i.e., write $\mathbb{H}_n(W, f)$ for the distribution of the random graph-signal $(W(\boldsymbol{X}), f(\boldsymbol{X}))$ and $\mathbb{G}_n(W, f)$ for the distribution of $(\mathbb{G}(W, \boldsymbol{X}), f(\boldsymbol{X}))$, where $\boldsymbol{X} \sim U(0,1)^n$ and $(W, f) \in \mathcal{WL}_r$.

Properties of Graphon-Signal Space. Levie (2023) proceeds by showing a regularity lemma for graphon-signals similar to the classical case. This is further used to prove compactness of the graphon-signal space:

Theorem 7 *(Levie, 2023, Theorem 3.6) The metric space $(\widetilde{\mathcal{WL}}_r, \delta_\square)$ is compact. Moreover, given $r > 0$ and $c > 1$, for every sufficiently small $\varepsilon > 0$, the space $\widetilde{\mathcal{WL}}_r$ can be covered by $\kappa(\varepsilon) = 2^{k^2}$ balls of radius ε, where $k = \lceil 2^{\frac{9c}{4\varepsilon^2}} \rceil$.*

Having established the compactness, Levie (2023) proves the following sampling lemma:

Theorem 8 *(Levie, 2023, Theorem 3.7, Sampling lemma for graphon-signals) Let $r > 1$. There exists a constant $K_0 > 0$ that depends on r, such that for every $k \ge K_0$, $(W, f) \in \mathcal{WL}_r$, we have*

$$\mathbb{E}\big[\delta_\square((W, f), \mathbb{H}_k(W, f))\big] < \frac{15}{\sqrt{\log(k)}}, \quad \mathbb{E}\big[\delta_\square((W, f), \mathbb{G}_k(W, f))\big] < \frac{15}{\sqrt{\log(k)}}. \tag{11}$$

As MPNNs are Lipschitz continuous w.r.t. $\delta_\square$, these theorems are then used by Levie (2023) for stability of MPNNs w.r.t. graph-signal subsampling and to show transferability and generalization theorems for MPNNs.

Appendix B. Details on Signal-Weighted Homomorphism Densities

In this section, we provide further background and explanations regarding our extension of homomorphism densities to graphon-signal space. We also state the counting lemma mentioned in the main body, and prove Theorem 1 and Corollary 2.

Homomorphism densities are of great importance when analyzing graphon space, as convergence in the cut distance is equivalent to *left convergence* (convergence of all $t(F, \cdot)$), i.e., in other words, the topology induced by cut distance is the initial topology of $\{t(F, \cdot)\}_F$, with F ranging over all simple graphs. Hence, also $\delta_\square(W, V) = 0$ iff $t(F, W) = t(F, V)$ for all F. 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). Further, expressive power of a GNN model can also be judged by its *homomorphism expressivity*, i.e., its ability to calculate homomorphism densities. For example, standard MPNNs can precisely distinguish graphs for which the values of $t(F, \cdot)$ differ if F are trees (Böker et al., 2023) or multigraphs of bounded treewidth for higher-order GNNs (Böker, 2023). Homomorphism densities are the continuous analogue to homomorphism counts for finite graphs, which are related to subgraph counts that also play a major role in GNN expressivity analyses (Zhang et al., 2024).

At first, we restate our definition of the extended homomorphism densities. Let F be a multigraph with $k := v(F)$, $\boldsymbol{d} = (d_1, \ldots, d_k) \in \mathbb{N}_0^k$, and $(W, f) \in \mathcal{WL}_r$. We define the corresponding *signal-weighted homomorphism density* as

$$t(F, \boldsymbol{d}, (W, f)) \;:=\; \int_{[0,1]^k} \left(\prod_{i \in V(F)} f(x_i)^{d_i} \right) \left(\prod_{\{i,j\} \in E(F)} W(x_i, x_j) \right) \mathrm{d}\lambda^k(\boldsymbol{x}). \tag{12}$$

Similarly to the probablistic interpretation of $t(F, W) = \mathbb{P}_{G \sim \mathbb{G}_k(W)} \left(E(G) \supseteq E(F) \right)$ (where F and G are seen as labeled graphs), we can rewrite the definition to

$$t(F, \boldsymbol{d}, (W, f)) \;=\; \mathbb{E}_{(G, \boldsymbol{f}) \sim \mathbb{G}_k(W, f)} \left[\prod_{i \in V(F)} f_i^{d_i} \cdot \mathbb{1}\Big\{ E(G) \supseteq E(F) \Big\} \right]. \tag{13}$$

B.1. Counting Lemma for Signal-Weighted Homomorphism Densities

We proceed by proving Lipschitz continuity of $t(F, \boldsymbol{d}, \cdot)$ w.r.t. $\delta_\square$, i.e., a statement similar to the counting lemma:

Lemma 9 (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 $D := \sum_{i \in V(F)} d_i$,*

$$\Big| t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g)) \Big| \;\leq\; 2r^{D-1} \Big(2r \cdot e(F) \left\| W - V \right\|_\square + D \left\| f - g \right\|_\square \Big). \tag{14}$$

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

$$\begin{aligned}
&\left|t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g))\right| \leq \\
&\quad \left| \int_{[0,1]^k} \left(\prod_{i \in V(F)} f(x_i)^{d_i} \right) \left(\prod_{\{i,j\} \in E(F)} W(x_i, x_j) - \prod_{\{i,j\} \in E(F)} V(x_i, x_j) \right) \mathrm{d}\lambda^k(\boldsymbol{x}) \right| \qquad (15) \\
&+ \left| \int_{[0,1]^k} \left(\prod_{\{i,j\} \in E(F)} V(x_i, x_j) \right) \left(\prod_{i \in V(F)} f(x_i)^{d_i} - \prod_{i \in V(F)} g(x_i)^{d_i} \right) \mathrm{d}\lambda^k(\boldsymbol{x}) \right|. \qquad (16)
\end{aligned}$$

For Equation (15), we set $D := \sum_i d_i$ and observe that for all $\boldsymbol{x} \in [0,1]^k$

$$\frac{1}{r^D} \prod_{i \in V(F)} f(x_i)^{d_i} \in [-1, 1], \qquad (17)$$

and hence similarly to the standard proof of the classical counting lemma (see, e.g., Zhao (2023)) we can bound Equation (15) by $r^D e(F) \|W - V\|_{\square,2} \leq 4r^D e(F) \|W - V\|_{\square}$. In comparison to the standard proof, the usage of $\|\cdot\|_{\square,2}$, an alternative definition of the cut norm, stems from the fact that function values appearing in the integral in Equation (15) (renormalizing by r^D) are not necessarily in $[0,1]$, but $[-1,1]$. See also Equations (4.3), (4.4) in Janson (2013). For Equation (16), we bound the $\mathcal{L}^1$ difference of the terms involving f and g:

$$\begin{aligned}
&\left| \int_{[0,1]^k} \left(\prod_{\{i,j\} \in E(F)} V(x_i, x_j) \right) \left(\prod_{i \in V(F)} f(x_i)^{d_i} - \prod_{i \in V(F)} g(x_i)^{d_i} \right) \mathrm{d}\lambda^k(\boldsymbol{x}) \right| \qquad (18) \\
&\leq \int_{[0,1]^k} \left| \prod_{\{i,j\} \in E(F)} V(x_i, x_j) \right| \left| \prod_{i \in V(F)} f(x_i)^{d_i} - \prod_{i \in V(F)} g(x_i)^{d_i} \right| \mathrm{d}\lambda^k(\boldsymbol{x}) \qquad (19) \\
&\leq \sum_{i \in V(F)} \int_{[0,1]^k} \left| f(x_i)^{d_i} - g(x_i)^{d_i} \right| \left| \prod_{j<i} f(x_i)^{d_i} \prod_{j>i} g(x_i)^{d_i} \right| \mathrm{d}\lambda^k(\boldsymbol{x}) \qquad (20) \\
&\leq \sum_{i \in V(F)} r^{\sum_{j \neq i} d_i} \int_{[0,1]} \left| f(x_i)^{d_i} - g(x_i)^{d_i} \right| \mathrm{d}\lambda(x) \qquad (21) \\
&\overset{(*)}{\leq} \sum_{i \in V(F)} r^{\sum_{j \neq i} d_i} \cdot d_i r^{d_i - 1} \|f - g\|_1 \;=\; D r^{D-1} \|f - g\|_1 \;\leq\; 2 D r^{D-1} \|f - g\|_{\square}, \qquad (22)
\end{aligned}$$

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, we obtain

$$\left|t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g))\right| \;\leq\; 4r^D e(F) \|W - V\|_{\square} + 2Dr^{D-1} \|f - g\|_{\square}, \qquad (23)$$

which yields the claim. ■

B.2. Proof of Theorem 1

To begin, we extend the characterization of $\delta_\square$ from Equation (6) 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).

Two norms $N = (N_1, N_2)$ on $\mathcal{L}^\infty[0,1]$ and $\mathcal{W}$ are called smooth if $f_n \to f \in \mathcal{L}^\infty[0,1]$, $W_n \to W \in \mathcal{W}$ almost everywhere implies $N_1(f_n) \to N_1(f)$ and $N_2(W_n) \to N_2(W)$. They are invariant if $N_1(f^\varphi) = N_1(f)$, $N_2(W^\varphi) = N_2(W)$ for all $\varphi \in \overline{S}_{[0,1]}$, $W \in \mathcal{W}$, $f \in \mathcal{L}^\infty[0,1]$. Both conditions clearly apply to $\|\cdot\|_\square$ (with the one-dimensional definition from Equation (9)) 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

$$\delta_N((W,f),(V,g)) := \inf_{\varphi \in S_{[0,1]}} \big(N_2(W - V^\varphi) + N_1(f - g^\varphi)\big). \tag{24}$$

Lemma 10 (Minima vs. Infima for 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:*

$$\delta_N((W,f),(V,g)) = \inf_{\varphi \in \overline{S}_{[0,1]}} \big(N_2(W - V^\varphi) + N_1(f - g^\varphi)\big) \tag{25}$$

$$= \min_{\varphi,\psi \in \overline{S}_{[0,1]}} \big(N_2(W^\varphi - V^\psi) + N_1(f^\varphi - g^\psi)\big). \tag{26}$$

Proof 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 10 justifies our definition in Equation (24), as $\delta_\square$ on the graphon-signal space was defined 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), see Equation (10). As $S_{[0,1]} \subseteq S'_{[0,1]} \subseteq \overline{S}_{[0,1]}$, this coincides with the definition of $\delta_\square$ in Equation (24) by the first equality of Lemma 10.

Proof [Proof of Theorem 1] For (1) $\Leftrightarrow$ (2), Lemma 10 implies that for any $p \in [1,\infty)$

$$\delta_\square((W,f),(V,g)) = 0 \Leftrightarrow \exists \varphi,\psi \in \overline{S}_{[0,1]} : (W,f)^\varphi = (V,g)^\psi \Leftrightarrow \delta_p((W,f),(V,g)) = 0, \tag{27}$$

where equality in the middle holds in an λ^2-a.s. sense. (2) $\Rightarrow$ (3) follows immediately from Lemma 9. For (3) $\Rightarrow$ (4), 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

$$\mathbb{P}_{(G,\boldsymbol{f})\sim\mathbb{G}_k(W,f)}(G \cong F), \tag{28}$$

$$\mathbb{P}_{(G,\boldsymbol{f})\sim\mathbb{G}_k(W,f)}(\boldsymbol{f} \in \cdot | G \cong F), \tag{29}$$

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 Equation (28), we remark that the standard homomorphism densities w.r.t. 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 Equation (28). With a similar inclusion-exclusion argument, we see that for any F

$$\mathbb{1}\{\mathbb{G}_k(W) \cong F\} \;=\; \sum_{F' \supseteq F} (-1)^{e(F')-e(F)} \mathbb{1}\{\mathbb{G}_k(W) \supseteq F'\} \tag{30}$$

and therefore

$$\mathbb{E}_{(G,\boldsymbol{f})\sim\mathbb{G}_k(W,f)}\left[\prod_{i \in V(F)} f_i^{d_i} \middle| G \cong F\right] \;=\; \frac{\sum_{F' \supseteq F} (-1)^{e(F')-e(F)}\, t(F', \boldsymbol{d}, (W, f))}{\mathbb{P}_{(G,\boldsymbol{f})\sim\mathbb{G}_k(W,f)}(G \cong F)} \tag{31}$$

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 ($\|\boldsymbol{f}\|_\infty \leq r$ a.s.), its distribution is uniquely determined by its multidimensional moments, i.e., precisely the expressions from Equation (31) (in the case of more general random variables/vectors, this is known as the *moment problem*, but under boundedness it is trivial and can, e.g., be proven via characteristic functions). Thus, we can conclude $\mathbb{G}_k(W, f) \overset{\mathcal{D}}{=} \mathbb{G}_k(V, g)$ for all $k \in \mathbb{N}$. Finally, (4) $\Rightarrow$ (2) is a straightforward application of Theorem 8: If (4) holds, we can bound

$$\delta_\square((W, f), (V, g)) \;\leq\; \mathbb{E}\big[\delta_\square((W, f), \mathbb{G}_k(W, f))\big] + \mathbb{E}\big[\delta_\square((V, g), \mathbb{G}_k(W, f))\big] \tag{32}$$

$$= \mathbb{E}\big[\delta_\square((W, f), \mathbb{G}_k(W, f))\big] + \mathbb{E}\big[\delta_\square((V, g), \mathbb{G}_k(V, g))\big] \;\to\; 0 \tag{33}$$

as $k \to \infty$. ∎

B.3. Proof of Corollary 2

Proof The proof idea is essentially the same as in the classical graphon case (for example, see Section 4.9 in Zhao (2023)): An application of Theorem 1 that uses compactness of the graphon-signal space. For the sake of completeness, we restate the argument.

"$\Rightarrow$" follows immediately from the counting lemma (Lemma 9). For "$\Leftarrow$", let $(W_n, f_n)_n$ be a sequence of graphon-signals that left-converges to $(W, f) \in \mathcal{WL}_r$. By compactness (Theorem 7), 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

$$t(F, \boldsymbol{d}, (W, f)) \;=\; t(F, \boldsymbol{d}, (V, g)) \quad \forall F, \boldsymbol{d} \in \mathbb{N}_0^{v(F)}. \tag{34}$$

Theorem 1 yields $\delta_\square((W, f), (V, g)) = 0$, i.e., also $(W_n, f_n) \to (W, f)$ in cut distance. ∎

B.4. Related Definitions in the Literature

In the discrete setting, homomorphisms between graphs with signals are typically defined as follows: For $(F, \boldsymbol{f})$ and $(G, \boldsymbol{g})$, $h : V(F) \to V(G)$ is a homomorphism if $f_i = g_{h(i)}$ for all $i \in V(F)$ and $\{h(i), h(j)\} \in E(G)$ for all $\{i, j\} \in E(F)$. This concept does not extend to graphon-signals, as for a signal $f \in \mathcal{L}^\infty[0, 1]$, its level sets $\{f = \alpha\}$ for $\alpha \in \mathbb{R}$ might *all* be sets of Lebesgue measure zero. In contrast to other common approaches in the GNN literature, only considering $\boldsymbol{d} = \mathbf{1}$ 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 and 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.

Appendix C. Details on Linear Equivariant Layers

C.1. Characterization of the IGN Basis

In this section, we restate the characterization of the IGN basis introduced by Cai and Wang (2022). As described in the original IGN paper by Maron et al. (2018), $\dim \mathrm{LE}_{k\to l}^{[n]} = \mathrm{bell}(k+l)$, i.e., the number of partitions Γ_{k+l} of the set $[k+l]$. In the basis of Cai and Wang (2022), each basis element $L_\gamma^{(n)}$ associated with a partition $\gamma \in \Gamma_{k+l}$ can be characterized as a sequence of basic operations. First, divide γ into 3 subsets $\gamma_1 := \{A \in \gamma \,|\, A \subseteq [k]\}$, $\gamma_2 := \{A \in \gamma \,|\, A \subseteq k + [l]\}$, $\gamma_3 := \gamma \setminus (\gamma_1 \cup \gamma_2)$. Here, the numbers $1, \dots, k$ are associated with the input axes and $k + 1, \dots, k + l$ with the output axes respectively.

① (**Selection**: $W \mapsto W_\gamma$). In a first step, we specify which part of the input tensor $W \in \mathbb{R}^{n^k}$ 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 W_γ by selecting the diagonal of the k-tensor W corresponding with the partition $\gamma\big|_{[k]}$.

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

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

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

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

C.2. Proof of Theorem 3

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

Lemma 11 (Fixed Points of Measure Preserving Functions) *If $k \in \mathbb{N}_0$ and $W \in \mathcal{L}^2[0,1]^k$ such that $W^{\varphi} = W$ for all $\varphi \in \overline{S}_{[0,1]}$, then W 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 $W \in \mathcal{L}^2[0,1]^k$ such that W is invariant under all measure preserving functions, and suppose that W is not constant λ^k-almost everywhere. Then, there exist $a < b$ such that $A := W^{-1}((-\infty, a])$, $B := W^{-1}([b, \infty))$ 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 $\mathcal{P}_k^{(n)} := \{I_{j_1}^{(n)} \times \dots \times I_{j_k}^{(n)} \,|\, j_1, \dots, j_k \in \{1, \dots, n\}\}$. First, note that we have

$$\lambda^k(Q \cap A) \neq \alpha \lambda^k(Q \cap B) \tag{35}$$

for some $m \in \mathbb{N}$ and $Q \in \mathcal{P}_k^{(m)}$. Otherwise, equality in Equation (35) 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 Equation (35). As

$$\sum_{S \in \mathcal{P}_k^{(m)}} \lambda^k(S \cap A) \,=\, \lambda^k(A) \,=\, \alpha\lambda^k(B) \,=\, \sum_{S \in \mathcal{P}_k^{(m)}} \alpha\lambda^k(S \cap B), \tag{36}$$

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

$$\Delta_k \,:=\, \{\boldsymbol{x} \in [0,1]^k \,|\, |\{x_1, \dots, x_k\}| < k\}, \quad \Delta_k^{(n)} \,:=\, \{Q \in \mathcal{P}_k^{(n)} \,|\, Q \cap \Delta \neq \emptyset\} \tag{37}$$

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

$$\lambda^k(Q^* \cap A) > \alpha\lambda^k(Q^* \cap B), \quad \lambda^k(R^* \cap A) < \alpha\lambda^k(R^* \cap B). \tag{38}$$

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 W under all measure-preserving functions, we get

$$\lambda^k(R^* \cap A) \,=\, \lambda^k\left((\varphi^{\otimes k})^{-1}(R^* \cap A)\right) \,=\, \lambda^k(Q^* \cap A), \tag{39}$$

$$\lambda^k(R^* \cap B) \,=\, \lambda^k\left((\varphi^{\otimes k})^{-1}(R^* \cap B)\right) \,=\, \lambda^k(Q^* \cap B), \tag{40}$$

which contradicts Equation (38). Hence, W must be λ^k-a.e. constant. ∎

Lemma 12 (Invariance under Discretization) *Let $k, l \in \mathbb{N}_0$ and $n \in \mathbb{N}$. Let $I_j^{(n)} := [\frac{j-1}{n}, \frac{j}{n})$ for $j \in \{1, \ldots, n-1\}$ and $I_n^{(n)} := [\frac{n-1}{n}, 1]$ be a partition of $[0,1]$ into regular intervals. Let $\mathcal{A}_n := \sigma\left(\{I_1^{(n)}, \ldots, I_n^{(n)}\}\right)$ denote the σ-algebra generated by this partition and let $\mathcal{F}_k^{(n)} := \{W \in \mathcal{L}^2[0,1]^k \,|\, W$ is $\mathcal{A}_n^{\otimes k}$-measurable$\}$. Then, any $L \in \mathrm{LE}_{k\to l}$ is* invariant under discretization*, which means that*

$$L(\mathcal{F}_k^{(n)}) \subseteq \mathcal{F}_l^{(n)}, \tag{41}$$

where the inclusion should be understood up to sets of measure zero.

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

$$\varphi(I_j^{(n)}) \subseteq I_j^{(n)} \tag{42}$$

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

$$\varphi^*(x) := x \operatorname{div} 1/n + \varphi\left(x \bmod 1/n\right), \tag{43}$$

which clearly satisfies Equation (42), and thus $L(W)^{\varphi^*} = L(W)$ almost everywhere. Since now

$$L(W)\big|_Q = L(W)^{\varphi^*}\Big|_Q = \left(L(W)\big|_Q\right)^\varphi, \tag{44}$$

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 11 that $L(W)\big|_Q$ is constant λ^l-almost everywhere. As Q was chosen arbitrarily, this implies the statement of the lemma. ■

Proof [Proof of Theorem 3] Let $n \in N$ and $L \in \mathrm{LE}_{k\to l}$. By Lemma 12, we know that $L(\mathcal{F}_k^{(n)}) \subseteq \mathcal{F}_l^{(n)}$. Since $\mathcal{F}_k^{(n)} \cong (\mathbb{R}^n)^{\otimes k} \cong \mathbb{R}^{n^k}$, we can regard $L\big|_{\mathcal{F}_k^{(n)}} : \mathcal{F}_k^{(n)} \to \mathcal{F}_l^{(n)}$ as a linear operator $\mathbb{R}^{n^k} \to \mathbb{R}^{n^l}$. 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 Appendix C.1.

Note that for any $n, m \in \mathbb{N}$ we have $\mathcal{F}_k^{(n)}, \mathcal{F}_k^{(m)} \subseteq \mathcal{F}_k^{(nm)}$ 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

$$L_\gamma^{(nm)}\Big|_{\mathcal{F}_k^{(n)}} = L_\gamma^{(n)}, \quad L_\gamma^{(nm)}\Big|_{\mathcal{F}_k^{(m)}} = L_\gamma^{(m)}. \tag{45}$$

Hence, the coefficients of $L\big|_{\mathcal{F}_k^{(n)}}$ w.r.t. the canonical basis $\{L_\gamma^{(n)}\}_{\gamma\in\Gamma_{k+l}}$ 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 $\gamma^* \in \Gamma_{k+l}$ 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 $\gamma_1^* := \{A \in \gamma^* \,|\, A \subseteq [k]\}$, $\gamma_2^* := \{A \in \gamma^* \,|\, A \subseteq k + [l]\}$, $\gamma_3^* := \gamma^* \setminus (\gamma_1^* \cup \gamma_2^*)$. For the constant $W \equiv 1 \in \mathcal{F}_k^{(1)} \subseteq \mathcal{L}^2[0,1]^k$, $L_{\gamma^*}^{(n)}(W) \neq 0$ 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 $|A \cap (k + [l])| \leq 1$ for all $A \in \gamma^*$. Now suppose that the input only depends on a diagonal. Denoting the restriction of the constant $W \equiv 1$ to the diagonal under discretization of $[0,1]$ into n pieces by $W_{\gamma^*}^{(n)}$,

$$L_{\gamma^*}^{(n)}(W_{\gamma^*}^{(n)}) \;=\; L_{\gamma^*}^{(n)}(W) \;=\; L_{\gamma^*}^{(1)}(W) \;\neq\; 0 \tag{46}$$

is constant, but $\|W_{\gamma^*}^{(n)}\|_2 \to 0$ for $n \to \infty$, which contradicts boundedness (i.e., continuity) of the operator $L \in \mathcal{B}(\mathcal{L}^2[0,1]^k, \mathcal{L}^2[0,1]^l)$. 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 $\gamma \in \Gamma_{k+l}$ whose sets $A \in \gamma$ contain at most one element from $[k]$ and $k + [l]$ respectively. In the following Lemma 13 we will check (and generalize) 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 $L_\gamma \in \mathcal{B}(\mathcal{L}^2[0,1]^k, \mathcal{L}^2[0,1]^l)$ 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 w.r.t. $\|\cdot\|_2$, 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 trivial application of the martingale convergence theorem: Considering $[0,1]^k$ with Lebesgue measure λ^k as a probability space and $W \in L^2[0,1]^k$ as a random variable, we have $\mathbb{E}[W|\mathcal{A}_n^{\otimes k}] \in \mathcal{F}_k^{(n)}$. 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 $\mathbb{E}[W|\mathcal{A}_n^{\otimes k}] \to \mathbb{E}[W|\mathcal{B}([0,1]^k)] = W$ in $\mathcal{L}^2$.

It is now straightforward to show that

$$\Big|\, \{\gamma \in \Gamma_{k+l} \,|\, \forall A \in \gamma : |A \cap [k]| \leq 1, |A \cap k + [l]| \leq 1\} \,\Big| \;=\; \sum_{s=0}^{\min\{k,l\}} s! \binom{k}{s}\binom{l}{s}, \tag{47}$$

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

C.3. Continuity of Linear Equivariant Layers

Lemma 13 (Continuity of Linear Equivariant Layers w.r.t. $\|\cdot\|_p$) *Fix $k, l \in \mathbb{N}_0$. Let $L \in \mathrm{LE}_{k\to l}$ and $p \in [1,\infty]$. Then, L can also be regarded as a bounded linear operator $\mathcal{L}^p[0,1]^k \to \mathcal{L}^p[0,1]^l$. Furthermore, all of the canonical basis elements from the proof of Theorem 3 have operator norm $\|L\|_{p\to p} = 1$.*

Proof If suffices to show boundedness of all canonical basis elements. Let $\gamma \in \Gamma_{k+l}$ be a partition corresponding to a basis element in $\mathrm{LE}_{k\to l}$, and suppose that γ contains r sets of size 2 $\{i_1, j_1\}, \dots, \{i_r, j_r\}$ with $i_1, \dots, i_r \in [k]$, $j_1, \dots, j_r \in k + [l]$, and set $A = (i_1, \dots, i_r)$, $B = (j_1, \dots, j_r)$. Then, we can write L as

$$L(W) := \left[[0,1]^l \ni \boldsymbol{y} \mapsto \int_{[0,1]^{k-r}} W(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A}) \, \mathrm{d}\lambda^{k-r}(\boldsymbol{x}_{[k]\setminus A}) \bigg|_{\boldsymbol{x}_A = \boldsymbol{y}_B} \right]. \tag{48}$$

Consider at first $p < \infty$. Clearly, Equation (48) is also well-defined for $W \in \mathcal{L}^p[0,1]^k$ and

$$\begin{aligned}
\|L(W)\|_p^p &= \int_{[0,1]^l} \left| \int_{[0,1]^{k-r}} W(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A}) \, \mathrm{d}\lambda^{k-r}(\boldsymbol{x}_{[k]\setminus A}) \bigg|_{\boldsymbol{x}_A = \boldsymbol{y}_B} \right|^p \mathrm{d}\lambda^l(\boldsymbol{y}) && (49)\\
&\le \int_{[0,1]^l} \int_{[0,1]^{k-r}} \left| W(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A}) \right|^p \, \mathrm{d}\lambda^{k-r}(\boldsymbol{x}_{[k]\setminus A}) \bigg|_{\boldsymbol{x}_A = \boldsymbol{y}_B} \mathrm{d}\lambda^l(\boldsymbol{y}) && (50)\\
&= \int_{[0,1]^r} \int_{[0,1]^{k-r}} \left| W(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A}) \right|^p \, \mathrm{d}\lambda^{k-r}(\boldsymbol{x}_{[k]\setminus A}) \, \mathrm{d}\lambda^l(\boldsymbol{x}_A) = \|W\|_p^p, && (51)
\end{aligned}$$

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

$$\begin{aligned}
\|L(W)\|_\infty &= \operatorname*{ess\,sup}_{\boldsymbol{y} \in [0,1]^l} \left| \int_{[0,1]^{k-r}} W(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A}) \, \mathrm{d}\lambda^{k-r}(\boldsymbol{x}_{[k]\setminus A}) \bigg|_{\boldsymbol{x}_A = \boldsymbol{y}_B} \right| && (52)\\
&\le \operatorname*{ess\,sup}_{\boldsymbol{y} \in [0,1]^l} \int_{[0,1]^{k-r}} \underbrace{\left| W(\boldsymbol{x}_A, \boldsymbol{x}_{[k]\setminus A}) \right|}_{\le \|W\|_\infty \text{ a.e.}} \, \mathrm{d}\lambda^{k-r}(\boldsymbol{x}_{[k]\setminus A}) \Bigg|_{\boldsymbol{x}_A = \boldsymbol{y}_B} && (53)\\
&\le \|W\|_\infty, && (54)
\end{aligned}$$

again with equality for $W \equiv 1$. ■

Appendix D. Details on Invariant Graphon Networks

D.1. Proof of Theorem 4

Proof [Proof of Theorem 4] First show statement (1), i.e., continuity w.r.t. $\|\cdot\|_p$. By Lemma 13, $L_s^{(1)}$, $\widetilde{L}_s^{(1)}$, and $L_s^{(2)}$ for $s \in \{1, \dots, S\}$ are clearly Lipschitz continuous w.r.t. $\|\cdot\|_p$ on the respective input and output spaces. Hence, it suffices to check that $\mathcal{L}^p[0,1]^k \ni W \mapsto \varrho(W)$ is Lipschitz continuous for every $k \in \mathbb{N}_0$, where ϱ is applied elementwise (which can later be applied to the graphon and the signal of $(W, f) \in \mathcal{WL}_r$ individually). For $p < \infty$, we have

$$\|\varrho(U) - \varrho(W)\|_p^p = \int_{[0,1]^k} |\varrho(U) - \varrho(W)|^p \, \mathrm{d}\lambda^k \le \int_{[0,1]^k} C_\varrho^p \, |U - W|^p \, \mathrm{d}\lambda^k = C_\varrho^p \|U - W\|_p^p, \tag{55}$$

and a similar argument shows the claim for $\|\cdot\|_\infty$.

Now we focus on statement (2), i.e., discontinuity of IWNs w.r.t. the cut norm. Specifically, we show that the assignment

$$\mathcal{W}_0 \ni W \;\mapsto\; \varrho(W) \in \mathcal{W}, \tag{56}$$

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

$$\left|\int_{[0,1]^2} W \,\mathrm{d}\lambda^2\right| \;\leq\; \sup_{S,T\subseteq[0,1]} \left|\int_{S\times T} W \,\mathrm{d}\lambda^2\right| \;=\; \|W\|_\square. \tag{57}$$

Let $\varrho : [0,1] \to \mathbb{R}$ such that $W \mapsto \varrho(W)$ is continuous and let $p \in [0,1]$. Then, also $W \mapsto \int_{[0,1]^2} \varrho(W) \,\mathrm{d}\lambda^2$ 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 metric. But

$$\int_{[0,1]^2} \varrho(W_{G_p^{(n)}}) \,\mathrm{d}\lambda^2 \;\to\; p \cdot \varrho(1) + (1-p)\cdot\varrho(0) \tag{58}$$

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

$$\forall p \in (0,1):\; \varrho(p) = p\cdot\varrho(1) + (1-p)\cdot\varrho(0), \tag{59}$$

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. ■

D.2. Proof of Theorem 5

The core idea of the proof is that we can approximate signal-weighted homomorphism densities (Equation (1)) with arbitrary precision, which we discuss in the following lemmas.

Lemma 14 (Approximation of Homomorphism Densities with cos**)** *Let F be a simple graph with $v(F) = k$, and let $\boldsymbol{d} \in \mathbb{N}_0^k$. Then, for any $\varepsilon > 0$ there exists an IWN $\mathcal{N}_{F,\boldsymbol{d},\varepsilon}$ with nonlinearity $\varrho = \cos$ such that*

$$\sup_{(W,f)\in\mathcal{WL}_r} |t(F,\boldsymbol{d},(W,f)) - \mathcal{N}_{F,\boldsymbol{d},\varepsilon}(W,f)| \;\leq\; \varepsilon, \tag{60}$$

i.e., $\mathcal{N}_{F,\boldsymbol{d},\varepsilon}$ uniformly approximates the signal-weighted homomorphism density $t(F,\boldsymbol{d},\cdot)$.

In the proof, we model the homomorphism densities explicitly while keeping track of which linear equivariant layers are being used.

Proof Fix a simple graph F with $v(F) = k$, $\boldsymbol{d} \in \mathbb{N}_0^k$, and $\delta > 0$. Fix an arbitrary graphon-signal $(W,f) \in \mathcal{WL}_r$. Let $\boldsymbol{x} = (x_1,\ldots,x_k) \in [0,1]^k$. Set $m := v(F) + e(F)$, and enumerate $V(F) = \{1,\ldots,v(F)\}$, $E(F) = \{e_1,\ldots,e_{e(F)}\}$. Define

$$\boldsymbol{\gamma} := (f(x_1),\ldots,f(x_{v(F)}), W(\boldsymbol{x}_{e_1}),\ldots,W(\boldsymbol{x}_{e_{e(F)}})) \;\in\; [-r,r]^{v(F)} \times [0,1]^{e(F)}. \tag{61}$$

Let

$$\mathcal{N}_\ell : \mathbb{R} \to \mathbb{R}, \quad \gamma_\ell \mapsto \sum_{j=1}^{L_\ell} c_j^{(\ell)} \cos\left(a_j^{(\ell)} \gamma_\ell + b_j^{(\ell)}\right) + d^{(\ell)} \tag{62}$$

be a neural network with one hidden layer for each $\ell \in [m]$ and cos as nonlinearity, approximating $g_\ell(x) := x^{d_\ell}$ if $\ell \leq v(F)$ on $[-r, r]$ and the identity $g_\ell(x) := x$ if $\ell > v(F)$ on $[0, 1]$, each uniformly with error at most δ. The standard universal approximation theorem ensures that such networks $\mathcal{N}_\ell$ exist. Further, we remark that for the identity, $L_\ell = 1$ can be chosen by differentiability of cos and zooming in and out appropriately at a point of non-vanishing derivative.

We 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

$$\prod_{j=1}^{m} \cos(x_j) \;=\; \frac{1}{2^m} \sum_{\sigma \in \{\pm 1\}^m} \cos\left(\sum_{j=1}^{m} \sigma_j x_j\right). \tag{63}$$

We now aim to approximate $[0,1]^k \ni \boldsymbol{x} \mapsto \left(\prod_{i \in V(F)} f(x_i)^{d_i}\right)\left(\prod_{\{i,j\} \in E(F)} W(x_i, x_j)\right) = \prod_{\ell=1}^{m} g_\ell(\gamma_\ell)$ by

$$\begin{aligned}
\prod_{\ell=1}^{m} \mathcal{N}_\ell(\gamma_\ell) \;&=\; \prod_{\ell=1}^{m} \left(\sum_{j=1}^{L_\ell} c_j^{(\ell)} \cos\left(a_j^{(\ell)} \gamma_\ell + b_j^{(\ell)}\right) + d^{(\ell)}\right) && (64)\\
&= \sum_{A \subseteq [m]} \left(\prod_{\ell \in [m] \setminus A} d^{(l)}\right) \sum_{\substack{j_\ell \in [L_\ell] \\ \ell \in A}} \left(\prod_{\ell \in A} c_{j_\ell}^{(\ell)} \cos\left(a_{j_\ell}^{(\ell)} \gamma_\ell + b_{j_\ell}^{(\ell)}\right)\right) && (65)\\
&= \sum_{A \subseteq [m]} \left(\prod_{\ell \in [m] \setminus A} d^{(l)}\right) \sum_{\substack{j_\ell \in [L_\ell] \\ \ell \in A}} \frac{\prod_{\ell \in A} c_{j_\ell}^{(\ell)}}{2^{|A|}} \sum_{\sigma \in \{\pm 1\}^A} \cos\left(\sum_{\ell \in A} \sigma_\ell \left(a_{j_\ell}^{(\ell)} \gamma_\ell + b_{j_\ell}^{(\ell)}\right)\right), && (66)
\end{aligned}$$

where we used Equation (63) for the last equality. By taking a look at the terms inside the cosines in Equation (66), it is straightforward to see that $\int_{[0,1]^k} \prod_{\ell=1}^{m} \mathcal{N}_\ell(\gamma_\ell) \,\mathrm{d}\lambda^k$ can be implemented by an IWN. It remains to be shown that Equation (66) does indeed approximate $t(F, \boldsymbol{d}, \cdot)$ uniformly on $\mathcal{WL}_r$. Note that

$$\begin{aligned}
\left|\int_{[0,1]^k} \prod_{\ell=1}^{m} \mathcal{N}_\ell(\gamma_\ell) \,\mathrm{d}\lambda^k(\boldsymbol{x}) - t(F, \boldsymbol{d}, (W, f))\right| \;&=\; \left|\int_{[0,1]^k} \left(\prod_{\ell=1}^{m} \mathcal{N}_\ell(\gamma_\ell) - \prod_{\ell=1}^{m} g_\ell(\gamma_\ell)\right) \mathrm{d}\lambda^k\right| && (67)\\
\leq \left\|\prod_{\ell=1}^{m} \mathcal{N}_\ell(\gamma_\ell) - \prod_{\ell=1}^{m} g_\ell(\gamma_\ell)\right\|_\infty, && && (68)
\end{aligned}$$

and $\{\mathcal{N}_\ell(\gamma_\ell)\}_\ell, \{g_\ell(\gamma_\ell)\}_\ell \in [-r-\delta, r+\delta]^{v(F)} \times [-1-\delta, 1+\delta]^{e(F)}$, $\|\mathcal{N}_\ell(\gamma_\ell) - g_\ell(\gamma_\ell)\|_\infty \leq \delta$ for all $\ell \in [m]$, and $\boldsymbol{\gamma} \mapsto \prod_{\ell=1}^{m} \gamma_\ell$ is Lipschitz-continuous on compact domains. Hence, Equation (68) converges to zero as $\delta \to 0$ uniformly in $(W, f) \in \mathcal{WL}_r$, and thus can be made $\leq \varepsilon$ if $\delta > 0$ is chosen small enough. This concludes the proof. ■

Until now, we have only worked with $\varrho = \cos$ in Lemma 14. Extending this result to an arbitrary nonlinearity boils down to a simple application of the standard universal approximation theorem for feedforward neural networks.

Lemma 15 *The statement of Lemma 14 holds for any continuous non-polynomial $\varrho : \mathbb{R} \to \mathbb{R}$.*

Proof We first show that we can approximate any IWN with cosine nonlinearity up to arbitrary precision using ϱ. Fix $\delta > 0$ and let

$$\mathcal{N}(W, f) := \sum_{s=1}^{S} L_s^{(2)}\Big(\cos(L_s^{(1)}(W) + \widetilde{L}_s^{(1)}(f) + b_s^{(1)})\Big) + b^{(2)} \tag{69}$$

be an IWN as in Equation (3). Set $M := \max_{s\in\{1,\dots,S\}} \|\cos(L_s^{(1)}(W) + \widetilde{L}_s^{(1)}(f) + b_s^{(1)})\|_\infty$ and let C_s be the Lipschitz constant of $L_s^{(2)}$ w.r.t. $\|\cdot\|_\infty$. Then, by universal approximation, there exists a feedforward NN $\mathcal{N}_{\cos} : \mathbb{R} \to \mathbb{R}$ with ϱ as nonlinearity such that

$$\sup_{x\in[-M,M]} |\mathcal{N}_{\cos}(x) - \cos(x)| \leq \delta. \tag{70}$$

Clearly, replacing cos with $\mathcal{N}_{\cos}$ in Equation (69) yields a valid IWN with ϱ as nonlinearity, and for any $(W, f) \in \mathcal{WL}_r$, we obtain

$$\left|\sum_{s=1}^{S} L_s^{(2)}\Big(\mathcal{N}_{\cos}(L_s^{(1)}(W) + \widetilde{L}_s^{(1)}(f) + b_s^{(1)})\Big) - \sum_{s=1}^{S} L_s^{(2)}\Big(\cos(L_s^{(1)}(W) + \widetilde{L}_s^{(1)}(f) + b_s^{(1)})\Big)\right| \tag{71}$$

$$\leq \sum_{s=1}^{S} C_s \left\|\mathcal{N}_{\cos}(L_s^{(1)}(W) + \widetilde{L}_s^{(1)}(f) + b_s^{(1)}) - \cos(L_s^{(1)}(W) + \widetilde{L}_s^{(1)}(f) + b_s^{(1)})\right\|_\infty \tag{72}$$

$$\overset{\text{Equation (70)}}{\leq} \delta \cdot \sum_{s=1}^{S} C_s, \tag{73}$$

which converges to zero as $\delta \to 0$. Hence, this bound also holds when taking the supremum over all $(W, f) \in \mathcal{WL}_r$. By combining this with Lemma 14, we obtain that any homomorphism density $t(F, \boldsymbol{d}, \cdot)$ on $\mathcal{WL}_r$ can be approximated by an IWN with any continuous nonlinearity ϱ up to any desired accuracy. ■

Proof [Proof of Theorem 5] We show this statement by applying the Stone-Weierstrass theorem. In principle, proving approximation of the signal-weighted homomorphism densities (Lemma 14 and Lemma 15) 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.,

$$\mathcal{D} := \operatorname{span}\{t(F, \boldsymbol{d}, \cdot) \mid F \text{ simple graph}, \boldsymbol{d} \in \mathbb{N}_0^{v(F)}\} \subseteq C(K, \mathbb{R}). \tag{74}$$

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, $\boldsymbol{d}_1 \in \mathbb{N}_0^{v(F_1)}$, $\boldsymbol{d}_2 \in \mathbb{N}_0^{v(F_2)}$,

$$t(F_1, \boldsymbol{d}_1, \cdot) \cdot t(F_2, \boldsymbol{d}_2, \cdot) \;=\; t(F_1 \sqcup F_2, \boldsymbol{d}_1 \| \boldsymbol{d}_2, \cdot) \;\in\; \mathcal{D}, \tag{75}$$

i.e., the product of homomorphism densities w.r.t. two simple graphs can be rewritten as the homomorphism density w.r.t. their disjoint union. By Theorem 1, $\mathcal{D}$ also separates points, and we can apply Stone-Weierstrass to conclude that $\mathcal{D} \subseteq C(K, \mathbb{R})$ is dense. However, by Lemma 15, any element of $\mathcal{D}$ can be approximated with arbitrary precision by $\mathcal{IWN}_\varrho$, and thus

$$C(K, \mathbb{R}) \;=\; \overline{\mathcal{D}} \;\subseteq\; \overline{\mathcal{IWN}_\varrho} \;\subseteq\; C(K, \mathbb{R}). \tag{76}$$

This concludes the proof. ■

D.3. Point Separation of IWNs

Corollary 16 (Point Separation of IWNs) *Let $(W, f), (V, g) \in \mathcal{WL}_r$ be not weakly isomorphic, i.e., $\delta_p((W, f), (V, g)) > 0$ for any $p \in [1, \infty)$ (or equivalently $\delta_\square((W, f), (V, g)) > 0$). Then, there exists an IWN $\mathcal{N}$ such that*

$$\mathcal{N}(W, f) \;\neq\; \mathcal{N}(V, g). \tag{77}$$

Proof By Theorem 1, for any two graphon-signals $(W, f), (V, g)$ we have $\delta_p((W, f), (V, g)) = 0$ ($\Leftrightarrow$ $\delta_\square((W, f), (V, g)) = 0$) if and only if $t(F, \boldsymbol{d}, (W, f)) = t(F, \boldsymbol{d}, (V, g))$ for any simple graph F and $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$. Hence, if $\delta_p((W, f), (V, g)) > 0$, there exists a simple graph F and $\boldsymbol{d} \in \mathbb{N}_0^{v(F)}$ such that $\varepsilon := |t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g))| > 0$. By Lemma 15, take an IWN $\mathcal{N}$ such that $\sup_{(U,h) \in \mathcal{WL}_r} |\mathcal{N}(U, h) - t(F, \boldsymbol{d}, (U, h))| \leq \varepsilon/3$. We obtain

$$\begin{aligned}
&|\mathcal{N}(W, f) - \mathcal{N}(V, g)| && (78)\\
&\;=\; |\mathcal{N}(W, f) - t(F, \boldsymbol{d}, (W, f)) + t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g)) + t(F, \boldsymbol{d}, (V, g)) - \mathcal{N}(V, g)| && (79)\\
&\;\geq\; |t(F, \boldsymbol{d}, (W, f)) - t(F, \boldsymbol{d}, (V, g))| && (80)\\
&\qquad - |\mathcal{N}(W, f) - t(F, \boldsymbol{d}, (W, f))| - |t(F, \boldsymbol{d}, (V, g)) - \mathcal{N}(V, g)| && (81)\\
&\;\geq\; \varepsilon - \varepsilon/3 - \varepsilon/3 \;=\; \varepsilon/3 > 0, && (82)
\end{aligned}$$

which yields the claim. ■

D.4. Proof of Theorem 6

Proof [Proof of Theorem 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

$$\|\mathcal{N}_p - \mathcal{N}\|_{\infty, \mathcal{WL}_r} \;:=\; \sup_{(W,f) \in \mathcal{WL}_r} |\mathcal{N}_p(W, f) - \mathcal{N}(W, f)| \;\leq\; \varepsilon/2. \tag{83}$$

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 the proof of Theorem 5.

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

$$\mathcal{N}_p(W, f) = \sum_i \alpha_i \cdot t(F_i, \boldsymbol{d}_i, (W, f)) \tag{84}$$

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

$$\widetilde{\mathcal{N}}_p(W, f) = \sum_i \alpha_i \cdot t(F_i^{\text{simple}}, \boldsymbol{d}_i, (W, f)), \tag{85}$$

where $F_i \mapsto F_i^{\text{simple}}$ removes parallel edges. $\widetilde{\mathcal{N}}_p$ is Lipschitz continuous in the cut distance by Lemma 9 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, Theorem 3.7), we can bound

$$\begin{aligned}
&\mathbb{E}\big[|\mathcal{N}(G_n, \boldsymbol{f}_n) - \mathcal{N}(G_m, \boldsymbol{f}_m)|\big] && (86)\\
&\le \underbrace{\mathbb{E}\big[|\mathcal{N}(G_n, \boldsymbol{f}_n) - \mathcal{N}_p(G_n, \boldsymbol{f}_n)|\big]}_{\le \varepsilon/2} + \mathbb{E}\big[|\widetilde{\mathcal{N}}_p(G_n, \boldsymbol{f}_n) - \widetilde{\mathcal{N}}_p(G_m, \boldsymbol{f}_m)|\big] && (87)\\
&\quad + \underbrace{\mathbb{E}\big[|\mathcal{N}_p(G_m, \boldsymbol{f}_m) - \mathcal{N}(G_m, \boldsymbol{f}_m)|\big]}_{\le \varepsilon/2} && (88)\\
&\le \varepsilon + M \cdot \mathbb{E}\left[\delta_\square((G_n, \boldsymbol{f}_n), (G_m, \boldsymbol{f}_m))\right] && (89)\\
&\le \varepsilon + M \cdot \Big(\mathbb{E}\big[\delta_\square((G_n, \boldsymbol{f}_n), (W, f))\big] + \mathbb{E}\big[\delta_\square((W, f), (G_m, \boldsymbol{f}_m))\big]\Big) && (90)\\
&\overset{(*)}{\le} \varepsilon + M\left(\frac{15}{\sqrt{\log n}} + \frac{15}{\sqrt{\log m}}\right) \le \varepsilon + \underbrace{15M}_{=:C_{\varepsilon,\mathcal{N}}}\left(\frac{1}{\sqrt{\log n}} + \frac{1}{\sqrt{\log m}}\right), && (91)
\end{aligned}$$

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

Appendix E. Comparison to Results of Cai and Wang (2022)

Cai and Wang (2022) already studied the convergence of IGNs using the full IGN basis and a *partition norm*, which is for $W \in \mathcal{L}^2[0,1]^k$ a 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 is a fact that also follows directly from our observation on continuity (Theorem 4).

As a remedy, Cai and 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. Note the fact that while their condition of IGN-small considers the entire multilayer neural network, we impose our boundedness condition on the individual linear equivariant layers. As a notable difference, Cai and Wang (2022) use multilayer continuous IGNs with multiple channels, i.e., neural networks $\mathcal{N}$ of the form

$$\mathcal{N} = \mathbf{L}^{(T)} \circ \varrho \circ \cdots \circ \varrho \circ \mathbf{L}^{(1)}, \tag{92}$$

in which each $\mathbf{L}^{(t)}$, $t \in [T]$, is an equivariant affine linear map from d_{t-1}-channeled kernels of "tensor order" k_{t-1} (i.e., functions $\mathbf{W} : [0,1]^{k_{t-1}} \to \mathbb{R}^{d_{t-1}}$) to d_t-channeled kernels of order k_t. Here, $(k_0, d_0) = (2,2)$ and $(k_T, d_T) = (0,1)$. For a graphon-signal $(W, f) \in \mathcal{WL}_r$, the input to such an IGN is $(x_1, x_2) \mapsto (W(x_1,x_2), \mathbb{1}\{x_1 = x_2\} f(x_1))$, but for compatibility with our method, $(x_1,x_2) \mapsto (W(x_1,x_2), f(x_1))$ yields the same function class. It is straightforward to see that any IWN as defined in Equation (3) can be represented as in Cai and Wang (2022): Padding all tensor orders to their maximum, we can indeed rewrite Equation (3) to Equation (92) with $T = 2$ layers, and the hidden dimension being the former number of addends $d_1 = S$. Alternatively, any IWN could be straightforwardly approximated up to arbitrary precision with $\mathcal{O}(S)$ layers of hidden dimension $\mathcal{O}(1)$.

Lemma 12 immediately implies that the representation of an IWN in the framework of Cai and Wang (2022) as described above does indeed yield a network in IGN-small. The converse, i.e., that individual linear equivariant layers in IGN-small have to be in $\mathrm{LE}_{k\to l}$, does not necessarily 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.

Consequently, Theorem 5 also applies to IGN-small in the sense that IGN-small of arbitrary tensor order can distinguish any pair of non-weakly isomorphic graphons, which is a significantly stronger statement than the ability of (2-)IGN-small being able to approximate spectral GNNs as demonstrated by Cai and Wang (2022). Further, our approach may appear more intuitive as it leads to less representation ambiguity, generalizes to arbitrary measure spaces (as shown in Equation (2)), and might align with defining a similar notion for IGNs in graphon space without prior knowledge of the discrete version, ensuring only well-behaved functions w.r.t. the underlying space of interest.

Appendix F. IWNs of Bounded Order and k-WL

We emphasize that our approach, just as Keriven and Peyré (2019), depends on having arbitrarily large tensor orders at disposal. In the discrete setting, multilayer k-IGNs of the form of Equation (92) are known to be at least as powerful as the k-dimensional Weisfeiler-Lehman (WL) test (Maron et al., 2019; Azizian and Lelarge, 2021). k-WL, in turn, has been shown to distinguish homomorphism densities w.r.t. precisely *all* graphs of treewidth bounded by k (Dvořák, 2010; Dell et al., 2018). In contrast, in our proof we use tensor orders of n to model homomorphism densities of graphs up to *size* n, which suggests that restricting the order in Equation (3) will not yield any particularly useful function class hierarchy. With multilayer IWNs similar to Equation (92) of order bounded by k, it can be shown that such a model can approximate any signal-weighted homomorphism density w.r.t. multigraphs of treewidth k, using the tree decomposition of a graph as in Böker (2023). As a result, multilayer IWNs of order k are at least as powerful as k-WL for graphons.

Crucially, we also note that cut distance discontinuity is *not* specific to IWNs, but inherently linked to k-WL. The k-WL test for graphons (Böker, 2023) 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. Future versions of this work will address this in more detail.

Appendix G. Asymptotic Dimension Analysis of $\mathrm{LE}_{k\to l}^{[0,1]}$

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

Recall that

$$\dim \mathrm{LE}_{k\to l}^{[n]} = \mathrm{bell}(k+l), \tag{93}$$

$$\dim \mathrm{LE}_{k\to l}^{[0,1]} = \sum_{s=0}^{\min\{k,l\}} s!\binom{k}{s}\binom{l}{s}. \tag{94}$$

For a comparison of the dimensions for the first few pairs (k, l), see Table 1.

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

$\dim \mathrm{LE}_{k\to l}^{[n]}$	**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

$\dim \mathrm{LE}_{k\to l}^{[0,1]}$	**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 1 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 $l \to \infty$, we have $\mathrm{bell}(k + l) \to \infty$ superexponentially.

However, for the case of $[0,1]$, suppose w.l.o.g. that only $k \to \infty$ and $l = \mathcal{O}(1)$ remains constant. Then, the corresponding dimension growth is bounded by

$$\dim \mathrm{LE}_{k \to l}^{[0,1]} \;=\; \dim \mathrm{LE}_{l \to k}^{[0,1]} \;=\; \mathcal{O}(k^l), \tag{95}$$

as Equation (94) is dominated by $\binom{k}{l}$ in this case.

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

$$\dim \mathrm{LE}_{k \to k}^{[n]} \;=\; \mathrm{bell}(2k), \quad \dim \mathrm{LE}_{k \to k}^{[0,1]} = \sum_{s=0}^{k} s! \binom{k}{s}^2. \tag{96}$$

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

$$\mathrm{bell}(n) \;\sim\; \frac{1}{\sqrt{n}} \left(\frac{n}{W(n)} \right)^{n+1/2} \exp\left(\frac{n}{W(n)} - n - 1 \right), \tag{97}$$

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

$$\left(\frac{1}{e} \frac{n}{\log n} \right)^n \;\leq\; \mathrm{bell}(n) \;\leq\; \left(\frac{3}{4} \frac{n}{\log n} \right)^n, \tag{98}$$

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

$$\dim \mathrm{LE}_{k \to k}^{[n]} \;\geq\; \left(\frac{1}{e} \frac{2k}{\log 2k} \right)^{2k}. \tag{99}$$

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

$$\dim \mathrm{LE}_{k \to k}^{[0,1]} \;\geq\; k! \;\geq\; \mathrm{bell}(k) \tag{100}$$

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

$$\frac{n^n}{e^{n-1}} \;\leq\; n! \;\leq\; \frac{n^{n+1}}{e^{n-1}}, \tag{101}$$

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

$$\dim \mathrm{LE}_{k \to k}^{[0,1]} \;=\; \sum_{s=0}^{k} s! \binom{k}{s}^2 \tag{102}$$

$$\leq\; (k+1) k! \binom{k}{k/2}^2 \;=\; (k+1) \frac{k!^3}{(k/2)!^4} \tag{103}$$

$$\overset{\text{Equation (101)}}{\leq} \;\; (k+1) \frac{k^{3k+3}}{e^{3k-3}} \frac{e^{2k-4}}{(k/2)^{2k}} \;=\; \frac{1}{e} (k+1) k^3 \left(\frac{4}{e} k \right)^k, \tag{104}$$

which still grows significantly slower than Equation (99).