Weisfeiler-Leman Indistinguishability of Graphons

Jan Böker
RWTH Aachen University, Aachen, Germany
boeker@informatik.rwth-aachen.de
ORCID: 0000-0003-4584-121X

December 17, 2021

Abstract

The color refinement algorithm is mainly known as a heuristic method for graph isomorphism testing. It has surprising but natural characterizations in terms of, for example, homomorphism counts from trees and solutions to a system of linear equations. Grebík and Rocha (2021) have recently shown that color refinement and some of its characterizations generalize to graphons, a natural notion for the limit of a sequence of graphs. In particular, they show that these characterizations are still equivalent in the graphon case. The k-dimensional Weisfeiler-Leman algorithm (k-WL) is a more powerful variant of color refinement that colors k-tuples instead of single vertices, where the terms 1-WL and color refinement are often used interchangeably since they compute equivalent colorings. We show how to adapt the result of Grebík and Rocha to k-WL or, in other words, how k-WL and its characterizations generalize to graphons. In particular, we obtain characterizations in terms of homomorphism densities from *multi*graphs of bounded treewidth and linear equations. We give a simple example that parallel edges make a difference in the graphon case, which means that the equivalence between 1-WL and color refinement is lost. We also show how to define a variant of k-WL that corresponds to homomorphism densities from simple graphs of bounded treewidth.

1 Introduction

The color refinement algorithm is usually used as an efficient heuristic in graph isomorphism testing [12] even though it has more applications, e.g., in machine learning. It iteratively colors the vertices of a (simple) graph, where initially all vertices get the same color. Then, in every refinement round, two vertices v and w of the same color get assigned different colors if there is some color c such that v and w have a different number of neighbors of color c. If these color patterns computed for two graphs G and H do not match, G and H are said to be *distinguished* by color refinement.

Indistinguishability by color refinement has various characterizations: A result of Dvořák states two graphs G and H are not distinguished by color refinement if and only if the number of homomorphisms $\mathsf{hom}(T, G)$ from T to G equals the correspondence number $\mathsf{hom}(T, H)$ from T to H for every tree T [7], see also [5]. An older result due to Tinhofer [20, 19] states that G and H are not distinguished by color refinement if and only if they are *fractionally isomorphic*, i.e., there is a doubly stochastic matrix X such that $AX = XB$, where A and B are the adjacency matrices of G and H, respectively. A characterization that is more closely related to the color refinement algorithm itself is given by *stable* partitions of the vertex set $V(G)$ of a graph G, which are partitions where all vertices in the same class have the same number of neighbors in every other class. The term *equitable* is also sometimes used for this but may not be confused with equitable partitions from Szemerédi's regularity lemma. The partition induced by the colors of color refinement is the *coarsest* stable partition, and graphs G and H are fractionally isomorphic if and only if their coarsest stable partitions have the same parameters, i.e., there is a bijection between the partitions that preserves the size of every class C and the numbers of neighbors a vertex in C has in some other class D [20]. This, in turn, is equivalent to there being *some* stable partitions of G and H with the same parameters [18]. We collect all these characterizations in Theorem 1. It is worth mentioning that fractional isomorphism can also be seen from the perspective of logic; it corresponds to equivalence in the logic C^2, the 2-variable fragment of first-order logic with counting quantifiers [13]. This, however, does not play a role in this paper, which is why we omit it.

Theorem 1 ([20, 19, 18, 7, 5])**.** *Let G, H be graphs with adjacency matrices A, B, respectively. The following are equivalent:*

1. $\mathsf{hom}(T, G) = \mathsf{hom}(T, H)$ *for every tree T.*
2. *Color refinement does not distinguish G and H.*
3. *The coarsest stable partitions of $V(G)$ and $V(H)$ have the same parameters.*
4. *There is a doubly stochastic X such that $AX = XB$.*
5. *There are stable partitions of $V(G)$ and $V(H)$ with the same parameters.*

The *k-dimensional Weisfeiler-Leman algorithm (k-WL)* is a variant of color refinement that colors k-tuples of vertices instead of single vertices; here and also throughout the paper, k is an integer with $k \geq 1$. See [4] for an overview of the history of k-WL. Usually, no distinction is made between 1-WL and color refinement as they, in some sense, compute equivalent colorings. All of the previously described characterizations of color refinement generalize to k-WL: First of all, k-WL does not distinguish graphs G and H if and only the number of homomorphisms $\mathsf{hom}(F, G)$ from F to G is equal to the corresponding number $\mathsf{hom}(F, H)$ from F to H for every graph F of treewidth at most k [7, 5]. The concept of fractional isomorphisms via non-negative solutions to the following system $\mathsf{L}^k_{\text{iso}}(G, H)$ of linear equations, which has a variable X_π for every set $\pi \subseteq V(G) \times V(H)$ of size $|\pi| \leq k$. Such a set π is called a *partial isomorphism* if the mapping it induces is injective and preserves (non-)adjacency. The equivalence of k-WL to precisely this system of linear equations is from [5], although it is already implicit in earlier work [13, 1, 11].

$$\mathsf{L}^k_{\text{iso}}(G,H)\colon \begin{cases} \displaystyle\sum_{v \in V(G)} X_{\pi \cup \{(v,w)\}} = X_\pi & \text{for every } \pi \subseteq V(G) \times V(H) \text{ of size} \\ & |\pi| \leq k-1 \text{ and every } w \in V(H) \\ \displaystyle\sum_{w \in V(H)} X_{\pi \cup \{(v,w)\}} = X_\pi & \text{for every } \pi \subseteq V(G) \times V(H) \text{ of size} \\ & |\pi| \leq k-1 \text{ and every } v \in V(G) \\ X_\varnothing = 1 & \\ X_\pi = 0 & \text{for every } \pi \subseteq V(G) \times V(H) \text{ of size } |\pi| \leq k \\ & \text{that is not a partial isomorphism} \end{cases}$$

Stable partitions of the vertex set $V(G)$ of a graph G easily generalize to stable partitions of $V(G)^k$. The coloring computed by k-WL on G induces the coarsest stable partition of $V(G)^k$ and two graphs G and H are not distinguished by k-WL if and only if the coarsest stable partitions of $V(G)^k$ and $V(H)^k$ have the same parameters, which again is equivalent to there being some stable partitions with the same parameters. See, for example, [11], where this is implicitly treated. Also note that equivalence in the logic C^2 generalizes to equivalence in C^{k+1}, the $k+1$-variable fragment of first-order logic with counting quantifiers [4]. Let us state the generalization of Theorem 1 to k-WL as Theorem 2.

Theorem 2 ([7, 5])**.** *Let $k \geq 1$ and G, H be graphs. The following are equivalent:*

1. $\mathsf{hom}(F, G) = \mathsf{hom}(F, H)$ *for every graph of treewidth at most k.*
2. *k-WL does not distinguish G and H.*
3. *The coarsest k-stable partitions of $V(G)^k$ and $V(H)^k$ have the same parameters.*
4. $\mathsf{L}_{\text{iso}}^{k+1}(G, H)$ *has a non-negative real solution.*
5. *There are k-stable partitions of $V(G)^k$ and $V(H)^k$ with the same parameters.*

Graphons emerged in the theory of graph limits as limit objects of sequences of dense graphs; see the book of Lovász [16] for a detailed introduction to the theory of graph limits. Formally, a graphon is a symmetric measurable function $W\colon [0,1] \times [0,1] \to [0,1]$, although it can be quite useful to consider more general underlying spaces than the unit interval with the Lebesgue measure. Grebík and Rocha recently generalized Theorem 1 to graphons [9]. A substantial part of their work involves showing how to even state the characterizations of color refinement that are found in Theorem 1 for graphons. Note that graphs and, more generally, (vertex- and edge-)weighted graphs can be viewed as graphons by partitioning $[0,1]$ into one interval for each vertex, cf. [16, Section 7.1]. This means that Theorem 1 and also a variant for weighted graphs can in fact be restored from their result. In this paper, we show how to marry their result with k-WL to obtain a variant of Theorem 2 for graphons. In the remainder of the introduction, we get more formal with the goal of giving the reader a clear understanding of the results of this paper without going into details too much. A reader interested in these details can then continue with the main part of the paper. In Section 1.1, we first state and explain the result of Grebík and Rocha, before we state and discuss our result and the structure of the main part of this paper in Section 1.2.

1.1 Fractional Isomorphism of Graphons

Let us briefly give a formal definition of graphs, homomorphisms, and color refinement. A *(simple) graph* is a pair $G = (V, E)$, where V is a set of *vertices* and $E \subseteq \binom{V}{2}$ a set of *edges*. We usually write $V(G) \coloneqq V$ and $E(G) \coloneqq E$. A *homomorphism* from a graph F to a graph G is a mapping $h\colon V(F) \to V(G)$ such that $uv \in E(F)$ implies $h(u)h(v) \in E(G)$. The number of homomorphisms from F to G is denoted by $\mathsf{hom}(F, G)$, and $t(F, G) \coloneqq \mathsf{hom}(F, G)/|V(G)|^{|V(F)|}$ is the *homomorphism density of F in G*. Now, let us turn our attention to color refinement. The initial coloring of the vertices of a graph G is obtained by letting $\mathsf{cr}_{G,0}(v) \coloneqq 1$ for every vertex $v \in V(G)$. Then, for every $n \geq 0$, let

$$\mathsf{cr}_{G,n+1}(v) \coloneqq (\mathsf{cr}_{G,n}(v), \{\!\{\mathsf{cr}_{G,n}(w) \mid wv \in E(G)\}\!\})$$

for every $v \in V(G)$. Here, $\{\!\{\cdot\}\!\}$ is used as the notion for a *multiset*. We say that *color refinement does not distinguish two graphs G and H* if $\{\!\{\mathsf{cr}_{G,n}(v) \mid v \in V(G)\}\!\} = \{\!\{\mathsf{cr}_{H,n}(v) \mid v \in V(H)\}\!\}$ for every $n \geq 0$.

Instead of the unit interval with the Lebesgue measure, we follow Grebík and Rocha, and throughout the whole paper, let $(X, \mathcal{B})$ denote a *standard Borel space* and μ a *Borel probability measure* on X; this has the advantage that we later can consider quotient spaces. We think of $(X, \mathcal{B}, \mu)$ as *atom free*, i.e., that there is no singleton set of positive measure, but do not formally require it. A *kernel* is a $(\mathcal{B} \otimes \mathcal{B})$-measurable map $W\colon X \times X \to [0,1]$, A symmetric kernel is called a *graphon*. Grebík and Rocha have shown the following generalization of Theorem 1 to graphons, whose characterizations we elaborate one by one.

Theorem 3 ([9])**.** *Let $U, W\colon X \times X \to [0,1]$ be graphons. The following are equivalent:*

1. *$t(T,U) = t(T,W)$ for every tree T.*
2. *$\nu_U = \nu_W$.*
3. *$W/\mathcal{C}(W)$ and $U/\mathcal{C}(U)$ are isomorphic.*
4. *There is a Markov operator $S\colon L^2(X,\mu) \to L^2(X,\mu)$ such that $T_U \circ S = S \circ T_W$.*
5. *There are U- and W-invariant μ-relatively complete sub-σ-algebras $\mathcal{C}$ and $\mathcal{D}$, respectively, such that $U_{\mathcal{C}}$ and $W_{\mathcal{D}}$ are weakly isomorphic.*

For Characterization 2, the *homomorphism density of a graph F in a graphon* $W\colon X \times X \to [0,1]$ is

$$t(F,W) \coloneqq \int_{X^{V(F)}} \prod_{ij \in E(F)} W(x_i,x_j)\, d\mu^{\otimes V(F)}(\bar{x}). \tag{1}$$

Note that this coincides with the previous definition for graphs, i.e., when viewing a graph G as a graphon W_G we have $t(F,G) = t(F,W_G)$ [16, (7.2)].

Characterization 2 generalizes color refinement to graphons and requires more formal precision than in the case of graphs. Grebík and Rocha first define the standard Borel space $\mathbb{M}$ of *iterated degree measures*, which can be seen as the space of colors used by color refinement; Its elements are sequences $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots)$ of colors after $0,1,2,\dots$ refinement rounds. Then, for a graphon $W\colon X \times X \to [0,1]$, they define the measurable function $\mathsf{cr}_W\colon X \to \mathbb{M}$ mapping every $x \in X$ to such a sequence $(\alpha_0, \alpha_1, \alpha_2, \dots)$. Then, the *distribution on iterated degree measures (DIDM)* ν_W defined by $\nu_W(A) \coloneqq \mu(\mathsf{cr}_W^{-1}(A))$ for every $A \in \mathcal{B}(\mathbb{M})$, i.e., as the *push-forward of μ via* cr_W, is a probability measure on the space $\mathbb{M}$. Note the similarity between Characterization 2 and color refinement not distinguishing two graphs: The multisets used in the definition of color refinement indistinguishability can be seen as maps mapping a color to a natural number stating how often it occurs in the graph. Intuitively, a DIDM does the same for a set of colors and a number in $[0,1]$.

Characterization 3 generalizes the coarsest stable partitions of the vertex set $V(G)$ of a graph G to the *minimum W-invariant μ-relatively complete sub-σ-algebra* $\mathcal{C}_W$ for a graphon $W\colon X \times X \to [0,1]$. Let us break down this term bit by bit, starting with μ-relatively complete sub-σ-algebras of $\mathcal{B}$. Let $L^2(X,\mu) \coloneqq L^2(X,\mathcal{B},\mu)$ denote the Hilbert space of all measurable real-valued functions on X with $\|f\|_2 < \infty$ modulo equality μ-almost everywhere. For a sub-σ-algebra $\mathcal{C}$ of $\mathcal{B}$, we want to consider the subspace of all $\mathcal{C}$-measurable functions of $L^2(X,\mathcal{C},\mu)$. To make this statement formally precise, a sub-σ-algebra $\mathcal{C} \subseteq \mathcal{B}$ of $\mathcal{B}$ is called *μ-relatively complete* if $Z \in \mathcal{C}$ for all $Z \in \mathcal{B}, Z_0 \in \mathcal{C}$ with $\mu(Z \triangle Z_0) = 0$. The set of all μ-relatively complete sub-σ-algebras of $\mathcal{B}$ is denoted by $\Theta(\mathcal{B},\mu)$. As an example, the smallest μ-relatively complete sub-σ-algebra that includes $\{\varnothing, X\}$ corresponds to the trivial partition of the vertex set of a graph. A kernel $W\colon X \times X \to [0,1]$ defines the *kernel operator* $T_W\colon L^2(X,\mu) \to L^2(X,\mu)$ by setting

$$(T_W f)(x) \coloneqq \int_X W(x,y) f(y)\, d\mu(y)$$

for every $f \in L^2(X,\mu)$ and every $x \in X$. It is a well-defined Hilbert-Schmidt operator[16, Section 7.5], and if W is a graphon, then T_W is self-adjoint. In general, for an operator

$T\colon L^2(X,\mu) \to L^2(X,\mu)$, a $\mathcal{C} \in \Theta(\mathcal{B},\mu)$ is called *T-invariant* if $L^2(X,\mathcal{C},\mu)$ is T-invariant, i.e., $T(L^2(X,\mathcal{C},\mu)) \subseteq L^2(X,\mathcal{C},\mu)$. Then, a $\mathcal{C} \in \Theta(\mathcal{B},\mu)$ is called *W-invariant* if it is T_W-invariant.

Grebík and Rocha show that, for a graphon $W\colon X \times X \to [0,1]$, the minimum W-invariant μ-relatively complete sub-σ-algebra $\mathcal{C}_W$ of $\mathcal{B}$ can be obtained by iterative applications of T_W when starting from $\{\varnothing, X\}$. From this, they define a quotient graphon $W/\mathcal{C}_W$. Formally, for every $\mathcal{C} \in \Theta(\mathcal{B},\mu)$, there is a corresponding quotient space, i.e., a standard Borel space $(X/\mathcal{C},\mathcal{C}')$ with a Borel probability measure $\mu/\mathcal{C}$ on $X/\mathcal{C}$, and $W/\mathcal{C}_W$ is defined on the space $X/\mathcal{C} \times X/\mathcal{C}$. Then, saying that two such quotient graphons are isomorphic corresponds to saying that two coarsest stable partitions have the same parameters. As a side note, in their proof, Grebík and Rocha show that every DIDM ν defines a kernel $\mathbb{M} \times \mathbb{M} \to [0,1]$. They show that, for a graphon $W\colon X \times X \to [0,1]$ and its DIDM ν_W, this kernel on $\mathbb{M} \times \mathbb{M}$ is actually isomorphic to $W/\mathcal{C}_W$. Intuitively, we can view this as a canonical representation of W on the space of all colors.

Characterization 5 is similar to Characterization 3. Just as the coarsest stable partitions of the vertex sets of two graphs have the same parameters if and only if there are some stable partitions with the same parameters, the minimum U- and W-invariant μ-relatively complete sub-σ-algebras can be replaced by some U- and W-invariant μ-relatively complete sub-σ-algebras $\mathcal{C}$. Note that there is a subtle difference in the way Grebík and Rocha phrase Characterization 5 as they use the conditional expectation instead of the quotient spaces: $W_\mathcal{C}$ is defined as the conditional expectation of W given $\mathcal{C} \times \mathcal{C}$. Intuitively, $W_\mathcal{C}$ is obtained by averaging over the color classes of $\mathcal{C}$, while $W/\mathcal{C}$ is obtained by averaging over the color classes of $\mathcal{C}$ and then identifying all elements of a color class. Then, the resulting graphons are required to be weakly isomorphic, where two graphons $U, W\colon X \times X \to [0,1]$ are called *weakly isomorphic* if $t(F,U) = t(F,W)$ for every simple graph F. This is the usual notion of isomorphism used for graphons, and two graphons are weakly isomorphic if and only if they have *cut distance* zero, cf. [16, Section 10.7].

Finally, Characterization 4 generalizes fractional isomorphisms. For standard Borel spaces $(X,\mathcal{B})$ and $(Y,\mathcal{D})$ with Borel probability measures μ and ν on X and Y, respectively, an operator $S\colon L^2(X,\mu) \to L^2(Y,\nu)$ is called a *Markov operator* if $Sf \geq 0$ for every $f \in L^2(X,\mu)$ with $f \geq 0$, $S\mathbf{1}_X = \mathbf{1}_Y$, and $S^*\mathbf{1}_Y = \mathbf{1}_X$. Here, $\mathbf{1}_X$ and $\mathbf{1}_Y$ denote the all-one functions on X and Y, respectively, and S^* denotes the *Hilbert adjoint* of S, which is the unique operator $S^*\colon L^2(Y,\nu) \to L^2(X,\mu)$ satisfying $\langle Sf, g\rangle = \langle f, S^*g\rangle$ for all $f \in L^2(X,\mu), g \in L^2(Y,\nu)$. Markov operators are simply the infinite-dimensional analogue to doubly stochastic matrices. With this in mind, the connection of Characterization 4 to the graph case is obvious.

1.2 Weisfeiler-Leman Indistinguishability of Graphons

Let us first state the definition of k-WL, which is important as there actually are two non-equivalent definitions to be found in the literature. Following Grohe [10], we refer to these distinct definitions as *k-WL* and *oblivious k-WL*. Both k-WL and oblivious k-WL operate on k-tuples of vertices, but in terms of expressive power, k-WL is equivalent to oblivious $k+1$-WL in the sense that they distinguish the same graphs. Hence, from an efficiency point of view, k-WL is more interesting as it needs less memory to achieve the same expressive power, but in our case, oblivious k-WL is more interesting as the connections to other characterizations are much cleaner, cf. the mismatch between the k in k-WL and the $k+1$ in the system $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$ of linear equations in Theorem 2 or the $k+1$ in the logic C^{k+1}. The reason that the k in k-WL matches the k in "treewidth k" is just that one is subtracted from the bag width in the definition of treewidth.

Let us start with k-WL. Let G be a graph. The atomic type $\mathsf{atp}_G(\bar{v})$ of a tuple $\bar{v} = (v_1,\dots,v_k) \in V(G)^k$ of vertices of G is the $k \times k$-matrix A with entries $A_{ij} = 2$ if $v_i = v_j$, $A_{ij} = 1$ if $v_iv_j \in E(G)$, and $A_{ij} = 0$ otherwise. Then, let $\mathsf{wl}^k_{G,0}(\bar{v}) \coloneqq \mathsf{atp}_G(\bar{v})$ and, for every

$n \geq 0$, define

$$\mathsf{wl}^k_{G,n+1}(\bar{v}) := \Big(\mathsf{wl}^k_{G,n+1}(\bar{v}), \{\!\{(\mathsf{atp}_G(\bar{v}w), \big(\mathsf{wl}^k_{G,n+1}(\bar{v}[w/j])\big)_{j\in[k]}) \mid w \in V(G)\}\!\}\Big) \tag{2}$$

for every $\bar{v} \in V(G)^k$. Here, $\bar{v}[w/j]$ denotes the k-tuple obtained from $\bar{v}$ by replacing the jth component by w; the k-tuple $\bar{v}[w/j]$ is usually called a *j-neighbor* of $\bar{v}$. We say that *k-WL does not distinguish graphs G and H* if $\{\!\{\mathsf{wl}^k_{G,n}(\bar{v}) \mid \bar{v} \in V(G)^k\}\!\} = \{\!\{\mathsf{wl}^k_{H,n}(\bar{v}) \mid \bar{v} \in V(H)^k\}\!\}$ for every $n \geq 0$. The colorings computed by 1-WL and color refinement induce the same partition and, in particular, 1-WL distinguishes two graphs if and only if color refinement does [10, Proposition V.4]. For oblivious k-WL, we also let $\mathsf{owl}^k_{G,0}(\bar{v}) := \mathsf{atp}_G(\bar{v})$, but then for every $n \geq 0$, we define

$$\mathsf{owl}^k_{G,n+1}(\bar{v}) := \Big(\mathsf{owl}^k_{G,n+1}(\bar{v}), \big(\{\!\{\mathsf{owl}^k_{G,n+1}(\bar{v}[w/j] \mid w \in V(G))\}\!\}\big)_{j\in[k]}\Big) \tag{3}$$

for every $\bar{v} \in V(G)^k$. We say that *oblivious k-WL does not distinguish graphs G and H* if $\{\!\{\mathsf{owl}^k_{G,n}(\bar{v}) \mid \bar{v} \in V(G)^k\}\!\} = \{\!\{\mathsf{owl}^k_{H,n}(\bar{v}) \mid \bar{v} \in V(H)^k\}\!\}$ for every $n \geq 0$. As mentioned before, k-WL is equivalent to oblivious $k+1$-WL in the sense that two graphs are distinguished by k-WL if and only if they are distinguished by oblivious $k+1$-WL [10, Corollary V.7]. This equivalence becomes clearer when diving into the details of this paper: intuitively, given a tree decomposition of width k, we may dissect it into parts at bags of size k or at bags of size $k+1$.

Let us state our main theorem, Theorem 4, before explaining its characterizations one by one. As mentioned before, it is based on oblivious k-WL, so there is a mismatch by one when comparing it to Theorem 2.

Theorem 4. *Let $k \geq 1$ and $U, W: X \times X \to [0,1]$ be graphons. The following are equivalent:*

1. *$t(F,U) = t(F,W)$ for every multigraph of treewidth at most $k-1$.*
2. *$\nu^k_U = \nu^k_W$.*
3. *There is a (permutation-inv.) Markov iso. $R: L^2(X^k/\mathcal{C}^k_W, \mu^{\otimes k}/\mathcal{C}^k_W) \to L^2(X^k/\mathcal{C}^k_U, \mu^{\otimes k}/\mathcal{C}^k_U)$ such that $\mathbb{T}^k_U/\mathcal{C}^k_U \circ R = R \circ \mathbb{T}^k_W/\mathcal{C}^k_W$.*
4. *There is a (permutation-inv.) Markov operator $S: L^2(X^k, \mu^{\otimes k}) \to L^2(X^k, \mu^{\otimes k})$ such that $\mathbb{T}^k_U \circ S = S \circ \mathbb{T}^k_W$.*
5. *There are $\mu^{\otimes k}$-relatively complete sub-σ-algebras $\mathcal{C}, \mathcal{D}$ of $\mathcal{B}^{\otimes k}$ that are U-invariant and W-invariant, respectively, and a Markov iso. $R: L^2(X^k/\mathcal{D}, \mu^{\otimes k}/\mathcal{D}) \to L^2(X^k/\mathcal{C}, \mu^{\otimes k}/\mathcal{C})$ such that $\mathbb{T}^k_U/\mathcal{C} \circ R = R \circ \mathbb{T}^k_W/\mathcal{D}$.*

First, let us examine Characterization 1, which uses *multi*graph homomorphism densities. A *multigraph* $G = (V, E)$ is defined like a graph with the exception that E is a multiset of edges from $\binom{V}{2}$. For a graphon $W: X \times X \to [0,1]$, the definition (1) of the homomorphism density $t(F,W)$ of F in W also makes sense for a multigraph F. We define the treewidth of a multigraph analogously to the case of simple graphs, i.e., we do not take the edge multiplicities into account. Note that, since the class of multigraphs of treewidth k is closed under taking disjoint unions, we could always assume the graphs in Characterization 1 to be connected. For example, in the case $k = 2$, it can also be phrased in terms of trees with parallel edges.

Two graphons U, W are weakly isomorphic, i.e., $t(F,U) = t(F,W)$ for every graph F, if and only if $t(F,U) = t(F,W)$ for every multigraph F [16, Corollary 10.36]. When restricting the treewidth, however, parallel edges do make a difference, cf. Figure 1: These weighted graphs have the same tree homomorphism densities as the coarsest stable partition of the graph on

Figure 1: Two fractionally isomorphic weighted graphs that are distinguished by oblivious 2-WL.

the left is the trivial partition, and the graph on the right is obtained by averaging the edge weights, cf. Characterization 5 of Theorem 3. However, already the multigraph C_2, i.e., two vertices connected by two parallel edges, distinguishes these weighted graphs, i.e., graphons that are not distinguished by oblivious 2-WL (in the sense of Theorem 4) are also not distinguished by color refinement (in the sense of Theorem 3), but the converse does not hold. Hence, while the difference between color refinement and 1-WL (corresponding to oblivious 2-WL) usually is neglected in the case of graphs, it is important to make a distinction in the more general case of graphons. Another way to phrase this is that color refinement and oblivious 2-WL are two different notions that coincide on the special case of simple graphs: if F is a multigraph and G a simple graph, then $t(F, G)$ is unaffected if we merge parallel edges of F into single edges since they have to be mapped to the same edges of G anyway. That is, just as Theorem 1 can be recovered from Theorem 3, Theorem 2 can be recovered from Theorem 4.

Characterization 2 generalizes oblivious k-WL. First, we define the the standard Borel space $\mathbb{M}^k$, which again can be seen as the space of colors used by oblivious k-WL. Also in this case, its elements $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots)$ are sequences of colors after $0, 1, 2, \dots$ refinement rounds. Based on the definition (3) of oblivious k-WL for graphs, we define the measurable function $\mathsf{i}_W^k \colon X^k \to \mathbb{M}^k$ mapping an $\bar{x} \in X^k$ to a sequence $(\alpha_0, \alpha_1, \alpha_2, \dots)$. In particular, α_0 corresponds to the "atomic type" of $\bar{x}$, which also further explains why oblivious 2-WL distinguishes the weighted graphs in Figure 1: For the weighted graph on the right, α_0 always contains the edge weight of $\frac{2}{3}$ which is nowhere to be found in the graph on the left. Hence, already the initial coloring distinguishes them. To continue, we then use i_W^k to define the *k-WL distribution (k-WLD)* ν_W^k as the push-forward of $\mu^{\otimes k}$ via i_W^k, a probability measure on X^k which again corresponds to the multiset of colors computed by oblivious k-WL.

The operator $T_W \colon L^2(X, \mu) \to L^2(X, \mu)$ of a graphon $W \colon X \times X \to [0, 1]$ plays an important role throughout Theorem 3, although it only becomes really apparent in the characterization via Markov operators. In Theorem 4, we replace this single operator by a whole family $\mathbb{T}_W^k$ of operators on the product space $L^2(X^k, \mu^{\otimes k}) \coloneqq L^2(X^k, \mathcal{B}^{\otimes k}, \mu^{\otimes k})$. We define a set $\mathcal{F}^k$ of *bi-labeled graphs* that serve as building blocks to construct precisely the graphs of treewidth at most $k-1$, and every such bi-labeled graph $\boldsymbol{F} \in \mathcal{F}^k$ together with a graphon $W \colon X \times X \to [0, 1]$ defines the *graphon operator* $T_{\boldsymbol{F} \to W}$. Then, $\mathbb{T}_W^k \coloneqq (T_{\boldsymbol{F} \to W})_{\boldsymbol{F} \in \mathcal{F}^k}$ denotes the family of all these operators. Characterization 4 states that there is a Markov operator on the product space $L^2(X^k, \mu^{\otimes k})$ that "commutes" with all operators in the families $\mathbb{T}_U^k$ and $\mathbb{T}_W^k$ simultaneously. Moreover, this operator can be assumed to be *permutation-invariant*, i.e., reordering the k components of X^k yields the same operator, an assumption that is implicitly made in the system $\mathsf{L}_{\text{iso}}^k$ of linear equations as its variables are indexed by sets. Permutation invariance can be left out without changing the equivalence to the other characterizations, i.e., if there is a (not necessarily permutation-invariant) Markov operator S satisfying Characterization 4, then there also is a permutation invariant one.

Characterizations 3 and 5 generalize (coarsest) stable partitions of $V(G)^k$. For a graphon

$W\colon X \times X \to [0,1]$, a $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ is called W-invariant if it is $\mathbb{T}^k_W$*-invariant*, i.e., T-invariant for every operator T in the family $\mathbb{T}^k_W$. In the case $k = 1$, this conflicts with the definition of Grebík and Rocha, but it will always be clear from the context what we mean. We show that the minimum W-invariant $\mu^{\otimes k}$-relatively complete sub-σ-algebra $\mathcal{C}^k_W$ of $\mathcal{B}^{\otimes k}$ can be obtained by iterative applications of the operators in $\mathbb{T}^k_W$. Then, Characterization 3 states that there is a *Markov isomorphism* from one quotient space to the other that "commutes" with all operators in the families of quotient operators $\mathbb{T}^k_W/\mathcal{C}^k_W$ and $\mathbb{T}^k_U/\mathcal{C}^k_U$ simultaneously; intuitively, for a $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ and an operator T on $L^2(X, \mu)$, its quotient operator $T/\mathcal{C}$ on $L^2(X/\mathcal{C}, \mu/\mathcal{C})$ is defined by going from $L^2(X/\mathcal{C}, \mu/\mathcal{C})$ to $L^2(X, \mu)$, applying T, and then going back to $L^2(X/\mathcal{C}, \mu/\mathcal{C})$. A Markov operator is called a *Markov embedding* if it is an isometry, and a *Markov isomorphism* is a surjective Markov embedding. There is a one-to-one correspondence between Markov isomorphisms and measure-preserving almost bijections, cf. [9, Theorem E.3], but for the ease of presentation, we stick to Markov isomorphisms.

Note that, in contrast to Theorem 3, there are no quotient graphons involved in Theorem 4, just quotient operators. The reason for this is that, unlike T_W, the operators in the family $\mathbb{T}^k_W$ are not integral operators. For our proof, this also means that we do not have a canonical representation of a graphon $W\colon X \times X \to [0,1]$ as a graphon $\mathbb{M}^k \times \mathbb{M}^k \to [0,1]$ (or as multiple such graphons). Instead, we define canonical representations of the operators in $\mathbb{T}^k_W$ on the space $L^2(\mathbb{M}^k, \nu^k_W)$ by hand.

In Section 2, the preliminaries, we collect some more definitions and basics we need. Section 3 introduces *bi-labeled graphs* and *graphon operators*, which are the key to our main theorem. In particular, we define the set $\mathcal{F}^k$ of bi-labeled graphs from which we are able to construct precisely the multigraphs of treewidth k. For a graphon W, this set of bi-labeled graphs defines the family of graphon operators $\mathbb{T}^k_W$ that takes the place of the usual integral operator T_W. Section 4 is the main section of this paper and closely follows Grebík and Rocha [9] in the definition of all notions in and the proof of Theorem 4. In Section 5, we show that it is also possible to define a variant of k-WL, which we call *simple k-WL*, that leads to a variant of Theorem 4 where the characterization by multigraph homomorphism densities is replaced by simple graph homomorphism densities. This variant of Theorem 4, however, is less elegant and has an artificial touch to it. Most of the proofs are left out as they are mostly analogous to the ones in Section 4. We draw some conclusions and discuss some open problems in Section 6.

2 Preliminaries

2.1 Product Spaces

Recall that, throughout the whole paper, $(X, \mathcal{B})$ denotes a *standard Borel space*, i.e., $\mathcal{B}$ is the Borel σ-algebra of a Polish space, and μ a *Borel probability measure* on X. We often consider the space $(X^k, \mathcal{B}^{\otimes k}, \mu^{\otimes k})$ with the product σ-algebra $\mathcal{B}^{\otimes k}$ of $\mathcal{B}$ and the product measure $\mu^{\otimes k}$ of μ for $k \geq 1$. The product of a countable family of standard Borel spaces is again a standard Borel space [15, Section 12.B]. Moreover, for a countable family of standard Borel spaces, its product σ-algebra is actually equal to the Borel σ-algebra of the product topology of the underlying Polish spaces as Polish spaces are second countable [15, Section 11.A]. Hence, the product space $(X^k, \mathcal{B}^{\otimes k})$ is again a standard Borel space and $\mathcal{B}^{\otimes k}$ is equal to the Borel σ-algebra of the product topology of the Polish space underlying $(X, \mathcal{B})$. For simplicity, we identify the products $X \times X \times X$ and $(X \times X) \times X$ in the usual way. Then, also $\mathcal{B} \otimes \mathcal{B} \otimes \mathcal{B} = (\mathcal{B} \otimes \mathcal{B}) \otimes \mathcal{B}$ and $\mu \otimes \mu \otimes \mu = (\mu \otimes \mu) \otimes \mu$ [2, Section 18]. We treat higher-order products in the same way.

We often use the Tonelli-Fubini theorem, cf. [6, Theorem 4.4.5] and also [2, Theorem 18.3], which states that, for σ-finite measure spaces $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \nu)$ and a non-negative function

f on $X \times Y$ that is measurable for $\mathcal{S} \otimes \mathcal{T}$, we have

$$\int_{X\times Y} f \, d(\mu \times \nu) = \int_X \int_Y f(x,y) \, d\nu(y) \, d\mu(x) = \int_Y \int_X f(x,y) \, d\mu(x) \, d\nu(y).$$

In particular, the functions $x \mapsto \int_Y f(x,y)\, d\nu(y)$ and $y \mapsto \int_X f(x,y)\, d\mu(x)$ are measurable for $\mathcal{S}$ and $\mathcal{T}$, respectively. If f is not necessarily non-negative but integrable with respect to $\mu \times \nu$, then the same equations hold and the aforementioned functions are measurable on sets X' and Y' with $\mu(X \setminus X') = 0$ and $\nu(Y \setminus Y') = 0$, respectively.

2.2 Markov Operators

In general, for a measure space $(X, \mathcal{S}, \mu)$ and $1 \leq p \leq \infty$, the space $\mathcal{L}^p(X, \mu) \coloneqq \mathcal{L}^p(X, \mathcal{S}, \mu)$ consists of all measurable real-valued functions on X with $\|f\|_p < \infty$, and $L^p(X, \mu) \coloneqq L^p(X, \mathcal{S}, \mu)$ is obtained from $\mathcal{L}^p(X, \mu)$ by identifying functions that are equal μ-almost everywhere. The space $L^2(X, \mu)$ plays a special role among these spaces as it is a Hilbert space with the inner product given by $\langle f, g\rangle \coloneqq \int_X fg \, d\mu$. Besides $L^2(X, \mu)$, the space $L^\infty(X, \mu)$ also plays an important role in this paper. Note that, if μ is a probability measure, then we have $\|f\|_2 \leq \|f\|_\infty$ and, in particular, the inclusion $L^\infty(X, \mu) \subseteq L^2(X, \mu)$.

Given two normed linear spaces $(X, \|\cdot\|)$ and $(Y, |\cdot|)$, a function $T \colon X \to Y$ is called a (bounded linear) *operator* if it is Lipschitz and linear. If $(X, \|\cdot\|) = (Y, |\cdot|)$, then we just say that T is an *operator on* X. The *operator norm of* T is given by $\|T\| \coloneqq \sup\{|T(x)| \mid \|x\| \leq 1\} < \infty$, and if $\|T\| \leq 1$, then T is called a *contraction*. For probability spaces $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \nu)$ and an operator $T \colon L^2(X, \mu) \to L^2(Y, \nu)$, we call T an L^∞-contraction if its restriction to $L^\infty(X, \mu)$ yields a well-defined contraction $L^\infty(X, \mu) \to L^\infty(Y, \nu)$. To clearly distinguish this from T being a contraction $L^2(X, \mu) \to L^2(Y, \nu)$, we sometimes use the term L^2-contraction for this. Observe that the composition of two contractions yields a contraction, and in particular, the composition of L^2- and L^∞- contractions yields a L^2- and a L^∞-contraction, respectively.

For measure spaces $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \nu)$, the *Hilbert adjoint* of an operator $T \colon L^2(X, \mu) \to L^2(Y, \nu)$, is the unique operator $T^* \colon L^2(Y, \nu) \to L^2(X, \mu)$ satisfying $\langle Sf, g\rangle = \langle f, S^* g\rangle$ for all $f \in L^2(X, \mu), g \in L^2(Y, \nu)$. For standard Borel spaces $(X, \mathcal{B})$ and $(Y, \mathcal{D})$ with Borel probability measures μ and ν on X and Y, respectively, an operator $S \colon L^2(X, \mu) \to L^2(Y, \nu)$ is called a *Markov operator* if $Sf \geq 0$ for every $f \in L^2(X, \mu)$ with $f \geq 0$, $S\mathbf{1}_X = \mathbf{1}_Y$, and $S^*\mathbf{1}_Y = \mathbf{1}_X$. Markov operators are both L^2- and L^∞-contractions [8, Theorem 13.2 b)]. A Markov operator is called a *Markov embedding* if it is an isometry. For example, the *Koopman operator* $T_\varphi \colon L^2(X, \mu) \to L^2(X, \mu)$ of a measure-preserving measurable map $\varphi \colon X \to X$, defined by $T_\varphi f \coloneqq f \circ \varphi$ for every $f \in L^2(X, \mu)$, is a Markov embedding [8, Example 13.1]. A *Markov isomorphism* is a surjective Markov embedding. Note that every Markov isomorphism S satisfies $S^{-1} = S^*$ [8, Corollary 13.14]. Moreover, there is a one-to-one correspondence between Markov isomorphisms and measure-preserving almost bijections, cf. [9, Theorem E.3]. See [8] for a thorough treatment of Markov operators. There, the results are stated for complex L^p-spaces, but this usually does not make a difference by the positivity of Markov operators, cf. [8, Lemma 7.5].

2.3 Quotient Spaces

Recall that a sub-σ-algebra $\mathcal{C} \subseteq \mathcal{B}$ of $\mathcal{B}$ is called *μ-relatively complete* if $Z \in \mathcal{C}$ for all $Z \in \mathcal{B}, Z_0 \in \mathcal{C}$ with $\mu(Z \triangle Z_0) = 0$. Note that requiring $Z \in \mathcal{C}$ for every $Z \in \mathcal{B}$ with $\mu(Z) = 0$ instead would yield an equivalent definition. The set of all μ-relatively complete sub-σ-algebras of $\mathcal{B}$ is denoted by $\Theta(\mathcal{B}, \mu)$ and clearly includes $\mathcal{B}$ itself. For a non-empty $\Phi \subseteq \Theta(\mathcal{B}, \mu)$, we have $\bigcap \Phi := \bigcap_{\mathcal{C}\in\Phi} \mathcal{C} \in \Theta(\mathcal{B}, \mu)$ [9, Claim 5.4]. Hence, for a set $\mathcal{X} \subseteq \mathcal{B}$, there is a smallest μ-relatively complete sub-σ-algebra including $\mathcal{X}$, which we denote by $\langle \mathcal{X} \rangle$. Note that $\langle \mathcal{C} \rangle = \{A \triangle Z \mid A \in \mathcal{C}, Z \in \mathcal{B} \text{ with } \mu(Z) = 0\}$ for a sub-σ-algebra $\mathcal{C} \subseteq \mathcal{B}$. Given $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$, we let $L^2(X, \mathcal{C}, \mu) \subseteq L^2(X, \mu)$ denote the subset of all functions that are $\mathcal{C}$-measurable. It is a standard fact that, for $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$, the linear hull of $\{\mathbf{1}_A\}_{A\in\mathcal{C}}$ is dense in $L^2(X, \mathcal{C}, \mu)$.

Claim 5 (Conditional Expectation, [2, Section 34])**.** *Let $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$. Then, $L^2(X, \mathcal{C}, \mu)$ is a closed linear subspace of $L^2(X, \mu)$ and $\mathbb{E}(- \mid \mathcal{C}) \colon L^2(X, \mu) \to L^2(X, \mu)$ is a self-adjoint operator such that*

1. *$\mathbb{E}(- \mid \mathcal{C})$ is the orthogonal projection onto $L^2(X, \mathcal{C}, \mu)$,*
2. *$\int_A f \, d\mu = \int_A \mathbb{E}(f \mid \mathcal{C}) \, d\mu$ for every $A \in \mathcal{C}$ and every $f \in L^2(X, \mu)$, and*
3. *$\int_X f \cdot \mathbb{E}(g \mid \mathcal{C}) \, d\mu = \int_X \mathbb{E}(f \mid \mathcal{C}) \cdot g \, d\mu$ for all $f, g \in L^2(X, \mu)$.*

Let $k \geq 1$ and consider $L^2(X^k, \mu^{\otimes k})$. Every permutation $\pi \colon [k] \to [k]$ induces a measure-preserving measurable map $\pi \colon X^k \to X^k$ by setting $\pi(x_1, \dots, x_k) := (x_{\pi(1)}, \dots, x_{\pi(k)})$ for all $x_1, \dots, x_k \in X$, which allows us to consider its Koopman operator T_π on $L^2(X^k, \mu^{\otimes k})$. Clearly, the adjoint of T_π is given by $T_{\pi^{-1}}$. We call a $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ *permutation invariant* if $\mathcal{C}$ is T_π-invariant for every permutation $\pi \colon [k] \to [k]$. It is easy to see that this is the case if and only if $\pi(\mathcal{C}) \subseteq \mathcal{C}$ for every permutation $\pi \colon [k] \to [k]$, which again is equivalent to $\pi(\mathcal{C}) = \mathcal{C}$ for every permutation $\pi \colon [k] \to [k]$. A trivial example of such a permutation-invariant sub-σ-algebra is $\mathcal{B}^{\otimes k}$ itself.

Given a measure space $(X, \mathcal{S}, \mu)$, a measurable space $(Y, \mathcal{T})$, and a measurable function $g \colon X \to Y$, the *push-forward* $g_*\mu$ is the measure on Y defined by $g_*\mu(A) := \mu(g^{-1}(A))$ for every $A \in \mathcal{T}$. For a measurable function $f \colon Y \to [-\infty, \infty]$, we then have $\int_Y f \, d(g_*\mu) = \int_X f \circ g \, d\mu$ [6, Theorem 4.1.11]. The following claim states the existence of quotient spaces.

Claim 6 ([9, Theorem E.1])**.** *Let $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$. There is a standard Borel space $(X/\mathcal{C}, \mathcal{C}')$, a Borel probability measure $\mu/\mathcal{C}$ on $X/\mathcal{C}$, a measurable surjection $q_\mathcal{C} \colon X \to X/\mathcal{C}$, and Markov operators $S_\mathcal{C} \colon L^2(X, \mu) \to L^2(X/\mathcal{C}, \mu/\mathcal{C})$ and $I_\mathcal{C} \colon L^2(X/\mathcal{C}, \mu/\mathcal{C}) \to L^2(X, \mu)$ such that*

1. *$I_\mathcal{C}$ is the Koopman operator of $q_\mathcal{C}$,*
2. *$\mu/\mathcal{C}$ is the push-forward of μ via $q_\mathcal{C}$,*
3. *$S_\mathcal{C}^* = I_\mathcal{C}$,*
4. *$S_\mathcal{C} \circ \mathbb{E}(- \mid \mathcal{C}) = S_\mathcal{C}$,*
5. *$I_\mathcal{C}$ is an isometry onto $L^2(X, \mathcal{C}, \mu)$,*
6. *$I_\mathcal{C} \circ S_\mathcal{C} = \mathbb{E}(- \mid \mathcal{C})$, and*
7. *$S_\mathcal{C} \circ I_\mathcal{C}$ is the identity on $L^2(X/\mathcal{C}, \mu/\mathcal{C})$.*

Claim 7 essentially states that the quotient space $(X/\mathcal{C}, \mathcal{C}')$ is unique up to sets of measure zero.

Claim 7 ([9, Corollary E.2])**.** *Let $(X, \mathcal{B})$ and $(Y, \mathcal{D})$ be standard Borel spaces. Let μ be a Borel probability measure on X and $f \colon X \to Y$ be a measurable function. Let $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$ be the minimum μ-relatively complete sub-σ-algebra that makes f measurable. Then, for every $g_0 \in L^2(X, \mathcal{C}, \mu)$, there is a measurable map $g_1 \colon Y \to \mathbb{R}$ such that $g_0(x) = (g_1 \circ f)(x)$ for μ-almost every $x \in X$.*

For $\mathcal{C}, \mathcal{D} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$, an operator $T \colon L^2(X^k/\mathcal{C}, \mu^{\otimes k}/\mathcal{C}) \to L^2(X^k/\mathcal{D}, \mu^{\otimes k}/\mathcal{D})$ is called *permutation invariant* if $T_\pi/\mathcal{D} \circ T = T \circ T_\pi/\mathcal{C}$ for every permutation $\pi \colon [k] \to [k]$. For the special case $\mathcal{C} = \mathcal{D} = \mathcal{B}^{\otimes k}$, this means that an operator T on $L^2(X^k, \mu^{\otimes k})$ is permutation invariant if $T_\pi \circ T = T \circ T_\pi$ for every permutation $\pi \colon [k] \to [k]$. Of course, this notion depends on the underlying space $(X, \mathcal{B}, \mu)$, i.e., if we consider $(X^k, \mathcal{B}^{\otimes k}, \mu^{\otimes k})$ as the underlying space, then all these operators mentioned before are trivially permutation invariant. However, since the intended underlying space is always clear from the context, we just use the term permutation invariant. It is not hard to prove that, if $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ is permutation invariant, then so are $S_{\mathcal{C}}$ and $I_{\mathcal{C}}$, i.e., $T_\pi/\mathcal{C} \circ S_{\mathcal{C}} = S_{\mathcal{C}} \circ T_\pi$ and $T_\pi \circ I_{\mathcal{C}} = I_{\mathcal{C}} \circ T_\pi/\mathcal{C}$ for every permutation $\pi\colon [k] \to [k]$.

2.4 Quotient Operators

For $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$ and an operator $T\colon L^2(X, \mu) \to L^2(X, \mu)$, we use the conditional expectation to define the operators $T_{\mathcal{C}}\colon L^2(X, \mu) \to L^2(X, \mu)$ and $T/\mathcal{C}\colon L^2(X/\mathcal{C}, \mu/\mathcal{C}) \to L^2(X/\mathcal{C}, \mu/\mathcal{C})$ by

$$T_{\mathcal{C}} := \mathbb{E}(- \mid \mathcal{C}) \circ T \circ \mathbb{E}(- \mid \mathcal{C}) \qquad \text{and} \qquad T/\mathcal{C} := S_{\mathcal{C}} \circ T \circ I_{\mathcal{C}},$$

respectively. These definitions reflect the same concept of a quotient operator via different languages. The following lemma states some basic properties and shows how both definitions are related.

Lemma 8. *Let $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$ and $T\colon L^2(X, \mu) \to L^2(X, \mu)$ be an operator. Then,*

1. *$(T_{\mathcal{C}})^* = (T^*)_{\mathcal{C}}$ and $(T/\mathcal{C})^* = T^*/\mathcal{C}$,*
2. *if T is self-adjoint, then so are $T_{\mathcal{C}}$, $T/\mathcal{C}$,*
3. *$I_{\mathcal{C}} \circ T/\mathcal{C} = T_{\mathcal{C}} \circ I_{\mathcal{C}}$,*
4. *$T/\mathcal{C} \circ S_{\mathcal{C}} = S_{\mathcal{C}} \circ T_{\mathcal{C}}$,*
5. *if $\mathcal{C}$ is T-invariant, then $T_{\mathcal{C}} = T \circ \mathbb{E}(- \mid \mathcal{C})$ and $I_{\mathcal{C}} \circ T/\mathcal{C} = T \circ I_{\mathcal{C}}$, and*
6. *if T is self-adjoint and $\mathcal{C}$ is T-invariant, then $T/\mathcal{C} \circ S_{\mathcal{C}} = S_{\mathcal{C}} \circ T$.*

Proof. For 1, we have $(T_{\mathcal{C}})^* = \mathbb{E}(- \mid \mathcal{C})^* \circ T^* \circ \mathbb{E}(- \mid \mathcal{C})^* = \mathbb{E}(- \mid \mathcal{C}) \circ T^* \circ \mathbb{E}(- \mid \mathcal{C}) = T_{\mathcal{C}}$ by Claim 5 and $(T/\mathcal{C})^* = I_{\mathcal{C}}^* \circ T^* \circ S_{\mathcal{C}}^* = S_{\mathcal{C}} \circ T^* \circ I_{\mathcal{C}} = T^*/\mathcal{C}$ by 3 of Claim 6. This also immediately yields 2. For 3, we have

$$I_{\mathcal{C}} \circ T/\mathcal{C} = I_{\mathcal{C}} \circ S_{\mathcal{C}} \circ T \circ I_{\mathcal{C}} = \mathbb{E}(- \mid \mathcal{C}) \circ T \circ I_{\mathcal{C}} = \mathbb{E}(- \mid \mathcal{C}) \circ T \circ \mathbb{E}(- \mid \mathcal{C}) \circ I_{\mathcal{C}} = T_{\mathcal{C}} \circ I_{\mathcal{C}}$$

by 6 and 4 of Claim 6 and Claim 5. For 4, we have

$$T/\mathcal{C} \circ S_{\mathcal{C}} = S_{\mathcal{C}} \circ T \circ I_{\mathcal{C}} \circ S_{\mathcal{C}} = S_{\mathcal{C}} \circ \mathbb{E}(- \mid \mathcal{C}) \circ T \circ \mathbb{E}(- \mid \mathcal{C}) = S_{\mathcal{C}} \circ T_{\mathcal{C}}$$

by 4 and 6 of Claim 6.

For 5, assume that $\mathcal{C}$ is T-invariant. By Claim 5, the expectation $\mathbb{E}(- \mid \mathcal{C})$ is the orthogonal projection onto $L^2(X, \mathcal{C}, \mu)$. Hence, $(T \circ \mathbb{E}(- \mid \mathcal{C}))(L^2(X, \mu)) = T(L^2(X, \mathcal{C}, \mu)) \subseteq L^2(X, \mathcal{C}, \mu)$ and, as $\mathbb{E}(- \mid \mathcal{C})$ is the identity on $L^2(X, \mathcal{C}, \mu)$, the first claim $T_{\mathcal{C}} = T \circ \mathbb{E}(- \mid \mathcal{C})$ follows. Then, continuing with 3, we get $I_{\mathcal{C}} \circ T/\mathcal{C} = T_{\mathcal{C}} \circ I_{\mathcal{C}} = T \circ \mathbb{E}(- \mid \mathcal{C}) \circ I_{\mathcal{C}} = T \circ I_{\mathcal{C}}$ by 4 of Claim 6 and Claim 5. Now, 6 follows from 2, 5, and 3 of Claim 6. ◻

The following lemma is an application of the Mean Ergodic Theorem for Hilbert spaces to Markov operators [8, Theorem 8.6, Example 13.24] and is essentially the essence of the proof of the direction " 4 $\implies$ 5" of Theorem 3 by Grebík and Rocha [9].

Lemma 9. *Let $S\colon L^2(X, \mu) \to L^2(X, \mu)$ be a Markov operator. There are $\mathcal{C}, \mathcal{D} \in \Theta(\mathcal{B}, \mu)$ with*

1. *$L^2(X, \mathcal{C}, \mu) = \{f \in L^2(X, \mu) \mid (S \circ S^*)f = f\}$,*
2. *$L^2(X, \mathcal{D}, \mu) = \{f \in L^2(X, \mu) \mid (S^* \circ S)f = f\}$,*
3. *$\mathbb{E}(- \mid \mathcal{C}) \circ S = S \circ \mathbb{E}(- \mid \mathcal{D})$,*
4. *$R := S_{\mathcal{C}} \circ S \circ I_{\mathcal{D}}\colon L^2(X/\mathcal{D}, \mu/\mathcal{D}) \to L^2(X/\mathcal{C}, \mu/\mathcal{C})$ is a Markov isomorphism, and*
5. *for all operators $T_1, T_2\colon L^2(X, \mu) \to L^2(X, \mu)$ with $T_1 \circ S = S \circ T_2$ and $S^* \circ T_1 = T_2 \circ S^*$,*

(a) $\mathcal{C}$ is T_1-invariant, *(b) $\mathcal{D}$ is T_2-invariant, and* *(c) $T_1/\mathcal{C} \circ R = R \circ T_2/\mathcal{D}$.*

Proof. The proof of the existence of $\mathcal{C}, \mathcal{D} \in \Theta(\mathcal{B}, \mu)$ satisfying 1 to 4 uses the Mean Ergodic Theorem and is identical to the the proof of Theorem 1.2, (4) $\Rightarrow$ (5), in [9]; we leave it out here. To prove 5, let $T_1, T_2 \colon L^2(X, \mu) \to L^2(X, \mu)$ be bounded linear operators satisfying $T_1 \circ S = S \circ T_2$ and $S^* \circ T_1 = T_2 \circ S^*$. We get $T_1 \circ (S \circ S^*) = S \circ T_2 \circ S^* = (S \circ S^*) \circ T_1$. Then, for $f \in L^2(X, \mathcal{C}, \mu)$, we have $(S \circ S^*)f = f$ by 1 and get $T_1 f = (T_1 \circ S \circ S^*)f = (S \circ S^* \circ T_1)f = (S \circ S^*)(T_1 f)$, which, again by 1, implies $T_1 f \in L^2(X, \mathcal{C}, \mu)$. Therefore, $\mathcal{C}$ is T_1-invariant, which proves 5a. Analogously, we get that $T_2 \circ (S^* \circ S) = (S^* \circ S) \circ T_2$ and that $\mathcal{D}$ is T_2 invariant, which proves 5b. Now, we use 3 and the T_2-invariance of $\mathcal{D}$ to obtain to obtain

$$
\begin{aligned}
T_1/\mathcal{C} \circ R = S_{\mathcal{C}} \circ T_1 \circ I_{\mathcal{C}} \circ S_{\mathcal{C}} \circ S \circ I_{\mathcal{D}} &= S_{\mathcal{C}} \circ T_1 \circ \mathbb{E}(- \mid \mathcal{C}) \circ S \circ I_{\mathcal{D}} && \text{(Claim 6 6)} \\
&= S_{\mathcal{C}} \circ T_1 \circ S \circ \mathbb{E}(- \mid \mathcal{D}) \circ I_{\mathcal{D}} && (3) \\
&= S_{\mathcal{C}} \circ T_1 \circ S \circ I_{\mathcal{D}} && \text{(Claim 6 3, 4)} \\
&= S_{\mathcal{C}} \circ S \circ T_2 \circ I_{\mathcal{D}} \\
&= S_{\mathcal{C}} \circ S \circ I_{\mathcal{D}} \circ T_2/\mathcal{D} && \text{(Lemma 8 5)} \\
&= R \circ T_2/\mathcal{D}.
\end{aligned}
$$

□

3 Graphon Operators

In this section, we present the key ingredient to Theorem 4. The key insight to go from color refinement to k-WL is, for a graphon W, to replace the operator T_W on $L^2(X, \mu)$ by a family $\mathbb{T}_W^k$ of operators on the product space $L^2(X^k, \mu^{\otimes k})$. This idea is somewhat already present in the work of Grohe and Otto [11, Section 5.1], where they define a family of graphs and consider a matrix X such that X is a fractional isomorphism between all these graphs simultaneously. The graphon setting shows that the step of defining these graphs for the sake of them having the right adjacency matrix is rather artificial; the operators we define are not integral operators defined by a graphon.

The family $\mathbb{T}_W^k$ we define is closely related to oblivious k-WL and tree decompositions, or more precisely, tree-decomposed graphs. In Section 3.1, we follow the approach of [17] of using a set of *bi-labeled graphs* as building blocks that are then glued together to form larger graphs. From our set $\mathcal{F}^k$ of bi-labeled graphs, we obtain precisely the multigraphs of treewidth at most $k-1$. In Section 3.2, we adapt the concept of *homomorphism matrices* of bi-labeled graphs from [17] by defining the *graphon operator* of a bi-labeled graph and a graphon. The graphon operators of our building blocks then yield the family $\mathbb{T}_W^k$. We show how this family is related to homomorphisms: on the level of bi-labeled graphs, we obtain all multigraphs of treewidth at most $k-1$, while we obtain all *homomorphism functions* of multigraphs of treewidth at most $k-1$ on the operator level.

3.1 Bi-Labeled Graphs

A *bi-labeled graph* $\boldsymbol{G}$ is a triple $(G, \boldsymbol{a}, \boldsymbol{b})$, where G is a multigraph and $\boldsymbol{a} \in V(G)^k$, $\boldsymbol{b} \in V(G)^\ell$ for $k, \ell \geq 0$ are vectors of vertices such that both the entries of $\boldsymbol{a}$ and the entries of $\boldsymbol{b}$ are pairwise distinct. When there is no fear of ambiguity, we sometimes just use the term *graph* to refer to a bi-labeled graph. The multigraph G is called the *underlying graph* of $\boldsymbol{G}$, and the vectors $\boldsymbol{a}$ and

$\boldsymbol{b}$ are called the vectors of *input* and *output vertices*, respectively. That is, a bi-labeled graph is a multigraph where additionally *input* and *output labels* are assigned to the vertices with every vertex having at most one label of each type. Note that one usually does not require that every vertex has at most one label of each type, cf. [17]., but this is needed to ensure that graphon operators are well defined; the reason is that the diagonal in the product space $(X^k, \mathcal{B}^{\otimes k}, \mu^{\otimes k})$ has measure zero (as long as our standard Borel space is atom free), a problem which one does not face in the finite-dimensional case.

Two bi-labeled graphs $\boldsymbol{G} = (G, \boldsymbol{a}, \boldsymbol{b})$ and $\boldsymbol{G'} = (G', \boldsymbol{a'}, \boldsymbol{b'})$ are *isomorphic* if there is an isomorphism $\varphi\colon V(G) \to V(G')$ from G to G' such that $\varphi(\boldsymbol{a}) = \boldsymbol{a'}$ and $\varphi(\boldsymbol{b}) = \boldsymbol{b'}$. For $k, \ell \geq 0$, let $\mathcal{M}^{k,\ell}$ denote the set of all (isomorphism types of) bi-labeled graphs with k input and ℓ output vertices, and let $\mathcal{G}^{k,\ell} \subseteq \mathcal{M}^{k,\ell}$ be the subset whose underlying graphs are simple. Let $\mathcal{M} \coloneqq \cup_{k,\ell \geq 0} \mathcal{M}^{k,\ell}$ and $\mathcal{G} \coloneqq \cup_{k,\ell \geq 0} \mathcal{G}^{k,\ell}$.

The *transpose* of a bi-labeled graph $\boldsymbol{G} = (G, \boldsymbol{a}, \boldsymbol{b}) \in \mathcal{M}^{k,\ell}$ is the bi-labeled graph $\boldsymbol{G}^* \coloneqq (G, \boldsymbol{b}, \boldsymbol{a}) \in \mathcal{M}^{\ell,k}$, and $\boldsymbol{G}$ is called *symmetric* if $\boldsymbol{G}^* = \boldsymbol{G}$. The *composition* of two bi-labeled graphs $\boldsymbol{F_1} = (F_1, \boldsymbol{a_1}, \boldsymbol{b_1}) \in \mathcal{M}^{k,m}$ and $\boldsymbol{F_2} = (F_2, \boldsymbol{a_2}, \boldsymbol{b_2}) \in \mathcal{M}^{m,\ell}$ is the bi-labeled graph $\boldsymbol{F_1} \circ \boldsymbol{F_2} \coloneqq (F, \boldsymbol{a_1}, \boldsymbol{b_2}) \in \mathcal{M}^{k,\ell}$, where F is obtained from the disjoint union of F_1 and F_2 by identifying vertices $b_{1,i}$ and $a_{2,i}$ for every $i \in [m]$. The *Schur product* of two bi-labeled graphs without output labels $\boldsymbol{F_1} = (F_1, \boldsymbol{a_1}, ()), \boldsymbol{F_2} = (F_2, \boldsymbol{a_2}, ()) \in \mathcal{M}^{k,0}$ is the bi-labeled graph $\boldsymbol{F_1} \cdot \boldsymbol{F_2} \coloneqq (F, \boldsymbol{a_1}, ()) \in \mathcal{M}^{k,0}$, where F is obtained from the disjoint union of F_1 and F_2 by identifying vertices $a_{1,i}$ and $a_{2,i}$ for every $i \in [m]$. One usually defines the Schur product for general bi-labeled graphs in $\mathcal{M}^{k,\ell}$ by also identifying output vertices, cf. [17]. This, however, can result in vertices with multiple input or output labels, which we do not allow by our definition of a bi-labeled graph as remarked earlier.

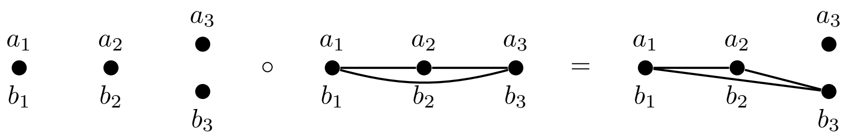

Figure 2: Composition of bi-labeled graphs.

Treewidth is a graph parameter that measures how "tree-like" a graph is. Too see how the concept is related to the just introduced bi-labeled graphs, let us first recall the usual definition of treewidth via *tree decompositions*. Formally, a *tree decomposition* of a multigraph G is a pair (T, β), where T is a tree and $\beta\colon V(T) \to 2^{V(G)}$ such that,

1. for every $v \in V(G)$, the set $\{t \mid v \in \beta(t)\}$ is non-empty and connected and,
2. for every $uv \in E(G)$, there is a $t \in V(T)$ such that $u, v \in \beta(t)$.

For every $t \in V(T)$, the set $\beta(t)$ is called the *bag* at t. The *width* of the tree decomposition (T, β) is $\max\{|\beta(t)| \mid t \in V(T)\} - 1$. The *treewidth* $\mathrm{tw}(G)$ of a multigraph G is the minimum of the widths of all tree decompositions of G. Note that treewidth is usually defined for simple graphs and not for multigraphs, but for us, ignoring the edge multiplicities like in the previous definition yields just the right notion for multigraphs. For the sake of completeness, note that *path decompositions* and *pathwidth* of a multigraph G can be defined analogously by only considering tree decomposition (T, β) where T is a path.

General tree decompositions are impractical to work with, and we rather use the following restricted form of a tree decomposition: a *nice tree decomposition* of a multigraph G is a triple (T, r, β) where (T, β) is a tree decomposition of G and $r \in V(T)$ a vertex of T, which we view as the root of T, such that

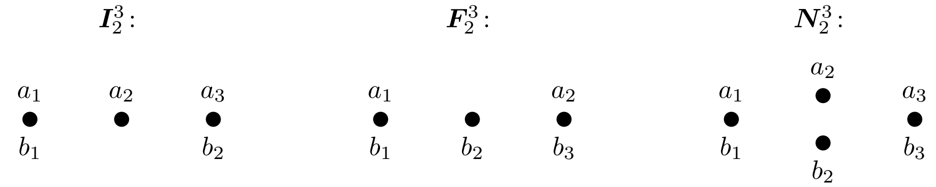

Figure 3: The bi-labeled graphs $\boldsymbol{I}_2^3$, $\boldsymbol{F}_2^3$, and $\boldsymbol{N}_2^3$.

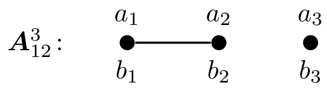

Figure 4: The bi-labeled graph $\boldsymbol{A}_{12}^3$.

1. $\beta(r) = \varnothing$ and $\beta(t) = \varnothing$ for every leaf t of (T, r) and
2. every internal node $s \in V(T)$ of T is of one of the following three types:

   Introduce node: s has exactly one child t with $\beta(s) = \beta(t) \cup v$ for some $v \in V(G) \setminus \beta(t)$.

   Forget node: s has exactly one child t with $\beta(s) \cup v = \beta(t)$ for some $v \in V(G) \setminus \beta(s)$.

   Join node: s has exactly two children t_1, t_2 with $\beta(s) = \beta(t_1) = \beta(t_2)$.

The *width* of (T, r, β) is the width of (T, β). It is well-known that every graph G has a nice tree decomposition of width $\mathrm{tw}(G)$.

Nice tree decompositions can be interpreted in terms of bi-labeled graphs: The vertices with input labels (and also the vertices with output labels) form a bag. An introduce node adds a fresh vertex with an input label. A forget node removes an input label from a vertex. A join node glues the input vertices of a bi-labeled graph to the input vertices of another bi-labeled graph. Hence, a join node is just the Schur product of the two bi-labeled graphs. The behavior of introduce and forget nodes corresponds to the composition with certain bi-labeled graphs, which we call *introduce* and *forget* graphs for this reason.

Definition 10 (Introduce, Forget, and Neighbor Graphs)**.** *Let $k \geq 1$. For $j \in [k]$, define*

1. *the j-introduce graph $\boldsymbol{I}_j^k := (([k], \varnothing), (1, \dots, k), (1, \dots, j-1, j+1, \dots, k)) \in \mathcal{G}^{k,k-1}$,*
2. *the j-forget graph $\boldsymbol{F}_j^k := {\boldsymbol{I}_j^k}^* \in \mathcal{G}^{k-1,k}$, and*
3. *the j-neighbor graph $\boldsymbol{N}_j^k := \boldsymbol{I}_j^k \circ \boldsymbol{F}_j^k \in \mathcal{G}^{k,k}$.*

Then, let $\mathcal{N}^k := \{\boldsymbol{N}_1^k, \dots, \boldsymbol{N}_k^k\} \subseteq \mathcal{G}^{k,k}$ be the set of all neighbor graphs.

Neighbor graphs correspond to a forget node that is immediately followed by an introduce node for the very same label. Considering these neighbor graphs instead of individual introduce and forget graphs has the advantage that our bi-labeled graphs always have both k input and k output labels, which means that we can restrict ourselves to the space $L^2(X^k, \mu^{\otimes k})$ later on. For our purposes, this is not a restriction as we can always add isolated vertices to a graph without affecting its homomorphism density in a graphon. Moreover, it is also not a restriction that the fresh vertex has to use the same label as the forgotten vertex since we may just inductively re-label the whole bi-labeled graph.

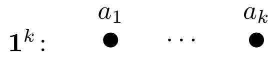

Figure 5: The bi-labeled graph $\mathbf{1}^k$.

By viewing bi-labeled graphs constructed from neighbor graphs by composition and the Schur product as tree decompositions, we are only halfway at our goal as we are missing a multigraph that is being decomposed. We rather have to view these bi-labeled graphs as tree-decomposed graphs, which we achieve by adding edges—but only between vertices in the same bag. Formally, we can add such an edge by the composition with an *adjacency graph*, a bi-labeled graph consisting just of a single edge and some isolated vertices.

Definition 11 (Adjacency Graphs)**.** *Let $k \geq 1$. For $i \neq j \in [k]$, define the ij-adjacency graph*

$$\boldsymbol{A}^k_{ij} \coloneqq (([k], \{ij\}), (1, \dots, k), (1, \dots, k)) \in \mathcal{G}^{k,k}.$$

Then, let $\mathcal{A}^k \coloneqq \left\{ \boldsymbol{A}^k_{ij} \mid i \neq j \in [k] \right\} \subseteq \mathcal{G}^{k,k}$ be the set of all adjacency graphs.

Having defined the set $\mathcal{N}^k$ of neighbor graphs and the set $\mathcal{A}^k$ of adjacency graphs, we can formalize our view of tree-decomposed graphs as *terms* built from these bi-labeled graphs by composition and the Schur product. For the sake of brevity, we define $\mathcal{F}^k \coloneqq \mathcal{N}^k \cup \mathcal{A}^k$, and for simplicity, we additionally define the *all-one graph*

$$\mathbf{1}^k \coloneqq (([k], \varnothing), (1, \dots, k), ()) \in \mathcal{G}^{k,0}.$$

for $k \geq 1$. It introduces k fresh vertices with input labels and serves as the leaves of our tree decompositions; this is much simpler than using k individual introduce graphs.

Definition 12. *Let $k \geq 1$. For a set $\mathcal{F} \subseteq \mathcal{M}^{k,k}$ of bi-labeled graphs with k input and k output labels, let $\langle \mathcal{F} \rangle_{\circ,\cdot}$ denote the smallest set of terms such that*

1. *$\mathbf{1}^k \in \langle \mathcal{F} \rangle_{\circ,\cdot}$,*
2. *$\boldsymbol{F} \circ \mathbb{F} \in \langle \mathcal{F} \rangle_{\circ,\cdot}$ for all $\boldsymbol{F} \in \mathcal{F}$, $\mathbb{F} \in \langle \mathcal{F} \rangle_{\circ,\cdot}$, and*
3. *$\mathbb{F}_1 \cdot \mathbb{F}_2 \in \langle \mathcal{F} \rangle_{\circ,\cdot}$ for all $\mathbb{F}_1, \mathbb{F}_2 \in \langle \mathcal{F} \rangle_{\circ,\cdot}$.*

Similarly, let $\langle \mathcal{F} \rangle_{\circ} \subseteq \langle \mathcal{F} \rangle_{\circ,\cdot}$ be the smallest set of terms satisfying 1 and 2. For a term $\mathbb{F} \in \langle \mathcal{F} \rangle_{\circ,\cdot}$, let $[\![\mathbb{F}]\!]$ denote the bi-labeled graph obtained from evaluating it.

Note that, for a set $\mathcal{F} \subseteq \mathcal{M}^{k,k}$ and a term $\mathbb{F} \in \langle \mathcal{F} \rangle_{\circ,\cdot}$, the bi-labeled graph $[\![\mathbb{F}]\!]$ is well-defined as we always have $[\![\mathbb{F}]\!] \in \mathcal{M}^{k,0}$. For the specific set $\mathcal{F}^k$ of neighbor and adjacency graphs, a term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ is essentially a tree-decomposed graph, where the tree decomposition is rooted, the multigraph being decomposed is the bi-labeled graph underlying $[\![\mathbb{F}]\!]$, and the bag at the root is given by the input vertices of $[\![\mathbb{F}]\!]$. As mentioned before, in terms of nice tree decompositions, the Schur product corresponds to a join node, composition with a neighbor graph corresponds to an introduce node followed by a forget node (when viewed from the root), and the composition with an adjacency graph adds an edge to a bag. The *height* $h(\mathbb{F})$ of a term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ is inductively defined by letting $h(\mathbf{1}^k) \coloneqq 0$, $h(\boldsymbol{N} \circ \mathbb{F}) \coloneqq h(\mathbb{F}) + 1$ for all $\boldsymbol{N} \in \mathcal{N}^k$, $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$, $h(\boldsymbol{A} \circ \mathbb{F}) \coloneqq h(\mathbb{F})$ for all $\boldsymbol{A} \in \mathcal{A}^k$, $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$, and $h(\mathbb{F}_1 \cdot \mathbb{F}_2) \coloneqq \max\{h(\mathbb{F}_1), h(\mathbb{F}_2)\}$ for all $\mathbb{F}_1, \mathbb{F}_2 \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$. Then, the height of $\mathbb{F}$ corresponds to the height of the tree of the tree decomposition when viewing $\mathbb{F}$ as a tree-decomposed graph.

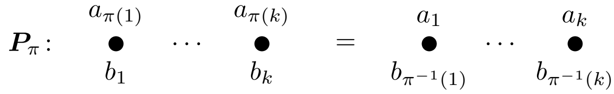

Figure 6: The graph $\boldsymbol{P}_\pi$.

Lemma 13. *The underlying graphs of the bi-labeled graphs obtained by evaluating the terms in $\langle \mathcal{F}^k \rangle_\circ$ and $\langle \mathcal{F}^k \rangle_{\circ,\cdot}$ are, up to isolated vertices, precisely the multigraphs of pathwidth and treewidth at most $k-1$, respectively.*

Proof. It is easy to see that a term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ encodes a tree decomposition of the underlying graph of $[\![\mathbb{F}]\!]$ of width $k-1$. Conversely, a nice tree decomposition (T, r, β) of a graph G of width at most $k-1$ can be turned into a term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ in a bottom-up fashion such that the underlying graph of $[\![\mathbb{F}]\!]$ is G with some additional isolated vertices: Note that a term fixes an ordering of the vertices of the graph, which we have to keep in mind in the following. First, pad the bag of every leaf to size k by adding k fresh isolated vertices. At an introduce node, add a forget node below that removes one of the isolated vertices. At a forget node, add an introduce node above adding a fresh isolated vertex. At a join node, re-order the vertices in one of the terms such that the original vertices of G are at the same positions in both terms and, then, identify every additional isolated vertex with the one at the same position in the other term. ☐

Lemma 13 would have been simplified if we included more graphs in $\mathcal{F}^k$: With individual introduce and forget graphs, we would not have to deal with isolated vertices. However, the price for this would be that we have to consider all product spaces $L^2(X^1, \mu^{\otimes 1}), \dots, L^2(X^k, \mu^{\otimes k})$ instead of just $L^2(X^k, \mu^{\otimes k})$. Similarly, we could have included graphs in $\mathcal{F}^k$ that allow to re-label input vertices; then we would not have to inductively re-label whole terms. But, also in this case it pays off to keep the set $\mathcal{F}^k$ as simple as possible. Let us briefly define these *permutation graphs* nevertheless since they come in handy when proving that the operators and sub-σ-algebras we define are *permutation invariant*. Formally, for $k \geq 1$ and a permutation $\pi \colon [k] \to [k]$, we define the *permutation graph*

$$\boldsymbol{P}_\pi := (([k], \varnothing), (1, \dots, k), (\pi(1), \dots, \pi(k)) \in \mathcal{G}^{k,k}.$$

Moreover, for a tuple $\boldsymbol{a} \in V(F)^k$ of vertices of a graph F, let $\pi(\boldsymbol{a}) := (a_{\pi(1)}, \dots, a_{\pi(k)})$. Then, for a bi-labeled graph $(F, \boldsymbol{a}, \boldsymbol{b}) \in \mathcal{M}^{k,\ell}$, we have $\boldsymbol{P}_\pi \circ (F, \boldsymbol{a}, \boldsymbol{b}) = (F, \pi^{-1}(\boldsymbol{a}), \boldsymbol{b})$ for every permutation $\pi \colon [k] \to [k]$ and $(F, \boldsymbol{a}, \boldsymbol{b}) \circ \boldsymbol{P}_\pi = (F, \boldsymbol{a}, \pi(\boldsymbol{b}))$ for every permutation $\pi \colon [\ell] \to [\ell]$.

3.2 Graphon Operators

Graphon operators generalize the homomorphism density $t(F, W)$ of a multigraph F in a graphon $W \colon X \times X \to [0,1]$ to bi-labeled graphs. To this end, let $\boldsymbol{F} = (F, \boldsymbol{a}, \boldsymbol{b}) \in \mathcal{M}^{k,\ell}$ be a bi-labeled graph. To simplify notation, let $t(\boldsymbol{F}, W) := t(F, W)$ denote the homomorphism density of the underlying graph of $\boldsymbol{F}$ in W, i.e., we ignore both the input and output labels. Now, let us first take the input labels of $\boldsymbol{F}$ into account, that is, we view $\boldsymbol{F}$ as a multi-rooted multigraph and the homomorphism density becomes a function by not fixing the vertices that have an input label. Formally, the *homomorphism function* of $\boldsymbol{F}$ in W is the function $f_{\boldsymbol{F} \to W} \colon X^k \to [0,1]$ defined by

$$f_{\boldsymbol{F} \to W}(x_{a_1}, \dots, x_{a_k}) := \int_{X^{V(F) \setminus \boldsymbol{a}}} \prod_{ij \in E(F)} W(x_i, x_j) \, d\mu^{\otimes V(F) \setminus \boldsymbol{a}}(\bar{x}) \tag{4}$$

for all $x_{a_1}, \dots, x_{a_k} \in X$. The Tonelli-Fubini theorem immediately yields that

$$\langle \mathbf{1}_{X^k}, f_{\boldsymbol{F}\to W} \rangle = t(\boldsymbol{F}, W).$$

Then, when taking both input and output labels of $\boldsymbol{F}$ into account, we obtain an operator $T_{\boldsymbol{F}\to W}$ instead of a function $f_{\boldsymbol{F}\to W}$ by, intuitively, "gluing" a given function f to the output vertices of $\boldsymbol{F}$ to obtain the function $T_{\boldsymbol{F}\to W} f$. The point of this definition is that an application of $T_{\boldsymbol{F}\to W}$ to a homomorphism function $f_{\boldsymbol{G}\to W}$ yields the homomorphism function $f_{\boldsymbol{F}\circ\boldsymbol{G}\to W}$. Formally, the *$\boldsymbol{F}$-operator of W* is the mapping $T_{\boldsymbol{F}\to W} \colon L^2(X^\ell, \mu^{\otimes \ell}) \to L^2(X^k, \mu^{\otimes k})$ defined by

$$(T_{\boldsymbol{F}\to W} f)(x_{a_1}, \dots, x_{a_k}) \coloneqq \int_{X^{V(F)\setminus \boldsymbol{a}}} \prod_{ij\in E(F)} W(x_i, x_j) \cdot f(x_{b_1}, \dots, x_{b_\ell})\, d\mu^{\otimes V(F)\setminus \boldsymbol{a}}(\bar{x}) \tag{5}$$

for every $f \in L^2(X^\ell, \mu^{\otimes \ell})$ and all $x_{a_1}, \dots, x_{a_k} \in X$. Note that $f_{\boldsymbol{F}\to W} = T_{\boldsymbol{F}\to W} \mathbf{1}_{X^\ell}$ as an element of $L^\infty(X^k, \mu^{\otimes k})$ and, in particular,

$$\langle \mathbf{1}_{X^k}, T_{\boldsymbol{F}\to W} \mathbf{1}_{X^\ell} \rangle = t(\boldsymbol{F}, W).$$

The Tonelli-Fubini theorem and the Cauchy-Schwarz inequality allow to verify that Equation (5) indeed yields a well-defined contraction. We stress that it is important that no vertex of $\boldsymbol{F}$ has multiple input or output vertices.

Lemma 14. *Let $\boldsymbol{F} \in \mathcal{M}^{k,\ell}$ be a bi-labeled graph and $W \colon X \times X \to [0,1]$ be a graphon. Then, $T_{\boldsymbol{F}\to W} \colon L^2(X^\ell, \mu^{\otimes \ell}) \to L^2(X^k, \mu^{\otimes k})$ is a well-defined L^2- and L^∞-contraction.*

Proof. Let $\boldsymbol{F} = (F, \boldsymbol{a}, \boldsymbol{b})$. First, note that $\bar{x} \mapsto W(x_i, x_j)$ for $ij \in E(F)$ is a function in $\mathcal{L}^\infty(X^{V(F)}, \mu^{\otimes V(F)})$: the measurability follows from the definition of the product σ-algebra and the measurability of W. Then, since W is bounded by 1 by definition, we get that it is a function in in $\mathcal{L}^\infty(X^{V(F)}, \mu^{\otimes V(F)})$. More precisely, its $\|\cdot\|_\infty$-norm is at most $\|W\|_\infty$ since $\boldsymbol{F}$ does not have loops, i.e., $i \neq j$. Now, consider an $f \in \mathcal{L}^2(X^\ell, \mu^{\otimes \ell})$. Then, $\bar{x} \mapsto f(x_{b_1}, \dots, x_{b_\ell})$ is a function in $\mathcal{L}^2(X^{V(F)}, \mu^{\otimes V(F)})$: Again, the measurability of these functions follows from the definition of the product σ-algebra. Then, by the Tonelli-Fubini theorem, we get that the $\|\cdot\|_2$-norm of this function function is just $\|f\|_2$, which means that it is in $\mathcal{L}^2(X^{V(F)}, \mu^{\otimes V(F)})$. Note that, at this point, it is important that the entries of $\boldsymbol{b}$ are pairwise distinct.

Define the function g on $X^{V(F)}$ by

$$g(\bar{x}) \coloneqq \prod_{ij\in E(F)} W(x_i, x_j) \cdot f(x_{b_1}, \dots, x_{b_\ell})$$

for every $\bar{x} \in X^{V(F)}$. By the previous considerations, $g \in \mathcal{L}^2(X^{V(F)}, \mu^{\otimes V(F)})$ with

$$\|g\|_2 \le \|W\|_\infty^{\mathsf{e}(F)} \cdot \|f\|_2 < \infty.$$

Then, the function being integrated in (5), which is obtained from g by fixing $x_{a_1}, \dots, x_{a_k} \in X$, is also measurable (see also [2, Theorem 18.1]). By the Tonelli-Fubini theorem, we have

$$\begin{aligned} \int_{X^{\boldsymbol{a}}} \int_{X^{V(F)\setminus \boldsymbol{a}}} g(\bar{x})^2 d\mu^{\otimes V(F)\setminus \boldsymbol{a}}(\bar{x})\, d\mu^{\otimes \boldsymbol{a}}(\bar{x}) &= \int_{X^{V(F)}} g(\bar{x})^2 d\mu^{\otimes V(F)}(\bar{x}) \\ &\le (\|W\|_\infty^{\mathsf{e}(F)} \cdot \|f\|_2)^2 < \infty, \end{aligned}$$

where

$$\int_{X^{V(F)\setminus \boldsymbol{a}}} g(\bar{x})^2 d\mu^{\otimes V(F)\setminus \boldsymbol{a}}(\bar{x})$$

is defined and finite for $\mu^{\otimes a}$-almost all $x_{a_1}, \dots, x_{a_k} \in X$. Hence, for $\mu^{\otimes \boldsymbol{a}}$-almost all $x_{a_1}, \dots, x_{a_k} \in X$, we obtain a function in $\mathcal{L}^2(X^{V(F)\setminus \boldsymbol{a}}, \mu^{\otimes V(F)\setminus \boldsymbol{a}})$, to which the Cauchy-Schwarz inequality is applicable, from g by fixing $x_{a_1}, \dots, x_{a_k}$.

Again by the Tonelli-Fubini theorem and since the entries of $\boldsymbol{a}$ are pairwise distinct, $T_{\boldsymbol{F}\to W} f$ is a measurable function defined almost everywhere, and we get

$$\begin{aligned}
\|T_{\boldsymbol{F}\to W} f\|_2^2 &= \int_{X^{\boldsymbol{a}}} \Big(\int_{X^{V(F)\setminus \boldsymbol{a}}} g(\bar{x}) d\mu^{\otimes V(F)\setminus \boldsymbol{a}}(\bar{x}) \Big)^2 d\mu^{\otimes \boldsymbol{a}}(\bar{x}) \\
&\le \int_{X^{\boldsymbol{a}}} \int_{X^{V(F)\setminus \boldsymbol{a}}} g(\bar{x})^2 d\mu^{\otimes V(F)\setminus \boldsymbol{a}}(\bar{x})\, d\mu^{\otimes \boldsymbol{a}}(\bar{x}) && \text{(Cauchy-Schwarz)} \\
&= \|g\|_2^2 \le (\|W\|_\infty^{\mathsf{e}(F)} \cdot \|f\|_2)^2 < \infty.
\end{aligned}$$

Hence, $T_{\boldsymbol{F}\to W} f$ is a function in $\mathcal{L}^2(X^k, \mu^{\otimes k})$. Now, for a function $f' \in \mathcal{L}^2(X^\ell, \mu^{\otimes \ell})$ such that f and f' are equal $\mu^{\otimes \ell}$-almost everywhere, define g' analogously to g. Then, g and g' are equal $\mu^{\otimes V(F)}$-almost everywhere and, with the previous considerations, another application of the Cauchy-Schwarz inequality and the Tonelli-Fubini theorem yields that

$$\|T_{\boldsymbol{F}\to W} f - T_{\boldsymbol{F}\to W} f'\|_2^2 \le \|g - g'\|_2^2 = 0.$$

Therefore, $T_{\boldsymbol{F}\to W}$ is a well-defined mapping $L^2(X^\ell, \mu^{\otimes \ell}) \to L^2(X^k, \mu^{\otimes k})$. Verifying the linearity of $T_{\boldsymbol{F}\to W}$ is straight-forward, and as seen before, we have

$$\|T_{\boldsymbol{F}\to W} f\|_2 \le \|W\|_\infty^{\mathsf{e}(F)} \cdot \|f\|_2,$$

i.e., $T_{\boldsymbol{F}\to W}$ is bounded since $\boldsymbol{F}$ and W are fixed.

Finally, note that if $f \in \mathcal{L}^\infty(X^\ell, \mu^{\otimes \ell})$, then $\|g\|_\infty \le \|W\|_\infty^{\mathsf{e}(F)} \cdot \|f\|_\infty$. From the previous considerations, we may even assume that g is bounded by $\|W\|_\infty^{\mathsf{e}(F)} \cdot \|f\|_\infty$. Then, the definition of $T_{\boldsymbol{F}\to W}$ immediately yields that $\|T_{\boldsymbol{F}\to W} f\|_\infty \le \|g\|_\infty \le \|W\|_\infty^{\mathsf{e}(F)} \cdot \|f\|_\infty$. □

Note that the definition of $T_{\boldsymbol{F}\to W}$ only depends on the isomorphism type of $\boldsymbol{F}$, i.e., isomorphic bi-labeled graphs $\boldsymbol{F}$ and $\boldsymbol{F'}$ define the same operator $T_{\boldsymbol{F}\to W} = T_{\boldsymbol{F'}\to W}$. Moreover, if $\boldsymbol{F}$ does not have any edges, then the definition of $T_{\boldsymbol{F}\to W}$ is independent of W and we just write $T_{\boldsymbol{F}}$. We just have to be a bit careful since $T_{\boldsymbol{F}}$ is still dependent on the standard Borel space $(X, \mathcal{B})$ and the Borel probability measure μ.

Example 15. *1. Define $\boldsymbol{A} := (([2], \{12\}), (1), (2)) \in \mathcal{G}^{1,1}$ to be the edge with one input and one output vertex. Let $W\colon X \times X \to [0,1]$ be a graphon. Then, $T_{\boldsymbol{A}\to W} = T_W$.*

2. Let $k \ge 1$ and $\pi\colon [k] \to [k]$ be a permutation. Then, $T_{\boldsymbol{P}_\pi}$ is equal to the Koopman operator T_π of the measure-preserving measurable map $X^k \to X^k$ induced by π.

The operator $T_{\boldsymbol{F}\to W}$ was defined such that the application to a homomorphism function $f_{\boldsymbol{G}\to W}$ yields the homomorphism function $f_{\boldsymbol{F}\circ\boldsymbol{G}\to W}$. The following lemma formalizes this by stating that the composition of bi-labeled graphs corresponds to the composition of graphon operators. Moreover, the analogous correspondence holds between the transpose and the Hilbert adjoint and between the Schur product and the point-wise product.

Lemma 16. *Let $W\colon X \times X \to [0,1]$ be a graphon. Then,*

1. $T_{\boldsymbol{F}^* \to W} = T^*_{\boldsymbol{F} \to W}$ *for every* $\boldsymbol{F} \in \mathcal{M}$,
2. *if* $\boldsymbol{F} \in \mathcal{M}$ *is symmetric, then* $T_{\boldsymbol{F} \to W}$ *is self-adjoint,*
3. $T_{\boldsymbol{F_1} \circ \boldsymbol{F_2} \to W} = T_{\boldsymbol{F_1} \to W} \circ T_{\boldsymbol{F_2} \to W}$ *for all* $\boldsymbol{F_1} \in \mathcal{M}^{k,m}$, $\boldsymbol{F_2} \in \mathcal{M}^{m,\ell}$, $k, \ell, m \geq 0$, *and*
4. $T_{\boldsymbol{F_1} \cdot \boldsymbol{F_2} \to W}(c_1 \cdot c_2) = (T_{\boldsymbol{F_1} \to W} c_1) \cdot (T_{\boldsymbol{F_2} \to W} c_2)$ *for all* $c_1, c_2 \in \mathbb{R}$, $\boldsymbol{F_1}, \boldsymbol{F_2} \in \mathcal{M}^{k,0}$, $k \geq 0$.

Proof. 1: We have

$$
\begin{aligned}
&\langle T_{\boldsymbol{F} \to W} f, g \rangle \\
&= \int_{X^{\boldsymbol{a}}} \Big(\int_{X^{V(F) \setminus \boldsymbol{a}}} \prod_{ij \in E(F)} W(x_i, x_j) \cdot f(x_{b_1}, \dots, x_{b_\ell}) \, d\mu^{\otimes V(F) \setminus \boldsymbol{a}}(\bar{x}) \Big) \cdot g(x_{a_1}, \dots, x_{a_k}) \, d\mu^{\otimes \boldsymbol{a}}(\bar{x}) \\
&= \int_{X^{V(F)}} \prod_{ij \in E(F)} W(x_i, x_j) \cdot f(x_{b_1}, \dots, x_{b_\ell}) \cdot g(x_{a_1}, \dots, x_{a_k}) \, d\mu^{\otimes V(F)}(\bar{x}) \\
&= \int_{X^{\boldsymbol{b}}} f(x_{b_1}, \dots, x_{b_\ell}) \cdot \Big(\int_{X^{V(F) \setminus \boldsymbol{b}}} \prod_{ij \in E(F)} W(x_i, x_j) \cdot g(x_{a_1}, \dots, x_{a_k}) \, d\mu^{\otimes V(F) \setminus \boldsymbol{b}}(\bar{x}) \Big) \, d\mu^{\otimes \boldsymbol{b}}(\bar{x}) \\
&= \langle f, T_{\boldsymbol{F}^* \to W} g \rangle
\end{aligned}
$$

for all $f \in L^2(X^\ell, \mu^{\otimes \ell}), g \in L^2(X^k, \mu^{\otimes k})$ by the Tonelli-Fubini theorem, which is applicable since the product being integrated is a function in $L^1(X^{V(F)}, \mu^{\otimes V(F)})$ by the Cauchy-Schwarz inequality.

2: By 1, we have $T^*_{\boldsymbol{F} \to W} = T_{\boldsymbol{F}^* \to W} = T_{\boldsymbol{F} \to W}$.

3: Let $\boldsymbol{F_1} = (F_1, \boldsymbol{a_1}, \boldsymbol{b_1})$, $\boldsymbol{F_2} = (F_2, \boldsymbol{a_2}, \boldsymbol{b_2})$, and $\boldsymbol{F_1} \circ \boldsymbol{F_2} = (F, \boldsymbol{a_1}, \boldsymbol{b_2})$. In the following, we identify vertices $b_{1,1}, \dots, b_{1,m}$ with $a_{2,1}, \dots, a_{2,m}$. Note that the sets $V(F_1) \setminus \boldsymbol{a_1}$ and $V(F_2) \setminus \boldsymbol{b_1} = V(F_2) \setminus \boldsymbol{a_2}$ form a partition of $V(F_1 \circ F_2) \setminus \boldsymbol{a_1}$. Then, we have

$$
\begin{aligned}
&(T_{\boldsymbol{F_1} \to W}(T_{\boldsymbol{F_2} \to W} f))(x_{a_{1,1}}, \dots, x_{a_{1,k}}) \\
&= \int_{X^{V(F_1) \setminus \boldsymbol{a_1}}} \prod_{ij \in E(F_1)} W(x_i, x_j) \cdot (T_{\boldsymbol{F_2} \to W} f)(x_{b_{1,1}}, \dots, x_{b_{1,m}}) \, d\mu^{\otimes V(F_1) \setminus \boldsymbol{a_1}}(\bar{x}) \\
&= \int_{X^{V(F_1) \setminus \boldsymbol{a_1}}} \Big(\int_{X^{V(F_2) \setminus \boldsymbol{a_2}}} \prod_{ij \in E(F)} W(x_i, x_j) \cdot f(x_{b_{2,1}}, \dots, x_{b_{2,\ell}}) \, d\mu^{\otimes V(F_2) \setminus \boldsymbol{a_2}}(\bar{x}) \Big) \, d\mu^{\otimes V(F_1) \setminus \boldsymbol{a_1}}(\bar{x}) \\
&= \int_{X^{V(F) \setminus \boldsymbol{a_1}}} \prod_{ij \in E(F)} W(x_i, x_j) \cdot f(x_{b_{2,1}}, \dots, x_{b_{2,\ell}}) \, d\mu^{\otimes V(F) \setminus \boldsymbol{a_1}}(\bar{x})
\end{aligned}
$$

for every $f \in L^2(X^\ell, \mu^{\otimes \ell})$ and $\mu^{\otimes \boldsymbol{a_1}}$-almost all $x_{a_{1,1}}, \dots, x_{a_{1,k}} \in X$ by the Tonelli-Fubini theorem.

4: We have

$$
\begin{aligned}
&(T_{\boldsymbol{F_1}\to W} c_1) \cdot (T_{\boldsymbol{F_2}\to W} c_2)(x_{a_{1,1}}, \dots, x_{a_{1,k}}) \\
&= \Big(\int_{X^{V(F_1)\setminus \boldsymbol{a_1}}} \prod_{ij \in E(F_1)} W(x_i, x_j) \cdot c_1 \, d\mu^{\otimes V(F_1)\setminus \boldsymbol{a_1}}(\bar{x}) \Big) \cdot \Big(\int_{X^{V(F_2)\setminus \boldsymbol{a_2}}} \prod_{ij \in E(F_2)} W(x_i, x_j) \cdot c_2 \, d\mu^{\otimes V(F_2)\setminus \boldsymbol{a_2}}(\bar{x}) \Big) \\
&= \int_{X^{V(F_1)\setminus \boldsymbol{a_1}}} \int_{X^{V(F_2)\setminus \boldsymbol{a_2}}} \prod_{ij \in E(F_1)} W(x_i, x_j) \cdot \prod_{ij \in E(F_2)} W(x_i, x_j) \cdot c_1 \cdot c_2 \, d\mu^{\otimes V(F_2)\setminus \boldsymbol{a_2}}(\bar{x}) \, d\mu^{\otimes V(F_1)\setminus \boldsymbol{a_1}}(\bar{x}) \\
&= \int_{X^{V(F)\setminus \boldsymbol{a_1}}} \prod_{ij \in E(F)} W(x_i, x_j) \cdot c_1 \cdot c_2 \, d\mu^{\otimes V(F)\setminus \boldsymbol{a_1}}(\bar{x}) \\
&= T_{\boldsymbol{F_1}\cdot \boldsymbol{F_2}\to W}(c_1 \cdot c_2)
\end{aligned}
$$

for all $c_1, c_2 \in \mathbb{R}$ and $\mu^{\otimes \boldsymbol{a_1}}$-almost all $x_{a_{1,1}}, \dots, x_{a_{1,k}} \in X$ by the Tonelli-Fubini theorem. □

For a set $\mathcal{F} \subseteq \mathcal{M}^{k,k}$, every graphon $W \colon X \times X \to [0,1]$ induces a family of L^∞-contractions $\mathbb{T}_{\mathcal{F}\to W} := (T_{\boldsymbol{F}\to W})_{\boldsymbol{F}\in\mathcal{F}}$ on $L^2(X^k, \mu^{\otimes k})$, cf. Lemma 14. When handling such families of operators, we often use notation like $\mathbb{T}_{\mathcal{F}\to W} \circ T$ for an L^∞-contraction T or $\mathbb{T}_{\mathcal{F}\to W}/\mathcal{C}$ for a $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ to denote the family obtained by applying the operation to every operator in the family; for these examples, we obtain the families $(T_{\boldsymbol{F}\to W} \circ T)_{\boldsymbol{F}\in\mathcal{F}}$ and $(T_{\boldsymbol{F}\to W}/\mathcal{C})_{\boldsymbol{F}\in\mathcal{F}}$. Moreover, if the graphs in $\mathcal{F}$ do not have any edges, we again abbreviate $\mathbb{T}_{\mathcal{F}} := (T_{\boldsymbol{F}})_{\boldsymbol{F}\in\mathcal{F}}$. Recall that $\mathcal{F}^k$ is the set of all neighbor and adjacency graphs with k input and output labels. Let us finally define the family

$$\mathbb{T}_W^k := \mathbb{T}_{\mathcal{F}^k \to W},$$

that replaces the single operator T_W in Theorem 4, our characterization of oblivious k-WL.

Let us explore the connection between the family $\mathbb{T}_W^k$ and treewidth $k-1$ homomorphism functions: Recall that the terms in $\mathcal{F}^k$ correspond to the tree-decomposed multigraphs of treewidth at most $k-1$ by Lemma 13. Given such a term $\mathbb{F} \in \mathcal{F}^k$, we can use the correspondence of bi-labeled graph operations to their operator counterparts, cf. Lemma 16, to inductively compute the homomorphism function $f_{[\![\mathbb{F}]\!]\to W}$ of $[\![\mathbb{F}]\!]$ in a graphon W using the operators $\mathbb{T}_W^k$. Hence, the operators in $\mathbb{T}_W^k$ yield all homomorphism functions of multigraphs of treewidth at most $k-1$ in W. An important part of the proof of Theorem 4 consists of defining different families of L^∞-contractions indexed by $\mathcal{F}^k$ that we may use instead of $\mathbb{T}_W^k$ and still yield the same homomorphism functions. For example, we may replace $\mathbb{T}_W^k$ by the quotient operators $\mathbb{T}_W^k/\mathcal{C}$ for an appropriate $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$. This leads to the following definition.

Definition 17. *Let $k \geq 1$ and $\mathbb{T} = (T_F)_{F\in\mathcal{F}}$ be a family of L^∞-contractions indexed by a set $\mathcal{F} \subseteq \mathcal{M}^{k,k}$. For every term $\mathbb{F} \in \langle \mathcal{F} \rangle_{\circ,\cdot}$, the homomorphism function of $\mathbb{F}$ in $\mathbb{T}$ is the function $f_{\mathbb{F}\to\mathbb{T}} \in L^\infty(X,\mu)$ with $\|f_{\mathbb{F}\to\mathbb{T}}\|_\infty \leq 1$ defined inductively by*

1. *$f_{\mathbb{F}\to\mathbb{T}} := \boldsymbol{1}_X$ for $\mathbb{F} = \boldsymbol{1}^k$,*
2. *$f_{\mathbb{F}\to\mathbb{T}} := T_{\boldsymbol{F}} f_{\mathbb{F}'\to\mathbb{T}}$ for $\mathbb{F} = \boldsymbol{F} \circ \mathbb{F}'$, where $\boldsymbol{F} \in \mathcal{F}$, and*
3. *$f_{\mathbb{F}\to\mathbb{T}} := f_{\mathbb{F}_1\to\mathbb{T}} \cdot f_{\mathbb{F}_2\to\mathbb{T}}$ for $\mathbb{F} = \mathbb{F}_1 \cdot \mathbb{F}_2$.*

Moreover, the homomorphism density of $\mathbb{F}$ in $\mathbb{T}$ is defined as $t(\mathbb{F}, \mathbb{T}) := \langle \boldsymbol{1}_X, f_{\mathbb{F}\to\mathbb{T}} \rangle$.

As remarked above, given a term $\mathbb{F} \in \mathcal{F}^k$, we can use the correspondence of bi-labeled graph operations to their operator counterparts to inductively compute the homomorphism function $f_{[\![\mathbb{F}]\!]\to W}$ and, in particular, the homomorphism density $t([\![\mathbb{F}]\!], W)$ of $[\![\mathbb{F}]\!]$ in a graphon W using the operators in $\mathbb{T}_W^k$.

Lemma 18. *Let $k \geq 1$. Let $W\colon X \times X \to [0,1]$ be a graphon. Then,*

$$f_{\mathbb{F}\to\mathbb{T}^k_W} = f_{[\![\mathbb{F}]\!]\to W} \qquad \text{and} \qquad t(\mathbb{F}, \mathbb{T}^k_W) = t([\![\mathbb{F}]\!], W)$$

for every $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$.

Proof. We show that $f_{\mathbb{F}\to\mathbb{T}^k_W} = f_{[\![\mathbb{F}]\!]\to W}$ by induction on $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$. Then,

$$t(\mathbb{F}, \mathbb{T}^k_W) = \langle \mathbf{1}_X, f_{\mathbb{F}\to\mathbb{T}^k_W} \rangle = \langle \mathbf{1}_X, f_{[\![\mathbb{F}]\!]\to W} \rangle = t([\![\mathbb{F}]\!], W)$$

by definition of $t(\mathbb{F}, \mathbb{T}^k_W)$ and $f_{[\![\mathbb{F}]\!]\to W}$.

For the induction basis $\mathbb{F} = \mathbf{1}^k$, we have $f_{\mathbb{F}\to\mathbb{T}^k_W} = \mathbf{1}_{X^k}$ by Definition 17 and $f_{\mathbb{F}\to W} = T_{\mathbf{1}^k\to W}\mathbf{1}_{X^k} = \mathbf{1}_{X^k}$ by the definition of $T_{\mathbf{1}^k\to W}$. For the first case of the inductive step $\mathbb{F} = \boldsymbol{F} \circ \mathbb{F}'$, where $\boldsymbol{F} \in \mathcal{F}^k$ and $[\![\mathbb{F}']\!] \in \mathcal{M}^{k,0}$, we have

$$\begin{aligned} f_{\mathbb{F}\to\mathbb{T}^k_W} = T_{\boldsymbol{F}\to W} f_{\mathbb{F}'\to\mathbb{T}^k_W} = T_{\boldsymbol{F}\to W} f_{[\![\mathbb{F}']\!]\to W} &= T_{\boldsymbol{F}\to W}(T_{[\![\mathbb{F}']\!]\to W}\mathbf{1}_{X^\ell}) \\ &= T_{\boldsymbol{F}\circ[\![\mathbb{F}']\!]\to W}\mathbf{1}_{X^\ell} \\ &= T_{[\![\mathbb{F}]\!]\to W}\mathbf{1}_{X^\ell} \\ &= f_{[\![\mathbb{F}]\!]\to W} \end{aligned}$$

by Definition 17, the induction hypothesis, the definition of $T_{[\![\mathbb{F}]\!]\to W}$, and Lemma 16 3. For the second case of the inductive step $\mathbb{F} = \mathbb{F}_1 \cdot \mathbb{F}_2$, where $[\![\mathbb{F}]\!], [\![\mathbb{F}_1]\!], [\![\mathbb{F}_2]\!] \in \mathcal{M}^{k,0}$. Then, we have

$$\begin{aligned} f_{\mathbb{F}\to\mathbb{T}^k_W} = f_{\mathbb{F}_1\to\mathbb{T}^k_W} \cdot f_{\mathbb{F}_2\to\mathbb{T}^k_W} = f_{[\![\mathbb{F}_1]\!]\to W} \cdot f_{[\![\mathbb{F}_2]\!]\to W} &= (T_{[\![\mathbb{F}_1]\!]\to W}\mathbf{1}_{X^0}) \cdot (T_{[\![\mathbb{F}_2]\!]\to W}\mathbf{1}_{X^0}) \\ &= T_{[\![\mathbb{F}_1]\!]\cdot[\![\mathbb{F}_2]\!]\to W}\mathbf{1}_{X^0} \\ &= T_{[\![\mathbb{F}]\!]\to W}\mathbf{1}_{X^0} \\ &= f_{[\![\mathbb{F}]\!]\to W} \end{aligned}$$

by Definition 17, the induction hypothesis, the definition of $T_{[\![\mathbb{F}]\!]\to W}$, and Lemma 16 4. ☐

As remarked above, an essential ingredient of the proof of Theorem 4 is the definition of families of L^∞-contractions that replace $\mathbb{T}^k_W$ but still yield the same homomorphism functions. The following lemma gives a sufficient condition under which this is possible. Recall that a Markov embedding is a Markov operator that is an isometry. Unlike Markov operators in general, Markov embeddings are compatible with point-wise products of functions, cf. [8, Theorem 13.9, Remark 13.10]. This is crucial since we need the point-wise product of functions to get from bounded pathwidth to bounded treewidth homomorphism functions.

Lemma 19. *Let $k \geq 1$. Let $(X_1, \mathcal{B}_1)$ and $(X_2, \mathcal{B}_2)$ be standard Borel spaces with Borel probability measures μ_1 and μ_2 on X_1 and X_2, respectively. Let $\mathbb{T}_1$ and $\mathbb{T}_2$ be families of L^∞-contractions on $L^2(X_1, \mu_1)$ and $L^2(X_2, \mu_2)$, respectively, indexed by $\mathcal{F}^k$. If $I\colon L^2(X_2,\mu_2) \to L^2(X_1,\mu_1)$ is a Markov embedding such that $\mathbb{T}_1 \circ I = I \circ \mathbb{T}_2$, then*

$$I f_{\mathbb{F}\to\mathbb{T}_2} = f_{\mathbb{F}\to\mathbb{T}_1} \qquad \text{and} \qquad t(\mathbb{F}, \mathbb{T}_1) = t(\mathbb{F}, \mathbb{T}_2)$$

for every $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$.

Proof. We show that $I f_{\mathbb{F}\to\mathbb{T}_1} = f_{\mathbb{F}\to\mathbb{T}_2}$ by induction on $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$. Then, also

$$t(\mathbb{F}, \mathbb{T}_1) = \langle \mathbf{1}_{X_1}, f_{\mathbb{F}\to\mathbb{T}_1} \rangle = \langle \mathbf{1}_{X_1}, I f_{\mathbb{F}\to\mathbb{T}_2} \rangle = \langle I^*\mathbf{1}_{X_2}, f_{\mathbb{F}\to\mathbb{T}_2} \rangle = \langle \mathbf{1}_{X_2}, f_{\mathbb{F}\to\mathbb{T}_2} \rangle = t(\mathbb{F}, \mathbb{T}_2).$$

For the induction basis $\mathbb{F} = \mathbf{1}^k$, we have

$$If_{\mathbb{F}\to\mathbb{T}_2} = I\mathbf{1}_{X_2} = \mathbf{1}_{X_1} = f_{\mathbb{F}\to\mathbb{T}_1}.$$

For $\mathbb{F} = \boldsymbol{F} \circ \mathbb{F}'$, where $\boldsymbol{F} \in \mathcal{F}^k$, we have

$$If_{\mathbb{F}\to\mathbb{T}_2} = (I \circ (\mathbb{T}_2)_{\boldsymbol{F}})f_{\mathbb{F}'\to\mathbb{T}_2} = ((\mathbb{T}_1)_{\boldsymbol{F}} \circ I)f_{\mathbb{F}'\to\mathbb{T}_2} = (\mathbb{T}_1)_{\boldsymbol{F}} f_{\mathbb{F}'\to\mathbb{T}_1} = f_{\mathbb{F}\to\mathbb{T}_1}$$

by the assumption and the induction hypothesis. Finally, for $\mathbb{F} = \mathbb{F}_1 \cdot \mathbb{F}_2$, we use that I is a Markov embedding and, hence, satisfies $I(f \cdot g) = If \cdot Ig$ for all $f, g \in L^\infty(X_2, \mu_2)$ [8, Theorem 13.9]. We have

$$If_{\mathbb{F}\to\mathbb{T}_2} = I(f_{\mathbb{F}_1\to\mathbb{T}_2} \cdot f_{\mathbb{F}_2\to\mathbb{T}_2}) = If_{\mathbb{F}_1\to\mathbb{T}_2} \cdot If_{\mathbb{F}_2\to\mathbb{T}_2} = f_{\mathbb{F}_1\to\mathbb{T}_1} \cdot f_{\mathbb{F}_2\to\mathbb{T}_1} = f_{\mathbb{F}\to\mathbb{T}_1}$$

by the induction hypothesis. ☐

An important application of Lemma 19 is to replace the family $\mathbb{T}_W^k$ by the quotient operators $\mathbb{T}_W^k/\mathcal{C}$ for an appropriate $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$. To this end, we call a $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ *W-invariant* if $\mathcal{C}$ is invariant for every operator in the family $\mathbb{T}_W^k$, i.e., $\mathcal{C}$ is $T_{\boldsymbol{F}\to W}$-invariant for every $\boldsymbol{F} \in \mathcal{F}^k$.

Corollary 20. *Let $k \geq 1$. Let $W\colon X \times X \to [0,1]$ be a graphon and $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ be W-invariant. Then,*

$$t(\mathbb{F}, (\mathbb{T}_W^k)_{\mathcal{C}}) = t(\mathbb{F}, \mathbb{T}_W^k/\mathcal{C}) = t(\mathbb{F}, \mathbb{T}_W^k) = t([\![\mathbb{F}]\!], W)$$

for every $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$.

Proof. The last equation is just Lemma 18. By Lemma 8 4 and 5, we have $I_{\mathcal{C}} \circ \mathbb{T}_W^k/\mathcal{C} = (\mathbb{T}_W^k)_{\mathcal{C}} \circ I_{\mathcal{C}}$ and $I_{\mathcal{C}} \circ \mathbb{T}_W^k/\mathcal{C} = \mathbb{T}_W^k \circ I_{\mathcal{C}}$, where $I_{\mathcal{C}}$ is a Markov embedding by Claim 6 5, Therefore, Lemma 19 yields the first two equations. ☐

4 Weisfeiler-Leman and Graphons

In Section 4.1 to Section 4.5 we closely follow Grebík and Rocha [9] to prove Theorem 4 and formally define all notions appearing in it. Many, but not all, of their proofs transfer without too many changes. In Section 4.1, we start off by showing that the minimum W-invariant $\mu^{\otimes k}$-relatively complete sub-σ-algebra $\mathcal{C}_W^k$ of $\mathcal{B}^{\otimes k}$ for a graphon W can be obtained by iterative applications of the operators $\mathbb{T}_W^k$. Section 4.2 defines define the space $\mathbb{M}^k$, i.e., the space of all colors used by oblivious k-WL, and *k-WL distributions*, which generalize multisets of colors. In Section 4.3, we define the function $\mathsf{owl}_W^k\colon X^k \to \mathbb{M}^k$ and the k-WL distribution ν_W^k for a graphon W. In Section 4.4, we deviate from Grebík and Rocha [9]: They show that every distribution on iterative degree measures ν defines a graphon on the space $\mathbb{M}$; this graphon for ν_W is then isomorphic to the quotient graphon $W/\mathcal{C}_W$. Since the operators in $\mathbb{T}_W^k$ are not integral operators, we take the different route of showing that a k-WL distribution ν defines a family of operators $\mathbb{T}_\nu$ on $L^2(\mathbb{M}^k, \nu)$; the family $\mathbb{T}_{\nu_W^k}$ then corresponds to $\mathbb{T}_W^k$. These operators are essential in the proof of Theorem 4 in Section 4.5.

Section 4.6 shows that one can combine all k-WL distributions $\nu_W^1, \nu_W^2, \ldots$ of a graphon W into a single distribution to obtain a new characterization of weak isomorphism. Section 4.7 further explains how the characterization of Theorem 4 using Markov operators corresponds to the system $\mathsf{L}_{\text{iso}}^k$ of linear equations.

4.1 The Minimum W-Invariant Sub-σ-Algebra

For a family $\mathbb{T} = (T_i)_{i \in I}$ of operators $T_i \colon L^2(X, \mu) \to L^2(X, \mu)$, where $i \in I$, and a $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$, define

$$\mathbb{T}(\mathcal{C}) \coloneqq \bigcap \left\{ \mathcal{D} \in \Theta(\mathcal{B}, \mu) \mid \mathcal{D} \supseteq \mathcal{C} \text{ and } T_i(L^2(X, \mathcal{C}, \mu)) \subseteq L^2(X, \mathcal{D}, \mu) \text{ for every } i \in I \right\}.$$

Then, $\mathbb{T}(\mathcal{C}) \in \Theta(\mathcal{B}, \mu)$, cf. Section 2.3, and $\mathcal{C}$ is called $\mathbb{T}$-*invariant* if $\mathbb{T}(\mathcal{C}) \subseteq \mathcal{C}$, which is equivalent to requiring that $\mathcal{C}$ is T_i-invariant for every $i \in I$. Note that this operation is monotonous, i.e., for all $\mathcal{C}, \mathcal{D} \in \Theta(\mathcal{B}, \mu)$ with $\mathcal{C} \subseteq \mathcal{D}$, we have $\mathbb{T}(\mathcal{C}) \subseteq \mathbb{T}(\mathcal{D})$. By definition, the family $\mathbb{T}_W^k$ consists of the two families $\mathbb{T}_{\mathcal{A}^k \to W}$ and $\mathbb{T}_{\mathcal{N}^k}$. The following definition uses these two individual families to define the sub-σ-algebra $\mathcal{C}_W^k$ of $\mathcal{B}^{\otimes k}$. Already at this point, one should notice the connection to oblivious k-WL, cf. Section 1.2: the operators in $\mathbb{T}_{\mathcal{A}^k \to W}$ capture the concept of atomic types while the operators in $\mathbb{T}_{\mathcal{N}^k}$ correspond to the refinement rounds via j-neighbors used in oblivious k-WL.

Definition 21. *Let $k \geq 1$ and $W \colon X \times X \to [0,1]$ be a graphon. Define $\mathcal{C}_{W,n}^k \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ for every $n \in \mathbb{N}$ by setting $\mathcal{C}_{W,0}^k \coloneqq \mathbb{T}_{\mathcal{A}^k \to W}(\langle \{\varnothing, X^k\} \rangle)$, $\mathcal{C}_{W,n+1}^k \coloneqq \mathbb{T}_{\mathcal{N}^k}(\mathcal{C}_{W,n}^k)$ for every $n \in \mathbb{N}$, and $\mathcal{C}_W^k \coloneqq \mathcal{C}_{W,\infty}^k \coloneqq \langle \bigcup_{n \in \mathbb{N}} \mathcal{C}_{W,n}^k \rangle$.*

Verifying that $\mathcal{C}_W^k$ is in fact the minimum W-invariant $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$ is mostly analogous to [9, Proposition 5.13]. A difference is given by the operators in $\mathbb{T}_{\mathcal{A}^k \to W}$, which are multiplicative, which implies that a single initial application guarantees $\mathbb{T}_{\mathcal{A}^k \to W}$-invariance for all subsequent sub-σ-algebras in the sequence. Moreover, we also verify that $\mathcal{C}_W^k$ is permutation invariant, i.e., $\mathcal{C}_W^k$ is T_π-invariant for every permutation $\pi \colon [k] \to [k]$.

Lemma 22. *Let $k \geq 1$ and $W \colon X \times X \to [0,1]$ be a graphon. Then,*

1. $\mathcal{C}_{W,0}^k = \left\langle \bigcup_{\boldsymbol{A} \in \mathcal{A}^k} \left\{ (T_{\boldsymbol{A} \to W} \mathbf{1}_{X^k})^{-1}(A) \mid A \in \mathcal{B}([0,1]) \right\} \right\rangle$,
2. $\mathcal{C}_{W,0}^k$ *is the minimum $\mathbb{T}_{\mathcal{A}^k \to W}$-invariant $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$,*
3. $\mathcal{C}_{W,n+1}^k = \left\langle \mathcal{C}_{W,n}^k \cup \bigcup_{\boldsymbol{N} \in \mathcal{N}^k} \left\{ (T_{\boldsymbol{N}} \mathbf{1}_A)^{-1}(B) \mid A \in \mathcal{C}_{W,n}^k,\, B \in \mathcal{B}([0,1]) \right\} \right\rangle$ *for every $n \in \mathbb{N}$,*
4. $\mathcal{C}_{W,n}^k$ *is $\mathbb{T}_{\mathcal{A}^k \to W}$-invariant for every $n \in \mathbb{N} \cup \{\infty\}$,*
5. $\mathcal{C}_W^k$ *is the minimum W-invariant $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$, and*
6. $\mathcal{C}_{W,n}^k$ *is permutation invariant for every $n \in \mathbb{N} \cup \{\infty\}$.*

Proof. 1 and 2: Let $\mathcal{C}$ denote the minimum $\mathbb{T}_{\mathcal{A}^k \to W}$-invariant $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$ and $\mathcal{D}$ denote the $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$ from 1. We prove that $\mathcal{C} = \mathcal{D} = \mathcal{C}_{W,0}^k$. We start by proving $\mathcal{C} \subseteq \mathcal{D}$. Let $\boldsymbol{A} \in \mathcal{A}^k$. The function $T_{\boldsymbol{A} \to W} \mathbf{1}_{X^k}$ is $\mathcal{D}$-measurable by definition of $\mathcal{D}$. Hence, for a $\mathcal{D}$-measurable function $g \in L^2(X^k, \mu^{\otimes k})$, their product $(T_{\boldsymbol{A} \to W} \mathbf{1}_{X^k}) \cdot g = T_{\boldsymbol{A} \to W} g$ is again $\mathcal{D}$-measurable, where the equality holds since $T_{\boldsymbol{A} \to W}$ is a multiplicative operator. That is, $\mathcal{D}$ is $\mathbb{T}_{\mathcal{A}^k \to W}$-invariant, which yields $\mathcal{C} \subseteq \mathcal{D}$. For the inclusion $\mathcal{D} \subseteq \mathcal{C}$ on the other hand, $\mathbf{1}_{X^k}$ is trivially $\mathcal{C}$-measurable and, since $\mathcal{C}$ is $\mathbb{T}_{\mathcal{A}^k \to W}$-invariant, the function $T_{\boldsymbol{A} \to W} \mathbf{1}_{X^k}$ is $\mathcal{C}$-measurable for every $\boldsymbol{A} \in \mathcal{A}^k$. Hence, $\mathcal{D} \subseteq \mathcal{C}$. We have established $\mathcal{C} = \mathcal{D}$ and it remains to prove that these are also equal to $\mathcal{C}_{W,0}^k$. We have $\langle \{\varnothing, X^k\} \rangle \subseteq \mathcal{C}$ and, hence, $\mathcal{C}_{W,0}^k = \mathbb{T}_{\mathcal{A}^k \to W}(\langle \{\varnothing, X^k\} \rangle) \subseteq \mathbb{T}_{\mathcal{A}^k \to W}(\mathcal{C}) \subseteq \mathcal{C}$. On the other hand, for every $\boldsymbol{A} \in \mathcal{A}^k$, the function $T_{\boldsymbol{A} \to W} \mathbf{1}_{X^k}$ is $\mathcal{C}_{W,0}^k$-measurable. Hence, $\mathcal{D} \subseteq \mathcal{C}_{W,0}^k$.

3: Let $\mathcal{D}$ denote the $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$ from 3, i.e., $\mathcal{D}$ is the minimum $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$ that contains $\mathcal{C}^k_{W,n}$ and makes the maps $T_{\boldsymbol{N}}\mathbf{1}_A$ for $\boldsymbol{N} \in \mathcal{N}^k$ and $A \in \mathcal{C}^k_{W,n}$ measurable. It is easy to see that $\mathcal{D} \subseteq \mathcal{C}^k_{W,n+1}$: We have $\mathcal{C}^k_{W,n} \subseteq \mathcal{C}^k_{W,n+1}$ by definition of $\mathcal{C}^k_{W,n+1}$. Moreover, for $\boldsymbol{N} \in \mathcal{N}^k$ and $A \in \mathcal{B}(\mathcal{C}^k_{W,n})$, the function $\mathbf{1}_A$ is $\mathcal{C}^k_{W,n}$-measurable and, hence by definition of $\mathcal{C}^k_{W,n+1}$, the function $T_{\boldsymbol{N}}\mathbf{1}_A$ is then $\mathcal{C}^k_{W,n+1}$ measurable. It remains to prove that $\mathcal{C}^k_{W,n+1} \subseteq \mathcal{D}$, i.e., that $\mathcal{C}^k_{W,n} \subseteq \mathcal{D}$ and $T_{\boldsymbol{N}}(L^2(X^k, \mathcal{C}^k_{W,n}, \mu^{\otimes k})) \subseteq L^2(X^k, \mathcal{D}, \mu^{\otimes k})$ for every $\boldsymbol{N} \in \mathcal{N}^k$. We have $\mathcal{C}^k_{W,n} \subseteq \mathcal{D}$ by definition of $\mathcal{D}$. Let $\boldsymbol{N} \in \mathcal{N}^k$. We have $T_{\boldsymbol{N}}\mathbf{1}_A \in L^2(X^k, \mathcal{D}, \mu^{\otimes k})$ for $A \in \mathcal{C}^k_{W,n}$ by definition of $\mathcal{D}$. Since the linear hull of $\{\mathbf{1}_A\}_{A \in \mathcal{C}^k_{W,n}}$ is dense in subspace $L^2(X^k, \mathcal{C}^k_{W,n}, \mu^{\otimes k})$ and since $L^2(X^k, \mathcal{D}, \mu^{\otimes k})$ is closed, linearity and continuity of $T_{\boldsymbol{N}}$ then yields that $T_{\boldsymbol{N}}(L^2(X^k, \mathcal{C}^k_{W,n}, \mu^{\otimes k})) \subseteq L^2(X^k, \mathcal{D}, \mu^{\otimes k})$.

4: Let $n \in \mathbb{N} \cup \{\infty\}$ and $\boldsymbol{A} \in \mathcal{A}^k$. We have $\mathcal{C}^k_{W,0} \subseteq \mathcal{C}^k_{W,n}$, which means that the function $T_{\boldsymbol{A}\to W}\mathbf{1}_{X^k}$ is $\mathcal{C}^k_{W,n}$-measurable. Then, the claim follows as $T_{\boldsymbol{A}\to W}$ is multiplicative, cf. the proof of 1 and 2.

5: We first show that $\mathcal{C}^k_W \subseteq \mathcal{C}$ for every $\mathbb{T}^k_W$-invariant $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$. We have $\langle\{\varnothing, X^k\}\rangle \subseteq \mathcal{C}$ and, hence, $\mathcal{C}^k_{W,0} = \mathbb{T}_{\mathcal{A}^k\to W}(\langle\{\varnothing, X^k\}\rangle) \subseteq \mathbb{T}_{\mathcal{A}^k\to W}(\mathcal{C}) \subseteq \mathcal{C}$. From there on, induction yields $\mathcal{C}^k_{W,n+1} = \mathbb{T}_{\mathcal{N}^k}(\mathcal{C}^k_{W,n}) \subseteq \mathbb{T}_{\mathcal{N}^k}(\mathcal{C}) \subseteq \mathcal{C}$ for every $n \in \mathbb{N}$. Hence, $\mathcal{C}^k_W \subseteq \mathcal{C}$.

It remains to prove that $\mathcal{C}^k_W$ is $\mathbb{T}^k_W$-invariant. By 4, it suffices to show that that $\mathcal{C}^k_W$ is $T_{\boldsymbol{N}}$-invariant for $\boldsymbol{N} \in \mathcal{N}^k$. This is essentially Proposition 5.13 of [9]: We first show that $T_{\boldsymbol{N}}\mathbf{1}_A \in L^2(X^k, \mathcal{C}^k_W, \mu^{\otimes k})$ for $A \in \mathcal{C}^k_W$. To this end, note that $\bigcup_{n\in\mathbb{N}} \mathcal{C}^k_{W,n}$ is an algebra and the σ-algebra generated by it is $\mathcal{C}^k_W$. Hence, from [6, Theorem 3.1.10], it easily follows that we can approximate every set in $\mathcal{C}^k_W$ by a set in $\bigcup_{n\in\mathbb{N}} \mathcal{C}^k_{W,n}$ w.r.t. the measure of their symmetric difference. This implies that, for every $A \in \mathcal{C}^k_W$, there is a sequence $(A_n)_{n\in\mathbb{N}}$ with $A_n \in \mathcal{C}^k_{W,n}$ such that $\mathbf{1}_{A_n} \to \mathbf{1}_A$ in $L^2(X^k, \mu^{\otimes k})$. Let $\boldsymbol{N} \in \mathcal{N}^k$. By continuity of $T_{\boldsymbol{N}}$, we have $T_{\boldsymbol{N}}\mathbf{1}_{A_n} \to T_{\boldsymbol{N}}\mathbf{1}_A$. Note that, for $n \in \mathbb{N}$, we have $T_{\boldsymbol{N}}\mathbf{1}_{A_n} \in L^2(X^k, \mathcal{C}^k_{W,n+1}, \mu^{\otimes k}) \subseteq L^2(X^k, \mathcal{C}^k_W, \mu^{\otimes k})$, which is a closed subspace by Claim 5. Hence, $T_{\boldsymbol{N}}\mathbf{1}_A \in L^2(X^k, \mathcal{C}^k_W, \mu^{\otimes k})$. Since the linear hull of $\{\mathbf{1}_A\}_{A\in\mathcal{C}^k_W}$ is dense in the closed subspace $L^2(X^k, \mathcal{C}^k_W, \mu^{\otimes k})$, linearity and continuity of $T_{\boldsymbol{N}}$ then yields that $L^2(X^k, \mathcal{C}^k_W, \mu^{\otimes k})$ is $T_{\boldsymbol{N}}$-invariant.

6: First, recall that $\mathcal{B}^{\otimes k}$ is permutation invariant. Moreover, if $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$, then $\pi(\mathcal{C}) \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ for every permutation $\pi\colon [k] \to [k]$. This implies that, if $\mathcal{X} \subseteq \mathcal{B}^{\otimes k}$ is a set with $\pi(\mathcal{X}) \subseteq \mathcal{X}$ for every permutation $\pi\colon [k] \to [k]$, then $\langle\mathcal{X}\rangle$ is permutation invariant. Hence, $\langle\{\varnothing, X^k\}\rangle$ is permutation invariant, and it suffices to show that, for a permutation-invariant $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$, both $\mathbb{T}_{\mathcal{A}^k\to W}(\mathcal{C})$ and $\mathbb{T}_{\mathcal{N}^k}(\mathcal{C})$ are permutation-invariant. Then, induction yields that $\mathcal{C}^k_{W,n}$ is permutation invariant for every $n \in \mathbb{N}$ and, hence, also $\mathcal{C}^k_W$ since $\pi\left(\bigcup_{n\in\mathbb{N}} \mathcal{C}^k_{W,n}\right) = \bigcup_{n\in\mathbb{N}} \pi(\mathcal{C}^k_{W,n}) \subseteq \bigcup_{n\in\mathbb{N}} \mathcal{C}^k_{W,n}$ for every permutation $\pi\colon [k] \to [k]$.

It remains to show that, for a permutation-invariant $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$, both $\mathbb{T}_{\mathcal{A}^k\to W}(\mathcal{C})$ and $\mathbb{T}_{\mathcal{N}^k}(\mathcal{C})$ are permutation-invariant. We prove the statement for $\mathbb{T}_{\mathcal{A}^k\to W}(\mathcal{C})$; the proof for $\mathbb{T}_{\mathcal{N}^k}(\mathcal{C})$ is analogous. To this end, we show that, for an arbitrary $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$, we have

$$\pi(\mathbb{T}_{\mathcal{A}^k\to W}(\mathcal{C})) = \mathbb{T}_{\mathcal{A}^k\to W}(\pi(\mathcal{C})) \tag{6}$$

for every permutation $\pi\colon [k] \to [k]$. Then, if $\mathcal{C}$ is permutation invariant, we get $\pi(\mathbb{T}_{\mathcal{A}^k\to W}(\mathcal{C})) = \mathbb{T}_{\mathcal{A}^k\to W}(\pi(\mathcal{C})) = \mathbb{T}_{\mathcal{A}^k\to W}(\mathcal{C})$ for every permutation $\pi\colon [k] \to [k]$.

To prove Equation (6), let $\pi\colon [k] \to [k]$ be a permutation and observe that $T_\pi \circ T_{\boldsymbol{A}^k_{ij}\to W} \circ T_{\pi^{-1}} = T_{\boldsymbol{A}^k_{\pi(i)\pi(j)}\to W}$ for all $i \neq j \in [k]$. As a side note, the analogous observation for $\mathbb{T}_{\mathcal{N}^k}(\mathcal{C})$ is

$T_\pi \circ T_{\boldsymbol{N}_j^k \to W} \circ T_{\pi^{-1}} = T_{\boldsymbol{N}_{\pi(j)}^k \to W}$ for every $j \in [k]$. We get that

$$\begin{aligned} T_{\boldsymbol{A}_{ij}^k \to W}(L^2(X^k, \pi(\mathcal{C}), \mu^{\otimes k})) &= T_{\boldsymbol{A}_{ij}^k \to W}(T_{\pi^{-1}}(L^2(X^k, \mathcal{C}, \mu^{\otimes k}))) \\ &= T_{\pi^{-1}}(T_{\boldsymbol{A}_{\pi(i)\pi(j)}^k \to W}(L^2(X^k, \mathcal{C}, \mu^{\otimes k}))). \end{aligned}$$

Hence, for $\mathcal{D} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$, we have

$$\begin{aligned} & T_{\boldsymbol{A}_{ij}^k \to W}(L^2(X^k, \pi(\mathcal{C}), \mu^{\otimes k})) \subseteq L^2(X^k, \mathcal{D}, \mu^{\otimes k}) \\ \iff & T_{\boldsymbol{A}_{ij}^k \to W}(L^2(X^k, \mathcal{C}, \mu^{\otimes k})) \subseteq T_\pi(L^2(X^k, \mathcal{D}, \mu^{\otimes k})) \\ \iff & T_{\boldsymbol{A}_{ij}^k \to W}(L^2(X^k, \mathcal{C}, \mu^{\otimes k})) \subseteq L^2(X^k, \pi^{-1}(\mathcal{D}), \mu^{\otimes k}). \end{aligned}$$

As the mapping $\boldsymbol{A}_{ij}^k \mapsto \boldsymbol{A}_{\pi(i)\pi(j)}^k$ is a permutation of $\mathcal{A}^k$ and we also have $\mathcal{D} \supseteq \pi(\mathcal{C}) \iff \pi^{-1}(\mathcal{D}) \supseteq \mathcal{C}$, this implies Equation (6). ☐

4.2 Weisfeiler-Leman Measures and Distributions

Before defining the mapping $\mathsf{owl}_W^k \colon X^k \to \mathbb{M}^k$, we have to define the space $\mathbb{M}^k$, which can be seen as the space of all colors used by oblivious k-WL. To this end, we have to state some facts regarding spaces of measures first. For a separable metrizable space $(X, \mathcal{T})$, let $\mathscr{P}(X)$ denote the set of all Borel probability measures on X. Let $C_b(X)$ denote the set of bounded continuous real-valued functions on X. We endow $\mathscr{P}(X)$ with the topology generated by the maps $\mu \mapsto \int f \, d\mu$ for $f \in C_b(X)$. Then, for $(\mu_i)_{i \in \mathbb{N}}$ with $\mu_i \in \mathscr{P}(X)$ and $\mu \in \mathscr{P}(X)$, the Portmanteau theorem states that the following three are equivalent [15, Theorem 17.20]:

1. $\mu_i \to \mu$.
2. $\int f d\mu_i \to \mu$ for every $f \in C_b(X)$.
3. $\int f d\mu_i \to \mu$ for every $f \in U_d(X)$.

Here, $U_d(X)$ denotes the set of bounded d-uniformly continuous real-valued functions on X and may clearly be replaced by some uniformly dense subset. If $(X, \mathcal{T})$ is compact, which is the case for the spaces we define, then $U_d(X) = C_b(X) = C(X)$, where $C(X)$ denotes the set of continuous real-valued functions on X. The Borel σ-algebra $\mathcal{B}(\mathscr{P}(X))$ is then generated by the maps $\mu \mapsto \mu(A)$ for $A \in \mathcal{B}(X)$ and also by the maps $\mu \mapsto \int f \, d\mu$ for bounded Borel real-valued functions f [15, Theorem 17.24]. If $(X, \mathcal{T})$ is Polish, then so is $P(X)$ [15, Theorem 17.23], which means that $(\mathscr{P}(X), \mathcal{B}(\mathscr{P}(X)))$ is again a standard Borel space for a standard Borel space $(X, \mathcal{B})$.

It is a standard fact that a compact metrizable space $K = (X, \mathcal{T})$ is separable [15, Proposition 4.6]. Hence, if we let $\mathcal{B}$ be denote the Borel σ-algebra generated by $\mathcal{T}$, then $(X, \mathcal{B})$ is a standard Borel space. The topological space $\mathscr{P}(X)$ is again compact metrizable [15, Theorem 17.22].

We are ready to define the space $\mathbb{M}^k$. One should pay attention to the connection to oblivious k-WL, cf. Section 1.2: Here, $P_0^k = [0,1]^{\binom{[k]}{2}}$ is the space of possible "edge weights" of a tuple $\bar{x} \in X^k$, generalizing possible atomic types. Moreover, oblivious k-WL defines k multisets of colors in every refinement, which results in k probability measures on the previous space $\mathbb{M}_n^k$ in the following definition.

Definition 23 (The Spaces $\mathbb{M}^k$ and $\mathbb{P}^k$)**.** *Let $k \geq 1$. Let $P_0^k \coloneqq [0,1]^{\binom{[k]}{2}}$ and inductively define $\mathbb{M}_n^k \coloneqq \prod_{i \leq n} P_i^k$ and $P_{n+1}^k \coloneqq \left(\mathscr{P}\left(\mathbb{M}_n^k\right)\right)^k$ for every $n \in \mathbb{N}$. Let $\mathbb{M}^k \coloneqq \mathbb{M}_\infty^k \coloneqq \prod_{n \in \mathbb{N}} P_i^k$ and, for $n \leq m \leq \infty$, let $p_{m,n} \colon \mathbb{M}_m^k \to \mathbb{M}_n^k$ be the natural projection. Finally, define*

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

As a product of a sequence of metrizable compact spaces, $\mathbb{M}^k$ is metrizable [6, Proposition 2.4.4] and also compact by Tychonoff's Theorem [6, Theorem 2.2.8]. Moreover, as $\mathbb{M}^k$ is a product of a sequence of second-countable spaces, the Borel σ-algebra of $\mathbb{M}^k$ and the product of the Borel σ-algebras of its factors are the same, cf. Section 2.1.

Note that $\alpha_{n+1} \in P^k_{n+1} = \left(\mathscr{P}\left(\mathbb{M}^k_n\right)\right)^k$ and $\alpha_{n+2} \in P^k_{n+2} = \left(\mathscr{P}\left(\mathbb{M}^k_{n+1}\right)\right)^k$ for $\alpha \in \mathbb{M}^k$ in the definition of $\mathbb{P}^k$, i.e., $\mathbb{P}^k$ is well-defined. This condition expresses that $\alpha_{n+2} \in \mathbb{P}^k$, which can be thought of as a coloring after $n+2$ refinement rounds, is consistent with α_{n+1} for every $n \in \mathbb{N}$, but it does not require that α_0 is consistent with α_1. One could add the additional consistency condition that, for $ij \in \binom{[k]}{2}$ and $u \notin ij$, the push-forward of $(\alpha_1)_u$ via the projection to component ij is the Dirac measure of $(\alpha_0)_{ij}$, but this would introduce an inconsistency in the case $k = 2$ where there is no such u. For simplicity, we just leave this out; it does not cause any problems for us.

In terms of graphs, an element $(\alpha_0, \alpha_1, \dots)$ of $\mathbb{M}^k$ can be thought of as a sequence of unfoldings of a graph, cf. [5], of heights $0, 1, 2, \dots$. These unfoldings, however, do not have to be related in any way. The subspace $\mathbb{P}^k$ contains these sequences where each unfolding is a continuation of the previous one. These sequences can also be viewed as a single, infinite unfolding: By the Kolmogorov Consistency Theorem [15, Exercise 17.16], for all $\alpha \in \mathbb{P}^k$ and $j \in [k]$, there is a unique measure $\mu^\alpha_j \in \mathscr{P}(\mathbb{M}^k)$ such that $(p_{\infty,n})_* \mu^\alpha_j = (\alpha_{n+1})_j$ for every $n \in \mathbb{N}$. Moreover, one can verify that this mapping $\alpha \mapsto \mu^\alpha_j$ is continuous, cf. [9, Claim 6.2].

Lemma 24. *$\mathbb{P}^k$ is closed in $\mathbb{M}^k$ and $\mathbb{P}^k \to \mathscr{P}(\mathbb{M}^k)$, $\alpha \mapsto \mu^\alpha_j$ is continuous for every $j \in [k]$.*

Proof. To prove that $\mathbb{P}^k$ is closed, let $\alpha_i \to \alpha$ with $\alpha_i \in \mathbb{P}^k$ for every $i \in \mathbb{N}$ and $\alpha \in \mathbb{M}^k$. Let $j \in [k]$ and $n \in \mathbb{N}$. By definition of the product topology, we have $((\alpha_i)_{n+2})_j \to (\alpha_{n+2})_j$, which yields

$$\begin{aligned}\int_{\mathbb{M}^k_n} f\, d((\alpha_i)_{n+1})_j \overset{\alpha_i \in \mathbb{P}^k}{=} \int_{\mathbb{M}^k_n} f\, d(p_{n+1,n})_*((\alpha_i)_{n+2})_j &= \int_{\mathbb{M}^k_{n+1}} f \circ p_{n+1,n}\, d((\alpha_i)_{n+2})_j \\ &\to \int_{\mathbb{M}^k_{n+1}} f \circ p_{n+1,n}\, d(\alpha_{n+2})_j \\ &= \int_{\mathbb{M}^k_n} f\, d(p_{n+1,n})_*(\alpha_{n+2})_j.\end{aligned}$$

for every $f \in C(\mathbb{M}^k_n)$. Therefore, $((\alpha_i)_{n+1})_j \to (p_{n+1,n})_*(\alpha_{n+2})_j$. Since also $((\alpha_i)_{n+1})_j \to (\alpha_{n+1})_j$ and the metrizable space $\mathscr{P}(\mathbb{M}^k_n)$ is Hausdorff, we get $(\alpha_{n+1})_j = (p_{n+1,n})_*(\alpha_{n+2})_j$. Hence, $\alpha \in \mathbb{P}^k$.

Let $j \in [k]$. Let $\alpha_i \to \alpha$ with $\alpha_i \in \mathbb{P}^k$ for every $i \in \mathbb{N}$ and $\alpha \in \mathbb{P}^k$. To prove that $\mu^{\alpha_i}_j \to \mu^\alpha_j$, we observe that

$$\begin{aligned}\int_{\mathbb{M}^k} f \circ p_{\infty,n}\, d\mu^{\alpha_i}_j = \int_{\mathbb{M}^k_n} f\, d(p_{\infty,n})_*\mu^{\alpha_i}_j = \int_{\mathbb{M}^k_n} f\, d((\alpha_i)_{n+1})_j \to \int_{\mathbb{M}^k_n} f\, d(\alpha_{n+1})_j &= \int_{\mathbb{M}^k_n} f\, d(p_{\infty,n})_*\mu^\alpha_j \\ &= \int_{\mathbb{M}^k} f \circ p_{\infty,n}\, d\mu^\alpha_j\end{aligned}$$

for every $n \in \mathbb{N}$ and every $f \in C(\mathbb{M}^k_n)$. This already proves the claim as the set $\bigcup_{n \in \mathbb{N}} C(\mathbb{M}^k_n) \circ p_{\infty,n}$ is uniformly dense in $C(\mathbb{M}^k)$ by the Stone-Weierstrass theorem [6, Theorem 2.4.11]; in particular, this set separates points by the definition of the product topology and the fact that every metrizable space is completely Hausdorff. ☐

Lemma 24 implies that $\mathbb{P}^k \in \mathcal{B}(\mathbb{M}^k)$ and that $\mathbb{P}^k \to \mathbb{R}$, $\alpha \mapsto \int f \, d\mu_j^\alpha$ is measurable for every bounded measurable real-valued function f on $\mathbb{M}^k$ and every $j \in [k]$, cf. the definition of $\mathscr{P}(\mathbb{M}^k)$. This justifies the following definition of a k-WL distribution, which intuitively generalizes the concept of a multiset of colors with the additional constraints that, first, that the non-consistent sequences $\alpha \in \mathbb{M}^k$ have measure zero and, second, it satisfies a variant of the Tonelli-Fubini theorem w.r.t. the measures given by the mappings $\mathbb{P}^k \to \mathscr{P}(\mathbb{M}^k)$, $\alpha \mapsto \mu_j^\alpha$.

Definition 25. *Let $k \geq 1$. A measure $\nu \in \mathscr{P}(\mathbb{M}^k)$ is called a k-Weisfeiler-Leman distribution (k-WLD) if*

1. $\nu(\mathbb{P}^k) = 1$ *and*
2. $\int_{\mathbb{M}^k} f \, d\nu = \int_{\mathbb{M}^k} \left(\int_{\mathbb{M}^k} f \, d\mu_j^\alpha \right) d\nu(\alpha)$ *for every bounded measurable $f \colon \mathbb{M}^k \to \mathbb{R}$ and every $j \in [k]$.*

4.3 The Mapping owl_W^k

Having defined the compact metrizable space $\mathbb{M}^k$, we can finally define the mapping $\mathsf{owl}_W^k \colon X^k \to \mathbb{M}^k$ and the k-WL distribution ν_W^k for a graphon W. To this end, let us first recall that oblivious k-WL for a graph G initially colors a k-tuple $\bar{v} \in V(G)^k$ by its atomic type, which includes the information of which vertices in $\bar{v}$ are equal and which are connected by an edge. In our case, this becomes somewhat simpler since we do deal with the case that entries of a k-tuple $\bar{x} \in X^k$ are equal; if our standard Borel space is atom free, such diagonal sets have measure zero in the product space and do not matter. Hence, we only include the information $W(x_i, x_j)$ for every $ij \in \binom{[k]}{2}$. Notice the connection to the operators $\mathbb{T}_{\mathcal{A}^k \to W}$: by definition, we have $(T_{\boldsymbol{A}_{ij}^k \to W} f)(\bar{x}) = W(x_i, x_j) \cdot f(\bar{x})$ for every $f \in L^2(X^k, \mu^{\otimes k})$ and $\mu^{\otimes k}$-almost every $\bar{x} \in X^k$.

Let us also take a look at the substitution operation in the refinement rounds of oblivious k-WL. Fix $\bar{x} \in X^k$ and $j \in [k]$ in the following. Define $\bar{x}[/j] \coloneqq (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k) \in X^{k-1}$ to be the tuple obtained from $\bar{x}$ by removing the jth component, and for $y \in X$, also $\bar{x}[y/j] \coloneqq (x_1, \dots, x_{j-1}, y, x_{j+1}, \dots, x_k) \in X^k$, which is the tuple obtained from $\bar{x}$ by replacing the jth component by y. The preimage of a set $A \subseteq X^k$ under the map $\bar{x}[\cdot/j] \colon X \to X^k$, $y \mapsto \bar{x}[y/j]$ is

$$\bar{x}[\cdot/j]^{-1}(A) = \{y \in X \mid \bar{x}[y/j] \in A\} =: A_{\bar{x}[/j]},$$

which we call the *section of A determined by $\bar{x}[/j]$*. Note that, technically, $A_{\bar{x}[/j]}$ also depends on j and not only on the $(k-1)$-tuple $\bar{x}[/j] \in X^{k-1}$, but we nevertheless stick to this notation. The mapping $\bar{x}[\cdot/j]$ is measurable, i.e., we have $A_{\bar{x}[/j]} \in \mathcal{B}$ for every $A \in \mathcal{B}^{\otimes k}$ [2, Theorem 18.1 (i)]. If we let $p_j \colon X^k \to X$ denote the projection to the jth component, which is measurable by definition of $\mathcal{B}^{\otimes k}$, then, the mapping $\bar{x}[\cdot/j] \circ p_j \colon X^k \to X^k, \bar{y} \mapsto \bar{x}[y_j/j]$ is measurable as the composition of measurable functions and we have $(\bar{x}[\cdot/j] \circ p_j)_* \mu^{\otimes k} = \bar{x}[\cdot/j]_* \mu$. To see the connection to the operators $\mathbb{T}_{\mathcal{N}^k}$, note that the definition of $T_{\boldsymbol{N}_j^k}$ yields that

$$(T_{\boldsymbol{N}_j^k} f)(\bar{x}) = \int_X f(\bar{x}[y/j]) \, d\mu(y) = \int_X f \circ \bar{x}[\cdot/j] \, d\mu = \int_{X^k} f d\left(\bar{x}[\cdot/j]_* \mu\right) \tag{7}$$

for every $f \in L^2(X^k, \mu^{\otimes k})$ and $\mu^{\otimes k}$-almost every $\bar{x} \in X^k$.

Definition 26 (The Mapping owl_W^k)**.** *Let $k \geq 1$ and $W\colon X \times X \to [0,1]$ be a graphon. Define $\mathsf{owl}_{W,0}^k\colon X^k \to \mathbb{M}_0^k$ by*

$$\mathsf{owl}_{W,0}^k(\bar{x}) \coloneqq \big(W(x_i, x_j)\big)_{ij \in \binom{[k]}{2}}$$

for every $\bar{x} \in X^k$. Inductively define $\mathsf{owl}_{W,n+1}^k\colon X^k \to \mathbb{M}_{n+1}^k$ by

$$\mathsf{owl}_{W,n+1}^k(\bar{x}) \coloneqq \left(\mathsf{owl}_{W,n}^k(\bar{x}), \left(\left(\mathsf{owl}_{W,n}^k \circ \bar{x}[\cdot/j]\right)_* \mu\right)_{j\in[k]}\right)$$

for every $\bar{x} \in X^k$. Then, let $\mathsf{owl}_W^k = \mathsf{owl}_{W,\infty}^k\colon X^k \to \mathbb{M}^k$ be the mapping defined by $(\mathsf{owl}_W^k(\bar{x}))_n \coloneqq (\mathsf{owl}_{W,\infty}^k(\bar{x}))_n \coloneqq (\mathsf{owl}_{W,n}^k(\bar{x}))_n$ for all $n \in \mathbb{N}$, $\bar{x} \in X^k$. Finally, let $\nu_W^k \coloneqq \mathsf{owl}_{W*}^k \mu^{\otimes k} \in \mathscr{P}(\mathbb{M}^k)$ be the push-forward of $\mu^{\otimes k}$ via owl_W^k.

An immediate consequence of Definition 26, which we often use, is that ${\mathsf{owl}_{W,m}^k}^{-1}(p_{m,n}^{-1}(A)) = {\mathsf{owl}_{W,n}^k}^{-1}(A)$ holds for all $1 \leq n < m \leq \infty$ and every $A \in \mathcal{B}(\mathbb{M}_n^k)$. In particular, we use it to prove that the mapping $\mathsf{owl}_{W,n}^k$ is measurable for every $n \in \mathbb{N} \cup \{\infty\}$, which actually is needed for everything in Definition 26 to be well defined. Lemma 27 states not only that $\mathsf{owl}_{W,n}^k$ is measurable but also that the minimum $\mu^{\otimes k}$-relatively complete sub-σ-algebra that makes it measurable is given by $\mathcal{C}_{W,n}^k$, cf. [9, Proposition 6.6].

Lemma 27. *Let $k \geq 1$ and $W\colon X \times X \to [0,1]$ be a graphon. For $n \in \mathbb{N} \cup \{\infty\}$,*

$$\mathcal{C}_{W,n}^k = \left\langle \left\{ {\mathsf{owl}_{W,n}^k}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}_n^k) \right\} \right\rangle.$$

Proof. Let $\mathcal{D}_n \coloneqq \langle \{ {\mathsf{owl}_{W,n}^k}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}_n^k) \} \rangle$. First, we prove $\mathcal{C}_{W,n}^k = \mathcal{D}_n$ for every $n \in \mathbb{N}$ by induction on n. For the induction basis $n = 0$, we have

$$\mathcal{D}_0 = \left\langle \left\{ {\mathsf{owl}_{W,0}^k}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}_0^k) \right\} \right\rangle = \left\langle \left\{ {\mathsf{owl}_{W,0}^k}^{-1}(A) \mid A \in \mathcal{B}([0,1]^{\binom{[k]}{2}}) \right\} \right\rangle$$

The Borel σ-algebra $\mathcal{B}([0,1]^{\binom{[k]}{2}})$ is generated by the sets of the form $\prod_{ij \in \binom{[k]}{2}} A_{ij}$ where $A_{ij} \in \mathcal{B}([0,1])$ and $A_{ij} = [0,1]$ for all but at most one ij [15, Section 10.B]. Since it suffices to check measurability of a function for a generating set [6, Theorem 4.1.6], we may replace $\mathcal{B}([0,1]^{\binom{[k]}{2}})$ by a generating set in the definition of $\mathcal{D}_0$, which yields that

$$\begin{aligned}
\mathcal{D}_0 &= \left\langle \left\{ {\mathsf{owl}_{W,0}^k}^{-1}(A) \mid A \in \mathcal{B}([0,1]^{\binom{[k]}{2}}) \right\} \right\rangle \\
&= \left\langle \left\{ \left\{ \bar{x} \in X^k \mid (W(x_i, x_j))_{ij \in \binom{[k]}{2}} \in A \right\} \mid A \in \mathcal{B}([0,1]^{\binom{[k]}{2}}) \right\} \right\rangle \\
&= \left\langle \left\{ \left\{ \bar{x} \in X^k \mid W(x_i, x_j) \in A \right\} \mid A \in \mathcal{B}([0,1]),\, ij \in \tbinom{[k]}{2} \right\} \right\rangle \\
&= \left\langle \textstyle\bigcup_{ij \in \binom{[k]}{2}} \left\{ \left\{ \bar{x} \in X^k \mid W(x_i, x_j) \in A \right\} \mid A \in \mathcal{B}([0,1]) \right\} \right\rangle \\
&= \left\langle \textstyle\bigcup_{\boldsymbol{A} \in \mathcal{A}^k} \left\{ (T_{\boldsymbol{A} \to W})^{-1}(A) \mid A \in \mathcal{B}([0,1]) \right\} \right\rangle \\
&= \mathcal{C}_{W,0}^k. && \text{(Lemma 22 and Lemma 22 2)}
\end{aligned}$$

For the inductive step, let $n \in \mathbb{N}$. We have to prove that $\mathcal{C}_{W,n+1}^k = \mathcal{D}_{n+1}$. Recall that we have $\mathbb{M}_{n+1}^k = \mathbb{M}_n^k \times \left(\mathscr{P}\left(\mathbb{M}_n^k\right)\right)^k$ by definition and that the Borel σ-algebra $\mathcal{B}(\mathscr{P}\left(\mathbb{M}_n^k\right))$ is generated by the maps $\mu \mapsto \mu(A)$ for $A \in \mathcal{B}(\mathbb{M}_n^k)$ [15, Theorem 17.24]. Hence, by definition of the product σ-algebra and since it suffices to check measurability of a function for a generating set [6, Theorem 4.1.6], $\mathcal{B}(\mathbb{M}^k_{n+1})$ is the smallest σ-algebra containing $\{p^{-1}_{n+1,n}(A) \mid A \in \mathcal{B}(\mathbb{M}^k_n)\}$ and making the maps

$$\mathbb{M}^k_{n+1} \ni \alpha \mapsto ((\alpha)_{n+1})_j(A)$$

for $A \in \mathcal{B}(\mathbb{M}^k_n)$ and $j \in [k]$ measurable. Again by [6, Theorem 4.1.6], this means that $\mathcal{D}_{n+1}$ is the smallest $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$ containing

$$\left\{ {\mathsf{owl}^k_{W,n+1}}^{-1}(p^{-1}_{n+1,n}(A)) \mid A \in \mathcal{B}(\mathbb{M}^k_n)\right\} = \left\{ {\mathsf{owl}^k_{W,n}}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}^k_n)\right\}$$

and making the maps

$$\begin{aligned} X^k \ni \bar{x} \mapsto ((\mathsf{owl}^k_{W,n+1}(\bar{x}))_{n+1})_j(A) &= \left(\left(\mathsf{owl}^k_{W,n} \circ \bar{x}[\cdot/j]\right)_* \mu\right)(A) \\ &= \int_{\mathbb{M}^k_n} \mathbf{1}_A d \left(\mathsf{owl}^k_{W,n} \circ \bar{x}[\cdot/j]\right)_* \mu \\ &= \int_{X^k} \mathbf{1}_A \circ \mathsf{owl}^k_{W,n} \, d\,\bar{x}[\cdot/j]_* \mu \\ &= (T_{\boldsymbol{N}^k_j} \mathbf{1}_A \circ \mathsf{owl}^k_{W,n})(\bar{x}), \end{aligned}$$

for $A \in \mathcal{B}(\mathbb{M}^k_n)$ and $j \in [k]$ measurable, where the equalities hold $\mu^{\otimes k}$-almost everywhere, cf. also Equation (7).

To see that $\mathcal{D}_{n+1} \subseteq \mathcal{C}^k_{W,n+1}$, we verify that $\mathcal{C}^k_{W,n+1}$ contains the aforementioned sets and that the aforementioned maps are measurable for it. We have

$$\left\{ {\mathsf{owl}^k_{W,n}}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}^k_n)\right\} \overset{\text{def.}}{\subseteq} \mathcal{D}_n \overset{\text{IH}}{\subseteq} \mathcal{C}^k_{W,n} \overset{\text{def.}}{\subseteq} \mathcal{C}^k_{W,n+1}.$$

By the induction hypothesis, $\mathsf{owl}^k_{W,n}$ is $\mathcal{C}^k_{W,n}$-measurable, and since $A \in \mathcal{B}(\mathbb{M}^k_n)$, so is $\mathbf{1}_A \circ \mathsf{owl}^k_{W,n}$. Hence, by definition of $\mathcal{C}^k_{W,n+1}$, $T_{\boldsymbol{N}^k_j} \mathbf{1}_A \circ \mathsf{owl}^k_{W,n}$ is $\mathcal{C}^k_{W,n+1}$-measurable, which is just what we wanted to prove.

It remains to verify that $\mathcal{C}^k_{W,n+1} \subseteq \mathcal{D}_{n+1}$. By Lemma 22 3, it suffices to prove that $\mathcal{D}_{n+1}$ contains $\mathcal{C}^k_{W,n}$ and makes the functions $T_{\boldsymbol{N}} \mathbf{1}_A$ for $\boldsymbol{N} \in \mathcal{N}^k$ and $A \in \mathcal{C}^k_{W,n}$ measurable. We have

$$\mathcal{C}^k_{W,n} \overset{\text{IH}}{\subseteq} \mathcal{D}_n = \left\langle \left\{ {\mathsf{owl}^k_{W,n}}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}^k_n)\right\} \right\rangle \subseteq \mathcal{D}_{n+1}.$$

Let $A \in \mathcal{C}^k_{W,n}$. By the induction hypothesis, we have $A \in \mathcal{D}_n$. Since the preimage of a σ-algebra is a σ-algebra, we have $A = \mathsf{owl}^k_{W,n}(B) \triangle Z$ for some $B \in \mathcal{B}(\mathbb{M}^k_n)$ and $Z \in \mathcal{B}^{\otimes k}$ with $\mu^{\otimes k}(Z) = 0$. Then, $\bar{x} \in A \iff \mathsf{owl}^k_{W,n}(\bar{x}) \in B$ for every $\bar{x} \notin Z$, i.e., $\mathbf{1}_B \circ \mathsf{owl}^k_{W,n} = \mathbf{1}_A$, where the equality holds $\mu^{\otimes k}$-almost everywhere. Let $j \in [k]$. We know that $\mathcal{D}_{n+1}$ makes the map $T_{\boldsymbol{N}^k_j} \mathbf{1}_B \circ \mathsf{owl}^k_{W,n} = T_{\boldsymbol{N}^k_j} \mathbf{1}_A$ measurable, but this is already what we wanted to show.

It remains to prove that

$$\mathcal{C}^k_W = \left\langle \left\{ {\mathsf{owl}^k_W}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}^k)\right\} \right\rangle,$$

where, by definition, we have $\mathcal{C}_W^k = \langle \bigcup_{n\in\mathbb{N}} \mathcal{C}_{W,n}^k \rangle$. It is easy to see that the Borel σ-algebra $\mathcal{B}(\mathbb{M}^k)$ is generated by the projections $p_{\infty,n}$. Hence, by [6, Theorem 4.1.6],

$$\begin{aligned}
\mathcal{C}_W^k = \left\langle \bigcup_{n\in\mathbb{N}} \mathcal{C}_{W,n}^k \right\rangle &= \left\langle \bigcup_{n\in\mathbb{N}} \left\{ {\mathsf{owl}_{W,n}^k}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}_n^k) \right\} \right\rangle \\
&= \left\langle \left\{ {\mathsf{owl}_{W,n}^k}^{-1}(A) \mid n \in \mathbb{N},\, A \in \mathcal{B}(\mathbb{M}_n^k) \right\} \right\rangle \\
&= \left\langle \left\{ {\mathsf{owl}_{W}^k}^{-1}(p_{\infty,n}^{-1}(A)) \mid n \in \mathbb{N},\, A \in \mathcal{B}(\mathbb{M}_n^k) \right\} \right\rangle \\
&= \left\langle \left\{ {\mathsf{owl}_{W}^k}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}^k) \right\} \right\rangle
\end{aligned}$$

□

By Lemma 27, $\mathcal{C}_W^k$ is the minimum $\mu^{\otimes k}$-relatively complete sub-σ-algebra that makes owl_W^k measurable. Hence $\mathsf{owl}_W^k \colon X^k \to \mathbb{M}^k$ is a measurable and measure-preserving mapping from the measure space $(X^k, \mathcal{B}^{\otimes k}, \mu^{\otimes k})$ to $(\mathbb{M}^k, \mathcal{B}(\mathbb{M}^k), \nu_W^k)$ and we can consider the Koopman operator $T_{\mathsf{owl}_W^k} \colon L^2(\mathbb{M}^k, \nu_W^k) \to L^2(X^k, \mu^{\otimes k})$ of owl_W^k, which is a Markov embedding [8, Theorem 7.20]. More precisely, by Claim 7, it is an isometry onto $L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$. In addition, the operator $S_{\mathcal{C}_W^k}$ of Claim 6 becomes a Markov isomorphism when restricted to $L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$ which means that $R_W^k \coloneqq S_{\mathcal{C}_W^k} \circ T_{\mathsf{owl}_W^k}$ is a Markov isomorphism.

Corollary 28. *Let $k \geq 1$ and $W \colon X \times X \to [0,1]$ be a graphon. Then, $R_W^k \colon L^2(\mathbb{M}^k, \nu_W^k) \to L^2(X^k/\mathcal{C}_W^k, \mu^{\otimes k}/\mathcal{C}_W^k)$ is a Markov isomorphism.*

It remains to verify that ν_W^k is in fact a k-WLD. The following lemma can also be seen as a justification of the definition of a k-WLD. In particular, it shows that Tonelli-Fubini-like requirement in Definition 25 actually stems from the Tonelli-Fubini theorem. In other words, the definition of a k-WLD is chosen such that it captures the essential properties of ν_W^k that make it possible to define the analogue of $\mathbb{T}_W^k$ on the space $L^2(\mathbb{M}^k, \nu_W^k)$. In the next section, we define these operators on $L^2(\mathbb{M}^k, \nu)$ for an arbitrary k-WLD ν.

Lemma 29. *Let $k \geq 1$ and $W \colon X \times X \to [0,1]$ be a graphon. Then,*

1. $\mu_j^{\mathsf{owl}_W^k(\bar{x})} = (\mathsf{owl}_W^k \circ \bar{x}[\cdot/j])_* \mu$ *for all* $j \in [k]$, $\bar{x} \in X^k$,
2. $\mathsf{owl}_W^k(X^k) \subseteq \mathbb{P}^k$, *and*
3. ν_W^k *is a k-WLD.*

Proof. 1: For $n \in \mathbb{N}$ and $A \in \mathcal{B}(\mathbb{M}_n^k)$, we have

$$\begin{aligned}
\mu_j^{\mathsf{owl}_W^k(\bar{x})}(p_{\infty,n}^{-1}(A)) = (p_{\infty,n})_* \mu_j^{\mathsf{owl}_W^k(\bar{x})}(A) &= ((\mathsf{owl}_W^k(\bar{x}))_{n+1})_j(A) && (\text{Definition } \mu_j^{\mathsf{owl}_W^k(\bar{x})}) \\
&= ((\mathsf{owl}_{W,n+1}^k(\bar{x}))_{n+1})_j(A) && (\text{Definition } \mathsf{owl}_W^k) \\
&= (\mathsf{owl}_{W,n}^k \circ \bar{x}[\cdot/j])_* \mu(A) && (\text{Definition } \mathsf{owl}_{W,n+1}^k) \\
&= (\mathsf{owl}_W^k \circ \bar{x}[\cdot/j])_* \mu(p_{\infty,n}^{-1}(A)).
\end{aligned}$$

That is, $\mu_j^{\mathsf{owl}_W^k(\bar{x})}$ and $(\mathsf{owl}_W^k \circ \bar{x}[\cdot/j])_* \mu$ both are probability measures that agree on the set $\bigcup_{n\in\mathbb{N}} \{p_{\infty,n}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}_n^k)\}$, which generates $\mathcal{B}(\mathbb{M}^k)$. By the π-λ theorem [15, Theorem 10.1 iii)], they agree on all of $\mathcal{B}(\mathbb{M}^k)$.

2: Let $\bar{x} \in X^k$. For $j \in [k]$ and $n \in \mathbb{N}$, we have

$$\begin{aligned}
(p_{n+1,n})_*((\mathsf{owl}^k_W(\bar{x}))_{n+2})_j &= (p_{n+1,n})_*((\mathsf{owl}^k_{W,n+2}(\bar{x}))_{n+2})_j && \text{(Definition } \mathsf{owl}^k_W\text{)}\\
&= (p_{n+1,n})_*((\mathsf{owl}^k_{W,n+1} \circ \bar{x}[\cdot/j])_*\mu) && \text{(Definition } \mathsf{owl}^k_{W,n+2}\text{)}\\
&= (\mathsf{owl}^k_{W,n} \circ \bar{x}[\cdot/j])_*\mu\\
&= ((\mathsf{owl}^k_{W,n+1}(\bar{x}))_{n+1})_j && \text{(Definition } \mathsf{owl}^k_{W,n+1}\text{)}\\
&= ((\mathsf{owl}^k_W(\bar{x}))_{n+1})_j. && \text{(Definition } \mathsf{owl}^k_W\text{)}
\end{aligned}$$

Hence, $\mathsf{owl}^k_W(\bar{x}) \in \mathbb{P}^k$.

3: By 2, we have $\nu^k_W(\mathbb{P}^k) = \mu^{\otimes k}({\mathsf{owl}^k_W}^{-1}(\mathbb{P}^k)) = \mu^{\otimes k}(X^k) = 1$. Let $j \in [k]$. Let $f \colon \mathbb{M}^k \to \mathbb{R}$ be bounded and measurable. We have

$$\begin{aligned}
\int\limits_{\mathbb{M}^k} f \, d\nu^k_W \overset{\text{def. } \nu^k_W}{=} \int\limits_{\mathbb{M}^k} f \, d\mathsf{owl}^k_{W*}\mu^{\otimes k} &= \int\limits_{X^k} f \circ \mathsf{owl}^k_W \; d\mu^{\otimes k}\\
&\overset{\text{T.-F.}}{=} \int\limits_{X^{k-1}} \Big(\int\limits_{X} f \circ \mathsf{owl}^k_W(\bar{x}[y/j]) \, d\mu(y) \Big) \, d\mu^{\otimes k-1}(\bar{x}[/j])\\
&\qquad\qquad (x_j \in X \text{ arb.})\\
&= \int\limits_{X^k} \Big(\int\limits_{X} f \circ \mathsf{owl}^k_W(\bar{x}[y/j]) \, d\mu(y) \Big) \, d\mu^{\otimes k}(\bar{x})\\
&= \int\limits_{X^k} \Big(\int\limits_{\mathbb{M}^k} f \, d(\mathsf{owl}^k_W \circ \bar{x}[\cdot/j])_*\mu \Big) \, d\mu^{\otimes k}(\bar{x})\\
&\overset{1}{=} \int\limits_{X^k} \Big(\int\limits_{\mathbb{M}^k} f \, d\mu_j^{\mathsf{owl}^k_W(\bar{x})} \Big) \, d\mu^{\otimes k}(\bar{x})\\
&= \int\limits_{\mathbb{M}^k} \Big(\int\limits_{\mathbb{M}^k} f \, d\mu^\alpha_j \Big) \, d\mathsf{owl}^k_{W*}\mu \overset{\text{def. } \nu^k_W}{=} \int\limits_{\mathbb{M}^k} \Big(\int\limits_{\mathbb{M}^k} f \, d\mu^\alpha_j \Big) \, d\nu^k_W.
\end{aligned}$$

□

4.4 Operators and Weisfeiler-Leman Measures

For a graphon W, the operator $R^k_W \colon L^2(\mathbb{M}^k, \nu^k_W) \to L^2(X^k/\mathcal{C}^k_W, \mu^{\otimes k}/\mathcal{C}^k_W)$ is a Markov isomorphism by Corollary 28. Hence, if U is another graphon with $\nu^k_U = \nu^k_W$, then $R^k_U \circ (R^k_W)^*$ is a Markov isomorphism from $L^2(X^k/\mathcal{C}^k_W, \mu^{\otimes k}/\mathcal{C}^k_W)$ to $L^2(X^k/\mathcal{C}^k_U, \mu^{\otimes k}/\mathcal{C}^k_U)$. However, we still lack that this Markov isomorphism "maps" the family $\mathbb{T}^k_W/\mathcal{C}^k_W$ to $\mathbb{T}^k_U/\mathcal{C}^k_U$. To close this gap, we show that we can define a family $\mathbb{T}_{\nu^k_W}$ of operators on $L^2(\mathbb{M}^k, \nu^k_W)$ such that R^k_W "maps" $\mathbb{T}_{\nu^k_W}$ to $\mathbb{T}^k_W/\mathcal{C}^k_W$. This replaces the graphon $\mathbb{M} \times \mathbb{M} \to [0, 1]$ defined by Grebík and Rocha [9]. Let us begin with operators for neighbor graphs as this is the interesting case; in particular, it shows why we have the Tonelli-Fubini-like requirement in the definition of a k-WLD.

Lemma 30. *Let $k \geq 1$, and let $\nu \in \mathscr{P}(\mathbb{M}^k)$ be a k-WLD. Let $j \in [k]$. Setting*

$$(T_{\boldsymbol{N}^k_j \to \nu} f)(\alpha) \coloneqq \int_{\mathbb{M}^k} f \, d\mu^\alpha_j$$

for all $f \in L^\infty(\mathbb{M}^k, \nu)$, $\alpha \in \mathbb{M}^k$ defines an L^∞-contraction that uniquely extends to an L^2-contraction $L^2(\mathbb{M}^k, \nu) \to L^2(\mathbb{M}^k, \nu)$.

Proof. We show that the definition yields a well-defined contraction $T_{\boldsymbol{N}_j^k \to \nu}$ on $L^\infty(\mathbb{M}^k, \nu)$ such that $\|T_{\boldsymbol{N}_j^k \to \nu} f\|_2 \le \|f\|_2$ for every $f \in L^\infty(\mathbb{M}^k, \nu)$. Then, $T_{\boldsymbol{N}_j^k \to \nu}$ uniquely extends to a contraction on $L^2(\mathbb{M}^k, \nu)$ since $L^\infty(\mathbb{M}^k, \nu)$ is dense in $L^2(\mathbb{M}^k, \nu)$.

The definition of a k-WLD immediately yields that, if $A \in \mathcal{B}(\mathbb{M}^k)$ with $\nu(A) = 0$, then $\mu_j^\alpha(A) = 0$ for ν-almost every $\alpha \in \mathbb{M}^k$. Hence, if a property holds ν-almost everywhere, it holds μ_j^α-almost everywhere for ν-almost every $\alpha \in \mathbb{M}^k$. Let $f \in \mathcal{L}^\infty(\mathbb{M}^k, \nu)$. Then, $|f| \le \|f\|_\infty$ ν-almost everywhere, and hence, $|f| \le \|f\|_\infty$ holds μ_j^α-almost everywhere for ν-almost every $\alpha \in \mathbb{M}^k$. Thus, for ν-almost every $\alpha \in \mathbb{M}^k$, we have

$$\Big| \int_{\mathbb{M}^k} f \, d\mu_j^\alpha \Big| \le \int_{\mathbb{M}^k} |f| \, d\mu_j^\alpha \le \int_{\mathbb{M}^k} \|f\|_\infty \, d\mu_j^\alpha = \|f\|_\infty,$$

which yields that $\|T_{\boldsymbol{N}_j^k \to \nu} f\|_\infty \le \|f\|_\infty$. In particular, if $f, g \in \mathcal{L}^\infty(\mathbb{M}^k, \nu)$ are equal ν-almost everywhere, then

$$\|T_{\boldsymbol{N}_j^k \to \nu} f - T_{\boldsymbol{N}_j^k \to \nu} g\|_\infty = \|T_{\boldsymbol{N}_j^k \to \nu}(f - g)\|_\infty \le \|f - g\|_\infty = 0,$$

that is, $T_{\boldsymbol{N}_j^k \to \nu} f$ and $T_{\boldsymbol{N}_j^k \to \nu} g$ are equal ν-almost everywhere. Here we used that the mapping $T_{\boldsymbol{N}_j^k \to \nu}$ is linear, which follows directly from the linearity of the integral. Recall that $\mathbb{P}^k \to \mathbb{R}, \alpha \mapsto \int f \mu_j^\alpha$ is measurable for every bounded measurable $\mathbb{R}$-valued function f on $\mathbb{M}^k$ by Lemma 24 and the definition of $\mathscr{P}(\mathbb{P}^k)$. Since $\mathbb{P}^k \in \mathcal{B}(\mathbb{M}^k)$ by Lemma 24 and $\nu(\mathbb{P}^k) = 1$, this combined with the previous considerations yields that $T_{\boldsymbol{N}_j^k \to \nu}$ is a well-defined mapping $L^\infty(\mathbb{M}^k, \nu) \to L^\infty(\mathbb{M}^k, \nu)$.

It remains to show that $\|T_{\boldsymbol{N}_j^k \to \nu} f\|_2 \le \|f\|_2$ for every $f \in L^\infty(\mathbb{M}^k, \nu)$. We have

$$\begin{aligned}
\|T_{\boldsymbol{N}_j^k \to \nu} f\|_2^2 = \int_{\mathbb{M}^k} \Big(\int_{\mathbb{M}^k} f \, d\mu_j^\alpha \Big)^2 d\nu(\alpha) &\overset{\text{C.-S.}}{\le} \int_{\mathbb{M}^k} \Big(\int_{\mathbb{M}^k} f^2 \, d\mu_j^\alpha \Big) d\nu(\alpha) && \text{(Cauchy-Schwarz)} \\
&= \int_{\mathbb{M}^k} f^2 \, d\nu && \text{(def. } k\text{-WLD)} \\
&= \|f\|_2^2
\end{aligned}$$

Note that we again used that $|f| \le \|f\|_\infty$ holds μ_j^α-almost everywhere for ν-almost every $\alpha \in \mathbb{M}^k$ in order to apply the Cauchy-Schwarz inequality. ☐

The following lemma states that Lemma 30 is indeed the right definition.

Lemma 31. *Let $k \ge 1$ and $W : X \times X \to [0, 1]$ be a graphon. For every $\boldsymbol{N} \in \mathcal{N}^k$,*

1. $T_{\boldsymbol{N}} \circ T_{\mathsf{owl}_W^k} = T_{\mathsf{owl}_W^k} \circ T_{\boldsymbol{N} \to \nu_W^k}$,
2. $(T_{\boldsymbol{N}})_{\mathcal{C}_W^k} \circ T_{\mathsf{owl}_W^k} = T_{\mathsf{owl}_W^k} \circ T_{\boldsymbol{N} \to \nu_W^k}$, *and*
3. $T_{\boldsymbol{N}} / \mathcal{C}_W^k \circ R_W^k = R_W^k \circ T_{\boldsymbol{N} \to \nu_W^k}$.

Proof. 1: Let $j \in [k]$. We have

$$\begin{aligned}
(T_{\boldsymbol{N}_j^k} \circ T_{\mathsf{owl}_W^k} f)(\bar{x}) = (T_{\boldsymbol{N}_j^k} f \circ \mathsf{owl}_W^k)(\bar{x}) &= \int_X f \circ \mathsf{owl}_W^k(\bar{x}[y/j]) \, d\mu(y) && \text{(def.)} \\
&= \int_{\mathbb{M}^k} f \, d(\mathsf{owl}_W^k \circ \bar{x}[\cdot/j])_* \mu \\
&= \int_{\mathbb{M}^k} f \, d\mu_j^{\mathsf{owl}_W^k(\bar{x})} && \text{(Lemma 29 1)} \\
&= (T_{\boldsymbol{N}_j^k \to \nu_W^k} f)(\mathsf{owl}_W^k(\bar{x})) && \text{(def. } T_{\boldsymbol{N}_j^k \to \nu_W^k}\text{)} \\
&= (T_{\mathsf{owl}_W^k} \circ T_{\boldsymbol{N}_j^k \to \nu_W^k} f)(\bar{x}) && \text{(def.)}
\end{aligned}$$

for $\mu^{\otimes k}$-almost every $\bar{x} \in X^k$ and every $f \in L^\infty(\mathbb{M}^k, \nu_W^k)$. This already proves the claim as $L^\infty(\mathbb{M}^k, \nu_W^k)$ is dense in $L^2(\mathbb{M}^k, \nu_W^k)$.

2: We have

$$\begin{aligned}
(T_{\boldsymbol{N}})_{\mathcal{C}_W^k} \circ T_{\mathsf{owl}_W^k} &= T_{\boldsymbol{N}} \circ \mathbb{E}(- \mid \mathcal{C}_W^k) \circ T_{\mathsf{owl}_W^k} && \text{(Lemma 8 5 and Lemma 22 5)} \\
&= T_{\boldsymbol{N}} \circ T_{\mathsf{owl}_W^k} && \text{(cf. Corollary 28)} \\
&= T_{\mathsf{owl}_W^k} \circ T_{\boldsymbol{N} \to \nu_W^k}. && (1)
\end{aligned}$$

3: We have

$$\begin{aligned}
T_{\boldsymbol{N}}/\mathcal{C}_W^k \circ R_W^k &= S_{\mathcal{C}_W^k} \circ T_{\boldsymbol{N}} \circ I_{\mathcal{C}_W^k} \circ S_{\mathcal{C}_W^k} \circ T_{\mathsf{owl}_W^k} && \text{(def.)} \\
&= S_{\mathcal{C}_W^k} \circ \mathbb{E}(- \mid \mathcal{C}_W^k) \circ T_{\boldsymbol{N}} \circ \mathbb{E}(- \mid \mathcal{C}_W^k) \circ T_{\mathsf{owl}_W^k} && \text{(Claim 6 4 and 6)} \\
&= S_{\mathcal{C}_W^k} \circ (T_{\boldsymbol{N}})_{\mathcal{C}_W^k} \circ T_{\mathsf{owl}_W^k} && \text{(def.)} \\
&= S_{\mathcal{C}_W^k} \circ T_{\mathsf{owl}_W^k} \circ T_{\boldsymbol{N} \to \nu_W^k} && (2) \\
&= R_W^k \circ T_{\boldsymbol{N} \to \nu_W^k}. && \text{(def.)}
\end{aligned}$$

□

Defining the operators for adjacency graphs is much simpler. Intuitively, every $\alpha \in \mathbb{M}^k$ contains the values $W(x_i, x_j)$ for every $ij \in \binom{[k]}{2}$ at position 0.

Lemma 32. *Let $k \geq 1$, and let $\nu \in \mathscr{P}(\mathbb{M}^k)$ be a k-WLD. Let $ij \in \binom{[k]}{2}$. Setting*

$$(T_{\boldsymbol{A}_{ij}^k \to \nu} f)(\alpha) \coloneqq (\alpha_0)_{ij} \cdot f(\alpha)$$

for all $f \in L^2(\mathbb{M}^k, \nu)$, $\alpha \in \mathbb{M}^k$ defines an L^∞- and L^2-contraction $L^2(\mathbb{M}^k, \nu) \to L^2(\mathbb{M}^k, \nu)$.

Proof. The mapping $\alpha \mapsto (\alpha_0)_{ij}$ is measurable by definition of the product σ-algebra Hence, $T_{\boldsymbol{A}_{ij}^k \to \nu} f$ for $f \in L^2(\mathbb{M}^k, \nu)$ is measurable as the product of measurable functions. Moreover, by definition of $\mathbb{M}^k$, the function $\alpha \mapsto (\alpha_0)_{ij}$ is bounded by 1, which immediately yields that $\|T_{\boldsymbol{A}_{ij}^k \to \nu} f\|_2 \leq \|f\|_2$ for $f \in L^2(\mathbb{M}^k, \nu)$ and $\|T_{\boldsymbol{A}_{ij}^k \to \nu} f\|_\infty \leq \|f\|_\infty$ for $f \in L^\infty(\mathbb{M}^k, \nu)$. Moreover, $T_{\boldsymbol{A}_{ij}^k \to \nu}$ is linear as a multiplicative operator. □

Analogously to Lemma 31, one can verify that Lemma 32 is in fact the right definition.

Lemma 33. *Let $k \geq 1$ and $W \colon X \times X \to [0, 1]$ be a graphon. For every $\boldsymbol{A} \in \mathcal{A}^k$,* *1.* $T_{\boldsymbol{A}\to W} \circ T_{\mathsf{owl}^k_W} = T_{\mathsf{owl}^k_W} \circ T_{\boldsymbol{A}\to\nu^k_W}$*,*

2. $(T_{\boldsymbol{A}\to W})_{\mathcal{C}^k_W} \circ T_{\mathsf{owl}^k_W} = T_{\mathsf{owl}^k_W} \circ T_{\boldsymbol{A}\to\nu^k_W}$*, and*

3. $T_{\boldsymbol{A}\to W}/\mathcal{C}^k_W \circ R^k_W = R^k_W \circ T_{\boldsymbol{A}\to\nu^k_W}$*.*

Proof. 1: Let $ij \in \binom{[k]}{2}$. We have

$$\begin{aligned}
(T_{\boldsymbol{A}^k_{ij}\to W} \circ T_{\mathsf{owl}^k_W} f)(\bar{x}) &= (T_{\boldsymbol{A}^k_{ij}\to W} f \circ \mathsf{owl}^k_W)(\bar{x}) \\
&= W(x_i, x_j) \cdot (f \circ \mathsf{owl}^k_W)(\bar{x}) && \text{(def.)} \\
&= ((\mathsf{owl}^k_W(\bar{x}))_0)_{ij} \cdot (f \circ \mathsf{owl}^k_W)(\bar{x}) && \text{(def. } \mathsf{owl}^k_W) \\
&= (T_{\boldsymbol{A}^k_{ij}\to\nu^k_W} f)(\mathsf{owl}^k_W(\bar{x})) && \text{(def. } T_{\boldsymbol{A}^k_{ij}\to\nu^k_W}) \\
&= (T_{\mathsf{owl}^k_W} \circ T_{\boldsymbol{A}^k_{ij}\to\nu^k_W} f)(\bar{x}) && \text{(def.)}
\end{aligned}$$

for $\mu^{\otimes k}$-almost every $\bar{x} \in X^k$ and every $f \in L^2(\mathbb{M}^k, \nu^k_W)$.

2 and 3: Analogous to the proof of 2 and 3 of Lemma 31, respectively. □

For a k-WLD $\nu \in \mathscr{P}(\mathbb{M}^k)$, define the family of L^∞-contractions $\mathbb{T}_\nu := (T_{\boldsymbol{F}\to\nu})_{\boldsymbol{F}\in\mathcal{F}^k}$. Lemma 31 3 and Lemma 33 3 can then be rephrased as the following corollary.

Corollary 34. *Let* $k \geq 1$ *and* $W \colon X \times X \to [0,1]$ *be a graphon. Then,* $\mathbb{T}^k_W/\mathcal{C}^k_W \circ R^k_W = R^k_W \circ \mathbb{T}_{\nu^k_W}$*.*

Recall Definition 17, i.e., the homomorphism density of a term in a family of L^∞-contractions. In particular, this definition applies to the family $\mathbb{T}_{\nu^k_W}$ of the k-WLD ν^k_W of a graphon W. Lemma 19 with the previous corollary yields that $\mathbb{T}_{\nu^k_W}$ and $\mathbb{T}^k_W/\mathcal{C}^k_W$ give us the same homomorphism densities (and also functions), which are just the original homomorphism densities in W.

Corollary 35. *Let* $k \geq 1$*. Let* $W \colon X \times X \to [0,1]$ *be a graphon and* $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ *be* W*-invariant. Then,* $t(\mathbb{F}, \mathbb{T}_{\nu^k_W}) = t(\llbracket\mathbb{F}\rrbracket, W)$ *for every* $\mathbb{F} \in \langle\mathcal{F}^k\rangle_{\circ,\cdot}$*.*

Proof. By Corollary 20, we have $t(\mathbb{F}, \mathbb{T}^k_W/\mathcal{C}^k_W) = t(\llbracket\mathbb{F}\rrbracket, W)$ since $\mathcal{C}^k_W$ is W-invariant by Lemma 22. By Corollary 34, we have $\mathbb{T}^k_W/\mathcal{C}^k_W \circ R^k_W = R^k_W \circ \mathbb{T}_{\nu^k_W}$, where R^k_W is a Markov isomorphism by Corollary 28. Then, Lemma 19 yields $t(\mathbb{F}, \mathbb{T}^k_W/\mathcal{C}^k_W) = t(\mathbb{F}, \mathbb{T}_{\nu^k_W})$. □

A permutation $\pi \colon [k] \to [k]$ extends to a measurable bijection $\pi \colon \mathbb{M}^k \to \mathbb{M}^k$ as follows: We obtain a measurable bijection $\pi \colon P^k_0 \to P^k_0$ by setting $\pi((y_{ij})_{ij}) := (y_{\pi(i)\pi(j)})_{ij}$ for $(y_{ij})_{ij} \in [0,1]^{\binom{[k]}{2}}$. From there on, π inductively extends to a measurable bijection $\pi \colon \mathbb{M}^k_n \to \mathbb{M}^k_n$ by component-wise application and, then, to a measurable bijection $\pi \colon P^k_{n+1} \to P^k_{n+1}$ by setting $\pi((\mu_j)_{j\in[k]}) = (\pi_*\mu_{\pi(j)})_{j\in[k]}$ for every $(\mu_j)_{j\in[k]} \in P^k_{n+1}$. Finally, we obtain the measurable bijection $\pi \colon \mathbb{M}^k_n \to \mathbb{M}^k_n$ by component-wise application.

Let $\nu \in \mathscr{P}(\mathbb{M}^k)$ be a k-WLD and $\pi \colon [k] \to [k]$ be a permutation. By definition of $\pi_*\nu$, the extension $\pi \colon \mathbb{M}^k \to \mathbb{M}^k$ is a measure-preserving map from $(\mathbb{M}^k, \mathcal{B}(\mathbb{M}^k), \nu)$ to $(\mathbb{M}^k, \mathcal{B}(\mathbb{M}^k), \pi_*\nu)$ by definition. The k-WLD ν is called π*-invariant* if $\pi_*\nu = \nu$, in which case we can view the Koopman operator of π as an operator $T_{\pi\to\nu} \colon L^2(\mathbb{M}^k, \nu) \to L^2(\mathbb{M}^k, \nu)$. The notation $T_{\pi\to\nu}$ avoids confusion with the Koopman operator of π when viewing it as a map $X^k \to X^k$, which we denote just by T_π. If we call a k-WLD $\nu \in \mathscr{P}(\mathbb{M}^k)$ *permutation-invariant* if it is π-invariant for every permutation $\pi \colon [k] \to [k]$, then Lemma 36 yields that the k-WLD ν^k_W of a graphon W is permutation invariant.

Lemma 36. *Let* $k \geq 1$ *and* $W \colon X \times X \to [0,1]$ *be a graphon. For every permutation* $\pi \colon [k] \to [k]$*,*

1. $\pi \circ \mathsf{owl}^k_W = \mathsf{owl}^k_W \circ \pi$,
2. ν^k_W *is* π*-invariant,*
3. $T_\pi \circ T_{\mathsf{owl}^k_W} = T_{\mathsf{owl}^k_W} \circ T_{\pi \to \nu^k_W}$,
4. $(T_\pi)_{\mathcal{C}^k_W} \circ T_{\mathsf{owl}^k_W} = T_{\mathsf{owl}^k_W} \circ T_{\pi \to \nu^k_W}$, *and*
5. $T_\pi / \mathcal{C}^k_W \circ S_{\mathcal{C}^k_W} \circ T_{\mathsf{owl}^k_W} = S_{\mathcal{C}^k_W} \circ T_{\mathsf{owl}^k_W} \circ T_{\pi \to \nu^k_W}$.

Proof. (1): We prove that $\pi \circ \mathsf{owl}^k_{W,n} = \mathsf{owl}^k_{W,n} \circ \pi$ by induction on $n \in \mathbb{N}$. This yields $(\pi \circ \mathsf{owl}^k_W(\bar{x}))_n = (\mathsf{owl}^k_W \circ \pi(\bar{x}))_n$ for every $\bar{x} \in X^k$ by induction on $n \in \mathbb{N}$, which then implies the claim. For the base case, we have

$$\pi(\mathsf{owl}^k_{W,0}(\bar{x})) = \Big((\mathsf{owl}^k_{W,0}(\bar{x}))_{\pi(i)\pi(j)}\Big)_{ij \in \binom{[k]}{2}} = \big(W(x_{\pi(i)}, x_{\pi(j)})\big)_{ij \in \binom{[k]}{2}} = \mathsf{owl}^k_{W,0}(\pi(\bar{x}))$$

for every $\bar{x} \in X^k$. For the inductive step, the induction hypothesis yields $(\pi(\mathsf{owl}^k_{W,n+1}(\bar{x})))_i = (\mathsf{owl}^k_{W,n+1}(\pi(\bar{x})))_i$ for every $\bar{x} \in X^k$ and every $i \leq n$. Moreover, we have

$$\begin{aligned}
(\pi(\mathsf{owl}^k_{W,n+1}(\bar{x})))_{n+1} &= \Big(\pi_*((\mathsf{owl}^k_{W,n+1}(\bar{x}))_{n+1})_{\pi(j)}\Big)_{j \in [k]} \\
&= \Big(\pi_*\Big(\Big(\mathsf{owl}^k_{W,n} \circ \bar{x}[\cdot/\pi(j)]\Big)_* \mu\Big)\Big)_{j \in [k]} \\
&= \Big(\Big(\pi \circ \mathsf{owl}^k_{W,n} \circ \bar{x}[\cdot/\pi(j)]\Big)_* \mu\Big)_{j \in [k]} \\
&= \Big(\Big(\pi \circ \mathsf{owl}^k_{W,n} \circ \pi^{-1} \circ \pi(\bar{x})[\cdot/j]\Big)\Big)_* \mu\Big)_{j \in [k]} \\
&= \Big(\Big(\mathsf{owl}^k_{W,n} \circ \pi(\bar{x})[\cdot/j]\Big)\Big)_* \mu\Big)_{j \in [k]} \qquad \text{(IH)} \\
&= (\mathsf{owl}^k_{W,n+1}(\pi(\bar{x})))_{n+1}
\end{aligned}$$

for every $\bar{x} \in X^k$.

(2): We have

$$\begin{aligned}
\pi_* \nu^k_W = \pi_*(\mathsf{owl}^k_{W*} \mu^{\otimes k}) = (\pi \circ \mathsf{owl}^k_W)_* \mu^{\otimes k} \overset{(1)}{=} (\mathsf{owl}^k_W \circ \pi)_* \mu^{\otimes k} = \mathsf{owl}^k_{W*}(\pi_* \mu^{\otimes k}) &= \mathsf{owl}^k_{W*} \mu^{\otimes k} \\
&= \nu^k_W.
\end{aligned}$$

(3): We have

$$T_\pi \circ T_{\mathsf{owl}^k_W} f = f \circ \mathsf{owl}^k_W \circ \pi \overset{(1)}{=} f \circ \pi \circ \mathsf{owl}^k_W = T_{\mathsf{owl}^k_W} \circ T_{\pi \to \nu^k_W} f$$

for every $f \in L^2(\mathbb{M}^k, \nu^k_W)$.

(4) and (5): Analogous to the proof of (2) and (3) of Lemma 31, respectively. ☐

4.5 Homomorphism Functions and Weisfeiler-Leman Measures

For the proof of Theorem 4, Corollary 35 allows us to get from k-WLDs to homomorphism densities, but getting to the other characterizations from there is arguably the most involved part of the proof. As Grebík and Rocha have shown [9], the key tool needed for this is the Stone-Weierstrass theorem: It yields that the set of homomorphism functions on $\mathbb{M}^k$, which is yet to be defined, is dense in the set $C(\mathbb{M}^k)$ of continuous functions on $\mathbb{M}^k$. Then, the Portmanteau theorem implies that equal homomorphism densities already imply equal k-WLDs.

To apply the Stone-Weierstrass theorem, we have to define the homomorphism function of a term on the set $\mathbb{M}^k$. Recall that an $\alpha \in \mathbb{M}^k$ is a sequence $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots)$ that, intuitively, corresponds to a sequence of unfoldings of heights $0, 1, 2, \dots$ of a graphon. However, as the components $\alpha_0, \alpha_1, \alpha_2$ do not have to be consistent, cf. the definition of $\mathbb{P}^k$, using different components may lead to different functions. Hence, we define a whole set of functions for a single term by considering all ways in which we may use the components to define a homomorphism function. We could avoid this by defining homomorphism functions just on $\mathbb{P}^k$ instead of $\mathbb{M}^k$; this, however, would complicate things further down the road, which is why we just accept this small inconvenience. Note the similarity between the following definition and the operators defined in the previous section.

Definition 37. *Let $k \geq 1$. For every term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ and every $n \in \mathbb{N}$ with $n \geq h(\mathbb{F})$, we inductively define the set $F_n^{\mathbb{F}}$ of functions $\mathbb{M}_n^k \to [0,1]$ as the smallest set such that*

1. $\mathbf{1}_{\mathbb{M}_n^k} \in F_n^{\mathbf{1}^k}$,
2. $\alpha \mapsto (\alpha_0)_{ij} \cdot f(\alpha) \in F_n^{\boldsymbol{A}_{ij}^k \circ \mathbb{F}}$ *for every* $f \in F_n^{\mathbb{F}}$,
3. $\alpha \mapsto \int_{\mathbb{M}_n^k} f \, d(\alpha_{n+1})_j \in F_{n+1}^{\boldsymbol{N}_j^k \circ \mathbb{F}}$ *for every* $f \in F_n^{\mathbb{F}}$ *and every* $j \in [k]$,
4. $f_1 \cdot f_2 \in F_n^{\mathbb{F}_1 \cdot \mathbb{F}_2}$ *for all* $f_1 \in F_n^{\mathbb{F}_1}, f_2 \in F_n^{\mathbb{F}_2}$, *and*
5. $f \circ p_{n,m} \in F_n^{\mathbb{F}}$ *for every* $f \in F_m^{\mathbb{F}}$ *and every* $m \in \mathbb{N}$ *with* $n > m \geq h(\mathbb{F})$.

Moreover, for every term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$, define the set $F^{\mathbb{F}}$ of functions $\mathbb{M}^k \to [0,1]$ by

$$F^{\mathbb{F}} := F_\infty^{\mathbb{F}} := \bigcup_{h(\mathbb{F}) \leq n < \infty} F_n^{\mathbb{F}} \circ p_{\infty,n}.$$

With a simple induction, one can verify that for every term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ and every $n \in \mathbb{N} \cup \{\infty\}$ with $n \geq h(\mathbb{F})$, the set $F_n^{\mathbb{F}}$ is non-empty and all functions in it are well-defined and continuous. Recall that, for a term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ and a k-WLD $\nu \in \mathscr{P}(\mathbb{M}^k)$, the operators $\mathbb{T}_\nu$ already define the homomorphism function $f_{\mathbb{F} \to \mathbb{T}_\nu} \in L^\infty(\mathbb{M}^k, \nu)$ by Definition 17. Note that the k-WLD ν satisfying $\nu(\mathbb{P}^k) = 1$ is the reason why we only have this single function $f_{\mathbb{F} \to \mathbb{T}_\nu}$. Then, it should come at no surprise that this single function is equal to all of the previous defined functions ν-almost everywhere.

Lemma 38. *Let $k \geq 1$ and $\nu \in \mathscr{P}(\mathbb{M}^k)$ be a k-WLD. Let $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ be a term and $n \in \mathbb{N}$ with $n \geq h(\mathbb{F})$. Then, every function in $F_n^{\mathbb{F}} \circ p_{\infty,n}$ is equal to $f_{\mathbb{F} \to \mathbb{T}_\nu}$ ν-almost everywhere.*

Proof. We prove the statement by induction on $\mathbb{F}$ and n. For the base case, we have $\mathbf{1}_{\mathbb{M}_n^k} \circ p_{\infty,n} = \mathbf{1}_{\mathbb{M}^k} = f_{\mathbf{1}^k \to \mathbb{T}_\nu}$ ν-almost everywhere. For the inductive step, first consider $\alpha \mapsto (\alpha_0)_{ij} \cdot f(\alpha) \in F_n^{\boldsymbol{A}_{ij}^k \circ \mathbb{F}}$ for an $f \in F_n^{\mathbb{F}}$, where we have

$$(\alpha_0)_{ij} \cdot f(p_{\infty,n}(\alpha)) = (T_{\boldsymbol{A}_{ij}^k \to \nu} f \circ p_{\infty,n})(\alpha) \overset{\text{IH}}{=} (T_{\boldsymbol{A}_{ij}^k \to \nu} f_{\mathbb{F} \to \mathbb{T}_\nu})(\alpha) = f_{\boldsymbol{A}_{ij}^k \circ \mathbb{F} \to \nu}(\alpha)$$

for ν-almost every $\alpha \in \mathbb{M}^k$. Next, consider $\alpha \mapsto \int_{\mathbb{M}_n^k} f \, d(\alpha_{n+1})_j \in F_{n+1}^{\boldsymbol{N}_j^k \circ \mathbb{F}}$ for an $f \in F_n^{\mathbb{F}}$ and a

$j \in [k]$. Since ν is a k-WLD, we have $\nu(\mathbb{P}^k) = 1$, which yields that

$$\begin{aligned}\int_{\mathbb{M}^k_n} f \, d(\alpha_{n+1})_j = \int_{\mathbb{M}^k_n} f \, d(p_{\infty,n})_* \mu^\alpha_j = \int_{\mathbb{M}^k} f \circ p_{\infty,n} \, d\mu^\alpha_j &= (T_{\boldsymbol{N}^k_j \to \nu} f \circ p_{\infty,n})(\alpha) \\ &\overset{\text{IH}}{=} (T_{\boldsymbol{N}^k_j \to \nu} f_{\mathbb{F} \to \nu})(\alpha) \\ &= f_{\boldsymbol{N}^k_j \circ \mathbb{F} \to \nu}(\alpha)\end{aligned}$$

for ν-almost every $\alpha \in \mathbb{M}^k$. For the product $f_1 \cdot f_2 \in F^{\mathbb{F}_1 \cdot \mathbb{F}_2}_n$ of $f_1 \in F^{\mathbb{F}_1}_n, f_2 \in F^{\mathbb{F}_2}_n$, we have

$$(f_1 \cdot f_2) \circ p_{\infty,n} = (f_1 \circ p_{\infty,n}) \cdot (f_2 \circ p_{\infty,n}) \overset{\text{IH}}{=} f_{\mathbb{F}_1 \to \mathbb{T}_\nu} \cdot f_{\mathbb{F}_2 \to \mathbb{T}_\nu} = f_{\mathbb{F}_1 \cdot \mathbb{F}_2 \to \mathbb{T}_\nu}$$

ν-almost everywhere. Finally, consider $f \circ p_{n,m} \in F^{\mathbb{F}}_n$ for $f \in F^{\mathbb{F}}_m$ and $m \in \mathbb{N}$ with $n > m \geq h(\mathbb{F})$. Then, $f \circ p_{n,m} \circ p_{\infty,n} = f \circ p_{\infty,m} = f_{\mathbb{F} \to \mathbb{T}_\nu}$ holds ν-almost everywhere by the inductive hypothesis. ☐

Corollary 35 yields the following corollary to the previous lemma.

Corollary 39. *Let $k \geq 1$ and $W \colon X \times X \to [0,1]$ be a graphon. For every term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ and every function $f \in F^{\mathbb{F}}$, we have*

$$t([\![\mathbb{F}]\!], W) = \int_{\mathbb{M}^k} f \nu^k_W.$$

For every $n \in \mathbb{N} \cup \{\infty\}$, define $\mathcal{T}^k_n := \bigcup_{\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}, h(\mathbb{F}) \leq n} F^{\mathbb{F}}_n$ and abbreviate $\mathcal{T}^k := \mathcal{T}^k_\infty$. By induction, we can use the Stone-Weierstrass theorem and the Portmanteau theorem to show that the Stone-Weierstrass is actually applicable to all of these sets and, in particular, to $\mathcal{T}^k$, cf. [9, Proposition 7.5].

Lemma 40. *Let $k \geq 1$. For every $n \in \mathbb{N} \cup \{\infty\}$, the set $\mathcal{T}^k_n$ is closed under multiplication, contains $\mathbf{1}_{\mathbb{M}^k_n}$, and separates points of $\mathbb{M}^k_n$.*

Proof. First, consider the case that $n \in \mathbb{N}$. We trivially have $\mathbf{1}_{\mathbb{M}^k_n} \in F^{\mathbf{1}^k}_n \subseteq \mathcal{T}^k_n$ by definition. Moreover, for $f_1, f_2 \in \mathcal{T}^k_n$, there are terms $\mathbb{F}_1, \mathbb{F}_2 \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ with $h(\mathbb{F}_1) \leq n$ and $h(\mathbb{F}_2) \leq n$. such that $f_1 \in F^{\mathbb{F}_1}_n$ and $f_2 \in F^{\mathbb{F}_2}_n$. Then, $f_1 \cdot f_2 \in F^{\mathbb{F}_1 \cdot \mathbb{F}_2}_n \subseteq \mathcal{T}^k_n$ as $h(\mathbb{F}_1 \cdot \mathbb{F}_2) = \max\{h(\mathbb{F}_1), h(\mathbb{F}_2)\} \leq n$. We prove that $\mathcal{T}^k_n$ separates points of $\mathbb{M}^k_n$ by induction on n. For the base case $n = 0$, let $\beta \neq \gamma \in \mathbb{M}^k_0$. Then, there is an $ij \in \binom{[k]}{2}$ such that $\beta_{ij} \neq \gamma_{ij}$, and the function $\alpha \mapsto (\alpha_0)_{ij} \in F^{\boldsymbol{A}^k_{ij} \circ \mathbf{1}^k}_0$ separates β and γ.

For the inductive step, assume that $\mathcal{T}^k_n$ separates points of $\mathbb{M}^k_n$. Let $\beta \neq \gamma \in \mathbb{M}^k_{n+1}$. If there is an $0 \leq m \leq n$ such that $\beta_m \neq \gamma_m$, then $p_{n+1,n}(\beta) \neq p_{n+1,n}(\gamma) \in \mathbb{M}^k_n$. Hence, by the inductive hypothesis, there is an $f \in \mathcal{T}^k_n$ such that $f(p_{n+1,n}(\beta)) \neq f(p_{n+1,n}(\gamma))$. By definition, $f \in F^{\mathbb{F}}_n$ for some term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ with $h(\mathbb{F}) \leq n$. Therefore, $f \circ p_{n+1,n} \in F^{\mathbb{F}}_{n+1} \subseteq \mathcal{T}^k_{n+1}$ is a function that separates β and γ.

For the remaining case, assume that $\beta_{n+1} \neq \gamma_{n+1}$. Then, there is an $ij \in \binom{[k]}{2}$ such that $(\beta_{n+1})_{ij} \neq (\gamma_{n+1})_{ij}$. By the inductive hypothesis and the Stone-Weierstrass theorem [6, Theorem 2.4.11], the linear hull of $\mathcal{T}^k_n$ is uniformly dense in $C(\mathbb{M}^k_n)$. Since $\mathbb{M}^k_n$ is Hausdorff, it then follows from the Portmanteau theorem [15, Theorem 17.20] that there is an $f \in \mathcal{T}^k_n$ such that $\int_{\mathbb{M}^k_n} f \, d(\beta_{n+1})_{ij} \neq \int_{\mathbb{M}^k_n} f \, d(\gamma_{n+1})_{ij}$. By definition, $f \in F^{\mathbb{F}}_n$ for some term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ with $h(\mathbb{F}) \leq n$. Then, $\alpha \mapsto \int_{\mathbb{M}^k_n} f \, d(\alpha_{n+1})_j \in F^{\boldsymbol{N}^k_j \circ \mathbb{F}}_{n+1} \subseteq \mathcal{T}^k_{n+1}$ is a function that separates β and γ.

Having proven the statement for every $n \in \mathbb{N}$, one can also easily see that it holds in the case $n = \infty$ from the definitions, cf. also the first case of the induction. ☐

A final application of the Stone-Weierstrass theorem and the Portmanteau theorem yields that, for a sequence of graphons, convergence of their k-WLDs is equivalent to convergence of treewidth $k-1$ multigraph homomorphism densities.

Lemma 41. *Let $k \geq 1$. Let $(W_n)_n$ and $W\colon X \times X \to [0,1]$ be a sequence of graphons and a graphon, respectively. Then, $\nu^k_{W_n} \to \nu^k_W$ if and only if $t(F, W_n) \to t(F, W)$ for every multigraph F of treewidth at most $k-1$.*

Proof. Note that the linear hull of $\mathcal{T}^k$ is uniformly dense in $C(\mathbb{M}^k)$ by Lemma 40 and the Stone-Weierstrass theorem [6, Theorem 2.4.11]. Hence, we have

$$\begin{aligned}
\nu^k_{W_n} \to \nu^k_W &\iff \int_{\mathbb{M}^k} f\, d\nu^k_{W_n} \to \int_{\mathbb{M}^k} f\, d\nu^k_W \text{ for every } f \in C(\mathbb{M}^k) && \text{(Portmanteau theorem)}\\
&\iff \int_{\mathbb{M}^k} f\, d\nu^k_{W_n} \to \int_{\mathbb{M}^k} f\, d\nu^k_W \text{ for every } f \text{ in the linear hull of } \mathcal{T}^k\\
&\iff \int_{\mathbb{M}^k} f\, d\nu^k_{W_n} \to \int_{\mathbb{M}^k} f\, d\nu^k_W \text{ for every } f \in \mathcal{T}^k && \text{(Linearity of the integral)}\\
&\iff t([\![\mathbb{F}]\!], W_n) \to t([\![\mathbb{F}]\!], W) \text{ for every } \mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot} && \text{(Corollary 39)}\\
&\iff t(F, W_n) \to t(F, W) \text{ for every multigraph } F \text{ of tw.} \leq k-1. && \text{(Lemma 13)}
\end{aligned}$$

□

Since $\mathbb{M}^k$ is Hausdorff, this in particular means that the k-WLDs of two graphons are equal if and only if their homomorphism densities are.

Corollary 42. *Let $k \geq 1$ and $U, W\colon X \times X \to [0,1]$ be graphons, Then, $\nu^k_U = \nu^k_W$ if and only if $t(F, U) = t(F, W)$ for every multigraph F of treewidth at most $k-1$.*

We are finally ready to prove Theorem 4. The majority of the work is already done and, at this point, it is just about putting all the previous results together.

(Proof of Theorem 4). 1 $\implies$ 2: This is just Corollary 42.

2 $\implies$ 3: Let $R \coloneqq R^k_U \circ (R^k_W)^*$. By the assumption, R is well defined, and by Corollary 28, it is a Markov isomorphism as the composition of two Markov isomorphisms. By Corollary 34, we have

$$\begin{aligned}
\mathbb{T}^k_U/\mathcal{C}^k_U \circ R = \mathbb{T}^k_U/\mathcal{C}^k_U \circ R^k_U \circ (R^k_W)^* = R^k_U \circ \mathbb{T}^k_{\nu^k_U} \circ (R^k_W)^* &= R^k_U \circ \mathbb{T}^k_{\nu^k_W} \circ (R^k_W)^*\\
&= R^k_U \circ (R^k_W)^* \circ \mathbb{T}^k_W/\mathcal{C}^k_W\\
&= R \circ \mathbb{T}^k_W/\mathcal{C}^k_W.
\end{aligned}$$

Similarly, Lemma 36 yields that R is permutation invariant.

3 $\implies$ 4: Set $S \coloneqq I_{\mathcal{C}^k_U} \circ R \circ S_{\mathcal{C}^k_W}$, which is a Markov operator as the composition of Markov operators. By Lemma 22 5, $\mathcal{C}^k_U$ and $\mathcal{C}^k_W$ are $\mathbb{T}^k_U$- and $\mathbb{T}^k_W$-invariant, respectively. Hence,

$$\begin{aligned}
\mathbb{T}^k_U \circ S = \mathbb{T}^k_U \circ I_{\mathcal{C}^k_U} \circ R \circ S_{\mathcal{C}^k_W} = I_{\mathcal{C}^k_U} \circ \mathbb{T}^k_U/\mathcal{C}^k_U \circ R \circ S_{\mathcal{C}^k_W} &= I_{\mathcal{C}^k_U} \circ R \circ \mathbb{T}^k_W/\mathcal{C}^k_W \circ S_{\mathcal{C}^k_W}\\
&= I_{\mathcal{C}^k_U} \circ R \circ S_{\mathcal{C}^k_W} \circ \mathbb{T}^k_W\\
&= S \circ \mathbb{T}^k_W.
\end{aligned}$$

by Lemma 8 5, 6. In a similar fashion, Lemma 22 6 implies that, if R is permutation invariant, then so is S.

4 $\implies$ 5: Follows immediately from Lemma 9.

5 $\implies$ 1: We have $t([\![\mathbb{F}]\!], U) = t(\mathbb{F}, \mathbb{T}^k_U/\mathcal{C}) = t(\mathbb{F}, \mathbb{T}^k_W/\mathcal{D}) = t([\![\mathbb{F}]\!], W)$, for every $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ by Corollary 20 and Lemma 19. Then, Lemma 13 yields the claim. □

4.6 Measure Hierarchies

Theorem 4 implies that the sequence $\nu_W^1, \nu_W^2, \dots$ of k-WLDs of a graphon W characterizes W up to weak isomorphism since every graph has some finite treewidth. Let us explore this a bit more in depth by combining all these k-WLDs into a single measure.

First, for $\infty > k \geq \ell \geq 1$, let $p^{k,\ell}$ denote the projection from $\mathbb{M}^k$ to $\mathbb{M}^\ell$ defined as follows: Inductively, define $p^{k,\ell}\colon P_n^k \to P_n^\ell$, which also directly extends to $p^{k,\ell}\colon \mathbb{M}_n^k \to \mathbb{M}_n^\ell$ by applying the function component-wise. For $n = 0$, let $p^{k,\ell}\colon P_0^k \to P_0^\ell$ be defined by $p^{k,\ell}((w_{ij})_{ij \in \binom{[k]}{2}}) \coloneqq (w_{ij})_{ij \in \binom{[\ell]}{2}}$. For the inductive step, $p^{k,\ell}\colon P_{n+1}^k \to P_{n+1}^\ell$ is defined by $p^{k,\ell}((\nu_j)_{j \in [k]}) \coloneqq (p^{k,\ell}{}_* \nu_j)_{j \in [\ell]}$. It is not hard to see that this is well-defined as every $p^{k,\ell}$ is continuous. Finally, again by applying the function component-wise, $p^{k,\ell}$ extends to a continuous function $p^{k,\ell}\colon \mathbb{M}^k \to \mathbb{M}^\ell$. Then, consider the *inverse limit* of the spaces $\mathbb{M}^k$ and the projections $p^{k+1,k}$ for $k \geq 1$ defined by

$$\mathbb{M}^\infty \coloneqq \Big\{ (\alpha^k)_{k \geq 1} \in \prod_{k \geq 1} \mathbb{M}^k \;\Big|\; p^{k+1,k}(\alpha^{k+1}) = \alpha^k \text{ for every } k \geq 1 \Big\}$$

with the σ-algebra $\mathcal{B}(\mathbb{M}^\infty)$ generated by the projections $p^{\infty,k}\colon \mathbb{M}^\infty \to \mathbb{M}^k, \alpha \mapsto \alpha^k$ for every $k \geq 1$. Note that this notation is justified as $\mathbb{M}^\infty$ is again a standard Borel space [15, Exercise 17.16]. As a product of a sequence of metrizable compact spaces, $\prod_{k \geq 1} \mathbb{M}^k$ is metrizable [6, Proposition 2.4.4] and also compact by Tychonoff's Theorem [6, Theorem 2.2.8]. Since $p^{k+1,k}$ is continuous, this implies that $\mathbb{M}^\infty$ is closed and, hence, a metrizable compact space. Let

$$\mathbb{WL} \coloneqq \Big\{ (\nu^k)_{k \geq 1} \in \prod_{k \geq 1} \mathbb{WL}^k \;\Big|\; \nu^k = p^{k+1,k}{}_* \nu^{k+1} \text{ for every } k \geq 1 \Big\},$$

where $\mathbb{WL}^k$ denotes the set of all k-WLDs. Then, by the Kolmogorov Consistency Theorem [15, Exercise 17.16], for every $\nu \in \mathbb{WL}$, there is a unique $\nu^\infty \in \mathscr{P}(\mathbb{M}^\infty)$ such that $p^{\infty,k}{}_* \nu^\infty = \nu^k$ for every $k \geq 1$.

Lemma 43. *Let $(\nu_n)_n$ be a sequence with $\nu_n \in \mathbb{WL}$ and $\nu \in \mathbb{WL}$. Then, $\nu_n^\infty \to \nu^\infty$ if and only if $\nu_n^k \to \nu^k$ for every $k \geq 1$.*

Proof. The set $\bigcup_{1 \leq k < \infty} C(\mathbb{M}^k) \circ p^{\infty,k}$ is uniformly dense in $C(\mathbb{M}^\infty)$ by the Stone-Weierstrass theorem [6, Theorem 2.4.11], cf. also the proof of Lemma 24. Hence, we have

$$\begin{aligned}
\nu_n^\infty \to \nu^\infty \iff& \int_{\mathbb{M}^\infty} f \, d\nu_n^\infty \to \int_{\mathbb{M}^\infty} f \, d\nu^\infty \text{ for every } f \in C(\mathbb{M}^\infty) && \text{(Portmanteau theorem)}\\
\iff& \int_{\mathbb{M}^\infty} f \circ p^{\infty,k} \, d\nu_n^\infty \to \int_{\mathbb{M}^\infty} f \circ p^{\infty,k} \, d\nu^\infty \text{ for all } k \geq 1,\, f \in C(\mathbb{M}^k)\\
\iff& \int_{\mathbb{M}^\infty} f \, dp^{\infty,k}{}_* \nu_n^\infty \to \int_{\mathbb{M}^\infty} f \, dp^{\infty,k}{}_* \nu^\infty \text{ for all } k \geq 1,\, f \in C(\mathbb{M}^k)\\
\iff& \int_{\mathbb{M}^\infty} f \, d\nu_n^k \to \int_{\mathbb{M}^\infty} f \, d\nu^k \text{ for all } k \geq 1,\, f \in C(\mathbb{M}^k)\\
\iff& \nu_n^k \to \nu^k \text{ for every } k \geq 1. && \text{(Portmanteau theorem)}
\end{aligned}$$

□

One can show that, for every graphon $W\colon X \times X \to [0,1]$, the sequence $(\nu_W^k)_{k \geq 1}$ of its k-WLDs is in $\mathbb{WL}$ and, hence, yields a measure $\nu_W^\infty \in \mathscr{P}(\mathbb{M}^\infty)$. Together, Lemma 41 and Lemma 43 imply that these measures induce the same topology on the space of graphons as multigraph homomorphism densities; note that this topology is different from the one induced by simple graph homomorphism densities, cf. [16, Exercise 10.26] or [14, Lemma C.2].

Corollary 44. *Let* $(W_n)_n$ *and* $W\colon X \times X \to [0,1]$ *be a sequence of graphons and a graphon, respectively. Then, the following are equivalent:*

1. $\nu^\infty_{W_n} \to \nu^\infty_W$.
2. $t(F, W_n) \to t(F, W)$ *for every multigraph* F.

While simple graph and multigraph homomorphism densities yield different topologies, two graphons are nevertheless weakly isomorphic if and only if they have the same multigraph homomorphism densities [16, Corollary 10.36]. Since $\mathscr{M}_{\leq 1}(\mathbb{M}^\infty)$ is Hausdorff, this yields the following corollary.

Corollary 45. *Let* $U, W\colon X \times X \to [0,1]$ *be graphons. Then,* $\nu^\infty_U = \nu^\infty_W$ *if and only if* U *and* W *are weakly isomorphic.*

4.7 Operator Hierarchies

Recall the system $\mathsf{L}^k_{\mathrm{iso}}$ of linear equations from the introduction: two simple graphs G and H are not distinguished by oblivious k-WL if and only if $\mathsf{L}^k_{\mathrm{iso}}(G, H)$ has a non-negative real solution. Let us take a closer look at $\mathsf{L}^k_{\mathrm{iso}}(G, H)$ to see that it is much closer related to the characterization $\mathbb{T}^k_U \circ S = S \circ \mathbb{T}^k_W$ from Theorem 4 than it might seem at first glance. The variables of $\mathsf{L}^k_{\mathrm{iso}}(G, H)$, which are indexed by sets $\pi \subseteq V(G) \times V(H)$ of size $|\pi| \leq k$, can be interpreted as permutation-invariant matrices on $V(G)^1 \times V(H)^1, \dots, V(G)^k \times V(H)^k$. Instead of permutation-invariant operators on all spaces $L^2(X^1, \mu^{\otimes 1}), \dots, L^2(X^k, \mu^{\otimes k})$, we only have a single permutation-invariant Markov operator S on $L^2(X^k, \mu^{\otimes k})$. For an operator S on $L^2(X^k, \mu^{\otimes k})$, defining

$$S{\downarrow} := T_{\boldsymbol{F}^k_k} \circ S \circ T_{\boldsymbol{I}^k_k}$$

yields an operator on $L^2(X^{k-1}, \mu^{\otimes k-1})$. It is easy to see that $(S{\downarrow})^* = S^*{\downarrow}$ since the adjoint of a forget graph is the corresponding introduce graph and vice versa. Moreover, as long as S is permutation-invariant, this definition is independent of the specific pair of forget and introduce graphs, i.e., we have $S{\downarrow} = T_{\boldsymbol{F}^k_j} \circ S \circ T_{\boldsymbol{I}^k_j}$ for every $j \in [k]$ since $T_{\boldsymbol{F}^k_k} \circ T_{(k\dots j)} = T_{\boldsymbol{F}^k_j}$ and $T_{(j\dots k)} \circ T_{\boldsymbol{I}^k_k} = T_{\boldsymbol{I}^k_j}$.

Lemma 46. *Let* $k \geq 1$ *and* S *be a permutation-invariant Markov operator on* $L^2(X^k, \mu^{\otimes k})$. *Then,* $S{\downarrow}$ *is a permutation-invariant Markov operator. Moreover, if* $T_{\boldsymbol{N}^k_k} \circ S = S \circ T_{\boldsymbol{N}^k_k}$, *then*

1. $S \circ T_{\boldsymbol{I}^k_k} = T_{\boldsymbol{I}^k_k} \circ S{\downarrow}$,
2. $T_{\boldsymbol{F}^k_k} \circ S = S{\downarrow} \circ T_{\boldsymbol{F}^k_k}$, *and*
3. $T_{\boldsymbol{N}^{k-1}_{k-1}} \circ S{\downarrow} = S{\downarrow} \circ T_{\boldsymbol{N}^{k-1}_{k-1}}$.

Proof. First note that

$$S{\downarrow}\mathbf{1}_{X^{k-1}} = (T_{\boldsymbol{F}^k_k} \circ S \circ T_{\boldsymbol{I}^k_k})\mathbf{1}_{X^{k-1}} = (T_{\boldsymbol{F}^k_k} \circ S)\mathbf{1}_{X^k} = T_{\boldsymbol{F}^k_k}\mathbf{1}_{X^k} = \mathbf{1}_{X^{k-1}},$$

where the last equality holds since μ is a probability measure. Since S^* is also a Markov operator, we also obtain $(S{\downarrow})^*\mathbf{1}_{X^{k-1}} = S^*{\downarrow}\mathbf{1}_{X^{k-1}} = \mathbf{1}_{X^{k-1}}$. Let $f \in L^2(X^{k-1}, \mu^{\otimes k-1})$ with $f \geq 0$. Then, $T_{\boldsymbol{I}^k_k} f = f \otimes \mathbf{1}_X \geq 0$, and hence, $(S \circ T_{\boldsymbol{I}^k_k})f \geq 0$. Therefore, also $S{\downarrow} f = (T_{\boldsymbol{F}^k_k} \circ S \circ T_{\boldsymbol{I}^k_k})f \geq 0$. Hence, $S{\downarrow}$ is a Markov operator. For a permutation $\pi\colon [k-1] \to [k-1]$, we define the permutation $\pi'\colon [k] \to [k]$ by $\pi'(i) := \pi(i)$ for $i \in [k-1]$ and $\pi'(k) := k$. Then,

$$\begin{aligned}
T_\pi \circ S{\downarrow} = T_\pi \circ T_{\boldsymbol{F}^k_k} \circ S \circ T_{\boldsymbol{I}^k_k} = T_{\boldsymbol{F}^k_k} \circ T_{\pi'} \circ S \circ T_{\boldsymbol{I}^k_k} &= T_{\boldsymbol{F}^k_k} \circ S \circ T_{\pi'} \circ T_{\boldsymbol{I}^k_k} \\
&= T_{\boldsymbol{F}^k_k} \circ S \circ T_{\boldsymbol{I}^k_k} \circ T_\pi \\
&= S{\downarrow} \circ T_\pi.
\end{aligned}$$

Hence, $S{\downarrow}$ is permutation invariant. Now, assume that $T_{\boldsymbol{N}_k^k} \circ S = S \circ T_{\boldsymbol{N}_k^k}$. Then,

$$\begin{aligned} T_{\boldsymbol{I}_k^k} \circ S{\downarrow} = T_{\boldsymbol{I}_k^k} \circ T_{\boldsymbol{F}_k^k} \circ S \circ T_{\boldsymbol{I}_k^k} = T_{\boldsymbol{N}_k^k} \circ S \circ T_{\boldsymbol{I}_k^k} &= S \circ T_{\boldsymbol{N}_k^k} \circ T_{\boldsymbol{I}_k^k} \\ &= S \circ T_{\boldsymbol{I}_k^k} \circ T_{\boldsymbol{F}_k^k} \circ T_{\boldsymbol{I}_k^k} \\ &= S \circ T_{\boldsymbol{I}_k^k}, \end{aligned}$$

where the last equality holds since μ is a probability measure. Then, we also obtain 2 by considering S^* and $S^*{\downarrow}$ and then taking adjoints. Finally, note that the permutation invariance of S yields that we also have $T_{\boldsymbol{N}_{k-1}^k} \circ S = S \circ T_{\boldsymbol{N}_{k-1}^k}$. Moreover, observe that $\boldsymbol{N}_{k-1}^{k-1} \circ \boldsymbol{F}_k^k = \boldsymbol{F}_k^k \circ \boldsymbol{N}_{k-1}^k$. Hence,

$$\begin{aligned} T_{\boldsymbol{N}_{k-1}^{k-1}} \circ S{\downarrow} = T_{\boldsymbol{N}_{k-1}^{k-1}} \circ T_{\boldsymbol{F}_k^k} \circ S \circ T_{\boldsymbol{I}_k^k} = T_{\boldsymbol{F}_k^k} \circ T_{\boldsymbol{N}_{k-1}^k} \circ S \circ T_{\boldsymbol{I}_k^k} &= T_{\boldsymbol{F}_k^k} \circ S \circ T_{\boldsymbol{N}_{k-1}^k} \circ T_{\boldsymbol{I}_k^k} \\ &= T_{\boldsymbol{F}_k^k} \circ S \circ T_{\boldsymbol{I}_k^k} \circ T_{\boldsymbol{N}_{k-1}^{k-1}} \\ &= S{\downarrow} \circ T_{\boldsymbol{N}_{k-1}^{k-1}}. \end{aligned}$$

□

Given a permutation-invariant Markov operator S on $L^2(X^k, \mu^{\otimes k})$, repeated applications of Lemma 46 yield a sequence $S_0, \dots, S_k$ of permutation-invariant Markov operators S_i on $L^2(X^i, \mu^{\otimes i})$ by letting $S_k \coloneqq S$ and $S_{i-1} \coloneqq S_i{\downarrow}$ for $i \in [k]$, which we call the *operator hierarchy defined by S*. If S satisfies $T_{\boldsymbol{N}_k^k} \circ S = S \circ T_{\boldsymbol{N}_k^k}$, then Lemma 46 yields that

1. $S_i(f \otimes \mathbf{1}_X) = S_{i-1}(f) \otimes \mathbf{1}_X$ for every $f \in L^2(X^{i-1}, \mu^{\otimes i-1})$ and every $i \in [k]$,
2. $S_i^*(f \otimes \mathbf{1}_X) = S_{i-1}^*(f) \otimes \mathbf{1}_X$ for every $f \in L^2(X^{i-1}, \mu^{\otimes i-1})$ and every $i \in [k]$,
3. S_0 is the identity operator, and
4. $S_i \geq 0$ for every $i \in [k]$.

Note that, by definition of $\boldsymbol{I}_i^i$, the first condition just states that $S_i \circ T_{\boldsymbol{I}_i^i} = T_{\boldsymbol{I}_i^i} \circ S_{i-1}$; the second condition is the analogous statement for forget graphs. With this observation, one also gets that the converse holds, i.e., if $S_0, \dots, S_k$ is a sequence of permutation-invariant operators S_i on $L^2(X^i, \mu^{\otimes i})$, then $S_0, \dots, S_k$ are Markov operators satisfying $T_{\boldsymbol{N}_i^i} \circ S_i = S_i \circ T_{\boldsymbol{N}_i^i}$.

As a final remark, note that in addition to Lemma 46, one can also easily prove that, if $T_{\boldsymbol{A}_{12}^k \to U} \circ S = S \circ T_{\boldsymbol{A}_{12}^k \to W}$ holds for graphons $U, W \colon X \times X \to [0,1]$ and $k \geq 3$, then we also have $T_{\boldsymbol{A}_{12}^{k-1} \to U} \circ S{\downarrow} = S{\downarrow} \circ T_{\boldsymbol{A}_{12}^{k-1} \to W}$. This inductively extends to operator hierarchies, and it is not hard to see that this requirement corresponds to the equations for partial isomorphisms in $\mathsf{L}_{\text{iso}}^k$; we are just missing injectivity, which is not important as long as our standard Borel space is atom-free.

5 Simple Weisfeiler-Leman

Theorem 4 show that oblivious k-WL corresponds to bounded treewidth *multi*graph homomorphism densities. The reason for this are the atomic types used by k-WL, or more accurately in our setting, the adjacency graphs since subsequent applications of the same adjacency graph $\boldsymbol{A}_{ij}^k$ to a term result in parallel edges. This cannot be prevented by simply disallowing such subsequent applications: for the application of the Stone-Weierstrass theorem in the proof of Theorem 4, it is crucial that the set $\mathcal{T}^k$ of homomorphism functions is closed under multiplications. To

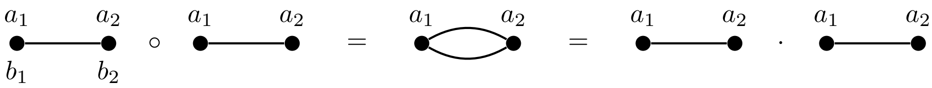

Figure 7: Two ways of introducing parallel edges.

Figure 8: The graphs $\boldsymbol{S}^2_{2,\{1\}}$ and $\boldsymbol{S}^3_{2,\{1,3\}}$.

achieve this, we have to close the set of terms under Schur product, which may also introduce parallel edges if we have edges between input vertices, cf. Figure 7. To prevent this we have to prevent edges from being added between input vertices in the first place. In the following, we show how Theorem 4 and its proof has to be adapted for simple graph homomorphism densities. In particular, what we refer to as *simple (oblivious) k-WL* is introduced. Not surprisingly, the definitions become more similar to color refinement and the ones of Grebík and Rocha [9]. Only proofs that significantly differ from their counterpart in Section 4 are included. At the end of this section, we also briefly show how *simple non-oblivious k-WL* can be defined.

To prevent edges from being added between input vertices, we only allow certain combinations of adjacency and neighbor graphs; after a bunch of adjacency graphs connecting a vertex j to other vertices, we immediately follow up with a j-neighbor graph. Formally, for every (j, V) in the set $S^k := \{(j, V) \mid j \in [k] \text{ and } V \subseteq [k] \setminus \{j\}\}$, define the bi-labeled graph

$$\boldsymbol{S}^k_{j,V} := \boldsymbol{N}^k_j \circ \bigcirc_{i \in V} \boldsymbol{A}^k_{ij} \in \mathcal{G}^{k,k}.$$

Then, let $\mathcal{F}^{\mathsf{s}k} := \left\{\boldsymbol{S}^k_{j,V} \mid (j, V) \in S^k\right\} \subseteq \mathcal{G}^{k,k}$ be the set of all these bi-labeled graphs. We have to be a bit cautious as, in general, these graphs are not symmetric and, hence, their graphon operators are not self-adjoint; in general, the set $\mathcal{F}^{\mathsf{s}k}$ is not even closed under transposition. Note that, by definition, the $\boldsymbol{S}^k_{j,V}$-graphon operator of a graphon W is given by

$$(T_{\boldsymbol{S}^k_{j,V} \to W} f)(\bar{x}) = \int\limits_X \Big(\prod_{i \in V} W(x_i, y)\Big) \cdot f \circ \bar{x}[y/j] \, d\mu(y)$$

for $\mu^{\otimes k}$-almost every $\bar{x} \in X^k$. Analogously to Lemma 13, one can observe that the underlying graphs of $[\![\mathbb{F}]\!]$ for terms $\mathbb{F} \in \langle \mathcal{F}^{\mathsf{s}k} \rangle_{\circ,\cdot}$ are, again up to isolated vertices, precisely the simple graphs of treewidth at most $k - 1$. Basically, when constructing a term from a nice tree decomposition, we just add all missing edges when a vertex is forgotten. This way, every edge is added the graph as the bag at the root node of a nice tree decomposition is the empty set.

For the sake of brevity, we write $\mathbb{T}^{\mathsf{s}k}_W := \mathbb{T}_{\mathcal{F}^{\mathsf{s}k} \to W}$ for a graphon W. Define $\mathcal{C}^{\mathsf{s}k}_{W,n} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ for every $n \in \mathbb{N}$ by setting $\mathcal{C}^{\mathsf{s}k}_{W,0} := \left\langle\left\{\varnothing, X^k\right\}\right\rangle$, $\mathcal{C}^{\mathsf{s}k}_{W,n+1} := \mathbb{T}^{\mathsf{s}k}_W(\mathcal{C}^{\mathsf{s}k}_{W,n})$ for every $n \in \mathbb{N}$, and finally, $\mathcal{C}^{\mathsf{s}k}_W := \mathcal{C}^{\mathsf{s}k}_{W,\infty} := \langle \bigcup_{n \in \mathbb{N}} \mathcal{C}^{\mathsf{s}k}_{W,n} \rangle$. Then, analogously to Lemma 22, one can show that $\mathcal{C}^{\mathsf{s}k}_W$ is permutation-invariant and the minimum $\mathbb{T}^{\mathsf{s}k}_W$-invariant $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$. We now deviate a bit from the definition of W-invariance and call a $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ *simply W-invariant* if $\mathcal{C}$ is invariant for every operator in the family $(\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W}$, i.e., $\mathcal{C}$ is $(T_{\boldsymbol{F} \to W})_{\mathcal{C}^{\mathsf{s}k}_W}$-invariant for every $\boldsymbol{F} \in \mathcal{F}^{\mathsf{s}k}$. The reason for this is that, since $\mathbb{T}^{\mathsf{s}k}_W$ is not closed under taking adjoints, $\mathcal{C}_W^{\mathsf{s}k}$ might not be invariant under these adjoints. In contrast, $\mathcal{C}_W^{\mathsf{s}k}$ is trivially both $(\mathbb{T}_W^{\mathsf{s}k})_{\mathcal{C}_W^{\mathsf{s}k}}$-invariant and $(\mathbb{T}_W^{\mathsf{s}k})^*_{\mathcal{C}_W^{\mathsf{s}k}}$-invariant. In fact, it is easy to see that $\mathcal{C}_W^{\mathsf{s}k}$ is also the minimum simply W-invariant $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$.

For a separable metrizable space $(X, \mathcal{T})$, let $\mathscr{M}_{\leq 1}(X)$ denote the set of all measures of total mass at most 1. We endow $\mathscr{M}_{\leq 1}(X)$ with a topology analogously to $\mathscr{P}(X)$, i.e., with the topology generated by the maps $\mu \mapsto \int f\, d\mu$ for $f \in C_b(X)$. Then, for measures that all have the same total mass, the Portmanteau theorem is still applicable as we can scale them to have total mass of one. Let $P_0^{\mathsf{s}k} := \{1\}$ be the one-point space and inductively define

$$\mathbb{M}_n^{\mathsf{s}k} := \prod_{i \leq n} P_i^{\mathsf{s}k} \text{ and } P_{n+1}^{\mathsf{s}k} := \left(\mathscr{M}_{\leq 1}\left(\mathbb{M}_n^{\mathsf{s}k}\right)\right)^{S^k}$$

for every $n \in \mathbb{N}$. Let $\mathbb{M}^{\mathsf{s}k} := \mathbb{M}_\infty^{\mathsf{s}k} := \prod_{n \in \mathbb{N}} P_i^{\mathsf{s}k}$ and, for $n \leq m \leq \infty$, let $p_{m,n} \colon \mathbb{M}_m^{\mathsf{s}k} \to \mathbb{M}_n^{\mathsf{s}k}$ be the natural projection. Finally, define

$$\mathbb{P}^{\mathsf{s}k} := \left\{\alpha \in \mathbb{M}^{\mathsf{s}k} \mid (\alpha_{n+1})_{(j,V)} = (p_{n+1,n})_*(\alpha_{n+2})_{(j,V)} \text{ for all } (j, V) \in S^k, n \in \mathbb{N}\right\}.$$

By the Kolmogorov Consistency Theorem [15, Exercise 17.16], for all $\alpha \in \mathbb{P}^k$ and $(j, V) \in S^k$, there is a unique measure $\mu^\alpha_{(j,V)} \in \mathscr{P}(\mathbb{M}^k)$ such that $(p_{\infty,n})_*\mu^\alpha_{(j,V)} = (\alpha_{n+1})_{(j,V)}$ for every $n \in \mathbb{N}$. Analogously to Lemma 24, the set $\mathbb{P}^{\mathsf{s}k}$ is closed in $\mathbb{M}^{\mathsf{s}k}$ and, for every $(j, V) \in S^k$, the mapping $\mathbb{P}^{\mathsf{s}k} \to \mathscr{P}(\mathbb{M}^{\mathsf{s}k}), \alpha \mapsto \mu^\alpha_{(j,V)}$ is continuous. To adapt the definition of k-WLD, we add a third requirement of absolute continuity and Radon-Nikodym derivatives, cf. the definition of distributions over iterated degree measures [9].

Definition 47. *Let $k \geq 1$. A measure $\nu \in \mathscr{P}(\mathbb{M}^{\mathsf{s}k})$ is called a simple k-Weisfeiler-Leman distribution (simple k-WLD) if*

1. $\nu(\mathbb{P}^{\mathsf{s}k}) = 1$,
2. $\int_{\mathbb{M}^{\mathsf{s}k}} f\, d\nu = \int_{\mathbb{M}^{\mathsf{s}k}} \left(\int_{\mathbb{M}^{\mathsf{s}k}} f\, d\mu^\alpha_{j,\varnothing}\right) d\nu(\alpha)$ *for all bounded measurable $f \colon \mathbb{M}^{\mathsf{s}k} \to \mathbb{R}$, $j \in [k]$, and*
3. $\mu^\alpha_{j,V} \preccurlyeq \mu^\alpha_{j,\varnothing}$ *and* $0 \leq \frac{d\mu^\alpha_{j,V}}{d\mu^\alpha_{j,\varnothing}} \leq 1$ *for ν-almost every $\alpha \in \mathbb{M}^{\mathsf{s}k}$ and every $(j, V) \in S^k$.*

Let $W \colon X \times X \to [0, 1]$ be a graphon. Define $\mathsf{owl}^{\mathsf{s}k}_{W,0} \colon X^k \to \mathbb{M}_0^{\mathsf{s}k}$ by $\mathsf{owl}^{\mathsf{s}k}_{W,0}(\bar{x}) := 1$ for every $\bar{x} \in X^k$. Inductively define $\mathsf{owl}^{\mathsf{s}k}_{W,n+1} \colon X^k \to \mathbb{M}_{n+1}^{\mathsf{s}k}$ by

$$\mathsf{owl}^{\mathsf{s}k}_{W,n+1}(\bar{x}) := \left(\mathsf{owl}^{\mathsf{s}k}_{W,n}(\bar{x}), \left(A \mapsto \int_{{\mathsf{owl}^{\mathsf{s}k}_{W,n}}^{-1}(A)_{\bar{x}[/j]}} \prod_{i \in V} W(x_i, y)\, d\mu(y)\right)_{(j,V) \in S^k}\right)$$

for every $\bar{x} \in X^k$. Then, let $\mathsf{owl}^{\mathsf{s}k}_W = \mathsf{owl}^{\mathsf{s}k}_{W,\infty} \colon X^k \to \mathbb{M}^{\mathsf{s}k}$ be the mapping defined by $(\mathsf{owl}^{\mathsf{s}k}_W(\bar{x}))_n := (\mathsf{owl}^{\mathsf{s}k}_{W,\infty}(\bar{x}))_n := (\mathsf{owl}^{\mathsf{s}k}_{W,n}(\bar{x}))_n$ for all $n \in \mathbb{N}$, $\bar{x} \in X^k$. Finally, let $\nu^{\mathsf{s}k}_W := \mathsf{owl}^{\mathsf{s}k}_{W*}\mu^{\otimes k} \in \mathscr{P}(\mathbb{M}^{\mathsf{s}k})$ be the push-forward of $\mu^{\otimes k}$ via $\mathsf{owl}^{\mathsf{s}k}_W$. Analogously to Lemma 27, one can show that

$$\mathcal{C}^{\mathsf{s}k}_{W,n} = \left\langle \left\{ {\mathsf{owl}^{\mathsf{s}k}_{W,n}}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}_n^{\mathsf{s}k}) \right\} \right\rangle.$$

for $n \in \mathbb{N} \cup \{\infty\}$. Defining $R^{\mathsf{s}k}_W := S_{\mathcal{C}^{\mathsf{s}k}_W} \circ T_{\mathsf{owl}^{\mathsf{s}k}_W}$ yields a Markov isomorphism from $L^2(\mathbb{M}^{\mathsf{s}k}, \nu^{\mathsf{s}k}_W)$ to $L^2(X^k/\mathcal{C}^{\mathsf{s}k}_W, \mu^{\otimes k}/\mathcal{C}^{\mathsf{s}k}_W)$, cf. Corollary 28. Let us explicitly state the adaptation of Lemma 29 since the proof requires some additional work.

Lemma 48. *Let $k \geq 1$ and $W\colon X \times X \to [0,1]$ be a graphon. Then,*

1. $\mu_{j,\varnothing}^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})} = (\mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j])_* \mu$ *for all* $j \in [k]$, $\bar{x} \in X^k$,
2. $\mathsf{owl}^{\mathsf{s}k}_W(X^k) \subseteq \mathbb{P}^{\mathsf{s}k}$, *and*
3. $\nu^{\mathsf{s}k}_W$ *is a simple* k*-WLD.*

Proof. For 1, observe that $\mu_{j,\varnothing}^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}$ is a probability measure. Then, the proof is analogous to Lemma 29 1. The proof of 2 is analogous to Lemma 29 2. For 3, we get $\nu^{\mathsf{s}k}_W(\mathbb{P}^{\mathsf{s}k}) = 1$ and $\int_{\mathbb{M}^{\mathsf{s}k}} f \, d\nu^{\mathsf{s}k} = \int_{\mathbb{M}^{\mathsf{s}k}} \left(\int_{\mathbb{M}^{\mathsf{s}k}} f \, d\mu^{\alpha}_{j,\varnothing} \right) d\nu^{\mathsf{s}k}_W(\alpha)$ for every bounded measurable $f\colon \mathbb{M}^{\mathsf{s}k} \to \mathbb{R}$ and every $j \in [k]$ as in the proof of Lemma 29 3. Let $(j,V) \in S^k$. Let $\bar{x} \in X^k$ and let

$$\mathcal{C} := \left\langle \left\{ \bar{x}[\cdot/j]^{-1}(\mathsf{owl}^{\mathsf{s}k^{-1}}_W(A)) \mid A \in \mathcal{B}(\mathbb{M}^{\mathsf{s}k}) \right\} \right\rangle$$

be the minimum μ-relatively complete sub-σ-algebra that makes $\mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j]$ measurable. Then, $\mathbb{E}(y \mapsto \prod_{i \in V} W(x_i, y) \mid \mathcal{C}) \in L^2(X, \mathcal{C}, \mu)$ and hence, by Claim 7, there is a measurable function $g\colon X \to \mathbb{R}$ such that $\mathbb{E}(y \mapsto \prod_{i \in V} W(x_i, y) \mid \mathcal{C}) = g \circ \mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j]$ μ-almost everywhere. Note that $0 \leq g \leq 1$ holds μ-almost everywhere. For every $n \in \mathbb{N}$ and every $A \in \mathcal{B}(\mathbb{M}^{\mathsf{s}k}_n)$, we have

$$\begin{aligned}
\mu_{j,V}^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}(p_{\infty,n}^{-1}(A)) = (p_{\infty,n})_* \mu_{j,V}^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}(A) &= (\mathsf{owl}^{\mathsf{s}k}_W(\bar{x}))_{n+1}(A) \\
&= (\mathsf{owl}^{\mathsf{s}k}_{W,n+1}(\bar{x}))_{n+1}(A) \\
&= \int_{\mathsf{owl}^{\mathsf{s}k^{-1}}_{W,n}(A)_{\bar{x}[/j]}} \prod_{i \in V} W(x_i, y) \, d\mu(y) \\
&= \int_{\bar{x}[\cdot/j]^{-1}(\mathsf{owl}^{\mathsf{s}k^{-1}}_W(p_{\infty,n}^{-1}(A)))} \prod_{i \in V} W(x_i, y) \, d\mu(y) \\
&= \int_{\bar{x}[\cdot/j]^{-1}(\mathsf{owl}^{\mathsf{s}k^{-1}}_W(p_{\infty,n}^{-1}(A)))} \mathbb{E}(y \mapsto \prod_{i \in V} W(x_i, y) \mid \mathcal{C}) \, d\mu \\
&\qquad \text{(Claim 5)} \\
&= \int_{\bar{x}[\cdot/j]^{-1}(\mathsf{owl}^{\mathsf{s}k^{-1}}_W(p_{\infty,n}^{-1}(A)))} g \circ \mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j] \, d\mu \\
&= \int_{p_{\infty,n}^{-1}(A)} g \, d(\mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j])_* \mu \\
&= \int_{p_{\infty,n}^{-1}(A)} g \, d\mu_{j,\varnothing}^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}.
\end{aligned}$$

Since $\bigcup_{n \in \mathbb{N}} \left\{ p_{\infty,n}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}^{\mathsf{s}k}_n) \right\}$ generates $\mathcal{B}(\mathbb{M}^{\mathsf{s}k})$, the π-λ theorem [15, Theorem 10.1 iii)] yields that $\mu_{j,V}^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}(A) = \int_A g \, d\mu_{j,\varnothing}^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}$ for every $A \in \mathcal{B}(\mathbb{M}^{\mathsf{s}k})$. Therefore, $\mu^{\alpha}_{j,V} \preccurlyeq \mu^{\alpha}_{j,\varnothing}$ and $0 \leq \frac{d\mu^{\alpha}_{j,V}}{d\mu^{\alpha}_{j,\varnothing}} \leq 1$ for every $\alpha \in \mathsf{owl}^{\mathsf{s}k}_W(X^k)$. By definition of $\nu^{\mathsf{s}k}_W$, this holds $\nu^{\mathsf{s}k}_W$-almost everywhere. Hence, $\nu^{\mathsf{s}k}_W$ is a simple k-WLD. ☐

Let $\nu \in \mathscr{P}(\mathbb{M}^{\mathsf{s}k})$ be a simple k-WLD and $(j,V) \in S^k$. By definition of a k-WLD, we have $0 \leq \frac{d\mu^{\alpha}_{j,V}}{d\mu^{\alpha}_{j,\varnothing}} \leq 1$ for ν-almost every $\alpha \in \mathbb{M}^{\mathsf{s}k}$. Hence, analogously to Lemma 30, one can show that setting

$$(T_{\boldsymbol{S}^k_{j,V} \to \nu} f)(\alpha) := \int_{\mathbb{M}^{\mathsf{s}k}} \frac{d\mu^{\alpha}_{j,V}}{d\mu^{\alpha}_{j,\varnothing}} \cdot f \, d\mu^{\alpha}_{j,\varnothing} = \int_{\mathbb{M}^{\mathsf{s}k}} f \, d\mu^{\alpha}_{j,V}$$

for all $f \in L^\infty(\mathbb{M}^{\mathsf{s}k}, \nu)$, $\alpha \in \mathbb{M}^{\mathsf{s}k}$ defines an L^∞-contraction that uniquely extends to an L^2-contraction $L^2(\mathbb{M}^{\mathsf{s}k}, \nu) \to L^2(\mathbb{M}^{\mathsf{s}k}, \nu)$.

Lemma 49. *Let $k \geq 1$ and $W\colon X \times X \to [0,1]$ be a graphon. For every $\boldsymbol{S} \in \mathcal{F}^{\mathsf{s}k}$,*

1. $T_{\boldsymbol{S}\to W} \circ T_{\mathsf{owl}^{\mathsf{s}k}_W} = T_{\mathsf{owl}^{\mathsf{s}k}_W} \circ T_{\boldsymbol{S}\to \nu^{\mathsf{s}k}_W}$,
2. $(T_{\boldsymbol{S}\to W})_{\mathcal{C}^k_W} \circ T_{\mathsf{owl}^{\mathsf{s}k}_W} = T_{\mathsf{owl}^{\mathsf{s}k}_W} \circ T_{\boldsymbol{S}\to \nu^{\mathsf{s}k}_W}$, *and*
3. $T_{\boldsymbol{S}\to W}/\mathcal{C}^k_W \circ R^{\mathsf{s}k}_W = R^{\mathsf{s}k}_W \circ T_{\boldsymbol{S}\to \nu^{\mathsf{s}k}_W}$.

Proof. Let $(j, V) \in S^k$ such that $\boldsymbol{S} = \boldsymbol{S}^k_{j,V}$. For $\bar{x} \in X^k$, let $\mathcal{C}_{\bar{x}}$ denote the minimum μ-relatively complete sub-σ-algebra that makes $\mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j]$ measurable. As seen in the proof of Lemma 48, we have

$$\mathbb{E}(y \mapsto \prod_{i\in V} W(x_i, y) \mid \mathcal{C}_{\bar{x}}) = \frac{d\mu^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}_{j,V}}{d\mu^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}_{j,\varnothing}} \circ \mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j]$$

μ-almost everywhere. Then, we have

$$\begin{aligned}
(T_{\mathsf{owl}^{\mathsf{s}k}_W} \circ T_{\boldsymbol{S}\to \nu^{\mathsf{s}k}_W} f)(\bar{x}) &= \int_{\mathbb{M}^{\mathsf{s}k}} \frac{d\mu^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}_{j,V}}{d\mu^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}_{j,\varnothing}} \cdot f \, d(\mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j])_*\mu \qquad \text{(Definition and Lemma 48 1)}\\
&= \int_X \mathbb{E}(y \mapsto \prod_{i\in V} W(x_i, y) \mid \mathcal{C}_{\bar{x}}) \cdot f \circ \mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j] \, d\mu\\
&= \int_X \prod_{i\in V} W(x_i, y) \cdot \mathbb{E}(f \circ \mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j] \mid \mathcal{C}_{\bar{x}})(y) \, d\mu(y) \qquad \text{(Claim 5)}\\
&= \int_X \prod_{i\in V} W(x_i, y) \cdot f \circ \mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[y/j] \, d\mu(y)\\
&= (T_{\boldsymbol{S}\to W} \circ T_{\mathsf{owl}^{\mathsf{s}k}_W} f)(\bar{x})
\end{aligned}$$

for every $f \in L^\infty(\mathbb{M}^{\mathsf{s}k}, \nu)$ and $\mu^{\otimes k}$-almost every $\bar{x} \in X^k$. As $L^\infty(\mathbb{M}^{\mathsf{s}k}, \nu^{\mathsf{s}k}_W)$ is dense in $L^2(\mathbb{M}^{\mathsf{s}k}, \nu^{\mathsf{s}k}_W)$, this implies 1. From there on, 2 and 3 are analogous to Lemma 31 2 and 3, respectively. ☐

For $k \geq 1$ and a simple k-WL distribution $\nu \in \mathscr{P}(\mathbb{M}^{\mathsf{s}k})$, let $\mathbb{T}_\nu := (T_{\boldsymbol{S}\to\nu})_{\boldsymbol{S}\in\mathcal{F}^{\mathsf{s}k}}$. Then, for a graphon $W\colon X \times X \to [0,1]$, we have

$$\mathbb{T}^{\mathsf{s}k}_W/\mathcal{C}^{\mathsf{s}k}_W \circ R^{\mathsf{s}k}_W = R^{\mathsf{s}k}_W \circ \mathbb{T}_{\nu^{\mathsf{s}k}_W} \qquad \text{and} \qquad {\mathbb{T}^{\mathsf{s}k}_W}^*/\mathcal{C}^{\mathsf{s}k}_W \circ R^{\mathsf{s}k}_W = R^{\mathsf{s}k}_W \circ \mathbb{T}^*_{\nu^{\mathsf{s}k}_W},$$

where the first equation is just Lemma 49 and the second equation follows from the first since $R^{\mathsf{s}k}$ is a Markov isomorphism. As before, a permutation $\pi\colon [k] \to [k]$ naturally extends to a measurable bijection $\pi\colon \mathbb{M}^{\mathsf{s}k} \to \mathbb{M}^{\mathsf{s}k}$, and the π-invariance, and more general the permutation invariance, of a simple k-WLD can be defined analogously to Section 4.4. The analogous result to Lemma 36 holds as well; in particular, $\nu^{\mathsf{s}k}_W$ is permutation invariant for a graphon W. Let $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ be simply W-invariant; recall that this definition is a bit quirky as it means that $\mathcal{C}$ is $(\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W}$-invariant. Corollary 20 can then be adapted to the also somewhat quirky statement, that

$$t(\mathbb{F}, \mathbb{T}_{\nu^{\mathsf{s}k}_W}) = t(\mathbb{F}, ((\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W})_{\mathcal{C}}) = t(\mathbb{F}, (\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W}/\mathcal{C}) = t(\mathbb{F}, \mathbb{T}^{\mathsf{s}k}_W) = t(\llbracket \mathbb{F} \rrbracket, W)$$

holds for every $\mathbb{F} \in \langle \mathcal{F}^{\mathsf{s}k} \rangle_{\circ,\cdot}$. To prove this, one has to apply Lemma 19 twice this time: first, to get from $\mathbb{T}^{\mathsf{s}k}_W$ to $(\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W}$ and, second, to get from there to $((\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W})_{\mathcal{C}}$ and $(\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W}/\mathcal{C}$.

For a term $\mathbb{F} \in \langle \mathcal{F}^{\mathsf{s}k} \rangle_{\circ,\cdot}$ and every $n \in \mathbb{N}$ with $n \geq h(\mathbb{F})$, the set $F_n^{\mathbb{F}}$ of functions $\mathbb{M}_n^{\mathsf{s}k} \to [0,1]$ is defined similarly to Definition 37. More precisely, while we could just use the old definition, it can actually be simplified as the distinct cases for adjacency and neighbor graphs can be subsumed by the function

$$\alpha \mapsto \int_{\mathbb{M}_n^{\mathsf{s}k}} f \, d(\alpha_{n+1})_{(j,V)} \in F_{n+1}^{\boldsymbol{S}_{j,V}^k \circ \mathbb{F}}$$

for every $f \in F_n^{\mathbb{F}}$ and every $j \in [k]$. From there, we analogously obtain the set $F^{\mathbb{F}}$ of continuous functions $\mathbb{M}^{\mathsf{s}k} \to [0,1]$. Lemma 38 and Corollary 39 adapt in a straight-forward fashion.

For every $n \in \mathbb{N} \cup \{\infty\}$, define $\mathcal{T}_n^{\mathsf{s}k} := \bigcup_{\mathbb{F} \in \langle \mathcal{F}^{\mathsf{s}k} \rangle_{\circ,\cdot}, h(\mathbb{F}) \leq n} F_n^{\mathbb{F}}$ and abbreviate $\mathcal{T}^{\mathsf{s}k} := \mathcal{T}_\infty^{\mathsf{s}k}$. Lemma 40 also adapts easily, i.e., for every $n \in \mathbb{N} \cup \{\infty\}$, the set $\mathcal{T}_n^{\mathsf{s}k}$ is closed under multiplication, contains $\mathbf{1}_{\mathbb{M}_n^{\mathsf{s}k}}$, and separates points of $\mathbb{M}_n^{\mathsf{s}k}$. Here, one has to observe that the all-one function distinguishes two measures if their total mass is different, which means that the Portmanteau theorem is still applicable in this case. From there, we obtain the following analogue to Lemma 41.

Lemma 50. *Let $k \geq 1$. Let $(W_n)_n$ and $W \colon X \times X \to [0,1]$ be a sequence of graphons and a graphon, respectively. Then, $\nu_{W_n}^{\mathsf{s}k} \to \nu_W^{\mathsf{s}k}$ if and only if $t(F, W_n) \to t(F, W)$ for every simple graph F of treewidth at most $k-1$.*

Since $\mathscr{P}(\mathbb{M}^{\mathsf{s}k})$ is Hausdorff, this also means that the simple k-WLDs of two graphons are equal if and only if their treewidth $k-1$ simple graph homomorphism densities are. With the Counting Lemma [16, Lemma 10.23], we also obtain the following additional corollary, which does *not* hold for k-WLDs as the Counting Lemma does not hold for multigraphs.

Corollary 51. *Let $k \geq 1$. The mapping $\mathcal{W}_0 \to \mathscr{P}(\mathbb{M}^{\mathsf{s}k}), W \mapsto \nu_W^{\mathsf{s}k}$ is continuous when $\mathcal{W}_0$ is endowed with the cut distance.*

Having outlined the necessary changes for simple graphs, we obtain the following variant of Theorem 4 for simple graph homomorphism densities. Note the quirky characterization via Markov operators, which is quite artificial in this case; this again stems from the fact that the family $\mathbb{T}_W^{\mathsf{s}k}$ of operators is not closed under taking adjoints.

Theorem 52. *Let $k \geq 1$ and $U, W \colon X \times X \to [0,1]$ be graphons. The following are equivalent:*

1. *$t(F, U) = t(F, W)$ for every simple graph of treewidth at most $k-1$.*
2. *$\nu_U^{\mathsf{s}k} = \nu_W^{\mathsf{s}k}$.*
3. *There is a (permutation-inv.) Markov iso. $R \colon L^2(X^k/\mathcal{C}_W^{\mathsf{s}k}, \mu^{\otimes k}/\mathcal{C}_W^{\mathsf{s}k}) \to L^2(X^k/\mathcal{C}_U^{\mathsf{s}k}, \mu^{\otimes k}/\mathcal{C}_U^{\mathsf{s}k})$ such that $\mathbb{T}_U^{\mathsf{s}k}/\mathcal{C}_U^{\mathsf{s}k} \circ R = R \circ \mathbb{T}_W^{\mathsf{s}k}/\mathcal{C}_W^{\mathsf{s}k}$.*
4. *There is a (permutation-inv.) Markov operator $S \colon L^2(X^k, \mu^{\otimes k}) \to L^2(X^k, \mu^{\otimes k})$ such that $(\mathbb{T}_U^{\mathsf{s}k})_{\mathcal{C}_U^{\mathsf{s}k}} \circ S = S \circ (\mathbb{T}_W^{\mathsf{s}k})_{\mathcal{C}_W^{\mathsf{s}k}}$ and $S^* \circ (\mathbb{T}_U^{\mathsf{s}k})_{\mathcal{C}_U^{\mathsf{s}k}} = (\mathbb{T}_W^{\mathsf{s}k})_{\mathcal{C}_W^{\mathsf{s}k}} \circ S^*$.*
5. *There are $\mu^{\otimes k}$-rel. comp. sub-σ-algebras $\mathcal{C}, \mathcal{D}$ of $\mathcal{B}^{\otimes k}$ that are simply U-invariant and simply W-invariant, respectively, and a Markov iso. $R \colon L^2(X^k/\mathcal{D}, \mu^{\otimes k}/\mathcal{D}) \to L^2(X^k/\mathcal{C}, \mu^{\otimes k}/\mathcal{C})$ such that $(\mathbb{T}_U^{\mathsf{s}k})_{\mathcal{C}_U^{\mathsf{s}k}}/\mathcal{C} \circ R = R \circ (\mathbb{T}_W^{\mathsf{s}k})_{\mathcal{C}_W^{\mathsf{s}k}}/\mathcal{D}$.*

Proof. 1 $\implies$ 2: Follows from Lemma 50.

2 $\implies$ 3: Analogous to Theorem 4 as we have both $\mathbb{T}_U^{\mathsf{s}k}/\mathcal{C}_U^{\mathsf{s}k} \circ R_U^{\mathsf{s}k} = R_U^{\mathsf{s}k} \circ \mathbb{T}_{\nu_U^{\mathsf{s}k}}$ and $(R_W^{\mathsf{s}k})^* \circ \mathbb{T}_W^{\mathsf{s}k}/\mathcal{C}_W^{\mathsf{s}k} = \mathbb{T}_{\nu_W^{\mathsf{s}k}} \circ (R_W^{\mathsf{s}k})^*$ since $R_W^{\mathsf{s}k}$ is a Markov isomorphism.

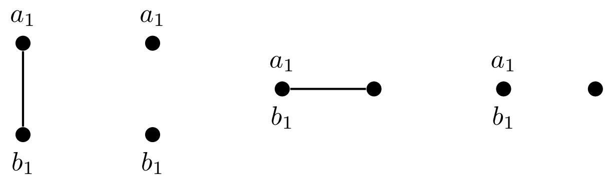

Figure 9: The (isomorphism types of) graphs in $\mathcal{F}^{\mathsf{ns}1}$.

3 $\implies$ 4: Set $S := I_{\mathcal{C}^{\mathsf{s}k}_U} \circ R \circ S_{\mathcal{C}^{\mathsf{s}k}_W}$, which is a Markov operator as the composition of Markov operators. Then,

$$\begin{aligned}
(\mathbb{T}^{\mathsf{s}k}_U)_{\mathcal{C}^{\mathsf{s}k}_U} \circ S = (\mathbb{T}^{\mathsf{s}k}_U)_{\mathcal{C}^{\mathsf{s}k}_U} \circ I_{\mathcal{C}^{\mathsf{s}k}_U} \circ R \circ S_{\mathcal{C}^{\mathsf{s}k}_W} &= I_{\mathcal{C}^{\mathsf{s}k}_U} \circ \mathbb{T}^{\mathsf{s}k}_U / \mathcal{C}^{\mathsf{s}k}_U \circ R \circ S_{\mathcal{C}^{\mathsf{s}k}_W} && \text{(Lemma 8 3)}\\
&= I_{\mathcal{C}^{\mathsf{s}k}_U} \circ R \circ \mathbb{T}^{\mathsf{s}k}_W / \mathcal{C}^{\mathsf{s}k}_W \circ S_{\mathcal{C}^{\mathsf{s}k}_W}\\
&= I_{\mathcal{C}^{\mathsf{s}k}_U} \circ R \circ S_{\mathcal{C}^{\mathsf{s}k}_W} \circ (\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W} && \text{(Lemma 8 4)}\\
&= S \circ (\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W}.
\end{aligned}$$

Note that we neither used that $\mathcal{C}^{\mathsf{s}k}_U$ is $\mathbb{T}^{\mathsf{s}k}_U$-invariant nor that $\mathcal{C}^{\mathsf{s}k}_W$ is $\mathbb{T}^{\mathsf{s}k}_W$-invariant. Since R is a Markov isomorphism, we also have ${\mathbb{T}^{\mathsf{s}k}_U}^* / \mathcal{C}^{\mathsf{s}k}_U \circ R = R \circ {\mathbb{T}^{\mathsf{s}k}_W}^* / \mathcal{C}^{\mathsf{s}k}_W$, which means that we obtain $({\mathbb{T}^{\mathsf{s}k}_U}^*)_{\mathcal{C}^{\mathsf{s}k}_U} \circ S = S \circ ({\mathbb{T}^{\mathsf{s}k}_W}^*)_{\mathcal{C}^{\mathsf{s}k}_W}$ in an analogous fashion. This implies the claim. Moreover, analogously to Theorem 4, if R is permutation invariant, then so is S.

4 $\implies$ 5: Follows immediately from Lemma 9.

5 $\implies$ 1: Analogous to Theorem 4. □

Also in this case, it is possible to define the space $\mathbb{M}^{\mathsf{s}\infty}$ and, for a graphon $W \colon X \times X \to [0, 1]$, the measure $\nu^{\mathsf{s}\infty}_W \in \mathscr{P}(\mathbb{M}^{\mathsf{s}\infty})$. Then, one obtains the following lemma corresponding to Corollary 44, where we now have a third characterization in terms of the cut distance $\delta_\square$, cf. [16, Theorem 11.5].

Lemma 53. *Let $(W_n)_n$ and $W \colon X \times X \to [0, 1]$ be a sequence of graphons and a graphon, respectively. Then, the following are equivalent:*

1. $\nu^{\mathsf{s}\infty}_{W_n} \to \nu^{\mathsf{s}\infty}_W$.
2. $t(F, W_n) \to t(F, W)$ *for every simple graph F.*
3. $W_n \xrightarrow{\delta_\square} W$.

One can easily adapt the definitions of this section to obtain a non-oblivious variant of simple k-WL. To this end, let $\mathcal{F}^{\mathsf{ns}k}$ to be the set of all bi-labeled graphs

$$\boldsymbol{F}_{k+1,j_1} \circ \bigcirc_{i \in V} \boldsymbol{A}_{k+1,ij_1} \circ \boldsymbol{I}_{k+1,j_2} \in \mathcal{G}^{k,k}$$

for $j_1, j_2 \in [k+1]$, $V \subseteq [k+1] \setminus \{j_1\}$. Note that every term in $\langle \mathcal{F}^{\mathsf{s}k+1} \rangle_{\circ,\cdot}$ can be turned into a term in $\langle \mathcal{F}^{\mathsf{ns}k} \rangle_{\circ,\cdot}$ by essentially re-grouping the introduce and forget graphs. For $k = 1$, the isomorphism types in $\mathcal{F}^{\mathsf{ns}k}$ are shown in Figure 9; they all are symmetric in this special case. All definitions and results from this section transfer to the set $\mathcal{F}^{\mathsf{ns}k}$ and, in particular, one can obtain a variant of Theorem 52 without the mismatch of the k of simple k-WL and the k of the treewidth.

6 Conclusions

We have shown how oblivious k-WL and the work of Grebík and Rocha [9] can be married, or in other words, how oblivious k-WL and some of its characterizations generalize to graphons. In particular, we obtained that oblivious k-WL characterizes graphons in terms of their homomorphism densities from multigraphs of treewidth at most $k-1$. This was made possible by using a special set of bi-labeled graphs as building blocks for the multigraphs of treewidth $k-1$ and considering the graphon operators these bi-labeled graphs. Additionally, we have shown how oblivious k-WL can be modified to obtain a characterization via simple graphs: simple oblivious k-WL corresponds to homomorphism densities from simple graphs of treewidth at most $k-1$. However, the characterizations obtained this way are less elegant as the set of bi-labeled graphs one uses as building blocks is not closed under transposition, i.e., the corresponding family of operators is not closed under taking Hilbert adjoints.

The original goal of this work was to define a k-WL distance of graphons and to prove that it yields the same topology as treewidth k homomorphism densities, cf. [3], where the result of Grebík and Rocha is used to prove such a result for the *tree distance*. However, this does not work out as hoped since multigraph homomorphism densities define a topology different from the one obtained by the cut distance, cf. [16, Exercise 10.26] or [14, Lemma C.2]. Moreover, the quirky characterization of simple k-WL via Markov operators, which stems from the non-symmetric bi-labeled graphs used as building blocks, is also not well-suited to define such a distance. Hence, it remains an open problem to define such a distance.

Acknowledgments: We would like to thank Jan Grebík for pointing out Section 4.6.

References

[1] Albert Atserias and Elitza Maneva. “Sherali–Adams Relaxations and Indistinguishability in Counting Logics”. In: *SIAM Journal on Computing* 42.1 (2013), pp. 112–137. DOI: 10.1137/120867834. eprint: https://doi.org/10.1137/120867834.

[2] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 1995. ISBN: 9780471007104.

[3] Jan Böker. “Graph Similarity and Homomorphism Densities”. In: *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*. Ed. by Nikhil Bansal, Emanuela Merelli, and James Worrell. Vol. 198. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 32:1–32:17. ISBN: 978-3-95977-195-5. DOI: 10.4230/LIPIcs.ICALP.2021.32.

[4] Jin Yi Cai, Martin Fürer, and Neil Immerman. “An optimal lower bound on the number of variables for graph identification”. English (US). In: *Combinatorica* 12.4 (Dec. 1992), pp. 389–410. ISSN: 0209-9683. DOI: 10.1007/BF01305232.

[5] Holger Dell, Martin Grohe, and Gaurav Rattan. “Lovász Meets Weisfeiler and Leman”. In: *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*. Ed. by Ioannis Chatzigiannakis et al. Vol. 107. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018, 40:1–40:14. ISBN: 978-3-95977-076-7. DOI: 10.4230/LIPIcs.ICALP.2018.40.

[6] R. M. Dudley. *Real Analysis and Probability*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002. DOI: 10.1017/CBO9780511755347.

[7] Zdeněk Dvořák. “On recognizing graphs by numbers of homomorphisms”. In: *Journal of Graph Theory* 64.4 (2010), pp. 330–342. DOI: 10.1002/jgt.20461.

[8] T. Eisner et al. *Operator Theoretic Aspects of Ergodic Theory*. Graduate Texts in Mathematics. Springer International Publishing, 2015. ISBN: 9783319168982.

[9] Jan Grebík and Israel Rocha. *Fractional Isomorphism of Graphons*. Accepted to Combinatorica. 2021. arXiv: 1909.04122 [math.CO].

[10] Martin Grohe. “The Logic of Graph Neural Networks”. In: *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*. IEEE, 2021, pp. 1–17. DOI: 10.1109/LICS52264.2021.9470677.

[11] Martin Grohe and Martin Otto. “Pebble Games and Linear Equations”. In: *The Journal of Symbolic Logic* 80.3 (2015), pp. 797–844. DOI: 10.1017/jsl.2015.28.

[12] Martin Grohe et al. “Color Refinement and Its Applications”. In: *An Introduction to Lifted Probabilistic Inference*. The MIT Press, Aug. 2021. ISBN: 9780262365598. DOI: 10.7551/mitpress/10548.003.00

[13] Neil Immerman and Eric Lander. “Describing Graphs: A First-Order Approach to Graph Canonization”. In: *Complexity Theory Retrospective: In Honor of Juris Hartmanis on the Occasion of His Sixtieth Birthday, July 5, 1988*. Ed. by Alan L. Selman. New York, NY: Springer New York, 1990, pp. 59–81. ISBN: 978-1-4612-4478-3. DOI: 10.1007/978-1-4612-4478-3_5.

[14] Svante Janson. “Graphons, cut norm and distance, couplings and rearrangements”. In: *New York Journal of Mathematics*. NYJM Monographs 4 (2013).

[15] A.S. Kechris. *Classical Descriptive Set Theory*. Graduate texts in mathematics. Springer-Verlag, 1995. ISBN: 9780387943749.

[16] László Lovász. *Large Networks and Graph Limits*. Vol. 60. Colloquium Publications. American Mathematical Society, 2012, pp. I–XIV, 1–475. ISBN: 978-0-8218-9085-1. DOI: 10.1090/coll/060.

[17] Laura Mančinska and David E. Roberson. “Quantum isomorphism is equivalent to equality of homomorphism counts from planar graphs”. In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. 2020, pp. 661–672. DOI: 10.1109/FOCS46700.2020.00067.

[18] Motakuri V. Ramana, Edward R. Scheinerman, and Daniel Ullman. “Fractional isomorphism of graphs”. In: *Discrete Mathematics* 132.1 (1994), pp. 247–265. ISSN: 0012-365X. DOI: https://doi.org/10.1016/0012-365X(94)90241-0.

[19] Gottfried Tinhofer. “A note on compact graphs”. In: *Discrete Applied Mathematics* 30.2 (1991), pp. 253–264. ISSN: 0166-218X. DOI: 10.1016/0166-218X(91)90049-3.

[20] Gottfried Tinhofer. “Graph Isomorphism and Theorems of Birkhoff-Type”. In: *Computing* 36.4 (June 1986), pp. 285–300. ISSN: 0010-485X. DOI: 10.1007/BF02240204.