Weisfeiler-Leman Indistinguishability of Graphons

Jan Böker

Submitted: Jan 14, 2022; Accepted: Oct 18, 2023; Published: Dec 1, 2023

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 (2022) have recently shown how color refinement and notions that characterize it generalize to graphons, which emerged as limit objects in the theory of dense graph limits. 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 more general case of graphons, which means that, there, the equivalence between 1-WL and color refinement does not hold anymore. We also show how this equivalence can be recovered by defining a variant of k-WL that corresponds to homomorphism densities from simple graphs of bounded treewidth.

Mathematics Subject Classifications: 05C80, 05C50, 05C60

1 Introduction

Color refinement is a polynomial-time algorithm best known as an efficient heuristic for graph isomorphism testing even though it has more applications, e.g., as graph kernels in machine learning [13]. It iteratively computes a coloring of the vertices of a simple graph, and two graphs are called *indistinguishable by color refinement* if the resulting multisets of colors match. For isomorphic graphs, the resulting multisets of colors are necessarily the same, but there are non-isomorphic graphs that still produce the same multisets of colors. The *k-dimensional Weisfeiler-Leman algorithm (k-WL)* is a generalization of color refinement that colors k-dimensional tuples of vertices instead of single vertices. This

RWTH Aachen University, Aachen, Germany (boeker@informatik.rwth-aachen.de).

again yields a heuristic for graph isomorphism testing by comparing the resulting multisets of colors, where two graphs are called *indistinguishable by k-WL* if the resulting multisets of colors match. Starting from 1-WL, which is equivalent to color refinement, it yields a hierarchy of ever-more-powerful polynomial-time algorithms, none of which actually decides graph isomorphism [6].

There are various seemingly unrelated characterizations of indistinguishability by color refinement, for example, by homomorphism counts from trees [9, 7] or by rational solutions to a certain system of linear equations encoding graph isomorphism called *fractional isomorphisms* [25, 24]. Similar characterizations exist for the k-dimensional Weisfeiler-Leman algorithm in terms of homomorphism counts from graphs of bounded treewidth [9, 7] and Sherali-Adams relaxations of the system of linear equations encoding graph isomorphism [7, 16, 1, 14]. Grebík and Rocha recently investigated the graphon counterpart of fractional isomorphism by first providing graphon counterparts of the most important notions used as characterizations for fractional isomorphism of graphs and then proving that they are all equivalent [11]. *Graphons* emerged as limit objects for sequences of graphs in the theory of dense graph limits developed by Borgs, Chayes, Lovász, Sós, Szegedy, and Vesztergombi [21, 4, 5]. In this theory, *homomorphism densities* play a crucial role as a starting point for a notion of convergence of a sequence of graphs and lead to the *cut distance*, a (pseudo-)metric on graphons. The book of Lovász [20] provides a comprehensive overview.

In this paper, we provide graphon counterparts of the most important notions used as characterizations for k-WL indistinguishability and also prove their equivalence. As Grebík and Rocha stress, for fractional isomorphism of graphons, both defining the corresponding notions and proving their equivalence turned out to be surprisingly difficult. This is no different in the case of k-WL indistinguishability, where somewhat unsurprisingly, even more difficulties and technical hurdles arise. In particular, it turns out that there is no clear single generalization of k-WL indistinguishability to graphons but multiple non-equivalent variants. The arguably most interesting characterization we obtain is in terms of homomorphism densities: The starting point of Grebík and Rocha was to call two graphons U and W *fractionally isomorphic* if the homomorphism density $t(T, U)$ of T in U equals the homomorphism density of T in W for every (finite simple) tree T. Based on this and the characterizations of k-WL for graphs, it is only natural to ask what kind of similarity U and W have to satisfy for $t(F, U) = t(F, W)$ to hold for every (finite simple) graph F of treewidth at most k. While we give an answer to this, we also show that a much more elegant characterization and direct correspondence to the usual definition of k-WL is obtained if we require $t(F, U) = t(F, W)$ to hold for every *multi*graph F of treewidth at most k instead.

1.1 Finite Graphs

In this section, we give a brief description of color refinement and k-WL for finite graphs. Moreover, we briefly present the notions characterizing them that are relevant to us. A *(finite 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 $E(G) \coloneqq E$. The initial coloring of color refinement for the vertices of a graph G is defined by simply letting $\mathsf{cr}_{G,0}(v) \coloneqq 1$ for every vertex $v \in V(G)$. Then, for every $n \geqslant 0$, 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 $v \in V(G)$, where $\{\!\{\cdot\}\!\}$ is used to denote a *multiset*. Hence, the new color of a vertex v is determined by aggregating the colors of all neighbors of v, and in particular, two vertices u and v get different colors if they have a different number of neighbors of some color c. Every coloring $\mathsf{cr}_{G,n}$ induces a partition of $V(G)$, and after a finite number of steps, the coloring we obtain is *stable*, i.e., the next and all further colorings induce the same partition. For graphs G and H, we say that *G and H are indistinguishable by color refinement* if $\{\!\{\mathsf{cr}_{G,n}(v) \mid v \in V(G)\}\!\} = \{\!\{\mathsf{cr}_{H,n}(v) \mid v \in V(H)\}\!\}$ holds for every $n \geqslant 0$.

The notions characterizing indistinguishability by color refinement that are important for us are tree homomorphisms, fractional isomorphisms, and stable partitions. 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*. Then, 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 corresponding number $\mathsf{hom}(T,H)$ from T to H for every tree T [9], see also [7]. An older result due to Tinhofer [25, 24] 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. One can show that the partition induced by the colors of color refinement is the *coarsest stable partition* and that 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 [25]. This, in turn, is equivalent to there being *some* stable partitions of G and H with the same parameters [23].

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 $k \geqslant 1$. See [6] 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. However, when stating the formal definition of k-WL, it is important to note that already for graphs there actually are two non-equivalent definitions to be found in the literature. Following Grohe [12], we refer to these distinct definitions as *(non-oblivious) 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. In this paper, we nearly always consider oblivious k-WL and only briefly define non-oblivious k-WL at the end of the paper in Section 5.2. However, to avoid any confusion, we nevertheless continue to explicitly use the term oblivious k-WL from here on and use the term k-WL only for non-oblivious k-WL.

Let G be a graph. To define oblivious k-WL, we first have to define the *atomic type* $\mathsf{atp}_G(\bar{v})$ of a tuple $\bar{v} = (v_1, \ldots, v_k) \in V(G)^k$ of vertices of G, which is the $k \times k$-matrix A with entries $A_{ij} = 2$ if $v_i = v_j$, $A_{ij} = 1$ if $v_i v_j \in E(G)$, and $A_{ij} = 0$ otherwise. We let $\mathsf{owl}^k_{G,0}(\bar{v}) \coloneqq \mathsf{atp}_G(\bar{v})$ for every $\bar{v} \in V(G)^k$, and then for every $n \geqslant 0$, we define

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

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; this k-tuple is usually called a *j-neighbor* of $\bar{v}$. Hence, the new color of a tuple $\bar{v}$ is determined by aggregating the colors of all j-neighbors of $\bar{v}$ for every $j \in [k]$, and in particular, two tuples $\bar{u}$ and $\bar{v}$ get different colors if, for some $j \in [k]$, they have a different number of j-neighbors of some color c. 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 $n \geqslant 0$.

The previously described notions that characterize indistinguishability by color refinement generalize to oblivious k-WL: First of all, oblivious $(k+1)$-WL does not distinguish graphs G and H if and only if 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 [9, 7]. A system of linear equations $\mathsf{L}^k_{\text{iso}}(G, H)$, which is closely related to the Sherali-Adams relaxations of the system of linear equations encoding graph isomorphism, generalizes the concept of fractional isomorphisms: oblivious k-WL does not distinguish G and H if and only if $\mathsf{L}^k_{\text{iso}}(G, H)$ has a non-negative real solution, cf. [7] and also [16, 1, 14]. The precise formulation of $\mathsf{L}^k_{\text{iso}}(G, H)$ is given in Section 4.8. Stable partitions of the vertex set $V(G)$ of a graph G can be generalized to *stable partitions of* $V(G)^k$, where tuples with different atomic types are in different classes and, for every $j \in [k]$, all tuples in the same class have the same number of j-neighbors in every other class. One can again show that 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 [14], where this is implicitly treated.

Remark 1. Immerman and Lander [16] first showed that fractional isomorphism of graphs 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. More generally, indistinguishability by oblivious k-WL corresponds to equivalence in C^k, the k-variable fragment of first-order logic with counting quantifiers [6]. However, this perspective does not play a further role in this paper.

1.2 Graphons and Homomorphism Densities

Graphons emerged in the theory of graph limits as limit objects of sequences of dense graphs; we refer to the book of Lovász [20] for a comprehensive treatment of this topic. Formally, a graphon is a symmetric Borel- or Lebesgue-measurable (this usually does not make a difference) function $W\colon [0,1] \times [0,1] \to [0,1]$, although it can be useful to consider more general underlying spaces than the unit interval with the Lebesgue measure. A graph G can be viewed as a graphon W_G by partitioning $[0,1]$ into intervals $I_1, \dots, I_n$ of the same size—one for each vertex—and setting $W_G(x,y)$ either to 1 or 0 for all $x \in I_i$, $y \in I_j$ depending on whether ij is an edge in G or not. Similarly, a vertex- and edge-weighted graph H with edge weights in $[0,1]$ can be viewed as graphon W_H, cf. [20, Section 7.1]. This allows one to restore statements about graphs and weighted graphs from statements about graphons, e.g., the equivalence between the notions characterizing fractional isomorphism of graphs from the results of Grebík and Rocha.

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; using this as the underlying space for graphons 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 $W\colon X \times X \to [0,1]$, and a symmetric kernel is called a *graphon*. An important way to view a kernel $W\colon X \times X \to [0,1]$ as an *(bounded linear) operator* is by defining the *kernel operator* $T_W\colon L^2(X,\mu) \to L^2(X,\mu)$ by setting

$$(T_W f)(x) \coloneqq \int_X W(x,y) f(y) \, d\mu(y) \tag{2}$$

for every $f \in L^2(X,\mu)$ and every $x \in X$. It is a well-defined Hilbert-Schmidt operator[20, Section 7.5], and if W is a graphon, then T_W is self-adjoint, i.e., its *Hilbert adjoint* T_W^* satisfies $T_W^* = T_W$; in general, the Hilbert adjoint of an operator $S\colon L^2(X,\mu) \to L^2(Y,\nu)$, where $(X,\mathcal{B})$ and $(Y,\mathcal{D})$ are standard Borel spaces with Borel probability measures μ and ν on X and Y, respectively, 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 $f \in L^2(X,\mu), g \in L^2(Y,\nu)$.

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

$$t(F,W) \coloneqq \int_{X^{V(F)}} \prod_{ij \in E(F)} W(x_i,x_j) \, d\mu^{\otimes V(F)}(\bar{x}), \tag{3}$$

where $\bar{x}$ denotes the vector of all variables x_i for $i \in V(F)$. This coincides with the previous definition for graphs, i.e., for the graphon W_G obtained from a graph G as described above, we have $t(F,G) = t(F,W_G)$ [20, (7.2)]. 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 has various characterizations. For example, two graphons U and W are weakly isomorphic if and only if their *cut distance* $\delta_\square(U,W)$ is zero, cf. [20, Section 10.7] for this result and the definition of the cut distance. The definition of weak isomorphism via homomorphism densities is robust in the following sense: A *multigraph* is a tuple $G = (V,E)$ where V is set of vertices and E is a *multiset* of

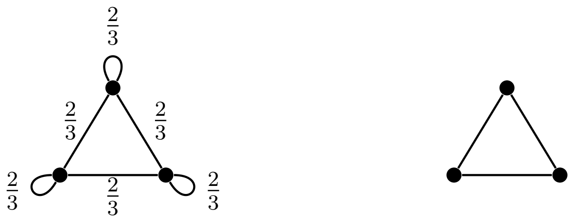

Figure 1: Two fractionally isomorphic weighted graphs that are 1-WL distinguishable.

edges from $\binom{V}{2}$. The definition of the homomorphism density $t(F, W)$ of F in a graphon $W\colon X \times X \to [0,1]$ in Equation (3) extends to the case in which F is a multigraph, where we slightly abuse notation and assume that each factor $W(x_i, x_j)$ occurs as often in the product $\prod_{ij \in E(F)} W(x_i, x_j)$ as ij is contained in $E(F)$. Then, two graphons U and W are weakly isomorphic if and only if $t(F, U) = t(F, W)$ for every multigraph F [20, Corollary 10.36], i.e., the definition of weak isomorphism remains unchanged if we use multigraphs instead of simple graphs.

Coming from the definition of weak isomorphism, the first definition in the paper of Grebík and Rocha is to call two graphons U and W *fractionally isomorphic* if $t(T, U) = t(T, W)$ holds for every tree T. We follow suit with the following definition, where the treewidth of a multigraph is defined analogously to the case of simple graphs, i.e., the edge multiplicities are not taken into account and parallel edges count as a single edge.

Definition 2. Two graphons $U, W\colon X \times X \to [0,1]$ are called *k-WL indistinguishable* if $t(T, U) = t(T, W)$ holds for every multigraph of treewidth at most k, and U and W are called *simply k-WL indistinguishable* if $t(T, U) = t(T, W)$ holds for every simple graph of treewidth at most k.

In these terms, two graphons U and W are weakly isomorphic if and only if they are k-WL indistinguishable for every k, which again is equivalent to them being simply k-WL indistinguishable for every $k \geqslant 1$. However, while k-WL indistinguishability clearly implies simple k-WL indistinguishability, the converse does not hold in general. To illustrate this, let E_ℓ denote the multigraph that consists of two vertices connected by ℓ parallel edges. Then, the homomorphism densities of E_ℓ in a graphon $W\colon X \times X \to [0,1]$ for $\ell \geqslant 0$ are precisely the moments of W, i.e., we have

$$t(E_\ell, W) = \int_{X \times X} W(x,y)^\ell d(\mu \times \mu)(x,y).$$

Hence, for graphons U and W to be 1-WL indistinguishable, it is already necessary that they have the same moments, while this is not required for U and W to be fractionally isomorphic. For example, the constant graphon $W_c\colon X \times X \to [0,1], (x,y) \mapsto c$ for $c \in [0,1]$ is fractionally isomorphic to any c-regular graphon, i.e., a graphon $W\colon X \times X \to [0,1]$ satisfying $\int_X W(x,y) d\mu(y) = c$ for every $x \in X$, but these may not have the same moments. A concrete example is given by the weighted graphs in Figure 1.

Since 1-WL indistinguishability is a more restrictive notion than fractional isomorphism, Definition 2 also introduces simple k-WL indistinguishability. Simple 1-WL indistinguishability is just fractional isomorphism since, for any class of graphs closed under connected components, the homomorphism densities of the connected components determine the homomorphism densities of all graphs in that class, cf. [20, (7.6)]. However counter-intuitive it may seem at this point, the characterizations we obtain for k-WL indistinguishability are much more natural than these for simple k-WL indistinguishability. In particular, the adaption of oblivious $(k+1)$-WL to graphons corresponds to k-WL indistinguishability and not simple k-WL indistinguishability.

As a last remark, we note that for $\{0,1\}$-graphons, i.e., graphons that only take the values 0 and 1, parallel edges do not make a difference for homomorphism densities since powers of 1 are just 1. Hence, for $\{0,1\}$-graphons—of which graphs, or more precisely graphons obtained from graphs, are a special case—, the two notions of k-WL indistinguishability and simple k-WL indistinguishability coincide. In particular, $\{0,1\}$-graphons are 1-WL indistinguishable if and only if they are fractionally isomorphic. Hence, while the terms color refinement and 1-WL are usually used synonymously in the literature, it is important to not confuse these concepts as they differ in the more general case of graphons.

1.3 Fractional Isomorphism of Graphons

In this section, we try to give a more formal but still high-level overview over the notions introduced by Grebík and Rocha that all characterize fractional isomorphism of graphons. We deem this essential for understanding our results.

Stating the color-refinement algorithm for graphons requires more formalism 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. For a graphon $W \colon X \times X \to [0,1]$, they define a measurable function $\mathsf{cr}_W \colon X \to \mathbb{M}$ (denoted i_W in their work) 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 is a probability measure on $\mathbb{M}$ defined as the *push-forward of* μ *via* cr_W, i.e., by $\nu_W(A) := \mu(\mathsf{cr}_W^{-1}(A))$ for every $A \in \mathcal{B}(\mathbb{M})$. Intuitively, this is the distribution of all colors assigned to the points of the graphon W and corresponds to the multiset of all colors used in the definition of color-refinement indistinguishability of graphs.

Grebík and Rocha show that the graphon analogue to stable partitions of the vertex set of a graph are *sub-σ-algebras* that satisfy certain properties. To gain some intuition, consider the sub-σ-algebra $\{\varnothing, X\}$ of $\mathcal{B}$: in some way, it corresponds to the partition of the vertex set of a graph that consists of a single class containing all vertices. Formally, Grebík and Rocha consider *μ-relatively complete* sub-σ-algebras, where 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 $\Theta(\mathcal{B}, \mu)$. Then, for our example, the smallest μ-relatively complete sub-σ-algebra that includes $\{\varnothing, X\}$

corresponds to the partition of the vertex set of a graph that consists of a single class containing all vertices.

Let $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$ be a μ-relatively complete sub-σ-algebra $\mathcal{C}$ of $\mathcal{B}$. If we 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, we can then consider the subspace $L^2(X, \mathcal{C}, \mu)$ of $L^2(X, \mu)$ consisting of all $\mathcal{C}$-measurable functions since $\mathcal{C}$ is μ-relatively complete. Moreover, there is a quotient space corresponding to $\mathcal{C}$, i.e., a standard Borel space $(X/\mathcal{C}, \mathcal{C}')$ with a Borel probability measure $\mu/\mathcal{C}$ on $X/\mathcal{C}$. Alternatively, the *conditional expectation* $\mathbb{E}(- \mid \mathcal{C})$, i.e., the orthogonal projection onto $L^2(X, \mathcal{C}, \mu)$, yields a different but equivalent perspective on quotient spaces.

Now, to connect sub-σ-algebras to *stable* partitions, for an operator $T\colon L^2(X, \mu) \to L^2(X, \mu)$, a μ-relatively complete sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$ 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, for a graphon $W\colon X \times X \to [0, 1]$, the μ-relatively complete sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$ is called *W-invariant* if it is T_W-invariant, where we recall that T_W is the operator defined by Equation (2). Grebík and Rocha show that there is a minimum W-invariant μ-relatively complete sub-σ-algebra $\mathcal{C}_W$ of $\mathcal{B}$ (denoted $\mathcal{C}(W)$ in their work), which corresponds to the coarsest stable partition of the vertex set of a graph.

Finally, for a graphon $W\colon X \times X \to [0, 1]$ and a W-invariant μ-relatively complete sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$, Grebík and Rocha define the *quotient graphon* $W/\mathcal{C}$ on the space $X/\mathcal{C} \times X/\mathcal{C}$. Then, for graphons $U, W\colon X \times X \to [0, 1]$, saying that the two quotient graphons $U/\mathcal{C}_U$ and $W/\mathcal{C}_W$ are isomorphic corresponds to saying that two coarsest stable partitions have the same parameters. Alternatively to the quotient graphon $W/\mathcal{C}$, one can also consider $W_{\mathcal{C}} \coloneqq \mathbb{E}(W \mid \mathcal{C} \times \mathcal{C})$, the conditional expectation of W given $\mathcal{C} \times \mathcal{C}$. Intuitively, the difference is that $W_{\mathcal{C}}$ is obtained by simply averaging over the color classes of $\mathcal{C}$, while $W/\mathcal{C}$ is obtained by first averaging over the color classes of $\mathcal{C}$ and then identifying all elements of a color class.

Remark 3. Grebík and Rocha show that every DIDM ν defines a kernel $\mathbb{M} \times \mathbb{M} \to [0, 1]$ and 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, this can be viewed as a canonical representation of W on the space of all colors.

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 $Sf \geqslant 0$ for every $f \in L^2(X, \mu)$ with $f \geqslant 0$, $S\mathbf{1}_X = \mathbf{1}_Y$, 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. The Markov operator S is called a *Markov embedding* if it is an isometry, i.e., $\|Sf\|_2 = \|f\|_2$ for every $f \in L^2(X, \mu)$, and a *Markov isomorphism* if it is a surjective Markov embedding. Markov operators are simply the infinite-dimensional analogue to doubly stochastic matrices and yield the graphon analogue to fractional isomorphisms. The main result of Grebík and Rocha then is the following Theorem 4.

Theorem 4 ([11])**.** *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.

Recall the various notions characterizing fractional isomorphism of graphs presented in Section 1.1. Characterization (1) corresponds to homomorphism numbers of trees and is the definition of fractional isomorphism of graphons, Characterization (2) corresponds to color refinement not distinguishing two graphs, Characterization (3) corresponds to the coarsest stable partitions of two graphs having the same parameters, Characterization (4) generalizes fractional isomorphisms, and Characterization (5) corresponds to some stable partitions of two graphs having the same parameters. We remark that there is a subtle difference in the way Grebík and Rocha phrase Characterization (3) and (5): the former uses quotient spaces and the stronger notion of isomorphism while the latter uses conditional expectation and weak isomorphism.

1.4 Weisfeiler-Leman Indistinguishability of Graphons

We continue in the vein of Section 1.3 and give an overview of the notions characterizing k-WL indistinguishability of graphons before stating our main result, Theorem 5. To generalize the oblivious k-WL to graphons, we first define the standard Borel space $\mathbb{M}^k$, which may not be confused with the product of k copies of $\mathbb{M}$. $\mathbb{M}^k$ is the k-dimensional analogue to $\mathbb{M}$ and can again be seen as the space of colors used by oblivious k-WL. Its elements $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots)$ are sequences, which can be viewed as the colors assigned to points of a graphon after $0, 1, 2, \dots$ refinement rounds. Based on the definition of oblivious k-WL for graphs, we define the measurable function $\mathsf{owl}^k_W\colon X^k \to \mathbb{M}^k$ mapping a k-tuple $\bar{x} \in X^k$ to a sequence $(\alpha_0, \alpha_1, \alpha_2, \dots)$. In particular, α_0 corresponds to the atomic type of a tuple of vertices and contains the values $W(x_i, x_j)$ for $ij \in \binom{[k]}{2}$. This further explains the difference between 1-WL indistinguishability and fractional isomorphism: Already the first step of oblivious 2-WL, which corresponds to 1-WL indistinguishability, completely determines the distribution of values attained by a graphon. In contrast, these distributions do not have to be equal for graphons to be fractionally isomorphic, cf. Section 1.2 and Figure 1.

To get an intuition of how the refinement step of oblivious k-WL adapts to graphons, recall how oblivious k-WL computes the new color $\mathsf{owl}^k_{G,n+1}(\bar{v})$ of a tuple $\bar{v} \in V(G)^k$ for a graph G in Equation (1): $\mathsf{owl}^k_{G,n+1}(\bar{v})$ is a tuple consisting of the old color $\mathsf{owl}^k_{G,n}(\bar{v})$ of $\bar{v}$ and, for every $j \in [k]$, the multiset

$$\{\!\{\mathsf{owl}^k_{G,n+1}(\bar{v}[w/j] \mid w \in V(G))\}\!\}$$

of colors of all j-neighbors of $\bar{v}$. For a graphon W, this multiset becomes the probability measure

$$A \mapsto \mu(\{y \in X \mid \mathsf{owl}^k_{W,n}(\bar{x}[y/j]) \in A\}),$$

where A is a set of "colors" used in the nth refinement step, i.e., we determine the mass of the j-neighbors of $\bar{x}$ having a color in A. Compiling these probability measures for every $j \in [k]$ into a single tuple then yields the new color $\mathsf{owl}^k_{W,n+1}(\bar{x})$ of $\bar{x}$. The sequence of all these colors $\mathsf{owl}^k_{W,n}(\bar{x})$ for $n = 0, 1, 2, \dots$ then yields the mapping $\mathsf{owl}^k_W \colon X^k \to \mathbb{M}^k$, and we define the *$k$-WL distribution ($k$-WLD)* ν^k_W as the push-forward of the product measure $\mu^{\otimes k}$ via owl^k_W. Then, ν^k_W is a probability measure on $\mathbb{M}^k$ corresponding to the multiset of colors computed by oblivious k-WL on a graph.

The central idea for getting from fractional isomorphism to k-WL indistinguishability is to replace the single operator $T_W \colon L^2(X, \mu) \to L^2(X, \mu)$ of a graphon $W \colon X \times X \to [0, 1]$ by a family $\mathbb{T}^k_W$ of operators on the product space $L^2(X^k, \mu^{\otimes k}) := L^2(X^k, \mathcal{B}^{\otimes k}, \mu^{\otimes k})$. This family $\mathbb{T}^k_W$ is indexed by a set $\mathcal{F}^k$ of *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 $k, \ell \geqslant 0$ are tuples of vertices such that both the entries of $\boldsymbol{a}$ and the entries of $\boldsymbol{b}$ are pairwise distinct; $\boldsymbol{a}$ and $\boldsymbol{b}$ may however overlap. The set $\mathcal{F}^k$ is carefully chosen such that its bi-labeled graphs, together with specific operations, serve as building blocks to construct precisely the graphs of treewidth at most $k - 1$. It contains two types of bi-labeled graphs: *adjacency graphs* and *j-neighbor graphs*, where intuitively, adjacency graphs insert an edge into a bag of a tree decomposition and j-neighbor graphs move from one bag of a tree decomposition to another by replacing a vertex by a fresh one. For a simple example, consider Figure 2, where $k = 3$ and a tree decomposition of the cycle C_4 is dissected into a sequence of bi-labeled graphs. Here, $\boldsymbol{A}^3_{12}$ and $\boldsymbol{A}^3_{23}$ are specific instances of adjacency graphs, while $\boldsymbol{N}^3_2$ is a j-neighbor graph. By "gluing" the output vertices of one bi-labeled graph to the input vertices of the next bi-labeled graph in the depicted order, we obtain $\mathbf{C}_4$, a bi-labeled variant of C_4 with input labels on the vertices from the upper bag and output labels on the vertices from the lower bag: going bottom up, $\boldsymbol{A}^3_{12}$ and $\boldsymbol{A}^3_{23}$ first insert the edges v_1v_4 and v_4v_3, then $\boldsymbol{N}^3_2$ replaces v_4 by v_2, and finally, $\boldsymbol{A}^3_{12}$ and $\boldsymbol{A}^3_{23}$ insert the edges v_1v_2 and v_2v_3.

Every bi-labeled graph $\boldsymbol{F} \in \mathcal{F}^k$ together with a graphon $W \colon X \times X \to [0, 1]$ defines a *graphon operator* $T_{\boldsymbol{F}\to W}$ on $L^2(X^k, \mu^{\otimes k})$. Then, $\mathbb{T}^k_W := (T_{\boldsymbol{F}\to W})_{\boldsymbol{F}\in\mathcal{F}^k}$ denotes the family of all these operators for bi-labeled graphs in $\mathcal{F}^k$. The graphon operators of adjacency graphs just multiply a given function $f(\bar{x})$ by the value $W(x_i, x_j)$ for fixed i, j. The graphon operator of a j-neighbor graph, on the other hand, averages in the direction j, i.e., integrates the given function $f(\bar{x})$ over the jth component. As an example, the graphon operators of the bi-labeled graphs from Figure 2 for a graphon $W \colon X \times X \to [0, 1]$ are given by $(T_{\boldsymbol{A}^3_{12}\to W} f)(x_1, x_2, x_3) := W(x_1, x_2) \cdot f(x_1, x_2, x_3)$, $(T_{\boldsymbol{A}^3_{23}\to W} f)(x_1, x_2, x_3) := W(x_2, x_3) \cdot f(x_1, x_2, x_3)$, and

$$(T_{\boldsymbol{N}^3_2\to W} f)(x_1, x_2, x_3) := \int_X f(x_1, y, x_3)\, d\mu(y)$$

for all $x_1, x_2, x_3 \in X$. At this point, the reader might already note the connection of these graphon operators to the description of oblivious k-WL for graphons given above as, intuitively, the graphon operators of adjacency graphs are used in the initial coloring and the graphon operators of j-neighbors are used in the refinement steps.

We will see that the composition of the graphon operators corresponding to the sequence

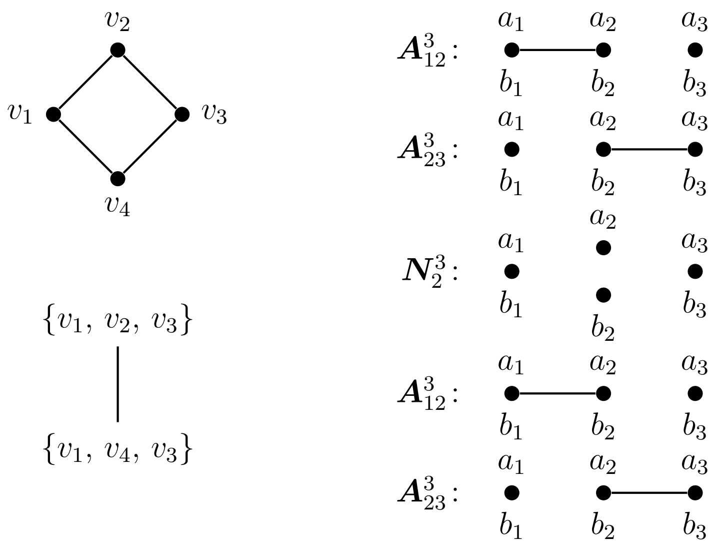

Figure 2: The cycle C_4 on four vertices with a tree decomposition. On the right, this tree-decomposed graph is written as a sequence of bi-labeled graphs.

of bi-labeled graphs in Figure 2,

$$T_{\boldsymbol{A}^3_{12}\to W} \circ T_{\boldsymbol{A}^3_{23}\to W} \circ T_{\boldsymbol{N}^3_{2}\to W} \circ T_{\boldsymbol{A}^3_{12}\to W} \circ T_{\boldsymbol{A}^3_{23}\to W},$$

is precisely the graphon operator $T_{\mathbf{C}_4\to W}$ of $\mathbf{C}_4$, i.e., the bi-labeled variant of C_4 obtained by gluing together this sequence of bi-labeled graphs. Furthermore, one can verify that

$$\int_{X^3} T_{\mathbf{C}_4\to W}\mathbf{1}_{X^3}\, d\mu^{\otimes 3} = t(C_4, W),$$

where $\mathbf{1}_{X^3}$ is the all-one function on X^3, i.e., the homomorphism density of C_4 in W is determined by the operator $T_{\mathbf{C}_4\to W}$. In greater generality, the homomorphism density of a bi-labeled graph in a graphon can always be expressed via the graphon operator.

Since we are now working with operators on $L^2(X^k, \mathcal{B}^{\otimes k}, \mu^{\otimes k})$, we consider $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$ for an analogue to partitions of $V(G)^k$ for some graph G. For a graphon $W\colon X \times X \to [0,1]$, a $\mu^{\otimes k}$-relatively complete sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ of $\mathcal{B}^{\otimes k}$ 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 earlier definition of Grebík and Rocha, but it will always be clear from the context what we mean. We show that there is a minimum W-invariant $\mu^{\otimes k}$-relatively complete sub-σ-algebra $\mathcal{C}^k_W$ of $\mathcal{B}^{\otimes k}$. The partitions of $V(G)^k$ induced by the colors of oblivious k-WL are invariant under permutations, i.e., reordering the vertices of a tuple yields a tuple in the same class. Similarly, we show that $\mathcal{C}^k_W$ is *permutation invariant* and also define the notion of *permutation-invariant* operators, i.e., operators where a reordering of the k components of X^k yields the same operator. This reflects the fact that, in the system $\mathsf{L}^k_{\text{iso}}$ of linear equations characterizing oblivious k-WL, variables are indexed by sets and not by tuples.

We cannot give a meaningful definition of the quotient of a graphon $W\colon X \times X \to [0,1]$ w.r.t. a $\mu^{\otimes k}$-relatively complete sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ if $k \geqslant 2$. Instead, we consider quotients of operators. Intuitively, for an operator T on $L^2(X^k, \mu^{\otimes k})$, its *quotient operator w.r.t.* $\mathcal{C}$ on $L^2(X^k/\mathcal{C}, \mu^{\otimes k}/\mathcal{C})$, denoted by $T/\mathcal{C}$, is defined by going from $L^2(X^k/\mathcal{C}, \mu^{\otimes k}/\mathcal{C})$ to $L^2(X^k, \mu^{\otimes k})$, applying T, and then going back to $L^2(X^k/\mathcal{C}, \mu^{\otimes k}/\mathcal{C})$. Again, a different but equivalent definition can also be given via conditional expectations by letting $T_{\mathcal{C}} \coloneqq \mathbb{E}(- \mid \mathcal{C}) \circ T \circ \mathbb{E}(- \mid \mathcal{C})$. Then, we can consider the families $\mathbb{T}_W^k/\mathcal{C}$ and $(\mathbb{T}_W^k)_{\mathcal{C}}$ of quotient operators w.r.t. $\mathcal{C}$.

We now state our main theorem, Theorem 5. As mentioned before, it is based on oblivious k-WL, so there is a mismatch between the k in the treewidth, i.e., the k in k-WL indistinguishability, and the other characterizations. We note that, since there are no quotient graphons involved in Theorem 5, we also do not obtain 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 $\mathbb{T}_W^k$ on the space $L^2(\mathbb{M}^k, \nu_W^k)$ by hand.

Theorem 5. *Let $k \geqslant 1$ 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 multigraph of treewidth at most $k-1$.*
2. *$\nu_U^k = \nu_W^k$.*
3. *There is a (permutation-invariant) Markov isomorphism $R\colon L^2(X^k/\mathcal{C}_W^k, \mu^{\otimes k}/\mathcal{C}_W^k) \to L^2(X^k/\mathcal{C}_U^k, \mu^{\otimes k}/\mathcal{C}_U^k)$ such that $\mathbb{T}_U^k/\mathcal{C}_U^k \circ R = R \circ \mathbb{T}_W^k/\mathcal{C}_W^k$.*
4. *There is a (permutation-invariant) Markov operator $S\colon L^2(X^k, \mu^{\otimes k}) \to L^2(X^k, \mu^{\otimes k})$ such that $\mathbb{T}_U^k \circ S = S \circ \mathbb{T}_W^k$.*
5. *There are $\mu^{\otimes k}$-relatively complete sub-σ-algebras $\mathcal{C}$ and $\mathcal{D}$ of $\mathcal{B}^{\otimes k}$ that are U-invariant and W-invariant, respectively, and a Markov isomorphism $R\colon L^2(X^k/\mathcal{D}, \mu^{\otimes k}/\mathcal{D}) \to L^2(X^k/\mathcal{C}, \mu^{\otimes k}/\mathcal{C})$ such that $\mathbb{T}_U^k/\mathcal{C} \circ R = R \circ \mathbb{T}_W^k/\mathcal{D}$.*

The characterizations of Theorem 5 are listed in the same order as these in Theorem 4. Recall the various notions characterizing k-WL indistinguishability of graphs presented in Section 1.1. Characterization (1) corresponds to homomorphism numbers of graphs of treewidth at most $k-1$ and is the definition of $(k-1)$-WL indistinguishability of graphons. We note that, as in the case of simple graphs, one could always assume the multigraphs in Characterization (1) to be connected, cf. [20, (7.6)]. For example, in the case $k = 2$, it could equivalently be phrased in terms of homomorphism densities of trees with parallel edges. Characterization (2) states that the k-WLDs of the graphons are the same and corresponds to oblivious k-WL not distinguishing two graphs. Characterization (3) –(5) look very similar, but have a different focus: Characterization (4) generalizes (non-negative real) solutions to the system $\mathsf{L}_{\text{iso}}^k(G,H)$ of linear equations by stating that there is a Markov operator on the product space $L^2(X^k, \mu^{\otimes k})$ that intertwines all operators in the families $\mathbb{T}_U^k$ and $\mathbb{T}_W^k$ simultaneously. 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. Characterization (3) and (5) on the other hand, correspond to the (coarsest) stable partitions of vertex-tuples of two graphs having the same parameters. We note that there is a one-to-one correspondence between Markov isomorphisms and measure-preserving almost bijections, cf. [11, Theorem E.3], but for the ease of presentation, we stick to Markov isomorphisms.

1.5 Overview and Further Remarks

In Section 2, the preliminaries, we collect some basics we need: we briefly visit product spaces, Markov operators, and quotient spaces before defining quotient operators. Section 3 formally introduces bi-labeled graphs and graphon operators, which are the key to both stating and proving Theorem 5. 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 $k-1$, and then, for a graphon W, the family of graphon operators $\mathbb{T}^k_W$. Section 4 is the main section of this paper containing the formal definitions of all notions in Theorem 5 and, of course, its proof, for which we follow the structure of Grebík and Rocha [11]. Let us give a brief overview of the proof and how the set $\mathcal{F}^k$ and the corresponding family of graphon operators $\mathbb{T}^k_W$ are used in it:

- (1) $\implies$ (2): We use the set $\mathcal{F}^k$ of bi-labeled graphs to construct expressions that correspond to precisely the tree-decomposed graphs of treewidth at most $k-1$ (Section 3.1). This allows us to define a set $\mathcal{T}^k \subseteq C(\mathbb{M}^k, \mathbb{R})$ of functions on $\mathbb{M}^k$ that corresponds to homomorphism densities of these graphs (Section 4.5). The Stone-Weierstrass Theorem yields that $\mathcal{T}^k$ is dense in $C(\mathbb{M}^k, \mathbb{R})$, which implies that k-WLDs are determined by homomorphism densities.
- (2) $\implies$ (3): We show that a k-WLD ν defines a family $\mathbb{T}_\nu$ of operators on $L^2(\mathbb{M}^k, \nu)$ (Section 4.4). Then, we proceed to show that, for every graphon W, the family $\mathbb{T}_{\nu^k_W}$ is in some sense isomorphic to the family $\mathbb{T}^k_W / \mathcal{C}^k_W$ of quotient operators of $\mathbb{T}^k_W$.
- (3) $\implies$ (4): We combine facts we establish on quotient operators (Section 2.4) together with the fact that $\mathcal{C}^k_U$ and $\mathcal{C}^k_W$ are $\mathbb{T}^k_U$- and $\mathbb{T}^k_W$-invariant, respectively.
- (4) $\implies$ (5): This is a refined variant of the argument by Grebík and Rocha in [11], which uses the Mean Ergodic Theorem for Hilbert spaces to Markov operators [10, Theorem 8.6, Example 13.24]. We have condensed this argument into a standalone lemma, Lemma 10.
- (5) $\implies$ (1): We again use the fact that the set $\mathcal{F}^k$ of bi-labeled graphs allows to construct expressions that correspond to precisely the tree-decomposed graphs of treewidth at most $k-1$ (Section 3.1). We use this to show that homomorphism densities of graphs of treewidth at most $k-1$ in a graphon W can be expressed purely by expressions built from the operators in $\mathbb{T}^k_W$ (Section 3.2). Then, we use that

Markov embeddings are compatible with point-wise products of functions, which for us means that intertwining Markov embeddings preserve homomorphism densities (Lemma 22).

In Section 5, we show how Theorem 5 can be modified to obtain a characterization of simple k-WL indistinguishability, i.e., indistinguishability w.r.t. homomorphism densities of simple graphs of treewidth at most k instead of multigraphs. However, the corresponding analogue to Theorem 5 obtained this way is less elegant and has an artificial touch to it. The reason for this is that the set of bi-labeled graphs one uses instead of $\mathcal{F}^k$ is not closed under *transposition*, which implies that the corresponding family of operators is not closed under taking Hilbert adjoints. Most of the proofs in Section 5 are left out as they are mostly analogous to the ones in Section 4.

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*, which is based on the characterization of fractional isomorphism via Markov operators. However, the approach in [3] does not go well together with Theorem 5 as multigraph homomorphism densities define a non-compact topology that is different from the one obtained by the cut distance, cf. [20, Exercise 10.26] or [17, Lemma C.2]. Moreover, the characterization of simple k-WL indistinguishability via Markov operators is also not well-suited for this as the corresponding family of operators is not closed under Hilbert adjoints. Hence, it remains an open problem to define such a distance.

A different open problem is given by the contemporaneous work of Grohe, Rattan, and Seppelt [15]: They do not focus on a specific set of bi-labeled graphs, but use bi-labeled graphs to give a unified framework to characterize graphs in terms of homomorphism numbers. In particular, they obtain a characterization of homomorphism numbers from graphs of bounded pathwidth. It would be interesting to see if their framework, or at least their work on graphs of bounded pathwidth, generalizes to graphons.

2 Preliminaries

In this section, we briefly collect some facts that we use throughout the paper. Section 2.1 concerns product spaces; for a more complete reference, we refer to [8, 2]. The definitions and results regarding Markov operators in Section 2.2 are taken from [10]. The treatment of quotient spaces in Section 2.3 is based on that of Grebík and Rocha [11]. We then use these quotient spaces to define *quotient operators* in Section 2.4.

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 $k \geqslant 1$. The product of a countable family of standard Borel spaces is again a standard Borel space [18, 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 [18, 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. [8, 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 $1 \leqslant p \leqslant \infty$, the space $\mathcal{L}^p(X, \mu) := \mathcal{L}^p(X, \mathcal{S}, \mu)$ consists of all measurable real-valued functions on X with $\|f\|_p < \infty$, and $L^p(X, \mu) := L^p(X, \mathcal{S}, \mu)$ 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 $\langle f, g \rangle := \int_X fg \, d\mu$. 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 $\|f\|_2 \leqslant \|f\|_\infty$ and, in particular, the inclusion $L^\infty(X, \mu) \subseteq L^2(X, \mu)$ holds.

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 $\|T\| := \sup\{|T(x)| \mid \|x\| \leqslant 1\} < \infty$, and if $\|T\| \leqslant 1$, 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 two 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 $\langle Tf, g \rangle = \langle f, T^* g \rangle$ for all $f \in L^2(X, \mu), g \in L^2(Y, \nu)$. For two 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 $Sf \geqslant 0$ for every

$f \in L^2(X, \mu)$ with $f \geqslant 0$, $S\mathbf{1}_X = \mathbf{1}_Y$, and $S^*\mathbf{1}_Y = \mathbf{1}_X$. Markov operators are both L^2- and L^∞-contractions [10, 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 [10, Example 13.1]. A *Markov isomorphism* is a surjective Markov embedding. Note that every Markov isomorphism S satisfies $S^{-1} = S^*$ [10, Corollary 13.14]. Moreover, there is a one-to-one correspondence between Markov isomorphisms and measure-preserving almost bijections, cf. [11, Theorem E.3]. See [10] 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. [10, 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 $\mu(Z \triangle Z_0) = 0$. 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 $\bigcap \Phi \coloneqq \bigcap_{\mathcal{C}\in\Phi} \mathcal{C} \in \Theta(\mathcal{B},\mu)$ [11, 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 $\langle \mathcal{X} \rangle$. If $\mathcal{C} \subseteq \mathcal{B}$ is a sub-σ-algebra, then one can show that $\langle \mathcal{C} \rangle = \{A \triangle Z \mid A \in \mathcal{C}, Z \in \mathcal{B} \text{ with } \mu(Z) = 0\}$. 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 $L^2(X,\mathcal{C},\mu)$. The *conditional expectation* $\mathbb{E}(- \mid \mathcal{C})$ is the orthogonal projection onto the closed linear subspace $L^2(X,\mathcal{C},\mu)$ of $L^2(X,\mu)$.

Proposition 6 (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 there is a self-adjoint operator $\mathbb{E}(- \mid \mathcal{C})\colon L^2(X,\mu) \to L^2(X,\mu)$ 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 $f, g \in L^2(X,\mu)$.*

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 $g_*\mu(A) \coloneqq \mu(g^{-1}(A))$ for every $A \in \mathcal{T}$. For a measurable function $f\colon Y \to [-\infty, \infty]$, we then have $\int_Y f \, d(g_*\mu) = \int_X f \circ g \, d\mu$ [8, Theorem 4.1.11]. The following proposition then guarantees the existence of a quotient space of $(X, \mathcal{B}, \mu)$ w.r.t. a μ-relatively complete sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B},\mu)$.

Proposition 7 ([11, 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 $L^2(X/\mathcal{C}, \mu/\mathcal{C})$.*

Proposition 8 is a technical result that intuitively states that the quotient space $(X/\mathcal{C}, \mathcal{C}')$ is unique and the same as $L^2(X, \mathcal{C}, \mu)$ up to sets of measure zero.

Proposition 8 ([11, 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 $x \in X$.*

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

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

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 9. *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 $T_{\mathcal{C}}$ and $T/\mathcal{C}$,*
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 Proposition 6 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 Proposition 7. 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 Proposition 7 and Proposition 6. 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 Proposition 7. For (5), assume that $\mathcal{C}$ is T-invariant. By Proposition 6, 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 Proposition 7 and Proposition 6. Now, (6) follows from (2), (5), and (3) of Proposition 7. □

The following lemma is an application of the Mean Ergodic Theorem for Hilbert spaces to Markov operators [10, Theorem 8.6, Example 13.24] and is the essence of the proof of the direction "(4) $\implies$ (5)" of Theorem 4 by Grebík and Rocha [11].

Lemma 10. *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. $R \coloneqq S_{\mathcal{C}} \circ S \circ I_{\mathcal{D}} \colon L^2(X/\mathcal{D}, \mu/\mathcal{D}) \to L^2(X/\mathcal{C}, \mu/\mathcal{C})$ *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) $\implies$ (5), in [11]; 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

$$\begin{aligned}
T_1/\mathcal{C} \circ R = S_{\mathcal{C}} \circ T_1 \circ I_{\mathcal{C}} \circ S_{\mathcal{C}} \circ S \circ I_{\mathcal{D}} &= S_{\mathcal{C}} \circ T_1 \circ \mathbb{E}(- \mid \mathcal{C}) \circ S \circ I_{\mathcal{D}} && \text{(Proposition 7 (6))} \\
&= S_{\mathcal{C}} \circ T_1 \circ S \circ \mathbb{E}(- \mid \mathcal{D}) \circ I_{\mathcal{D}} && \text{((3))} \\
&= S_{\mathcal{C}} \circ T_1 \circ S \circ I_{\mathcal{D}} && \text{(Proposition 7 (3) and (4))} \\
&= S_{\mathcal{C}} \circ S \circ T_2 \circ I_{\mathcal{D}} \\
&= S_{\mathcal{C}} \circ S \circ I_{\mathcal{D}} \circ T_2/\mathcal{D} && \text{(Lemma 9 (5))} \\
&= R \circ T_2/\mathcal{D}.
\end{aligned}$$

□

2.5 Permutation Invariance

Let $k \geqslant 1$ 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 $\pi(x_1, \ldots, x_k) \coloneqq (x_{\pi(1)}, \ldots, x_{\pi(k)})$ for all $x_1, \ldots, x_k \in X$, 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 $\mu^{\otimes k}$-relatively complete sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ *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.

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 $\pi\colon [k] \to [k]$.

3 Graphon Operators

In this section, we present the main ingredient to Theorem 5. 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 [14, Section 5.1], where they define a family of graphs and consider a matrix that is a fractional isomorphism between all these graphs simultaneously. The graphon setting will show that the step of defining these graphs for the sake of them having the right adjacency matrix is rather artificial and only works in the setting of (finite-dimensional) matrices: 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 [22] 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 [22] by defining the *graphon operator* of a bi-labeled graph and a graphon. The graphon operators of the bi-labeled graphs in $\mathcal{F}^k$ and a graphon W 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.

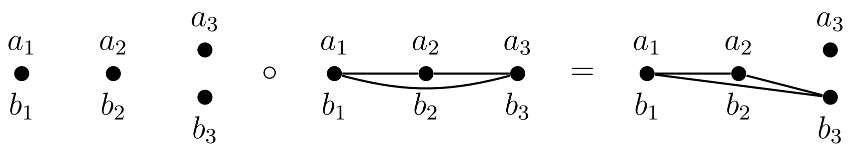

Figure 3: Composition of bi-labeled graphs.

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 $k, \ell \geqslant 0$ are tuples of vertices such that both the entries of $\boldsymbol{a}$ and the entries of $\boldsymbol{b}$ are pairwise distinct; $\boldsymbol{a}$ and $\boldsymbol{b}$ may however overlap. 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 tuples $\boldsymbol{a}$ and $\boldsymbol{b}$ are called the tuples 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. [22], but this is needed to ensure that graphon operators are well defined; the precise reason for this will be seen later when graphon operators are defined.

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 $k, \ell \geqslant 0$, 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 $\mathcal{M} := \cup_{k,\ell \geqslant 0} \mathcal{M}^{k,\ell}$ and $\mathcal{G} := \cup_{k,\ell \geqslant 0} \mathcal{G}^{k,\ell}$.

The *transpose* of a bi-labeled graph $\boldsymbol{G} = (G, \boldsymbol{a}, \boldsymbol{b}) \in \mathcal{M}^{k,\ell}$ is the bi-labeled graph $\boldsymbol{G}^* := (G, \boldsymbol{b}, \boldsymbol{a}) \in \mathcal{M}^{\ell,k}$, 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 $\boldsymbol{F_1} \circ \boldsymbol{F_2} := (F, \boldsymbol{a_1}, \boldsymbol{b_2}) \in \mathcal{M}^{k,\ell}$, 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 $i \in [m]$. An example is given in Figure 3. 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 $\boldsymbol{F_1} \cdot \boldsymbol{F_2} := (F, \boldsymbol{a_1}, ()) \in \mathcal{M}^{k,0}$, 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. [22]. 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. Both the composition and the Schur product of bi-labeled graph may introduce parallel edges, cf. Figure 4, which means that the set $\mathcal{G}$ is neither closed under composition nor under Schur products.

Treewidth is a graph parameter that measures how "tree-like" a graph is. Too see how the concept is related to bi-labeled graphs, let us first recall the standard 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,

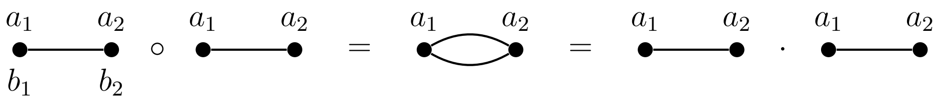

Figure 4: Both composition and the Schur product may introduce parallel edges.

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: First, a *rooted tree decomposition* 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. Then, a *nice tree decomposition* of a multigraph G is a rooted tree decomposition (T, r, β) such that

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). Nice tree decompositions are attractive from an algorithmic point of view because of their simplified structure, which allows one to specify dynamic-programming algorithms by a simple case distinction based on the node type. We do not design such an algorithm here but use nice tree decompositions to obtain a simple set of bi-labeled graphs that serve as building blocks for all graphs of treewidth at most k; nice tree decompositions do not pose a restriction since every graph G with treewidth k has a nice tree decomposition of width k.

Lemma 11 ([19, Lemma **13.1.2**])**.** *Every graph G with treewidth k has a nice tree decomposition of width k.*

We now want to view a bi-labeled graph $\boldsymbol{G}$ that is decomposed by a nice tree decomposition as a term built from atomic terms, where these atomic terms act as building blocks for the decomposition by providing elementary operations like adding an edge to a

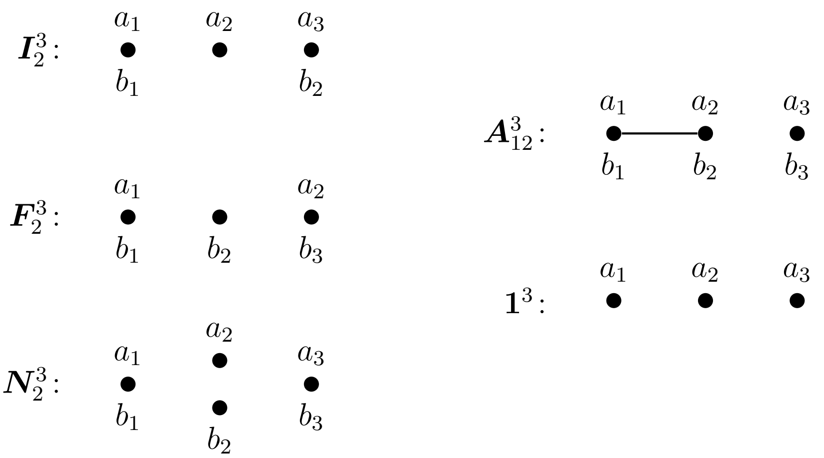

Figure 5: The bi-labeled graphs $\boldsymbol{I}_2^3$, $\boldsymbol{F}_2^3$, $\boldsymbol{N}_2^3$, $\boldsymbol{A}_{12}^3$, and $\boldsymbol{1}^3$.

bag or moving between bags of the decomposition. Such a term can then be evaluated to obtain $\boldsymbol{G}$, and in the next section, we show that by defining *graphon operators* for each atomic term, we can alternatively evaluate this expression to a function describing the homomorphism density of $\boldsymbol{G}$ in a graphon. The following definition gives us this set $\mathcal{F}^k$ of building blocks.

Definition 12. Let $k \geqslant 1$. Define

1. the *ij-adjacency graph* $\boldsymbol{A}_{ij}^k := (([k], \{ij\}), (1, \dots, k), (1, \dots, k)) \in \mathcal{G}^{k,k}$ for $i \neq j \in [k]$,
2. the *j-introduce graph* $\boldsymbol{I}_j^k := (([k], \varnothing), (1, \dots, k), (1, \dots, j-1, j+1, \dots, k)) \in \mathcal{G}^{k,k-1}$,
3. the *j-forget graph* $\boldsymbol{F}_j^k := \boldsymbol{I}_j^{k^*} \in \mathcal{G}^{k-1,k}$, and
4. the *j-neighbor graph* $\boldsymbol{N}_j^k := \boldsymbol{I}_j^k \circ \boldsymbol{F}_j^k \in \mathcal{G}^{k,k}$ for $j \in [k]$, and finally,
5. the *all-one graph* $\boldsymbol{1}^k := (([k], \varnothing), (1, \dots, k), ()) \in \mathcal{G}^{k,0}$.

Let $\mathcal{A}^k := \left\{\boldsymbol{A}_{ij}^k \mid i \neq j \in [k]\right\} \subseteq \mathcal{G}^{k,k}$ and $\mathcal{N}^k := \left\{\boldsymbol{N}_j^k \mid j \in [k]\right\} \subseteq \mathcal{G}^{k,k}$ be the sets of all adjacency graphs and all neighbor graphs, respectively. Finally, let $\mathcal{F}^k := \mathcal{A}^k \cup \mathcal{N}^k$.

The set $\mathcal{F}^k$ of adjacency and neighbor graphs together with $\boldsymbol{1}^k$ suffices to construct essentially every graph of treewidth at most k. Let us first illustrate this with an example: Consider the tree-decomposed graph in Figure 6, which can be translated to the language of bi-labeled graphs as the expression

$$(\boldsymbol{N}_3^3 \circ \boldsymbol{A}_{12}^3 \circ \boldsymbol{A}_{13}^3 \circ \boldsymbol{N}_1^3 \circ \boldsymbol{A}_{12}^3 \circ \boldsymbol{A}_{13}^3 \circ \boldsymbol{1}^3) \cdot (\boldsymbol{N}_1^3 \circ \boldsymbol{A}_{13}^3 \circ \boldsymbol{A}_{23}^3 \circ \boldsymbol{N}_3^3 \circ \boldsymbol{A}_{13}^3 \circ \boldsymbol{A}_{23}^3 \circ \boldsymbol{1}^3). \quad (4)$$

Consider the left subtree and the subexpression left of the Schur product. The all-one graph $\boldsymbol{1}^3$ corresponds to the leaf containing v_6, v_2, and v_4 in the left subtree, the adjacency graphs $\boldsymbol{A}_{12}^3$ and $\boldsymbol{A}_{13}^3$ insert the edges v_6v_2 and v_6v_4, respectively, and the neighbor graph $\boldsymbol{N}_1^3$ moves to the new bag containing v_1, v_2, and v_4 as it replaces the first vertex v_6 by v_1. Then, the adjacency graphs $\boldsymbol{A}_{12}^3$ and $\boldsymbol{A}_{13}^3$ insert the edges v_1v_2 and v_1v_4, respectively,

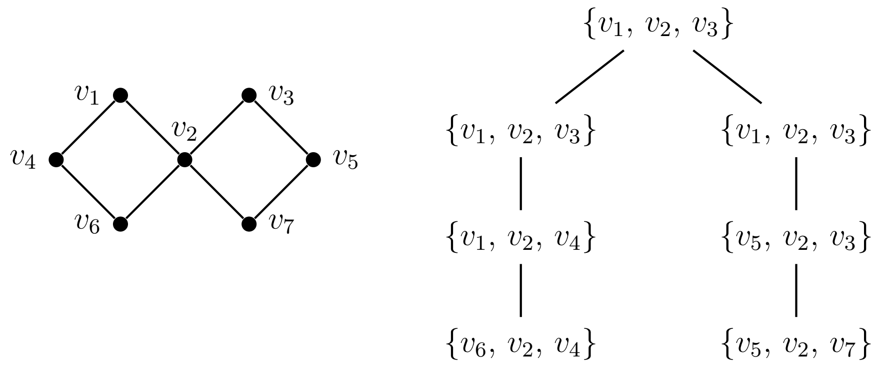

Figure 6: A graph and a tree decomposition of it.

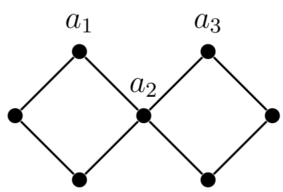

Figure 7: The bi-labeled graph obtained by evaluation of (4).

before the neighbor graph $\boldsymbol{N}_3^3$ moves to the new bag containing v_1, v_2, and v_3 as it replaces the third vertex v_4 by v_1. The right subexpression is constructed analogously from the right subtree, and finally, the Schur product corresponds to the join node of the tree decomposition and glues the two bi-labeled graphs obtained from the two subexpressions together. Then, evaluating the expression in Equation (4) yields the bi-labeled graph in Figure 7, i.e., its underlying graph is the graph in Figure 6, its input labels are v_1, v_2, and v_3, and it has no output labels. The expression in (4) is an example of a *term*, and we can formalize this view of tree-decomposed graphs as expressions built from bi-labeled graphs by composition and the Schur product.

Definition 13. Let $k \geqslant 1$ and $\mathcal{F} \subseteq \mathcal{M}^{k,k}$ be a set of bi-labeled graphs with k input and k output labels. The set $\langle \mathcal{F} \rangle_{\circ,\cdot}$ of *$\mathcal{F}$-terms (terms)* is the smallest set of expressions 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 Conditions (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 $[\![\mathbb{F}]\!] \in \mathcal{M}^{k,0}$ since every term starts with $\mathbf{1}^k$. As demonstrated before, for the specific set $\mathcal{F}^k$ of adjacency and neighbor 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 $[\![\mathbb{F}]\!]$.

Theorem 14. *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. The proof is quite simple and we only sketch it here. The main idea is the following: For a bi-labeled graph $\boldsymbol{G} \in \mathcal{M}^{k,\ell}$, the composition $\boldsymbol{A}^k_{ij} \circ \boldsymbol{G}$ of the ij-adjacency graph $\boldsymbol{A}^k_{ij}$ with $\boldsymbol{G}$ adds an edge between the vertices in $\boldsymbol{G}$ with the ith and the jth input label. When viewing the k input vertices of $\boldsymbol{G}$ as the "current" bag of a tree decomposition, we can compose adjacency graphs with $\boldsymbol{G}$ to add any edge between the vertices of this bag. Similarly, the composition $\boldsymbol{N}^k_j \circ \boldsymbol{G}$ of the j-neighbor graph $\boldsymbol{N}^k_j$ with $\boldsymbol{G}$ moves to a new bag obtained by replacing the vertex with the jth input label by a fresh vertex. In terms of nice tree decompositions, this corresponds to an introduce node that immediately follows on a forget node.

Following the interpretation sketched above, a term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ can inductively be translated into a rooted tree decomposition for the underlying graph of $[\![\mathbb{F}]\!]$ of width $k-1$: For $\mathbf{1}^k$, we take a single leaf containing k fresh vertices. For $\boldsymbol{A}^k_{ij} \circ \mathbb{F}$, we just take the tree decomposition for $\mathbb{F}$. For $\boldsymbol{N}^k_j \circ \mathbb{F}$, we take the tree decomposition for $\mathbb{F}$ and connect a new node to its root and replace the vertex specified by the jth input label in $[\![\mathbb{F}]\!]$ by the fresh vertex introduced by $\boldsymbol{N}^k_j$. Finally, for the Schur product $\mathbb{F}_1 \cdot \mathbb{F}_2$ we take the tree decompositions of $\mathbb{F}_1$ and $\mathbb{F}_2$ and turn them into a single tree decomposition by connecting their roots using a join node.

For the converse direction, we inductively turn a nice tree decomposition (T, r, β) of a graph G of width at most $k-1$ into a term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ such that the underlying graph of $[\![\mathbb{F}]\!]$ is G with some additional isolated vertices. This direction requires a bit more thought since we have to make sure that the set $\mathcal{F}^k$ is not too restrictive. We do this by modifying the nice tree decomposition (T, r, β) as follows, where we have to keep in mind that a term fixes an ordering of the vertices of the graph: 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 directly 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. Then, on this modified tree decomposition, we can apply the inverse construction to the one described in the previous paragraph. $\square$

The *height* $h(\mathbb{F})$ of a term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ is inductively defined by letting $h(\mathbf{1}^k) := 0$, $h(\boldsymbol{N} \circ \mathbb{F}) := h(\mathbb{F}) + 1$ for all $\boldsymbol{N} \in \mathcal{N}^k$, $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$, $h(\boldsymbol{A} \circ \mathbb{F}) := h(\mathbb{F})$ for all $\boldsymbol{A} \in \mathcal{A}^k$, $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$, and $h(\mathbb{F}_1 \cdot \mathbb{F}_2) := \max \{h(\mathbb{F}_1), h(\mathbb{F}_2)\}$ 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.

$$\boldsymbol{P}_\pi\colon \quad \begin{matrix} a_{\pi(1)} & & a_{\pi(k)} \\ \bullet & \cdots & \bullet \\ b_1 & & b_k \end{matrix} \quad = \quad \begin{matrix} a_1 & & a_k \\ \bullet & \cdots & \bullet \\ b_{\pi^{-1}(1)} & & b_{\pi^{-1}(k)} \end{matrix}$$

Figure 8: Two representations of the graph $\boldsymbol{P}_\pi$. In the first representation, the vertices are sorted by their output labels, and in the second, by their their input labels.

Remark 15. In terms of nice tree decompositions, a j-neighbor graph corresponds to an introduce node that immediately follows on a forget node. This is also nicely reflected in the definition of $\boldsymbol{N}_j^k$: it is a j-introduce graph composed with a j-forget graph, which correspond to an introduce node and a forget node, respectively. Theorem 14 and its proof would have been simplified if we included individual j-introduce and j-forget graphs, in $\mathcal{F}^k$: we would not have to deal with isolated vertices. However, just considering j-neighbor graphs instead has the advantage that all bi-labeled graphs in $\mathcal{F}^k$ have both k input and k output labels, which means that we can restrict ourselves to the single product space $L^2(X^k, \mu^{\otimes k})$ later on instead of having to deal with all product spaces $L^2(X^1, \mu^{\otimes 1}), \dots, L^2(X^k, \mu^{\otimes k})$. For this very same reason, we use the all-one graph $\mathbf{1}^k$ instead of defining a *leaf graph* as an empty graph and then using k individual introduce graphs, each introducing a single vertex, to obtain a bag of size k. The downside of these restrictions is that we may have to add isolated vertices to a graph, cf. the statement and proof of Theorem 14. Since additional isolated vertices do not affect the homomorphism density of a graph in a graphon, this is perfectly fine for us. Moreover, it is also not a restriction that, in a j-neighbor graph, both the forgotten and introduced vertex use the jth label since we may just inductively re-order the vertices of the whole term afterwards to make sure that the newly introduced vertex has the desired label; this is also done in the proof of Theorem 14.

In the proof of Theorem 14, we use that we can re-label input vertices by inductively re-labeling a whole term. This could have been simplified by including *permutation graphs* in $\mathcal{F}^k$: for $k \geqslant 1$ and a permutation $\pi\colon [k] \to [k]$, we define the *permutation graph*

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

Moreover, for a tuple $\boldsymbol{a} \in V(F)^k$ of vertices of a graph F, let $\pi(\boldsymbol{a}) \coloneqq (a_{\pi(1)}, \dots, a_{\pi(k)})$. 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 $\pi\colon [\ell] \to [\ell]$. In order to keep the set $\mathcal{F}^k$ simple, we do not include permutations graphs in it. Nevertheless, they come in handy later when proving that the operators and sub-σ-algebras we define are permutation invariant.

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 $t(\boldsymbol{F}, W) \coloneqq t(F, W)$ 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

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

for all $x_{a_1},\dots,x_{a_k}\in X$. We note that we again slightly abuse notation and assume that each factor $W(x_i,x_j)$ occurs as often in the product $\prod_{ij\in E(F)} W(x_i,x_j)$ as ij is contained in $E(F)$. The Tonelli-Fubini theorem immediately yields that

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

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 $T_{\boldsymbol{F}\to W}f$. 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

$$(T_{\boldsymbol{F}\to W}f)(x_{a_1},\dots,x_{a_k}) \coloneqq \int\limits_{X^{V(F)\setminus \boldsymbol{a}}} \prod_{ij\in E(F)} W(x_i,x_j)\cdot f(x_{b_1},\dots,x_{b_\ell})\, d\mu^{\otimes V(F)\setminus \boldsymbol{a}}(\bar{x}) \tag{6}$$

for every $f\in L^2(X^\ell,\mu^{\otimes \ell})$ and all $x_{a_1},\dots,x_{a_k}\in X$, where we again slightly abuse notation w.r.t. the product over all $ij\in E(F)$. The goal of this definition is that an application of $T_{\boldsymbol{F}\to W}$ to a homomorphism function $f_{\boldsymbol{G}\to W}$ should yield the homomorphism function $f_{\boldsymbol{F}\circ\boldsymbol{G}\to W}$; as we will see in Lemma 19, this is the case. We 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 definition of $T_{\boldsymbol{F}\to W}$ only depends on the isomorphism type of $\boldsymbol{F}$, i.e., isomorphic bi-labeled graphs define the same operator.

Lemma 16. *Let $\boldsymbol{F} = (F,\boldsymbol{a},\boldsymbol{b}), \boldsymbol{F'} = (F',\boldsymbol{a'},\boldsymbol{b'}) \in \mathcal{M}^{k,\ell}$ be isomorphic bi-labeled graphs and $W\colon X\times X\to[0,1]$ be a graphon. Then, $T_{\boldsymbol{F}\to W} = T_{\boldsymbol{F'}\to W}$.*

Proof. Let $\varphi\colon V(F)\to V(F')$ be an isomorphism from F to F' such that $\varphi(\boldsymbol{a}) = \boldsymbol{a'}$ and $\varphi(\boldsymbol{b}) = \boldsymbol{b'}$. Note that $\prod_{ij\in E(F')} W(x_i,x_j) = \prod_{ij\in E(F)} W(x_{\varphi(i)},x_{\varphi(j)})$ for $\mu^{\otimes V(F')}$-almost every $\bar{x}\in X^{V(F')}$. Hence, by substituting variable $x_{\varphi(i)}$ by x_i for every $i\in V(F)$ in (6), we immediately get $T_{\boldsymbol{F}\to W} = T_{\boldsymbol{F'}\to W}$. $\square$

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 17. 1. Define $\boldsymbol{A} \coloneqq (([2], \{12\}), (1), (2)) \in \mathcal{G}^{1,1}$ to be the edge with one input and one output vertex. Let $W\colon X \times X \to [0,1]$ be a graphon and $f \in L^2(X, \mu)$. Then,

$$(T_{\boldsymbol{A}\to W} f)(x_1) = \int_X W(x_1, x_2) \cdot f(x_2)\, d\mu(x_2) = (T_W f)(x_1)$$

for every $x_1 \in X$, i.e., $T_{\boldsymbol{A}\to W} = T_W$.

2. Define $\boldsymbol{D} \coloneqq (([2], \{12\}), (1), (1)) \in \mathcal{G}^{1,1}$ to be the edge where one vertex has both an input and an output label. Let $W\colon X \times X \to [0,1]$ be a graphon and $f \in L^2(X, \mu)$. Then,

$$(T_{\boldsymbol{D}\to W} f)(x_1) = \int_X W(x_1, x_2) \cdot f(x_1)\, d\mu(x_2) = f(x_1) \cdot \deg_W(x_1)$$

for every $x_1 \in X$, where $\deg_W(x) \coloneqq \int_X W(x,y)\, d\mu(y)$ for every $x \in X$.

3. Let $k \geqslant 1$ and $\pi\colon [k] \to [k]$ be a permutation. Since $\boldsymbol{P}_\pi$ does not have any edges, $T_{\boldsymbol{P}_\pi \to W}$ is independent of a specific graphon $W\colon X \times X \to [0,1]$ and we simply denote it by $T_{\boldsymbol{P}_\pi}$ as agreed on before. The operator $T_{\boldsymbol{P}_\pi}$ is equal to the Koopman operator T_π of the measure-preserving measurable map $X^k \to X^k$ induced by π: Let $f \in L^2(X^k, \mu^{\otimes k})$. Then,

$$(T_{\boldsymbol{P}_\pi} f)(x_1, \dots, x_k) = f(x_{\pi(1)}, \dots, x_{\pi(k)}) = (f \circ \pi)(x_1, \dots, x_k) = (T_\pi f)(x_1, \dots, x_k)$$

for all $x_1, \dots, x_k \in X$.

The Tonelli-Fubini theorem and the Cauchy-Schwarz inequality allow us to verify that Equation (6) is indeed is a well-defined operator and, furthermore, a contraction, where it is crucial that we made the somewhat unusual assumption that, in a bi-labeled graph, every vertex has at most one label of each type. The intuitive reason for this is that the diagonal $(X^k, \mathcal{B}^{\otimes k}, \mu^{\otimes k})$ has measure zero (as long as our standard Borel space is atom free), a problem which one does not face in the case of (finite-dimensional) matrices.

Lemma 18. *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 well-defined and an L^2- and L^∞-contraction.*

Proof. Let $\boldsymbol{F} = (F, \boldsymbol{a}, \boldsymbol{b})$. For $f \in \mathcal{L}^2(X^\ell, \mu^{\otimes \ell})$, define S by

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

for every $\bar{x} \in X^{V(F)}$. Then, S is clearly linear, and we claim that

$$\|Sf\|_2 \leqslant \|W\|_\infty^{\mathsf{e}(F)} \cdot \|f\|_2 < \infty,$$

which not only implies that $Sf \in \mathcal{L}^2(X^{V(F)}, \mu^{\otimes V(F)})$ but also that S is a well-defined mapping $L^2(X^\ell, \mu^{\otimes \ell}) \to L^2(X^{V(F)}, \mu^{\otimes V(F)})$ that additionally is an L^2-contraction.

To prove the claim, we first note that it is easy to see that $\bar{x} \mapsto \prod_{ij \in E(F)} W(x_i, x_j)$ is a function in $\mathcal{L}^\infty(X^{V(F)}, \mu^{\otimes V(F)})$ with $\|\cdot\|_\infty$-norm at most $\|W\|_\infty$ since $\boldsymbol{F}$ does not have loops, i.e., $i \neq j$. Second, $\bar{x} \mapsto f(x_{b_1}, \dots, x_{b_\ell})$ is a function in $\mathcal{L}^2(X^{V(F)}, \mu^{\otimes V(F)})$ and, by the Tonelli-Fubini theorem, its $\|\cdot\|_2$-norm is $\|f\|_2$, where it is important that the entries of $\boldsymbol{b}$ are pairwise distinct. Then, the claim easily follows. Similarly, for $f \in \mathcal{L}^\infty(X^\ell, \mu^{\otimes \ell})$, we have $\|Sf\|_\infty \leqslant \|W\|_\infty^{\mathsf{e}(F)} \cdot \|f\|_\infty$, i.e., S is also an L^∞-contraction.

Now, $T_{\boldsymbol{F}\to W} f$ is the function obtained from Sf by integrating out the variables x_i for $i \in V(F) \setminus \boldsymbol{a}$. The Cauchy-Schwarz inequality together with the Tonelli-Fubini Theorem yields that this is an L^2-contraction, i.e., $\|T_{\boldsymbol{F}\to W} f\|_2 \leqslant \|Sf\|_2$. Moreover, this is trivially an L^∞-contraction, i.e., $\|T_{\boldsymbol{F}\to W} f\|_\infty \leqslant \|Sf\|_\infty$ for $f \in \mathcal{L}^\infty(X^\ell, \mu^{\otimes \ell})$. Since this operation is also linear, this finishes the proof. $\square$

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 19. *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 $\boldsymbol{F_1} \in \mathcal{M}^{k,m}$, $\boldsymbol{F_2} \in \mathcal{M}^{m,\ell}$, $k, \ell, m \geqslant 0$, 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 $c_1, c_2 \in \mathbb{R}$, $\boldsymbol{F_1}, \boldsymbol{F_2} \in \mathcal{M}^{k,0}$, $k \geqslant 0$.*

Proof. (1): We have

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

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

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

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): Let $\boldsymbol{F_1} = (F_1, \boldsymbol{a_1}, ())$, $\boldsymbol{F_2} = (F_2, \boldsymbol{a_2}, ())$, and $\boldsymbol{F_1} \cdot \boldsymbol{F_2} = (F, \boldsymbol{a_1}, ())$. In the following, we identify vertices $a_{1,1}, \dots, a_{1,k}$ with $a_{2,1}, \dots, a_{2,k}$. Then, we have

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

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. $\square$

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 18. 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 $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ 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

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

that replaces the single operator T_W in Theorem 5, our characterization of oblivious k-WL.

Let us explore the connection between the family $\mathbb{T}^k_W$ and treewidth $k-1$ homomorphism functions: Recall that the terms in $\langle \mathcal{F}^k \rangle_{\circ,\cdot}$ correspond to the tree-decomposed multigraphs of treewidth at most $k-1$ by Theorem 14. Given such a term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$, we can use the correspondence of bi-labeled graph operations to their operator counterparts, cf. Lemma 19, to inductively compute the homomorphism function $f_{[\![\mathbb{F}]\!]\to W}$ of $[\![\mathbb{F}]\!]$ in a graphon W using the operators $\mathbb{T}^k_W$. Hence, the operators in $\mathbb{T}^k_W$ yield all homomorphism functions of multigraphs of treewidth at most $k-1$ in W. An important part of the proof of Theorem 5 consists of defining different families of L^∞-contractions indexed by $\mathcal{F}^k$ that we may use instead of $\mathbb{T}^k_W$ and still yield the same homomorphism functions. For example, we may replace $\mathbb{T}^k_W$ by the quotient operators $\mathbb{T}^k_W/\mathcal{C}$ for an appropriate $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$. This leads to the following general definition, where we consider families of operators on a space $L^2(X, \mu)$ where $(X, \mathcal{B}, \mu)$ may not necessarily be a product space. For example, it could be a quotient space.

Definition 20. Let $(Y, \mathcal{D})$ be a standard Borel space and ν be a Borel probability measure on Y. Let $k \geqslant 1$ and $\mathbb{T} = (T_{\boldsymbol{F}})_{\boldsymbol{F} \in \mathcal{F}}$ be a family of L^∞-contractions on $L^2(Y, \nu)$ 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 $f_{\mathbb{F} \to \mathbb{T}} \in L^\infty(Y, \nu)$ with $\|f_{\mathbb{F} \to \mathbb{T}}\|_\infty \leqslant 1$ defined inductively by

1. $f_{\mathbb{F} \to \mathbb{T}} := \mathbf{1}_Y$ for $\mathbb{F} = \mathbf{1}^k$,
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 $t(\mathbb{F}, \mathbb{T}) := \langle \mathbf{1}_Y, f_{\mathbb{F} \to \mathbb{T}} \rangle$.

As remarked above, given a term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$, 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 $\mathbb{T}^k_W$.

Lemma 21. *Let $k \geqslant 1$. 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. $\mathbb{T}^k_W$ is a family of L^∞-contractions on $L^2(X^k, \mu^{\otimes k})$. 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,

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

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 20 and $f_{[\![\mathbb{F}]\!] \to W} = T_{\mathbf{1}^k \to W} \mathbf{1}_{X^0} = \mathbf{1}_{X^k}$ 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

$$\begin{aligned} f_{\mathbb{F} \to \mathbb{T}^k_W} = T_{\boldsymbol{F} \to W} f_{\mathbb{F}' \to \mathbb{T}^k_W} = T_{\boldsymbol{F} \to W} f_{[\![\mathbb{F}']\!] \to W} &= T_{\boldsymbol{F} \to W} (T_{[\![\mathbb{F}']\!] \to W} \mathbf{1}_{X^0}) \\ &= T_{\boldsymbol{F} \circ [\![\mathbb{F}']\!] \to W} \mathbf{1}_{X^0} \\ &= T_{[\![\mathbb{F}]\!] \to W} \mathbf{1}_{X^0} \\ &= f_{[\![\mathbb{F}]\!] \to W} \end{aligned}$$

by Definition 20, the induction hypothesis, the definition of $T_{[\![\mathbb{F}]\!] \to W}$, and Lemma 19 (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 20, the induction hypothesis, the definition of $T_{[\![\mathbb{F}]\!] \to W}$, and Lemma 19 (4). $\square$

The following lemma gives a sufficient condition under which two families of L^∞-contractions yield the same homomorphism densities. 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. [10, 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 22. *Let $k \geqslant 1$. 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

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

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

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

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

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

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 $f, g \in L^\infty(X_2, \mu_2)$ [10, Theorem 13.9]. We have

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

by the induction hypothesis. $\square$

An important application of Lemma 22 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 $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ *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 $\boldsymbol{F} \in \mathcal{F}^k$, which in turn means that $T_{\boldsymbol{F}\to W}(L^2(X^k, \mathcal{C}, \mu^{\otimes k})) \subseteq L^2(X^k, \mathcal{C}, \mu^{\otimes k})$ for every $\boldsymbol{F} \in \mathcal{F}^k$.

Corollary 23. *Let $k \geqslant 1$. 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 equality is just Lemma 21. By Lemma 9 (3) 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 $I_{\mathcal{C}} \circ \mathbb{T}_W^k/\mathcal{C} = \mathbb{T}_W^k \circ I_{\mathcal{C}}$, respectively, where $I_{\mathcal{C}}$ is a Markov embedding by Proposition 7 (5), Therefore, Lemma 22 yields the first two equalities. $\square$

4 Weisfeiler-Leman and Graphons

In Section 4.1 to Section 4.5, we closely follow Grebík and Rocha [11] to prove Theorem 5 and formally define all notions appearing in it. We start off by defining 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 via the family of operators $\mathbb{T}_W^k$. in Section 4.1. Then, in Section 4.2, we 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 [11] by a larger margin: 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 defined by a graphon (intuitively, these graphons would have to be non-zero only on the diagonal, which has measure zero), 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 $\mathbb{T}_W^k$. In Section 4.5, we define the set $\mathcal{T}^k$ of *homomorphism functions* on $\mathbb{M}^k$ and use the Stone-Weierstrass Theorem to show that it is dense in $C(\mathbb{M}^k)$ before we finally prove Theorem 5 in Section 4.6. The remaining sections discuss some implications of Theorem 5: Section 4.7 shows that one can combine all k-WL distributions $\nu_W^1, \nu_W^2, \dots$ of a graphon W into a single distribution to obtain a new characterization of weak isomorphism. Section 4.8 explains how the characterization of Theorem 5 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 sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}, \mu)$, define

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

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 operators from 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 1.4: 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 24. Let $k \geqslant 1$ 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 $\mathcal{C}_{W,0}^k := \mathbb{T}_{\mathcal{A}^k \to W}(\langle \{\varnothing, X^k\} \rangle)$, $\mathcal{C}_{W,n+1}^k := \mathbb{T}_{\mathcal{N}^k}(\mathcal{C}_{W,n}^k)$ for every $n \in \mathbb{N}$, and $\mathcal{C}_W^k := \mathcal{C}_{W,\infty}^k := \langle \bigcup_{n \in \mathbb{N}} \mathcal{C}_{W,n}^k \rangle$.

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 [11, Proposition 5.13]. A difference is given by the operators in $\mathbb{T}_{\mathcal{A}^k \to W}$, which are *multiplication operators*, i.e., they multiply their arguments with a fixed function. This 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 $\mathcal{C}^k_W$ is permutation invariant, i.e., $\mathcal{C}^k_W$ is T_π-invariant for every permutation $\pi\colon [k] \to [k]$.

Lemma 25. *Let $k \geqslant 1$ and $W\colon X \times X \to [0,1]$ be a graphon. Then,*

1. $\mathcal{C}^k_{W,0} = \left\langle \bigcup_{\boldsymbol{A} \in \mathcal{A}^k} \{(T_{\boldsymbol{A}\to W}\mathbf{1}_{X^k})^{-1}(A) \mid A \in \mathcal{B}([0,1])\} \right\rangle$,
2. $\mathcal{C}^k_{W,0}$ *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. $\mathcal{C}^k_{W,n+1} = \left\langle \mathcal{C}^k_{W,n} \cup \bigcup_{\boldsymbol{N}\in\mathcal{N}^k} \left\{(T_{\boldsymbol{N}}\mathbf{1}_A)^{-1}(B) \mid A \in \mathcal{C}^k_{W,n}, B \in \mathcal{B}([0,1])\right\} \right\rangle$ *for every* $n \in \mathbb{N}$,
4. $\mathcal{C}^k_{W,n}$ *is* $\mathbb{T}_{\mathcal{A}^k \to W}$*-invariant for every* $n \in \mathbb{N} \cup \{\infty\}$,
5. $\mathcal{C}^k_W$ *is the minimum* W*-invariant* $\mu^{\otimes k}$*-relatively complete sub-*σ*-algebra of* $\mathcal{B}^{\otimes k}$*, and*
6. $\mathcal{C}^k_{W,n}$ *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 let $\mathcal{D}$ denote the $\mu^{\otimes k}$-relatively complete sub-σ-algebra of $\mathcal{B}^{\otimes k}$ on the right-hand side of the equality in (1). We prove that $\mathcal{C} = \mathcal{D} = \mathcal{C}^k_{W,0}$. 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 $g \in L^2(X^k, \mu^{\otimes k})$, the 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 multiplication 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 $\mathcal{C}^k_{W,0}$. We have $\left\langle \left\{\varnothing, X^k\right\} \right\rangle \subseteq \mathcal{C}$ and, hence, $\mathcal{C}^k_{W,0} = \mathbb{T}_{\mathcal{A}^k \to W}(\left\langle \left\{\varnothing, X^k\right\} \right\rangle) \subseteq \mathbb{T}_{\mathcal{A}^k \to W}(\mathcal{C}) \subseteq \mathcal{C}$. On the other hand, for every $\boldsymbol{A} \in \mathcal{A}^k$, the function $T_{\boldsymbol{A}\to W}\mathbf{1}_{X^k}$ is $\mathcal{C}^k_{W,0}$-measurable. Hence, $\mathcal{D} \subseteq \mathcal{C}^k_{W,0}$.

(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 yield 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 a multiplication operator, cf. the proof of (1) and (2).

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

It remains to prove that $\mathcal{C}_W^k$ is $\mathbb{T}_W^k$-invariant. By (4), it suffices to show that that $\mathcal{C}_W^k$ is $T_{\boldsymbol{N}}$-invariant for $\boldsymbol{N} \in \mathcal{N}^k$. This is essentially Proposition 5.13 of [11]: We first show that $T_{\boldsymbol{N}} \mathbf{1}_A \in L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$ for $A \in \mathcal{C}_W^k$. To this end, note that $\bigcup_{n \in \mathbb{N}} \mathcal{C}_{W,n}^k$ is an algebra and the σ-algebra generated by it is $\mathcal{C}_W^k$. Hence, from [8, Theorem 3.1.10], it easily follows that we can approximate every set in $\mathcal{C}_W^k$ by a set in $\bigcup_{n \in \mathbb{N}} \mathcal{C}_{W,n}^k$ w.r.t. the measure of their symmetric difference. This implies that, for every $A \in \mathcal{C}_W^k$, there is a sequence $(A_n)_{n \in \mathbb{N}}$ with $A_n \in \mathcal{C}_{W,n}^k$ 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 $T_{\boldsymbol{N}} \mathbf{1}_{A_n} \in L^2(X^k, \mathcal{C}_{W,n+1}^k, \mu^{\otimes k}) \subseteq L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$, which is a closed subspace by Proposition 6. Hence, $T_{\boldsymbol{N}} \mathbf{1}_A \in L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$. Since the linear hull of $\{\mathbf{1}_A\}_{A \in \mathcal{C}_W^k}$ is dense in the closed subspace $L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$, linearity and continuity of $T_{\boldsymbol{N}}$ then yields that $L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$ 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, $\langle \left\{ \varnothing, X^k \right\} \rangle$ is permutation invariant, and it suffices to show that, for a permutation-invariant sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$, 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 $\mathcal{C}_{W,n}^k$ is permutation invariant for every $n \in \mathbb{N}$ and, hence, also $\mathcal{C}_W^k$ since $\pi \left(\bigcup_{n \in \mathbb{N}} \mathcal{C}_{W,n}^k \right) = \bigcup_{n \in \mathbb{N}} \pi(\mathcal{C}_{W,n}^k) \subseteq \bigcup_{n \in \mathbb{N}} \mathcal{C}_{W,n}^k$ for every permutation $\pi \colon [k] \to [k]$.

It remains to show that, for a permutation-invariant sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$, 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 sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$, we have

$$\pi(\mathbb{T}_{\mathcal{A}^k \to W}(\mathcal{C})) = \mathbb{T}_{\mathcal{A}^k \to W}(\pi(\mathcal{C})) \tag{7}$$

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

To prove Equation (7), let $\pi \colon [k] \to [k]$ be a permutation and observe that $T_\pi \circ T_{\boldsymbol{A}_{ij}^k \to W} \circ T_{\pi^{-1}} = T_{\boldsymbol{A}_{\pi(i)\pi(j)}^k \to W}$ 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

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

Let $\mathcal{D} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$. Then, we have

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

As the mapping $\boldsymbol{A}^k_{ij} \mapsto \boldsymbol{A}^k_{\pi(i)\pi(j)}$ 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 (7). $\square$

4.2 Weisfeiler-Leman Measures and Distributions

Before defining the mapping $\mathsf{owl}^k_W \colon X^k \to \mathbb{M}^k$, we have to define the space $\mathbb{M}^k$, which can be seen as the space of all colors used by oblivious k-WL. We 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 $f \in C_b(X)$. This leads to the notion of *weak convergence of measures* of which the *Portmanteau Theorem* gives many equivalent characterizations [18, Theorem 17.20]. We only use that, for $(\mu_i)_{i \in \mathbb{N}}$ with $\mu_i \in \mathscr{P}(X)$ and $\mu \in \mathscr{P}(X)$, we have $\mu_i \to \mu$ if and only if

$$\int f d\mu_i \to \int f d\mu$$

for every $f \in C_b(X)$, where we may replace $C_b(X)$ by a dense subset, i.e., a subset that is dense for $d_{\sup}(f, g) \coloneqq \sup|f - g|$. If $(X, \mathcal{T})$ is compact, which is the case for the spaces we define, then $C_b(X) = C(X)$, where $C(X)$ denotes the set of continuous real-valued functions on X. The Borel σ-algebra $\mathcal{B}(\mathscr{P}(X))$ is 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 [18, Theorem 17.24]. If $(X, \mathcal{T})$ is Polish, then so is $P(X)$ [18, Theorem 17.23], which means that, if $(X, \mathcal{B})$ is a standard Borel space, then so is $(\mathscr{P}(X), \mathcal{B}(\mathscr{P}(X)))$. We note that every compact metrizable space $K = (X, \mathcal{T})$ is separable [18, Proposition 4.6], which means that $(X, \mathcal{B})$ is a standard Borel space, where $\mathcal{B}$ is the Borel σ-algebra generated by $\mathcal{T}$. Additionally, in the case of such a K, the topological space $\mathscr{P}(X)$ is again compact metrizable [18, 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 1.4: Here, $P^k_0 = [0,1]^{\binom{[k]}{2}}$ 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 $\mathbb{M}^k_n$ in the following definition, where we recall that $f_*\mu$ denotes the push-forward of μ via f.

Definition 26 (The Spaces $\mathbb{M}^{\boldsymbol{k}}$ and $\mathbb{P}^{\boldsymbol{k}}$)**.** Let $k \geqslant 1$. Let $P^k_0 \coloneqq [0,1]^{\binom{[k]}{2}}$ and inductively define $\mathbb{M}^k_n \coloneqq \prod_{i \leqslant n} P^k_i$ and $P^k_{n+1} \coloneqq \left(\mathscr{P}\left(\mathbb{M}^k_n\right)\right)^k$ for every $n \in \mathbb{N}$. Let $\mathbb{M}^k \coloneqq \mathbb{M}^k_\infty \coloneqq$

$\prod_{n\in\mathbb{N}} P_i^k$ and, for $n \leqslant m \leqslant \infty$, let $p_{m,n}\colon \mathbb{M}_m^k \to \mathbb{M}_n^k$ be the natural projection, i.e., the restriction to the first n components. Finally, define

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

As a product of a sequence of metrizable compact spaces, $\mathbb{M}^k$ is metrizable [8, Proposition 2.4.4] and also compact by Tychonoff's Theorem [8, 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.

Consider the requirement $(\alpha_{n+1})_j = (p_{n+1,n})_*(\alpha_{n+2})_j$ in the definition of $\mathbb{P}^k$, and note that $\alpha_{n+1} \in P_{n+1}^k = \left(\mathscr{P}\left(\mathbb{M}_n^k\right)\right)^k$ and $\alpha_{n+2} \in P_{n+2}^k = \left(\mathscr{P}\left(\mathbb{M}_{n+1}^k\right)\right)^k$ for $\alpha \in \mathbb{M}^k$, i.e., $\mathbb{P}^k$ is well-defined. This requirement intuitively expresses that α_{n+2}, 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. [7], 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 [18, Exercise 17.16], for all $\alpha \in \mathbb{P}^k$ and $j \in [k]$, there is a unique measure $\mu_j^\alpha \in \mathscr{P}(\mathbb{M}^k)$ such that $(p_{\infty,n})_*\mu_j^\alpha = (\alpha_{n+1})_j$ for every $n \in \mathbb{N}$. Moreover, one can verify that this mapping $\alpha \mapsto \mu_j^\alpha$ is continuous, cf. [11, Claim 6.2].

Lemma 27. *$\mathbb{P}^k$ is closed in $\mathbb{M}^k$ and $\mathbb{P}^k \to \mathscr{P}(\mathbb{M}^k)$, $\alpha \mapsto \mu_j^\alpha$ 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

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

for every $f \in C(\mathbb{M}_n^k)$. 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 $\mathscr{P}(\mathbb{M}_n^k)$ 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 $\mu_j^{\alpha_i} \to \mu_j^{\alpha}$, we observe that

$$\int\limits_{\mathbb{M}^k} f \circ p_{\infty,n}\, d\mu_j^{\alpha_i} = \int\limits_{\mathbb{M}_n^k} f\, d(p_{\infty,n})_* \mu_j^{\alpha_i} = \int\limits_{\mathbb{M}_n^k} f\, d((\alpha_i)_{n+1})_j \to \int\limits_{\mathbb{M}_n^k} f\, d(\alpha_{n+1})_j = \int\limits_{\mathbb{M}_n^k} f\, d(p_{\infty,n})_* \mu_j^{\alpha}$$
$$= \int\limits_{\mathbb{M}^k} f \circ p_{\infty,n}\, d\mu_j^{\alpha}$$

for every $n \in \mathbb{N}$ and every $f \in C(\mathbb{M}_n^k)$. This already proves the claim as the set $\bigcup_{n\in\mathbb{N}} C(\mathbb{M}_n^k) \circ p_{\infty,n}$ is uniformly dense in $C(\mathbb{M}^k)$ by the Stone-Weierstrass Theorem [8, 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 27 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 (k-WLD), which intuitively generalizes the concept of a multiset of colors with the additional constraints that, first, 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 $\mathbb{P}^k \to \mathscr{P}(\mathbb{M}^k)$, $\alpha \mapsto \mu_j^{\alpha}$. This definition only become fully clear in the next subsections: we will show that every graphon W has a natural k-WLD ν_W^k associated with it that satisfies both conditions and that these conditions guarantee the existence of certain operators associated with the k-WLD.

Definition 28. Let $k \geqslant 1$. 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. $\int\limits_{\mathbb{M}^k} f\, d\nu = \int\limits_{\mathbb{M}^k} \Big(\int\limits_{\mathbb{M}^k} f\, d\mu_j^{\alpha} \Big)\, d\nu(\alpha)$ for all bounded measurable $f \colon \mathbb{M}^k \to \mathbb{R}$, $j \in [k]$.

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 not have to 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 $j \in [k]$. 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

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

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

$$(T_{\boldsymbol{N}_j^k} f)(\bar{x}) = \int_X f(\bar{x}[y/j]) \, d\mu(y) = \int_X f \circ \bar{x}[\cdot/j] \, d\mu = \int_{X^k} f d \left(\bar{x}[\cdot/j]_* \mu\right) \tag{8}$$

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

Definition 29 (The Mapping owl_W^k)**.** Let $k \geqslant 1$ 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

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

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 29 is the following lemma. 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 29 to be well defined.

Lemma 30. *Let $k \geqslant 1$ and $W \colon X \times X \to [0,1]$ be a graphon. Then, ${\mathsf{owl}_{W,m}^k}^{-1}(p_{m,n}^{-1}(A)) = {\mathsf{owl}_{W,n}^k}^{-1}(A)$ for all $1 \leqslant n < m \leqslant \infty$ and every $A \in \mathcal{B}(\mathbb{M}_n^k)$.*

Proof. Let $A \in \mathcal{B}(\mathbb{M}_n^k)$. We have

$$\begin{aligned}
{\mathsf{owl}_{W,m}^{k}}^{-1}(p_{m,n}^{-1}(A)) &= \left\{\bar{x} \in X^k \mid \left(\left(\mathsf{owl}_{W,m}^{k}(\bar{x})\right)_1, \dots, \left(\mathsf{owl}_{W,m}^{k}(\bar{x})\right)_n\right) \in A\right\} \\
&= \left\{\bar{x} \in X^k \mid \left(\left(\mathsf{owl}_{W,n}^{k}(\bar{x})\right)_1, \dots, \left(\mathsf{owl}_{W,n}^{k}(\bar{x})\right)_n\right) \in A\right\} \\
&= {\mathsf{owl}_{W,n}^{k}}^{-1}(A)
\end{aligned}$$

by definition of $\mathsf{owl}_{W,m}^{k}$ and $\mathsf{owl}_{W,n}^{k}$. □

Lemma 31 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. [11, Proposition 6.6].

Lemma 31. *Let $k \geqslant 1$ 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 $\mathcal{D}_n := \langle \{ {\mathsf{owl}_{W,n}^{k}}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}_n^k)\}\rangle$. 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 [18, Section 10.B]. Since it suffices to check measurability of a function for a generating set [8, 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

$$\begin{aligned}
\mathcal{D}_0 &= \left\langle \left\{ {\mathsf{owl}_{W,0}^{k}}^{-1}(A) \mid A \in \mathcal{B}([0,1]^{\binom{[k]}{2}}) \right\} \right\rangle \\
&= \left\langle \left\{ \left\{ \bar{x} \in X^k \mid (W(x_i, x_j))_{ij \in \binom{[k]}{2}} \in A \right\} \mid A \in \mathcal{B}([0,1]^{\binom{[k]}{2}}) \right\} \right\rangle \\
&= \left\langle \left\{ \left\{ \bar{x} \in X^k \mid W(x_i, x_j) \in A \right\} \mid A \in \mathcal{B}([0,1]),\, ij \in \tbinom{[k]}{2} \right\} \right\rangle \\
&= \left\langle \textstyle\bigcup_{ij \in \binom{[k]}{2}} \left\{ \left\{ \bar{x} \in X^k \mid W(x_i, x_j) \in A \right\} \mid A \in \mathcal{B}([0,1]) \right\} \right\rangle \\
&= \left\langle \textstyle\bigcup_{\boldsymbol{A} \in \mathcal{A}^k} \left\{ (T_{\boldsymbol{A} \to W} \mathbf{1}_{X^k})^{-1}(A) \mid A \in \mathcal{B}([0,1]) \right\} \right\rangle \\
&= \mathcal{C}_{W,0}^k. && \text{(Lemma 25 (1))}
\end{aligned}$$

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 $A \in \mathcal{B}(\mathbb{M}_n^k)$ [18, 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 [8, 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 [8, 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

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

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 (8).

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 25 (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 $A = {\mathsf{owl}^k_{W,n}}^{-1}(B) \triangle Z$ 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

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

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 [8, 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 31, $\mathcal{C}_W^k$ 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 $T_{\mathsf{owl}_W^k} : L^2(\mathbb{M}^k, \nu_W^k) \to L^2(X^k, \mu^{\otimes k})$ of owl_W^k.

Corollary 32. *Let $k \geqslant 1$ and $W \colon X \times X \to [0,1]$ be a graphon. Then,*

1. *the Koopman operator $T_{\mathsf{owl}_W^k} : L^2(\mathbb{M}^k, \nu_W^k) \to L^2(X^k, \mu^{\otimes k})$ of owl_W^k is a Markov embedding onto $L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$,*
2. *the operator $S_{\mathcal{C}_W^k} : L^2(X^k, \mu^{\otimes k}) \to L^2(X^k/\mathcal{C}_W^k, \mu^{\otimes k}/\mathcal{C}_W^k)$ becomes a Markov isomorphism when restricted to $L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$, and*
3. *$R_W^k \coloneqq S_{\mathcal{C}_W^k} \circ T_{\mathsf{owl}_W^k} : L^2(\mathbb{M}^k, \nu_W^k) \to L^2(X^k/\mathcal{C}_W^k, \mu^{\otimes k}/\mathcal{C}_W^k)$ is a Markov isomorphism.*

Proof. (1): Consider the measure spaces $(X^k, \mathcal{B}^{\otimes k}, \mu^{\otimes k})$ and $(\mathbb{M}^k, \mathcal{B}(\mathbb{M}^k), \nu_W^k)$. The mapping $\mathsf{owl}_W^k \colon X^k \to \mathbb{M}^k$ is measurable by Lemma 31 and measure-preserving by definition of ν_W^k. Hence, its Koopman operator $T_{\mathsf{owl}_W^k} : L^2(\mathbb{M}^k, \nu_W^k) \to L^2(X^k, \mu^{\otimes k})$ is a Markov embedding [10, Theorem 7.20]. By Proposition 8, it is an isometry onto $L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$.

(2): By Proposition 7 (5), the operator $S_{\mathcal{C}_W^k} : L^2(X^k, \mu^{\otimes k}) \to L^2(X^k/\mathcal{C}_W^k, \mu^{\otimes k}/\mathcal{C}_W^k)$ becomes a Markov isomorphism when restricted to $L^2(X^k, \mathcal{C}_W^k, \mu^{\otimes k})$.

(3): Follows immediately from (1) and (2). □

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 28 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 33. *Let $k \geqslant 1$ and $W\colon X \times X \to [0,1]$ be a graphon. Then,*

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

Proof. (1): Let $\bar{x} \in X^k$. For $j \in [k]$ and $n \in \mathbb{N}$, we have

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

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

(2): 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 $\bigcup_{n\in\mathbb{N}} \left\{p_{\infty,n}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}_n^k)\right\}$, which generates $\mathcal{B}(\mathbb{M}^k)$. By the π-λ Theorem [18, Theorem 10.1 iii)], they agree on all of $\mathcal{B}(\mathbb{M}^k)$.

(3): By (1), we have $\nu_W^k(\mathbb{P}^k) = \mu^{\otimes k}({\mathsf{owl}_W^k}^{-1}(\mathbb{P}^k)) = \mu^{\otimes k}(X^k) = 1$. Let $j \in [k]$. Let

$f\colon \mathbb{M}^k \to \mathbb{R}$ be bounded and measurable. We have

$$\begin{aligned}
\int_{\mathbb{M}^k} f \, d\nu_W^k \overset{\text{def. } \nu_W^k}{=} \int_{\mathbb{M}^k} f \, d\mathsf{owl}^k_{W*}\mu^{\otimes k} = \int_{X^k} f \circ \mathsf{owl}^k_W \; d\mu^{\otimes k} \\
&\overset{\text{T.-F.}}{=} \int_{X^{k-1}} \left(\int_X f \circ \mathsf{owl}^k_W(\bar{x}[y/j]) \, d\mu(y) \right) d\mu^{\otimes k-1}(\bar{x}[/j]) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (x_j \in X \text{ arb.}) \\
&= \int_{X^k} \left(\int_X f \circ \mathsf{owl}^k_W(\bar{x}[y/j]) \, d\mu(y) \right) d\mu^{\otimes k}(\bar{x}) \\
&= \int_{X^k} \left(\int_{\mathbb{M}^k} f \, d(\mathsf{owl}^k_W \circ \bar{x}[\cdot/j])_*\mu \right) d\mu^{\otimes k}(\bar{x}) \\
&\overset{(2)}{=} \int_{X^k} \left(\int_{\mathbb{M}^k} f \, d\mu_j^{\mathsf{owl}^k_W(\bar{x})} \right) d\mu^{\otimes k}(\bar{x}) \\
&\overset{\text{push-f.}}{=} \int_{\mathbb{M}^k} \left(\int_{\mathbb{M}^k} f \, d\mu_j^{\alpha} \right) d\mathsf{owl}^k_{W*}\mu^{\otimes k}(\alpha) \\
&\overset{\text{def. } \nu_W^k}{=} \int_{\mathbb{M}^k} \left(\int_{\mathbb{M}^k} f \, d\mu_j^{\alpha} \right) d\nu_W^k(\alpha).
\end{aligned}$$

□

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 32. 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 [11]. 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 34. *Let $k \geqslant 1$, and let $\nu \in \mathscr{P}(\mathbb{M}^k)$ be a k-WLD. Let $j \in [k]$. Setting*

$$(T_{\boldsymbol{N}_j^k \to \nu} f)(\alpha) := \int_{\mathbb{M}^k} f \, d\mu_j^\alpha$$

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 $\|T_{\boldsymbol{N}_j^k \to \nu} f\|_2 \leqslant \|f\|_2$ 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, $|f| \leqslant \|f\|_\infty$ ν-almost everywhere, and hence, $|f| \leqslant \|f\|_\infty$ 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

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

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

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

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 27 and the definition of $\mathscr{P}(\mathbb{P}^k)$. Since $\mathbb{P}^k \in \mathcal{B}(\mathbb{M}^k)$ by Lemma 27 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 $\|T_{\boldsymbol{N}_j^k \to \nu} f\|_2 \leqslant \|f\|_2$ for every $f \in L^\infty(\mathbb{M}^k, \nu)$. We have

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

Note that we again used that $|f| \leqslant \|f\|_\infty$ 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 34 is indeed the right definition.

Lemma 35. *Let $k \geqslant 1$ and $W \colon X \times X \to [0, 1]$ be a graphon. For every $\boldsymbol{N} \in \mathcal{N}^k$,*

1. $T_{\boldsymbol{N}} \circ T_{\mathsf{owl}^k_W} = T_{\mathsf{owl}^k_W} \circ T_{\boldsymbol{N} \to \nu^k_W}$,

2. $(T_{\boldsymbol{N}})_{\mathcal{C}^k_W} \circ T_{\mathsf{owl}^k_W} = T_{\mathsf{owl}^k_W} \circ T_{\boldsymbol{N} \to \nu^k_W}$, *and* 3. $T_{\boldsymbol{N}}/\mathcal{C}^k_W \circ R^k_W = R^k_W \circ T_{\boldsymbol{N} \to \nu^k_W}$.

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

$$\begin{aligned}
(T_{\boldsymbol{N}^k_j} \circ T_{\mathsf{owl}^k_W} f)(\bar{x}) = (T_{\boldsymbol{N}^k_j} f \circ \mathsf{owl}^k_W)(\bar{x}) &= \int_X f \circ \mathsf{owl}^k_W(\bar{x}[y/j])\, d\mu(y) && \text{(def.)}\\
&= \int_{\mathbb{M}^k} f\, d(\mathsf{owl}^k_W \circ \bar{x}[\cdot/j])_* \mu \\
&= \int_{\mathbb{M}^k} f\, d\mu_j^{\mathsf{owl}^k_W(\bar{x})} && \text{(Lemma 33 (2))}\\
&= (T_{\boldsymbol{N}^k_j \to \nu^k_W} f)(\mathsf{owl}^k_W(\bar{x})) && \text{(def. } T_{\boldsymbol{N}^k_j \to \nu^k_W})\\
&= (T_{\mathsf{owl}^k_W} \circ T_{\boldsymbol{N}^k_j \to \nu^k_W} f)(\bar{x}) && \text{(def.)}
\end{aligned}$$

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

(2): We have

$$\begin{aligned}
(T_{\boldsymbol{N}})_{\mathcal{C}^k_W} \circ T_{\mathsf{owl}^k_W} &= T_{\boldsymbol{N}} \circ \mathbb{E}(- \mid \mathcal{C}^k_W) \circ T_{\mathsf{owl}^k_W} && \text{(Lemma 9 (5) and Lemma 25 (5))}\\
&= T_{\boldsymbol{N}} \circ T_{\mathsf{owl}^k_W} && \text{(cf. Corollary 32)}\\
&= T_{\mathsf{owl}^k_W} \circ T_{\boldsymbol{N} \to \nu^k_W}. && \text{((1))}
\end{aligned}$$

(3): We have

$$\begin{aligned}
T_{\boldsymbol{N}}/\mathcal{C}^k_W \circ R^k_W &= S_{\mathcal{C}^k_W} \circ T_{\boldsymbol{N}} \circ I_{\mathcal{C}^k_W} \circ S_{\mathcal{C}^k_W} \circ T_{\mathsf{owl}^k_W} && \text{(def.)}\\
&= S_{\mathcal{C}^k_W} \circ \mathbb{E}(- \mid \mathcal{C}^k_W) \circ T_{\boldsymbol{N}} \circ \mathbb{E}(- \mid \mathcal{C}^k_W) \circ T_{\mathsf{owl}^k_W} && \text{(Proposition 7 (4) and (6))}\\
&= S_{\mathcal{C}^k_W} \circ (T_{\boldsymbol{N}})_{\mathcal{C}^k_W} \circ T_{\mathsf{owl}^k_W} && \text{(def.)}\\
&= S_{\mathcal{C}^k_W} \circ T_{\mathsf{owl}^k_W} \circ T_{\boldsymbol{N} \to \nu^k_W} && \text{((2))}\\
&= R^k_W \circ T_{\boldsymbol{N} \to \nu^k_W}. && \text{(def.)}
\end{aligned}$$

□

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

$$(T_{\boldsymbol{A}^k_{ij} \to \nu} f)(\alpha) := (\alpha_0)_{ij} \cdot f(\alpha)$$

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, $T_{\boldsymbol{A}^k_{ij}\to\nu} f$ 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 $\|T_{\boldsymbol{A}^k_{ij}\to\nu} f\|_2 \leqslant \|f\|_2$ for $f \in L^2(\mathbb{M}^k, \nu)$ and $\|T_{\boldsymbol{A}^k_{ij}\to\nu} f\|_\infty \leqslant \|f\|_\infty$ for $f \in L^\infty(\mathbb{M}^k, \nu)$. Moreover, $T_{\boldsymbol{A}^k_{ij}\to\nu}$ is linear as a multiplication operator. □

Analogously to Lemma 35, one can verify that Lemma 36 is in fact the right definition.

Lemma 37. *Let $k \geqslant 1$ and $W\colon X \times X \to [0,1]$ be a graphon. For every $\boldsymbol{A} \in \mathcal{A}^k$,*

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

For a k-WLD $\nu \in \mathscr{P}(\mathbb{M}^k)$, we define the family of L^∞-contractions $\mathbb{T}_\nu := (T_{\boldsymbol{F}\to\nu})_{\boldsymbol{F}\in\mathcal{F}^k}$. Lemma 35 (3) and Lemma 37 (3) can then be rephrased as the following corollary.

Corollary 38. *Let $k \geqslant 1$ 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 20, 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 22 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 39. *Let $k \geqslant 1$. 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, $t(\mathbb{F}, \mathbb{T}_{\nu^k_W}) = t([\![\mathbb{F}]\!], W)$ for every $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$.*

Proof. By Corollary 23, we have $t(\mathbb{F}, \mathbb{T}^k_W/\mathcal{C}^k_W) = t([\![\mathbb{F}]\!], W)$ since $\mathcal{C}^k_W$ is W-invariant by Lemma 25. By Corollary 38, 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 32. Then, Lemma 22 yields $t(\mathbb{F}, \mathbb{T}^k_W/\mathcal{C}^k_W) = t(\mathbb{F}, \mathbb{T}_{\nu^k_W})$. □

The mapping $\mathsf{owl}_W^k \colon X^k \to \mathbb{M}^k$ assigns an element of $\mathbb{M}^k$ to a k-tuple of elements from X. The order of these elements from X in the tuple has no deeper meaning and should, intuitively, not matter. While we already have defined what it means for an operator and a sub-σ-algebra to be permutation invariant, we now also formalize this for k-WLDs, and in particular, show that the k-WLD ν_W^k of a graphon W, which is defined via owl^k, is permutation invariant. This allows us to show that, in Theorem 5, Markov operators and Markov isomorphisms can always be assumed to be permutation invariant and, hence, confirms the intuition that the order of the elements in a k-tuple from X does not matter. This becomes important for the connection to linear equations in Section 4.8, where this assumption of permutation invariance is always implicitly present.

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 $\pi \colon P_0^k \to P_0^k$ 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 $\pi \colon \mathbb{M}_n^k \to \mathbb{M}_n^k$ by component-wise application and, then, to a measurable bijection $\pi \colon P_{n+1}^k \to P_{n+1}^k$ by setting $\pi((\mu_j)_{j\in[k]}) = (\pi_* \mu_{\pi(j)})_{j\in[k]}$ for every $(\mu_j)_{j\in[k]} \in P_{n+1}^k$. Finally, we obtain the measurable bijection $\pi \colon \mathbb{M}_n^k \to \mathbb{M}_n^k$ 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. 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 40 yields that the k-WLD ν_W^k of a graphon W is permutation invariant.

Lemma 40. *Let $k \geqslant 1$ and $W \colon X \times X \to [0,1]$ be a graphon. For every permutation $\pi \colon [k] \to [k]$,*

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

Proof. (1): We prove that $\pi \circ \mathsf{owl}_{W,n}^k = \mathsf{owl}_{W,n}^k \circ \pi$ by induction on $n \in \mathbb{N}$. This yields $(\pi \circ \mathsf{owl}_W^k(\bar{x}))_n = (\mathsf{owl}_W^k \circ \pi(\bar{x}))_n$ 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

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

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

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

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 35, respectively. $\square$

4.5 Homomorphism Functions and Weisfeiler-Leman Measures

For the proof of Theorem 5, Corollary 39 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 [11], 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 $\mathbb{M}^k$. This then means 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 41. Let $k \geqslant 1$. For every term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ and every $n \in \mathbb{N}$ with $n \geqslant h(\mathbb{F})$, 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 $n > m \geqslant h(\mathbb{F})$.

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

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

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 $n \geqslant h(\mathbb{F})$, 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 20. 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 previously defined functions ν-almost everywhere.

Lemma 42. *Let $k \geqslant 1$ 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 $n \geqslant h(\mathbb{F})$. 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

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

for ν-almost every $\alpha \in \mathbb{M}^k$. For the product $f_1 \cdot f_2 \in F_n^{\mathbb{F}_1 \cdot \mathbb{F}_2}$ of $f_1 \in F_n^{\mathbb{F}_1}, f_2 \in F_n^{\mathbb{F}_2}$, 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 $f \circ p_{n,m} \in F_n^{\mathbb{F}}$ for $f \in F_m^{\mathbb{F}}$ and $m \in \mathbb{N}$ with $n > m \geqslant h(\mathbb{F})$. 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 39 yields the following corollary to the previous lemma.

Corollary 43. *Let $k \geqslant 1$ and $W : X \times X \to [0, 1]$ 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*

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

For every $n \in \mathbb{N} \cup \{\infty\}$, define $\mathcal{T}_n^k := \bigcup_{\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}, h(\mathbb{F}) \leqslant n} F_n^{\mathbb{F}}$ and abbreviate $\mathcal{T}^k := \mathcal{T}_\infty^k$. By induction, we can use the Stone-Weierstrass Theorem and the Portmanteau Theorem to show that the Stone-Weierstrass Theorem is actually applicable to all of these sets and, in particular, to $\mathcal{T}^k$, cf. [11, Proposition 7.5].

Lemma 44. *Let $k \geqslant 1$. For every $n \in \mathbb{N} \cup \{\infty\}$, the set $\mathcal{T}_n^k$ is closed under multiplication, contains $\mathbf{1}_{\mathbb{M}_n^k}$, and separates points of $\mathbb{M}_n^k$.*

Proof. First, consider the case that $n \in \mathbb{N}$. We trivially have $\mathbf{1}_{\mathbb{M}_n^k} \in F_n^{\mathbf{1}^k} \subseteq \mathcal{T}_n^k$ by definition. Moreover, for $f_1, f_2 \in \mathcal{T}_n^k$, there are terms $\mathbb{F}_1, \mathbb{F}_2 \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ with $h(\mathbb{F}_1) \leqslant n$ and $h(\mathbb{F}_2) \leqslant n$. such that $f_1 \in F_n^{\mathbb{F}_1}$ and $f_2 \in F_n^{\mathbb{F}_2}$. Then, $f_1 \cdot f_2 \in F_n^{\mathbb{F}_1 \cdot \mathbb{F}_2} \subseteq \mathcal{T}_n^k$ as $h(\mathbb{F}_1 \cdot \mathbb{F}_2) = \max\{h(\mathbb{F}_1), h(\mathbb{F}_2)\} \leqslant n$. We prove that $\mathcal{T}_n^k$ separates points of $\mathbb{M}_n^k$ by induction on n. For the base case $n = 0$, let $\beta \neq \gamma \in \mathbb{M}_0^k$. Then, there is an $ij \in \binom{[k]}{2}$ such that $\beta_{ij} \neq \gamma_{ij}$, and the function $\alpha \mapsto (\alpha_0)_{ij} \in F_0^{\boldsymbol{A}_{ij}^k \circ \mathbf{1}^k}$ separates β and γ.

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

For the remaining case, assume that $\beta_{n+1} \neq \gamma_{n+1}$. Then, there is a $j \in [k]$ such that $(\beta_{n+1})_j \neq (\gamma_{n+1})_j$. By the inductive hypothesis and the Stone-Weierstrass Theorem [8, Theorem 2.4.11], the linear hull of $\mathcal{T}_n^k$ is uniformly dense in $C(\mathbb{M}_n^k)$. Since $\mathbb{M}_n^k$ is Hausdorff, it then follows from the Portmanteau Theorem [18, Theorem 17.20] that there is an $f \in \mathcal{T}_n^k$ such that $\int_{\mathbb{M}_n^k} f \, d(\beta_{n+1})_j \neq \int_{\mathbb{M}_n^k} f \, d(\gamma_{n+1})_j$. By definition, $f \in F_n^{\mathbb{F}}$ for some term $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ with $h(\mathbb{F}) \leqslant n$. Then, $\alpha \mapsto \int_{\mathbb{M}_n^k} f \, d(\alpha_{n+1})_j \in F_{n+1}^{\boldsymbol{N}_j^k \circ \mathbb{F}} \subseteq \mathcal{T}_{n+1}^k$ 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 45. *Let $k \geqslant 1$. 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 44 and the Stone-Weierstrass Theorem [8, Theorem 2.4.11]. Hence, we have

$$\begin{aligned}
\nu^k_{W_n} \to \nu^k_W \iff & \int_{\mathbb{M}^k} f \, d\nu^k_{W_n} \to \int_{\mathbb{M}^k} f \, d\nu^k_W \text{ for every } f \in C(\mathbb{M}^k) \\
& \hfill \text{(Portmanteau Theorem)} \\
\iff & \int_{\mathbb{M}^k} f \, d\nu^k_{W_n} \to \int_{\mathbb{M}^k} f \, d\nu^k_W \text{ for every } f \text{ in the linear hull of } \mathcal{T}^k \\
\iff & \int_{\mathbb{M}^k} f \, d\nu^k_{W_n} \to \int_{\mathbb{M}^k} f \, d\nu^k_W \text{ for every } f \in \mathcal{T}^k \quad \text{(Linearity of the integral)} \\
\iff & t([\![\mathbb{F}]\!], W_n) \to t([\![\mathbb{F}]\!], W) \text{ for every } \mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot} \quad \text{(Corollary 43)} \\
\iff & t(F, W_n) \to t(F, W) \text{ for every multigraph } F \text{ of tw.} \leqslant k-1. \\
& \hfill \text{(Theorem 14)}
\end{aligned}$$

□

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 46. *Let $k \geqslant 1$ 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 $k-1$.*

4.6 Proof of Theorem 5

We are finally ready to prove Theorem 5. The majority of the work is already done and, at this point, it is just about putting all the previous results together. For easier readability, we restate the theorem here.

Theorem 5. *Let $k \geqslant 1$ 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 multigraph of treewidth at most $k-1$.*
2. *$\nu^k_U = \nu^k_W$.*
3. *There is a (permutation-invariant) Markov isomorphism $R\colon L^2(X^k/\mathcal{C}^k_W, \mu^{\otimes k}/\mathcal{C}^k_W) \to L^2(X^k/\mathcal{C}^k_U, \mu^{\otimes k}/\mathcal{C}^k_U)$ 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-invariant) Markov operator* $S\colon L^2(X^k,\mu^{\otimes k}) \to L^2(X^k,\mu^{\otimes k})$ *such that* $\mathbb{T}_U^k \circ S = S \circ \mathbb{T}_W^k$.

5. *There are* $\mu^{\otimes k}$*-relatively complete sub-*σ*-algebras* $\mathcal{C}$ *and* $\mathcal{D}$ *of* $\mathcal{B}^{\otimes k}$ *that are* U*-invariant and* W*-invariant, respectively, and a Markov isomorphism* $R\colon L^2(X^k/\mathcal{D},\mu^{\otimes k}/\mathcal{D}) \to L^2(X^k/\mathcal{C},\mu^{\otimes k}/\mathcal{C})$ *such that* $\mathbb{T}_U^k/\mathcal{C} \circ R = R \circ \mathbb{T}_W^k/\mathcal{D}$.

Proof. (1) $\implies$ (2): This is just Corollary 46.

(2) $\implies$ (3): Let $R := R_U^k \circ (R_W^k)^*$. By the assumption, R is well defined, and by Corollary 32, it is a Markov isomorphism as the composition of two Markov isomorphisms. By Corollary 38, we have

$$\begin{aligned}\mathbb{T}_U^k/\mathcal{C}_U^k \circ R = \mathbb{T}_U^k/\mathcal{C}_U^k \circ R_U^k \circ (R_W^k)^* = R_U^k \circ \mathbb{T}_{\nu_U^k}^k \circ (R_W^k)^* &= R_U^k \circ \mathbb{T}_{\nu_W^k}^k \circ (R_W^k)^* \\ &= R_U^k \circ (R_W^k)^* \circ \mathbb{T}_W^k/\mathcal{C}_W^k \\ &= R \circ \mathbb{T}_W^k/\mathcal{C}_W^k.\end{aligned}$$

Similarly, Lemma 40 yields that R is permutation invariant.

(3) $\implies$ (4): Set $S := I_{\mathcal{C}_U^k} \circ R \circ S_{\mathcal{C}_W^k}$, which is a Markov operator as the composition of Markov operators. By Lemma 25 (5), $\mathcal{C}_U^k$ and $\mathcal{C}_W^k$ are $\mathbb{T}_U^k$- and $\mathbb{T}_W^k$-invariant, respectively. Hence,

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

by Lemma 9 (5) and (6). In a similar fashion, Lemma 25 (6) implies that, if R is permutation invariant, then so is S.

(4) $\implies$ (5): Follows immediately from Lemma 10.

(5) $\implies$ (1): We have $t(\llbracket \mathbb{F} \rrbracket, U) = t(\mathbb{F}, \mathbb{T}_U^k/\mathcal{C}) = t(\mathbb{F}, \mathbb{T}_W^k/\mathcal{D}) = t(\llbracket \mathbb{F} \rrbracket, W)$, for every $\mathbb{F} \in \langle \mathcal{F}^k \rangle_{\circ,\cdot}$ by Corollary 23 and Lemma 22. Then, Theorem 14 yields the claim. ☐

4.7 Measure Hierarchies

Theorem 5 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.

For $1 \leqslant \ell \leqslant k < \infty$, we define the projection $p^{k,\ell}$ from $\mathbb{M}^k$ to $\mathbb{M}^\ell$ as follows: First, inductively define the function $p^{k,\ell}\colon P_n^k \to P_n^\ell$ by defining $p^{k,\ell}\colon P_0^k \to P_0^\ell$ by $p^{k,\ell}((w_{ij})_{ij\in\binom{[k]}{2}}) := (w_{ij})_{ij\in\binom{[\ell]}{2}}$ and, for the inductive step, by defining $p^{k,\ell}\colon P_{n+1}^k \to P_{n+1}^\ell$ by $p^{k,\ell}((\nu_j)_{j\in[k]}) := (p^{k,\ell}{}_*\nu_j)_{j\in[\ell]}$. This is well-defined as every $p^{k,\ell}$ is continuous. Second, the function $p^{k,\ell}\colon P_n^k \to P_n^\ell$ directly extends to a function $p^{k,\ell}\colon \mathbb{M}_n^k \to \mathbb{M}_n^\ell$ by applying it component wise. Finally, by then applying this function component-wise, $p^{k,\ell}$ extends to a continuous function $p^{k,\ell}\colon \mathbb{M}^k \to \mathbb{M}^\ell$.

Consider the *inverse limit* of the spaces $\mathbb{M}^k$ and the projections $p^{k+1,k}$ for $k \geqslant 1$ defined by

$$\mathbb{M}^\infty := \left\{ (\alpha^k)_{k\geqslant 1} \in \prod_{k\geqslant 1} \mathbb{M}^k \,\middle|\, p^{k+1,k}(\alpha^{k+1}) = \alpha^k \text{ for every } k \geqslant 1 \right\}$$

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 $k \geqslant 1$. Note that this notation is justified as $\mathbb{M}^\infty$ is again a standard Borel space [18, Exercise 17.16]. As a product of a sequence of metrizable compact spaces, $\prod_{k\geqslant 1} \mathbb{M}^k$ is metrizable [8, Proposition 2.4.4] and also compact by Tychonoff's Theorem [8, 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

$$\mathbb{WL} := \left\{ (\nu^k)_{k\geqslant 1} \in \prod_{k\geqslant 1} \mathbb{WL}^k \,\middle|\, \nu^k = {p^{k+1,k}}_* \nu^{k+1} \text{ for every } k \geqslant 1 \right\},$$

where $\mathbb{WL}^k$ denotes the set of all k-WLDs. Then, by the Kolmogorov Consistency Theorem [18, 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 $k \geqslant 1$.

Lemma 47. *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 $k \geqslant 1$.*

Proof. The set $\bigcup_{1\leqslant k<\infty} C(\mathbb{M}^k) \circ p^{\infty,k}$ is uniformly dense in $C(\mathbb{M}^\infty)$ by the Stone-Weierstrass Theorem [8, Theorem 2.4.11], cf. also the proof of Lemma 27. Hence, we have

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

□

One can show that, for every graphon $W\colon X \times X \to [0,1]$, the sequence $(\nu_W^k)_{k\geqslant 1}$ of its k-WLDs is in $\mathbb{WL}$ and, hence, yields a measure $\nu_W^\infty \in \mathscr{P}(\mathbb{M}^\infty)$. Together, Lemma 45 and Lemma 47 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. [20, Exercise 10.26] or [17, Lemma C.2].

Corollary 48. *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, they do not make a difference for weak isomorphism: two graphons are weakly isomorphic if and only if they have the same multigraph homomorphism densities [20, Corollary 10.36]. Since $\mathscr{M}_{\leqslant 1}(\mathbb{M}^{\infty})$ is Hausdorff, this yields the following corollary.

Corollary 49. *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.*

4.8 Linear Equations

In this section, we elaborate the connection between Characterization (4) of Theorem 5 and the system $\mathsf{L}^k_{\text{iso}}(G,H)$ of linear equations mentioned in the introduction. To this end, we first describe $\mathsf{L}^k_{\text{iso}}(G,H)$: G and H are graphs and $\mathsf{L}^k_{\text{iso}}(G,H)$ has a variable X_π for every set $\pi \subseteq V(G) \times V(H)$ of size $|\pi| \leqslant k$. Such a set π is called a *partial isomorphism* if the mapping it induces is injective and preserves adjacency and non-adjacency. Then, $\mathsf{L}^k_{\text{iso}}(G,H)$ is given by the following equations:

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

The graphs G and H are not distinguished by oblivious k-WL if and only if $\mathsf{L}^k_{\text{iso}}(G,H)$ has a non-negative real solution. The equivalence to precisely this system of linear equations is from [7], although it is already implicit in earlier work [16, 1, 14].

$\mathsf{L}^k_{\text{iso}}(G,H)$ is much closer related to Characterization (4) of Theorem 5 than it might seem at first glance: The variables of $\mathsf{L}^k_{\text{iso}}(G,H)$ can be interpreted as permutation-invariant matrices on $V(G)^1 \times V(H)^1, \dots, V(G)^k \times V(H)^k$. In Theorem 5, 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 $L^2(X^k, \mu^{\otimes k})$. In general, 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, if S

is permutation-invariant, this definition is independent of the specific pair of forget and introduce graphs, i.e., we have $S{\downarrow} = T_{\boldsymbol{F}_j^k} \circ S \circ T_{\boldsymbol{I}_j^k}$ for every $j \in [k]$ since $T_{\boldsymbol{F}_k^k} \circ T_{(k \dots j)} = T_{\boldsymbol{F}_j^k}$ and $T_{(j \dots k)} \circ T_{\boldsymbol{I}_k^k} = T_{\boldsymbol{I}_j^k}$.

Lemma 50. *Let $k \geqslant 1$ 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 $f \geqslant 0$. Then, $T_{\boldsymbol{I}_k^k} f = f \otimes \mathbf{1}_X \geqslant 0$, and hence, $(S \circ T_{\boldsymbol{I}_k^k}) f \geqslant 0$. Therefore, also $S{\downarrow} f = (T_{\boldsymbol{F}_k^k} \circ S \circ T_{\boldsymbol{I}_k^k}) f \geqslant 0$. 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 $\pi'(i) \coloneqq \pi(i)$ for $i \in [k-1]$ and $\pi'(k) \coloneqq k$. 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 50 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. The following lemma shows that the equation $T_{\boldsymbol{N}_k^k} \circ S = S \circ T_{\boldsymbol{N}_k^k}$ of Theorem 5 is just a way of formulating graphon analogues for the first three equations in the definition of $\mathsf{L}_{\text{iso}}^k$.

Lemma 51. *If S is a permutation-invariant Markov operator on $L^2(X^k, \mu^{\otimes k})$ such that $T_{\boldsymbol{N}_k^k} \circ S = S \circ T_{\boldsymbol{N}_k^k}$, then its operator hierarchy satisfies*

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. *$S_i \geqslant 0$ for every $i \in [k]$.*

Conversely, if $S_0, \dots, S_k$ is a sequence of permutation-invariant operators S_i on $L^2(X^i, \mu^{\otimes i})$ satisfying the four conditions above, then $S_0, \dots, S_k$ are Markov operators satisfying $T_{\boldsymbol{N}_i^i} \circ S_i = S_i \circ T_{\boldsymbol{N}_i^i}$.

Proof. The forward direction follows inductively from Lemma 50. For backward direction, 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. This immediately yields the backward direction. ☐

As a final remark, we note that in addition to Lemma 50, 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 $k \geqslant 3$, 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 easy to see that this requirement corresponds to the fourth equation in the definition of $\mathsf{L}_{\text{iso}}^k$ (concerning partial isomorphisms); we are just missing the injectivity that a partial isomorphism requires, which is not important as long as our standard Borel space is atom-free. Together with Lemma 51, this shows how to restore the characterization of oblivious k-WL indistinguishability by $\mathsf{L}_{\text{iso}}^k$ for graphs G and H from Theorem 5.

5 Simple Weisfeiler-Leman Indistinguishability

Theorem 5 shows 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 5, it is crucial that the set $\mathcal{T}^k$ of homomorphism functions is closed under multiplications. However, to achieve this closure under multiplications, we

Figure 9: The graphs $\boldsymbol{S}^2_{2,\{1\}}$ and $\boldsymbol{S}^3_{2,\{1,3\}}$.

close the set of terms under Schur products, which may introduce parallel edges if we have edges between input vertices, cf. Figure 4. To prevent this way of introducing parallel edges, we have to prevent edges from being added between input vertices in the first place.

In Section 5.1, we show how Theorem 5 and its proof have to be adapted for simple graph homomorphism densities. To this end, we introduce *simple (oblivious) k-WL*. Not surprisingly, the definitions become more similar to color refinement and the ones of Grebík and Rocha [11]. For the sake of brevity, we only include proofs that significantly differ from their counterpart in Section 4. We also briefly show how *simple non-oblivious k-WL* can be defined in Section 5.2.

5.1 Simple Oblivious Weisfeiler-Leman

To prevent edges from being added between input vertices, we only allow certain combinations of adjacency and neighbor graphs; after a sequence 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

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

We note that this is well-defined since the composition of adjacency graphs is associative. 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

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

for $\mu^{\otimes k}$-almost every $\bar{x} \in X^k$. Analogously to Theorem 14, 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, 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 edges that are missing for a vertex whenever that vertex is forgotten. We note that we do not miss any edges this way, i.e., every edge of the original graph is present in the term and added at some point, since 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. Then, 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 $\mathcal{C}^{\mathsf{s}k}_{W,0} := \left\langle \left\{\varnothing, X^k\right\} \right\rangle$, $\mathcal{C}^{\mathsf{s}k}_{W,n+1} :=$

$\mathbb{T}_W^{\mathsf{s}k}(\mathcal{C}_{W,n}^{\mathsf{s}k})$ for every $n \in \mathbb{N}$, and finally, $\mathcal{C}_W^{\mathsf{s}k} \coloneqq \mathcal{C}_{W,\infty}^{\mathsf{s}k} \coloneqq \langle \bigcup_{n\in\mathbb{N}} \mathcal{C}_{W,n}^{\mathsf{s}k} \rangle$. Then, analogously to Lemma 25, one can show that $\mathcal{C}_W^{\mathsf{s}k}$ is permutation-invariant and the minimum $\mathbb{T}_W^{\mathsf{s}k}$-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 $\mu^{\otimes k}$-relatively complete sub-σ-algebra $\mathcal{C} \in \Theta(\mathcal{B}^{\otimes k}, \mu^{\otimes k})$ *simply W-invariant* if $\mathcal{C}$ is invariant for every operator in the family $(\mathbb{T}_W^{\mathsf{s}k})_{\mathcal{C}_W^{\mathsf{s}k}}$, i.e., $\mathcal{C}$ is $(T_{\boldsymbol{F}\to W})_{\mathcal{C}_W^{\mathsf{s}k}}$-invariant for every $\boldsymbol{F} \in \mathcal{F}^{\mathsf{s}k}$. The reason for this is that, since $\mathbb{T}_W^{\mathsf{s}k}$ 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 $\mathscr{M}_{\leqslant 1}(X)$ denote the set of all measures of total mass at most 1. We endow $\mathscr{M}_{\leqslant 1}(X)$ 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 $P_0^{\mathsf{s}k} \coloneqq \{1\}$ be the one-point space and inductively define

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

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

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

By the Kolmogorov Consistency Theorem [18, Exercise 17.16], for all $\alpha \in \mathbb{P}^k$ and $(j,V) \in S^k$, there is a unique measure $\mu_{j,V}^\alpha \in \mathscr{P}(\mathbb{M}^k)$ such that $(p_{\infty,n})_*\mu_{j,V}^\alpha = (\alpha_{n+1})_{(j,V)}$ for every $n \in \mathbb{N}$. Analogously to Lemma 27, 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 $\mathbb{P}^{\mathsf{s}k} \to \mathscr{P}(\mathbb{M}^{\mathsf{s}k}), \alpha \mapsto \mu_{j,V}^\alpha$ is continuous. To adapt the definition of a k-WLD, we add a third requirement of absolute continuity and Radon-Nikodym derivatives, cf. the definition of distributions over iterated degree measures [11].

Definition 52. Let $k \geqslant 1$. 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. $\int_{\mathbb{M}^{\mathsf{s}k}} f \, d\nu = \int_{\mathbb{M}^{\mathsf{s}k}} \left(\int_{\mathbb{M}^{\mathsf{s}k}} f \, d\mu_{j,\varnothing}^\alpha\right) d\nu(\alpha)$ for all bounded measurable $f \colon \mathbb{M}^{\mathsf{s}k} \to \mathbb{R}$, $j \in [k]$, and
3. $\mu_{j,V}^\alpha \preccurlyeq \mu_{j,\varnothing}^\alpha$ and $0 \leqslant \frac{d\mu_{j,V}^\alpha}{d\mu_{j,\varnothing}^\alpha} \leqslant 1$ 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 $\mathsf{owl}_{W,0}^{\mathsf{s}k} \colon X^k \to \mathbb{M}_0^{\mathsf{s}k}$ by $\mathsf{owl}_{W,0}^{\mathsf{s}k}(\bar{x}) \coloneqq 1$ for every $\bar{x} \in X^k$. Inductively define $\mathsf{owl}_{W,n+1}^{\mathsf{s}k} \colon X^k \to \mathbb{M}_{n+1}^{\mathsf{s}k}$ by

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

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 $(\mathsf{owl}^{\mathsf{s}k}_W(\bar{x}))_n \coloneqq (\mathsf{owl}^{\mathsf{s}k}_{W,\infty}(\bar{x}))_n \coloneqq (\mathsf{owl}^{\mathsf{s}k}_{W,n}(\bar{x}))_n$ for all $n \in \mathbb{N}$, $\bar{x} \in X^k$. Finally, let $\nu^{\mathsf{s}k}_W \coloneqq \mathsf{owl}^{\mathsf{s}k}_{W*}\mu^{\otimes k} \in \mathscr{P}(\mathbb{M}^{\mathsf{s}k})$ be the push-forward of $\mu^{\otimes k}$ via $\mathsf{owl}^{\mathsf{s}k}_W$. Analogously to Lemma 31, one can show that

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

for $n \in \mathbb{N} \cup \{\infty\}$. Defining $R^{\mathsf{s}k}_W \coloneqq S_{\mathcal{C}^{\mathsf{s}k}_W} \circ T_{\mathsf{owl}^{\mathsf{s}k}_W}$ 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 32. Let us explicitly state the adaptation of Lemma 33 since the proof requires some additional work.

Lemma 53. *Let $k \geqslant 1$ and $W \colon X \times X \to [0,1]$ be a graphon. Then,*

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

Proof. The proof of (1) is analogous to Lemma 33 (1). For (2), observe that $\mu^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}_{j,\varnothing}$ is a probability measure. Then, the proof is analogous to Lemma 33 (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 33 (3).

Let $(j, V) \in S^k$. Let $\bar{x} \in X^k$ and let

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

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 Proposition 8, 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 $0 \leqslant g \leqslant 1$ holds μ-almost everywhere. For every $n \in \mathbb{N}$

and every $A \in \mathcal{B}(\mathbb{M}_n^{\mathsf{s}k})$, we have

$$\begin{aligned}
\mu_{j,V}^{\mathsf{owl}_W^{\mathsf{s}k}(\bar{x})}(p_{\infty,n}^{-1}(A)) &= (p_{\infty,n})_* \mu_{j,V}^{\mathsf{owl}_W^{\mathsf{s}k}(\bar{x})}(A) \\
&= ((\mathsf{owl}_W^{\mathsf{s}k}(\bar{x}))_{n+1})_{(j,V)}(A) \\
&= ((\mathsf{owl}_{W,n+1}^{\mathsf{s}k}(\bar{x}))_{n+1})_{(j,V)}(A) \\
&= \int_{\mathsf{owl}_{W,n}^{\mathsf{s}k\,-1}(A)_{\bar{x}[/j]}} \prod_{i \in V} W(x_i, y)\, d\mu(y) \\
&= \int_{\bar{x}[\cdot/j]^{-1}(\mathsf{owl}_W^{\mathsf{s}k\,-1}(p_{\infty,n}^{-1}(A)))} \prod_{i \in V} W(x_i, y)\, d\mu(y) \\
&= \int_{\bar{x}[\cdot/j]^{-1}(\mathsf{owl}_W^{\mathsf{s}k\,-1}(p_{\infty,n}^{-1}(A)))} \mathbb{E}\left(y \mapsto \prod_{i \in V} W(x_i, y) \mid \mathcal{C}\right) d\mu \\
&\hphantom{=} \hfill \text{(Proposition 6)} \\
&= \int_{\bar{x}[\cdot/j]^{-1}(\mathsf{owl}_W^{\mathsf{s}k\,-1}(p_{\infty,n}^{-1}(A)))} g \circ \mathsf{owl}_W^{\mathsf{s}k} \circ \bar{x}[\cdot/j]\, d\mu \\
&= \int_{p_{\infty,n}^{-1}(A)} g\, d(\mathsf{owl}_W^{\mathsf{s}k} \circ \bar{x}[\cdot/j])_* \mu \\
&= \int_{p_{\infty,n}^{-1}(A)} g\, d\mu_{j,\varnothing}^{\mathsf{owl}_W^{\mathsf{s}k}(\bar{x})}.
\end{aligned}$$

Since $\bigcup_{n \in \mathbb{N}} \left\{p_{\infty,n}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{M}_n^{\mathsf{s}k})\right\}$ generates $\mathcal{B}(\mathbb{M}^{\mathsf{s}k})$, the π-λ Theorem [18, Theorem 10.1 iii)] yields that $\mu_{j,V}^{\mathsf{owl}_W^{\mathsf{s}k}(\bar{x})}(A) = \int_A g\, d\mu_{j,\varnothing}^{\mathsf{owl}_W^{\mathsf{s}k}(\bar{x})}$ for every $A \in \mathcal{B}(\mathbb{M}^{\mathsf{s}k})$. Therefore, $\mu_{j,V}^\alpha \preccurlyeq \mu_{j,\varnothing}^\alpha$ and $0 \leqslant \frac{d\mu_{j,V}^\alpha}{d\mu_{j,\varnothing}^\alpha} \leqslant 1$ for every $\alpha \in \mathsf{owl}_W^{\mathsf{s}k}(X^k)$. By definition of $\nu_W^{\mathsf{s}k}$, this holds $\nu_W^{\mathsf{s}k}$-almost everywhere. Hence, $\nu_W^{\mathsf{s}k}$ is a simple k-WLD. $\square$

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 $0 \leqslant \frac{d\mu_{j,V}^\alpha}{d\mu_{j,\varnothing}^\alpha} \leqslant 1$ for ν-almost every $\alpha \in \mathbb{M}^{\mathsf{s}k}$. Hence, analogously to Lemma 34, one can show that setting

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

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

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

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 53, 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

$$\begin{aligned}
(T_{\mathsf{owl}^{\mathsf{s}k}_W} \circ T_{\boldsymbol{S} \to \nu^{\mathsf{s}k}_W} f)(\bar{x}) &= \int_{\mathbb{M}^{\mathsf{s}k}} \frac{d\mu^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}_{j,V}}{d\mu^{\mathsf{owl}^{\mathsf{s}k}_W(\bar{x})}_{j,\varnothing}} \cdot f \, d(\mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j])_* \mu \\
&\qquad \text{(Definition and Lemma 53 (2))} \\
&= \int_X \mathbb{E}(y \mapsto \prod_{i \in V} W(x_i, y) \mid \mathcal{C}_{\bar{x}}) \cdot f \circ \mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j] \, d\mu \\
&= \int_X \prod_{i \in V} W(x_i, y) \cdot \mathbb{E}(f \circ \mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[\cdot/j] \mid \mathcal{C}_{\bar{x}})(y) \, d\mu(y) \\
&\qquad \text{(Proposition 6)} \\
&= \int_X \prod_{i \in V} W(x_i, y) \cdot f \circ \mathsf{owl}^{\mathsf{s}k}_W \circ \bar{x}[y/j] \, d\mu(y) \\
&= (T_{\boldsymbol{S} \to W} \circ T_{\mathsf{owl}^{\mathsf{s}k}_W} f)(\bar{x})
\end{aligned}$$

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 35 (2) and (3), respectively. ☐

For $k \geqslant 1$ and a simple k-WL distribution $\nu \in \mathscr{P}(\mathbb{M}^{\mathsf{s}k})$, let $\mathbb{T}_\nu \coloneqq (T_{\boldsymbol{S} \to \nu})_{\boldsymbol{S} \in \mathcal{F}^{\mathsf{s}k}}$. 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 54 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 40 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 artificial as it means that $\mathcal{C}$ is $(\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W}$-invariant. Corollary 23 can then be adapted to the also somewhat convoluted 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 22 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 $n \geqslant h(\mathbb{F})$, the set $F_n^{\mathbb{F}}$ of functions $\mathbb{M}_n^{\mathsf{s}k} \to [0,1]$ is defined similarly to Definition 41. 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 42 and Corollary 43 adapt in a straight-forward fashion.

For every $n \in \mathbb{N} \cup \{\infty\}$, define $\mathcal{T}_n^{\mathsf{s}k} := \bigcup_{\mathbb{F} \in \langle \mathcal{F}^{\mathsf{s}k} \rangle_{\circ,\cdot}, h(\mathbb{F}) \leqslant n} F_n^{\mathbb{F}}$ and abbreviate $\mathcal{T}^{\mathsf{s}k} := \mathcal{T}_\infty^{\mathsf{s}k}$. Lemma 44 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 45.

Lemma 55. *Let $k \geqslant 1$. 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 [20, Lemma 10.23], we also obtain the following additional corollary.

Corollary 56. *Let $k \geqslant 1$. 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.*

We note that the same reasoning does not work for multigraphs as the Counting Lemma does not hold for multigraphs. Moreover, the above corollary does not hold for multigraphs since convergence of simple graph homomorphism densities does not imply convergence of multigraph homomorphism densities, cf. [20, Exercise 10.26] or [17, Lemma C.2].

Having outlined the necessary changes for simple graphs, we obtain the following variant of Theorem 5 for simple graph homomorphism densities. Note the inelegant 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 57. *Let $k \geqslant 1$ 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-invariant) Markov isomorphism* $R\colon L^2(X^k/\mathcal{C}^{\mathsf{s}k}_W, \mu^{\otimes k}/\mathcal{C}^{\mathsf{s}k}_W) \to L^2(X^k/\mathcal{C}^{\mathsf{s}k}_U, \mu^{\otimes k}/\mathcal{C}^{\mathsf{s}k}_U)$ *such that* $\mathbb{T}^{\mathsf{s}k}_U/\mathcal{C}^{\mathsf{s}k}_U \circ R = R \circ \mathbb{T}^{\mathsf{s}k}_W/\mathcal{C}^{\mathsf{s}k}_W$.

4. *There is a (permutation-invariant) Markov operator* $S\colon L^2(X^k, \mu^{\otimes k}) \to L^2(X^k, \mu^{\otimes k})$ *such that* $(\mathbb{T}^{\mathsf{s}k}_U)_{\mathcal{C}^{\mathsf{s}k}_U} \circ S = S \circ (\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W}$ *and* $S^* \circ (\mathbb{T}^{\mathsf{s}k}_U)_{\mathcal{C}^{\mathsf{s}k}_U} = (\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W} \circ S^*$.

5. *There are* $\mu^{\otimes k}$*-relatively complete sub-*σ*-algebras* $\mathcal{C}$ *and* $\mathcal{D}$ *of* $\mathcal{B}^{\otimes k}$ *that are simply* U*-invariant and simply* W*-invariant, respectively, and a Markov isomorphism* $R\colon L^2(X^k/\mathcal{D}, \mu^{\otimes k}/\mathcal{D}) \to L^2(X^k/\mathcal{C}, \mu^{\otimes k}/\mathcal{C})$ *such that* $(\mathbb{T}^{\mathsf{s}k}_U)_{\mathcal{C}^{\mathsf{s}k}_U}/\mathcal{C} \circ R = R \circ (\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W}/\mathcal{D}$.

Proof. (1) $\implies$ (2): Follows from Lemma 55.

(2) $\implies$ (3): Analogous to Theorem 5 as we have both $\mathbb{T}^{\mathsf{s}k}_U/\mathcal{C}^{\mathsf{s}k}_U \circ R^{\mathsf{s}k}_U = R^{\mathsf{s}k}_U \circ \mathbb{T}_{\nu^{\mathsf{s}k}_U}$ and $(R^{\mathsf{s}k}_W)^* \circ \mathbb{T}^{\mathsf{s}k}_W/\mathcal{C}^{\mathsf{s}k}_W = \mathbb{T}_{\nu^{\mathsf{s}k}_W} \circ (R^{\mathsf{s}k}_W)^*$ since $R^{\mathsf{s}k}_W$ is a Markov isomorphism.

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

$$\begin{aligned}
(\mathbb{T}^{\mathsf{s}k}_U)_{\mathcal{C}^{\mathsf{s}k}_U} \circ S = (\mathbb{T}^{\mathsf{s}k}_U)_{\mathcal{C}^{\mathsf{s}k}_U} \circ I_{\mathcal{C}^{\mathsf{s}k}_U} \circ R \circ S_{\mathcal{C}^{\mathsf{s}k}_W} &= I_{\mathcal{C}^{\mathsf{s}k}_U} \circ \mathbb{T}^{\mathsf{s}k}_U/\mathcal{C}^{\mathsf{s}k}_U \circ R \circ S_{\mathcal{C}^{\mathsf{s}k}_W} && \text{(Lemma 9 (3))}\\
&= I_{\mathcal{C}^{\mathsf{s}k}_U} \circ R \circ \mathbb{T}^{\mathsf{s}k}_W/\mathcal{C}^{\mathsf{s}k}_W \circ S_{\mathcal{C}^{\mathsf{s}k}_W}\\
&= I_{\mathcal{C}^{\mathsf{s}k}_U} \circ R \circ S_{\mathcal{C}^{\mathsf{s}k}_W} \circ (\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W} && \text{(Lemma 9 (4))}\\
&= S \circ (\mathbb{T}^{\mathsf{s}k}_W)_{\mathcal{C}^{\mathsf{s}k}_W}.
\end{aligned}$$

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 5, if R is permutation invariant, then so is S.

(4) $\implies$ (5): Follows immediately from Lemma 10.

(5) $\implies$ (1): Analogous to Theorem 5. ☐

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 48, where we now have a third characterization in terms of the cut distance, denoted by $\delta_\square$, cf. [20, Theorem 11.5].

Lemma 58. *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. $W_n \xrightarrow{\delta_\square} W$.

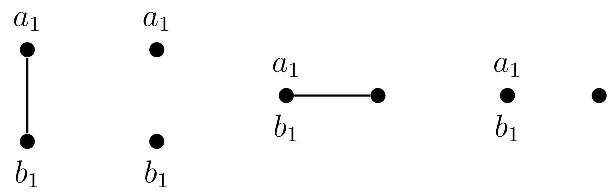

Figure 10: The (isomorphism types of) graphs in $\mathcal{F}^{\mathsf{ns}1}$.

5.2 Non-Oblivious Simple Weisfeiler-Leman

As mentioned in the introduction, there are two non-equivalent variants of k-WL for graphs: oblivious k-WL and (non-oblivious) k-WL, where 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 [12, Corollary V.7]. (Non-oblivious) k-WL is usually considered in the graph setting since it needs less memory to achieve the same expressive power, but the connections of oblivious k-WL to other characterizations are much cleaner. Examples of this are the system $\mathsf{L}^k_{\text{iso}}(G, H)$ of linear equations, where the k directly corresponds to the k of oblivious k-WL, the logic C^k, the k-variable fragment of first-order logic with counting quantifiers, and the maximum bag size in a tree decomposition, although the latter is usually hidden by the fact that one subtracts one from the maximum bag size in a tree decomposition to get the width of such a decomposition. Tree decompositions also give an explanation of the difference between oblivious k-WL and non-oblivious k-WL: intuitively, given a tree decomposition of width k, we may dissect it into parts at bags of size $k+1$ or at bags of size k yielding oblivious k-WL and non-oblivious k-WL, respectively.

Let us formally define non-oblivious k-WL. Let G be a graph and recall that 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 $v_i v_j \in E(G)$, and $A_{ij} = 0$ otherwise. Then, let $\mathsf{wl}^k_{G,0}(\bar{v}) \coloneqq \mathsf{atp}_G(\bar{v})$ and, for every $n \geqslant 0$, define

$$\mathsf{wl}^k_{G,n+1}(\bar{v}) \coloneqq \Big(\mathsf{wl}^k_{G,n}(\bar{v}), \{\!\!\{ \big(\mathsf{atp}_G(\bar{v}w), \big(\mathsf{wl}^k_{G,n}(\bar{v}[w/j])\big)_{j \in [k]}\big) \mid w \in V(G) \}\!\!\}\Big) \tag{9}$$

for every $\bar{v} \in V(G)^k$. 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 $n \geqslant 0$. Recall that the j-neighbor $\bar{v}[w/j]$ denotes the k-tuple obtained from $\bar{v}$ by replacing the jth component by w. 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 [12, Proposition V.4].

Following the intuition that a tree decomposition of width k can be dissect into parts either at bags of size $k+1$ or at bags of size k, one can adapt the definitions of this section to obtain a variant of simple k-WL akin to non-oblivious k-WL. To this end, recall the definition of forget, adjacency, and introduce graphs from Definition 12 and 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 $j_1, j_2 \in [k+1]$, $V \subseteq [k+1] \setminus \{j_1\}$. Then, a 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. 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 57 without the mismatch of the k of simple k-WL and the k of the treewidth. Since it is so similar, however, we do not state it here.

For a last remark, consider the special case of $k = 1$ of fractional isomorphism. The isomorphism types in $\mathcal{F}^{\mathsf{ns}1}$ are shown in Figure 10; they all are symmetric in this special case. We note that the graph on the far left is $\boldsymbol{A}$, the edge with one input and one output vertex, cf. Example 17, which satisfies $T_{\boldsymbol{A} \to W} = T_W$. Among the bi-labeled graphs in $\mathcal{F}^{\mathsf{ns}1}$, this is actually the only interesting graph since it is necessary and already sufficient for the construction of all trees. In other words, we can leave out the other bi-labeled graphs from $\mathcal{F}^{\mathsf{ns}1}$. Then, the resulting characterizations are essentially Theorem 4, the result of Grebík and Rocha.

Acknowledgements

The author would like to thank Jan Grebík for mentioning the idea of combining all k-WL distributions of a graphon into a single object to obtain a new characterization of weak isomorphism, which led to Section 4.7. Moreover, the author is grateful to the anonymous reviewers for their many helpful suggestions and, in particular, for their advice on how to make the introduction more accessible.

Funded by the European Union (ERC, SymSim, 101054974). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] A. Atserias and E. Maneva. "Sherali–Adams relaxations and indistinguishability in counting logics". In: *SIAM Journal on Computing* 42.1 (2013), pp. 112–137. DOI: 10.1137/120867834.

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

[3] J. Böker. "Graph similarity and homomorphism densities". In: *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*. Ed. by N. Bansal, E. Merelli, and J. 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] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. "Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing". In: *Advances in Mathematics* 219.6 (Dec. 2008), pp. 1801–1851. ISSN: 0001-8708. DOI: 10.1016/j.aim.2008.07.008.

[5] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. "Convergent sequences of dense graphs II. Multiway cuts and statistical physics". In: *Annals of Mathematics* 176.1 (2012), pp. 151–219. ISSN: 0003-486X. DOI: 10.4007/annals.2012.176.1.2.

[6] J. Y. Cai, M. Fürer, and N. Immerman. "An optimal lower bound on the number of variables for graph identification". In: *Combinatorica* 12.4 (Dec. 1992), pp. 389–410. ISSN: 0209-9683. DOI: 10.1007/BF01305232.

[7] H. Dell, M. Grohe, and G. Rattan. "Lovász meets Weisfeiler and Leman". In: *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*. Ed. by I. Chatzigiannakis, C. Kaklamanis, D. Marx, and D. Sannella. 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.

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

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

[10] T. Eisner, B. Farkas, M. Haase, and R. Nagel. *Operator Theoretic Aspects of Ergodic Theory*. Springer, Nov. 2015. ISBN: 978-3-319-16898-2. DOI: 10.1007/978-3-319-16898-2.

[11] J. Grebík and I. Rocha. "Fractional isomorphism of graphons". In: *Combinatorica* 42.3 (June 2022), pp. 365–404. ISSN: 1439-6912. DOI: 10.1007/s00493-021-4336-9.

[12] M. Grohe. "The logic of graph neural networks". In: *Proceedings of the 36th Annual ACM/IEEE Symposium on Logic in Computer Science*. LICS '21. New York, NY, USA: Association for Computing Machinery, Nov. 2021, pp. 1–17. ISBN: 978-1-66544-895-6. DOI: 10.1109/LICS52264.2021.9470677.

[13] M. Grohe, K. Kersting, M. Mladenov, and P. Schweitzer. "Color refinement and its applications". In: *An Introduction to Lifted Probabilistic Inference*. The MIT Press, Aug. 2021. ISBN: 9780262365598. DOI: 10.7551/mitpress/10548.003.0023.

[14] M. Grohe and M. Otto. "Pebble games and linear equations". In: *The Journal of Symbolic Logic* 80.3 (2015), pp. 797–844. DOI: 10.1017/jsl.2015.28.

[15] M. Grohe, G. Rattan, and T. Seppelt. "Homomorphism tensors and linear equations". In: *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*. Ed. by M. Bojańczyk, E. Merelli, and D. P. Woodruff. Vol. 229. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 70:1–70:20. ISBN: 978-3-95977-235-8. DOI: 10.4230/LIPIcs.ICALP.2022.70.

[16] N. Immerman and E. 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 A. 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.

[17] S. Janson. "Graphons, cut norm and distance, couplings and rearrangements". In: *New York Journal of Mathematics*. NYJM Monographs 4 (2013).

[18] A. S. Kechris. *Classical Descriptive Set Theory*. Graduate texts in mathematics. Springer-Verlag, 1995. ISBN: 978-0-387-94374-9. DOI: 10.1007/978-1-4612-4190-4.

[19] T. Kloks. *Treewidth: Computations and Approximations*. Springer Science & Business Media, Aug. 1994. ISBN: 978-3-540-58356-1. DOI: 10.1007/BFb0045375.

[20] L. 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.

[21] L. Lovász and B. Szegedy. "Limits of dense graph sequences". In: *Journal of Combinatorial Theory, Series B* 96.6 (Nov. 2006), pp. 933–957. ISSN: 0095-8956. DOI: 10.1016/j.jctb.2006.05.002.

[22] L. Mančinska and D. 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.

[23] M. V. Ramana, E. R. Scheinerman, and D. 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.

[24] G. 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.

[25] G. Tinhofer. "Graph isomorphism and theorems of Birkhoff-type". In: *Computing* 36.4 (June 1986), pp. 285–300. ISSN: 0010-485X. DOI: 10.1007/BF02240204.