Graph Similarity and Homomorphism Densities

Jan Böker
RWTH Aachen University, Germany

Abstract

We introduce the tree distance, a new distance measure on graphs. The tree distance can be computed in polynomial time with standard methods from convex optimization. It is based on the notion of fractional isomorphism, a characterization based on a natural system of linear equations whose integer solutions correspond to graph isomorphism. By results of Tinhofer (1986, 1991) and Dvořák (2010), two graphs G and H are fractionally isomorphic if and only if, for every tree T, the number of homomorphisms from T to G equals the corresponding number from T to H, which means that the tree distance of G and H is zero. Our main result is that this correspondence between the equivalence relations "fractional isomorphism" and "equal tree homomorphism densities" can be extended to a correspondence between the associated distance measures. Our result is inspired by a similar result due to Lovász and Szegedy (2006) and Borgs, Chayes, Lovász, Sós, and Vesztergombi (2008) that connects the cut distance of graphs to their homomorphism densities (over all graphs), which is a fundamental theorem in the theory of graph limits. We also introduce the path distance of graphs and take the corresponding result of Dell, Grohe, and Rattan (2018) for exact path homomorphism counts to an approximate level. Our results answer an open question of Grohe (2020) and help to build a theoretical understanding of vector embeddings of graphs.

The distance measures we define turn out be closely related to the cut distance. We establish our main results by generalizing our definitions to graphons, which are limit objects of sequences of graphs, as this allows us to apply techniques from functional analysis. We prove the fairly general statement that, for every "reasonably" defined graphon pseudometric, an exact correspondence to homomorphism densities can be turned into an approximate one. We also provide an example of a distance measure that violates this reasonableness condition. This incidentally answers an open question of Grebík and Rocha (2021).

2012 ACM Subject Classification Mathematics of computing → Graph theory

Keywords and phrases graph similarity, homomorphism densities, cut distance

Digital Object Identifier 10.4230/LIPIcs.ICALP.2021.32

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: `https://arxiv.org/abs/2104.14213` [4]

Acknowledgements The author would like to thank Yufei Zhao for pointing out how a non-quantitative inverse counting lemma for the cut distance follows from the compactness of the graphon space and the anonymous reviewers for their helpful comments.

1 Introduction

Vector representations of graphs allow to apply standard machine learning techniques to graphs, and a variety of methods to generate such embeddings has been studied in the machine learning literature. However, from a theoretical point of view, these embeddings have not received much attention and are not well understood. Some machine learning methods only implicitly operate on such vector representations as they only access the inner products of these vectors. These methods are known as *kernel methods* and most graph kernels are based on counting occurrences of certain substructures, e.g., walks or trees. See [15] for a recent survey on vector embeddings.

Many kinds of substructure counts in a graph such as graph motifs are actually just *homomorphism* counts "in disguise", and hence, homomorphisms provide a formal and flexible framework for counting all kinds of substructures in graphs [5]; a homomorphism from a graph F to a graph G is a mapping from the vertices of F to the vertices of G such that every edge of F is mapped to an edge of G. A theorem of Lovász from 1967 [18], which states that two graphs G and H are isomorphic if and only if, for every graph F, the number $\mathsf{hom}(F, G)$ of homomorphisms from F to G equals the corresponding number $\mathsf{hom}(F, H)$ from F to H, led to the development of the theory of graph limits [2, 20], where one considers convergent sequences of graphs and their limit objects, graphons. In terms of the *homomorphism vector* $\mathsf{Hom}(G) \coloneqq (\mathsf{hom}(F, G))_{F \text{ graph}}$ of a graph G, the result of Lovász states that graphs are mapped to the same vector if and only if they are isomorphic.

Computing an entry of $\mathsf{Hom}(G)$ is $\#P$-complete and recent results have mostly focused on restrictions $\mathsf{Hom}_{\mathcal{F}}(G) \coloneqq (\mathsf{hom}(F, G))_{F \in \mathcal{F}}$ of these vectors to classes $\mathcal{F}$ for which computing these entries is actually tractable. Under a natural assumption from parameterized complexity theory, this is the case for precisely the classes $\mathcal{F}$ of bounded tree width [6]. This has led to various surprisingly clean results, e.g., for trees and, more general, graphs of bounded treewidth [10], cycles and paths [7], planar graphs [22], and, most recently, graphs of bounded tree-depth [14]. These results only show what it means for graphs to be mapped to the same homomorphism vector; they do not say anything about the similarity of two graphs if the homomorphism vectors are not exactly the same but *close*. Grohe formulated the vague hypothesis that, for suitable classes $\mathcal{F}$, the embedding $\mathsf{Hom}_{\mathcal{F}}$ combined with a suitable inner product on the latent space induces a natural similarity measure on graphs [15]. This is supported by initial experiments, which show that homomorphism vectors in combination with support vector machines perform well on standard graph classification. Our results further support this hypothesis from a theoretical standpoint by showing that tree homomorphism counts provide a robust similarity measure.

For the class $\mathcal{T}$ of trees and two graphs G and H, we have $\mathsf{Hom}_{\mathcal{T}}(G) = \mathsf{Hom}_{\mathcal{T}}(H)$ if and only if G and H are not distinguished by *color refinement* (also known as the 1-*dimensional Weisfeiler-Leman algorithm*) [10], a popular heuristic for graph isomorphism. Another characterization of this equivalence due to Tinhofer is that of *fractional isomorphism* [26, 27]. Let $A \in \mathbb{R}^{V(G) \times V(G)}$ and $B \in \mathbb{R}^{V(H) \times V(H)}$ be the adjacency matrices of G and H, respectively, and consider the following system $\mathsf{F}_{\text{iso}}(G, H)$ of linear equations:

$$\mathsf{F}_{\text{iso}}(G, H) : \begin{cases} AX = XB \\ X\mathbf{1}_{V(H)} = \mathbf{1}_{V(G)} \\ \mathbf{1}_{V(G)}^T X = \mathbf{1}_{V(H)}^T \end{cases}$$

Here, X denotes a $(V(G) \times V(H))$-matrix of variables, and $\mathbf{1}_U$ denotes the all-1 vector over the index set U. The non-negative integer solutions to $\mathsf{F}_{\text{iso}}(G, H)$ are precisely the permutation matrices that describe isomorphisms between G and H. The non-negative real solutions are called *fractional isomorphisms* of G and H. Tinhofer proved that G and H are not distinguished by the color refinement algorithm if and only if there is a fractional isomorphism of G and H. Grohe proposed to define a similarity measure based on this characterization [15]: For a matrix norm $\|\cdot\|$ that is invariant under permutations of the rows and columns, consider

$$\operatorname{dist}_{\|\cdot\|}(G, H) \coloneqq \min_{\substack{X \in [0,1]^{V(G) \times V(H)}, \\ X \text{ doubly stochastic}}} \|AX - XB\|.$$

Most graph distance measures based on matrix norms are highly intractable as the problem of their computation is related to notoriously hard maximum quadratic assignment problem [23]. This hardness, which stems from the minimization over the set of all permutation matrices, motivated Grohe to propose $\mathrm{dist}_{\|\cdot\|}$, where the set of all permutation matrices is relaxed to the the convex set of doubly stochastic matrices, yielding a convex optimization problem. With the results of Tinhofer and Dvořák, we know that the graphs of distance zero w.r.t. $\mathrm{dist}_{\|\cdot\|}$ are precisely those that cannot be distinguished by tree homomorphism counts.

So far, the only known connection between a graph distance measure based on matrix norms and graph homomorphisms is between the *cut distance* and normalized homomorphism numbers (called *homomorphism densities*) [2]. Grohe asks whether a similar correspondence between $\mathrm{dist}_{\|\cdot\|}$ and restricted homomorphism vectors can be established, and we give a positive answer to this question. We introduce the *tree distance* $\delta^{\mathcal{T}}$ of graphs, which is a normalized variant of $\mathrm{dist}_{\|\cdot\|}$ and show the following theorem, which is stated here only informally. We also introduce the *path distance* $\delta^{\mathcal{P}}$ of graphs and prove the analogous theorem to Theorem 1 for $\delta^{\mathcal{P}}$ and normalized path homomorphism counts.

▶ **Theorem 1** (Informal Theorem 6 and Theorem 7)**.** *Two graphs G and H are similar w.r.t. $\delta^{\mathcal{T}}$ if and only if the homomorphism densities $t(T, G)$ and $t(T, H)$ are close for trees T.*

In the theory of graph limits, *graphons* serve as limit objects for sequences of graphs. By defining distance measures on the more general graphons, we are able to use techniques from functional analysis to show that any "reasonably" defined pseudometric on graphons satisfying an exact correspondence to homomorphism densities also has to satisfy an approximate one. As an application, we get that both the tree and the path distance satisfy this correspondence to tree and path homomorphism densities, respectively. For the case of trees, we rely on a generalization of the notion of fractional isomorphism to graphons by Grebík and Rocha [13]. For the case of paths, we prove this generalization of the result of Dell, Grohe, and Rattan [7] by ourselves.

This paper is organized as follows. In the preliminaries, Section 2, we collect the definitions of graphs, the space $L_2[0,1]$, graphons, and the cut distance. In Section 3, we define the tree distance and the path distance for graphs and formally state Theorem 1 and its path counterpart. In Section 4, we state and prove the theorems that allow us to show these correspondences for graphon pseudometrics. Section 5 provides the first application of these tools for the tree distance: we first state the needed result of fractional isomorphism of graphons due to Grebík and Rocha and then use this to define the tree distance of graphons. These definitions and results specialize to the ones presented in Section 3 for graphs. The treatment of the path distance for graphons is similar to the one of the tree distance, except for the fact that we prove a characterization of graphons with the same path homomorphism densities ourselves, and can be found in Section 6. In Section 7, we define another distance measure on graphs based on the invariant computed by the color refinement algorithm and show that it only satisfies one direction of the approximate correspondence to tree homomorphism densities. Our counterexample incidentally answers an open question of Grebík and Rocha [13]. Section 8 poses some interesting open questions that come up during the study of these distance measures. All missing proofs can be found in the full version of the paper.

2 Preliminaries

2.1 Graphs

By the term graph, we refer to a simple, undirected, and finite graph. For a graph G, we denote its vertex set by $V(G)$ and its edge set by $E(G)$, and we let $\mathsf{v}(G) \coloneqq |V(G)|$ and $\mathsf{e}(G) \coloneqq |E(G)|$. We usually view the adjacency matrix A of a graph G as a matrix $A \in \mathbb{R}^{V(G)\times V(G)}$, i.e., it is indexed by the vertices of G. Sometimes, we assume without loss of generality that the vertex set of a graph is $[n] \coloneqq \{1,\dots,n\}$, where $n \in \mathbb{N}$ is a natural number. A homomorphism from a graph F to a graph G is a mapping $\varphi\colon V(F) \to V(G)$ such that $\varphi(u)\varphi(v) \in E(G)$ for every $uv \in E(F)$. We denote the number of homomorphisms from F to G by $\mathsf{hom}(F,G)$. The homomorphism density from F to G is given by $t(F,G) \coloneqq \mathsf{hom}(F,G)/\mathsf{v}(G)^{\mathsf{v}(F)}$.

A *weighted graph* $G = (V, a, B)$ consists of a vertex set V, a positive real vector $a = (\alpha_v)_{v\in V} \in \mathbb{R}^V$ of vertex weights and a real symmetric matrix $B = (\beta_{uv}) \in [0,1]^{V\times V}$ of edge weights; that is, we restrict ourselves to edge weights from $[0,1]$. We write $\mathsf{v}(G) = |V|$, $V(G) = V$, $\alpha_v(G) = \alpha_v$, $\alpha_G = \sum_{v\in V(G)} \alpha_v(G)$ and $\beta_{uv}(G) = \beta_{uv}$. A weighted graph is called *normalized* if $\alpha_G = 1$. For a simple graph F and a weighted graph G, we define the *homomorphism number*

$$\mathsf{hom}(F,G) = \sum_{\varphi\colon V(F)\to V(G)} \prod_{v\in V(F)} \alpha_{\varphi(v)}(G) \prod_{uv\in E(F)} \beta_{\varphi(u)\varphi(v)}(G)$$

and the *homomorphism density* $t(F,G) = \mathsf{hom}(F,G)/\alpha_G^{\mathsf{v}(F)}$. When viewing a graph as a weighted graph in the obvious way, these notions coincide with the ones for graphs.

2.2 The Space $L_2[0,1]$ and Graphons

A detailed introduction to functional analysis can be found in [8]; here, we only repeat some notions we use throughout the main body of the paper. Let $L_2[0,1]$ denote the space of $\mathbb{R}$-valued 2-integrable functions on $[0,1]$ (modulo equality almost anywhere). We could consider consider a standard Borel space with a Borel probability measure, cf. [13], but for the sake of convenience, we stick to $[0,1]$ with the Lebesgue measure just as [20]. The space $L_2[0,1]$ is a Hilbert space with the inner product defined by $\langle f,g\rangle \coloneqq \int_{[0,1]} f(x)g(x)\,dx$ for functions $f,g \in L_2[0,1]$. Let $T\colon L_2[0,1] \to L_2[0,1]$ be a bounded linear operator, or operator for short. We write $\|T\|_{2\to2}$ for its operator norm, i.e., $\|T\|_{2\to2} = \sup_{\|g\|_2\le1}\|Tg\|_2$. The *Hilbert adjoint* of T is the unique operator $T^*\colon L_2[0,1] \to L_2[0,1]$ such that $\langle Tf,g\rangle = \langle f,T^*g\rangle$ for all $f,g \in L_2[0,1]$, and T is called self-adjoint if $T^* = T$.

Let $\mathcal{W}$ denote the set of all bounded symmetric measurable functions $W\colon [0,1]^2 \to \mathbb{R}$, called *kernels*. Let $\mathcal{W}_0 \subseteq \mathcal{W}$ denote all such W that satisfy $0 \le W \le 1$; such a W is called a *graphon*. Every kernel $W \in \mathcal{W}$ defines a self-adjoint operator $T_W\colon L_2[0,1] \to L_2[0,1]$ by setting $(T_W f)(x) = \int_{[0,1]} W(x,y)f(y)\,dy$ for every $x \in [0,1]$, which then is a Hilbert-Schmidt operator, and in particular, compact [20].

A kernel $W \in \mathcal{W}$ is called a *step function* if there is a partition $S_1 \cup \dots \cup S_k$ of $[0,1]$ such that W is constant on $S_i \times S_j$ for all $i,j \in [k]$. For a weighted graph H on $[n]$, one can define a step function $W_H \in \mathcal{W}$ by splitting $[0,1]$ into n intervals $I_1,\dots,I_n$, where I_i has length $\lambda(I_i) = \alpha_i(H)/\alpha(H)$ for every $i \in [n]$, and letting $W_H(x,y) \coloneqq \beta_{ij}(H)$ for all $x \in I_i, y \in I_j$ and $i,j \in [n]$. Of course, W_H depends on the labeling of the vertices of H. Note that W_H is a graphon, and in particular, W_G is a graphon for every graph G.

2.3 The Cut Distance

See [20] for a thorough introduction to the cut distance. The usual definition of the cut distance involves the *blow-up* $G(k)$ of a graph G by $k \geq 0$, where every vertex of G is replaced by k identical copies, to get graphs on the same number of vertices. Going this route is rather cumbersome, and we directly define the cut distance for weighted graphs via *fractional overlays*; this definition also applies to graphs in the straightforward way. A *fractional overlay* of weighted graphs G and H is a matrix $X \in \mathbb{R}^{V(G) \times V(H)}$ such that $X_{uv} \geq 0$ for all $u \in V(G)$, $v \in V(H)$, $\sum_{v \in V(H)} X_{uv} = \alpha_u(G)/\alpha_G$ for every $u \in V(G)$, and $\sum_{u \in V(G)} X_{uv} = \alpha_v(H)/\alpha_H$ for every $v \in V(H)$. Let $\mathcal{X}(G, H)$ denote the set of all fractional overlays of G and H. Note that, for graphs G and H, the second and third condition just say that the row and column sums of X are $1/\mathsf{v}(G)$ and $1/\mathsf{v}(H)$, respectively. For weighted graphs G and H and a fractional overlay $X \in \mathcal{X}(G, H)$, let

$$d_\square(G, H, X) \coloneqq \max_{Q,R \subseteq V(G) \times V(H)} \Big| \sum_{\substack{iu \in Q,\\ jv \in R}} X_{iu} X_{jv} (\beta_{ij}(G) - \beta_{uv}(H)) \Big|.$$

Then, define the *cut distance* $\delta_\square(G, H) \coloneqq \min_{X \in \mathcal{X}(G,H)} d_\square(G, H, X)$.

Defining the cut distance of graphons is actually much simpler. Define the *cut norm* on the linear space $\mathcal{W}$ of kernels by $\|W\|_\square \coloneqq \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W(x, y)\, dx\, dy \right|$ for $W \in \mathcal{W}$; here, as in the whole of the paper, we tacitly assume sets (and functions) we take an infimum or supremum over to be measurable. Let $S_{[0,1]}$ denote the group of all invertible measure-preserving maps $\varphi\colon [0, 1] \to [0, 1]$. For a kernel $W \in \mathcal{W}$ and a $\varphi \in S_{[0,1]}$, let W^φ be the kernel defined by $W^\varphi(x, y) \coloneqq W(\varphi(x), \varphi(y))$. For kernels $U, W \in \mathcal{W}$, define their *cut distance* by setting $\delta_\square(U, W) \coloneqq \inf_{\varphi \in S_{[0,1]}} \|U - W^\varphi\|_\square$. This coincides with the previous definition when viewing weighted graphs as graphons [20, Lemma 8.9]. We can also express $\delta_\square(U, W)$ via the kernel operator as $\delta_\square(U, W) = \inf_{\varphi \in S_{[0,1]}} \sup_{f,g\colon [0,1] \to [0,1]} |\langle f, T_{U - W^\varphi} g\rangle|$ [20, Lemma 8.10]. The definition of the cut distance is quite robust. For example, allowing f and g in the previous definition to be complex-valued or choosing a different operator norm does not make a difference in most cases [16, Appendix E].

For a graph F and a kernel $W \in \mathcal{W}$, define the *homomorphism density*

$$t(F, W) \coloneqq \int_{[0,1]^{V(F)}} \prod_{ij \in E(F)} W(x_i, x_j) \prod_{i \in V(F)} dx_i,$$

which coincides with the previous definition when viewing weighted graphs as graphons [20, Equation (7.2)]. Lemma 2 and Lemma 3 state the connection between the cut distance and homomorphism densities: Informally, the Lemma 2 states that graphons that are close in the cut distance have similar homomorphism densities, while Lemma 3 states that graphs that have similar homomorphism densities are close in the cut distance. We refer to such statements as a *counting lemma* and an *inverse counting lemma*, respectively.

▶ **Lemma 2** (Counting Lemma [21])**.** *Let F be a simple graph, and let $U, W \in \mathcal{W}_0$ be graphons. Then, $|t(F, U) - t(F, W)| \leq \mathsf{e}(F) \cdot \delta_\square(U, W)$.*

▶ **Lemma 3** (Inverse Counting Lemma [3, 20])**.** *Let $k > 0$, let $U, W \in \mathcal{W}_0$ be graphons, and assume that, for every graph F on k vertices, we have $|t(F, U) - t(F, W)| \leq 2^{-k^2}$. Then, $\delta_\square(U, W) \leq 50/\sqrt{\log k}$.*

In particular, graphons U and W have cut distance zero if and only if, for every graph F, we have $t(F, U) = t(F, W)$. Call a sequence $(W_n)_{n \in \mathbb{N}}$ of graphons *convergent* if, for every graph F, the sequence $(t(F, W_n))_{n\in\mathbb{N}}$ is Cauchy. The two theorems above yield that $(W_n)_{n\in\mathbb{N}}$ is convergent if and only if $(W_n)_{n\in\mathbb{N}}$ is Cauchy in $\delta_\square$. Let $\widetilde{\mathcal{W}}_0$ be obtained from $\mathcal{W}_0$ by identifying graphons with cut distance zero; such graphons are called *weakly isomorphic*. One of the main results from graph limit theory is the compactness of the space $(\widetilde{\mathcal{W}}_0, \delta_\square)$.

▶ **Theorem 4** ([19]). *The space $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is compact.*

3 Similarity Measures of Graphs

In this section, we define the tree and path distances of graphs and formally state the correspondences to tree and path homomorphism densities, respectively. All presented results are specializations of the results for graphons proven in Section 5 and Section 6.

3.1 The Tree Distance of Graphs

Recall that two graphs G and H have the same tree homomorphism counts if and only if the system $\mathsf{F}_{\text{iso}}(G, H)$ of linear equations has a non-negative solution. Based on this, Grohe proposed $\text{dist}_{\|\cdot\|}$ as a similarity measure of graphs. This is nearly what we define as the tree distance of graphs. What is missing is, first, a more general definition for graphs with different numbers of vertices and, second, an appropriate choice of a matrix norm with an appropriate normalization factor; analogously to the cut distance, we normalize the tree distance to values in $[0, 1]$. As in the definition of the cut distance in the preliminaries, we handle graphs on different numbers of vertices by considering fractional overlays instead of blow-ups (and doubly stochastic matrices). Recall that a fractional overlay of graphs G and H is a matrix $X \in \mathbb{R}^{V(G)\times V(H)}$ such that $X_{uv} \geq 0$ for all $u \in V(G)$, $v \in V(H)$, $\sum_{v\in V(H)} X_{uv} = 1/\mathsf{v}(G)$ for every $u \in V(G)$, and $\sum_{u\in V(G)} X_{uv} = 1/\mathsf{v}(H)$ for every $v \in V(H)$. If $\mathsf{v}(G) = \mathsf{v}(H)$, then the difference between a fractional overlay and a doubly stochastic matrix is just a factor of $\mathsf{v}(G)$. Also recall that $\mathcal{X}(G, H)$ denotes the set of all fractional overlays of G and H.

We consider two matrix norms for the tree distance: First, just like in the definition of the cut distance, we use the *cut norm* for matrices, introduced by Frieze and Kannan [12], defined as $\|A\|_\square := \max_{S\subseteq[m],T\subseteq[n]} |\sum_{i\in S,j\in T} A_{ij}|$ for $A \in \mathbb{R}^{m\times n}$. Second, we also consider the more standard spectral norm $\|A\|_2 := \sup_{x\in\mathbb{R}^n,\|x\|_2\leq 1} \|Ax\|_2$ of a matrix $A \in \mathbb{R}^{m\times n}$. From a computational point of view, the Frobenius norm might also be appealing, but this would lead to a different topology, cf. [16, Appendix E].

▶ **Definition 5** (Tree Distance of Graphs). *Let G and H be graphs with adjacency matrices $A \in \mathbb{R}^{V(G)\times V(G)}$ and $B \in \mathbb{R}^{V(H)\times V(H)}$, respectively. Then, define*

$$\delta_\square^{\mathcal{T}}(G, H) := \inf_{X\in\mathcal{X}(G,H)} \frac{1}{\mathsf{v}(G)\cdot\mathsf{v}(H)} \|\mathsf{v}(H)\cdot AX - \mathsf{v}(G)\cdot XB\|_\square \text{ and}$$
$$\delta_2^{\mathcal{T}}(G, H) := \inf_{X\in\mathcal{X}(G,H)} \frac{1}{\sqrt{\mathsf{v}(G)\mathsf{v}(H)}} \|\mathsf{v}(H)\cdot AX - \mathsf{v}(G)\cdot XB\|_2.$$

Note that the spectral norm requires an adapted normalization factor in Definition 5. The advantage of $\delta_\square^{\mathcal{T}}$ is the close connection to the cut distance, which also utilizes the cut norm. However, the crucial advantage of the spectral norm is that minimization of the spectral norm of a matrix is a standard application of interior-point methods in convex optimization. In particular, an ε-solution to $\delta_2^{\mathcal{T}}$ can be computed in polynomial time [24, Section 6.3.3]. For $\delta_\square^{\mathcal{T}}$, it is not clear whether this is possible.

From the results of Section 5, we get that $\delta_\square^{\mathcal{T}}$ and $\delta_2^{\mathcal{T}}$ are pseudometrics (Lemma 16) and that two graphs have distance zero if and only if their tree homomorphism densities are the same (Lemma 18). Moreover, we have $\delta_\square^{\mathcal{T}} \leq \delta_\square$ (Lemma 19), and these pseudometrics are invariant under blow-ups. Finally, we get the following counting lemma (Corollary 20) and inverse counting lemma (Corollary 21).

▶ **Theorem 6** (Counting Lemma for $\delta^{\mathcal{T}}$, Graphs)**.** *Let $\delta^{\mathcal{T}} \in \{\delta_\square^{\mathcal{T}}, \delta_2^{\mathcal{T}}\}$. For every tree T and every $\varepsilon > 0$, there is an $\eta > 0$ such that, for all graphs G and H, if $\delta^{\mathcal{T}}(G, H) \leq \eta$, then $|t(T, G) - t(T, H)| \leq \varepsilon$.*

▶ **Theorem 7** (Inverse Counting Lemma for $\delta^{\mathcal{T}}$, Graphs)**.** *Let $\delta^{\mathcal{T}} \in \{\delta_\square^{\mathcal{T}}, \delta_2^{\mathcal{T}}\}$. For every $\varepsilon > 0$, there are $k > 0$ and $\eta > 0$ such that, for all graphs G and H, if $|t(T, G) - t(T, H)| \leq \eta$ for every tree T on at most k vertices, then $\delta^{\mathcal{T}}(G, H) \leq \varepsilon$.*

3.2 The Path Distance of Graphs

Dell, Grohe, and Rattan proved that two graphs G and H have the same path homomorphism counts if and only if the system $\mathsf{F}_{\text{iso}}(G, H)$ of linear equations has a real solution [7]. This transfers to the definition of the path distance, i.e., we define the path distance analogously to the tree distance but relax the non-negativity condition of fractional overlays. For graphs G and H, we call a matrix $X \in \mathbb{R}^{V(G)\times V(H)}$ a *signed fractional overlay* of G and H if $\|Xy\|_2 \leq \|y\|_2/\sqrt{\mathsf{v}(G)\mathsf{v}(H)}$ for every $y \in \mathbb{R}^{V(H)}$, $\sum_{v\in V(H)} X_{uv} = 1/\mathsf{v}(G)$ for every $u \in V(G)$, and $\sum_{u\in V(G)} X_{uv} = 1/\mathsf{v}(H)$ for every $v \in V(H)$. Let $\mathcal{S}(G, H)$ denote the set of all signed fractional overlays of G and H. The first condition requires that X is a contraction (up to a scaling factor) in the spectral norm; we need this to guarantee that our definition of the path distance actually yields a pseudometric. This restriction to the spectral norm stems from the fact that the proof of Dell, Grohe, and Rattan [7] (and our generalization thereof to graphons) only guarantees that the constructed solution is a contraction in the spectral norm, cf. Section 6 for the details.

▶ **Definition 8** (Path Distance of Graphs)**.** *Let G and H be graphs with adjacency matrices $A \in \mathbb{R}^{V(G)\times V(G)}$ and $B \in \mathbb{R}^{V(H)\times V(H)}$, respectively. Then, define*

$$\delta_2^{\mathcal{P}}(G, H) := \inf_{X\in\mathcal{S}(G,H)} \frac{1}{\sqrt{\mathsf{v}(G)\mathsf{v}(H)}} \|\mathsf{v}(H) \cdot AX - \mathsf{v}(G) \cdot XB\|_2.$$

From Section 6, we get that $\delta_2^{\mathcal{P}}$ is a pseudometric (Lemma 25) that is invariant under blow-ups and that has as graphs of distance zero precisely these with the same path homomorphism densities. Moreover, we get the following (quantitative) counting lemma (Corollary 27) and inverse counting lemma (Corollary 30).

▶ **Theorem 9** (Counting Lemma for $\delta_2^{\mathcal{P}}$, Graphs)**.** *Let P be a path, and let G and H be graphs. Then, $|t(P, G) - t(P, H)| \leq \mathsf{e}(P) \cdot \delta_2^{\mathcal{P}}(G, H)$.*

▶ **Theorem 10** (Inverse Counting Lemma for $\delta_2^{\mathcal{P}}$, Graphs)**.** *For every $\varepsilon > 0$, there are $k > 0$ and $\eta > 0$ such that, for all graphs G and H, if $|t(P, G) - t(P, H)| \leq \eta$ for every path P on at most k vertices, then $\delta_2^{\mathcal{P}}(G, H) \leq \varepsilon$.*

4 Graphon Pseudometrics and Homomorphism Densities

In this section, we provide the main tools we need to prove the correspondences between the tree and path distances and tree and path homomorphism densities, respectively. Consider a pseudometric δ on graphons. We say that δ *is compatible with* $\delta_\square$ if, for every sequence of graphons $(U_n)_n$, $U_n \in \mathcal{W}_0$, and every graphon $\widetilde{U} \in \mathcal{W}_0$, $\delta_\square(U_n, \widetilde{U}) \xrightarrow{n\to\infty} 0$ implies

$\delta(U_n, \widetilde{U}) \xrightarrow{n\to\infty} 0$. For example, this is the case if $\delta \leq \delta_\square$, i.e., graphons only get closer if we consider δ instead of $\delta_\square$. We anticipate that the pseudometrics we are interested in, the tree distance and the path distance, are compatible with $\delta_\square$.

Together, the next two theorems state that every pseudometric that is compatible with $\delta_\square$ and whose graphons of distance zero can be characterized by homomorphism densities from a class of graphs $\mathcal{F}$ already has to satisfy both a counting lemma and an inverse counting lemma for this class $\mathcal{F}$. The proof of these theorems is a simple compactness argument, utilizing the compactness of the graphon space, Theorem 4, and the counting lemma for $\delta_\square$, Lemma 2. Therefore, it is absolutely crucial that we consider a pseudometric defined on graphons as the limit of a sequence of graphs may not be a graph.

▶ **Theorem 11** (Counting Lemma for $\mathcal{F}$). *Let $\mathcal{F}$ be a class of graphs, and let $\delta^{\mathcal{F}}$ be a pseudometric on graphons such that (1) $\delta^{\mathcal{F}}$ is compatible with $\delta_\square$ and (2), for all graphons $U, W \in \mathcal{W}_0$, $\delta^{\mathcal{F}}(U, W) = 0$ implies $t(F, U) = t(F, W)$ for every graph $F \in \mathcal{F}$. Then, for every graph $F \in \mathcal{F}$ and every $\varepsilon > 0$, there is an $\eta > 0$ such that, for all graphons $U, W \in \mathcal{W}_0$, if $\delta^{\mathcal{F}}(U, W) \leq \eta$, then $|t(F, U) - t(F, W)| \leq \varepsilon$.*

Proof of Theorem 11. We proceed by contradiction and assume that the statement does not hold. Then, there is a graph $F \in \mathcal{F}$ and an $\varepsilon > 0$ such that, for every $\eta > 0$, there are graphons $U, W \in \mathcal{W}_0$ such that $\delta^{\mathcal{F}}(U, W) \leq \eta$ and $|t(F, U) - t(F, W)| > \varepsilon$.

Let $k > 0$. Then, by choosing $\eta = \frac{1}{k}$, we get that there are graphons $U_k, W_k \in \mathcal{W}_0$ such that $\delta^{\mathcal{F}}(U_k, W_k) \leq \frac{1}{k}$ and $|t(F, U_k) - t(F, W_k)| > \varepsilon$. By the compactness theorem, Theorem 4, we get that the sequence $(U_k)_k$ has a convergent subsequence $(U_{k_i})_i$ converging to a graphon $\widetilde{U}$ in the metric $\delta_\square$. By another application of that theorem, we get that $(W_{k_i})_i$ has a convergent subsequence $(W_{\ell_i})_i$ converging to a graphon $\widetilde{W}$ in the metric $\delta_\square$. Then, $(U_{\ell_i})_i$ and $(W_{\ell_i})_i$ are sequences converging to $\widetilde{U}$ and $\widetilde{W}$ in the metric $\delta_\square$, respectively.

Now, for every $i > 0$, we have

$$\delta^{\mathcal{F}}(\widetilde{U}, \widetilde{W}) \leq \delta^{\mathcal{F}}(\widetilde{U}, U_{\ell_i}) + \delta^{\mathcal{F}}(U_{\ell_i}, W_{\ell_i}) + \delta^{\mathcal{F}}(W_{\ell_i}, \widetilde{W}).$$

By assumption, we have $\delta^{\mathcal{F}}(U_{\ell_i}, W_{\ell_i}) \leq \frac{1}{\ell_i}$, which means that $\delta^{\mathcal{F}}(U_{\ell_i}, W_{\ell_i}) \xrightarrow{i\to\infty} 0$. Since $\delta_\square(U_{\ell_i}, \widetilde{U}) \xrightarrow{i\to\infty} 0$ and $\delta_\square(W_{\ell_i}, \widetilde{W}) \xrightarrow{i\to\infty} 0$, the first assumption about $\delta^{\mathcal{F}}$ yields that we also have $\delta^{\mathcal{F}}(U_{\ell_i}, \widetilde{U}) \xrightarrow{i\to\infty} 0$ and $\delta^{\mathcal{F}}(W_{\ell_i}, \widetilde{W}) \xrightarrow{i\to\infty} 0$. Hence, we must have $\delta^{\mathcal{F}}(\widetilde{U}, \widetilde{W}) = 0$.

Since $\delta^{\mathcal{F}}(\widetilde{U}, \widetilde{W}) = 0$, we have $t(F, \widetilde{U}) = t(F, \widetilde{W})$ by the second assumption about $\delta^{\mathcal{F}}$. By the Counting Lemma, Lemma 2, we get that $|t(F, U_{\ell_i}) - t(F, \widetilde{U})| \xrightarrow{i\to\infty} 0$ and $|t(F, \widetilde{W}) - t(F, W_{\ell_i})| \xrightarrow{i\to\infty} 0$. Now, for every $i > 0$, we have

$$|t(F, U_{\ell_i}) - t(F, W_{\ell_i})| \leq |t(F, U_{\ell_i}) - t(F, \widetilde{U})| + |t(F, \widetilde{U}) - t(F, \widetilde{W})| + |t(F, \widetilde{W}) - t(F, W_{\ell_i})|$$

Hence, $|t(F, U_{\ell_i}) - t(F, W_{\ell_i})| \xrightarrow{i\to\infty} 0$. This contradicts the fact that $|t(F, U_{\ell_i}) - t(F, W_{\ell_i})| > \varepsilon$ for every i. ◀

Just as the proof of Theorem 11, the proof of Theorem 12 only relies on the compactness of the graphon space and the counting lemma for $\delta_\square$, and not on a counting lemma for a specific class of graphs or the inverse counting lemma for $\delta_\square$.

▶ **Theorem 12** (Inverse Counting Lemma for $\mathcal{F}$). *Let $\mathcal{F}$ be a class of graphs, and let $\delta^{\mathcal{F}}$ be a pseudometric on graphons such that (1) $\delta^{\mathcal{F}}$ is compatible with $\delta_\square$ and (2), for all graphons $U, W \in \mathcal{W}_0$, $t(F, U) = t(F, W)$ for every graph $F \in \mathcal{F}$ implies $\delta^{\mathcal{F}}(U, W) = 0$. Then, for every $\varepsilon > 0$, there are $k > 0$ and $\eta > 0$ such that, for all graphons $U, W \in \mathcal{W}_0$, if $|t(F, U) - t(F, W)| \leq \eta$ for every graph $F \in \mathcal{F}$ on at most k vertices, then $\delta^{\mathcal{F}}(U, W) \leq \varepsilon$.*

Proof. We proceed by contradiction and assume that the statement does not hold. Then, there is an $\varepsilon > 0$ such that, for every $k > 0$ and every $\eta > 0$, there are graphons $U, W \in \mathcal{W}_0$ such that $|t(F, U) - t(F, W)| \leq \eta$ for every graph $F \in \mathcal{F}$ on at most k vertices but $\delta^{\mathcal{F}}(U, W) > \varepsilon$.

Let $k > 0$. Then, by choosing $\eta = \frac{1}{k}$, we get that there are graphons $U_k, W_k \in \mathcal{W}_0$ such that $|t(F, U_k) - t(F, W_k)| \leq \frac{1}{k}$ for every graph $F \in \mathcal{F}$ on at most k vertices and $\delta^{\mathcal{F}}(U_k, W_k) > \varepsilon$. By the compactness theorem, Theorem 4, we get that the sequence $(U_k)_k$ has a convergent subsequence $(U_{k_i})_i$ converging to a graphon $\widetilde{U}$ in the metric $\delta_\square$. By another application of that theorem, we get that $(W_{k_i})_i$ has a convergent subsequence $(W_{\ell_i})_i$ converging to a graphon $\widetilde{W}$ in the metric $\delta_\square$. Then, $(U_{\ell_i})_i$ and $(W_{\ell_i})_i$ are sequences converging to $\widetilde{U}$ and $\widetilde{W}$ in the metric $\delta_\square$, respectively.

Let $F \in \mathcal{F}$ be a graph. Now, for every $i > 0$, we have

$$|t(F, \widetilde{U}) - t(F, \widetilde{W})| \leq |t(F, \widetilde{U}) - t(F, U_{\ell_i})| + |t(F, U_{\ell_i}) - t(F, W_{\ell_i})| + |t(F, W_{\ell_i}) - t(F, \widetilde{W})|$$

By the counting lemma for $\delta_\square$, Lemma 2, we get that $|t(F, \widetilde{U}) - t(F, U_{\ell_i})| \xrightarrow{i \to \infty} 0$ and $|t(F, W_{\ell_i}) - t(F, \widetilde{W})| \xrightarrow{i \to \infty} 0$. Moreover, by assumption, we have $|t(F, U_{\ell_i}) - t(F, W_{\ell_i})| \leq \frac{1}{\ell_i}$ for large enough i, which means that also $|t(F, U_{\ell_i}) - t(F, W_{\ell_i})| \xrightarrow{i \to \infty} 0$. Hence, we must have $t(F, \widetilde{U}) = t(F, \widetilde{W})$.

As we have $t(F, \widetilde{U}) = t(F, \widetilde{W})$ for every graph $F \in \mathcal{F}$, the second assumption about $\delta^{\mathcal{F}}$ yields that $\delta^{\mathcal{F}}(\widetilde{U}, \widetilde{W}) = 0$. Since $\delta_\square(U_{\ell_i}, \widetilde{U}) \xrightarrow{i \to \infty} 0$ and $\delta_\square(W_{\ell_i}, \widetilde{W}) \xrightarrow{i \to \infty} 0$, we also have $\delta^{\mathcal{F}}(U_{\ell_i}, \widetilde{U}) \xrightarrow{i \to \infty} 0$ and $\delta^{\mathcal{F}}(W_{\ell_i}, \widetilde{W}) \xrightarrow{i \to \infty} 0$ by the first assumption about $\delta^{\mathcal{F}}$. Now, for every $i > 0$, we have

$$\delta^{\mathcal{F}}(U_{\ell_i}, W_{\ell_i}) \leq \delta^{\mathcal{F}}(U_{\ell_i}, \widetilde{U}) + \delta^{\mathcal{F}}(\widetilde{U}, \widetilde{W}) + \delta^{\mathcal{F}}(\widetilde{W}, W_{\ell_i}).$$

Hence, $\delta^{\mathcal{F}}(U_{\ell_i}, W_{\ell_i}) \xrightarrow{i \to \infty} 0$. This contradicts the fact that $\delta^{\mathcal{F}}(U_{\ell_i}, W_{\ell_i}) > \varepsilon$ for every i. ◀

5 Homomorphisms from Trees

In this section, we define the tree distance of graphons. To use the results from Section 4, we prove that the graphons of distance zero are precisely those with the same tree homomorphism densities (Lemma 18) and that the tree distance is compatible with the cut distance (Lemma 19). As for graphs, we define two variants of the tree distance, which yield the same topology (Lemma 17): one using the analogue of the cut norm and one using the analogue of the spectral norm.

5.1 Fractional Isomorphism of Graphons

Recall that two graphs G and H with adjacency matrices $A \in \mathbb{R}^{V(G) \times V(G)}$ and $B \in \mathbb{R}^{V(H) \times V(H)}$, respectively, are called fractionally isomorphic if there is a doubly stochastic matrix $X \in \mathbb{R}^{V(G) \times V(H)}$ such that $AX = XB$. Grebík and Rocha proved Theorem 13, which generalizes this to graphons [13]; doubly stochastic matrices become *Markov operators* [11]. An operator $S\colon L_2[0,1] \to L_2[0,1]$ is called a Markov operator if $S \geq 0$, i.e., $f \geq 0$ implies $S(f) \geq 0$, $S(\mathbf{1}) = \mathbf{1}$, and $S^*(\mathbf{1}) = \mathbf{1}$, where $\mathbf{1}$ is the all-one function on $[0,1]$. We denote the set of all Markov operators $S\colon L_2[0,1] \to L_2[0,1]$ by $\mathcal{M}$.

▶ **Theorem 13** ([13], Part of Theorem 1.2). *Let $U, W \in \mathcal{W}_0$ be graphons. There is a Markov operator $S\colon L_2[0,1] \to L_2[0,1]$ such that $T_U \circ S = S \circ T_W$ if and only if $t(T, U) = t(T, W)$ for every tree T.*

5.2 The Tree Distance

Recall that, for graphons $U, W \in \mathcal{W}_0$, the cut distance of U and W can be written as $\delta_\square(U,W) = \inf_{\varphi \in S_{[0,1]}} \sup_{f,g\colon [0,1]\to[0,1]} \left|\langle f, T_{U-W^\varphi} g\rangle\right|$. We obtain the tree distance of U and W by relaxing measure-preserving maps to Markov operators.

▶ **Definition 14** (Tree Distance). *Let $U, W \in \mathcal{W}_0$ be graphons. Then, define*

$$\delta_\square^{\mathcal{T}}(U,W) := \inf_{S \in \mathcal{M}} \sup_{f,g\colon [0,1]\to[0,1]} |\langle f, (T_U \circ S - S \circ T_W) g\rangle| \text{ and}$$

$$\delta_{2\to2}^{\mathcal{T}}(U,W) := \inf_{S \in \mathcal{M}} \|T_U \circ S - S \circ T_W\|_{2\to2}.$$

As the notation $\delta_\square^{\mathcal{T}}$ indicates, the definition of $\delta_\square^{\mathcal{T}}$ is based (although not explicitly) on the cut norm, while $\delta_{2\to2}^{\mathcal{T}}$ is defined via the operator norm $\|\cdot\|_{2\to2}$, which corresponds to the spectral norm for matrices. One can verify that these definitions specialize to the ones for graphs from Section 3.1.

▶ **Lemma 15.** *Let G and H be graphs. Then, $\delta_\square^{\mathcal{T}}(G,H) = \delta_\square^{\mathcal{T}}(W_G, W_H)$ and $\delta_2^{\mathcal{T}}(G,H) = \delta_{2\to2}^{\mathcal{T}}(W_G, W_H)$.*

We verify that the tree distance actually is a pseudometric. To prove the triangle inequality for $\delta_\square^{\mathcal{T}}$ and $\delta_{2\to2}^{\mathcal{T}}$, we use that a Markov operator is a contraction on $L_\infty[0,1]$ and $L_2[0,1]$, respectively [11, Theorem 13.2 b)].

▶ **Lemma 16.** *$\delta_\square^{\mathcal{T}}$ and $\delta_{2\to2}^{\mathcal{T}}$ are pseudometrics on $\mathcal{W}_0$.*

The Riesz-Thorin Interpolation Theorem (see, e.g., [1, Theorem 1.1.1]) allows to prove that both variants of the tree distance define the same topology.

▶ **Lemma 17.** *Let $U, W \in \mathcal{W}_0$ be graphons. Then, $\delta_\square^{\mathcal{T}}(U,W) \le \delta_{2\to2}^{\mathcal{T}}(U,W) \le 4\delta_\square^{\mathcal{T}}(U,W)^{1/2}$.*

To be able to apply the results from Section 4, we need that the tree distance of two graphons is zero if and only if their tree homomorphism densities are the same. Let $U, W \in \mathcal{W}_0$ be graphons. From the respective definitions, it is not immediately clear that $\delta_\square^{\mathcal{T}}(U,W) = 0$ or $\delta_{2\to2}^{\mathcal{T}}(U,W) = 0$ implies $t(T,U) = t(T,W)$ for every tree T since the infimum over all Markov operators might not be attained. Here, we can use a continuity argument as the set of Markov operators is compact in the weak operator topology [11, Theorem 13.8]. However, we have to take a detour via a third variant of the tree distance where compactness in the weak operator topology suffices. All the details can be found in the full version of the paper.

▶ **Lemma 18.** *Let $U, W \in \mathcal{W}_0$ be graphons. Then, $\delta_\square^{\mathcal{T}}(U,W) = 0$ if and only if $t(T,U) = t(T,W)$ for every tree T.*

The *Koopman operator* $T_\varphi\colon f \mapsto f \circ \varphi$ of a measure-preserving map $\varphi\colon [0,1] \to [0,1]$ is a Markov operator [11, Example 13.1, 3)]. Hence, the tree distance can be seen as the relaxation of the cut distance obtained by relaxing measure-preserving maps to Markov operators. In particular, this means that the tree distance is compatible with the cut distance.

▶ **Lemma 19.** *Let $U, W \in \mathcal{W}_0$ be graphons. Then, $\delta_\square^{\mathcal{T}}(U,W) \le \delta_\square(U,W)$.*

With Lemma 18 and Lemma 19 we can apply the theorems of Section 4 and get both a counting lemma and an inverse counting lemma for the tree distance.

▶ **Corollary 20** (Counting Lemma for $\delta^{\mathcal{T}}$). *Let $\delta^{\mathcal{T}} \in \{\delta_\square^{\mathcal{T}}, \delta_{2\to2}^{\mathcal{T}}\}$. For every tree T and every $\varepsilon > 0$, there is an $\eta > 0$ such that, for all graphons $U, W \in \mathcal{W}_0$, if $\delta^{\mathcal{T}}(U,W) \le \eta$, then $|t(T,U) - t(T,W)| \le \varepsilon$.*

▶ **Corollary 21** (Inverse Counting Lemma for $\delta^{\mathcal{T}}$). *Let $\delta^{\mathcal{T}} \in \{\delta^{\mathcal{T}}_{\square}, \delta^{\mathcal{T}}_{2\to 2}\}$. For every $\varepsilon > 0$, there are $k > 0$ and $\eta > 0$ such that, for all graphons $U, W \in \mathcal{W}_0$, if $|t(T,U) - t(T,W)| \le \eta$ for every tree T on k vertices, then $\delta^{\mathcal{T}}(U,W) \le \varepsilon$.*

6 Homomorphisms from Paths

In this section, we define the path distance of graphons. We prove a quantitative counting lemma for it (Corollary 27) and only rely on the results from Section 4 to obtain an inverse counting lemma. To this end, we we prove that the graphons of distance zero are precisely those with the same path homomorphism densities (Lemma 28) and that the path distance is compatible with the cut distance (Lemma 29). Since there is no existing characterization of graphons with the same path homomorphism densities that we can rely on, we first generalize the result of Dell, Grohe, and Rattan to graphons (Theorem 22).

6.1 Path Densities and Graphons

Dell, Grohe, and Rattan have shown the surprising fact that G and H have the same path homomorphism counts if and only if the system $\mathsf{F}_{\text{iso}}(G,H)$ has a real solution [7]. We need a generalization of their characterization to graphons in order to define the path distance of graphons and apply the results from Section 4. If two graphons $U, W \in \mathcal{W}_0$ have the same path homomorphism densities, the proof of Theorem 22 yields an operator $S\colon L_2[0,1] \to L_2[0,1]$ such that $S(\mathbf{1}) = \mathbf{1}$ and $S^*(\mathbf{1}) = \mathbf{1}$, which generalizes the result of [7] in a straight-forward fashion. An important detail is that the proof also yields that S is an L_2-contraction; this guarantees that the path distance satisfies the triangle inequality, i.e., that it is a pseudometric in the first place. For the sake of brevity, we call an operator $S\colon L_2[0,1] \to L_2[0,1]$ a *signed Markov operator* if S is an L_2-contraction, i.e., $\|Sf\|_2 \le \|f\|_2$ for every $f \in L_2[0,1]$, $S(\mathbf{1}) = \mathbf{1}$, and $S^*(\mathbf{1}) = \mathbf{1}$. Let $\mathcal{S}$ denote the set of all signed Markov operators. It is easy to see that $\mathcal{S}$ is closed under composition and Hilbert adjoints.

▶ **Theorem 22.** *Let $U, W \in \mathcal{W}_0$. There is a signed Markov operator $S\colon L_2[0,1] \to L_2[0,1]$ such that $T_U \circ S = S \circ T_W$ if and only if $t(P,U) = t(P,W)$ for every path P.*

Homomorphism densities from paths can be expressed in terms of operator powers. For $\ell \ge 0$, let P_ℓ denote the path of length ℓ. Then, for a graphon U, we have

$$t(P_\ell, U) = \int_{[0,1]^{\ell+1}} \prod_{i\in[\ell]} U(x_i, x_{i+1}) \prod_{i\in[\ell+1]} dx_i = \langle \mathbf{1}, T_U^\ell \mathbf{1}\rangle$$

for every $\ell \ge 0$. The proof of Theorem 22 utilizes the Spectral Theorem for compact operators on Hilbert spaces to express $\mathbf{1}$ as a sum of orthogonal eigenfunctions. For a kernel $W \in \mathcal{W}$, $T_W\colon L_2[0,1] \to L_2[0,1]$ is a Hilbert-Schmidt operator and, hence, compact [20]. Since $L_2[0,1]$ is separable and T_W is compact and self-adjoint, the Spectral Theorem yields that there is a countably infinite orthonormal basis $\{f'_i\}$ of $L_2[0,1]$ consisting of eigenfunctions of T_W with the corresponding multiset of eigenvalues $\{\lambda_n\} \subseteq \mathbb{R}$ such that $\lambda_n \xrightarrow{n\to\infty} 0$ (see, e.g., [9]). If graphons U and W have the same path homomorphism densities, an interpolation argument yields that the lengths of the eigenvectors in the decomposition of $\mathbf{1}$ and their eigenvalues have to be the same. Then, one can define the operator S from these eigenfunctions of U and W. The detailed proof can be found in the full version of the paper.

6.2 The Path Distance

We define the *path distance* of graphons can analogously to the tree distance. However, as the proof Theorem 22 does not yield that the resulting operator is an L_∞-contraction, we are limited in our choice of norms.

▶ **Definition 23** (Path Distance)**.** *Let $U, W \in \mathcal{W}_0$ be graphons. Then, define*

$$\delta^{\mathcal{P}}_{2\to 2}(U, W) := \inf_{S \in \mathcal{S}} \|T_U \circ S - S \circ T_W\|_{2\to 2}.$$

One can verify that this defines a pseudometric that specializes to the one for graphs from Section 3.2.

▶ **Lemma 24.** *Let G and H be graphs. Then, $\delta^{\mathcal{P}}_2(G, H) = \delta^{\mathcal{P}}_{2\to 2}(W_G, W_H)$.*

▶ **Lemma 25.** *$\delta^{\mathcal{P}}_{2\to 2}$ is a pseudometric on $\mathcal{W}_0$.*

To apply the theorems of Section 4, we need that two graphons have distance zero in the path distance if and only if their path homomorphism densities are the same and that $\delta^{\mathcal{P}}_{2\to 2}$ is compatible with $\delta_\square$. For the former, we deviate from the way we proceeded for the tree distance as we actually can prove a quantitative counting lemma.

▶ **Theorem 26** (Counting Lemma for Paths)**.** *Let P be a path, and let $U, W \in \mathcal{W}_0$ be graphons. Then, for every operator $S \colon L_2[0, 1] \to L_2[0, 1]$ with $S(\mathbf{1}) = \mathbf{1}$ and $S^*(\mathbf{1}) = \mathbf{1}$,*

$$|t(P, U) - t(P, W)| \le \mathsf{e}(P) \cdot \sup_{f,g\colon [0,1]\to[0,1]} |\langle f, (T_U \circ S - S \circ T_W)g\rangle|.$$

Proof. Let $\ell \in \mathbb{N}$ and $S \in \mathcal{S}$. Then,

$$\begin{aligned}
|t(P_\ell, U) - t(P_\ell, W)| &= |\langle \mathbf{1}, T_U^\ell(S\mathbf{1})\rangle - \langle (S^*\mathbf{1}), T_W^\ell \mathbf{1}\rangle| \\
&= \Big| \sum_{i\in[\ell]} \big(\langle \mathbf{1}, (T_U^{\ell-i+1} \circ S \circ T_W^{i-1})\mathbf{1}\rangle - \langle \mathbf{1}, (T_U^{\ell-i} \circ S \circ T_W^{i})\mathbf{1}\rangle\big) \Big| \\
&= \Big| \sum_{i\in[\ell]} \langle T_U^{\ell-i}\mathbf{1}, (T_U \circ S - S \circ T_W)(T_W^{i-1}\mathbf{1})\rangle \Big| \\
&\le \ell \cdot \sup_{f,g\colon [0,1]\to[0,1]} |\langle f, (T_U \circ S - S \circ T_W)g\rangle|.
\end{aligned}$$ ◀

Theorem 26 suggests that, for graphons $U, W \in \mathcal{W}_0$, one should define

$$\delta^{\mathcal{P}}_\square(U, W) := \inf_{S\in\mathcal{S}} \sup_{f,g\colon [0,1]\to[0,1]} |\langle f, (T_U \circ S - S \circ T_W)g\rangle|.$$

Then, we have $|t(P, U) - t(P, W)| \le \mathsf{e}(P) \cdot \delta^{\mathcal{P}}_\square(U, W)$ for every path P. However, as mentioned before, we cannot verify that $\delta^{\mathcal{P}}_\square$ is a pseudometric as the operator S might not be an L_∞-contraction.

▶ **Corollary 27** (Counting Lemma for $\delta^{\mathcal{P}}_{2\to 2}$)**.** *Let P be a path, and let $U, W \in \mathcal{W}_0$ be graphons. Then, $|t(P, U) - t(P, W)| \le \mathsf{e}(P) \cdot \delta^{\mathcal{P}}_{2\to 2}(U, W)$.*

Proof. By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\sup_{f,g\colon [0,1]\to[0,1]} |\langle f, (T_U \circ S - S \circ T_W)g\rangle| &\le \sup_{f,g\colon [0,1]\to[0,1]} \|f\|_2 \|(T_U \circ S - S \circ T_W)g\|_2 \\
&\le \sup_{g\colon [0,1]\to[0,1]} \|T_U \circ S - S \circ T_W\|_{2\to 2} \|g\|_2 \\
&\le \|T_U \circ S - S \circ T_W\|_{2\to 2}
\end{aligned}$$

for every operator $S \colon L_2[0, 1] \to L_2[0, 1]$. Hence, the statement follows from Theorem 26. ◀

With this explicit counting lemma, we obtain that two graphons have distance zero in the path distance if and only if their path homomorphism densities are the same.

▶ **Lemma 28.** *Let $U, W \in \mathcal{W}_0$ be graphons. Then, $\delta^{\mathcal{P}}_{2\to 2}(U, W) = 0$ if and only if $t(P, U) = t(P, W)$ for every path P.*

Proof. If $\delta^{\mathcal{P}}_{2\to 2}(U, W) = 0$, then Corollary 27 yields that $t(P, U) = t(P, W)$ for every path P. On the other hand, if $t(P, U) = t(P, W)$ for every path P, then there is a signed Markov operator $S \in \mathcal{S}$ with $T_U \circ S = S \circ T_W$ by Theorem 22. Then, $\delta^{\mathcal{P}}_{2\to 2}(U, W) = 0$ follows immediately from the definition. ◀

By definition, the path distance is bounded from above by the tree distance (with the appropriate norm), which means that it also is compatible with the cut distance.

▶ **Lemma 29.** *Let $U, W \in \mathcal{W}_0$ be graphons. Then, $\delta^{\mathcal{P}}_{2\to 2}(U, W) \le \delta^{\mathcal{T}}_{2\to 2}(U, W)$.*

With these lemmas, we can apply Theorem 12 and obtain the following inverse counting lemma for the path distance.

▶ **Corollary 30** (Inverse Counting Lemma for $\delta^{\mathcal{P}}_{2\to 2}$). *For every $\varepsilon > 0$, there are $k > 0$ and $\eta > 0$ such that, for all graphons $U, W \in \mathcal{W}_0$, if $|t(P, U) - t(P, W)| \le \eta$ for every path P on at most k vertices, then $\delta^{\mathcal{P}}_{2\to 2}(U, W) \le \varepsilon$.*

7 The Color Distance

Color Refinement, also known as the *1-dimensional Weisfeiler-Leman algorithm*, is a heuristic graph isomorphism test. It computes a coloring of the vertices of a graph in a sequence of *refinement rounds*; we say that color refinement *distinguishes* two graphs if the computed color patterns differ. Formally, for a graph G, we let $C^G_0(u) = 1$ for every $u \in V(G)$ and $C^G_{i+1}(u) = \{\!\{C^G_i(v) \mid uv \in E(G)\}\!\}$ for every $i \ge 0$. Let $C^G_\infty = C^G_i$ for the smallest i such that $C^G_i(u) = C^G_i(v) \iff C^G_{i+1}(u) = C^G_{i+1}(v)$ for all $u, v \in G$ ("C_i is stable"). Then, color refinement distinguishes two graphs G and H if there is an $i \ge 0$ such that $\{\!\{C^G_i(v) \mid v \in V(G)\}\!\} \ne \{\!\{C^H_i(v) \mid v \in V(H)\}\!\}$. It is well-known that the partition $\{C^{-1}_\infty(i) \mid i \in C_\infty(V(G))\}$ is the *coarsest equitable partition* of $V(G)$, where a partition Π of $V(G)$ is called *equitable* if for all $P, Q \in \Pi$ and $u, v \in P$, the vertices u and v have the same number of neighbors in Q.

For a graph G, we can define a weighted graph G/C^G_∞ by letting $V(G/C^G_\infty) := \{C^{-1}_\infty(i) \mid i \in C_\infty(V(G))\}$, $\alpha_C(G/C^G_\infty) := |C|$ for $C \in V(G/C^G_\infty)$, and $\beta_{CD}(G/C^G_\infty) := M^G_{CD}/|D|$ for all $C, D \in V(G/C^G_\infty)$, where M^G_{CD} is the number of neighbors a vertex from C has in D, which is the same for all vertices in C as the partition induced by the colors of C^G_∞ is equitable. Note that we have $|C|M^G_{CD} = |D|M^G_{DC}$ as both products describe the number of edges between C and D, i.e., G/C^G_∞ is well-defined. Usually, when talking about the invariant $\mathcal{I}^2_C$ computed by color refinement (see, e.g., [17]), one does not normalize M^G_{CD} by $|D|$. However, by doing so, we do not only get a weighted graph (with symmetric edge weights), but the graphs G and G/C^G_∞ actually have the same tree homomorphism counts. Grebík and Rocha already introduced the graphon analogue $U/\mathcal{C}(U)$ of G/C^G_∞ and proved the same fact for it [13, Corollary 4.3]; hence, we omit the proof.

▶ **Lemma 31.** *Let T be a tree, and let G be a graph. Then, $\mathsf{hom}(T, G) = \mathsf{hom}(T, G/C^G_\infty)$.*

By the result of Dvořák [10], G/C^G_∞ and H/C^H_∞ are isomorphic if and only if G and H have the same tree homomorphism counts. Hence, it is tempting to define a tree distance-like similarity measure on graphs by simply considering the cut distance of G/C^G_∞ and H/C^H_∞.

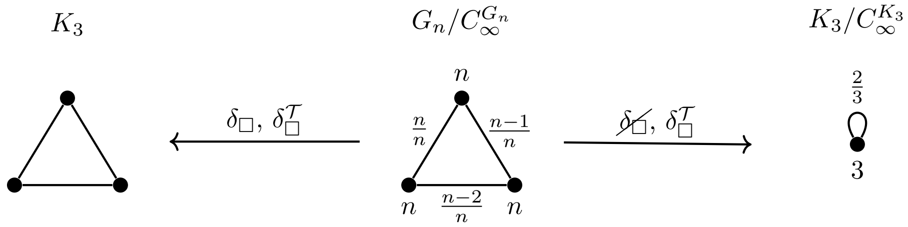

Figure 1 An example separating the color distance from the tree distance.

For graphs G and H, we call $\delta_\square^{\mathcal{C}}(G, H) := \delta_\square(G/C_\infty^G, H/C_\infty^H)$ the *color distance* of G and H. As the cut distance $\delta_\square$ is a pseudometric on graphs, so is $\delta_\square^{\mathcal{C}}$. For $\delta_\square^{\mathcal{C}}$, we immediately obtain a quantitative counting lemma from Lemma 2 and Lemma 31.

▶ **Corollary 32** (Counting Lemma for $\delta_\square^{\mathcal{C}}$). *Let T be a tree, and let G and H be graphs. Then, $|t(T, G) - t(T, H)| \leq |E(T)| \cdot \delta_\square^{\mathcal{C}}(G, H)$.*

Clearly, $\delta_\square^{\mathcal{T}}$ and $\delta_\square^{\mathcal{C}}$ have the same graphs of distance zero. Moreover, one can easily verify that the tree distance is bounded from above by the color distance.

▶ **Lemma 33.** *Let G and H be graphs. Then, $\delta_\square^{\mathcal{T}}(G, H) \leq \delta_\square^{\mathcal{C}}(G, H)$.*

Proof. We have $\delta_\square^{\mathcal{T}}(G, H) = \delta_\square^{\mathcal{T}}(G/C_\infty^G, H/C_\infty^H) \leq \delta_\square(G/C_\infty^G, H/C_\infty^H) = \delta_\square^{\mathcal{C}}(G, H)$ by Lemma 31 and Lemma 19. ◀

Now, the obvious question is whether these pseudometrics are the same or, at least, define the same topology. But it is not hard to find a counterexample; the color distance sees differences between graphs that the tree distance and tree homomorphisms do not see. In particular, an inverse counting lemma cannot hold for the color distance. See Figure 1, and for the moment, assume that we can construct a sequence $(G_n)_n$ of graphs such that $G_n/C_\infty^{G_n}$ is as depicted. It is easy to verify that $\delta_\square(G_n/C_\infty^{G_n}, K_3) \xrightarrow{n\to\infty} 0$, and thus, both $\delta_\square^{\mathcal{T}}(G_n, K_3) \xrightarrow{n\to\infty} 0$ and $|t(T, G_n) - t(T, K_3)| \xrightarrow{n\to\infty} 0$ for every tree T. But, $\delta_\square^{\mathcal{C}}(G_n, K_3) \geq \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$ for every n since $G_n/C_\infty^{G_n}$ has a vertex without a loop.

The existence of graphs G_n such that $G_n/C_\infty^{G_n}$ is as depicted in Figure 1 follows easily from inversion results for the color refinement invariant $\mathcal{I}_C^2$. Otto first proved that $\mathcal{I}_C^2$ admits polynomial time inversion on structures [25], and Kiefer, Schweitzer, and Selman gave a simple construction to show that $\mathcal{I}_C^2$ admits linear-time inversion on the class of graphs [17]. Basically, we partition $3n$ vertices into three sets of size n and add edges between these partitions such that they induce n-, $(n-1)$-, and $(n-2)$-regular bipartite graphs.

The example in Figure 1 actually answers an open question of Grebík and Rocha [13, Question 3.1]. They ask whether the set $\{W/\mathcal{C}(W) \mid W \in \widetilde{\mathcal{W}}_0\}$ is closed in $\widetilde{\mathcal{W}}_0$: it is not. With a more refined argument, we can actually show that $\{W_{G/C_\infty^G} \mid G \text{ graph}\}$ is already dense in $\widetilde{\mathcal{W}}_0$. By properly rounding the weights of a given weighted graph, we can turn the inversion result of [17] into a statement about approximate inversion.

▶ **Theorem 34.** *Let H be a weighted graph. For every $n \geq 2 \cdot \mathsf{v}(H)$, there is a graph G on n^2 vertices such that $\delta_\square(G/C_\infty^G, H) \leq 3 \cdot \mathsf{v}(H)/n + \frac{1}{4} \cdot (\mathsf{v}(H)/n)^2$.*

In Theorem 34, the size of the resulting graph depends on how close we want it to be to the input graph. A simple consequence of the compactness of the graphon space is that, for $\varepsilon > 0$, we can approximate any graphon with an error of ε in $\delta_\square$ by a graph on $N(\varepsilon)$

vertices, where $N(\varepsilon)$ is independent of the graphon [20, Corollary 9.25]. With Theorem 34, this implies that the same is possible with the weighted graphs G/C^G_∞. This also means that the closure of the set $\{W_{G/C^G_\infty} \mid G \text{ graph}\}$ is already $\widetilde{\mathcal{W}}_0$.

8 Conclusions

We have introduced similarity measures for graphs that can be formulated as convex optimization problems and shown surprising correspondences to tree and path homomorphism densities. This takes previous results on the "expressiveness" of homomorphism counts from an exact to an approximate level. Moreover, it helps to give a theoretical understanding of kernel methods in machine learning, which are often based on counting certain substructures in graphs. Proving the correspondences to homomorphism densities was made possible by introducing our similarity measures for the more general case of graphons, where tools from functional analysis let us prove the general statement that every "reasonably defined" pseudometric has to satisfy a correspondence to homomorphism densities.

Various open questions remain. The compactness argument used in Section 4 only yields non-quantitative statements. Hence, we do not know how close the graphs have to be in the pseudometric for their homomorphism densities to be close and vice versa. Only for paths we were able to prove a quantitative counting lemma, which uses the same factor $\mathsf{e}(F)$ as the counting lemma for general graphs. It seems conceivable that a quantitative counting lemma for trees that uses the same factor $\mathsf{e}(T)$ also holds. As the proof of the quantitative inverse counting lemma is quite involved [3, 20], proving such statements for trees and paths should not be easy.

More in reach seems to be the question of how the tree distance generalizes to the class $\mathcal{T}_k$ of graphs of treewidth at most k. Homomorphism counts from graphs in $\mathcal{T}_k$ can also be characterized in terms of linear equations in the case of graphs [10] (see also [7]). How does such a characterization for graphons look like? And how does one define a distance measure from this?

Another open question concerns further characterizations of fractional isomorphism, e.g., the color refinement algorithm, which gives a characterization based on equitable partitions. Can one prove a correspondence between the tree distance and, say, ε-equitable partitions? It is not hard to come up with a definition for such partitions; the hard part is to prove that graphs that are similar in the tree distance possess such a partition.

References

1 Jöran Bergh and Jörgen Löfström. *Interpolation Spaces: An Introduction*. Die Grundlehren Der Mathematischen Wissenschaften in Einzeldarstellungen. Springer-Verlag, 1976. `doi:10.1007/978-3-642-66451-9`.

2 Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, Balázs Szegedy, and Katalin Vesztergombi. Graph limits and parameter testing. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '06, page 261–270, New York, NY, USA, 2006. Association for Computing Machinery. `doi:10.1145/1132516.1132556`.

3 Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801–1851, 2008. `doi:10.1016/j.aim.2008.07.008`.

4 Jan Böker. Graph similarity and homomorphism densities, 2021. `arXiv:2104.14213`.

5 Radu Curticapean, Holger Dell, and Dániel Marx. Homomorphisms are a good basis for counting small subgraphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, page 210–223, New York, NY, USA, 2017. Association for Computing Machinery. `doi:10.1145/3055399.3055502`.

6 Víctor Dalmau and Peter Jonsson. The complexity of counting homomorphisms seen from the other side. *Theoretical Computer Science*, 329(1):315–323, 2004. doi:10.1016/j.tcs.2004.08.008.

7 Holger Dell, Martin Grohe, and Gaurav Rattan. Lovász Meets Weisfeiler and Leman. In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*, volume 107 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 40:1–40:14, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.ICALP.2018.40.

8 Richard M. Dudley. *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2002. doi:10.1017/CBO9780511755347.

9 Nelson Dunford and Jacob T. Schwartz. *Linear Operators, Part 2: Spectral Theory, Self Adjoint Operators in Hilbert Space*. Linear Operators. Interscience Publishers, 1963.

10 Zdeněk Dvořák. On recognizing graphs by numbers of homomorphisms. *Journal of Graph Theory*, 64(4):330–342, 2010. doi:10.1002/jgt.20461.

11 Tanja Eisner, Bálint Farkas, Markus Haase, and Rainer Nagel. *Operator Theoretic Aspects of Ergodic Theory*. Graduate Texts in Mathematics. Springer International Publishing, 2015. doi:10.1007/978-3-319-16898-2.

12 Alan Frieze and Ravindran Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19:175–220, February 1999. doi:10.1007/s004930050052.

13 Jan Grebík and Israel Rocha. Fractional isomorphism of graphons, 2021. Accepted to Combinatorica. arXiv:1909.04122.

14 Martin Grohe. Counting bounded tree depth homomorphisms. In *Proceedings of the 35th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '20, page 507–520, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3373718.3394739.

15 Martin Grohe. Word2vec, node2vec, graph2vec, x2vec: Towards a theory of vector embeddings of structured data. In *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, PODS'20, page 1–16, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3375395.3387641.

16 Svante Janson. Graphons, cut norm and distance, couplings and rearrangements. *New York Journal of Mathematics*, 4, 2013.

17 Sandra Kiefer, Pascal Schweitzer, and Erkal Selman. Graphs identified by logics with counting. In Giuseppe F Italiano, Giovanni Pighizzini, and Donald T. Sannella, editors, *Mathematical Foundations of Computer Science 2015*, pages 319–330, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg. doi:10.1007/978-3-662-48057-1_25.

18 László Lovász. Operations with structures. *Acta Mathematica Hungarica*, 18(3-4):321–328, 1967. doi:10.1007/BF02280291.

19 László Lovász and Balázs Szegedy. Szemerédi's lemma for the analyst. *GAFA Geometric And Functional Analysis*, 17:252–270, 2007. doi:10.1007/s00039-007-0599-6.

20 László Lovász. *Large Networks and Graph Limits*, volume 60 of *Colloquium Publications*. American Mathematical Society, 2012. doi:10.1090/coll/060.

21 László Lovász and Balázs Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006. doi:10.1016/j.jctb.2006.05.002.

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

23 Viswanath Nagarajan and Maxim Sviridenko. On the maximum quadratic assignment problem. *Mathematics of Operations Research*, 34(4):859–868, 2009. doi:10.1287/moor.1090.0418.

24 Yurii Nesterov and Arkadii Nemirovskii. *Interior-point Polynomial Algorithms in Convex Programming*. Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, 1994. doi:10.1137/1.9781611970791.

25 Martin Otto. Canonization for two variables and puzzles on the square. *Annals of Pure and Applied Logic*, 85(3):243–282, 1997. doi:10.1016/S0168-0072(96)00047-4.

26 Gottfried Tinhofer. Graph isomorphism and theorems of birkhoff-type. *Computing*, 36(4):285–300, 1986. doi:10.1007/BF02240204.

27 Gottfried Tinhofer. A note on compact graphs. *Discrete Applied Mathematics*, 30(2):253–264, 1991. doi:10.1016/0166-218X(91)90049-3.