Homomorphism Tensors and Linear Equations

Martin Grohe
RWTH Aachen University, Germany

Gaurav Rattan
RWTH Aachen University, Germany

Tim Seppelt
RWTH Aachen University, Germany

Abstract

Lovász (1967) showed that two graphs G and H are isomorphic if and only if they are *homomorphism indistinguishable* over the class of all graphs, i.e. for every graph F, the number of homomorphisms from F to G equals the number of homomorphisms from F to H. Recently, homomorphism indistinguishability over restricted classes of graphs such as bounded treewidth, bounded treedepth and planar graphs, has emerged as a surprisingly powerful framework for capturing diverse equivalence relations on graphs arising from logical equivalence and algebraic equation systems.

In this paper, we provide a unified algebraic framework for such results by examining the linear-algebraic and representation-theoretic structure of tensors counting homomorphisms from labelled graphs. The existence of certain linear transformations between such homomorphism tensor subspaces can be interpreted both as homomorphism indistinguishability over a graph class and as feasibility of an equational system. Following this framework, we obtain characterisations of homomorphism indistinguishability over several natural graph classes, namely trees of bounded degree, graphs of bounded pathwidth (answering a question of Dell et al. (2018)), and graphs of bounded treedepth.

2012 ACM Subject Classification Mathematics of computing → Discrete mathematics

Keywords and phrases homomorphisms, labelled graphs, treewidth, pathwidth, treedepth, linear equations, Sherali–Adams relaxation, Wiegmann–Specht Theorem, Weisfeiler–Leman

Digital Object Identifier 10.4230/LIPIcs.ICALP.2022.70

Category Track A: Algorithms, Complexity and Games

Related Version *Preprint*: https://arxiv.org/abs/2111.11313 [22]

Funding *Gaurav Rattan*: DFG Research Grants Program RA 3242/1-1 via project number 411032549.
Tim Seppelt: German Research Council (DFG) within Research Training Group 2236 (UnRAVeL).

Acknowledgements We thank Andrei Bulatov for many fruitful discussions on the foundations of theory developed here and its relation to the algebraic theory of valued constraint satisfaction problems. Moreover, we thank Jan Böker for discussions about linear systems of equations and basal graphs. Finally, we thank the anonymous reviewers for suggestions for improvement.

1 Introduction

Representations in terms of homomorphism counts provide a surprisingly rich view on graphs and their properties. Homomorphism counts have direct connections to logic [14, 17, 27], category theory [12, 34], the graph isomorphism problem [13, 14, 27], algebraic characterisations of graphs [13], and quantum groups [32]. Counting subgraph patterns in graphs has a wide range of applications, for example in graph kernels (see [24]) and motif counting (see [1, 33]). Homomorphism counts can be used as a flexible basis for counting all kinds of substructures [11], and their complexity has been studied in great detail (e.g. [8, 9, 11, 39]). It has been argued in [18] that homomorphism counts are well-suited as a theoretical foundation for analysing graph embeddings and machine learning techniques on graphs, both indirectly through their connection with graph neural networks via the Weisfeiler–Leman algorithm [14, 35, 46] and directly as features for machine learning on graphs. The latter has also been confirmed experimentally [4, 25, 37].

The starting point of the theory is an old result due to Lovász [27]: two graphs G, H are isomorphic if and only if for every graph F, the number $\hom(F, G)$ of homomorphisms from F to G equals $\hom(F, H)$. For a class $\mathcal{F}$ of graphs, we say that G and H are *homomorphism indistinguishable* over $\mathcal{F}$ if and only if $\hom(F, G) = \hom(F, H)$ for all $F \in \mathcal{F}$. A beautiful picture that has only emerged in the last few years shows that homomorphism indistinguishability over natural graph classes, such as paths, trees, or planar graphs, characterises a variety of natural equivalence relations on graphs.

Broadly speaking, there are two types of such results, the first relating homomorphism indistinguishability to logical equivalence, and the second giving algebraic characterisations of homomorphism equivalence derived from systems of linear (in)equalities for graph isomorphism. Examples of logical characterisations of homomorphism equivalence are the characterisation of homomorphism indistinguishability over graphs of treewidth at most k in terms of the $(k + 1)$-variable fragment of first-order logic with counting [14] and the characterisation of homomorphism indistinguishability over graphs of treedepth at most k in terms of the quantifier-rank-k fragment of first-order logic with counting [17]. Results of this type have also been described in a general category theoretic framework [12, 34]. Examples of equational characterisations are the characterisation of homomorphism indistinguishability over trees in terms of fractional isomorphism [13, 14, 44], which may be viewed as the LP relaxation of a natural ILP for graph isomorphism, and a generalisation to homomorphism indistinguishability over graph of bounded treewidth in terms of the Sherali–Adams hierarchy over that ILP [3, 13, 21, 14, 31]. Further examples include a characterisation of homomorphism indistinguishability over paths in terms of the same system of equalities by dropping the non-negativity constraints of fractional isomorphism [13], and a characterisation of homomorphism indistinguishability over planar graphs in terms of quantum isomorphism [32]. Remarkably, quantum isomorphism is derived from interpreting the same system of linear equations over C*-algebras [2].

1.1 Results

Two questions that remained open in [13] are (1) whether the equational characterisation of homomorphism indistinguishability over paths can be generalised to graphs of bounded pathwidth in a similar way as the characterisation of homomorphism indistinguishability over trees can be generalised to graphs of bounded treewidth, and (2) whether homomorphism indistinguishability over graphs of bounded degree suffices to characterise graphs up to isomorphism. In this paper, we answer the first question affirmatively.

▶ **Theorem 1.** *For every $k \geq 1$, the following are equivalent for two graphs G and H:*
1. *G and H are homomorphism indistinguishable over graphs of pathwidth at most k.*
2. *The $(k+1)$-st level relaxation $\mathsf{L}^{k+1}_{\text{iso}}(G, H)$ of the standard ILP for graph isomorphism has a rational solution.*

The detailed description of the system $\mathsf{L}^{k+1}_{\text{iso}}(G, H)$ is provided in Section 5. In fact, we also devise an alternative system of linear equations $\mathsf{PW}^{k+1}(G, H)$ characterising homomorphism indistinguishability over graphs of pathwidth at most k. The definition of this system turns out to be very natural from the perspective of homomorphism counting, and as we explain later, it forms a fruitful instantiation of a more general representation-theoretic framework for homomorphism indistinguishability.

Moreover, we obtain an equational characterisation of homomorphism indistinguishability over graphs of bounded treedepth. The resulting system $\mathsf{TD}^k(G, H)$ is very similar to $\mathsf{L}^k_{\text{iso}}(G, H)$ and $\mathsf{PW}^k(G, H)$, except that variables are indexed by (ordered) k-tuples of variables rather than sets of at most k variables, which reflects the order induced by the recursive definition of treedepth.

▶ **Theorem 2.** *For every $k \geq 1$, the following are equivalent for two graphs G and H:*

1. *G and H are homomorphism indistinguishable over graphs of treedepth at most k,*
2. *The linear systems of equations $\mathsf{TD}^k(G, H)$ has a non-negative rational solution,*
3. *The linear systems of equations $\mathsf{TD}^k(G, H)$ has a rational solution.*

Along with [17], the above theorem implies that the logical equivalence of two graphs G and H over the quantifier-rank-k fragment of first-order logic with counting can be characterised by the feasibility of the system $\mathsf{TD}^k(G, H)$ of linear equations.

We cannot answer the second open question from [13], but we prove a partial negative result: homomorphism indistinguishability over trees of bounded degree is strictly weaker than homomorphism indistinguishability over all trees.

▶ **Theorem 3.** *For every integer $d \geq 1$, there exist graphs G and H such that G and H are homomorphism indistinguishable over trees of degree at most d, but G and H are not homomorphism indistinguishable over the class of all trees.*

In conjunction with [13], the above theorem yields the following corollary: counting homomorphisms from trees of bounded degree is strictly less powerful than the classical Colour Refinement algorithm [20], in terms of their ability to distinguish non-isomorphic graphs.

To prove these results, we develop a general theory that enables us to derive some of the existing results as well as the new results in a unified algebraic framework exploiting a duality between algebraic varieties of "tensor maps" derived from homomorphism counts over families of rooted graphs and equationally defined equivalence relations, which are based on transformations of graphs in terms of unitary or, more often, pseudo-stochastic or doubly-stochastic matrices. (We call a matrix over the complex numbers *pseudo-stochastic* if its row and column sums are all 1, and we call it *doubly-stochastic* if it is pseudo-stochastic and all its entries are non-negative reals.) The foundations of this theory have been laid in [13] and, mainly, [32]. Some ideas can also be traced back to the work on homomorphism functions and connection matrices [15, 28, 29, 40], and a similar duality, called Galois connection there, that is underlying the algebraic theory of constraint satisfaction problems [6, 7, 42, 47].

1.2 Techniques

To explain our core new ideas, let us start from a simple and well-known result: two symmetric real matrices A, B are co-spectral if and only if for every $k \geq 1$ the matrices A^k and B^k have the same trace. If A, B are the adjacency matrices of two graph G, H, the latter can be phrased graph theoretically as: for every k, G and H have the same number of closed walks of length k, or equivalently, the numbers of homomorphisms from a cycle C_k of length k to G and to H are the same. Thus, G and H are homomorphism indistinguishable over the class of all cycles if and only if they are co-spectral. Note next that the graphs, or their adjacency matrices A, B, are co-spectral if and only if there is a unitary matrix U (or orthogonal matrix, but we need to work over the complex numbers) such that $UA = BU$. Now, in [14, 13] it was proved that G, H are homomorphism indistinguishable over the class of all paths if and only if there is a pseudo-stochastic matrix X such that $XA = BX$, and they are homomorphism indistinguishable over the class of all trees if and only if there is a doubly-stochastic matrix X such that $XA = BX$. From an algebraic perspective, the transition from a unitary matrix in the cycle result to a pseudo-stochastic in the path result is puzzling: where unitary matrices are very natural, pseudo-stochastic matrices are much less so from an algebraic point of view. Moving on to the tree result, we suddenly add non-negativity constraints – where do they come from? Our theory presented in Section 3 provides a uniform and very transparent explanation for the three results. It also allows us to analyse homomorphism indistinguishability over d-ary trees, for every $d \geq 1$, and to prove that it yields a strict hierarchy of increasingly finer equivalence relations.

Now suppose we want to extend these results to edge coloured graphs. Each edge-coloured graph corresponds to a family of matrices, one for each colour. Theorems due to Specht [41] and Wiegmann [45] characterise families of matrices that are simultaneously equivalent with respect to a unitary transformation. Interpreted over coloured graphs, the criterion provided by these theorems can be interpreted as homomorphism indistinguishability over coloured cycles. One of our main technical contributions is a variant of these theorems that establishes a correspondence between simultaneous equivalence with respect to pseudo-stochastic transformations and homomorphism indistinguishability over coloured paths. The proof is based on basic representation theory, in particular the character theory of semisimple algebras.

Interpreting graphs of bounded pathwidth in a "graph-grammar style" over coloured paths using graphs of bounded size as building blocks, we give an equational characterisation of homomorphism indistinguishability over graphs of pathwidth at most k. After further manipulations, we even obtain a characterisation in terms of a system of equations that are derived by lifting the basic equations for paths in a Sherali–Adams style. (The basic idea of these lifted equations goes back to [3].) This answers the open question from [13] stated above. In the same way, we can lift the characterisations of homomorphism indistinguishability over trees to graphs of treewidth k, and we can also establish a characterisation of homomorphism indistinguishability over graphs of "cyclewidth" k, providing a uniform explanation for all these results. Finally, we combine these techniques to prove a characterisation of homomorphism indistinguishability over graphs of treedepth k in terms of a novel system of linear equations.

2 Preliminaries

We briefly state the necessary definitions and, along the way, introduce our notation. We assume familiarity with elementary definitions from graph theory and linear algebra. As usual, let $\mathbb{N} = \{1, 2, 3, \dots\}$, $[n] = \{1, \dots, n\}$, and $(n) = (1, \dots, n)$. All mentioned graphs are simple, loopless, and undirected.

2.1 Labelled Graphs and Tensor Maps

Labelled and Bilabelled Graphs. For $\ell \in \mathbb{N}$, an *ℓ-labelled* graph $\boldsymbol{F}$ is a tuple $\boldsymbol{F} = (F, \boldsymbol{v})$ where F is a graph and $\boldsymbol{v} \in V(F)^\ell$. The vertices in $\boldsymbol{v}$ are not necessarily distinct, i.e. vertices may have several labels.

The operation of *gluing* two ℓ-labelled graphs $\boldsymbol{F} = (F, \boldsymbol{u})$ and $\boldsymbol{F'} = (F', \boldsymbol{u}')$ yields the ℓ-labelled graph $\boldsymbol{F} \odot \boldsymbol{F'}$ obtained by taking the disjoint union of F and F' and pairwise identifying the vertices $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ to become the i-th labelled vertex, for $i \in [\ell]$, and removing any multiedges in the process. In fact, since we consider homomorphisms into simple graphs, multiedges can always be omitted. Likewise, self-loops can also be disregarded since the number of homomorphisms $F \to G$ where F has a self-loop and G does not is always zero. We henceforth tacitly assume that all graphs are simple.

For $\ell_1, \ell_2 \in \mathbb{N}$, an (ℓ_1, ℓ_2)-*bilabelled graph* $\boldsymbol{F}$ is a tuple $(F, \boldsymbol{u}, \boldsymbol{v})$ for $\boldsymbol{u} \in V(F)^{\ell_1}$, $\boldsymbol{v} \in V(F)^{\ell_2}$. If $\boldsymbol{u} = (u_1, \dots, u_{\ell_1})$ and $\boldsymbol{v} = (v_1, \dots, v_{\ell_2})$, it is usual to say that the vertex u_i, resp. v_i, is labelled with the i-th *in-label*, resp. *out-label*.

The *reverse* of an (ℓ_1, ℓ_2)-bilabelled graph $\boldsymbol{F} = (F, \boldsymbol{u}, \boldsymbol{v})$ is defined to be the (ℓ_2, ℓ_1)-bilabelled graph $\boldsymbol{F}^* = (F, \boldsymbol{v}, \boldsymbol{u})$ with roles of in- and out-labels interchanged. The *concatenation* or *series composition* of an (ℓ_1, ℓ_2)-bilabelled graph $\boldsymbol{F} = (F, \boldsymbol{u}, \boldsymbol{v})$ and an (ℓ_2, ℓ_3)-bilabelled graph $\boldsymbol{F}' = (F', \boldsymbol{u}', \boldsymbol{v}')$, $\ell_3 \in \mathbb{N}$, denoted by $\boldsymbol{F} \cdot \boldsymbol{F}'$ is the (ℓ_1, ℓ_3)-bilabelled graph obtained by taking the disjoint union of F and F' and identifying for all $i \in [\ell_2]$ the vertices $\boldsymbol{v}_i$ and $\boldsymbol{u}'_i$. The in-labels of $\boldsymbol{F} \cdot \boldsymbol{F}'$ lie on $\boldsymbol{u}$ while its out-labels are positioned on $\boldsymbol{v}'$. The *parallel composition* of (ℓ_1, ℓ_2)-bilabelled graphs $\boldsymbol{F} = (F, \boldsymbol{u}, \boldsymbol{v})$ and $\boldsymbol{F}' = (F', \boldsymbol{u}', \boldsymbol{v}')$ denoted by $\boldsymbol{F} \odot \boldsymbol{F}'$ is obtained by taking the disjoint union of F and F' and identifying $\boldsymbol{u}_i$ with $\boldsymbol{u}'_i$, and $\boldsymbol{v}_j$ with with $\boldsymbol{v}'_j$ for $i \in [\ell_1]$ and $j \in [\ell_2]$.

Tensors and Tensor Maps. For a set V and $k \in \mathbb{N}$, the set of all functions $X \colon V^k \to \mathbb{C}$ forms a complex vector space denoted by $\mathbb{C}^{V^k}$. We call the elements of $\mathbb{C}^{V^k}$ the k-*dimensional tensors* over V. We identify 0-dimensional tensors with scalars, i.e. $\mathbb{C}^{V^0} = \mathbb{C}$. Furthermore, 1-dimensional tensors are vectors in $\mathbb{C}^V$, 2-dimensional tensors are matrices in $\mathbb{C}^{V \times V}$, et cetera.

A k-*dimensional tensor map on graphs* is a function φ that maps graphs G to k-dimensional tensors $\varphi_G \in \mathbb{C}^{V(G)^k}$. A k-dimensional tensor map φ is *equivariant* if for all isomorphic graphs G and H, all isomorphisms f from G to H, and all $\boldsymbol{v} \in V(G)^k$ it holds that $\varphi_G(\boldsymbol{v}) = \varphi_H(f(\boldsymbol{v}))$.

Homomorphism Tensors and Homomorphism Tensor Maps. For graphs F and G, let $\hom(F, G)$ denote the number of homomorphisms from F to G, i.e. the number of mappings $h \colon V(F) \to V(G)$ such that $v_1 v_2 \in E(F)$ implies $h(v_1)h(v_2) \in E(G)$. For an ℓ-labelled graph $\boldsymbol{F} = (F, \boldsymbol{v})$ and $\boldsymbol{w} \in V(G)^\ell$, let $\hom(\boldsymbol{F}, G, \boldsymbol{w})$ denote the number of homomorphisms h from F to G such that $h(\boldsymbol{v}_i) = \boldsymbol{w}_i$ for all $i \in [\ell]$. Analogously, for an (ℓ_1, ℓ_2)-bilabelled graph $\boldsymbol{F}' = (F', \boldsymbol{u}, \boldsymbol{v})$ and $\boldsymbol{x} \in V(G)^{\ell_1}$, $\boldsymbol{y} \in V(G)^{\ell_2}$, let $\hom(F', G, \boldsymbol{x}, \boldsymbol{y})$ denote the number of homomorphisms $h \colon F' \to G$ such that $h(\boldsymbol{u}_i) = \boldsymbol{x}_i$ and $h(\boldsymbol{v}_j) = \boldsymbol{y}_j$ for all $i \in [\ell_1]$, $j \in [\ell_2]$. More succinctly, we write $\boldsymbol{F}_G \in \mathbb{C}^{V(G)^\ell}$ for the *homomorphism tensor* defined by letting $\boldsymbol{F}_G(\boldsymbol{w}) \coloneqq \hom(\boldsymbol{F}, G, \boldsymbol{w})$ for all $\boldsymbol{w} \in V(G)^\ell$. Similarly, for a bilabelled graph $\boldsymbol{F}'$, $\boldsymbol{F}'_G \in \mathbb{C}^{V(G)^{\ell_1} \times V(G)^{\ell_2}}$ is the matrix defined as $\boldsymbol{F}'_G(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \hom(\boldsymbol{F}, G, \boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{x} \in V(G)^{\ell_1}$, $\boldsymbol{y} \in V(G)^{\ell_2}$.

Letting this construction range over all right-hand side graphs G, the map $G \mapsto \boldsymbol{F}_G$ becomes a tensor map, the *homomorphism tensor map* induced by $\boldsymbol{F}$. It is easy to see that homomorphism tensor maps are equivariant.

Homomorphism tensors give rise to the complex vector spaces of our main interest and their endomorphisms. For a set $\mathcal{R}$ of ℓ-labelled graphs, the tensors $\boldsymbol{R}_G$ for $\boldsymbol{R} \in \mathcal{R}$ span a subspace of $\mathbb{C}^{V(G)^\ell}$, which is denoted by $\mathbb{C}\mathcal{R}_G$. Moreover, the tensors $\boldsymbol{S}_G$ for an (ℓ, ℓ)-bilabelled graph $\boldsymbol{S}$ induces an endomorphisms of $\mathbb{C}^{V(G)^\ell}$.

▶ **Example 4.** For $k \geq 1$, let $\boldsymbol{1}^k$ denote the labelled graph consisting of k isolated vertices with distinct labels $(1, \dots, k)$. Then, $\boldsymbol{1}^k_G$ is the uniform tensor in $\mathbb{C}^{V(G)^k}$ with every entry equal to 1. Let $\boldsymbol{A}$ denote the $(1,1)$-bilabelled graph $\Big(\overset{1}{\bullet}\!\!-\!\!\!-\!\!\overset{2}{\bullet}, (1), (2)\Big)$. For every graph G, the matrix $\boldsymbol{A}_G$ is the adjacency matrix of G.

Algebraic and Combinatorial Operations on Homomorphism Tensor Maps. Tensor maps naturally admit a variety of algebraic operations. These include linear combination, complex conjugation, and permutation of coordinates, which are readily defined. Crucially, many operations when applied to homomorphism tensor maps correspond to operations on (bi)labelled graphs. This observation due to [30, 32] is illustrated by the following examples.

- *Sum of Entries = Dropping Labels.* Given a k-labelled graph $\boldsymbol{F} = (F, \boldsymbol{u})$, let $\mathrm{soe}(\boldsymbol{F})$ denote the 0-labelled graph $(F, ())$. Then, for every graph G, $\mathrm{soe}(\boldsymbol{F})_G = \hom(F, G) = \sum_{\boldsymbol{v} \in V(G)^k} \boldsymbol{F}_G(\boldsymbol{v}) =: \mathrm{soe}(\boldsymbol{F}_G)$.
- *Matrix Product = Series Composition.* Let an (ℓ_1, ℓ_2)-bilabelled graph $\boldsymbol{F} = (F, \boldsymbol{u}, \boldsymbol{v})$ and an (ℓ_2, ℓ_3)-bilabelled graph $\boldsymbol{F'} = (F', \boldsymbol{u'}, \boldsymbol{v'})$ be given. Then for every graph G, vertices $\boldsymbol{x} \in V(G)^{\ell_1}$, and $\boldsymbol{y} \in V(G)^{\ell_2}$, $(\boldsymbol{F} \cdot \boldsymbol{F'})_G(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{w} \in V(G)^{\ell_2}} \boldsymbol{F}_G(\boldsymbol{x}, \boldsymbol{w}) \boldsymbol{F'}_G(\boldsymbol{w}, \boldsymbol{y}) =: (\boldsymbol{F}_G \cdot \boldsymbol{F'}_G)(\boldsymbol{x}, \boldsymbol{y})$. A similar operation corresponds to the matrix-vector product, where $\boldsymbol{F'}$ is assumed to be ℓ_2-labelled.
- *Schur Product = Parallel Composition.* The parallel composition $\boldsymbol{F} \odot \boldsymbol{F'}$ of two k-labelled graphs $\boldsymbol{F} = (F, \boldsymbol{u})$ and $\boldsymbol{F'} = (F', \boldsymbol{u'})$ corresponds to the Schur product of the homomorphism tensors. That is, for every graph G and $\boldsymbol{v} \in V(G)^k$, $(\boldsymbol{F} \odot \boldsymbol{F})_G(\boldsymbol{v}) = \boldsymbol{F}_G(\boldsymbol{v}) \boldsymbol{F'}_G(\boldsymbol{v}) =: (\boldsymbol{F}_G \odot \boldsymbol{F'}_G)(\boldsymbol{v})$. Moreover, the *inner-product* of ℓ-labelled graphs $\boldsymbol{F}$, $\boldsymbol{F'}$ can be defined by $\langle \boldsymbol{F}, \boldsymbol{F'} \rangle := \mathrm{soe}(\boldsymbol{F} \odot \boldsymbol{F'})$. It corresponds to the standard inner-product on the tensor space.

2.2 Representation Theory of Involution Monoids

We recall standard notions from representation theory, cf. [26]. A *monoid* Γ is a possibly infinite set equipped with an associative binary operation and an identity element denoted by 1_Γ. An example for a monoid is the *endomorphism monoid* $\operatorname{End} V$ for a vector space V over $\mathbb{C}$ with composition as binary operation and id_V as identity element. A *monoid representation* of Γ is a map $\varphi \colon \Gamma \to \operatorname{End} V$ such that $\varphi(1_\Gamma) = \mathrm{id}_V$ and $\varphi(gh) = \varphi(g)\varphi(h)$ for all $g, h \in \Gamma$. The representation is *finite-dimensional* if V is finite-dimensional. For every monoid Γ, there exists a representation, for example the *trivial representation* $\Gamma \to \operatorname{End}\{0\}$ given by $g \mapsto \mathrm{id}_{\{0\}}$.

Let $\varphi \colon \Gamma \to \operatorname{End}(V)$ and $\psi \colon \Gamma \to \operatorname{End}(W)$ be two representations. Then φ and ψ are *equivalent* if there exists a vector space isomorphism $X \colon V \to W$ such that $X\varphi(g) = \psi(g)X$ for all $g \in \Gamma$. Moreover, φ is a *subrepresentation* of ψ if $V \leq W$ and $\psi(g)$ restricted to V equals $\varphi(g)$ for all $g \in \Gamma$. A representation φ is *simple* if its only subrepresentations are the trivial representation and φ itself. The *direct sum* of φ and ψ denoted by $\varphi \oplus \psi \colon \Gamma \to \operatorname{End}(V \oplus W)$ is the representation that maps $g \in \Gamma$ to $\varphi(g) \oplus \psi(g) \in \operatorname{End}(V) \oplus \operatorname{End}(W) \leq \operatorname{End}(V \oplus W)$. A representation φ is *semisimple* if it is the direct sum of simple representations.

Let $\varphi \colon \Gamma \to \operatorname{End} V$ be a representation with subrepresentations $\psi' \colon \Gamma \to \operatorname{End} V'$ and $\psi'' \colon \Gamma \to \operatorname{End} V''$. Then the restriction of φ to $V' \cap V''$ is a representation as well, called the *intersection of* ψ' *and* ψ''. For a set $S \subseteq V$, define the *subrepresentation of* φ *generated by* S as the intersection of all subrepresentations $\psi' \colon \Gamma \to \operatorname{End} V'$ of φ such that $S \subseteq V'$.

The *character* of a representation φ is the map $\chi_\varphi \colon \Gamma \to \mathbb{C}$ defined as $g \mapsto \operatorname{tr}(\varphi(g))$. Its significance stems from the following theorem, which can be traced back to Frobenius and Schur [16]. For a contemporary proof, consult [26] from whose Theorem 7.19 the statement follows.

▶ **Theorem 5** (Frobenius–Schur [16]). *Let* Γ *be a monoid. Let* $\varphi \colon \Gamma \to \operatorname{End}(V)$ *and* $\psi \colon \Gamma \to \operatorname{End}(W)$ *be finite-dimensional semisimple representations. Then* φ *and* ψ *are equivalent if and only if* $\chi_\varphi = \chi_\psi$.

The monoids studied in this work are equipped with an additional structure which ensures that their finite-dimensional representations are always semisimple: An *involution monoid* is a monoid Γ with a unary operation $^*\colon \Gamma \to \Gamma$ such that $(gh)^* = h^*g^*$ and $(g^*)^* = g$ for all $g, h \in \Gamma$. Note that $\operatorname{End} V$ is an involution monoid with the adjoint operation $X \mapsto X^*$. Representations of involution monoids must preserve the involution operations. Thereby, they correspond to representations of $*$-algebras.

▶ **Lemma 6.** *Let Γ be an involution monoid. Every finite-dimensional representation of Γ is semisimple.*

Proof. Let $\varphi\colon \Gamma \to \operatorname{End} V$ be a finite-dimensional representation of Γ. It suffices to show that for every subrepresentation $\psi\colon \Gamma \to \operatorname{End} W$ of φ there exists a subrepresentation $\psi'\colon \Gamma \to \operatorname{End} W'$ of φ such that $\varphi = \psi \oplus \psi'$, i.e. φ acts as ψ on W and as ψ' on W'. Set W' to be the orthogonal complement of W in V. It has to be shown that $\varphi(g) \in \operatorname{End} V$ for every $g \in \Gamma$ can be restricted to an endomorphism of W'. Let $w \in W$ and $w' \in W'$ be arbitrary. Then $\langle \varphi(g)w', w\rangle = \langle w', \varphi(g)^*w\rangle = \langle w', \varphi(g^*)w\rangle = 0$ since $\varphi(g^*)$ maps $W \to W$ and $W \perp W'$. Hence, the $\operatorname{im}\varphi(g)$ is contained in the orthogonal complement of W, which equals W'. Clearly, $\varphi = \psi \oplus \psi'$. ◀

2.3 Path and Cycle Decompositions of Bilabelled Graphs

We recall the well-studied notions of path and tree decompositions. For illustrating subsequent arguments, we introduce cycle decompositions.

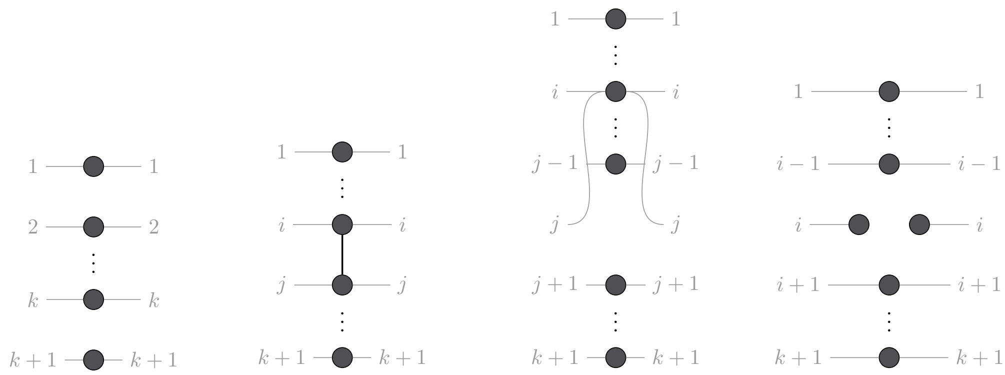
▶ **Definition 7.** *A* decomposition *of a graph G is a pair (F, β) where F is a graph and β is map $V(F) \to 2^{V(G)}$ such that*

1. *the union of the $\beta(v)$ for $v \in V(F)$ is equal to $V(G)$,*
2. *for every edge $e \in E(G)$ there exists $v \in V(F)$ such that $e \subseteq \beta(v)$,*
3. *for every vertex $u \in V(G)$ the set of vertices $v \in V(F)$ such that $u \in \beta(v)$ is connected in F.*

The sets $\beta(v)$ for $v \in V(F)$ are called the *bags of* (F, β). The *width* of (F, β) is the maximum over all $|\beta(v)| + 1$ for $v \in V(F)$. A decomposition (F, β) is called a *tree decomposition* if F is a tree, a *path decomposition* if F is a path, and a *cycle decomposition* if F is a cycle. The *tree- / path- / cyclewidth* of a graph G is the minimum width of a tree/path/cycle decomposition of G.

Let $k \in \mathbb{N}$. A *leaf bag* of a path decomposition (P, β) is a bag $\beta(v)$ such that $v \in V(P)$ has degree 1. A *path decomposition* of a $(k+1, k+1)$-bilabelled graph $\boldsymbol{F} = (F, \boldsymbol{u}, \boldsymbol{v})$ is a path decomposition (P, β) of the underlying graph F such that the leaf bags consist precisely of the vertices occurring (possibly repeatedly) in $\boldsymbol{u}$ and in $\boldsymbol{v}$, respectively. A $(k+1, k+1)$-bilabelled graph $\boldsymbol{F}$ is said to be of *pathwidth at most k* if its underlying graph admits a path decomposition of width at most k with this property.

Let $\mathcal{PW}^k$ denote the set of all $(k+1, k+1)$-bilabelled graphs of pathwidth at most k. Every unlabelled graph F of pathwidth at most k can be turned into a $(k+1, k+1)$-bilabelled graph $\boldsymbol{F} = (F, \boldsymbol{u}, \boldsymbol{v})$ of pathwidth at most k by assigning labels to the vertices $\boldsymbol{u}, \boldsymbol{v} \in V(F)^{k+1}$ in the leaf bags. The set $\mathcal{PW}^k$ is closed under concatenation and taking reverses. The *identity graph* $\boldsymbol{I} = (I, (1, \dots, k+1), (1, \dots, k+1))$ with $V(I) = [k+1]$, $E(I) = \emptyset$ is the multiplicative identity under concatenation. Hence, $\mathcal{PW}^k$ forms an involution monoid. A generating set for $\mathcal{PW}^k$ under these operations is called a k-basal set:



(a) Identity graph $\boldsymbol{I}$. **(b)** Adjacency graph $\boldsymbol{A}^{ij}$. **(c)** Identification graph $\boldsymbol{I}^{ij}$. **(d)** Forgetting graph $\boldsymbol{F}^{i}$.

Figure 1 Basal graphs from Lemma 9 in wire notation of [32]: A vertex carries in-label (out-label) i if it is connected to the number i on the left (right) by a wire. Actual edges and vertices of the graph are depicted in black.

▶ **Definition 8.** *A finite set $\mathcal{B}^k$ of $(k+1, k+1)$-bilabelled graphs is called a k-*basal set *if it satisfies the following properties:*
1. $\mathcal{B}^k \subseteq \mathcal{PW}^k$*,*
2. *the identity graph $\boldsymbol{I}$ is contained in $\mathcal{B}^k$,*
3. *for every $\boldsymbol{B} \in \mathcal{B}$, the reverse graph $\boldsymbol{B}^*$ also belongs to $\mathcal{B}$, and,*
4. *every $\boldsymbol{P} \in \mathcal{PW}^k$ can be obtained by concatenating a sequence of elements from $\mathcal{B}$.*

Concrete examples of k-basal sets can be constructed for every k as described in Lemma 9 and Figure 1. In fact, in all what follows, every k-basal set can be assumed to be this particular k-basal set.

▶ **Lemma 9.** *The set $\mathcal{B}^k$ consisting of the following $(k+1, k+1)$-bilabelled graphs is k-basal. For $1 \leq i \neq j \leq k+1$,*
- *the* identity graph $\boldsymbol{I} = (I, (1, \dots, k+1), (1, \dots, k+1))$ *with* $V(I) = [k+1]$, $E(I) = \emptyset$*,*
- *the* adjacency graphs $\boldsymbol{A}^{ij} = (A^{ij}, (k+1), (k+1))$ *with* $V(A^{ij}) = [k+1]$ *and* $E(A) = \{ij\}$*,*
- *the* identification graphs $\boldsymbol{I}^{ij} = (I^{ij}, (1, \dots, i, i+1, \dots, j-1, i, j+1, \dots, k+1), (1, \dots, i, i+1, \dots, j-1, i, j+1, \dots, k+1))$ *with* $V(I^{ij}) = [k+1] \setminus \{j\}$ *and* $E(I^{ij}) = \emptyset$*, and*
- *the* forgetting graphs $\boldsymbol{F}^{i} = (F^i, (1, \dots, k+1), (1, \dots, i-1, i', i+1, \dots, k+1))$ *with* $V(F^i) = [k+1] \cup \{i'\}$ *and* $E(F^i) = \emptyset$*.*

Proof. Items 1 and 3 of Definition 8 are clear. For Item 4, observe that every $\boldsymbol{P} = (P, \boldsymbol{v}, \boldsymbol{v}) \in \mathcal{PW}^k$ such that all vertices of $\boldsymbol{P}$ are labelled with corresponding in- and out-labels coinciding can be written as the concatenation of $\prod_{ij \in I} \boldsymbol{A}^{ij}$ for $I = E(P)$ with the $\boldsymbol{I}^{ij}$ for all $i \neq j$ such that $\boldsymbol{v}_i = \boldsymbol{v}_j$. Arbitrary $\boldsymbol{Q} \in \mathcal{PW}^k$ can then be obtained as the concatenation of such $\boldsymbol{P}$ interleaved with $\boldsymbol{F}^i$ for certain i. This corresponds to linking adjacent bags of the path decomposition together. ◀

Crucial is the following proposition which is immediate from the above observations:

▶ **Proposition 10.** *Let $\mathcal{B}^k$ denote a k-basal set. If F is a graph of pathwidth at most k then there exist $\boldsymbol{B}^1, \dots, \boldsymbol{B}^r \in \mathcal{B}^k$ such that $\hom(F, G) = \operatorname{soe}(\boldsymbol{B}^1_G \cdots \boldsymbol{B}^r_G)$ for all graphs G.*

The constructions for graphs of bounded pathwidth carry over to graphs of bounded cyclewidth (Definition 7). Let $\boldsymbol{F} = (F, \boldsymbol{u}, \boldsymbol{v})$ be a $(k+1, k+1)$-bilabelled graph of pathwidth at most k. Let $\boldsymbol{F}^{\text{id}}$ denote the $(k+1)$-labelled graph obtained by identifying the elements of $\boldsymbol{u}$ and $\boldsymbol{v}$ element-wise. Every unlabelled graph C of cyclewidth k can be associated with a $(k+1, k+1)$-bilabelled graph $\boldsymbol{C}$ of pathwidth k such that C is the unlabelled graph underlying $\boldsymbol{C}^{\mathrm{id}}$. Observe that $\operatorname{soe}(\boldsymbol{F}^{\mathrm{id}}) = \operatorname{tr}(\boldsymbol{F})$.

▶ **Proposition 11.** *Let $\mathcal{B}^k$ denote a k-basal set. If F is a graph of cyclewidth at most k then there exist $\boldsymbol{B}^1, \dots, \boldsymbol{B}^r \in \mathcal{B}^k$ such that $\hom(F, G) = \operatorname{tr}(\boldsymbol{B}^1_G \cdots \boldsymbol{B}^r_G)$ for all graphs G.*

3 Homomorphisms from Trees, Paths, and Trees of Bounded Degree

Two graphs G, H with adjacency matrices $\boldsymbol{A}_G$, $\boldsymbol{A}_H$ are isomorphic if and only if there is a matrix X over the non-negative integers such that $X\boldsymbol{A}_G = \boldsymbol{A}_H X$ and $X\mathbf{1} = X^T\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the all-ones vector. Writing the constraints as linear equations whose variables are the entries of X, we obtain a system $\mathsf{F}_{\mathrm{iso}}(G, H)$ that has a non-negative integer solution if and only if G and H are isomorphic. A combination of results from [44] and [14] shows that $\mathsf{F}_{\mathrm{iso}}(G, H)$ has a non-negative rational solution if and only if G and H are homomorphism indistinguishable over the class of trees, and by [13], $\mathsf{F}_{\mathrm{iso}}(G, H)$ has an arbitrary rational solution if and only if G and H are homomorphism indistinguishable over the class of paths. We devise a more general framework connecting homomorphism indistinguishability and $\mathsf{F}_{\mathrm{iso}}(G, H)$-style equations, with paths and trees as two special cases. On the way, we characterise homomorphism indistinguishability over trees of bounded degree.

The prime objects of our study are sets $\mathcal{R}$ of 1-labelled graphs. For a fixed target graph G, the corresponding homomorphism tensors yield a subspace $\mathbb{C}\mathcal{R}_G \leq \mathbb{C}^{V(G)}$. Since algebraic operations on 1-dimensional tensors (i.e. vectors) and combinatorial operations on 1-labelled graphs correspond to each other, cf. Section 2.1, the existence of *linear transformations* $X \colon \mathbb{C}\mathcal{R}_G \to \mathbb{C}\mathcal{R}_H$ respecting these algebraic operations is central to conceptualising homomorphism indistinguishability of G and H as solvability of *linear equations*.

▶ **Definition 12.** *Recall the definition of $\boldsymbol{A}$ from Example 4. Let $\mathcal{R}$ denote a set of 1-labelled graphs containing the one-vertex graph.*

1. *The set $\mathcal{R}$ is* $\boldsymbol{A}$-invariant *if for all $\boldsymbol{R} \in \mathcal{R}$ also $\boldsymbol{A} \cdot \boldsymbol{R} \in \mathcal{R}$. Combinatorially, $\boldsymbol{A}$-invariance means that for every labelled graph $\boldsymbol{R} = (R, u) \in \mathcal{R}$, the labelled graph $\boldsymbol{A} \cdot \boldsymbol{R}$ obtained by adding a fresh vertex u' to R, adding the edge uu', and placing the label on u', is also in $\mathcal{R}$.*
2. *The set $\mathcal{R}$ is* inner-product compatible *if for all $\boldsymbol{R}, \boldsymbol{S} \in \mathcal{R}$ there exists $\boldsymbol{T} \in \mathcal{R}$ such that $\langle \boldsymbol{R}, \boldsymbol{S} \rangle = \operatorname{soe}(\boldsymbol{T})$. Combinatorially, the homomorphism counts from the graph obtained by gluing $\boldsymbol{R}$ and $\boldsymbol{S}$ and forgetting labels, are equal to the homomorphism counts from another graph in $\mathcal{R}$.*

Examples of sets satisfying the above two properties include the set $\mathcal{P}$ of 1-labelled paths where the label is placed on a vertex of degree at most 1, the set $\mathcal{T}$ of 1-labelled trees where the label is placed on an arbitrary vertex, and the set $\mathcal{T}^d$ of 1-labelled d-ary trees with label on a vertex of degree at most 1, where a tree is *d-ary* if its vertices have degree at most $d+1$.

▶ **Theorem 13.** *Let $\mathcal{R}$ be an inner-product compatible set of 1-labelled graphs containing the one-vertex graph. Let G and H be two graphs. Then the following are equivalent:*

1. *G and H are homomorphism indistinguishable over $\mathcal{R}$, that is, for all $\boldsymbol{R} = (R, u) \in \mathcal{R}$,* $\hom(R, G) = \hom(R, H)$.
2. *There exists a unitary*[1] *map $U \colon \mathbb{C}\mathcal{R}_G \to \mathbb{C}\mathcal{R}_H$ such that $U\boldsymbol{R}_G = \boldsymbol{R}_H$ for all $\boldsymbol{R} \in \mathcal{R}$.*

[1] A map $U \colon V \to W$ is *unitary* if $U^*U = \mathrm{id}_V$ and $UU^* = \mathrm{id}_W$ for $U^* \colon W \to V$ the adjoint of U.

3. *There exists a pseudo-stochastic*[2] *map* $X\colon \mathbb{C}\mathcal{R}_G \to \mathbb{C}\mathcal{R}_H$ *such that* $X\boldsymbol{R}_G = \boldsymbol{R}_H$ *for all* $\boldsymbol{R} \in \mathcal{R}$.

If furthermore $\mathcal{R}$ *is* $\boldsymbol{A}$*-invariant then the conditions above are equivalent to the following:*

4. *There exists a pseudo-stochastic matrix* $X \in \mathbb{Q}^{V(H)\times V(G)}$ *such that* $X\boldsymbol{A}_G = \boldsymbol{A}_H X$ *and* $X\boldsymbol{R}_G = \boldsymbol{R}_H$ *for all* $\boldsymbol{R} \in \mathcal{R}$.

Proof. Suppose that Item 1 holds. Since $\mathcal{R}$ is inner-product compatible, for all $\boldsymbol{R}, \boldsymbol{S} \in \mathcal{R}$ there exists $\boldsymbol{T} = (T, v) \in \mathcal{R}$ such that

$$\langle \boldsymbol{R}_G, \boldsymbol{S}_G\rangle = \langle \boldsymbol{R}, \boldsymbol{S}\rangle_G = (\operatorname{soe} \boldsymbol{T})_G = \hom(T, G) = \hom(T, H) = \langle \boldsymbol{R}_H, \boldsymbol{S}_H\rangle .$$

Thus, by a Gram–Schmidt argument, there exists U with the properties in Item 2. Conversely, supposing that Item 2 holds, let $\boldsymbol{R} = (R, u) \in \mathcal{R}$. It holds that $U\boldsymbol{1}_G = \boldsymbol{1}_H$, because $\mathcal{R}$ contains the one-vertex graph $\boldsymbol{1}$. Since U is unitary, $\boldsymbol{1}_G = U^*\boldsymbol{1}_H$. Hence,

$$\hom(R, G) = \langle \boldsymbol{1}_G, \boldsymbol{R}_G\rangle = \langle \boldsymbol{1}_H, U\boldsymbol{R}_G\rangle = \langle \boldsymbol{1}_H, \boldsymbol{R}_H\rangle = \hom(R, H).$$

This shows that Items 1 and 2 are equivalent. Inspecting the above arguments more closely shows that Item 3 is also equivalent with these.

Now suppose additionally that $\mathcal{R}$ is $\boldsymbol{A}$-invariant. It remains to show that in this case Item 4 is equivalent with the first three assertions. For all graphs G, $\mathbb{C}\mathcal{R}_G$ is an $\boldsymbol{A}_G$-invariant subspace of $\mathbb{C}^{V(G)}$. Since $\boldsymbol{A}_G$ is symmetrical, it preserves the direct sum decomposition $\mathbb{C}^{V(G)} = \mathbb{C}\mathcal{R}_G \oplus (\mathbb{C}\mathcal{R}_G)^\perp$. Given U as in Item 2, define X as the map acting like U on $\mathbb{C}\mathcal{R}_G$ and annihilating $(\mathbb{C}\mathcal{R}_G)^\perp$. Let $\boldsymbol{R} \in \mathcal{R}$ be arbitrary. Then $\boldsymbol{A} \cdot \boldsymbol{R} \in \mathcal{R}$ and hence,

$$X\boldsymbol{A}_G\boldsymbol{R}_G = U\boldsymbol{A}_G\boldsymbol{R}_G = \boldsymbol{A}_H\boldsymbol{R}_H = \boldsymbol{A}_H U\boldsymbol{R}_G = \boldsymbol{A}_H X\boldsymbol{R}_G.$$

For $v \in (\mathbb{C}\mathcal{R}_G)^\perp$, $X\boldsymbol{A}_G v = 0 = \boldsymbol{A}_H X v$. Thus, $X\boldsymbol{A}_G = \boldsymbol{A}_H X$.

Finally, $X\boldsymbol{1}_G = U\boldsymbol{1}_G = \boldsymbol{1}_H$ since $\mathcal{R}$ contains the one-vertex graph, and $X^T\boldsymbol{1}_H = \overline{U^*\boldsymbol{1}_H} = \overline{\boldsymbol{1}_G} = \boldsymbol{1}_G$, so X is pseudo-stochastic. The just constructed matrix X may a priori have non-rational entries. However, since Item 4 is essentially a linear system of equations with rational coefficients, it holds that whenever it has a complex solution, it also has a solution over the rationals. This is a consequence of Cramer's rule. The converse, i.e. that Item 4 implies Item 1, follows analogously to the implication from Item 2 to Item 1. ◀

As an easy application of Theorem 13, we recover the characterisation of indistinguishability with respect to path homomorphisms [13].

▶ **Corollary 14.** *Two graphs* G *and* H *are homomorphism indistinguishable over the class of paths if and only if there exists a pseudo-stochastic* $X \in \mathbb{Q}^{V(H)\times V(G)}$ *such that* $X\boldsymbol{A}_G = \boldsymbol{A}_H X$.

The classical characterisation [43] of homomorphism indistinguishability over trees involves a non-negativity condition on the matrix X. While such an assumption appears natural from the viewpoint of solving the system of equations for fractional isomorphism, it lacks an algebraic or combinatorial interpretation. Using Theorem 13, we reprove this known characterisation and give an alternative description that emphasises its graph theoretic origin.

[2] Let I and J be finite sets. Fix vector spaces $V \leq \mathbb{C}^I$ and $W \leq \mathbb{C}^J$ such that the all-ones vectors $\boldsymbol{1}_I \in V$ and $\boldsymbol{1}_J \in W$. Then a map $X\colon V \to W$ is *pseudo-stochastic* if $X\boldsymbol{1}_I = \boldsymbol{1}_J$ and $X^*\boldsymbol{1}_J = \boldsymbol{1}_I$ for X^* the adjoint of X.

▶ **Corollary 15.** *Let G and H be two graphs. G and H are homomorphism indistinguishable over the class of trees if and only if there exists a pseudo-stochastic matrix $X \in \mathbb{Q}^{V(H)\times V(G)}$ such that $X\boldsymbol{A}_G = \boldsymbol{A}_H X$ and one of the following equivalent conditions holds:*

1. *$X \geq 0$ entry-wise,*
2. *$X\boldsymbol{T}_G = \boldsymbol{T}_H$ for all 1-labelled trees $\boldsymbol{T} \in \mathcal{T}$,*
3. *X preserves the Schur product on $\mathbb{C}\mathcal{T}_G$, i.e. $X(u \odot v) = (Xu) \odot (Xv)$ for all $u, v \in \mathbb{C}\mathcal{T}_G$.*

The key graph-theoretic observation is the following: Every labelled tree can be obtained from the one-vertex graph $\boldsymbol{1}$ by adding edges and identifying trees at their labels. Put algebraically, the set $\mathcal{T}$ of 1-labelled trees is the closure of $\{\boldsymbol{1}\}$ under Schur products and multiplication with $\boldsymbol{A}$. Hence, Items 2 and 3 are equivalent. Moreover, Theorem 13 implies the equivalence between Item 2 and homomorphism indistinguishability over the class of trees. The missing equivalence between Items 1 and 2 is deferred to the full version.

Finally, Theorem 13 also gives a characterisation of homomorphism indistinguishability over the class of bounded degree trees. Let $d \geq 1$. The set $\mathcal{T}^d$ of d-ary trees with label on a vertex of degree one or zero is closed under *guarded Schur products*, i.e. the d-ary operation $\circledast^d$ defined as $\circledast^d(\boldsymbol{R}^1, \dots, \boldsymbol{R}^d) \coloneqq \boldsymbol{A} \cdot (\boldsymbol{R}^1 \odot \cdots \odot \boldsymbol{R}^d)$ for $\boldsymbol{R}^1, \dots, \boldsymbol{R}^d \in \mathcal{T}^d$. This operation induces a d-ary operation on $\mathbb{C}\mathcal{T}^d_G$ for every graph G, i.e. $\circledast^d_G(u_1, \dots, u_d) \coloneqq \boldsymbol{A}_G(u_1 \odot \cdots \odot u_d)$ for $u_1, \dots, u_d \in \mathbb{C}\mathcal{T}^d_G$.

▶ **Corollary 16.** *Let $d \geq 1$. Let G and H be graphs. Then the following are equivalent:*

1. *G and H are homomorphism indistinguishable over the class of d-ary trees.*
2. *There exists a pseudo-stochastic matrix $X \in \mathbb{Q}^{V(H)\times V(G)}$ such that $X\boldsymbol{A}_G = \boldsymbol{A}_H X$ and $X\boldsymbol{T}_G = \boldsymbol{T}_H$ for all $\boldsymbol{T} \in \mathcal{T}^d$.*
3. *There exists a pseudo-stochastic matrix $X \in \mathbb{Q}^{V(H)\times V(G)}$ such that X preserves $\circledast^d$ on $\mathbb{C}\mathcal{T}^d_G$, i.e. $X(\circledast^d_G(u_1, \dots, u_d)) = \circledast^d_H(Xu_1, \dots, Xu_d)$ for all $u_1, \dots, u_d \in \mathbb{C}\mathcal{T}^d_G$.*

The systems of equations in Corollary 16 are parametrised by the nested subspaces $\mathbb{C}\mathcal{T}^d_G$ for $d \geq 1$. The following theorem asserts that there exist graphs G for which the inclusions in the chain

$$\mathbb{C}\mathcal{P}_G = \mathbb{C}\mathcal{T}^1_G \subseteq \mathbb{C}\mathcal{T}^2_G \subseteq \cdots \subseteq \mathbb{C}\mathcal{T}^d_G \subseteq \mathbb{C}\mathcal{T}^{d+1}_G \subseteq \cdots \subseteq \mathbb{C}\mathcal{T}_G \subseteq \mathbb{C}^{V(G)}$$

are strict. Conceptually, this is due to the fact that $\mathcal{T}^d$ is only closed under the guarded Schur product $\circledast^d$ while $\mathcal{T}$ is closed under arbitrary Schur products.

▶ **Theorem 17.** *For every integer $d \geq 1$, there exists a graph H such that $\mathbb{C}\mathcal{T}^d_H \neq \mathbb{C}\mathcal{T}^{d+1}_H$.*

The proof of the above theorem can be modified to show that homomorphism indistinguishability over trees of bounded degree is a strictly weaker notion than homomorphism indistinguishability over trees. As a consequence, it is not possible to simulate the 1-dimensional Weisfeiler–Leman algorithm (also known as Colour Refinement, [19]) by counting homomorphisms from trees of any fixed bounded degree.

▶ **Theorem 3.** *For every integer $d \geq 1$, there exist graphs G and H such that G and H are homomorphism indistinguishable over trees of degree at most d, but G and H are not homomorphism indistinguishable over the class of all trees.*

The key step underlying the proofs of Theorems 3 and 17 is the construction of graphs whose (adjacency matrix) eigenspaces behave nicely with respect to the Schur product. Both proofs are deferred to the full version.

4 Representations of Involution Monoids and Homomorphism Indistinguishability

Let $\boldsymbol{F}_1, \dots, \boldsymbol{F}_m$ be (ℓ, ℓ)-bilabelled graphs for some $\ell \in \mathbb{N}$. The closure of $\{\boldsymbol{F}_1, \dots, \boldsymbol{F}_m\}$ under concatenation and taking reverses gives rise to an involution monoid $\mathcal{F}$. If a target graph G is fixed, every bilabelled graph $\boldsymbol{F} \in \mathcal{F}$ yields a homomorphism tensor $\boldsymbol{F}_G$. The association $\boldsymbol{F} \mapsto \boldsymbol{F}_G$ is thus a representation[3] of the involution monoid $\mathcal{F}$. This representation-theoretic viewpoint constitutes a compelling framework for analysing homomorphism tensors.

Recall from Section 2.3 that for an (ℓ, ℓ)-bilabelled graph $\boldsymbol{F}$, the ℓ-labelled graph obtained from $\boldsymbol{F}$ by identifying the in- and out-labels pairwise is denoted by $\boldsymbol{F}^{\text{id}}$. Let $\mathcal{F}^{\text{id}}$ denote the set of all unlabelled graphs underlying graphs of the form $\boldsymbol{F}^{\text{id}}$, $\boldsymbol{F} \in \mathcal{F}$. Then, the character χ_G of the representation of $\mathcal{F}$ induced by G tabulates all homomorphism numbers of the form $\hom(F, G)$ for $F \in \mathcal{F}^{\text{id}}$. Given two target graphs G and H, the equality of characters χ_G and χ_H coincides thus with homomorphism indistinguishability over the class $\mathcal{F}^{\text{id}}$.

On the other hand, equality of characters is, under mild representation-theoretic assumptions, a necessary and sufficient condition for two representations to be equivalent. The equivalence of the representation induced by G and H, when explicitly stated, yields a system of linear equations $X\boldsymbol{F}_G = \boldsymbol{F}_H X$ with $\boldsymbol{F} \in \{\boldsymbol{F}_1, \dots, \boldsymbol{F}_m\}$ where the desired solution X is a unitary matrix. This interpretation forms a useful template for homomorphism indistinguishability results: homomorphism indistinguishability over the class $\mathcal{F}^{\text{id}}$ is equivalent to the existence of a unitary matrix satisfying a suitably defined system of linear equations. Indeed, this template yields Theorem 19 below, a characterisation of homomorphism indistinguishability over graphs of bounded cyclewidth, by setting the generators $\boldsymbol{F}_1, \dots, \boldsymbol{F}_m$ to form a k-basal set.

The following theorem about involution monoid representations due to Specht [41], in particular its generalisation due to Wiegmann [45], forms the main tool for obtaining the results of this section. Let $\mathsf{A} = (A_1, \dots, A_m)$ be a sequence of matrices in $\mathbb{C}^{I \times I}$ for some finite index set I. Let Σ_{2m} denote the finite alphabet $\{x_i, y_i \mid i \in [m]\}$. Let Γ_{2m}denote the infinite set of all words over Σ_{2m}. Equipped with the extension to Γ_{2m} of the map swapping x_i and y_i for all $i \in [m]$, Γ_{2m} can be thought of as a *free involution monoid.* For a word $w \in \Gamma_{2m}$, let w_{A} denote the matrix obtained by substituting $x_i \mapsto A_i$ and $y_i \mapsto A_i^*$ for all $i \in [m]$ and evaluating the matrix product. The substitution $w \mapsto w_{\mathsf{A}}$ is a representation of Γ_{2m}.

▶ **Theorem 18** (Specht [41], Wiegmann [45])**.** *Let I and J be finite index sets. Let $\mathsf{A} = (A_1, \dots, A_m)$ and $\mathsf{B} = (B_1, \dots, B_m)$ be two sequences of matrices such that $A_i \in \mathbb{C}^{I \times I}$ and $B_i \in \mathbb{C}^{J \times J}$ for $i \in [m]$. Then the following are equivalent:*

1. *There exists a unitary $U \in \mathbb{C}^{J \times I}$ such that $UA_i = B_iU$ and $UA_i^* = B_i^*U$ for every $i \in [m]$.*
2. *For every word $w \in \Gamma_{2m}$, $\operatorname{tr}(w_{\mathsf{A}}) = \operatorname{tr}(w_{\mathsf{B}})$.*

Note that the given matrices need not be symmetric. Moreover, it is easy to see that trace equality for words of length at most $\mathcal{O}(n^2)$ suffice to imply trace equality for all words in Γ_{2m}. See [38] for a tighter bound. Finally note that although Theorem 18 is stated as a result involving matrices, the underlying bases are in fact irrelevant.

[3] Phrased in the language of [28], this representation of an involution monoid is a representation of the *concatenation algebra $\mathcal{F}$.*

Our first result follows by applying Wiegmann's theorem to the $(k+1, k+1)$-bilabelled graphs of a k-basal set (Definition 8). This yields an equational characterisation of homomorphism indistinguishability over the class of graphs of cyclewidth at most k.

▶ **Theorem 19.** *Let $k \geq 1$. Let $\mathcal{B}^k$ denote a k-basal set. Let G and H be a graphs. Then the following are equivalent:*

1. *G and H are homomorphism indistinguishable over the class of cyclewidth at most k.*
2. *There exists a unitary matrix $U \in \mathbb{C}^{V(H)^{k+1} \times V(G)^{k+1}}$ such that $U\boldsymbol{B}_G = \boldsymbol{B}_H U$ for all $\boldsymbol{B} \in \mathcal{B}^k$.*

Proof. Let $\boldsymbol{B}^1, \dots, \boldsymbol{B}^m$ be an enumeration of the finite set $\mathcal{B}^k$. Define A and B by setting $A_i \coloneqq \boldsymbol{B}^i_G$ and $B_i \coloneqq \boldsymbol{B}^i_H$ for $i \in [m]$. Recall that the k-basal set $\mathcal{B}^k$ is closed under taking reverses. The theorem immediately follows by an application of Theorem 18 on the matrix sequences A and B, along with Proposition 11. ◀

In contrast to $\mathcal{F}^{\mathrm{id}}$, let $\mathcal{F}^{\mathrm{un}}$ denote the set of all unlabelled graphs underlying graphs $\boldsymbol{F} \in \mathcal{F}$. Although the class $\mathcal{F}^{\mathrm{un}}$ is combinatorially more natural than $\mathcal{F}^{\mathrm{id}}$, it invokes the operator soe on the representations instead of the tr operator, which is algebraically better understood. This technical difficulty is overcome by considering, instead of the original involution monoid representation, its subrepresentation generated by the all-ones vector. In this manner, the *useful spectrum* used in [13] to characterise homomorphism indistinguishability over paths receives an algebraic interpretation. The equivalence of these subrepresentations amounts to the desired solutions being pseudo-stochastic matrices instead of unitary matrices. Formally, we prove the following sum-of-entries analogue of Theorem 18.

▶ **Theorem 20.** *Let I and J be finite index sets. Let $\mathsf{A} = (A_1, \dots, A_m)$ and $\mathsf{B} = (B_1, \dots, B_m)$ be two sequences of matrices such that $A_i \in \mathbb{C}^{I \times I}$ and $B_i \in \mathbb{C}^{J \times J}$ for $i \in [m]$. Then the following are equivalent:*

1. *There exists a pseudo-stochastic matrix $X \in \mathbb{C}^{J \times I}$ such that $XA_i = B_i X$ and $XA_i^* = B_i^* X$ for all $i \in [m]$.*
2. *For every word $w \in \Gamma_{2m}$, $\operatorname{soe}(w_{\mathsf{A}}) = \operatorname{soe}(w_{\mathsf{B}})$.*

Theorem 20 is implied by Lemma 21, which provides a sum-of-entries analogue of Theorem 5. As it establishes a character-theoretic interpretation of the function soe, it may be of independent interest.

▶ **Lemma 21.** *Let Γ be an involution monoid. Let I and J be finite index sets. Let $\varphi \colon \Gamma \to \mathbb{C}^{I \times I}$ and $\varphi \colon \Gamma \to \mathbb{C}^{J \times J}$ be representations of Γ. Let $\varphi' \colon \Gamma \to V$ and $\psi' \colon \Gamma \to W$ denote the subrepresentations of φ and of ψ generated by $\mathbf{1}_I$ and $\mathbf{1}_J$, respectively. Then the following are equivalent:*

1. *For all $g \in \Gamma$, $\operatorname{soe}\varphi(g) = \operatorname{soe}\psi(g)$.*
2. *There exists a unitary pseudo-stochastic $U \colon V \to W$ such that $U\varphi'(g) = \psi'(g)U$ for all $g \in \Gamma$.*
3. *There exists a pseudo-stochastic $X \in \mathbb{C}^{J \times I}$ such that $X\varphi(g) = \psi(g)X$ for all $g \in \Gamma$.*

Proof. Suppose that Item 1 holds. The space V is spanned by the vectors $\varphi(g)\mathbf{1}_I$ for $g \in \Gamma$ while W is spanned by $\psi(g)\mathbf{1}_J$ for $g \in \Gamma$. For $g, h \in \Gamma$ it holds that

$$\langle \varphi(g)\mathbf{1}_I, \varphi(h)\mathbf{1}_I \rangle = \langle \mathbf{1}_I, \varphi(g^*h)\mathbf{1}_I \rangle = \operatorname{soe}\varphi(g^*h) = \operatorname{soe}\psi(g^*h) = \langle \psi(g)\mathbf{1}_J, \psi(h)\mathbf{1}_J \rangle .$$

Hence, V and W are spanned by vectors whose pairwise inner-products are respectively the same. Thus, by a Gram–Schmidt argument, there exists a unitary $U\colon V \to W$ such that $U\varphi(g)\mathbf{1}_I = \psi(g)\mathbf{1}_J$ for all $g \in \Gamma$. This immediately implies that $U\varphi'(g) = \psi'(g)U$ for $g \in \Gamma$. Furthermore, $U\mathbf{1}_I = U\varphi(1_\Gamma)\mathbf{1}_I = \psi(1_\Gamma)\mathbf{1}_J = \mathbf{1}_J$ and $U^*\mathbf{1}_J = \mathbf{1}_I$ since U is unitary. Thus, Item 2 holds.

Suppose now that Item 2 holds. By Lemma 6, write $\varphi = \varphi' \oplus \varphi''$ and $\psi = \psi' \oplus \psi''$. By assumption, there exists a unitary $U\colon V \to W$ such that $U\varphi'(g) = \psi'(g)U$ for all $g \in \Gamma$. Extend U to X by letting it annihilate $V^\perp$. Then $X\varphi(g) = (U\oplus 0)(\varphi'\oplus\varphi'')(g) = U\varphi'(g)\oplus 0 = \psi'(g)U \oplus 0 = \psi(g)X$ for all $g \in \Gamma$. Since U is pseudo-stochastic and $\mathbf{1}_I \in V$ and $\mathbf{1}_J \in W$, X is pseudo-stochastic as well. Hence, Item 3 holds. That Item 3 implies Item 1 is immediate. ◀

Paralleling Theorem 19, we now apply the sum-of-entries version of Specht's theorem to characterise homomorphism indistinguishability over the class of graphs of bounded pathwidth.

▶ **Theorem 22.** *Let $k \geq 1$. Let $\mathcal{B}^k$ denote a k-basal set. Let G and H be a graphs. Then the following are equivalent:*

1. *G and H are homomorphism indistinguishable over the class graphs of pathwidth at most k,*
2. *There exists a pseudo-stochastic matrix $X \in \mathbb{Q}^{V(H)^{k+1}\times V(G)^{k+1}}$ such that $X\boldsymbol{B}_G = \boldsymbol{B}_H X$ for all $\boldsymbol{B} \in \mathcal{B}^k$.*

Let $\mathsf{PW}^{k+1}(G,H)$ denote the system of linear equations in Item 2 above with the basal set from Lemma 9. It comprises $n^{\mathcal{O}(k^2)}$ variables and $|\mathcal{B}^k| \cdot n^{\mathcal{O}(k^2)} = \mathcal{O}(k^2 \cdot n^{\mathcal{O}(k^2)})$ equations.

5 Comparison with Known Systems of Linear Equations

Towards understanding the power and limitations of convex optimisation approaches to the graph isomorphism problem, the level-k Sherali–Adams relaxation of $\mathsf{F}_{\text{iso}}(G,H)$, denoted by $\mathsf{F}^k_{\text{iso}}(G,H)$, was studied in [3]. The system $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$ is another closely related system of interest [21]. Every solution for $\mathsf{F}^{k+1}_{\text{iso}}(G,H)$ yields a solution to $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$, and every solution to $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$ yields a solution to $\mathsf{F}^{k}_{\text{iso}}(G,H)$ [21]. In [3, 21], it was shown that the system $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$ has a non-negative solution if and only if G and H are indistinguishable by the k-dimensional Weisfeiler–Leman algorithm. Following the results of [14, 13], the feasibility of $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$ is thus equivalent to homomorphism indistinguishability over graphs of treewidth at most k.

Dropping non-negativity constraints in $\mathsf{F}_{\text{iso}}(G,H)$ yields a system of linear equations whose feasibility characterises homomorphism indistinguishability over the class of paths [13]. It was conjectured ibidem that dropping non-negativity constraints in $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$ analogously characterises homomorphism indistinguishability over graphs of pathwidth at most k. One direction of this conjecture was shown in [13]: the existence of a rational solution to $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$ implies homomorphism indistinguishability over graphs of pathwidth at most k.

We resolve the aforementioned conjecture by showing that the system of equations $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$ is feasible if and only if the system of equations $\mathsf{PW}^{k+1}(G,H)$ stated in Theorem 22 is feasible. The proof repeatedly makes use of the observation that the equations in $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$ can be viewed as equations in $\mathsf{PW}^{k+1}(G,H)$ where specific k-basal graphs model the continuity and compatibility equations of $\mathsf{L}^{k+1}_{\text{iso}}(G,H)$. Building on Theorem 22, we obtain the following theorem implying Theorem 1.

▶ **Theorem 23.** *For $k \geq 1$ and graphs G and H, the following are equivalent:*
1. *G and H are homomorphism indistinguishable over the class of graphs of pathwidth at most k.*
2. *The system of equations $\mathsf{PW}^{k+1}(G, H)$ has a rational solution.*
3. *The system of equations $\mathsf{L}_{\text{iso}}^{k+1}(G, H)$ has a rational solution.*

Moreover, we show that $\mathsf{PW}^{k+1}(G, H)$ has a non-negative rational solution if and only if $\mathsf{L}_{\text{iso}}^{k+1}(G, H)$ has a non-negative rational solution. Consequently, the system of linear equations $\mathsf{PW}^{k+1}(G, H)$ has a non-negative rational solution if and only if G and H are homomorphism indistinguishable over graphs of treewidth at most k. Hence, the systems of equations $\mathsf{PW}^{k+1}(G, H)$, for $k \in \mathbb{N}$, form an alternative well-motivated hierarchy of linear programming relaxations of the graph isomorphism problem. The details are deferred to the full version.

6 Multi-Labelled Graphs and Homomorphisms from Graphs of Bounded Treewidth and -depth

By considering k-labelled graphs, we complete the picture emerging in Sections 3 and 4, where respectively 1-labelled and (k, k)-bilabelled graphs were considered. In virtue of a generalisation of Theorem 13, a representation-theoretic characterisation of indistinguishability with respect to the k-dimensional Weisfeiler–Leman algorithm (k-WL, see [19] for its definition) is obtained. Amounting [14] to a characterisation of homomorphism indistinguishability over the class of graphs of treewidth at most k, this goal is achieved by constructing, given a k-WL colouring, a representation-theoretic object, the *colouring algebra*, such that two graphs are not distinguished by k-WL if and only if the associated colouring algebras are isomorphic. It turns out that the well-known algebraic characterisation of 2-WL indistinguishability formulated in the language of coherent algebras [10] is a special case of this correspondence. Finally, a combination of the techniques developed in this article yields an equational characterisation of homomorphism indistinguishability over graphs of bounded treedepth. We set off by generalising Definition 12:

▶ **Definition 24.** *A set of k-labelled graphs $\mathcal{R}$ is* inner-product compatible *if $\mathbf{1}^k \in \mathcal{R}$ and for all $\boldsymbol{R}, \boldsymbol{S} \in \mathcal{R}$ there exists $\boldsymbol{T} \in \mathcal{R}$ such that $\langle \boldsymbol{R}, \boldsymbol{S} \rangle = \operatorname{soe} \boldsymbol{T}$.*

For example, the class $\mathcal{TW}^k$ of k-labelled graphs of treewidth k with all labels in a single bag is inner-product compatible. Another example is the class of 2-labelled planar graphs with labels placed on neighbouring vertices of the boundary of a single face. The following main theorem for k-labelled graphs can be derived analogously to Theorem 13.

▶ **Theorem 25.** *Let $k \geq 1$. Let $\mathcal{R}$ be an inner-product compatible set of k-labelled graphs. Let G and H be two graphs. Then the following are equivalent:*
1. *G and H are homomorphism indistinguishable over $\mathcal{R}$, that is for all $\boldsymbol{R} = (R, \boldsymbol{v}) \in \mathcal{R}$, $\hom(R, G) = \hom(R, H)$.*
2. *There exists a unitary $U \colon \mathbb{C}\mathcal{R}_G \to \mathbb{C}\mathcal{R}_H$ such that $U\boldsymbol{R}_G = \boldsymbol{R}_H$ for all $\boldsymbol{R} \in \mathcal{R}$.*
3. *There exists a pseudo-stochastic $X \in \mathbb{Q}^{V(H)^k \times V(G)^k}$ such that $X\boldsymbol{R}_G = \boldsymbol{R}_H$ for all $\boldsymbol{R} \in \mathcal{R}$.*

It turns out that Theorem 25 yields a characterisation of 2-WL indistinguishability in terms of coherent algebras (see [10, 23]): Given a graph G, let $X = (V(G); R_1, \dots, R_s)$ denote the binary relational structure encoding the 2-WL colouring of G. More precisely, each relation $R_i \subseteq V(G) \times V(G)$ corresponds to one of the 2-WL colour classes of G. The *adjacency algebra* $\mathbb{C}\mathcal{A}_G$ *of* G is the $\mathbb{C}$-span of the matrices A_i with $A_i(u, v) = 1$ iff $(u, v) \in R_i$ and zero otherwise. It follows from the properties of 2-WL [10, Theorem 2.3.6] that $\mathbb{C}\mathcal{A}_G$ is closed under matrix products, Schur products, and Hermitian conjugations. In other words, it forms a *coherent algebra*. This construction yields the following algebraic characterisation of 2-WL indistinguishability [10, Proposition 2.3.17]: Two graphs G and H are 2-WL indistinguishable if and only if $\mathbb{C}\mathcal{A}_G$ and $\mathbb{C}\mathcal{A}_H$ are *isomorphic as coherent algebras*, i.e. there exists a vector space isomorphism $X\colon \mathbb{C}\mathcal{A}_G \to \mathbb{C}\mathcal{A}_H$ such that X respects the matrix and the Schur product. That is, for all $A, B \in \mathbb{C}\mathcal{A}_G$, $X(A \cdot B) = X(A) \cdot X(B)$ and $X(A \odot B) = X(A) \odot X(B)$.

Along these lines, it may be argued that adjacency algebras as coherent algebras are the adequate algebraic objects to capture 2-WL indistinguishability. For higher-dimensional WL we propose a similar construction: Informally, G and H are k-WL indistinguishable if and only if certain involution monoid representations closed under Schur products are isomorphic. The aforementioned characterisation of 2-WL will be recovered as a special case in Corollary 28.

More precisely, given a graph G with k-ary relational structure $X = (V(G); R_1, \dots, R_s)$ corresponding to its k-WL colouring, define its *k-WL colouring algebra* $\mathbb{C}\mathcal{A}_G^k$ as the $\mathbb{C}$-span of the tensors $A_i \in \mathbb{C}^{V(G)^k}$ with $A_i(\boldsymbol{u}) = 1$ iff $\boldsymbol{u} \in R_i$ and zero otherwise. The colouring algebra has a rich algebraic structure and is closed under various operations. In particular, it has an interpretation in terms of homomorphism tensors: Let $\mathcal{TW}^k$ denote the set of all k-labelled graphs of treewidth at most k where the labelled vertices all lie in the same bag. Furthermore, let $\mathcal{PWS}^k$ denote the set of (k,k)-bilabelled graphs $\boldsymbol{F} = (F, \boldsymbol{u}, \boldsymbol{v})$ such that F has a path decomposition of width at most k with $\boldsymbol{u}$ and $\boldsymbol{v}$ representing respectively the vertices in the leaf bags.[4] As before, $\mathcal{PWS}^k$ forms an involution monoid under concatenation and taking reverses. These observations are summarised in the following Theorem 26.

▶ **Theorem 26.** *Let G be a graph and let $k \geq 1$. Then*
1. $\mathbb{C}\mathcal{TW}_G^k = \mathbb{C}\mathcal{A}_G^k$,
2. $\mathbb{C}\mathcal{TW}_G^k$ *is closed under Schur products,*
3. *The map $\mathcal{PWS}^k \to \operatorname{End}(\mathbb{C}\mathcal{A}_G^k)$ is a subrepresentation of the involution monoid representation $\mathcal{PWS}^k \to \mathbb{C}^{V(G)^k \times V(G)^k}$ defined as $\boldsymbol{P} \mapsto \boldsymbol{P}_G$.*

The involved proof of Theorem 26 is deferred to the full version. Finally, a representation-theoretic characterisation of k-WL indistinguishability extending [14] can be obtained.

▶ **Theorem 27.** *Let $k \geq 1$. Let G and H be two graphs. Then the following are equivalent:*
1. *G and H are k-WL indistinguishable.*
2. *G and H are homomorphism indistinguishable over the class of graphs of treewidth at most k.*
3. *There exists an isomorphism of $\mathcal{PWS}^k$-representations $X\colon \mathbb{C}\mathcal{A}_G^k \to \mathbb{C}\mathcal{A}_H^k$ respecting the Schur product. That is, for all $A, B \in \mathbb{C}\mathcal{A}_G^k$ and $\boldsymbol{P} \in \mathcal{PWS}^k$, $X(A \odot B) = X(A) \odot X(B)$ and $X(\boldsymbol{P}_G A) = \boldsymbol{P}_H X(A)$.*

To illustrate that the colouring algebra for 2-WL coincides with the coherent algebra, we conclude with inferring Corollary 28 from Theorem 27.

▶ **Corollary 28** (e.g. [10, Proposition 2.3.17])**.** *Let G and H be graphs. Then G and H are 2-WL indistinguishable if and only if there exists a vector space isomorphism $X\colon \mathbb{C}\mathcal{A}_G \to \mathbb{C}\mathcal{A}_H$ such that X respects the matrix and the Schur product. That is, for all $A, B \in \mathbb{C}\mathcal{A}_G$, $X(A \cdot B) = X(A) \cdot X(B)$ and $X(A \odot B) = X(A) \odot X(B)$.*

[4] Observe that this is in contrast to Section 2.3, where the set $\mathcal{PW}^k$ of $(k+1, k+1)$-bilabelled graphs with underlying graphs of pathwidth at most k was considered. There, the labels are carried by vertices in the intersection of two adjacent bags, while here the labelled vertices must only lie in the same bag.

As a final application of our theory, we infer an equational characterisation of homomorphism indistinguishability over graphs of bounded treedepth. The *treedepth* [36] of a graph F is defined as the minimum height of an elimination forest of F, i.e. of a rooted forest T with $V(T) = V(F)$ such that every edge in F connects vertices that are in an ancestor-descendent relationship in T. In [17], it was shown that homomorphism indistinguishable over graphs of treedepth at most k corresponds to equivalence over the quantifier-rank-k fragment of first order logic with counting quantifiers. We extend this characterisation by proposing a linear system of equations very similar to the one for bounded pathwidth.

Let $k \geq 1$. For graphs G and H, consider the following system of equations $\mathsf{TD}^k(G, H)$ with variables $X(\boldsymbol{w}, \boldsymbol{v})$ for every pair of tuples $\boldsymbol{w} \in V(H)^\ell$ and $\boldsymbol{v} \in V(G)^\ell$ for $0 \leq \ell \leq k$. A length-ℓ pair $(\boldsymbol{w}, \boldsymbol{v})$ is said to be a *partial pseudo-isomorphism* if $\boldsymbol{v}_i = \boldsymbol{v}_{i+1} \iff \boldsymbol{w}_i = \boldsymbol{w}_{i+1}$ for all $i \in [\ell - 1]$ and $\{\boldsymbol{v}_i, \boldsymbol{v}_j\} \in E(G) \iff \{\boldsymbol{w}_i, \boldsymbol{w}_j\} \in E(H)$ for all $i, j \in [\ell]$. Note that in contrary to the partial isomorphisms appearing in [13], partial pseudo-isomorphism only need to preserve the equality of consecutive vertices in the domain tuple.

$\mathsf{TD}^k(G, H)$

$$\sum_{v' \in V(G)} X(\boldsymbol{w}w, \boldsymbol{v}v') = X(\boldsymbol{w}, \boldsymbol{v}) \quad \text{for all } w \in V(H) \text{ and } \boldsymbol{v} \in V(G)^\ell,\ \boldsymbol{w} \in V(H)^\ell \text{ where } 0 \leq \ell < k. \tag{TD1}$$

$$\sum_{w' \in V(H)} X(\boldsymbol{w}w', \boldsymbol{v}v) = X(\boldsymbol{w}, \boldsymbol{v}) \quad \text{for all } v \in V(G) \text{ and } \boldsymbol{v} \in V(G)^\ell,\ \boldsymbol{w} \in V(H)^\ell \text{ where } 0 \leq \ell < k. \tag{TD2}$$

$$X((), ()) = 1 \tag{TD3}$$

$$X(\boldsymbol{w}, \boldsymbol{v}) = 0 \quad \text{whenever } (\boldsymbol{w}, \boldsymbol{v}) \text{ is not a partial pseudo-isomorphism} \tag{TD4}$$

The proof of the following theorem is deferred to the full version.

▶ **Theorem 2.** *For every $k \geq 1$, the following are equivalent for two graphs G and H:*

1. *G and H are homomorphism indistinguishable over graphs of treedepth at most k,*
2. *The linear systems of equations $\mathsf{TD}^k(G, H)$ has a non-negative rational solution,*
3. *The linear systems of equations $\mathsf{TD}^k(G, H)$ has a rational solution.*

7 Concluding Remarks

We have developed an algebraic theory of homomorphism indistinguishability that allows us to reprove known results in a unified way and derive new characterisations of homomorphism indistinguishability over bounded degree trees, graphs of bounded treedepth, graphs of bounded cyclewidth, and graphs of bounded pathwidth. The latter answers an open question from [13].

Homomorphism indistinguishabilities over various graph classes can be viewed as similarity measures for graphs, and our new results as well as many previous results show that these are natural and robust. Yet homomorphism indistinguishability only yields equivalence relations, or families of equivalence relations, and not a “quantitative” distance measure. For many applications of graph similarity, such quantitative measures are needed. Interestingly, we can derive distance measure both from homomorphism indistinguishability and from the equational characterisations we study here. For a class $\mathcal{F}$ of graphs, we can consider the *homomorphism embedding* that maps graphs G to the vector in $\mathbb{R}^{\mathcal{F}}$ whose entries are the numbers $\hom(F, G)$ for graphs $F \in \mathcal{F}$. Then a norm on the space $\mathbb{R}^{\mathcal{F}}$ induces a graph (pseudo)metric. Such metrics give a generic family of graph kernels (see [18]). On the equational side, a notion like fractional isomorphism induces a (pseudo)metric on graphs where the distance between graphs G and H is $\min_X \|X\boldsymbol{A}_G - \boldsymbol{A}_H X\|$, where X ranges over all doubly-stochastic matrices. It is a very interesting question whether the correspondence between the equivalence relations for homomorphism indistinguishability and feasibility of the systems of equations can be extended to the associated metrics. In the special case of isomorphism and homomorphism indistinguishability over all graphs, the theory of graph limits provides some answers [28]. This has recently been extended to fractional isomorphism and homomorphism indistinguishability over trees [5].

References

1 Noga Alon, Phuong Dao, Iman Hajirasouliha, Fereydoun Hormozdiari, and Süleyman Cenk Sahinalp. Biomolecular network motif counting and discovery by color coding. In *Proceedings 16th International Conference on Intelligent Systems for Molecular Biology (ISMB), Toronto, Canada, July 19-23, 2008*, pages 241–249, 2008. doi:10.1093/bioinformatics/btn163.

2 Albert Atserias, Laura Mančinska, David E. Roberson, Robert Sámal, Simone Severini, and Antonios Varvitsiotis. Quantum and non-signalling graph isomorphisms. *J. Comb. Theory, Ser. B*, 136:289–328, 2019. doi:10.1016/j.jctb.2018.11.002.

3 Albert Atserias and Elitza N. Maneva. Sherali–Adams Relaxations and Indistinguishability in Counting Logics. *SIAM J. Comput.*, 42(1):112–137, 2013. doi:10.1137/120867834.

4 Paul Beaujean, Florian Sikora, and Florian Yger. Scaling up graph homomorphism for classification via sampling. *ArXiv*, 2104.04040, 2021. doi:10.48550/arXiv.2104.04040.

5 Jan Böker. Graph Similarity and Homomorphism Densities. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 32:1–32:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ICALP.2021.32.

6 A. Bulatov, P. Jeavons, and A. Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM Journal on Computing*, 34:720–742, 2005. doi:10.1137/S0097539700376676.

7 A.A. Bulatov. A Dichotomy Theorem for Nonuniform CSPs. In *Proceedings of the 58th IEEE Annual Symposium on Foundations of Computer Science*, pages 319–330, 2017. doi:10.1109/FOCS.2017.37.

8 Andrei A. Bulatov. The complexity of the counting constraint satisfaction problem. *J. ACM*, 60(5):34:1–34:41, 2013. doi:10.1145/2528400.

9 Andrei A. Bulatov and Stanislav Zivný. Approximate counting CSP seen from the other side. *ACM Trans. Comput. Theory*, 12(2):11:1–11:19, 2020. doi:10.1145/3389390.

10 Gang Chen and Ilia Ponomarenko. *Lectures on Coherent Configurations.* Central China Normal University Press, Wuhan, 2018. URL: https://www.pdmi.ras.ru/~inp/ccNOTES.pdf.

11 Radu Curticapean, Holger Dell, and Dániel Marx. Homomorphisms are a good basis for counting small subgraphs. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 210–223. ACM, 2017. doi:10.1145/3055399.3055502.

12 Anuj Dawar, Tomás Jakl, and Luca Reggio. Lovász-type theorems and game comonads. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*, pages 1–13. IEEE, 2021. doi:10.1109/LICS52264.2021.9470609.

13 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, July 9-13, 2018, Prague, Czech Republic*, volume 107 of *LIPIcs*, pages 40:1–40:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.ICALP.2018.40.

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

15 M. Freedman, L. Lovász, and A. Schrijver. Reflection positivity, rank connectivity, and homomorphism of graphs. *Journal of the American Mathematical Society*, 20:37–51, 2007. doi:10.1090/S0894-0347-06-00529-7.

16 G. Frobenius and I. Schur. Über die Äquivalenz der Gruppen linearer Substitutionen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, 1:209–217, 1906. URL: https://www.biodiversitylibrary.org/item/93364.

17 Martin Grohe. Counting Bounded Tree Depth Homomorphisms. In Holger Hermanns, Lijun Zhang, Naoki Kobayashi, and Dale Miller, editors, *LICS '20: 35th Annual ACM/IEEE Symposium on Logic in Computer Science, Saarbrücken, Germany, July 8-11, 2020*, pages 507–520. ACM, 2020. doi:10.1145/3373718.3394739.

18 Martin Grohe. word2vec, node2vec, graph2vec, x2vec: Towards a theory of vector embeddings of structured data. In Dan Suciu, Yufei Tao, and Zhewei Wei, editors, *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2020, Portland, OR, USA, June 14-19, 2020*, pages 1–16. ACM, 2020. doi:10.1145/3375395.3387641.

19 Martin Grohe. The Logic of Graph Neural Networks. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*, pages 1–17. IEEE, 2021. doi:10.1109/LICS52264.2021.9470677.

20 Martin Grohe, Kristian Kersting, Martin Mladenov, and Pascal Schweitzer. Color Refinement and Its Applications. In *An Introduction to Lifted Probabilistic Inference*. The MIT Press, August 2021. doi:10.7551/mitpress/10548.003.0023.

21 Martin Grohe and Martin Otto. Pebble Games and Linear Equations. *J. Symb. Log.*, 80(3):797–844, 2015. doi:10.1017/jsl.2015.28.

22 Martin Grohe, Gaurav Rattan, and Tim Seppelt. Homomorphism tensors and linear equations. *ArXiv*, 2111.11313, 2021. doi:10.48550/ARXIV.2111.11313.

23 Donald G. Higman. Coherent algebras. *Linear Algebra and its Applications*, 93:209–239, August 1987. doi:10.1016/S0024-3795(87)90326-0.

24 Nils M. Kriege, Fredrik D. Johansson, and Christopher Morris. A survey on graph kernels. *Appl. Netw. Sci.*, 5(1):6, 2020. doi:10.1007/s41109-019-0195-3.

25 Pascal Kühner. Graph embeddings based on homomorphism counts. Master's thesis, Department of Computer Science, RWTH Aachen University, 2021.

26 Tsit-Yuen Lam. *A First Course in Noncommutative Rings*. Number 131 in Graduate texts in mathematics. Springer, New York, NY, 2. ed edition, 2001. doi:10.1007/978-1-4419-8616-0.

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

28 László Lovász. *Large Networks and Graph Limits*. American Mathematical Society, 2012. doi:10.1090/coll/060.

29 László Lovász and Alexander Schrijver. Graph parameters and semigroup functions. *Eur. J. Comb.*, 29(4):987–1002, 2008. doi:10.1016/j.ejc.2007.11.008.

30 László Lovász and Balázs Szegedy. Contractors and connectors of graph algebras. *Journal of Graph Theory*, 60(1):11–30, 2009. doi:10.1002/jgt.20343.

31 Peter N. Malkin. Sherali–Adams Relaxations of Graph Isomorphism Polytopes. *Discrete Optimization*, 12:73–97, 2014. doi:10.1016/j.disopt.2014.01.004.

32 Laura Mančinska and David E. Roberson. Quantum isomorphism is equivalent to equality of homomorphism counts from planar graphs. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 661–672. IEEE, 2020. doi:10.1109/FOCS46700.2020.00067.

33 Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002. doi:10.1126/science.298.5594.824.

34 Yoàv Montacute and Nihil Shah. The pebble-relation comonad in finite model theory. *ArXiv*, 2110.08196, 2021. doi:10.48550/arXiv.2110.08196.

35 Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 4602–4609. AAAI Press, 2019. doi:10.1609/aaai.v33i01.33014602.

36 Jaroslav Nešetřil and Patrice Ossona de Mendez. *Sparsity: graphs, structures, and algorithms*. Number 28 in Algorithms and combinatorics. Springer, Heidelberg ; New York, 2012. OCLC: ocn773019512. doi:10.1007/978-3-642-27875-4.

37 Hoang Nguyen and Takanori Maehara. Graph homomorphism convolution. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 7306–7316. PMLR, 2020. URL: http://proceedings.mlr.press/v119/nguyen20c.html.

38 Christopher J Pappacena. An Upper Bound for the Length of a Finite-Dimensional Algebra. *Journal of Algebra*, 197(2):535–545, 1997. doi:10.1006/jabr.1997.7140.

39 Marc Roth and Philip Wellnitz. Counting and finding homomorphisms is universal for parameterized complexity theory. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 2161–2180. SIAM, 2020. doi:10.1137/1.9781611975994.133.

40 Alexander Schrijver. Graph invariants in the spin model. *Journal of Combinatorial Theory, Series B*, 99:502–511, 2009. doi:10.1016/j.jctb.2008.10.003.

41 Wilhelm Specht. Zur Theorie der Matrizen. II. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 50:19–23, 1940. URL: http://gdz.sub.uni-goettingen.de/dms/load/toc/?PPN=PPN37721857X_0050&DMDID=dmdlog6.

42 J. Thapper and S. Živný. The complexity of finite-valued csps. *Journal of the ACM*, 63(4), 2016. doi:10.1145/2974019.

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

44 Gottfried Tinhofer. A note on compact graphs. *Discret. Appl. Math.*, 30(2-3):253–264, 1991. doi:10.1016/0166-218X(91)90049-3.

45 N. A. Wiegmann. Necessary and sufficient conditions for unitary similarity. *Journal of the Australian Mathematical Society*, 2(1):122–126, 1961. Edition: 2009/04/09 Publisher: Cambridge University Press. doi:10.1017/S1446788700026422.

46 Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL: https://openreview.net/forum?id=ryGs6iA5Km.

47 Dmitriy Zhuk. A proof of CSP dichotomy conjecture. In *Proceedings of the 58th IEEE Annual Symposium on Foundations of Computer Science*, pages 331–342, 2017. doi:10.1109/FOCS.2017.38.