Equi-Rank Homomorphism Preservation Theorem on Finite Structures

Benjamin Rossman
Duke University, Durham, NC, USA

Abstract

The Homomorphism Preservation Theorem (HPT) of classical model theory states that a first-order sentence is preserved under homomorphisms if, and only if, it is equivalent to an existential-positive sentence. This theorem remains valid when restricted to finite structures, as demonstrated by the author in [33, 34] via distinct model-theoretic and circuit-complexity based proofs. In this paper, we present a third (and significantly simpler) proof of the finitary HPT based on a generalized Cai-Fürer-Immerman construction. This method establishes a tight correspondence between syntactic parameters of a homomorphism-preserved sentence (quantifier rank, variable width, alternation height) and structural parameters of its minimal models (tree-width, tree-depth, decomposition height). Consequently, we prove a conjectured "equi-rank" version of the finitary HPT. In contrast, previous versions of the finitary HPT possess additional properties, but incur blow-ups in the quantifier rank of the equivalent existential-positive sentence.

2012 ACM Subject Classification Theory of computation → Finite Model Theory

Keywords and phrases finite model theory, preservation theorems, quantifier rank

Digital Object Identifier 10.4230/LIPIcs.CSL.2025.6

Acknowledgements I thank Deepanshu Kush, William He and Ken-ichi Kawarabayashi for stimulating conversions that prompted me to revisit the topic of this paper. Anuj Dawar provided helpful comments on a preliminary draft of this paper. I am grateful to the anonymous referees for numerous suggestions that improved the clarity of this paper. This paper was partially written during a visit to the National Institute of Informations in Tokyo, Japan.

1 Introduction

A first-order sentence φ is said to be *preserved under homomorphisms* if, for any model $\mathfrak{A} \models \varphi$ and any structure $\mathfrak{B}$ such that there exists a homomorphism $\mathfrak{A} \to \mathfrak{B}$, it holds that $\mathfrak{B} \models \varphi$. One class of first-order sentences that are always preserved under homomorphisms are the *existential-positive sentences*, which are built from atomic formulas (of the form $x_1 = x_2$ and $R(x_1, \dots, x_r)$ where R is an r-ary relation symbol) via conjunction $\varphi_1 \wedge \varphi_2$, disjunction $\varphi_1 \vee \varphi_2$, and existential quantification $\exists x\, \varphi$ (that is, without negation $\neg\varphi$ or universal quantification $\forall x\, \varphi$).[1]

The Homomorphism Preservation Theorem (HPT) of classical model theory, attributed to Łoś, Lyndon and Tarski [30, 39, 31], states that existential-positive sentences are – up to logical equivalence – the only first-order sentences that are preserved under homomorphisms.[2]

▶ **Theorem 1.1** (HPT). *A first-order sentence is preserved under homomorphisms if, and only if, it is equivalent to an existential-positive sentence.*

[1] A *sentence* is a formula with no free variables. Although the definition of *preserved under homomorphisms* extends to formulas with free variables, we speak of sentences for simplicity sake. We further restrict attention to *relational languages* (without functions or constant symbols), even though most definitions and results in this paper extend to general first-order languages.

[2] Here the semantic notions of *logical equivalence* and *preserved under homomorphisms* are with respect to all (finite or infinite) structures.

33rd EACSL Annual Conference on Computer Science Logic (CSL 2025).
Editors: Jörg Endrullis and Sylvain Schmitz; Article No. 6; pp. 6:1–6:17
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The closely related Łoś-Tarski and Lyndon Preservation Theorems state that a first-order sentence that is preserved under embeddings (respectively: surjective homomorphisms) if, and only if, it is equivalent to an existential sentence (respectively: positive sentence). In each of these classical preservation theorems, the "if" direction follows directly from the semantics of first-order logic, while the "only if" direction was originally proved non-constructively using the Compactness Theorem.

A line of work in finite model theory initiated by Gurevich [20] studies the question of which theorems of classical model theory remain valid, and which become false, when restricted to finite structures. For example, the Compactness Theorem is easily seen to be false on finite structures. Counterexamples in [38, 20] witness the failure on finite structure of the Łoś-Tarski and Lyndon Theorems (on preservation under embeddings and surjective homomorphisms). In contrast, previous work of the author [33, Theorem 1.7] and [34, Theorem 6] showed that the classical HPT remains valid when restricted to finite structures.

▶ **Theorem 1.2** (HPT on finite structures)**.** *Every first-order sentence that is preserved under homomorphisms* ***on finite structures*** *is equivalent* ***on finite structures*** *to an existential-positive sentence.*

Articles [33, 34] provide two entirely different proofs of Theorem 1.2, which we discuss in §2.1 and §2.2. Unfortunately, both proofs incur large blow-ups from the quantifier rank r of a homomorphism-preserved first-order sentence φ to the quantifier rank $r^{\exists+}$ ($\gg r$) of the equivalent existential-positive sentence $\varphi^{\exists+}$. The blow-up in [33] is non-elementary: $r^{\exists+}$ is a tower-of-exponentials of height r. (With respect to the *length* of φ and $\varphi^{\exists+}$, a non-elementary blow-up is necessary [33, Theorem 6.1].) The method of [34] improved the quantifier rank blow-up to merely polynomial: $r^{\exists+} = O(r^3 \log r)$.

At the same time, an additional result of [33, Theorem 1.6] showed the classical HPT (Theorem 1.1) requires *no blow-up at all* in quantifier rank:

▶ **Theorem 1.3** (Equi-rank HPT)**.** *Every first-order sentence that is preserved under homomorphisms (on all structures) is equivalent to an existential-positive sentence* ***with the same quantifier rank****.*

The proof of Theorem 1.3 avoids the non-constructive Compactness Theorem, using instead a method of $\exists^+$-saturated co-retracts (see the discussion in §2.1). It was left as an open question whether Theorem 1.3 remains valid on finite structures. The main result of the present paper answers this question in the affirmative, finally unifying the finitary and equi-rank versions of the HPT.

▶ **Theorem 1.4** (Equi-rank HPT on finite structures)**.** *Every first-order sentence that is preserved under homomorphisms on finite structures is equivalent on finite structures to an existential-positive sentence with the same quantifier rank, variable width, and alternation height.*

Theorem 1.4 is simultaneously tight with respect to three different parameters: quantifier rank, variable width, and alternation height (see §3.2 for definitions). The surprisingly simple proof utilizes a generalized Cai-Fürer-Immerman construction [6] on finite relational structures (see §2.3). In particular, we consider two CFI structures $\mathfrak{C}_{\mathrm{even}}$ and $\mathfrak{C}_{\mathrm{odd}}$ over a finite core $\mathfrak{C}$ (i.e., structure such that every homomorphism $\mathfrak{C} \to \mathfrak{C}$ is an isomorphism). We show that

$$\mathfrak{C}_{\mathrm{even}} \rightleftarrows \mathfrak{C} \quad \text{and} \quad \mathfrak{C}_{\mathrm{odd}} \to \mathfrak{C} \not\to \mathfrak{C}_{\mathrm{odd}}$$

where $\to$ denotes the existence of a homomorphism (Lemma 5.5). Theorem 1.4 then follows from a characterization of existential-positive definability in terms of the minimal cores of a homomorphism-closed class of finite structures (Lemma 4.6).

Related work

Our proof of Theorem 1.4 combines a few standard techniques that appear in many prior works. Variants of the Cai-Fürer-Immerman construction have found numerous applications similar in nature to our main result. In particular, Fürer [16] and Neuen [32] considered very similar generalization of CFI structures as given in Definition 5.4, only with different applications in mind.

The correspondence between quantifier rank/variable width, tree-width/tree-depth, and the # of moves/# of cops parameters of the cops-and-robber game, has been developed in several works including [1, 3, 12, 15, 19]. Variants of Lemmas 4.6 can be found in many of these papers.

Anuj Dawar (personal communication) independently identified the "core" homomorphism property of CFI structures (Lemma 5.5) in the context of a different problem. It is perhaps surprising that the application of the CFI construction to prove the Homomorphism Preservation Theorem remained unrecognized for such an extended period.

Versions of the HPT relativized to various classes of structures have been investigated in [4, 5, 10, 21] (see also [11], which corrects some claims in articles [5, 10]). Versions of the HPT for different fragments, as well as extensions, of first-order logic have been studied in [2, 7, 14].

2 Comparing the three proofs of the finitary HPT

In this section, we discuss both previous proofs of the Homomorphism Preservation Theorem on finite structures (Theorem 1.2) from articles [33, 34]. We then give a brief overview of our new proof using the Cai-Fürer-Immerman construction.

2.1 First proof via $\exists^+$-indistinguishable co-retracts

The original proof of the finitary HPT actually establishes the following stronger result.

▶ **Theorem 2.1** ([33, Theorem 5.15]). *For every finite relational signature σ and integer $r \geq 0$, there exists an integer $r^{\exists+}$ $(\geq r)$ and an operation*

$$\mathfrak{A} \longmapsto \widehat{\mathfrak{A}} : \{\ \sigma\text{-structures}\} \longrightarrow \{\ \sigma\text{-structures}\}$$

with the following properties:

(I) *$\widehat{\mathfrak{A}}$ is a co-retract of $\mathfrak{A}$ (i.e., $\mathfrak{A}$ is a substructure of $\widehat{\mathfrak{A}}$ and there is a homomorphism $\widehat{\mathfrak{A}} \to \mathfrak{A}$ that fixes each element of $\mathfrak{A}$).*

(II) *Whenever $\mathfrak{A}$ is finite, so is $\widehat{\mathfrak{A}}$.*

(III) *Whenever $\mathfrak{A}$ and $\mathfrak{B}$ satisfy the same existential-positive sentences of quantifier rank $r^{\exists+}$, their co-retracts $\widehat{\mathfrak{A}}$ and $\widehat{\mathfrak{B}}$ satisfy the same first-order sentences of quantifier rank r.*

The finitary HPT follows straightforwardly from Theorem 2.1, although inheriting the same blow-up in quantifier rank from r to $r^{\exists+}$ (a tower-of-exponentials of height r). It is unknown whether every operation $\mathfrak{A} \longmapsto \widehat{\mathfrak{A}}$ satisfying properties (I), (II) and (III) requires a large blow-up from r to $r^{\exists+}$; the method of the present paper sheds no light on this question.

An additional result in [33, Theorem 4.11] shows that the optimal value $r^{\exists+} = r$ can be achieved by sacrificing property (II), that is, allowing $\widehat{\mathfrak{A}}$ to be infinite even when $\mathfrak{A}$ is finite. This yields the equi-rank version of the classical HPT (Theorem 1.3). In this version of the hat operation, $\widehat{\mathfrak{A}}$ is an (infinite) $\exists^+$-saturated co-retract of $\mathfrak{A}$. The manner of "finitizing" this operation in [33] is responsible for the tower-of-exponentials blow-up in Theorem 2.1.

▶ Remark 2.2. The method of $\exists^+$-saturated co-retracts was recently generalized by Abramsky and Reggio [2], though the lens of game comonads and arboreal categories. They give general conditions leading to equi-resource homomorphism preservation theorems, both with respect to fragments of first-order logic and relativized to classes of structures satisfying certain axioms. Our proof of Theorem 1.4 does not follow this template, but it would be interesting to know if a $\exists^+$-saturation proof is possible in the finite setting.

2.2 Second proof via AC^0 lower bounds

Subsequent work of the author [34] provides an entirely different proof of the finitary HPT, with a merely polynomial blow-up in quantifier rank, based on lower bounds in circuit complexity.

▶ **Theorem 2.3** (Main result of [34]). *Let $\mathcal{C}$ be a homomorphism-closed class of finite structures.*

(1) *If membership in $\mathcal{C}$ is decidable on structures of size n by non-uniform AC^0 formulas of size $O(n^r)$, then $\mathcal{C}$ is definable on finite structures (of all sizes) by a single existential-positive sentence of quantifier rank $O(r^3 \log r)$.*

(2) *If membership in $\mathcal{C}$ is decidable on structures of size n by non-uniform AC^0 circuits of size $O(n^s)$, then $\mathcal{C}$ is definable on finite structures (of all sizes) by a single existential-positive sentence of variable width $O(s \log s)$.*

The proof of Theorem 2.3 relies on three different lower bounds in circuit complexity [27, 28, 35], in addition to a result in graph minor theory on the excluded-minor approximation of tree-depth [9, 25].[3] We remark that Theorem 2.3 has an equivalent *descriptive complexity* formulation via the well-known correspondence [13, 24] between the non-uniform complexity class AC^0 and the logic $\mathsf{FO}[\mathsf{Arb}]$ (first-order logic with arbitrary background predicates).

▶ **Corollary 2.4.** *Let $\mathcal{C}$ be a homomorphism-closed class of finite structures.*

(1) *If $\mathcal{C}$ is definable on finite structures by an $\mathsf{FO}[\mathsf{Arb}]$ sentence of quantifier rank r, then $\mathcal{C}$ is definable on finite structures by an existential-positive sentence of quantifier rank $O(r^3 \log r)$.*

(2) *If $\mathcal{C}$ is definable on finite structures by an $\mathsf{FO}[\mathsf{Arb}]$ sentence of quantifier width s, then $\mathcal{C}$ is definable on finite structures by an existential-positive sentence of variable width $O(s \log s)$.*

Corollary 2.4 strengthens Theorem 1.2 by expanding the hypothesis from first-order sentences to the more expressive class of $\mathsf{FO}[\mathsf{Arb}]$ sentences, while the equivalent existential-positive sentences remain first-order (without background predicates). The results of the present paper do not directly improve Corollary 2.4, but do show that improvements would

[3] Part (1) of Theorem 2.3 is stated in [34] with a weaker polynomial bound $O(r^5 \log r)$. The improvement to $O(r^3 \log r)$ relies on a subsequent result of Czerwinski, Nadara, and Pilipczuk [9]. Part (2) of Theorem 2.3 is not explicitly stated in [34], but follows by a similar argument to part (1) using the AC^0 circuit lower bound of Li, Razborov and the author [28].

follow from strong enough lower bounds on the AC^0 complexity of distinguishing "even" and "odd" CFI structures over any base graph. Recent size-depth tradeoffs for AC^0-Frege refutations of Tseitin formulas [22, 17] might be relevant to this question.

2.3 New proof via the Cai-Fürer-Immerman construction

An influential article of Cai, Fürer and Immerman [6] introduced a construction of non-isomorphic simple graphs of order n that are indistinguishable by $o(n)$-variable counting logic (equivalently, by the $o(n)$-dimensional Weisfeiler-Leman algorithm). The general construction associates any 3-regular graph G with a pair of non-isomorphic graphs G_{even} and G_{odd}, which are hard to distinguish when the "base graph" G is an expander. The CFI construction is closely related to 3-XOR-CNF formulas associated with G, studied by Tseitin [40] in the setting of proof complexity. The Tseitin formulas are a system of linear equations modulo 2, with a variable X_e for each edge and a constraint for each vertex v (either $X_e \oplus X_f \oplus X_g = 0$ or $X_e \oplus X_f \oplus X_g = 1$, where e, f, g are the edges incident to v). This system is satisfiable if, and only if, the number of inhomogeneous constraints is even. The CFI graphs G_{even} and G_{odd} encode the two different (satisfiable and unsatisfiable) Tseitin formulas over G, corresponding to the 0-homomology classes of G over $\mathbb{Z}_2$.

The CFI construction generalizes to arbitrary (non-3-regular, non-connected) simple graphs G, as well as to abelian coordinate groups other than $\mathbb{Z}_2$. Generalizations of Tseitin formulas and the CFI construction have found numerous applications in finite model theory and proof complexity. In this paper we consider a natural version of the CFI construction on finite structures with a fixed relational signature (Definition 5.4). For any finite "base" structure $\mathfrak{A}$, this construction produces a pair of non-isomorphic finite structures $\mathfrak{A}_{\text{even}}$ and $\mathfrak{A}_{\text{odd}}$. Like the original CFI graphs, these structures are indistinguishable by first-order sentences whose quantifier rank / number of variables is less than the tree-depth / tree-width of $\mathfrak{A}$ (Lemma 5.6).

Unlike some versions of the CFI construction (such as the original CFI graphs [6]), structures $\mathfrak{A}_{\text{even}}$ and $\mathfrak{A}_{\text{odd}}$ project homomorphically to the base structure $\mathfrak{A}$. In the key special case of a finite core $\mathfrak{C}$ (where every homomorphism $\mathfrak{C} \to \mathfrak{C}$ is an isomorphism), we show that $\mathfrak{C}$ admits a homomorphism to $\mathfrak{C}_{\text{even}}$ but not to $\mathfrak{C}_{\text{odd}}$ (Lemma 5.5). Our main result, the equi-rank finitary HPT (Theorem 1.4), then follows by essentially well-known arguments.

3 Preliminaries

This section includes all relevant definitions pertaining to structures, homomorphisms, and first-order logic. See [29, 24] for additional background on finite model theory.

3.1 Structures and homomorphisms

▶ **Definition 3.1.** A *(relational) signature* is a set σ of relation symbols, each with an associated positive integer "arity". A *σ-structure* $\mathfrak{A}$ consists a set A (called the *universe* of $\mathfrak{A}$) together with an *interpretation* $R^{\mathfrak{A}} \subseteq A^t$ for each t-ary relation symbol R in σ. A structure is *finite* if its signature and universe are both finite.

▶ **Definition 3.2** (Homomorphism and isomorphism).

- For structures $\mathfrak{A}, \mathfrak{B}$ with the same signature, a *homomorphism* $h : \mathfrak{A} \to \mathfrak{B}$ is a function from the universe of $\mathfrak{A}$ to the universe of $\mathfrak{B}$, which maps each tuple in each relation of $\mathfrak{A}$ to a tuple in the corresponding relation of $\mathfrak{B}$. An *isomorphism* is a homomorphism h which is a bijection and whose inverse h^{-1} is a homomorphism.

- Notation $\mathfrak{A} \to \mathfrak{B}$ expresses that there exists a homomorphism from $\mathfrak{A}$ to $\mathfrak{B}$. We say that $\mathfrak{A}$ and $\mathfrak{B}$ are *homomorphically equivalent*, denoted $\mathfrak{A} \rightleftarrows \mathfrak{B}$, if $\mathfrak{A} \to \mathfrak{B}$ and $\mathfrak{B} \to \mathfrak{A}$.
- Notation $\mathfrak{A} \to *$ denotes the class of all (finite or infinite) structures $\mathfrak{B}$ such that $\mathfrak{A} \to \mathfrak{B}$.

▶ **Definition 3.3** (The core of a finite structure).
- A structure $\mathfrak{C}$ is a *core* if every homomorphism $h : \mathfrak{C} \to \mathfrak{C}$ is an isomorphism.
- Every finite structure $\mathfrak{A}$ is homomorphically equivalent to a unique core (up to isomorphism), which we call "the" *core* of $\mathfrak{A}$ and denote by $\mathsf{Core}(\mathfrak{A})$. $\mathsf{Core}(\mathfrak{A})$ is isomorphic to an induced substructure of $\mathfrak{A}$, namely any minimal retract of $\mathfrak{A}$ [23].

▶ **Definition 3.4** (Gaifman graph). The *Gaifman graph* of a structure $\mathfrak{A}$, denoted $\mathsf{Gaif}(\mathfrak{A})$, is the simple graph with vertex set $V(\mathsf{Gaif}(\mathfrak{A})) = A$ (the universe of $\mathfrak{A}$) and undirected edge set

$$E(\mathsf{Gaif}(\mathfrak{A})) = \left\{ \{v, w\} \in \binom{A}{2} : v, w \text{ occur together in any tuple of any relation of } \mathfrak{A} \right\}.$$

3.2 First-order logic

Definitions in this subsection are with respect to an arbitrary fixed relational signature (i.e., a finite set of relation symbols, each with an associated positive integer "arity").

▶ **Definition 3.5** (First-order formulas). *Formulas* of first-order logic (denoted by φ, ψ, θ) are constructed from
- *atomic formulas* $x = y$ and $Rx_1 \dots x_t$ (for a t-ary relation symbol R) via
- *Boolean connectives* $\varphi \wedge \psi$ and $\varphi \vee \psi$ and $\neg\varphi$ and
- *quantifiers* $\exists x\, \varphi$ and $\forall x\, \varphi$.

(Here $x, y, x_1, \dots, x_t$ are arbitrary variable symbols.)

A formula is said to be:
- a *sentence* if it contains no free variables (i.e., if every occurrence of every variable symbol is bounded by a quantifier),
- *positive* if it contains no negations ($\neg$),
- *existential-positive* if it contains no negations ($\neg$) or universal quantifiers ($\forall$),
- *primitive-positive* if it contains no negations ($\neg$), universal quantifiers ($\forall$) or disjunction ($\vee$).

In other words, primitive-positive formulas are constructed from atomic formulas via existential quantification ($\exists$) and conjunction ($\wedge$) only; existential-positive formulas additionally allow disjunction ($\vee$).

▶ **Definition 3.6** (Quantifier rank and variable width). Two important parameters of first-order formulas are:
- *quantifier rank*, defined as the maximum nesting depth of quantifiers, and
- *variable width*, defined as the maximum number of free variables in any subformula.

▶ **Definition 3.7** (Alternation height). A first-order formula φ has alternation height 0 iff it is quantifier-free (i.e., a Boolean combination of atomic formulas). For $d \geq 1$, φ has alternation height at most d iff it is a Boolean combination of finitely many first-order formulas $\psi_1, \dots, \psi_m$, each of the form $\exists x_1 \dots \exists x_k\, \theta$ or $\forall x_1 \dots \forall x_k\, \theta$ some $k \geq 0$ and θ with alternation height at most $d - 1$.

▶ **Example 3.8.** To illustrate various tradeoffs in parameters, we present four different primitive-positive sentences, all which define the class $\vec{P}_9 \to *$ where $\vec{P}_9$ is the directed path graph of order 9.

(a) quantifier rank 5, variable width 3, alternation height 3

$$\exists x_0\,\exists x_8\,\exists x_4 \left(\begin{array}{r}\exists x_2 \Big(\exists x_1\,(Ex_0x_1 \wedge Ex_1x_2) \wedge \exists x_3\,(Ex_2x_3 \wedge Ex_3x_4)\Big)\\ \wedge\ \exists x_6 \Big(\exists x_5\,(Ex_4x_5 \wedge Ex_5x_6) \wedge \exists x_7\,(Ex_6x_7 \wedge Ex_7x_8)\Big)\end{array}\right)$$

(b) quantifier rank 4 (the minimum possible), variable width 3, alternation height 4

$$\exists x_4 \left(\begin{array}{r}\exists x_2 \Big(\exists x_1\,(\exists x_0\, Ex_0x_1 \wedge Ex_1x_2) \wedge \exists x_3\,(Ex_2x_3 \wedge Ex_3x_4)\Big)\\ \wedge\ \exists x_6 \Big(\exists x_5\,(Ex_4x_5 \wedge Ex_5x_6) \wedge \exists x_7\,(Ex_6x_7 \wedge \exists x_8\, Ex_7x_8)\Big)\end{array}\right)$$

(c) quantifier rank 9, variable width 2 (the minimum possible), alternation height 8

$$\exists x_0\,\exists x_1\,(Ex_0x_1 \wedge \exists x_2\,(Ex_1x_2 \wedge \exists x_3\,(Ex_2x_3 \wedge \cdots \exists x_7\,(Ex_6x_7 \wedge \exists x_8\, Ex_7x_8)\cdots)))$$

(d) quantifier rank 9, variable width 9, alternation height 1 (the minimum possible)

$$\exists x_0\,\exists x_1\,\exists x_2\,\cdots\,\exists x_7\,\exists x_8\,(Ex_0x_1 \wedge (Ex_1x_2 \wedge (Ex_2x_3 \wedge \cdots (Ex_6x_7 \wedge Ex_7x_8)\cdots)))$$

4 Characterization of $\exists^+$ definability via the cops-and-robber game

The main result of this section (Lemma 4.6) provides a useful characterization of the classes of structures that are definable by existential-positive sentences with a given quantifier rank r, variable width s, and alternation height d. This characterization uses the well-known cops-and-robber game, which is also a means of defining the graph parameters *tree-width* and *tree-depth*.

4.1 Cops-and-robber game

Let $G = (V, E)$, $E \subseteq \binom{V}{2}$, be a finite simple graph. *Tree-width* and *tree-depth*, denoted by $\mathbf{tw}(G)$ and $\mathbf{td}(G)$, are well-studied graphs parameters that are usually defined in terms of tree-like decompositions of G. Below, we present an alternative definition of these parameters in terms of a two-player pursuit-evasion game. This "cops-and-robbers" characterization of $\mathbf{tw}(G)$ and $\mathbf{td}(G)$ is better suited to our purposes in this paper.

▶ **Definition 4.1** (Cops-and-robber game). For any $d \geq 0$, the *height-d cops-and-robber game* on a graph G is a pursuit-evasion between two players: a team of *cops* and a sole *robber* (each with complete knowledge of the other's moves). The game is played in a sequence of d rounds as follows:

- Initially, the robber positions himself on any vertex of his choice.
- In round 1 of the game, any number of cops take up positions on their choice of vertices. The robber then moves to any vertex in the same connected component (bypassing any newly positioned cops that stand in the way). The game ends immediately only if the robber moves to a position guarded by a cop.
- In round 2 of the game, any subset of the assigned cops remain on their stationed vertices, and any number of (reassigned or additional) cops take up new positions on their choice of vertices. The robber then moves to any vertex in that reachable via a path that avoids the stationary cops (but may bypass any newly (re)positioned cops).
- The game proceeds in this manner for up to d rounds, ending immediately if the robber ever occupies the same vertex as a cop at the end of a round. This situation is a win for the cop team; otherwise the robber wins if not caught after d rounds.

The team of cops clearly have a winning strategy for any $d \geq 1$: in the first round, simply occupy all vertices in the connected component of the robber. The two questions that concern us are:

- *How many distinct cops are required to catch the robber?*
- *How many distinct cop "moves" (i.e., instances of positioning a cop on a vertex) are required to catch the robber?*

When the "height" of the game (i.e., number of rounds) is unbounded, the answers to these questions respectively characterize the *tree-width* and *tree-depth* of G [37, 18].

Taking into account the height d, we get two hierarchies of parameters

$$\mathbf{tw}_1(G) \geq \mathbf{tw}_2(G) \geq \cdots \quad \text{and} \quad \mathbf{td}_1(G) \geq \mathbf{td}_2(G) \geq \cdots$$

defined as follows:

- $\mathbf{tw}_d(G)$ is the maximum $s \geq 0$ such that the **robber** has a winning strategy in the height-d ∞-move s-cops-and-robber game on G.
- $\mathbf{td}_d(G)$ is the minimum $r \geq 1$ such that the **cops** have a winning strategy in height-d r-move ∞-cops-and-robber game.

Observe that

$$\mathbf{tw}_1(G) + 1 = \mathbf{td}_1(G) = \text{maximum \# of vertices in a connected component of } G.$$

Also note that $\mathbf{tw}_d(G) + 1 \leq \mathbf{td}_d(G)$ for all d. Finally, note that $\mathbf{tw}_{|V(G)|}(G) = \mathbf{tw}(G)$ and $\mathbf{td}_{|V(G)|}(G) = \mathbf{td}(G)$.

▶ **Example 4.2.** With respect to the path graph P_k of order k, it is well-known that $\mathbf{tw}(P_k) = 2$ and $\mathbf{td}(P_k) = \log_2(k) + O(1)$. Moreover it is not hard to show that

$$\mathbf{tw}_d(P_k) = k^{1/d} + O(1) \quad \text{and} \quad \mathbf{td}_d(P_k) = (d + o(d))k^{1/d}$$

for all $d \leq \log_2(k)$.

In the remainder of this paper, we are not interested in $\mathbf{tw}_d(G)$ and $\mathbf{td}_d(G)$ *per se*, but rather in tradeoffs among all three parameters in the cops-and-robbers game: the numbers of cops, cop moves, and height. That is, for any triple of parameters r, s, d, which side has a winning strategy in the ***height-d r-move s-cops-and-robber game*** on G?

▶ Remark 4.3. Tradeoffs between r and s (when d is unbounded) were recently studied in [15], and monotonicity of an optimal cops strategy in the r-move s-cops-and-robber game was established in [3]. In both of those articles, the r-move s-cops-and-robber game is called the "r-round s-cops-and-robber game". Optimizing s with respect to fixed r characterizes a parameter called the *depth-r tree-width* of G.

In the present article, we use terminology "height-d" instead of "d-round" to avoid confusion with the previous terminology. We propose names *height-d tree-width* and *height-d tree-depth* for parameters $\mathbf{tw}_d(G)$ and $\mathbf{td}_d(G)$.

4.2 $\exists^+$ definability of homomorphism-closed classes

We shall now review a well-known characterization of the parameters (quantifier rank and variable width) required to define the class of structures $\mathfrak{C} \to *$ by a primitive-positive sentence, for finite core $\mathfrak{C}$. In fact, we slightly extend this characterization by including alternation height as a third parameter.

▶ **Lemma 4.4.** *For any finite core $\mathfrak{C}$ and integer $d, r, s \geq 0$, the following statements are equivalent:*

(i) *The class $\mathfrak{C} \to *$ is definable by a primitive-positive sentence with quantifier rank r, variable width s, and alternation height d.*

(ii) *The* **cops** *have a winning strategy in the height-d r-move s-cops-and-robber game on* $\mathsf{Gaif}(\mathfrak{C})$.

Results very similar to Lemma 4.4, which consider only one or two of the parameters d, r, s, have appeared before in the literature [1, 3, 12, 15, 18, 19, 37]. Lemma 4.4 may be used to characterize the parameters of existential-positive sentences that define a given homomorphism-closed class of finite structures.

▶ **Definition 4.5** (Minimal cores of a homomorphism-closed class of finite structures).

- A class of finite structures $\mathcal{C}$ is *homomorphism-closed* if

$$(\mathfrak{A} \in \mathcal{C} \text{ and } \mathfrak{A} \to \mathfrak{B}) \implies \mathfrak{B} \in \mathcal{C}$$

for all finite structures $\mathfrak{A}$ and $\mathfrak{B}$.

- A *minimal core* in $\mathcal{C}$ is a core $\mathfrak{C} \in \mathcal{C}$ such that

$$(\mathfrak{A} \in \mathcal{C} \text{ and } \mathfrak{A} \to \mathfrak{C}) \implies \mathfrak{A} \rightleftarrows \mathfrak{C}$$

for all finite structures $\mathfrak{A}$.

Note that $\mathcal{C}$ is determined by its set of minimal cores: a finite structures $\mathfrak{A}$ belongs to $\mathcal{C}$ if, and only if, there is a homomorphism to $\mathfrak{A}$ from at least one minimal core in $\mathcal{C}$.

▶ **Lemma 4.6.** *Let $\mathcal{C}$ be a homomorphism-closed class of finite structures. For any $d, r, s \geq 0$, the following statements are equivalent:*

(i) *$\mathcal{C}$ is definable on finite structures by an existential-positive sentence with quantifier rank r, variable width s, and alternation height d.*

(ii) *The* **cops** *have a winning strategy in the height-d r-move s-cops-and-robber game on* $\mathsf{Gaif}(\mathfrak{C})$*, for every minimal core $\mathfrak{C}$ in $\mathcal{C}$.*

Lemma 4.6 follows from Lemma 4.4 by standard arguments (see [34, Proposition 2.16]). The only minor subtlety in the proof is a reliance on the fact that, for any given finite relational signature and $r \geq 0$, there are only a finite number of non-isomorphic cores with tree-depth r [23]. This is required so that the disjunction of primitive-positive sentences defining $\mathfrak{C} \to *$, over the non-isomorphic minimal cores $\mathfrak{C}$ in $\mathcal{C}$, constitutes a well-defined (finite length) existential-positive sentence.

5 Generalized Cai-Fürer-Immerman construction

In this section we present the generalized Cai-Fürer-Immerman construction discussed in §2.3 and establish its key properties given by Lemmas 5.5 and 5.6.

▶ **Notation 5.1.** Let $\mathbb{Z}_2$ denote the group $\{0, 1\}$ with addition modulo 2.

We will make use of the following basic lemma of graph homology, stated here over the coefficient group $\mathbb{Z}_2$.

▶ **Lemma 5.2.** *For any graph $G = (V, E)$, $E \subseteq \binom{V}{2}$, and function $\xi : V \to \mathbb{Z}_2$, the following statements are equivalent:*

(i) *$\sum_{u \in U} \xi(u) = 0$ for every connected component $U \subseteq V$.*

(ii) *ξ is a "1-boundary", that is, there exists a function $\varepsilon : E \to \mathbb{Z}_2$ such that for every $v \in V$, we have $\sum_{e \in E \,:\, v \in e} \varepsilon(e) = \xi(v)$.*

▶ **Notation 5.3.** Let $\mathfrak{A}$ be a structure with Gaifman graph $G = (A, E)$. For an element $v \in A$, let E_v, N_v and $N_v^\bullet$ respectively denote the incident-edge set, neighbor set and 1-neighborhood of v in G. That is,

$$E_v := \Big\{\{v, w\} : w \in A \text{ such that } \{v, w\} \text{ is an edge in } G\Big\},$$
$$N_v := \Big\{w \in A : \{v, w\} \text{ is an edge in } G\Big\},$$
$$N_v^\bullet := N_v \cup \{v\}.$$

▶ **Definition 5.4** (Generalized Cai-Fürer-Immerman structures $\mathfrak{A}_\xi$)**.** For any relational structure $\mathfrak{A}$ with universe A and any function $\xi : A \to \mathbb{Z}_2$, we define a structure $\mathfrak{A}_\xi$ (with the same signature) as follows:

- $\mathfrak{A}_\xi$ has universe

$$A_\xi := \Big\{\langle v, \alpha\rangle : v \in A \text{ and } \alpha : N_v^\bullet \to \mathbb{Z}_2 \text{ such that } \alpha(v) = 0 \text{ and } \sum_{w \in N_v} \alpha(w) = \xi(v)\Big\}.$$

  (Here for readability sake we use a distinctive notation $\langle v, \alpha\rangle$ for the ordered pair (v, α).)
- For each t-ary relation $R \subseteq A^t$ in $\mathfrak{A}$, the corresponding relation $R_\xi \subseteq A_\xi^t$ in $\mathfrak{A}_\xi$ is defined by

$$R_\xi := \left\{(\langle v_1, \alpha_1\rangle, \dots, \langle v_t, \alpha_t\rangle) \in A_\xi^t : \begin{array}{l}(v_1, \dots, v_t) \in R \text{ and} \\ \alpha_i(v_j) = \alpha_j(v_i) \text{ for all } i, j \in \{1, \dots, t\}\end{array}\right\}.$$

Note that the projection $\langle v, \alpha\rangle \mapsto v$ is a homomorphism $\mathfrak{A}_\xi \to \mathfrak{A}$. Also note that this homomorphism need not be surjective (for instance, if $v \in A$ is an isolated vertex in $\mathsf{Gaif}(\mathfrak{A})$ and $\xi(v) = 1$).

Our first key lemma gives a necessary and sufficient condition for the existence of a homomorphism in the opposite direction in the special case that $\mathfrak{A}$ is a core.

▶ **Lemma 5.5.** *Let $\mathfrak{C}$ be a finite core with Gaifman graph $G = (C, E)$, and let $\xi : C \to \mathbb{Z}_2$. There exists a homomorphism $\mathfrak{C} \to \mathfrak{C}_\xi$ if, and only if, $\sum_{u \in U} \xi(u) = 0$ for each connected component U of G.*

Proof of Lemma 5.5. We first prove the "if" direction. Assume that $\sum_{u \in U} \xi(u) = 0$ for each connected component U of $\mathfrak{C}$. By Lemma 5.2, there exists a function $\varepsilon : E \to \mathbb{Z}_2$ such that for all $v \in C$, we have

$$\sum_{e \in E_v} \varepsilon(e) = \xi(v).$$

For each $v \in C$, define $\alpha_v : N_v \to \mathbb{Z}_2$ by $\alpha_v(v) := 0$ and $\alpha_v(w) := \varepsilon(\{v, w\})$ for all $w \in N_v \setminus \{v\}$. Note that $\langle v, \alpha_v\rangle \in C_\xi$. The function $h : v \mapsto \langle v, \alpha_v\rangle$ is the desired homomorphism $\mathfrak{C} \to \mathfrak{C}_\xi$.

We now prove the "only if" direction. Assume that h is an arbitrary homomorphism $\mathfrak{C} \to \mathfrak{C}_\xi$. Consider the projection homomorphism $\langle v, \alpha\rangle \mapsto v : \mathfrak{C}_\xi \to \mathfrak{C}$ that maps $\langle v, \alpha\rangle$ to v, and let f be the composition

$$f = (\langle v, \alpha\rangle \mapsto v) \circ h : \mathfrak{C} \to \mathfrak{C}.$$

Since $\mathfrak{C}$ is a core, f is an isomorphism. In particular, note that f restricts to a bijection from N_v to $N_{f(v)}$ for each $v \in C$.

For each $v \in C$, we have $h(v) = \langle f(v), \alpha_v \rangle$ for some $\alpha_v : N^{\bullet}_{f(v)} \to \mathbb{Z}_2$. Since $h(v) \in C_\xi$, we have

$$\alpha_v(f(v)) = 0 \quad \text{ and } \quad \sum_{x \in N_{f(v)}} \alpha_v(x) = \xi(f(v)).$$

We now define $\widetilde{\alpha}_v : N^{\bullet}_v \to \mathbb{Z}_2$ by

$$\widetilde{\alpha}_v(w) := \alpha_v(f(w)).$$

Using the fact that f maps N_v bijectively to $N_{f(v)}$, we have

$$\widetilde{\alpha}_v(v) = \alpha_v(f(v)) = 0 \quad \text{ and } \quad \sum_{w \in N_v} \widetilde{\alpha}_v(w) = \sum_{w \in N_v} \alpha_v(f(w)) = \sum_{x \in N_{f(v)}} \alpha_v(x) = \xi(f(v)).$$

Therefore, we have $\langle v, \widetilde{\alpha}_v \rangle \in C_{\widetilde{\xi}}$ where $\widetilde{\xi} : C \to \mathbb{Z}_2$ is the function $\widetilde{\xi}(v) := \xi(f(v))$.

Let us next consider the function $\widetilde{h} : C \to C_{\widetilde{\xi}}$ defined by

$$\widetilde{h}(v) := \langle v, \widetilde{\alpha}_v \rangle.$$

We claim that $\widetilde{h}$ is a homomorphism $\mathfrak{C} \to \mathfrak{C}_{\widetilde{\xi}}$. (We prove this claim only in order to show that $\widetilde{\alpha}_v(w) = \widetilde{\alpha}_w(v)$ for all $\{v, w\} \in E$.) To see why, consider any tuple $(v_1, \dots, v_t) \in R$ in any relation of $\mathfrak{C}$. Since h is a homomorphism $\mathfrak{C} \to \mathfrak{C}_\xi$, we have

$$(h(v_1), \dots, h(v_r)) = (\langle f(v_1), \alpha_{v_1} \rangle, \dots, \langle f(v_t), \alpha_{v_t} \rangle) \in R_\xi.$$

By definition of R_ξ, for all $i, j \in \{1, \dots, t\}$, we have $\alpha_{v_i}(f(v_j)) = \alpha_{v_j}(f(v_i))$ and hence

$$\widetilde{\alpha}_{v_i}(v_j) = \alpha_{v_i}(f(v_j)) = \alpha_{v_j}(f(v_i)) = \widetilde{\alpha}_{v_j}(v_i).$$

So we see that (by definition of $R_{\widetilde{\xi}}$)

$$(\widetilde{h}(v_1), \dots, \widetilde{h}(v_t)) = (\langle v_1, \widetilde{\alpha}_{v_1} \rangle, \dots, \langle v_t, \widetilde{\alpha}_{v_t} \rangle) \in R_{\widetilde{\xi}}.$$

This argument shows that $\widetilde{h}$ is a homomorphism $\mathfrak{C} \to \mathfrak{C}_{\widetilde{\xi}}$ as claimed.

We now define a function $\varepsilon : E \to \mathbb{Z}_2$ by

$$\varepsilon(\{v, w\}) := \widetilde{\alpha}_v(w).$$

This is well-defined, since (as established in previous paragraph) we have $\widetilde{\alpha}_v(w) = \widetilde{\alpha}_w(v)$ for all $\{v, w\} \in E$ (i.e., for all distinct $v, w \in C$ that appear together in any tuple of any relation of $\mathfrak{C}$).

For all $v \in C$, we have

$$\sum_{e \in E_v} \varepsilon(e) = \sum_{w \in N_v} \varepsilon(\{v, w\}) = \sum_{w \in N_v} \widetilde{\alpha}_v(w) = \xi(f(v)) = \widetilde{\xi}(v).$$

By Lemma 5.2, it follows that $\sum_{u \in U} \widetilde{\xi}(u) = 0$ for each connected component U of C. Since $\mathfrak{C}$ is a core, $f : \mathfrak{C} \to \mathfrak{C}$ restricts to a bijection on each connected component. We conclude that

$$\sum_{u \in U} \xi(u) = \sum_{u \in U} \xi(f(u)) = \sum_{u \in U} \widetilde{\xi}(u) = 0$$

as required. ◀

The second key lemma concerns the parameters of first-order sentences that distinguish any two structures in the class $\{\mathfrak{A}_\xi : \xi \text{ is a function from } A \text{ to } \mathbb{Z}_2\}$.

▶ **Lemma 5.6.** *Let $\mathfrak{A}$ be a finite structure with Gaifman graph $G = (A, E)$. Assume that the* **robber** *has a winning strategy starting on vertex $u \in A$ in the height-d r-move s-ccops-and-robber game on G. Further assume that $\xi, \zeta : A \to \mathbb{Z}_2$ and $\varepsilon : E \to \mathbb{Z}_2$ are functions satisfying*

$$\xi(v) + \zeta(v) + \sum_{e \in E_v} \varepsilon(e) = \mathbb{1}[\ v = u\ ] \text{ for all } v \in A.$$

Then structures $\mathfrak{A}_\xi$ and $\mathfrak{A}_\zeta$ are indistinguishable by first-order sentences with quantifier rank r, variable width s, and alternation height d.

We obtain Lemma 5.6 as the $k = 0$ case of the following more general lemma, whose statement is suited for proof by induction on the alternation height d.

▶ **Lemma 5.7.** *Let $\mathfrak{A}$ be a finite structure with Gaifman graph $G = (A, E)$. Assume that the* **robber** *has a winning strategy in the height-d r-move s-ccops-and-robber game on G with k ($\le s$) cops initially positioned at vertices $v_1, \dots, v_k \in A$ and the robber initially positioned at vertex $u \in A \setminus \{v_1, \dots, v_k\}$. Further assume that $\xi, \zeta : A \to \mathbb{Z}_2$ and $\varepsilon : E \to \mathbb{Z}_2$ and $\alpha_i, \beta_i : N^\bullet_{v_i} \to \mathbb{Z}_2$ are functions satisfying*

$$\begin{aligned}
\langle v_1, \alpha_1\rangle, \dots, \langle v_k, \alpha_k\rangle &\in A_\xi,\\
\langle v_1, \beta_1\rangle, \dots, \langle v_k, \beta_k\rangle &\in A_\zeta,\\
\alpha_i(w) + \beta_i(w) + \varepsilon(\{v_i, w\}) &= 0 \text{ for all } i \in \{1, \dots, k\} \text{ and } w \in N_{v_i},\\
\xi(v) + \zeta(v) + \sum_{e \in E_v} \varepsilon(e) &= \mathbb{1}[\ v = u\ ] \text{ for all } v \in A.
\end{aligned}$$

Then for every first-order formula $\varphi(x_1, \dots, x_k)$ with quantifier rank r, variable width s, and alternation height d, we have

$$\mathfrak{A}_\xi \models \varphi(\langle v_1, \alpha_1\rangle, \dots, \langle v_k, \alpha_k\rangle) \iff \mathfrak{A}_\zeta \models \varphi(\langle v_1, \beta_1\rangle, \dots, \langle v_k, \beta_k\rangle).$$

Proof. We argue by induction on d. The base case $d = 0$ is equivalent to showing that $\mathfrak{A}_\xi$ and $\mathfrak{A}_\zeta$ satisfy the same quantifier-free formulas. Here it suffices to consider only the atomic formulas. That is, we must show

- $\langle v_i, \alpha_i\rangle = \langle v_j, \alpha_j\rangle \iff \langle v_i, \beta_i\rangle = \langle v_j, \beta_j\rangle$ for all indices $i, j \in \{1, \dots, k\}$, and
- $(\langle v_{i_1}, \alpha_{i_1}\rangle, \dots, \langle v_{i_t}, \alpha_{i_t}\rangle) \in R_\xi \iff (\langle v_{i_1}, \beta_{i_1}\rangle, \dots, \langle v_{i_t}, \beta_{i_t}\rangle) \in R_\zeta$ for every t-ary relation symbol R and indices $i_1, \dots, i_t \in \{1, \dots, k\}$.

Both equivalences follow from our assumptions on $\xi, \zeta, \alpha_i, \beta_i, \varepsilon$. In particular, the second equivalence follows from the definition of relations R_ξ, R_ζ and the observation that

$$\alpha_i(v_j) = \alpha_j(v_i) \iff \alpha_i(v_j) + \varepsilon(\{v_i, v_j\}) = \alpha_j(v_i) + \varepsilon(\{v_i, v_j\}) \iff \beta_i(v_j) = \beta_j(v_i)$$

for all $i, j \in \{1, \dots, k\}$.

For the induction step, assume that $d \ge 1$. By definition of having alternation height d, $\varphi(x_1, \dots, x_k)$ is a Boolean combination of finitely many first-order formulas $\psi(x_{i_1}, \dots, x_{i_j})$, each of the form

$$\exists y_1 \dots \exists y_\ell\ \theta(x_{i_1}, \dots, x_{i_j}, y_1, \dots, y_\ell) \quad \text{or} \quad \forall y_1 \dots \forall y_\ell\ \theta(x_{i_1}, \dots, x_{i_j}, y_1, \dots, y_\ell)$$

for some $j, \ell \geq 0$ and $1 \leq i_1 < \cdots < i_j \leq k$ and first-order formula θ with quantifier rank (at most) $r - \ell$, variable width (at most) s, and alternation depth (at most) $d-1$. Consider any such formula $\psi(x_1, \ldots, x_j)$, without loss of generality of the form $\exists y_1 \ldots \exists y_\ell\ \theta(x_1, \ldots, x_j, y_1, \ldots, y_\ell)$ where $(i_1, \ldots, i_j) = (1, \ldots, j)$. It suffices to show that

$$\mathfrak{A}_\xi \models \psi(\langle v_1, \alpha_1\rangle, \ldots, \langle v_j, \alpha_j\rangle) \iff \mathfrak{A}_\zeta \models \psi(\langle v_1, \beta_1\rangle, \ldots, \langle v_j, \beta_j\rangle).$$

We will prove the implication $\Longrightarrow$; the reverse implication follows by a symmetric argument.

Assume that $\mathfrak{A}_\xi \models \psi(\langle v_1, \alpha_1\rangle, \ldots, \langle v_j, \alpha_j\rangle)$ and fix a choice of $\langle \widehat{v}_1, \widehat{\alpha}_1\rangle, \ldots, \langle \widehat{v}_\ell, \widehat{\alpha}_\ell\rangle \in A_\xi$ such that

$$\mathfrak{A}_\xi \models \theta(\langle v_1, \alpha_1\rangle, \ldots, \langle v_j, \alpha_j\rangle, \langle \widehat{v}_1, \widehat{\alpha}_1\rangle, \ldots, \langle \widehat{v}_\ell, \widehat{\alpha}_\ell\rangle).$$

In the remainder of this proof, we will show that there exist functions $\widehat{\beta}_i : N^\bullet_{\hat{v}_\ell} \to \mathbb{Z}_2$ with $\langle \widehat{v}_i, \widehat{\beta}_i\rangle \in A_\zeta$ $(i \in \{1, \ldots, \ell\})$ such that

$$\mathfrak{A}_\zeta \models \theta(\langle v_1, \alpha_1\rangle, \ldots, \langle v_j, \alpha_j\rangle, \langle \widehat{v}_1, \widehat{\beta}_1\rangle, \ldots, \langle \widehat{v}_\ell, \widehat{\beta}_\ell\rangle).$$

It then follows that $\mathfrak{A}_\zeta \models \psi(\langle v_1, \beta_1\rangle, \ldots, \langle v_j, \beta_j\rangle)$, which establishes the required implication

$$\mathfrak{A}_\xi \models \psi(\langle v_1, \alpha_1\rangle, \ldots, \langle v_j, \alpha_j\rangle) \implies \mathfrak{A}_\zeta \models \psi(\langle v_1, \beta_1\rangle, \ldots, \langle v_j, \beta_j\rangle).$$

In order to define suitable functions $\widehat{\beta}_i$, we invoke the robber's winning strategy in the height-d r-move s-ccops-and-robber game on G with cops starting at $v_1, \ldots, v_k$ and the robber starting at u. Suppose that in the round 1 of the game, the first j cops remain at $v_1, \ldots, v_j$ while the next ℓ cops redeploy to $\widehat{v}_1, \ldots, \widehat{v}_\ell$. There exists $\widehat{u} \in V \setminus \{v_1, \ldots, v_j, \widehat{v}_1, \ldots, \widehat{v}_\ell\}$ and a path $u = p_0, p_1, \ldots, p_m = \widehat{u}$ in G (with $m \geq 0$ and $\{p_{i-1}, p_i\} \in E$ for all $1 \leq i \leq m$) such that $\{v_1, \ldots, v_j\} \cap \{p_0, \ldots, p_m\} = \emptyset$ and the robber has a winning strategy in the $d-1$-round $r-\ell$-move s-ccops-and-robber game on G with cops starting at $v_1, \ldots, v_j, \widehat{v}_1, \ldots, \widehat{v}_\ell$ and the robber starting at $\widehat{u}$.

Define $\widehat{\varepsilon} : E \to \mathbb{Z}_2$ and $\widehat{\beta}_i : N^\bullet_{\hat{v}_i} \to \mathbb{Z}_2$ $(i \in \{1, \ldots, \ell\})$ by

$$\widehat{\varepsilon}(e) := \varepsilon(e) + \sum_{i=1}^{m} \mathbb{1}[\ e = \{p_{i-1}, p_i\}\ ],$$

$$\widehat{\beta}_i(w) := \begin{cases} 0 & \text{if } w = \widehat{v}_i, \\ \widehat{\alpha}_i(w) + \widehat{\varepsilon}(\{\widehat{v}_i, w\}) & \text{if } w \in N_{\hat{v}_i}. \end{cases}$$

We will next show that $\xi, \zeta, \alpha_1, \ldots, \alpha_j, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_\ell, \beta_1, \ldots, \beta_j, \widehat{\beta}_1, \ldots, \widehat{\beta}_\ell$ and $\widehat{e}$ satisfy the conditions of the lemma with respect to $v_1, \ldots, v_j, \widehat{v}_1, \ldots, \widehat{v}_\ell, \widehat{u}$ and the first-order formula θ.

First, note that

$$\langle v_1, \alpha_1\rangle, \ldots, \langle v_j, \alpha_j\rangle, \langle \widehat{v}_1, \widehat{\alpha}_1\rangle, \ldots, \langle \widehat{v}_\ell, \widehat{\alpha}_\ell\rangle \in A_\xi.$$

Second, to establish that

$$\langle v_1, \beta_1\rangle, \ldots, \langle v_j, \beta_j\rangle, \langle \widehat{v}_1, \widehat{\beta}_1\rangle, \ldots, \langle \widehat{v}_\ell, \widehat{\beta}_\ell\rangle \in A_\zeta,$$

we observe that for each $i \in \{1, \ldots, \ell\}$,

$$\begin{aligned}
\sum_{w \in N_{\hat{v}_i}} \widehat{\beta}_i(w) &= \sum_{w \in N_{\hat{v}_i}} \Big(\widehat{\alpha}_i(w) + \widehat{\varepsilon}(\{\widehat{v}_i, w\})\Big) \\
&= \sum_{w \in N_{\hat{v}_i}} \Big(\widehat{\alpha}_i(w) + \varepsilon(\{\widehat{v}_i, w\}) + \sum_{i=1}^{m} \mathbb{1}[\ \{\widehat{v}_i, w\} = \{p_{i-1}, p_i\}\ ]\Big) \\
&= \xi(\widehat{v}_i) + \sum_{e \in E_{\widehat{v}_i}} \varepsilon(e) + \sum_{w \in N_{\hat{v}_i}} \sum_{i=1}^{m} \mathbb{1}[\ \{\widehat{v}_i, w\} = \{p_{i-1}, p_i\}\ ] \\
&= \zeta(\widehat{v}_i) + \mathbb{1}[\ \widehat{v}_i = u\ ] + \mathbb{1}[\ \widehat{v}_i = p_0\ ] + \mathbb{1}[\ \widehat{v}_i = \widehat{u}\ ] \\
&= \zeta(\widehat{v}_i) \quad (\text{since } p_0 = u \text{ and } \widehat{u} \notin \{v_1, \ldots, v_j, \widehat{v}_1, \ldots, \widehat{v}_\ell\}).
\end{aligned}$$

Third, by definition of $\widehat{\beta}_i$, we have

$$\widehat{\alpha}_i(w) + \widehat{\beta}_i(w) + \widehat{\varepsilon}(\{\widehat{v}_i, w\}) = 0 \text{ for all } i \in \{1, \ldots, \ell\} \text{ and } w \in N_{\hat{v}_i}.$$

Fourth and finally, for all $v \in A$, we have

$$\begin{aligned}
\xi(v) + \zeta(v) + \sum_{e \in E_v} \widehat{\varepsilon}(e) &= \xi(v) + \zeta(v) + \sum_{e \in E_v} \varepsilon(e) + \sum_{w \in N_v} \sum_{i=1}^{m} \mathbb{1}[\ \{v, w\} = \{p_{i-1}, p_i\}\ ] \\
&= \mathbb{1}[\ v = u\ ] + \mathbb{1}[\ v = p_0\ ] + \mathbb{1}[\ v = p_m\ ] \\
&= \mathbb{1}[\ v = \widetilde{u}\ ].
\end{aligned}$$

By the induction hypothesis applied to θ, we conclude that

$$\mathfrak{A}_\zeta \models \theta(\langle v_1, \alpha_1 \rangle, \ldots, \langle v_j, \alpha_j \rangle, \langle \widehat{v}_1, \widehat{\beta}_1 \rangle, \ldots, \langle \widehat{v}_\ell, \widehat{\beta}_\ell \rangle),$$

finishing the proof. ◀

6 Proof of the equi-rank finitary HPT

Proof of Theorem 1.4. Let φ be a first-order sentence that is preserved under homomorphisms on finite structures. Let r, s and d be the quantifier rank, variable width and alternation height of φ.

Assume that φ has at least one finite model, since otherwise the theorem is trivial (allowing $\perp$ as a special primitive-positive sentence with no models). Consider any minimal core $\mathfrak{C}$ (with universe C) in the class of finite models of φ. By Lemma 4.6, it suffices to show that the **cops** have a winning strategy in the height-d r-move s-cops-and-robber game on $\mathsf{Gaif}(\mathfrak{C})$, starting from an (adversarial) choice of initial position $u \in C$ for the robber.

Let $\mathfrak{C}_0$ be the CFI structure where 0 stands for the all-zero function $C \to \mathbb{Z}_2$, and let $\mathfrak{C}_{1_u}$ be the CFI structure where 1_u stands for the function $C \to \mathbb{Z}_2$ with value 1 at u and 0 elsewhere. Additionally, let ε be the all-zero function $E \to \mathbb{Z}_2$. Note that

$$0(v) + 1_u(v) + \sum_{e \in E_v} \varepsilon(e) = 0 + \mathbb{1}[\ v = u\ ] + \sum_{e \in E_v} 0 = \mathbb{1}[\ v = u\ ] \text{ for all } v \in A.$$

By Lemma 5.5, we have $\mathfrak{C} \to \mathfrak{C}_0$. Since φ is closed under homomorphisms on finite structures, it follows that $\mathfrak{C}_0 \models \varphi$. Lemma 5.5 also implies $\mathfrak{C} \not\to \mathfrak{C}_{1_u}$. Since $\mathfrak{C}_{1_u} \to \mathfrak{C}$, our assumption that $\mathfrak{C}$ is a minimal core in the class of finite models of φ implies that $\mathfrak{C}_{1_u} \not\models \varphi$.

We have established that φ is a first-order sentence with quantifier rank r, variable width s and alternation height d, which distinguishes the pair of structures $\mathfrak{C}_0$ and $\mathfrak{C}_{1_u}$. Therefore, by (the contrapositive of) Lemma 5.6, the **cops** have a winning strategy with the robber starting on $u \in C$ in the height-d r-move s-cops-and-robber game on $\mathsf{Gaif}(\mathfrak{C})$.

By Lemma 4.6, we conclude that φ is equivalent on finite structures to an existential-positive sentence with quantifier rank r, variable width s, and alternation height d, as required. ◀

7 Open questions

As discussed in §2, it remains an open question whether the quantifier-rank blow-up can be eliminated or significantly reduced in either Theorem 2.1 or 2.3 (the main results of [33, 34]).

Another interesting question is to investigate tradeoffs in Theorems 2.1 or 2.3 involving alternation depth d. There is a natural correspondence between *primitive-positive sentences* and *monotone* SAC^0 *circuits* (with unbounded $\bigvee$ gates and fan-in 2 $\wedge$ gates). For any finite graph G, this correspondence gives the following upper bounds on the COLORED G-SUBGRAPH ISOMORPHISM problem (equivalent to the G-HOMOMORPHISM problem when G is a core).

▶ **Proposition 7.1.** *For any finite graph G, the* COLORED G-SUBGRAPH ISOMORPHISM *problem, as a sequence of monotone Boolean functions $\{0,1\}^{|E(G)|\cdot n^2} \to \{0,1\}$, is computable for all $d \geq 1$ by both*

- *monotone SAC^0 formulas with $\bigvee$-depth d and size $n^{\mathbf{td}_d(G)+O(1)}$, and*
- *monotone SAC^0 circuits with $\bigvee$-depth d and size $n^{\mathbf{tw}_d(G)+O(1)}$.*

In the arithmetic setting, the corresponding set-multilinear polynomials are computable by monotone arithmetic SAC^0 formulas and circuits with $\sum$-depth d and size $n^{\mathbf{td}_d(G)+O(1)}$ and $n^{\mathbf{tw}_d(G)+O(1)}$, respectively.

It would be interesting to establish lower bounds in circuit complexity that nearly match the size-depth tradeoffs of Proposition 7.1. Recent work of the author [36] takes a step in this direction by establishing $n^{\Omega(\mathbf{td}_d(G))}$ and $n^{\Omega(\mathbf{tw}_d(G))}$ lower bounds in the case of path graphs $G = P_k$. With respect to monotone arithmetic circuits and formulas, even tighter $n^{\mathbf{td}_d(G)-O(1)}$ and $n^{\mathbf{tw}_d(G)-O(1)}$ lower bounds for general graphs G might be possible using the technique of Komarath, Pandey and Rahul [26].

References

1 Samson Abramsky, Anuj Dawar, and Pengming Wang. The pebbling comonad in finite model theory. In *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–12. IEEE, 2017. doi:10.1109/LICS.2017.8005129.

2 Samson Abramsky and Luca Reggio. Arboreal categories and equi-resource homomorphism preservation theorems. *Annals of Pure and Applied Logic*, 175(6):103423, 2024. doi:10.1016/J.APAL.2024.103423.

3 Isolde Adler and Eva Fluck. Monotonicity of the Cops and Robber Game for Bounded Depth Treewidth. In *49th International Symposium on Mathematical Foundations of Computer Science (MFCS 2024)*, volume 306, pages 6:1–6:18, 2024. doi:10.4230/LIPICS.MFCS.2024.6.

4 Albert Atserias. On digraph coloring problems and treewidth duality. *European Journal of Combinatorics*, 29(4):796–820, 2008. doi:10.1016/J.EJC.2007.11.004.

5 Albert Atserias, Anuj Dawar, and Phokion G. Kolaitis. On preservation under homomorphisms and unions of conjunctive queries. *J. ACM*, 53(2):208–237, 2006. doi:10.1145/1131342.1131344.

6 Jin-Yi Cai, Martin Fürer, and Neil Immerman. An optimal lower bound on the number of variables for graph identification. *Combinatorica*, 12(4):389–410, 1992. doi:10.1007/BF01305232.

7 James Carr. Homomorphism preservation theorems for many-valued structures. *arXiv preprint arXiv:2403.00217*, 2024.

8 Hubie Chen. On the complexity of existential positive queries. *ACM Transactions on Computational Logic (TOCL)*, 15(1):1–20, 2014. doi:10.1145/2559946.

9 Wojciech Czerwinski, Wojciech Nadara, and Marcin Pilipczuk. Improved bounds for the excluded-minor approximation of treedepth. In *27th Annual European Symposium on Algorithms (ESA 2019)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019.

10 Anuj Dawar. Homomorphism preservation on quasi-wide classes. *Journal of Computer and System Sciences*, 76(5):324–332, 2010. doi:10.1016/J.JCSS.2009.10.005.

11 Anuj Dawar and Ioannis Eleftheriadis. Preservation theorems on sparse classes revisited. In *Proceedings of the 49th International Symposium on Mathematical Foundations of Computer Science*, 2024.

12 Anuj Dawar, Tomáš Jakl, and Luca Reggio. Lovász-type theorems and game comonads. In *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–13. IEEE, 2021.

13 Larry Denenberg, Yuri Gurevich, and Saharon Shelah. Definability by constant-depth polynomial-size circuits. *Information and control*, 70(2-3):216–240, 1986. doi:10.1016/S0019-9958(86)80006-7.

14 Tomás Feder and Moshe Y. Vardi. Homomorphism closed vs. existential positive. In *Proceedings of the 18th IEEE Symposium on Logic in Computer Science*, pages 310–320, 2003. doi:10.1109/LICS.2003.1210071.

15 Eva Fluck, Tim Seppelt, and Gian Luca Spitzer. Going deep and going wide: Counting logic and homomorphism indistinguishability over graphs of bounded treedepth and treewidth. In *32nd EACSL Annual Conference on Computer Science Logic (CSL 2024)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2024.

16 Martin Fürer. On the combinatorial power of the Weisfeiler-Lehman algorithm. In *International Conference on Algorithms and Complexity*, pages 260–271. Springer, 2017. doi:10.1007/978-3-319-57586-5_22.

17 Nicola Galesi, Dmitry Itsykson, Artur Riazanov, and Anastasia Sofronova. Bounded-depth Frege complexity of Tseitin formulas for all graphs. *Annals of Pure and Applied Logic*, 174(1):103166, 2023. doi:10.1016/J.APAL.2022.103166.

18 Archontia C Giannopoulou, Paul Hunter, and Dimitrios M Thilikos. LIFO-search: A min–max theorem and a searching game for cycle-rank and tree-depth. *Discrete Applied Mathematics*, 160(15):2089–2097, 2012. doi:10.1016/J.DAM.2012.03.015.

19 Martin Grohe. The complexity of homomorphism and constraint satisfaction problems seen from the other side. *Journal of the ACM*, 54(1):1–24, 2007. doi:10.1145/1206035.1206036.

20 Yuri Gurevich. Toward logic tailored for computational complexity. In M. M. Richter et al., editor, *Computation and Proof Theory*, pages 175–216. Springer Lecture Notes in Mathematics, 1984.

21 Lucy Ham. Relativised homomorphism preservation at the finite level. *Studia Logica*, 105(4):761–786, 2017. doi:10.1007/S11225-017-9710-7.

22 Johan Håstad. On small-depth Frege proofs for Tseitin for grids. *Journal of the ACM (JACM)*, 68(1):1–31, 2020. doi:10.1145/3425606.

23 Pavol Hell and Jaroslav Nešetřil. The core of a graph. *Discrete Math.*, 109:117–126, 1992. doi:10.1016/0012-365X(92)90282-K.

24 Neil Immerman. *Descriptive Complexity Theory.* Graduate Texts in Computer Science. Springer, New York, 1999.

25 Ken-ichi Kawarabayashi and Benjamin Rossman. A polynomial excluded-minor characterization of treedepth. In *29th ACM-SIAM Symposium on Discrete Algorithms*, pages 234–246, 2018.

26 Balagopal Komarath, Anurag Pandey, and Chengot Sankaramenon Rahul. Monotone arithmetic complexity of graph homomorphism polynomials. *Algorithmica*, 85(9):2554–2579, 2023. `doi:10.1007/S00453-023-01108-0`.

27 Deepanshu Kush and Benjamin Rossman. Tree-depth and the formula complexity of subgraph isomorphism. *SIAM Journal on Computing*, 52(1):273–325, 2023. `doi:10.1137/20M1372925`.

28 Yuan Li, Alexander Razborov, and Benjamin Rossman. On the AC^0 complexity of subgraph isomorphism. *SIAM Journal on Computing*, 46(3):936–971, 2017.

29 Leonid Libkin. *Elements of Finite Model Theory*. Springer-Verlag, 2004.

30 Jerzy Łoś. On the extending of models (I). *Fundamenta Mathematicae*, 42:38–54, 1955.

31 Roger C. Lyndon. Properties preserved under homomorphism. *Pacific J. Math.*, 9:129–142, 1959.

32 Daniel Neuen. Homomorphism-distinguishing closedness for graphs of bounded tree-width. In *41st International Symposium on Theoretical Aspects of Computer Science (STACS 2024)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024.

33 Benjamin Rossman. Homomorphism preservation theorems. *Journal of the ACM*, 55(3):15, 2008.

34 Benjamin Rossman. An improved homomorphism preservation theorem from lower bounds in circuit complexity. In *8th Innovations in Theoretical Computer Science*, volume 67 of *LIPIcs*, pages 27:1–17, 2017. `doi:10.4230/LIPICS.ITCS.2017.27`.

35 Benjamin Rossman. Formulas versus circuits for small distance connectivity. *SIAM Journal on Computing*, 47(5):1986–2028, 2018. `doi:10.1137/15M1027310`.

36 Benjamin Rossman. Formula size-depth tradeoffs for iterated sub-permutation matrix multiplication. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1386–1395, 2024. `doi:10.1145/3618260.3649628`.

37 Paul D Seymour and Robin Thomas. Graph searching and a min-max theorem for tree-width. *Journal of Combinatorial Theory, Series B*, 58(1):22–33, 1993. `doi:10.1006/JCTB.1993.1027`.

38 William W. Tait. A counterexample to a conjecture of Scott and Suppes. *J. Symbolic Logic*, 24:15–16, 1959. `doi:10.2307/2964569`.

39 Alfred Tarski. Contributions to the theory of models. I. In *Indagationes Mathematicae (Proceedings)*, volume 57, pages 572–581. Elsevier BV, 1954.

40 Grigori S Tseitin. On the complexity of derivation in propositional calculus. *Automation of reasoning: 2: Classical papers on computational logic 1967–1970*, pages 466–483, 1983.