

ABOUT THE EXPRESSIVE POWER AND COMPLEXITY OF ORDER-INVARIANCE WITH TWO VARIABLES

BARTOSZ BEDNARCZYK ^a AND JULIEN GRANGE ^b

^a Institute of Computer Science, University of Wrocław
e-mail address: bartosz.bednarczyk@cs.uni.wroc.pl

^b LACL, Université Paris-Est Créteil, France
e-mail address: julien.grange@lacl.fr

ABSTRACT. Order-invariant first-order logic is an extension of first-order logic (FO) where formulae can make use of a linear order on the structures, under the proviso that they are order-invariant, *i.e.* that their truth value is the same for all linear orders. We continue the study of the two-variable fragment of order-invariant first-order logic initiated by Zeume and Harwath, and study its complexity and expressive power. We first establish `CONEXPTIME`-completeness for the problem of deciding if a given two-variable formula is order-invariant, which tightens and significantly simplifies the `CON2EXPTIME` proof by Zeume and Harwath. Second, we address the question of whether every property expressible in order-invariant two-variable logic is also expressible in first-order logic without the use of a linear order. We suspect that the answer is “no”. To justify our claim, we present a class of finite tree-like structures (of unbounded degree) in which a relaxed variant of order-invariant two-variable FO expresses properties that are not definable in plain FO. By contrast, we show that if one restricts their attention to classes of structures of bounded degree, then the expressive power of order-invariant two-variable FO is contained within FO.

1. INTRODUCTION

The main goal of finite model theory is to understand formal languages describing finite structures: their complexity and their expressive power. Such languages are ubiquitous in computer science, starting from descriptive complexity, where they are used to provide machine-independent characterisations of complexity classes, and ending up on database theory and knowledge representation, where formal languages serve as fundamental querying formalism. A classical idea in finite model theory is to employ invariantly-used relations, capturing the data-independence principle in databases: it makes sense to give queries the ability to exploit the presence of the order in which the data is stored in the memory, but at the same time we would like to make query results independent of this specific ordering. It is not immediately clear that the addition of an invariantly-used linear order to first-order logic (FO) allow us to gain anything from the standpoint of expressive power. Indeed, as long as we consider arbitrary (*i.e.* not necessarily finite) structures it does not, which is a direct consequence of FO having the Craig Interpolation Property. However, as it was first shown

Key words and phrases: Finite model theory, order-invariance, two-variable logic, complexity.

by Gurevich [Lib04, Thm. 5.3], the claim holds true over finite structures: order-invariant FO is more expressive than plain FO (already the four-variable fragment of FO can express properties that are not FO-definable).

Unfortunately, order-invariant FO is poorly understood. As stated in [BL16], one of the reasons why progress in understanding order-invariance is rather slow is the lack of logical toolkit. The classical model-theoretic methods based on types were proposed only recently [BL16], and order-invariant FO is not even a logic in the classical sense, as its syntax is undecidable. Moreover, the availability of locality-based methods is limited: order-invariant FO is known to be Gaifman-local [GS00, Thm. 2] but the status of its Hanf-locality remains open. This suggests that a good way to understand order-invariant FO is to first look at its fragments, *e.g.* the fragments with a limited number of variables.

Our contribution. We continue the line of research initiated in [ZH16], which aims to study the complexity and the expressive power of order-invariant FO^2 , the two-variable fragment of order-invariant FO. From a complexity point of view, it is known that order-invariant FO^2 has a CONEXPTIME -complete validity problem (which is inherited from FO^2 with a single linear order, see [Ott01, Thm. 1.2]), and that whether a given FO^2 -formula is order-invariant is decidable in CON2EXPTIME [ZH16, Thm. 12]. From an expressive power point of view, order-invariant FO^2 is more expressive than plain FO^2 as it can count globally, see [ZH16, Example 2]. It remains open [ZH16, Sec. 7], however, whether it is true that every order-invariant FO^2 -formula is equivalent to an FO-formula without the linear order predicate. This paper contributes to the field in the three following ways:

- We provide a tight bound for deciding order-invariance for FO^2 ; namely, we show that this problem is CONEXPTIME -complete. Our proof method relies on establishing an exponential-size counter-model property for order-invariance, and is significantly easier than the proof of Zeume and Harwath [ZH16, Thm. 12].
- We present a class \mathcal{C}_{tree} of tree-like structures, inspired by [Pot94], and show that there exists an FO^2 -formula that is *order-invariant over \mathcal{C}_{tree}* (but not over all finite structures!) which is not equivalent to any FO-formula without the linear order predicate. This leads us to believe that the answer to the question of [ZH16, Sec. 7] of whether the expressive power of order-invariant FO^2 lies inside FO is “no”. The problem remains open, though.
- In stark contrast to the previous result, we show that order-invariant FO^2 cannot express properties beyond the scope of FO over classes of structures of bounded degree. We show that this upper bound remains when adding counting to FO^2 .

This work is an extended version of [Bed22] and [Gra23].

2. PRELIMINARIES

We employ standard terminology from finite model theory, assuming that the reader is familiar with the syntax and the semantics of first-order logic (FO) [Lib04, Sec. 2.1], basics on computability and complexity [Lib04, Secs. 2.2–2.3], and order-invariant queries [Lib04, Secs. 5.1–5.2]. By $\text{FO}(\Sigma)$ we denote the first-order logic with equality (written FO when Σ is clear from the context) on a finite signature Σ composed of relation and constant symbols. We say that Σ is *purely relational* if it contains no constant symbol. By FO^2 we denote the fragment of FO in which only two variables x and y are used.

As usual in mathematics, for a given set S we write $|S|$ to denote its cardinality.

Structures. Structures are denoted by calligraphic upper-case letters \mathcal{A}, \mathcal{B} and their domains are denoted by the corresponding Roman letters A, B . We assume that structures have non-empty, *finite* domains. We write $\varphi[R/S]$ to denote the formula obtained from φ by replacing each occurrence of the symbol R with S . We write $\varphi(\bar{x})$ to indicate that all the free variables of φ are in \bar{x} . A sentence is a formula without free variables. By $\mathcal{A}|_{\Delta}$ we denote the substructure of the structure \mathcal{A} restricted to the set $\Delta \subseteq A$. Given a Σ -structure \mathcal{A} and a relational vocabulary $\tau \subseteq \Sigma$, by the τ -*reduct* of \mathcal{A} we mean the unique τ -structure \mathcal{B} such that $A = B$ and for every predicate R from τ we have $R^{\mathcal{A}} = R^{\mathcal{B}}$.

Order-invariance. A sentence $\varphi \in \text{FO}^2(\Sigma \cup \{<\})$, where $<$ is a binary relation symbol not belonging to Σ , is said to be *order-invariant* if for every finite Σ -structure \mathcal{A} , and every pair of strict linear orders $<_0$ and $<_1$ on A , $(\mathcal{A}, <_0) \models \varphi$ if and only if $(\mathcal{A}, <_1) \models \varphi$. It is then convenient to omit the interpretation for the symbol $<$, and to write $\mathcal{A} \models \varphi$ if $(\mathcal{A}, <) \models \varphi$ for any (or, equivalently, every) linear order $<$. Note that φ is *not* order-invariant if there is a structure \mathcal{A} and two linear orders $<_0, <_1$ on A such that $(\mathcal{A}, <_0) \models \varphi$ and $(\mathcal{A}, <_1) \not\models \varphi$. The set of order-invariant sentences using two variables is denoted $<$ -inv FO^2 . While determining whether an FO-sentence is order-invariant is undecidable [Lib04, Ex. 9.3], the situation improves when we allow only two variables: checking order-invariance for FO^2 -formulae was shown to be in CON2EXPTIME in [ZH16, Thm. 12].¹

Decision problems. The *finite satisfiability problem* for a logic \mathcal{L} asks whether an input sentence φ from \mathcal{L} is satisfied in some finite structure. The *finite satisfiability problem* for a logic \mathcal{L} asks whether an input sentence φ from \mathcal{L} is satisfied by all finite structure. Both the finite satisfiability and validity problems for FO are undecidable [Tur38, Tra50], while for FO^2 they are NEXPTIME -complete and CONEXPTIME -complete, see [GGY97, Thm. 5.3] and [Für83, Thm. 3], respectively. Note that φ is finitely valid iff $\neg\varphi$ is finitely unsatisfiable.

Definability and similarity. Let $\mathcal{L}, \mathcal{L}'$ be two logics defined over the same signature, and \mathcal{C} be a class of finite structures on this signature. We say that a property $\mathcal{P} \subseteq \mathcal{C}$ is *definable* (or *expressible*) in \mathcal{L} on \mathcal{C} if there exists an \mathcal{L} -sentence φ such that $\mathcal{P} = \{\mathcal{A} \in \mathcal{C} : \mathcal{A} \models \varphi\}$. When \mathcal{C} is the class of all finite structures, we omit it. We say that $\mathcal{L} \subseteq \mathcal{L}'$ on \mathcal{C} if every property on \mathcal{C} definable in \mathcal{L} is also definable in \mathcal{L}' . Since a sentence which does not mention the linear order predicate is trivially order-invariant, we get the inclusion $\text{FO}^2 \subseteq <$ -inv FO^2 . This inclusion is strict [ZH16, Example 2].

The *quantifier rank* of a formula is the maximal number of quantifiers in a branch of its syntactic tree. Given two Σ -structures \mathcal{A}_0 and \mathcal{A}_1 , and \mathcal{L} being one of FO, FO^2 and $<$ -inv FO^2 , we write $\mathcal{A}_0 \equiv_k^{\mathcal{L}} \mathcal{A}_1$ if \mathcal{A}_0 and \mathcal{A}_1 satisfy the same \mathcal{L} -sentences of quantifier rank at most k . In this case, we say that \mathcal{A}_0 and \mathcal{A}_1 are \mathcal{L} -*similar at depth k* .

We write $\mathcal{A}_0 \simeq \mathcal{A}_1$ if \mathcal{A}_0 and \mathcal{A}_1 are isomorphic.

Atomic types. An (atomic) *1-type* over Σ is a maximal satisfiable set of atoms or negated atoms over Σ with a free variable x . Similarly, an (atomic) *2-type* over Σ is a maximal satisfiable set of atoms or negated atoms with variables x, y . We stress that it is not necessary that all variables are mentioned in an atom. Note that the total number of atomic 1- and 2-types over τ is bounded exponentially in $|\Sigma|$. We often identify a type with the conjunction of all its elements. The sets of 1-types and 2-types over the signature consisting of the

¹The authors of [ZH16] incorrectly stated the complexity in their Thm. 12, mistaking “invariance” with “non-invariance”.

symbols appearing in φ are denoted by α_φ and β_φ , respectively. Given a structure \mathcal{A} and an element $a \in A$ we say that a *realises* a 1-type α if α is the unique 1-type such that $\mathcal{A} \models \alpha[a]$ (a satisfies α in \mathcal{A}). We then write $\text{tp}_{\mathcal{A}}(a)$ to refer to this type. Similarly, for (non-necessarily distinct) $a, b \in A$, we denote by $\text{tp}_{\mathcal{A}}(a, b)$ the unique 2-type *realised* by the pair (a, b) , *i.e.* the 2-type β such that $\mathcal{A} \models \beta[a, b]$ (the pair a, b satisfies β in \mathcal{A}). Finally, given a linearly ordered Σ -structure $(\mathcal{A}, <)$, we split $\text{tp}_{(\mathcal{A}, <)}(a, b)$ into $\text{tp}_{\mathcal{A}}(a, b)$ and $\text{tp}_{<}(a, b)$, where $\text{tp}_{<}(a, b)$ contains (and, over constant-free vocabularies, is fully determined by) one of the following atoms “ $x < y$ ”, “ $x > y$ ” and “ $x = y$ ”.

The two-pebble Ehrenfeucht-Fraïssé game. A common way to prove similarity (with respect to some logic) between two structures is to use Ehrenfeucht-Fraïssé games. Such a game is played between two players: one tries to highlight the differences between the two structures, while the other attempts to show that they are identical. There exists a whole family of those games, each characterizing a specific logic, in the sense that the second player wins exactly when the two structures cannot be distinguished by the corresponding logic. These games differ by the kind of moves allowed, which reflect the particulars of the logic they characterize.

As we are interested in two-variable first-order logic, let us define the two-pebble FO game, which captures exactly the expressive power of FO^2 . This flavor of Ehrenfeucht-Fraïssé game has been introduced by Immerman and Kozen [IK89].

Let us consider two structures \mathcal{A}_0 and \mathcal{A}_1 . The k -round *two-pebble Ehrenfeucht-Fraïssé game on \mathcal{A}_0 and \mathcal{A}_1* is played by two players: the spoiler and the duplicator. The spoiler tries to expose differences between the two structures, while the duplicator tries to establish their indistinguishability.

There are two pebbles associated with each structure: p_0^x and p_0^y on \mathcal{A}_0 , and p_1^x and p_1^y on \mathcal{A}_1 . Formally, these pebbles can be seen as the interpretations in each structure of two new constant symbols, but it will be convenient to see them as moving pieces.

At the start of the game, the duplicator places p_0^x and p_0^y on elements of \mathcal{A}_0 , and p_1^x and p_1^y on elements of \mathcal{A}_1 . The spoiler wins if the duplicator is unable to ensure that $\text{tp}_{\mathcal{A}_0}(p_0^x, p_0^y) = \text{tp}_{\mathcal{A}_1}(p_1^x, p_1^y)$. Otherwise, the proper game starts. Note that in the usual definition of the starting position, the pebbles are not on the board; however, it will be convenient to have them placed in order to uniformize our invariants in Section 5.4. This change is not profound and does not affect the properties of the game.

For each of the k rounds, the spoiler starts by choosing a structure and a pebble in this structure, and places this pebble on an element of the chosen structure. In turn, the duplicator must place the corresponding pebble in the other structure on an element of that structure. The spoiler wins at once if $\text{tp}_{\mathcal{A}_0}(p_0^x, p_0^y) \neq \text{tp}_{\mathcal{A}_1}(p_1^x, p_1^y)$. Otherwise, another round is played. If the spoiler has not won after k rounds, then the duplicator wins.

The main interest of these games is that they capture the expressive power of FO^2 . We will only need the fact that these games are correct:

Theorem 2.1. [IK89, Lemma 12] *If the duplicator has a winning strategy in the k -round two-pebble Ehrenfeucht-Fraïssé game on \mathcal{A}_0 and \mathcal{A}_1 , then $\mathcal{A}_0 \equiv_k^{\text{FO}^2} \mathcal{A}_1$.*

Gaifman graphs and degree. The *Gaifman graph* $\mathcal{G}_{\mathcal{A}}$ of a structure \mathcal{A} is the simple graph with vertices in A and undirected edges between any pair of distinct elements that appear in the same tuple of some relation of \mathcal{A} . By $\text{dist}_{\mathcal{A}}(a, b)$ we denote the distance between a and b

in $\mathcal{G}_{\mathcal{A}}$, defined in the usual way. The distance between an element a and a set B of elements is defined as $\min_{b \in B}(\text{dist}_{\mathcal{A}}(a, b))$. For $B \subseteq A$, we let $N_{\mathcal{A}}(B)$ denote the set of elements at distance exactly 1 from B in $\mathcal{G}_{\mathcal{A}}$. In particular, $B \cap N_{\mathcal{A}}(B) = \emptyset$. The *degree* of \mathcal{A} is the maximal degree of its Gaifman graph. The class \mathcal{C} of Σ -structures is said to have *bounded degree* if there exists some $d \in \mathbb{N}$ such that the degree of every $\mathcal{A} \in \mathcal{C}$ is at most d .

3. COMPLEXITY OF THE INVARIANCE PROBLEM

We study the complexity of the problem of deciding if an input sentence φ of FO^2 is order-invariant and establish its CONEXPTIME -completeness. We start from a lower bound, inspired by slides of Schweikardt [Sch13, Slide 9]. The following lemma is crucial:

Lemma 3.1. *Fix an FO^2 sentence φ , and let $\psi_{<} := \exists x (P(x) \wedge \forall y (y \leq x))$ contain a fresh unary predicate P . Then $(\neg\varphi) \rightarrow \psi_{<}$ is $<$ -invariant if and only if φ is satisfied by all finite structures with at least two elements.*

Proof. Suppose that φ is satisfied by every structure with at least two elements. To prove that $(\neg\varphi) \rightarrow \psi_{<}$ is $<$ -invariant, we establish that $(\neg\varphi) \rightarrow \psi_{<}$ is equivalent to \top over all structures with at least two elements. Indeed, over such structures $\neg\varphi$ is equivalent to \perp (and hence $(\neg\varphi) \rightarrow \psi_{<}$ is logically equivalent to \top). Thus, $(\neg\varphi) \rightarrow \psi_{<}$ is trivially $<$ -invariant. For the other direction, suppose that there exists an \mathcal{A} with at least two elements (say, a_1, a_2, \dots, a_n) that satisfies $\neg\varphi$. Let \mathcal{B} be the finite structure with $P^{\mathcal{B}} = \{a_n\}$ whose reduct is \mathcal{A} . Clearly $\mathcal{B} \models \neg\varphi$, since the predicate P is fresh. Finally, let $<_1, <_2$ be two linear orders on B : the first one orders the elements a_i of B according to their subscript numbering, while the second ordering is like the first one but swaps the positions of a_1 and a_n . Now note that $(\mathcal{B}, <_1) \models (\neg\varphi) \rightarrow \psi_{<}$ but $(\mathcal{B}, <_2) \not\models (\neg\varphi) \rightarrow \psi_{<}$ (the formula $\psi_{<}$ basically states that the maximal element of $<_i^{\mathcal{B}}$ satisfies P). Hence, the formula $(\neg\varphi) \rightarrow \psi_{<}$ is not $<$ -invariant. \square

The above lemma provides a reduction from finite FO^2 -validity (over structures of size at least two) to checking order-invariance of FO^2 -sentences. By CONEXPTIME -hardness² of finite validity problem for FO^2 [Für83, Thm. 3], we conclude:

Theorem 3.2. *Checking whether an FO^2 formula is order-invariant is CONEXPTIME -hard.*

Our upper bound uses the following fact, immediate from the definition of order-invariance.

Fact 3.3. An FO^2 sentence φ is *not* order-invariant iff the sentence $\varphi[</<_0] \wedge \neg\varphi[</<_1]$ is finitely satisfiable over structures interpreting $<_0$ and $<_1$ as linear orders over the domain.

Let $\text{FO}^2_{<}[<_0, <_1]$ be composed of sentences of the shape $\varphi[</<_0] \wedge \neg\varphi[</<_1]$ for all sentences $\varphi \in \text{FO}^2(\Sigma \cup \{<\})$ over all signatures Σ . To simplify the reasoning about such formulae, we rewrite them into the so-called *Scott Normal Form*.

Definition 3.4. We say that an $\text{FO}^2_{<}[<_0, <_1]$ -sentence is in *Scott's normal form* if

- its signature is constant-free and employs only predicates of arity 1 or 2, and

²The finite validity problem is equally hard to the validity problem over structures of size at least 2. Let φ be an FO^2 -sentence, P be a fresh unary predicate, and let φ^P be the P -relativisation of φ , i.e. the formula obtained from φ by replacing all its subformulae of the form $\exists x.\psi$ with $\exists x(P(x) \wedge \psi(x))$ and $\forall x.\psi$ with $\forall x(P(x) \rightarrow \psi(x))$ (and analogously for the variable y). One can verify that φ is valid if and only if $(\exists x.P(x) \wedge \exists x.\neg P(x)) \rightarrow \varphi^P$ is satisfied in every structure of size at least 2.

- it has the form

$$\bigwedge_{i=0}^1 \left(\forall x \forall y \chi_i(x, y) \wedge \bigwedge_{j=1}^{m_i} \forall x \exists y \gamma_i^j(x, y) \right),$$

where the decorated χ and γ are quantifier-free and for each $i \in \{0, 1\}$ the symbol $<_i$ does not appear in χ_{1-i} .

The following lemma justifies the use of formulae in Scott's normal form.

Lemma 3.5. *Given an $FO^2_{<_0, <_1}$ -sentence φ one can compute in time polynomial in $|\varphi|$ an $FO^2_{<_0, <_1}$ -sentence φ_{scott} in Scott's normal form (which is of size polynomial in $|\varphi|$) such that φ is finitely satisfiable over structures interpreting $<_0$ and $<_1$ as linear orders over the domain if and only if φ_{scott} is.*

Proof sketch. As the translation we are going to provide is folklore and appears almost in every paper regarding FO^2 , we only sketch its proof here. As usual in the context of two-variable logic, we can confine our attention only to vocabularies composed solely of predicates of arity one and two. A pedantic proof of this fact can be found in the recent textbook of Ian Pratt-Hartmann [PH23, Sec. 3, Lemma 3.4].³ The remainder of the proof goes via a routine renaming by replacing formulae with a single free variable with a fresh unary predicate and sharpening its interpretation. To do so, we one-by-one take the maximally nested subformula of φ starting from a single quantifier, say $\exists x \lambda(x, y)$. We then replace the formula $\exists x \lambda(x, y)$ with $A(y)$ for a fresh unary predicate A , and append the conjunct $\forall y (\exists x \lambda(x, y) \leftrightarrow A(y))$ to φ . Note that λ does not simultaneously contain $<_0$ and $<_1$. Finally, we split the equivalence into two implications, yielding (by shifting atoms) the formulae of the form $\forall \forall \theta$ and $\forall \exists \xi$. The resulting formula is clearly equisatisfiable to the original one. By repeating this procedure and regrouping and reordering the obtained formulae (and adding vacuous quantifiers if needed), we obtain the desired equivalent formula in normal form. \square

For brevity we speak about $\forall \forall$ and $\forall \exists$ -conjuncts of φ in the normal form given above, meaning the appropriate subformulae of φ starting from the mentioned prefix of quantifiers. Given a model $\mathcal{A} \models \varphi$ of a formula φ in Scott's normal form given above and elements $a, b \in A$ such that $\mathcal{A} \models \gamma_i^j(a, b)$, we call b a γ_i^j -witness for a (or simply a witness). The core of our upper bound proof is the following small model theorem, employing the circular witnessing scheme by Grädel, Kolaitis, and Vardi [GGY97, Thm. 4.3].

Lemma 3.6. *If an $FO^2_{<_0, <_1}$ sentence φ (in normal form of Lemma 3.5) has a finite model (that interprets $<_0, <_1$ as linear orders over the domain) then there exists such a finite model with $\mathcal{O}(|\varphi|^3 \cdot 2^{|\varphi|})$ elements.*

Proof. Let $M := m_0 + m_1$ be the total number of $\forall \exists$ -conjuncts of φ , and let α_φ be the set of all 1-types over the set of unary symbols of φ . Recall that M is polynomial in $|\varphi|$ and $|\alpha_\varphi|$ is exponential in $|\varphi|$. Suppose that \mathcal{A} is a finite model of φ that interprets $<_0, <_1$ as linear orders over the domain, and let us explain our proof plan.

Intuition and proof plan. In what follows, we employ a variant of the circular witnessing scheme by Grädel, Kolaitis, and Vardi [GGY97, Thm. 4.3] in order to distinguish three disjoint sets of domain elements of \mathcal{A} , namely W_0, W_1, W_2 whose union $W := W_0 \cup W_1 \cup W_2$ will serve as a “base” of a new “small” model of φ . In order to satisfy φ , the universal ($\forall \forall$) and the

³Pratt-Hartmann's proof leaves nullary symbols in the vocabulary, but one can easily get rid of them. More precisely, given a nullary symbol p and a formula φ , we replace φ with $(\forall x P(x) \vee \forall x \neg P(x)) \wedge \varphi[p/\forall x P(x)]$.

existential ($\forall\exists$) conjuncts of φ must be fulfilled. In our construction, we first distinguish a set W_0 of elements of \mathcal{A} that contains elements of \mathcal{A} with 1-types that do not appear “frequently” in \mathcal{A} together with the first and last M realisations w.r.t. $<_0$ and w.r.t. $<_1$ of each other (*i.e.* frequent) 1-type needed later in the construction. Unfortunately, \mathcal{A} restricted to W_0 may not be a model of φ . Indeed, while such a structure satisfies the universal part of φ , some elements from W_0 may violate $\forall\exists$ -conjuncts of φ . To “repair” the resulting structure, we construct the set W_1 by taking the minimal number of elements in \mathcal{A} so that all elements from W_0 have all the required γ_i^j -witnesses in \mathcal{A} restricted to $W_0 \cup W_1$. Note that W_1 has no more than $M \cdot |W_0|$ elements. The situation is a bit better now: the \mathcal{A} restricted to $W_0 \cup W_1$ again satisfies the $\forall\forall$ -subformula of φ (as the universal formulae are preserved under taking substructures), all elements from W_0 have their required witnesses for subformulae γ_i^j , but there may be still some elements W_1 may violate $\forall\exists$ -conjuncts of φ . We “repair” the resulting structure one final time by constructing the set W_2 analogously to the case of W_1 , so that all the required γ_i^j -witnesses for the elements from W_1 in \mathcal{A} are in the set $W_0 \cup W_1 \cup W_2$. Unfortunately, for the same reason as before, some elements from W_2 may violate some $\forall\exists$ -subformulae of φ . To fix the problem, for each of such elements a , we suitably select M pairwise-different elements b_1, b_2, \dots, b_M from W_0 and alter binary relations between a and all of b_i in order to provide suitable witnesses for a . By our choice of W_0 such *different* elements will always exist. The resulting 2-type between a and other b_i will be one of the 2-types already present in \mathcal{A} , which ensures that we do not accidentally violate the $\forall\forall$ -subformulae of φ .⁴ Finally, by construction, all the selected elements have the required witnesses in $W_0 \cup W_1 \cup W_2$, and hence the $\forall\exists$ -subformulae of φ are also satisfied in the altered structure \mathcal{A} restricted to $W_0 \cup W_1 \cup W_2$. This yields a desired “small” model of the formula φ and concludes the proof.

The construction of W_0 . Call a 1-type $\alpha \in \alpha_\varphi$ *rare* if it is realised by at most $4M$ elements in \mathcal{A} (the factor 4 comes from the fact that we consider maximal and minimal elements w.r.t. $<_0$ and $<_1$ and M is the number of $\forall\exists$ -conjuncts of φ). For every 1-type $\alpha \in \alpha_\varphi$ that is not rare, we select M minimal and M maximal elements w.r.t. each of the orders $<_0$ and $<_1$ that realise α . Let W_0 be the set of the selected elements from \mathcal{A} together with all elements of \mathcal{A} realising rare 1-types.

The construction of W_1 and W_2 . We next close W_0 twice under taking witnesses. More precisely, let W_1 be any \subseteq -minimal subset of \mathcal{A} so that all elements from W_0 have all the required γ_i^j -witnesses in $W_0 \cup W_1$. Similarly, we define W_2 to be any \subseteq -minimal subset of \mathcal{A} so that all elements from $W_0 \cup W_1$ have all the required γ_i^j -witnesses in $W_0 \cup W_1 \cup W_2$. As they are only M different γ_i^j -formulae in φ , we have that:

$$|W_1| \leq M|W_0| \leq 4M^2|\alpha_\varphi| \text{ and } |W_2| \leq M|W_0 \cup W_1| \leq M \cdot (4M + 4M^2)|\alpha_\varphi| \leq 8M^3|\alpha_\varphi|.$$

The construction of \mathcal{B} . Consider the structure $\mathcal{B} := \mathcal{A} \upharpoonright_{W_0 \cup W_1 \cup W_2}$. We see that:

$$|\mathcal{B}| \leq |W_0| + |W_1| + |W_2| \leq (4M + 4M^2 + 8M^3) \cdot |\alpha_\varphi| \leq 16M^3 \cdot |\alpha_\varphi| \leq 16 |\varphi|^3 \cdot 2^{|\varphi|}.$$

Hence, the size of \mathcal{B} is as desired. Note that universal formulae are preserved under substructures, thus $<_0^{\mathcal{B}}, <_1^{\mathcal{B}}$ (defined as $<_0^{\mathcal{A}} \cap \mathcal{B}$ and $<_1^{\mathcal{A}} \cap \mathcal{B}$) are linear orders over \mathcal{B} and \mathcal{B} satisfies the $\forall\forall$ -conjuncts of φ . Hence, by the construction of W_1 and W_2 , the only reason for \mathcal{B} to not be a model of φ is the lack of required γ_i^j -witnesses for elements from the set W_2 . We fix this issue by reinterpreting binary relational symbols between the sets W_2 and W_0 .

⁴It would be unwise to apply the same approach to W_1 , as such elements are already constrained by the construction of W_0 (while W_2 are not!).

Reinterpreting binary relations in \mathcal{B} . Let us fix an element a from W_2 that violate some of the $\forall\exists$ -conjuncts of φ in \mathcal{B} , as well as a $\forall\exists$ -conjunct $\psi := \forall x\exists y \gamma_i^j(x, y)$ whose satisfaction is violated by a . We stress that there are at most M such conjuncts, and γ_i^j does not use the $<_{1-i}$ predicate. Since $\mathcal{A} \models \varphi$ we know that there is an element $b \in A$ (different from a) such that b is a γ_i^j -witness for a and γ_i^j in \mathcal{A} . Let α be the 1-type of b in \mathcal{A} . Note that α is not rare. Indeed, otherwise $b \in W_0$, and hence $b \in B$. Moreover either $b <_i^A a$ or $a <_i^A b$. Suppose that $b <_i^A a$ (the other case is analogous), and pick an element a' from W_0 among M -minimal (w.r.t. $<_i$) realisations of the 1-type α . In case a' was already selected to provide a missing witness for a , we take yet another one. Note that such an element exists due to the fact that there are only M $\forall\exists$ -conjuncts in φ and we selected M minimal and maximal realisations of every non-rare 1-type for each of the orders $<_0$ and $<_1$. We now alter the binary relations between a and a' in \mathcal{B} so that the equality $\text{tp}_{\mathcal{A}}(a, b) = \text{tp}_{\mathcal{B}}(a, b)$ holds (which can be done as a' and b have equal 1-type). We repeat the process to provide all missing witnesses for a . Observe that all elements from B that had γ_i^j -witnesses before our redefinition of certain 2-types, still do have them (as we did not touch 2-types between them and their witnesses), \mathcal{B} still satisfies the $\forall\forall$ -component of φ (since the modified 2-type does not violate φ in \mathcal{A} it does not violate φ in \mathcal{B}) and a has all required witnesses. By repeating the strategy for all the other elements from W_2 violating φ , we obtain the desired “small” model of φ . \square

Lemma 3.6 yields an NEXPTIME algorithm for deciding satisfiability of $\text{FO}_-^2[<_0, <_1]$ formulae: convert an input into Scott’s normal form, guess its exponential size model and verify the modelhood with a standard model-checking algorithm (in PTIME [GO99, Prop. 4.1]). After applying Fact 3.3 and Theorem 3.2 we conclude:

Theorem 3.7. *Checking if an FO^2 -formula is order-invariant is CONEXPTIME-complete.*

4. CAN ORDER-INVARIANT FO^2 EXPRESS PROPERTIES BEYOND THE SCOPE OF FO ?

While we do not solve the question stated in the heading of this section, we provide a partial solution. Let \mathcal{C} be some class of finite structures. A sentence $\varphi \in \text{FO}^2(\Sigma \cup \{<\})$, where $<$ is a binary relation symbol not belonging to Σ , is said to be *order-invariant over \mathcal{C}* if for every finite Σ -structure \mathcal{A} in \mathcal{C} , and every pair of strict linear orders $<_0$ and $<_1$ on A , $(\mathcal{A}, <_0) \models \varphi$ iff $(\mathcal{A}, <_1) \models \varphi$. Note that this is a weakening of the classical condition of order-invariance, and that the usual definition is recovered when \mathcal{C} is the class of all finite structures.

In what follows, we present a class \mathcal{C} over the vocabulary $\Sigma := \{T, D, S\}$ (for binary predicates T , D and S) of tree-like finite structures, and a sentence $\varphi \in \text{FO}^2(\Sigma \cup \{<\})$ “expressing even depth” that is order-invariant over \mathcal{C} but not equivalent to any first-order sentence over Σ . To do so, we employ *dendroids*, namely finite Σ -structures \mathcal{A} that, intuitively, are complete directed binary trees decorated with a binary parent-child relation $T^{\mathcal{A}}$, a descendant relation $D^{\mathcal{A}}$, and a sibling relation $S^{\mathcal{A}}$. Their formal definition comes next.

Definition 4.1. A Σ -structure \mathcal{A} is a *dendroid* if there is a positive integer n such that

- $A = \{0, 1\}^{\leq n}$ (i.e. the set of all binary words of length at most n),
- $T^{\mathcal{A}} = \{(w, w0), (w, w1) \mid w \in A, |w| < n\}$,
- $D^{\mathcal{A}} = (T^{\mathcal{A}})^+$ (i.e. $D^{\mathcal{A}}$ is the transitive closure of $T^{\mathcal{A}}$), and
- $S^{\mathcal{A}} = \{(w0, w1), (w1, w0) \mid w \in A, |w| < n\}$.

We call the number n the *depth* of \mathcal{A} , and call the length of a node $v \in \mathcal{A}$ the *level* of v . We also use the terms “root” and “leaf” in the usual way.

The following is a variant of a classical exercise stating that for every threshold q we have that trees that are exponentially deep w.r.t. q are FO_q -equivalent. Indeed:

Lemma 4.2. *For all $q \in \mathbb{N}$, if \mathcal{A}, \mathcal{B} are dendroids of depth $\geq 2^{q+1}$ then $\mathcal{A} \equiv_q^{\text{FO}} \mathcal{B}$.*

Proof. This is a tedious generalisation of the winning strategy for the duplicator in the q -round Ehrenfeucht-Fraïssé games on linear orders [Lib04, Thm 3.6 Proof #1]. \square

As an immediate corollary we get:

Corollary 4.3. *There is no $\text{FO}(\Sigma)$ -formula φ_{even} such that for every $\mathcal{A} \in \mathcal{C}$ we have $\mathcal{A} \models \varphi_{\text{even}}$ iff the depth of \mathcal{A} is even.*

In contrast to Corollary 4.3, we show that the even depth query can be defined with an $\text{FO}^2(\Sigma \cup \{<\})$ -formula which is order-invariant over \mathcal{C} (but unfortunately not over the class of all finite structures). Henceforth we considered *ordered dendroids*, i.e. dendroids that are additionally linearly-ordered by $<$. Given such an ordered dendroid \mathcal{T} , and an element c with children a, b we say that a is the *left child* of c iff $a <^{\mathcal{T}} b$ holds. Otherwise we call a the *right child* of c . A *zig-zag* in \mathcal{T} is a sequence of elements a_0, a_1, \dots, a_n , where a_n is a leaf of \mathcal{T} , a_0 is the root of \mathcal{T} , a_{2i+1} is the right child of a_{2i} for any $i \geq 0$ and a_{2i} is the left child of a_{2i-1} for any $\frac{n}{2} \geq i \geq 1$. A zig-zag is *even* if its last element is the left child of its parent, and *odd* otherwise. The underlying trees in dendroids are complete and binary, thus:

Observation 4.4. An ordered dendroid \mathcal{T} has an even zig-zag iff \mathcal{T} is of even depth. Moreover, if \mathcal{T} is a dendroid of even and odd depth then for any linear order $<$ over its domain the ordered dendroid $(\mathcal{T}, <)$ has an even and odd zig-zag, respectively.

Proof. Immediate by induction after observing that $\mathcal{A} \upharpoonright_{\{0,1\}^{\leq n}}$, for any positive integer n smaller than the depth of \mathcal{A} , is also a dendroid. \square

The above lemma suggests that a good way to express the evenness of the depth of a dendroid is to state the existence of an even zig-zag; this is precisely the property that we are going to describe with an $\text{FO}^2(\Sigma \cup \{<\})$ -formula. Let us first introduce a few useful macros:

$$\begin{aligned} \text{ROOT}(x) &:= \neg \exists y T(y, x) & \text{LEAF}(x) &:= \neg \exists y T(x, y) & \text{2nd}(x) &:= \exists y T(y, x) \wedge \text{ROOT}(y) \\ \text{LS}(x) &:= \exists y S(x, y) \wedge x < y & \text{RS}(x) &:= \exists y S(x, y) \wedge y < x \end{aligned}$$

The first two macros have an obvious meaning. The third macro identifies a child of the root, while the last two macros identify the left and the right siblings (according to the linear order $<$), respectively. Our desired formula $\varphi_{\text{even-zig-zag}}$ is then

$$\begin{aligned} &\exists x \left([\text{LEAF}(x) \wedge \text{LS}(x)] \wedge [\forall y (2\text{nd}(y) \wedge D(y, x)) \rightarrow \text{RS}(y)] \right. \\ &\quad \wedge [\forall y (\neg \text{ROOT}(y) \wedge \neg 2\text{nd}(y) \wedge D(y, x) \wedge \text{RS}(y)) \rightarrow \exists x T(x, y) \wedge \text{LS}(x)] \\ &\quad \left. \wedge [\forall y (\neg \text{ROOT}(y) \wedge \neg 2\text{nd}(y) \wedge D(y, x) \wedge \text{LS}(y)) \rightarrow \exists x T(x, y) \wedge \text{RS}(x)] \right). \end{aligned}$$

Note that the above formula, by fixing a leaf, fixes the whole path from such a leaf to the root (since root-to-leaf paths in trees are unique). To say that such a path is an even zig-zag, we need a base of induction (the first line) stating that the selected leaf is a left child and the root’s child lying on this path is its right one, as well as an inductive step stating that every left and right child on the path has a parent which is itself a right and left child, respectively, with the obvious exception of the root and its child. From there, it is easily shown that:

Proposition 4.5. *An ordered dendroid \mathcal{T} satisfies $\varphi_{\text{even-zig-zag}}$ iff it has even depth.*

Proof. To prove the right-to-left implication, we use Observation 4.4 to infer the existence of an even zig-zag a_0, a_1, \dots, a_{2n} in \mathcal{T} . Taking a_{2n} as a witness for the existential quantifier in front of $\varphi_{\text{even-zig-zag}}$ and going back to the definition of an even zig-zag, we get $\mathcal{T} \models \varphi_{\text{even-zig-zag}}$. For the other direction, consider a leaf a satisfying the properties enforced in $\varphi_{\text{even-zig-zag}}$. There is a unique path $\rho = a_0, a_1, \dots, a_n = a$ from the root of \mathcal{T} to a . The first line of $\varphi_{\text{even-zig-zag}}$ guarantees that a_n is a left child and a_1 is a right child. We then show by induction, relying on the last two lines of $\varphi_{\text{even-zig-zag}}$, that for any $i \geq 0$, a_{2i+1} is the right child of a_{2i} , and for $i \geq 1$, a_{2i} is the left child of a_{2i-1} . Thus ρ is an even zig-zag. By invoking Observation 4.4 again, we get that \mathcal{T} has even depth. \square

As a direct consequence of the previous statement, observe that our formula $\varphi_{\text{even-zig-zag}}$ is order-invariant over \mathcal{C} : whether an ordered dendroid has even depth only depends on the underlying dendroid, and not on the particulars of its linear order. Recalling Corollary 4.3, we conclude the following:

Theorem 4.6. *There exists a class of finite structures \mathcal{C} and an $\text{FO}^2(\Sigma \cup \{<\})$ -sentence which is order-invariant over \mathcal{C} , but is not equivalent to any $\text{FO}(\Sigma)$ sentence.*

5. EXPRESSIVE POWER WHEN THE DEGREE IS BOUNDED

We have seen in the previous section that if we relax the order-invariant constraint (namely, by requiring invariance only on a restricted class of structures), then one is able to define, with two variables, properties that lie beyond the expressive power of FO. We conjecture that this is still the case when requiring invariance over the class of all finite structures.

In this section, we go the other way, and show that when one considers only classes of bounded degree, then $<$ -inv FO^2 can only express FO-definable properties. Note that although the class \mathcal{C} from Section 4 contains tree-like structures, the descendant relation makes this a dense class of structures (as its Gaifman graph contains cliques of arbitrarily large size), and in particular \mathcal{C} does not have a bounded degree.

5.1. Overview of the result. We give an upper bound to the expressive power of order-invariant FO^2 when the degree is bounded:

Theorem 5.1. *Let \mathcal{C} be a class of structures of bounded degree on a purely relational vocabulary. Then $<$ -inv $\text{FO}^2 \subseteq \text{FO}$ on \mathcal{C} .*

For the remainder of this section, we fix a purely relational vocabulary Σ , an integer d and a class \mathcal{C} of Σ -structures of degree at most d . The constraint on Σ is merely to ease the presentation of the proof. We discuss in Paragraph 5.6 how to extend Theorem 5.1 to vocabulary containing constants.

Let us now detail the outline of the proof. The technical parts of the proof will be the focus of Sections 5.3 and 5.4. Our general strategy is to show the existence of a function $f: \mathbb{N} \rightarrow \mathbb{N}$ such that every formula $\varphi \in <$ -inv FO^2 of quantifier rank k is equivalent on \mathcal{C} (i.e. satisfied by the same structures of \mathcal{C}) to an FO-formula ψ of quantifier rank at most $f(k)$.

To prove the existence of such an equivalent FO-formula ψ to any order-invariant FO^2 -formula φ on \mathcal{C} , we will pick $f(k)$ large enough so that for any two structures $\mathcal{A}_0, \mathcal{A}_1 \in \mathcal{C}$ such that $\mathcal{A}_0 \equiv_{f(k)}^{\text{FO}} \mathcal{A}_1$, we have $\mathcal{A}_0 \equiv_k^{<$ -inv $\text{FO}^2} \mathcal{A}_1$. In other words, we show that the equivalence

relation $\equiv_{f(k)}^{\text{FO}}$ is more fine-grained than the equivalence relation $\equiv_k^{<\text{-inv FO}^2}$. Let us show that this indeed guarantees the existence of ψ . By definition of $\equiv_k^{<\text{-inv FO}^2}$, the class of structures satisfying a formula φ of $<\text{-inv FO}^2$ of quantifier rank k is a union of equivalence classes for the equivalence relation $\equiv_k^{<\text{-inv FO}^2}$, whose intersection with \mathcal{C} is consequently the intersection of \mathcal{C} with a union of equivalence classes c_1, \dots, c_l for $\equiv_{f(k)}^{\text{FO}}$. It is folklore (see, *e.g.*, [Lib04, Cor. 3.16]) that there are only a finite number of equivalence classes for $\equiv_{f(k)}^{\text{FO}}$, each of which being definable by an FO-sentence of quantifier rank $f(k)$. Then the disjunction ψ of the sentences defining c_1, \dots, c_l (which is again an FO-sentence of quantifier rank $f(k)$) is indeed equivalent to φ on \mathcal{C} , thus showing that the addition of a built-in order does not allow to express more first-order properties on \mathcal{C} , although it may allow more concise descriptions of these properties.

In order to show that $\mathcal{A}_0 \equiv_k^{<\text{-inv FO}^2} \mathcal{A}_1$ for any structures \mathcal{A}_0 and \mathcal{A}_1 of \mathcal{C} such that $\mathcal{A}_0 \equiv_{f(k)}^{\text{FO}} \mathcal{A}_1$, we will construct in Section 5.3 two particular different orders on \mathcal{A}_0 and \mathcal{A}_1 (denoted $<$ on both structures, whose domain can assumed to be disjoint), and we will prove in Section 5.4 that

$$(\mathcal{A}_0, <) \equiv_k^{\text{FO}^2} (\mathcal{A}_1, <). \quad (5.1)$$

This concludes the proof, since any sentence of $<\text{-inv FO}^2$ with quantifier rank at most k holds in \mathcal{A}_0 iff it holds in $(\mathcal{A}_0, <)$ (by definition of order-invariance), iff it holds in $(\mathcal{A}_1, <)$ (by Equation (5.1)), iff it holds in \mathcal{A}_1 .

In the process of constructing the linear orders on \mathcal{A}_0 and \mathcal{A}_1 , it will be convenient to sort the elements of these structures based on the isomorphism type of their *neighbourhood* (for $k \in \mathbb{N}$, the k -neighbourhood of some element a is the substructure induced by the set of elements that are at distance at most k from a). This notion is central in finite model theory, and is the basic tool of the locality methods to prove inexpressibility results. For an introduction on the subject, see, *e.g.*, [Lib04, Chapter 4].

Section 5.2 is devoted to the notion of neighbourhood: it is defined in Section 5.2.1, where we recall that, when the degree is bounded, first-order logic is able to count (up to some threshold) the number of occurrences of each isomorphism type of neighbourhood, and nothing more. In Section 5.2.2, we establish a dichotomy between neighbourhood types that appear frequently in a structure, and those that have only a few occurrences. This distinction will highlight which elements play a particular role in the structures \mathcal{A}_0 and \mathcal{A}_1 , and should thus be treated with special care when constructing the linear orders. By contrast, we can be much more liberal with the way we treat elements with a frequent neighbourhood type; in particular, we will use them as padding between critical parts of the orders.

Lastly, we show in Section 5.5 that when the degree is bounded, the inclusion of $<\text{-inv FO}^2$ in FO still holds when one adds counting quantifiers to FO^2 .

5.2. Neighbourhoods.

5.2.1. Definition and properties of neighbourhoods. Let us start by defining the notion of neighbourhood of an element in a structure.

Let c be a constant symbol, and let $\mathcal{A} \in \mathcal{C}$. For $k \in \mathbb{N}$ and $a \in A$, the (pointed) k -neighbourhood $\mathcal{N}_{\mathcal{A}}^k(a)$ of a in \mathcal{A} is the $(\Sigma \cup \{c\})$ -structure whose restriction to the vocabulary Σ is the substructure of \mathcal{A} induced by the set $N_{\mathcal{A}}^k(a) := \{b \in A : \text{dist}_{\mathcal{A}}(a, b) \leq k\}$, and where c is interpreted as a . In other words, it consists of all the elements at distance at most k

from a in \mathcal{A} , together with the relations they share in \mathcal{A} ; the center a being marked by the constant c . We sometimes refer to $N_{\mathcal{A}}^k(a)$ as the k -neighbourhood of a in \mathcal{A} as well, but the context will always make it clear whether we refer to the whole substructure or only its domain. The k -neighbourhood type $\tau = \text{neigh-tp}_{\mathcal{A}}^k(a)$ of a in \mathcal{A} is the isomorphism class of its k -neighbourhood. We say that τ is a k -neighbourhood type over Σ , and that a is an occurrence of τ . We denote by $|\mathcal{A}|_{\tau}$ the number of occurrences of τ in \mathcal{A} (i.e. the number of elements with k -neighbourhood τ in \mathcal{A}), and we write $\llbracket \mathcal{A}_0 \rrbracket_k =^t \llbracket \mathcal{A}_1 \rrbracket_k$ to mean that for every k -neighbourhood type τ , $|\mathcal{A}_0|_{\tau}$ and $|\mathcal{A}_1|_{\tau}$ are either equal, or both larger than t .

Let NEIGHTYPE_k^d denote the set of k -neighbourhood types over Σ occurring in structures of degree at most d . Note that NEIGHTYPE_k^d is a finite set.

The interest of this notion resides in the fact that when the degree is bounded, FO is exactly able to count the number of occurrences of neighbourhood types up to some threshold [FJY95]. We will only use one direction of this characterization, namely:

Proposition 5.2. *For all integers k and t , there exists some $\hat{f}(k, t) \in \mathbb{N}$ (which also depends on the bound d on the degree of structures in \mathcal{C}) such that for all structures $\mathcal{A}_0, \mathcal{A}_1 \in \mathcal{C}$, $\mathcal{A}_0 \equiv_{\hat{f}(k, t)}^{\text{FO}} \mathcal{A}_1$ implies $\llbracket \mathcal{A}_0 \rrbracket_k =^t \llbracket \mathcal{A}_1 \rrbracket_k$.*

Remark 5.3. Assume the existence of a function $\Upsilon: \mathbb{N} \rightarrow \mathbb{N}$ such that whenever $\mathcal{A}_0, \mathcal{A}_1 \in \mathcal{C}$ are such that $\llbracket \mathcal{A}_0 \rrbracket_k =^{\Upsilon(k)} \llbracket \mathcal{A}_1 \rrbracket_k$, one can construct $<$ satisfying Equation (5.1).

Summing up the above discussion, we show that the existence of such a Υ is enough to end the proof of Theorem 5.1. Indeed, Proposition 5.2 ensures that for $f: k \mapsto \hat{f}(k, \Upsilon(k))$, if $\mathcal{A}_0 \equiv_{f(k)}^{\text{FO}} \mathcal{A}_1$ then there exists $<$ such that $(\mathcal{A}_0, <) \equiv_k^{\text{FO}^2} (\mathcal{A}_1, <)$, entailing $\mathcal{A}_0 \equiv_k^{<-inv \text{FO}^2} \mathcal{A}_1$.

In other words, this means the equivalence relation $\equiv_{f(k)}^{\text{FO}}$ is more fine-grained than $\equiv_k^{<-inv \text{FO}^2}$ on \mathcal{C} . Given an $<-inv \text{FO}^2$ -sentence of quantifier rank k , one can decompose it along the equivalence classes of $\equiv_k^{<-inv \text{FO}^2}$, and thus, in turn, decompose it according to $\equiv_{f(k)}^{\text{FO}}$ (i.e. as a disjunction of FO-sentences of rank $f(k)$ defining the relevant $\equiv_{f(k)}^{\text{FO}}$ classes). This gives a translation on \mathcal{C} from $<-inv \text{FO}^2$ -sentences of quantifier rank k to FO-sentences of rank $f(k)$.

Let us now explain how the function Υ is chosen.

5.2.2. Frequency of a neighbourhood type. Let us write $\#\text{neigh}_k^d := |\text{NEIGHTYPE}_k^d|$, the number of types of k -neighbourhoods possibly appearing in structures of \mathcal{C} (which have degree at most d).

In this section, we show that if we consider the set $\text{FREQ}[\mathcal{A}]_k$ of k -neighbourhood types that have enough occurrences in some $\mathcal{A} \in \mathcal{C}$ (where “enough” will be given a precise meaning later on), each type in $\text{FREQ}[\mathcal{A}]_k$ must have many occurrences that are scattered across \mathcal{A} . Not only that, but we can also make sure that such occurrences are far from all the occurrences of every k -neighbourhood type not in $\text{FREQ}[\mathcal{A}]_k$, which by definition have few occurrences in \mathcal{A} .

Such a dichotomy was introduced in [Gra21]; we simply accommodate this construction to our needs. In the remainder of this section, we recall and adapt this construction.

The following lemma, which heavily relies on the degree boundedness assumption, states that, when given a small set B and a bounded number of large enough sets C_1, \dots, C_n , it is always possible to find many elements in each C_j which are far apart from one another, as well as far from B . It is a strengthening of [NO10, Lemma 3.2]).

Lemma 5.4. *Given three positive integers m, δ, s , there exists a threshold $g(m, \delta, s) \in \mathbb{N}$ such that for all $\mathcal{A} \in \mathcal{C}$, all $B \subseteq A$ of size at most s , and all subsets $C_1, \dots, C_n \subseteq A$ (with $n \leq \#\text{neigh}_k^d$) of size at least $g(m, \delta, s)$, it is possible to find elements $c_j^1, \dots, c_j^m \in C_j$ for all $j \in \{1, \dots, n\}$, such that for all $j, j' \in \{1, \dots, n\}$ and $i, i' \in \{1, \dots, m\}$, $\text{dist}_{\mathcal{A}}(c_j^i, B) > \delta$ and $\text{dist}_{\mathcal{A}}(c_j^i, c_{j'}^{i'}) > \delta$ if $(j, i) \neq (j', i')$.*

Furthermore, such a function g can be chosen such that:

- for all integers m, δ, s , $g(m, \delta, s) \geq s$, and
- g is non-decreasing in its third argument, i.e. for every integers m, δ, s, s' such that $s \leq s'$, $g(m, \delta, s) \leq g(m, \delta, s')$.

Proof. Let us fix $m, \delta, s \in \mathbb{N}$ and $\mathcal{A} \in \mathcal{C}$, which has degree at most d by assumption.

Let us start by proving that for any $p \in \mathbb{N}$ and sets of elements $X, Y \subseteq A$ such that $|X| \leq p$ and $|Y| \geq (p + m)(d^\delta + 1)$, there exist elements $y^1, \dots, y^m \in Y$ such that for all $i, i' \in \{1, \dots, m\}$, $\text{dist}_{\mathcal{A}}(y^i, Y) > \delta$ and $\text{dist}_{\mathcal{A}}(y^i, y^{i'}) > \delta$ if $i \neq i'$.

Let us consider the δ -th power $\mathcal{G}_{\mathcal{A}}^\delta$ of the Gaifman graph of \mathcal{A} , whose vertex set is the same as $\mathcal{G}_{\mathcal{A}}$ and where there is an edge between two vertices iff their distance in $\mathcal{G}_{\mathcal{A}}$ is at most δ . The degree of $\mathcal{G}_{\mathcal{A}}^\delta$ can be coarsely bounded by d^δ . Thus, there are at most $p(d^\delta + 1)$ elements in or adjacent to X in $\mathcal{G}_{\mathcal{A}}^\delta$, and at least $m(d^\delta + 1)$ elements in $Y \setminus (X \cup N_{\mathcal{G}_{\mathcal{A}}^\delta}(X))$. Consider an arbitrary $(d^\delta + 1)$ -coloring of $\mathcal{G}_{\mathcal{A}}^\delta$ (where two adjacent vertices have different colors), which is guaranteed to exist since the degree of $\mathcal{G}_{\mathcal{A}}^\delta$ is at most d^δ . By the pigeonhole principle, one can find m elements y^1, \dots, y^m in $Y \setminus (X \cup N_{\mathcal{G}_{\mathcal{A}}^\delta}(X))$ with the same color. Since these elements do not belong to $X \cup N_{\mathcal{G}_{\mathcal{A}}^\delta}(X)$, they are at distance at least $\delta + 1$ from X in $\mathcal{G}_{\mathcal{A}}$. Furthermore, they form an independent set in $\mathcal{G}_{\mathcal{A}}^\delta$ (by definition of a coloring), and are thus at distance at least $(\delta + 1)$ from one another in $\mathcal{G}_{\mathcal{A}}$. This proves the claim.

We are now ready to prove the lemma: let B be a subset of A of size at most s . Define $g_n(m, \delta, s) := (s + nm)(d^\delta + 1)$. We show by induction on n that for every $C_1, \dots, C_n \subseteq A$ of size at least $g_n(m, \delta, s)$, one can find elements $c_j^i \in C_j$, for $j \in \{1, \dots, n\}$ and $i \in \{1, \dots, m\}$, which are at distance at least $\delta + 1$ from each other and from B in $\mathcal{G}_{\mathcal{A}}$. For $n = 0$, there is nothing to show. For the inductive step from n to $n + 1$: since $g_{n+1}(m, \delta, s) \geq g_n(m, \delta, s)$, we start by using the induction hypothesis to get suitable elements $c_j^i \in C_j$ for $j \in \{1, \dots, n\}$ and $i \in \{1, \dots, m\}$. We can then rely on the previous claim with $X := B \cup \bigcup_{j=1}^n \{c_j^1, \dots, c_j^m\}$ (i.e. $p = s + nm$) and $Y := C_{n+1}$ (which, by choice of $g_{n+1}(m, \delta, s)$, satisfy the cardinality condition of the claim), and pick the y^1, \dots, y^m as our $c_{n+1}^1, \dots, c_{n+1}^m$.

Setting $g(m, \delta, s) := g_{\#\text{neigh}_k^d}(m, \delta, s)$ concludes the proof: it is routine to check that such a function g satisfies the last two properties of the lemma. \square

Remark 5.5. Note that the size of (an instance of) a neighbourhood type $\tau \in \text{NEIGHTYPE}_k^d$ is at most

$$M_k^d := \begin{cases} 2k + 1 & \text{if } d = 2 \\ 1 + d^{\frac{(d-1)^k - 1}{d-2}} & \text{otherwise} \end{cases}$$

and thus can be bounded by a function of k (recall that the signature Σ and the degree d are fixed).

We now explain how to use Lemma 5.4 in order to define the notion of k -neighbourhood frequency in a structure \mathcal{A} . Let us consider

$$m := 2(k + 1) \cdot M_k^d! \quad \text{and} \quad \delta := 4k. \quad (5.2)$$

Our goal is, given a structure $\mathcal{A} \in \mathcal{C}$, to partition the k -neighbourhood types into two categories: the frequent types, and the rare types. The property we wish to enforce is that there exist in \mathcal{A} at least m occurrences of each one of the frequent k -neighbourhood types which are both

- at distance greater than δ from one another, and
- at distance greater than δ from every occurrence of a rare k -neighbourhood type.

To that end, for $t \in \mathbb{N}$ and $\mathcal{A} \in \mathcal{C}$ let $\text{FREQ}[\mathcal{A}]_k^{\geq t} \subseteq \text{NEIGHTYPE}_k^d$ denote the set of k -neighbourhood types which have at least t occurrences in \mathcal{A} .

Lemma 5.6. *Let $k \in \mathbb{N}$ and define m and δ as in Equation (5.2). There exists $\Upsilon(k) \in \mathbb{N}$ such that for every $\mathcal{A} \in \mathcal{C}$, there exists some $t \leq \Upsilon(k)$, for which*

$$t \geq g\left(m, \delta, \sum_{\tau \in \text{NEIGHTYPE}_k^d \setminus \text{FREQ}[\mathcal{A}]_k^{\geq t}} |\mathcal{A}|_{\tau}\right).$$

Proof. Define by induction $h_1(k) := 0$ and $h_{i+1}(k) := h_i(k) + g(m, \delta, h_i(k))$. This is well defined, as m and δ depend only on k . We now show that the lemma holds for

$$\Upsilon(k) := h_{\#\text{neigh}_k^d + 2}(k).$$

Fix some structure $\mathcal{A} \in \mathcal{C}$ and let $\tau_1, \dots, \tau_{\#\text{neigh}_k^d}$ be an enumeration of NEIGHTYPE_k^d such that for all $i < j$ we have $|\mathcal{A}|_{\tau_i} \leq |\mathcal{A}|_{\tau_j}$.

Consider the algorithm described in Procedure 1. The set X , which is empty at first, will eventually contain all types that will be declared rare. In words, this algorithm proceeds as follows:

- (1) First, we mark every k -neighbourhood type as frequent.
- (2) Among the types which are currently marked as frequent, let τ be one with the smallest number of occurrences in \mathcal{A} .
- (3) If $|\mathcal{A}|_{\tau}$ is at least $g(m, \delta, s)$ (g being the function from Lemma 5.4) where s is the total number of occurrences of all the k -neighbourhood types which are currently marked as rare, then we are done and the marking frequent/rare is final. Otherwise, mark τ as rare, and go back to step 2 if there remains at least one frequent k -neighbourhood type

Procedure 1: Determining the rare and frequent neighbourhood types

```

1  $X \leftarrow \emptyset$ 
2  $i \leftarrow 1$ 
3 while  $i \leq \#\text{neigh}_k^d$  and  $|\mathcal{A}|_{\tau_i} \leq g(m, \delta, \sum_{\tau \in X} |\mathcal{A}|_{\tau})$ 
4      $X \leftarrow X \cup \{\tau_i\}$ 
5      $i++$ 
6 end ;           // the rare types are those in  $X$ ; types outside  $X$  are frequent types
```

Let us establish the following invariant for the “while” loop:

$$\sum_{\tau \in X} |\mathcal{A}|_{\tau} \leq h_i(k). \tag{5.3}$$

Before the first execution of the body of the loop, both sides of Equation (5.3) are zero. Suppose now that Equation (5.3) is true for some $i \leq \#\text{neigh}_k^d$ at the start of the loop, and

let us prove that Equation (5.3) still holds after executing the body of the loop. We went from i to $i + 1$. Let X_i denote the old value of X , and X_{i+1} its new value; in other words, $X_{i+1} = X_i \cup \{\tau_i\}$. The left part of Equation (5.3) has been augmented by $|\mathcal{A}|_{\tau_i}$, which, because the “while” condition was satisfied, is less than $g(m, \delta, \sum_{\tau \in X_i} |\mathcal{A}|_{\tau})$. Thus,

$$\sum_{\tau \in X_{i+1}} |\mathcal{A}|_{\tau} \leq \sum_{\tau \in X_i} |\mathcal{A}|_{\tau} + g\left(m, \delta, \sum_{\tau \in X_i} |\mathcal{A}|_{\tau}\right).$$

By assumption, $\sum_{\tau \in X_i} |\mathcal{A}|_{\tau} \leq h_i(k)$, and since g is non-decreasing in its third variable, we get $\sum_{\tau \in X_{i+1}} |\mathcal{A}|_{\tau} \leq h_{i+1}(k)$, which means that Equation (5.3) is preserved by the body of the loop.

Let us now look at the two ways we can exit the “while” loop, and prove that the statement of the lemma holds whichever the case:

- If we exit the loop because the condition “ $i \leq \# \text{neigh}_k^d$ ” is false, then we have $i = \# \text{neigh}_k^d + 1$ and $X = \text{NEIGHTYPE}_k^d$. By Equation (5.3), $\sum_{\tau \in \text{NEIGHTYPE}_k^d} |\mathcal{A}|_{\tau} \leq h_{\# \text{neigh}_k^d + 1}(k)$, and we can choose $t := g(m, \delta, h_{\# \text{neigh}_k^d + 1}(k))$. We then indeed get $t \leq \Upsilon(k)$ and, because g is non-decreasing in its third argument,

$$t \geq g\left(m, \delta, \sum_{\tau \in \text{NEIGHTYPE}_k^d} |\mathcal{A}|_{\tau}\right) = g\left(m, \delta, \sum_{\tau \in \text{NEIGHTYPE}_k^d \setminus \text{FREQ}[\mathcal{A}]_k^{\geq t}} |\mathcal{A}|_{\tau}\right).$$

- Otherwise, $i \leq \# \text{neigh}_k^d$ and it is “ $|\mathcal{A}|_{\tau_i} \leq g(m, \delta, \sum_{\tau \in X} |\mathcal{A}|_{\tau})$ ” which fails. We can then choose $t := g(m, \delta, \sum_{\tau \in X} |\mathcal{A}|_{\tau}) + 1$ (which is, again, smaller than $\Upsilon(k)$). Let us prove that X is exactly the set of k -neighbourhood types that have less than t occurrences in \mathcal{A} . For $\tau \in X$, we have $|\mathcal{A}|_{\tau} \leq \sum_{\tau' \in X} |\mathcal{A}|_{\tau'} \leq g(m, \delta, \sum_{\tau' \in X} |\mathcal{A}|_{\tau'}) < t$ (the second inequality follows from the good properties of g , cf. Lemma 5.4). Conversely, if $\tau \notin X$, it means that $|\mathcal{A}|_{\tau} \geq |\mathcal{A}|_{\tau_i}$ (recall that the τ_i are examined in number of occurrences in \mathcal{A}), which in turn is at least t .

Then we have

$$t \geq g\left(m, \delta, \sum_{\tau \in X} |\mathcal{A}|_{\tau}\right) = g\left(m, \delta, \sum_{\tau \in \text{NEIGHTYPE}_k^d \setminus \text{FREQ}[\mathcal{A}]_k^{\geq t}} |\mathcal{A}|_{\tau}\right). \quad \square$$

We are now ready to formally establish the notion of frequent type:

Definition 5.7. Fix an integer k and a structure $\mathcal{A} \in \mathcal{C}$, and let $\text{FREQ}[\mathcal{A}]_k := \text{FREQ}[\mathcal{A}]_k^{\geq t}$ for the smallest threshold $t \leq \Upsilon(k)$ such that

$$t \geq g\left(m, \delta, \sum_{\tau \in \text{NEIGHTYPE}_k^d \setminus \text{FREQ}[\mathcal{A}]_k^{\geq t}} |\mathcal{A}|_{\tau}\right),$$

whose existence is guaranteed by Lemma 5.6. Some k -neighbourhood type $\tau \in \text{NEIGHTYPE}_k^d$ is said to be *frequent in \mathcal{A}* if it belongs to $\text{FREQ}[\mathcal{A}]_k$; that is, if $|\mathcal{A}|_{\tau} \geq t$. Otherwise, τ is said to be *rare in \mathcal{A}* .

Remark 5.8. A structure \mathcal{A} without frequent type has size bounded by $\Upsilon(k) \cdot \# \text{neigh}_k^d$ since, by Definition 5.7, each $\tau \in \text{NEIGHTYPE}_k^d$ has at most $\Upsilon(k)$ occurrences in \mathcal{A} .

Recall how m and δ were defined in Equation (5.2). Combining Lemmas 5.4–5.6, we get:

Proposition 5.9. *Let $k \in \mathbb{N}$. In every structure $\mathcal{A} \in \mathcal{C}$, one can find $2(k+1) \cdot M_k^d!$ occurrences of each frequent k -neighbourhood type in \mathcal{A} which are at distance at least $4k+1$ from one another and from the set of occurrences of every rare k -neighbourhood type in \mathcal{A} .*

With these notions of frequent and rare neighbourhood types established, we now turn to the construction of the two linear orders.

5.3. Constructing linear orders on \mathcal{A}_0 and \mathcal{A}_1 . Recall that our goal is to find a function f such that, given two structures $\mathcal{A}_0, \mathcal{A}_1$ in \mathcal{C} (assumed without loss of generality to have disjoint domains) such that $\mathcal{A}_0 \equiv_{f(k)}^{\text{FO}} \mathcal{A}_1$, we are able to construct linear orders on \mathcal{A}_0 and \mathcal{A}_1 (both denoted $<$) such that $(\mathcal{A}_0, <) \equiv_k^{\text{FO}^2} (\mathcal{A}_1, <)$. In this section, we define f and we detail the construction of such orders. The proof of FO^2 -similarity between $(\mathcal{A}_0, <)$ and $(\mathcal{A}_1, <)$ will be the focus of Section 5.4.

For every integer k , we choose an $f(k)$ large enough so that

- for every structures $\mathcal{A}_0, \mathcal{A}_1 \in \mathcal{C}$, $\mathcal{A}_0 \equiv_{f(k)}^{\text{FO}} \mathcal{A}_1$ entails $\llbracket \mathcal{A}_0 \rrbracket_k =^{\Upsilon(k)} \llbracket \mathcal{A}_1 \rrbracket_k$, where $\Upsilon(k)$ is defined in Lemma 5.6,
- $f(k) \geq \Upsilon(k) \cdot \#\text{neigh}_k^d + 1$,
- f satisfies an additional lower bound made explicit later in this section, in the paragraph *Transfer on \mathcal{A}_1* below.

In what follows, we fix an integer $k \in \mathbb{N}$, as well as two structures \mathcal{A}_0 and \mathcal{A}_1 of \mathcal{C} such that

$$\mathcal{A}_0 \equiv_{f(k)}^{\text{FO}} \mathcal{A}_1 \tag{5.4}$$

for such a choice of f .

By choice of f , we get $\llbracket \mathcal{A}_0 \rrbracket_k =^{\Upsilon(k)} \llbracket \mathcal{A}_1 \rrbracket_k$. This ensures, for all $t \leq \Upsilon(k)$, that $\text{FREQ}[\mathcal{A}_0]_k^{\geq t} = \text{FREQ}[\mathcal{A}_1]_k^{\geq t}$, and that

$$\sum_{\tau \in \text{NEIGHTYPE}_k^d \setminus \text{FREQ}[\mathcal{A}_0]_k^{\geq t}} |\mathcal{A}_0|_\tau = \sum_{\tau \in \text{NEIGHTYPE}_k^d \setminus \text{FREQ}[\mathcal{A}_1]_k^{\geq t}} |\mathcal{A}_1|_\tau.$$

The smallest t such that

$$t \geq g\left(m, \delta, \sum_{\tau \in \text{NEIGHTYPE}_k^d \setminus \text{FREQ}[\mathcal{A}_0]_k^{\geq t}} |\mathcal{A}_0|_\tau\right)$$

is therefore also the smallest t such that

$$t \geq g\left(m, \delta, \sum_{\tau \in \text{NEIGHTYPE}_k^d \setminus \text{FREQ}[\mathcal{A}_1]_k^{\geq t}} |\mathcal{A}_1|_\tau\right),$$

meaning exactly (according to Definition 5.7) that $\text{FREQ}[\mathcal{A}_0]_k = \text{FREQ}[\mathcal{A}_1]_k$. In other words, the notion of frequent and rare k -neighbourhood types coincide on \mathcal{A}_0 and \mathcal{A}_1 . From this point on, we thus refer to *frequent* and *rare* k -neighbourhood types without mentioning the structure involved.

Furthermore, if $\text{FREQ}[\mathcal{A}_0]_k$ (and thus $\text{FREQ}[\mathcal{A}_1]_k$) is empty, then according to Remark 5.8 we have $|\mathcal{A}_0| \leq \Upsilon(k) \cdot \#\text{neigh}_k^d$. As we chose $f(k) \geq \Upsilon(k) \cdot \#\text{neigh}_k^d + 1$, Equation (5.4) entails that $\mathcal{A}_0 \simeq \mathcal{A}_1$, and in particular that $\mathcal{A}_0 \equiv_k^{\text{FO}^2} \mathcal{A}_1$. From now on, we thus suppose that there is at least one frequent k -neighbourhood type.

To construct the two linear orders, we need to define the notion of k -environment: given $\mathcal{A} \in \mathcal{C}$, a linear order $<$ on A , $k \in \mathbb{N}$ and an element $a \in A$, we define the k -environment $\text{Env}_{(\mathcal{A}, <)}^k(a)$ of a in $(\mathcal{A}, <)$ as the restriction of $(\mathcal{A}, <)$ to the k -neighbourhood of a in \mathcal{A} , where a is the interpretation of the constant symbol c . Note that the order is not taken into account when determining the domain of the substructure (it would otherwise be A , given that any two distinct elements are adjacent for $<$). The k -environment type $\text{env-tp}_{(\mathcal{A}, <)}^k(a)$ is the isomorphism class of $\text{Env}_{(\mathcal{A}, <)}^k(a)$. In other words, $\text{env-tp}_{(\mathcal{A}, <)}^k(a)$ contains the information of $\mathcal{N}_{\mathcal{A}}^k(a)$ together with the order of its elements in $(\mathcal{A}, <)$.

Given $\tau \in \text{NEIGHTYPE}_k^d$, we define $\text{ENV}(\tau)$ as the set of k -environment types whose underlying k -neighbourhood type is τ .

For each $i \in \{0, 1\}$, we aim to partition A_i into $2(2k + 1) + 2$ segments:

$$A_i = X_i \uplus \biguplus_{j=0}^{2k} (L_i^j \uplus R_i^j) \uplus M_i.$$

Once we have set a linear order on each segment, the linear order $<$ on A_i will result from the concatenation of the orders on the segments as follows:

$$(A_i, <) := X_i \cdot L_i^0 \cdot L_i^1 \cdots L_i^{2k} \cdot M_i \cdot R_i^{2k} \cdots R_i^1 \cdot R_i^0.$$

Formally, we order the segments as

$$X_i < L_i^0 < L_i^1 < \cdots < L_i^{2k} < M_i < R_i^{2k} < \cdots < R_i^1 < R_i^0,$$

and for $x, y \in A_i$, we let $x < y$ either if x 's segment is smaller than y 's segment, or if both belong to the same segment and x is less than y according to the order for this segment.

Each segment L_i^j , for $j \in \{0, \dots, 2k\}$, is itself decomposed into two segments NL_i^j and UL_i^j , and the order on L_i^j will be defined as the concatenation of the order on NL_i^j with the order on UL_i^j (meaning that all elements from NL_i^j will be smaller than all elements from UL_i^j). The UL_i^j for $j \in \{k+1, \dots, 2k\}$ will be empty; they are defined solely in order to keep the notations uniform. The 'N' stands for "neighbour" and the 'U' for "universal", for reasons that will soon become apparent. Symmetrically, each R_i^j is decomposed into $UR_i^j \cdot NR_i^j$ (and thus all elements of UR_i^j will be smaller than all elements of NR_i^j), with UR_i^j being empty as soon as $j \geq k+1$.

For $i \in \{0, 1\}$ and $r \in \{0, \dots, 2k\}$, we define S_i^r as

$$S_i^r := X_i \cup \bigcup_{j=0}^r (L_i^j \cup R_i^j).$$

Let us now explain how the segments are constructed in \mathcal{A}_0 ; see Figure 1 for an illustration.

For every $\tau \in \text{FREQ}[\mathcal{A}_0]_k$, let $\tau_1, \dots, \tau_{|\text{ENV}(\tau)|}$ be an enumeration of $\text{ENV}(\tau)$. Recall the definition of M_k^d in Remark 5.5: we have $|\text{ENV}(\tau)| \leq M_k^d!$ for every $\tau \in \text{NEIGHTYPE}_k^d$.

In particular, Proposition 5.9 ensures that we are able to pick, for every $\tau \in \text{FREQ}[\mathcal{A}_0]_k$, every $l \in \{1, \dots, |\text{ENV}(\tau)|\}$ and every $j \in \{0, \dots, k\}$, two elements $a[\tau_l]_L^j$ and $a[\tau_l]_R^j$ which have τ as k -neighbourhood type in \mathcal{A}_0 , such that all the $a[\tau_l]_L^j$ and the $a[\tau_l]_R^j$, for $\tau \in \text{FREQ}[\mathcal{A}_0]_k$, $l \in \{1, \dots, |\text{ENV}(\tau)|\}$ and $j \in \{0, \dots, k\}$, are at distance at least $4k+1$ from each other and from any occurrence of a rare k -neighbourhood type in \mathcal{A}_0 .

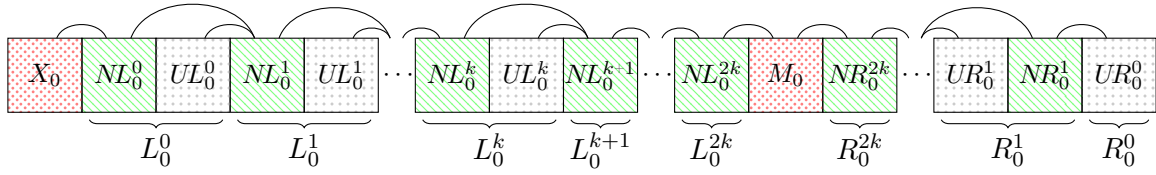


Figure 1: Elements from two distinct segments can only be adjacent if there is a black curvy edge between those segments. The order $<$ grows from the left to the right. Note that NR_0^0 is empty and thus left out of the picture.

As their names suggest, $a[\tau_l]_L^j$ will appear in UL_0^j and $a[\tau_l]_R^j$ in UR_0^j , in order to make sure that each environment (i.e., ordering) of each frequent type appears in each universal segment UL_0^j, UR_0^j – which are called “universal” for that very reason.

Construction of X_0 and NL_0^0 . We start with the set X_0 , which contains all the occurrences of rare k -neighbourhood types, together with their k -neighbourhoods. Formally, X_0 is $\bigcup_{a \in \mathcal{A}_0: \text{neigh-tp}_{\mathcal{A}_0}^k(a) \in \text{NEIGHTYPE}_k^d \setminus \text{FREQ}[\mathcal{A}_0]_k} N_{\mathcal{A}_0}^k(a)$. We set $NL_0^0 := N_{\mathcal{A}_0}(X_0)$ (the set of neighbours of elements of X_0), and define the order $<$ on X_0 and on NL_0^0 in an arbitrary way.

Construction of UL_0^j . If $k < j \leq 2k$, then we set $UL_0^j := \emptyset$. Otherwise, for $j \in \{0, \dots, k\}$, once we have constructed L_0^0, \dots, L_0^{j-1} and NL_0^j , we construct UL_0^j as follows.

The elements of UL_0^j are $\bigcup_{\tau \in \text{FREQ}[\mathcal{A}_0]_k} \bigcup_{l=1}^{|\text{ENV}(\tau)|} N_{\mathcal{A}_0}^k(a[\tau_l]_L^j)$. Note that UL_0^j does not intersect the previously constructed segments, by choice of the $a[\tau_l]_L^j$. Furthermore, the $N_{\mathcal{A}_0}^k(a[\tau_l]_L^j)$ are pairwise disjoint, hence we can fix $<$ freely and independently on each of them. Unsurprisingly, we order each $N_{\mathcal{A}_0}^k(a[\tau_l]_L^j)$ so that $\text{env-tp}_{(\mathcal{A}_0, <)}^k(a[\tau_l]_L^j) = \tau_l$. This is possible because for every $\tau \in \text{FREQ}[\mathcal{A}_0]_k$ and each l , $\text{neigh-tp}_{\mathcal{A}_0}^k(a[\tau_l]_L^j) = \tau$ by choice of $a[\tau_l]_L^j$.

Once each $N_{\mathcal{A}_0}^k(a[\tau_l]_L^j)$ is ordered according to τ_l , the linear order $<$ on UL_0^j can be completed in an arbitrary way. Note that every possible k -environment type extending a frequent k -neighbourhood type in \mathcal{A}_0 occurs in each UL_0^j . The UL_0^j are *universal* in that sense.

Construction of NL_0^j . Now, let us see how the NL_0^j are constructed. For $j \in \{1, \dots, 2k\}$, suppose that we have constructed L_0^0, \dots, L_0^{j-1} . The domain of NL_0^j consists of all the neighbours (in \mathcal{A}_0) of the elements of L_0^{j-1} not already belonging to the construction so far. Formally, $N_{\mathcal{A}_0}(L_0^{j-1}) \setminus (X_0 \cup \bigcup_{m=0}^{j-2} L_0^m)$.

The order $<$ on NL_0^j is chosen arbitrarily.

Construction of R_0^j . We construct similarly the R_0^j , for $j \in \{0, \dots, 2k\}$, starting with $NR_0^0 := \emptyset$, then UR_0^0 which contains each $a[\tau_l]_R^0$ together with its k -neighbourhood in \mathcal{A}_0 ordered according to τ_l , then $NR_0^1 := N_{\mathcal{A}_0}(R_0^0)$, then UR_0^1 , etc.

Note that the $a[\tau_l]_R^j$ have been chosen so that they are far enough in \mathcal{A}_0 from all the segments that have been constructed so far, allowing us once more to order their k -neighbourhood in \mathcal{A}_0 as we see fit.

Construction of M_0 . Let M_0 contain all the elements of \mathcal{A}_0 besides those already belonging to S_0^{2k} . The order $<$ chosen on M_0 is arbitrary.

Transfer on \mathcal{A}_1 . Suppose that we have constructed S_0^{2k} . We can make sure, retrospectively, that the index $f(k)$ in Equation (5.4) is large enough so that there exists a set $S \subseteq \mathcal{A}_1$ so that $\mathcal{A}_0 \upharpoonright_{S_0^{2k} \cup N_{\mathcal{A}_0}(S_0^{2k})} \simeq \mathcal{A}_1 \upharpoonright_S$. This is ensured as long as $f(k) \geq |S_0^{2k} \cup N_{\mathcal{A}_0}(S_0^{2k})| + 1$, whose right side can be bounded by a function of k , independent of \mathcal{A}_0 and \mathcal{A}_1 . Indeed, all elements of $S_0^{2k} \cup N_{\mathcal{A}_0}(S_0^{2k})$ are at distance at most $3k + 2$ from X_0 or one of the $a[\tau_l]_*^j$. Since the number of such elements is bounded by a function of k (remember that, by Definition 5.7, the number of occurrences of each rare neighbourhood types is at most $\Upsilon(k)$) and since the degree is at most d (which is fixed), $|S_0^{2k} \cup N_{\mathcal{A}_0}(S_0^{2k})|$ can be bounded by a function of k .

Let $\phi_0 : \mathcal{A}_0 \upharpoonright_{S_0^{2k}} \rightarrow \mathcal{A}_1 \upharpoonright_{S'}$ be the restriction to S_0^{2k} of said isomorphism, and let ϕ_1 be its converse. By construction, the k -neighbourhood of every $a \in S_0^k$ is included in S_0^{2k} ; hence every such a has the same k -neighbourhood type in \mathcal{A}_0 as has $\phi_0(a)$ in \mathcal{A}_1 .

We transfer alongside ϕ_0 all the segments besides M_0 , with their order, from $(\mathcal{A}_0, <)$ to \mathcal{A}_1 , thus defining $X_1, NL_1^j, UL_1^j, \dots$ as the respective images by ϕ_0 of $X_0, NL_0^j, UL_0^j, \dots$, and we define M_1 as the set of remaining elements of \mathcal{A}_1 . It remains to define the restriction of $<$ on M_1 : this is done in an arbitrary way. Note that the properties concerning neighbourhood are transferred; *e.g.* all the neighbours of an element in L_1^j , $1 \leq j < 2k$, belong to $L_1^{j-1} \cup L_1^j \cup L_1^{j+1}$.

By construction, we get the following lemma:

Lemma 5.10. *For each $a \in S_0^k$, we have $\text{env-tp}_{(\mathcal{A}_0, <)}^k(a) = \text{env-tp}_{(\mathcal{A}_1, <)}^k(\phi_0(a))$.*

Lemma 5.10 has two immediate consequences:

- The set X_1 contains the occurrences in \mathcal{A}_1 of all the rare k -neighbourhood types (just forget about the order on the k -environments, and remember that \mathcal{A}_0 and \mathcal{A}_1 have the same number of occurrences of each rare k -neighbourhood type).
- All the universal segments UL_1^j and UR_1^j , for $0 \leq j \leq k$, contain at least one occurrence of each environment in $\text{ENV}(\tau)$, for each $\tau \in \text{FREQ}[\mathcal{A}_0]_k$.

By virtue of ϕ_0 being an isomorphism, and because the order $<$ was constructed on S' as the image by ϕ_0 of $<$, we get:

Lemma 5.11. *For each $a, b \in S_0^k$, we have $\text{tp}_{(\mathcal{A}_0, <)}(a, b) = \text{tp}_{(\mathcal{A}_1, <)}(\phi_0(a), \phi_0(b))$.*

In particular, for $a = b \in S_0^k$, we have $\text{tp}_{(\mathcal{A}_0, <)}(a) = \text{tp}_{(\mathcal{A}_1, <)}(\phi_0(a))$.

5.4. Proof of the FO²-similarity of $(\mathcal{A}_0, <)$ and $(\mathcal{A}_1, <)$. In this section, we aim to show the following result:

Proposition 5.12. $(\mathcal{A}_0, <) \equiv_k^{\text{FO}^2} (\mathcal{A}_1, <)$

Recall Theorem 2.1: in order to prove Proposition 5.12, we show that the duplicator wins the k -round two-pebble Ehrenfeucht-Fraïssé game on $(\mathcal{A}_0, <)$ and $(\mathcal{A}_1, <)$. For that, let us show by a decreasing induction on r from k to 0 that the duplicator can ensure, after $k - r$ rounds, that the three following properties (described below) hold. Recall that ϕ_0 and

ϕ_1 are the isomorphisms between $\mathcal{A}_0 \upharpoonright_{\mathcal{A}_0 \setminus M_0}$ and $\mathcal{A}_1 \upharpoonright_{\mathcal{A}_1 \setminus M_1}$ along which the segments were transferred from \mathcal{A}_0 to \mathcal{A}_1 .

$$\forall i \in \{0, 1\} \forall \alpha \in \{x, y\} p_i^\alpha \in S_i^r \text{ implies } p_{1-i}^\alpha = \phi_i(p_i^\alpha) \quad (\mathcal{S}_r)$$

$$\forall \alpha \in \{x, y\} \text{env-tp}_{(\mathcal{A}_0, <)}^r(p_0^\alpha) = \text{env-tp}_{(\mathcal{A}_1, <)}^r(p_1^\alpha) \quad (\mathcal{E}_r)$$

$$\text{tp}_{(\mathcal{A}_0, <)}(p_0^x, p_0^y) = \text{tp}_{(\mathcal{A}_1, <)}(p_1^x, p_1^y) \quad (\mathcal{T}_r)$$

Let us stress that in this context, p_0^x, p_0^y, p_1^x and p_1^y denote the elements of \mathcal{A}_0 and \mathcal{A}_1 on which the pebbles have been played after $k - r$ rounds, and thus depend on r .

The first property, (\mathcal{S}_r) , guarantees that if a pebble is close (in a sense that depends on the number of rounds left in the game) to one of the $<$ -minimal or $<$ -maximal elements, the corresponding pebble in the other structure is located at the same position with respect to this $<$ -extremal element.

As for (\mathcal{E}_r) , it states that two corresponding pebbles are always placed on elements sharing the same r -environment type. Once again, the safety distance decreases with each round that goes.

Finally, (\mathcal{T}_r) controls that both pebbles have the same relative position (both with respect to the order and the original vocabulary) in the two ordered structures. In particular, the duplicator wins the game if (\mathcal{T}_r) is satisfied at the beginning of the game, and after each of the k rounds of the game.

5.4.1. *Base case: proofs of (\mathcal{S}_k) , (\mathcal{E}_k) and (\mathcal{T}_k) .* We start by proving (\mathcal{S}_k) , (\mathcal{E}_k) and (\mathcal{T}_k) . At the start of the game, the duplicator places both p_0^x and p_0^y on the $<$ -minimal element of $(\mathcal{A}_0, <)$, and both p_1^x and p_1^y on the $<$ -minimal element of $(\mathcal{A}_1, <)$. In particular,

$$p_1^x = p_1^y = \phi_0(p_0^x) = \phi_0(p_0^y).$$

This ensures that (\mathcal{S}_k) holds, while (\mathcal{E}_k) and (\mathcal{T}_k) follow from Lemmas 5.10–5.11.

5.4.2. *Winning strategy for the duplicator.* We now describe the duplicator's strategy to ensure that (\mathcal{S}_r) , (\mathcal{E}_r) and (\mathcal{T}_r) hold no matter how the spoiler plays.

Suppose that we have (\mathcal{S}_{r+1}) , (\mathcal{E}_{r+1}) and (\mathcal{T}_{r+1}) for some $0 \leq r < k$, after $k - r - 1$ rounds of the game. Without loss of generality, we may assume that, in the $(k - r)$ -th round of the Ehrenfeucht-Fraïssé game between $(\mathcal{A}_0, <)$ and $(\mathcal{A}_1, <)$, the spoiler moves p_0^x in $(\mathcal{A}_0, <)$.

- (I) If $p_0^x \in S_0^r$, then the duplicator responds by placing p_1^x on $\phi_0(p_0^x)$.

This corresponds to a tit-for-tat strategy. If the duplicator does not respond this way, then the spoiler will be able to expose the difference between $(\mathcal{A}_0, <)$ and $(\mathcal{A}_1, <)$ in the subsequent moves, by forcing the duplicator to play closer and closer to the $<$ -minimal or $<$ -maximal elements.

- (II) Else, if $p_0^x \notin S_0^r$, and $p_0^x \in N_{\mathcal{A}_0}^1(p_0^y)$, then (\mathcal{E}_{r+1}) ensures that there exists an isomorphism $\psi : \text{Env}_{(\mathcal{A}_0, <)}^{r+1}(p_0^y) \rightarrow \text{Env}_{(\mathcal{A}_1, <)}^{r+1}(p_1^y)$. The duplicator responds by placing p_1^x on $\psi(p_0^x)$ – in other words, on the element whose relative position to p_1^y is the same as the relative position of p_0^x with respect to p_0^y .
- (III) Else suppose that $(\mathcal{A}_0, <) \models p_0^x < p_0^y$ and $p_0^y \notin L_0^{r+1}$. In other words, the spoiler places p_0^x to the left of p_0^y , when p_0^y is not in the leftmost segment outside of S_0^r (namely, L_0^{r+1}). The situation where $p_0^y \in L_0^{r+1}$ is covered by Case (IV).

We have $\tau := \text{neigh-tp}_{\mathcal{A}_0}^k(p_0^x) \in \text{FREQ}[\mathcal{A}_0]_k$ as $p_0^x \notin X_0$. Let $\tau_l := \text{env-tp}_{(\mathcal{A}_0, <)}^k(p_0^x)$.

The duplicator responds by placing p_1^x on $\phi_0(a[\tau_l]_L^{r+1})$.

- (IV) Else, if $(\mathcal{A}_0, <) \models p_0^x < p_0^y$ and $p_0^y \in L_0^{r+1}$, then the duplicator moves p_1^x on $\phi_0(p_0^x)$ (by (\mathcal{S}_{r+1}) , p_0^x indeed belongs to the domain of ϕ_0).
- (V) Else, suppose that $(\mathcal{A}_0, <) \models p_0^y < p_0^x$ and $p_0^y \notin R_0^{r+1}$. This case is symmetric to Case (III).

Similarly, the duplicator opts to play p_1^x on $\phi_0(a[\tau_l]_R^{r+1})$, where $\tau_l := \text{env-tp}_{(\mathcal{A}_0, <)}^k(p_0^x)$.

- (VI) If we are in none of the cases above, it means that the spoiler has placed p_0^x to the right of p_0^y , and that $p_0^y \in R_0^{r+1}$. This case is symmetric to Case (IV).

Once again, the duplicator places p_1^x on $\phi_0(p_0^x)$.

It remains to show that this strategy satisfies our invariants: under the inductive assumption that (\mathcal{S}_{r+1}) , (\mathcal{E}_{r+1}) and (\mathcal{T}_{r+1}) hold, for some $0 \leq r < k$, we need to show that this strategy ensures that (\mathcal{S}_r) , (\mathcal{E}_r) and (\mathcal{T}_r) hold.

We treat each case in its own section: Section 5.4.3 is devoted to Case (I) while Section 5.4.4 covers Case (II). Both Cases (III) and (IV) are treated in Section 5.4.5. Cases (V) and (VI), being their exact symmetric counterparts, are left to the reader.

Remark 5.13. Note that some properties need no verification. Since p_0^y and p_1^y are left untouched by the players, (\mathcal{S}_{r+1}) ensures that half of (\mathcal{S}_r) automatically holds, namely that

$$\forall i \in \{0, 1\}, \quad p_i^y \in S_i^r \quad \rightarrow \quad p_{1-i}^y = \phi_i(p_i^y).$$

Similarly, the part of (\mathcal{E}_r) concerning p_0^y and p_1^y follows from (\mathcal{E}_{r+1}) :

$$\text{env-tp}_{(\mathcal{A}_0, <)}^r(p_0^y) = \text{env-tp}_{(\mathcal{A}_1, <)}^r(p_1^y).$$

Lastly, notice that once we have shown that (\mathcal{E}_r) holds, it follows that $\text{tp}_{\mathcal{A}_0}(p_0^x) = \text{tp}_{\mathcal{A}_1}(p_1^x)$ and $\text{tp}_{\mathcal{A}_0}(p_0^y) = \text{tp}_{\mathcal{A}_1}(p_1^y)$. This remark will help us prove (\mathcal{T}_r) .

5.4.3. When the spoiler plays near the endpoints: Case (I). In this section, we treat the case where the spoiler places p_0^x near the $<$ -minimal or $<$ -maximal element of $(\mathcal{A}_0, <)$. Obviously, what “near” means depends on the number of rounds left in the game; the more rounds remain, the more the duplicator must be cautious regarding the possibility for the spoiler to reach an endpoint and potentially expose a difference between $(\mathcal{A}_0, <)$ and $(\mathcal{A}_1, <)$.

As we have stated in Case (I), with r rounds left, we consider a move on p_0^x by the spoiler to be near the endpoints if it is made in S_0^r . In that case, the duplicator responds along the tit-for-tat strategy, namely by placing p_1^x on $\phi_0(p_0^x)$.

Let us now prove that this strategy guarantees that (\mathcal{S}_r) , (\mathcal{E}_r) and (\mathcal{T}_r) hold. Recall from Remark 5.13 that part of the task is already taken care of.

Proof of (\mathcal{S}_r) in Case (I). We have to show that $\forall i \in \{0, 1\}, p_i^x \in S_i^r \rightarrow p_{1-i}^x = \phi_i(p_i^x)$. This follows directly from the duplicator’s strategy, since $p_1^x = \phi_0(p_0^x)$ (thus $p_0^x = \phi_1(p_1^x)$).

Proof of (\mathcal{E}_r) in Case (I). We need to prove that $\text{env-tp}_{(\mathcal{A}_0, <)}^r(p_0^x) = \text{env-tp}_{(\mathcal{A}_1, <)}^r(p_1^x)$, which is a consequence of Lemma 5.10 given that $p_1^x = \phi_0(p_0^x)$ and $r < k$.

Proof of (\mathcal{T}_r) in Case (I). First, suppose that $p_0^y \in S_0^{r+1}$. By (\mathcal{S}_{r+1}) , we know that $p_1^y = \phi_0(p_0^y)$. Thus, Lemma 5.11 allows us to conclude that $\text{tp}_{(\mathcal{A}_0, <)}(p_0^x, p_0^y) = \text{tp}_{(\mathcal{A}_1, <)}(p_1^x, p_1^y)$.

Otherwise, $p_0^y \notin S_0^{r+1}$ and (\mathcal{S}_{r+1}) entails that $p_1^y \notin S_1^{r+1}$. We have two points to establish:

$$\text{tp}_{\mathcal{A}_0}(p_0^x, p_0^y) = \text{tp}_{\mathcal{A}_1}(p_1^x, p_1^y), \quad (5.5)$$

$$\text{tp}_{<}(p_0^x, p_0^y) = \text{tp}_{<}(p_1^x, p_1^y). \quad (5.6)$$

By construction, the neighbours in \mathcal{A}_i of an element of S_i^r all belong to S_i^{r+1} , thus p_0^x and p_0^y never appear in the same tuple of some relation, and neither do p_1^x and p_1^y . In other words, the only atoms that hold in \mathcal{A}_0 about p_0^x and p_0^y concern independently p_0^x and p_0^y . Similarly, the only atoms that hold in \mathcal{A}_1 about p_1^x and p_1^y concern independently p_1^x and p_1^y . Equation (5.5) follows from this fact together with $\text{tp}_{\mathcal{A}_0}(p_0^x) = \text{tp}_{\mathcal{A}_1}(p_1^x)$ and $\text{tp}_{\mathcal{A}_0}(p_0^y) = \text{tp}_{\mathcal{A}_1}(p_1^y)$ (Remark 5.13).

As for Equation (5.6), either

$$p_0^x \in X_0 \cup \bigcup_{0 \leq j \leq r} L_0^j \quad \text{and} \quad p_1^x \in X_1 \cup \bigcup_{0 \leq j \leq r} L_1^j,$$

in which case $\text{tp}_{<}(p_0^x, p_0^y)$ and $\text{tp}_{<}(p_1^x, p_1^y)$ both contain the atom “ $x < y$ ” (as p_0^x and p_1^x appear in segments to the left of the segments containing p_0^y and p_1^y , respectively) and are thus equal, or

$$p_0^x \in \bigcup_{0 \leq j \leq r} R_0^j \quad \text{and} \quad p_1^x \in \bigcup_{0 \leq j \leq r} R_1^j,$$

in which case $\text{tp}_{<}(p_0^x, p_0^y)$ and $\text{tp}_{<}(p_1^x, p_1^y)$ both contain the atom “ $x > y$ ”, and are still equal.

5.4.4. *When the spoiler plays next to the other pebble: Case (II).* Suppose now that the spoiler places p_0^x next to the other pebble in \mathcal{A}_0 (i.e. $p_0^x \in N_{\mathcal{A}_0}^1(p_0^y)$), but not in S_0^r (for that move would fall under the jurisdiction of Case (I)). In that case, the duplicator must place p_1^x so that the relative position of p_1^x and p_1^y is the same as that of p_0^x and p_0^y .

For that, we can use (\mathcal{E}_{r+1}) , which guarantees that $\text{env-tp}_{(\mathcal{A}_0, <)}^{r+1}(p_0^y) = \text{env-tp}_{(\mathcal{A}_1, <)}^{r+1}(p_1^y)$. Thus there exists an isomorphism ψ between $\text{Env}_{(\mathcal{A}_0, <)}^{r+1}(p_0^y)$ and $\text{Env}_{(\mathcal{A}_1, <)}^{r+1}(p_1^y)$. Note that this isomorphism is unique, as we are dealing with linear orders.

The duplicator’s response is to place p_1^x on $\psi(p_0^x)$. Let us now prove that this strategy is correct with respect to our invariants (\mathcal{S}_r) , (\mathcal{E}_r) and (\mathcal{T}_r) .

Proof of (\mathcal{S}_r) in Case (II). Because the spoiler’s move does not fall under Case (I), we know that $p_0^x \notin S_0^r$. Let us now show that p_1^x is not near the endpoints either, i.e. that $p_1^x \notin S_1^r$.

We consider two cases, depending on whether $p_1^y \in S_1^{r+1}$. Since p_1^x and p_1^y are neighbours in \mathcal{A}_1 , if $p_1^y \notin S_1^{r+1}$, then by construction $p_1^x \notin S_1^r$. Otherwise, $p_1^y \in S_1^{r+1}$ and we know by (\mathcal{S}_{r+1}) that $p_0^y = \phi_1(p_1^y)$. Because ψ is the unique isomorphism between $\text{Env}_{(\mathcal{A}_0, <)}^{r+1}(p_0^y)$ and $\text{Env}_{(\mathcal{A}_1, <)}^{r+1}(p_1^y)$, ψ is equal to the restriction $\widetilde{\phi}_0$ of ϕ_0 :

$$\widetilde{\phi}_0 : \text{Env}_{(\mathcal{A}_0, <)}^{r+1}(p_0^y) \rightarrow \text{Env}_{(\mathcal{A}_1, <)}^{r+1}(p_1^y).$$

Thus $p_0^x = \psi^{-1}(p_1^x) = \widetilde{\phi}_0^{-1}(p_1^x) = \phi_1(p_1^x)$, and by definition of the segments on $(\mathcal{A}_1, <)$, which are just a transposition of the segments of $(\mathcal{A}_0, <)$ via ϕ_0 , $p_0^x \notin S_0^r$ then entails $p_1^x \notin S_1^r$.

Since we neither have $p_0^x \in S_0^r$ nor $p_1^x \in S_1^r$, (\mathcal{S}_r) holds – recall from Remark 5.13 that the part concerning p_0^y and p_1^y is always satisfied.

Proof of (\mathcal{E}_r) in Case (II). Recall that the duplicator placed p_1^x on the image of p_0^x by the isomorphism

$$\psi : \text{Env}_{(\mathcal{A}_0, <)}^{r+1}(p_0^y) \rightarrow \text{Env}_{(\mathcal{A}_1, <)}^{r+1}(p_1^y).$$

It is easy to check that the restriction $\tilde{\psi}$ of ψ : $\tilde{\psi} : \text{Env}_{(\mathcal{A}_0, <)}^r(p_0^x) \rightarrow \text{Env}_{(\mathcal{A}_1, <)}^r(p_1^x)$ is well-defined, and is indeed an isomorphism.

This ensures that $\text{env-tp}_{(\mathcal{A}_0, <)}^r(p_0^x) = \text{env-tp}_{(\mathcal{A}_1, <)}^r(p_1^x)$, thus completing the proof of (\mathcal{E}_r) .

Proof of (\mathcal{T}_r) in Case (II). This follows immediately from the fact that the isomorphism ψ maps p_0^x to p_1^x and p_0^y to p_1^y : all the atomic facts about these elements are preserved.

5.4.5. *When the spoiler plays to the left: Cases (III) and (IV).* We now treat our last case, which covers both Cases (III) and (IV), *i.e.* the instances where the spoiler places p_0^x to the left of p_0^y (formally: such that $(\mathcal{A}_0, <) \models p_0^x < p_0^y$), which do not already fall in Cases (I)–(II).

Note that the scenario in which the spoiler plays to the right of the other pebble is the exact symmetric of this one (since the X_i play no role in this case, left and right can be interchanged harmlessly).

The idea here is very simple: since the spoiler has placed p_0^x to the left of p_0^y , but neither in S_0^r nor in $N_{\mathcal{A}_0}^1(p_0^y)$, the duplicator responds by placing p_1^x on an element of UL_1^{r+1} (the leftmost universal segment not in S_1^r) sharing the same k -environment. This is possible by construction of the universal segments: if $\tau_l := \text{env-tp}_{(\mathcal{A}_0, <)}^k(p_0^x)$ (which must extend a frequent k -neighbourhood type, since $p_0^x \notin X_0$), then $\phi_0(a[\tau_l]_L^{r+1})$ satisfies the requirements.

There is one caveat to this strategy. If p_1^y is itself in L_1^{r+1} , two problems may arise: first, it is possible for p_1^x and p_1^y to be in the wrong order (*i.e.* such that $(\mathcal{A}_1, <) \models p_1^x > p_1^y$). Second, it may be the case that p_1^x and p_1^y are neighbours in \mathcal{A}_1 , which, together with the fact that p_0^x and p_0^y are not neighbours in \mathcal{A}_0 , would break (\mathcal{T}_r) .

This is why the duplicator's strategy depends on whether $p_1^y \in L_1^{r+1}$:

- if this is not the case, then the duplicator places p_1^x on $\phi_0(a[\tau_l]_L^{r+1})$. This corresponds to Case (III).
- if $p_1^y \in L_1^{r+1}$, then (\mathcal{S}_{r+1}) guarantees that $p_0^y \in L_0^{r+1}$. Hence p_0^x , which is located to the left of p_0^y , is in the domain of ϕ_0 : the duplicator moves p_1^x to $\phi_0(p_0^x)$. This situation corresponds to Case (IV).

Let us prove that (\mathcal{S}_r) , (\mathcal{E}_r) and (\mathcal{T}_r) hold in both of these instances.

Proof of (\mathcal{S}_r) in Case (III). Since the spoiler's move does not fall under Case (I), we have that $p_0^x \notin S_0^r$. By construction, $a[\tau_l]_L^{r+1} \in L_0^{r+1}$, thus $\phi_0(a[\tau_l]_L^{r+1}) \in L_1^{r+1}$, and $p_1^x \notin S_1^r$.

Proof of (\mathcal{E}_r) in Case (III). It follows from $\text{env-tp}_{(\mathcal{A}_0, <)}^k(a[\tau_l]_L^{r+1}) = \tau_l$ together with Lemma 5.10 that

$$\text{env-tp}_{(\mathcal{A}_0, <)}^k(p_0^x) = \text{env-tp}_{(\mathcal{A}_1, <)}^k(p_1^x).$$

A fortiori, $\text{env-tp}_{(\mathcal{A}_0, <)}^r(p_0^x) = \text{env-tp}_{(\mathcal{A}_1, <)}^r(p_1^x)$.

Proof of (\mathcal{T}_r) in Case (III). Because the spoiler's move does not fall under Case (II), $p_0^x \notin N_{\mathcal{A}_0}^1(p_0^y)$. In other words, the (positive) atoms occurring in $\text{tp}_{\mathcal{A}_0}(p_0^x, p_0^y)$ are exactly those appearing in $\text{tp}_{\mathcal{A}_0}(p_0^x)$ and $\text{tp}_{\mathcal{A}_0}(p_0^y)$, as no atom involving both p_0^x and p_0^y holds.

Recall the construction of UL_0^{r+1} : the whole k -neighbourhood of $a[\tau_l]_L^{r+1}$ was included in this segment. In particular, $N_{\mathcal{A}_1}^1(p_1^x) = N_{\mathcal{A}_1}^1(\phi_0(a[\tau_l]_L^{r+1})) \subseteq UL_1^{r+1}$. By assumption, $p_1^y \notin L_1^{r+1}$, which entails that the atoms in $\text{tp}_{\mathcal{A}_1}(p_1^x, p_1^y)$ are those occurring in $\text{tp}_{\mathcal{A}_1}(p_1^x)$ and $\text{tp}_{\mathcal{A}_1}(p_1^y)$. It then follows from the last observation of Remark 5.13 that $\text{tp}_{\mathcal{A}_0}(p_0^x, p_0^y) = \text{tp}_{\mathcal{A}_1}(p_1^x, p_1^y)$.

Let us now prove that $\text{tp}_{<}(p_1^x, p_1^y)$ contains the atom “ $x < y$ ”. We claim that $p_1^y \notin X_1 \cup \bigcup_{0 \leq j \leq r+1} L_1^j$. Suppose otherwise: (\mathcal{S}_{r+1}) would entail that $p_0^y \in X_0 \cup \bigcup_{0 \leq j \leq r+1} L_0^j$ which, together with the assumption $p_0^y \notin L_0^{r+1}$ and $p_0^x < p_0^y$, would result in p_0^x being in S_0^r , which is a contradiction. Both $\text{tp}_{<}(p_1^x, p_1^y)$ and $\text{tp}_{<}(p_0^x, p_0^y)$ contain the atom “ $x < y$ ” and thus must be equal, which concludes the proof of (\mathcal{T}_r) .

Proof of (\mathcal{S}_r) , (\mathcal{E}_r) and (\mathcal{T}_r) in Case (IV). Let us now move to the case where $p_1^y \in L_1^{r+1}$. Recall that under this assumption, $p_0^y = \phi_1(p_1^y) \in L_0^{r+1}$ and since $p_0^x < p_0^y$ and $p_0^x \notin S_0^r$, we have that $p_0^x \in L_0^{r+1}$. The duplicator places the pebble p_1^x on $\phi_0(p_0^x)$; in particular, $p_1^x \in L_1^{r+1}$.

The proof of (\mathcal{S}_r) follows from the simple observation that $p_0^x \notin S_0^r$ and $p_1^x \notin S_1^r$. As for (\mathcal{E}_r) and (\mathcal{T}_r) , they follow readily from Lemmas 5.10 and 5.11 and the fact that $p_1^x = \phi_0(p_0^x)$ and $p_1^y = \phi_0(p_0^y)$.

5.5. Counting quantifiers. We now consider the natural extension C^2 of FO^2 , where one is allowed to use counting quantifiers of the form $\exists^{\geq i} x$ and $\exists^{\geq i} y$, for $i \in \mathbb{N}$. Such a quantifier, as expected, expresses the existence of at least i elements satisfying the formula which follows it. This logic C^2 has been extensively studied. From an expressiveness standpoint, C^2 strictly extends FO^2 (which cannot express the existence of three elements over the empty vocabulary), and contrary to the latter, C^2 does not enjoy the small model property (meaning that contrary to FO^2 , there exist satisfiable C^2 -sentences which do not have small, or even finite, models). However, the satisfiability problem for C^2 is still decidable [GOR97, Pra07, Pra10]. To the best of our knowledge, it is not known whether $<$ -inv C^2 has a decidable syntax. Let us now explain how the proof of Theorem 5.1 can be adapted to show the following stronger version:

Theorem 5.14. *Let \mathcal{C} be a class of structures of bounded degree on a purely relational vocabulary. Then $<$ -inv $C^2 \subseteq FO$ on \mathcal{C} .*

Proof. The proof is very similar to that of Theorem 5.1. The difference is that we now need to consider C^2 -similarity instead of FO^2 -similarity. To that end, we consider two families of equivalence relations, denoted $\equiv_{k,k}^{C^2}$ and $\equiv_{k,k}^{<\text{-inv } C^2}$ (for $k \in \mathbb{N}$), such that $\mathcal{A}_0 \equiv_{k,k}^{C^2} \mathcal{A}_1$ and $\mathcal{A}_0 \equiv_{k,k}^{<\text{-inv } C^2} \mathcal{A}_1$ iff \mathcal{A}_0 and \mathcal{A}_1 agree on all C^2 -sentences and $<$ -inv C^2 -sentences, respectively, of quantifier rank at most k and with counting indexes at most k .

Let us rephrase Remark 5.3 in the case of C^2 . Assume we manage to show the existence of some function Υ such that for $\mathcal{A}_0, \mathcal{A}_1 \in \mathcal{C}$, one can construct $<$ with

$$(\mathcal{A}_0, <) \equiv_{k,k}^{C^2} (\mathcal{A}_1, <) \quad (5.7)$$

whenever $\llbracket \mathcal{A}_0 \rrbracket_k = \Upsilon(k) \llbracket \mathcal{A}_1 \rrbracket_k$. Once again considering $f : k \mapsto \hat{f}(k, \Upsilon(k))$, Proposition 5.2 guarantees that $\mathcal{A}_0 \equiv_{f(k)}^{FO} \mathcal{A}_1$ implies $\mathcal{A}_0 \equiv_{k,k}^{<\text{-inv } C^2} \mathcal{A}_1$. This amounts to saying that the

equivalence relation $\equiv_{f(k)}^{\text{FO}}$ is more fine-grained than $\equiv_{k,k}^{<\text{inv } \text{C}^2}$ on \mathcal{C} . An $<$ -inv C^2 -sentence of quantifier rank k and with counting indexes at most k can first be decomposed according to the equivalence classes of $\equiv_{k,k}^{<\text{inv } \text{C}^2}$, and then rewritten on \mathcal{C} as a disjunction of FO-sentences of rank $f(k)$ defining $\equiv_{f(k)}^{\text{FO}}$ classes, which concludes the proof.

It only remains to show the existence of such a function Υ and to detail the strategy for the construction of two linear orders satisfying Equation (5.7).

In order to prove Equation (5.7), we need an Ehrenfeucht-Fraïssé game capturing $\equiv_{k,k}^{\text{C}^2}$. It is not hard to derive such a game from the Ehrenfeucht-Fraïssé game for C^2 [IL90]. This game only differs from the two-pebble Ehrenfeucht-Fraïssé game in that in each round, once the spoiler has chosen a structure (say $(\mathcal{A}_0, <)$) and a pebble to move (say p_0^x), the spoiler picks not only one element of that structure, but a set P_0 of up to k elements. Then the duplicator must respond with a set P_1 of same cardinality in $(\mathcal{A}_1, <)$. The spoiler then places p_1^x on any element of P_1 , to which the duplicator responds by placing p_0^x on some element of P_0 . As usual, the spoiler wins after this round if $\text{tp}_{(\mathcal{A}_0, <)}(p_0^x, p_0^y) \neq \text{tp}_{(\mathcal{A}_1, <)}(p_1^x, p_1^y)$. Otherwise, the game goes on until k rounds are played.

It is not hard to establish that this game indeed captures $\equiv_{k,k}^{\text{C}^2}$, in the sense that $(\mathcal{A}_0, <) \equiv_{k,k}^{\text{C}^2} (\mathcal{A}_1, <)$ if and only if the duplicator has a winning strategy for k rounds of this game. The restriction on the cardinality of the set chosen by the spoiler (which is at most k) indeed corresponds to the fact that the counting indexes of the formulae are at most k . As for the number of rounds (namely, k), it corresponds as usual to the quantifier rank. This can be easily derived from a proof of Theorem 5.3 in [IL90], and is left to the reader.

Let us now explain how to modify the construction of $<$ presented in Section 5.3 in order for the duplicator to maintain similarity for k -round in such a game. The only difference lies in the choice of the universal elements. Recall that in the previous construction, we chose, for each k -environment type τ_l extending a frequent k -neighbourhood type and each segment UL_0^j , an element $a[\tau_l]_L^j$ whose k -environment type in $(\mathcal{A}_0, <)$ is destined to be τ_l (and similarly for UR_0^j and $a[\tau_l]_R^j$).

In the new construction, we pick k such elements, instead of just one. Just as previously, all these elements must be far enough from one another in the Gaifman graph of \mathcal{A}_0 . Once again, this condition can be met by virtue of the k -neighbourhood type τ underlying τ_l being frequent, and thus having many occurrences scattered across \mathcal{A}_0 (remember that we have a bound on the degree of \mathcal{A}_0 , thus all the occurrences of τ cannot be concentrated). We only need to multiply the value of m by k in Equation (5.2). This in turn means that Υ must be larger than for the case of FO^2 .

When the spoiler picks a set of elements of size at most k in one of the structures (say P_0 in $(\mathcal{A}_0, <)$), the duplicator responds by selecting, for each one of the elements of P_0 , an element in $(\mathcal{A}_1, <)$ along the strategy for the FO^2 -game explained in Section 5.4.2. All that remains to be shown is that it is possible for the duplicator to answer each element of P_0 with a different element in $(\mathcal{A}_1, <)$.

Note that if the duplicator follows the strategy from Section 5.4.2, they will never answer two moves by the spoiler falling under different cases among Cases (I)-(VI) with the same element. Thus we can treat separately each one of these cases; and for each case, we show that if the spoiler chooses up to k elements in $(\mathcal{A}_0, <)$ falling under this case in P_0 , then the duplicator can find the same number of elements in $(\mathcal{A}_1, <)$, following the aforementioned strategy.

- For Case (I), this is straightforward, since the strategy is based on the isomorphism between the borders of the linear orders. The same goes for Cases (II), (IV) and (VI), as the strategy in these cases also relies on an isomorphism argument.
- Suppose now that $p_0^y \notin L_0^{r+1}$, and assume that the spoiler chooses several elements to the left of p_0^y , but outside of S_0^r and not adjacent to p_0^y . This corresponds to Case (III). Recall that our new construction guarantees, for each k -environment type extending a frequent k -neighbourhood type, the existence in L_1^{r+1} of k elements having this environment. This lets us choose, in L_1^{r+1} , a distinct answer for each element in the set selected by the spoiler, sharing the same k -environment type. Case (V) is obviously symmetric.

This concludes the proof of Theorem 5.14. \square

5.6. Constant symbols. Theorem 5.1 and 5.14 assume the vocabulary is purely relational. We now explain how to lift this constraint.

To extend these theorems to a vocabulary Σ containing constant symbols, it suffices to alter the definition of k -neighborhoods in the proof. Instead of including only elements at distance at most k from a , $\mathcal{N}_A^k(a)$ now also includes the interpretations of all the (finitely many) constant symbols in Σ . The only notable difference in the proof is an increase in the bound on the size of neighborhoods – more precisely, M_k^d has to be increased by the number of constant symbols in Remark 5.5. This increase has no bearing on the proof, as M_k^d is still a function of Σ , k and d .

6. CONCLUSION

In this paper, we made significant progress towards a better understanding of the two-variable fragment of order-invariant first-order logic:

- From a complexity point of view, we established the CONEXPTIME-completeness of the problem of deciding if a given FO^2 -sentence is order-invariant (Theorem 3.7), significantly simplifying and improving the result by Zeume and Harwath [ZH16, Thm. 12].
- From an expressivity point of view, we addressed the question of whether every property definable in order-invariant FO^2 can also be expressed in plain FO. We failed short of fully answering the question, but provided two interesting results. The first one (namely, Theorem 4.6) establishes that under a more relaxed notion of order-invariance, the answer to the above question is “no”. While this does not bring a fully-satisfactory answer to the problem, this leads us to believe that order-invariant FO^2 can indeed express properties beyond the scope of FO. The second one (Theorem 5.1) states that when the degree is bounded, every property expressible in order-invariant FO^2 is definable in FO without the use of the order. This is an important step towards resolving the conjecture that order-invariant FO over classes of structures of bounded degree cannot express properties beyond the reach of FO.

Results of Section 5 also apply to the case of the two-variable logic with counting, C^2 . While order-invariant C^2 has decidable satisfiability and validity problems [CW16, Theorem 6.20], it is open if it has a decidable syntax (*i.e.* whether the problem of determining if a given C^2 -sentence is order-invariant is decidable). Unfortunately the techniques introduced in Section 3 are of no use here, as C^2 lacks the finite model property. Finally, it might be a good idea to study order-invariant FO^2 over graph classes beyond classes of bounded-degree, *e.g.* planar graphs or nowhere-dense classes of graphs.

ACKNOWLEDGEMENTS

Bartosz Bednarczyk was supported by the ERC Consolidator Grant No. 771779 (DeciGUT). He would like to thank Antti Kuusisto and Anna Karykowska for many insightful discussions on the problem. The authors want to thank Thomas Schwentick and the anonymous reviewers for their helpful comments.

REFERENCES

- [Bed22] Bartosz Bednarczyk. Order-Invariance of Two-Variable Logic is coNExpTime-complete. *ArXiv*, 2022.
- [BL16] Pablo Barceló and Leonid Libkin. Order-Invariant Types and Their Applications. *Logical Methods in Computer Science*, 12(1), 2016.
- [CW16] Witold Charatonik and Piotr Witkowski. Two-variable Logic with Counting and a Linear Order. *Logical Methods in Computer Science*, 12(2), 2016.
- [FJY95] Ronald Fagin, Larry J. Stockmeyer, and Moshe Y. Vardi. On Monadic NP vs. Monadic co-NP. *Information and Computation*, 120(1):78–92, 1995.
- [Für83] Martin Fürer. The Computational Complexity of the Unconstrained Limited Domino Problem (with Implications for Logical Decision Problems). In Egon Börger, Gisbert Hasenjaeger, and Dieter Rödding, editors, *Logic and Machines: Decision Problems and Complexity, Proceedings of the Symposium "Rekursive Kombinatorik" held from May 23-28, 1983 at the Institut für Mathematische Logik und Grundlagenforschung der Universität Münster/Westfalen*, volume 171 of *Lecture Notes in Computer Science*, pages 312–319. Springer, 1983.
- [GGY97] Erich Grädel, Phokion G. Kolaitis, and Moshe Y. Vardi. On the Decision Problem for Two-Variable First-Order Logic. *Bulletin of Symbolic Logic*, 3(1):53–69, 1997.
- [GO99] Erich Grädel and Martin Otto. On Logics with Two Variables. *Theoretical Computer Science*, 224(1-2):73–113, 1999.
- [GOR97] Erich Grädel, Martin Otto, and Eric Rosen. Two-Variable Logic with Counting is Decidable. In *Proceedings, 12th Annual IEEE Symposium on Logic in Computer Science, Warsaw, Poland, June 29 - July 2, 1997*, pages 306–317. IEEE Computer Society, 1997.
- [Gra21] Julien Grange. Successor-Invariant First-Order Logic on Classes of Bounded Degree. *Logical Methods in Computer Science*, 17(3), 2021.
- [Gra23] Julien Grange. Order-Invariance in the Two-Variable Fragment of First-Order Logic. In Bartek Klin and Elaine Pimentel, editors, *31st EACSL Annual Conference on Computer Science Logic, CSL 2023, February 13-16, 2023, Warsaw, Poland*, volume 252 of *LIPICs*, pages 23:1–23:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.
- [GS00] Martin Grohe and Thomas Schwentick. Locality of order-invariant first-order formulas. *ACM Transactions on Computational Logic*, 1(1):112–130, 2000.
- [IK89] Neil Immerman and Dexter Kozen. Definability with Bounded Number of Bound Variables. *Information and Computation*, 83(2):121–139, 1989.
- [IL90] Neil Immerman and Eric Lander. *Describing graphs: A first-order approach to graph canonization*, pages 59–81. Springer New York, New York, NY, 1990.
- [Lib04] Leonid Libkin. *Elements of Finite Model Theory*. Springer, 2004.
- [NO10] Jaroslav Nesetril and Patrice Ossona de Mendez. First Order Properties on Nowhere Dense Structures. *Journal of Symbolic Logic*, 75(3):868–887, 2010.
- [Ott01] Martin Otto. Two Variable First-Order Logic over Ordered Domains. *Journal of Symbolic Logic*, 66(2):685–702, 2001.
- [PH23] Ian Pratt-Hartmann. *Fragments of First-Order Logic*. Oxford University Press, 2023.
- [Pot94] Andreas Potthoff. *Logische Klassifizierung regulärer Baumsprachen*. PhD thesis, Christian-Albrechts-Universität Kiel, 1994.
- [Pra07] Ian Pratt-Hartmann. Complexity of the Guarded Two-variable Fragment with Counting Quantifiers. *Journal of Logic and Computation*, 17(1):133–155, 2007.
- [Pra10] Ian Pratt-Hartmann. The Two-Variable Fragment with Counting Revisited. In Anuj Dawar and Ruy J. G. B. de Queiroz, editors, *Logic, Language, Information and Computation, 17th International*

- Workshop, WoLLIC 2010, Brasilia, Brazil, July 6-9, 2010. Proceedings*, volume 6188 of *Lecture Notes in Computer Science*, pages 42–54. Springer, 2010.
- [Sch13] Nicole Schweikardt. A Tutorial on Order- and Arb-Invariant Logics. <https://www.irif.fr/~steiner/jifp/schweikardt.pdf>, 2013.
- [Tra50] Boris A Trakhtenbrot. Impossibility of an Algorithm for the Decision Problem for Finite Classes. In *Doklady Akademii Nauk SSSR*, volume 70, page 569, 1950.
- [Tur38] Alan Mathison Turing. On computable numbers, with an application to the Entscheidungsproblem. A correction. *Proceedings of the London Mathematical Society*, 2(1):544–546, 1938.
- [ZH16] Thomas Zeume and Frederik Harwath. Order-Invariance of Two-Variable Logic is Decidable. In Martin Grohe, Eric Koskinen, and Natarajan Shankar, editors, *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16, New York, NY, USA, July 5-8, 2016*, pages 807–816. ACM, 2016.