

---

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

BARTOSZ BEDNARCZYK<sup>a</sup> AND JULIEN GRANGE<sup>b</sup>

<sup>a</sup> Computational Logic Group, TU Dresden & Institute of Computer Science, University of Wrocław  
e-mail address: bartosz.bednarczyk@cs.uni.wroc.pl

<sup>b</sup> 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. While we were not able to provide a satisfactory answer to the question, 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. On the other hand, 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 on the standpoint of expressive power. And indeed, as long as we consider arbitrary (*i.e.* not necessarily finite) structures it does not, which is a direct

---

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

consequence of FO having the Craig Interpolation Property. However, as it was first shown by Gurevich [Lib04, Thm. 5.3], the claim holds true over finite structures: order-invariant FO is more expressive than plain FO.

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  $\text{FO}^2$ , the two-variable fragment of order-invariant FO. From a complexity point of view, it is known that order-invariant  $\text{FO}^2$  has a  $\text{CONEXPTIME}$ -complete validity problem (which is inherited from  $\text{FO}^2$  with a single linear order, see [Ott01, Thm. 1.2]), and that whether a given  $\text{FO}^2$ -formula is order-invariant is decidable in  $\text{CON2EXPTIME}$  [ZH16, Thm. 12]. From an expressive power point of view, order-invariant  $\text{FO}^2$  is more expressive than plain  $\text{FO}^2$  as it can count globally, *cf.* [ZH16, Example 2]. It remains open [ZH16, Sec. 7], however, whether it is true that every order-invariant  $\text{FO}^2$ -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  $\text{FO}^2$ ; namely, we show that this problem is  $\text{CONEXPTIME}$ -complete. Our proof method relies on establishing an exponential-size counter-model property, and is significantly easier than the proof of [ZH16, Thm. 12].
- We present a class  $\mathcal{C}_{tree}$  of tree-like structures, inspired by [Pot94], and show that there exists an  $\text{FO}^2$ -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  $\text{FO}^2$  lies inside FO is “no”. The problem remains open, though.
- In stark contrast to the previous result, we show that order-invariant  $\text{FO}^2$  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  $\text{FO}^2$ .

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  $\Sigma$  is clear from the context) on a finite signature  $\Sigma$  composed of relation and constant symbols. By  $\text{FO}^2$  we denote the fragment of FO in which the only two variables are  $x$  and  $y$ .

**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  $\varphi$  by replacing each occurrence of the symbol  $R$  with  $S$ . We write  $\varphi(\bar{x})$  to indicate that all the free variables of  $\varphi$  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  $\Delta \subseteq A$ .

**Order-invariance.** A sentence  $\varphi \in \text{FO}^2(\Sigma \cup \{<\})$ , where  $<$  is a binary relation symbol not belonging to  $\Sigma$ , is said to be *order-invariant* if for every finite  $\Sigma$ -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  $\varphi$  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  $<\text{-inv FO}^2$ . 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  $\text{FO}^2$ -formulae was shown to be in  $\text{CON2EXPTIME}$  in [ZH16, Thm. 12].<sup>1</sup>

**Decision problems.** The *finite satisfiability* (resp. *validity*) *problem* for a logic  $\mathcal{L}$  asks whether an input sentence  $\varphi$  from  $\mathcal{L}$  is satisfied in some (resp. every) finite structure. Recall that the finite satisfiability and validity for FO are undecidable [Tur38, Chu36], while for  $\text{FO}^2$  they are respectively  $\text{NEXPTIME}$ -complete and  $\text{CONEXPTIME}$ -complete, cf. [GKV97, Thm. 5.3] and [Für83, Thm. 3]. Note that  $\varphi$  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  $\varphi$  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  $\text{FO}^2 \subseteq <\text{-inv FO}^2$ . 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  $\Sigma$ -structures  $\mathcal{A}_0$  and  $\mathcal{A}_1$ , and  $\mathcal{L}$  being one of FO,  $\text{FO}^2$  and  $<\text{-inv FO}^2$ , 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  $\Sigma$  is a maximal satisfiable set of atoms or negated atoms from  $\Sigma$  with a free variable  $x$ . Similarly, an (atomic) *2-type* over  $\Sigma$  is a maximal satisfiable set of atoms or negated atoms with free variables  $x, y$ . Note that the total number of atomic 1- and 2-types over  $\tau$  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 symbols appearing in  $\varphi$  are respectively denoted  $\alpha_\varphi$  and  $\beta_\varphi$ . Given a structure  $\mathcal{A}$  and an element  $a \in A$  we say that  $a$  *realises* a 1-type  $\alpha$  if  $\alpha$  is the unique 1-type such that  $\mathcal{A} \models \alpha[a]$ . We then write  $\text{tp}_{\mathcal{A}}^0(a)$  to refer to this type. Similarly, for (non-necessarily distinct)  $a, b \in A$ , we denote by  $\text{tp}_{\mathcal{A}}^0(a, b)$  the unique 2-type *realised*

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



- Finally, we reduce  $\varphi[</<_0] \wedge \neg\varphi[</<_1]$  to a Scott-like *normal form*, cf. [GKV97, Sec. 4], [Ott01, Sec. 3.1]. It suffices to apply [ZH16, Lemma 1] (providing such a form for  $FO^2$  with a linear order predicate) to both  $\varphi[</<_0]$  and  $\neg\varphi[</<_1]$  and take their conjunction.

By summarizing all the above steps, we conclude:

**Corollary 3.3.** *For any  $FO^2[<_0, <_1]$ -sentence there is an equi-satisfiable, polynomial-time computable  $FO^2[<_0, <_1]$ -sentence (over a purely relational signature composed of predicates of arity only 1 or 2) having 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  $\chi$  and  $\gamma$  are quantifier-free and the  $<_i$  do not appear in  $\chi_{1-i}$  and  $\gamma_{1-i}^j$ .

Given a model  $\mathcal{A} \models \varphi$  of a formula  $\varphi$  in normal form and elements  $a, b \in A$  such that  $\mathcal{A} \models \gamma_i^j(a, b)$ , we call  $b$  a  $\gamma_i^j$ -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 [GKV97, Thm. 4.3].

**Lemma 3.4.** *Any finitely satisfiable sentence  $\varphi \in FO^2[<_0, <_1]$  in normal form has a model with  $\mathcal{O}(|\varphi|^3 \cdot 2^{|\varphi|})$  elements.*

*Proof.* Let  $M := \max(m_0, m_1)$  and let  $\mathcal{A}$  be a model of  $\varphi$ . We are going to construct from  $\mathcal{A}$  a model  $\mathcal{B} \models \varphi$  whose domain  $B := W_0 \cup W_1 \cup W_2 \cup W_3$  (where the sets  $W_i$  are constructed below) has cardinality at most  $224 |\varphi|^3 \cdot 2^{|\varphi|}$ .

Call a 1-type *rare* if it is realised by at most  $32M$  elements in  $\mathcal{A}$ . Let the set  $S$  be composed of all elements of  $\mathcal{A}$  of rare 1-types, and of the  $8M$  minimal and  $8M$  maximal (w.r.t. each of  $<_0^A, <_1^A$ ) realisations of each non-rare 1-type in  $\mathcal{A}$ . Define  $W_0$  as the set composed of all elements realising rare 1-types, as well as the  $M$  minimal and  $M$  maximal (w.r.t. each of  $<_0^A$  and  $<_1^A$ ) realisations of each non-rare 1-type in  $\mathcal{A}$ . Put the rest of elements of  $S$  to  $W_1$ . We clearly have  $|W_0 \cup W_1| \leq 32M \cdot |\alpha_\varphi|$ . The idea behind  $W_0$  is that this set contains “dangerous” elements, *i.e.* the ones for which  $\mathcal{A}|_{W_0}$  may be uniquely determined by  $\varphi$ . Elements from  $W_1$  will help to restore the satisfaction of  $\forall\exists$  conjuncts. According to the terminology from [GKV97], such elements would be called kings and the royal court.

We next close  $W_0 \cup W_1$  twice under taking witnesses. More precisely, let  $W_2$  be any  $\subseteq$ -minimal subset of  $A$  so that all elements from  $W_0 \cup W_1$  have all the required  $\gamma_i^j$ -witnesses in  $W_0 \cup W_1 \cup W_2$ . Similarly, we define  $W_3$  to be any  $\subseteq$ -minimal subset of  $A$  so that all elements from  $W_0 \cup W_1 \cup W_2$  have all the required  $\gamma_i^j$ -witnesses in  $W_0 \cup W_1 \cup W_2 \cup W_3$ . Observe that:

$$|W_2| \leq 2M|W_0 \cup W_1| \leq 2M \cdot 32M|\alpha_\varphi| = 64M^2|\alpha_\varphi| \text{ and } |W_3| \leq 2M|W_2| \leq 128M^3|\alpha_\varphi|.$$

Consider the structure  $\mathcal{B} := \mathcal{A}|_{W_0 \cup W_1 \cup W_2 \cup W_3}$ . We see that:

$$|B| \leq |W_0 \cup W_1| + |W_2| + |W_3| \leq (32M + 64M^2 + 128M^3)|\alpha_\varphi| \leq 224M^3|\alpha_\varphi| \leq 224 |\varphi|^3 \cdot 2^{|\varphi|}.$$

Note that universal formulae are preserved under substructures, thus  $<_1^B, <_2^B$  are linear orders over  $B$  and  $\mathcal{B}$  satisfies the  $\forall\forall$ -conjuncts of  $\varphi$ . Hence, the only reason for  $\mathcal{B}$  to not be a model of  $\varphi$  is the lack of required  $\gamma_i^j$ -witnesses for elements from the set  $W_3$ . We fix this issue by reinterpreting binary relations between the sets  $W_3$  and  $W_1$ .

Before we start, we are going to collect, for each non-rare 1-type  $\alpha$ , pairwise-disjoint sets of  $M$  minimal and  $M$  maximal (w.r.t. each of  $<_0^A, <_1^A$ ) realisations of  $\alpha$  from  $W_1$ . Formally: Fix a non-rare  $\alpha$ . Let  $V_\alpha^0$  be composed of the first  $M$   $<_0$ -minimal elements from  $\mathcal{A}|_{W_1}$ . Next, let  $V_\alpha^1$  be composed of the last  $M$   $<_0$ -maximal elements from  $\mathcal{A}|_{W_1 \setminus V_\alpha^0}$ . Similarly, let  $V_\alpha^2$  be composed of the first  $M$   $<_1$ -minimal elements from  $\mathcal{A}|_{W_1 \setminus (V_\alpha^0 \cup V_\alpha^1)}$ . Finally let  $V_\alpha^3$  be composed of the last  $M$   $<_1$ -maximal elements from  $\mathcal{A}|_{W_1 \setminus (V_\alpha^0 \cup V_\alpha^1 \cup V_\alpha^2)}$ . Put  $V_\alpha := \bigcup_{k=0}^3 V_\alpha^k$ . Notice that all the components of  $V_\alpha$  are pairwise disjoint (by construction), and they are well-defined since we included sufficiently many elements in  $W_1$ .

Going back to the proof, we fix any element  $a$  from  $W_3$  that violate some of the  $\forall\exists$ -conjuncts of  $\varphi$ , and fix any  $\forall\exists$ -conjunct  $\psi := \forall x \exists y \gamma_i^j(x, y)$  whose satisfaction is violated by  $a$ . Since  $\mathcal{A} \models \varphi$  we know that there is an element  $b \in A$  such that  $b$  is a  $\gamma_i^j$ -witness for  $a$  and  $\gamma_i^j$  in  $\mathcal{A}$  and let  $\alpha$  be the 1-type of  $b$  in  $\mathcal{A}$ . Observe that  $\alpha$  is not rare (otherwise  $b \in W_0$ , and hence  $b \in B$ ), and  $a \neq b$ . Moreover either  $b <_i^A a$  or  $a <_i^A b$  holds. Thus, we take  $V_\alpha^{2i+k}$  (where  $k$  equals 0 if  $b <_i^A a$  and 1 otherwise) to be the corresponding set of  $M$  minimal/maximal  $<_i$  realisations of  $\alpha$  in the same direction to  $a$  as  $b$  is. Now it suffices to take the  $j$ -th element  $b_j$  from  $V_\alpha^{2i+k}$  and change the binary relations between  $a$  and  $b_j$  in  $\mathcal{B}$  so that the equality holds  $\text{tp}_A^0(a, b) = \text{tp}_B^0(a, b_j)$  holds (which can be done as  $b$  and  $b_j$  have the same 1-type). We repeat the process for all remaining  $\gamma_i^j$  formulae violated by  $a$ . We stress that it is not a coincidence that we use the  $j$ -th element  $b_j$  from the corresponding set  $V_\alpha^{2i+k}$  to be a fresh  $\gamma_i^j$ -witness for  $a$ : this guarantees that we never redefine connection between  $a$  and some element twice.

Observe that all elements from  $B$  that had  $\gamma_i^j$ -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  $\varphi$  (since the modified 2-type does not violate  $\varphi$  in  $\mathcal{A}$  it does not violate  $\varphi$  in  $\mathcal{B}$ ) and  $a$  has all required witnesses. By repeating the strategy for all the other elements from  $W_3$  violating  $\varphi$ , we obtain the desired “small” model of  $\varphi$ .  $\square$

Lemma 3.4 yields an NEXPTIME algorithm for deciding satisfiability of  $\text{FO}^2[<_0, <_1]$  formulae: convert an input into 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 3.2 and Corollary 3.1 we conclude:

**Theorem 3.5.** *Checking if an  $\text{FO}^2$ -formula is order-invariant is CONEXPTIME-complete.*

#### 4. CAN ORDER-INVARIANT $\text{FO}^2$ EXPRESS PROPERTIES BEYOND THE SCOPE OF $\text{FO}$ ?

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  $\Sigma$ , is said to be *order-invariant over  $\mathcal{C}$*  if for every finite  $\Sigma$ -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  $\mathcal{C}_{tree}$  over the vocabulary  $\Sigma_{\mathcal{C}_{tree}} := \{T, D, S\}$  of tree-like finite structures, and a sentence  $\varphi \in \text{FO}^2[\Sigma_{\mathcal{C}_{tree}} \cup \{<\}]$  “expressing even depth” that is order-invariant over  $\mathcal{C}_{tree}$  but not equivalent to any first-order sentence over  $\Sigma_{\mathcal{C}_{tree}}$ .

A *dendroid* is a finite  $\Sigma_{\mathcal{C}_{tree}}$ -structure  $\mathcal{A}$  that, intuitively, is a complete directed binary tree decorated with a binary parent-child relation  $T^A$ , a descendant relation  $D^A$ , and a

sibling relation  $S^{\mathcal{A}}$ . Formally, a  $\Sigma_{\mathcal{C}_{tree}}$ -structure  $\mathcal{A}$  is called 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  $v \in A$  the *level* of  $v$ . We also use the terms “root” and “leaf” in the usual way.

**Lemma 4.1.** *If  $\mathcal{A}, \mathcal{B}$  are dendroids of depth  $\geq 2^{q+1}$  then  $\mathcal{A} \equiv_q \mathcal{B}$ .*

*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 4.2.** *There is no  $FO(\Sigma_{\mathcal{C}_{tree}})$ -formula  $\varphi_{even}$  such that for every  $\mathcal{A} \in \mathcal{C}_{tree}$  we have  $\mathcal{A} \models \varphi_{even}$  iff the depth of  $\mathcal{A}$  is even.*

In contrast to the above corollary, we will show that the even depth query can be defined as an  $FO^2(\Sigma_{\mathcal{C}_{tree}} \cup \{<\})$ -formula which is order-invariant over  $\mathcal{C}_{tree}$  (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 the ordered  $\mathcal{T}$  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  $i \geq 1$ . 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 4.3.** 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 (resp. odd) depth then for any linear order  $<$  over its domain the ordered dendroid  $(\mathcal{T}, <)$  has an even (resp. odd) zig-zag.

*Proof.* Immediate by induction after observing that  $\mathcal{A}|_{\{0,1\}^{\leq n}}$ , 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  $FO^2$ -formula. Let us first introduce a few useful macros:

$$\text{ROOT}(x) := \neg \exists y T(y, x) \quad \text{LEAF}(x) := \neg \exists y T(x, y) \quad \text{2nd}(x) := \exists y T(y, x) \wedge \text{ROOT}(y)$$

$$\text{LS}(x) := \exists y S(x, y) \wedge x < y \quad \text{RS}(x) := \exists y S(x, y) \wedge y < x$$

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

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

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 (resp. right) child on the path has a parent which is itself a right (resp. left) child, with the obvious exception of the root and its child. From there, it is easily shown that:

**Proposition 4.4.** *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 4.3 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  $\rho$  is an even zig-zag. By invoking Observation 4.3 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  $\mathcal{C}_{\text{tree}}$ : whether an ordered dendroid has even depth only depends on the underlying dendroid, and not on the particulars of its linear order. Recalling Corollary 4.2, we conclude the following:

**Theorem 4.5.** *There exists a class of finite structures  $\mathcal{C}_{\text{tree}}$  and an  $\text{FO}^2(\Sigma_{\mathcal{C}_{\text{tree}}} \cup \{<\})$ -sentence which is order-invariant over  $\mathcal{C}_{\text{tree}}$ , but is not equivalent to any  $\text{FO}(\Sigma_{\mathcal{C}_{\text{tree}}})$  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  $<\text{-inv FO}^2$  can only express FO-definable properties. Note that although the class  $\mathcal{C}_{\text{tree}}$  from Section 4 contains tree-like structures, the descendant relation makes this a dense class of structures (as it contains cliques of arbitrarily large size), and in particular  $\mathcal{C}_{\text{tree}}$  does not have a bounded degree.

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

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

For the remainder of this section, we fix a signature  $\Sigma$ , an integer  $d$  and a class  $\mathcal{C}$  of  $\Sigma$ -structures of degree at most  $d$ .

Let us now show the skeleton of our proof. The technical part of the proof will be the focus of Sections 5.2 and 5.3. Our general strategy is to show the existence of a function  $f : \mathbb{N} \rightarrow \mathbb{N}$  such that every formula  $\varphi \in <\text{-inv FO}^2$  of quantifier rank  $k$  is equivalent on  $\mathcal{C}$  (i.e. satisfied by the same structures of  $\mathcal{C}$ ) to an FO-formula  $\psi$  of quantifier rank at most  $f(k)$ .



To prove this, it is enough to show 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  $\mathcal{A}_0 \equiv_k^{<\text{inv FO}^2} \mathcal{A}_1$ . Indeed, the class of structures satisfying a formula  $\varphi$  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 in turn the intersection of  $\mathcal{C}$  with a union of equivalence classes for  $\equiv_{f(k)}^{\text{FO}}$ . It is folklore (see, *e.g.*, [Lib04, Cor. 3.16]) that the equivalence relation  $\equiv_{f(k)}^{\text{FO}}$  has finite index, and that each of its equivalence classes is definable by an FO-sentence of quantifier rank  $f(k)$ . Then  $\psi$  is just the finite disjunction of these FO-sentences.

In order to show that  $\mathcal{A}_0 \equiv_k^{<\text{inv FO}^2} \mathcal{A}_1$ , we will construct in Section 5.2 two particular orders  $<_0, <_1$  on these respective structures, and we will prove in Section 5.3 that

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

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

**5.2. 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  $\mathcal{C}$  such that

$$\mathcal{A}_0 \equiv_{f(k)}^{\text{FO}} \mathcal{A}_1, \quad (5.2)$$

we are able to construct two linear orders  $<_0, <_1$  such that  $(\mathcal{A}_0, <_0) \equiv_k^{\text{FO}^2} (\mathcal{A}_1, <_1)$ .

In this section, we define  $f$  and we detail the construction of such orders. The proof of  $<\text{-inv FO}$ -similarity between  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$  will be the focus of Section 5.3.

Let us now explain how we define  $f$ . For that, we need to introduce the notion of neighbourhood and neighbourhood type. These notions are defined in Section 5.2.1. We then explain in Section 5.2.2 how to divide neighbourhood types into rare ones and frequent ones. Finally, the details of the construction are given in Section 5.2.3.

**5.2.1. Neighbourhoods.** Let us now define the notion of neighbourhood of an element in a structure.

Let  $c$  be a new 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  $\Sigma$  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 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  $\tau$  is a  $k$ -neighbourhood type over  $\Sigma$ , and that  $a$  is an occurrence of  $\tau$ . We denote by  $|\mathcal{A}|_{\tau}$  the number of occurrences of  $\tau$  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  $\tau$ ,  $|\mathcal{A}_0|_{\tau}$  and  $|\mathcal{A}_1|_{\tau}$  are either equal, or both larger than  $t$ .

Let  $\text{NEIGHTYPE}_k^d$  denote the set of  $k$ -neighbourhood types over  $\Sigma$  occurring in structures of degree at most  $d$ . Note that  $\text{NEIGHTYPE}_k^d$  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 [FSV95]. 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  $\mathcal{A}_0, \mathcal{A}_1 \in \mathcal{C}$ ,*

$$\mathcal{A}_0 \equiv_{\hat{f}(k, t)}^{FO} \mathcal{A}_1 \quad \rightarrow \quad \llbracket \mathcal{A}_0 \rrbracket_k =^t \llbracket \mathcal{A}_1 \rrbracket_k.$$

We now exhibit a function  $\Theta : \mathbb{N} \rightarrow \mathbb{N}$  such that, if  $\llbracket \mathcal{A}_0 \rrbracket_k =^{\Theta(k)} \llbracket \mathcal{A}_1 \rrbracket_k$ , then one can construct  $<_0, <_1$  satisfying (5.1). Proposition 5.2 then ensures that  $f : k \mapsto \hat{f}(k, \Theta(k))$  fits the bill. Let us now explain how the function  $\Theta$  is chosen.

**5.2.2. Frequency of a neighbourhood type.** Let us denote  $|\text{NEIGHTYPE}_k^d|$  as  $N$ .

Recall that every  $\mathcal{A} \in \mathcal{C}$  has degree at most  $d$ . What this means is that if we consider the set  $\text{FREQ}[\mathcal{A}]_k$  of  $k$ -neighbourhood types that have enough occurrences in  $\mathcal{A}$  (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  $\mathcal{A}$ . Since the degree is bounded,  $N$  is bounded too, which prevents our distinction (which will be formalized later on) between rare neighbourhood types and frequent neighbourhood types from being circular.

Such a dichotomy is introduced and detailed in [Gra21]; we simply adapt this construction to our needs. In the remainder of this section, we describe this construction at a high level, and leave the technical details (such as the exact bounds) to the reader.

The proof of the following lemma (in the vein of [ADG08]) is straightforward, and relies on the degree boundedness hypothesis. Intuitively, Lemma 5.3 states that when the degree is bounded, it is not possible for all the elements of large sets to be concentrated in one corner of the structure, thus making it possible to pick elements in each set that are scattered across the structure.

**Lemma 5.3.** *Given three integers  $m, \delta, s$ , 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  $n \leq N$ ) 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  $(j, i) \neq (j', i')$ .*

Note that the  $N$  in this lemma could be replaced by any constant.

Our goal is, given a structure  $\mathcal{A} \in \mathcal{C}$ , to partition the  $k$ -neighbourhood types into two classes: the frequent types, and the rare types. The property we wish to ensure is that there exist in  $\mathcal{A}$  some number  $m$  (which will be made precise later on, but only depends on  $k$ ) of occurrences of each one of the frequent  $k$ -neighbourhood types which are both

- at distance greater than  $\delta$  (which, as for  $m$ , is a function of  $k$  and will be fixed in the following) from one another, and
- at distance greater than  $\delta$  from every occurrence of a rare  $k$ -neighbourhood type.

To establish this property, we would like to use Lemma 5.3, with  $s$  being the total number of occurrences of all the rare  $k$ -neighbourhood types, and  $C_1, \dots, C_n$  being the sets of occurrences of the  $n$  distinct frequent  $k$ -neighbourhood types.

The number  $N$  of different  $k$ -neighbourhood types of degree at most  $d$  is bounded by a function of  $k$  (as well as  $\Sigma$  and  $d$ , which are fixed). Hence, we can proceed according to the following (terminating) algorithm to make the distinction between frequent and rare types:

- (1) First, let us mark every  $k$ -neighbourhood type as frequent.
- (2) Among the types which are currently marked as frequent, let  $\tau$  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 5.3) 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  $\tau$  as rare, and go back to step 2 if there remains at least one frequent  $k$ -neighbourhood type.

Notice that we can go at most  $N$  times through step 2, where  $N$  depends only on  $k$ . Furthermore, each time we add a type to the set of rare  $k$ -neighbourhood types, we have the guarantee that this type has few occurrences (namely, less than  $g(m, \delta, s)$ , where  $s$  can be bounded by a function of  $k$ ).

It is thus apparent that the threshold  $t$  such that a  $k$ -neighbourhood type  $\tau$  is frequent in  $\mathcal{A}$  iff  $|\mathcal{A}|_\tau \geq t$  can be bounded by some  $T$  depending only on  $k$  - importantly,  $T$  is the same for all structures of  $\mathcal{C}$ .

Let us now make the above more formal. 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  $\mathcal{A}$ . By applying the procedure presented above, we derive the following lemma:

**Lemma 5.4.** *Let  $k, m, \delta \in \mathbb{N}$ . There exists  $T \in \mathbb{N}$  such that for every  $\mathcal{A} \in \mathcal{C}$ , there exists some  $t \leq T$  such that*

$$t \geq g(m, \delta, \sum_{\tau \notin \text{FREQ}[\mathcal{A}]_k^{\geq t}} |\mathcal{A}|_\tau).$$

Let  $\text{FREQ}[\mathcal{A}]_k := \text{FREQ}[\mathcal{A}]_k^{\geq t}$  for the smallest threshold  $t$  given in Lemma 5.4. Some  $k$ -neighbourhood type  $\tau \in \text{NEIGHTYPE}_k^d$  is said to be *frequent* in  $\mathcal{A} \in \mathcal{C}$  if it belongs to  $\text{FREQ}[\mathcal{A}]_k$ ; that is, if  $|\mathcal{A}|_\tau \geq t$ . Otherwise,  $\tau$  is said to be *rare*. With the definition of  $g$  in mind, Lemma 5.4 can then be reformulated as follows: in every structure  $\mathcal{A} \in \mathcal{C}$ , one can find  $m$  occurrences of each frequent  $k$ -neighbourhood type which are at distance greater than  $\delta$  from one another and from the set of occurrences of every rare  $k$ -neighbourhood type.

All that remains is for us to give a value (depending only on  $k$ ) to the integers  $m$  and  $\delta$ : let  $M := \max\{|\tau| : \tau \in \text{NEIGHTYPE}_k^d\}$  ( $M$  indeed exists, and is a function of  $k$  - recall that the signature  $\Sigma$  and the degree  $d$  are assumed to be fixed). Let us consider

$$m := 2 \cdot (k + 1) \cdot M! \quad \text{and} \quad \delta := 4k. \quad (5.3)$$

We then define  $\Theta(k)$  as the integer  $T$  provided by Lemma 5.4 for these values of  $m$  and  $\delta$ . The threshold  $\Theta(k)$  indeed only depends on  $k$ . Finally, notice that if  $\llbracket \mathcal{A}_0 \rrbracket_k =^{\Theta(k)} \llbracket \mathcal{A}_1 \rrbracket_k$ , then  $\text{FREQ}[\mathcal{A}_0]_k = \text{FREQ}[\mathcal{A}_1]_k$ .

As discussed in Section 5.2.1, there exists a function  $f$  such that  $\mathcal{A}_0 \equiv_{f(k)}^{\text{FO}} \mathcal{A}_1$  entails  $\llbracket \mathcal{A}_0 \rrbracket_k =^{\Theta(k)} \llbracket \mathcal{A}_1 \rrbracket_k$ . We also make sure that  $f(k) \geq \Theta(k) \cdot N + 1$  for every  $k$ .

Let us now consider  $\mathcal{A}_0, \mathcal{A}_1 \in \mathcal{C}$  such that  $\mathcal{A}_0 \equiv_{f(k)}^{\text{FO}} \mathcal{A}_1$  for such an  $f$ . If  $\text{FREQ}[\mathcal{A}_0]_k = \emptyset$ , then  $|\mathcal{A}_0| \leq \Theta(k) \cdot N$ . This guarantees that  $\mathcal{A}_0 \simeq \mathcal{A}_1$ , and in particular that  $\mathcal{A}_0 \equiv_k^{<\text{inv FO}^2} \mathcal{A}_1$ . From now on, we suppose that there is at least one frequent  $k$ -neighbourhood type.

The construction of two linear orders  $<_0$  and  $<_1$  satisfying  $(\mathcal{A}_0, <_0) \equiv_k^{\text{FO}^2} (\mathcal{A}_1, <_1)$  is the object of Section 5.2.3.

5.2.3. *Construction of  $<_0$  and  $<_1$ .* This section is dedicated to the definition of two linear orders  $<_0, <_1$  on  $\mathcal{A}_0, \mathcal{A}_1 \in \mathcal{C}$ . We then prove in Section 5.3 that  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$  are FO<sup>2</sup>-similar at depth  $k$ .

Recall that by hypothesis,  $\mathcal{A}_0$  and  $\mathcal{A}_1$  are FO-similar at depth  $f(k)$ , which entails that they have the same number of occurrences of each  $\tau \in \text{NEIGHTYPE}_k^d$  up to a threshold  $\Theta(k)$ .

To construct our 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  $<_0$  as the set of  $k$ -environment types whose underlying  $k$ -neighbourhood type is  $\tau$ .

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

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

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

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

Each segment  $L_i^j$ , for  $j \in \{0, \dots, 2k\}$  is itself decomposed into two segments  $NL_i^j \cdot UL_i^j$ . 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  $UR_i^j \cdot NR_i^j$ , with empty  $UR_i^j$  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.

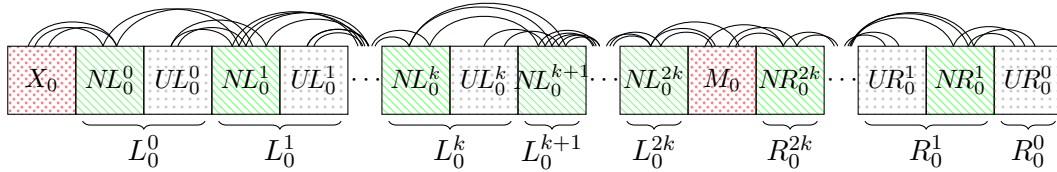


Figure 1: The black curly edges represent the edges between elements belonging to different segments. Edges between elements of the same segment are not represented here. The order  $<_0$  grows from the left to the right.

For every  $\tau \in \text{FREQ}[\mathcal{A}_0]_k$ , let  $\tau_1, \dots, \tau_{|\text{ENV}(\tau)|}$  be an enumeration of  $<_0$ . Recall that we defined  $M$  in Section 5.2.2 as  $\max\{|\tau| : \tau \in \text{NEIGHTYPE}_k^d\}$ . Thus, we have  $|\text{ENV}(\tau)| \leq M!$  for every  $\tau \in \text{NEIGHTYPE}_k^d$ .

In particular, by definition of frequency, and by choice of  $m$  and  $\delta$  in (5.3), Lemma 5.4 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  $\tau$  as  $k$ -neighbourhood type in  $\mathcal{A}_0$ , such that all the  $a[\tau_l]_*$  are at distance at least  $4k + 1$  from each other and from any occurrence of a rare  $k$ -neighbourhood type in  $\mathcal{A}_0$ .

**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, the domain of  $X_0$  is  $\bigcup_{a \in A_0: \text{neigh-tp}_{\mathcal{A}_0}^k(a) \notin \text{FREQ}[\mathcal{A}_0]_k} N_{\mathcal{A}_0}^k(a)$ .

We set  $NL_0^0 := N_{\mathcal{A}_0}(X_0)$  (the set of neighbours of elements of  $X_0$ ), and define the order  $<_0$  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  $a[\tau_l]_L^j$  and of  $\delta = 4k$  in (5.3). Furthermore, the  $N_{\mathcal{A}_0}^k(a[\tau_l]_L^j)$  are pairwise disjoint, hence we can fix  $<_0$  freely and independently on each of them. Unsurprisingly, we order each  $N_{\mathcal{A}_0}^k(a[\tau_l]_L^j)$  so that  $\text{env-tp}_{(\mathcal{A}_0, <_0)}^k(a[\tau_l]_L^j) = \tau_l$ . 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  $\tau_l$ , the linear order  $<_0$  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  $<_0$  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  $\tau_l$ , 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  $M_0$ .**  $M_0$  contains all the elements of  $A_0$  besides those already belonging to  $S_0^{2k}$ . The order  $<_0$  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 (5.2) is large enough so that there exists a set  $S \subseteq A_1$  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  $f(k) \geq |S_0^{2k} \cup N_{\mathcal{A}_0}(S_0^{2k})| + 1$ , which can be bounded by a function of  $k$ , independent of  $\mathcal{A}_0$  and  $\mathcal{A}_1$ ).

Let  $\varphi_0 : \mathcal{A}_0 \upharpoonright_{S_0^{2k}} \rightarrow \mathcal{A}_1 \upharpoonright_S$  be the restriction to  $S_0^{2k}$  of said isomorphism, and let  $\varphi_1$  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  $\varphi_0(a)$  in  $\mathcal{A}_1$ .

We transfer alongside  $\varphi_0$  all the segments, with their order, from  $(\mathcal{A}_0, <_0)$  to  $\mathcal{A}_1$ , thus defining  $X_1, NL_1^j, UL_1^j, \dots$  as the respective images by  $\varphi_0$  of  $X_0, NL_0^j, UL_0^j, \dots$ , and define  $M_1$  as the counterpart to  $M_0$ . 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 5.5.** *For each  $a \in S_0^k$ , we have  $\text{env-tp}_{(\mathcal{A}_0, <_0)}^k(a) = \text{env-tp}_{(\mathcal{A}_1, <_1)}^k(\varphi_0(a))$ .*

Lemma 5.5 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  $<_0$ , for each  $\tau \in \text{FREQ}[\mathcal{A}_0]_k$ .

Our construction also guarantees the following result:

**Lemma 5.6.** *For each  $a, b \in S_0^k$ , we have  $\text{tp}_{(\mathcal{A}_0, <_0)}^0(a, b) = \text{tp}_{(\mathcal{A}_1, <_1)}^0(\varphi_0(a), \varphi_0(b))$ .*

In particular, for  $a = b \in S_0^k$ , we have  $\text{tp}_{(\mathcal{A}_0, <_0)}^0(a) = \text{tp}_{(\mathcal{A}_1, <_1)}^0(\varphi_0(a))$ .

**5.3. Proof of the  $\text{FO}^2$ -similarity of  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$ .** In this section, we aim to show the following result:

**Proposition 5.7.** *We have that  $(\mathcal{A}_0, <_0) \equiv_k^{\text{FO}^2} (\mathcal{A}_1, <_1)$ .*

**5.3.1. The two-pebble Ehrenfeucht-Fraïssé game.** To establish Proposition 5.7, we use Ehrenfeucht-Fraïssé games with two pebbles. These games have been introduced by Immerman and Kozen [IK89]. Let us adapt their definition to our context.

The  $k$ -round two-pebble Ehrenfeucht-Fraïssé game on  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$  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  $(\mathcal{A}_0, <_0)$ , and  $p_1^x$  and  $p_1^y$  on  $(\mathcal{A}_1, <_1)$ . 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  $(\mathcal{A}_0, <_0)$ , and  $p_1^x$  and  $p_1^y$  on elements of  $(\mathcal{A}_1, <_1)$ . The spoiler wins if the duplicator is unable to ensure that  $\text{tp}_{(\mathcal{A}_0, <_0)}^0(p_0^x, p_0^y) = \text{tp}_{(\mathcal{A}_1, <_1)}^0(p_1^x, p_1^y)$ . 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 invariant. 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  $\text{tp}_{(\mathcal{A}_0, <_0)}^0(p_0^x, p_0^y) \neq \text{tp}_{(\mathcal{A}_1, <_1)}^0(p_1^x, p_1^y)$ . 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  $\text{FO}^2$  [IK89]. We will only need the fact that these games are correct:

**Theorem 5.8.** *If the duplicator has a winning strategy in the  $k$ -round two-pebble Ehrenfeucht-Fraïssé game on  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$ , then  $(\mathcal{A}_0, <_0) \equiv_k^{\text{FO}^2} (\mathcal{A}_1, <_1)$ .*

Thus, in order to prove Proposition 5.7, we show that the duplicator wins the  $k$ -round two-pebble Ehrenfeucht-Fraïssé game on  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$ . For that, let us show by a decreasing induction on  $r = k, \dots, 0$  that the duplicator can ensure, after  $k - r$  rounds, that the three following properties (described below) hold:

$$\forall i \in \{0, 1\}, \forall \alpha \in \{x, y\}, p_i^\alpha \in S_i^r \rightarrow p_{1-i}^\alpha = \varphi_i(p_i^\alpha) \quad (S_r)$$

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

$$\text{tp}_{(\mathcal{A}_0, <_0)}^0(p_0^x, p_0^y) = \text{tp}_{(\mathcal{A}_1, <_1)}^0(p_1^x, p_1^y) \quad (T_r)$$

The first property,  $(S_r)$ , 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  $<_i$ -minimal or  $<_i$ -maximal elements, the corresponding pebble in the other structure is located at the same position with respect to this  $<_i$ -extremal element.

As for  $(E_r)$ , 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,  $(T_r)$  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  $(T_r)$  is satisfied at the beginning of the game, and after each of the  $k$  rounds of the game.

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

$$p_1^x = p_1^y = \varphi_0(p_0^x) = \varphi_0(p_0^y).$$

This ensures that  $(S_k)$  holds, while  $(E_k)$  and  $(T_k)$  respectively follow from Lemmas 5.5—5.6.

**5.3.3. Strategy for the duplicator.** We now describe the duplicator's strategy to ensure that  $(S_r)$ ,  $(E_r)$  and  $(T_r)$  hold no matter how the spoiler plays.

Suppose that we have  $(S_{r+1})$ ,  $(E_{r+1})$  and  $(T_{r+1})$  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  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$ , the spoiler moves  $p_0^x$  in  $(\mathcal{A}_0, <_0)$ . Let us first explain informally the general idea behind the duplicator's strategy.

- (1) If the spoiler plays around the endpoints (by which we mean the elements that are  $<_i$ -minimal and maximal), the duplicator has no choice but to play a tit-for-tat strategy, *i.e.* to respond to the placement of  $p_i^\alpha$  near the endpoints by moving  $p_{1-i}^\alpha$  on  $\varphi_i(p_i^\alpha)$ .

If the duplicator does not respond this way, then the spoiler will be able to expose the difference between  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$  in the subsequent moves, by forcing the duplicator to play closer and closer to the endpoint, which will prove to be impossible at some point.

On top of that, the occurrences of rare neighbourhood types are located in  $(\mathcal{A}_i, <_i)$  near the  $<_i$ -minimal element. If the duplicator does not play according to  $\varphi_0$  in this area, it will be easy enough for the spoiler to win the game.

The reason we introduced the segments  $NL_i^j, UL_i^j, NR_i^j$  and  $UR_i^j$  is precisely to bound the area in which the duplicator must implement the tit-for-tat strategy. Indeed, as soon as a pebble is placed in  $M_i$ , there is no way for the spoiler to join the endpoints in less than  $k$  moves while forcing the duplicator's hand.

The case where the spoiler plays near the endpoints corresponds to Case ((I)) below, and is detailed in Section 5.3.4.

- (2) Next, suppose that the spoiler places a pebble, say  $p_0^x$ , next (in  $\mathcal{A}_0$ ) to  $p_0^y$ , *i.e.* such that  $p_0^x \in N_{\mathcal{A}_0}^1(p_0^y)$ . The duplicator must place  $p_1^x$  on an element whose relative position to  $p_1^y$  is the same as the relative position of  $p_0^x$  with respect to  $p_0^y$ . Note that once this is done, the spoiler can change variable, and place  $p_0^y$  (or  $p_1^y$ , if they decide to play in  $(\mathcal{A}_1, <_1)$ ) in  $N_{\mathcal{A}_0}^1(p_0^x)$ , thus forcing the duplicator to play near  $p_1^x$ . In order to prevent the spoiler from being able, in  $k$  such moves, to expose the difference between  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$ , the duplicator must make sure, with  $r$  rounds left, that  $p_0^x$  and  $p_1^x$  (as well as  $p_0^y$  and  $p_1^y$ ) share the same  $r$ -environment in  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$ . This will guarantee that the duplicator can play along if the spoiler decides to take  $r$  moves adjacent (in  $\mathcal{A}_i$ ) to one another.

The case where the spoiler places a pebble next (in the structure without ordering) to the other pebble is our Case ((II)), and is treated in Section 5.3.5.

- (3) Suppose now that the spoiler's move does not fall under the previous templates. Let us assume that the spoiler plays in  $(\mathcal{A}_0, <_0)$ , and moves  $p_0^x$  to the left of  $p_0^y$  (*i.e.* such that  $(\mathcal{A}_0, <_0) \models p_0^x < p_0^y$ ).

In order to play according to the remarks from Cases 1 and 2, the duplicator must place  $p_1^x$  on an element which shares the same  $r$ -environment with  $p_0^x$  (where  $r$  is the number of rounds left in the game), which is not near the endpoints.

It must be the case that the  $k$ -neighbourhood type of  $p_0^x$  in  $\mathcal{A}_0$  is frequent, since it is not near the endpoints of  $(\mathcal{A}_0, <_0)$ , hence not in  $X_0$ . By construction, every universal segment  $UL_1^j$ , for  $0 \leq j \leq k$ , contains elements of each  $k$ -environment type extending any frequent  $k$ -neighbourhood type. In particular, it contains an element having the same  $r$ -environment as  $p_0^x$ . The duplicator will place  $p_1^x$  on such an element in the leftmost segment  $UL_1^j$  which is not considered to be near the endpoints (this notion depends on the number  $r$  of rounds left in the game). This is detailed in Cases ((III)) and ((V)) (for the symmetrical case where  $p_0^x$  is placed to the right of  $p_0^y$ ) below.

However, we have to consider a subcase, where  $p_1^y$  is itself in the leftmost segment  $L_1^j$  which is not near the endpoints. Indeed, in this case, placing  $p_1^x$  as discussed may result in  $p_1^x$  being to the right of  $p_1^y$ , or being in  $N_{\mathcal{A}_1}^1(p_1^y)$ ; either of which being game-losing to the duplicator. However, since  $p_1^y$  was considered to be near the endpoints in the



previous round of the game, we know that the duplicator played a tit-for-tat strategy at that point, which allows us to replicate the placement of  $p_0^x$  according to  $\varphi_0$ . This subcase, as well as the equivalent subcase where the spoiler places  $p_0^x$  to the right of  $p_0^y$ , are formalized in Cases ((IV)) and ((VI)) below.

We are now ready to describe formally the strategy implemented by the duplicator:

- (I) If  $p_0^x \in S_0^r$ , then the duplicator responds by placing  $p_1^x$  on  $\varphi_0(p_0^x)$ .  
This corresponds to the tit-for-tat strategy implemented when the spoiler plays near the endpoints, as discussed in Case 1.
- (II) Else, if  $p_0^x \notin S_0^r$ , and  $p_0^x \in N_{\mathcal{A}_0}^1(p_0^y)$ , then  $(E_{r+1})$  ensures that there exists an isomorphism  $\psi : \mathcal{Env}_{(\mathcal{A}_0, <_0)}^{r+1}(p_0^y) \rightarrow \mathcal{Env}_{(\mathcal{A}_1, <_1)}^{r+1}(p_1^y)$ . The duplicator responds by placing  $p_1^x$  on  $\psi(p_0^x)$ .  
This makes formal the duplicator's response to a move next to the other pebble, as discussed in Case 2 above.
- (III) Else suppose that  $(\mathcal{A}_0, <_0) \models p_0^x < p_0^y$  and  $p_0^y \notin L_0^{r+1}$ . Note that  $\tau := \text{neigh-tp}_{\mathcal{A}_0}^k(p_0^x) \in \text{FREQ}[\mathcal{A}_0]_k$ , since  $p_0^x \notin X_0$ . Let  $\tau_l := \text{env-tp}_{(\mathcal{A}_0, <_0)}^k(p_0^x)$ .  
The duplicator responds by placing  $p_1^x$  on  $\varphi_0(a[\tau_l]_L^{r+1})$ .
- (IV) Else, if  $(\mathcal{A}_0, <_0) \models p_0^x < p_0^y$  and  $p_0^y \in L_0^{r+1}$ , then the duplicator moves  $p_1^x$  on  $\varphi_0(p_0^x)$  (by  $(S_{r+1})$ ,  $p_0^x$  indeed belongs to the domain of  $\varphi_0$ ).
- (V) Else, suppose that  $(\mathcal{A}_0, <_0) \models p_0^y < p_0^x$  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  $\varphi_0(a[\tau_l]_R^{r+1})$ , where  $\tau_l := \text{env-tp}_{(\mathcal{A}_0, <_0)}^k(p_0^x)$ .
- (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  $\varphi_0(p_0^x)$ .

It remains to show that this strategy satisfies our invariants: under the inductive assumption that  $(S_{r+1})$ ,  $(E_{r+1})$  and  $(T_{r+1})$  hold, for some  $0 \leq r < k$ , we need to show that this strategy ensures that  $(S_r)$ ,  $(E_r)$  and  $(T_r)$  hold.

We treat each case in its own section: Section 5.3.4 is devoted to Case ((I)) while Section 5.3.5 covers Case ((II)). Both Cases ((III)) and ((IV)) are treated in Section 5.3.6. Cases ((V)) and ((VI)), being their exact symmetric counterparts, are left to the reader.

**Remark 5.9.** Note that some properties need no verification. Since  $p_0^y$  and  $p_1^y$  are left untouched by the players,  $(S_{r+1})$  ensures that half of  $(S_r)$  automatically holds, namely that

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

Similarly, the part of  $(E_r)$  concerning  $p_0^y$  and  $p_1^y$  follows from  $(E_{r+1})$ :

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

Lastly, notice that once we have shown that  $(E_r)$  holds, it follows that

$$\begin{cases} \text{tp}_{\mathcal{A}_0}^0(p_0^x) = \text{tp}_{\mathcal{A}_1}^0(p_1^x) \\ \text{tp}_{\mathcal{A}_0}^0(p_0^y) = \text{tp}_{\mathcal{A}_1}^0(p_1^y) \end{cases}$$

5.3.4. *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  $<_0$ -minimal or  $<_0$ -maximal element of  $(\mathcal{A}_0, <_0)$ . 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  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$ .

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  $\varphi_0(p_0^x)$ .

Let us now prove that this strategy guarantees that  $(S_r)$ ,  $(E_r)$  and  $(T_r)$  hold. Recall from Note 5.9 that part of the task is already taken care of.

**Proof of  $(S_r)$  in Case ((I)).** We have to show that  $\forall i \in \{0, 1\}, p_i^x \in S_i^r \rightarrow p_{1-i}^x = \varphi_i(p_i^x)$ . This follows directly from the duplicator’s strategy, since  $p_1^x = \varphi_0(p_0^x)$  (thus  $p_0^x = \varphi_1(p_1^x)$ ).

**Proof of  $(E_r)$  in Case ((I)).** We need to prove that  $\text{env-tp}_{(\mathcal{A}_0, <_0)}^r(p_0^x) = \text{env-tp}_{(\mathcal{A}_1, <_1)}^r(p_1^x)$ , which is a consequence of Lemma 5.5 given that  $p_1^x = \varphi_0(p_0^x)$  and  $r < k$ .

**Proof of  $(T_r)$  in Case ((I)).** First, suppose that  $p_0^y \in S_0^{r+1}$ . By  $(S_{r+1})$ , we know that  $p_1^y = \varphi_0(p_0^y)$ . Thus, Lemma 5.6 allows us to conclude that  $\text{tp}_{(\mathcal{A}_0, <_0)}^0(p_0^x, p_0^y) = \text{tp}_{(\mathcal{A}_1, <_1)}^0(p_1^x, p_1^y)$ .

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

$$\text{tp}_{\mathcal{A}_0}^0(p_0^x, p_0^y) = \text{tp}_{\mathcal{A}_1}^0(p_1^x, p_1^y) \quad (5.4)$$

$$\text{tp}_{<_0}^0(p_0^x, p_0^y) = \text{tp}_{<_1}^0(p_1^x, p_1^y) \quad (5.5)$$

Notice that

$$\begin{cases} \text{tp}_{\mathcal{A}_0}^0(p_0^x, p_0^y) = \text{tp}_{\mathcal{A}_0}^0(p_0^x) \cup \text{tp}_{\mathcal{A}_0}^0(p_0^y) \\ \text{tp}_{\mathcal{A}_1}^0(p_1^x, p_1^y) = \text{tp}_{\mathcal{A}_1}^0(p_1^x) \cup \text{tp}_{\mathcal{A}_1}^0(p_1^y) \end{cases}$$

This is because, by construction, the neighbours in  $\mathcal{A}_i$  of an element of  $S_i^r$  all belong to  $S_i^{r+1}$ . Equation (5.4) follows from this remark and Note 5.9.

As for Equation (5.5), 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  $\text{tp}_{<_0}^0(p_0^x, p_0^y) = \{x < y\} = \text{tp}_{<_1}^0(p_1^x, p_1^y)$ , 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  $\text{tp}_{<_0}^0(p_0^x, p_0^y) = \{x > y\} = \text{tp}_{<_1}^0(p_1^x, p_1^y)$ .

5.3.5. *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  $(E_{r+1})$ , which guarantees that  $\text{env-tp}_{(\mathcal{A}_0, <_0)}^{r+1}(p_0^y) = \text{env-tp}_{(\mathcal{A}_1, <_1)}^{r+1}(p_1^y)$ . Thus there exists an isomorphism  $\psi$  between  $\mathcal{Env}_{(\mathcal{A}_0, <_0)}^{r+1}(p_0^y)$  and  $\mathcal{Env}_{(\mathcal{A}_1, <_1)}^{r+1}(p_1^y)$ . Note that this isomorphism is unique, by virtue of  $<_0$  and  $<_1$  being 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  $(S_r)$ ,  $(E_r)$  and  $(T_r)$ .

**Proof of  $(S_r)$  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: suppose that  $p_1^x \in S_1^r$ . By construction, since  $p_1^x$  and  $p_1^y$  are neighbours in  $\mathcal{A}_1$ , this entails that  $p_1^y \in S_1^{r+1}$ . But then, we know by  $(S_{r+1})$  that  $p_0^y = \varphi_1(p_1^y)$ ; and because  $\psi$  is the unique isomorphism between  $\mathcal{Env}_{(\mathcal{A}_0, <_0)}^{r+1}(p_0^y)$  and  $\mathcal{Env}_{(\mathcal{A}_1, <_1)}^{r+1}(p_1^y)$ ,  $\psi$  is equal to the restriction  $\widetilde{\varphi}_0$  of  $\varphi_0$ :

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

Thus  $p_0^x = \psi^{-1}(p_1^x) = \widetilde{\varphi}_0^{-1}(p_1^x) = \varphi_1(p_1^x)$ , and by definition of the segments on  $(\mathcal{A}_1, <_1)$ , which are just a transposition of the segments of  $(\mathcal{A}_0, <_0)$  via  $\varphi_0$ ,  $p_1^x \in S_1^r$  then entails that  $p_0^x \in S_0^r$ , which is clearly a contradiction.

Since we neither have  $p_0^x \in S_0^r$  nor  $p_1^x \in S_1^r$ ,  $(S_r)$  holds - recall from Note 5.9 that the part concerning  $p_0^y$  and  $p_1^y$  is always satisfied.

**Proof of  $(E_r)$  in Case ((II)).** Recall that the duplicator placed  $p_1^x$  on the image of  $p_0^x$  by the isomorphism

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

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

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

**Proof of  $(T_r)$  in Case ((II)).** This follows immediately from the fact that the isomorphism  $\psi$  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.

5.3.6. *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  $(\mathcal{A}_0, <_0) \models p_0^x < p_0^y$ ), which do not already fall in Cases ((I)) and ((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  $\tau_l := \text{env-tp}_{(\mathcal{A}_0, <_0)}^k(p_0^x)$  (which must extend a frequent  $k$ -neighbourhood type, since  $p_0^x \notin X_0$ ), then  $\varphi_0(a[\tau_l]_L^{r+1})$  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  $(\mathcal{A}_1, <_1) \models p_1^x > p_1^y$ ). 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 orthogonal in  $\mathcal{A}_0$  (*i.e.*  $\text{tp}_{\mathcal{A}_0}^0(p_0^x, p_0^y) = \text{tp}_{\mathcal{A}_0}^0(p_0^x) \cup \text{tp}_{\mathcal{A}_0}^0(p_0^y)$ ), would break  $(T_r)$ .

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  $\varphi_0(a[\tau_l]_L^{r+1})$ . This corresponds to Case ((III)).
- if  $p_1^y \in L_1^{r+1}$ , then  $(S_{r+1})$  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  $\varphi_0$ : the duplicator moves  $p_1^x$  to  $\varphi_0(p_0^x)$ . This situation corresponds to Case ((IV)).

Let us prove that  $(S_r)$ ,  $(E_r)$  and  $(T_r)$  hold in both of these instances.

**Proof of  $(S_r)$  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  $\varphi_0(a[\tau_l]_L^{r+1}) \in L_1^{r+1}$ , and  $p_1^x \notin S_1^r$ .

**Proof of  $(E_r)$  in Case ((III)).** It follows from  $\text{env-tp}_{(\mathcal{A}_0, <_0)}^k(a[\tau_l]_L^{r+1}) = \tau_l$  together with Lemma 5.5 that

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

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

**Proof of  $(T_r)$  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,

$$\text{tp}_{\mathcal{A}_0}^0(p_0^x, p_0^y) = \text{tp}_{\mathcal{A}_0}^0(p_0^x) \cup \text{tp}_{\mathcal{A}_0}^0(p_0^y).$$

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,  $N_{\mathcal{A}_1}^1(p_1^x) = N_{\mathcal{A}_1}^1(\varphi_0(a[\tau_l]_L^{r+1})) \subseteq UL_1^{r+1}$ . By assumption,  $p_1^y \notin L_1^{r+1}$ , which entails that  $\text{tp}_{\mathcal{A}_1}^0(p_1^x, p_1^y) = \text{tp}_{\mathcal{A}_1}^0(p_1^x) \cup \text{tp}_{\mathcal{A}_1}^0(p_1^y)$ .

It then follows from the last observation of Note 5.9 that  $\text{tp}_{\mathcal{A}_0}^0(p_0^x, p_0^y) = \text{tp}_{\mathcal{A}_1}^0(p_1^x, p_1^y)$ .

Let us now prove that  $\text{tp}_{<_1}^0(p_1^x, p_1^y) = \{x < y\}$ .

We claim that  $p_1^y \notin X_1 \cup \bigcup_{0 \leq j \leq r+1} L_1^j$ . Suppose otherwise:  $(S_{r+1})$  would entail that  $p_0^y \in X_0 \cup \bigcup_{0 \leq j \leq r+1} L_0^j$  which, together with the hypothesis  $p_0^y \notin L_0^{r+1}$  and  $p_0^x < p_0^y$ , would result in  $p_0^x$  being in  $S_0^r$ , which is absurd.

Thus,  $\text{tp}_{<_1}^0(p_1^x, p_1^y) = \{x < y\} = \text{tp}_{<_0}^0(p_0^x, p_0^y)$ , which concludes the proof of  $(T_r)$ .

**Proof of  $(S_r)$ ,  $(E_r)$  and  $(T_r)$  in Case ((IV)).** Let us now move to the case where  $p_1^y \in L_1^{r+1}$ . Recall that under this assumption,  $p_0^y = \varphi_1(p_1^y) \in L_0^{r+1}$  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  $\varphi_0(p_0^x)$ ; in particular,  $p_1^x \in L_1^{r+1}$ .

The proof of  $(S_r)$  follows from the simple observation that  $p_0^x \notin S_0^r$  and  $p_1^x \notin S_1^r$ .

As for  $(E_r)$  and  $(T_r)$ , they follow readily from Lemma 5.5 and 5.6 and the fact that  $p_1^x = \varphi_0(p_0^x)$  and  $p_1^y = \varphi_0(p_0^y)$ .

**5.4. 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. On an expressiveness standpoint,  $C^2$  strictly extends  $FO^2$  (which cannot count up to three), 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 5.10.** *Let  $\mathcal{C}$  be a class of structures of bounded degree.  
Then  $<$ -inv  $C^2 \subseteq FO$  on  $\mathcal{C}$ .*

*Proof.* The proof is very similar as to that of Theorem 5.1. The difference is that we now need to show, at the end of the construction, that the structures  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$  are not only  $FO^2$ -similar, but  $C^2$ -similar. More precisely, we show that for every  $k \in \mathbb{N}$ , there exists some  $f(k) \in \mathbb{N}$  such that if  $\mathcal{A}_0 \equiv_{f(k)}^{FO} \mathcal{A}_1$ , then it is possible to construct two linear orders  $<_0$  and  $<_1$  such that  $(\mathcal{A}_0, <_0)$  and  $(\mathcal{A}_1, <_1)$  agree on all  $C^2$ -sentences of quantifier rank at most  $k$ , and with counting indexes at most  $k$ , which we denote  $(\mathcal{A}_0, <_0) \equiv_{k,k}^{C^2} (\mathcal{A}_1, <_1)$ . This is enough to complete the proof, as these classes of  $C^2$ -sentences cover all the  $C^2$ -definable properties.

In order to prove that  $(\mathcal{A}_0, <_0) \equiv_{k,k}^{C^2} (\mathcal{A}_1, <_1)$ , we need an Ehrenfeucht-Fraïssé-game capturing  $\equiv_{k,k}^{C^2}$ . 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  $(\mathcal{A}_0, <_0)$ ) 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  $(\mathcal{A}_1, <_1)$ . 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  $tp_{(\mathcal{A}_0, <_0)}^0(p_0^x, p_0^y) \neq tp_{(\mathcal{A}_1, <_1)}^0(p_1^x, p_1^y)$ . Otherwise, the game goes on until  $k$  rounds are played.

It is not hard to establish that this game indeed captures  $\equiv_{k,k}^{C^2}$ , in the sense that  $(\mathcal{A}_0, <_0) \equiv_{k,k}^{C^2} (\mathcal{A}_1, <_1)$  if and only if the duplicator has a winning strategy for  $k$  rounds of this game. The restriction on the cardinal 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  $<_0$  and  $<_1$  presented in Section 5.2 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  $\tau_l$  extending a frequent  $k$ -neighbourhood type and each segment  $UL_0^j$ , an element  $a[\tau_l]_L^j$  whose  $k$ -environment type in  $(\mathcal{A}_0, <_0)$  is destined to be  $\tau_l$  (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  $\tau$  underlying  $\tau_l$  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  $\tau$  cannot be concentrated). We only need to multiply the value of  $m$  by  $k$  in (5.3).

When the spoiler picks a set of elements of size at most  $k$  in one of the structures (say  $P_0$  in  $(\mathcal{A}_0, <_0)$ ), the duplicator responds by selecting, for each one of the elements of  $P_0$ , an element in  $(\mathcal{A}_1, <_1)$  along the strategy for the  $\text{FO}^2$ -game explained in Section 5.3.3. 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  $(\mathcal{A}_1, <_1)$ .

Note that if the duplicator follows the strategy from Section 5.3.3, 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  $(\mathcal{A}_0, <_0)$  falling under this case in  $P_0$ , then the duplicator can find the same number of elements in  $(\mathcal{A}_1, <_1)$ , 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 5.10.  $\square$

## 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  $\text{CONEXPTIME}$ -completeness of the problem of deciding if a given  $\text{FO}^2$ -sentence is order-invariant (Theorem 3.5), 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  $\text{FO}^2$  can also be expressed in plain  $\text{FO}$ . We failed short of fully answering the question, but provided two interesting results. The first one (namely, Theorem 4.5) 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  $\text{FO}^2$  can indeed express properties beyond the scope of  $\text{FO}$ . The second one (Theorem 5.1) states that when the degree is bounded, every property expressible in order-invariant  $\text{FO}^2$  is definable in  $\text{FO}$  without the use of the order. This is an important step towards resolving the conjecture that order-invariant  $\text{FO}$  over classes of structures of bounded degree cannot express properties beyond the reach of  $\text{FO}$ .

Results of Section 5 also apply to the case of the two-variable logic with counting,  $\text{C}^2$ . While order-invariant  $\text{C}^2$  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  $\text{C}^2$ -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.

#### REFERENCES

- [ADG08] Albert Atserias, Anuj Dawar, and Martin Grohe. Preservation under extensions on well-behaved finite structures. *SIAM Journal on Computing*, 2008.
- [Bed22] Bartosz Bednarczyk. Order-invariance of two-variable logic is conexptime-complete. *CoRR*, 2022.
- [BL16] Pablo Barceló and Leonid Libkin. Order-Invariant Types and Their Applications. *Log. Methods Comput. Sci.*, 12(1), 2016.
- [Chu36] Alonzo Church. A note on the Entscheidungsproblem. *The journal of symbolic logic*, 1(1):40–41, 1936.
- [CW16] Witold Charatonik and Piotr Witkowski. Two-variable logic with counting and a linear order. *Log. Methods Comput. Sci.*, 12(2), 2016.
- [FSV95] Ronald Fagin, Larry J. Stockmeyer, and Moshe Y. Vardi. On monadic NP vs. monadic co-NP. *Inf. Comput.*, 1995.
- [Für83] Martin Fürer. The Computational Complexity of the Unconstrained Limited Domino Problem (with Implications for Logical Decision Problems). *Logic and Machines: Decision Problems and Complexity*, 1983.
- [GKV97] Erich Grädel, Phokion G. Kolaitis, and Moshe Y. Vardi. On the Decision Problem for Two-Variable First-Order Logic. *Bull. Symb. Log.*, 1997.
- [GO99] Erich Grädel and Martin Otto. On Logics with Two Variables. *Theor. Comput. Sci.*, 224(1-2):73–113, 1999.
- [GOR97] Erich Grädel, Martin Otto, and Eric Rosen. Two-variable logic with counting is decidable. In *LICS*, 1997.
- [Gra21] Julien Grange. Successor-invariant first-order logic on classes of bounded degree. *Log. Methods Comput. Sci.*, 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 Trans. Comput. Log.*, 1(1):112–130, 2000.
- [IK89] Neil Immerman and Dexter Kozen. Definability with bounded number of bound variables. *Inf. Comput.*, 1989.
- [IL90] Neil Immerman and Eric Lander. Describing graphs: A first-order approach to graph canonization. In *Complexity theory retrospective*. Springer, 1990.
- [Lib04] Leonid Libkin. *Elements of Finite Model Theory*. Springer, 2004.
- [Ott01] Martin Otto. Two Variable First-Order Logic over Ordered Domains. *J. Symb. Log.*, 66(2):685–702, 2001.
- [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. *J. Log. Comput.*, 2007.
- [Pra10] Ian Pratt-Hartmann. The two-variable fragment with counting revisited. In *WoLLIC*, 2010.

- [Sch13] Nicole Schweikardt. A Tutorial on Order- and Arb-Invariant Logics. <https://www.irif.fr/~steiner/jifp/schweikardt.pdf>, 2013.
- [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. *LICS*, 2016.