How many variables are needed to express an existential positive query?

Simone Bova[1] **and Hubie Chen**[2]

1 **TU Wien (Austria)**
`simone.bova@ac.tuwien.ac.at`
2 **Birkbeck, University of London**

Abstract

The number of variables used by a first-order query is a fundamental measure which has been studied in numerous contexts, and which is known to be highly relevant to the task of query evaluation. In this article, we study this measure in the context of existential positive queries. Building on previous work, we present a combinatorial quantity defined on existential positive queries; we show that this quantity not only characterizes the minimum number of variables needed to express a given existential positive query by another existential positive query, but also that it characterizes the minimum number of variables needed to express a given existential positive query, over all first-order queries. Put differently and loosely, we show that for any existential positive query, no variables can ever be saved by moving out of existential positive logic to first-order logic. One component of this theorem's proof is the construction of a winning strategy for a certain Ehrenfeucht-Fraïssé type game.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes; F.4.1 Mathematical Logic; G.2.2 Graph Theory; H.2.1 Logical Design

Keywords and phrases Existential positive queries, finite-variable logics, first-order logic, query optimization

Digital Object Identifier 10.4230/LIPIcs...

1 Introduction

Background. The number of variables used by a first-order query is recognized as a highly useful and fundamental measure, and has been studied in numerous settings, including descriptive complexity, finite model theory, and query evaluation. By the number of variables used by a query, we refer to the total number of variables that appear in the query. Note that this measure is, in essence, equivalent to the *width* of a query, which is defined as the maximum number of free variables over all subformulas of the query: a query having width k can be rewritten, just by syntactically renaming variables, as a query using k variables; and, a query using k variables clearly has width at most k. Within this article, all queries dealt with are relational and first-order.

In the setting of query evaluation, the number of variables is a measure of prime and crucial interest. A first reason for this is that the natural bottom-up algorithm for evaluating a first-order query on a finite structure exhibits, in general, an exponential dependence on the number of variables; it also runs in polynomial-time when a constant bound is placed on the number of variables [22]. Furthermore, there are complexity classification theorems [18, 15, 10, 11] on classes of Boolean queries in which the number of variables emerges as the decisive measure for describing whether or not a class of Boolean queries enjoys tractable query evaluation; in particular, these classification theorems show that, if a class of queries enjoys tractable query evaluation at all, then there exists a constant $k \geq 1$ such that each query in the class can be expressed by (that is, is logically equivalent to) a query using at most k variables. Let us remark that these classification theorems are given in a parameterized complexity setting in which a query can be preprocessed independently of the structure on which it is to be evaluated; and, that these theorems concern classes of queries having bounded arity.[1] (We refer the reader to the cited articles for precise theorem statements and more information.)

Given the computational relevance of *the number of variables* as a query measure, it is natural to inquire, given a query, to what extent the number of variables can be minimized; indeed, it is a natural desire to attempt to rewrite/optimize a given query as one that uses the fewest number of variables (and which retains logical equivalence to the original query). In this article, we study this question on existential positive queries. They include and are semantically equivalent to the so-called *unions of conjunctive queries*, also known as *select-project-join-union queries*; these have been argued to be the most common database queries [1]. Previous work [6] due to the present authors yields a combinatorial characterization (Theorem 21) of the minimum number of variables needed to express a given Boolean existential positive query, *by another existential positive query*. Let FO denote the class of first-order queries, let EP denote the class of existential positive queries, and let FO^k and EP^k denote the restrictions of these classes to queries using at most k variables, respectively; say that a query ϕ is *L-expressible* if there exists a query $\psi \in L$ that is logically equivalent to ϕ. Rephrasing, the combinatorial characterization yields, given a Boolean existential positive query ϕ, the minimum value k such that ϕ is EP^k-expressible. This characterization thus indicates how to minimize the number of variables *within* the class of existential positive queries. However, this characterization does not preclude the possibility that a query requiring k variables to be expressed as an existential positive query, could be expressed by a first-order query that uses strictly fewer than k variables.

Contributions. We prove that the just-mentioned possibility can never occur. We generalize the aforementioned combinatorial quantity so that it is defined on all existential positive queries (both Boolean and non-Boolean), and dub this quantity the *combinatorial width*. Our primary theorem states that, for *any* existential positive query ϕ, when k is set equal to the combinatorial width,

- ϕ is can be expressed by an existential positive query using k variables, but
- ϕ cannot be expressed by any first-order query using a number of variables that is strictly less than k.

That is, the combinatorial width not only gives the minimum value k such that ϕ is EP^k-expressible, it in fact more sharply gives the minimum value k such that ϕ is FO^k-expressible. This theorem can be viewed as a collapse result, namely, that for existential positive queries, FO^k-expressibility implies (and hence coincides with) EP^k-expressibility. We want to emphasize that the theorem applies individually to every single existential positive query; in our view, the theorem contains a certain element of surprise, since it states (essentially) that there is *no existential positive query whatsoever* for which one can save variables by moving out of existential positive logic to the more general first-order logic.

One corollary of our development is that deciding FO^k-expressibility of existential positive sentences is complete for the class Π_2^p of the polynomial hierarchy (Corollary 25); this follows

[1] A class of queries has *bounded arity* if there is a constant upper bound on the arity of all relation symbols appearing in a query of the class.

from the present theory in conjunction with a previous theorem on the complexity of deciding EP^k-expressibility of existential positive sentences ([6, Theorem 6]). Let us remark that FO^k-expressibility is undecidable on first-order sentences ([2, Remark 5.3]), and that (to our knowledge) prior to this work, FO^k-expressibility of existential positive sentences was not even known to be decidable.

To establish the inexpressibility portion of our primary theorem, for each Boolean existential positive query ϕ, we show how to construct two finite structures $\mathbf{B}$, $\mathbf{B}'$ on which the query differs, but which are not distinguishable from each other by any first-order query using a number of variables strictly less than the combinatorial width of ϕ. To show this non-distinguishability, we make use of a known Ehrenfeucht-Fraïssé type game [5, 20] designed for showing non-FO^m-expressibility. We in fact first perform this construction for Boolean conjunctive queries (phrased in terms of homomorphisms; see Section 3.1, Theorem 6); after this, we observe that this result extends to Boolean existential positive queries (Theorem 22), and then build on this understanding to treat general existential positive queries (Theorem 23).

The construction of the aforementioned two structures is based on a construction due to Atserias et al. [3]. This previous work characterized, for each finite structure $\mathbf{A}$, the number of pebbles needed for the *existential pebble game* [21] to act as a solution procedure for deciding if there exists a homomorphism from $\mathbf{A}$ to n given structure $\mathbf{B}$ (or, equivalently, if the conjunctive query corresponding to $\mathbf{A}$ evaluates to true on a given structure $\mathbf{B}$). As we show in the present article (see the discussion of Theorem 28 in Section 5), this previous characterization theorem can be readily derived from our primary theorem, and hence our primary theorem provides a strengthening of and broader perspective on this previous theorem.

Let us mention that, in related work, there are numerous articles that investigate the applicability of pebble games to query evaluation problems, which issue was a motivation for the Atserias et al. article [3]. As examples, we mention the work of Dalmau et al. on conjunctive queries and the existential pebble game [16]; the works of Chen and Dalmau on quantified conjunctive queries [14, 9]; the work of Chen and Dalmau on conjunctive queries and generalized hypertree width [13]; and, the work of Barceló et al. [4] on semantically acyclic query evaluation under database constraints.

A conference version of this article appeared in the proceedings of ICDT 2017 [7].

2 Preliminaries

For an integer $k \geq 0$, we use $[k]$ to denote the set $\{1, \ldots, k\}$, with the convention that $[0] = \emptyset$. For $i = 1, 2$, we freely let π_i denote the ith projection both over pairs and over sets of pairs, so for instance $\pi_i((a_1, a_2)) = a_i$ and $\pi_i(A_1 \times A_2) = A_i$. For an integer n, we let $n \pmod 2$ denote the value $b \in \{0, 1\}$ such that n and b are congruent modulo 2, that is, such that $n - b$ is an integer multiple of 2.

When $f\colon A \to B$ and $g\colon B \to C$ are functions, we use $g(f)$ to denote their composition. When h is a partial function, we use $\mathsf{dom}(h)$ to denote the domain of h.

Graphs, Structures, and Logic

Graphs. All graphs $G = (V, E)$ in this article are undirected and simple, that is, E is a set of 2-element subsets of V.

A *walk* in G is a sequence $W = (a_1, \ldots, a_m) \in V^m$, $m \geq 0$, such that $a_1, \ldots, a_m \in V$ and $\{a_i, a_{i+1}\} \in E$ for all $i \in [m-1]$. Relative to a walk $W = (a_1, \ldots, a_m)$, we use the following terminology. We say that W: *contains* $a \in V$ if $a = a_i$ for some $i \in [m]$; is *from* $s \in V$ *to* $t \in V$ if $a_1 = s$ and $a_m = t$; is *from* $s \in V$ *to* $T \subseteq V$ if $a_1 = s$ and $a_m \in T$; is s*-cyclic* if $a_1 = a_m = s$. A graph $G = (V, E)$ is *connected* if for every two vertices v and v' in V there exists a walk from v to v'.

A *tree decomposition* of a graph $G = (V, E)$ consists of a tree T where each node u is associated to a nonempty subset B_u of V (also called *bag*) such that the following holds:

- For each vertex $v \in V$, the nodes u of T such that $v \in B_u$ form a non-empty connected subtree of T.
- For each edge $e \in E$, there exists a node u in T such that $e \subseteq B_u$.

The *width* of a tree decomposition T of a graph G is defined as the maximum size attained by its bags minus 1, that is, $\max_{u \in T} |B_u| - 1$. The *treewidth* of a graph G, denoted by $\mathsf{tw}(G)$, is the minimum width over all tree decompositions of G.

Let $G = (V, E)$ be a graph. We say that two subsets U and U' of V *touch* if $U \cap U' \neq \emptyset$ or there exist $u \in U$, $u' \in U'$ such that $\{u, u'\} \in E$. A set $\mathcal{M}$ of mutually touching connected subsets of V is called a *bramble* of G. A subset H of V *hits* $\mathcal{M}$ if $H \cap M \neq \emptyset$ for all $M \in \mathcal{M}$. The *order* of a bramble is the minimum size attained over its hitting sets. We will make use of the following duality theorem.

▶ **Theorem 1.** *(refer to [17]) For $k \geq 1$, a graph has treewidth $\geq k$ if and only if it has a bramble of order $> k$.*

Structures. A *relational vocabulary* σ is a set of *relation symbols* R, each of which has an associated natural number $\mathsf{ar}(R)$ called its *arity*.

Let σ be a relational vocabulary. A σ*-structure* $\mathbf{A}$ is specified by a nonempty set A, called the *universe* of $\mathbf{A}$ and denoted by the corresponding italic letter, and a relation $R^{\mathbf{A}} \subseteq A^{\mathsf{ar}(R)}$ for each relation symbol $R \in \sigma$. A structure is *finite* if its universe is finite.

Let $\mathbf{A}$ and $\mathbf{B}$ be σ-structures. A *homomorphism* from $\mathbf{A}$ to $\mathbf{B}$ is a mapping $h : A \to B$ such that for each symbol $R \in \sigma$: if the tuple $(a_1, \ldots, a_{\mathsf{ar}(R)})$ is in $R^{\mathbf{A}}$, then the tuple $(h(a_1), \ldots, h(a_{\mathsf{ar}(R)}))$ is in $R^{\mathbf{B}}$. We write $\mathbf{A} \to \mathbf{B}$ to indicate that there exists a homomorphism from $\mathbf{A}$ to $\mathbf{B}$. $\mathbf{A}$ and $\mathbf{B}$ are *homomorphically equivalent* if $\mathbf{A} \to \mathbf{B}$ and $\mathbf{B} \to \mathbf{A}$ both hold. An *endomorphism* of $\mathbf{A}$ is a homomorphism from $\mathbf{A}$ to $\mathbf{A}$. An *automorphism* of $\mathbf{A}$ is a bijective mapping $h \colon A \to A$ such that $(a_1, \ldots, a_{\mathsf{ar}(R)}) \in R^{\mathbf{A}}$ if and only if $(h(a_1), \ldots, h(a_{\mathsf{ar}(R)})) \in R^{\mathbf{A}}$, for all $R \in \sigma$ and $(a_1, \ldots, a_{\mathsf{ar}(R)}) \in A^{\mathsf{ar}(R)}$; note that the inverse of an automorphism is an automorphism. A structure $\mathbf{A}$ is a *core* if every endomorphism of $\mathbf{A}$ is an automorphism of $\mathbf{A}$.

The structure $\mathbf{B}$ is a *substructure* of the structure $\mathbf{A}$ if $B \subseteq A$ and $R^{\mathbf{B}} \subseteq R^{\mathbf{A}}$ for all relation symbols $R \in \sigma$. When $\mathbf{B}$ is a substructure of $\mathbf{A}$, there exists a homomorphism h from $\mathbf{A}$ to $\mathbf{B}$, and h fixes each element $b \in B$, the mapping h is said to be a *retraction* from $\mathbf{A}$ to $\mathbf{B}$; when there exists a retraction from $\mathbf{A}$ to $\mathbf{B}$, it is said that $\mathbf{A}$ *retracts* to $\mathbf{B}$. A *core of a structure* $\mathbf{A}$ is a structure $\mathbf{C}$ such that $\mathbf{A}$ retracts to $\mathbf{C}$, but $\mathbf{A}$ does not retract to any proper substructure of $\mathbf{C}$. It is well known that each finite structure has a core and all cores of a finite structure are isomorphic [19]; we therefore freely refer to *the* core of a finite structure $\mathbf{A}$, and denote this object by $\mathsf{core}(\mathbf{A})$. The following facts are known and straightforwardly verified: when $\mathbf{A}$ is a core, then $\mathbf{A}$ is a core of $\mathbf{A}$; and, when $\mathbf{C}$ is a core of $\mathbf{A}$, then $\mathbf{C}$ is a core.

The *Gaifman graph* of a structure $\mathbf{A}$ is the graph with vertex set A and having an edge $\{a, a'\}$ if and only if $a \neq a'$ and a and a' cooccur in a tuple of $\mathbf{A}$. The treewidth of a structure $\mathbf{A}$, denoted by $\mathsf{tw}(\mathbf{A})$, is defined as the treewidth of its Gaifman graph.

Logic. In this article, we deal with first-order logic. An *atom* (over vocabulary σ) is an equality of variables ($x = y$) or is a predicate application $R(x_1, \ldots, x_r)$, where $x_1, \ldots, x_r$ are variables, and $R \in \sigma$ is a relation symbol of arity r. A *formula* (over vocabulary σ) is built from atoms (over σ), negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$), universal quantification ($\forall$), and existential quantification ($\exists$). We define $\mathsf{free}(\phi)$ to be the set of free variables of a formula ϕ. A *sentence* is a formula having no free variables.

We let FO denote the class of first-order formulas. An *existential positive* formula (over vocabulary σ) is a formula built from atoms (over σ) using conjunction, disjunction, and existential quantification; we let EP denote the class of existential positive formulas. A *primitive positive* formula (over vocabulary σ) is a formula built from atoms (over σ) using conjunction and existential quantification; we let PP denote the class of primitive positive formulas. Let $\mathrm{L} \subseteq \mathrm{FO}$. A formula $\phi \in \mathrm{FO}$ is called an L-formula (respectively, an L-sentence) if ϕ is in L (respectively, if ϕ is a sentence in L).

When $\mathbf{A}$ is a structure, f is an assignment of variables in A, and ϕ is a formula over the vocabulary of $\mathbf{A}$, we write $\mathbf{A}, f \models \phi$ to indicate that ϕ is true in $\mathbf{A}$ under f; if ϕ is a sentence, we simply write $\mathbf{A} \models \phi$. Let ϕ and ψ be formulas over the vocabulary σ having the same free variables. We write $\phi \models \psi$ to indicate that ϕ entails ψ, that is, for all σ-structures $\mathbf{A}$ and assignments f in A it holds that, $\mathbf{A}, f \models \phi$ implies $\mathbf{A}, f \models \psi$. We say that ϕ and ψ are *logically equivalent* (denoted $\phi \equiv \psi$) if $\phi \models \psi$ and $\psi \models \phi$. Let ϕ be an FO-formula and let $\mathrm{L} \subseteq \mathrm{FO}$. We say that ϕ is L-*expressible* if there exists an L-formula ϕ' such that $\phi \equiv \phi'$.

For any PP-sentence ϕ over σ, we let $\mathbf{C}[\phi]$ denote the canonical structure induced by ϕ. The *canonical structure* of ϕ is obtained by first prenexing the quantifiers and eliminating the equalities in ϕ, obtaining a logically equivalent PP-sentence ϕ' in prenex form and equality free; and second by defining $\mathbf{C}[\phi]$ to be the structure having a universe element for each existentially quantified variable in ϕ', and where, for each $R \in \sigma$, the relation $R^{\mathbf{C}[\phi]}$ contains $(x_1, \ldots, x_r)$ if and only if $R(x_1, \ldots, x_r)$ appears in the quantifier free part of ϕ'.

For a finite σ-structure $\mathbf{A}$, we let $Q[\mathbf{A}]$ denote the *canonical query* of $\mathbf{A}$, namely, if $A = \{a_1, \ldots, a_n\}$, then

$$Q[\mathbf{A}] = \exists a_1 \ldots \exists a_n \bigwedge_{R \in \sigma} \bigwedge_{(a'_1, \ldots, a'_k) \in R^{\mathbf{A}}} R(a'_1, \ldots, a'_k).$$

It is straightforward to verify that any PP-sentence ϕ is logically equivalent to $Q[\mathbf{C}[\phi]]$, and that every finite structure $\mathbf{A}$ is homomorphically equivalent to $\mathbf{C}[Q[\mathbf{A}]]$. We will use the following known fact [8].

▶ **Theorem 2** (Chandra and Merlin [8])**.** *Let ϕ be a* PP*-sentence and let $\mathbf{A}$ be a finite structure, such that $\phi = Q[\mathbf{A}]$ or $\mathbf{A} = \mathbf{C}[\phi]$. Then, for any structure $\mathbf{B}$, it holds that $\mathbf{A} \to \mathbf{B}$ if and only if $\mathbf{B} \models \phi$.*

Finite Variable Logics

For each class L of first-order formulas and each integer $k \geq 1$, we let L^k denote those formulas in L that use at most k distinct variable symbols. Fragments of first-order logic using only finitely many variables, called *finite variable logics*, are central in finite model theory. Equivalence in these fragments can be characterized by Ehrenfeucht-Fraïssé style games called *pebble games* [20], which we now introduce.

▶ **Definition 3.** Let σ be a relational vocabulary and let $\mathbf{A}$ and $\mathbf{B}$ be σ-structures. A *partial isomorphism from* $\mathbf{A}$ *to* $\mathbf{B}$ is an injective partial function h from A to B such that for all $R \in \sigma$ and $a_1, \ldots, a_{\mathsf{ar}(R)} \in \mathsf{dom}(h)$ it holds that $(a_1, \ldots, a_{\mathsf{ar}(R)}) \in R^{\mathbf{A}}$ if and only if $(h(a_1), \ldots, h(a_{\mathsf{ar}(R)})) \in R^{\mathbf{B}}$.

The k-pebble game is played by two players, a spoiler and a duplicator, over two (finite) relational structures $\mathbf{A}$ and $\mathbf{B}$ over the same vocabulary. A position of the game is a subset of $A \times B$ of size at most k. The game starts in the empty position and continues in a sequence of rounds. In each round of the game, the spoiler removes a pair from the current position if its size is k, and then selects an element $a \in A$ or $b \in B$; the duplicator answers by selecting an element $b \in B$ or $a \in A$, respectively. The new position is defined by adding the pair (a, b) to the old position. The duplicator wins the game if each position occurring along the rounds is a partial isomorphism from $\mathbf{A}$ to $\mathbf{B}$.

▶ **Definition 4.** [20] Let $k \geq 0$. A *duplicator winning strategy in the k-pebble game on* $\mathbf{A}$ *and* $\mathbf{B}$ is a family $\mathcal{S}$ of partial isomorphisms h from $\mathbf{A}$ to $\mathbf{B}$ with $|\mathsf{dom}(h)| \leq k$ such that:
(S1) $\emptyset \to \emptyset$ is in $\mathcal{S}$.
(S2) If $h \in \mathcal{S}$ and $|\mathsf{dom}(h)| < k$, then:
  (S2.F) For every $a \in A$ there exists $b \in B$ such that $h \cup \{(a, b)\}$ is in $\mathcal{S}$.
  (S2.B) For every $b \in B$ there exists $a \in A$ such that $h \cup \{(a, b)\}$ is in $\mathcal{S}$.
(S3) If $h \in \mathcal{S}$ and $a \in \mathsf{dom}(h)$, then $h|_{\mathsf{dom}(h) \setminus \{a\}}$ is in $\mathcal{S}$.

It is clear that the duplicator wins the above described k-pebble game on $\mathbf{A}$ and $\mathbf{B}$ if and only if the game admits a duplicator winning strategy.

We say that two structures $\mathbf{A}$, $\mathbf{B}$ are *indistinguishable by* FO^k*-sentences* (in short FO^k-*indistinguishable*), if for each FO^k-sentence ϕ it holds that $\mathbf{A} \models \phi$ if and only if $\mathbf{B} \models \phi$. As anticipated, k-pebble games characterize expressibility in the k-variable fragment of first-order logic.

▶ **Theorem 5** (Immerman [20])**.** *Let $k \geq 0$ and let $\mathbf{A}$ and $\mathbf{B}$ be relational structures on the same vocabulary. The following are equivalent.*
1. *There exists a duplicator winning strategy in the k-pebble game on $\mathbf{A}$ and $\mathbf{B}$.*
2. *$\mathbf{A}$ and $\mathbf{B}$ are FO^k-indistinguishable.*

3 Construction of Structures

In this section, we show a theorem implying that certain PP-sentences, namely those corresponding (via Chandra-Merlin, Theorem 2) to cores of treewidth at least k, cannot be expressed by FO-sentences using k variables (Theorem 6). This inexpressibility result allows us to later derive our primary theorem (Theorem 23).

▶ **Theorem 6.** *Let $\mathbf{A}$ be a core on the relational vocabulary σ such that $\mathsf{tw}(\mathbf{A}) \geq k \geq 1$. There exist σ-structures $\mathbf{B}$ and $\mathbf{B}'$ such that $\mathbf{B} \to \mathbf{A}$; $\mathbf{A} \to \mathbf{B}'$; $\mathbf{A} \not\to \mathbf{B}$; and, $\mathbf{B}$ and $\mathbf{B}'$ are FO^k-indistinguishable.*

The remainder of the current section is devoted to the construction (Section 3.1) and the study (Section 3.2 and Section 3.3) of the structures $\mathbf{B}$ and $\mathbf{B}'$ mentioned in Theorem 6. We remark that the structure $\mathbf{B}$ is essentially equal to the structure defined by Atserias et al. in [3, Section 4].

We give a proof of Theorem 6 that makes forward references; this proof might serve as a guide to the layout of this section.

Proof of Theorem 6. It is readily verified that $\mathsf{tw}(\mathbf{A}) \geq k \geq 1$ implies the existence of a connected component (C, E) in the Gaifman graph of $\mathbf{A}$ such that $\mathsf{tw}((C, E)) \geq k \geq 1$. Let s be an arbitrary but fixed vertex in C. Let $\mathbf{B}$ and $\mathbf{B}'$ be the two σ-structures defined relative to $\mathbf{A}$ and s in Section 3.1; then $\mathbf{B} \to \mathbf{A}$ by Observation 9 in Section 3.1. In Section 3.2, Lemma 11 and Lemma 12 show that $\mathbf{A} \to \mathbf{B}'$ and $\mathbf{A} \not\to \mathbf{B}$, respectively. In Section 3.3, Lemma 16 gives a duplicator winning strategy in the k-pebble game on $(\mathbf{B}, \mathbf{B}')$, which suffices to yield the theorem, via Theorem 5. ◀

▶ Notation 7. The following names are reserved throughout the current section:

- $\mathbf{A}$ denotes a core on the relational vocabulary σ, with universe A, such that $\mathsf{tw}(\mathbf{A}) \geq k \geq 1$.
- $G = (C, E)$ denotes a connected component in the Gaifman graph of $\mathbf{A}$ such that $\mathsf{tw}(G) \geq k \geq 1$. Note that, in particular, $|C| \geq 2$.
- s denotes an arbitrary but fixed vertex in C.
- $\mathcal{M}$ denotes an arbitrary but fixed bramble of G having order $> k$ (such a bramble exists by Theorem 1). Recall that, therefore, any hitting set for $\mathcal{M}$ has size at least $k + 1$.

3.1 Construction of $\mathbf{B}$ and $\mathbf{B}'$

In this section, relative to $\mathbf{A}$ and s, we define two σ-structures $\mathbf{B}$ and $\mathbf{B}'$ as follows.

For all $a \in A$, let E_a denote the edges incident on a in the Gaifman graph of $\mathbf{A}$. Let:

$$U_C = \left\{ (a, f) \;\middle|\; a \in C, f \colon E_a \to \{0, 1\} \text{ is such that } \sum_{e \in E_a} f(e) \pmod 2 = \begin{cases} 0 & \text{if } a \neq s \\ 1 & \text{if } a = s \end{cases} \right\},$$

$$U'_C = \left\{ (a, f) \;\middle|\; a \in C, f \colon E_a \to \{0, 1\} \text{ is such that } 0 = \sum_{e \in E_a} f(e) \pmod 2 \right\},$$

$$U_{A \setminus C} = \{(a, f \colon E_a \to \{0\}) \mid a \in A \setminus C\}.$$

Then $\mathbf{B}$ and $\mathbf{B}'$ have universes B and B' defined as follows:

$$B = U_C \cup U_{A \setminus C},$$
$$B' = U'_C \cup U_{A \setminus C}.$$

The vocabulary is interpreted as follows. For all $R \in \sigma$, let $(a_1, f_1), \ldots, (a_r, f_r)$ be elements of $\mathbf{B}$ (respectively, of $\mathbf{B}'$), where $r = \mathsf{ar}(R)$. Then $((a_1, f_1), \ldots, (a_r, f_r))$ is in $R^{\mathbf{B}}$ (respectively, in $R^{\mathbf{B}'}$) if and only if

- $(a_1, \ldots, a_r) \in R^{\mathbf{A}}$;
- for all $i, j \in [r]$, if $e = \{a_i, a_j\} \in E$ then $f_i(e) = f_j(e)$.

For the sake of intuition, suppose that $\mathbf{A}$ is a connected graph, so that $\mathbf{A}$ is isomorphic to its Gaifman graph and $C = A$. The universes of $\mathbf{B}$ and $\mathbf{B}'$ are formed by pairs (a, f) where a is a vertex of $\mathbf{A}$ and f is a Boolean labelling of the edges incident on a. The Boolean labellings have even parity with the only exception of those paired with s in B which have odd parity. Moreover there is an edge $\{(a, f), (a', f')\}$ in $\mathbf{B}$ (respectively, $\mathbf{B}'$) if and only if the edge $\{a, a'\}$ is in $\mathbf{A}$ and the labellings f and f' agree on $\{a, a'\}$. A concrete example follows.

▶ **Example 8.** Let $\mathbf{A} = (A, E^{\mathbf{A}})$ where $A = \{a, s\}$ and $E^{\mathbf{A}} = \{(a, s), (s, a)\}$, that is, $\mathbf{A}$ is the graph formed by the single edge $\{a, s\}$. Let $f_i \colon \{\{a, s\}\} \to \{0, 1\}$ be such that $f_i(\{a, s\}) = i$ for $i = 0, 1$. Then $\mathbf{B} = (B, E^{\mathbf{B}})$ where $B = \{(a, f_0), (s, f_1)\}$ and $E^{\mathbf{B}} = \emptyset$, and $\mathbf{B}' = (B', E^{\mathbf{B}'})$ where $B' = \{(a, f_0), (s, f_0)\}$ and $E^{\mathbf{B}'} = \{((a, f_0), (s, f_0)), ((s, f_0), (a, f_0))\}$.

Note that, by construction of $\mathbf{B}$, the map $\pi_1\colon B \to A$ is a homomorphism from $\mathbf{B}$ to $\mathbf{A}$. We inline this observation for later use.

▶ Observation 9. $\mathbf{B} \to \mathbf{A}$.

3.2 A Treats B and B' Differently

Let $\mathbf{B}$ and $\mathbf{B}'$ be the two σ-structures defined relative to $\mathbf{A}$ and s in Section 3.1. We show that $\mathbf{A}$ maps homomorphically to $\mathbf{B}'$ (Lemma 11) but not to $\mathbf{B}$ (Lemma 12).

▶ **Example 10.** Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{B}'$ be as in Example 8. Then the mapping h defined by $h(a') = (a', f_0)$ for all $a' \in A$ is a homomorphism from $\mathbf{A}$ to $\mathbf{B}'$, but $\mathbf{A}$ has no homomorphisms to $\mathbf{B}$ by direct inspection, as $E^{\mathbf{B}} = \emptyset$.

We show that $\mathbf{A}$ maps homomorphically to $\mathbf{B}'$.

▶ **Lemma 11.** $\mathbf{A} \to \mathbf{B}'$.

The essence of the proof is to conduct a rather direct inspection of the construction in Section 3.1; this shows that a copy of $\mathbf{A}$ sits inside $\mathbf{B}'$, by looking at the pairs in B' carrying an identically 0 labelling.

Proof. Let $h\colon A \to B'$ be defined as follows. For all $a \in A$, let $h(a) = (a, f\colon E_a \to \{0\}) \in B'$. We claim that h is a homomorphism from $\mathbf{A}$ to $\mathbf{B}'$. Let $R \in \sigma$, $\mathsf{ar}(R) = r$, and let $(a_1, \ldots, a_r) \in R^{\mathbf{A}}$.

Note that by the definition of Gaifman graph, either $\{a_1, \ldots, a_r\} \cap C = \emptyset$ or $\{a_1, \ldots, a_r\} \subseteq C$. In the former case, $(h(a_1), \ldots, h(a_r))$ is trivially in $R^{\mathbf{B}'}$ because no $i, j \in [r]$ satisfy $\{a_i, a_j\} \in E$. In the latter case, let $i, j \in [r]$ be such that $i \neq j$. Then, $\{a_i, a_j\} \in E$; set $e = \{a_i, a_j\}$. We have $\pi_2(h(a_i))(e) = 0 = \pi_2(h(a_j))(e)$ by definition of h. Hence $(h(a_1), \ldots, h(a_r))$ is in $R^{\mathbf{B}'}$. ◀

On the other hand, $\mathbf{B}$ has no homomorphisms from $\mathbf{A}$.

▶ **Lemma 12.** *[3, Lemma 1]* $\mathbf{A} \not\to \mathbf{B}$.

For the sake of understanding this lemma intuitively, let $\mathbf{A}$ be a connected graph. A homomorphism from $\mathbf{A}$ to $\mathbf{B}$ maps A to B preserving all edges. By construction (here we use that $\mathbf{A}$ is a core), the homomorphic image of $\mathbf{A}$ in $\mathbf{B}$ is a copy of $\mathbf{A}$ where all edges carry *two* Boolean labels *equal* to each other. Summing these Boolean labels in two ways, edgewise and vertexwise, we get the contradiction that the edgewise sum has even parity but the vertexwise sum, by the contribution of the labels of s, has odd parity.

We give a proof for completeness.

Proof. Assume for a contradiction that h is a homomorphism from $\mathbf{A}$ to $\mathbf{B}$. By a standard argument, we may assume, without loss of generality, that $\pi_1(h(a)) = a$ for all $a \in A$.[2]

[2] This can be justified as follows. We write $g \circ f = g(f)$ to denote the composition of g and f, with f applied first. We have that π_1 is a homomorphism from $\mathbf{B}$ to $\mathbf{A}$. Hence $\pi_1 \circ h$ is an endomorphism of $\mathbf{A}$, and since $\mathbf{A}$ is a core, $\pi_1 \circ h$ is an automorphism of $\mathbf{A}$. By associativity of function composition,

$$\mathrm{id}_A = (\pi_1 \circ h) \circ (\pi_1 \circ h)^{-1} = \pi_1 \circ (h \circ (\pi_1 \circ h)^{-1}) = \pi_1 \circ h'$$

that is, if $h' = h \circ (\pi_1 \circ h)^{-1}$, then $\pi_1(h'(a)) = a$ for all $a \in A$. Moreover, h' is a homomorphism from $\mathbf{A}$ to $\mathbf{B}$ because $(\pi_1 \circ h)^{-1}$ is an automorphism of $\mathbf{A}$ and h is a homomorphism from $\mathbf{A}$ to $\mathbf{B}$.

We claim that for all $e = \{a', a''\} \in E$ it holds that

$$\pi_2(h(a'))(e) = \pi_2(h(a''))(e). \tag{1}$$

Let $e = \{a', a''\} \in E$. Note that $e \in E_{a'} \cap E_{a''}$. By the definition of the Gaifman graph of $\mathbf{A}$, there exist $R \in \sigma$ and $(a_1, \ldots, a_i, \ldots, a_j, \ldots, a_r) \in R^{\mathbf{A}}$ with $a_i = a'$ and $a_j = a''$. Then, since $\mathbf{A} \to \mathbf{B}$ via h,

$$(h(a_1), \ldots, h(a_r)) = ((a_1, f_1), \ldots, (a_r, f_r)) \in R^{\mathbf{B}}.$$

By construction, for all $i, j \in [r]$, if $e = \{a_i, a_j\} \in E$, then $f_i(e) = f_j(e)$. In particular,

$$\pi_2(h(a'))(e) = \pi_2(h(a_i))(e) = f_i(e) = f_j(e) = \pi_2(h(a_j))(e) = \pi_2(h(a''))(e),$$

and we have that (1) holds.

We conclude observing that, by construction,

$$1 = \sum_{e \in E_s} \pi_2(h(s))(e) \pmod 2$$

and for all $a \in A \setminus \{s\}$,

$$0 = \sum_{e \in E_a} \pi_2(h(a))(e) \pmod 2.$$

Then, letting b_e denote the quantity in (1), we have:

$$\begin{aligned} 1 &= \sum_{e \in E_s} \pi_2(h(s))(e) + \sum_{a \in A \setminus \{s\}} \sum_{e \in E_a} \pi_2(h(a))(e) \pmod 2 \\ &= 2 \cdot \sum_{e \in E} b_e \pmod 2 && \text{by (1)} \\ &= 0, \end{aligned}$$

a contradiction. ◀

3.3 B and B′ are FO^k-Indistinguishable

Let $\mathbf{B}$ and $\mathbf{B}'$ be the two σ-structures defined relative to $\mathbf{A}$ and s in Section 3.1. We show that $\mathbf{B}$ and $\mathbf{B}'$ are FO^k-indistinguishable (Lemma 16).

We describe informally a winning strategy for the duplicator in the k-pebble game on $\mathbf{B}$ and $\mathbf{B}'$. For the sake of illustration, consider the simple case where $\mathbf{B}$ and $\mathbf{B}'$ are constructed relative to a *connected graph* $\mathbf{A}$ so that $\mathbf{A} = G$ (the lemma lifts the idea to arbitrary relational structures $\mathbf{A}$, whose Gaifman graphs are possibly disconnected). The duplicator fixes a bramble $\mathcal{M}$ of G and maintains along the rounds of the game the following position ($i = 0, 1, \ldots, k$):

- i pebble pairs are placed on elements $(a_1, f_1), \ldots, (a_i, f_i) \in B$ and correspondingly on elements $(a_1, f'_1), \ldots, (a_i, f'_i) \in B'$ such that for some walk W in G from s to a bramble set in $\mathcal{M}$ avoiding $a_1, \ldots, a_i$ it holds that $f_j(e)$ equals the parity of $f'_j(e)$ plus the number of times e occurs in W (for all $j \in [i]$ and $e \in E_{a_j}$).

That such a position exists and is a partial isomorphism is the content of Lemma 14; that the duplicator can maintain such a position along the rounds of the game is the content of Lemma 16.

We now start proving our statement. Recall Notation 7 for the meaning of G, $\mathcal{M}$, et cetera. We further prepare the following notation.

▶ Notation 13. Let $W = (a_1, \dots, a_m)$ be a walk in G. For $e \in E$, we let

$$\mathsf{occ}_W(e) = |\{i \in [m-1] \mid e = \{a_i, a_{i+1}\}\}|$$

denote the number of times the edge e is used in W. Moreover, for every $S \subseteq A$, let

$$\mathsf{avoid}_{\mathcal{M}}(S) = \bigcup_{\{M \in \mathcal{M} \mid M \cap S = \emptyset\}} M$$

be the union of the bramble sets in $\mathcal{M}$ disjoint from S.

▶ **Lemma 14.** *Let $0 \leq i \leq k$ and let $\{(a_1, f_1), \dots, (a_i, f_i)\} \subseteq B$. Let W be a walk in G from s to $\mathsf{avoid}_{\mathcal{M}}(\{a_1, \dots, a_i\})$. For all $j \in [i]$, let $f'_j : E_{a_j} \to \{0, 1\}$ be defined by $f'_j(e) = f_j(e) + \mathsf{occ}_W(e) \pmod 2$. Then the mapping h sending (a_j, f_j) to (a_j, f'_j) for all $j \in [i]$ is a partial isomorphism from $\mathbf{B}$ to $\mathbf{B}'$.*

Towards proving Lemma 14, we claim the following.

▶ Claim 15. Let $a \in A$ and let W be a walk in G from s to $t \neq a$. Then:

- If $(a, f) \in B$ and $f'(e) = f(e) + \mathsf{occ}_W(e) \pmod 2$ for all $e \in E_a$, then $(a, f') \in B'$.
- If $(a, f') \in B'$ and $f(e) = f'(e) - \mathsf{occ}_W(e) \pmod 2$ for all $e \in E_a$, then $(a, f) \in B$.

The idea underlying the claim is that, for $(a, f) \in B$ with $a = s$, both the sum of the labellings of E_a under f, and the number of occurrences of edges in E_a in a walk W that starts at a and does not end at a, are odd. For $(a, f) \in B$ with $a \neq s$ both the sum of the labellings under f of edges incident on a, and the number of occurrences of edges incident on a in a walk W that neither starts nor ends at a, are even. It follows that $(a, f') \in B'$ by construction.

Proof of Claim 15. For the first item, let $a \in A$, let $(a, f) \in B$, and let $f' \colon E_a \to \{0, 1\}$ be such that $f'(e) = f(e) + \mathsf{occ}_W(e) \pmod 2$ for all $e \in E_a$.

If $a \in A \setminus C$, then $(a, f) \in B$ implies that $f(e) = \mathsf{occ}_W(e) = 0$ for all $e \in E_a$, so that $f'(e) = 0$ for all $e \in E_a$. By construction, $(a, f') \in B'$. Otherwise, if $a \in C$, we distinguish two cases.

Case: If $a = s$, then

$$1 = \sum_{e \in E_s} \mathsf{occ}_W(e) \pmod 2$$

because W starts at s and ends at $t \neq s = a$ (and G does not contain loops). On the other hand, by definition of $\mathbf{B}$,

$$1 = \sum_{e \in E_s} f(e) \pmod 2,$$

so $0 = \sum_{e \in E_s} f'(e) \pmod 2$, and $(s, f') = (a, f') \in B'$ by definition of $\mathbf{B}'$.

Case: If $a \neq s$, then

$$0 = \sum_{e \in E_a} \mathsf{occ}_W(e) \pmod 2$$

because W neither starts nor ends at a (and G does not contain loops). On the other hand, by definition,

$$0 = \sum_{e \in E_a} f(e) \pmod 2,$$

so $0 = \sum_{e \in E_a} f'(e) \pmod 2$ and $(a, f') \in B'$.

For the second item, if $a \in A \setminus C$ we similarly have that f is identically 0 on E_a and therefore $(a, f) \in B$. Otherwise, if $a \in C$, let $(a, f') \in B'$, and let $f \colon E_a \to \{0, 1\}$ be such that $f(e) = f'(e) - \mathsf{occ}_W(e) \pmod 2$. We distinguish two cases. If $a = s$, then $1 = \sum_{e \in E_s} \mathsf{occ}_W(e) \pmod 2$ and $0 = \sum_{e \in E_s} f'(e) \pmod 2$, therefore $1 = \sum_{e \in E_s} f(e) \pmod 2$ and $(s, f) = (a, f) \in B$. If $a \neq s$, then $0 = \sum_{e \in E_a} \mathsf{occ}_W(e) \pmod 2$ and $0 = \sum_{e \in E_a} f'(e) \pmod 2$ imply $0 = \sum_{e \in E_a} f(e) \pmod 2$, so that $(a, f) \in B$ and we are done. ◀

We are now ready to prove the lemma. For the sake of intuition, consider the case where $\mathbf{A}$ is a connected graph, so that $\mathbf{A} = G$. If the edge $\{(a, f), (b, g)\}$ is in $\mathbf{B}$ and $h((a, f)) = (a, f')$, $h((b, g)) = (b, g')$, then on the one hand $f(\{a, b\}) = g(\{a, b\})$, which is the content of (4); and on the other hand $f'(\{a, b\})$ and $g'(\{a, b\})$ equal the parity of the sum of $f(\{a, b\})$ and $g(\{a, b\})$, respectively, and the number of occurrences of $\{a, b\}$ in a fixed walk W, which is the content of (3). Therefore $f'(\{a, b\}) = g'(\{a, b\})$ and the edge $\{(a, f), (b, g)\}$ is in $\mathbf{B}'$. The converse is symmetric.

Proof of Lemma 14. By Claim 15, $(a_j, f'_j) \in B'$ for all $j \in [i]$, hence h is a partial function from B to B'; moreover, h is injective by definition. Let $R \in \sigma$, let $\mathsf{ar}(R) = r$, and let $(b_1, f_1), \ldots, (b_r, f_r) \in \mathsf{dom}(h)$. It is sufficient to show that

$$((b_1, f_1), \ldots, (b_r, f_r)) \in R^{\mathbf{B}} \iff (h(b_1, f_1), \ldots, h(b_r, f_r)) \in R^{\mathbf{B}'}$$

($\Longrightarrow$) Assume $((b_1, f_1), \ldots, (b_r, f_r)) \in R^{\mathbf{B}}$. Then by construction $(b_1, \ldots, b_r) \in R^{\mathbf{A}}$. We distinguish two cases.

Case: $\{b_1, \ldots, b_r\} \cap C = \emptyset$. Note that if $b_j \in A \setminus C$, then $f_j(e) = 0$ for all $e \in E_{b_j}$ by construction and $\mathsf{occ}_W(e) = 0$ for all $e \in E_{b_j}$ because W lies entirely in C. Then $f'_j(e) = 0$ for all $e \in E_{b_j}$. Therefore, by construction, $(h(b_1, f_1), \ldots, h(b_r, f_r)) \in R^{\mathbf{B}'}$.

Case: If $\{b_1, \ldots, b_r\} \subseteq C$, then let $e = \{b_j, b_{j'}\} \in E$, $j, j' \in [r]$. We claim that $f'_j(e) = f'_{j'}(e)$. By hypothesis, there exists a walk W in G from s to $t \in \mathsf{avoid}_{\mathcal{M}}(\{a_1, \ldots, a_i\})$ such that for all $j \in [i]$ and all $e \in E_{a_j}$ it holds that

$$f'_j(e) = f_j(e) + \mathsf{occ}_W(e) \pmod 2. \tag{2}$$

It follows from (2) that, for all $j \in [r]$ and $e \in E_{b_j}$,

$$f'_j(e) = f_j(e) + \mathsf{occ}_W(e) \pmod 2. \tag{3}$$

Moreover, by construction, if $j, j' \in [r]$ and $e = \{b_j, b_{j'}\} \in E$, then

$$f_j(e) = f_{j'}(e). \tag{4}$$

Therefore,

$$\begin{aligned} f'_j(e) &= f_j(e) + \mathsf{occ}_W(e) \pmod 2 && \text{by (3)} \\ &= f_{j'}(e) + \mathsf{occ}_W(e) \pmod 2 && \text{by (4)} \\ &= f'_{j'}(e) && \text{by (3)} \end{aligned}$$

We conclude that $((b_1, f'_1), \ldots, (b_r, f'_r)) = (h(b_1, f_1), \ldots, h(b_r, f_r)) \in R^{\mathbf{B}'}$.

($\Longleftarrow$) Assume $((b_1, f'_1), \ldots, (b_r, f'_r)) \in R^{\mathbf{B}'}$. Then $(b_1, \ldots, b_r) \in R^{\mathbf{A}}$. If $\{b_1, \ldots, b_r\} \cap C = \emptyset$, then $(h(b_1, f'_1), \ldots, h(b_r, f'_r)) \in R^{\mathbf{B}}$ reasoning as above. If $\{b_1, \ldots, b_r\} \subseteq C$, then let $j, j' \in [r]$ and $e = \{b_j, b_{j'}\} \in E$. We claim that $f_j(e) = f_{j'}(e)$. By hypothesis, there exists a walk W in G from s to $t \in \mathsf{avoid}_{\mathcal{M}}(\{a_1, \ldots, a_i\})$ such that for all $j \in [i]$ and all $e \in E_{a_j}$ it holds that

$$f_j(e) = f'_j(e) - \mathsf{occ}_W(e) \pmod 2. \tag{5}$$

By (5) we have that for all $j \in [r]$ and $e \in E_{b_j}$

$$f_j(e) = f'_j(e) - \mathsf{occ}_W(e) \pmod 2. \tag{6}$$

Moreover, by construction, if $j, j' \in [r]$ and $e = \{b_j, b_{j'}\} \in E$, then

$$f'_j(e) = f'_{j'}(e). \tag{7}$$

Therefore,

$$\begin{aligned} f_j(e) &= f'_j(e) - \mathsf{occ}_W(e) \pmod 2 && \text{by (6)} \\ &= f'_{j'}(e) - \mathsf{occ}_W(e) \pmod 2 && \text{by (7)} \\ &= f_{j'}(e) && \text{by (6)} \end{aligned}$$

We conclude that $((b_1, f_1), \dots, (b_r, f_r)) \in R^{\mathbf{B}}$. ◀

We now formalize the strategy informally described above and show that, indeed, it is a winning strategy for the duplicator in the k-pebble game on $\mathbf{B}$ and $\mathbf{B}'$.

▶ **Lemma 16.** *Let $\mathcal{S}$ be the family of partial isomorphisms from $\mathbf{B}$ to $\mathbf{B}'$ that contains an injective map $h\colon B \to B'$ when $|\mathsf{dom}(h)| \le k$ and there exists a walk W in G from s to $\mathsf{avoid}_{\mathcal{M}}(\pi_1(\mathsf{dom}(h)))$ such that, for all $(a, f) \in \mathsf{dom}(h)$, it holds that $h((a, f)) = (a, f')$ where*

$$f'(e) = f(e) + \mathsf{occ}_W(e) \pmod 2$$

for all $e \in E_a$. Then $\mathcal{S}$ is a duplicator winning strategy in the k-pebble game on $\mathbf{B}$ and $\mathbf{B}'$.

The crux of the proof is the following. Suppose that $i < k$ pebble pairs are placed on elements $(a_1, f_1), \dots, (a_i, f_i) \in B$ and correspondingly on elements $(a_1, f'_1), \dots, (a_i, f'_i) \in B'$ such that for some walk W in G from s to t in a bramble set in $\mathcal{M}$ avoiding $a_1, \dots, a_i$ it holds that $f'_j(e)$ equals the parity of $f_j(e)$ plus the number of times e occurs in W (for all $j \in [i]$ and $e \in E_{a_j}$).

The spoiler pebbles, say, $(a_{i+1}, f_{i+1}) \in B$. The duplicator obtains a walk W' in G from s to a vertex t' lying in a bramble set of $\mathcal{M}$ that avoids $a_1, \dots, a_i, a_{i+1}$ (such a set exists because the bramble has order greater than $k \ge i + 1$) by walking from s to t over W and then from t to t' using only vertices in the bramble sets of t and t' (which is feasible by the properties of the bramble). The duplicator pebbles $(a_{i+1}, f'_{i+1}) \in B$, where $f'_{i+1}(e)$ equals the parity of $f_{i+1}(e)$ plus the number of times e occurs in W' for all $e \in E_{a_{i+1}}$; thus maintaining its winning position, because the number of occurrences of edges incident to any of $a_1, \dots, a_i$ does not change in passing from W to W'.

Proof of Lemma 16. We check that $\mathcal{S}$ satisfies Definition 4 relative to $\mathbf{B}$ and $\mathbf{B}'$.

For (S1), we show that the function $h\colon \emptyset \to \emptyset$ is in $\mathcal{S}$. Since any hitting set of the bramble $\mathcal{M}$ has size at least $k + 1 \ge 2$, there exists a set M in $\mathcal{M}$ such that $s \notin M$. Let $a \in M$. Then $a \in \mathsf{avoid}_{\mathcal{M}}(\pi_1(\mathsf{dom}(h)))$. Moreover, G is connected, hence there is a walk W in G from s to a. Thus $h \in \mathcal{S}$.

We now verify that $\mathcal{S}$ satisfies (S2.F) and (S2.B). Let $h \in \mathcal{S}$ and let $|\mathsf{dom}(h)| < k$. By definition of $\mathcal{S}$, there exists a walk W in G from s to $t \in \mathsf{avoid}_{\mathcal{M}}(\pi_1(\mathsf{dom}(h)))$ such that, for all $(a, f) \in \mathsf{dom}(h)$ it holds that $h((a, f)) = (a, f')$ where $f'(e) = f(e) + \mathsf{occ}_W(e) \pmod 2$ for all $e \in E_a$.

For (S2.F), let $(a, f) \in B$. Since $|\mathsf{dom}(h)| < k$, we have that $|\pi_1(\mathsf{dom}(h)) \cup \{a\}| \le k$. So, by the observation about any hitting set of the bramble $\mathcal{M}$, there exists a set $M' \in \mathcal{M}$ such that $M' \cap (\pi_1(\mathsf{dom}(h)) \cup \{a\}) = \emptyset$. Let $t \in M \in \mathcal{M}$. Obtain a walk W' in G from s to $t' \in \mathsf{avoid}_{\mathcal{M}}(\pi_1(\mathsf{dom}(h)) \cup \{a\})$ by concatenating W and a walk in G from $t \in M$ to $t' \in M'$ containing only vertices in M and M'. Let (a, f') be such that $f'(e) = f(e) + \mathsf{occ}_{W'}(e) \pmod 2$ for all $e \in E_a$. Now define $h' = h \cup \{((a, f), (a, f'))\}$.

We want to show that h' is in $\mathcal{S}$. We claim that, for all $(b, f_b) \in \mathsf{dom}(h')$, if $h'((b, f_b)) = (b, f'_b)$, then $f'_b(e) = f_b(e) + \mathsf{occ}_{W'}(e) \pmod 2$ for all all $e \in E_b$. It follows that h' is a partial isomorphism from $\mathbf{B}$ to $\mathbf{B}'$ by Lemma 14, so that h' is in $\mathcal{S}$ witnessed by W'. Hence $h' \in \mathcal{S}$.

For the claim, let $(b, f_b) \in \mathsf{dom}(h')$ and let $h'((b, f_b)) = (b, f'_b)$. If $b = a$, then $(b, f_b) = (a, f)$ and $h'((a, f)) = (a, f')$ such that $f'(e) = f(e) + \mathsf{occ}_{W'}(e) \pmod 2$ for all $e \in E_a$ by construction. If $b \neq a$, then notice that $b \notin M$ and $b \notin M'$, so that for every $e \in E_b$ it holds that $\mathsf{occ}_W(e) = \mathsf{occ}_{W'}(e)$, and therefore,

$$
\begin{aligned}
f'_b(e) &= f_b(e) + \mathsf{occ}_W(e) \pmod 2 && \text{by hypothesis on } h \\
&= f_b(e) + \mathsf{occ}_{W'}(e) \pmod 2
\end{aligned}
$$

and the claim is settled.

For (S2.B), let $(a, f') \in B'$. Along the lines above, we obtain a walk W' in G from s to $\mathsf{avoid}_{\mathcal{M}}(\pi_1(\mathsf{dom}(h)) \cup \{a\})$, and $f \colon E_a \to \{0, 1\}$ such that $f'(e) = f(e) + \mathsf{occ}_{W'}(e) \pmod 2$ for all $e \in E_a$; we put $h' = h \cup \{((a, f), (a, f'))\}$, and show that $h' \in \mathcal{S}$ appealing to Lemma 14.

We conclude by verifying that $\mathcal{S}$ satisfies (S3). Let $h \in \mathcal{S}$ and let $(a, f) \in \mathsf{dom}(h)$. We want to show that the restriction of h to $\mathsf{dom}(h) \setminus \{(a, f)\}$, namely, $h' = h|_{\mathsf{dom}(h) \setminus \{(a,f)\}}$ is in $\mathcal{S}$. Partial isomorphisms are closed under restrictions, hence h' is a partial isomorphism from $\mathbf{B}$ to $\mathbf{B}'$ of domain size at most k. Let W be a walk in G from s to $\mathsf{avoid}_{\mathcal{M}}(\pi_1(\mathsf{dom}(h)))$ witnessing that h is in $\mathcal{S}$. We claim that W also witnesses that h' is in $\mathcal{S}$. By definition, W is from s to $\mathsf{avoid}_{\mathcal{M}}(\pi_1(\mathsf{dom}(h))) \subseteq \mathsf{avoid}_{\mathcal{M}}(\pi_1(\mathsf{dom}(h')))$. Moreover, if $h'((a', f')) = (a', f'')$, then $h((a', f')) = (a', f'')$ and for all $e \in E_{a'}$ it holds that $f''(e) = f'(e) + \mathsf{occ}_W(e) \pmod 2$. ◀

4 Existential Positive Logic

In this section, we present *combinatorial width*, the combinatorial measure on EP-formulas. While the specialization of this measure to EP-sentences is due (implicitly) to previous work, we here give a definition that applies to all EP-formulas.

In order to define our measure, we first associate a structure to each PP-formula, as follows. In the following definition, one should conceive of ψ as a disjunct of an EP-formula which is a disjunction of PP-formulas and which has free variables $v_1, \dots, v_\ell$.

▶ **Definition 17.** For each vocabulary σ and each integer $\ell \geq 1$, we fix $U_{\sigma,\ell}$ to be a relation symbol of arity ℓ outside of σ.

For each vocabulary σ, each list $v_1, \dots, v_\ell$ of variables, and each PP-formula ψ over vocabulary σ with $\mathsf{free}(\psi) \subseteq \{v_1, \dots, v_\ell\}$, we define a structure $\mathbf{C}[\psi; \sigma; v_1, \dots, v_\ell]$ as follows:

- Define ψ' as the formula obtained from ψ by prenexing ψ and then renaming quantified variables (if necessary) so that none of the variables $v_1, \dots, v_\ell$ are quantified.
- Define ψ^+ as $\exists v_1 \dots v_\ell (U_{\sigma,\ell}(v_1, \dots, v_\ell) \wedge \psi')$ if $\ell > 0$, and as ψ' if $\ell = 0$.
- Define $\mathbf{C}[\psi; \sigma; v_1, \dots, v_\ell]$ as $\mathbf{C}[\psi^+]$.

Note that, in the case that $\ell = 0$, it holds that $\mathbf{C}[\psi; \sigma; v_1, \dots, v_\ell]$ is isomorphic to $\mathbf{C}[\psi]$, since in this case ψ' and ψ are identical up to renaming variables, and $\psi^+ = \psi'$.

In essence, the structure just defined is the canonical structure obtained by conjoining, to the PP-formula, the atom $U_{\sigma,\ell}(v_1, \dots, v_\ell)$, where the symbol $U_{\sigma,\ell}$ is a fresh one; and then existentially quantifying all free variables.

We now define a notion of *normalized* EP-formula.

▶ **Definition 18.** (extends [6, Definition 3]) Let ϕ be an EP-formula over vocabulary σ and whose free variables are $v_1, \ldots, v_\ell$. We say that ϕ is *normalized* if it is equal to a disjunction $\bigvee_{i\in[m]} \psi_i$ (with $m \geq 0$) where:

- each ψ_i is a prenex PP-formula,
- for each $i \in [m]$, the structure $\mathbf{C}[\psi_i; \sigma; v_1, \ldots, v_\ell]$ is a core, and
- for each $i, j \in [m]$ with $i \neq j$, it holds that $\mathbf{C}[\psi_i; \sigma; v_1, \ldots, v_\ell] \not\to \mathbf{C}[\psi_j; \sigma; v_1, \ldots, v_\ell]$.

▶ Proposition 19. There exists an algorithm that, given as input an EP-formula ϕ, outputs a normalized EP-formula $\phi' = \bigvee_{i\in[m]} \psi'_i$ that is logically equivalent to ϕ, and such that (for any $k \geq 0$) if ϕ is EP^k-expressible, then ψ'_i is PP^k-expressible for every $i \in [m]$.

This proposition was observed in the particular case of sentences by [6, Section 3]; the construction is, in essence, a solution to a classic exercise in database theory [1, Exercise 6.14(c)].

Proof. Let σ denote the vocabulary of ϕ, and let $v_1, \ldots, v_\ell$ denote the free variables of ϕ.

By the proof of [6, Lemma 4(3)], ϕ can be syntactically transformed (while preserving logical equivalence) to a disjunction $\bigvee_{i\in[m]} \psi_i$ where each ψ_i is a PP-formula, via transformations that do not introduce variables. Hence, each ψ_i is PP^k-expressible assuming that ϕ was EP^k-expressible.

For each $i \in [m]$, we may further assume (by rewriting ψ_i if necessary) that $\psi_i = \psi'_i$, where ψ'_i is as defined in Definition 17. As this operation preserves logical equivalence, it does not affect whether or not the disjunction $\bigvee_{i\in[m]} \psi_i$ is EP^k-expressible.

Suppose that $i, j \in [m]$ and h are such that h is a homomorphism from $\mathbf{C}[\psi_i; \sigma; v_1, \ldots, v_\ell]$ to $\mathbf{C}[\psi_j; \sigma; v_1, \ldots, v_\ell]$. Due to the interpretation of $U_{\sigma,\ell}$ by these structures as $\{(v_1, \ldots, v_\ell)\}$, it holds that h fixes each of $v_1, \ldots, v_\ell$. We claim that when $\mathbf{B}$ is a structure and when $f : \{v_1, \ldots, v_\ell\} \to B$ is an assignment, it is straightforward to verify that $\mathbf{B}, f \models \psi_j$ implies $\mathbf{B}, f(h) \models \psi_i$, which in turn implies $\mathbf{B}, f \models \psi_i$.

In the case that there exists such a homomorphism for $i, j \in [m]$ with $i \neq j$, we have that ψ_j entails ψ_i, and hence removing ψ_j from the disjunction preserves logical equivalence of the disjunction. In the case that there exists such a homomorphism that is not surjective for $i = j$, then set ψ to be the modification of ψ_i where one removes all atoms whose variables are not in the image of h, as well as the quantifications of such variables; then, we have $\mathbf{C}[\psi_i; \sigma; v_1, \ldots, v_\ell] \to \mathbf{C}[\psi; \sigma; v_1, \ldots, v_\ell]$; from this, it follows that ψ and ψ_i are logically equivalent, and ψ_i can be replaced with ψ in the disjunction. By iteratively performing such removals and replacements until none can be performed, a normalized EP-formula is obtained. Moreover, these removals and replacements preserve PP^k-expressibility of all disjuncts. ◀

We now define the notion of *combinatorial width*. Although we only define it directly on normalized EP-formulas, the definition can be naturally extended to all EP-formulas in light of Proposition 19.

▶ **Definition 20.** Let $\phi = \bigvee_{i\in[m]} \psi_i$ be a normalized EP-formula with free variables $v_1, \ldots, v_\ell$ and over vocabulary σ. We define $\mathsf{comb\text{-}width}(\phi) = \max_{i\in[m]}(\mathsf{tw}(\mathbf{C}[\psi_i; \sigma; v_1, \ldots, v_\ell]) + 1)$. Note that, in the case that ϕ is a sentence (equivalently, when $\ell = 0$), it holds that $\mathsf{comb\text{-}width}(\phi) = \max_{i\in[m]}(\mathsf{tw}(\mathbf{C}[\psi_i]) + 1)$.

The notion of the *combinatorial width* of a normalized sentence was studied implicitly in previous work; see Proposition 3.4 of [10] and Section 3 of [6]. The following theorem was known [16, 6].

▶ **Theorem 21.** *Let $\phi = \bigvee_{i \in [m]} \psi_i$ be a normalized* EP*-sentence. Let $k =$ comb-width(ϕ). The sentence ϕ is* EP^k*-expressible, but (assuming $k > 0$) is not* EP^{k-1}*-expressible.*

Proof. We will use the fact, which follows from [16, Theorem 12], that (when $w \geq 1$) a PP-sentence ψ has $\mathsf{tw}(\mathsf{core}(\mathbf{C}[\psi])) < w$ if and only if ψ is PP^w-expressible.

We have that ϕ is EP^k-expressible via this fact, since each disjunct ψ_i has $\mathsf{tw}(\mathbf{C}[\psi_i]) <$ comb-width(ϕ).

For the non-expressibility result, suppose that ϕ is EP^n-expressible; we prove that $n \geq k$. It follows from Proposition 19 that ϕ is logically equivalent to a normalized EP-formula that is a disjunction $\bigvee_{i \in [m']} \psi'_i$ of PP^n-formulas. We have that each $\mathbf{C}[\psi'_i]$ is a core; by the fact, we obtain that $\mathsf{tw}(\mathbf{C}[\psi'_i]) + 1 \leq n$. By Lemma 4(2) of [6], it follows that $k = \max_{i \in [m']}(\mathsf{tw}(\mathbf{C}[\psi'_i]) + 1)$. We thus have $k \leq n$. ◀

5 Main Theorems and Consequences

We first prove our number-of-variables characterization for EP-sentences.

▶ **Theorem 22.** *Let $\phi = \bigvee_{i \in [m]} \psi_i$ be a normalized* EP*-sentence; let $w =$ comb-width(ϕ). The sentence ϕ is* EP^w*-expressible, but (assuming $w > 1$) is not* FO^{w-1}*-expressible.*

This theorem is transparently seen to be a strengthening of Theorem 21 (when the stated assumption holds).

Proof. That the sentence ϕ is EP^w-expressible follows directly from Theorem 21, so we prove that ϕ is not FO^{w-1}-expressible. Choose $i \in [m]$ such that comb-width$(\phi) = \mathsf{tw}(\mathbf{C}[\psi_i]) + 1$; set $\mathbf{A} = \mathbf{C}[\psi_i]$. By the definition of *normalized*, we have that $\mathbf{A}$ is a core. Let $\mathbf{B}$, $\mathbf{B}'$ be the structures provided by Theorem 6 relative to $\mathbf{A}$ and $k = \mathsf{tw}(\mathbf{A})$; note that $k = w - 1$.

We show in the next paragraph that $\mathbf{B}' \models \phi$ and $\mathbf{B} \not\models \phi$. Then, since $\mathbf{B}'$ and $\mathbf{B}$ are FO^k-indistinguishable by Theorem 6, and since ϕ distinguishes between $\mathbf{B}'$ and $\mathbf{B}$, it follows that ϕ is not FO^k-expressible.

We prove the claim. We have that $\mathbf{A} \to \mathbf{B}'$, hence $\mathbf{B}' \models \psi_i$ by Theorem 2, and $\mathbf{B}' \models \phi$. Now, assume for a contradiction that $\mathbf{B} \models \phi$. Then $\mathbf{B} \models \psi_j$ for some $j \in [m]$. Then $\mathbf{C}[\psi_j] \to \mathbf{B}$ by Theorem 2. We have $\mathbf{B} \to \mathbf{A} = \mathbf{C}[\psi_i]$ by Theorem 6. Hence $\mathbf{C}[\psi_j] \to \mathbf{C}[\psi_i]$, so $i = j$ by the hypothesis that ϕ is normalized. But we also have $\mathbf{A} = \mathbf{C}[\psi_i] \not\to \mathbf{B}$, hence $i \neq j$, a contradiction. ◀

We now extend the previous theorem to address all normalized EP-formulas. The following is our primary theorem.

▶ **Theorem 23** (Primary theorem)**.** *Let $\phi = \bigvee_{i \in [m]} \psi_i$ be a normalized* EP*-formula; let $w =$ comb-width(ϕ). The formula ϕ is* EP^w*-expressible, but (assuming $w > 1$) is not* FO^{w-1}*-expressible.*

Proof. Let σ denote the vocabulary of ϕ. Let $v_1, \ldots, v_\ell$ denote the free variables of ϕ. We assume that $\ell \geq 1$ (otherwise, the theorem follows from Theorem 22).

We first establish that ϕ is EP^w-expressible. It suffices to prove that, for each $i \in [m]$, the formula ψ_i is expressible using $b = \mathsf{tw}(\mathbf{C}[\psi_i; \sigma; v_1, \ldots, v_\ell]) + 1$ variables. By definition of treewidth, there exists a tree decomposition of $\mathbf{C}[\psi_i; \sigma; v_1, \ldots, v_\ell]$ where each bag has size b or less. It is readily verified that this tree decomposition is a tree decomposition of $\mathbf{C}[\exists v_1 \ldots \exists v_\ell \psi'_i]$ (where ψ'_i is derived from ψ_i as described in Definition 17) which has a bag B_s with $v_1, \ldots, v_\ell \in B_s$. The result now essentially follows from the argument of

Lemma 5.11 of [11, 12]. We give an explanation for the sake of completeness. Add a vertex r adjacent to s to the tree and define $B_r = \{v_1, \ldots, v_\ell\}$. Let T be the resulting object, which is straightforwardly verified to be a tree decomposition of $\mathbf{C}[\exists v_1 \ldots \exists v_\ell \psi_i']$ where each bag has size b or less.

We now describe how to construct the desired formula. Each variable of ψ_i', other than $v_1, \ldots, v_\ell$, will be existentially quantified exactly once in the formula; and all atoms of ψ_i', will appear in the formula. In this way, the constructed formula will be clearly logically equivalent to ψ_i', and hence ψ_i.

- Root the tree T at r; note that r has a single child, s.
- For each non-root vertex t of T, we define a PP-formula θ_t inductively, as follows. Define θ_t as $\exists w_1 \ldots w_m \theta_t'$ where $w_1, \ldots, w_m$ is a list of variables that are in the bag B_t of t but not in the bag of t's parent, and θ_t' is the conjunction of θ_u over all children u of t and all atoms $R(z_1, \ldots, z_k)$ of ψ_i' where $\{z_1, \ldots, z_k\} \subseteq B_t$.
- The desired formula is θ_s.

Observe (by induction) that, for each non-root vertex t of T, it holds that $\mathsf{free}(\theta_t') \subseteq B_t$ and $\mathsf{free}(\theta_t) \subseteq B_p$ where p is the parent of t. It follows that each formula θ_t, and in particular θ_s, has width $\leq b$, and is thus PP^b-expressible.

We now establish that ϕ is not FO^{w-1}-expressible.

If $w \leq \ell$, then it is straightforward to verify that the formula ϕ is not FO^{w-1}-expressible, since $w-1$ is strictly lower than the number of free variables of ϕ. (We can remark that $w \geq \ell$, since each structure $\mathbf{C}[\psi_i; \sigma; v_1, \ldots, v_\ell]$ interprets $U_{\sigma,\ell}$ as $\{(v_1, \ldots, v_\ell)\}$, and so the treewidth of each such structure plus 1 is ℓ or more.)

We thus assume in the sequel that $w > \ell$. For each $i \in [m]$, we may assume without loss of generality that each ψ_i is prenexed and that in each ψ_i, none of the variables $v_1, \ldots, v_\ell$ are quantified. Define ψ_i^+ as in Definition 17; we then have $\psi_i^+ = \exists v_1 \ldots \exists v_\ell (U_{\sigma,\ell}(v_1, \ldots, v_\ell) \wedge \psi_i)$. Define ϕ^+ as $\bigvee_{i \in [m]} \psi_i^+$. We have that $\mathbf{C}[\psi_i; \sigma; v_1, \ldots, v_\ell] = \mathbf{C}[\psi_i^+]$. We have

$$\mathsf{comb\text{-}width}(\phi) = \max_{i \in [m]} (\mathsf{tw}(\mathbf{C}[\psi_i; \sigma; v_1, \ldots, v_\ell]) + 1) = \max_{i \in [m]} (\mathsf{tw}(\mathbf{C}[\psi_i^+]) + 1) = \mathsf{comb\text{-}width}(\phi^+),$$

where the last equality holds by the note in Definition 20.

Observe that

$$\begin{aligned} \phi^+ &= \bigvee_{i \in [m]} \psi_i^+ = \bigvee_{i \in [m]} \exists v_1 \ldots \exists v_\ell (U_{\sigma,\ell}(v_1, \ldots, v_\ell) \wedge \psi_i) \\ &\equiv \exists v_1 \ldots \exists v_\ell \bigvee_{i \in [m]} (U_{\sigma,\ell}(v_1, \ldots, v_\ell) \wedge \psi_i) \equiv \exists v_1 \ldots \exists v_\ell (U_{\sigma,\ell}(v_1, \ldots, v_\ell) \wedge (\bigvee_{i \in [m]} \psi_i)) \\ &= \exists v_1 \ldots \exists v_\ell (U_{\sigma,\ell}(v_1, \ldots, v_\ell) \wedge \phi). \end{aligned}$$

Suppose, for a contradiction, that ϕ is FO^{w-1}-expressible. Then, ϕ can be expressed just using the variables $x_1, \ldots, x_{w-1}$ (recall the assumption that $w > \ell$, which gives $w - 1 \geq \ell$). Since $\phi^+ \equiv \exists v_1 \ldots \exists v_\ell (U_{\sigma,\ell}(v_1, \ldots, v_\ell) \wedge \phi)$, this immediately implies that ϕ^+ can be expressed just using the variables $x_1, \ldots, x_{w-1}$, and that ϕ^+ is FO^{w-1}-expressible. As $w = \mathsf{comb\text{-}width}(\phi) = \mathsf{comb\text{-}width}(\phi^+)$, this contradicts Theorem 22. ◀

▶ **Corollary 24.** *For each $k \geq 1$, FO^k-expressibility and EP^k-expressibility coincide on existential positive formulas; that is, an existential positive formula is FO^k-expressible if and only if it is EP^k-expressible.*

Proof. Clearly, EP^k-expressibility implies FO^k-expressibility. For the other direction, suppose that an existential positive formula ϕ is FO^k-expressible. Let ϕ' be a normalized EP-formula that is logically equivalent to ϕ (which exists by Proposition 19). We have that ϕ' is FO^k-expressible. We have that $k \geq \mathsf{comb\text{-}width}(\phi')$; if not, we would have $1 \leq k < \mathsf{comb\text{-}width}(\phi')$, contradicting Theorem 23. It follows from Theorem 21 that ϕ' is EP^k-expressible. ◀

▶ **Corollary 25.** *The problem of deciding* FO^k*-expressibility of* EP*-sentences is* Π_2^p*-complete. By this, we refer to the problem of deciding, given an* EP*-sentence* ϕ *and an integer* $k \geq 1$*, whether* ϕ *is* FO^k*-expressible.*

Proof. The problem of deciding, given an EP-sentence ϕ and an integer $k \geq 1$, whether ϕ is EP^k-expressible, is Π_2^p-complete by [6, Theorem 6]. The present corollary thus follows from Corollary 24. ◀

We conclude the article by explaining how the main result of [3] follows from our development. We first present the necessary definitions.

▶ **Definition 26.** Let σ be a relational vocabulary and let $\mathbf{A}$ and $\mathbf{B}$ be σ-structures. A *partial homomorphism from* $\mathbf{A}$ *to* $\mathbf{B}$ is a partial function h from A to B such that, for all $R \in \sigma$ and all $a_1, \ldots, a_{\mathsf{ar}(R)} \in \mathsf{dom}(h)$, it holds that $(a_1, \ldots, a_{\mathsf{ar}(R)}) \in R^{\mathbf{A}}$ implies $(h(a_1), \ldots, h(a_{\mathsf{ar}(R)})) \in R^{\mathbf{B}}$.

▶ **Definition 27.** [21] Let $k \geq 0$. A *duplicator winning strategy in the existential* k*-pebble game on* $(\mathbf{A}, \mathbf{B})$ is a family $\mathcal{S}$ of partial homomorphisms h from $\mathbf{A}$ to $\mathbf{B}$ with $|\mathsf{dom}(h)| \leq k$ such that:

(E1) $\emptyset \to \emptyset$ is in $\mathcal{S}$.

(E2) If $h \in \mathcal{S}$ and $|\mathsf{dom}(h)| < k$, then for every $a \in A$ there exists $b \in B$ such that $h \cup \{(a, b)\}$ is in $\mathcal{S}$.

(E3) If $h \in \mathcal{S}$ and $a \in \mathsf{dom}(h)$, then $h|_{\mathsf{dom}(h) \setminus \{a\}}$ is in $\mathcal{S}$.

▶ **Theorem 28.** *(Main theorem of [3]) Let* $\mathbf{A}$ *be a core on the relational vocabulary* σ *such that* $\mathsf{tw}(\mathbf{A}) \geq k \geq 1$*. There exists a* σ*-structure* $\mathbf{B}$ *such that* $\mathbf{A} \not\to \mathbf{B}$ *and there exists a duplicator winning strategy in the existential* k*-pebble game on* $(\mathbf{A}, \mathbf{B})$*.*

To prove Theorem 28, we will make use of the following transitivity property.

▶ **Lemma 29.** *Let* $k \geq 0$*. If there exist duplicator winning strategies in the existential* k*-pebble games on* $(\mathbf{A}, \mathbf{B})$ *and* $(\mathbf{B}, \mathbf{C})$*, then there exists a duplicator winning strategy in the existential* k*-pebble game on* $(\mathbf{A}, \mathbf{C})$*.*

Proof. Suppose that G is a duplicator winning strategy in the existential k-pebble game on $(\mathbf{A}, \mathbf{B})$, and that H is a duplicator winning strategy in the existential k-pebble game on $(\mathbf{B}, \mathbf{C})$. Let F be the set containing each mapping of the form $h(g)$ where $g \in G$, $h \in H$, and $\mathsf{dom}(h) = g(A)$. It is straightforward to verify that F is a duplicator winning strategy in the existential k-pebble game on $(\mathbf{A}, \mathbf{C})$. ◀

We now give a proof of Theorem 28 using the construction of this article.

Proof of Theorem 28. Let $\mathbf{A}$ be a σ-structure. By hypothesis, $\mathbf{A}$ is a core and $\mathsf{tw}(\mathbf{A}) \geq k$; so, by Theorem 6 there exist σ-structures $\mathbf{B}$ and $\mathbf{B}'$, such that $\mathbf{A} \not\to \mathbf{B}$, $\mathbf{A} \to \mathbf{B}'$, and (via Theorem 5) the duplicator has a winning strategy in the k-pebble game on $\mathbf{B}$ and $\mathbf{B}'$.

Therefore the duplicator has a winning strategy $\mathcal{S}$ in the existential k-pebble game on $(\mathbf{B}', \mathbf{B})$, because duplicator winning strategies in the k-pebble game on $\mathbf{B}$ and $\mathbf{B}'$ are also duplicator winning strategies in the existential k-pebble game on $(\mathbf{B}', \mathbf{B})$. Moreover, there exists a homomorphism g from $\mathbf{A}$ to $\mathbf{B}'$. It is straightforward to verify that the family containing each restriction of g to a subset $S \subseteq A$ with $|S| \leq k$ is a duplicator winning strategy in the existential k-pebble game on $(\mathbf{A}, \mathbf{B}')$. It follows, from Lemma 29 that there exists a duplicator winning strategy existential k-pebble game on $(\mathbf{A}, \mathbf{B})$. ◀

Acknowledgments. The first author was supported by the FWF Austrian Science Fund (Parameterized Compilation, P26200). The second author was supported by the Spanish Project MINECO COMMAS TIN2013-46181-C2-R, Basque Project GIU15/30, and Basque Grant UFI11/45.

References

1 S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases.* Addison-Wesley, 1995.

2 I. Adler and M. Weyer. Tree-Width for First-Order Formulae. *Logical Methods in Computer Science*, 8(1), 2012.

3 A. Atserias, A.A. Bulatov, and V. Dalmau. On the power of k -consistency. In *Proceedings of the 34th International Colloquium on Automata, Languages and Programming (ICALP 2007)*, pages 279–290, 2007.

4 P. Barceló, G. Gottlob, and A. Pieris. Semantic Acyclicity under Constraints. In *Proceedings of the 35th Symposium on Principles of Database Systems (PODS 2016)*, pages 343–354, 2016.

5 J. Barwise. On Moschovakis Closure Ordinals. *Journal of Symbolic Logic*, 42:292–296, 1977.

6 S. Bova and H. Chen. The Complexity of Width Minimization for Existential Positive Queries. In *Proceedings of the 17th International Conference on Database Theory (ICDT 2014)*, pages 235–244, 2014.

7 Simone Bova and Hubie Chen. How Many Variables Are Needed to Express an Existential Positive Query? In *20th International Conference on Database Theory (ICDT) 2017*, pages 9:1–9:16, 2017.

8 A.K. Chandra and P.M. Merlin. Optimal Implementation of Conjunctive Queries in Relational Data Bases. In *Proceedings of the 9th Symposium on Theory of Computing (STOC 1977)*, pages 77–90, 1977.

9 H. Chen. Beyond Q-Resolution and Prenex Form: A Proof System for Quantified Constraint Satisfaction. *Logical Methods in Computer Science*, 10(4), 2014.

10 H. Chen. On the Complexity of Existential Positive Queries. *ACM Transactions on Computational Logic*, 15(1), 2014.

11 H. Chen. The Tractability Frontier of Graph-Like First-Order Query Sets. In *Proceedings of the Joint Meeting of the 23rd Conference on Computer Science Logic and the 29th Symposium on Logic in Computer Science (CSL-LICS 2014)*, pages 31:1–31:9, 2014.

12 H. Chen. The Tractability Frontier of Graph-Like First-Order Query Sets. *CoRR*, abs/1407.3429, 2014.

13 H. Chen and V. Dalmau. Beyond Hypertree Width: Decomposition Methods Without Decompositions. In *Proceedings of the 11th International Conference on Principles and Practice of Constraint Programming (CP 2015)*, pages 167–181, 2005.

14 H. Chen and V. Dalmau. From Pebble Games to Tractability: An Ambidextrous Consistency Algorithm for Quantified Constraint Satisfaction. In *Proceedings of the 19th Conference on Computer Science Logic (CSL 2005)*, pages 232–247, 2005.

15 H. Chen and M. Müller. One Hierarchy Spawns Another: Graph Deconstructions and the Complexity Classification of Conjunctive Queries. In *Proceedings of the Joint Meeting of the 23rd Conference on Computer Science Logic and the 29th Symposium on Logic in Computer Science (CSL-LICS 2014)*, pages 32:1–32:10, 2014.

16 V. Dalmau, P.G. Kolaitis, and M.Y. Vardi. Constraint Satisfaction, Bounded Treewidth, and Finite-Variable Logics. In *Proceedings of the 8th International Conference on Principles and Practice of Constraint Programming (CP 2002)*, pages 310–326, 2002.

17 R. Diestel. *Graph Theory*. Springer, 2012.

18 M. Grohe. The Complexity of Homomorphism and Constraint Satisfaction Problems Seen from the Other Side. *Journal of the ACM*, 54(1), 2007.

19 P. Hell and J. Nesetril. The Core of a Graph. *Discrete Mathematics*, 109:117–126, 1992.

20 N. Immerman. Upper and Lower Bounds for First-Order Expressibility. *Journal of Computer and System Sciences*, 25:76–98, 1982.

21 P.G. Kolaitis and M.Y. Vardi. On the Expressive Power of Datalog: Tools and a Case Study. *Journal of Computer and System Sciences*, 51(1):110–134, 1995.

22 M.Y. Vardi. On the Complexity of Bounded-Variable Queries. In *Proceedings of the 14th Symposium on Principles of Database Systems (PODS 1995)*, pages 266–276, 1995.