The Journal of Symbolic Logic
Volume 89, Number 2, June 2024

FORBIDDEN INDUCED SUBGRAPHS AND THE ŁOŚ–TARSKI THEOREM

YIJIA CHEN AND JÖRG FLUM

Abstract. Let $\mathcal{C}$ be a class of finite and infinite graphs that is closed under induced subgraphs. The well-known Łoś–Tarski Theorem from classical model theory implies that $\mathcal{C}$ is definable in first-order logic by a sentence φ if and only if $\mathcal{C}$ has a finite set of forbidden induced finite subgraphs. This result provides a powerful tool to show nontrivial characterizations of graphs of small vertex cover, of bounded tree-depth, of bounded shrub-depth, etc. in terms of forbidden induced finite subgraphs. Furthermore, by the Completeness Theorem, we can compute from φ the corresponding forbidden induced subgraphs. This machinery fails on finite graphs as shown by our results:

– There is a class $\mathcal{C}$ of finite graphs that is definable in first-order logic and closed under induced subgraphs but has no finite set of forbidden induced subgraphs.
– Even if we only consider classes $\mathcal{C}$ of finite graphs that can be characterized by a finite set of forbidden induced subgraphs, such a characterization cannot be computed from a first-order sentence φ that defines $\mathcal{C}$ and the size of the characterization cannot be bounded by $f(|\varphi|)$ for any computable function f.

Besides their importance in graph theory, the above results also significantly strengthen similar known theorems for arbitrary structures.

§1. Introduction. Many classes of graphs can be defined by a finite set of forbidden induced finite subgraphs. One of the simplest examples is the class of graphs of bounded degree. Let $d \geq 1$ and let the set $\mathcal{F}_d$ consist of all graphs with vertex set $\{1, \dots, d+2\}$ and maximum degree $d+1$. Then a graph G has degree at most d if and only if no graph in $\mathcal{F}_d$ is isomorphic to an induced subgraph of G. Less trivial examples include classes of graphs of small vertex cover (attributed to Lovász [12]), of bounded tree-depth [7], and of bounded shrub-depth [16]. As a matter of fact, understanding forbidden induced subgraphs for those graph classes is an important question in structural graph theory [10, 14, 15, 28]. However, a straightforward adaptation of a result in [13] shows that it is in general impossible to compute the forbidden induced subgraphs from a description of classes of finite graphs by Turing machines.

Łoś [19] and Tarski [26] proved the first so-called preservation theorem of classical model theory. In its simplest form it says for classes of graphs that the class $\textsc{Graph}(\varphi)$ of finite and infinite graphs that are models of a sentence φ of first-order logic (FO) is closed under induced subgraphs (or, that φ is preserved under induced subgraphs) if and only if there is a universal FO-sentence μ with $\textsc{Graph}(\varphi) = \textsc{Graph}(\mu)$. Recall

Received November 20, 2021.
2020 *Mathematics Subject Classification.* Primary 03B70, 03C13.
Key words and phrases. preservation theorem for graphs, finite model theory.

0022-4812/24/8902-0004
DOI:10.1017/jsl.2023.99

that a universal sentence μ is a sentence of the form $\forall x_1 \dots \forall x_k \, \mu_0$, where μ_0 is quantifier-free.

For a class $\mathcal{F}$ of graphs let $\textsc{Forb}(\mathcal{F})$ consist of all graphs that do not contain an induced subgraph isomorphic to a graph in $\mathcal{F}$. For any class $\mathcal{C}$ of graphs closed under isomorphism and under induced subgraphs we have $\mathcal{C} = \textsc{Forb}(\mathcal{F})$ where $\mathcal{F}$ consists of all graphs not in $\mathcal{C}$. Observe that $\mathcal{F}$ is an infinite class that contains infinite graphs if $\mathcal{C}$ is not the class of all graphs (later on we only will consider $\textsc{Forb}(\mathcal{F})$ for *sets* $\mathcal{F}$).

It is folklore (see, e.g., [20]) that for a class $\mathcal{C}$ of graphs its definability by a universal sentence of first-order logic is equivalent to its characterization by finitely many forbidden induced finite subgraphs, i.e., equivalent to $\mathcal{C} = \textsc{Forb}(\mathcal{F})$ for some finite set $\mathcal{F}$ of finite graphs. In fact, for a universal sentence $\mu := \forall x_1 \dots \forall x_k \, \mu_0$ with quantifier-free μ_0 we have (see Proposition 2.2)

$$\textsc{Graph}(\mu) = \textsc{Forb}\big(\mathcal{F}_k(\mu)\big). \tag{1}$$

Here for any FO-sentence φ and $k \geq 1$ by $\mathcal{F}_k(\varphi)$ we denote the class of graphs that are models of $\neg\varphi$ and whose universe is $\{1, \dots, \ell\}$ for some ℓ with $1 \leq \ell \leq k$. Clearly, $\mathcal{F}_k(\varphi)$ is finite.

We say that a class $\mathcal{C}$ of finite and infinite graphs is *definable by a finite set of forbidden induced finite subgraphs* if there is a finite set $\mathcal{F}$ of finite graphs such that $\mathcal{C} = \textsc{Forb}(\mathcal{F})$. Hence the Łoś–Tarski Theorem (for classes of graphs) can be restated in the form:

For a class $\mathcal{C}$ of finite and infinite graphs the following are equivalent:

(i) $\mathcal{C}$ is closed under induced subgraphs and FO-axiomatizable.
(ii) $\mathcal{C}$ is axiomatizable by a universal sentence.
(iii) $\mathcal{C}$ is definable by a finite set of forbidden induced finite subgraphs.

This version of the Łoś–Tarski Theorem is already contained, at least implicitly, in the article [27] of Vaught published in 1954.

Clearly (1) implies $\textsc{Graph}_{\text{fin}}(\mu) = \textsc{Forb}_{\text{fin}}\big(\mathcal{F}_k(\mu)\big)$ where $\textsc{Graph}_{\text{fin}}(\mu)$ and $\textsc{Forb}_{\text{fin}}\big(\mathcal{F}_k(\mu)\big)$ denote the class of finite graphs in $\textsc{Graph}(\mu)$ and in $\textsc{Forb}\big(\mathcal{F}_k(\mu)\big)$, respectively. Hence the equivalence between (ii) and (iii) holds too if we only consider classes of finite graphs.

Note that we have repeatedly mentioned that in the Łoś–Tarski Theorem graphs are allowed to be infinite. This is not merely a technicality. In [3], to obtain the forbidden induced subgraph characterization of graphs of bounded shrub-depth using the Łoś–Tarski Theorem, one simple but vital step is to extend the notion of shrub-depth to infinite graphs. Indeed, Tait [25] exhibited a class $\mathcal{C}$ of finite structures (which might be understood as colored directed graphs) that is closed under induced substructures and FO-axiomatizable. Yet, $\mathcal{C}$ is not definable by any universal sentence, thus cannot be characterized by a finite set of forbidden induced finite substructures. In [1] the authors present a class $\mathcal{C}$ of finite directed graphs with loops with the same properties, i.e., $\mathcal{C}$ is closed under induced substructures and FO-axiomatizable (even by a sentence without equality) but not axiomatizable by a universal sentence. Of course, the class $\mathcal{C}'$ of graphs in $\mathcal{C}$ is closed under induced subgraphs but $\mathcal{C}'$ is axiomatizable by a universal sentence (as $\mathcal{C}'$ is empty).

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

The first result of this paper strengthens the preceding "negative results" to graphs; more precisely we show the following theorem:

THEOREM 1.1. *There is a class $\mathcal{C}$ of finite graphs that is closed under induced subgraphs and* FO*-axiomatizable but not definable by a finite set of forbidden induced finite subgraphs* (*i.e., there is no finite set $\mathcal{F}$ of finite graphs such that $\mathcal{C} = \text{FORB}_{\text{fin}}(\mathcal{F})$*).

Even though we are interested in structural and algorithmic results for classes of *finite* graphs, we see that in order to apply the Łoś–Tarski Theorem for such purposes we have to consider classes of finite and infinite graphs. Hence, in this paper "graph" means finite or infinite graph. As in the preceding result we mention it explicitly if we only consider finite graphs.

Complementing Theorem 1.1 we show that it is even undecidable whether a given FO-definable class of finite graphs that is closed under induced subgraphs can be characterized by a finite set of forbidden induced finite subgraphs. More precisely, we prove:

THEOREM 1.2. *There is no algorithm that for any* FO*-sentence φ such that $\text{GRAPH}_{fin}(\varphi)$ is closed under induced subgraphs decides whether φ is equivalent to a universal sentence on finite graphs.*

For a first-order definable class of graphs closed under induced subgraphs, often it is preferable to have an explicit construction of a finite set of forbidden induced finite subgraphs. This however turns out to be difficult for many natural classes of graphs. Let us consider the k-vertex cover problem for a constant $k \geq 1$. It asks whether a given graph has a vertex cover (i.e., a set of vertices that contains at least one endpoint of every edge) of size at most k. The class of all YES-instances of this problem, finite and infinite, is closed under induced subgraphs and FO-axiomatizable by the FO-sentence

$$\varphi^k_{\text{VC}} := \varphi_{\text{GRAPH}} \wedge \exists x_1 \ldots \exists x_k \forall y \forall z \Big(Eyz \rightarrow \bigvee_{1 \leq \ell \leq k} (x_\ell = y \vee x_\ell = z) \Big),$$

where φ_{GRAPH} axiomatizes the class of graphs. Hence, by the Łoś–Tarski Theorem there is a universal sentence μ equivalent to φ^k_{VC}. As the reader will notice, it is by no means trivial to find such a μ. On the other hand, using the Completeness Theorem, we eventually will get such a μ. Then we can extract corresponding forbidden induced subgraphs from μ as in (1). For the reader familiar with parameterized complexity [8], to get a μ we can alternatively use that a graph with vertex cover of size at most k admits a kernel with at most k^2 edges. Observe that this approach involves the co-NP-hard problem of deciding whether an input graph *does not* contain a vertex cover of size at most k.

By [7], also the class of finite graphs of tree-depth at most k is definable by a finite set of forbidden induced finite subgraphs. However, forbidden induced subgraphs are only known for $k \leq 3$ [10].

We prove two "negative" results that explain the hardness of constructing forbidden induced subgraphs.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

THEOREM 1.3. *There is no algorithm that for any* FO*-sentence* φ *which is equivalent to a universal sentence* μ *on finite graphs computes such a* μ*. Or equivalently, there is no algorithm that for any* FO*-sentence* φ *such that*

$$\textsc{Graph}_{fin}(\varphi) = \textsc{Forb}_{fin}(\mathcal{F})$$

for a finite set $\mathcal{F}$ *of finite graphs computes such an* $\mathcal{F}$*.*

THEOREM 1.4. *Let* $f : \mathbb{N} \to \mathbb{N}$ *be a computable function. Then there is a class* $\mathcal{C}$ *of finite graphs and an* FO*-sentence* φ *such that*:

(i) $\mathcal{C} = \textsc{Graph}_{fin}(\varphi)$.
(ii) $\mathcal{C} = \textsc{Graph}_{fin}(\mu)$ *for some universal sentence* μ*, in particular* $\mathcal{C}$ *is closed under induced subgraphs.*
(iii) *For every universal sentence* μ *with* $\mathcal{C} = \textsc{Graph}_{fin}(\mu)$ *we have* $|\mu| \geq f(|\varphi|)$.

Theorem 1.3 significantly strengthens the aforementioned result of [13]: even if a class $\mathcal{C}$ of finite graphs definable by a finite set of forbidden induced finite subgraphs is given by an FO-sentence φ with $\mathcal{C} = \textsc{Graph}_{\text{fin}}(\varphi)$, instead of a (much more powerful) Turing machine deciding $\mathcal{C}$, we still cannot compute an appropriate finite set of forbidden induced finite subgraphs for $\mathcal{C}$ from φ. On top of it, Theorem 1.4 implies that the size of forbidden subgraphs for $\mathcal{C}$ cannot be bounded by any computable function in terms of the size of φ. There is an important precursor for Theorem 1.4:

THEOREM 1.5 (Gurevich's Theorem [17]). *Let* $f : \mathbb{N} \to \mathbb{N}$ *be computable. Then there is an* FO*-sentence* φ *such that the class* $\textsc{Mod}(\varphi)$ *of models of* φ *is closed under induced substructures but for every universal sentence* μ *with* $\textsc{Mod}_{fin}(\mu) = \textsc{Mod}_{fin}(\varphi)$ *we have* $|\mu| \geq f(|\varphi|)$.

Hence, Theorem 1.4 can be viewed as the graph-theoretic version of Theorem 1.5.

Besides its importance in graph theory, Theorem 1.4 is also relevant in the context of algorithmic model theory. For algorithmic applications, the Łoś–Tarski theorem provides a normal form (i.e., a universal sentence) for any FO-sentence preserved under induced substructures. In [5, Theorem 6.1], it is shown that on *labelled trees* there is no *elementary bound* on the length of the equivalent universal sentence in terms of the original one. We should point out that Theorem 1.4 is not comparable to this result, since our lower bound is uncomputable (and thus, much higher than non-elementary) while the classes of graphs we construct in the proof are dense (thus very far from trees).

Our technical contributions. For every vocabulary it is well-known that the class of structures of this vocabulary is FO-interpretable in the class of graphs (see, for example, [11]). Hence one might expect that Theorems 1.1 and 1.4 can be derived easily from Tait's Theorem and Gurevich's Theorem using the standard FO-interpretations. However, an easy analysis shows that those interpretations yield classes of graphs that are not closed under induced subgraphs. So we introduce the notion of *strongly existential interpretation* that translates any class of structures preserved under induced substructures and relevant to our investigations to a class of graphs closed under induced subgraphs. A lot of care is needed to construct strongly existential interpretations.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

Related research. Let us briefly mention some further results related to the Łoś–Tarski Theorem. Essentially one could divide them into three categories (a), (b), and (c).

(a) The *positive results* showing that for certain classes $\mathcal{C}$ of finite structures the analogue of the Łoś–Tarski Theorem holds if we restrict to structures in $\mathcal{C}$. For example, this is the case if $\mathcal{C}$ is the class of all finite structures of tree-width at most k for some $k \in \mathbb{N}$ [2] or if $\mathcal{C}$ is the class of all finite structures whose hypergraph satisfies certain properties [9].

(b) Both just mentioned papers contain also *negative results*, i.e., classes for which the analogue of the Łoś–Tarski Theorem fails. For example, in [2] this is shown for the class of finite planar graphs, a class not axiomatizable in FO (cf. Remark 5.7(b)).

(c) The third category contains generalizations of the Łoś–Tarski Theorem or of its failure on finite structures. For example, in [24] the authors for every $k \in \mathbb{N}$ derive a preservation theorem for Σ_2-sentences of the form $\exists x_1 \dots \exists x_k\, \mu$ with universal μ. For $k = 0$ it coincides with the Łoś–Tarski Theorem; see Remark 3.5 for the precise statement. The paper [18] contains a further extension of the Łoś–Tarski Theorem. In [6] the authors show that for every $n \geq 1$ there is a Π_{2n+1}-sentence whose class of finite models is closed under induced substructures and that is not equivalent to a Σ_{2n+1}-sentence in the finite.

Most classical preservation theorems fail in the finite (see [21] for an exception). The question whether a preservation theorem fails for finite graphs is specially relevant for the Łoś–Tarski Theorem due to its connection to forbidden induced subgraphs.

Organization of this paper. In Section 2 we fix some notation and recall or derive some results about universal sentences we need in this paper. In Section 3 we include a proof of Tait's result (essentially as done in [1]). Moreover, we prove a technical result (Proposition 3.11) that is an important tool in the proof of Gurevich's Theorem. We introduce the concept of strongly existential interpretation in Section 4 and show that the results of the preceding section remain true under such interpretations. We present an appropriate strongly existential interpretation for graphs (in Section 5). Hence, we get the results of Section 3 for graphs. In Section 6 we first derive Gurevich's Theorem and apply our interpretations to get the corresponding results for graphs. Finally, in Section 7, we prove that various problems related to our results are undecidable.

This paper is the full version of our conference paper [4].

§2. Preliminaries.

We denote by $\mathbb{N}$ the set of natural numbers greater or equal to 0. For $n \in \mathbb{N}$ let $[n] := \{1, 2, \dots, n\}$.

2.1. First-order logic FO. A *vocabulary* τ is a finite set of relation symbols. Each relation symbol has an *arity*. A *structure* $\mathcal{A}$ of vocabulary τ, or *τ-structure*, consists of a (finite or infinite) nonempty set A, called the *universe* of $\mathcal{A}$, and of an interpretation $R^{\mathcal{A}} \subseteq A^r$ of each r-ary relation symbol $R \in \tau$. If $\mathcal{A}$ and $\mathcal{B}$ are τ-structures, then $\mathcal{A}$ *is a substructure of* $\mathcal{B}$, denoted by $\mathcal{A} \subseteq \mathcal{B}$, if $A \subseteq B$ and $R^{\mathcal{A}} \subseteq R^{\mathcal{B}}$, and $\mathcal{A}$ *is an induced*

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

substructure of $\mathcal{B}$, denoted by $\mathcal{A} \subseteq_{\text{ind}} \mathcal{B}$, if $A \subseteq B$ and $R^{\mathcal{A}} = R^{\mathcal{B}} \cap A^r$, where r is the arity of R. A substructure $\mathcal{A}$ of $\mathcal{B}$ is *proper* if $\mathcal{A} \neq \mathcal{B}$. By $\text{STR}[\tau]$ ($\text{STR}_{\text{fin}}[\tau]$) we denote the class of all (of all finite) τ-structures.

> *If we speak of a class of structures, we assume that it is closed under isomorphism. On the other hand, note that every nonempty set of structures does not have this closure property.*

Formulas φ of *first-order logic* FO of vocabulary τ are built up from *atomic formulas* $x_1 = x_2$ and $Rx_1 \dots x_r$ (where $R \in \tau$ is of arity r and $x_1, x_2, \dots, x_r$ are variables) using the boolean connectives $\neg$, $\wedge$, and $\vee$ and the universal $\forall$ and existential $\exists$ quantifiers. A relation symbol R *is positive (negative) in* φ if all atomic subformulas $R \dots$ in φ appear in the scope of an *even (odd)* number of negation symbols. By the notation $\varphi(\bar{x})$ with $\bar{x} = x_1, \dots, x_e$ we indicate that the variables free in φ are among $x_1, \dots, x_e$. If $\mathcal{A}$ is a τ-structure and $a_1, \dots, a_e \in A$, then $\mathcal{A} \models \varphi(a_1, \dots, a_e)$ means that $\varphi(\bar{x})$ holds in $\mathcal{A}$ if x_i is interpreted by a_i for $i \in [e]$.

A *sentence* is a formula without free variables. For a sentence φ we denote by $\text{MOD}(\varphi)$ the class of models of φ and $\text{MOD}_{\text{fin}}(\varphi)$ is its subclass consisting of the finite models of φ. Sentences φ and ψ are *equivalent* if $\text{MOD}(\varphi) = \text{MOD}(\psi)$ and *finitely equivalent* if $\text{MOD}_{\text{fin}}(\varphi) = \text{MOD}_{\text{fin}}(\psi)$.

2.2. Graphs. Let $\tau_E := \{E\}$ with binary E. For *all* τ_E-structures we use the notation $G = (V(G), E(G))$ common in graph theory. Here $V(G)$, the universe of G, is the set of vertices, and $E(G)$, the interpretation of the relation symbol E, is the set of edges. The τ_E-structure $G = (V(G), E(G))$ is a *directed graph* if $E(G)$ does not contain loops, i.e., $(v, v) \notin E(G)$ for all $v \in V(G)$. If moreover $(u, v) \in E(G)$ implies $(v, u) \in E(G)$ for all pairs (u, v), then G is an (undirected) *graph*. We denote by GRAPH and $\text{GRAPH}_{\text{fin}}$ the class of graphs and the class of finite graphs, respectively. Furthermore, for an $\text{FO}[\tau_E]$-sentence φ by $\text{GRAPH}(\varphi)$ and $(\text{GRAPH}_{\text{fin}}(\varphi))$ we denote the class of graphs (and the class of finite graphs) that are models of φ.

2.3. Universal sentences and forbidden induced substructures. An FO-formula is *universal* if it is built up from atomic and negated atomic formulas by means of the connectives $\wedge$ and $\vee$ and the universal quantifier $\forall$. Often we say that a formula containing, for example, the connective $\rightarrow$ is universal if by replacing $\varphi \rightarrow \psi$ by $\neg\varphi \vee \psi$ (and "simple manipulations") we get an equivalent universal formula. Every universal sentence μ is equivalent to a sentence μ' of the form $\forall x_1 \dots \forall x_k \, \mu'_0$ for some $k \geq 1$ and some quantifier-free μ'_0; moreover the length $|\mu'|$ of μ' is at most $|\mu|$. If in the definition of universal formula we replace the universal quantifier by the existential one, we get the definition of an *existential formula*.

One easily verifies that the class of models of a set of universal sentence is closed under induced substructures. As already mentioned in the Introduction for classes of graphs, Łoś [19] and Tarski [26] proved the following theorem:

THEOREM 2.1 (Łoś–Tarski Theorem). *Let* τ *be a vocabulary and* φ *an* $\text{FO}[\tau]$*-sentence. Then* $\text{MOD}(\varphi)$ *is closed under induced substructures if and only if* φ *is equivalent to a universal sentence.*

We first recall the relationship between the axiomatizability of a class of structures by a universal sentence and its definability by a finite set of forbidden finite induced

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

substructures. We fix a vocabulary τ. Let $\mathcal{F}$ be a *set* of τ-structures and denote by $\text{FORB}(\mathcal{F})$ and $(\text{FORB}_{\text{fin}}(\mathcal{F}))$ the class of structures (of finite structures) that do not contain an induced substructure isomorphic to a structure in $\mathcal{F}$. Clearly for sets $\mathcal{F}$ and $\mathcal{F}'$ of τ-structures we have

$$\text{if } \mathcal{F} \subseteq \mathcal{F}', \text{then } \text{FORB}(\mathcal{F}') \subseteq \text{FORB}(\mathcal{F}). \tag{2}$$

Furthermore, for a class $\mathcal{C}$ of τ-structures closed under induced substructures one easily verifies that

$$\mathcal{C} = \text{FORB}(\text{“}\{\mathcal{A} \in \text{STR}[\tau] \mid \mathcal{A} \notin \mathcal{C}\}\text{”}) \quad \text{and} \quad \mathcal{C}_{\text{fin}} = \text{FORB}_{\text{fin}}(\text{“}\{\mathcal{A} \in \text{STR}_{\text{fin}}[\tau] \mid \mathcal{A} \notin \mathcal{C}\}\text{”}). \tag{3}$$

We have put the corresponding $\mathcal{F}$'s in brackets as they are not sets but classes. However for $\mathcal{C}_{\text{fin}}$ we can repair this by considering only structures whose universe is an initial segment of natural numbers, i.e.,

$$\mathcal{C}_{\text{fin}} = \text{FORB}_{\text{fin}}\big(\{\mathcal{A} \in \text{STR}_{\text{fin}}[\tau] \mid \mathcal{A} \notin \mathcal{C} \text{ and } A = [\ell] \text{ for some } \ell \geq 1\}\big).$$

So we see that for every class of finite structures a denumerable set $\mathcal{F}$ suffices. When is a finite $\mathcal{F}$ enough?

We say that a class $\mathcal{C}$ of τ-structures (of finite τ-structures) is *definable by a finite set of forbidden induced finite substructures* if there is a *finite* set $\mathcal{F}$ of *finite* structures such that $\mathcal{C} = \text{FORB}(\mathcal{F})$ $(\mathcal{C} = \text{FORB}_{\text{fin}}(\mathcal{F}))$. Recall that $\tau_E = \{E\}$ with binary E. The sentences

$$\varphi_{\text{DG}} := \forall x \neg Exx \text{ and } \varphi_{\text{GRAPH}} := \forall x \neg Exx \wedge \forall x \forall y (Exy \to Eyx)$$

axiomatize the classes of directed graphs and of graphs, respectively. Define the τ_E-structures $H_0 = \big(V(H_0), E(H_0)\big)$ and $H_1 = \big(V(H_1), E(H_1)\big)$ by

$$V(H_0) := \{1\},\ E(H_0) := \big\{(1,1)\big\} \text{ and } V(H_1) := \{1,2\},\ E(H_1) := \big\{(1,2)\big\}.$$

Then $\text{FORB}\big(\{H_0\}\big)$ and $\text{FORB}\big(\{H_0, H_1\}\big)$ are the class of directed graphs and the class of graphs, respectively, i.e., $\text{MOD}(\varphi_{\text{DG}}) = \text{FORB}\big(\{H_0\}\big)$ and $\text{MOD}(\varphi_{\text{GRAPH}}) = \text{FORB}\big(\{H_0, H_1\}\big)$.

The following result (Proposition 2.2) generalizes this simple fact and establishes the equivalence between axiomatizability by a universal sentence and definability by a finite set of forbidden induced finite substructures. For an arbitrary vocabulary τ, an FO[τ]-sentence φ, and $k \geq 1$ let

$$\mathcal{F}_k(\varphi) := \big\{\mathcal{A} \in \text{STR}[\tau] \;\big|\; \mathcal{A} \models \neg\varphi \text{ and } A = [\ell] \text{ for some } \ell \in [k]\big\}.$$

Thus, $\mathcal{F}_k(\varphi)$ is, up to isomorphism, the class of structures with at most k elements that fail to be a model of φ. Note that $\mathcal{F}_1(\varphi_{\text{DG}}) = \{H_0\}$. By (2) and (3) we have

$$\text{if } \text{MOD}(\varphi) \text{ is closed under induced substructures,} \quad \text{then } \text{MOD}(\varphi) \subseteq \text{FORB}(\mathcal{F}_k(\varphi)) \text{ for all } k \geq 1. \tag{4}$$

PROPOSITION 2.2. *For a class $\mathcal{C}$ of τ-structures and $k \geq 1$ the statements* (i) *and* (ii) *are equivalent.*

(i) $\mathcal{C} = \text{MOD}(\mu)$ *for some universal sentence $\mu := \forall x_1 \ldots \forall x_k\, \mu_0$ with quantifier-free μ_0.*

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

(ii) $\mathcal{C} = \text{FORB}(\mathcal{F})$ *for some finite set* $\mathcal{F}$ *of structures, all of at most* k *elements.*

If (i) *holds for* μ, *then* $\mathcal{C} = \text{FORB}(\mathcal{F}_k(\mu))$.

The main step of the proof of (ii) $\Rightarrow$ (i) is contained in the following lemma.

LEMMA 2.3. *Let* $\mathcal{A}$ *be a finite* τ*-structure and* $k := |A|$. *There is a universal sentence* $\mu_{\mathcal{A}\not\to} = \forall x_1 \dots \forall x_k \, \mu_0$ *with quantifier-free* μ_0 *such that for every* τ*-structure* $\mathcal{B}$,

$$\mathcal{B} \models \mu_{\mathcal{A}\not\to} \iff \mathcal{B} \text{ has no induced substructure isomorphic to } \mathcal{A},$$

i.e.,

$$\text{MOD}(\mu_{\mathcal{A}\not\to}) = \text{FORB}(\{\mathcal{A}\}). \tag{5}$$

PROOF. Let $A = \{a_1, \dots, a_k\}$. Let $\delta(x_1, \dots, x_k)$ be the conjunction of all literals (i.e., atomic or negated atomic formulas) $\lambda(x_1, \dots, x_k)$ such that $\mathcal{A} \models \lambda(a_1, \dots, a_k)$. Then for every τ-structure $\mathcal{B}$ and $b_1, \dots, b_k \in B$ we have

$$\mathcal{B} \models \delta(b_1, \dots, b_k) \iff \text{the clauses } \pi(a_i) = b_i \text{ for } i \in [k] \text{ define an isomorphism from } \mathcal{A} \text{ onto } [b_1, \dots, b_k]^{\mathcal{B}}$$

(recall that $[b_1, \dots, b_k]^{\mathcal{B}}$ denotes the substructure of $\mathcal{B}$ induced on $\{b_1, \dots, b_k\}$). Thus we can set

$$\mu_{\mathcal{A}\not\to} := \forall x_1 \dots \forall x_k \neg\delta(x_1, \dots, x_k). \qquad \dashv$$

PROOF OF PROPOSITION 2.2. (ii) $\Rightarrow$ (i): Let $\mathcal{C} = \text{FORB}(\mathcal{F})$ for some finite set $\mathcal{F}$ of structures, all of at most k elements. If $\mathcal{F}$ is empty, then $\mathcal{C} = \text{MOD}(\forall x \, x = x)$. Otherwise, by (ii) and (5),

$$\mathcal{C} = \text{MOD}\left(\bigwedge_{\mathcal{A}\in\mathcal{F}} \mu_{\mathcal{A}\not\to}\right).$$

As the conjunction of finitely many universal sentences of the form $\forall x_1 \dots \forall x_\ell \, \mu_0$ with quantifier-free sentence μ_0 and with $\ell \leq k$ is equivalent to such a sentence, we get the desired result.

(i) $\Rightarrow$ (ii): Let $\mathcal{C} = \text{MOD}(\mu)$ for μ as in (i). Then $\text{MOD}(\mu)$ is closed under induced substructures and hence, $\mathcal{C} \subseteq \text{FORB}\big(\mathcal{F}_k(\mu)\big)$ by (4). Now assume that $\mathcal{A} \notin \mathcal{C}$. Then $\mathcal{A} \models \neg\mu$ and hence there are $a_1, \dots, a_k \in A$ with $\mathcal{A} \models \neg\mu_0(a_1, \dots, a_k)$. For $\mathcal{B} := [a_1, \dots, a_k]^{\mathcal{A}}$ we have $\mathcal{B} \models \neg\mu_0(a_1, \dots, a_k)$ (as μ_0 is quantifier-free) and thus, $\mathcal{B} \models \neg\mu$. Therefore, $\mathcal{B}$ is isomorphic to a structure in $\mathcal{F}_k(\mu)$ and therefore, $\mathcal{A} \notin \text{FORB}\big(\mathcal{F}_k(\mu)\big)$. $\dashv$

COROLLARY 2.4. *Let* φ *be a* τ*-sentence and* $k \geq 1$. *Then*

$$\text{MOD}(\varphi) = \text{FORB}\big(\mathcal{F}_k(\varphi)\big) \iff \varphi \text{ is equivalent to a universal sentence of the form } \forall x_1 \dots \forall x_k \, \mu_0 \text{ with quantifier-free } \mu_0.$$

By (2) and (4) we get the following corollaries:

COROLLARY 2.5. *If* $\text{MOD}(\mu) = \text{FORB}\big(\mathcal{F}_k(\mu)\big)$ *for some universal* μ *and some* $k \geq 1$, *then* $\text{MOD}(\mu) = \text{FORB}\big(\mathcal{F}_\ell(\mu)\big)$ *for all* $\ell \geq k$.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

COROLLARY 2.6. *It is decidable whether two universal sentences are equivalent.*

PROOF. Let μ and μ' be universal sentences. W.l.o.g. we may assume that $\mu = \forall x_1 \dots \forall x_k\, \mu_0$ and $\mu' = \forall x_1 \dots \forall x_\ell\, \mu'_0$ with $1 \leq k \leq \ell$ and quantifier-free μ_0 and μ'_0. By Corollaries 2.4 and 2.5, we have

$$\mathrm{MOD}(\mu) = \mathrm{FORB}\big(\mathcal{F}_\ell(\mu)\big) \text{ and } \mathrm{MOD}(\mu') = \mathrm{FORB}\big(\mathcal{F}_\ell(\mu')\big).$$

Thus μ and μ' are equivalent if and only if $\mathcal{F}_\ell(\mu) = \mathcal{F}_\ell(\mu')$. The right-hand side of this equivalence is clearly decidable. ⊣

The last equivalence of the preceding proof shows the following:

COROLLARY 2.7. *For universal sentences μ and μ' we have*

$$\mu \text{ and } \mu' \text{ are equivalent} \iff \mu \text{ and } \mu' \text{ are finitely equivalent}.$$

The next result generalizes this corollary.

LEMMA 2.8. *Let Φ be a set of universal sentences and ν a Π_2-sentence, i.e., a sentence of the form $\forall x_1 \dots \forall x_k \exists y_1 \dots \exists y_\ell\, \nu_0$ for some $k, \ell \in \mathbb{N}$ and some quantifier-free ν_0.*
If $\Phi \models_{fin} \nu$, then $\Phi \models \nu$ and thus there exists a finite $\Phi_0 \subseteq \Phi$ with $\Phi_0 \models_{fin} \nu$.

PROOF. If not $\Phi \models \nu$, then there is a (finite or infinite) structure $\mathcal{A}$ with $\mathcal{A} \models \Phi \cup \{\neg\nu\}$. Note that $\neg\nu$ is equivalent to a sentence of the form

$$\exists x_1 \dots \exists x_k\, \mu$$

with universal μ. Choose elements $a_1, \dots, a_k$ with $\mathcal{A} \models \mu(a_1, \dots, a_k)$. Then $\mathcal{B} := [a_1, \dots, a_k]^{\mathcal{A}}$ is a model of $\Phi \cup \{\neg\nu\}$ and thus shows $\Phi \not\models_{\mathrm{fin}} \nu$. ⊣

As the class of all at most countable τ-structures shows, not every class closed under induced substructures is the class of models of a set of universal sentences. In contrast, for classes of finite structures, we have the following:

LEMMA 2.9. *Let $\mathcal{C}$ be a class of finite τ-structures closed under induced substructures and define the set of universal sentences $\Phi_{\mathcal{C}}$ by*

$$\Phi_{\mathcal{C}} := \{\mu_{\mathcal{A}\not\to} \mid \mathcal{A} \in \mathrm{STR}_{fin}[\tau] \text{ and } \mathcal{A} \notin \mathcal{C}\}.$$

Then,

$$\mathcal{C} = \mathrm{FORB}_{fin}(\text{“}\{\mathcal{A} \in \mathrm{STR}_{fin}[\tau] \mid \mathcal{A} \notin \mathcal{C}\}\text{”}) = \mathrm{MOD}_{fin}(\Phi_{\mathcal{C}}).$$

PROOF. The first equality immediately follows from (3) using the closure of $\mathcal{C}$ under induced substructures and the second equality follows from (5). ⊣

We use the two preceding results to prove (see [17]):

THEOREM 2.10 (Compton's Theorem). *Let $\mathcal{C}$ be a class of finite τ-structures closed under induced substructures and* FO*-axiomatizable by a Π_2-sentence ν. Then $\mathcal{C}$ is already axiomatizable by a universal sentence.*

PROOF. By assumption and the preceding lemma, we have $\mathcal{C} = \mathrm{MOD}_{\mathrm{fin}}(\nu) = \mathrm{MOD}_{\mathrm{fin}}(\Phi_{\mathcal{C}})$, in particular, $\Phi_{\mathcal{C}} \models_{\mathrm{fin}} \nu$. By Lemma 2.8 there is a finite subset Φ_0 of $\Phi_{\mathcal{C}}$ such that $\Phi_0 \models_{\mathrm{fin}} \nu$. Thus,

$$\mathcal{C} = \mathrm{MOD}_{\mathrm{fin}}(\Phi_{\mathcal{C}}) \subseteq \mathrm{MOD}_{\mathrm{fin}}(\Phi_0) \subseteq \mathrm{MOD}_{\mathrm{fin}}(\nu) = \mathcal{C}.$$

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

Hence, $\mathcal{C}$ is axiomatizable by the conjunction of the sentences in Φ_0, a universal sentence. ⊣

Recall that our main goal is to find a class of finite graphs with the properties (a)–(c).

(a) The class is closed under induced subgraphs.
(b) The class is FO-axiomatizable (by a single sentence).
(c) The class is not FO-axiomatizable by a universal sentence.

Compton's Theorem tells us that w.r.t. the quantifier prefix the simplest possible FO-axiomatization of such a class is by a Σ_2-sentence, i.e., by a sentence of the form $\exists x_1 \dots \exists x_k \forall y_1 \dots \forall y_\ell\, \rho$ with $k, \ell \in \mathbb{N}$ and quantifier-free ρ.

The following consequence of Corollary 2.2 will be used in the next section.

COROLLARY 2.11. *Let $m, k \in \mathbb{N}$ with $m > k$ and let ψ_0 and ψ_1 be* FO[τ]*-sentences. Assume that $\mathcal{A}$ is a finite model of $\psi_0 \wedge \psi_1$ with at least m elements and all its induced substructures with at most k elements are models of $\psi_0 \wedge \neg\psi_1$. Then $\psi_0 \wedge \neg\psi_1$ is not finitely equivalent to a universal sentence of the form $\mu := \forall x_1 \dots \forall x_k\, \mu_0$ with quantifier-free μ_0.*

PROOF. As there is no universal *sentence* μ as above for $k = 0$, we can assume $k \geq 1$. For a contradiction assume $\mathrm{MOD}_{\mathrm{fin}}(\psi_0 \wedge \neg\psi_1) = \mathrm{MOD}_{\mathrm{fin}}(\mu)$ for $\mu := \forall x_1 \dots \forall x_k\, \mu_0$ with quantifier-free μ_0. As $\mathrm{MOD}(\mu) = \mathrm{FORB}\big(\mathcal{F}_k(\mu)\big)$ by Proposition 2.2, we get (applying the finite equivalence of $\psi_0 \wedge \neg\psi_1$ and μ to obtain the last equality)

$$\mathrm{MOD}_{\mathrm{fin}}(\psi_0 \wedge \neg\psi_1) = \mathrm{MOD}_{\mathrm{fin}}(\mu) = \mathrm{FORB}_{\mathrm{fin}}\big(\mathcal{F}_k(\mu)\big) = \mathrm{FORB}_{\mathrm{fin}}\big(\mathcal{F}_k(\psi_0 \wedge \neg\psi_1)\big).$$

However, by the assumptions the structure $\mathcal{A}$ is not contained in $\mathrm{MOD}_{\mathrm{fin}}(\psi_0 \wedge \neg\psi_1)$ but in the class $\mathrm{FORB}_{\mathrm{fin}}(\mathcal{F}_k(\psi_0 \wedge \neg\psi_1))$. ⊣

REMARK 2.12. Let $\mathcal{C}$ be a class of τ-structures closed under induced substructures. For an FO[τ]-sentence φ we set $\mathrm{MOD}_{\mathcal{C}}(\varphi) := \{\mathcal{A} \in \mathcal{C} \mid \mathcal{A} \models \varphi\}$. We say that the Łoś-Tarski *Theorem holds for* $\mathcal{C}$ if for every FO[τ]-sentence φ such that the class $\mathrm{MOD}_{\mathcal{C}}(\varphi)$ is closed under induced substructures there is a universal sentence μ such that $\mathrm{MOD}_{\mathcal{C}}(\varphi) = \mathrm{MOD}_{\mathcal{C}}(\mu)$. The following holds:

> *Let $\mathcal{C}$ and $\mathcal{C}'$ be classes of τ-structures closed under induced substructures with $\mathcal{C}' \subseteq \mathcal{C}$. Furthermore assume that there is a universal sentence μ_0* such that *$\mathcal{C}' = \mathrm{MOD}_{\mathcal{C}}(\mu_0)$. If the* Łoś–Tarski *Theorem holds for $\mathcal{C}$, then it holds for $\mathcal{C}'$, too.*

In fact, for every FO[τ]-sentence φ we have $\mathrm{MOD}_{\mathcal{C}'}(\varphi) = \mathrm{MOD}_{\mathcal{C}}(\mu_0 \wedge \varphi)$. Hence, if $\mathrm{MOD}_{\mathcal{C}'}(\varphi)$ is closed under induced substructures, then by assumption there is a universal μ such that $\mathrm{MOD}_{\mathcal{C}}(\mu_0 \wedge \varphi) = \mathrm{MOD}_{\mathcal{C}}(\mu)$. Therefore, $\mathrm{MOD}_{\mathcal{C}'}(\varphi) = \mathrm{MOD}_{\mathcal{C}}(\mu_0 \wedge \varphi) = \mathrm{MOD}_{\mathcal{C}}(\mu \wedge \mu_0) = \mathrm{MOD}_{\mathcal{C}'}(\mu)$.

However, as examples mentioned in the Introduction show, in general the failure of the Łoś–Tarski Theorem on a class does not imply its failure on every subclass.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

§3. Basic ideas underlying the classical results. This section contains a proof of Tait's Theorem telling us that the analogue of the Łoś–Tarski-Theorem fails if we only consider finite structures. Afterwards we refine the argument to derive a generalization, namely Proposition 3.11, which is a key result to get Gurevich's Theorem.

Our counterexample to the Łoś–Tarski-Theorem on finite structures essentially is the one given by Alechina and Gurevich in [1], which in turn simplifies the one given by Gurevich and Shelah in [17]. All these counterexamples use the main idea of the original counterexample due to Tait [25].

We consider the vocabulary $\tau_0 := \{<, U_{\min}, U_{\max}, S\}$, where $<$ and S (the "successor relation") are binary relation symbols and $U_{\min}$ and $U_{\max}$ are unary.

Let φ_0 be the conjunction of the universal sentences:

- $\forall x \neg x < x$, $\quad \forall x \forall y (x < y \vee x = y \vee y < x)$, $\qquad \forall x \forall y \forall z ((x < y \wedge y < z) \rightarrow x < z)$, i.e., "$<$ is an ordering."
- $\forall x \forall y \big(U_{\min} x \rightarrow (x = y \vee x < y)\big)$, i.e., "every element in $U_{\min}$ is a minimum w.r.t. $<$."
- $\forall x \forall y \big(U_{\max} x \rightarrow (x = y \vee y < x)\big)$, i.e., "every element in $U_{\max}$ is a maximum w.r.t. $<$."
- $\forall x \forall y (Sxy \rightarrow x < y)$.
- $\forall x \forall y \forall z ((x < y \wedge y < z) \rightarrow \neg Sxz)$.

Note that in models of φ_0 there is at most one element in $U_{\min}$, at most one in $U_{\max}$, and that S is a subset of the successor relation w.r.t. $<$. We call the models of φ_0 τ_0-*orderings*.

For a vocabulary τ with $< \in \tau$ and τ-structures $\mathcal{A}$ and $\mathcal{B}$ we write $\mathcal{B} \subseteq_< \mathcal{A}$ and say that $\mathcal{B}$ is a $<$-*substructure* of $\mathcal{A}$ if $\mathcal{B}$ is a substructure of $\mathcal{A}$ with $<^{\mathcal{B}} = <^{\mathcal{A}} \cap (B \times B)$.

We remark that the relation symbols $U_{\min}$, $U_{\max}$, and S are negative in φ_0. Therefore we have the following:

LEMMA 3.1. *Let $\mathcal{A}$ and $\mathcal{B}$ be τ_0-structures with $\mathcal{B} \subseteq_< \mathcal{A}$. If $\mathcal{A} \models \varphi_0$, then $\mathcal{B} \models \varphi_0$.*

Let

$$\varphi_1 := \exists x\, U_{\min} x \wedge \exists x U_{\max} x \wedge \forall x \forall y (x < y \rightarrow \exists z S x z). \tag{6}$$

We call models of $\varphi_0 \wedge \varphi_1$ *complete* τ_0-*orderings*. Clearly, for every $k \geq 1$ there is a unique, up to isomorphism, complete τ_0-ordering with exactly k elements. The next lemma shows that all its proper $<$-substructures are models of $\varphi_0 \wedge \neg\varphi_1$.

LEMMA 3.2. *Let $\mathcal{A}$ and $\mathcal{B}$ be τ_0-structures. Assume that $\mathcal{A} \models \varphi_0$ and $\mathcal{B}$ is a finite $<$-substructure of $\mathcal{A}$ that is a model of φ_1. Then $\mathcal{B} = \mathcal{A}$ (in particular, $\mathcal{A} \models \varphi_1$).*

PROOF. By the previous lemma we know that $\mathcal{B} \models \varphi_0$. Let $B := \{b_1, \ldots, b_n\}$. As $<^{\mathcal{B}}$ is an ordering, we may assume that

$$b_1 <^{\mathcal{B}} b_2 <^{\mathcal{B}} \cdots <^{\mathcal{B}} b_{n-1} <^{\mathcal{B}} b_n.$$

As $\mathcal{B} \models (\varphi_0 \wedge \varphi_1)$, we have $U^{\mathcal{B}}_{\min} b_1$, $U^{\mathcal{B}}_{\max} b_n$, and $S^{\mathcal{B}} b_i b_{i+1}$ for $i \in [n-1]$. As $\mathcal{B} \subseteq \mathcal{A}$, everywhere we can replace the upper index $^{\mathcal{B}}$ by $^{\mathcal{A}}$.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

We show $A = B$ (then $\mathcal{A} = \mathcal{B}$ follows from $\mathcal{A} \models \varphi_0$): Let $a \in A$. By $\mathcal{A} \models \varphi_0$, we have $b_1 \leq^{\mathcal{A}} a \leq^{\mathcal{A}} b_n$. Let $i \in [n]$ be maximal with $b_i \leq^{\mathcal{A}} a$. If $i = n$, then $b_n = a$. Otherwise, $b_i \leq^{\mathcal{A}} a <^{\mathcal{A}} b_{i+1}$. As $S^{\mathcal{A}} b_i b_{i+1}$, we see that $b_i = a$ (by the last conjunct of φ_0). ⊣

COROLLARY 3.3. *Every finite proper $<$-substructure of a model of $\varphi_0 \wedge \varphi_1$ is a model of $\varphi_0 \wedge \neg\varphi_1$.*

The Łoś–Tarski Theorem does not remain valid when restricted to finite structures. In fact, the class of finite τ_0-orderings that are not complete is closed under induced substructures but not axiomatizable by a universal sentence:

THEOREM 3.4 (Tait's Theorem). *The class $\mathrm{MOD}_{fin}(\varphi_0 \wedge \neg\varphi_1)$ is closed under $<$-substructures (and hence, closed under induced substructures) but $\varphi_0 \wedge \neg\varphi_1$ is not finitely equivalent to a universal sentence.*[1]

By Compton's Theorem (Theorem 2.10) the sentence $\varphi_0 \wedge \neg\varphi_1$ is not even equivalent to a Π_2-sentence. However, note that $\varphi_0 \wedge \neg\varphi_1$ is (equivalent to) a Σ_2-sentence.

PROOF OF THEOREM 3.4. $\mathrm{MOD}_{\mathrm{fin}}(\varphi_0 \wedge \neg\varphi_1)$ is closed under $<$-substructures: If $\mathcal{A} \models \varphi_0 \wedge \neg\varphi_1$ and $\mathcal{B}$ is a finite $<$-substructure of $\mathcal{A}$, then $\mathcal{B} \models \varphi_0$ (by Lemma 3.1). If $\mathcal{B} \models \neg\varphi_1$, we are done. If $\mathcal{B} \models \varphi_1$, then $\mathcal{A} \models \varphi_1$ by Lemma 3.2, which contradicts our assumption $\mathcal{A} \models \neg\varphi_1$.

Let $k \in \mathbb{N}$. It is clear that there is a finite model $\mathcal{A}$ of $\varphi_0 \wedge \varphi_1$ with at least $k + 1$ elements. By Corollary 3.3 every proper induced substructure of $\mathcal{A}$ is a model of $\varphi_0 \wedge \neg\varphi_1$. Therefore, by Corollary 2.11, the sentence $\varphi_0 \wedge \neg\varphi_1$ is not finitely equivalent to a universal sentence of the form $\forall x_1 \ldots \forall x_k\, \mu_0$ with quantifier-free μ_0. As k was arbitrary, we get our claim. ⊣

REMARK 3.5. As $\mathrm{MOD}_{\mathrm{fin}}(\varphi_0 \wedge \neg\varphi_1)$ is closed under induced substructures but $\varphi_0 \wedge \neg\varphi_1$ is not finitely equivalent and hence not equivalent to a universal sentence, the class $\mathrm{MOD}(\varphi_0 \wedge \neg\varphi_1)$ of finite and infinite models of $\varphi_0 \wedge \neg\varphi_1$ is not closed under induced substructures. Nevertheless, as $\varphi_0 \wedge \neg\varphi_1$ is equivalent to a Σ_2-sentence, a result in [24] tells us that this class has a "local Łoś–Tarski property." More precisely, for $k \in \mathbb{N}$ we say that a class $\mathcal{C}$ of τ-structures has *k-cores* if for every $\mathcal{A} \in \mathcal{C}$ there is a subset C of A of at most k elements such that every induced substructure $\mathcal{B}$ of $\mathcal{A}$ with $C \subseteq B$ is in the class $\mathcal{C}$. Note that $\mathrm{MOD}(\varphi_0 \wedge \neg\varphi_1)$ has 2-cores. In fact, let $\mathcal{A}$ be in this class. If $\mathcal{A} \models \neg\forall x \forall y(x < y \to \exists z Sxz)$, then set $C = \{a, b\}$ where $a <^{\mathcal{A}} b$ and for all $a' \in A$ not $S^{\mathcal{A}} aa'$. If $\mathcal{A} \models \forall x \forall y(x < y \to \exists z Sxz)$, then choose as C the empty set.

In [24] the authors showed: Let $k \in \mathbb{N}$ and $\mathcal{C}$ be an FO-axiomatizable class of τ-structures. Then $\mathcal{C}$ has k-cores if and only if $\mathcal{C}$ is axiomatizable by a Σ_2-sentence of the form $\exists x_1 \ldots \exists x_k\, \mu$ with universal μ. Note that for $k = 0$ we get the Łoś–Tarski Theorem.

We turn to a refinement of Theorem 3.4 that will be helpful to get Gurevich's Theorem.

[1] As already mentioned, Tait showed this result in [25] for a sentence different from $\varphi_0 \wedge \neg\varphi_1$.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

DEFINITION 3.6. (a) Let τ be obtained from the vocabulary τ_0 by adding finitely many relation symbols "in pairs," the *standard* R together with its *complement* R^{comp} (intended as the complement of R). The symbols R and R^{comp} have the same arity and for our purposes we can restrict ourselves to unary or binary relation symbols (even though all results can be generalized to arbitrary arities). We briefly say that τ *is obtained from* τ_0 *by adding pairs*.

(b) Let τ be obtained from τ_0 by adding pairs. We say that $\varphi_{0\tau} \in \mathrm{FO}[\tau]$ *is an extension of* φ_0 (where φ_0 is as above) if it is a universal sentence such that:
   (i) the sentence φ_0 is a conjunct of $\varphi_{0\tau}$,
   (ii) the sentence $\bigwedge_{R \text{ standard}} \forall \bar{x}(\neg R\bar{x} \vee \neg R^{\mathrm{comp}}\bar{x})$ is a conjunct of $\varphi_{0\tau}$,
   (iii) besides $<$ all relation symbols are negative in $\varphi_{0\tau}$ (if this is not the case for some new R or R^{comp}, the idea is to replace any positive occurrence of R or R^{comp} by $\neg R^{\mathrm{comp}}$ and $\neg R$, respectively). For instance, we replace a subformula

$$x < y \wedge Rxy \quad \text{by} \quad x < y \wedge \neg R^{\mathrm{comp}}xy.$$

(c) Let τ be obtained from τ_0 by adding pairs. Then we set

$$\varphi_{1\tau} := \varphi_1 \wedge \bigwedge_{R \text{ standard}} \forall \bar{x}(R\bar{x} \vee R^{\mathrm{comp}}\bar{x}),$$

where φ_1 is as above see (6).

For a τ-structure $\mathcal{B}$ with $\mathcal{B} \models \varphi_{0\tau} \wedge \varphi_{1\tau}$ we have

$$\mathcal{B} \models \bigwedge_{R \text{ standard}} \Big(\forall \bar{x}(\neg R\bar{x} \vee \neg R^{\mathrm{comp}}\bar{x}) \wedge \forall \bar{x}(R\bar{x} \vee R^{\mathrm{comp}}\bar{x})\Big).$$

Hence, for standard $R \in \tau$ of arity r, we have

$$\text{if } \mathcal{B} \models \varphi_{0\tau} \wedge \varphi_{1\tau}, \text{ then } (R^{\mathrm{comp}})^{\mathcal{B}} = B^r \setminus R^{\mathcal{B}}. \tag{7}$$

Now we derive the analogues of Lemma 3.1—Theorem 3.4 essentially by the same proofs. In all these results the vocabulary τ is obtained from τ_0 by adding pairs and $\varphi_{0\tau}$ is an extension of φ_0.

LEMMA 3.7. *If* $\mathcal{B} \subseteq_< \mathcal{A}$ *and* $\mathcal{A} \models \varphi_{0\tau}$, *then* $\mathcal{B} \models \varphi_{0\tau}$.

PROOF. By Definition 3.6, the sentence $\varphi_{0\tau}$ is universal and all relation symbols distinct from $<$ are negative in $\varphi_{0\tau}$. ⊣

LEMMA 3.8. *Assume that* $\mathcal{A} \models \varphi_{0\tau}$ *and that a finite* $<$*-substructure* $\mathcal{B}$ *of* $\mathcal{A}$ *is a model of* $\varphi_{1\tau}$. *Then* $\mathcal{B} = \mathcal{A}$ *in particular,* $\mathcal{A} \models \varphi_{1\tau}$.

PROOF. Let $\mathcal{A} \restriction \tau_0$ and $\mathcal{B} \restriction \tau_0$ be the τ_0-structures obtained from $\mathcal{A}$ and from $\mathcal{B}$ by removing all relations in $\tau \setminus \tau_0$. By Lemma 3.2 we know that $\mathcal{B} \restriction \tau_0 = \mathcal{A} \restriction \tau_0$. Furthermore, $\mathcal{B} \models \varphi_{0\tau}$ by the previous lemma; thus, $\mathcal{B} \models \varphi_{0\tau} \wedge \varphi_{1\tau}$. Hence, by (7), $(R^{\mathrm{comp}})^{\mathcal{B}}$ is the complement of $R^{\mathcal{B}}$ for standard R. Clearly, $R^{\mathcal{B}} \subseteq R^{\mathcal{A}}$ and $(R^{\mathrm{comp}})^{\mathcal{B}} \subseteq (R^{\mathrm{comp}})^{\mathcal{A}}$. As $A = B$ and $\mathcal{A}$ is a model of the sentence $\bigwedge_{R \text{ standard}} \forall \bar{x}(\neg R\bar{x} \vee \neg R^{\mathrm{comp}}\bar{x})$, we get $R^{\mathcal{B}} = R^{\mathcal{A}}$ and $(R^{\mathrm{comp}})^{\mathcal{B}} = (R^{\mathrm{comp}})^{\mathcal{A}}$. ⊣

COROLLARY 3.9. *Every proper* $<$*-substructure of a finite model of* $\varphi_{0\tau} \wedge \varphi_{1\tau}$ *is a model of* $\varphi_{0\tau} \wedge \neg\varphi_{1\tau}$.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

By replacing in the proof of Tait's Theorem the use of Lemma 3.1, Lemma 3.2, and Corollary 3.3 by Lemma 3.7, Lemma 3.8, and Corollary 3.9, respectively, we get the following:

LEMMA 3.10. *The class* $\mathrm{MOD}_{fin}(\varphi_{0\tau} \wedge \neg\varphi_{1\tau})$ *is closed under* $<$*-substructures (and hence, closed under induced substructures) but* $\varphi_{0\tau} \wedge \neg\varphi_{1\tau}$ *is not finitely equivalent to a universal sentence.*

Perhaps the reader will ask why we do not introduce for $<$ the "complement relation symbol" $<^{\mathrm{comp}}$ and add the corresponding conjuncts to $\varphi_{0\tau}$ and $\varphi_{1\tau}$ (or, to φ_0 and φ_1) in order to get a result of the type of Lemma 3.8 (or already of the type of Lemma 3.2) where we can replace "$<$-substructure" by "substructure." The reader will realize that corresponding proofs of $B = A$ break down.

The next proposition, the core of the proof of Gurevich's Theorem, provides a uniform way to construct FO-sentences that are only equivalent to universal sentences of large size.

PROPOSITION 3.11. *Again let* τ *be obtained from* τ_0 *by adding pairs and* $\varphi_{0\tau}$ *be an extension of* φ_0. *Let* $m \geq 1$ *and* γ *be an* $\mathrm{FO}[\tau]$*-sentence such that*

$$\varphi_{0\tau} \wedge \varphi_{1\tau} \wedge \gamma \text{ has no infinite model but a finite model with at least } m \text{ elements.} \quad (8)$$

For $\chi := \varphi_{0\tau} \wedge (\varphi_{1\tau} \to \neg\gamma)$ *the statements* (a) *and* (b) *hold.*

(a) *The class* $\mathrm{MOD}(\chi)$ *is closed under* $<$*-substructures.*
(b) *If* $\mu := \forall x_1 \ldots \forall x_k\, \mu_0$ *with quantifier-free* μ_0 *is finitely equivalent to* χ, *then* $k \geq m$.

PROOF. (a) Let $\mathcal{A} \models \chi$ and $\mathcal{B} \subseteq_< \mathcal{A}$. Thus, $\mathcal{B} \models \varphi_{0\tau}$. If $\mathcal{B} \not\models \varphi_{1\tau}$, we are done. Assume $\mathcal{B} \models \varphi_{1\tau}$. In case B is infinite, by (8) we know that $\mathcal{B}$ is a model of $\neg\gamma$ and hence of χ. Otherwise, B is finite; then $\mathcal{B} = \mathcal{A}$ (by Lemma 3.8) and thus, $\mathcal{B} \models \chi$.

(b) According to (8) there is a finite model $\mathcal{A}$ of $\varphi_{0\tau} \wedge \varphi_{1\tau} \wedge \gamma$, i.e., of $\varphi_{0\tau} \wedge \neg(\varphi_{1\tau} \to \neg\gamma)$, with at least m elements. By Corollary 3.9 every proper induced substructure of $\mathcal{A}$ is not a model of $\varphi_{1\tau}$ and therefore, it is a model of $\varphi_{0\tau} \wedge (\varphi_{1\tau} \to \neg\gamma)$. Hence by Corollary 2.11, $\varphi_{0\tau} \wedge (\varphi_{1\tau} \to \neg\gamma)$ is not finitely equivalent to a universal sentence of the form $\mu := \forall x_1 \ldots \forall x_k\, \mu_0$ with $k < m$ and quantifier-free μ_0. ⊣

REMARK 3.12. We can strengthen the statement (b) of the preceding proposition to:

> *If the* Π_2*-sentence* $\nu = \forall x_1 \ldots \forall x_k \exists y_1 \ldots \exists y_\ell\, \nu_0$ *with quantifier-free* ν_0 *is finitely equivalent to* χ, *then* $k \geq m$.

In fact, assume that $\mathrm{MOD}_{\mathrm{fin}}(\chi) = \mathrm{MOD}_{\mathrm{fin}}(\nu)$ with ν as above. We first show that k cannot be 0. In fact, if $k = 0$, then ν is an existential sentence. By assumption there is a finite model $\mathcal{A}$ of $\varphi_{0\tau} \wedge \varphi_{1\tau} \wedge \gamma$ with at least m elements. Obtain the τ-structure $\mathcal{B}$ from $\mathcal{A}$ by setting $U_{\min}^{\mathcal{B}} = U_{\max}^{\mathcal{B}} = \emptyset$. Clearly, then $\mathcal{B} \models \neg\varphi_{1\tau}$ and $\mathcal{B} \models \varphi_{0\tau}$ as $U_{\min}$ and $U_{\max}$ are negative in $\varphi_{0\tau}$. Hence, $\mathcal{B} \models \chi$ and thus, $\mathcal{B} \models \nu$. Let $\mathcal{C}$ be a finite extension of $\mathcal{B}$ (i.e., $\mathcal{B}$ is an induced substructure of $\mathcal{C}$) with an element c such that $(c, c) \in <^{\mathcal{C}}$. Then, $\mathcal{C} \models \nu$ as ν is existential. However, $\mathcal{C} \models \neg\varphi_{0\tau}$ as $<^{\mathcal{C}}$ is not an ordering. Thus, $\mathcal{C} \models \neg\chi$, a contradiction.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

So we know that $k \geq 1$ and now show that $\mathrm{MOD}_{\mathrm{fin}}(\chi) = \mathrm{MOD}_{\mathrm{fin}}(\nu)$ implies $k \geq m$. For a contradiction assume $k < m$. By (8) there is a finite model $\mathcal{A}$ of $\varphi_{0\tau} \wedge \varphi_{1\tau} \wedge \gamma$ with at least m elements. Then $\mathcal{A} \not\models \nu$. Hence there are $a_1, \dots, a_k \in A$ with $\mathcal{A} \models \neg\exists y_1 \dots \exists y_\ell\, \nu_0(a_1, \dots, a_k)$. Then $\mathcal{B} \models \neg\exists y_1 \dots \exists y_\ell\, \nu_0(a_1, \dots, a_k)$, where $\mathcal{B} := [a_1, \dots, a_k]^{\mathcal{A}}$ is the substructure of $\mathcal{A}$ induced by $a_1, \dots, a_k$. Hence, $\mathcal{B} \not\models \nu$ and therefore, $\mathcal{B} \not\models \varphi_{0\tau} \wedge \neg\varphi_{1\tau}$. As $k < m$, the structure $\mathcal{B}$ is a proper induced substructure of $\mathcal{A}$. Thus, $\mathcal{B} \models \varphi_{0\tau} \wedge \neg\varphi_{1\tau}$ by Corollary 3.9, a contradiction.

§4. The general machinery: strongly existential interpretations. We show that appropriate interpretations preserve the validity of Tait's theorem and of the statement of Proposition 3.11. Later on these interpretations will allow us to get versions of the results for graphs.

Let $\tau_E := \{E\}$ with binary E. As already remarked in the Preliminaries for *all* τ_E-structures we use the notation $G = (V(G), E(G))$ common in graph theory.

Let τ be obtained from τ_0 by adding pairs. Furthermore, let $I := (\varphi_{\mathrm{uni}}, (\varphi_T)_{T\in\tau})$ be an interpretation of *width* 2 (we only need this case) of τ-structures in τ_E-structures. This means that φ_{uni} and the φ_T's are FO[τ_E]-formulas with $\varphi_{\mathrm{uni}} = \varphi_{\mathrm{uni}}(x_1, x_2)$, $\varphi_T = \varphi_T(x_1, x_2)$ for every unary relation symbol $T \in \tau$, and $\varphi_T = \varphi_T(x_1, x_2, y_1, y_2)$ for every binary relation symbol $T \in \tau$.

Then for every τ_E-structure G we set

$$O_I(G) := \big\{\bar{a} \in V(G) \times V(G) \;\big|\; G \models \varphi_{\mathrm{uni}}(\bar{a})\big\}.$$

If $O_I(G) \neq \emptyset$, i.e., if $G \models \exists\bar{x}\varphi_{\mathrm{uni}}(\bar{x})$, then the interpretation I assigns to G a τ-structure with universe $O_I(G)$, which we denote by $\mathcal{O}_I(G)$ [2], given by:

- $T^{O_I(G)} := \big\{\bar{a} \in O_I(G) \;\big|\; G \models \varphi_T(\bar{a})\big\}$ for unary $T \in \tau$.
- $T^{O_I(G)} := \big\{(\bar{a}, \bar{b}) \in O_I(G) \times O_I(G) \;\big|\; G \models \varphi_T(\bar{a}, \bar{b})\big\}$ for binary $T \in \tau$.

As the interpretation I is of width 2, we have

$$|O_I(G)| \leq |V(G)|^2. \tag{9}$$

Recall that for every sentence $\varphi \in \mathrm{FO}[\tau]$ there is a sentence $\varphi^I \in \mathrm{FO}[\tau_E]$ such that for all τ_E-structures G with $G \models \exists\bar{x}\varphi_{\mathrm{uni}}(\bar{x})$ we have

$$\mathcal{O}_I(G) \models \varphi \iff G \models \varphi^I. \tag{10}$$

For example, for the sentence $\varphi = \forall x \forall y\, Txy$ we have

$$\varphi^I = \forall\bar{x}\Big(\varphi_{\mathrm{uni}}(\bar{x}) \to \forall\bar{y}\big(\varphi_{\mathrm{uni}}(\bar{y}) \to \varphi_T(\bar{x}, \bar{y})\big)\Big).$$

Furthermore there is a constant $c_I \in \mathbb{N}$ such that for all $\varphi \in \mathrm{FO}[\tau]$,

$$|\varphi^I| \leq c_I \cdot |\varphi|. \tag{11}$$

From time to time we will make use of the following lemma.

[2] As for the interpretations I and the graphs G we are interested in, the structure $\mathcal{O}_I(G)$ is an *o*rdered structure, we use the notation $\mathcal{O}_I(G)$

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

LEMMA 4.1. *Let* $I := \big(\varphi_{uni}, (\varphi_T)_{T\in\tau}\big)$ *be an interpretation of* τ*-structures in* τ_E*-structures. For* τ*-sentences* ψ_1 *and* ψ_2,

$$\text{MOD}(\psi_1) = \text{MOD}(\psi_2) \Rightarrow \text{GRAPH}(\forall\bar{x}\neg\varphi_{uni}(\bar{x}) \vee {\psi_1}^I) = \text{GRAPH}(\forall\bar{x}\neg\varphi_{uni}(\bar{x}) \vee {\psi_2}^I) \tag{12}$$

and the same implication holds if we restrict to finite structures, i.e.,

$$\begin{aligned}&\text{MOD}_{fin}(\psi_1) = \text{MOD}_{fin}(\psi_2)\\ &\quad\Rightarrow \text{GRAPH}_{fin}(\forall\bar{x}\neg\varphi_{uni}(\bar{x}) \vee {\psi_1}^I) = \text{GRAPH}_{fin}(\forall\bar{x}\neg\varphi_{uni}(\bar{x}) \vee {\psi_2}^I).\end{aligned} \tag{13}$$

If for every finite τ*-structure* $\mathcal{A}$ *there is a finite graph* G *with* $\mathcal{O}_I(G) \cong \mathcal{A}$*, then*

$$\begin{aligned}&\text{MOD}_{fin}(\psi_1) = \text{MOD}_{fin}(\psi_2)\\ &\quad\iff \text{GRAPH}_{fin}(\forall\bar{x}\neg\varphi_{uni}(\bar{x}) \vee {\psi_1}^I) = \text{GRAPH}_{fin}(\forall\bar{x}\neg\varphi_{uni}(\bar{x}) \vee {\psi_2}^I).\end{aligned} \tag{14}$$

PROOF. The implications in (12) and (13) follow immediately from (10). We still have to show the implication from right to left in (14). So let $\mathcal{A}$ be a finite τ-structure. By assumption there is a finite graph G with $\mathcal{O}_I(G) \cong \mathcal{A}$. As $A \neq \emptyset$, we have $G \models \neg\forall\bar{x}\neg\varphi_{\text{uni}}(\bar{x})$. By the equality on the right-hand side, thus we know that $\big(G \models {\psi_1}^I \iff G \models {\psi_2}^I\big)$. Hence, by (10), $\mathcal{A} \in \text{MOD}_{\text{fin}}(\psi_1) \iff \mathcal{A} \in \text{MOD}_{\text{fin}}(\psi_2)$. ⊣

DEFINITION 4.2. Let τ be obtained from τ_0 by adding pairs. An interpretation I of τ-structures in τ_E-structures is *strongly existential* if all formulas of I (i.e., φ_T for $T \in \tau$ and φ_{uni}) are existential and in addition $\varphi_<$ is quantifier-free.

LEMMA 4.3. *Let* τ *be obtained from* τ_0 *by adding pairs and let* $\varphi_{0\tau}$ *be an extension of* φ_0*. Then for every strongly existential interpretation* I *the sentence* ${\varphi_{0\tau}}^I$ *is (equivalent to) a universal sentence.*

PROOF. The claim holds as all relation symbols distinct from $<$ are negative in $\varphi_{0\tau}$. For example, for $\varphi := \forall x\forall y\big(U_{\min} x \to (x = y \vee x < y)\big)$, we have

$$\varphi^I = \forall\bar{x}\Big(\varphi_{\text{uni}}(\bar{x}) \to \forall\bar{y}\big(\varphi_{\text{uni}}(\bar{y}) \to (\varphi_{U_{\min}}(\bar{x}) \to ((x_1 = y_1 \wedge x_2 = y_2) \vee \varphi_<(\bar{x},\bar{y})))\big)\Big).$$

⊣

The following result shows that strongly existential interpretations transform induced subgraphs into $<$-substructures; this will be crucial to transfer the results of the preceding section to graphs.

LEMMA 4.4. *Assume that* I *is strongly existential. Then for all* τ_E*-structures* G *and* H *with* $H \subseteq_{ind} G$ *and* $O_I(H) \neq \emptyset$*, we have* $\mathcal{O}_I(H) \subseteq_< \mathcal{O}_I(G)$.

PROOF. As φ_{uni} is existential, we have $O_I(H) \subseteq O_I(G)$. Let $T \in \tau$ be distinct from $<$ and $\bar{b} \in T^{\mathcal{O}_I(H)}$. Then $H \models \varphi_T(\bar{b})$. As φ_T is existential, $G \models \varphi_T(\bar{b})$ and thus, $\bar{b} \in T^{\mathcal{O}_I(G)}$. Moreover, for $\bar{b}, \bar{b}' \in O_I(H)$ we have

$$\begin{aligned}\bar{b} <^{\mathcal{O}_I(H)} \bar{b}' &\iff H \models \varphi_<(\bar{b},\bar{b}')\\ &\iff G \models \varphi_<(\bar{b},\bar{b}') \qquad (\text{as } H \subseteq_{\text{ind}} G \text{ and } \varphi_< \text{ is quantifier-free})\\ &\iff \bar{b} <^{\mathcal{O}_I(G)} \bar{b}'.\end{aligned}$$

Putting all together we see that $\mathcal{O}_I(H) \subseteq_< \mathcal{O}_I(G)$. ⊣

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

COROLLARY 4.5. *Assume I is strongly existential and let ψ be a τ-sentence. If* $\mathrm{MOD}(\psi)$ *(resp.* $\mathrm{MOD}_{fin}(\psi)$*) is closed under $<$-substructures, then* $\mathrm{MOD}(\forall\bar{x}\neg\varphi_{uni}(\bar{x}) \vee \psi^I)$ *(resp.* $\mathrm{MOD}_{fin}(\forall x\neg\varphi_{uni}(\bar{x}) \vee \psi^I)$*) is closed under induced substructures.*

PROOF. Let G and H be τ_E-structures, $H \subseteq_{\mathrm{ind}} G$, and $G \models \forall\bar{x}\neg\varphi_{\mathrm{uni}}(\bar{x}) \vee \psi^I$. If $H \models \forall\bar{x}\neg\varphi_{\mathrm{uni}}(\bar{x})$, we are done. Otherwise, also $G \not\models \forall\bar{x}\neg\varphi_{\mathrm{uni}}(\bar{x})$ and thus, $G \models \psi^I$. Hence, $\mathcal{O}_I(G) \models \psi$ and $\mathcal{O}_I(H) \subseteq_< \mathcal{O}_I(G)$ by the previous lemma. Therefore, by assumption, $\mathcal{O}_I(H) \models \psi$ and thus, $H \models \psi^I$. ⊣

We obtain from Lemma 3.8 the corresponding result in our framework.

LEMMA 4.6. *Let I be strongly existential and let $\varphi_{0\tau}$ be an extension of φ_0. Assume that the τ_E-structure G is a model of ${\varphi_{0\tau}}^I$ and that $H \subseteq_{ind} G$ with finite $O_I(H)$, is a model of ${\varphi_{1\tau}}^I$. Then $\mathcal{O}_I(H) = \mathcal{O}_I(G)$ and $G \models {\varphi_{1\tau}}^I$.*

PROOF. As $H \models {\varphi_{1\tau}}^I$, we have $H \models (\exists x\, U_{\min}\, x)^I$ holds and thus, $O_I(H) \neq \emptyset$. Therefore, $\mathcal{O}_I(H) \subseteq_< \mathcal{O}_I(G)$ by Lemma 4.4. By assumption and (10), $\mathcal{O}_I(G) \models \varphi_{0\tau}$ and $\mathcal{O}_I(H) \models \varphi_{1\tau}$. As $O_I(H)$ is finite, Lemma 3.8 implies $\mathcal{O}_I(H) = \mathcal{O}_I(G)$, and in particular $\mathcal{O}_I(G) \models \varphi_{1\tau}$. Hence, $G \models {\varphi_{1\tau}}^I$ by (10). ⊣

We now prove for strongly existential interpretations two results, Proposition 4.7 corresponds to Tait's Theorem (Theorem 3.4) and Proposition 4.8 corresponds to Proposition 3.11 (relevant to Gurevich's Theorem).

PROPOSITION 4.7. *Assume that the interpretation I of τ_0-structures in τ_E-structures is strongly existential. Furthermore, assume that for every finite complete τ_0-ordering $\mathcal{A}$, i.e., $\mathcal{A} \models \varphi_0 \wedge \varphi_1$, there is a finite graph G with $\mathcal{O}_I(G) \cong \mathcal{A}$. Then for*

$$\varphi := \forall\bar{x}\neg\varphi_{uni}(\bar{x}) \vee \big(\varphi_0 \wedge \neg\varphi_1\big)^I$$

the class $\mathrm{GRAPH}_{fin}(\varphi)$ *is closed under induced subgraphs, but φ is not equivalent to a universal sentence in finite graphs.*

PROOF. By Theorem 3.4, we know that $\mathrm{MOD}_{\mathrm{fin}}(\varphi_0 \wedge \neg\varphi_1)$ is closed under $<$-substructures. Hence, $\mathrm{GRAPH}_{\mathrm{fin}}\big(\forall\bar{x}\neg\varphi_{\mathrm{uni}}(\bar{x}) \vee \big(\varphi_0 \wedge \neg\varphi_1\big)^I\big)$ is closed under induced subgraphs by Corollary 4.5.

Now we show that for every $k \geq 1$ the sentence φ is not equivalent in finite graphs to a sentence of the form $\mu = \forall z_1 \ldots \forall z_k\, \mu_0$ with quantifier-free μ_0. Let $\mathcal{A} := \big(A, <^{\mathcal{A}}, U^{\mathcal{A}}_{\min}, U^{\mathcal{A}}_{\max}, S^{\mathcal{A}}\big)$ be a complete τ_0-ordering with at least $k^2 + 1$ elements. In particular, $\mathcal{A} \models \varphi_0 \wedge \varphi_1$. By assumption there is a finite graph G such that $\mathcal{O}_I(G) \cong \mathcal{A}$. Then $\mathcal{O}_I(G) \models \varphi_0 \wedge \varphi_1$, hence, $G \models {\varphi_0}^I \wedge {\varphi_1}^I$. Thus $G \models \neg\varphi$. As $|O_I(G)| = |A| \geq k^2 + 1$, the graph G must contain more than k vertices by (9).

We want to show that every induced subgraph of G with at most k vertices is a model of φ. Then the result follows from Corollary 2.11 for $\psi_0 := \varphi_{\mathrm{GRAPH}} \wedge \big(\forall\bar{x}\neg\varphi_{\mathrm{uni}}(\bar{x}) \vee {\varphi_0}^I\big)$ and $\psi_1 := \exists\bar{x}\varphi_{\mathrm{uni}}(\bar{x}) \wedge {\varphi_1}^I$.

So let H be an induced subgraph of G with at most k vertices. Clearly, $H \models {\varphi_0}^I$. If $H \models \forall\bar{x}\neg\varphi_{\mathrm{uni}}(\bar{x})$ or $H \models \neg{\varphi_1}^I$, we are done. Otherwise $O_I(H) \neq \emptyset$ and $H \models {\varphi_1}^I$. Then, Lemma 4.6 implies $O_I(H) = O_I(G)$. Recall $|V(H)| \leq k$, so $O_I(H)$ has at most k^2 elements by (9), a contradiction as $|O_I(G)| \geq k^2 + 1$. ⊣

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

PROPOSITION 4.8. *Let τ be obtained from τ_0 by adding pairs and let $\varphi_{0\tau}$ be an extension of φ_0. Assume that I is a strongly existential interpretation of τ-structures in τ_E-structures with the property that for every finite τ-structure $\mathcal{A}$ that is a model of $\varphi_{0\tau} \wedge \varphi_{1\tau}$ there is a finite graph G with $\mathcal{O}_I(G) \cong \mathcal{A}$ and $G \models \psi$.*

Let $m \geq 1$ and γ be an FO[τ]*-sentence such that*

$$\varphi_{0\tau} \wedge \varphi_{1\tau} \wedge \gamma \text{ has no infinite model but a finite model with at least } m \text{ elements.} \quad (15)$$

For

$$\rho := \forall \bar{x} \neg \varphi_{uni}(\bar{x}) \vee \big(\varphi_{0\tau} \wedge (\varphi_{1\tau} \to \neg\gamma)\big)^I,$$

the statements (a) *and* (b) *hold.*

(a) *The class* GRAPH(ρ) *is closed under induced subgraphs.*
(b) *If $\mu := \forall x_1 \ldots \forall x_k \, \mu_0$ with quantifier-free μ_0 is equivalent in finite graphs to ρ, then $k^2 \geq m$.*

PROOF. Again (a) follows from Proposition 3.11(a) by Corollary 4.5.

(b) By (15) there is a finite model $\mathcal{A}$ of $\varphi_{0\tau} \wedge \varphi_{1\tau} \wedge \gamma$ with at least m elements. By assumption there is a finite graph G with $\mathcal{O}_I(G) \cong \mathcal{A}$ and $G \models \psi$. Clearly, $G \models \neg\forall\bar{x}\neg\varphi_{\text{uni}}(\bar{x})$ and $G \models (\varphi_{0\tau} \wedge \varphi_{1\tau} \wedge \gamma)^I$. Hence, $G \models \neg\rho$. Assume that $k^2 < m$. We want to show that every induced subgraph of G with at most k elements is a model of ρ. Then the claim (b) follows from Corollary 2.11 (with $\psi_0 := \forall x\, x = x$ and $\psi_1 := \neg\rho$).

So let H be an induced subgraph of G with at most k elements. Clearly, $H \models {\varphi_{0\tau}}^I$. If $H \models \forall\bar{x}\neg\varphi_{\text{uni}}(\bar{x})$ or $H \models \neg{\varphi_{1\tau}}^I$, we are done. Otherwise $O_I(H) \neq \emptyset$ and $H \models {\varphi_{1\tau}}^I$. Then, $O_I(H) = O_I(G)$ by Lemma 4.6. This leads to a contradiction, as $O_I(H)$ has at most k^2 elements by (10), while $O_I(G)$ has m elements and we assumed $k^2 < m$. ⊣

REMARK 4.9. (a) The result corresponding to Remark 3.12 is valid for Proposition 4.7 too.

(b) By Compton's Theorem (Theorem 2.10) the sentence $\forall\bar{x}\neg\varphi_{\text{uni}}(\bar{x}) \vee \big(\varphi_0 \wedge \neg\varphi_1\big)^I$ is not equivalent to a Π_2-sentence. However, $\forall\bar{x}\neg\varphi_{\text{uni}}(\bar{x}) \vee \big(\varphi_0 \wedge \neg\varphi_1\big)^I$ itself is equivalent to a Σ_2-sentence. In fact, as all relation symbols besides $<$ are negative in φ_0, the sentence ${\varphi_0}^I$ is universal. Moreover, as $U_{\min}$, $U_{\max}$, and S are positive in φ_1, the sentence ${\varphi_1}^I$ (as φ_1) is equivalent to a Π_2-sentence. Hence $\forall\bar{x}\neg\varphi_{\text{uni}}(\bar{x}) \vee \big(\varphi_0 \wedge \neg\varphi_1\big)^I$ is equivalent to a Σ_2-sentence.

§5. Tait's Theorem for finite graphs. We present strongly existential interpretations that allow us to get Tait's Theorem for graphs in this section and Gurevich's Theorem for graphs in Section 6.

We first introduce a further concept. Let G be a graph and $a, b \in V(G)$. For $r, s \geq 3$ *a path from vertex a to vertex b of length r with an s-ear* is a path between a and b with a cycle of length s; one vertex of this cycle is adjacent to the vertex adjacent to b on the path; path and cycle have no vertex in common. Figure 1 is a path from a to b of length 6 with a 4-ear.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

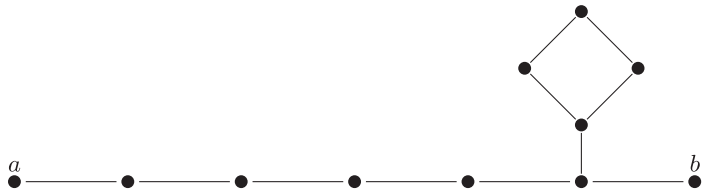

FIGURE 1. A path of length 6 with a 4-ear.

LEMMA 5.1. *For $r, s \geq 3$ there are quantifier-free formulas $\varphi_{c,r}(x, \bar{z})$ and $\varphi_{pe,r,s}(x, y, \bar{z}, \bar{w})$ such that for all graphs G we have*:

(a) $G \models \varphi_{c,r}(a, \bar{u}) \iff \bar{u}$ *is a cycle of length r containing a.*

(b) $G \models \varphi_{pe,r,s}(a, b, \bar{u}, \bar{v}) \iff \bar{u}$ *is path from a to b of length r with the s-ear $\bar{v}$.*

PROOF. (a) We can take as $\varphi_{c,r}(x, z_1, \dots, z_r)$ the formula

$$\bigwedge_{1 \leq i < r} Ez_i z_{i+1} \wedge Ez_r z_1 \wedge \bigwedge_{1 \leq i < j \leq r} \neg z_i = z_j \wedge \bigvee_{i \in [r]} x = z_i.$$

(b) We can take as $\varphi_{pe,r,s}(x, y, z_0, \dots, z_r, w_1, \dots, w_s)$ the formula

$$x = z_0 \wedge y = z_r \wedge \bigwedge_{0 \leq i < r-1} Ez_i z_{i+1} \wedge \bigwedge_{0 \leq i < j \leq r} \neg z_i = z_j \wedge \bigwedge_{0 \leq i \leq r,\ j \in [s]} \neg z_i = w_j \wedge \bigvee_{i \in [s]} (\varphi_{c,s}(w_i, w_1, \dots, w_s) \wedge Ez_{r-1} w_i).$$

⊣

To understand better how we obtain the desired interpretation we first assign to every complete τ_0-ordering $\mathcal{A}$, i.e., to every model of $\varphi_0 \wedge \varphi_1$, a τ_E-structure $G := G(\mathcal{A})$ that is a graph.

In a first step we extend $\mathcal{A}$ to a τ_0^*-structure $\mathcal{A}^*$, where $\tau_0^* := \tau_0 \cup \{B, C, L, F\}$ in the following way. Here B, C are unary and L, F are binary relation symbols.

For every *original* (or, *basic*) element a, i.e., for every $a \in A$, we introduce a new element a', the *companion of a*. We set

– $A^* := A \cup \{a' \mid a \in A\}$,
– $B^{\mathcal{A}^*} := A, \qquad C^{\mathcal{A}^*} := \{a' \mid a \in A\}$,
– $L^{\mathcal{A}^*} := \big\{(a, a') \bigm| a \in A\big\}, \qquad F^{\mathcal{A}^*} := \big\{(a', b), (b, a') \bigm| a, b \in A,\ a <^{\mathcal{A}} b\big\}$.

Note that the relation F is irreflexive and symmetric, i.e., $\big(A^*, F^{\mathcal{A}^*}\big)$ is already a graph, which is illustrated by Figure 2. Observe that F contains the whole information of the ordering $<^{\mathcal{A}}$ up to isomorphism.

We use $\mathcal{A}^*$ to define the desired graph $G = G(\mathcal{A})$. The vertex set $V(G)$ contains the elements of A^* and the edge relation $E(G)$ contains $F^{\mathcal{A}^*}$. Furthermore G contains just all the vertices and edges required by the "gadgets" introduced by the following clauses:

– To $a \in U^{\mathcal{A}}_{\min}$ we add a cycle of length 5 through a, all the other vertices of the cycle are new, i.e., not in A^*.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

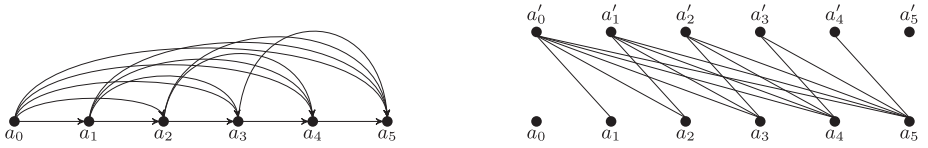

FIGURE 2. Turning an ordering to the relation F.

- To $a \in U^{\mathcal{A}}_{\max}$ we add a cycle of length 7 through a, all the other vertices of the cycle are new.
- To $a \in B^{\mathcal{A}^*}$ we add a cycle of length 9 through a, all the other vertices of the cycle are new.
- To $a \in C^{\mathcal{A}^*}$ we add a cycle of length 11 through a, all the other vertices of the cycle are new.
- To $(a, b) \in S^{\mathcal{A}}$ we add a path from a to b of length 17 with a 13-ear consisting of new vertices (besides a and b).
- To $(a, a') \in L^{\mathcal{A}^*}$ we add a path from a to a' of length 17 with a 15-ear consisting of new vertices (besides a and a').

Hereby we meant by "add a cycle" or "add a path with an ear" that we only add the edges required by the corresponding formulas in Lemma 5.1.

To ease the discussion, we divide cycles in G $(= G(\mathcal{A}))$ into four categories.

[*F-cycle*] These are the cycles in $\left(A^*, F^{\mathcal{A}^*}\right)$, i.e., the cycles using only edges of $F^{\mathcal{A}^*}$.

[*T-cycle*] For every $T \in \left\{U_{\min}, U_{\max}, B, C\right\}$ and $a \in T^{\mathcal{A}}$ the cycle introduced for a is a T-cycle.

[*ear-cycle*] These are the cycles that are the ears on the gadgets introduced for the pairs of the relations $S^{\mathcal{A}^*}$ and $L^{\mathcal{A}^*}$.

[*mixed-cycle*] All the other cycles are *mixed*.

For example, we get a mixed cycle if we start with a_2, a'_0, a_1 in Figure 2 and then add the path introduced for $(a_1, a_2) \in S^{\mathcal{A}}$ (ignoring the ear).

A number of observations for these types of cycles are in order.

LEMMA 5.2. (i) *All the F-cycles are of even length.*
(ii) *Every* U_{min}*-,* U_{max}*-, B-, and C-cycle is of length* 5, 7, 9, *and* 11, *respectively.*
(iii) *Every ear-cycle is of length* 13 *or* 15.
(iv) *Every mixed-cycle neither uses new vertices of any T-cycle for* $T \in \left\{U_{min}, U_{max}, B, C\right\}$ *nor any vertex of any ear-cycle.*
(v) *Every mixed-cycle has length at least* 17.

PROOF. (i) follows easily from the fact that $\left(A^*, F^{\mathcal{A}^*}\right)$ is a bipartite graph; (ii) and (iii) are trivial. For (iv) assume that a mixed-cycle uses a *new* vertex b of a T-cycle $\mathcal{C}$ introduced for some $a \in T^{\mathcal{A}^*}$, where $T \in \left\{U_{\min}, U_{\max}, B, C\right\}$. As $\mathcal{C}$ is mixed, it must contain a vertex $c \notin T^{\mathcal{A}^*}$. To reach b from c the mixed cycle must pass through a and hence must contain one of the two segments of $\mathcal{C}$ between b and a. Therefore, in order for the mixed-cycle to go back from b to c, it must also use the other segment of $\mathcal{C}$ between a and b. This means that it must be the T-cycle $\mathcal{C}$ itself, instead of a

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

mixed one. A similar argument shows that mixed cycles do not contain vertices of any ear-cycle.

To prove (v), let $\mathcal{C}$ be a mixed-cycle. By (iv), $\mathcal{C}$ must contain all vertices of a (at least one) path introduced for a pair $(a, a') \in L^{\mathcal{A}*}$ or $(a, b) \in S^{\mathcal{A}^*}$ (ignoring the ear). As these paths have length 17, we get our claim. ⊣

We want to recover $\mathcal{A}$ (up to isomorphism) from $G(\mathcal{A})$ by means of a strongly existential interpretation. Let G be any graph. First we define a τ_0-structure $\mathcal{O}(G)$, possibly the "empty structure" (and then we show that $\mathcal{O}(G) = \mathcal{O}_I(G)$ for some strongly existential interpretation I). For the definitions of "cycle" and of "path with ear" see Lemma 5.1.

- $O(G) := \big\{(a_1, a_2) \in V(G) \times V(G) \bigm| a_1$ is a member of a cycle of length 9, a_2 is a memberof a cycle of length 11, and there is a path from a_1 to a_2 of length 17 with a 15-ear$\big\}$
- $<^{\mathcal{O}(G)} := \big\{((a_1, a_2), (b_1, b_2)) \in O(G) \times O(G) \bigm| \{a_2, b_1\} \in E(G)\big\}$
- $U_{\min}^{\mathcal{O}(G)} := \big\{(a_1, a_2) \in O(G) \bigm| a_1$ is a member of a cycle of length 5$\big\}$
- $U_{\max}^{\mathcal{O}(G)} := \big\{(a_1, a_2) \in O(G) \bigm| a_1$ is a member of a cycle of length 7$\big\}$
- $S^{\mathcal{O}(G)} := \big\{((a_1, a_2), (b_1, b_2)) \in O(G) \times O(G) \mid$ there is a path from a_1 to b_1 of length 17 with a 13-ear$\big\}$.

LEMMA 5.3. *For every complete τ_0-ordering $\mathcal{A}$ we have $\mathcal{O}(G(\mathcal{A})) \cong \mathcal{A}$.*

PROOF. Let $G := G(\mathcal{A})$ and $\mathcal{A}^+ := \mathcal{O}(G)$. We claim that the mapping $h : A \to A^+$ defined by

$$h(a) := (a, a') \quad \text{for } a \in A$$

is an isomorphism from $\mathcal{A}$ to $\mathcal{A}^+$. To that end, we first prove that

$$A^+ = \big\{(a, a') \bigm| a \in A\big\},$$

which implies that h is well defined and a bijection. For every $a \in A$ it is easy to see that $(a, a') \in O(G)$ $(= A^+)$. For the converse, let $(a_1, a_2) \in O(G)$. In particular, a_1 is a member of a cycle of length 9. By Lemma 5.2, this must be a B-cycle that contains some $a \in A$. Using the same argument, a_2 is a member of a C-cycle that contains a vertex b' being the companion of some $b \in A$. Furthermore, there is a path from a_1 to a_2 of length 17 with a 15-ear. The 15-ear is a cycle of length 15. Again by Lemma 5.2 this cycle is an ear-cycle that belongs to the gadget we introduced for some $(c, c') \in L^{\mathcal{A}^*}$ with $c \in A$. Then it is easy to see that $a = c = b$. This finishes the proof that h is a bijection from A to A^+.

Similarly, we can prove that h preserves all the relations. ⊣

We show that we can obtain $\mathcal{O}(G)$ from G by a strongly existential FO-interpretation I of width 2. We set

$$\varphi_{\text{uni}}(x, x') := \exists \bar{x} \exists \bar{x}' \exists \bar{z} \exists \bar{w}\, \eta(x, x', \bar{x}, \bar{x}', \bar{z}, \bar{w}).$$

Here $\eta(x, x', \bar{x}, \bar{x}', \bar{z}, \bar{w})$ is the formula

$$\varphi_{c,9}(x, \bar{x}) \wedge \varphi_{c,11}(x', \bar{x}') \wedge \varphi_{pe,17,15}(x, x' \bar{z}, \bar{w})$$

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

that expresses "$\bar{x}$ is a cycle of length 9 containing x, $\bar{x}'$ is a cycle of length 11 containing x', and $\bar{z}$ is a path from x to x' of length 17 with the 15-ear $\bar{w}$." Furthermore we define:

- $\varphi_<(x, x', y, y') := Ex'y$,
- $\varphi_{U_{\min}}(x, x') := \exists \bar{z}\, \varphi_{c,5}(x, \bar{z})$,
- $\varphi_{U_{\max}}(x, x') := \exists \bar{z}\, \varphi_{c,7}(x, \bar{z})$,
- $\varphi_S(x, x', y, y') := \exists \bar{z} \exists \bar{w} \varphi_{pe,17,13}(x, y, \bar{z}, \bar{w})$.

Then we have the following:

LEMMA 5.4. *$I := (\varphi_{uni}, \varphi_<, \varphi_{U_{min}}, \varphi_{U_{max}}, \varphi_S)$ is a strongly existential interpretation of τ_0-structures in τ_E-structures. For every complete τ_0-ordering $\mathcal{A}$ we have $\mathcal{O}_I(G(\mathcal{A})) = \mathcal{O}(G(\mathcal{A}))$ and hence, by Lemma 5.3,*

$$\mathcal{O}_I(G(\mathcal{A})) \cong \mathcal{A}.$$

We get from Proposition 4.7:

THEOREM 5.5 (Tait's Theorem for graphs). *There is a τ_E-sentence φ such that $\textsc{Graph}_{fin}(\varphi)$, the class of finite graphs that are models of φ, is closed under induced subgraphs but φ is not equivalent to a universal sentence in finite graphs.*

In this section we presented a strongly existential interpretation of τ_0-structures in τ_E-structures (more precisely, in graphs) and applied it to finite complete τ_0-orderings, i.e., to models of $\varphi_0 \wedge \varphi_1$. A straightforward generalization of the preceding proofs allows us to show the following result for vocabularies obtained from τ_0 by adding pairs. We shall use it in Section 6.

LEMMA 5.6. *Let τ be obtained from τ_0 by adding pairs. There is a strongly existential interpretation I $(= I_\tau)$ that for every extension $\varphi_{0\tau}$ of φ_0 assigns to every τ-structure $\mathcal{A}$ that is a model of $\varphi_{0\tau} \wedge \varphi_{1\tau}$ a graph $G(\mathcal{A})$ with $\mathcal{O}_I(G(\mathcal{A})) \cong \mathcal{A}$. For finite $\mathcal{A}$ the graph $G(\mathcal{A})$ is finite.*

PROOF. We get the graph $G(\mathcal{A})$ as in the case $\tau := \tau_0$: For the elements of new unary relations we add cycles such that the lengths of the cycles are odd and distinct for distinct unary relations in τ. Let c be the maximal length of these cycles. Then we add paths with ears to the tuples of binary relations as above. For distinct binary relations the ears should have distinct length and again this length should be odd and greater than c. On the other hand, the length of added new paths can be the same for all binary relations but should be greater than the length of all the cycles. ⊣

REMARK 5.7. (a) Let $\mathcal{C} := \textsc{Mod}_{\text{fin}}(\forall x \neg Exx)$ be the class of finite directed graphs. Then $\mathcal{C}' := \textsc{Graph}_{\text{fin}}$, the class of finite graphs, is a subclass of $\mathcal{C}$ closed under induced substructures and definable in $\mathcal{C}$ by the universal sentence $\forall x \forall y(Exy \to Eyx)$. As the Łoś–Tarski Theorem fails for the class of finite graphs, it fails for the class of directed graphs by Remark 2.12.

(b) Let $\mathcal{C}' := \textsc{Planar}_{\text{fin}}$ be the class of finite planar graphs, a subclass of $\mathcal{C} := \textsc{Graph}_{\text{fin}}$ closed under induced subgraphs. As mentioned in the Introduction, in [2] it is shown that the Łoś–Tarski Theorem fails for $\textsc{Planar}_{\text{fin}}$. As $\textsc{Planar}_{\text{fin}}$ is not axiomatizable in $\textsc{Graph}_{\text{fin}}$ by a universal sentence, not even by a first-order sentence, we do not get the failure of the Łoś–Tarski Theorem for the class of

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

finite graphs (i.e., Theorem 5.5) by applying the result of Remark 2.12. We show that $\textsc{Planar}_{\text{fin}} = \textsc{Forb}_{\text{fin}}(\mathcal{F})$ for a finite set $\mathcal{F}$ of finite graphs (or, equivalently, $\textsc{Planar}_{\text{fin}} = \textsc{Mod}_{\text{fin}}(\mu)$ for a universal μ) leads to a contradiction. Let k be the maximum size of the set of vertices of graphs in $\mathcal{F}$. Let G be the graph obtained from the clique K_5 of five vertices by subdividing each edge $k+1$ times. Clearly, $G \notin \textsc{Planar}_{\text{fin}}$. However, every subgraph of G induced on at most k vertices is planar. Hence, $G \in \textsc{Forb}_{\text{fin}}(\mathcal{F})$.

(c) Let τ be any vocabulary with at least one at least binary relation T. Then the Łoś–Tarski Theorem fails for the class $\mathcal{C} := \textsc{Str}_{\text{fin}}[\tau]$, the class of all finite τ-structures. By Remark 2.12 it suffices to show the existence of a universally definable subclass $\mathcal{C}'$ of $\mathcal{C}$ which "essentially is the class of graphs." We set

$$\mu := \forall x \forall \bar{u} \neg T x x \bar{u} \wedge \forall x \forall y \forall \bar{u} \forall \bar{v} (T x y \bar{u} \to T y x \bar{v}) \wedge \bigwedge_{R \in \tau,\ R \neq T} \forall \bar{u} \neg R \bar{u}$$

and let $\mathcal{C}'$ be $\textsc{Mod}_{\text{fin}}(\mu)$.

If τ only contains unary relation symbols, the Łoś–Tarski Theorem holds for $\textsc{Str}_{\text{fin}}[\tau]$. It is easy to see for an $\text{FO}(\tau)$-sentence φ that the closure under induced substructures of $\textsc{Mod}_{\text{fin}}(\varphi)$ implies that of $\textsc{Mod}(\varphi)$.

§6. Gurevich's Theorem. The following discussion will eventually lead to a proof of Gurevich's Theorem, i.e., Theorem 1.5. Our proof essentially follows Gurevich's proof in [17], but it contains some elements of Rossman's proof of the same result in [22].[3] Afterwards we show that it remains true if we restrict ourselves to graphs.

Our main tool is Proposition 3.11: the goal is to construct a formula γ satisfying (8) and whose size is much smaller than the number m. Basically γ will describe a very long computation of a Turing machine on a short input. We fix a universal Turing machine M operating on a one-way infinite tape, the tape alphabet is $\{0,1\}$, where 0 is also considered as blank and Q is the set of states of M. The initial state is q_0 and q_h is the halting state; thus $q_0, q_h \in Q$ and we assume that $q_0 \neq q_h$. An instruction of M has the form

$$qapbd,$$

where $q, p \in Q$, $a, b \in \{0,1\}$ and $d \in \{-1, 0, 1\}$. It indicates that if M is in state q and the head of M reads an a, then M changes to state p, the head replaces a by b and moves to the left (if $d = -1$), stays still (if $d = 0$), or moves to the right (if $d = 1$). In order to describe computations of M by FO-formulas we introduce binary predicates $H_q(x,t)$ for $q \in Q$ to indicate that at time t the machine M is in state q and the head scans cell x, and a binary predicate $C_0(x,t)$ to indicate that the content of cell x at time t is 0.

The vocabulary τ_M is obtained from τ_0 by adding pairs (see Definition 3.6(a)),

$$\tau_M := \tau_0 \cup \{H_q, H_q^{\text{comp}} \mid q \in Q\} \cup \{C_0, C_0^{\text{comp}}\}.$$

[3] The reader of [17] will realize that the definition of φ^n on page 190 of [17] must be modified in order to ensure that the class of models of φ^n is closed under induced substructures.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

Intuitively, $H_q^{\text{comp}}(x,t)$ says that "at time t the machine is not in state q or the head does not scan cell x;" and $C_0^{\text{comp}}(x,t)$ says that "at time t the content of cell x is (not 0 and thus is) 1." Sometimes we write C_1 instead of C_0^{comp} (e.g., below in φ_2 if $a=1$ or $b=0$).

Let φ_0 and φ_1 be the sentences already introduced in Section 3. For $w \in \{0,1\}^*$ the sentence φ_{0w} will be an extension of φ_0 (compare Definition 3.6(b)). Hence, φ_{0w} will be a universal sentence and all relations symbols besides $<$ are negative in φ_{0w}; in particular, it contains as conjuncts φ_0 and

$$\forall x \forall t \big(\neg C_0(x,t) \vee \neg C_0^{\text{comp}}(x,t)\big) \wedge \bigwedge_{q \in Q} \forall x \forall t \big(\neg H_q(x,t) \vee \neg H_q^{\text{comp}}(x,t)\big).$$

Finally, φ_{0w} will contain the following sentences φ_2 and φ_w as conjuncts. The sentence φ_2 describes one computation step. It contains for each instruction of M one conjunct. For example, the instruction $qapb1$ contributes the conjunct

$$\begin{aligned}\forall x \forall x' \forall t \forall t' \forall y \Big(& \big(H_q(x,t) \wedge C_a(x,t) \wedge S(x,x') \wedge S(t,t')\big) \\ & \to \Big((\neg C_{1-b}(x,t') \wedge \neg H_p^{\text{comp}}(x',t')) \\ & \quad \wedge (y \neq x' \to \bigwedge_{r \in Q} \neg H_r(y,t')) \\ & \quad \wedge (y \neq x \to ((C_0(y,t) \to \neg C_0^{\text{comp}}(y,t')) \wedge (C_0^{\text{comp}}(y,t) \to \neg C_0(y,t'))))\Big)\Big).\end{aligned}$$

For $w \in \{0,1\}^*$ the sentence φ_w describes the initial configuration of M with input w (if $w = w_1 \dots w_{|w|}$, the first $|w|$ cells (if present) contain $w_1, \dots, w_{|w|}$, the remaining cells contain 0, and the head scans the first cell in the starting state q_0). Taking into account that models of $\varphi_0 \wedge \varphi_1$ might contain less than $|w|$ elements, as φ_w we can take the conjunction of

- $\forall x_1 \dots \forall x_{|w|} \big((U_{\min} x_1 \to \neg C_{1-w_1}(x_1,x_1)) \wedge \bigwedge_{i \in [|w|-1]} (S x_i x_{i+1} \to \neg C_{1-w_{i+1}}(x_{i+1},x_1))\big)$
- $\forall x_1 \dots \forall x_{|w|} \forall x \big((U_{\min} x_1 \wedge \bigwedge_{i \in [|w|-1]} S x_i x_{i+1} \wedge x_{|w|} < x) \to \neg C_0^{\text{comp}}(x,x_1)\big)$
- $\forall x \forall y \big(U_{\min} x \to (\neg H_{q_0}^{\text{comp}}(x,x) \wedge (y \neq x \to \bigwedge_{q \in Q} \neg H_q(y,x)))\big)$.

Note that besides $<$ all relation symbols of τ_M are negative in φ_{0w}. We set $\varphi_{1M} := \varphi_{1\tau_M}$; recall that by Definition 3.6(c),

$$\varphi_{1M} = \varphi_1 \wedge \forall x \forall t \big(C_0(x,t) \vee C_0^{\text{comp}}(x,t)\big) \wedge \bigwedge_{q \in Q} \forall x \forall t \big(H_q(x,t) \vee H_q^{\text{comp}}(x,t)\big). \tag{16}$$

Let $w \in \{0,1\}^*$ and $r \in \mathbb{N}$. Furthermore, let $\mathcal{A}$ be a τ_M-structure where $<^{\mathcal{A}}$ is an ordering and $|A| \geq r+1$. Let $a_0, \dots, a_r$ be the first $r+1$ elements of $<^{\mathcal{A}}$. Assume that M on the input $w \in \{0,1\}^*$ runs at least r steps. We say that $\mathcal{A}$ *correctly encodes r steps of the computation of M on w* if for i, j with $0 \leq i, j \leq r$,

$$(a_i, a_j) \in C_0^{\mathcal{A}} \iff \text{the content of cell } i \text{ after } j \text{ steps is } 0 \tag{17}$$

and for $q \in Q$,

$$(a_i, a_j) \in H_q^{\mathcal{A}} \iff \text{after } j \text{ steps } M \text{ is in state } q \text{ and the head scans cell } i. \tag{18}$$

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

LEMMA 6.1. *Let $w \in \{0,1\}^*$ and $r \in \mathbb{N}$.*

(a) *Let $\mathcal{A} \models \varphi_{0w} \wedge \varphi_{1M}$ and $r+1 \leq |A|$ (this holds if A is infinite). If M on w runs at least r steps, then $\mathcal{A}$ correctly encodes r steps of the computation of M on w.*

(b) *There is a finite model of $\varphi_{0w} \wedge \varphi_{1M}$ with $r+1$ elements. If M runs at least r steps, then this model is unique up to isomorphism.*

PROOF. (a) holds by the definitions of φ_{0w} and φ_{1M}. For (b) let $A = \{a_0, \dots, a_r\}$ with pairwise distinct a_i's. Assume first that M on w runs at least r steps. We can interpret (17) and (18) as defining relations $C_0^{\mathcal{A}}$ and $H_q^{\mathcal{A}}$ on A equipped with the "natural" ordering and its corresponding relations $U_{\min}$, $U_{\max}$, and S. If furthermore we let $(C_0^{\mathrm{comp}})^{\mathcal{A}}$ and $(H_q^{\mathrm{comp}})^{\mathcal{A}}$ be the complements in $A \times A$ of $C_0^{\mathcal{A}}$ and $H_q^{\mathcal{A}}$, respectively, we get a model of $\varphi_{0w} \wedge \varphi_{1M}$ with $r+1$ elements. By (a), this model is unique up to isomorphism.

If M on input w halts, say in $h(w)$ steps, with $h(w) < r$, we get a model $\mathcal{A}$ of $\varphi_{0w} \wedge \varphi_{1M}$ with $A = \{0, 1, \dots, r\}$, for example "by repeating the configuration reached after $h(w)$ steps". This means, if T is any of the relations $C_0, H_q, C_0^{\mathrm{comp}}, H_q^{\mathrm{comp}}$, we set for j with $h(w) < j \leq r$ and $i = 0, \dots, r$,

$$(i,j) \in T^{\mathcal{A}} \iff (i, h(w)) \in T^{\mathcal{A}}. \qquad \dashv$$

Let γ_M be a sentence expressing that "M reaches the halting state q_h in exactly 'max' steps," e.g., we let γ_M be

$$\exists t \exists x \big(U_{\max} t \wedge H_{q_h}(x,t) \wedge \forall t' \forall y (t' < t \to \neg H_{q_h}(y,t')) \big). \tag{19}$$

As a consequence of the preceding lemma, we obtain the following:

COROLLARY 6.2. *Let $w \in \{0,1\}^*$ and set*

$$\pi_w := \varphi_{0w} \wedge \varphi_{1M} \wedge \gamma_M.$$

(a) *If M on w does not halt, then π_w has no finite model.*

(b) *Assume M on w eventually halts, say in $h(w)$ steps. Then π_w has a unique model up to isomorphism. This model is finite and has exactly $h(w)+1$ elements.*

We set

$$\chi_w := \varphi_{0w} \wedge (\varphi_{1M} \to \neg \gamma_M). \tag{20}$$

Applying Proposition 3.11 to part (b) of the preceding corollary, we get the following:

LEMMA 6.3. *Let M on w halt in $h(w)$ steps. Then*:

(a) $\mathrm{MOD}(\chi_w)$ *is closed under $<$-substructures.*

(b) *If χ_w is finitely equivalent to a universal sentence μ, then $|\mu| \geq h(w)+1$.*

Now we show the following version of Gurevich's Theorem.

THEOREM 6.4. *Let $f : \mathbb{N} \to \mathbb{N}$ be a computable function. Then there is a $w \in \{0,1\}^*$ such that $\mathrm{MOD}(\chi_w)$ is closed under $<$-substructures (and hence equivalent to a universal sentence) but χ_w is not finitely equivalent to a universal sentence of length less than $f(|\chi_w|)$.*

Note that by Corollary 2.7 the conclusion of this theorem is only apparently stronger than "χ_w is not *equivalent* to a universal sentence of length less than $f(|\chi_w|)$." A similar remark applies to Theorem 6.6.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

PROOF OF THEOREM 6.4. By the previous lemma it suffices to find a $w \in \{0,1\}^*$ such that M on input w halts in $h(w)$ steps with

$$h(w) \geq f(|\chi_w|).$$

W.l.o.g. we assume that f is increasing. An analysis of the formula χ_w shows that for some $c_M \in \mathbb{N}$ we have for all $w \in \{0,1\}^*$,

$$|\chi_w| \leq c_M \cdot |w|. \tag{21}$$

We define $g : \mathbb{N} \to \mathbb{N}$ by

$$g(k) := f(5 \cdot c_M \cdot k).$$

Let M_0 be a Turing machine computing g, more precisely, the function $1^k \mapsto 1^{g(k)}$. We code M_0 and 1^k by a $\{0,1\}$-string $code(M_0, 1^k)$ such that M on $code(M_0, 1^k)$ simulates the computation of M_0 on 1^k.

Choose the least k such that for $w := code(M_0, 1^k)$ we have

$$|w| \leq 5k. \tag{22}$$

The universal Turing machine M on input w computes $1^{g(k)}$ and thus runs at least $g(k)$ steps, say, exactly $h(w)$ steps. By (21) and (22)

$$h(w) \geq g(k) = f(5 \cdot c_M \cdot k) \geq f(c_M \cdot |w|) \geq f(|\chi_w|). \qquad \dashv$$

Finally we prove Gurevich's Theorem for graphs. For $\tau := \tau_M$ let I be an interpretation according to Lemma 5.6. For $w \in \{0,1\}^*$ we consider the sentence

$$\rho_w := \forall \bar{x} \neg \varphi_{\text{uni}}(\bar{x}) \vee (\varphi_{0w} \wedge (\varphi_{1M} \to \neg \gamma_M))^I = \forall \bar{x} \neg \varphi_{\text{uni}}(\bar{x}) \vee {\chi_w}^I. \tag{23}$$

That is, for $G \models \rho_w$, either the graph G interprets an "empty τ_M-structure," or a τ_M-structure that is a model of χ_w. If M halts in $h(w)$ steps on input w, then $\varphi_{0w} \wedge \varphi_{1M} \wedge \gamma_M$ has no infinite model but a finite model with $h(w)+1$ elements by Corollary 6.2(b). Hence, by Proposition 4.8 we get the following analogue of Lemma 6.3.

LEMMA 6.5. *Let M on input w halt in $h(w)$ steps. Then*:

(a) GRAPH(ρ_w), *the class of graphs that are models of ρ_w, is closed under induced subgraphs.*

(b) *If ρ_w is equivalent in the class of finite graphs to the universal sentence μ, then $|\mu|^2 \geq h(w)$.*

THEOREM 6.6 (Gurevich's Theorem for graphs). *Let $f : \mathbb{N} \to \mathbb{N}$ be a computable function. Furthermore, let ρ_w be defined by (23), where I is an interpretation for $\tau := \tau_M$ according to Lemma 5.6. Then there is a $w \in \{0,1\}^*$ such that* GRAPH(ρ_w) *is closed under induced subgraphs (and hence equivalent in the class of graphs to a universal sentence) but ρ_w is not equivalent in the class of finite graphs to a universal sentence of length less than $f(|\rho_w|)$.*

PROOF. Again we assume that f is increasing. By the previous lemma it suffices to find a $w \in \{0,1\}^*$ such that M on input w halts in $h(w)$ steps with

$$h(w) \geq f(|\rho_w|)^2.$$

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

There is a $c \in \mathbb{N}$, which depends on I but not on w, such that for c_I as in (11) and c_M as in (21) we have for $d_M := c + c_I \cdot c_M$,

$$|\rho_w| \leq c + c_I \cdot |\chi_w| \leq c + c_I \cdot c_M \cdot |w| \leq d_M \cdot |w|. \tag{24}$$

We define $g : \mathbb{N} \to \mathbb{N}$ by

$$g(k) := f(5 \cdot d_M \cdot k)^2 \tag{25}$$

and then proceed as in the proof of Theorem 6.4. Let M_0 be a Turing machine computing the function $1^k \mapsto 1^{g(k)}$. We code M_0 and 1^k by a $\{0,1\}$-string $code(M_0, 1^k)$ such that M on $code(M_0, 1^k)$ simulates the computation of M_0 on 1^k.

Choose the least k such that for $w := code(M_0, 1^k)$ we have

$$|w| \leq 5k. \tag{26}$$

The universal Turing machine M on input w computes $1^{g(k)}$ and thus runs at least $g(k)$ steps, say, exactly $h(w)$ steps. We have

$$h(w) \geq g(k) = f(5 \cdot d_M \cdot k)^2 \geq f(d_M \cdot |w|)^2 \geq f(|\rho_w|)^2$$

by (24)–(26). ⊣

REMARK 6.7. Using previous remarks (Remarks 3.12 and 4.9) one can even show that for every computable function $f : \mathbb{N} \to \mathbb{N}$ the sentence χ_w is not finitely equivalent to a Π_2-sentence of length less than $f(|\chi_w|)$ and the sentence ρ_w is not finitely equivalent in graphs to a Π_2-sentence of length less than $f(|\chi_w|)$. Moreover, χ_w and ρ_w are equivalent to Σ_2-sentences. To verify this note that in models of φ_{0w} the sentence γ_M is equivalent to

$$\exists t \exists x \big(U_{\max} t \wedge H_{q_h}(x,t)\big) \wedge \forall t_1 \forall t_2 \forall y \big(t_1 < t_2 \to \neg H_{q_h}(y,t_2)\big),$$

and hence equivalent to a Σ_2 and to a Π_2-sentence. One easily verifies that the same holds for γ_M^I.

§7. Some undecidable problems. In this section we show that various problems related to the results of the preceding sections are undecidable. Among others, these results explain why it might be hard, in fact impossible in general, to algorithmically obtain forbidden induced subgraphs for various classes of graphs.

A simple application of Gurevich's Theorem for graphs yields:

PROPOSITION 7.1. *There is no algorithm that applied to any* FO$[\tau_E]$*-sentence* φ *decides whether the class* GRAPH(φ) *is closed under induced subgraphs.*

PROOF. Assume $\mathbb{A}$ is such an algorithm. By the Completeness Theorem there is an algorithm $\mathbb{B}$ that assigns to every sentence φ such that GRAPH(φ) is closed under induced subgraphs a universal sentence equivalent to φ in graphs. Define the function g by

$$g(\varphi) := \begin{cases} 0, & \text{if } \mathbb{A} \text{ rejects } \varphi, \\ m, & \mathbb{B} \text{ needs } m \text{ steps to produce a universal sentence equivalent to } \varphi \text{ in graphs,} \end{cases}$$

and set $f(k) := \max\{g(\varphi) \mid |\varphi| \leq k\}$. Then f would contradict Theorem 6.6. ⊣

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

COROLLARY 7.2. *There is no algorithm that applied to any* FO[τ_E]*-sentence* φ *either reports that* GRAPH(φ) *is not closed under induced subgraphs or it computes for* GRAPH(φ) *a finite set of forbidden induced finite subgraphs.*

PROOF. Otherwise we could use this algorithm as a decision algorithm for the previous result. ⊣

The following proposition is the analogue of Proposition 7.1 for classes of finite graphs. We state it for FO[τ_E]-sentences and graphs even though we prove it for FO[τ_M]-sentences. One gets the version for graphs using the machinery we developed in previous sections similarly as we get Corollary 7.5 along the lines of the proof of Proposition 7.4.

We write $M : w \mapsto \infty$ for the universal Turing machine M and a word $w \in \{0,1\}^*$ if M on input w does not halt. We make use of the sentences φ_{0w}, φ_{1M}, and γ_M defined in the previous section.

PROPOSITION 7.3. *There is no algorithm that applied to any* FO[τ_E]*-sentence* φ *decides whether the class* GRAPH$_{fin}$(φ) *is closed under induced subgraphs.*

PROOF. For the universal Turing machine M and a word $w \in \{0,1\}^*$ consider the sentence

$$\pi_w = \varphi_{0w} \wedge \varphi_{1M} \wedge \gamma_M$$

introduced in Corollary 6.2. Then

$$M : w \mapsto \infty \iff \text{MOD}_{\text{fin}}(\pi_w) \text{ is closed under induced substructures.} \tag{27}$$

In fact, if $M : w \mapsto \infty$, then $\text{MOD}_{\text{fin}}(\pi_w) = \emptyset$ (see Corollary 6.2(a)), hence $\text{MOD}_{\text{fin}}(\pi_w)$ is trivially closed under induced substructures. If M on input w halts after $h(w)$ steps, then, up to isomorphism, there is a unique model $\mathcal{A}_w$ of π_w and it has $h(w)+1$ elements (see Corollary 6.2(b)). Take an induced substructure of $\mathcal{A}_w$ with $h(w)$ elements (note that $h(w) \geq 1$). Hence this substructure is not a model of π_w and thus $\text{MOD}_{\text{fin}}(\pi_w)$ is not closed under induced substructures. As the halting problem for every universal Turing machine is not decidable, by (27) we get our claim. ⊣

PROPOSITION 7.4. *There is no algorithm that applied to any* FO[τ_M]*-sentence that is finitely equivalent to a universal sentence computes such a universal sentence.*

PROOF. Assume that there exists such an algorithm $\mathbb{A}$. It suffices to show for every $w \in \{0,1\}^*$ the statements (a) and (b) for

$$\chi_w = \varphi_{0w} \wedge (\varphi_{1M} \to \neg\gamma_M)$$

defined in (20).

(a) $\text{MOD}_{\text{fin}}(\chi_w) = \text{MOD}_{\text{fin}}(\mu)$ for some universal μ.
(b) $M : w \to \infty \iff \text{MOD}_{\text{fin}}(\chi_w) = \text{MOD}_{\text{fin}}(\varphi_{0w})$.

Then we can decide the halting problem for M by checking whether the universal sentence produced by the claimed algorithm $\mathbb{A}$ is finitely equivalent to the universal sentence φ_{0w}. This can be decided effectively by Corollaries 2.6 and 2.7, which leads to a contradiction.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

If M halts on w, say in $h(w)$ steps, then we get (a) by Lemma 6.3(a) and the Łoś–Tarski Theorem. Furthermore, by Corollary 6.2(b) we know that there is a finite structure with $h(w)+1$ elements that is a model of $\varphi_{0w} \wedge \varphi_{1M} \wedge \gamma_M$ and thus of $\varphi_{0w} \wedge \neg\chi_w$. Hence this structure is a model of $\varphi_{0w} \wedge \neg\mu$. In particular, μ (and hence, χ_w) is not finitely equivalent to φ_{0w}. Thus, also (b) holds if M halts on w.

If $M : w \to \infty$, then we show that $\text{MOD}_{\text{fin}}(\chi_w) = \text{MOD}_{\text{fin}}(\varphi_{0w})$ (this implies (a) and (b) in this case). Clearly, $\text{MOD}_{\text{fin}}(\chi_w) \subseteq \text{MOD}_{\text{fin}}(\varphi_{0w})$. Now let $\mathcal{A}$ be a finite model of φ_{0w}. If $\mathcal{A} \not\models \varphi_{1M}$, then $\mathcal{A} \models \chi_w$. Otherwise $\mathcal{A} \models \varphi_{1M}$ and then $\mathcal{A}$ correctly represents the first $|A|-1$ steps of the computation of M on w by Lemma 6.1. Thus $\mathcal{A}$ is a model of $\neg\gamma_M$ as M does not halt on w. Therefore, $\mathcal{A}$ is a model of χ_w. ⊣

COROLLARY 7.5. *There is no algorithm that applied to any* $\text{FO}[\tau_E]$*-sentence* φ *such that* $\text{GRAPH}_{fin}(\varphi)$ *has a finite set of forbidden induced finite subgraphs computes such a set.*

PROOF. Equivalently we show that there is no algorithm that applied to any $\text{FO}[\tau_E]$-sentence φ such that $\text{GRAPH}_{\text{fin}}(\varphi) = \text{GRAPH}_{\text{fin}}(\mu)$ for some universal sentence μ computes such a μ.

For graphs let I $(= I_{\tau_M})$ be a strongly existential interpretation of τ_M-structures in graphs according to Lemma 5.6. For $w \in \{0,1\}^*$ we consider the sentence

$$\rho_w = \forall\bar{x}\neg\varphi_{\text{uni}}(\bar{x}) \vee \chi_w{}^I$$

defined in (23) and show (a') and (b'), the analogues of (a) and (b) of the preceding proof.

(a') $\text{GRAPH}_{\text{fin}}(\rho_w) = \text{GRAPH}_{\text{fin}}(\mu)$ for some universal μ.
(b') $M : w \to \infty \iff \text{GRAPH}_{\text{fin}}(\rho_w) = \text{GRAPH}_{\text{fin}}(\forall x\neg\varphi_{\text{uni}}(\bar{x}) \vee \varphi_{0w}{}^I)$.

Then we get the claim of the corollary arguing as in the previous proof.

(a') holds by Lemma 6.5(a).

By Lemma 5.6 for every finite τ_M-structure $\mathcal{A}$ there is a finite graph G with $\mathcal{O}_I(G) \cong \mathcal{A}$. Therefore,

$$\begin{aligned} M : w \to \infty &\iff \text{MOD}_{\text{fin}}(\chi_w) = \text{MOD}_{\text{fin}}(\varphi_{0w}) && \text{(by (a) of the preceding proof)} \\ &\iff \text{GRAPH}_{\text{fin}}(\rho_w) = \text{GRAPH}_{\text{fin}}(\forall x\neg\varphi_{\text{uni}}(\bar{x}) \vee \varphi_{0w}{}^I) && \text{(by (14))}. \end{aligned}$$

⊣

Observe that Corollary 7.5 is precisely Theorem 1.3 as stated in the Introduction. Finally we prove Theorem 1.2, which is equivalent to the following result.

THEOREM 7.6. *There is no algorithm that applied to an* $\text{FO}[\tau_E]$*-sentence* φ *such that* $\text{GRAPH}_{fin}(\varphi)$ *is closed under induced subgraphs decides whether there is a finite set* $\mathcal{F}$ *of finite graphs such that*

$$\text{GRAPH}_{fin}(\varphi) = \text{FORB}_{fin}(\mathcal{F}).$$

PROOF. Again we prove the corresponding result for τ_M-sentences and τ_M-structures and leave it to the reader to translate it to graphs as in the previous

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

proof. That is, we show:

> *There is no algorithm that applied to an* $\mathrm{FO}[\tau_M]$*-sentence* φ *such that* $\mathrm{MOD}_{\mathrm{fin}}(\varphi)$ *is closed under induced substructures decides whether there is a finite set* $\mathcal{F}$ *of* τ_M*-structures such that*
>
> $$\mathrm{MOD}_{\mathrm{fin}}(\varphi) = \mathrm{FORB}_{\mathrm{fin}}(\mathcal{F}).$$

For $w \in \{0,1\}^*$ set

$$\alpha_w := \varphi_{0w} \wedge (\varphi_{1M} \to \gamma_M).$$

It suffices to show that $\mathrm{MOD}_{\mathrm{fin}}(\alpha_w)$ is closed under induced substructures and that

$$M : w \to \infty \iff \alpha_w \text{ is not finitely equivalent to a universal sentence.}$$

Assume first that $M : w \to \infty$. Then $\varphi_{0w} \wedge \varphi_{1M} \wedge \gamma_M$ has no finite model by Lemma 6.1(a) and the definition (19) of γ_M. Therefore, $\mathrm{MOD}_{\mathrm{fin}}(\alpha_w) = \mathrm{MOD}_{\mathrm{fin}}(\varphi_{0w} \wedge \neg\varphi_{1M})$. By Lemma 6.1(b) the sentence $\varphi_{0w} \wedge \neg\varphi_{1M}$ has arbitrarily large finite models. Recall that φ_{0w} is an extension of φ_0 and $\varphi_{1M} = \varphi_{1\tau_M}$ (see (16)). Hence, by Lemma 3.10, we know that $\mathrm{MOD}_{\mathrm{fin}}(\varphi_{0w} \wedge \neg\varphi_{1M})$ is closed under induced substructures but not finitely equivalent to a universal sentence.

Now assume that M on input w halts in $h(w)$ steps. Then Corollary 6.2(b) guarantees that there is a unique model $\mathcal{A}_w$ of $\varphi_{0w} \wedge \varphi_{1M} \wedge \gamma_M$; moreover, $|A_w| = h(w) + 1$. We present a finite set $\mathcal{F}$ of finite τ_M-structures such that

$$\mathrm{MOD}_{\mathrm{fin}}(\alpha_w) = \mathrm{FORB}_{\mathrm{fin}}(\mathcal{F}). \tag{28}$$

As φ_{0w} is universal, there is a finite set $\mathcal{F}_0$ of finite τ_M-structures such that

$$\mathrm{MOD}_{\mathrm{fin}}(\varphi_{0w}) = \mathrm{FORB}_{\mathrm{fin}}(\mathcal{F}_0).$$

We define the sets $\mathcal{F}_1$ and $\mathcal{F}_2$ as follows: For every τ_M-structure $\mathcal{B}$,

$$\mathcal{B} \in \mathcal{F}_1 \quad \text{iff} \quad \mathcal{B} \models \varphi_{0w} \wedge \varphi_{1M} \text{ and } B = [\ell] \text{ for some } \ell \leq h(w),$$

$$\mathcal{B} \in \mathcal{F}_2 \quad \text{iff} \quad \mathcal{B} \models \varphi_{0w} \wedge \varphi_{1M}^* \wedge \forall t \forall t'(t < t' \to \forall y \neg H_{q_h}(y,t)) \text{ and } B = [h(w)+2].$$

Here φ_{1M}^* is obtained from φ_{1M} by replacing the conjunct φ_1 see (6) by

$$\varphi_1^* := \exists x U_{\min} x \wedge \forall x \forall y (x < y \to \exists z S x z).$$

The difference is that φ_1^* does not require the set $U_{\max}$ to be nonempty. Hence, φ_{1M}^* is the conjunction of φ_1^* with

$$\forall x \forall t \big((C_0(x,t) \vee C_0^{\mathrm{comp}}(x,t)) \wedge \bigwedge_{q \in Q} (H_q(x,t) \vee H_q^{\mathrm{comp}}(x,t)) \big).$$

Note that Lemma 6.1(a) remains true if in its statement we replace φ_{1M} by φ_{1M}^*.

For $\mathcal{F} := \mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2$ we show (28). Assume first that a finite structure $\mathcal{C}$ is a model of α_w. In particular, $\mathcal{C} \models \varphi_{0w}$ and therefore, $\mathcal{C}$ has no induced substructure isomorphic to a structure in $\mathcal{F}_0$.

Now, for a contradiction suppose that $\mathcal{B}$ is an induced substructure of $\mathcal{C}$ isomorphic to a structure in $\mathcal{F}_1$. Then $\mathcal{B} \models \varphi_{1M}$ and thus, by Lemma 3.8, $\mathcal{C} = \mathcal{B}$. As

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

$\mathcal{C} \models \alpha_w$, we get $\mathcal{C} \models \varphi_{0w} \wedge \varphi_{1M} \wedge \gamma_M$. Hence, $\mathcal{C} \cong \mathcal{A}_w$, a contradiction, as on the one hand $|C| = |B| \leq h(w)$ and on the other hand $|C| = |A_w| = h(w) + 1$.

Next we show that $\mathcal{C}$ has no induced substructure $\mathcal{B}$ isomorphic to a structure in $\mathcal{F}_2$. As $\mathcal{B} \models \varphi_{0w} \wedge \varphi^*_{1M}$ and has $h(w) + 2$ elements, the first $h(w) + 1$ elements of $\mathcal{B}$ correctly encode the first $h(w)$ steps of the computation of M on w, hence the full computation. As $|B| = h(w) + 2$, this contradicts $\mathcal{B} \models \forall t \forall t' \big(t < t' \rightarrow \forall y \neg H_{q_h}(y, t)\big)$.

As the final step let $\mathcal{C} \in \text{FORB}_{\text{fin}}(\mathcal{F})$. We show that $\mathcal{C} \models \alpha_w$. As any structure in $\mathcal{C}$ does not contain structures in $\mathcal{F}_0$ as induced substructures, we see that $\mathcal{C} \models \varphi_{0w}$. If $\mathcal{C} \not\models \varphi_{1M}$, we are done.

Recall that by Lemma 6.1(a) (more precisely, by the extension of Lemma 6.1(a) mentioned above) for finite models $\mathcal{B}$ of $\varphi_{0w} \wedge \varphi^*_{1M}$ we know:

(a) if $|B| \leq h(w) + 1$, then $\mathcal{B}$ encodes $|B| - 1$ steps of the computation of M on w,
(b) if $|B| > h(w) + 1$, then the first $h(w) + 1$ elements in the ordering $<^{\mathcal{B}}$ correctly encode the (full) computation of M on w.

Now assume that $\mathcal{C} \models \varphi_{1M}$, then (a) and (b) apply to $\mathcal{C}$. As no structure in $\mathcal{F}_1$ is isomorphic to an induced substructure of $\mathcal{C}$, we see that $|C| \geq h(w) + 1$. But $\mathcal{C}$ cannot have more than $h(w) + 1$ elements, as otherwise the substructure of $\mathcal{C}$ induced on the first $h(w) + 2$ elements would be isomorphic to a structure $\mathcal{B}$ in $\mathcal{F}_2$, a contradiction. Hence, $|C| = h(w) + 1$ and thus, $\mathcal{C} \models \alpha_w$. ⊣

REMARK 7.7. Mainly using Remark 6.7 one easily verifies that in all results but Proposition 7.3 of this section we can replace:

There is no algorithm that applied to an FO[τ_E]-sentence φ ...

by

There is no algorithm that applied to a Σ_2-sentence φ ....

In Proposition 7.3 we have to replace it by:

There is no algorithm that applied to a Π_2-sentence φ ...

as φ_{1M} (and φ^I_{1M}) are Π_2-sentences.

7.1. Open problem. The main result of this paper shows that the analogue of the Łoś–Tarski Theorem fails for the class of finite graphs. That is, there exist FO-axiomatizable classes of finite graphs closed under induced subgraphs that are not definable by a finite set of forbidden induced subgraphs. Often in graph theory one considers subgraphs instead of induced subgraphs. It is known that FO-axiomatizable classes of finite and infinite graphs are closed under subgraphs if and only if they are definable by a finite set of forbidden finite subgraphs. However, to the best of our knowledge it is still open whether FO-axiomatizable classes of finite graphs closed under subgraphs are definable by a finite set of forbidden subgraphs.

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press

Acknowledgment. We thank Abhisekh Sankaran for mentioning to the first author the question of whether Tait's Theorem generalizes to graphs (see also [23]).

Funding. The collaboration of the authors is funded by the Sino-German Center for Research Promotion (GZ 1518). Yijia Chen is supported by the National Natural Science Foundation of China (Project 62372291). He also likes to express his gratitude to Hong Xu and Liqun Zhang for offering a very cordial working environment at Fudan through the difficult year of 2020.

REFERENCES

[1] N. Alechina and Y. Gurevich, *Syntax vs. semantics on finite structures*, ***Structures in Logic and Computer Science, A Selection of Essays in Honor of Andrzej Ehrenfeucht*** (J. Mycielski, G. Rozenberg, and A. Salomaa, editors), Lecture Notes in Computer Science, 1261, Springer, Berlin, 1997, pp. 14–33.

[2] A. Atserias, A. Dawar, and M. Grohe, *Preservation under extensions on well-behaved finite structures*. ***SIAM Journal on Computing***, vol. 38 (2008), pp. 1364–1381.

[3] Y. Chen and J. Flum, *FO-definability of shrub-depth*. ***28th EACSL Annual Conference on Computer Science Logic, CSL 2020***, 13–16 January 2020, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Barcelona, Spain, 2020, pp. 1–15.

[4] ———, *Forbidden induced subgraphs and the Łoś–Tarski theorem*. ***36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021***. IEEE, 2021, pp. 1–13.

[5] A. Dawar, M. Grohe, S. Kreutzer, and N. Schweikardt, *Model theory makes formulas large*. ***Automata, Languages and Programming, 34th International Colloquium, ICALP 2007***, Springer, Wroclaw, Poland, 9–13 July 2007, 2007, pp. 913–924.

[6] A. Dawar and A. Sankaran, *Extension preservation in the finite and prefix classes of first order logic*. ***29th EACSL Annual Conference on Computer Science Logic, CSL 2021***, LIPIcs, 183, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021, pp. 1–18.

[7] G. Ding, *Subgraphs and well-quasi-ordering*. ***Journal of Graph Theory***, vol. 16 (1992), no. 5, pp. 489–502.

[8] R. G. Downey and M. R. Fellows, ***Parameterized Complexity***, Springer, New York, 1999.

[9] D. Duris, *Extension preservation theorems on classes of acyclic finite structures*. ***SIAM Journal on Computing***, vol. 39 (2010), no. 8, pp. 3670–3681.

[10] Z. Dvořák, A. C. Giannopoulou, and D. M. Thilikos, *Forbidden graphs for tree-depth*. ***European Journal of Combinatorics***, vol. 33 (2012), no. 5, pp. 969–979.

[11] H.-D. Ebbinghaus and J. Flum, ***Finite Model Theory***, Perspectives in Mathematical Logic, Springer, Berlin, 1999.

[12] M. R. Fellows, *Private communication*, 2019.

[13] M. R. Fellows and M. A. Langston, *On search, decision, and the efficiency of polynomial-time algorithms*. ***Journal of Computer and System Sciences***, vol. 49 (1994), no. 3, pp. 769–779.

[14] J. Gajarský and S. Kreutzer, *Computing shrub-depth decompositions*. ***37th International Symposium on Theoretical Aspects of Computer Science, STACS 2020***, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, pp. 1–56.

[15] R. Ganian, P. Hlinený, J. Nesetril, J. Obdrzálek, and P. Ossona de Mendez, *Shrub-depth: Capturing height of dense graphs*. ***Logical Methods in Computer Science***, vol. 15 (2019), no. 1, pp. 7:1–7:25.

[16] R. Ganian, P. Hlinený, J. Nesetril, J. Obdrzálek, P. Ossona de Mendez, and R. Ramadurai, *When trees grow low: Shrubs and fast*$_1$. ***Mathematical Foundations of Computer Science 2012 – 37th International Symposium, MFCS 2012***, Springer, Bratislava, Slovakia, 27–31 August 2012, 2012, pp. 419–430.

[17] Y. Gurevich, *Toward logic tailored for computational complexity*. ***Lecture Notes in Mathematics***, vol. 1104 (1984), pp. 175–216.

[18] A. Lopez, *When locality meets preservation*. ***LICS'22: 37th Annual ACM/IEEE Symposium on Logic in Computer Science, 2022***, ACM, 2022, pp. 1–46.

[19] J. Łoś, *On the extending of models I*. ***Fundamenta Mathematicae***, vol. 42 (1955), pp. 38–54.

[20] T. A. McKee, *Forbidden subgraphs in terms of forbidden quantifiers*. ***Notre Dame Journal of Formal Logic***, vol. 19 (1978), pp. 186–188.

[21] B. Rossman, *Homomorphism preservation theorems*. ***Journal of the ACM***, vol. 55 (2008), no. 3, pp. 1–15.

[22] ———, *Łoś–Tarski Theorem has non-recursive blow-up*, Unpublished manuscript, 2012, pp. 1–2.

[23] A. Sankaran, *Revisiting the generalized Łoś–Tarski Theorem*. ***Logic and Its Applications - 8th Indian Conference, ICLA 2019***, Delhi, 1–5 March 2019, Lecture Notes in Computer Science, 11600, Springer, 2019, pp. 76–88.

[24] A. Sankaran, B. Adsul, and S. Chakraborty, *A generalization of the Łoś–Tarski preservation theorem*. ***Annals of Pure and Applied Logic***, vol. 167 (2016), no. 3, pp. 189–210.

[25] W. W. Tait, *A counterexample to a conjecture of Scott and Suppes*, this Journal, vol. 24 (1959), no. 1, pp. 15–16.

[26] A. Tarski, *Contributions to the theory of models I–II*. ***Indagationes Mathematicae***, vol. 16 (1954), pp. 572–588.

[27] R. Vaught, *Remarks on universal classes of relational systems*. ***Indagationes Mathematicae***, vol. 16 (1954), pp. 589–591.

[28] T. Zaslavsky, *Forbidden induced subgraphs*. ***Electronic Notes in Discrete Mathematics***, vol. 63 (2017), pp. 3–10.

DEPARTMENT OF COMPUTER SCIENCE
SHANGHAI JIAO TONG UNIVERSITY
SHANGHAI, CHINA

E-mail: yijia.chen@cs.sjtu.edu.cn

MATHEMATISCHES INSTITUT
UNIVERSITÄT FREIBURG
FREIBURG, GERMANY

E-mail: joerg.flum@math.uni-freiburg.de

https://doi.org/10.1017/jsl.2023.99 Published online by Cambridge University Press