Forbidden Induced Subgraphs and the Łoś–Tarski Theorem

Yijia Chen
Department of Computer Science
Shanghai Jiao Tong University, China
Email: yijia.chen@cs.sjtu.edu.cn

Jörg Flum
Mathematisches Institut
Albert-Ludwigs-Universität Freiburg, Germany
Email: flum@uni-freiburg.de

2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS) | 978-1-6654-4895-6/20/$31.00 ©2021 IEEE | DOI: 10.1109/LICS52264.2021.9470742

***Abstract*—Let $\mathscr{C}$ 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 $\mathscr{C}$ is definable in first-order logic (FO) by a sentence φ if and only if $\mathscr{C}$ has a finite set of forbidden induced finite subgraphs. It 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. Our results (a) and (b) show that this machinery fails on finite graphs.**

(a) There is a class of finite graphs that is definable in FO and closed under induced subgraphs but has no finite set of forbidden induced subgraphs.

(b) Even if we only consider classes $\mathscr{C}$ of finite graphs that can be characterized by a finite set of forbidden induced subgraphs such a characterization cannot be computed from an FO-sentence φ that defines $\mathscr{C}$ and the size of the characterization cannot be bounded by $f(|\varphi|)$ for any computable function f.

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

***Index Terms*—Forbidden induced subgraphs, universal first-order sentences.**

I. 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 $\mathscr{F}_d$ consist of all graphs with vertex set $\{1, \ldots, d+2\}$ and maximum degree exactly $d+1$. Then a graph G has degree at most d if and only if no graph in $\mathscr{F}_d$ is isomorphic to an induced subgraph of G. Less trivial examples include classes of graphs of small vertex cover (attributed to Lovász [11]), of bounded tree-depth [6], and of bounded shrub-depth [15]. As a matter of fact, understanding forbidden induced subgraphs for those graph classes is an important question in structural graph theory [9], [13], [14], [24]. However, a straightforward adaptation of a result in [12] shows that it is in general impossible to compute the forbidden induced subgraphs from a description of classes of graphs by Turing machines.

Łoś [18] and Tarski [22] proved the first so-called preservation theorem of classical model theory. In its simplest form it says 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 that a universal sentence μ is a sentence (equivalent to a sentence) of the form $\forall x_1 \ldots \forall x_k \, \mu_0$ with quantifier-free μ_0.

It is folklore (e.g., see [17]) that for a class of graphs its definability by a universal sentence of first-order logic is equivalent to its characterization by finitely many forbidden induced finite subgraphs. For any class $\mathscr{C}$ of graphs closed under induced subgraphs we have $\mathscr{C} = \textsc{Forb}(\mathscr{F})$[1] where $\mathscr{F}$ consists of all graphs not in $\mathscr{C}$. Observe that $\mathscr{F}$ might be infinite and also contain some infinite graphs. However, for a universal sentence $\mu = \forall x_1 \ldots \forall x_k \, \mu_0$ as above we have

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

Here for any FO-sentence φ and $k \geq 1$ by $\mathscr{F}_k(\varphi)$ we denote the class of graphs that are models of $\neg\varphi$ and whose universe is $\{1, \ldots, \ell\}$ for some ℓ with $1 \leq \ell \leq k$. Clearly $\mathscr{F}_k(\varphi)$ consists of finitely many finite graphs. Moreover, (1) implies $\textsc{Graph}_{\text{fin}}(\mu) = \textsc{Forb}_{\text{fin}}\big(\mathscr{F}_k(\mu)\big)$ where $\textsc{Graph}_{\text{fin}}(\mu)$ and $\textsc{Forb}_{\text{fin}}\big(\mathscr{F}_k(\mu)\big)$ denote the class of finite graphs in $\textsc{Graph}(\mu)$ and in $\textsc{Forb}\big(\mathscr{F}_k(\mu)\big)$, respectively.

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

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

(i) $\mathscr{C}$ is closed under induced subgraphs and FO-axiomatizable.
(ii) $\mathscr{C}$ is axiomatizable by a universal sentence.
(iii) $\mathscr{C}$ 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 Vaught's article [23] of 1954. It is easy to see that the equivalence between (ii) and (iii) holds too if we only consider classes of finite graphs.

[1] $\textsc{Forb}(\mathscr{F})$ consists of all graphs that do not contain an induced subgraph isomorphic to a graph in $\mathscr{F}$.

978-1-6654-4895-6/21/$31.00

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 [2], to obtain the forbidden induced subgraph characterization of classes of graphs of bounded shrub-depth, one simple but vital step is to extend the notion of shrub-depth to classes of finite and infinite graphs in order to apply the Łoś–Tarski Theorem. Indeed, Tait [21] exhibited a class $\mathscr{C}$ of finite structures that is closed under induced substructures and FO-axiomatizable. Yet, $\mathscr{C}$ is not definable by any universal sentence, thus cannot be defined by a finite set of forbidden induced finite substructures. The first result of this paper strengthens Tait's result to graphs.

Theorem I.1. *There is a class $\mathscr{C}$ of finite graphs that is closed under induced subgraphs and* FO*-axiomatizable but not definable by a finite set of forbidden induced finite subgraphs.*

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. So 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 I.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 defined by a finite set of forbidden induced finite subgraphs. More precisely:

Theorem I.2. *There is no algorithm that for any* FO*-sentence φ such that $\textsc{Graph}_{\text{fin}}(\varphi)$ is closed under induced subgraphs decides whether φ is equivalent to a universal sentence on finite graphs.*

As mentioned at the beginning, for a class of finite graphs definable by a finite set of forbidden induced subgraphs, it is preferable to have an explicit construction of forbidden induced finite subgraphs. This however turns out to be difficult for many natural classes of graphs. For example, forbidden induced subgraphs are only known for tree-depth at most 3 [9]. 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 following FO-sentence φ^k_{VC}:

$$\varphi_{\textsc{Graph}} \wedge \exists x_1 \ldots \exists x_k \forall y \forall z \big(Eyz \rightarrow \bigvee_{1 \leq \ell \leq k} (x_\ell = y \vee x_\ell = z) \big)$$

where $\varphi_{\textsc{Graph}}$ 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 [7], to get a μ we can alternatively use that a graph with vertex cover number k admits a kernel with at most k^2 edges; this approach involves a co-NP hard problem.

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

Theorem I.3. *There is no algorithm that for any* FO*-sentence φ 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 $\textsc{Graph}_{\text{fin}}(\varphi) = \textsc{Forb}_{\text{fin}}(\mathscr{F})$ for a finite set $\mathscr{F}$ of finite graphs computes such an $\mathscr{F}$.*

Theorem I.4. *Let $f : \mathbb{N} \to \mathbb{N}$ be computable. Then there is a class $\mathscr{C}$ of finite graphs and an* FO*-sentence φ such that*

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

Theorem I.3 significantly strengthens the aforementioned result of [12]: even if a class $\mathscr{C}$ of finite graphs definable by a finite set of forbidden induced finite subgraphs is given by an FO-sentence φ with $\mathscr{C} = \textsc{Graph}_{\text{fin}}(\varphi)$, instead of a (much more powerful) Turing machine deciding $\mathscr{C}$, we still cannot compute an appropriate finite set of forbidden induced finite subgraphs for $\mathscr{C}$ from φ. On top of it, Theorem I.4 implies that the size of forbidden subgraphs for $\mathscr{C}$ cannot be bounded by any computable function in terms of the size of φ.

There is an important precursor for Theorem I.4,

Theorem I.5 (Gurevich's Theorem [16])**.** *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 $\textsc{Mod}_{\text{fin}}(\mu) = \textsc{Mod}_{\text{fin}}(\varphi)$ we have $|\mu| \geq f(|\varphi|)$.*

Hence, Theorem I.4 can be viewed as the graph-theoretic version of Theorem I.5.

Besides its importance in graph theory, Theorem I.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 [4], it is shown that on *labeled 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 I.4 is not comparable to Theorem 6.1 in [4], since our lower bound is uncomputable (and thus, much worse 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 [10]). So one might expect that Theorem I.1 and Theorem I.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*, which translates

any class of structures closed under induced substructures to a class of graphs closed under induced subgraphs. A lot of care is needed to construct strongly existential interpretations.

Related research. Let us briefly mention some further results related to the Łoś–Tarski Theorem. Essentially one could divide them into three categories: (a) The *positive results* showing that for certain classes $\mathscr{C}$ of finite structures the analogue of the Łoś–Tarski Theorem holds if we restrict to structures in $\mathscr{C}$. For example, this is the case if $\mathscr{C}$ is the class of all finite structures of tree-width at most k for some $k \in \mathbb{N}$ [1] or if $\mathscr{C}$ is the class of all finite structures whose hypergraph satisfies certain properties [8]. (b) The just mentioned papers contain also *negative results*, i.e, classes for which the analogue of the Łoś–Tarski Theorem fails: E. g., in [1] this is shown for the class of finite planar graphs, a not FO-axiomatizable class (cf. Remark V.7 (b)).(c) The third category contains generalizations of the Łoś–Tarski Theorem (see [5], [20]). For example, in [5] the authors strengthen Tait's result by showing that for every $n \geq 1$ there are first-order definable classes of finite structures closed under substructures that are not definable with n quantifier alternations.

Organization of this paper. In Section II we fix some notations and recall or derive some results about universal sentences we need in this paper. In Section III we present a proof of Tait's result. Moreover, we prove a technical result (Proposition III.11) that is an important tool in Gurevich's Theorem. We introduce the concept of strongly existential interpretation in Section IV 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 V. Hence, we get the results of Section III for graphs. In Section VI we first derive Gurevich's Theorem and apply our interpretations to get the corresponding result for graphs. Finally, in Section VII, we prove that various problems related to our results are undecidable.

Due to space limitations we only sketch some proofs or refer to the full version of the paper [3].

II. Preliminaries

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

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 $R \in \tau$. Let $\mathcal{A}$ and $\mathcal{B}$ be 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 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 $A \subsetneq B$. By $\text{STR}[\tau]$ ($\text{STR}_{\text{fin}}[\tau]$) we denote the class of all (of all finite) τ-structures.

Formulas φ of *first-order logic* FO of vocabulary τ are built up from *atomic formulas* $x_1 = x_2$ and $Rx_1 \ldots x_r$ (where $R \in \tau$ is of arity r and $x_1, x_2, \ldots, x_r$ 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 $R\ldots$ in φ appear in the scope of an *even (odd)* number of negation symbols. By the notation $\varphi(\bar{x})$ with $\bar{x} = x_1, \ldots, x_e$ we indicate that the variables free in φ are among $x_1, \ldots, x_e$. If $\mathcal{A}$ is a τ-structure and $a_1, \ldots, a_e \in A$, then $\mathcal{A} \models \varphi(a_1, \ldots, a_e)$ 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 $\text{MOD}_{\text{fin}}(\varphi) = \text{MOD}_{\text{fin}}(\psi)$.

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 self-loops, i.e., $(v, v) \notin E(G)$ for any $v \in V(G)$. If moreover $(u, v) \in E(G)$ implies $(v, u) \in E(G)$ for any pair (u, v), then G is an (undirected) *graph*. We denote by GRAPH and $\text{GRAPH}_{\text{fin}}$ the class of all graphs and the class of all 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.

Universal sentences and forbidden induced substructures. *All results of this section are more or less known. For the proofs we point to the literature or to the full version of this paper.*

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 $\forall x_1 \ldots \forall x_k\, \mu'_0$ 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 universal sentence is closed under induced substructures. Łoś [18] and Tarski [22] proved:

Theorem II.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 fix a vocabulary τ. Let $\mathscr{F}$ be a class of τ-structures and denote by $\text{FORB}(\mathscr{F})$ (and $\text{FORB}_{\text{fin}}(\mathscr{F})$) the class of structures (of finite structures) that do not contain an induced substructure isomorphic to a structure in $\mathscr{F}$. Clearly for two classes $\mathscr{F}$ and $\mathscr{F}'$ of τ-structures we have

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

We say that a class $\mathscr{C}$ of τ-structures (of finite τ-structures) is *definable by a finite set of forbidden induced finite substructures* if there is a *finite* set $\mathscr{F}$ of *finite* structures such that $\mathscr{C} = \text{FORB}(\mathscr{F})$ ($\mathscr{C} = \text{FORB}_{\text{fin}}(\mathscr{F})$).

Recall that $\tau_E = \{E\}$ with binary E. The sentences $\varphi_{\text{DG}} := \forall x \neg Exx$ and $\varphi_{\text{GRAPH}} := \forall x \neg Exx \wedge \forall x \forall y (Exy \rightarrow Eyx)$ axiomatize the classes of directed graphs and of graphs, respectively. Define the τ_E-structures H_0 and H_1 by

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

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, i.e., $\text{MOD}(\varphi_{\text{DG}}) = \text{FORB}(\{H_0\})$ and $\text{MOD}(\varphi_{\text{GRAPH}}) = \text{FORB}(\{H_0, H_1\})$.

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

$$\mathscr{F}_k(\varphi) := \{\mathcal{A} \in \text{STR}[\tau] \mid \mathcal{A} \models \neg\varphi \text{ and } A = [\ell] \text{ with } \ell \in [k]\}.$$

Thus, $\mathscr{F}_k(\varphi)$ is, up to isomorphism, the class of structures with at most k elements that fail to be a model of φ. Note that $\mathscr{F}_1(\varphi_{\text{DG}}) = \{H_0\}$. Clearly, for a τ-sentence we have:

$$\text{if } \text{MOD}(\varphi) \text{ is closed under induced substructures,} \\ \text{then } \text{MOD}(\varphi) \subseteq \text{FORB}(\mathscr{F}_k(\varphi)) \text{ for all } k \geq 1. \quad (3)$$

Proposition II.2. *For a class $\mathscr{C}$ of τ-structures and $k \geq 1$ the statements (i) and (ii) are equivalent.*

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

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

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

Corollary II.3. *Let φ be a τ-sentence and $k \geq 1$. Then $\text{MOD}(\varphi) = \text{FORB}(\mathscr{F}_k(\varphi))$ iff φ is equivalent to a universal sentence of the form $\forall x_1 \ldots \forall x_k\, \mu_0$ with quantifier-free μ_0.*

By (2) and (3) we get:

Corollary II.4. *If $\text{MOD}(\mu) = \text{FORB}\big(\mathscr{F}_k(\mu)\big)$ for some universal μ and some $k \geq 1$, then $\text{MOD}(\mu) = \text{FORB}\big(\mathscr{F}_\ell(\mu)\big)$ for all $\ell \geq k$.*

Corollary II.5. *It is decidable whether two universal sentences are equivalent.*

Corollary II.6. *For universal sentences μ and μ' we have*

μ and μ' are equivalent iff μ and μ' are finitely equivalent.

The following consequence of Proposition II.2 will be used in the next section.

Corollary II.7. *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 $\forall x_1 \ldots \forall x_k\, \mu_0$ with quantifier-free μ_0.*

Remark II.8. Let $\mathscr{C}$ be a class of τ-structures closed under induced substructures. For an FO[τ]-sentence φ we set $\text{MOD}_{\mathscr{C}}(\varphi) := \{\mathcal{A} \in \mathscr{C} \mid \mathcal{A} \models \varphi\}$. We say that the *Łoś–Tarski Theorem holds for $\mathscr{C}$* if for every FO[τ]-sentence φ such that the class $\text{MOD}_{\mathscr{C}}(\varphi)$ is closed under induced substructures there is a universal sentence μ such that $\text{MOD}_{\mathscr{C}}(\varphi) = \text{MOD}_{\mathscr{C}}(\mu)$. The following holds:

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

III. 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 III.11, which is a key result to get Gurevich's Theorem.

We consider the vocabulary $\tau_0 := \{<, U_{\text{min}}, U_{\text{max}}, S\}$, 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

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

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 τ_0-*orderings* the models of φ_0.

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 $\mathcal{A}$ is a substructure of $\mathcal{B}$ 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:

Lemma III.1. *Let $\mathcal{B} \subseteq_< \mathcal{A}$. If $\mathcal{A} \models \varphi_0$, then $\mathcal{B} \models \varphi_0$.*

We set

$$\varphi_1 := \exists x\, U_{\text{min}}\, x \wedge \exists x U_{\text{max}}\, x \wedge \forall x \forall y (x < y \rightarrow \exists z Sxz). \quad (4)$$

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.

Lemma III.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 Lemma III.1 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

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

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 superscript $^{\mathcal{B}}$ by $^{\mathcal{A}}$.

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 III.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 class of finite τ_0-orderings that are not complete is closed under $<$-substructures but not axiomatizable by a universal sentence:

Theorem III.4 (Tait's Theorem)**.** *The class $\mathrm{MOD}_{\mathrm{fin}}(\varphi_0 \wedge \neg\varphi_1)$ 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.*

Proof : $\mathrm{MOD}_{\mathrm{fin}}(\varphi_0 \wedge \neg\varphi_1)$ 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 III.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 III.2, a contradiction.

Let $k \in \mathbb{N}$. There is a finite model $\mathcal{A}$ of $\varphi_0 \wedge \varphi_1$ with at least $k+1$ elements. By Corollary III.3 every proper induced substructure of $\mathcal{A}$ is a model of $\varphi_0 \wedge \neg\varphi_1$. Thus, by Corollary II.7, 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 III.5. A slight generalization of the previous proof shows that $\mathrm{MOD}_{\mathrm{fin}}(\varphi_0 \wedge \neg\varphi_1)$ is not even axiomatizable by a Π_2-sentence, i.e., by a sentence of the form $\forall x_1 \ldots \forall x_k \exists y_1 \ldots \exists y_\ell \, \nu_0$ for some $k, \ell \in \mathbb{N}$ and some quantifier-free ν_0.

The fact that $\mathrm{MOD}_{\mathrm{fin}}(\varphi_0 \wedge \neg\varphi_1)$ is not axiomatizable by a Π_2-sentence immediately follows also from the result due to Compton (see [16]) that every Π_2-sentence whose class of finite models is closed under induced substructures is finitely equivalent to a universal sentence.

Note that $\varphi_0 \wedge \neg\varphi_1$ is (equivalent to) a Σ_2-sentence, i.e., equivalent to the negation of a Π_2-sentence.

We turn to a refinement of Theorem III.4 that will be helpful to get Gurevich's Theorem.

Definition III.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

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

(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 (4)).

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:

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

Now the analogues of Lemma III.1–Theorem III.4, namely Lemma III.8–Lemma III.10, can be derived essentially by the same proofs. Thereby we always assume that τ is obtained from τ_0 by adding pairs and $\varphi_{0\tau}$ is an extension of φ_0.

Lemma III.7. *If $\mathcal{B} \subseteq_< \mathcal{A}$ and $\mathcal{A} \models \varphi_{0\tau}$, then $\mathcal{B} \models \varphi_{0\tau}$.*

Lemma III.8. *Assume that $\mathcal{A} \models \varphi_{0\tau}$ and that the finite $<$-substructure $\mathcal{B}$ of $\mathcal{A}$ is a model of $\varphi_{1\tau}$. Then $\mathcal{B} = \mathcal{A}$ (in particular, $\mathcal{A} \models \varphi_{1\tau}$).*

Corollary III.9. *Every finite proper $<$-substructure of a model of $\varphi_{0\tau} \wedge \varphi_{1\tau}$ is a model of $\varphi_{0\tau} \wedge \neg\varphi_{1\tau}$.*

Lemma III.10. *Assume that the sentence $\varphi_{0\tau} \wedge \varphi_{1\tau}$ has arbitrary large finite models. Then the class $\mathrm{MOD}_{\mathrm{fin}}(\varphi_{0\tau} \wedge \neg\varphi_{1\tau})$ is closed under $<$-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 version of Lemma III.8 (or of Lemma III.2) where we can replace "$<$-substructure" by "substructure." The reader will realize that the 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 III.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* FO$[\tau]$*-sentence such that*

$$\varphi_{0\tau} \wedge \varphi_{1\tau} \wedge \gamma \text{ has no infinite model but a finite model with at least } m \text{ elements.} \quad (5)$$

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

(a) *The class* $\text{MOD}(\chi)$ *is closed under $<$-substructures.*

(b) *If $\forall x_1 \dots \forall x_k\, \mu_0$ 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 $\mathcal{B} \models \varphi_{1\tau}$. If B is infinite, by (5) we know that $\mathcal{B}$ is a model of $\neg\gamma$ and hence of χ. If B is finite, then $\mathcal{B} = \mathcal{A}$ (by Lemma III.8) and thus, $\mathcal{B} \models \chi$.

(b) By (5) 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 III.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 II.7, $\varphi_{0\tau} \wedge (\varphi_{1\tau} \to \neg\gamma)$ is not finitely equivalent to a universal sentence of the form $\forall x_1 \dots \forall x_k\, \mu_0$ with $k < m$ and quantifier-free μ_0. □

Remark III.12. We can strengthen the statement (b) of the preceding proposition to:

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

IV. The general machinery: strongly existential interpretations

We show that appropriate interpretations preserve the validity of Tait's theorem and of the statement of Proposition III.11. Later on these interpretations will allow us to get versions of these 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 be an interpretation of *width* 2 (we only need this case) of τ-structures in τ_E-structures. This means that I assigns to every unary relation symbol $T \in \tau$ an FO$[\tau_E]$-formula $\varphi_T(x_1, x_2)$ and to every binary relation symbol $T \in \tau$ an FO$[\tau_E]$-formula $\varphi_T(x_1, x_2, y_1, y_2)$; moreover, I selects an FO$[\tau_E]$-formula $\varphi_{\text{uni}}(x_1, x_2)$.

Then for every τ_E-structure G we set

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

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

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

As the interpretation I is of width 2, we have

$$|O_I(G)| \leq |V(G)|^2. \quad (6)$$

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

$$(G_I =)\ \mathcal{O}_I(G) \models \varphi \iff G \models \varphi^I. \quad (7)$$

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

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

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

$$|\varphi^I| \leq c_I \cdot |\varphi|.$$

Definition IV.1. Let τ be obtained from τ_0 by adding pairs. An interpretation I of τ-structures in τ_E 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 IV.2. *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.*

The following result shows that strongly existential interpretations preserve induced substructures; this will be crucial to transfer the results of the preceding section to graphs.

Lemma IV.3. *Assume that I is strongly existential. Then for all τ_E-structures G and H with $H \subseteq_{\text{ind}} G$ 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

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

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

We obtain from Lemma III.8 the corresponding result in our framework.

Lemma IV.4. *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 $H \subseteq_{\text{ind}} G$ 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 $H \models (\exists x\, U_{\min} x)^I$; thus, $O_I(H) \neq \emptyset$. Therefore, $\mathcal{O}_I(H) \subseteq_< \mathcal{O}_I(G)$ by Lemma IV.3. By assumption and (7), $\mathcal{O}_I(G) \models \varphi_{0\tau}$ and $\mathcal{O}_I(H) \models \varphi_{1\tau}$. As $O_I(H)$ is finite, Lemma III.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 (7). □

We now prove two results for strongly existential interpretations: Proposition IV.5 corresponds to Tait's Theorem

(Theorem III.4), and Proposition IV.6 corresponds to Proposition III.11 (relevant to Gurevich's Theorem). In our application of these results to graphs in the next section the sentence ψ will be the sentence φ_{GRAPH} axiomatizing the class of graphs.

Proposition IV.5. *Let ψ be a universal τ_E-sentence. Assume that the interpretation I of τ_0-structures in τ_E-structures is strongly existential. Furthermore, assume that for every sufficiently large finite complete τ_0-ordering $\mathcal{A}$ there is a finite τ_E-structure G with $\mathcal{O}_I(G) \cong \mathcal{A}$ and $G \models \psi$. Then there is an $\text{FO}[\tau_E]$-sentence φ such that $\text{MOD}_{\text{fin}}(\psi \wedge \varphi)$ is closed under induced substructures, but $\psi \wedge \varphi$ is not finitely equivalent to a universal sentence. As φ we can take the sentence*

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

(for the definition of φ_0 and φ_1 see the third paragraph of Section III and (4)*, respectively).*

Proof : First we verify that the class $\text{MOD}_{\text{fin}}(\psi \wedge \varphi)$ is closed under induced substructures. Assume that G is finite and $G \models \psi \wedge \varphi$ and $H \subseteq_{\text{ind}} G$. Since ψ is universal, we have $H \models \psi$. If $G \models \forall \bar{x} \neg \varphi_{\text{uni}}(\bar{x})$, then $H \models \forall \bar{x} \neg \varphi_{\text{uni}}(\bar{x})$. Now assume that $G \models \varphi_0^I \wedge \neg \varphi_1^I$. Then $H \models \varphi_0^I$, as φ_0^I is universal by Lemma IV.2. If $H \models \forall \bar{x} \neg \varphi_{\text{uni}}(\bar{x})$ or $H \models \neg \varphi_1^I$, we are done. Otherwise $O_I(H) \neq \emptyset$ and $H \models \varphi_1^I$. Then $G \models \varphi_1^I$ (see Lemma IV.4), a contradiction.

Finally we show that for every $k \geq 1$ the sentence $\psi \wedge \varphi$ is not finitely equivalent to a sentence of the form $\forall z_1 \dots \forall z_k \, \mu_0$ with quantifier-free μ_0.

Let $\mathcal{A} := \left(A, <^{\mathcal{A}}, U_{\text{min}}^{\mathcal{A}}, U_{\text{max}}^{\mathcal{A}}, S^{\mathcal{A}}\right)$ be a finite and complete τ_0-ordering with at least $k^2 + 1$ elements. In particular, $\mathcal{A} \models \varphi_0 \wedge \varphi_1$. By assumption we can choose $\mathcal{A}$ in such a way that there is a finite τ_E-structure G such that $\mathcal{O}_I(G) \cong \mathcal{A}$ and $G \models \psi$. Then $\mathcal{O}_I(G) \models \varphi_0 \wedge \varphi_1$, hence, $G \models \varphi_0^I \wedge \varphi_1^I$. Thus $G \models \psi \wedge \neg \varphi$. As $|O_I(G)| = |A| \geq k^2 + 1$, the graph G must contain more than k elements by (6).

We want to show that every induced substructure of G with at most k elements is a model of $\psi \wedge \varphi$. Then the result follows from Corollary II.7. So let H be an induced substructure of G with at most k elements. Clearly, $H \models (\psi \wedge \varphi_0^I)$. If $H \models \forall \bar{x} \neg \varphi_{\text{uni}}(\bar{x})$ or $H \models \neg \varphi_1^I$, we are done. Otherwise $O_I(H) \neq \emptyset$ and $H \models \varphi_1^I$. Then, Lemma IV.4 implies $O_I(H) = O_I(G)$. Recall $|V(H)| \leq k$, so $O_I(H)$ has at most k^2 elements by (6), a contradiction as $|O_I(G)| \geq k^2 + 1$. □

Proposition IV.6. *Assume that ψ is a universal τ_E-sentence. Let τ be obtained from τ_0 by adding pairs and let $\varphi_{0\tau}$ be an extension of φ_0. Assume 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 τ_E-structure G with $\mathcal{O}_I(G) \cong \mathcal{A}$ and $G \models \psi$.*

Let $m \geq 1$ and γ be an $\text{FO}[\tau]$-sentence such that

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

For

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

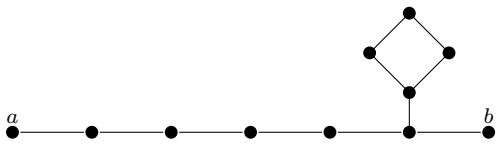

Fig. 1. A path of length 6 with a 4-ear.

the statements (a) and (b) hold.

(a) $\text{MOD}(\psi \wedge \rho)$ *is closed under induced substructures.*
(b) *If $\forall x_1 \dots \forall x_k \, \mu_0$ with quantifier-free μ_0 is finitely equivalent to $\psi \wedge \rho$, then $k^2 \geq m$.*

Remark IV.7. The results corresponding to Remark III.5 and Remark III.12 are valid for Proposition IV.5 and Proposition IV.6 too. In particular, $\psi \wedge \varphi$ $(= \psi \wedge \forall \bar{x} \neg \varphi_{\text{uni}}(\bar{x}) \vee (\varphi_0^I \wedge \neg \varphi_1^I))$ is not equivalent to a Π_2-sentence. Furthermore $\psi \wedge \varphi$ itself is equivalent to a Σ_2-sentence.

V. 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 VI.

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 of length r with a cycle of length s; one vertex of this cycle is adjacent to the vertex adjacent to b on the path. Figure 1 is a path from a to b of length 6 with a 4-ear.

Lemma V.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,*

(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 a 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

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

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

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

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 $G := G(\mathcal{A})$, which 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 $a \in A$, we introduce a new element a', the *companion of a*. We set

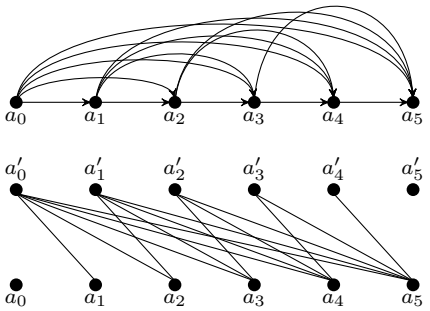

Fig. 2. Turning an ordering to the relation F.

- $A^* := A \cup \{a' \mid a \in A\}$,
- $B^{\mathcal{A}^*} := A$, $\quad C^{\mathcal{A}^*} := \{a' \mid a \in A\}$,
- $L^{\mathcal{A}^*} := \{(a, a') \mid a \in A\}$,
- $F^{\mathcal{A}^*} := \{(a', b), (b, a') \mid a, b \in A,\ a <^{\mathcal{A}} b\}$.

Note that the relation F is irreflexive and symmetric, i.e., $\left(A^*, F^{\mathcal{A}^*}\right)$ 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 following items:

- To $a \in U^{\mathcal{A}}_{\min}$ we add a cycle of length 5 consisting of new vertices, i.e., not in A^* (besides a).
- To $a \in U^{\mathcal{A}}_{\max}$ we add a cycle of length 7 consisting of new vertices (besides a).
- To $a \in B^{\mathcal{A}^*}$ we add a cycle of length 9 consisting of new vertices (besides a).
- To $a \in C^{\mathcal{A}^*}$ we add a cycle of length 11 consisting of new vertices (besides a).
- 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 mean by "add a cycle" or "add a path with an ear" that we only add the (undirected) edges required by the corresponding formulas in Lemma V.1.

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

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

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

[*ear-cycle*] These are the cycles that are the ears on the gadgets for 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 in Figure 2 we start with a_2, a'_0, a_1 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 V.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 $T \in \{U_{\min}, U_{\max}, B, C\}$ 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 $T \in \{U_{\min}, U_{\max}, B, C\}$. 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 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 this path has length 17, we get our claim. □

We want to recover $\mathcal{A}$ from $G(\mathcal{A})$ by means of a strongly existential interpretation. Let G be any graph. We define a τ_0-structure $\mathcal{O}(G)$, possibly the "empty structure". For the definitions of "cycle" and of "path with ear" see Lemma V.1.

- $O(G) := \{(a_1, a_2) \in V(G)^2 \mid a_1$ is a member of a cycle of length 9, a_2 is a member of a cycle of length 11, and there is a path from a_1 to a_2 of length 17 with a 15-ear$\}$
- $<^{\mathcal{O}(G)} := \{((a_1, a_2), (b_1, b_2)) \in O(G)^2 \mid \{a_2, b_1\} \in E(G)\}$
- $U^{\mathcal{O}(G)}_{\min} := \{(a_1, a_2) \in O(G) \mid a_1$ is a member of a cycle of 5 elements$\}$
- $U^{\mathcal{O}(G)}_{\max} := \{(a_1, a_2) \in O(G) \mid a_1$ is a member of a cycle of 7 elements$\}$
- $S^{\mathcal{O}(G)} := \{((a_1, a_2), (b_1, b_2)) \in O(G)^2 \mid$ there is a path from a_1 to b_1 of length 17 with a 13-ear$\}$.

Lemma V.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 $h(a) := (a, a')$ for $a \in A$ is an isomorphism from $\mathcal{A}$ to $\mathcal{A}^+$. To that end, we first prove that $A^+ = \{(a, a') \mid a \in A\}$, 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 V.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 V.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

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

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

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

Then we have:

Lemma V.4. *$I := (\varphi_{\text{uni}}, \varphi_<, \varphi_{U_{\min}}, \varphi_{U_{\max}}, \varphi_S)$ is a strongly existential 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 V.3,*

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

Setting $\psi := \varphi_{\text{GRAPH}}$, the sentence axiomatizing the class of graphs, we get from Proposition IV.5:

Theorem V.5 (Tait's Theorem for graphs)**.** *There is a τ_E-sentence φ such that $\text{GRAPH}_{\text{fin}}(\varphi)$, 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 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 to show the following result for vocabularies obtained from τ_0 by adding pairs. We shall use it in Section VI.

Lemma V.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 V.7. (a) Let $\mathscr{C} := \text{MOD}_{\text{fin}}(\forall x\neg Exx)$ be the class of finite directed graphs. Then $\mathscr{C}' := \text{GRAPH}_{\text{fin}}$, the class of finite graphs, is a subclass of $\mathscr{C}$ closed under induced substructures and definable in $\mathscr{C}$ 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 II.8.

(b) Let $\mathscr{C} := \text{GRAPH}_{\text{fin}}$ and $\mathscr{C}' := \text{PLANAR}_{\text{fin}}$ be the class of finite planar graphs, a subclass of $\mathscr{C}$ closed under induced subgraphs. In [1] it is shown that the Łoś–Tarski Theorem fails for $\text{PLANAR}_{\text{fin}}$. As $\text{PLANAR}_{\text{fin}}$ is not axiomatizable in $\text{GRAPH}_{\text{fin}}$ 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 finite graphs (i.e., Theorem V.5) by applying the result of Remark II.8. We show that $\text{PLANAR}_{\text{fin}} = \text{FORB}_{\text{fin}}(\mathscr{F})$ for a finite set $\mathscr{F}$ of finite graphs (or, equivalently, $\text{PLANAR}_{\text{fin}} = \text{MOD}_{\text{fin}}(\mu)$ for a universal μ) leads to a contradiction. Let k be the maximum size of the set of vertices of graphs in $\mathscr{F}$. Let G be the graph obtained from the clique K_5 of 5 vertices by subdividing each edge into $k + 1$ edges. Clearly, $G \notin \text{PLANAR}_{\text{fin}}$. However, every subgraph of G induced on at most k elements is planar. Hence, $G \in \text{FORB}_{\text{fin}}(\mathscr{F})$.

(c) Let τ be any vocabulary with at least one relation symbol T which is at least binary. Then the Łoś–Tarski Theorem fails for the class $\mathscr{C} := \text{STR}_{\text{fin}}[\tau]$, the class of all finite τ-structures. By Remark II.8 it suffices to show the existence of a universally definable subclass $\mathscr{C}'$ of $\mathscr{C}$ which "essentially is the class of graphs." We let $\mathscr{C}' = \text{MOD}_{\text{fin}}(\mu)$, where μ is

$$\forall x\bar{u}\neg Tx x\bar{u} \wedge \forall xy\forall\bar{u}\bar{v}(Txy\bar{u} \to Tyx\bar{v}) \wedge \bigwedge_{R\in\tau,\ R\neq T} \forall\bar{u}\neg R\bar{u}.$$

If τ only contains unary relation symbols, the Łoś–Tarski Theorem holds for $\text{STR}_{\text{fin}}[\tau]$ as for every $\text{FO}[\tau]$-sentence φ the closure under induced substructures of $\text{MOD}_{\text{fin}}(\varphi)$ implies that of $\text{MOD}(\varphi)$.

VI. Gurevich's Theorem

The following discussion will eventually lead to a proof of Gurevich's Theorem, i.e., Theorem I.5. Our proof essentially follows Gurevich's proof in [16], but it contains some elements of Rossman's proof of the same result in [19]. [2] Afterwards we show that it remains true if we restrict ourselves to graphs.

Our main tool is Proposition III.11: the goal is to construct a formula γ satisfying (5), 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 an 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

[2]The reader of [16] will realize that the definition of φ^n on page 190 of [16] must be modified in order to ensure that the class of models of φ^n is closed under induced substructures.

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 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 III.6 (a)),

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

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 III. For $w \in \{0,1\}^*$ the sentence φ_{0w} will be an extension of φ_0 (see Definition III.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

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

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

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

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. Hence, as φ_w we can take the conjunction of

– $\forall x_1 \dots \forall x_{|w|}\big((U_{\min} x_1 \rightarrow \neg C_{1-w_1}(x_1,x_1)) \wedge \bigwedge_{i\in[|w|-1]}(Sx_ix_{i+1} \rightarrow \neg C_{1-w_{i+1}}(x_{i+1},x_1))\big)$

– $\forall x_1 \dots \forall x_{|w|}\forall x((U_{\min} x_1 \wedge \bigwedge_{i\in[|w|-1]} Sx_ix_{i+1} \wedge x_{|w|} < x) \rightarrow \neg C_0^{\text{comp}}(x,x_1))$

– $\forall x\forall y\big(U_{\min} x \rightarrow (\neg H_{q_0}^{\text{comp}}(x,x) \wedge (y \neq x \rightarrow \bigwedge_{q\in Q} \neg H_q(y,x)))\big)$.

Note that $U_{\min}$, $U_{\max}$, and S are negative in φ_{0w}.

We set $\varphi_{1M} := \varphi_{1\tau_M}$; i.e., by Definition III.6 (c),

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

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 $0 \leq i, j \leq r$ and for $q \in Q$,

$$(a_i, a_j) \in C_0^{\mathcal{A}} \text{ iff the content of cell } i \text{ after } j \text{ steps is } 0 \quad (8)$$

$$(a_i, a_j) \in H_q^{\mathcal{A}} \text{ iff after } j \text{ steps } M \text{ is in state } q \text{ and the head scans cell } i. \quad (9)$$

Lemma VI.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 (8) and (9) 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 $(C_0^{\text{comp}})^{\mathcal{A}}$ and $(H_q^{\text{comp}})^{\mathcal{A}}$ 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 $C_0, H_q, C_0^{\text{comp}}, H_q^{\text{comp}}$, we set for j with $h(w) < j \leq r$ and $i = 0, \dots, r$,

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

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

$$\exists t\exists x\big(U_{\max} t \wedge H_{q_h}(x,t) \wedge \forall t'\forall y(t' < t \rightarrow \neg H_{q_h}(y,t'))\big). \quad (10)$$

As a consequence of the preceding lemma, we obtain:

Corollary VI.2. *Let $w \in \{0,1\}^*$ and set*

$$\pi_w := \varphi_{0w} \wedge \varphi_{1M} \wedge \gamma_M. \quad (11)$$

(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

$$\chi_w := \varphi_{0w} \wedge (\varphi_{1M} \rightarrow \neg\gamma_M). \quad (12)$$

Applying Proposition III.11 to part (b) of the preceding corollary, we get:

Lemma VI.3. *Let M on w halt in $h(w)$ steps. Then:*

(a) $\text{MOD}(\chi_w)$ *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 VI.4. *Let $f : \mathbb{N} \to \mathbb{N}$ be a computable function. Then there is a $w \in \{0,1\}^*$ such that $\text{MOD}(\chi_w)$ is closed under $<$-substructures but χ_w is not finitely equivalent to a universal sentence of length less than $f(|\chi_w|)$.*

Proof : By the previous lemma it suffices to find a $w \in \{0,1\}^*$ such that M on 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 χ_w shows that for some $c_M \in \mathbb{N}$ we have for all $w \in \{0,1\}^*$,

$$|\chi_w| \leq c_M \cdot |w|. \tag{13}$$

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

$$g(k) := f(5 \cdot c_M \cdot k). \tag{14}$$

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 $\mathit{code}(M_0, 1^k)$ such that M on $\mathit{code}(M_0, 1^k)$ simulates the computation of M_0 on 1^k.

Choose the least k such that for $w := \mathit{code}(M_0, 1^k)$,

$$|w| \leq 5k. \tag{15}$$

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 (13)–(15)

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

Finally we prove Gurevich's Theorem for graphs. For $\tau := \tau_M$ let I be an interpretation according to Lemma V.6. For $w \in \{0,1\}^*$ we consider the sentence $\rho_w := \forall \bar{x} \neg \varphi_{\text{uni}}(\bar{x}) \vee \chi_w^I$, i.e.,

$$\rho_w := \forall \bar{x} \neg \varphi_{\text{uni}}(\bar{x}) \vee (\varphi_{0w} \wedge (\varphi_{1M} \to \neg \gamma_M))^I. \tag{16}$$

That is, for $G \models \rho_w$, either the graph G interprets an "empty τ_M-structure," or a τ_M-structure which 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 VI.2(b). Hence taking in Proposition IV.6 as ψ the sentence φ_{GRAPH} axiomatizing the class of graphs, we get the following analogue of Lemma VI.3.

Lemma VI.5. *Let M on input w halt in $h(w)$ steps. Then:*

(a) $\text{GRAPH}(\rho_w)$*, the class of graphs that are model 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 $|\mu|^2 \geq h(w)$.*

Arguing as in the proof of Gurevich's Theorem we get:

Theorem VI.6 (Gurevich's Theorem for graphs)**.** *Let $f : \mathbb{N} \to \mathbb{N}$ be a computable function. Furthermore, let ρ_w be defined by* (16)*. Then there is a $w \in \{0,1\}^*$ such that $\text{GRAPH}(\rho_w)$ is closed under induced subgraphs but ρ_w is not equivalent to a universal sentence of length less than $f(|\rho_w|)$, not even in the class of finite graphs.*

Remark VI.7. Using previous remarks (Remark III.12 and Remark IV.7) one can even show that for every computable function $f : \mathbb{N} \to \mathbb{N}$ the sentence χ_w is not finitely equivalent in graphs 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

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

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

VII. 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 obtain forbidden induced subgraphs for various classes of graphs.

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

Proposition VII.1. *There is no algorithm that applied to any $\text{FO}[\tau_E]$-sentence φ decides whether the class $\text{GRAPH}(\varphi)$ is closed under induced subgraphs.*

Corollary VII.2. *There is no algorithm that applied to any $\text{FO}[\tau_E]$-sentence φ either reports that $\text{GRAPH}(\varphi)$ is not closed under induced subgraphs or it computes for $\text{GRAPH}(\varphi)$ a finite set of forbidden induced finite subgraphs.*

Proof : Otherwise we could use this algorithm as a decision algorithm for the previous result. $\square$

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. The core of the proof of the following proposition (the finite analog of Proposition VII.1) is the verification of the equivalence (see (11) for the definition of π_w)

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

Proposition VII.3. *There is no algorithm that applied to any $\text{FO}[\tau_E]$-sentence φ decides whether the class $\text{GRAPH}_{\text{fin}}(\varphi)$ is closed under induced subgraphs.*

Similarly we get the next result essentially by showing the equivalence (see (12) for the definition of χ_w)

$$M : w \mapsto \infty \iff \text{MOD}_{\text{fin}}(\chi_w) = \text{MOD}_{\text{fin}}(\varphi_{0w}).$$

Theorem VII.4. *There is no algorithm that applied to any* FO$[\tau_E]$*-sentence* φ *such that* $\textsc{Graph}_{\text{fin}}(\varphi)$ *is definable by a finite set of forbidden induced finite subgraphs computes such a set.*

This theorem is Theorem I.3 of the Introduction. Finally we prove Theorem I.2, which is equivalent to the following result.

Theorem VII.5. *There is no algorithm that applied to an* FO$[\tau_E]$*-sentence* φ *such that* $\textsc{Graph}_{\text{fin}}(\varphi)$ *is closed under induced subgraphs decides whether there is a finite set* $\mathscr{F}$ *of finite graphs such that* $\textsc{Graph}_{\text{fin}}(\varphi) = \textsc{Forb}_{\text{fin}}(\mathscr{F})$.

Proof : We prove the corresponding result for τ_M-sentences and τ_M-structures and leave it to the reader to translate it to graphs using the corresponding strongly existential interpretation. So we show:

There is no algorithm that applied to an FO$[\tau_M]$*-sentence* φ *with* $\textsc{Mod}_{\text{fin}}(\varphi)$ *closed under induced substructures decides whether there is a finite set* $\mathscr{F}$ *of finite* τ_M*-structures such that* $\textsc{Mod}_{\text{fin}}(\varphi) = \textsc{Forb}_{\text{fin}}(\mathscr{F})$.

For $w \in \{0,1\}^*$ set $\alpha_w := \varphi_{0w} \wedge (\varphi_{1M} \to \gamma_M)$. It suffices to show that $\textsc{Mod}_{\text{fin}}(\alpha_w)$ 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 VI.1(a) and the definition (10) of γ_M. Therefore, $\textsc{Mod}_{\text{fin}}(\alpha_w) = \textsc{Mod}_{\text{fin}}(\varphi_{0w} \wedge \neg\varphi_{1M})$. By Lemma VI.1(b) the sentence $\varphi_{0w} \wedge \neg\varphi_{1M}$ has arbitrarily large finite models. Hence, by Lemma III.10, we know that $\textsc{Mod}_{\text{fin}}(\varphi_{0w} \wedge \neg\varphi_{1M})$ 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 VI.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 $\mathscr{F}$ of finite τ_M-structures such that

$$\textsc{Mod}_{\text{fin}}(\alpha_w) = \textsc{Forb}_{\text{fin}}(\mathscr{F}). \tag{17}$$

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

$$\textsc{Mod}_{\text{fin}}(\varphi_{0w}) = \textsc{Forb}_{\text{fin}}(\mathscr{F}_0).$$

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

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

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

Here $\varphi^*_{1M}$is obtained from φ_{1M} by replacing the conjunct φ_1 (see (4)) 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

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

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

For $\mathscr{F} := \mathscr{F}_0 \cup \mathscr{F}_1 \cup \mathscr{F}_2$ we show (17). 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 $\mathscr{F}_0$.

Now, for a contradiction suppose that $\mathcal{B}$ is an induced substructure of $\mathcal{C}$ isomorphic to a structure in $\mathscr{F}_1$. Then $\mathcal{B} \models \varphi_{1M}$ and thus, by Lemma III.8, $\mathcal{C} = \mathcal{B}$. As $\mathcal{C} \models \alpha_w$, 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 $\mathscr{F}_2$. 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 $\mathcal{B} \models \forall t\forall t'\big(t < t' \to \forall y \neg H_{q_h}(y,t)\big)$.

As the final step let $\mathcal{C} \in \textsc{Forb}_{\text{fin}}(\mathscr{F})$. We show that $\mathcal{C} \models \alpha_w$. As $\mathcal{C}$ omits the structures in $\mathscr{F}_0$ 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 VI.1(a) (more precisely, by the extension of Lemma VI.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 $\mathscr{F}_1$ 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 $\mathscr{F}_2$, a contradiction. Hence, $|C| = h(w)+1$ and thus, $\mathcal{C} \models \alpha_w$. □

Remark VII.6. Mainly using Remark VI.7 one easily verifies that in all results but Proposition VII.3 of this section we can replace *There is no algorithm that applied to an* FO$[\tau_E]$*-sentence* φ … by *There is no algorithm that applied to a* Σ_2*-sentence* φ … In Proposition VII.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. Compton's result mentioned in Remark III.5 shows that Theorem VII.5 does not hold if we restrict ourselves to Π_2-sentences.

Acknowledgement. We thank Abhisekh Sankaran for mentioning to the first author the question of whether Tait's Theorem generalizes to graphs. 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 61872092). 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] A. Atserias, A. Dawar, and M. Grohe. Preservation under extensions on well-behaved finite structures. *SIAM Journal on Computing*, 38:1364–1381, 2008.

[2] Y. Chen and J. Flum. FO-definability of shrub-depth. In *28th EACSL Annual Conference on Computer Science Logic, CSL 2020, January 13-16, 2020, Barcelona, Spain*, pages 15:1–15:16, 2020.

[3] Y. Chen and J. Flum. Forbidden induced subgraphs and the Łoś-Tarski Theorem. *CoRR*, abs/2008.00420, 2020.

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

[5] A. Dawar and A. Sankaran. Extension preservation in the finite and prefix classes of first order logic. *CoRR*, abs/2007.05459, 2020.

[6] G. Ding. Subgraphs and well-quasi-ordering. *Journal of Graph Theory*, 16(5):489–502, 1992.

[7] R.G. Downey and M.R. Fellows. *Parameterized Complexity*. Springer, 1999.

[8] D. Duris. Extension preservation theorems on classes of acyclic finite structures. *SIAM Journal on Computing*, 39(8):3670–3681, 2010.

[9] Z. Dvorák, A. C. Giannopoulou, and D. M. Thilikos. Forbidden graphs for tree-depth. *European Journal of Combinatorics*, 33(5):969–979, 2012.

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

[11] M. R. Fellows. Private communication. 2019.

[12] M. R. Fellows and M. A. Langston. On search, decision, and the efficiency of polynomial-time algorithms. *Journal of Computer and System Sciences*, 49(3):769–779, 1994.

[13] J. Gajarský and S. Kreutzer. Computing shrub-depth decompositions. In *37th International Symposium on Theoretical Aspects of Computer Science, STACS 2020*, pages 56:1–56:17, 2020.

[14] 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*, 15(1), 2019.

[15] R. Ganian, P. Hlinený, J. Nesetril, J. Obdrzálek, P. Ossona de Mendez, and R. Ramadurai. When trees grow low: Shrubs and fast MSO_1. In *Mathematical Foundations of Computer Science 2012 - 37th International Symposium, MFCS 2012, Bratislava, Slovakia, August 27-31, 2012. Proceedings*, pages 419–430, 2012.

[16] Y. Gurevich. Toward logic tailored for computational complexity. *Lecture Notes in Mathematics*, 1104:175–216, 1984.

[17] T. A. McKee. Forbidden subgraphs in terms of forbidden quantifiers. *Notre Dame Journal of Formal Log.*, 19:186–188, 1978.

[18] J. Łoś. On the extending of models I. *Fundamenta Mathematicae*, 42:38–54, 1955.

[19] B. Rossman. Łoś-Tarski Theorem has non-recursive blow-up. *Unpublished manuscript*, pages 1–2, 2012.

[20] A. Sankaran, B. Adsul, and S. Chakraborty. A generalization of the Łoś-Tarski preservation theorem over classes of finite structures. In *Mathematical Foundations of Computer Science 2014 - MFCS 2014, Proceedings, Part I*, pages 474–485. Springer, 2014.

[21] W. W. Tait. A counterexample to a conjecture of Scott and Suppes. *The Journal of Symbolic Logic*, 24(1):15–16, 1959.

[22] A. Tarski. Contributions to the theory of models I, II. *Indagationes Mathematicae*, 16:589–588, 1954.

[23] R. Vaught. Remarks on universal classes of relational systems. *Indagationes Mathematicae*, 16:572–591, 1954.

[24] T. Zaslavsky. Forbidden induced subgraphs. *Electronic Notes in Discrete Mathematics*, 63:3–10, 2017.