A Generalization of the Łoś-Tarski Preservation Theorem over Classes of Finite Structures

Abhisekh Sankaran, Bharat Adsul, and Supratik Chakraborty

Dept. of Computer Science and Engg., IIT Bombay, India
{abhisekh,adsul,supratik}@cse.iitb.ac.in

Abstract. We present a logic-based combinatorial property of classes of finite structures that allows an effective generalization of the Łoś-Tarski preservation theorem to hold over classes satisfying the property. The well-studied classes of words and trees, and structures of bounded tree-depth are shown to satisfy the property. We also show that starting with classes satisfying this property, the classes generated by applying composition operations like disjoint union, cartesian and tensor products, inherit the property. We finally show that all classes of structures that are well-quasi-ordered under the embedding relation satisfy a natural generalization of our property.

1 Introduction

Preservation theorems in first-order logic (henceforth called FO) have been extensively studied in classical model theory [3]. An FO preservation theorem asserts that the collection of FO definable classes closed under a model-theoretic operation corresponds to the collection of classes definable by a syntactic fragment of FO. A classical preservation theorem is the Łoś-Tarski theorem, which states that over the class of all structures, the class defined by an FO sentence is preserved under substructures if, and only if, the sentence is equivalent to a universal sentence [3]. It was conjectured in [11], and subsequently proved in [10], that this theorem can be generalized using a simple yet delicate semantic notion of *preservation under substructures modulo k-sized "cruxes"* (in [11,10], 'cruxes' are called 'cores'). This notion reduces to the usual notion of preservation under substructures when k equals 0. The generalized Łoś-Tarski theorem, proved in [10], states that over the class of all structures and for all natural numbers k, the class defined by an FO sentence is preserved under substructures modulo k-sized cruxes if, and only if, the sentence is equivalent to an $\exists^k\forall^*$ sentence (i.e., a prenex sentence having quantifier prefix of the form $\exists^k\forall^*$). Since finite structures are important from a computational perspective, it is interesting to study preservation theorems over classes of finite structures. Unfortunately, most preservation theorems, including the Łoś-Tarski theorem, fail over the class of all finite structures. Earlier work [1,5,6,2] has therefore studied preservation theorems over special classes of finite structures. In this paper, we undertake a similar study for the generalized Łoś-Tarski theorem. Specfically, we identify a logic-based combinatorial property that allows the generalized Łoś-Tarski theorem to hold over any class of finite structures satisfying the property. We show that several well-studied

E. Csuhaj-Varjú et al. (Eds.): MFCS 2014, Part I, LNCS 8634, pp. 474–485, 2014.

classes satisfy this property. Furthermore, the property permits an *effective* translation of an FO sentence defining a class that is preserved under substructures modulo k-sized cruxes, to an equivalent $\exists^k\forall^*$ sentence.

In [1], Atserias, Dawar and Grohe considered classes of finite structures that are acyclic, of bounded degree (more generally, "wide") or of bounded tree-width. They showed that under suitable closure assumptions, each of these classes admits the Łoś-Tarski theorem. Subsequently, Harwath, Heimberg and Schweikardt [6] studied the bounds for an effective version of the Łoś-Tarski theorem over bounded degree structures. In [5], Duris showed that the Łoś-Tarski theorem holds for structures that are acyclic in a more general sense. Unfortunately, as discussed in Section 2, none of the above classes, in general, admits the generalized Łoś-Tarski theorem. This motivates us to ask: *Can we identify properties that allow an effective version of the generalized Łoś-Tarski theorem to hold, and are also satisfied by interesting classes of finite structures?* This paper answers this question affirmatively. Interestingly, the classes of structures studied here are incomparable to those studied in [1,6,5].

The primary results of this paper can be summarized as follows.

1. In Section 3, we present a parameterized logic-based combinatorial property of classes of finite structures, and show that this property entails an effective version of the generalized Łoś-Tarski theorem. Intuitively, if a class $\mathcal{S}$ satisfies this property for parameter k, denoted $\mathcal{P}_{logic}(\mathcal{S}, k)$, then for every natural number m, given any structure in $\mathcal{S}$ and k elements of it, there always exists an m-equivalent bounded substructure, containing these elements, that is in $\mathcal{S}$. Further, the bound is a computable function of m.
2. In Sections 4 and 5, we respectively show that the following interesting classes of structures satisfy $\mathcal{P}_{logic}(\cdot, k)$ for all k: (i) the classes of all words and trees over a finite alphabet, and (ii) any substructure-closed class of relational structures whose Gaifman graphs have bounded tree-depth.
3. In Section 6, we show that for all k, the property $\mathcal{P}_{logic}(\cdot, k)$ is preserved under natural composition operators on structures, like disjoint union, cartesian and tensor products. This allows us to construct additional classes of structures that satisfy the generalized Łoś-Tarski theorem, from known classes. Interesting examples of such constructed classes are grids of bounded dimension and various classes of co-graphs like all co-graphs, complete graphs, complete n-partite graphs for each n, threshold graphs etc. It is important to note that the classes considered in Sections 4, 5 and 6 lie beyond those studied in [1,6,5], and yet satisfy the Łoś-Tarski theorem.
4. In Section 7, we briefly discuss two other parameterized properties, denoted $\mathcal{P}_{wqo}(\cdot, k)$ and $\mathcal{P}^{gen}_{logic}(\cdot, k)$, each of which entails the generalized Łoś-Tarski theorem "(not necessarily an effective version though)". The property $\mathcal{P}_{wqo}(\cdot, k)$ is based on well-quasi-ordering of the embedding relation on structures, while $\mathcal{P}^{gen}_{logic}(\cdot, k)$ is a generalization of $\mathcal{P}_{logic}(\cdot, k)$. An interesting result is that $\mathcal{P}_{wqo}(\cdot, k)$ is subsumed by $\mathcal{P}^{gen}_{logic}(\cdot, k)$, yielding a logic-based tool to show that certain classes are not well-quasi-ordered under the embedding relation.

To prove the above results, we use a combination of techniques. For example, to show that trees and words satisfy $\mathcal{P}_{logic}(\cdot, k)$, we present a composition lemma for trees in Section 4, and use it to show that certain "prunings" of trees preserve m-equivalence. In Section 5, to prove that the class of structures with Gaifman graphs of tree-depth at most n satisfies $\mathcal{P}_{logic}(\cdot, k)$, we introduce the notion of a *twin of a structure with respect to a given element* and use inductive reasoning over n. In Section 6, the proof of closure of $\mathcal{P}_{logic}(\cdot, k)$ under natural composition operators uses a *tree representation* of structures generated by applying the operators, and uses results for trees proved earlier in Section 4. For lack of space, we defer the full proofs of our results to the journal version of the paper.

2 Notation and Preliminaries

Let $\mathbb{N}$ denote the natural numbers *including zero*. We assume that the reader is familiar with standard notation and terminology of first-order logic. We consider only finite vocabularies, represented by τ, that contain only predicate symbols of *positive* arity (and no constants or functions), unless explicitly stated otherwise. We denote by $FO(\tau)$ the set of all FO formulae over τ. A sequence $(x_1, \ldots, x_k)$ of variables is written as $\bar{x}$. We abbreviate a block of quantifiers $Qx_1 \ldots Qx_k$ by $Q^k\bar{x}$, where $Q \in \{\forall, \exists\}$. Given $k, p \in \mathbb{N}$, let $\exists^k\forall^p$ denote the set of all $FO(\tau)$ sentences in prenex normal form whose quantifier prefix has k existential quantifiers followed by p universal quantifiers. We use $\exists^k\forall^*$ to denote $\bigcup_{p\in\mathbb{N}} \exists^k\forall^p$.

Standard notions of τ-structures, substructures and extensions (see [3]) are used throughout. As in [3], by substructures, we mean *induced* substructures. Given a τ-structure $\mathfrak{A}$, we use $\mathsf{U}_{\mathfrak{A}}$ to denote the universe of $\mathfrak{A}$ and $|\mathfrak{A}|$ to denote its cardinality. Given τ-structures $\mathfrak{A}$ and $\mathfrak{B}$, we use $\mathfrak{A} \subseteq \mathfrak{B}$ to denote that $\mathfrak{A}$ is a substructure of $\mathfrak{B}$. Given a τ-structure $\mathfrak{A}$ and an $FO(\tau)$ sentence φ, if $\mathfrak{A} \models \varphi$, we say that $\mathfrak{A}$ is a *model* of φ. We focus only on recursive (or decidable) classes of finite τ-structures in this paper. All classes of τ-structures, and subclasses thereof, are also assumed to be closed under isomorphisms.

The following notion is central to our work.

Definition 1. *Let $\mathcal{S}$ be a class of τ-structures and $k \in \mathbb{N}$. A subclass $\mathcal{C}$ of $\mathcal{S}$ is said to be* preserved under substructures modulo k-sized cruxes over $\mathcal{S}$ *if every τ-structure $\mathfrak{A} \in \mathcal{C}$ has a subset* Crux *of $\mathsf{U}_{\mathfrak{A}}$ such that (i) $|\mathsf{Crux}| \leq k$, and (ii) for every $\mathfrak{B} \in \mathcal{S}$, if $\mathfrak{B} \subseteq \mathfrak{A}$ and $\mathsf{Crux} \subseteq \mathsf{U}_{\mathfrak{B}}$, then $\mathfrak{B} \in \mathcal{C}$. The set* Crux *is called a* k-crux of $\mathfrak{A}$ with respect to $\mathcal{C}$ over $\mathcal{S}$*.*

As an example, if $\mathcal{S}$ is the class of all graphs, then the subclass $\mathcal{C}$ of $\mathcal{S}$ comprising graphs containing a k-length cycle as a subgraph is preserved under substructures modulo k-sized cruxes over $\mathcal{S}$. Like Definition 1, most other definitions, discussions and results in this paper are stated with respect to an underlying class $\mathcal{S}$ of structures. When $\mathcal{S}$ is clear from the context, we omit the mention of $\mathcal{S}$. Note that Definition 1 is an adapted version of Definition 2 of [11]; the notion of 'core' in the latter is exactly the notion of 'crux' in the former when the underlying class $\mathcal{S}$ is the class of all structures. We avoid using the word 'core' for a crux to prevent confusion with existing notions of cores in the literature [2].

Given a class $\mathcal{S}$ of structures, let $PSC(k)$ denote the collection of the subclasses of $\mathcal{S}$ that are preserved under substructures modulo k-sized cruxes over $\mathcal{S}$, and that are definable over $\mathcal{S}$ by FO sentences. We interchangeably talk of $PSC(k)$ as a collection of classes and as a set of the defining FO sentences. Similarly, we interchangeably use $\exists^k\forall^p$ (and $\exists^k\forall^*$) to denote a set of FO sentences and the corresponding subclasses of $\mathcal{S}$ defined by these sentences. The generalized Łoś-Tarski theorem, proved in [10], can now be stated as follows.

Theorem 1. *Over the class of all structures, for all $k \in \mathbb{N}$, $PSC(k) = \exists^k\forall^*$.*

Note that the notion of cruxes, central to Theorem 1, differs from that of existential witnesses. If φ is an $\exists^k\forall^*$ sentence and $\mathfrak{A} \models \varphi$, then every witness of the existential variables of φ forms a k-crux of $\mathfrak{A}$. The converse, however, need not be true [11]. Specifically, let $\tau = \{E\}$, where E is a binary predicate. Consider the $FO(\tau)$ sentence $\varphi \equiv \exists x \forall y\, E(x, y)$, and the τ-structure $\mathfrak{A}$ defined by $\mathsf{U}_{\mathfrak{A}} = \{0, 1\}$ and $E^{\mathfrak{A}} = \{(0, 0), (0, 1), (1, 1)\}$. Clearly, $\mathfrak{A} \models \varphi$ and $\mathfrak{A}$ has only one witness for variable x of φ, viz. 0. Yet, both $\{0\}$ and $\{1\}$ are 1-cruxes of $\mathfrak{A}$!

Significantly, Theorem 1 fails, in general, over the classes studied in [1,6,5]. To see why this is so, let $\mathcal{S}$ be the class of graphs that are disjoint unions of undirected paths. Observe that $\mathcal{S}$ is closed under substructures and disjoint unions, is acyclic and has degree bounded by 2. Consider the subclass $\mathcal{C}$ of $\mathcal{S}$ comprising graphs containing at least 2 connected components. The subclass $\mathcal{C}$ is definable over $\mathcal{S}$ by an FO sentence ψ asserting that any model either has at least 3 end points or has at least 2 isolated vertices. Further, for any graph G in $\mathcal{C}$, any two vertices belonging to distinct components of G form a 2-crux of G; hence $\mathcal{C}$ is in $PSC(2)$. However, as shown in [11], there exists no $\exists^2\forall^*$ sentence that defines $\mathcal{C}$ over $\mathcal{S}$. Likewise, one can show that the class of all directed graphs of tree-width 1 fails to satisfy $PSC(2) = \exists^2\forall^*$. This motivates our quest for alternative properties of classes of finite structures over which Theorem 1 holds.

3 A Logic Based Combinatorial Property

We begin by recalling from standard FO terminology [7] that if m is a natural number, two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ are said to be *m-equivalent*, denoted $\mathfrak{A} \equiv_m \mathfrak{B}$, iff $\mathfrak{A}$ and $\mathfrak{B}$ agree on the truth of every $FO(\tau)$ sentence of quantifier rank at most m. We can now define a parameterized logic-based combinatorial property of classes of finite structures as follows.

Definition 2. *Let k be a natural number and $\mathcal{S}$ be a class of finite structures. We say that $\mathcal{P}_{logic}(\mathcal{S}, k)$ holds if there exists a* computable *function $\theta_k : \mathbb{N} \to \mathbb{N}$ such that for each $m \in \mathbb{N}$, for each structure $\mathfrak{A}$ of $\mathcal{S}$ and for each subset W of $\mathsf{U}_{\mathfrak{A}}$ of size at most k, there exists $\mathfrak{B} \subseteq \mathfrak{A}$ such that (i) $\mathfrak{B} \in \mathcal{S}$, (ii) $W \subseteq \mathsf{U}_{\mathfrak{B}}$, (iii) $|\mathfrak{B}| \leq \theta_k(m)$, and (iv) $\mathfrak{B} \equiv_m \mathfrak{A}$. We call θ_k a* witness function *of $\mathcal{P}_{logic}(\mathcal{S}, k)$.*

Remark: If $\mathcal{P}_{logic}(\mathcal{S}, k)$ holds and if $\mathcal{S}'$ is a subclass of $\mathcal{S}$ that is closed under substructures over $\mathcal{S}$, then it is easy to see that $\mathcal{P}_{logic}(\mathcal{S}', k)$ also holds.

We list below two simple examples of classes satisfying $\mathcal{P}_{logic}(\cdot, k)$ for every $k \in \mathbb{N}$. Various non-trivial examples are presented in Sections 4, 5 and 6.

1. Let $\mathcal{S}$ be a finite class of finite structures. Clearly, $\mathcal{P}_{logic}(\mathcal{S}, k)$ holds for all $k \in \mathbb{N}$, with $\theta_k(m)$ giving the size of the largest structure in $\mathcal{S}$.
2. Let $\mathcal{S}$ be the class of all finite linear orders. Then $\mathcal{P}_{logic}(\mathcal{S}, k)$ holds for all $k \in \mathbb{N}$, with $\theta_k(m) = \max(2^m, k)$.

The next theorem is one of the main results of this paper. Before stating the theorem, we make two observations. First, given a recursive class $\mathcal{S}$ of finite structures and a natural number n, the subclass of all structures in $\mathcal{S}$ of size at most n is definable by an effectively computable FO sentence in $\exists^n\forall^*$. We call this sentence $\xi_{\mathcal{S},n}$. Second, given a sentence ψ of $FO(\tau)$ and any sequence $\bar{x}$ of variables, one can effectively compute a quantifier-free $FO(\tau)$ formula with free variables $\bar{x}$ such that this formula evaluates to true in a τ-structure $\mathfrak{A}$ with $\bar{x}$ interpreted as $\bar{a}$ iff ψ holds in the substructure of $\mathfrak{A}$ induced by $\bar{a}$. Following notation in [11], we denote this formula as $\psi|_{\bar{x}}$, read as ψ *relativized to* $\bar{x}$.

Theorem 2. *Let $\mathcal{S}$ be a recursive class of finite structures and $k \in \mathbb{N}$ be such that $\mathcal{P}_{logic}(\mathcal{S}, k)$ holds. Then $PSC(k) = \exists^k\forall^*$ over $\mathcal{S}$, and the translation from $PSC(k)$ to $\exists^k\forall^*$ is effective. Specifically, if a witness function for $\mathcal{P}_{logic}(\mathcal{S}, k)$ is θ_k, then an FO sentence χ of quantifier rank m in $PSC(k)$ is equivalent (over $\mathcal{S}$) to the sentence $\exists^k\bar{x}\forall^p\bar{y}\ \psi|_{\bar{x}\bar{y}}$, where $p = \theta_k(m)$ and $\psi \equiv (\xi_{\mathcal{S},p} \rightarrow \chi)$.*

Proof: It is obvious that $\exists^k\forall^* \subseteq PSC(k)$ over $\mathcal{S}$. Towards the converse, consider a sentence χ, of quantifier rank m, in $PSC(k)$ over $\mathcal{S}$. Consider the sentence $\varphi \equiv \exists^k\bar{x}\forall^p\bar{y}\ \psi|_{\bar{x}\bar{y}}$, where p and ψ are as stated above. Since χ is in $PSC(k)$ over $\mathcal{S}$, every model $\mathfrak{A}$ of χ in $\mathcal{S}$ also satisfies φ. This is because the elements of any k-crux of $\mathfrak{A}$ can serve as witnesses of the existential quantifiers in φ. To show $\varphi \rightarrow \chi$ over $\mathcal{S}$, suppose $\mathfrak{A}$ is a model of φ in $\mathcal{S}$. Let W be a set of witnesses in $\mathfrak{A}$ for the k existential variables in φ. Clearly, $|W| \leq k$. Since $\mathcal{P}_{logic}(\mathcal{S}, k)$ holds, there exists $\mathfrak{B} \subseteq \mathfrak{A}$ such that (i) $\mathfrak{B} \in \mathcal{S}$, (ii) $W \subseteq \mathsf{U}_{\mathfrak{B}}$, (iii) $|\mathfrak{B}| \leq \theta_k(m) = p$, and (iii) $\mathfrak{B} \equiv_m \mathfrak{A}$. Since $\mathfrak{A} \models \varphi$, by instantiating the universal variables in φ with the elements of $\mathsf{U}_{\mathfrak{B}}$, we have $\mathfrak{B} \models \chi$. Since the quantifier rank of χ is m and $\mathfrak{B} \equiv_m \mathfrak{A}$, it follows that $\mathfrak{A} \models \chi$. Therefore, χ is equivalent to φ over $\mathcal{S}$. Finally, since m is effectively computable from χ, so are p, $\xi_{\mathcal{S},p}$ and φ. ■

4 Words and Trees over a Finite Alphabet

Given an alphabet Σ, let $Words(\Sigma)$ and $Trees(\Sigma)$ denote the set of all finite words and trees, respectively, over Σ. The key result of this section is as follows.

Theorem 3. *For every finite alphabet Σ, both $\mathcal{P}_{logic}(Words(\Sigma), k)$ and $\mathcal{P}_{logic}(Trees(\Sigma), k)$ hold for every natural number k.*

For purposes of our discussion, we use a poset-theoretic representation of trees. A tree is a finite poset $P = (A, \leq)$ with a unique minimal element (called "root"), and for every $a, b, c \in A$, $\big((a \leq c)\ \wedge\ (b \leq c)\big) \rightarrow \big(a \leq b\ \vee\ b \leq a\big)$. Informally, the Hasse diagram of P is an (inverted) tree with every parent p connected to its child c. A tree over Σ, henceforth called a Σ-tree, is a pair (P, λ) where

$P = (A, \leq)$ is a tree and $\lambda : A \to \Sigma$ is a labeling function. The elements of A are also called nodes (or elements) of the Σ-tree. In the special case where the underlying poset is a linear order, a Σ-tree is called a *Σ-word*. We denote trees by either $\mathfrak{s}$ or $\mathfrak{t}$. A *Σ-forest* $\mathfrak{f}$ is a (finite) disjoint union of Σ-trees.

Let τ be the vocabulary $\{\leq\} \cup \{Q_a \mid a \in \Sigma\}$, where $\leq$ is a binary predicate and each Q_a is a unary predicate. A Σ-tree $\mathfrak{s} = ((A_\mathfrak{s}, \leq_\mathfrak{s}), \lambda_\mathfrak{s})$ has a natural representation as a structure $\mathfrak{A}_\mathfrak{s}$ over τ. To represent a Σ-forest $\mathfrak{f}$ as a τ-structure $\mathfrak{A}_\mathfrak{f}$, we use the disjoint union (denoted $\sqcup$) of the τ-structures representing the Σ-trees in $\mathfrak{f}$. For clarity of exposition, we use $\mathfrak{s}$ (resp. $\mathfrak{f}$) to denote both a Σ-tree $\mathfrak{s}$ (resp. Σ-forest $\mathfrak{f}$) and its corresponding τ-structure $\mathfrak{A}_\mathfrak{s}$ (resp. $\mathfrak{A}_\mathfrak{f}$).

We use the standard notions of height, degree and subtree of a given tree. Given two Σ-trees $\mathfrak{s} = ((A_\mathfrak{s}, \leq_\mathfrak{s}), \lambda_\mathfrak{s})$ and $\mathfrak{t} = ((A_\mathfrak{t}, \leq_\mathfrak{t}), \lambda_\mathfrak{t})$ with disjoint sets of nodes, and an element e of $\mathfrak{s}$, the *join of* $\mathfrak{t}$ *to* $\mathfrak{s}$ *at* e, denoted $\mathfrak{s} \cdot_e \mathfrak{t}$, is the Σ-tree obtained from $\mathfrak{s}$ by adding $\mathfrak{t}$ as a new "child subtree" of the element e of $\mathfrak{s}$. Given a Σ-tree $\mathfrak{s}$, a Σ-forest $\mathfrak{f} = \bigsqcup_{i=1}^{n} \mathfrak{t}_i$ and an element e of $\mathfrak{s}$, the *join of* $\mathfrak{f}$ *to* $\mathfrak{s}$ *at* e, denoted $\mathfrak{s} \cdot_e \mathfrak{f}$, is the Σ-tree $((((\mathfrak{s} \cdot_e \mathfrak{t}_1) \cdot_e \mathfrak{t}_2) \cdots) \cdot_e \mathfrak{t}_n)$.

The proof of Theorem 3 uses two key auxiliary lemmas. The first is a *composition* lemma for trees. This lemma intuitively states that if $\mathfrak{t}$ is a tree, a is a node of $\mathfrak{t}$, and $\mathfrak{f}$ is a forest, then the $\equiv_m$ class of $\mathfrak{t} \cdot_a \mathfrak{f}$ is completely determined by the $\equiv_m$ classes of $(\mathfrak{t}, a)$ and $\mathfrak{f}$. Composition results of this kind were first studied by Feferman and Vaught, and subsequently by others (see [8] for a survey).

Lemma 1. *Let $\mathfrak{t}_i$ be a non-empty Σ-tree containing element a_i, and $\mathfrak{f}_i$ be a non-empty Σ-forest containing element b_i, for $i \in \{1, 2\}$. Let $\mathfrak{s}_i = \mathfrak{t}_i \cdot_{a_i} \mathfrak{f}_i$ for $i \in \{1, 2\}$. Suppose $(\mathfrak{t}_1, a_1) \equiv_m (\mathfrak{t}_2, a_2)$. Then the following hold.*

1. *If $(\mathfrak{f}_1, b_1) \equiv_m (\mathfrak{f}_2, b_2)$, then $(\mathfrak{s}_1, a_1, b_1) \equiv_m (\mathfrak{s}_2, a_2, b_2)$.*
2. *If $\mathfrak{f}_1 \equiv_m \mathfrak{f}_2$, then $(\mathfrak{s}_1, a_1) \equiv_m (\mathfrak{s}_2, a_2)$.*

The proof of Lemma 1 uses the Ehrenfeucht-Fraïssé theorem [7] and is similar to the proof of the composition lemma for words. Before stating the next auxiliary lemma, we introduce some notation. Given an alphabet Σ and a natural number m, let $\Delta(m, \Sigma)$ denote the set of all equivalence classes of the $\equiv_m$ relation over $Trees(\Sigma)$. Let Alph denote the set of all finite alphabets, and $g : \mathbb{N} \times \mathsf{Alph} \to \mathbb{N}$ be a computable function such that $g(m, \Sigma) \geq |\Delta(m, \Sigma)|$. It is known that g exists (see proof of Lemma 3.13 in [7]). We now state our next auxiliary lemma.

Lemma 2. *Let $\mathfrak{s}$ be a Σ-tree. For every $m \in \mathbb{N}$, each of the following exist.*

(a) A subtree $\mathfrak{t}_1$ of $\mathfrak{s}$ such that $\mathfrak{t}_1$ has degree $\leq m \cdot g(m, \Sigma)$ and $\mathfrak{t}_1 \equiv_m \mathfrak{s}$.
(b) A subtree $\mathfrak{t}_2$ of $\mathfrak{s}$ such that $\mathfrak{t}_2$ has height $\leq g(m, \Sigma)$ and $\mathfrak{t}_2 \equiv_m \mathfrak{s}$.

Proof Sketch: (a) Let d denote $m \cdot g(m, \Sigma)$. If each node of $\mathfrak{s}$ has at most d children, then taking $\mathfrak{t}_1$ to be $\mathfrak{s}$, we are done. Else, let a be a node of $\mathfrak{s}$, having $> d$ children. Let $\Gamma(a)$ denote the set of all subtrees of $\mathfrak{s}$ rooted at the children of a in $\mathfrak{s}$, and let $\mathfrak{f}$ be the forest whose trees are exactly the members of $\Gamma(a)$. Let $\mathfrak{t}$ be the tree such that $\mathfrak{s} = \mathfrak{t} \cdot_a \mathfrak{f}$. For every $\delta \in \Delta(m, \Sigma)$, let $\Gamma(a, \delta)$ be the set consisting of the members of $\Gamma(a)$ whose $\equiv_m$ class is δ. Construct the forest $\mathfrak{f}_1$

by performing the following operation on $\mathfrak{f}$ for each $\delta \in \Delta(m, \Sigma)$: retain $\Gamma(a, \delta)$ entirely in $\mathfrak{f}$ if $|\Gamma(a, \delta)| < m$, else retain exactly (any) m members of $\Gamma(a, \delta)$ in $\mathfrak{f}$ and remove the rest. It is easy to see that $\mathfrak{f} \equiv_m \mathfrak{f}_1$. Let $\mathfrak{s}_1 = \mathfrak{t} \cdot_a \mathfrak{f}_1$. Then using Lemma 1, we get that $\mathfrak{s}_1 \equiv_m \mathfrak{s}$. Observe that $\mathfrak{s}_1$ has strictly fewer nodes that have $> d$ children, compared to $\mathfrak{s}$. Recursing on $\mathfrak{s}_1$, we are eventually done.

(b) Let A be the underlying set of $\mathfrak{s}$. Define the function $h : A \to \Delta(m, \Sigma)$ such that $h(a)$ is the $\equiv_m$ class of the subtree of $\mathfrak{s}$ rooted at a, for every $a \in A$. If for each path in $\mathfrak{s}$, no two distinct elements on the path have the same h value, then the height of $\mathfrak{s}$ is at most $g(m, \Sigma)$. Then the desired subtree $\mathfrak{t}_2$ can be chosen to be $\mathfrak{s}$ itself. Otherwise, there exist distinct $a, b \in A$ such that (i) $\mathfrak{s} \models (a \leq b)$ and (ii) $h(a) = h(b)$. Let $\mathfrak{s}_1$ be the subtree of $\mathfrak{s}$ obtained by 'replacing' the subtree rooted at a with the subtree rooted at b. By Lemma 1, we get $\mathfrak{s}_1 \equiv_m \mathfrak{s}$. Also, $\mathfrak{s}_1$ has strictly fewer nodes than $\mathfrak{s}$. Recursing on $\mathfrak{s}_1$, we are eventually done. ∎

The proof of Theorem 3 for $Trees(\Sigma)$ is now completed as follows. Given $m \in \mathbb{N}$, a Σ-tree $\mathfrak{s} = (P, \lambda)$ and a set W of at most k elements of $\mathfrak{s}$, let $\mathfrak{s}' = (P, \lambda')$ be the tree over $\Sigma' = \Sigma \times \{0, 1\}$ such that for every $a \in P$, $\lambda'(a) = (\lambda(a), 1)$ if $a \in W$, and $\lambda'(a) = (\lambda(a), 0)$ otherwise. Let $n = \max(m, k)$. By Lemma 2, there exists a subtree $\mathfrak{t}'$ of $\mathfrak{s}'$ with degree at most $n \cdot g(n, \Sigma')$ and height at most $g(n, \Sigma')$, that is n-equivalent to $\mathfrak{s}'$. It is easy to check that by dropping the second component of the labels of all elements of $\mathfrak{t}'$, we get a subtree $\mathfrak{t}$ of $\mathfrak{s}$ containing W. In addition, $\mathfrak{t} \equiv_m \mathfrak{s}$ and $|\mathfrak{t}| \leq \theta_k(m)$, where $\theta_k(m) = (n \cdot g(n, \Sigma') + 1)^{g(n, \Sigma')+1}$ and $n = \max(m, k)$. Then $\mathcal{P}_{logic}(Trees(\Sigma), k)$ holds with θ_k as the witness function. Whence, $\mathcal{P}_{logic}(Words(\Sigma), k)$ holds by the remark following Definition 2. ∎

5 Structures of Bounded Tree-Depth

Nešetřil and de Mendez introduced the notion of *tree-depth* of an undirected graph in [9]. Intuitively, the tree-depth of a graph G, denoted $td(G)$, is a measure of how far G is from being a star. The following is an inductive definition of tree-depth, given by Lemma 2.2 of [9]. In this definition, $G = (V, E)$ denotes a graph, $\mathsf{Comp}(G)$ denote the set of all connected components of G, and $G \setminus v$ denotes the graph obtained by removing the vertex v from G.

$$td(G) = \begin{cases} 1 & \text{if } G \text{ has a single vertex} \\ 1 + \min_{v \in V} td(G \setminus v) & \text{if } G \text{ is connected and has multiple vertices} \\ \max_{G' \in \mathsf{Comp}(G)} td(G') & \text{if } G \text{ is disconnected} \end{cases}$$

The *Gaifman graph* $\mathcal{G}(\mathfrak{A})$ of a relational structure $\mathfrak{A}$ is an undirected graph whose nodes are the elements of $\mathfrak{A}$, and in which two nodes are adjacent if, and only if, they appear together in some tuple of some relation of $\mathfrak{A}$ [7]. We say that a structure $\mathfrak{A}$ is *connected* if $\mathcal{G}(\mathfrak{A})$ is connected, else we say $\mathfrak{A}$ is *disconnected.* A substructure $\mathfrak{B}$ of $\mathfrak{A}$ is said to be a *connected component* of $\mathfrak{A}$ if $\mathcal{G}(\mathfrak{B})$ is a connected component of $\mathcal{G}(\mathfrak{A})$. We say that $\mathfrak{A}$ has tree-depth n if $\mathcal{G}(\mathfrak{A})$ has tree-depth n. We say that a class $\mathcal{S}$ of structures has *bounded tree-depth* if there exists a natural number n such that all structures in $\mathcal{S}$ have tree-depth at most n. The main result of this section can now be stated as follows.

Theorem 4. *Let $\mathcal{S}$ be a substructure-closed class of finite structures, of bounded tree-depth. Then $\mathcal{P}_{logic}(\mathcal{S}, k)$ holds for every natural number k.*

In this section, we allow the vocabulary τ to contain 0-ary predicate symbols. To prove Theorem 4, we introduce the notion of *twin-structures*. Given a vocabulary τ and a predicate R in τ, let $\#R$ denote the arity of R. If $\#R > 0$, then for each subset T of $\{1, \ldots \#R\}$, let R_T denote a predicate of arity $\#R - |T|$. Define $\widehat{\tau}$ to be the vocabulary $\{R \mid R \in \tau, \#R = 0\} \cup \{R_T \mid R \in \tau, \#R > 0, T \subseteq \{1, \ldots, \#R\}\}$. Given a τ-structure $\mathfrak{A}$ and an element a of it, let $\mathfrak{A} \setminus a$ denote the substructure of $\mathfrak{A}$ induced by $\mathsf{U}_{\mathfrak{A}} \setminus \{a\}$. Given a predicate R in τ, a subset T of $\{1, \ldots, \#R\}$ and a $(\#R - |T|)$-tuple $\bar{b}$ from $\mathfrak{A} \setminus a$, let $\mathsf{ex}_a(\bar{b}, T)$ denote the expansion of $\bar{b}$ with a at the positions in T. Formally, $\mathsf{ex}_a(\bar{b}, T)$ is the $\#R$-tuple whose i^{th} component is a for each $i \in T$, and whose sub-tuple obtained by dropping all the a's, is exactly $\bar{b}$. Then the *twin-structure of* $\mathfrak{A}$ *with respect to* a, denoted $\mathsf{twin}(\mathfrak{A}, a)$, is a $\widehat{\tau}$-structure defined as: (i) The universe of $\mathsf{twin}(\mathfrak{A}, a)$ is $\mathsf{U}_{\mathfrak{A}} \setminus \{a\}$ (ii) For every 0-ary predicate R in τ, we have $\mathsf{twin}(\mathfrak{A}, a) \models R$ iff $\mathfrak{A} \models R$ (iii) For every predicate R_T in $\widehat{\tau}$ and for every $(\#R - |T|)$-tuple $\bar{b}$ of elements of $\mathsf{twin}(\mathfrak{A}, a)$, we have $\mathsf{twin}(\mathfrak{A}, a) \models R_T(\bar{b})$ iff $\mathfrak{A} \models R(\mathsf{ex}_a(\bar{b}, T))$. Observe that $\mathfrak{A}$ and $\mathsf{twin}(\mathfrak{A}, a)$ uniquely identify each other, upto isomorphism. The following lemma is easy.

Lemma 3. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be given structures and let a and b be elements of $\mathfrak{A}$ and $\mathfrak{B}$ respectively. Then given $m \in \mathbb{N}$, if $\mathsf{twin}(\mathfrak{A}, a) \equiv_m \mathsf{twin}(\mathfrak{B}, b)$, then $\mathfrak{A} \equiv_m \mathfrak{B}$.*

Let $\Delta(m, \tau)$ be the set of equivalence classes of the $\equiv_m$ relation over the class of all τ-structures. Let Vocab be the set of all finite vocabularies. Consider a computable function $g_1 : \mathbb{N} \times \mathsf{Vocab} \to \mathbb{N}$ such that $g_1(m, \tau) \geq |\Delta(m, \tau)|$. It is known that g_1 exists (see proof of Lemma 3.13 in [7]). Define the function $f_1 : \mathbb{N} \times \mathbb{N} \times \mathsf{Vocab} \to \mathbb{N}$ as $f_1(n, m, \tau) = 1 + m \cdot g_1(m, \widehat{\tau}) \cdot f_1(n-1, m, \widehat{\tau})$, where $f_1(1, m, \tau) = 1$. We now have the following lemma.

Lemma 4. *Given $m \in \mathbb{N}$ and a τ-structure $\mathfrak{A}$ that is connected and has tree-depth n, there exists $\mathfrak{B} \subseteq \mathfrak{A}$ such that (i) $\mathfrak{B} \equiv_m \mathfrak{A}$ and (ii) $|\mathfrak{B}| \leq f_1(n, m, \tau)$.*

Proof Sketch: The proof is by induction on n. The base case is trivial. Assume that the result holds for all tree-depths from 1 to $n-1$ and for all vocabularies τ. Let $\mathfrak{A}$ be as given. Since $\mathfrak{A}$ has tree-depth n, by definition, there exists $a \in \mathfrak{A}$ such that $\mathcal{G}(\mathfrak{A}) \setminus a$ has tree-depth at most $n-1$. Consider $\mathsf{twin}(\mathfrak{A}, a)$. It is easy to show that for any $\widehat{\tau}$-structure $\widehat{\mathfrak{D}}$, if $\widehat{\mathfrak{D}} \subseteq \mathsf{twin}(\mathfrak{A}, a)$, then $\widehat{\mathfrak{D}} = \mathsf{twin}(\mathfrak{C}, a)$ for some $\mathfrak{C} \subseteq \mathfrak{A}$, containing a. Then by a "degree reduction" argument similar to the one in the proof of Lemma 2(a), there exists $\mathfrak{A}' \subseteq \mathfrak{A}$, containing a such that (i) each connected component of $\mathsf{twin}(\mathfrak{A}', a)$ is a connected component of $\mathsf{twin}(\mathfrak{A}, a)$ (ii) the set Y of the connected components of $\mathsf{twin}(\mathfrak{A}', a)$ has size $\leq m \cdot g_1(m, \widehat{\tau})$, and (iii) $\mathsf{twin}(\mathfrak{A}', a) \equiv_m \mathsf{twin}(\mathfrak{A}, a)$. From Lemma 3, it follows that $\mathfrak{A}' \equiv_m \mathfrak{A}$.

Now for $\widehat{\mathfrak{D}} \in Y$, let $\mathfrak{C} \subseteq \mathfrak{A}'$ be such that $\mathfrak{C}$ contains a and $\widehat{\mathfrak{D}} = \mathsf{twin}(\mathfrak{C}, a)$. Observing that $\mathcal{G}(\mathfrak{C}) \setminus a$ and $\mathcal{G}(\mathsf{twin}(\mathfrak{C}, a))$ are the same graph, it follows that the tree-depth of $\widehat{\mathfrak{D}}$ is at most $n-1$. Applying the induction hypothesis on $\widehat{\mathfrak{D}}$, there exists $\widehat{\mathfrak{D}}_1 \subseteq \widehat{\mathfrak{D}}$ such that (i) $\widehat{\mathfrak{D}}_1 \equiv_m \widehat{\mathfrak{D}}$ and (ii) $|\widehat{\mathfrak{D}}_1| \leq f_1(n-1, m, \widehat{\tau})$.

If $\widehat{\mathfrak{D}}_2 = \bigsqcup_{\widehat{\mathfrak{D}} \in Y} \widehat{\mathfrak{D}}_1$, then $\widehat{\mathfrak{D}}_2 \equiv_m \mathsf{twin}(\mathfrak{A}', a)$. Since $\widehat{\mathfrak{D}}_2 \subseteq \mathsf{twin}(\mathfrak{A}', a)$, there exists $\mathfrak{A}'' \subseteq \mathfrak{A}'$ containing a such that $\widehat{\mathfrak{D}}_2 = \mathsf{twin}(\mathfrak{A}'', a)$. Invoking Lemma 3 on $\mathsf{twin}(\mathfrak{A}'', a)$ and $\mathsf{twin}(\mathfrak{A}', a)$, it follows that $\mathfrak{A}'' \equiv_m \mathfrak{A}'$. Then $\mathfrak{A}'' \subseteq \mathfrak{A}$ and $\mathfrak{A}'' \equiv_m \mathfrak{A}$. It is easy to see that $|\mathfrak{A}''| \leq f_1(n, m, \tau)$. Taking $\mathfrak{B}$ as $\mathfrak{A}''$, the proof is completed. ■

Proof of Theorem 4 : Let $\mathcal{S}$ be a substructure-closed class of τ-structures of tree-depth at most n. For notational convenience, let $f_2(n, m, \tau)$ denote $m \cdot g_1(m, \tau) \cdot f_1(n, m, \tau)$, for all $m, n \in \mathbb{N}$ and $\tau \in \mathsf{Vocab}$. We show below that $\mathcal{P}_{logic}(\mathcal{S}, k)$ holds for each $k \in \mathbb{N}$, with the witness function given by $\theta_k(m) = f_2(n, m, \tau)$ if $k = 0$, and by $\theta_k(m) = f_2(n, r, \nu)$ otherwise, where $r = \max(m, k)$ and $\nu = \tau \cup \{P\}$ for a unary predicate P not in τ. The result for $k = 0$ follows from Lemma 4 and from the fact that upto m-equivalence, any τ-structure has at most $m \cdot g_1(m, \tau)$ connected components. Suppose we are given $\mathfrak{A} \in \mathcal{S}$ and a subset W of at most k elements of $\mathfrak{A}$, for $k > 0$. Then consider the ν-structure $\mathfrak{A}'$ whose τ-reduct is $\mathfrak{A}$ and in which P is interpreted to be exactly W. By the result for $k = 0$, there exists $\mathfrak{B}' \subseteq \mathfrak{A}'$ such that (i) $\mathfrak{A}' \equiv_r \mathfrak{B}'$ and (ii) $|\mathfrak{B}'| \leq f_2(n, r, \nu)$. It is clear that the τ-reduct of $\mathfrak{B}'$, say $\mathfrak{B}$, is such that (i) $\mathfrak{B} \subseteq \mathfrak{A}$ (hence $\mathfrak{B} \in \mathcal{S}$) (ii) $W \subseteq \mathsf{U}_{\mathfrak{B}}$ (iii) $\mathfrak{A} \equiv_m \mathfrak{B}$ and (iv) $|\mathfrak{B}| \leq f_2(n, r, \nu) = \theta_k(m)$. Finally, since $g_1(\cdot, \cdot)$, $f_1(\cdot, \cdot, \cdot)$ and $f_2(\cdot, \cdot, \cdot)$ are easily seen to be computable, we are done. ■

Remark: The classes of bounded tree-depth considered in this section were not studied earlier in [1]. While these classes in general are not acyclic, nor of bounded degree (more generally, not wide too), they are certainly of bounded tree-width [9]. However, [1] talks only about the class of *all* structures of tree-width n for each $n \in \mathbb{N}$, and not about any subclasses of it.

6 Generating New Classes of Structures

We consider some natural ways of generating new classes of structures from a base class $\mathcal{S}$ of structures. The primary result of this section is that classes generated by these techniques inherit the $\mathcal{P}_{logic}(\cdot, k)$ property of the base class.

We focus on disjoint union ($\sqcup$), complement ($!$), cartesian product ($\times$) and tensor product ($\otimes$) of τ-structures coming from a base class $\mathcal{S}$. The definition of $\sqcup$ is standard. The definitions of $!$, $\times$ and $\otimes$ below are inspired by their definitions for graphs. Let $\mathfrak{A}$ and $\mathfrak{B}$ be τ-structures. The *complement* of $\mathfrak{A}$, denoted $!\mathfrak{A}$, is the τ-structure such that (i) $\mathsf{U}_{!\mathfrak{A}} = \mathsf{U}_{\mathfrak{A}}$, and (ii) for every n-ary predicate R in τ, for every n-tuple $(a_1, \ldots a_n) \in \mathsf{U}_{\mathfrak{A}}^n$, $!\mathfrak{A} \models R(a_1, \ldots a_n)$ iff $\mathfrak{A} \not\models R(a_1, \ldots a_n)$. The *cartesian product* of $\mathfrak{A}$ and $\mathfrak{B}$, denoted $\mathfrak{A} \times \mathfrak{B}$, is the structure $\mathfrak{C}$ defined as follows: (i) $\mathsf{U}_{\mathfrak{C}} = \mathsf{U}_{\mathfrak{A}} \times \mathsf{U}_{\mathfrak{B}}$, and (ii) for each n-ary predicate R in τ, for each n-tuple $\big((a_1, b_1), \ldots, (a_n, b_n)\big)$ of $\mathsf{U}_{\mathfrak{C}}$, we have $\mathfrak{C} \models R\big((a_1, b_1), \ldots, (a_n, b_n)\big)$ iff $\big((a_1 = \cdots = a_n \wedge \mathfrak{B} \models R(b_1, \ldots, b_n)) \bigvee (\mathfrak{A} \models R(a_1, \ldots, a_n) \wedge b_1 = \cdots = b_n)\big)$. The *tensor product* of $\mathfrak{A}$ and $\mathfrak{B}$, denoted $\mathfrak{A} \otimes \mathfrak{B}$, is defined similar to the cartesian product, except that $\mathfrak{A} \otimes \mathfrak{B} \models R\big((a_1, b_1), \ldots, (a_n, b_n)\big)$ iff $\mathfrak{A} \models R(a_1, \ldots, a_n)$ and $\mathfrak{B} \models R(b_1, \ldots, b_n)$.

Let Op be the set $\{\sqcup, !, \times, \otimes\}$. The following properties of operations in Op are important for our purposes. Let $\circledast$ be a binary operation in Op and $m \in \mathbb{N}$.

P1) If $\mathfrak{A}_1 \subseteq \mathfrak{B}_1$ and $\mathfrak{A}_2 \subseteq \mathfrak{B}_2$, then (i) $!\mathfrak{A}_1 \subseteq !\mathfrak{B}_1$ and (ii) $(\mathfrak{A}_1 \circledast \mathfrak{A}_2) \subseteq (\mathfrak{B}_1 \circledast \mathfrak{B}_2)$.
P2) If $\mathfrak{A}_1 \equiv_m \mathfrak{B}_1$ and $\mathfrak{A}_2 \equiv_m \mathfrak{B}_2$, then (i) $!\mathfrak{A}_1 \equiv_m !\mathfrak{B}_1$ and (ii) $(\mathfrak{A}_1 \circledast \mathfrak{A}_2) \equiv_m (\mathfrak{B}_1 \circledast \mathfrak{B}_2)$.

Given a class $\mathcal{S}$, let $!\mathcal{S}$ denote the class $\{!\mathfrak{A} \mid \mathfrak{A} \in \mathcal{S}\}$. Given classes $\mathcal{S}_1$ and $\mathcal{S}_2$ and a binary operation $\circledast \in \mathsf{Op}$, let $\mathcal{S}_1 \circledast \mathcal{S}_2$ denote the class $\{\mathfrak{A} \circledast \mathfrak{B} \mid \mathfrak{A} \in \mathcal{S}_1, \mathfrak{B} \in \mathcal{S}_2\}$. We now have the following important lemma.

Lemma 5. *Let $\mathcal{S}_1, \mathcal{S}_2$ be classes of structures. Let $\circledast$ be a binary operation in* Op *and $k \in \mathbb{N}$. If $\mathcal{P}_{logic}(\mathcal{S}_i, k)$ holds for $i \in \{1, 2\}$, then each of $\mathcal{P}_{logic}(!\mathcal{S}_1, k)$, $\mathcal{P}_{logic}(!\mathcal{S}_2, k)$ and $\mathcal{P}_{logic}(\mathcal{S}_1 \circledast \mathcal{S}_2, k)$ holds. In addition, $\mathcal{P}_{logic}(\mathcal{S}_1 \cup \mathcal{S}_2, k)$ holds.*

Given a class $\mathcal{S}$ of structures satisfying $\mathcal{P}_{logic}(\cdot, k)$, it follows that any class $\mathcal{S}'$ of structures obtained by finitely many applications of the operations in Op and by taking finite unions of the classes obtained thereof, also satisfies $\mathcal{P}_{logic}(\cdot, k)$. However, there are interesting classes that can be generated only by allowing *infinite* unions of these derived classes. For example, the class of all co-graphs is generated from the class of single vertex graphs by finitely many applications of $\sqcup$ and !, and then taking the infinite union of all the classes of graphs thus obtained. The rest of this section is motivated by such infinite unions.

Given a class $\mathcal{S}$ of structures and $O \subseteq \mathsf{Op}$, an *expression tree over* $(\mathcal{S}, O)$ is a tree over O whose leaf nodes are labelled with specific structures from $\mathcal{S}$ and internal nodes are labelled with operations from O (i.e. elements of O). [1] If $\mathfrak{s}$ is an expression tree over $(\mathcal{S}, O)$, let $\mathfrak{C}_{\mathfrak{s}}$ denote the structure naturally represented by $\mathfrak{s}$ upto isomorphism. We denote by $Z_{\mathcal{S},O}$ the class of all structures defined by all possible expression trees over $(\mathcal{S}, O)$.

Theorem 5. *Let $\mathcal{S}$ be a given class of structures and let $O = \{\sqcup, !\}$. For each $k \in \mathbb{N}$, if $\mathcal{P}_{logic}(\mathcal{S}, k)$ holds, then so does $\mathcal{P}_{logic}(Z_{\mathcal{S},O}, k)$.*

Proof Sketch: Consider $\mathfrak{A} \in Z_{\mathcal{S},O}$ and $m \in \mathbb{N}$. Let $W \subseteq \mathsf{U}_{\mathfrak{A}}$ be a set of size at most k. Let $\mathfrak{s}$ be an expression tree of $\mathfrak{A}$, i.e. $\mathfrak{C}_{\mathfrak{s}} = \mathfrak{A}$. The proof is in two parts:

(I) We first construct a bounded sized sub-expression-tree $\mathfrak{t}$ of $\mathfrak{s}$ such that (i) W is contained in the leaves of $\mathfrak{t}$, (ii) $\mathfrak{C}_{\mathfrak{t}} \subseteq \mathfrak{A}$ and (ii) $\mathfrak{C}_{\mathfrak{t}} \equiv_m \mathfrak{A}$. To do this, we label each node a of $\mathfrak{s}$ by the pair (δ, i), where δ is the $\equiv_m$ class of $\mathfrak{C}_{\mathfrak{s}_a}$, $\mathfrak{s}_a$ is the subtree of $\mathfrak{s}$ rooted at a, and i is the number of leaves of $\mathfrak{s}_a$ that contain any element of W. We then do a "height reduction" as in the proof of Lemma 2(b) to get $\mathfrak{t}$.
(II) We create a tree $\mathfrak{t}_1$ from $\mathfrak{t}$ by replacing the leaves of $\mathfrak{t}$ with m-equivalent bounded substructures ensuring that $\mathfrak{C}_{\mathfrak{t}_1}$ contains W (using the $\mathcal{P}_{logic}(\mathcal{S}, k)$ assumption). By a hierarchical compositional reasoning (using properties P1 and P2), we show that $\mathfrak{C}_{\mathfrak{t}_1}$ is the 'right' substructure of $\mathfrak{C}_{\mathfrak{s}}$. ■

We list below examples of classes of structures satisfying $\mathcal{P}_{logic}(\cdot, k)$ that can be generated by applying the above results to simple classes of structures. In all these examples, we assume a finite set of colours.

[1] We think of a tree in the poset-theoretic sense, as in Section 4. The number of children of any internal node is equal to the arity of the operation labeling the node.

1. The class of all coloured co-graphs, obtained using expression trees with $\sqcup$ and ! as operations at the internal nodes and coloured isolated nodes as leaves. By the remark following Definition 2, any class of coloured co-graphs closed under induced subgraphs is also an example. Special cases include the classes of coloured complete graphs, coloured complete n-partite graphs for any $n \in \mathbb{N}$ and coloured threshold graphs [4].
2. The class of r-dimensional grid posets for every $r \in \mathbb{N}$, where an r-dimensional grid poset is a tensor product of r linear orders.

Using ideas similar to those in the proof of Theorem 5, it can be shown that for $k = 0$ or 1, if $\mathcal{P}_{logic}(\mathcal{S}, k)$ holds, then so does $\mathcal{P}_{logic}(Z_{\mathcal{S},\mathsf{Op}}, k)$, where Op is as defined earlier. An interesting corollary of this result is that the class of all finite dimensional grid posets satisfies $\mathcal{P}_{logic}(\cdot, k)$ when $k = 0$ or 1.

7 Related Properties: $\mathcal{P}_{wqo}(\mathcal{S}, k)$ and $\mathcal{P}^{gen}_{logic}(\mathcal{S}, k)$

In this section, we investigate some other natural properties which are also sufficient to guarantee a generalization of the Łoś-Tarski theorem. Towards this, we first introduce the property $P^{gen}_{logic}(\mathcal{S}, k)$.

Definition 3. *Let $P^{gen}_{logic}(\mathcal{S}, k)$ be the property obtained by dropping the computability restriction of the witness function θ_k in the definition of $P_{logic}(\mathcal{S}, k)$.*

Clearly, $P_{logic}(\mathcal{S}, k)$ implies $P^{gen}_{logic}(\mathcal{S}, k)$. It is also clear from the proof of Theorem 2 that $P^{gen}_{logic}(\mathcal{S}, k)$ entails $PSC(k) = \exists^k\forall^*$, though the former need not entail an effective form of the latter. Also, it turns out that the converse of this entailment is not true in general; the class $\mathcal{S}$ of all undirected cycles satisfies $PSC(k) = \exists^k\forall^*$ for all k, but fails to satisfy $\mathcal{P}^{gen}_{logic}(\mathcal{S}, k)$ for any k.

We now turn our attention to another seemingly unrelated property. We begin with some notation. Given a vocabulary τ and $k \in \mathbb{N}$, let τ_k denote the vocabulary obtained by adding k new constant symbols to τ. Let $\mathcal{S}$ be a class of structures. We use $\mathcal{S}_k$ to denote the class of all τ_k-structures whose τ-reducts are structures in $\mathcal{S}$. Given $\mathfrak{A}, \mathfrak{B} \in \mathcal{S}_k$, we say that $\mathfrak{A}$ *embeds* in $\mathfrak{B}$ if $\mathfrak{A}$ is isomorphic to a substructure of $\mathfrak{B}$. Notationally, we represent this as $\mathfrak{A} \hookrightarrow \mathfrak{B}$. Observe that $(\mathcal{S}_k, \hookrightarrow)$ is a pre-order. We now define the property $\mathcal{P}_{wqo}(\mathcal{S}, k)$ via the notion of a pre-order being a *well-quasi-order* (w.q.o.) [4].

Definition 4. *We say that $\mathcal{P}_{wqo}(\mathcal{S}, k)$ holds if $(\mathcal{S}_k, \hookrightarrow)$ is a well-quasi-order.*

Basic examples of classes satisfying $\mathcal{P}_{wqo}(\mathcal{S}, k)$ are (i) a finite class, and (ii) the class of all finite linear orders. Let Σ be a finite alphabet. In our notation, the celebrated results such as Higman's lemma and Kruskal's tree theorem [4] simply say that $\mathcal{P}_{wqo}(Words(\Sigma), 0)$ and $\mathcal{P}_{wqo}(Trees(\Sigma), 0)$ respectively hold. *A priori*, there is no reason to expect any relation between the w.q.o.-based and the logic-based properties defined above. Surprisingly, we have the following result.

Theorem 6. *For any class $\mathcal{S}$ and any $k \in \mathbb{N}$, $\mathcal{P}_{wqo}(\mathcal{S}, k)$ implies $\mathcal{P}^{gen}_{logic}(\mathcal{S}, k)$.*

It turns out however that $\mathcal{P}_{logic}(\mathcal{S}, k)$ and $\mathcal{P}_{wqo}(\mathcal{S}, k)$ are mutually incomparable. We will present the proofs of these results in the journal version of the paper. An important consequence of the above discussion is that $\mathcal{P}_{wqo}(\mathcal{S}, k)$ entails $PSC(k) = \exists^k \forall^*$. Note that this entailment also need not necessarily give us an effective generalization of the Łoś-Tarski theorem. This highlights the importance of our central notion, namely $P_{logic}(\mathcal{S}, k)$.

8 Conclusion

The study of preservation theorems over special classes of finite structures has recently seen a revival of interest. This paper contributes to this line of work by studying a logic-based combinatorial property that permits an effective version of the generalized Łoś-Tarski theorem to hold over several well-studied classes of finite structures. As future work, we wish to understand better the boundaries of when the generalized Łoś-Tarski theorem, and more importantly, an effective version of it, holds over classes of finite structures. The notion of well-quasi-ordering has turned out to be of central importance in several areas of computer science. In this context, Theorem 6 provides a new logic-based tool for proving that certain classes are not well-quasi-ordered under the embedding relation on structures. It also suggests that our formulations of the logic-based properties might have applications even outside the realm of preservation theorems.

Acknowledgment. The authors thank Ajit A. Diwan for insightful discussions. The authors also thank the anonymous referees for their comments.

References

1. Atserias, A., Dawar, A., Grohe, M.: Preservation under extensions on well-behaved finite structures. SIAM J. Comput. 38(4), 1364–1381 (2008)
2. Atserias, A., Dawar, A., Kolaitis, P.G.: On preservation under homomorphisms and unions of conjunctive queries. J. ACM 53(2), 208–237 (2006)
3. Chang, C.C., Keisler, H.J.: Model Theory, 3rd edn. Elsevier Science Publishers (1990)
4. Diestel, R.: Graph Theory. Springer (2010)
5. Duris, D.: Extension preservation theorems on classes of acyclic finite structures. SIAM J. Comput. 39(8), 3670–3681 (2010)
6. Harwath, F., Heimberg, L., Schweikardt, N.: Preservation and decomposition theorems for bounded degree structures. In: Proc. of CSL-LICS (2014)
7. Libkin, L.: Elements of Finite Model Theory. Springer (2004)
8. Makowsky, J.A.: Algorithmic uses of the Feferman-Vaught theorem. Ann. Pure Appl. Logic 126(1-3), 159–213 (2004)
9. Nesetril, J., de Mendez, P.O.: Tree-depth, subgraph coloring and homomorphism bounds. Eur. J. Comb. 27(6), 1022–1041 (2006)
10. Sankaran, A., Adsul, B., Chakraborty, S.: Generalizations of the Łoś-Tarski preservation theorem. CoRR, abs/1302.4350 (2013)
11. Sankaran, A., Adsul, B., Madan, V., Kamath, P., Chakraborty, S.: Preservation under substructures modulo bounded cores. In: Ong, L., de Queiroz, R. (eds.) WoLLIC 2012. LNCS, vol. 7456, pp. 291–305. Springer, Heidelberg (2012)