FO-Definability of Shrub-Depth

Yijia Chen

Fudan University, Shanghai, China
yijiachen@fudan.edu.cn

Jörg Flum

Albert-Ludwigs-Universität Freiburg, Germany
flum@uni-freiburg.de

Abstract

Shrub-depth is a graph invariant often considered as an extension of tree-depth to dense graphs. We show that the model-checking problem of monadic second-order logic on a class of graphs of bounded shrub-depth can be decided by AC^0-circuits after a precomputation on the formula. This generalizes a similar result on graphs of bounded tree-depth [3]. At the core of our proof is the definability in first-order logic of tree-models for graphs of bounded shrub-depth.

2012 ACM Subject Classification Theory of computation → Graph algorithms analysis; Mathematics of computing → Graph theory

Keywords and phrases shrub-depth, model-checking, monadic second-order logic

Digital Object Identifier 10.4230/LIPIcs.CSL.2020.15

Funding The collaboration of the authors is funded by the Sino-German Center for Research Promotion (GZ 1518). Yijia Chen is also supported by National Natural Science Foundation of China (Project 61872092).

Acknowledgements We thank Abhisekh Sankaran for discussions concerning problems of the last section. Anonymous reviewers' comments also help to improve the presentation.

1 Introduction

In [15] Ganian et al. introduced the graph invariant shrub-depth with the goal to extend the invariant tree-depth in a similar way as clique-width extends tree-width. Shrub-depth turned out to be a quite robust notion as shown by the following result of [15].

For a class K of graphs the following are equivalent:

(i) *K has bounded shrub-depth.*

(ii) *K has an* MSO*-interpretation (i.e., an interpretation definable in monadic second-order logic* MSO*) of width one in a class of rooted labelled trees of bounded depth.*

(iii) *K has bounded* SC*-depth (subset-complementation depth).*

Let $p\text{-MC}(K, \text{MSO})$ denote the parameterized model-checking problem for MSO on the class K parameterized by the length of the formula. In [3] we showed that $p\text{-MC}(K, \text{MSO})$ is in para-AC^0 for every class K of graphs of bounded tree-depth. The parameterized circuit complexity class para-AC^0 is considered to be the parameterized analog of the circuit complexity class (dlogtime-uniform) AC^0. In fact, by definition, a parameterized problem is in para-AC^0 if it is in (dlogtime-uniform) AC^0 after a precomputation on the parameter. Recall that the class FPT (fixed-parameter-tractability) consists of the parameterized problems that are solvable in polynomial time after a precomputation on the parameter.

As the main result of this paper we extend our result on the MSO-model-checking for classes of bounded tree-depth to classes of bounded shrub-depth.

▶ **Theorem 1.** $p\text{-MC}(K, \text{MSO}) \in \text{para-}\mathrm{AC}^0$ *for every class of bounded shrub-depth.*

28th EACSL Annual Conference on Computer Science Logic (CSL 2020).
Editors: Maribel Fernández and Anca Muscholl; Article No. 15; pp. 15:1–15:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

It is well known that p-MC$(K, \mathrm{MSO}) \in \mathrm{FPT}$ if K has bounded clique-width (a result due to Courcelle et al. [5]). By [15] every class of bounded shrub-depth has bounded clique-width. Hence, p-MC$(K, \mathrm{MSO}) \in \mathrm{FPT}$ for K of bounded shrub-depth. However, there exist graph classes K of bounded clique-width with p-MC$(K, \mathrm{MSO}) \notin$ para-AC^0, e.g., the class of all graphs consisting of disjoint paths [3, Theorem 7.3]. Therefore, the algorithmic techniques via clique-width cannot be adapted to para-AC^0. Instead, we develop some combinatorial machinery on graphs of bounded shrub-depth which can be defined in first-order logic (FO).

We briefly explain some ingredients of this combinatorial machinery. Central to the definition of shrub-depth are tree-models of graphs. The tree-models are rooted trees of constant depth with colored leaves, the leaves being the vertices of the corresponding graph. Their FO-definability has presented some major challenges. To better understand tree-models, we find it more convenient to work with the SC-depth instead of the shrub-depth. Roughly speaking, the SC-depth SC(G) of a graph G is the minimum number of parallel subset complementations required to construct G from graphs without any edges (i.e, from graphs of isolated vertices). The equivalence between (i) and (iii) mentioned at the beginning tells us that the boundedness of the SC-depth of a class of graphs is equivalent to the boundedness of its shrub-depth. As a first step we prove that the complementation subsets (we call them flipping sets) underlying SC(G) can be uniquely determined in FO if we are given an *unambiguous representative system* of G (once having the flipping sets we can construct a tree-model as in the proof of the implication (iii) $\Rightarrow$ (i) in [15]). However the size of a representative system cannot be bounded in terms of the depth of the graph. Hence we cannot afford to guess such a system in FO. We show that every graph of bounded SC-depth has a representation as a *tiered graph.* For such graphs we can guess appropriate representative systems iteratively in FO. Once all the flipping subsets have been obtained, we can FO-define a tree-model. More precisely,we show:

► **Theorem 2.** *If K is a class of bounded shrub-depth (or of bounded* SC*-depth), there is an* FO*-interpretation that assigns to every ordered graph $(G, <)$ with $G \in K$ a tree-model.*

Barrington et al. [1] showed that the expressive power of FO with built-in arithmetic coincides exactly with dlogtime-uniform AC^0-computability. Using this fact we get Theorem 1 from Theorem 2 in the same way as we did for tree-depth in [3].

We obtain a further consequence of our proof of the FO-definability of tree-models. In fact, we get that every MSO-sentence is equivalent to an FO-sentence on *ordered* graphs of bounded shrub-depth. More precisely:

► **Proposition 3.** *Let K be a class of graphs of bounded shrub-depth. Then for every* MSO*-sentence φ there is an* FO*-sentence ψ such that for any ordered graph $(G, <)$ with $G \in K$,*

$$G \models \varphi \iff (G, <) \models \psi.$$

Observe that the above φ has no access to the order, while ψ does. So it is natural to ask whether we can eliminate all occurrences of $<$ in ψ, in other words, whether MSO = FO on K. The result was already claimed by Gajarský and Hliněný [11]:

► **Theorem 4.** MSO = FO *on every class of graphs of bounded shrub-depth.*

We prove this result from Proposition 3 using Craig's Interpolation Theorem. Craig's Interpolation Theorem [6] is a basic result in classical model theory. However it fails on finite models [17]. To circumvent this problem, in a straightforward way we generalize the notion of

SC-depth to infinite graphs and observe that our combinatorial characterization of bounded SC-depth carries over to infinite graphs as well. The excursion to the infinite yields new insights for finite graphs, e.g., we show the following effective version of [15, Corollary 5.6]:

▶ **Theorem 5.** *There is an algorithm that applied to d eventually stops and outputs a finite set F_d of graphs such that a graph has* SC*-depth $\leq d$ if and only if it excludes the graphs in F_d as induced subgraphs.*

Related work. Theorem 1 can be viewed as a part of the recent efforts to extend algorithmic meta-theorem to dense graphs. Algorithmic meta-theorems unify many algorithmic results on graph classes where the underlying computational problems can be defined in terms of logic. Most existing such meta-theorems concern sparse graph classes, i.e., graph classes where the number of edges is linearly bounded by the number of vertices. As examples we mention Courcelle's Theorem [4] that the p-MC(MSO, K) can be solved in fixed-parameter linear time provided that K has bounded tree-width and the result due to Grohe et al. [16] stating that p-MC(FO, K) $\in$ FPT if K is a nowhere dense class. The dependence of the parameter in Courcelle's Theorem is non-elementary as shown by Frick and Grohe [9]. Improvements of the dependence of the parameter are known for various classes of graphs (see e.g., Gajarský and Hliněný [10] and Lampis [19]). A similar better dependence of the parameter holds for graph classes of bounded shrub-depth if every graph is given alongside with a tree-model of corresponding depth [10]. As far as circuit complexity is concerned, Pilipczuk et. al showed [20] that the model-checking problem for FO on graphs of bounded expansion can be decided by circuits of size $f(k) \cdot n^{O(1)}$ and of depth $f(k) + O(\log n)$, where k is the size of the input formula and n the size of the graph.

Compared to sparse graph classes, much less is known for dense graphs. We have already mentioned that p-MC(K, MSO) $\in$ FPT if K has bounded clique-width. Recall that the class of cliques (i.e., complete graphs), which are obviously dense, has clique-width 1. For first-order logic, algorithmic meta-theorems are known e.g., for interval graphs [14], partial orders [12], and graphs FO-interpretable in bounded degree graphs [13].

In [8] Elberfeld et al. proved that MSO = FO on graphs of bounded tree-depth. Graphs of bounded tree-depth has bounded shrub-depth as well. Thus Theorem 4 generalizes this result. As already mentioned, Theorem 4 was first claimed in [11, Theorem 5.14]. One crucial tool is based on the proof of [11, Theorem 5.2]. However, we could not verify this proof. Besides that, our proof uses completely different techniques.

The MSO-sentence ψ in Proposition 3 contains a symbol for the order relation, however its validity in a graph G (of the class K) does not depend on what order of the set of vertices of G we choose. That is, by definition, the sentence φ is order-invariant on K. In a recent paper Eickmeyer et al. [7] obtain FPT-tractability results for the set of order-invariant MSO-sentences essentially for the same classes of graphs as in the unordered case. However the model-checking problem for order-invariant MSO on graphs of bounded tree-depth (thus on graphs of bounded shrub-depth) is not in para-AC^0. In fact, consider graphs consisting of disjoint triangles and isolated vertices. Then the parity of the number of the triangles can be expressed by an order-invariant MSO-sentence. On the other hand, it is easy to see that this property cannot be in AC^0 by PARITY $\notin \mathrm{AC}^0$.

Organization of this paper. In Section 2 we fix some notations. In Section 3 and Section 4 we recall the definitions and some basic properties of shrub-depth and of SC-depth, respectively, and show that the classes $\mathrm{TM}_m(d)$ and $\mathrm{SC}(d)$ are MSO-axiomatizable. In Sections 5–7 we stepwise develop the machinery which finally allows us to prove Theorem 2 (and hence, Theorem 1) in Section 8. Theorem 4 and Theorem 5 are shown in Section 9 and Section 10, respectively.

Due to space limitations we defer many proofs to the full version of this paper. Sometimes we indicate this by writing "Proof: full paper" at the end of the statement of a theorem, proposition,...

2 Preliminaries

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

First-order logic FO and monadic second-order logic MSO. 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 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 $R \in \tau$. In this paper all structures have a finite universe with the exception of Section 9 and Section 10.

Formulas φ of *first-order logic* FO of vocabulary τ are built up from *atomic formulas* $x_1 = x_2$ and $Rx_1 \dots x_r$ (where $R \in \tau$ is of arity r and $x_1, x_2, \dots, x_r$ are variables) using the boolean connectives $\neg$, $\wedge$, and $\vee$ and the universal $\forall$ and existential $\exists$ quantifiers. By the notation $\varphi(\bar{x})$ with $\bar{x} = x_1, \dots, x_e$ we indicate that the variables free in φ are among $x_1, \dots, x_e$. In addition to the individual variables of FO, *formulas* of *monadic second-order logic* MSO may also contain *set variables*. We use lowercase letters (usually x, y, z) to denote individual variables and uppercase letters (usually X, Y, Z) to denote set variables. To obtain MSO the syntax of FO is enhanced by new atomic formulas of the form Xy and quantification is also allowed over set variables.

Graphs and trees. In this paper *graphs* are always simple and undirected. When considering definability problems for graphs we view graphs as $\tau := \{E\}$-structures where the edge relation is an irreflexive and symmetric binary relation. Otherwise we use the notation G for a graph and view it as a pair $G = (V(G), E(G))$, where $V(G)$ is the set of vertices and $E(G)$ the set of edges. For graphs G and H with disjoint vertex sets we denote by $G \,\dot\cup\, H$ the graph with vertex set $V(G) \cup V(H)$ and edge set $E(G) \cup E(H)$.

We view *rooted trees with m labels* as $\tau_m := \{P, L_1, \dots, L_m\}$-structures $\mathcal{T} = \big(T, P^{\mathcal{T}}, L_1^{\mathcal{T}}, \dots, L_m^{\mathcal{T}}\big)$. Here P is a binary relation symbol and $L_1, \dots, L_m$ are unary. $P^{\mathcal{T}}$ is the *parent-child relation* of the tree. The root of the tree can be defined by the formula $root(x) := \forall y \neg Pyx$. The relations $L_1^{\mathcal{T}}, \dots, L_m^{\mathcal{T}}$ are the *labels*. Recall that the *depth of* $\mathcal{T}$ is the maximum length of a path from the root to a leaf. We denote by *leaves*$(\mathcal{T})$ the set of leaves of $\mathcal{T}$. For $m, d \in \mathbb{N}$ we denote by $\textsc{Tree}[m, d]$ the class of rooted trees with m labels and of depth d, where each root-to-leaf path is of length exactly d.

3 Shrub-depth

We recall the notion of the shrub-depth of a graph (introduced in [15]) and show that the classes $\mathrm{TM}_m(d)$ (with $m, d \in \mathbb{N}$) of bounded shrub-depth are axiomatizable in MSO.

▶ **Definition 6.** *Let $m, d \in \mathbb{N}$. A* tree-model of m labels and depth d *of a graph G is a pair $(\mathcal{T}, D)$ with $\mathcal{T} \in \textsc{Tree}[m, d]$ and $D \subseteq \{1, 2, \dots, m\}^2 \times \{1, 2, \dots, h\}$ for some $h \geq d$* [1] *(called*

[1] For technical reasons (in particular, for the proof of Proposition 12), we allow h to be greater than d.

the signature *of the tree-model) such that*

- $V(G) = leaves(\mathcal{T})$,
- *each leaf of T holds exactly one label from $\{P_1, \dots, P_m\}$ and no other node of T holds a label, i.e., $leaves(\mathcal{T}) = P_1^{\mathcal{T}} \dot\cup \cdots \dot\cup P_m^{\mathcal{T}}$,*
- *for any $i, j \in [m]$ and $s \in [d]$ if $(i, j, s) \in D$, then $(j, i, s) \in D$,*
- $E(G) = \big\{\{u, v\} \mid u, v \in V(G),\ u \neq v,\ u \in P_i^{\mathcal{T}},\ v \in P_j^{\mathcal{T}},\ \text{and}\ \big(i, j, dist^{\mathcal{T}}(u \wedge v, u)\big) \in D\big\}$.

By $dist^{\mathcal{T}}(u \wedge v, u)$ we denote the distance from the least common ancestor $u \wedge v$ of u and v to v. Note that $dist^{\mathcal{T}}(u \wedge v, u) = dist^{\mathcal{T}}(u \wedge v, v)$, as both u and v are leaves of $\mathcal{T}$ (of the same depth). In the context of tree-models we also speak of the colors *P_i and say that vertex v has color P_i if $v \in P_i^{\mathcal{T}}$.*

▶ **Definition 7** ([15])**.** *Let $\mathrm{TM}_m(d)$ denote the class of graphs with a tree-model of m labels and depth d. A class K of graphs has* shrub-depth d *if there exists m such that $K \subseteq \mathrm{TM}_m(d)$, while for all $m' \in \mathbb{N}$ we have $K \not\subseteq \mathrm{TM}_{m'}(d-1)$.*

The class K has bounded shrub-depth *if $K \subseteq \mathrm{TM}_m(d)$ for some $m, d \in \mathbb{N}$.*

The following lemma shows that the shrub-depth is relevant only to infinite classes of graphs and not to a single graph.

▶ **Lemma 8.** *For every graph G we have $G \in \mathrm{TM}_{|V(G)|}(1)$.*

Proof. Assume $V(G) = [m]$. Then a tree $\mathcal{T} \in \mathrm{TREE}[m, 1]$ with $P_i^{\mathcal{T}} = \{i\}$ for $i \in [m]$ together with the signature $D := \{(i, j, 1) \mid i, j \in [m] \text{ and } \{i, j\} \in E(G)\}$ is a tree-model for G. ◀

The shrub-depth hierarchy is strict:

▶ **Proposition 9.** *Let $d \in \mathbb{N}$. The class of graphs underlying rooted trees in $\mathrm{TREE}[m, d]$ for some (= all) m has shrub depth d.* Proof: full paper.

The following facts are easy to verify.

▶ **Lemma 10.**

(a) $\mathrm{TM}_m(d) \subseteq \mathrm{TM}_{m'}(d')$ *for $m \leq m'$ and $d \leq d'$.*
(b) $\mathrm{TM}_m(d)$ *is closed under induced subgraphs.*
(c) *Every graph which is a clique is in $\mathrm{TM}_1(1)$.*

A proof of the next result can be found in [15].

▶ **Proposition 11.** *There is a computable function $\ell : \mathbb{N}^2 \to \mathbb{N}$ such that for $G \in \mathrm{TM}_m(d)$ every two vertices, which are in the same connected component of G, have a distance $\leq \ell(m, d)$. Hence, for fixed $m, d \in \mathbb{N}$, we can express in* FO *that two vertices are in the same connected component in graphs of $\mathrm{TM}_m(d)$.*

By this lemma we see that the class of paths is not of bounded shrub-depth. As every path is a subgraph of some clique, Lemma 10(c) shows that the classes $\mathrm{TM}_m(d)$ for $m, d \geq 1$ are not closed under subgraphs.

By [15, Corollary 5.6] we know that for $m, d \in \mathbb{N}$ there is a finite set $F_{m,d}$ of graphs such that a graph G is in $\mathrm{TM}_m(d)$ if and only if "G excludes the graphs in $F_{m,d}$ as induced subgraphs", i.e, no graph in $F_{m,d}$ is isomorphic to an induced subgraph of G. Hence there is an FO-sentence $\rho(m, d)$ axiomatizing $\mathrm{TM}_m(d)$. However the proof of [15, Corollary 5.6] sheds no light on how to compute $\rho(m, d)$ from (m, d). We need the corresponding result for MSO in order to get such an effective FO-axiomatization of $\mathrm{TM}_m(d)$ in Section 9.

▶ **Proposition 12.** *We can effectively compute for $m, d \in \mathbb{N}$ an* MSO*-axiomatization of $\mathrm{TM}_m(d)$.* Proof: full paper.

4 SC-depth

Classes of graphs of bounded shrub-depth coincide with classes of graphs of bounded SC-depth. For our goal it is more convenient to work with the SC-depth. Let G be a graph and S a subset of its vertex set $V(G)$. Then G^S denotes the graph *obtained from G by flipping the set S*; that is, G^S has the vertex set $V(G)$ and edge set

$$\big\{\{u,v\} \in E(G) \mid u \notin S \text{ or } v \notin S\big\} \cup \big\{\{u,v\} \mid u,v \in S,\ u \neq v, \text{ and } \{u,v\} \notin E(G)\big\}.$$

Here, we deviate from the original notation $\bar{G}^S$, which might become cumbersome when there are several flipping sets. For subsets $S_1,\ldots,S_n$ of $V(G)$ we write $G^{S_1\cdots S_n}$ for $\big(\ldots((G^{S_1})^{S_2})\ldots\big)^{S_n}$. The following lemma contains some simple facts about $G \mapsto G^S$.

▶ **Lemma 13.** *Let $S \subseteq V(G)$, $T \subseteq V(G)$, and H be a further graph.*
(a) *If $|S| \leq 1$, then $G^S = G$;*
(b) *if $G^S = G^T$ and $|S| \geq 2$, then $S = T$;*
(c) *$G^{ST} = G^{TS}$;*
(d) *$G^{SS} = G$;*
(e) *$(G^S \mathbin{\dot\cup} H) = (G \mathbin{\dot\cup} H)^S$ (recall that $S \subseteq V(G)$).*

We introduce the class $\mathrm{SC}(d)$ of graphs of *complementation depth* $\leq d$ (or, SC-*depth* $\leq d$).

▶ **Definition 14.** *Let $d \in \mathbb{N}$. We define inductively the class $\mathrm{SC}(d)$.*

$\mathrm{SC}(0)$ *is the class of graphs whose vertex set is a singleton.*

Assume that $m \geq 1$ and the graphs $G_1,\ldots,G_m \in \mathrm{SC}(d)$ have pairwise disjoint vertex sets. Then for $S \subseteq V(G_1) \cup \cdots \cup V(G_m)$ we have

$$\big(G_1 \mathbin{\dot\cup} G_2 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} G_m\big)^S \in \mathrm{SC}(d+1).$$

A class of graphs is of bounded SC-depth *if it is contained in $\mathrm{SC}(d)$ for some $d \in \mathbb{N}$.*

As every clique has SC-depth ≤ 1 we see that the class of cliques has bounded SC-depth. The following lemma shows that every graph is in some $\mathrm{SC}(d)$. We define the SC-*depth* $\mathrm{SC}(G)$ of a graph G as the least $d \in \mathbb{N}$ such that $G \in \mathrm{SC}(d)$.

▶ **Lemma 15.** *$G \in \mathrm{SC}(|V(G)|-1)$ for every graph G.*

Proof. The proof is a simple induction on $|V(G)|$. A graph with only one vertex is in $\mathrm{SC}(0)$ by definition. Let $d \geq 1$ and let u be any vertex of a graph G with exactly $d+1$ vertices. Let H be the graph induced by G on $V(G) \setminus \{u\}$ and set $H_1 := H^{\{v \in V(H) \mid \{u,v\} \in E(G)\}}$. By induction hypothesis, $H_1 \in \mathrm{SC}(d-1)$ as H_1 has d elements. Let U denote the graph with $V(U) = \{u\}$. As $G = (H_1 \mathbin{\dot\cup} U)^{\{u\} \cup \{v \in V(H) \mid \{u,v\} \in E(G)\}}$, we get $G \in \mathrm{SC}(d)$. ◀

If we write $G \mathbin{\dot\cup} H$ we tacitly assume that the graphs G and H have disjoint vertex sets and if we write G^S we assume that $S \subseteq V(G)$.

The following basic properties of the classes $\mathrm{SC}(d)$ will be proven in the full paper.

▶ **Lemma 16.** *Let $d \in \mathbb{N}$.*
(a) $\mathrm{SC}(d) \subseteq \mathrm{SC}(d+1)$.
(b) $\mathrm{SC}(d)$ *is closed under taking induced subgraphs.*
(c) $\mathrm{SC}(d+1) = \left\{ \big(G_1 \mathbin{\dot\cup} G_2 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} G_m\big)^S \;\middle|\; \begin{array}{l} m \geq 1,\ G_1,\ \ldots,\ G_m \in \mathrm{SC}(d) \text{ are connected and } S \subseteq V(G_1) \cup \cdots \cup V(G_m) \end{array} \right\}.$

Moreover, assume that $H = \big(G_1 \mathbin{\dot\cup} G_2 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} G_m\big)^S$ with $G_i \in \mathrm{SC}(d)$. Then for the connected components H_{ij} of G_i we have $H_{ij} \in \mathrm{SC}(d)$ and $H = \big(\dot\bigcup H_{ij}\big)^S$.

Parts (a) and (b) of the following proposition show that a class of graph has bounded shrub-depth if and only if it has bounded SC-depth. Then its part (c) follows by Proposition 11.

▶ **Proposition 17** ([15]).

(a) *Let* $m, d \in \mathbb{N}$. *Then* $\mathrm{TM}_m(d) \subseteq \mathrm{SC}(d \cdot m \cdot (m+1))$.

(b) *Let* $d \in \mathbb{N}$. *Then* $\mathrm{SC}(d) \subseteq \mathrm{TM}_{2^d}(d)$.

(c) *There is a computable function* $\ell_{\mathrm{SC}} : \mathbb{N} \to \mathbb{N}$ *such that for* $G \in \mathrm{SC}(d)$ *every two vertices, which are in the same connected component of* G*, have a distance* $\leq \ell_{\mathrm{SC}}(d)$.

Again using the existence of a characterization of $\mathrm{SC}(d)$ in terms of excluding a finite set of induced subgraphs, one gets the FO-axiomatizability of $\mathrm{SC}(d)$. We will show the effective FO-axiomatizability (see Corollary 43(a)). Here we get (see the full paper for a proof):

▶ **Proposition 18.** *We can effectively compute for* $d \in \mathbb{N}$ *an* MSO*-axiomatization* ρ_d *of* $\mathrm{SC}(d)$.

5 Towers and representative systems

For $d \geq 1$ we denote by $\textsc{Tow}(d)$ the *class of towers* $\leq d$, i.e., the class of graphs which can be written in the form

$$G = \big(I(V(G))\big)^{S_1 \dots S_d}. \tag{1}$$

Here $I(X)$ denotes the graph with vertex X and no edges. By Lemma 13(a) we have $\textsc{Tow}(d) \subseteq \textsc{Tow}(d+1)$ and by Lemma 13(e), every graph is in $\textsc{Tow}(d)$ for some d. Note that $\textsc{Tow}(d) \subseteq \mathrm{SC}(d)$ for $d \geq 1$. However, already $\mathrm{SC}(2)$ is not contained in any class $\textsc{Tow}(d)$. In fact, the graphs $G_n := \Big(\{a_1, \dots, a_n, b_1, \dots, b_n\}, \big\{\{a_i, b_i\} \mid i \in \mathbb{N}\big\}\Big)$ for $n \geq 1$ are all contained in $\mathrm{SC}(2)$. Note that every two vertices of the graph in (1), which are in the same atom of the boolean algebra generated by $S_1, \dots, S_d$, "behave in the same way." Hence, $G_n \notin \textsc{Tow}(d)$ for $d < \log_2 n$. Readers familiar with [19] will realize that a class of graphs has bounded *neighborhood diversity* if and only if it is contained in $\textsc{Tow}(k)$ for some $k \in \mathbb{N}$.

In this section, as a first step towards the main results we show that the classes $\textsc{Tow}(d)$ are FO-axiomatizable, thereby getting familiar with some tools relevant to the general case.

We set $\mathbf{S} := S_1 \dots S_d$. For G as in (1) we associate with every $v \in G$ a "color"

$$\chi_{\mathbf{S}}(v) := (b_1, \dots, b_d) \in \{0,1\}^d, \quad \text{where } b_i = \begin{cases} 1 & \text{if } v \in S_i \\ 0 & \text{otherwise.} \end{cases}$$

For $b = (b_1, \dots, b_d) \in \{0,1\}^d$ and $b' = (b'_1, \dots, b'_d) \in \{0,1\}^d$ we define

$$\langle b, b' \rangle := \sum_{i \in [d]} b_i \cdot b'_i \mod 2.$$

▶ **Lemma 19.** *Let* $G = \big(I(X)\big)^{\mathbf{S}}$ *and* $v, w \in V(G)$ *with* $v \neq w$. *Then*

$$\{v, w\} \in E^G \iff \langle \chi_{\mathbf{S}}(v), \chi_{\mathbf{S}}(w) \rangle = 1.$$

Note that the mapping $\chi_{\mathbf{S}} : V(G) \to \{0,1\}^d$ is not necessarily surjective. Assume that $\chi_{\mathbf{S}}(V(G)) = \{b_1, \dots, b_m\}$ with pairwise distinct b_i's in $\{0,1\}^d$. For $i \in [m]$ choose a vertex $u_i \in V(G)$ such that $\chi_{\mathbf{S}}(u_i) = b_i$. Then, $(u_1, \dots, u_m; b_1, \dots, b_m)$ is a d-representative system for G in the sense of the following definition and $\chi_{\mathbf{S}}$ is a corresponding coloring.

▶ **Definition 20.** *Let G be a graph and $d \geq 1$. A d-*representative system *for G is a tuple*

$$\mathcal{R} := \big(u_1, \dots, u_m; b_1, \dots, b_m\big)$$

with $u_1, \dots, u_m \in V(G)$ and with pairwise distinct elements $b_1, \dots, b_m$ of $\{0,1\}^d$ if there is a "coloring" $\chi : V(G) \to \{b_1, \dots, b_m\}$ with (R1) and (R2).

(R1) *For every $i \in [m]$: $\chi(u_i) = b_i$.*

(R2) *For all $v, w \in V(G)$ with $v \neq w$: $\Big(\{v,w\} \in E^G \iff \langle \chi(v), \chi(w)\rangle = 1\Big)$.*

The vertices $u_1, \dots, u_m$ are then called representatives.

▶ **Proposition 21.** *A graph is in* $\textsc{Tow}(d)$ *if and only if it has a d-representative system.*

We prove this characterization of $\textsc{Tow}(d)$ in the full paper. It does not yield an FO-axiomatization of $\textsc{Tow}(d)$ as we need the coloring χ. In general this coloring is not uniquely determined (again see the full paper). This fact motivates the following definition.

▶ **Definition 22.**

(i) *Let $d \geq 1$ and $B \subseteq \{0,1\}^d$. The set B is* unambiguous *if for $b_1, b_2 \in B$, $\langle b_1, b\rangle = \langle b_2, b\rangle$ for all $b \in B$ implies $b_1 = b_2$.*

(ii) *Let G be a graph with $G = \big(I(V(G))\big)^{S_1 \dots S_d}$. Then $S_1, \dots, S_d$ is* unambiguous *if $\chi_{S_1 \dots S_d}(V(G))$ is unambiguous. A d-representative system $(u_1, \dots, u_m; b_1, \dots, b_m)$ for G is* unambiguous *if $\{b_1, \dots, b_m\}$ is* unambiguous.

Every representative system contains an unambiguous representative system.

▶ **Lemma 23.** *Let $\mathcal{R} := \big(u_1, \dots, u_m; b_1, \dots, b_m\big)$ be a d-representative system for a graph G. Then there is an $s \in [m]$ and $1 \leq i_1 < \ldots < i_s \leq m$ such that $\big(u_{i_1}, \dots, u_{i_s}; b_{i_1}, \dots, b_{i_s}\big)$ is an unambiguous d-representative system for G.* Proof: full paper.

Why is unambiguity an important property? The next result shows that for unambiguous representative systems there is a unique coloring. Its value for a vertex is already determined by its neighbors in the set of representatives.

▶ **Proposition 24.** *Let G be a graph, $\mathcal{R} := \big(u_1, \dots, u_m; b_1, \dots, b_m\big)$ be an unambiguous d-representative system for G, and χ a corresponding coloring. Then (by Definition 20 and unambiguity) for $v \in V(G) \setminus \{u_1, \dots, u_m\}$ the color $\chi(v)$ is the unique b_j with $j \in [m]$ such that for all $i \in [m]$ we have*

$$\{v, u_i\} \in E(H) \iff \langle b_j, b_i\rangle = 1.$$

Then $S_1, \dots, S_d \subseteq V(G)$ with $S_i := \{v \in V(G) \mid (\chi(v))_i = 1\}\big($by $(\chi(v))_i$ we denote the ith component of $\chi(v)\big)$ is unambiguous and $G = (I(V(G)))^{S_1 \dots S_d}$.

Proof. The second part follows from the fact that $\chi_{S_1 \dots S_d} = \chi$. ◀

Now we easily get the FO-axiomatizability of $\textsc{Tow}(d)$ (for a proof see the full paper).

▶ **Theorem 25.** *For $d \geq 1$ the class* $\textsc{Tow}(d)$ *is axiomatizable in* FO.

6 Tiered graphs

We introduce (d,q)-tiered graphs, a technical tool we use to obtain our main results. Every graph we considered in the previous section is $(0,q)$-tiered for some $q \in \mathbb{N}$. So in the preceding section we saw that for $(0,q)$-tiered graphs G we can FO-define flipping sets that applied to $V(G)$ yield G. (d,q)-tiered graphs G contain some distinguished sets of flipping sets $\mathbf{S}_0, \dots, \mathbf{S}_d$. In this section we show that essentially we can FO-define flipping sets $\mathbf{S}'_0, \dots, \mathbf{S}'_d$ with $G^{\mathbf{S}_0, \dots, \mathbf{S}_d} = G^{\mathbf{S}'_0, \dots, \mathbf{S}'_d}$ (see Corollary 32).

▶ **Definition 26.** *For a set X and $S_1, \dots, S_d \subseteq X$ we set*

$$\mathit{color}(X, S_1 \dots S_d) := \{\chi_{S_1 \dots S_d}(v) \mid v \in X\},$$

the set of colors of elements of X w.r.t. $S_1, \dots, S_d$.

▶ **Definition 27.** *Let $q, d \in \mathbb{N}$.*

(a) *If $G = \big(I(V(G))\big)^{\mathbf{S}}$ with $|\mathbf{S}| \leq q$, then G is a* $(0, q)$-tiered *graph.*

(b) *Assume $d \geq 1$ and let $G_0, \dots, G_d = G$ be graphs. If*

- **(i)** *$G_0 = \big(I(X)\big)^{\mathbf{S}_0}$ for some set X and some $\mathbf{S}_0$ with $|\mathbf{S}_0| \leq q$,*
- **(ii)** *for every $t \in [d]$ we have $G_t = \Big(G_{t-1} \dot\cup \dot\bigcup_{\delta \in \Delta_t,\ e \in F_\delta} H_{te}\Big)^{\mathbf{S}_t}$ for some $\mathbf{S}_t$ with $|\mathbf{S}_t| \leq q$, for some finite Δ_t and F_δ for $\delta \in \Delta_t$, and for some graphs H_{te} for $e \in F_\delta$ with $\delta \in \Delta_t$,*
- **(iii)** *for every $t \in [d]$ and every $\delta \in \Delta_t$ we have $|F_\delta| \geq 3$ and for all $e, e' \in F_\delta$,*

$$c(\delta) := \mathit{color}(V(H_{te}), \mathbf{S}_t \dots \mathbf{S}_d) = \mathit{color}(V(H_{te'}), \mathbf{S}_t \dots \mathbf{S}_d),$$

then G is a (d, q)-tiered *graph.*

Note that part (b)(iii) is the only restriction on the graphs H_{te} even though in our applications these graphs will be "simpler" than G. In part (b)(ii) we allow that on the right hand side of the equality at most one of the terms is missing. That is, it can be that either the term G_{t-1} is not present (G_{t-1} is the "empty" graph) or that $\Delta_t = \emptyset$ $\big(\big(\dot\bigcup_{\delta \in \Delta_t,\ e \in F_\delta} H_{te}\big)^{\mathbf{S}_t}$ is the "empty" graph$\big)$. If for $t = 1$ the term G_0 is not present, then X is empty in (b)(i).

In this section and the next one terms may represent the "empty" graph by similar reasons.

For G as in Definition 27(b)(ii) and $t \in \{0, 1, \dots, d\}$ we define the tth *tier* T_t by

$$T_0 = X \quad \text{and} \quad T_t = \bigcup_{\delta \in \Delta_t,\ e \in F_\delta} V(H_{te}) \ \text{ for all } t \in [d].$$

▶ **Lemma 28.** *Let $v \in T_t$ with $t \in \{0, \dots, d\}$. Then $\chi_{\mathbf{S}_0 \dots \mathbf{S}_{t-1}}(v) = \bar{0}$.*

By Definition 27 we have

$$\begin{aligned} G &= \Big(\cdots \big(I(T_0)\big)^{\mathbf{S}_0} \dot\cup \dot\bigcup_{\delta \in \Delta_1,\ e \in F_\delta} H_{1e}\Big)^{\mathbf{S}_1} \dot\cup \cdots \dot\cup \dot\bigcup_{\delta \in \Delta_d,\ e \in F_\delta} H_{de}\Big)^{\mathbf{S}_d} \\ &= \Big(I(T_0) \dot\cup \dot\bigcup_{t \in [d],\ \delta \in \Delta_t,\ e \in F_\delta} H_{te}\Big)^{\mathbf{S}_0 \mathbf{S}_1 \dots \mathbf{S}_d} \qquad \text{(by Lemma 13 (e)).} \end{aligned} \tag{2}$$

Our goal is to show that in G we can FO-define flipping sets "equivalent to" $\mathbf{S}_0, \mathbf{S}_1, \dots, \mathbf{S}_d$. To that end, we introduce an auxiliary graph

$$L := \Big(\dots \Big(\big(I(T_0)\big)^{\mathbf{S}_0} \dot\cup I(T_1)\Big)^{\mathbf{S}_1} \dot\cup \cdots \dot\cup I(T_d)\Big)^{\mathbf{S}_d} = \big(I(V(G))\big)^{\mathbf{S}_0 \mathbf{S}_1 \dots \mathbf{S}_d}. \tag{3}$$

We want to apply to $L = \big(I(V(G))\big)^{\mathbf{S}_0 \mathbf{S}_1 \dots \mathbf{S}_d}$ the results developed in the preceding section. First we show that relevant information on $E(L)$ can be FO-defined in G.

Let $t \in [d]$. For $\delta \in \Delta_t$ we fix pairwise distinct $e_1, e_2, e_3 \in F_\delta$. As for $i \in [3]$, $c(\delta) = \mathit{color}(V(H_{te_i}), \mathbf{S}_t \dots \mathbf{S}_d)$, we choose for $I(V(H_{te_i}))^{\mathbf{S}_t \dots \mathbf{S}_d}$ a representative system $\big(u^\delta_{i1}, \dots, u^\delta_{i|c(\delta)|}; \chi_{\mathbf{S}_t \dots \mathbf{S}_d}(u^\delta_{i1}), \dots, \chi_{\mathbf{S}_t \dots \mathbf{S}_d}(u^\delta_{i|c(\delta)|})\big)$ such that for $i, j \in [3]$ and $\ell \in \big[|c(\delta)|\big]$,

$$\chi_{\mathbf{S}_t \dots \mathbf{S}_d}(u^\delta_{i\ell}) = \chi_{\mathbf{S}_t \dots \mathbf{S}_d}(u^\delta_{j\ell}). \tag{4}$$

Then $VV_t := \{u^\delta_{i\ell} \mid \delta \in \Delta_t,\, i \in [3], \text{ and } \ell \in [|c(\delta)|]\}$ is the set of *voting vertices in* T_t. Let $c(0) := \mathit{color}(T_0, \mathbf{S}_0 \dots \mathbf{S}_d)$ and let $(u_1, \dots, u_{|c(0)|}; \chi_{\mathbf{S}_0\dots\mathbf{S}_d}(u_0), \dots, \chi_{\mathbf{S}_0\dots\mathbf{S}_d}(u_{|c(0)|}))$ be a representative system for $(I(T_0))^{\mathbf{S}_0\dots\mathbf{S}_d}$. We define the *set* VV *of all voting vertices* by

$$VV := \{u_1, \dots, u_{|c(0)|}\} \cup \bigcup_{t\in[d]} VV_t.$$

One easily shows:

▶ **Lemma 29.**

(a) $\chi_{\mathbf{S}_0\dots\mathbf{S}_d}(\{u \mid u \in VV\}) = \chi_{\mathbf{S}_0\dots\mathbf{S}_d}(V(L))$.

(b) *The size of* VV *can be bounded in terms of* d*,* q*, and* $\sum_{t\in[d]} |\Delta_t|$.

Part (a) of the next lemma shows that we can decide in G whether $\{u, v\} \in E(L)$ between a voting vertex u and any other vertex v (see the full paper for a proof). Essentially, the majority opinion of the voting vertices decides. Part (b) is an immediate consequence of (a).

▶ **Lemma 30.**

(a) *Let* $v \in V(G)$ *and* $u \in VV_t$ *for some* $t \geq 1$*, say* $u = u^\delta_{i\ell}$ *where* $\delta \in \Delta_t$*,* $i \in [3]$*, and* $\ell \in [|c(\delta)|]$*. If* $v \neq u$*, then*

$$\{v, u\} \in E(L) \iff \text{there are } 1 \leq i_1 < i_2 \leq 3 \text{ such that } \{v, u^\delta_{i_1\ell}\}, \{v, u^\delta_{i_2\ell}\} \in E(G).$$

For $u \in \{u_1, \dots, u_{|c(0)|}\}$ *with* $v \neq u$ *we have* $(\{v, u\} \in E(L) \iff \{v, u\} \in E(G))$.

(b) *In* G *we can express in* FO *with parameters for the elements of* VV *whether* $\{v, u\} \in E(L)$ *for* $v \in V(L)$ *and* $u \in VV$.

However the information of part (a) doesn't allow us to compute $\chi_{\mathbf{S}_0\dots\mathbf{S}_d}(v)$ for all $v \in V(L)$. Again we have the problem of ambiguity. We turn to this problem. By Lemma 29(a) we can choose vertices $u_1, \dots, u_m \in VV$ such that

$$\mathcal{R} := (u_1, \dots, u_m; \chi_{\mathbf{S}_0\dots\mathbf{S}_d}(u_1), \dots, \chi_{\mathbf{S}_0\dots\mathbf{S}_d}(u_m))$$

is a representative system for $L = (I(V(G)))^{\mathbf{S}_0\mathbf{S}_1\dots\mathbf{S}_d}$. In the preceding section we have seen how to obtain unambiguous $\mathbf{S}'_0, \dots, \mathbf{S}'_d$ with $L = (I(V(G)))^{\mathbf{S}'_0\mathbf{S}'_1\dots\mathbf{S}'_d}$. Let us recall how we did this. So assume $\mathcal{R}$ is not unambiguous. Then there are distinct $j_1, j_2 \in [m]$ such that

$$\text{for all } i \in [m]: \quad \chi_{\mathbf{S}_0\dots\mathbf{S}_d}(u_{j_1}) \oplus \chi_{\mathbf{S}_0\dots\mathbf{S}_d}(u_i) = \chi_{\mathbf{S}_0\dots\mathbf{S}_d}(u_{j_2}) \oplus \chi_{\mathbf{S}_0\dots\mathbf{S}_d}(u_i).$$

Then we gave all vertices of color $\chi_{\mathbf{S}_0\dots\mathbf{S}_d}(u_{j_2})$ the color $\chi_{\mathbf{S}_0\dots\mathbf{S}_d}(u_{j_1})$. This could be problematic as we also want to preserve (2), that is, we also aim at:

$$G = \Big(I(T_0) \,\dot\cup\, \dot{\bigcup}_{t\in[d],\ \delta\in\Delta_t,\ e\in F_\delta} H_{te}\Big)^{\mathbf{S}'_0\mathbf{S}'_1\dots\mathbf{S}'_d}. \tag{5}$$

If e.g. the vertex u_{j_2} is in H_{5e} and u_{j_1} is in $H_{3e'}$ and we give u_{j_2} the color of u_{j_1}, then already $\mathbf{S}_3$ (more precisely, $\mathbf{S}'_3$) could introduce edges between u_{j_2} and vertices in $H_{3e'}$ that destroy the validity of (5). In fact, the first equality of (2) implies that $\mathbf{S}_3$ cannot contain any vertex from H_{5e}. In the proof of our goal (the following proposition) in the full version we take care of this problem. In essence, we will always keep a vertex in its original tier.

▶ **Proposition 31.** *Let the* (d, q)*-tiered graph* G *and* L *be as above. There exist sequences of subsets* $\mathbf{S}'_0, \dots, \mathbf{S}'_d$ *of* $V(G) = V(L)$ *such that:*

(a) $G = \Big(\cdots(I(T_0)^{\mathbf{S}'_0} \dot\cup \dot\bigcup_{\delta\in\Delta_1,\ e\in F_\delta} H_{1e}\Big)^{\mathbf{S}'_1} \dot\cup \cdots \dot\cup \dot\bigcup_{\delta\in\Delta_d,\ e\in F_d} H_{de}\Big)^{\mathbf{S}'_d}$ *and thus,*

$$G = \Big(I(T_0) \dot\cup \dot\bigcup_{t\in[d],\ \delta\in\Delta,\ e\in F_\delta} H_{te}\Big)^{\mathbf{S}'_0\mathbf{S}'_1\ldots\mathbf{S}'_d}.$$

In particular, all Δ_t's, F_δ's, and H_{te}'s are the same as in (2) and for $t \in [d]$, $\delta \in \Delta_t$, and $e, e' \in F_\delta$ we have $color(V(H_{te}), \mathbf{S}'_t \ldots \mathbf{S}'_d) = color(V(H_{te'}), \mathbf{S}'_t \ldots \mathbf{S}'_d)$. Hence, also the first equality for G witnesses that G is a (d, q)-tiered graph.

(b) $L := \Big(\ldots\Big((I(T_0))^{\mathbf{S}'_0} \dot\cup I(T_1)\Big)^{\mathbf{S}'_1} \dot\cup \cdots \dot\cup I(T_d)\Big)^{\mathbf{S}'_d} = (I(V(G)))^{\mathbf{S}'_0\mathbf{S}'_1\ldots\mathbf{S}'_d}$. *Moreover, $\mathbf{S}'_0\mathbf{S}'_1\ldots\mathbf{S}'_d$ is unambiguous with respect to L.*

By (2) and Lemma 13(c), (d) we get the following immediate consequence of part (a).

▶ **Corollary 32.** $G^{\mathbf{S}'_0\mathbf{S}'_1\ldots\mathbf{S}'_d} = I(T_0) \dot\cup \dot\bigcup_{t\in[d],\ \delta\in\Delta_t,\ e\in F_\delta} H_{te} = G^{\mathbf{S}_0,\ldots,\mathbf{S}_d}$.

The main message of this section is the following: For (d, q)-tiered graphs once we have guessed the correct unambiguous representative system we can FO-define the edge relation of the graph $G^{\mathbf{S}'_0\ldots\mathbf{S}'_d}$.

7 From graphs of bounded SC-depth to tiered graphs

Here we reduce graphs of bounded SC-depth to tiered graphs (see Proposition 36). For this purpose it is useful to consider the generalized SC-depth of graphs obtained from the SC-depth by allowing "at the end" a bounded number of flipping sets.

▶ **Definition 33.** *Let $d, q \in \mathbb{N}$. By* GSC(d, q) *we denote the* class of graphs of q-generalized SC-depth $\le d$ *(here q refers to the bound for the number of flipping sets in the last step). The classes* GSC(d, q) *are defined as follows.*

(i) *If the vertex set $V(G)$ of the graph G is a singleton, then $G \in \mathrm{GSC}(0, q)$.*

(ii) *Assume $d \ge 1$ and $G = \big(\dot\bigcup_{e\in F} G_e\big)^{\mathbf{S}}$ with $|\mathbf{S}| \le q$ and $G_e \in \mathrm{SC}(d-1)$ for every $e \in F$. Then $G \in \mathrm{GSC}(d, q)$.*

▶ **Lemma 34.**

(a) $\mathrm{SC}(d) = \mathrm{GSC}(d, 1)$.

(b) *If $G \in \mathrm{GSC}(1, q)$, then G is a $(0, q)$-tiered graph.*

Let $d \ge 2$ and $q \ge 1$ and $G \in \mathrm{GSC}(d, q)$. Hence $G = \big(\dot\bigcup_{e\in F} G_e\big)^{\mathbf{S}}$, where $|\mathbf{S}| \le q$ and $G_e \in \mathrm{SC}(d-1)$ for every $e \in F$. We let F_0 be the set of $e \in F$ such that there is at most one $e' \in F$ with $e' \ne e$ and $color(V(G_{e'}), \mathbf{S}) = color(V(G_e), \mathbf{S})$. We partition $F \setminus F_0$ into sets $(F_\delta)_{\delta\in\Delta}$ such that for every $\delta \in \Delta$ there is a color $c(\delta)$ such that for all $e \in F_\delta$ we have $c(\delta) = color(V(G_e), \mathbf{S})$ and for distinct $\delta, \delta' \in \Delta$ we have $c(\delta) \ne c(\delta')$. By definition, $|\Delta| \le 2^{2^q}$ and $|F_\delta| \ge 3$ for all $\delta \in \Delta$ (note that Δ may be empty). Observe that

$$|F_0| \le 2 \cdot 2^{2^{|\mathbf{S}|}} = 2^{2^q+1}. \tag{6}$$

As $d \ge 2$, for every $e \in F_0$ the graph G_e can be written in the form $G_e = \Big(\dot\bigcup_{f\in F_e} G_{ef}\Big)^{S_e}$, where all G_{ef} are in $\mathrm{SC}(d-2)$ and $S_e \subseteq V(G_e)$. Let $\mathbf{T}$ be the sequence of all sets S_e with $e \in F_0$. By (6), $|\mathbf{T}| \le 2^{2^q+1}$. We define the graph G' by (note that F_0 may be empty)

$$G' := \Big(\dot\bigcup_{e\in F_0,\ f\in F_e} G_{ef}\Big)^{\mathbf{T}}.$$

The following statements result directly from the definitions.

▶ **Lemma 35.**
(a) $G' \in \mathrm{GSC}(d-1, 2^{2^q+1})$.
(b) $G = \Big(G' \dot\cup \dot\bigcup_{\delta\in\Delta,\ e\in F_\delta} G_e\Big)^{\mathbf{S}}$ *with* $|\Delta| \leq 2^{2^q}$, $G_e \in \mathrm{SC}(d-1)$ *for* $e \in F_\delta$ *and* $\delta \in \Delta$.

We define the function h by: $h(0,q) := 0$, $h(1,q) := q$, and $h(d+1,q) := h(d, 2^{2^q+1})$ for $d \geq 1$. Now a simple induction shows (for a proof see the full paper):

▶ **Proposition 36.** *Let* $d \geq 1$ *and* $G \in \mathrm{GSC}(d,q)$. *Then* G *is a* $(d-1, h(d,q))$*-tiered graph, i.e.,* G *can be written in the form*

$$G = \Big(\dots\Big(\big(I(T_0)\big)^{\mathbf{S}_1} \dot\cup \dot\bigcup_{\delta\in\Delta_1,\ e\in F_\delta} H_{1e}\Big)^{\mathbf{S}_2} \dot\cup \dots \dot\cup \dot\bigcup_{\delta\in\Delta_{d-1},\ e\in F_\delta} H_{d-1\ e}\Big)^{\mathbf{S}_d},$$

where $|\mathbf{S}_t| \leq h(d,q)$ *for* $t \in [d]$ *and where* $|\Delta_t| \leq h(d,q)$ *for* $t \in [d-1]$. *In addition, for* $d \geq 2$ *we have* $H_{te} \in \mathrm{SC}(t-1)$ *for* $t \in [d-1]$, $\delta \in \Delta_t$, *and* $e \in F_\delta$.

8 FO-definition of tree-models for graphs of bounded shrub-depth

Using the results of the preceding sections we first prove that there is a computable function $d \mapsto \varphi_d$ where φ_d is an FO-sentence whose class of models has bounded shrub-depth and contains $\mathrm{SC}(d)$. Then we show how using ideas from [15] we can refine this proof to obtain Theorem 2, i.e., the FO-definability of tree-models for graphs of bounded shrub-depth.

▶ **Proposition 37.** *Let* $d \in \mathbb{N}$ *and* $\Gamma := h(d,1)$. *There is an* FO*-sentence* φ_d *with (a) and (b).*
(a) *If* $G \in \mathrm{SC}(d)$, *then* $G \models \varphi_d$.
(b) *If* $G \models \varphi_d$, *then* $\mathrm{SC}(G) \leq \frac{d\cdot(d+1)\cdot\Gamma}{2}$.
For later purposes we assume that φ_d *also expresses that* E *is irreflexive and symmetric.*

Proof. We set $\varphi_0 := \forall x \forall y (x = y \wedge \neg Exy)$. Let $d \geq 1$ and G be a graph with $\mathrm{SC}(G) = d$. Hence $G \in \mathrm{GSC}(d,1)$ by Lemma 34(a). By Proposition 36 the graph G is $(d-1,\Gamma)$-tiered, thus G can be written in the form

$$G = \Big(\dots\Big(\big(I(T_0)\big)^{\mathbf{S}_1} \dot\cup \dot\bigcup_{\delta\in\Delta_1,\ e\in F_\delta} H_{1e}\Big)^{\mathbf{S}_2} \dot\cup \dots \dot\cup \dot\bigcup_{\delta\in\Delta_{d-1},\ e\in F_\delta} H_{d-1\ e}\Big)^{\mathbf{S}_d},$$

where in particular, $|\mathbf{S}_t| \leq \Gamma$ for $t \in [d]$ and $H_{te} \in \mathrm{SC}(t-1)$ for every $t \in [d-1]$, $\delta \in \Delta_t$ and $e \in F_\delta$. By Proposition 31 we can assume that for

$$L := \Big(\dots\Big(\big(I(T_0)\big)^{\mathbf{S}_1} \dot\cup I(T_1)\Big)^{\mathbf{S}_2} \dot\cup \dots \dot\cup I(T_{d-1})\Big)^{\mathbf{S}_d} = \big(I_{V(G)}\big)^{\mathbf{S}_1\dots\mathbf{S}_d}$$

$\mathbf{S}_1 \dots \mathbf{S}_d$ is unambiguous (with respect to L). Here $T_i = \dot\bigcup_{\delta\in\Delta_i,\ e\in F_\delta} V(H_{ie})$ for $i \in [d-1]$.

By Lemma 29(b) the size of voting vertices can be bounded in terms of d. Thus as φ_d we can take an FO-sentence which (existentially) guesses voting vertices for such an unambiguous $\mathbf{S}_1 \dots \mathbf{S}_d$ and guesses a subset of these vertices which togethr with their $\chi_{\mathbf{S}_1\dots\mathbf{S}_d}$-colors (which are also guessed) yield a representative system for L. Then it defines the $\mathbf{S}_1, \dots, \mathbf{S}_d$ and expresses that every connected component of $G^{\mathbf{S}_1\dots\mathbf{S}_d}$ satisfies φ_{d-1}. Now the validity of (a) should be clear.

We prove (b) by induction on d. Of course, (b) holds for $d = 0$. So assume that $d \geq 1$ and that the statement (b) is true for $d-1$. Let $G \models \varphi_d$. Then there are $\mathbf{S}_1, \dots, \mathbf{S}_d$ such that every connected component H of $G_1 := G^{\mathbf{S}_1\dots\mathbf{S}_d}$ satisfies φ_{d-1}. Hence, $\mathrm{SC}(H) \leq \frac{(d-1)\cdot d\cdot\Gamma}{2}$ by induction hypothesis. As $G = \Big(\dot\bigcup_{H \text{ connected component of } G_1} H\Big)^{\mathbf{S}_1\dots\mathbf{S}_d}$ by Lemma 16(c), we get

$$\begin{aligned}\mathrm{SC}(G) &\leq d\cdot\Gamma + \max\{\mathrm{SC}(H) \mid H \text{ connected component of } G_1\}\\ &\leq d\cdot\Gamma + \frac{(d-1)\cdot d\cdot\Gamma}{2} \leq \frac{d\cdot(d+1)\cdot\Gamma}{2}.\end{aligned}$$ ◀

We turn to a proof of Theorem 2. It suffices to show the following result (cf. Proposition 17).

▶ **Theorem 38.** *Let $d \in \mathbb{N}$. There is an* FO*-interpretation that assigns to every ordered graph $(G, <)$ with $G \in \mathrm{SC}(d)$ a tree-model.*

Proof. We know that $\mathrm{SC}(d) \subseteq \mathrm{TM}_{2^d}(d)$ (see Proposition 17(b)). We recall the proof of this result from [15, Theorem 3.6]. For an SC-derivation W witnessing $G \in \mathrm{SC}(d)$ it constructs a tree $\mathcal{T}(W) \in \mathrm{TREE}[2^d, d]$ (see page 4 for the definition of $\mathrm{TREE}[m, d]$), which together with a signature D will be a tree-model of G. We denote the labels by L_b with $b \in \{0,1\}^d$. Essentially the tree $\mathcal{T}(W)$ is the "tree of the SC-derivation": The leaves of $\mathcal{T}(W)$ are the vertices of G. Each internal node t of $\mathcal{T}(W)$ is associated with a flipping set S_t. By adding nodes with the empty flipping set we can assume that every path from the root to a leaf has length d. Let $v \in V(G)$ and let $t_0 = r, t_1, \ldots, t_d = v$ be the path from the root r of $\mathcal{T}(W)$ to v. Then v gets the color (= label) L_b if for $i \in [d]$ we have $(b_i = 1 \iff v \in S_{t_i})$. The pair $\{u, v\}$ is an edge of G if and only if u and v are simultaneously contained in an odd number of flipping sets S_t, where t ranges over all internal nodes. This can easily be determined from the colors of u and v, and from the depth of their least common ancestor $u \wedge v$. So it yields the definition of the corresponding signature $D(d)$ (note that $D(d)$ doesn't depend on the concrete SC-derivation but only on d).

Now let $<^G$ be an arbitrary order of $V(G)$. As $G \in \mathrm{SC}(d)$, the graph G is a model of the sentence φ_d of Proposition 37. The process described in φ_d shows how one gets an SC-derivation witnessing $\mathrm{SC}(G) \leq g(d) := \frac{d \cdot (d+1) \cdot \Gamma}{2}$. Using $<^G$ we can describe in FO such a derivation $W(<^G)$: According to φ_d first we guess voting vertices with certain properties, now we choose the lexicographically $<^G$-first voting vertices with these properties. Then by φ_d we guess a subset of these vertices together with their colors as representative system. Now we choose the lexicographically $<^G$-smallest such subset and the "smallest" colors which do the job. Then by φ_d we get $\mathbf{S}_1, \ldots, \mathbf{S}_d$ with $|\mathbf{S}_i| \leq \Gamma$ for $i \in [d]$. W.l.o.g. we may assume that $|\mathbf{S}_i| = \Gamma$. The root of the tree $\mathcal{T}(W(<^G))$ starts with a path of length $d \cdot \Gamma - 1$ ending with a node t_0. The nodes of the path are associated with the flipping sets in $\mathbf{S}_1, \ldots, \mathbf{S}_d$. Then for every $v \in V(G)$, which in $G^{\mathbf{S}_1 \ldots \mathbf{S}_d}$ is not in the connected component of an $<^G$-smaller element, we add an derivation for this component according to φ_{d-1} as one child of t_0. By this procedure we get a tree in $\mathcal{T}(W(<^G)) \in \mathrm{TREE}[2^{g(d)}, g(d)]$, where $g(d) := \frac{d \cdot (d+1) \cdot \Gamma}{2}$.

In this way we get an FO-interpretation $\mathcal{I}$ defining in $(G, <^G)$ with $G \models \varphi_d$ the tree-model $\mathcal{T}(W(<^G))$ of G. That is, we can present (it is tedious but straightforward) a tuple of FO-formulas $\mathcal{I} = \big(\varphi_{uni}(\bar{x}), \varphi_P(\bar{x}, \bar{y}), (\varphi_{L_b}(\bar{x}))_{b \in \{0,1\}^{g(d)}}\big)$, where $\bar{x} = x_0, \ldots, x_{g(d)}$ and $\bar{y} = y_0, \ldots, y_{g(d)}$ such that $(G, <^G)^{\mathcal{I}} := \big(\varphi_{uni}^{(G,<^G)}, \varphi_P^{(G,<^G)}, (\varphi_{L_b}^{(G,<^G)})_{b \in \{0,1\}^{g(d)}}\big)$ is isomorphic to the tree-model $\mathcal{T}(W(<^G))$. For example, $\varphi_{uni}^{(G,<^G)} := \big\{(v_0, \ldots, v_{g(d)}) \in G^{g(d)} \mid (G, <^G) \models \varphi_{uni}(\bar{v})\big\}$ is the universe of $(G, <^G)^{\mathcal{I}}$ and $\varphi_P^{(G,<^G)} := \big\{(\bar{v}, \bar{w}) \mid \bar{v}, \bar{w} \in \varphi_{uni}^{(G,<^G)}, (G, <^G) \models \varphi_P(\bar{v}, \bar{w})\big\}$ is the parent-child relation of $(G, <^G)^{\mathcal{I}}$. ◀

As already mentioned in the Introduction we get Theorem 1 from the preceding result in the same way as we did for tree-depth in [3]. For the sake of completeness let us recall that for a class K of graphs the parameterized model-checking p-MC(K, MSO) for MSO on K is defined by

Instance:	A graph $G \in K$ and an MSO-sentence φ.
Parameter:	$k \in \mathbb{N}$.
Problem:	Decide if $k = |\varphi|$ and $\mathcal{A} \models \varphi$.

In the next section we will apply a further consequence of Theorem 38, an improvement of Proposition 3, which again can be obtained as the corresponding result for tree-depth in [3]:

▶ **Proposition 39.** *Let $d \in \mathbb{N}$. There is an algorithm that assigns to every* MSO$[\{E\}]$*-sentence φ an* FO$\big[\{E,<\}\big]$*-sentence φ^+ such that for every ordered graph $(G,<)$ with $G \models \varphi_d$,*

$$G \models \varphi \iff (G,<^G) \models \varphi^+.$$

9 MSO = FO on classes of bounded shrub-depth

The goal of this section is to show that in models of φ_d every MSO$[\{E\}]$-sentence is equivalent to an FO$[\{E\}]$-sentence, i.e, that the we can omit the order relation used in Proposition 39. To get this result we use Craig's interpolation which is known to be true only if we consider finite and infinite models. However, we have introduced the notion of SC-depth for finite graphs only. So let us extend this concept to infinite graphs. *If not stated otherwise explicitly, in the following "graph" always means a finite or infinite graph.*

Let G be a graph and $S \subseteq V(G)$, then G^S is defined as in the finite case.

▶ **Definition 40.** *The class* SC(d) *(extending the "old"* SC(d)*) is defined by induction on d:*

SC(0) *is the class of graphs whose vertex set is a singleton.*

Assume that I is a set with $|I| \geq 1$ and that for $i \in I$ the graphs G_i are in SC(d) *and have pairwise disjoint vertex sets. Then $\big(\dot\bigcup_{i\in I} G_i\big)^S \in \mathrm{SC}(d+1)$ for every $S \subseteq \bigcup_{i\in I} V(G_i)$.*

The SC-depth *of a graph is the least $d \in \mathbb{N}$ such that $G \in \mathrm{SC}(d)$.*

Not every graph has an SC-depth, that is, the analogue of Lemma 15 fails. For example, an infinite path has no SC-depth. We leave it to the reader to generalize the notion of shrub-depth and to realize that Lemma 8, the analogue of Lemma 15 for shrub-depth, fails. But all other results in Section 3–Section 7 and Proposition 37 of Section 8 are true for graphs in the way stated (or with obvious changes). One exception: In the definition of tiered graph we have to require that the H_{te}'s have an SC-depth.

Let us look what happens with Theorem 38 of Section 8. On page 13 for the first time we considered orders on graphs. Once we have an order $<^G$ in a finite model G of the sentence φ_d we got a canonical SC-derivation $W(<^G)$ of G, which could be described by an interpretation $\mathcal{I}$. When defining the derivation $W(<^G)$ we used a few times the property that every nonempty subset of $V(G)$ contains a $<^G$-least element or the same property for subsets of $V(G)^r$ for some $r \geq 1$ with respect to the lexicographic order $<^G_{\mathrm{lex}}$ of r-tuples induced by $<^G$. All these subsets were definable by an FO$[\{E\}]$-formula $\psi(\bar x, \bar y)$, where $|\bar x| = r$ and the variables in $\bar y$ are parameters. Then we used the fact that the following sentence is true in every *finite* $(G,<^G)$:

$$\textit{least-element}(\psi) := \forall \bar y\Big(\exists \bar x \psi(\bar x,\bar y) \to \exists \bar x\big(\psi(\bar x,\bar y) \wedge \forall \bar x'(\psi(\bar x',\bar y) \to \bar x \leq_{\mathrm{lex}} \bar x')\big)\Big).$$

Here $|\bar x'| = |\bar x|$. Let

$$\mathrm{LE}(<) := \big\{\textit{least-element}(\psi) \mid \psi = \psi(\bar x,\bar y) \in \mathrm{FO}[\{E\}] \text{ and } |\bar x| \geq 1\big\}$$

be the set of least-element sentences for all FO$[\{E\}]$-formulas. Furthermore set

$$\mathrm{LE}^*(<) := \mathrm{LE}(<) \cup \big\{\text{"}E \text{ is irreflexive and symmetric"}\big\} \cup \big\{\text{"}< \text{ is an order"}\big\}.$$

Then we can reformulate Proposition 39 and extend it to arbitrary graphs:

▶ **Proposition 41.** *For every* MSO$[\{E\}]$*-sentence φ there exists an* FO$[\{E,<\}]$*-sentence φ^+ such that* $\mathrm{LE}^*(<) \cup \{\varphi_d\} \models (\varphi \leftrightarrow \varphi^+)$.

Now we turn to the main result of this section, which extends Theorem 4.

▶ **Theorem 42.** *Let $d \geq 1$. In models of φ_d (hence, in particular, in graphs in* $\mathrm{SC}(d)$*) every* $\mathrm{MSO}[\{E\}]$*-sentence φ is equivalent to an* $\mathrm{FO}[\{E\}]$*-sentence ψ. Moreover, there is an algorithm that on input φ yields ψ.* Proof: full paper.

▶ **Corollary 43.**

(a) *There is a computable function $d \mapsto \psi_d$, where $\psi_d \in \mathrm{FO}[\{E\}]$ axiomatizes* $\mathrm{SC}(d)$*.*

(b) *There is a computable function $(m, d) \mapsto \psi_{m,d}$, where $\psi_{m,d} \in \mathrm{FO}[\{E\}]$ axiomatizes* $\mathrm{TM}_m(d)$*.*

Proof.

(a) We know from Proposition 18 that there is a computable function $d \mapsto \rho_d$, where $\rho_d \in \mathrm{MSO}[\{E\}]$ axiomatizes $\mathrm{SC}(d)$. By the preceding theorem we effectively get a $\psi^d \in \mathrm{FO}[\{E\}]$ equivalent to ρ_d in models of φ_d. Then $\psi_d := \varphi_d \wedge \psi^d$ axiomatizes $\mathrm{SC}(d)$.

(b) As by Proposition 17 (a) we have $\mathrm{TM}_m(d) \subseteq \mathrm{SC}(d \cdot m \cdot (m+1))$, we can argue for $\mathrm{TM}_m(d)$ similarly, now using Proposition 12. ◀

10 The excursion to the infinite yields further results

The following result is an effective version of the result [15, Corollary 5.6] mentioned on page 5 and on page 7.

▶ **Theorem 44.** *There is an algorithm that applied to (m, d) eventually stops and outputs a finite set $F_{m,d}$ of finite graphs such that a graph is in* $\mathrm{TM}_m(d)$ *if and only if it excludes the graphs in $F_{m,d}$ as induced subgraphs. The anloguous result holds for* $\mathrm{SC}(d)$*.*

Proof. As $\mathrm{TM}_m(d)$ is closed under induced subgraphs, by the Łoś-Tarski Theorem of classical model theory (cf. [2, 18]) we effectively find (from $\psi_{m,d}$ of Corollary 43) a universal FO-sentence $\nu_{m,d}$ axiomatizing $\mathrm{TM}_m(d)$ (recall that a universal FO-sentence is a sentence of the form $\forall x_1 \ldots \forall x_n \chi$ with quantifier-free χ). Every universal sentence just expresses that there is a finite set of finite graphs that are excluded as induced subgraphs. ◀

Various applications of the same flavour may be obtained using the following lemma. We will prove this lemma and present various applications in the full version of this paper.

▶ **Lemma 45.** *Let $d \geq 1$ and $K \subseteq \mathrm{SC}(d)$ be a class closed under induced subgraphs. For every* MSO*-sentence φ, if the class of finite models of φ in K is closed under induced subgraphs, so is the class of models of φ in K.*

For the class of finite graphs Rossman [21] has proved the analogue of the result of classical model theory that a sentence preserved under homomorphisms is equivalent to an existential-positive FO-sentence (a sentence is *positive* if it does not contain the negation symbol). Along the previous lines one can show that this preservation theorem holds for $\mathrm{TM}_m(d)$: A sentence preserved under homomorphisms between finite graphs in $\mathrm{TM}_m(d)$ is equivalent to an existential-positive FO-sentence.

References

1 D. A. Mix Barrington, N. Immerman, and H. Straubing. On Uniformity within NC^1. *Journal of Computer and System Sciences*, 41(3):274–306, 1990.

2 C. C. Chang and H. J. Keisler. *Model Theory*. Studies in Logic and the Foundations of Mathematics. Elsevier, 3rd edition, 1992.

3 Y. Chen and J. Flum. Tree-depth, quantifier elimination, and quantifier rank. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK, July 09-12, 2018*, pages 225–234, 2018.

4 B. Courcelle. Graph Rewriting: An Algebraic and Logic Approach. In *Handbook of Theoretical Computer Science, Volume B:*, pages 193–242. Elsevier and MIT Press, 1990.

5 B. Courcelle, J. Makowsky, and U. Rotics. Linear Time Solvable Optimization Problems on Graphs of Bounded Clique-Width. *Theory of Computing Systems*, 33(2):125–150, 2000.

6 W. Craig. Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory. *Journal of Symbolic Logic*, 22(3), 1957.

7 K. Eickmeyer, J. van den Heuvel, K. Kawarabayashi, S. Kreutzer, P. Ossona de Mendez, M. Pilipczuk, D. Quiroz, R. Rabinovich, and S. Siebertz. Model-Checking on Ordered Structures. *CoRR*, abs/1812.08003, 2018.

8 M. Elberfeld, M. Grohe, and T. Tantau. Where First-Order and Monadic Second-Order Logic Coincide. *ACM Transactions on Computational Logic*, 17(4):25:1–25:18, 2016.

9 M. Frick and M. Grohe. The complexity of first-order and monadic second-order logic revisited. *Annals of Pure and Applied Logic*, 130(1-3):3–31, 2004.

10 J. Gajarský and P. Hliněný. Faster Deciding MSO Properties of Trees of Fixed Height, and Some Consequences. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2012, December 15-17, 2012, Hyderabad, India*, pages 112–123, 2012.

11 J. Gajarský and P. Hliněný. Kernelizing MSO Properties of Trees of Fixed Height, and Some Consequences. *Logical Methods in Computer Science*, 11(1), 2015.

12 J. Gajarský, P. Hliněný, D. Lokshtanov, J. Obdrzálek, S. Ordyniak, M. S. Ramanujan, and S. Saurabh. FO model checking on posets of bounded width. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 963–974, 2015.

13 J. Gajarský, P. Hliněný, J. Obdrzálek, D. Lokshtanov, and M. S. Ramanujan. A New Perspective on FO Model Checking of Dense Graph Classes. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS'16, New York, NY, USA, July 5-8, 2016*, pages 176–184, 2016.

14 R. Ganian, P. Hliněný, D. Král, J. Obdrzálek, J. Schwartz, and J. Teska. FO model checking of interval graphs. *Logical Methods in Computer Science*, 11(4), 2015.

15 R. Ganian, P. Hliněný, 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 M. Grohe, S. Kreutzer, R. Rabinovich, S. Siebertz, and K. Stavropoulos. Coloring and Covering Nowhere Dense Graphs. *SIAM Journal on Discrete Mathematics*, 32(4):2467–2481, 2018.

17 Y. Gurevich. Toward logic tailored for computational complexity. In *Computation and Proof Theory, Lecture Notes in Mathematics*, pages 1104:175–216, 1984.

18 W. Hodges. *Model theory*, volume 42 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, 1993.

19 M. Lampis. Algorithmic Meta-theorems for Restrictions of Treewidth. *Algorithmica*, 64(1):19–37, 2012.

20 M. Pilipczuk, S. Siebertz, and S. Toruńczyk. Parameterized circuit complexity of model-checking on sparse structures. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK, July 09-12, 2018*, pages 789–798, 2018.

21 B. Rossman. Homomorphism preservation theorems. *Journal of the ACM*, 55(3):15:1–15:53, 2008.