

Expressivity and Succinctness of Order-Invariant Logics on Depth-Bounded Structures

Kord Eickmeyer¹, Michael Elberfeld², and Frederik Harwath³

¹ TU Darmstadt, Darmstadt, Germany
eickmeyer@mathematik.tu-darmstadt.de

² RWTH Aachen University, Aachen, Germany
elberfeld@informatik.rwth-aachen.de

³ Goethe-Universität, Frankfurt am Main, Germany
harwath@cs.uni-frankfurt.de

Abstract. We study the expressive power and succinctness of order-invariant sentences of first-order (FO) and monadic second-order (MSO) logic on graphs of bounded tree-depth. Order-invariance is undecidable in general and, therefore, in finite model theory, one strives for logics with a decidable syntax that have the same expressive power as order-invariant sentences. We show that on graphs of bounded tree-depth, order-invariant FO has the same expressive power as FO, and order-invariant MSO has the same expressive power as the extension of FO with modulo-counting quantifiers. Our proof techniques allow for a fine-grained analysis of the succinctness of these translations. We show that for every order-invariant FO sentence there exists an FO sentence whose size is elementary in the size of the original sentence, and whose number of quantifier alternations is linear in the tree-depth. Our techniques can be adapted to obtain a similar quantitative variant of a known result that the expressive power of MSO and FO coincides on graphs of bounded tree-depth.

Keywords: *expressivity, succinctness, first-order logic, monadic second-order logic, order-invariance, tree-depth*

1 Introduction

Understanding the *expressivity* of logics on finite structures—the question of which properties are definable in a certain logic—plays an important role in database and complexity theory. In the former, logics are used to formulate queries; in the latter, they describe computational problems. Moreover, besides just studying a logic’s expressivity, understanding its *succinctness*—the question how complex definitions of properties like queries and problems must be—is a requirement towards (theoretical) expressivity results of (potential) practical importance. The present work studies the succinctness of first-order logic (FO) as well as its succinctness compared to extensions allowing for the use of a linear

$\varphi \in$	$\leq\text{-inv-FO}$	MSO	$\leq\text{-inv-MSO}$
$\psi \in$	FO	FO	FO+MOD
$\ \psi\ $	$(2d+1)\text{-EXP}(q)$	$\mathcal{O}(d^2)\text{-EXP}(\text{qr}(\varphi))$	non-elementary
$\text{qad}(\psi)$	$\mathcal{O}(d)$	$\mathcal{O}(d)$	$\mathcal{O}(d)$

Table 1. Summary of our results: A formula φ of quantifier rank q is translated into a formula ψ that is equivalent to φ on graphs of tree-depth at most d .

order and set quantifiers. This extends and refines recent studies on the expressivity of these logics [1,6] on restricted classes of structures. The structures we consider have bounded tree-depth, which is a graph invariant that measures how far a graph is from being a star in a similar way as tree-width measures how far a graph is from being a tree. Our results are summarised in Table 1.

In both database and complexity theory, one often assumes that structures come with a linear order and formulae are allowed to use this order as long as the properties defined by them do not depend on the concrete interpretation of the order in a structure. Such formulae are called *order-invariant*. Since testing order-invariance for given FO-formulae is undecidable in general, one tries to find logics that have the same expressive power as order-invariant formulae, but a decidable syntax. Several examples prove that order-invariant FO-formulae ($\leq\text{-inv-FO}$) are more expressive than FO-formulae without access to orders, cf. [11]. A common feature of these separating examples is that their Gaifman graphs contain large cliques, making them rather complicated from the point of view of graph structure theory. For tree structures, on the other hand, Benedikt and Segoufin [1] showed that the expressivity of FO and $\leq\text{-inv-FO}$ coincide. We extend and refine this result by showing equal expressivity and succinctness results for FO and $\leq\text{-inv-FO}$ on graphs of bounded tree-depth. The importance of the expressivity result is highlighted by the fact (proved in the full version of this paper) that order-invariance is undecidable even on graphs of tree-depth at most 2.

A logic that is commonly studied from the perspectives of algorithm design and language theory is monadic second-order logic (MSO), which extends FO-formulae by the ability to quantify over sets of elements instead of just single elements. While it has a rich expressivity that exceeds that of FO already on word structures, the expressive powers of FO and MSO coincide on any class of structures whose tree-depth is bounded [6]. We refine this by presenting a translation into succinct FO-formulae.

In [4], Courcelle raised the (still open) question whether $\leq\text{-inv-MSO}$ has the same expressive power on graphs of bounded *tree-width* as the extension of MSO by first-order modulo-counting quantifiers (CMSO). We prove a stronger statement for graphs of bounded tree-depth: $\leq\text{-inv-MSO}$ has the same expressive power as FO+MOD, i.e. the extension of FO by arbitrary first-order modulo-counting quantifiers.

Our results also have implications on FO itself. They imply that that the quantifier alternation hierarchy for FO of Chandra and Harel [3] collapses on graphs of bounded tree-depth, whereas by their result it is strict on trees of

unbounded height. That means, for graphs of bounded tree-depth, we are able to turn any FO-formula into a formula whose size is bounded by the quantifier depth of the original formula and whose quantifier alternation depth is bounded by a linear function in the tree-depth.

Proof techniques and relation to prior works. Our proofs are based on fundamental techniques from finite model theory like interpretation arguments, logical types, and games. Compared to prior works like [6], we enrich the application of these techniques by a quantitative analysis, which results in succinct translations instead of just equal expressivity results.

The proofs of [6] use an involved constructive variant of the Feferman–Vaught composition theorem, which complicates a straightforward analysis of the formula size in the translation from MSO to FO. We also use composition arguments, but we get along with an easier non-constructive variant. There is another proof of the result of [6] in [7], but it relies on involved combinatorial insights that seem unsuited for both a tight analysis of succinctness as well as an adaptation to the ordered setting.

While our proofs are based on techniques from finite model theory, the results of [1] about the expressivity of \leq -inv-FO on trees use automata-theoretic and algebraic methods, which seem unsuited to obtain succinct translations. Due to the following reason, even our equal expressivity result for \leq -inv-FO and FO on graphs of bounded tree-depth is interesting: Benedikt and Segoufin [1] proved that on graphs of bounded tree-width every \leq -inv-FO sentence is equivalent to an MSO sentence. With the results of [6] this would imply our expressivity result (not the succinctness result). However, the result from [1] relies on an earlier proof of how to define tree decompositions of bounded width in MSO whose correctness has been doubted by Courcelle and Engelfriet [5]. A proof of our expressivity result for \leq -inv-FO along this lines is nevertheless possible. A direct FO-construction of a tree decomposition of bounded width and depth for graphs of bounded tree-depth will appear in the full version of this paper.

Organisation of this paper. The paper continues with a background section and, then, the results related to \leq -inv-FO, MSO, and \leq -inv-MSO are proved in Sections 3, 4, and 5, respectively. Due to space restrictions, some parts of the paper, such as proofs and definitions, are only sketched or omitted.

2 Background

General notation. The sets of natural numbers with and without 0 are denoted respectively by \mathbb{N} and \mathbb{N}^+ . Let $[i, j] := \{i, \dots, j\}$ for all $i, j \in \mathbb{N}$ with $i \leq j$, and let $[j] := [1, j]$. For any $d \in \mathbb{N}$, the *class of functions that grow at most d -fold exponentially*, denoted by $d\text{-EXP}(n)$, is made up by all functions $f: \mathbb{N} \rightarrow \mathbb{N}$ with $f(n) \leq c + d\text{-exp}(n^c)$ for some $c \in \mathbb{N}$; where $d\text{-exp}: \mathbb{N} \rightarrow \mathbb{N}$ is recursively defined via $0\text{-exp}(n) := n$ and $(d + 1)\text{-exp}(n) := 2^{d\text{-exp}(n)}$ for $d \in \mathbb{N}$. If we say that a relation is an *order*, we implicitly assume that it is *linear*.

Logic. For a reference on notation and standard methods in finite model theory, we refer to the book of Libkin [9]. Besides the standard logics FO and MSO, we also consider the logic FO+MOD that is obtained from FO by allowing the use of *modulo-counting quantifiers* $\exists^{i \pmod p}$ for each $i \in \mathbb{N}, p \in \mathbb{N}^+$, where $\psi(\bar{y}) := \exists^{i \pmod p} x \varphi(x, \bar{y})$ means that for a structure \mathfrak{A} with universe A and a tuple of its elements \bar{a} , $\mathfrak{A} \models \psi(\bar{a})$ iff $|\{b \in A : \mathfrak{A} \models \varphi(b, \bar{a})\}| \equiv i \pmod p$.

We write $\text{qr}(\varphi)$ for the *quantifier rank* and $\|\varphi\|$ for the *size* (or *length*) of a formula φ . The *quantifier alternation depth* $\text{qad}(\varphi)$ of a formula φ in *negation normal form* (NNF, i.e. all negations of φ occur directly in front of atomic formulae) is the maximum number of alternations between \exists - and \forall -quantifiers on all directed paths in the syntax tree of φ . If φ is not in NNF, we first find an equivalent formula φ' in NNF using a fixed conversion procedure and, then, define $\text{qad}(\varphi) := \text{qad}(\varphi')$.

For any logic $L \in \{\text{FO}, \text{FO+MOD}, \text{MSO}\}$, we write $\mathfrak{A} \equiv_q^L \mathfrak{B}$ for $q \in \mathbb{N}$ to denote that structures \mathfrak{A} and \mathfrak{B} over the same signature σ satisfy the same $L[\sigma]$ -sentences of quantifier rank at most q . The \equiv_q^L -equivalence class of \mathfrak{A} is its (L, q) -*type* and denoted by $\text{tp}_q^L(\mathfrak{A})$. For $L \in \{\text{FO}, \text{MSO}\}$, each (L, q) -type τ is definable by an L -sentence φ_τ with $\text{qr}(\varphi_\tau) = q$, i.e. $\mathfrak{A} \models \varphi_\tau$ iff $\text{tp}_q^L(\mathfrak{A}) = \tau$; we identify each τ with one such sentence φ_τ . If the logic L has been fixed (as will be the case in most parts of this paper) or the concrete logic is not important for the discussion, we omit it in this and similar notation.

For every signature σ , we define the signature $\sigma^{\leq} := \sigma \cup \{\leq\}$, where \leq is a binary relation symbol. A sentence $\varphi \in \text{FO}[\sigma^{\leq}]$ is *order-invariant* exactly if the following holds for all finite σ -structures G and all linear orders \preceq, \preceq' on the universe of G :

$$(G, \preceq) \models \varphi \quad \text{iff} \quad (G, \preceq') \models \varphi.$$

The set of all order-invariant $\varphi \in \text{FO}[\sigma^{\leq}]$ is denoted by $\leq\text{-inv-FO}[\sigma]$, and for such a φ and a σ -structure G we write $G \models_{\leq} \varphi$ if $(G, \preceq) \models \varphi$ for some (equivalently, for every) linear order \preceq on G ; $\leq\text{-inv-MSO}$ is defined in the same way by using MSO instead of FO-formulae.

If ψ is a formula with a free variable z and φ is an arbitrary formula, then $\varphi|_{\psi}$ is the formula φ *relativised to* ψ . We construct $\varphi|_{\psi}$ by replacing subformulae $\exists z \varphi$ and $\forall z \varphi$ by $\exists z (\psi \wedge \varphi|_{\psi})$ and $\forall z (\neg \psi \vee \varphi|_{\psi})$, respectively.

Coloured and ordered graphs. The letter C will be used to denote a finite set of colours, and we define the signature $\sigma_C := \{E\} \cup \{P_c \mid c \in C\}$, where E is binary and every P_c is unary. A C -*coloured graph* is a σ_C -structure G with universe $V(G)$, symmetric and irreflexive edge relation $E(G)$, and such that the $P_c(G)$ form a partition of $V(G)$. We will simply speak of graphs when referring to C -coloured graphs, and write FO for $\text{FO}[\sigma_C]$ etc. An order on a graph is an order on its vertex set. An *ordered graph* is a σ_C^{\leq} -structure (G, \leq^G) where G is a graph and \leq^G is an order on G .

The *restriction* of a binary relation R on a set M to a subset $N \subseteq M$ is the relation $R|_N := \{(x, y) \in R : x, y \in N\}$. For ease of notation we will sometimes

drop the relativisation for orders on subgraphs and write (H, \preceq) for $(H, \preceq|_H)$. For two linear orders \preceq and \preceq' on disjoint sets M_1 and M_2 , respectively, we define a linear order $\preceq \cdot \preceq'$ on $M_1 \cup M_2$, the *concatenation* of \preceq and \preceq' , as $\preceq \cup \preceq' \cup (M_1 \times M_2)$.

Our formulae often speak about the distance between two vertices of a graph. To this end, we define existential FO-formulae $\varphi_{\text{dist} \leq \ell}(x, y)$ by $\varphi_{\text{dist} \leq 0}(x, y) := x = y$ and $\varphi_{\text{dist} \leq \ell}(x, y) := \exists z (\varphi_{\text{dist} \leq \ell - 1}(x, z) \wedge (Ezy \vee z = y))$ for each $\ell \geq \mathbb{N}^+$.

Tree-depth. The following inductive definition is one of several equivalent ways to define the *tree-depth* $\text{td}(G)$ of a graph (see [10] for a reference on tree-depth):

$$\text{td}(G) := \begin{cases} 1 & \text{if } |V(G)| = 1 \\ 1 + \min_{r \in V(G)} \text{td}(G \setminus r) & \text{if } G \text{ is connected and } |V(G)| > 1 \\ \max_{i \in [n]} \text{td}(K_i) & \text{if } G \text{ has components } K_1, \dots, K_n. \end{cases}$$

As an immediate consequence of this definition, each connected graph with $\text{td}(G) > 1$ contains a vertex r with $\text{td}(G \setminus r) = \text{td}(G) - 1$. We denote the set of all such vertices by $\text{roots}(G)$. Elements of $\text{roots}(G)$ are called *tree-depth roots* of G . Furthermore, graphs of tree-depth 1 contain only isolated vertices. Another fact about tree-depth that we need is that there are only paths of length at most 2^d in graphs G with $\text{td}(G) \leq d$. In particular, the diameter of such graphs is bounded by 2^d and hence the formula $\text{reach}_d(x, y) := \varphi_{\text{dist} \leq 2^d}(x, y)$ defines the relation containing all pairs $(u, v) \in V(G) \times V(G)$ such that u and v belong to the same (connected) component. Using this observation and the inductive definition of tree-depth, one can write down an FO-sentence $\varphi_{\text{td} \leq d}$ with $\|\varphi_{\text{td} \leq d}\| \in \mathcal{O}(d)$ that defines the class of graphs of tree-depth at most d on the class of all graphs, and an FO-formula $\varphi_{d\text{-roots}}(x)$ with $\|\varphi_{d\text{-roots}}\| \in \mathcal{O}(d)$ that defines the set $\text{roots}(G)$ in a connected graph G with $1 < \text{td}(G) \leq d$.

3 Order-invariant first-order logic

We prove the following theorem in the present section.

Theorem 3.1. *For every $d \in \mathbb{N}^+$ and \leq -inv-FO-sentence φ with $\text{qr}(\varphi) = q$, there is an FO-sentence ψ with $\|\psi\| \in (2d + 1)\text{-EXP}(q)$ and $\text{qad}(\psi) \in \mathcal{O}(d)$ that is equivalent to φ on (coloured) graphs of tree-depth at most d .*

Several definitions and lemmas of this section are given in greater generality than needed here, because we will reuse them in later sections. In this section, whenever notation refer to a logic L and we omit it, assume that $L = \text{FO}$. The main ingredient for the proof of Theorem 3.1 is the following lemma which states that q -types of q -ordered graphs of tree-depth at most d , i.e. ordered graphs where the order is a q -order, which we define below, can be defined by FO-formulae without referring to a linear order. Let $\mathcal{T}_{C, q, d}$ denote the set of all q -types τ over the signature $\sigma_C \leq$ such that there exists a q -ordered graph (G, \preceq) with $\text{td}(G) \leq d$ and $\text{tp}_q(G, \preceq) = \tau$.

Lemma 3.2. *For all $q, d \in \mathbb{N}^+$ and $\tau \in \mathcal{T}_{C, q, d}$, there is an FO-sentence $\varphi_{\tau, d}$ with $\|\varphi_{\tau, d}\| \in (2d)\text{-EXP}(q)$ and $\text{qad}(\varphi_{\tau, d}) \in \mathcal{O}(d)$ that defines τ on graphs of tree-depth at most d .*

Here an FO-sentence φ_{τ} defines τ on graphs of tree-depth at most d if for each graph G with $\text{td}(G) \leq d$, we have $G \models \varphi_{\tau}$ iff there exists a q -order \preceq such that $\text{tp}_q(G, \preceq) = \tau$.

Before we discuss how to prove Lemma 3.2, let us first sketch how Theorem 3.1 can be proved with its help: For a given \leq -inv-FO-sentence φ with $\text{qr}(\varphi) = q$, we let ψ be the disjunction over all FO-sentences $\varphi_{\tau, d}$ for $\tau \in \mathcal{T}_{C, q, d}$ that are types of q -ordered graphs of tree-depth at most d satisfying φ . We have $\|\psi\| \in (2d + 1)\text{-EXP}(q)$ and $\text{qad}(\psi) \in \mathcal{O}(d)$; since φ is order-invariant, ψ is equivalent to φ .

Encoding vertex information in extended colourings. During our proofs, we remove single vertices from a graph and encode information about them into colours of the remaining vertices. This allows us to recover the original graph using an FO-interpretation. Let $C' := C \times \{0, 1\}$. For a C -coloured graph G and $r \in V(G)$, define a C' -colouring of $G \setminus r$ by assigning to each vertex $v \in V(G \setminus r)$ of colour c in G the colour $(c, 1)$ if $\{r, v\} \in E(G)$, and $(c, 0)$, otherwise. The C' -coloured graph thus obtained is denoted by $G^{[r]}$. The following lemma is easy to prove following this definition.

Lemma 3.3. *Let $\mathbf{L} \in \{\text{FO}, \text{FO}+\text{MOD}\}$. For every $\mathbf{L}[\sigma_{C'}]$ -sentence φ there is an $\mathbf{L}[\sigma_C]$ -formula $\mathcal{I}(\varphi)(x)$ of the same quantifier rank and quantifier alternation depth such that*

$$G \models \mathcal{I}(\varphi)(r) \quad \text{iff} \quad G^{[r]} \models \varphi.$$

for all C -coloured graphs G and $r \in V(G)$.

Definition of q -orders. We fix orders $\preceq_{\mathbf{L}, q}$, for any logic \mathbf{L} , and \preceq_C on, respectively, the set of (\mathbf{L}, q) -types and any colour set C .

Definition 3.4 ((\mathbf{L}, q)-order). *An order \preceq of a graph G is an (\mathbf{L}, q) -order if the following conditions are satisfied:*

1. *If G is a connected graph, then it contains either only one vertex, or it contains more than one vertex and the \preceq -least element r is an element of $\text{roots}(G)$ whose colour is \preceq_C -minimal among the elements of $\text{roots}(G)$, and $\text{tp}_q^{\mathbf{L}}(G^{[r]}, \preceq) \preceq_{(\mathbf{L}, q)} \text{tp}_q^{\mathbf{L}}(G^{[r']}, \preceq)$ for all $r' \in \text{roots}(G)$ of the same colour. Furthermore, $\preceq|_{V(G \setminus r)}$ is an (\mathbf{L}, q) -order of $G \setminus r$.*
2. *Otherwise, if G has components H_1, \dots, H_ℓ , then, after suitably permuting the components, $\preceq = \preceq|_{H_1} \cdots \preceq|_{H_\ell}$, where each $\preceq|_{H_i}$ is an (\mathbf{L}, q) -order of H_i , and $\text{tp}_q^{\mathbf{L}}(H_i, \preceq) \preceq_{\mathbf{L}, q} \text{tp}_q^{\mathbf{L}}(H_j, \preceq)$ for $i \leq j$.*

The least element of a q -order \preceq is denoted by r_{\preceq} .

For each q -ordered C -coloured connected graph (G, \preceq) with $\text{td}(G) > 1$, we define an ordered C' -coloured graph

$$\tilde{G}_{\preceq} := (G^{[r_{\preceq}]}, \preceq).$$

Observe that $\text{td}(G^{[r_{\preceq}]}) < \text{td}(G)$ and that \tilde{G}_{\preceq} is q -ordered. The following lemma, which states that \tilde{G}_{\preceq} together with the colour of r_{\preceq} determine the q -type of (G, \preceq) , can be proved using standard EF-game-based arguments.

Lemma 3.5. *Let $\mathbb{L} \in \{\text{FO}, \text{MSO}\}$ and $q \in \mathbb{N}^+$. Let (G, \preceq_G) and (H, \preceq_H) be (\mathbb{L}, q) -ordered connected graphs such that $\text{td}(G), \text{td}(H) > 1$ and $r_{\preceq_G}, r_{\preceq_H}$ have the same colour. Then $\tilde{G}_{\preceq_G} \equiv_q^{\mathbb{L}} \tilde{H}_{\preceq_H}$ implies $(G, \preceq_G) \equiv_q^{\mathbb{L}} (H, \preceq_H)$.*

Using this, we can show that while there might be several q -orders of a given graph, up to $\equiv_q^{\mathbb{L}}$ they are all equivalent.

Lemma 3.6. *Let $\mathbb{L} \in \{\text{FO}, \text{MSO}\}$, $q \in \mathbb{N}^+$. For all (\mathbb{L}, q) -orders \preceq, \preceq' of a graph G , we have $(G, \preceq) \equiv_q^{\mathbb{L}} (G, \preceq')$.*

Threshold counting of components. We define an equivalence relation $\approx_{q,t}$ on ordered graphs that counts the number of components of different q -types up to a threshold value t . We show that there is a t depending on q , such that each \equiv_q^{FO} -equivalence class of q -ordered graphs is a union of $\approx_{q,t}$ -equivalence classes. Then we show, basically, that these equivalence classes are definable for graphs of bounded tree-depth. For every logic \mathbb{L} and \mathbb{L} -sentence φ , we let $n_{\varphi}(G)$ denote the number of components K of G such that $K \models \varphi$ and we let $n_{\varphi,t}(G) := \min\{n_{\varphi}(G), t\}$, for each $t \in \mathbb{N}$.

Definition 3.7 ($\approx_{\Phi,t}, \approx_{\mathbb{L},q,t}$). *Let Φ be a set of \mathbb{L} -sentences and $t \in \mathbb{N}$. We say that two graphs G and H are (Φ, t) -similar (written $G \approx_{\Phi,t} H$) if*

$$n_{\varphi,t}(G) = n_{\varphi,t}(H)$$

for each $\varphi \in \Phi$. In the special case that Φ is a set of \mathbb{L} -sentences containing one sentence that defines τ for each (\mathbb{L}, q) -type τ , we write $\approx_{\mathbb{L},q,t}$ instead of $\approx_{\Phi,t}$; whenever \mathbb{L} is fixed, we write $\approx_{q,t}$.

All these definitions are extended to ordered graphs by stipulating that a *component* of an ordered graph (G, \preceq) is an ordered graph (K, \preceq) where K is a component of G .

We show that in q -ordered graphs FO inherits its component counting capabilities from its capability to distinguish linear orders of different length. A proof of the lemma (based on different notation) is contained in the proof of [1, Thm. 5.5]; it requires only the fact that the components of a q -ordered graph are ordered according to their q -type.

Lemma 3.8. *Let $q \in \mathbb{N}^+$ and let $t := 2^q + 1$. If (G, \preceq_G) and (H, \preceq_H) are q -ordered graphs with $(G, \preceq_G) \approx_{q,t} (H, \preceq_H)$, then $(G, \preceq_G) \equiv_q^{\text{FO}} (H, \preceq_H)$.*

The following lemma shows that $\approx_{\Phi,t}$ -equivalence classes are definable for graphs of bounded tree-depth. It will be needed for the formula construction in the proof of Lemma 3.2 and in later sections.

Lemma 3.9. *Let $\mathbf{L} \in \{\text{FO}, \text{FO}+\text{MOD}\}$. For every $d, t \in \mathbb{N}^+$, set of \mathbf{L} -sentences $\Phi := \{\varphi_1, \dots, \varphi_\ell\}$, and $\bar{n} := (n_1, \dots, n_\ell) \in [0, t]^\ell$, there is an \mathbf{L} -sentence $\psi_{\bar{n},t}^\Phi$ such that for each graph G with $\text{td}(G) \leq d$, we have $G \models \psi_{\bar{n},t}^\Phi$ iff $n_{\varphi_i,t}(G) = n_i$ for each $i \in [\ell]$. Moreover, the sentence has size $\|\psi_{\bar{n},t}^\Phi\| \in \ell \cdot \mathcal{O}(d \max_{i \in [\ell]} n_i^2 \|\varphi_i\|)$ and $\text{qad}(\psi_{\bar{n},t}^\Phi) \leq \max_{i \in [\ell]} \text{qad}(\varphi_i) + 1$,*

Finally, we can proof our main lemma.

Proof of Lemma 3.2. The proof proceeds by induction on the tree-depth d . Let $\mathcal{T}_{C,q,d}^{\text{conn}}$ be defined analogously to $\mathcal{T}_{C,q,d}$ for q -ordered connected graphs.

Case 1: Connected graphs. As a first step, we prove the special case of the claim for connected graphs with a stronger upper bound on the formula size, i.e. we show that, on connected graphs of tree-depth at most d , each $\tau \in \mathcal{T}_{C,q,d}^{\text{conn}}$ is defined by an FO-sentence $\varphi_{\tau,d}^{\text{conn}}$ such that $\|\varphi_{\tau,d}^{\text{conn}}\| \in (2(d-1) + 1)\text{-EXP}(q)$ and $\text{qad}(\varphi_{\tau,d}^{\text{conn}}) \in \mathcal{O}(d)$. If $d = 1$, then any graph G of type τ consists of a single vertex of some colour $c \in C$; the FO-sentence $\varphi_{\tau,1}^{\text{conn}} := \exists x P_c(x) \wedge \forall y (x = y)$ defines τ since there is only one linear order on each such graph. Hence $\|\varphi_{\tau,1}^{\text{conn}}\|$ and $\text{qad}(\varphi_{\tau,1}^{\text{conn}})$ are constant.

Now suppose that $d > 1$ and $\tau \in \mathcal{T}_{C,q,d}^{\text{conn}}$. For each colour \hat{c} we define a set $R_{\hat{c}} \subseteq \mathcal{T}_{C',q,d-1}$ that contains a q -type θ iff $\text{tp}_q(H, \preceq) = \tau$ for a q -ordered C' -coloured connected graph (H, \preceq) with $1 < \text{td}(H) \leq d$ such that $\text{tp}_q(\tilde{H}_{\preceq}) = \theta$ and r_{\preceq} has colour \hat{c} . We obtain an FO[$\sigma_{C'}$]-sentence $\varphi_{\theta,d-1}$ by induction that defines θ on q -ordered C' -coloured graphs of tree-depth at most $d-1$, and that has size $\|\varphi_{\theta,d-1}\| \in (2(d-1))\text{-EXP}(q)$ and alternation-depth $\text{qad}(\varphi_{\theta,d-1}) \in \mathcal{O}(d)$. Let $\varphi_{\tau,1}$ be an FO-sentence with $\|\varphi_{\tau,1}\| \in 2\text{-EXP}(q)$ and $\text{qad}(\varphi_{\tau,1}) \in \mathcal{O}(d)$, also given by induction, that defines τ on graphs of tree-depth 1. Now consider the following FO-sentence

$$\varphi_{\tau,d}^{\text{conn}} := (\varphi_{\text{td} \leq 1} \wedge \varphi_{\tau,1}) \vee \bigvee_{\hat{c} \in C, \theta \in R_{\hat{c}}} \exists x \varphi_{d\text{-roots}}(x) \wedge P_{\hat{c}}(x) \wedge \mathcal{I}(\varphi_{\theta,d-1})(x),$$

where \mathcal{I} is the operator defined in Lemma 3.3. Note that the size of $\varphi_{\tau,d}^{\text{conn}}$ is $\|\varphi_{\tau,d}^{\text{conn}}\| \in (2(d-1) + 1)\text{-EXP}(q)$ (this is dominated by the maximal size of $|R_{\hat{c}}|$) and that $\text{qad}(\varphi_{\tau,d}^{\text{conn}}) \in \mathcal{O}(d)$. Using Lemma 3.3 and Lemma 3.5, it is not too hard to verify that $\varphi_{\tau,d}^{\text{conn}}$ defines τ .

Case 2: Disconnected Graphs. Let $\mathcal{T}_{C,q,d}^{\text{conn}} := \{\tau_1, \dots, \tau_\ell\}$. Let Φ be a set that contains the formulae $\varphi_i := \varphi_{\tau_i,d}^{\text{conn}}$ for each $i \in [\ell]$. For each graph G with $\text{td}(G) \leq d$ and each component K of G , we have $K \models \varphi_i$ iff there is a q -order \preceq of G such that $\text{tp}_q(K, \preceq) = \tau_i$; due to Lemma 3.6, this holds iff $\text{tp}_q(K, \preceq) = \tau_i$ for each q -order \preceq of G . Thus $n_{\varphi_i}(G) = n_{\tau_i}(G, \preceq)$ for each q -order \preceq of G . Let $t := 2^q + 1$ as in Lemma 3.8. For any ordered graph (G, \preceq) , let $\bar{n}(G, \preceq) := (n_{\tau_1,t}(G, \preceq), \dots, n_{\tau_\ell,t}(G, \preceq))$, i.e. $\bar{n}(G, \preceq) \in [0, t]^\ell$.

Now consider a $\tau \in \mathcal{T}_{C,q,d}$. Let $R \subseteq [0, t]^\ell$ such that for each $\bar{n} \in [0, t]^\ell$, $\bar{n} \in R$ iff there exists a q -ordered graph (G, \preceq) with $\text{td}(G) \leq d$ and $\text{tp}_q(G, \preceq) = \tau$ such that $\bar{n} = \bar{n}(G, \preceq)$. For each $\bar{n} \in R$, let $\psi_{\bar{n},t}^\Phi(\bar{x})$ be the formula of Lemma 3.9. Observe that $\|\psi_{\bar{n},t}^\Phi(\bar{x})\| \in \ell \cdot \mathcal{O}(dt^2 \max_{i \in [\ell]} \|\varphi_i\|)$. We hence have $\|\psi_{\bar{n},t}^\Phi(\bar{x})\| \in ((2d-1)+1)\text{-EXP}(q)$ and $\text{qad}(\psi_{\bar{n},t}^\Phi(\bar{x})) \in \mathcal{O}(d)$.

Define the FO-sentence $\varphi_{\tau,d} := \bigvee_{\bar{n} \in R} \psi_{\bar{n},t}^\Phi$. Since $|R| \in (2d)\text{-EXP}(q)$, also $\|\varphi_{\tau,d}\| \leq |R| \cdot \max_{\bar{n} \in R} \|\psi_{\bar{n},t}^\Phi\| \in (2d)\text{-EXP}(q)$ and $\text{qad}(\varphi_{\tau,d}) \in \mathcal{O}(d)$.

We prove that $\varphi_{\tau,d}$ defines τ on graphs of tree-depth at most d . Let G be such a graph. Suppose first that there is a q -order \preceq such that (G, \preceq) has type τ . By the definition of R_τ there is a tuple $\bar{n} \in R_\tau$ such that $\bar{n} = \bar{n}(G, \preceq)$, so $G \models \psi_{\bar{n},t}^\Phi$ by Lemma 3.9.

Suppose now that $G \models \varphi_{\tau,d}$, i.e. say $G \models \psi_{\bar{n},t}^\Phi$ for some tuple $\bar{n} \in R_\tau$. By the definition of R , an ordered graph (H, \preceq_H) with $(H, \preceq_H) \models \tau$, $\text{td}(H) \leq d$ and $\bar{n}(H, \preceq_H) = \bar{n}$ exists. By Lemma 3.9, we have $G \approx_{\Phi,t} H$. It follows from our choice of Φ that there is a q -order \preceq_G on G such that $(H, \preceq_H) \approx_{q,t} (G, \preceq_G)$. Now $(H, \preceq_H) \equiv_q^{\text{FO}} (G, \preceq_G)$ by Lemma 3.8. \square

4 Monadic second-order logic

The approach towards the results of the previous section can be adapted to obtain a quantitative variant of the result of [6] that MSO and FO have the same expressive power on the class of graphs of tree-depth at most d . Let $s(d) := \frac{d(d+1)}{2} + 2d$ for each $d \in \mathbb{N}$.

Theorem 4.1. *For each $d \in \mathbb{N}^+$ and MSO-sentence φ there is an FO-sentence ψ with $\|\psi\| \in (s(d)+1)\text{-EXP}(\text{qr}(\varphi))$ and $\text{qad}(\psi) \in \mathcal{O}(d)$ that is equivalent to φ on graphs of tree-depth at most d .*

Much of the proof of Theorem 4.1 follows the proof of Theorem 3.1, but we are spared of the complications that arose in connection with the ordering of graphs. On the other hand, the proof of an analogue to Lemma 3.8 becomes more complicated. In Lemma 3.8, we did not use the fact that we consider only graphs of bounded tree-depth. Here naively ignoring the bounded tree-depth would lead to a non-elementary dependence of the counting threshold on q . We use the following lemma to avoid this.

Lemma 4.2. *For every $d, q \in \mathbb{N}^+$, there is a $t \in d\text{-EXP}(q)$ such that, if G and H are graphs with $\text{td}(G), \text{td}(H) \leq d$ and $G \approx_{q,t} H$, then $G \equiv_q^{\text{MSO}} H$.*

5 Order-invariant monadic second-order logic

It is well-known that for each sentence in modulo-counting MSO (CMSO) there is an equivalent \leq -inv-MSO-sentence, and a conjecture of Courcelle implies that, on graphs of *bounded tree-width*, the converse of this statement is also true. In the special case where instead of bounded tree-width the graphs have bounded tree-depth, we show the following stronger result.

Theorem 5.1. *For every $d \in \mathbb{N}^+$ and \leq -inv-MSO-sentence φ there exists an FO+MOD-sentence ψ with $\text{qad}(\psi) \in \mathcal{O}(d)$ that is equivalent to φ on graphs of tree-depth at most d .*

Of course the analogue of this statement for more general classes of graphs is not true, e.g. graph connectivity is MSO-, but not FO+MOD-definable. In contrast to the previous sections, we do not analyse the formula size, because it is known from [8] that (plain) MSO can define the length of orders non-elementarily more succinct than FO. Again we need to understand \leq -inv-MSO's capabilities to count the components of a given q -type in q -ordered graphs. We say that ordered graphs (G, \preceq_G) and (H, \preceq_H) are (q, p) -similar, written $(G, \preceq_G) \cong_{q,p} (H, \preceq_H)$, if $n_\tau(G) \equiv n_\tau(H) \pmod{p}$, and $n_\tau(G) \geq p$ iff $n_\tau(H) \geq p$, for each q -type τ . The following lemma shows that MSO inherits its component counting capabilities on q -ordered graphs from its semilinear spectrum on linear orders.

Lemma 5.2. *For each $q \in \mathbb{N}^+$ a $p \in \mathbb{N}^+$ exists such that for all q -ordered graphs (G, \preceq_G) and (H, \preceq_H) , if $(G, \preceq_G) \cong_{q,p} (H, \preceq_H)$ then $(G, \preceq_G) \equiv_q^{\text{MSO}} (H, \preceq_H)$.*

The next lemma is a modulo-counting analogue of Lemma 3.9, and the two lemmas together can be used to define the $\cong_{q,p}$ -equivalence class of a graph G from given sentences that define the q -types of the components.

Lemma 5.3. *For each $d, p \in \mathbb{N}^+$, each set $\Phi := \{\varphi_1, \dots, \varphi_\ell\}$ of FO+MOD-sentences and each tuple of numbers $\bar{n} := (n_1, \dots, n_\ell) \in [0, p-1]^\ell$ there is an FO+MOD-sentence $\chi_{\bar{n}, p}^\Phi$ such that for each graph G with $\text{td}(G) \leq d$, we have $G \models \chi_{\bar{n}, p}^\Phi$ iff, for each $i \in [\ell]$, $n_{\varphi_i}(G) \equiv n_i \pmod{p}$. If $\text{qad}(\varphi_i) = \mathcal{O}(d)$ for each $i \in [\ell]$, then $\text{qad}(\chi_{\bar{n}, p}^\Phi) \in \mathcal{O}(d)$.*

To prove the lemma, at first, it is not clear at all how modulo-counting quantifiers can be used to count the number of components satisfying a given FO+MOD-sentence. But it is shown in [2, Lem. 7] that the number of tree-depth roots of each component of a graph is bounded in terms of its tree-depth. For each component, we can use its roots as FO-definable representatives that allow us to perform the necessary counting. Using the previous lemmas, we can prove an analogue to Lemma 3.2, i.e. that each (MSO, q) -type of (MSO, q) -ordered graphs is FO+MOD-definable on bounded tree-depth graphs, in a very similar way to Lemma 3.2. This makes it possible to prove Theorem 5.1.

6 Final remarks

We phrased our results for undirected (coloured) graphs to simplify notation, but their proofs generalise to structures with higher-arity relations (where the tree-depth of a structure is defined to be the tree-depth of its Gaifman graph). Furthermore, all our formula constructions imply algorithms to compute the formulas. It would be interesting to obtain corresponding lower bounds for our succinctness upper bounds.

Acknowledgements. We want to thank Isolde Adler for bringing the first two authors together with the third author, and Nicole Schweikardt for her helpful suggestions.

References

1. Benedikt, M.A., Segoufin, L.: Towards a characterization of order-invariant queries over tame graphs. *Journal of Symbolic Logic* 74(1), pp. 168–186 (2009)
2. Bouland, A., Dawar, A., Kopczynski, E.: On tractable parameterizations of graph isomorphism. In: *Proc. IPEC 2012*. pp. 218–230 (2012)
3. Chandra, A., Harel, D.: Structure and complexity of relational queries. *JCSS* 25(1), pp. 99–128 (1982)
4. Courcelle, B.: The monadic second-order logic of graphs x: linear orderings. *Theoretical Computer Science* 160(1–2), pp. 87–143 (1996)
5. Courcelle, B., Engelfriet, J.: *Graph Structure and Monadic Second-Order Logic – A Language-Theoretic Approach*. Cambridge University Press (2012)
6. Elberfeld, M., Grohe, M., Tantau, T.: Where first-order and monadic second-order logic coincide. In: *Proc. LICS 2012*. pp. 265–274. IEEE Computer Society (2012)
7. Gajarský, J., Hliněný, P.: Faster deciding MSO properties of trees of fixed height, and some consequences. In: *Proc. FSTTCS 2012*. pp. 112–123 (2012)
8. Grohe, M., Schweikardt, N.: The succinctness of first-order logic on linear orders. *Logical Methods in Computer Science* 1(1:6), pp. 1–25 (2005)
9. Libkin, L.: *Elements of Finite Model Theory*. Springer-Verlag (2004)
10. Nešetřil, J., Ossona de Mendez, P.: *Sparsity: Graphs, Structures, and Algorithms*. Springer-Verlag Berlin Heidelberg (2012)
11. Schweikardt, N.: A short tutorial on order-invariant first-order logic. In: *Proc. CSR 2013*. pp. 112–126 (2013)