

The Descriptive Complexity of Subgraph Isomorphism Without Numerics

Oleg Verbitsky^{1(✉)} and Maksim Zhukovskii²

¹ Institut für Informatik, Humboldt-Universität zu Berlin,
Unter den Linden 6, 10099 Berlin, Germany
verbitsk@informatik.hu-berlin.de

² Department of Discrete Mathematics,
Moscow Institute of Physics and Technology (State University),
Dolgoprudny, Moscow Region, Russia

Abstract. Let F be a connected graph with ℓ vertices. The existence of a subgraph isomorphic to F can be defined in first-order logic with quantifier depth no better than ℓ , simply because no first-order formula of smaller quantifier depth can distinguish between the complete graphs K_ℓ and $K_{\ell-1}$. We show that, for some F , the existence of an F subgraph in *sufficiently large* connected graphs is definable with quantifier depth $\ell - 3$. On the other hand, this is never possible with quantifier depth better than $\ell/2$. If we, however, consider definitions over connected graphs with *sufficiently large treewidth*, the quantifier depth can for some F be arbitrarily small comparing to ℓ but never smaller than the treewidth of F .

We also prove that any first-order definition of the existence of an *induced* subgraph isomorphic to F requires quantifier depth strictly more than the density of F , even over highly connected graphs. From this bound we derive a succinctness result for existential monadic second-order logic: A usage of just one monadic quantifier sometimes reduces the first-order quantifier depth at a super-recursive rate.

1 Introduction

For a fixed graph F on ℓ vertices, let $\mathcal{S}(F)$ denote the class of all graphs containing a subgraph isomorphic to F . The decision problem for $\mathcal{S}(F)$ is known as SUBGRAPH ISOMORPHISM problem. It is solvable in time $O(n^\ell)$ on n -vertex input graphs by exhaustive search. Nešetřil and Poljak [15] showed that $\mathcal{S}(F)$ can be recognized in time $O(n^{(\omega/3)\ell+2})$, where $\omega < 2.373$ is the exponent of fast square matrix multiplication. Moreover, the color-coding method by Alon, Yuster and Zwick [2] yields the time bound

$$2^{O(\ell)} \cdot n^{tw(F)+1} \log n,$$

The first author was supported by DFG grant VE 652/1–2. He is on leave from the IAPMM, Lviv. The second author was supported by grants No. 15-01-03530 and 16-31-60052 of Russian Foundation for Basic Research.

where $tw(F)$ denotes the treewidth of F . On the other hand, the decision problem for $\mathcal{S}(K_\ell)$, that is, the problem of deciding if an input graph contains a clique of ℓ vertices, cannot be solved in time $n^{o(\ell)}$ unless the Exponential Time Hypothesis fails.

We here are interested in the *descriptive complexity* of SUBGRAPH ISOMORPHISM. A sentence Φ defines a class of graphs \mathcal{C} if

$$G \models \Phi \iff G \in \mathcal{C}, \tag{1}$$

where $G \models \Phi$ means that Φ is true on G . For a logic \mathcal{L} , we let $D_{\mathcal{L}}(\mathcal{C})$ (resp. $W_{\mathcal{L}}(\mathcal{C})$) denote the minimum quantifier depth (resp. variable width) of $\Phi \in \mathcal{L}$ defining \mathcal{C} . Note that $W_{\mathcal{L}}(\mathcal{C}) \leq D_{\mathcal{L}}(\mathcal{C})$. We simplify notation by writing

$$W_{\mathcal{L}}(F) = W_{\mathcal{L}}(\mathcal{S}(F)) \text{ and } D_{\mathcal{L}}(F) = D_{\mathcal{L}}(\mathcal{S}(F)). \tag{2}$$

We are primarily interested in the first-order logic of graphs with relation symbols for adjacency and equality of vertices, that will be denoted by FO. We suppose that the vertex set of any n -vertex graph is $\{1, \dots, n\}$. Seeking the adequate logical formalism for various models of computation, descriptive complexity theory considers also more expressive logics involving numerical relations over the integers. Given a set \mathcal{N} of such relations, $\text{FO}[\mathcal{N}]$ is used to denote the extension of FO whose language contains symbols for each relation in \mathcal{N} . Of special interest are $\text{FO}[<]$, $\text{FO}[+, \times]$, and $\text{FO}[\text{Arb}]$, where Arb indicates that arbitrary relations are allowed. It is known [10, 14] that $\text{FO}[\text{Arb}]$ and $\text{FO}[+, \times]$ capture (non-uniform) AC^0 and DLOGTIME-uniform AC^0 respectively.

We will simplify the notation (2) further by writing $D(F) = D_{\text{FO}}(F)$ and $W(F) = W_{\text{FO}}(F)$. Dropping FO in the subscript, we also use notation like $D_{<}(F)$ or $W_{\text{Arb}}(F)$. In this way we obtain two hierarchies of width and depth parameters. In particular,

$$W_{\text{Arb}}(F) \leq W_{<}(F) \leq W(F) \text{ and } D_{\text{Arb}}(F) \leq D_{<}(F) \leq D(F).$$

The relation of $\text{FO}[\text{Arb}]$ to circuit complexity implies that $\mathcal{S}(F)$ is recognizable by bounded-depth unbounded-fan-in circuits of size $n^{W_{\text{Arb}}(F)+o(1)}$; see [10, 18]. The interplay between the two areas has been studied in [12, 13, 18, 19]. Noteworthy, the parameters $W_{\text{Arb}}(F)$ and $D_{\text{Arb}}(F)$ admit combinatorial upper bounds

$$W_{\text{Arb}}(F) \leq tw(F) + 3 \text{ and } D_{\text{Arb}}(F) \leq td(F) + 2 \tag{3}$$

in terms of the treewidth and treedepth of F ; see [20].¹

The focus of our paper is on FO without any background arithmetical relations. Our interest in this weakest setting is motivated by the prominent problem

¹ In his presentation [20], Benjamin Rossman states upper bounds $W_{\text{FO}}(F) \leq tw(F) + 1$ and $D_{\text{FO}}(F) \leq td(F)$ for the *colorful* version of SUBGRAPH ISOMORPHISM studied in [13]. It is not hard to observe that the auxiliary color predicates can be defined in $\text{FO}[\text{Arb}]$ at the cost of two extra quantified variables by the color-coding method developed in [2]; see also [3, Theorem 4.2].

on the power of encoding-independent computations; see, e.g., [9]. It is a long-standing open question in finite model theory as to whether there exists a logic capturing polynomial time on finite relational structures. The existence of a logic capturing polynomial time would mean that any polynomial-time computation could be made, in a sense, independent of the input encoding. If this is true, are the encoding-independent computations necessarily slower than the standard ones? This question admits the following natural variation. Suppose that a decision problem a priori admits an encoding-independent polynomial-time algorithm, say, being definable in FO, like SUBGRAPH ISOMORPHISM for a fixed pattern graph F . Is it always true that the running time of this algorithm can be improved in the standard encoding-dependent Turing model of computation?

A straightforward conversion of an FO sentence defining $\mathcal{S}(F)$ into an algorithm recognizing $\mathcal{S}(F)$ results in the time bound $O(n^{D(F)})$ for SUBGRAPH ISOMORPHISM, which can actually be improved to $O(n^{W(F)})$; see [14, Proposition 6.6]. The same applies to FO[<]. The last logic is especially interesting in the context of *order-invariant definitions*. It is well known [14, 21] that there are properties of (unordered) finite structures that can be defined in FO[<] but not in FO. Even if a property, like $\mathcal{S}(F)$, is definable in FO, one can expect that in FO[<] it can be defined much more succinctly. As a simple example, take F to be the star graph $K_{1,s}$ and observe that $D_{<}(K_{1,s}) \leq \log_2 s + 3$ and $W_{<}(K_{1,s}) \leq 3$ while $W(K_{1,s}) = s + 1$.

The main goal we pose in this paper is examining abilities and limitations of the “pure” FO in succinctly defining SUBGRAPH ISOMORPHISM. Actually, if a pattern graph F has ℓ vertices, then the trivial upper bound $D(F) \leq \ell$ cannot be improved. We have $W(F) = \ell$ simply because no first-order formula with less than ℓ variables can distinguish between the complete graphs K_ℓ and $K_{\ell-1}$. Is this, however, the only reason preventing more succinct definitions of $\mathcal{S}(F)$? How succinctly can $\mathcal{S}(F)$ be defined on large enough graphs? The question can be formalized as follows. We say that a sentence Φ defines $\mathcal{S}(F)$ on *sufficiently large connected graphs* if there is k such that (1) with $\mathcal{C} = \mathcal{S}(F)$ is true for all connected G with at least k vertices. Let $W_v(F)$ (resp. $D_v(F)$) denote the minimum variable width (resp. quantifier depth) of such Φ .

Throughout the paper, we assume that the fixed **pattern graph F is connected**. Therefore, F is contained in a host graph G if and only if it is contained in a connected component of G . By this reason, the decision problem for $\mathcal{S}(F)$ efficiently reduces to its restriction to connected input graphs. Since it suffices to solve the problem only on all sufficiently large inputs, $\mathcal{S}(F)$ is still recognizable in time $O(n^{W_v(F)})$, while $W_v(F) \leq W(F)$.

A further relaxation is motivated by Courcelle’s theorem [6] saying that every graph property definable by a sentence in monadic second-order logic can be efficiently decided on graphs of bounded treewidth. More precisely, for SUBGRAPH ISOMORPHISM Courcelle’s theorem implies that $\mathcal{S}(F)$ is decidable in time $f(\ell, tw(G)) \cdot n$, which means linear time for any class of input graphs having bounded treewidth.

Now, we say that a sentence Φ defines $\mathcal{S}(F)$ on *connected graphs with sufficiently large treewidth* if there is k such that (1) with $\mathcal{C} = \mathcal{S}(F)$ is true for all connected G with treewidth at least k . Denote the minimum variable width (resp. quantifier depth) of such Φ by $W_{tw}(F)$ (resp. $D_{tw}(F)$). Fix k that ensures the minimum value $W_{tw}(F)$ and recall that, by Courcelle’s theorem, the subgraph isomorphism problem is solvable on graphs with treewidth less than k in linear time. Note that, for a fixed k , whether or not $tw(G) < k$ is also decidable in linear time [4]. It follows that $\mathcal{S}(F)$ is recognizable even in time $O(n^{W_{tw}(F)})$, while $W_{tw}(F) \leq W_v(F)$.

The above discussion shows that the parameters $W_v(F)$, $D_v(F)$, $W_{tw}(F)$, and $D_{tw}(F)$ have clear algorithmic meaning. Analyzing this setting, we obtain the following results.

- We demonstrate that non-trivial definitions over sufficiently large graphs are possible by showing that $D_v(F) \leq v(F) - 3$ for some F , where $v(F)$ denotes the number of vertices in F . On the other hand, we show limitations of this approach by proving that $W_v(F) \geq (v(F) - 1)/2$ for all F .
- The last barrier (as well as any lower bound in terms of $v(F)$) can be overcome by definitions over graphs with sufficiently large treewidth. Specifically, for every ℓ and $a \leq \ell$ there is an ℓ -vertex F such that $D_{tw}(F) \leq a$ and, moreover, $tw(F) = a - 1$. On the other hand, $W_{tw}(F) \geq tw(F)$ for all F . Note that, along with (3), this implies that $W_{\text{Arb}}(F) \leq W_{tw}(F) + 3$.

We also address the descriptive complexity of the INDUCED SUBGRAPH ISOMORPHISM problem. Let $\mathcal{I}(F)$ denote the class of all graphs containing an *induced* subgraph isomorphic to F . The state-of-the-art of the algorithmics for INDUCED SUBGRAPH ISOMORPHISM is different from SUBGRAPH ISOMORPHISM. Floderus et al. [8] collected evidence in favor of the conjecture that $\mathcal{I}(F)$ for F with ℓ vertices cannot be recognized faster than $\mathcal{I}(K_{c\ell})$, where $c < 1$ is a constant.

Similarly to $D(F)$, we define $D[F] = D(\mathcal{I}(F))$, where the square brackets indicate that the case of induced subgraphs is considered. The trivial argument showing that $D(F) = v(F)$ does not work anymore unless F is a complete graph (whereas K_ℓ contains every ℓ -vertex F as a subgraph, it contains no *induced* copy of F unless $F = K_\ell$). Proving or disproving that $D[F] = v(F)$ seems to be a subtle problem. Our results on INDUCED SUBGRAPH ISOMORPHISM are as follows.

- We prove a general lower bound $D[F] > e(F)/v(F)$, where $e(F)$ denotes the number of edges in F . In fact, the bound holds true even for $D_{tw}[F]$.
- From this bound we derive a succinctness result for existential monadic second-order logic: A usage of just one monadic quantifier sometimes reduces the FO quantifier depth at a super-recursive rate. More precisely, let $D_{\exists\text{MSO}}[F]$ denote the minimum quantifier depth of a second-order sentence with a single existential monadic quantifier that defines $\mathcal{I}(F)$. Then $D_{\exists\text{MSO}}[F]$ can sometimes be so small comparing to $D[F] = D_{\text{FO}}[F]$ that there is no total recursive function f such that $f(D_{\exists\text{MSO}}[F]) \geq D[F]$ for all F .

2 Preliminaries

First-Order Complexity of Graph Properties. We consider first-order sentences about graphs in the language containing the adjacency and the equality relations. Let \mathcal{C} be a first-order definable class of graphs and π be a graph parameter. Let $D_\pi^k(\mathcal{C})$ denote the minimum quantifier depth of a first-order sentence Φ such that, for every connected graph G with $\pi(G) \geq k$, Φ is true on G exactly when G belongs to \mathcal{C} . Note that $D_\pi^k(\mathcal{C}) \geq D_\pi^{k+1}(\mathcal{C})$, and define $D_\pi(\mathcal{C}) = \min_k D_\pi^k(\mathcal{C})$. In other words, $D_\pi(\mathcal{C})$ is the minimum quantifier depth of a first-order sentence defining \mathcal{C} over connected graphs with sufficiently large values of π .

The *variable width* of a first-order sentence Φ is the number of first-order variables used to build Φ ; different occurrences of the same variable do not count. By $W_\pi(\mathcal{C})$ we denote the minimum variable width of Φ defining \mathcal{C} over connected graphs with sufficiently large π . Note that $W_\pi(\mathcal{C}) \leq D_\pi(\mathcal{C})$.

Recall that a graph is *k-connected* if it has more than k vertices, is connected, and remains connected after removal of any $k-1$ vertices. The *connectivity* $\kappa(G)$ of G is equal to the maximum k such that G is k -connected. We will consider the depth $D_\pi(\mathcal{C})$ and the width $W_\pi(\mathcal{C})$ for three parameters π , namely the number of vertices $v(G)$, the treewidth $tw(G)$, and the connectivity $\kappa(G)$. It is not hard to see that

$$D_v(\mathcal{C}) \geq D_{tw}(\mathcal{C}) \geq D_\kappa(\mathcal{C}) \text{ and } W_v(\mathcal{C}) \geq W_{tw}(\mathcal{C}) \geq W_\kappa(\mathcal{C}).$$

As it was discussed in Sect. 1, the values of $D_v(\mathcal{C})$ and $D_{tw}(\mathcal{C})$, as well as $W_v(\mathcal{C})$ and $W_{tw}(\mathcal{C})$, are related to the time complexity of the decision problem for \mathcal{C} . Consideration of $D_\kappa(\mathcal{C})$ and $W_\kappa(\mathcal{C})$ is motivated by the fact that some lower bounds we are able to show for $D_v(\mathcal{C})$ and $D_{tw}(\mathcal{C})$ actually hold for $D_\kappa(\mathcal{C})$ or even for $W_\kappa(\mathcal{C})$, and it is natural to present them in this stronger form.

Recall that $\mathcal{S}(F)$ denotes the class of graphs containing a subgraph isomorphic to F . Simplifying the notation, we write $D_v(F) = D_v(\mathcal{S}(F))$, $W_v(F) = W_v(\mathcal{S}(F))$, etc.

Given two non-isomorphic graphs G and H , let $D(G, H)$ (resp. $W(G, H)$) denote the minimum quantifier depth (resp. variable width) of a sentence that is true on one of the graphs and false on the other.

- Lemma 1.**
1. $D_\pi(\mathcal{C}) \geq d$ if there are connected graphs $G \in \mathcal{C}$ and $H \notin \mathcal{C}$ with arbitrarily large values of $\pi(G)$ and $\pi(H)$ such that $D(G, H) \geq d$.
 2. $W_\pi(\mathcal{C}) \geq d$ if there are connected graphs $G \in \mathcal{C}$ and $H \notin \mathcal{C}$ with arbitrarily large values of $\pi(G)$ and $\pi(H)$ such that $W(G, H) \geq d$.
 3. $D_\pi(\mathcal{C}) \leq d$ if $D(G, H) \leq d$ for all connected graphs $G \in \mathcal{C}$ and $H \notin \mathcal{C}$ with sufficiently large values of $\pi(G)$ and $\pi(H)$.

Lemma 1 reduces estimating $D_\pi(\mathcal{C})$ to estimating $D(G, H)$ over connected $G \in \mathcal{C}$ and $H \notin \mathcal{C}$ with large values of π . Also, proving lower bounds for $W_\pi(\mathcal{C})$ reduces to proving lower bounds for $W(G, H)$. For estimating $D(G, H)$ and $W(G, H)$ we will use the well-known characterization of these parameters in terms of the k -pebble Ehrenfeucht-Fraïssé game [10]:

1. $D(G, H)$ is equal to the minimum k such that Spoiler has a winning strategy in the k -round k -pebble game on G and H .
2. $W(G, H)$ is equal to the minimum k such that, for some d , Spoiler has a winning strategy in the d -round k -pebble game on G and H .

Graph-Theoretic Preliminaries. Recall that $v(G)$ denotes the number of vertices in a graph G . The treewidth of G is denoted by $tw(G)$. The *neighborhood* $N(v)$ of a vertex v consists of all vertices adjacent to v . The number $\deg v = |N(v)|$ is called the *degree* of v . The vertex of degree 1 is called *pendant*.

We use the standard notation K_n for complete graphs, P_n for paths, and C_n for cycles on n vertices. Furthermore, $K_{a,b}$ denotes the complete bipartite graph whose vertex classes have a and b vertices. In particular, $K_{1,n-1}$ is the star graph on n vertices. The subscript in the name of a graph will almost always denote the number of vertices. If a graph is indexed by two parameters, their sum is typically equal to the total number of vertices in the graph.

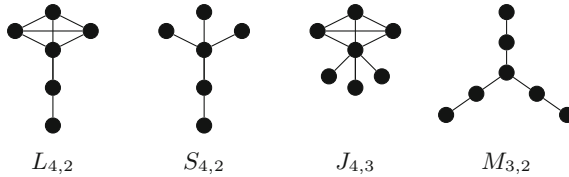


Fig. 1. Special graph families: Lollipops, sparklers, jellyfishes, and megastars.

The following definitions are illustrated in Fig. 1. Let $a \geq 3$ and $b \geq 1$. The *lollipop graph* $L_{a,b}$ is obtained from K_a and P_b by adding an edge between an end vertex of P_b and a vertex of K_a . We also make a natural convention that $L_{a,0} = K_a$. Furthermore, the *sparkler graph* $S_{a,b}$ is obtained from $K_{1,a-1}$ and P_b by adding an edge between an end vertex of P_b and the central vertex of $K_{1,a-1}$. The *jellyfish graph* $J_{a,b}$ is the result of attaching b pendant vertices to a vertex of K_a . Finally, the *megastar graph* $M_{s,t}$ is obtained from the star $K_{1,s}$ by subdividing each edge into P_{t+1} ; thus $v(M_{s,t}) = st + 1$.

3 Definitions over Sufficiently Large Graphs

Our first goal is to demonstrate that non-trivial definitions over large connected graphs are really possible. The lollipop graphs $L_{a,1}$ give simple examples of pattern graphs F with $D_v(F) \leq v(F) - 1$. Though not so easily, the same can be shown for the path graphs P_ℓ . We are able to show better upper bounds using sparkler graphs.

Theorem 2. *There is a graph F with $D_v(F) \leq v(F) - 3$. Specifically, $D_v(S_{4,4}) = 5$.*

For the proof we need two technical lemmas.

Lemma 3. *Suppose that a connected graph H contains the 4-star $K_{1,4}$ as a subgraph but does not contain any subgraph $S_{4,4}$. Then H contains a vertex of degree more than $(v(H)/2)^{1/7}$.*

Proof. H cannot contain P_{15} because, together with $K_{1,4}$, it would give an $S_{4,4}$ subgraph. Consider an arbitrary spanning tree T in H and denote its maximum vertex degree by d and its radius by r . Note that $v(T) \leq 1 + d + d(d - 1) + \dots + d(d - 1)^{r-1}$. Since T contains no P_{15} , we have $r \leq 7$. It follows that $v(H) = v(T) < 2d^7$. □

Let \sim denote the adjacency relation.

Lemma 4. *Let $y_0 \in V(H)$ and assume that*

- H is a sufficiently large connected graph,
- H does not contain $S_{4,4}$,
- $\deg y_0 \geq 4$,
- $y_0 y_1 y_2 y_3 y_4$ is a path in H .

Then (see Fig. 2)

1. $\deg y_0 = 4$,
2. $y_0 \sim y_2, y_0 \sim y_3, y_0 \sim y_4$,
3. if $N(y_0) = \{y_1, y_2, y', y''\}$, then $y_1 \sim y'$ and $y_1 \sim y''$.

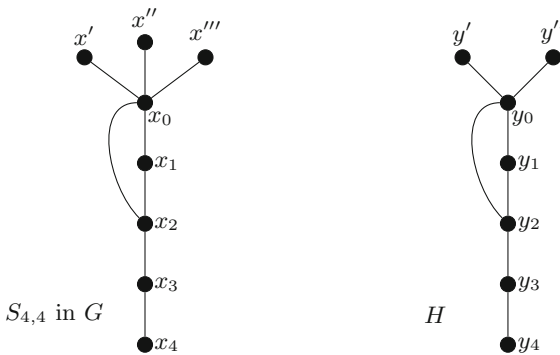


Fig. 2. Proof of Theorem 2.

Proof. By Lemma 3 we know that H must contain a vertex z of large degree, namely $\deg z \geq 7$. We have $y_0 \approx y_4$ for else H would contain a cycle C_5 and, together with z , this would give us a subgraph $S_{4,4}$ (because, by connectedness of H , we would have a path P_5 emanating from z). Therefore, y_0 has a neighbor $y' \notin \{y_1, y_2, y_3, y_4\}$. Furthermore, $y_0 \approx y_3$ for else, considering a path from z to one of the vertices $y', y_0, y_1, y_2, y_3, y_4$, we get a P_5 emanating from z and, hence, an $S_{4,4}$. Therefore, y_0 has another neighbor $y'' \notin \{y', y_1, y_2, y_3, y_4\}$. Furthermore, $y_0 \sim y_2$ for else y_0 would have three neighbors y', y'', y''' different from y_1, y_2, y_3, y_4 , which would give $S_{4,4}$. By the same reason, y_0 has no other neighbors, that is, $N(y_0) = \{y_1, y_2, y', y''\}$ and $\deg y_0 = 4$. Note that $z \in \{y_0, y_1, y_2, y_3, y_4\}$ for else we easily get an $S_{4,4}$ by considering a path from z to one of these vertices. It is also easy to see that $z \neq y_0, y_4, y_3, y_1$ (for example, if $\deg y_1 \geq 7$, then it would give an $S_{4,4}$ with tail $y_1 y_0 y_2 y_3 y_4$). Therefore, $z = y_2$. If $y_1 \sim y'$ or $y_1 \sim y''$, we would have an $S_{4,4}$ with tails $y_2 y_1 y' y_0 y''$ or $y_2 y_1 y'' y_0 y'$ respectively. \square

Proof (of Theorem 2). We are now ready to prove the upper bound $D_v(S_{4,4}) \leq 5$. Consider sufficiently large connected graphs G and H and suppose that G contains an $S_{4,4}$, whose vertices are labeled as in Fig. 2, and H contains no copy of $S_{4,4}$. We describe a winning strategy for Spoiler in the game on G and H .

1st round. Spoiler pebbles x_0 . Denote the response of Duplicator in H by y_0 . Assume that $\deg y_0 \geq 4$ for else Spoiler wins in the next 4 moves. Assume that $x_0 \sim x_2$ for else Spoiler wins by pebbling x_1, x_2, x_3, x_4 (if Duplicator responds with a path $y_0 y_1 y_2 y_3 y_4$, she loses by Condition 2 in Lemma 4).

2nd round. Spoiler pebbles x_1 . Denote the response of Duplicator in H by y_1 . Assume that there is a path $y_0 y_1 y_2 y_3 y_4$ for else Spoiler wins in the next 3 moves.

Case 1: x_1 is adjacent to any of the vertices x', x'', x''' , say, to x' . Spoiler pebbles x_2 and x' and wins. Indeed, Duplicator has to respond with two vertices in H both in $N(y_0) \cap N(y_1)$, which is impossible by Conditions 1 and 3 of Lemma 4.

Case 2: $x_1 \approx x', x_1 \approx x'', x_1 \approx x'''$. Spoiler wins by pebbling x', x'', x''' . Duplicator has to respond with three vertices in $N(y_0) \setminus N(y_1)$, which is impossible by Conditions 1 and 2 of Lemma 4.

This completes the proof of the upper bound. On the other hand, we have $D_v(S_{4,4}) > 4$ by considering the jellyfish graphs $G = J_{5,n}$ and $H = J_{4,n}$. \square

We now show general lower bounds for $D_v(F)$ and $W_v(F)$. For this, we need some definitions. Let $v_0 v_1 \dots v_t$ be an induced path in a graph G . We call it *pendant* if $\deg v_0 \neq 2$, $\deg v_t = 1$ and $\deg v_i = 2$ for all $1 \leq i < t$. Furthermore, let S be an induced star $K_{1,s}$ in G with the central vertex v_0 . We call S *pendant* if all its pendant vertices are pendant also in G , and in G there are no more than s pendant vertices adjacent to v_0 . The definition ensures that a pendant path (or star) cannot be contained in a larger pendant path (or star). As an example, note that the sparkler graph $S_{s+1,t}$ has a pendant P_{t+1} and a pendant $K_{1,s}$.

Let $p(F)$ denote the maximum t such that F has a pendant path P_{t+1} . Similarly, let $s(F)$ denote the maximum s such that F has a pendant star $K_{1,s}$. If F has no pendant vertex, then we set $p(F) = 0$ and $s(F) = 0$.

Theorem 5. $D_v(F) \geq (v(F) + 1)/2$ and $W_v(F) \geq (v(F) - 1)/2$ for every connected F unless $F = P_2$ or $F = P_3$.

Proof. Denote

$$\ell = v(F), \quad t = p(F) \text{ and } s = s(F).$$

We begin with noticing that

$$D_v(F) \geq \ell - t \text{ and } W_v(F) \geq \ell - t - 1. \tag{4}$$

Indeed, this is obvious if F is a path, that is, $F = P_{t+1}$. If F is not a path, we consider lollipop graphs $G = L_{\ell-t,n}$ and $H = L_{\ell-t-1,n}$ for each $n \geq t$ (note that $\ell \geq t + 3$ and, if $\ell = t + 3$, then $H = L_{2,n} = P_{n+2}$). Obviously, G contains F , and H does not. It remains to note that $D(G, H) \geq \ell - t$ and $W(G, H) \geq \ell - t - 1$.

We also claim that

$$D_v(F) \geq \ell - s \text{ and } W_v(F) \geq \ell - s - 1. \tag{5}$$

This is obvious if F is a star, that is, $F = K_{1,s}$. If F is not a star, we consider jellyfish graphs $G = J_{\ell-s,n}$ and $H = J_{\ell-s-1,n}$ for each $n \geq s$ (note that $\ell \geq s + 3$ and, if $\ell = s + 3$, then $H = J_{2,n} = K_{1,n+1}$). Clearly, G contains F , and H does not. It remains to observe that $D(G, H) \geq \ell - s$ and $W(G, H) \geq \ell - s - 1$.

Let $F = K_{1,\ell-1}$ or $F = P_\ell$, where $\ell \geq 4$. Using (4) and (5) respectively, we get $D_v(F) \geq \ell - 1 \geq \frac{\ell+1}{2}$ and, similarly, $W_v(F) \geq \ell - 2 \geq \frac{\ell-1}{2}$. Assume, therefore, that F is neither a star nor a path. In this case we claim that

$$t + s < \ell. \tag{6}$$

This is obviously true if F has no pendant vertex, that is, $t = s = 0$. Suppose that F has a pendant vertex and, therefore, both $t > 0$ and $s > 0$. Consider an arbitrary spanning tree T of F and note that T contains all pendant paths and stars of F . Fix a longest pendant path P and a largest pendant star S in F . If P and S share at most one common vertex, we readily get (6). If they share two vertices, then $S = K_{1,1}$, i.e., $s = 1$, and $t + 1 < \ell$ follows from the assumption that F is not a path.

The theorem readily follows from (4)–(6). □

4 Definitions over Graphs of Sufficiently Large Treewidth

Theorem 5 poses limitations on the succinctness of definitions over sufficiently large graphs. We now show that there are no such limitations for definitions over connected graphs with sufficiently large treewidth.

The Grid Minor Theorem says that every graph of large treewidth contains a large grid minor; see [7]. The strongest version of this result belongs to Chekuri and Chuzhoy [5] who proved that, for some $\epsilon > 0$, every graph G of treewidth

k contains the $m \times m$ grid as a minor with $m = \Omega(k^\epsilon)$. If $m > 2b$, then G must contain $M_{3,b}$ as a subgraph. This applies also to all subgraphs of $M_{3,b}$. The following result is based on the fact that a graph of large treewidth contains a long path.

Theorem 6. *For all a and ℓ such that $3 \leq a \leq \ell$ there is a graph F with $v(F) = \ell$ and $tw(F) = a - 1$ such that $D_{tw}(F) \leq a$. Specifically, $D_{tw}(L_{a,b}) = W_\kappa(L_{a,b}) = a$ if $a \geq 3$ and $b \geq 0$.*

Note for comparison that $W_v(L_{a,b}) \geq a + b - 2$, as follows from the bound (5) in the proof of Theorem 5.

Proof. We first prove the upper bound $D_{tw}(L_{a,b}) \leq a$. If a connected graph H of large treewidth does not contain $L_{a,b}$, it cannot contain even K_a for else K_a could be combined with a long path to give $L_{a,b}$. Therefore, Spoiler wins on $G \in \mathcal{S}(L_{a,b})$ and such H in a moves.

For the lower bound $W_\kappa(L_{a,b}) \geq a$, consider $G = K(a, n)$ and $H = K(a - 1, n)$, where $K(a, n)$ denotes the complete a -partite graph with each part having n vertices. Note that this graph is $(a - 1)n$ -connected. If $n > b$, then G contains $L_{a,b}$, while H for any n does not contain even K_a . It remains to note that $W(G, H) \geq a$ if $n \geq a - 1$. □

We now prove a general lower bound for $W_{tw}(F)$ in terms of the treewidth $tw(F)$. Using the terminology of [11, Chap. 5], we define the *core* F_0 of F to be the graph obtained from F by removing, consecutively and as long as possible, vertices of degree at most 1. If F is not a forest, then F_0 is nonempty; it consists of all cycles of F and the paths between them.

We will use the well-known fact that there are cubic graphs of arbitrary large treewidth. This fact dates back to Pinsker [17] who showed that a random cubic graph with high probability has good expansion properties, implying linear treewidth.

Theorem 7. *If F is connected, then*

1. $W_{tw}(F) \geq v(F_0)$, and
2. $W_{tw}(F) \geq tw(F) + 1$ unless F is contained in some 3-megastar $M_{3,b}$.

Note that the bound in part 2 of Theorem 7 is tight by Theorem 6.

Proof. 1. Denote $v(F) = \ell$ and $v(F_0) = \ell_0$. If F is a tree, then $\ell_0 = 0$, and the claim is trivial. Suppose, therefore, that F is not a tree. In this case, $\ell_0 \geq 3$.

We begin with a cubic graph B of as large treewidth $tw(B)$ as desired. Let $(B)_\ell$ denote the graph obtained from B by subdividing each edge by ℓ new vertices. Since B is a minor of $(B)_\ell$, we have $tw((B)_\ell) \geq tw(B)$; see [7].

Next, we construct a gadget graph A as follows. By a k -uniform tree we mean a tree of even diameter where every non-leaf vertex has degree k and all distances between a leaf and the central vertex are equal. The graph A is obtained by merging the ℓ -uniform tree of radius ℓ and $(B)_\ell$; merging is done by identifying one leaf of the tree and one vertex of $(B)_\ell$.

We now construct G by attaching a copy of A to each vertex of K_{ℓ_0} . Specifically, a copy A_u of A is created for each vertex u of K_{ℓ_0} , and u is identified with the central vertex of (the tree part of) A_u . Let H be obtained from G by shrinking its clique part to K_{ℓ_0-1} . Since both G and H contain copies of $(B)_\ell$, these two graphs have treewidth at least as large as $tw(B)$.

The clique part of G is large enough to host the core F_0 , and the remaining tree shoots of F fit into the A -parts of G . Therefore, G contains F as a subgraph. On the other hand, the clique part of H is too small for hosting F_0 , and no cycle of F fits into any A -part because A has larger girth than F . Therefore, H does not contain F . It remains to notice that $W(G, H) \geq \ell_0$.

2. Suppose first that F is not a tree. By part 1, we then have

$$W_{tw}(F) \geq v(F_0) \geq tw(F_0) + 1 = tw(F) + 1.$$

If F is a tree not contained in any 3-megastar, then there are connected graphs of arbitrarily large treewidth that do not contain F as a subgraph (for example, consider $(B)_\ell$ for a connected cubic graph B as in part 1). Trivially, there are also connected graphs of arbitrarily large treewidth that contain F as a subgraph. Since one pebble is not enough for Spoiler to distinguish the latter from the former, we have $W_{tw}(F) \geq 2 = tw(F) + 1$ in this case. \square

5 Induced Subgraphs: Trading Super-Recursively Many First-Order Quantifiers for a Single Monadic One

By $\mathcal{I}(F)$ we denote the class of all graphs containing an *induced* subgraph isomorphic to F . Similarly to $D(F)$, we use the notation $D[F] = D(\mathcal{I}(F))$, where the square brackets indicate that only induced subgraphs are considered. In the same vein, $D_\kappa[F] = D_\kappa(\mathcal{I}(F))$.

Unlike the case of (not necessarily induced) subgraphs, where the equality $D(F) = v(F)$ is trivial, determining and estimating the parameter $D[F]$ seems to be a subtle problem. In this section we prove a lower bound for $D[F]$ in terms of the density of F ; this bound actually holds for $D_\kappa[F]$. The proof will use known facts about random graphs in the Erdős-Rényi model $G(n, p)$, collected below. It should be stressed that, whenever the term *subgraph* stands alone, it refers to a *not necessarily induced* subgraph. *With high probability* means that the probability approaches 1 as $n \rightarrow \infty$.

The *density* of a graph K is defined to be the ratio $\rho(K) = e(K)/v(K)$. The maximum $\rho(K)$ over all subgraphs K of a graph F will be denoted by $\rho^*(F)$. The following fact from the random graph theory was used also in [13] for proving average-case lower bounds on the AC^0 complexity of SUBGRAPH ISOMORPHISM.

Lemma 8 (Subgraph Threshold, see [11, Chap. 3]).

1. If $\alpha = 1/\rho^*(F)$, then the probability that $G(n, n^{-\alpha})$ contains F as a subgraph converges to a limit different from 0 and 1 as $n \rightarrow \infty$.
2. If $\alpha > 1/\rho^*(F)$, then with high probability $G(n, n^{-\alpha})$ does not contain F as a subgraph.

Let $\alpha > 0$. Given a graph S and its subgraph K , we define $f_\alpha(S, K) = v(S) - v(K) - \alpha(e(S) - e(K))$.

Lemma 9 (Generic Extension, see [1, Chap.10]). *Let F be a graph with vertices v_1, \dots, v_ℓ and K be a subgraph of F with vertices v_1, \dots, v_k . Assume that $f_\alpha(S, K) > 0$ for every subgraph S of F containing K as a proper subgraph. Then with high probability every sequence of pairwise distinct vertices x_1, \dots, x_k in $G(n, n^{-\alpha})$ can be extended with pairwise distinct x_{k+1}, \dots, x_ℓ such that $x_i \sim x_j$ if and only if $v_i \sim v_j$ for all $i \leq \ell$ and $k < j \leq \ell$.*

Lemma 10 (Zero-One d -Law [23]). *Let $0 < \alpha < \frac{1}{d-2}$, and Ψ be a first-order statement of quantifier depth d . Then the probability that Ψ is true on $G(n, n^{-\alpha})$ converges either to 0 or to 1 as $n \rightarrow \infty$.*

We are now ready to prove our result.

Theorem 11. *If $e(F) > v(F)$, then $D_\kappa[F] \geq \frac{e(F)}{v(F)} + 2$ and $D_\kappa(F) \geq \frac{e(F)}{v(F)} + 2$.*

Proof. We prove the bound for $D_\kappa[F]$. The same proof works as well for $D_\kappa(F)$ (and is even simpler as the equality (7) below is only needed in the induced case).

Set $\alpha = 1/\rho^*(F)$ and denote $\mathbb{G}_n = G(n, n^{-\alpha})$. We begin with proving that

$$P[\mathbb{G}_n \in \mathcal{I}(F)] = P[\mathbb{G}_n \in \mathcal{S}(F)] - o(1). \tag{7}$$

Let K be a maximal subgraph of F with $\rho(K) = \rho^*(F)$. Note that K is an induced subgraph of F . Note also that, if F is balanced, i.e., $\rho^*(F) = \rho(F)$, then $K = F$. The graph K has less than $\binom{v(K)}{2}$ supergraphs K' obtainable by adding an edge to K , and every K' has density strictly larger than K , that is, $\rho(K') > 1/\alpha$. By part 2 of Lemma 8, each such K' appears as a subgraph in \mathbb{G}_n with probability $o(1)$. It follows that

$$P[\mathbb{G}_n \in \mathcal{I}(K)] = P[\mathbb{G}_n \in \mathcal{S}(K)] - o(1). \tag{8}$$

which readily implies (7) in the case that F is balanced.

Suppose now that F is not balanced. In this case, for every subgraph S of F containing K properly we have $v(S)/e(S) > \alpha$, which implies $f_\alpha(S, K) > 0$. Lemma 9 ensures that, with probability $1 - o(1)$, every induced copy of K in \mathbb{G}_n extends to an induced copy of F . Therefore,

$$\begin{aligned} P[\mathbb{G}_n \in \mathcal{S}(F)] &\geq P[\mathbb{G}_n \in \mathcal{I}(F)] \geq P[\mathbb{G}_n \in \mathcal{I}(K)] - o(1) \\ &\geq P[\mathbb{G}_n \in \mathcal{S}(K)] - o(1) \geq P[\mathbb{G}_n \in \mathcal{S}(F)] - o(1), \end{aligned} \tag{9}$$

where the last but one inequality is due to (8). Equality (7) is proved.

By part 1 of Lemma 8, $\lim_{n \rightarrow \infty} P[\mathbb{G}_n \in \mathcal{S}(F)]$ exists and equals neither 0 nor 1. It follows from (7) that $P[\mathbb{G}_n \in \mathcal{I}(F)]$ converges to the same limit, different from 0 and 1.

Now, assume that a first-order sentence Φ of quantifier depth d defines $\mathcal{S}(F)$ over k -connected graphs for all $k \geq k_0$. We have to prove that $d \geq \frac{e(F)}{v(F)} + 2$, whatever k_0 .

By the assumption of the theorem, $\rho^*(F) \geq \rho(F) > 1$. Fix k such that $1 + 1/k < \rho(F)$ and $k \geq k_0$. Lemma 9 implies that with high probability every two vertices in \mathbb{G}_n can be connected by k vertex-disjoint paths (of length k each). Therefore, \mathbb{G}_n is k -connected with high probability.

Since Φ correctly decides the existence of an induced copy of F on all k -connected graphs,

$$P[\mathbb{G}_n \models \Phi] = P[\mathbb{G}_n \in \mathcal{I}(F)] + o(1).$$

Therefore, $P[\mathbb{G}_n \models \Phi]$ converges to the same limit as $P[\mathbb{G}_n \in \mathcal{I}(F)]$, which, as we have seen, is different from 0 and 1. By Lemma 10, this implies that $\alpha \geq \frac{1}{d-2}$. From here we conclude that

$$d \geq \rho^*(F) + 2 \geq \frac{e(F)}{v(F)} + 2,$$

as required. □

We now turn to *existential monadic second-order logic*, denoted by $\exists\text{MSO}$, whose formulas are of the form

$$\exists X_1 \dots \exists X_m \Phi, \tag{10}$$

where a first-order subformula Φ is preceded by (second-order) quantification over unary relations (that is, we are now allowed to use existential quantifiers over subsets of vertices X_1, X_2, \dots). The second-order quantifiers contribute to the *quantifier depth* as well as the first-order ones. Thus, the quantifier depth of the sentence (10) is larger by m than the quantifier depth of the formula Φ . If a graph property \mathcal{C} is definable in $\exists\text{MSO}$, the minimum quantifier depth of a defining formula will be denoted by $D_{\exists\text{MSO}}(\mathcal{C})$. Furthermore, we define $D_{\exists\text{MSO}}[F] = D_{\exists\text{MSO}}(\mathcal{I}(F))$.

It is very well known that $\exists\text{MSO}$ is strictly more expressive than first-order logic. For example, the properties of a graph to be disconnected or to be bipartite are expressible in $\exists\text{MSO}$ but not in FO. We now show that $\exists\text{MSO}$ is also much more succinct than FO, which means that some properties of graphs that are expressible in FO can be expressed in $\exists\text{MSO}$ with significantly smaller quantifier depth. In fact, this can be demonstrated by considering the properties of containing a fixed induced subgraph. It turns out that, if we are allowed to use just one monadic second-order quantifier, the number of first-order quantifiers can sometimes be drastically reduced.

Theorem 12. *There is no total recursive function f such that*

$$f(D_{\exists\text{MSO}}[F]) \geq D[F]$$

for all graphs F . Moreover, this holds true even for the fragment of $\exists\text{MSO}$ where exactly one second-order quantifier is allowed.

The proof, which can be found in a long version of this paper [22], is based on Theorem 11 and [16, Theorem 4.2].

Acknowledgements. We would like to thank Tobias Müller for his kind hospitality during the Workshop on Logic and Random Graphs in the Lorentz Center (August 31–September 4, 2015), where this work was originated. We also thank the anonymous referee who provided us with numerous useful comments on the manuscript.

References

1. Alon, N., Spencer, J.H.: *The Probabilistic Method*. Wiley, Chichester (2016)
2. Alon, N., Yuster, R., Zwick, U.: Color-coding. *J. ACM* **42**(4), 844–856 (1995)
3. Amano, K.: k -Subgraph isomorphism on AC^0 circuits. *Comput. Complex.* **19**(2), 183–210 (2010)
4. Bodlaender, H.L.: A linear-time algorithm for finding tree-decompositions of small treewidth. *SIAM J. Comput.* **25**(6), 1305–1317 (1996)
5. Chekuri, C., Chuzhoy, J.: Polynomial bounds for the grid-minor theorem. In: *Proceedings of the 46th ACM Symposium on Theory of Computing (STOC 2014)*, pp. 60–69 (2014)
6. Courcelle, B.: The monadic second-order logic of graphs I. Recognizable sets of finite graphs. *Inf. Comput.* **85**(1), 12–75 (1990)
7. Diestel, R.: *Graph Theory*. Springer, New York (2000)
8. Floderus, P., Kowaluk, M., Lingas, A., Lundell, E.: Induced subgraph isomorphism: are some patterns substantially easier than others? *Theor. Comput. Sci.* **605**, 119–128 (2015)
9. Grädel, E., Grohe, M.: Is polynomial time choiceless? In: Beklemishev, L., Blass, A., Dershowitz, N., Finkbeiner, B., Schulte, W. (eds.) *Fields of Logic and Computation II*. LNCS, vol. 9300, pp. 193–209. Springer, Cham (2015). doi:[10.1007/978-3-319-23534-9_11](https://doi.org/10.1007/978-3-319-23534-9_11)
10. Immerman, N.: *Descriptive Complexity*. Springer, New York (1999)
11. Janson, S., Łuczak, T., Ruciński, A.: *Random Graphs*. Wiley, New York (2000)
12. Koucký, M., Lautemann, C., Poloczek, S., Thérien, D.: Circuit lower bounds via Ehrenfeucht-Fraïssé games. In: *Proceedings of the 21st Annual IEEE Conference on Computational Complexity (CCC 2006)*, pp. 190–201 (2006)
13. Li, Y., Razborov, A.A., Rossman, B.: On the AC^0 complexity of subgraph isomorphism. In: *Proceedings of the 55th IEEE Annual Symposium on Foundations of Computer Science (FOCS 2014)*, pp. 344–353. IEEE Computer Society (2014)
14. Libkin, L.: *Elements of Finite Model Theory*. Springer, Berlin (2004)
15. Nešetřil, J., Poljak, S.: On the complexity of the subgraph problem. *Commentat. Math. Univ. Carol.* **26**, 415–419 (1985)
16. Pikhurko, O., Spencer, J., Verbitsky, O.: Succinct definitions in the first order theory of graphs. *Ann. Pure Appl. Log.* **139**(1–3), 74–109 (2006)
17. Pinsker, M.S.: On the complexity of a concentrator. In: *Proceedings of the 7th Annual International Teletraffic Conference*, vol. 4, pp. 1–318 (1973)
18. Rossman, B.: On the constant-depth complexity of k -clique. In: *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC 2008)*, pp. 721–730. ACM (2008)
19. Rossman, B.: An improved homomorphism preservation theorem from lower bounds in circuit complexity (2016). Manuscript

20. Rossman, B.: Lower bounds for subgraph isomorphism and consequences in first-order logic. In: Talk in the Workshop on Symmetry, Logic, Computation at the Simons Institute, Berkeley, November 2016. <https://simons.berkeley.edu/talks/benjamin-rossman-11-08-2016>
21. Schweikardt, N.: A short tutorial on order-invariant first-order logic. In: Bulatov, A.A., Shur, A.M. (eds.) CSR 2013. LNCS, vol. 7913, pp. 112–126. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-38536-0_10](https://doi.org/10.1007/978-3-642-38536-0_10)
22. Verbitsky, O., Zhukovskii, M.: The descriptive complexity of subgraph isomorphism without numerics. E-print (2016). <http://arxiv.org/abs/1607.08067>
23. Zhukovskii, M.: Zero-one k -law. *Discrete Math.* **312**, 1670–1688 (2012)