Tight Bounds on the Asymptotic Descriptive Complexity of Subgraph Isomorphism

OLEG VERBITSKY, * Institut für Informatik, Humboldt-Universität zu Berlin, Germany
MAKSIM ZHUKOVSKI, Moscow Institute of Physics and Technology, Laboratory of Advanced Combinatorics and Network Applications, Russia

Let \( v(F) \) denote the number of vertices in a fixed connected pattern graph \( F \). We show an infinite family of patterns \( F \) such that the existence of a subgraph isomorphic to \( F \) is expressible by a first-order sentence of quantifier depth \( \frac{2}{3} v(F) + 1 \), assuming that the host graph is sufficiently large and connected. However, this is impossible for any \( F \) using less than \( \frac{2}{3} v(F) - 2 \) first-order variables.

CCS Concepts: • Theory of computation → Complexity theory and logic; Finite Model Theory;

Additional Key Words and Phrases: The subgraph isomorphism problem, descriptive and computational complexity, first-order logic, quantifier depth, and variable width

ACM Reference format:
Oleg Verbitsky and Maksim Zhukovskii. 2019. Tight Bounds on the Asymptotic Descriptive Complexity of Subgraph Isomorphism. ACM Trans. Comput. Logic 20, 2, Article 9 (March 2019), 18 pages. https://doi.org/10.1145/3303881

1 INTRODUCTION

We consider graph properties expressible in first-order logic over the vocabulary consisting of two binary relations symbols, \( \sim \) for adjacency and = for equality of vertices. Let \( F \) be a fixed pattern graph on the vertex set \( \{1, \ldots, \ell\} \). The Subgraph Isomorphism problem asks whether a given graph contains a subgraph isomorphic to \( F \), which can be expressed by the first-order sentence

\[
\exists x_1 \ldots \exists x_\ell \left( \bigwedge_{i \neq j} x_i \neq x_j \land \bigwedge_{(i,j) \in E(F)} x_i \sim x_j \right),
\]

where \( E(F) \) denotes the edge set of \( F \).

Consider the parameters \( D(F) \) and \( W(F) \) defined, respectively, as the minimum quantifier depth and the minimum variable width of a first-order sentence expressing Subgraph Isomorphism for

*On leave from the IAPMM, Lviv, Ukraine
Oleg Verbitsky is supported by DFG grant KO 1053/8–1. Maksim Zhukovskii is supported by grant No. 18-71-00069 of Russian Science Foundation.

Authors’ addresses: O. Verbitsky, Institut für Informatik, Humboldt-Universität zu Berlin, Unter den Linden 6, Berlin, Germany, D-10099; email: verbitsk@informatik.hu-berlin.de; M. Zhukovskii, Moscow Institute of Physics and Technology, Laboratory of Advanced Combinatorics and Network Applications, Institutskii per. 9, Dolgoprudny, Moscow region, Russia; email: zhukmax@gmail.com.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
1529-3785/2019/03-ART9 $15.00
https://doi.org/10.1145/3303881

ACM Transactions on Computational Logic, Vol. 20, No. 2, Article 9. Publication date: March 2019.
the pattern graph $F$. Note the relation $W(F) \leq D(F)$, following from the general fact that any first-order sentence of quantifier depth $d$ can be rewritten using at most $d$ variables. Since the sentence (1) has quantifier depth $\ell$, we have $D(F) \leq \ell$. However, note that $K_\ell$, the complete graph on $\ell$ vertices, contains $F$ as a subgraph, while the smaller complete graph $K_{\ell-1}$ does not. Since $\ell$ first-order variables are necessary to distinguish between $K_\ell$ and $K_{\ell-1}$, we have $W(F) \geq \ell$ and, therefore, $W(F) = D(F) = \ell$. In other words, the existence of an $F$ subgraph cannot in general be expressed more succinctly—with respect to the quantifier depth or the variable width—than by the exhaustive description (1).

Assume that the pattern graph $F$ is connected. The time complexity of Subgraph Isomorphism will not be affected if we restrict this problem to connected input graphs. The same holds true for the descriptive complexity: we still need $\ell$ variables to express Subgraph Isomorphism over connected graphs, because the “hard case” was given by connected graphs $K_\ell$ and $K_{\ell-1}$. Can it happen, however, that this pair is the only obstacle to expressing Subgraph Isomorphism over connected graphs more succinctly? In fact, it is natural (in accordance with computational complexity theory) to consider the asymptotic descriptive complexity of Subgraph Isomorphism over connected graphs. More precisely, let $D'(F)$ be the minimum quantifier depth of a first-order sentence correctly deciding whether or not a graph $G$ contains an $F$ subgraph over all sufficiently large connected graphs $G$; see Section 2.1 for more details. Similarly, let $W'(F)$ denote the asymptotic version of the width parameter $W(F)$. The question addressed in this article is whether the asymptotic descriptive complexity of Subgraph Isomorphism can be much lower over connected graphs than in general or, more formally, how much the asymptotic parameters $D'(F)$ and $W'(F)$ can differ from their standard counterparts $D(F)$ and $W(F)$.

In our earlier paper [19], we found an example of a pattern graph $F$ with $D'(F) \leq \ell - 3$ and observed, however, that $W'(F) \geq \frac{1}{2} \ell - \frac{1}{2}$ for all $F$. It remained unknown whether the difference between $W'(F)$ and $W(F) = \ell$ could be arbitrarily large. We now show an infinite family of pattern graphs $F$ with $W'(F) \leq c \cdot \ell$ for a constant factor $c < 1$ and determine the minimum value of $c$, for which such a bound is possible. More precisely, we even show that

$$D'(F) \leq \frac{2}{3} \ell + 1$$

for infinitely many $F$, where $\ell$ always denotes the number of vertices in $F$. We also prove that this upper bound is tight (up to a small additive constant) by accompanying it with a nearly matching bound

$$W'(F) > \frac{2}{3} \ell - 2 \text{ for every } F.$$

A general reason why the existence of an $\ell$-vertex subgraph $F$ can be defined in this setting with sharply less than $\ell$ first-order variables lies in the fact that the logical truth changes if we restrict our realm to large connected graphs. In particular, some special sentences about subgraph containment become logically valid in this realm. As an example, consider the following statement:

$$\Phi_\ell$$

A graph $G$ contains a subgraph on $\ell$ vertices isomorphic either to a path $P_\ell$ or to a star $K_{1,\ell-1}$.

The observation that, for each $\ell$, the statement $\Phi_\ell$ is true for all sufficiently large connected graphs $G$ serves as a starting point of the graph-theoretic work by Oporowski, Oxley, and Thomas [13]; other examples of this kind can be found in Chapter 9.4 of Diestel’s textbook [3].

Though $\Phi_\ell$ implies no impressive upper bounds for $W'(P_\ell)$ nor for $W'(K_{1,\ell-1})$,\(^1\) this property underlies our analysis of an important hybrid pattern graph. Specifically, the sparkler graph $S_{q,p}$ is obtained by adding an edge between an end vertex of a path $P_p$ and the central vertex of a star

---

\(^1\)These parameters are actually equal to $\ell - 2$ and $\ell - 1$, respectively; see Section 5.
We determine the values of $D'(S_{q,p})$ and $W'(S_{q,p})$ up to a small additive constant. Specifically,

$$D'(S_{q,p}) \leq \max(p + 2, \frac{1}{2} p + q), \text{ while}$$

$$W'(S_{q,p}) \geq \max(p, \frac{1}{2} p + q - \frac{5}{2}).$$

The right-hand side of Inequality (4) attains its minimum when $p = 2q - 4$, yielding the upper bound of Inequality (2). Our proof of Inequality (4) exploits a structural property of connected $S_{q,p}$-free graphs closely related to the aforementioned properties $\Phi_{\ell}$: The maximum vertex degree of such graphs is bounded either from above by a constant or from below by an increasing function of $\ell = q + p$ (see Lemma 3.2). Another important ingredient in our analysis is a dichotomy theorem by Pikhurko, Veith, and Verbitsky [14] about succinct definability of an individual graph, stated as Lemma 2.4 in Section 2.

The lower bound of Inequality (5) readily implies that $W'(S_{q,p}) > \frac{2}{3} \ell - 2$, irrespective of the parameters $q$ and $p$. This particular fact about the sparkler graphs plays a key role in the proof of the general lower bound of Inequality (3). Our argument for Inequality (3) actually reveals the structure of extremal patterns $F$ with $W'(F) = \frac{2}{3} \ell$: Every such $F$ either has one of a few simple combinatorial properties or is a sparkler graph $S_{q,p}$ with $p \approx 2q$.

We conclude the summary of our results with listing some reasons motivating investigation of the asymptotic descriptive complexity of subgraph containment problems and, perhaps, other first-order properties of graphs.

**Relation to Computational Complexity.** The model-checking problem for a fixed first-order sentence $\Phi$ is solvable in time $O(n^{W(\Phi)})$, where $n$ is the number of vertices in an input graph and $W(\Phi)$ denotes the variable width of $\Phi$ (Immerman [8], Vardi [17]). This implies the time bound $O(n^{W(F)}) = O(n^\ell)$ for Subgraph Isomorphism for an $\ell$-vertex pattern graph $F$, which, of course, corresponds to exhaustive search through all $\ell$-tuples of vertices in the input graph. If $F$ is connected, then Subgraph Isomorphism efficiently reduces to its restriction to connected inputs, and the time bound $O(n^{W(F)})$ can be replaced with a potentially better bound $O(n^{W'(F)})$. According to Inequality (2), for some $F$ this results in time $O(n^{\frac{1}{2} \ell + 1})$, which actually may not look so bad if compared to the general time bound $O(n^{(\omega/3)\ell + 2})$ established for Subgraph Isomorphism by Nešetřil and Poljak [12]; here, $\omega$ is the exponent of fast matrix multiplication, whose value is known [5] to lie between 2 and 2.373. However, all patterns $F$ with $W'(F) = \frac{2}{3} \ell$ have a large tree part and, for such graphs, the time bound $O(n^{\frac{1}{2} \ell + 1})$ cannot compete with other algorithmic techniques for Subgraph Isomorphism, such as the color-coding method by Alon, Yuster, and Zwick [1].

Nevertheless, there is apparently no general reason why the time bounds based on estimating the asymptotic descriptive complexity cannot be somewhat more successful in some other

\[\text{Fig. 1. Examples of useful graphs: a sparkler and a broken fan.}\]
situations. Consider, for example, the \textit{Induced} Subgraph Isomorphism problem, whose computational complexity seems very different from the not necessarily induced case. Here, it is unknown if the bound $O(n^{W(F)})$ can, for some patterns $F$, be comparable with the Nešetřil–Poljak time bound. The last bound applies also to Induced Subgraph Isomorphism, and no techniques achieving running time $O(n^{a(f)})$ with a sublinear exponent are known, not even for any particular sequence of patterns. We discuss this topic in Reference [18]; see also the concluding remarks in Section 6.

\textit{Encoding-independent Computations and Order-invariant Definitions}. The preceding discussion shows that any first-order sentence in the vocabulary \{\sim, =\} defining a graph property over sufficiently large graphs can be considered a weak computational model for the corresponding decision problem. The question of its efficiency is in the spirit of the eminent problem on the power of encoding-independent computations; see, e.g., Reference [6]. If we extend the vocabulary with the order relation $<$, then comparison of the two settings is interesting in the context of \textit{order-invariant definitions}; see, e.g., References [11, 15]. Note also that the setting where arbitrary numerical relations are allowed brings us into the field of circuit complexity; see References [9, 11]. In this context, the Subgraph Isomorphism problem has been studied in Reference [10]. Our result reveals the somewhat surprising fact that exhaustive search can be beaten for some pattern graphs even in the weakest possible situation.

\textit{Measurement of Succinctness}. A traditional question studied in finite model theory asks whether or not a graph property of interest is expressible in a certain logical formalism. If the expressibility is known, then it is reasonable to ask how succinctly the property can be expressed with respect to the length, the quantifier depth, or the variable width of a defining sentence, and the questions we address in this article are exactly of this kind. We refer the reader to Grohe and Schweikardt [7] and Turán [16] for instances of the problems studied in this line of research.

\textit{Organization of the Article}. Section 2 contains definitions and preliminary lemmas. The Bounds (4) and (5) for sparkler graphs are proved in Section 3. Our main result, the general bounds given by Inequalities (2) and (3), is obtained in Section 4. Furthermore, Section 5 is devoted to the particular case of path and star graphs, in which we are able to determine the values of $D'(F)$ and $W'(F)$ precisely. We conclude with a discussion of further questions in Section 6.

2 PRELIMINARIES

2.1 Basic Definitions

We consider the first-order language containing the adjacency and the equality relations. We say that a first-order sentence $\Phi$ defines a class of graphs $C$ \textit{asymptotically over connected graphs} if there is an integer $N$, such that

$$G \models \Phi \text{ iff } G \in C$$

for all connected graphs $G$ with at least $N$ vertices. The \textit{asymptotic logical depth} of $C$, denoted by $D'(C)$, is the minimum quantifier depth (rank) of such $\Phi$.

The \textit{variable width} of a first-order sentence $\Phi$ is the number of first-order variables used to build $\Phi$; different occurrences of the same variable do not count. The \textit{asymptotic logical width} of $C$, denoted by $W'(C)$, is the minimum variable width of a $\Phi$ defining $C$ asymptotically over connected graphs. Note that $W'(C) \leq D'(C)$.

2.2 Our Toolkit

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 distinguishing $G$ and $H$; that is, true on one of the graphs and false on the other.
Lemma 2.1.

1. $D'(C) \leq d$ if $D(G, H) \leq d$ for all sufficiently large connected graphs $G \in C$ and $H \notin C$.
2. $W'(C) \geq d$ if there are arbitrarily large connected graphs $G \in C$ and $H \notin C$ with $W(G, H) \geq d$.

Part 1 of this lemma can easily be deduced from the fact that, over a given vocabulary, there are only finitely many pairwise inequivalent first-order formulas of a fixed quantifier depth. Part 2 follows directly from the definitions.

Lemma 2.1 reduces estimating $D'(C)$ and $W'(C)$ to estimating, respectively, the parameters $D(G, H)$ and $W(G, H)$ over large connected $G \in C$ and $H \notin C$. For estimating $D(G, H)$ and $W(G, H)$, we have a very handy instrument.

The $k$-pebble Ehrenfeucht-Fraïssé game is played on two vertex-disjoint graphs $G$ and $H$. This is a two-person game; the players are called Spoiler and Duplicator. He and she have equal sets of $k$ pairwise different pebbles. In each round, Spoiler takes a pebble and puts it on a vertex in $G$ or in $H$; then, Duplicator has to put her copy of this pebble on a vertex of the other graph. The pebbles can be reused and change their positions during play. Duplicator’s objective is to ensure that the pebbling determines a partial isomorphism between $G$ and $H$ after each round; when she fails, she immediately loses. The proof of the following facts can be found in Immerman’s textbook [9]:

Lemma 2.2.

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 $H$.

For bounding $W'(C)$ from below based on Lemma 2.1, we need a supply of pairs of graphs $G, H$ with large $W(G, H)$. The following simple construction works quite efficiently sometimes:

Let $u, v$, and $w$ be three pairwise different vertices of a graph. We say that $w$ distinguishes $u$ and $v$ if $w$ is adjacent to exactly one of these two vertices. We call $u$ and $v$ twins if these vertices are indistinguishable by any vertex $w$ or, equivalently, if the transposition $(uv)$ is an automorphism of the graph. We consider $u$ and $v$ twins also if $u = v$. Being twins is an equivalence relation on the vertex set, and every equivalence class, called twin class, is a clique or an independent set.

Though completely elementary, the concept of the twin relation is amazingly useful in many different contexts; see, e.g., Reference [2, Section 2]. We exploit it in various ways while proving both our lower and upper bounds. For a vertex $v$ in a graph $G$, let $G - v$ denote the graph obtained by removing $v$ from $G$.

Lemma 2.3. Suppose that $G$ is a graph containing a twin class of size $t$. Let $v$ be one of these $t$ twins. Then, $W(G, G - v) \geq t$.

Proof. By Part 2 of Lemma 2.2, it suffices to notice that Duplicator survives in the $(t - 1)$-pebble game on $G$ and $G - v$, whatever the number of rounds is played. Whenever Spoiler plays outside the twin class under consideration, Duplicator just mirrors his moves in the other graph. Whenever Spoiler pebbles one of the twins, Duplicator makes the same in the other graph; the particular choice of a twin is immaterial. This is always possible, because the reduced twin class in $G - v$ contains as many vertices as the number of pebbles.

Let $\sigma(G)$ denote the maximum size of a twin class in a graph $G$. Lemma 2.3 implies that we need at least $\sigma(G)$ first-order variables to define $G$ as an individual graph, i.e., to distinguish $G$ from all non-isomorphic graphs. As a consequence, the quantifier depth needed for this purpose cannot be
smaller than \( \sigma(G) \). It turns out that this is the only reason why it can be large. More specifically, the following result shows that every \( n \)-vertex graph \( G \) either is definable with quantifier depth no larger than \( \frac{1}{2} n + \frac{1}{2} \) or has \( \sigma(G) > \frac{1}{2} n + \frac{1}{2} \), and in the latter case the minimum quantifier depth of a sentence defining \( G \) is very close to \( \sigma(G) \).

**Lemma 2.4 (Pikhurko, Veith, and Verbitsky [14, Theorem 4.1]).** If \( A \) and \( B \) are non-isomorphic graphs, then

\[
D(A, B) \leq \begin{cases} 
\frac{1}{2} v(A) + \frac{5}{2} & \text{if } \sigma(A) \leq \frac{1}{2} v(A) + \frac{1}{2}, \\
\sigma(A) + 2 & \text{if } \sigma(A) > \frac{1}{2} v(A) + \frac{1}{2}.
\end{cases}
\]

Moreover, if \( \sigma(A) \geq \frac{1}{2} v(A) + 1 \), then \( D(A, B) \leq \sigma(A) + 1 \) whenever the largest twin class in \( A \) is an inclusion-maximal clique or an inclusion-maximal independent set.

Here and throughout the article, \( v(G) \) denotes the number of vertices in a graph \( G \).

### 2.3 Useful Graphs

The neighborhood \( N(v) \) of a vertex \( v \) consists of all vertices adjacent to \( v \). The number of neighbors \( |N(v)| \) is called the degree of \( v \) and denoted by \( \deg v \). A vertex of degree \( \deg v(G) - 1 \) (i.e., adjacent to all other vertices) is called universal. A vertex of degree 1 is called pendant.

We use the standard notation \( P_n \) for paths and \( C_n \) for cycles on \( n \) vertices. Furthermore, \( K_{t,s} \) denotes the complete bipartite graph whose vertex classes have \( t \) and \( s \) vertices. In particular, \( K_{1,n-1} \) is the star graph on \( n \) vertices.

The sparkler graph \( S_{q,p} \), playing an important role in the article, is obtained from \( K_{1,q-1} \) and \( P_p \) by adding an edge between an end vertex of \( P_p \) and the central vertex of \( K_{1,q-1} \). The components \( K_{1,q-1} \) and \( P_p \) are referred to as the star and tail parts of the sparkler graph, respectively. By the central vertex of \( S_{q,p} \), we mean the central vertex of the star part. We will almost always assume that \( q \geq 3 \) and \( p \geq 2 \), though the smaller parameters still make sense representing \( S_{q,1} \equiv K_{1,q} \) and \( S_{2,p} \equiv P_{p+1} \).

The following graph will appear in Section 3.2 for a technical purpose. Recall that the \( n \)-vertex fan graph is obtained by adding a universal vertex \( y \) to the path graph \( P_{n-1} \). Let us “break” the edge between \( y \) and one of the end vertices of \( P_{n-1} \) by removing this edge from the graph and replacing it with an edge from \( y \) to a new vertex \( y' \). This results in an \( (n+1) \)-vertex graph that we call broken fan and denote by \( B_{n+1} \); see Figure 1.

### 2.4 First Bounds

Let \( S(F) \) denote the class of graphs containing a subgraph isomorphic to \( F \). Simplifying the general notation introduced in Subsection 2.1, we write \( D'(F) = D'(S(F)) \) and \( W'(F) = W'(S(F)) \). We now state simple combinatorial bounds for these parameters that were used already in Reference [19].

Let \( v_0 v_1 \ldots v_t \) be a path in a graph \( G \). We call it pendant if \( \deg v_0 \neq 2, \deg v_i = 1, \) 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 \( v_0 \). Note that any pendant path is an induced path and any pendant star is an induced star subgraph of \( G \). The definition ensures that a pendant path (or star) cannot be contained in a larger pendant path (or star).

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 \( K_{1,s} \). As an example, note that the sparkler graph \( S_{q,p} \) has a pendant \( P_{p+1} \) and a pendant \( K_{1,q-1} \) and, therefore, \( p(S_{q,p}) = p \) and \( s(S_{q,p}) = q - 1 \). If \( F \) has no pendant vertex, then we set \( p(F) = 0 \) and \( s(F) = 0 \).
LEMMA 2.5. For every connected graph $F$ with $\ell$ vertices,

1. $W'(F) \geq \ell - s(F) - 1$, and
2. $W'(F) \geq \ell - p(F) - 1$.

Proof. 1. Since $s(K_{1,\ell-1}) = \ell - 1$, the bound is trivial for $F = K_{1,\ell-1}$, and we assume that $F$ is not a star. By Part 2 of Lemma 2.1, it suffices to exhibit, for every sufficiently large $n$, a pair of graphs $G_n$ and $H_n$ with at least $n$ vertices each, such that $G_n$ contains $F$ as a subgraph, $H_n$ does not, and $W(G_n, H_n) \geq \ell - s(F) - 1$. For this purpose, consider $G_n$, consisting of the complete graph $K_{\ell-s(F)}$ with a pendant star $K_{1,n}$ and $H_n$ obtained in the same way from the smaller complete graph $K_{\ell-s(F)-1}$. Note that $H_n$ is obtainable from $G_n$ by removing one of $\ell - s(F) - 1$ twins. The lower bound for $W(G_n, H_n)$ follows by Lemma 2.3.

2. Assume that $F$ is not a path graph, because otherwise the bound is trivial. We proceed similarly to the first part, considering now $G_n$ consisting of $K_{\ell-s(F)}$ with a pendant path $P_n$ and $H_n$ obtained from $G_n$ by removing one of the twins. \hfill \square

3 BOUNDS FOR SPARKLER GRAPHS

We begin our analysis with a very instructive case of sparkler graphs, for which we are able to determine the asymptotic depth and width parameters with high precision.

THEOREM 3.1. Let $q \geq 3$ and $p \geq 2$. Let $\ell = q + p$ denote the number of vertices of the sparkler graph $S_{q,p}$. Then,

1. $W'(S_{q,p}) \geq \max(p, \ell - \frac{1}{2}p - 2 - \frac{1}{2}(p \bmod 2)).$
2. $D'(S_{q,p}) \leq \max(p, \ell - \frac{1}{2}p - 2 - \frac{1}{2}(p \bmod 2)) + 2.$

The proof of Theorem 3.1 occupies the rest of this section.

3.1 The Lower Bound

Applied to a sparkler graph, Part 1 of Lemma 2.5 yields $W'(S_{q,p}) \geq \ell - s(S_{q,p}) - 1 = \ell - q = p$. It remains to prove the lower bound

$$W'(S_{q,p}) \geq \ell - \frac{1}{2}p - 2 - \frac{1}{2}(p \bmod 2) = q + \frac{1}{2}p - 2 - \frac{1}{2}(p \bmod 2). \quad (6)$$

If $p = 2$ or $p = 3$, then this bound immediately follows from Part 2 of Lemma 2.5. We, therefore, assume that $p \geq 4$.

To prove the bound (6), we will construct two connected graphs $G_{a,b,n}$ and $H_{a,b,n}$ depending on integer parameters $a$, $b$, and $n$. Adjusting appropriately the parameters $a$ and $b$, we will ensure that

(a) $G_{a,b,n}$ contains $S_{q,p}$ as a subgraph,
(b) $H_{a,b,n}$ does not, and
(c) $W(G_{a,b,n}, H_{a,b,n}) \geq q + \frac{1}{2}p - 2 - \frac{1}{2}(p \bmod 2)$

for any choice of the parameter $n$. Moreover,

(d) both $G_{a,b,n}$ and $H_{a,b,n}$ have more than $n$ vertices

by construction. Properties (a)–(d) will yield Bound (6) by Part 2 of Lemma 2.1, as the parameter $n$ can be taken arbitrarily large.

Set $s = q - 1$. The construction of $G_{a,b,n}$ is illustrated in Figure 2. This graph is composed by two subgraphs sharing one common vertex. One of them is the complete bipartite graph $K_{a+1,a+s}$, which will be referred to as the $K$-component. The other is the sparkler graph $S_{n+1,b}$, which will be
Fig. 2. Construction of $G_{a,b,n}$ by gluing $K_{a+1,a+s}$ and $S_{n+1,b}$ at the central vertex $w$ of the latter. Like the shown subgraph $S_{3,7}$ (with central vertex $u$), the graph $G_{a,b,n}$ hosts a copy of $S_{s+1,2a+b}$ as a subgraph.

referred to as the $S$-component. The smaller part of the $K$-component includes the central vertex $w$ of the $S$-component.

Note that the larger part of the $K$-component is an independent set consisting of $a+s$ twins. The graph $H_{a,b,n}$ is obtained from $G_{a,b,n}$ by removing one vertex from this twin class. Lemma 2.3 implies that

$$W(G_{a,b,n}, H_{a,b,n}) \geq a + s. \quad (7)$$

Now, we fix the parameters $a$ and $b$, such that Conditions (a)–(c) are fulfilled. Specifically, we set

$$b = 2 + (p \mod 2) \text{ and } a = \frac{p - b}{2}.$$  

Note that $b \in \{2, 3\}$. The value of $a$ is an integer due to the choice of $b$. Our assumption that $p \geq 4$ ensures that $a \geq 1$. Note that Condition (c) now readily follows from Inequality (7).

As easily seen, the graph $G_{a,b,n}$ contains a copy of $S_{q,p}$. Indeed, we can put the central vertex at any vertex $u \neq w$ in the smaller part of the $K$-component. The larger part of the $K$-component is large enough to accommodate $s$ vertices adjacent to $u$ and, moreover, there still remains enough space to draw a path from $u$ to $w$ via $2a - 1$ intermediate vertices. Prolonging it along the tail part of the $S$-component, we obtain a path emanating from $u$ and passing through $2a + b = p$ further vertices.

This path is clearly destroyed by removing a vertex from the larger part of the $K$-component. It remains to prove that no copy of $S_{q,p}$ can be found in $H_{a,b,n}$ in any other way. Assume, to the contrary, that $H_{a,b,n}$ contains an isomorphic copy $S'$ of $S_{q,p}$ with central vertex $u'$. Consider several cases.

Case 1: $u' \neq w$ belongs to the smaller part of the $K$-component of $H_{a,b,n}$. After locating the star part of $S'$ in the larger part of the $K$-component, this part contains only $a-1$ unoccupied vertices. It follows that the longest path that can be drawn from $u'$ has length $2a - 2 + b = p - 2$.

Case 2: $u' = w$. Now the star part of $S'$ can be located among the $n$ pendant vertices of $H_{a,b,n}$ adjacent to $w$. However, by the assumption that $p \geq 4$, the tail part of $S'$ does not fit into the
tail part of the $S$-component of $H_{a,b,n}$. The longest path from $w$ in the $K$-component has length $2a + 1 < 2a + b = p$.

Case 3: $u'$ belongs to the larger part of the $K$-component of $H_{a,b,n}$. Suppose that $s < a + 1$, because otherwise the star part of $S'$ does not fit into the smaller part of the $K$-component. After locating the star part of $S'$ in this part, the longest path from $u'$ that can be found in the $K$-component has length $2(a + 1 - s) < 2a < p$ if it terminates in the larger part of the $K$-component. Otherwise, such a path has length at most $2(a + 1 - s) - 1$ and, arriving at $w$, can be prolonged in the $S$-component of $H_{a,b,n}$ to a path of total length at most $2(a + 1 - s) - 1 + b = p + 1 - 2s < p$.

We get a contradiction in each of the cases, which completes our analysis. The proof of Part 1 of Theorem 3.1 is complete.

3.2 The Upper Bound

We now turn to proving the upper bound of Theorem 3.1. We begin with a few simple properties of graphs without $S_{q,p}$ subgraphs. The following lemma generalizes a property of $S_{q,p}$-free graphs observed in Reference [19]. As usual, $\Delta(H)$ denotes the maximum vertex degree in a graph $H$.

**Lemma 3.2.** Suppose that a connected graph $H$ does not contain a subgraph $S_{q,p}$. Then, either $\Delta(H) < q$ or $\Delta(H) \geq (v(H)/3)^{1/(2qp)}$.

**Proof.** Assume that $\Delta(H) \geq q$; that is, $H$ contains a subgraph $K_{1,q}$. Then, $H$ cannot contain any subgraph $P_{2qp}$, because, together with $K_{1,q}$, it would give an $S_{q,p}$ subgraph. Indeed, consider a shortest path joining the central vertex of $K_{1,q}$ with a vertex of $P_{2qp}$. Combining this path with a part of $P_{2qp}$, we get a path of length at least $qp$ starting in the central vertex of $K_{1,q}$. If this path contains leaves of the $K_{1,q}$ subgraph, then they split the $P_{2qp}$ into parts, one of which is long enough to form the tail of an $S_{q,p}$ subgraph.

Consider an arbitrary spanning tree $T$ in $H$ and denote its maximum vertex degree $d$ and its radius by $r$. Note that

$$v(H) = v(T) \leq 1 + d + d(d - 1) + \cdots + d(d - 1)^{r-1} \leq 3d^r,$$

Since $r \leq 2qp$, we have $v(H) \leq 3d^{2qp}$. It follows that $\Delta(H) \geq d \geq (v(H)/3)^{1/(2qp)}$. □

**Lemma 3.3.** Suppose that a connected graph $H$ with at least $3(q + p)^2qp$ vertices does not contain a subgraph $S_{q,p}$. If $\Delta(H) \geq q$, then $H$ contains neither the cycle $C_{p+1}$ nor the broken fan $B_{p+2}$ as subgraphs.

**Proof.** By Lemma 3.2, $H$ has a vertex $z$ of degree at least $(v(H)/3)^{1/(2qp)} \geq q + p$. Assume, to the contrary, that $H$ contains $C_{p+1}$ or $B_{p+2}$. We show that such a subgraph could be used for building an $S_{q,p}$ subgraph with central vertex $z$ (which is impossible by assumption). Indeed, if $z$ belongs to a cycle of length $p + 1$, then the cycle provides the tail part of $S_{q,p}$, and $z$ has sufficiently many neighbors to build also the star part of $S_{q,p}$. If there is a $C_{p+1}$ subgraph not containing $z$, then we can even find $S_{q+1, p}$ in $H$ by considering an arbitrary path from $z$ to this cycle.

If $H$ contains a $B_{p+2}$ subgraph, denote its vertices by $y, y', y_1, y_2, \ldots, y_p$ where $y_1, y_2, \ldots, y_p$ appear (in this order) along the path part of $B_{p+2}$, then, $y$ is adjacent to all of these vertices but $y_p$, and $y'$ is adjacent to $y$. If $z = y_i$ with $1 < i < p - 1$, then the tail part of $S_{q,p}$ in $H$ is formed by the path $zy_{i-1} \cdots y_i y_{i+1} \cdots y_p$. The case of $z = y_1$ is similar. If $z = y_{p-1}$, then the tail part of $S_{q,p}$ is $zy_{p-2} \cdots y_i y_{i+1}$. If $z = y$, then the tail is $zy_{i-1} \cdots y_p$. If $z$ is $y'$ or $y_p$ or does not belong to the $B_{p+2}$ subgraph, then we can find a tail even longer than needed. □

We now restate Part 2 of Theorem 3.1 as

$$D'(S_{q,p}) \leq \max \left( p + 2, q + \frac{1}{2} p - \frac{1}{2} (p \mod 2) \right).$$

(8)
Let $G$ and $H$ be two graphs, such that $G$ contains $S_{q,p}$ as a subgraph and $H$ does not. Moreover, assume that $H$ is connected and sufficiently large; specifically, $\nu(H) \geq (q+p)^{2q/p}$. Using Part 1 of Lemma 2.1, we have to design a strategy allowing Spoiler to win the Ehrenfeucht game on $G$ and $H$ so fast that the number of moves corresponds to the Bound (8).

Let $S$ be a subgraph of $G$ isomorphic to $S_{q,p}$. Denote the central vertex of $S$ by $x$, and let Spoiler pebble $x$ in the first round. Let $y$ denote the vertex pebbled in response by Duplicator in $H$. If $\deg y < q$, then Spoiler wins in the next $q$ rounds by pebbling $q$ neighbors of $x$. We, therefore, suppose that

$$\deg y \geq q. \quad (9)$$

This implies that $\Delta(H) \geq q$ and makes Lemma 3.3 applicable to $H$.

Let $xx_1x_2 \ldots x_p$ be the tail path of $S$. If $x$ and $x_p$ are adjacent, then this yields a $C_{p+1}$ subgraph in $G$, and Spoiler wins by pebbling it due to Lemma 3.3. We, therefore, suppose that $x$ and $x_p$ are not adjacent.

Denote the number of neighbors of $x$ among $x_1, x_2, \ldots, x_{p-1}$ by $r$. We also suppose that Spoiler cannot win in $p+1$ moves just by pebbling the tail part of $S$; that is, Duplicator manages to keep a partial isomorphism between $G$ and $H$ in this case by pebbling some path $y_1y_2 \ldots y_p$ in $H$. Note that $y$ cannot be adjacent to $y_p$ and must have exactly $r$ neighbors among $y_1, y_2, \ldots, y_{p-1}$. Since $y$ cannot be the central vertex of any $S_{q,p}$ subgraph in $H$, this implies that

$$\deg y < r + q - 1. \quad (10)$$

By definition, we have $r \leq p - 1$. In fact, this inequality is strict under the additional condition that $q \geq p$. Indeed, the equality $r = p - 1$ means that $y$ is adjacent to all $y_1, y_2, \ldots, y_{p-1}$. In the case that $q \geq p$, Inequality (9) shows that $y$ also has a neighbor $y'$ different from any $y_i$, $1 \leq i \leq p$. This yields a $B_{p+2}$ subgraph in $H$, a contradiction.

Therefore, Inequality (10) implies that

$$\deg y \leq q + p - 3 \quad (11)$$

and, moreover,

$$\deg y \leq q + p - 4 \text{ if } q \geq p. \quad (12)$$

Let $G' = G[N(x)]$ and $H' = H[N(y)]$, where $G[X]$ denotes the induced subgraph of $G$ on a set of vertices $X$. Thus, $G'$ has $d x$ vertices and $H'$ has $d y$ vertices. Whenever Spoiler moves in $G'$, Duplicator is forced to respond in $H'$ and vice versa, because otherwise she loses immediately. For the graphs $G'$ and $H'$, we now apply Lemma 2.4.

Set $A = H'$ and $B = G'$, and consider several cases. Assume first that $\sigma(H') \leq \frac{1}{2} \nu(H') + \frac{1}{2}$. By Lemma 2.4,

$$D(H', G') \leq \frac{1}{2} \nu(H') + \frac{5}{2}.$$

It follows by Inequality (11) that $D(H', G') \leq \frac{1}{2} q + \frac{1}{2} p + 1$; that is, Spoiler wins the game on $G'$ and $H'$ making no more than $\frac{1}{2} q + \frac{1}{2} p + 1$ moves. This allows him to win the game on $G$ and $H$ making no more than $\frac{1}{2} q + \frac{1}{2} p + 2$ moves, implying

$$D(G, H) \leq \frac{1}{2} q + \frac{1}{2} p + 2. \quad (13)$$

This yields Bound (8) if $\frac{1}{2} q + \frac{1}{2} p + 2 \leq p + 2$. The last inequality is false exactly when $q > p$. But then we can use Inequality (12) and, similar to Inequality (13), obtain a slightly better bound, namely

$$D(G, H) \leq \frac{1}{2} q + \frac{1}{2} p + \frac{3}{2}.$$
Again, this implies Bound (8) if \( \frac{1}{2} q + \frac{1}{2} p + \frac{3}{2} \leq p + 2 \). The last inequality is false exactly when \( q \geq p + 2 \). In this case, Bound (8) is also true, because \( \frac{1}{2} q + \frac{1}{2} p + \frac{3}{2} \leq q + \frac{1}{2} p - \frac{1}{2} (p \mod 2) \). The last inequality is true, because otherwise we would have \( q < 3 + (p \mod 2) \leq 4 \), contradicting the condition \( q \geq p + 2 \geq 4 \).

Assume now that \( \sigma(H') \geq \frac{1}{2} v(H') + 1 \) and denote the largest twin class in \( H' \) by \( T \). By Lemma 2.4, we now have

\[
D(H', G') \leq \sigma(H') + 2.
\]

Our further analysis depends on whether \( T \) is a clique or an independent set.

Assume the former. Note that \( |T| \leq p - 1 \). Indeed, the clique \( T \) cannot contain more vertices, or else \( T \cup \{y\} \) would be a clique of size at least \( p + 1 \) in \( H \), and \( H \) would hence contain a \( C_{p+1} \) subgraph, contradicting Lemma 3.3. Thus, \( \sigma(H') \leq p - 1 \). Therefore, \( D(H', G') \leq p + 1 \) and \( D(H, G) \leq p + 2 \), and Bound (8) follows.

Assume now that the largest twin class in \( H' \) is an independent set. By the assumption we made previously, \( H \) contains a path \( y_1y_2 \ldots y_{p-1} \), such that at most \( q - 2 \) vertices of \( N(y) \) do not belong to this path. It follows that

\[
\sigma(H') \leq q - 2 + \alpha(P_{p-1}).
\]

Here, \( \alpha(K) \) denotes the independence number of a graph \( K \); that is,

\[
\alpha(P_{p-1}) = \frac{1}{2} p - \frac{1}{2} (p \mod 2).
\]

If Inequality (14) is strict, then we obtain \( D(H', G') \leq \sigma(H') + 2 \leq q - 1 + \alpha(P_{p-1}) \) and

\[
D(H, G) \leq q + \frac{1}{2} p - \frac{1}{2} (p \mod 2).
\]

It remains to consider the case that \( \sigma(H') = q - 2 + \alpha(P_{p-1}) \). This equality implies that the largest twin class in \( H' \) is an inclusion-maximal independent set. Under this condition, Lemma 2.4 gives us the better relation \( D(H', G') \leq \sigma(H') + 1 \), and the same bound for \( D(H, G) \) stays true.

The proof of Theorem 3.1 is complete.

## 4 GENERAL BOUNDS

We are now prepared to prove our main result.

**Theorem 4.1.**

1. \( D'(F) \leq \frac{2}{3} v(F) + 1 \) for infinitely many connected \( F \).
2. \( W'(F) > \frac{2}{3} v(F) - 2 \) for every connected \( F \).

We prove Theorem 4.1 in the rest of this section. Part 1 readily follows from Part 2 of Theorem 3.1 by setting \( q = t + 2 \) and \( p = 2t - 2 \).

To prove the lower bound, we introduce the following notion: Let \( v \) be a vertex of a connected graph \( F \). By a \( v \)-branch of \( F \), we mean a subgraph of \( F \) induced by the vertex set of a connected component of \( F \setminus v \) in union with the vertex \( v \) itself. Let \( S \) be a \( v \)-branch of \( F \). We call \( S \) a pendant sparkler subgraph \( S_{q, p} \) of \( F \) if
• $S$ is isomorphic to $S_{q,p}$.
• $v$ is the end vertex of the tail part of $S$, and
• $v$ has degree at least 3 or exactly 1 in $F$.

Note that $\deg v = 1$ means that $F = S$; that is, $F$ is a sparkler graph itself. As usual, we suppose that $q \geq 3$. However, we allow $p = 1$ ($S$ is a star $K_{1,q}$ attached at one of its leaves to the rest of the graph). In addition, we even allow $p = 0$ ($S$ is a pendant star $K_{1,q-1}$ attached at the central vertex). In the case that $F$ has at least one pendant sparkler subgraph, we write $sp(F)$ to denote the maximum $p$, such that $F$ has a pendant $S_{q,p}$ for some $q \geq 3$.

**Lemma 4.2.** Let $F$ be a connected graph with $\ell$ vertices. If $F$ has a pendant sparkler subgraph, then $W'(F) \geq \ell - sp(F) - 3$.

**Proof.** We suppose that $F$ is not a sparkler graph, because otherwise the required bound readily follows from Part 2 of Lemma 2.5.

Denote $p = sp(F)$. Let $n \geq 3$ be an integer parameter. Let $G$ be obtained from $K_{\ell}$ and $S_{n,p+1}$ by gluing these graphs at the end vertex of the tail of $S_{n,p+1}$. Thus, $G$ can be seen as a pendant $S_{n,p+1}$ attached to $K_{\ell}$. Let $H$ be obtained similarly from $K_{\ell-p-3}$ and $S_{n,p+1}$.

Note that $G$ contains $F$ as a subgraph only because the $K_{\ell}$ part of $G$ is large enough to contain any graph with $\ell$ vertices. However, $H$ does not contain an $F$ subgraph. Indeed, since the tail of the $S_{n,p+1}$ part of $H$ is larger than $p$, it can host at most a path of length $p + 2$ (which is useful if $F$ has a pendant path). However, the remaining part of $H$ has at least $\ell - p - 2$ vertices and does not fit into the $K_{\ell-p-3}$ part of $H$.

Now, note that the $K_{\ell}$ part of $G$ with the end vertex of $S_{n,p+1}$ excluded is a twin class in $G$. Moreover, $H$ is obtained from $G$ by removing $p + 3$ twins from this class. Similar to Lemma 2.3, we have $W(G,H) \geq \ell - p - 3$ (which can also be formally deduced from Lemma 2.3 by repeatedly applying it $p + 3$ times and using transitivity of the logical equivalence in the $(\ell - p - 4)$-variable logic). Since the parameter $n$ can be chosen arbitrarily large, we conclude that $W'(F) \geq \ell - p - 3$. \hfill \Box

We are now ready to prove Part 2 of Theorem 4.1. Let $\ell = v(F)$ and assume that $W'(F) < \frac{2}{3} \ell - 1$, as otherwise we are done. By Lemma 2.5, this implies that

\[
s(F) > \frac{1}{3} \ell \quad \text{and} \quad p(F) > \frac{1}{3} \ell.
\]

The former estimate implies that $F$ has a pendant sparkler subgraph. Lemma 4.2 is, therefore, applicable. It implies that

\[
sp(F) > \frac{1}{3} \ell - 2.
\]

From Inequalities (15)–(17), we conclude that $F$ contains a pendant star $F_1$, a pendant path $F_2$, and a pendant sparkler $F_3$, such that each of these subgraphs has more than $\frac{1}{3} \ell + 1$ vertices. If $F_1$ and $F_3$ are vertex-disjoint or share only one vertex, which must be the central vertex of $F_1$ and the end vertex of the tail part of $F_3$, then these subgraphs together occupy more than $\frac{2}{3} \ell + 1$ vertices and there does not remain enough space for the pendant path $F_2$. Therefore, $F_1$ and $F_3$ share at least two vertices, which is possible only if $F_1$ coincides with the star part of $F_3$. In this case, $F_3$ contains more than $\frac{2}{3} \ell - 1$ vertices. It follows that $F_3$ and $F_2$ cannot be disjoint, or else they together would occupy more than $\ell$ vertices. Therefore, $F_3$ and $F_2$ overlap, which is possible only if $F_2$ is the tail part of $F_3$; that is, $F = F_3$ is a sparkler graph.
Now, let $F = S_{q,p}$. By Part 1 of Theorem 3.1,
\[ W'(S_{q,p}) \geq \max\left(p, \ell - \frac{1}{2} p - \frac{5}{2}\right). \]
As easily seen, the minimum of the right-hand side over real $p \in (0, \ell)$ is attained at $p = \frac{2}{3} \ell - \frac{5}{3}$.
It follows that $W'(F) > \frac{2}{3} \ell - 2$, completing the proof.

5 Exact Values for Stars and Paths

Theorem 3.1 shows that, for sparkler graphs, the parameters $W'(S_{q,p})$ and $D'(S_{q,p})$ take values in a segment of three consecutive integers (even two if $p$ is odd). We now determine the exact values of $W'(F)$ and $D'(F)$ for the two kinds of “marginal” sparklers, namely for the star graphs $K_{1,\ell-1}$ and the path graphs $P_\ell$.

We begin with the stars. Note that $D'(K_{1,2}) = 1$, because every connected graph with at least 3 vertices contains $K_{1,2} = P_3$ as a subgraph.

**Theorem 5.1.** Let $s \geq 3$. Then,
\begin{enumerate}
  \item $D'(K_{1,s}) = s + 1$;
  \item $W'(K_{1,s}) = s$.
\end{enumerate}

**Proof.** Let $M_{s,t}$ denote the graph obtained by subdividing each edge of $K_{1,s}$ into $P_{t+1}$; thus, $v(M_{s,t}) = st + 1$.

1. Consider $G = M_{s,t}$ and $H = M_{s-1,t}$. Obviously, $G$ contains $K_{1,s}$, while $H$ does not. We now describe a strategy for Duplicator allowing her to win the $s$-round game on $G$ and $H$, irrespective of how large $t$ is.

Note that $G$ is obtained from $H$ by adding another, $s$th branch $P_t$. During the first $s - 1$ rounds, at least one branch of $G$ stays free of pebbles, and this allows Duplicator to use a mirror strategy according to an isomorphism between $H$ and a subgraph of $G$. Note that, according to this strategy, Duplicator pebbles the central vertex of one graph exactly when Spoiler pebbles the central vertex in the other graph. To describe the last, $s$th round of the game, we consider two cases:

Case 1: The central vertex was pebbled in the first $s - 1$ rounds. Then there is at least one completely free branch in $H$ and at least two such branches in $G$. Duplicator can use the mirror strategy in the $s$th round, too.

Case 2: The central vertex was not pebbled in the first $s - 1$ rounds. This case is different from the previous one only when each of the $s - 1$ branches of $H$ contains a pebbled vertex and exactly one branch remains free in $G$. Assuming this, we consider two subcases:

Subcase 2-a: Spoiler does not move in the free branch of $G$. The mirroring strategy is still available to Duplicator. Note that this case includes the possibility that Spoiler pebbles the central vertex in one of the graphs.

Subcase 2-b: Spoiler pebbles a vertex $u$ in the free branch of $G$. In this case, the central vertices of $G$ and $H$ are not pebbled. This implies that $u$ is not adjacent to any other vertex pebbled in $G$. If $t \geq 4$, then Duplicator is able to find a vertex $u'$ in $H$ not adjacent to the vertices previously pebbled in this graph.

2. The lower bound $W'(K_{1,s}) \geq s$. Like in the proof of Part 1, take $G = M_{s,t}$, $H = M_{s-1,t}$, and notice that Duplicator wins the $(s - 1)$-pebble game on $G$ and $H$ whatever the number of rounds.
The Upper Bound $W'(K_{1,s}) \leq s$. Using $s$ variables, we have to write down a sentence expressing the existence of a $K_{1,s}$ subgraph in a sufficiently large connected graph $G$. Consider $\Phi_s$ saying that

$$\exists x_1 \ldots \exists x_s \left( \bigwedge_{1 \leq i < j \leq s} x_i \neq x_j \land \bigwedge_{1 \leq i \leq s} \exists x_j \left( \bigwedge_{j \neq i} x_i \sim x_j \right) \right).$$

This sentence is obviously true on every graph containing $K_{1,s}$. Suppose that a graph $H$ is connected, has more than $2s$ vertices, and does not contain $K_{1,s}$. Let us prove that $\Phi_s$ is false on $H$. Assuming the opposite—that is, $H \models \Phi_s$—consider vertices $v_1, \ldots, v_s$ of $H$ whose existence is claimed by $\Phi_s$. Then, for each $i \leq s$, there is a vertex $u_i$ adjacent to all $v_1, \ldots, v_s$ but $v_i$. All $u_i$ are pairwise distinct (if $u_i = u_j$, then this vertex would have degree $s$). It follows that $\deg u_i = s - 1$ for each $i \leq s$. Since $H$ is connected and has more than $2s$ vertices, at least one of the vertices in $\{v_1, \ldots, v_s, u_1, \ldots, u_s\}$ must have a neighbor outside this set. Such a vertex must have degree at least $s$, a contradiction.

We now turn to the path graphs.

**Theorem 5.2.**

1. $D'(P_\ell) = \ell - 1$ for all $\ell \geq 4$.
2. $W'(P_\ell) = \ell - 2$ for all $\ell \geq 3$.

**Proof.** As we already mentioned, $D'(P_3) = W'(P_3) = 1$ by trivial reasons. The equalities $D'(P_4) = 3$ and $W'(P_4) = 2$ are proved in Reference [19]. We, therefore, suppose that $\ell \geq 5$.

The Upper Bound $D'(P_\ell) \leq \ell - 1$. Let $G$ and $H$ be two graphs, each with at least $\ell$ vertices. Suppose that $H$ contains no $P_\ell$ as a subgraph, while $G$ contains a path $a_1a_2\ldots a_\ell$. Let $m$ be the largest number, such that the vertices $a_1, a_2, \ldots, a_m$ form a clique. We split our analysis into three cases:

**Case 1:** $m \geq \ell - 1$. Spoiler pebbles the clique $a_1, a_2, \ldots, a_{\ell-1}$ and wins, because there is no $K_{\ell-1}$ in $H$ (having a $K_{\ell-1}$ and at least one more vertex, $H$ would contain $P_\ell$ by connectedness).

**Case 2:** $3 \leq m \leq \ell - 2$. By the definition of $m$, the vertex $a_{m+1}$ is not adjacent to at least one of the preceding vertices. Since the vertices $a_1, a_2, \ldots, a_m$ form a clique, they actually can appear in the initial segment of the path $a_1a_2\ldots a_\ell$ in an arbitrary order. Therefore, we can suppose without loss of generality that $a_1$ and $a_{m+1}$ are not adjacent (see Figure 3, where the proof is illustrated in the case of $\ell = 5$; in Case 2, we then have $m = 3$).

In the first $\ell - 2$ rounds, Spoiler pebbles the vertices $a_i$ for all $i$ except $i = 1$ and $i = m$. Suppose that Duplicator manages to respond in $H$ with two paths $b_2, \ldots, b_{m-1}$ and $b_{m+1}, \ldots, b_\ell$. Then, Spoiler wins in one extra move by pebbling either $a_1$ or $a_m$. Duplicator loses, because in $H$ there are no two vertices $b_1$ and $b_m$ with the adjacency patterns to the already pebbled vertices that $a_1$ and $a_m$ have in $G$. Indeed, if such $b_1$ and $b_m$ exist, then they should be distinct because of different adjacency to $b_{m+1}$. But then $b_1b_2\ldots b_\ell$ would be an $\ell$-path in $H$, a contradiction.

Fig. 3. Proof of the bound $D'(P_\ell) \leq \ell - 1$ in the case of $\ell = 5$: $a_1a_2a_3a_4a_5$ is a 5-path in $G$. Dashed segments are drawn between non-adjacent vertices.
Case 3: $m = 2$. Note that $a_1$ is not adjacent to $a_3$.

Subcase 3-a: $a_3$ and $a_4$ are not adjacent. In the first $\ell - 2$ rounds, Spoiler pebbles the vertices $a_i$ for all $i$ except $i = 2$ and $i = 4$. Suppose that Duplicator manages to pebble vertices $b_1, b_3, b_5, \ldots, b_\ell$ in $H$, such that $b_3 \ldots b_\ell$ is an $(\ell - 4)$-path. Then, Spoiler wins in one extra move by pebbling either $a_2$ or $a_4$. Duplicator loses, because in $H$ there are no two vertices $b_2$ and $b_4$ with the same adjacency pattern to the already pebbled vertices. Indeed, if such $b_2$ and $b_4$ exist, then they should be distinct because of different adjacency to $b_1$. This is impossible, as $b_1 b_2 \ldots b_\ell$ would be an $\ell$-path in $H$.

Subcase 3-b: $a_1$ and $a_4$ are adjacent; $a_2$ and $a_3$ are not. Spoiler wins in $\ell - 1$ moves using exactly the same strategy as in the preceding subcase (now $b_2$ and $b_4$ should be distinct because of different adjacency to $b_3$).

Subcase 3-c: $a_3$ and $a_4$ are adjacent; $a_2$ and $a_3$ are also adjacent. Spoiler pebbles the $\ell - 2$ vertices $a_3, a_4, \ldots, a_\ell$. Suppose that Duplicator succeeds to respond with an $(\ell - 2)$-path $b_3 \ldots b_\ell$ in $H$. Then, Spoiler wins in the next round by pebbling either $a_1$ or $a_2$. Duplicator loses, because in $H$ there are no two vertices $b_1$ and $b_2$ with the same adjacency pattern to the already pebbled vertices. Indeed, if such $b_1$ and $b_2$ exist, then they should be distinct because of different adjacency to $b_3$. Note, however, that such $b_1$ and $b_2$ need not be adjacent. In any case, $b_1 b_4 b_3 b_2 b_5 \ldots b_\ell$ would be an $\ell$-path in $H$.

The Upper Bound $W'(P_\ell) \leq \ell - 2$. Consider two connected graphs $G$ and $H$ both with at least $\ell$ vertices and such that $G$ contains a copy of $P_\ell$ while $H$ does not. Let us prove that Spoiler has a winning strategy in the $(\ell - 2)$-pebble game in at most $\ell + 1$ rounds. This will mean that $G$ and $H$ are distinguishable by a first-order sentence $\Phi_{G,H}$ with $\ell - 2$ variables of quantifier depth $\ell + 1$. Since there are only finitely many pairwise inequivalent sentences of a bounded quantifier depth, the inequivalent sentences $\Phi_{G,H}$ for various $G$ and $H$ can be combined into a single $(\ell - 2)$-variable sentence defining the existence of a $P_\ell$ subgraph over all sufficiently large connected graphs.

Assume first that $H$ has the following two properties:

(A) all pendant vertices have a common neighbor;
(B) there are at most $\ell - 2$ non-pendant vertices.

We show how Spoiler can win in this case by playing with $\ell - 2$ pebbles and making at most $\ell + 1$ moves.

If $G$ does not satisfy Condition (A), then Spoiler pebbles two pendant vertices of $G$ with distinct neighbors. If Duplicator responds with two pendant vertices in $H$, then Spoiler wins by pebbling their common neighbor. If one of Duplicator’s vertices in $H$ is not pendant, then Spoiler wins by pebbling two of its neighbors.

Suppose now that $G$ satisfies Condition (A). If $G$ has no pendant vertex, then Spoiler pebbles a pendant vertex in $H$ and wins in the next two moves. We, therefore, suppose that $G$ has at least one pendant vertex. Denote the common neighbor of all pendant vertices in $G$ by $z$ and in $H$ by $z'$. Condition (A) implies that $z$ is the central vertex of a pendant star in $G$. Denote the number of non-pendant vertices in $G$ by $t$. Since $G$ contains $P_\ell$ as a subgraph, we have $t \geq \ell - 1$. Let Spoiler pebble $\ell - 2$ non-pendant vertices in $G$ different from $z$. By Condition (B), Duplicator is forced to pebble either $z'$ or a pendant vertex in $H$. In the former case, Spoiler keeps the pebble on $z'$, moving one of the other pebbles to a pendant vertex in $H$. Since the counterpart of $z'$ in $G$ is a non-pendant vertex different from $z$, Duplicator is forced to respond with a non-pendant vertex in $G$. In each case, Spoiler forces pebbling a pendant vertex in one graph and a non-pendant vertex in the other graph, which allows him to win in the next two moves.
Now, we assume that $H$ does not satisfy Conditions (A) or (B) and show that then Spoiler is able to win with $\ell - 2$ pebbles in at most $\ell$ moves. Since $H$ contains no $P_\ell$, has at least $\ell$ vertices, and is connected, it contains also no $C_{\ell-1}$. Let $x_1, \ldots, x_{\ell-1}$ be a copy of $P_\ell$ in $G$. In the first $\ell - 2$ rounds, Spoiler pebbles the vertices $x_2, \ldots, x_{\ell-1}$. Suppose that Duplicator does not lose immediately and responds with vertices $y_2, \ldots, y_{\ell-1}$ that form a path $P$ in $H$. In the next round, Spoiler chooses one of the vertices $x_3, \ldots,x_{\ell-1}$ and moves the pebble from this vertex to $x_1$. Duplicator can survive under every choice of Spoiler only if there is a neighbor of $y_2$ outside $P$ or when all $y_3, \ldots, y_{\ell-1}$ are adjacent to $y_2$. By a symmetric reason, Duplicator does not lose in this round only if there is a neighbor of $y_{\ell-1}$ outside $P$ or all $y_2, \ldots, y_{\ell-2}$ are adjacent to $y_{\ell-1}$. It remains to show that these conditions imply Conditions (A) and (B).

It is impossible that both $y_2$ and $y_{\ell-1}$ have neighbors outside $P$, because then $H$ would contain either $C_{\ell-1}$ or $P_\ell$ as a subgraph. Therefore, we can assume that at least one of the vertices $y_2$ and $y_{\ell-1}$ (say, $y_{\ell-1}$) is adjacent to all the other vertices in $P$. Consider two cases:

Case 1: $y_2$ has a neighbor $y$ outside $P$. It is easily seen that if $y$ were adjacent to at least one of $y_3, \ldots, y_{\ell-1}$, then $H$ would contain $C_{\ell-1}$. Moreover, if at least one of $y_3, \ldots, y_{\ell-1}$ had a neighbor outside $\{y, y_2, \ldots, y_{\ell-1}\}$, then $H$ would contain $P_\ell$. A $P_\ell$ subgraph would exist also if a neighbor of $y_2$ outside $P$ had another neighbor outside $P$. It follows that, in this case, all the vertices outside $P$ are pendant in $H$ and are adjacent to $y_2$. We conclude that $H$ satisfies Conditions (A) and (B).

Case 2: $y_2$ does not have any neighbor outside $P$. In this case, the vertex $y_2$, like the vertex $y_{\ell-1}$, is adjacent to all the other vertices of $P$. Any vertex outside $P$ must be pendant in $H$, because otherwise $H$ would contain $P_\ell$ or $C_{\ell-1}$. Any two such vertices must have a common neighbor or else $H$ would contain $P_\ell$. Again, $H$ must satisfy Conditions (A) and (B).

The Lower Bounds $D'(P_\ell) \geq \ell - 1$ and $W'(P_\ell) \geq \ell - 2$. The bound for the width parameter follows directly from Part 1 of Lemma 2.5. To prove the bound for the depth parameter, consider the graphs $G_n$ and $H_n$ as in the proof of this lemma. That is, $G_n$ consists of $K_{\ell-1}$ with a pendant star $K_{1,n}$ and $H_n$ consists of $K_{\ell-2}$ also with a pendant star $K_{1,n}$, where the parameter $n$ can be chosen arbitrarily large. Obviously, $G_n$ contains $P_\ell$ as a subgraph, while $H_n$ does not. The vertex set of each of the graphs $G_n$ and $H_n$ can be split into three classes: the pendant vertices, the central vertex of the pendant $K_{1,n}$, and the remaining non-pendant vertices. Consider the Ehrenfeucht-Fraïssé game on $G_n$ and $H_n$. Duplicator does not lose as long as he respects the vertex classes. Spoiler can break this strategy of Duplicator only by pebbling all vertices of $G_n$ in the third class, for which he needs no less than $\ell - 2$ moves. If he does this in the first $\ell - 2$ rounds of the game, then Duplicator still does not lose if in the $(\ell - 2)$-th round she responds with the vertex in the second class of $H$. It follows that Spoiler cannot win by making only $\ell - 2$ moves.}

\section*{6 Further Questions}

Theorem 3.1 shows that, for any sparkler graph, the values of $W'(S_{q,p})$ and $D'(S_{q,p})$ can differ from each other by at most 2. The examples of path and star graphs show that $W'(F)$ and $D'(F)$ can have different values. In both of these cases, we have $D'(F) = W'(F) + 1$ by Theorems 5.1 and 5.2. How large can the gap between the two parameters be, in general? Do we have $D'(F) \leq W'(F) + O(1)$ for all $F$? If true, then this would be a quite interesting feature of Subgraph Isomorphism as, in general, first-order properties expressible with a bounded number of variables can require an arbitrarily large quantifier depth. Note in this respect that $D'(F) < \frac{1}{2} W'(F) + 3$ as a consequence of Part 2 of Theorem 4.1; that is, the gap between $W'(F)$ and $D'(F)$ cannot be superlinear.

As mentioned in Section 1, the questions studied in the article make perfect sense also in the case that we want to express the existence of an \textit{induced} subgraph isomorphic to the pattern graph $F$.\hfill $\Box$
Let $D[F]$, $W[F]$, $D'[F]$, and $W'[F]$ be the analogs of the parameters $D(F)$, $W(F)$, $D'(F)$, and $W'(F)$, where the bracket parentheses emphasize that Induced Subgraph Isomorphism is considered. The argument showing the equality $W(F) = \ell$ does not work anymore in the induced case and, indeed, $D[K_3 + e] = 3$ for the 4-vertex paw graph $K_3 + e$. We noticed this example in Reference [18], where we also proved a general lower bound $W[F] \geq \left( \frac{1}{2} - o(1) \right) \ell$. If this bound were tight for infinitely many $F$, then Induced Subgraph Isomorphism for these patterns would be solvable in time $O(n^{(\frac{1}{2} - o(1)) \ell})$, which would be interesting, because the Nešetřil-Poljak bound, based on fast matrix multiplication, could never be better than $O(n^{\frac{3}{2}} \ell)$. Recall in this respect that no known method solves the problem in time $O(n^{o(\ell)})$, even if we restrict it to an arbitrary infinite family of induced patterns; see Reference [4]. However, it is still not excluded that $W[F] \geq \ell - 1$ for all $F$, which makes the problem of determining or estimating the parameters $W[F]$ and $D[F]$ quite intriguing.

As for the asymptotic descriptive complexity over connected graphs, we observed in Reference [18] that $W'[F] = W[F]$ and $D'[F] = D[F]$ for a large class of pattern graphs, including all those without universal vertices. In general, $W'[F] \geq \left( \frac{4}{5} - o(1) \right) \ell$ for all $F$.

**ACKNOWLEDGMENTS**

The authors thank the anonymous referees for their detailed comments.

**REFERENCES**

[1] Noga Alon, Raphael Yuster, and Uri Zwick. 1995. Color-coding. J. ACM 42, 4 (1995), 844–856. DOI: https://doi.org/10.1145/210332.210337

[2] László Babai. 2016. Graph isomorphism in quasipolynomial time. In Proceedings of the 48th ACM SIGACT Symposium on Theory of Computing (STOC’16). ACM, New York, 684–697. DOI: https://doi.org/10.1145/2897518.2997542

[3] Reinhard Diestel. 2000. Graph Theory. Springer, New York.

[4] Peter Floderus, Miroslaw Kowaluk, Andrzej Lingas, and Eva-Marta Lundell. 2015. Induced subgraph isomorphism: Are some patterns substantially easier than others? Theor. Comput. Sci. 605 (2015), 119–128. DOI: https://doi.org/10.1016/j.tcs.2015.09.001

[5] François Le Gall. 2014. Powers of tensors and fast matrix multiplication. In Proceedings of the International Symposium on Symbolic and Algebraic Computation (ISSAC’14). ACM, New York, 296–303. DOI: https://doi.org/10.1145/2608628.2608664

[6] Erich Grädel and Martin Grohe. 2015. Is polynomial time choiceless? In Fields of Logic and Computation II, Essays Dedicated to Yuri Gurevich on the Occasion of His 75th Birthday (Lecture Notes in Computer Science), Vol. 9300. Springer, Berlin.

[7] Martin Grohe and Nicole Schweikardt. 2005. The succinctness of first-order logic on linear orders. Log. Meth. Comput. Sci. 1, 1 (2005), 1–25. DOI: https://doi.org/10.2168/LMCS-1(1:6)2005

[8] Neil Immelman. 1982. Upper and lower bounds for first-order expressibility. J. Comput. Syst. Sci. 25, 1 (1982), 76–98. DOI: https://doi.org/10.1016/0022-0000(82)90011-3

[9] Neil Immelman. 1999. Descriptive Complexity. Springer, New York. DOI: https://doi.org/10.1007/978-1-4612-0539-5

[10] Yuan Li, Alexander A. Razborov, and Benjamin Rossman. 2017. On the AC^0 complexity of subgraph isomorphism. SIAM J. Comput. 46, 3 (2017), 936–971. DOI: https://doi.org/10.1137/14099721X

[11] Leonid Libkin. 2004. Elements of Finite Model Theory. Springer, Berlin. DOI: https://doi.org/10.1007/978-3-662-07003-1

[12] Jaroslav Nešetřil and Svatomásl Poljak. 1985. On the complexity of the subgraph problem. Commentat. Math. Univ. Carol. 26 (1985), 415–419.

[13] Bogdan Oporowski, James G. Oxley, and Robin Thomas. 1993. Typical subgraphs of 3- and 4-connected graphs. J. Comb. Theor., Ser. B 57, 2 (1993), 239–257. DOI: https://doi.org/10.1016/j.jctb.1993.1019

[14] Oleg Pikhurko, Helmut Veith, and Oleg Verbitsky. 2006. The first-order definability of graphs: Upper bounds for quantifier depth. Disc. Appl. Math. 154, 17 (2006), 2511–2529. DOI: https://doi.org/10.1016/j.dam.2006.03.002

[15] Nicole Schweikardt. 2013. A short tutorial on order-invariant first-order logic. In Proceedings of the 8th International Computer Science Symposium in Russia (CSR’13) (Lecture Notes in Computer Science), Vol. 7913. Springer, Berlin, 112–126. DOI: https://doi.org/10.1007/978-3-642-38536-0_10

[16] György Turán. 1984. On the definability of properties of finite graphs. Disc. Math. 49, 3 (1984), 291–302. DOI: https://doi.org/10.1016/0012-365X(84)90166-3
[17] Moshe Y. Vardi. 1995. On the complexity of bounded-variable queries. In *Proceedings of the 14th ACM Symposium on Principles of Database Systems*. ACM Press, New York, 266–276. DOI: https://doi.org/10.1145/212433.212474

[18] Oleg Verbitsky and Maksim Zhukovskii. 2019. On the first-order complexity of induced subgraph isomorphism. *Log. Meth. Comput. Sci* 15, 1 (2019), 1–24.

[19] Oleg Verbitsky and Maksim Zhukovskii. 2018. The descriptive complexity of subgraph isomorphism without numerics. *Theor. Comput. Syst.* (2018). DOI: https://doi.org/10.1007/s00224-018-9864-3 In press.

Received July 2018; accepted January 2019