On the First-Order Complexity of Induced Subgraph Isomorphism*

Oleg Verbitsky[1] **and Maksim Zhukovskii**[2]

1 **Institut für Informatik, Humboldt-Universität zu Berlin, Berlin, Germany**
2 **Moscow Institute of Physics and Technology, Department of Discrete Mathematics, Moscow, Russia**

Abstract

Given a graph F, let $\mathcal{I}(F)$ be the class of graphs containing F as an induced subgraph. Let $W[F]$ denote the minimum k such that $\mathcal{I}(F)$ is definable in k-variable first-order logic. The recognition problem of $\mathcal{I}(F)$, known as Induced Subgraph Isomorphism (for the pattern graph F), is solvable in time $O(n^{W[F]})$. Motivated by this fact, we are interested in determining or estimating the value of $W[F]$. Using Olariu's characterization of paw-free graphs, we show that $\mathcal{I}(K_3+e)$ is definable by a first-order sentence of quantifier depth 3, where K_3+e denotes the paw graph. This provides an example of a graph F with $W[F]$ strictly less than the number of vertices in F. On the other hand, we prove that $W[F]=4$ for all F on 4 vertices except the paw graph and its complement. If F is a graph on ℓ vertices, we prove a general lower bound $W[F]>(1/2-o(1))\ell$, where the function in the little-o notation approaches 0 as ℓ increases. This bound holds true even for a related parameter $W^*[F]\le W[F]$, which is defined as the minimum k such that $\mathcal{I}(F)$ is definable in the infinitary logic $L^k_{\infty\omega}$. We show that $W^*[F]$ can be strictly less than $W[F]$. Specifically, $W^*[P_4]=3$ for P_4 being the path graph on 4 vertices.

1998 ACM Subject Classification F.4.1 Finite Model Theory

Keywords and phrases The induced subgraph isomorphism problem, descriptive and computational complexity, finite-variable first-order logic, quantifier depth and variable width

Digital Object Identifier 10.4230/LIPIcs.CSL.2017.40

1 Introduction

For a given graph F, let $\mathcal{I}(F)$ denote the class of all graphs containing a copy of F as an induced subgraph. We are interested in the descriptive complexity of $\mathcal{I}(F)$, for a fixed pattern graph F, in first-order logic whose vocabulary consists of the adjacency and the equality relations ($\sim$ and $=$ respectively). Let $D[F]$ denote the minimum quantifier depth of a sentence in this logic that defines $\mathcal{I}(F)$. Furthermore, let $W[F]$ denote the minimum variable width of a sentence defining $\mathcal{I}(F)$. Note that

$$W[F]\le D[F]\le \ell,$$

where here and throughout the paper ℓ denotes the number of vertices in F. It may come as some surprise that the parameter $D[F]$ can be strictly less than ℓ. To see an example, let

---

* O. Verbitsky was supported by DFG grant VE 652/1–2. He is on leave from the Institute for Applied Problems of Mechanics and Mathematics, Lviv, Ukraine. M. Zhukovskii was supported by grants No. 15-01-03530 and 16-31-60052 of Russian Foundation for Basic Research.

26th EACSL Annual Conference on Computer Science Logic (CSL 2017).
Editors: Valentin Goranko and Mads Dam; Article No. 40; pp. 40:1–40:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$F = K_3 + e$ be the paw graph . The following sentence uses three variables x_1, x_2, x_3 and has quantifier depth 3:

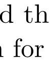
$$\exists x_1 \, (\exists x_2 \exists x_3 \, (x_1 \sim x_2 \wedge x_1 \sim x_3 \wedge x_2 \sim x_3) \wedge \\ \exists x_2 \, (x_1 \not\sim x_2 \wedge \exists x_3 \, (x_1 \sim x_3 \wedge x_3 \sim x_2) \wedge \exists x_3 \, (x_3 \sim x_1 \wedge x_3 \not\sim x_2))).$$

It says that a graph contains a vertex v that belongs to a triangle and can be accompanied with a vertex u at distance 2 from v such that there is a vertex w adjacent to v but non-adjacent to u. Obviously, this sentence is true on the paw graph and on every graph containing an induced paw subgraph. Olariu's characterization of paw-free graphs [21] implies that the sentence is false on every graph without an induced paw subgraph; see Section 4.1 for details.

The decision problem for $\mathcal{I}(F)$ is known as Induced Subgraph Isomorphism (for the pattern graph F). We denote it by ISI(F). Our interest in the parameters $D[F]$ and $W[F]$ is motivated by the fact that ISI(F) is solvable in time $O(n^{W[F]})$; see [19, Prop. 6.6]. Before stating our results on $D[F]$ and $W[F]$ we very briefly overview the known algorithmic results in this area.

1.1 Computational complexity of Induced Subgraph Isomorphism

Obviously, ISI(F) is solvable in time $O(n^\ell)$ on n-vertex input graphs by exhaustive search. We use the standard notation K_ℓ for complete graphs, P_ℓ for paths, and C_ℓ for cycles on ℓ vertices. Itai and Rodeh [14] observed that ISI(K_3) is solvable in time $O(n^\omega)$, where $\omega < 2.373$ is the exponent of fast square matrix multiplication [11]. Nešetřil and Poljak [20] showed, by a reduction of ISI(F) to ISI(K_3), that ISI(F) is solvable in time $O(n^{(\omega/3)\ell+c})$, where $c = 0$ if ℓ is divisible by 3 and $c \le \frac{2}{3}$ otherwise. For ℓ not divisible by 3, this time bound was improved by Eisenbrand and Grandoni [8] using fast rectangular matrix multiplication. On the other hand, ISI(K_ℓ) is unsolvable in time $n^{o(\ell)}$ unless the Exponential Time Hypothesis fails [3]. Floderus et al. [10] collected evidence in favour of the conjecture that ISI(F) for F with ℓ vertices cannot be solved faster than ISI($K_{c\ell}$), where $c < 1$ is a constant independent on F. Along with the Exponential Time Hypothesis, this would imply that the time complexity of ISI(F) is $n^{\Theta(\ell)}$. As an example of a particular result of [10] in this direction, ISI(P_{2a-1}) is not easier than ISI(K_a), and the same holds true for ISI(C_{2a}); see also the earlier work [4].

The induced subgraph isomorphism problem has been intensively studied for particular pattern graphs F with small number of vertices. Let $\overline{F}$ denote the complement graph of F and note that at least one of the graphs F and $\overline{F}$ is connected. Since ISI(F) and ISI($\overline{F}$) have the same time complexity, we can restrict our attention to connected pattern graphs. There are six such graphs on four vertices, namely K_4, P_4, C_4, $K_3 + e$, the claw graph $K_{1,3}$ (), and the diamond graph $K_4 \setminus e$ (). Corneil et al. [5] designed an $O(n + m)$ time algorithm for ISI(P_4), where m denotes the number of edges in an input graph. As noted in [17], the Olariu characterization of paw-free graphs reduces ISI($K_3 + e$) to ISI(K_3), showing that the former problem is also solvable in time $O(n^\omega)$. The same time bound is obtained by Vassilevska Williams et al. [28] for the diamond graph $K_4 \setminus e$ and, using randomization, for the other pattern graphs on 4 vertices except K_4. The best known time bound for ISI(K_4) is given by the methods of [8] and is currently $O(n^{3.257})$ [28].

1.2 Our results on the descriptive complexity of Induced Subgraph Isomorphism

In Section 3 we prove a general lower bound $W[F] > (1/2 - o(1))\ell$, where the function in the little-o notation approaches 0 as ℓ, the number of vertices in the pattern graph F, increases.

Whether or not it can be improved remains an intriguing open question. Note that this bound leaves a hypothetical possibility that the time bound $O(n^{W[F]})$ for Induced Graph Isomorphism can be better than the Nešetřil-Poljak bound $O(n^{(\omega/3)\ell+c})$ for infinitely many pattern graphs F.

Our approach uses a connection to the k-extension property of graphs, that is well known in finite model theory; see, e.g, [25]. We define the *extension index* of F, denoted by $E[F]$, as the minimum k such that the k-extension property forces the existence of an induced copy of F. It is easy to show that $W[F] \geq E[F]$. Our results about the parameter $E[F]$ may be interesting on its own. In particular, we show that $E[F] \geq \chi(F)$, where $\chi(F)$ denotes the chromatic number of F.

In Section 4, we determine the values of $D[F]$ and $W[F]$ for all pattern graphs with at most 4 vertices. We prove that $W[F] = 4$ for all F on 4 vertices except the paw graph and its complement.

Our lower bounds for $W[F]$ hold true also for a related parameter $W^*[F] \leq W[F]$, which is defined as the minimum k such that $\mathcal{I}(F)$ is definable in the infinitary logic $L^k_{\infty\omega}$. The only exception is $F = P_4$, the path on 4 vertices. In this case, we prove that $W^*[P_4] = 3$ while $W[P_4] = 4$. This shows that $W^*[F]$ can be strictly less than $W[F]$, that is, the infinitary logic can be more succinct when defining $\mathcal{I}(F)$.

In Section 5, we address a relaxation version of the parameter $W[F]$. Consider a simple example. Let $D_v[F]$ be the minimum quantifier depth of a sentence Φ defining $\mathcal{I}(F)$ over *sufficiently large connected* graphs. That is, it is required that there is a number s such that Φ correctly detects whether or not a graph G belongs to $\mathcal{I}(F)$ only if G is connected and has at least s vertices. Whereas $D_v[F] \leq D[F]$, it is clear that $\mathcal{I}(F)$ for a connected pattern graph F is still recognizable in time $O(n^{D_v[F]})$. Let $F = P_3$. As easily seen, P_3-free graphs are exactly disjoint unions of cliques. Therefore, connected P_3-free graphs are exactly the complete graphs, which readily implies that $D_v[P_3] \leq 2$, whereas $D[P_3] = 3$. As a further example, we remark that the existence of a *not necessarily induced* subgraph P_4 can be defined over sufficiently large connected graphs with just 2 variables; see [26] for this and further examples.

We can go further and define $W_{tw}[F]$ to be the minimum variable width of a sentence defining $\mathcal{I}(F)$ over connected graphs G of *sufficiently large treewidth* $tw(G)$. As a consequence of Courcelle's theorem [7], $\mathcal{I}(F)$ for a connected pattern graph F is recognizable in time $O(n^{W_{tw}[F]})$; cf. the discussion in [26].

The above discussion motivates the problem of proving lower bounds for the parameter $W_\kappa[F]$ which we define as the minimum variable width of a sentence defining $\mathcal{I}(F)$ over graphs G of *sufficiently large connectedness* $\kappa(G)$. Note that $W_\kappa[F] \leq W_{tw}[F] \leq W_v[F] \leq W[F]$. We prove that $W_\kappa[F] = W[F]$ for a large class of pattern graphs F. We also prove a general lower bound $W_\kappa[F] > (1/3 - o(1))\,\ell$ for all F on ℓ vertices.

1.3 Comparison to (not necessarily induced) Subgraph Isomorphism

The SUBGRAPH ISOMORPHISM problem is very different from its induced version. For infinitely many pattern graphs F, SUBGRAPH ISOMORPHISM can be solved in time $O(n^c)$ for a constant c. This follows from a result by Alon, Yuster and Zwick [1] who showed that the problem is solvable in time $2^{O(\ell)} \cdot n^{tw(F)+1} \log n$.

Let $W(F)$ and $D(F)$ be the analogs of $W[F]$ and $D[F]$ for the not-necessarily-induced case. Note that $W(F) = D(F) = \ell$ as a consequence of the trivial observation that K_ℓ contains F as a subgraph while $K_{\ell-1}$ does not. Nevertheless, $D_v[F]$ can be strictly smaller than ℓ for some connected pattern graphs. Moreover, $D_{tw}[F]$ can sometimes be arbitrarily small comparing to ℓ. This is the subject of our preceding paper [26].

1.4 Logic with numeric predicates

In the present paper, we consider the most laconic first-order language of graphs whose vocabulary has only the adjacency and the equality relations. If we assume that the vertex set of a graph is $\{1, 2, \ldots, n\}$ and additionally allow arbitrary numerical relations, this richer logic captures the non-uniform AC^0; see [13, 19]. Let $W_{\mathrm{Arb}}(F)$ denote the analog of the parameter $W(F)$ (the not-necessarily-induced case) for this logic, and $W_{\mathrm{Arb}}[F]$ denote the analog of the parameter $W[F]$ (the induced case). The known relations to circuit complexity [13, 24] imply that the (not necessarily induced) Subgraph Isomorphism is solvable by bounded-depth unbounded-fan-in circuits of size $n^{W_{\mathrm{Arb}}(F)+o(1)}$, and the similar is true also for Induced Subgraph Isomorphism. Whereas the parameter $W_{\mathrm{Arb}}(F)$ is studied in this context by Li, Razborov, and Rossman [18], the parameter $W_{\mathrm{Arb}}[F]$ remains unexplored.

2 Preliminaries

A *graph property* is a class of graphs $\mathcal{C}$ closed under isomorphism, that is, for isomorphic graphs G and H, $G \in \mathcal{C}$ iff $H \in \mathcal{C}$. We consider first-order sentences about graphs in the language containing the adjacency and the equality relations. A sentence Φ *defines* a graph property $\mathcal{C}$ if $G \in \mathcal{C}$ exactly when $G \models \Phi$, i.e., Φ is true on G. A graph property $\mathcal{C}$ is *first-order definable* if there is a first-order sentence defining $\mathcal{C}$.

Let $\mathcal{C}$ be a first-order graph property. The *logical depth* of $\mathcal{C}$, denoted by $D(\mathcal{C})$, is the minimum quantifier depth (rank) of a sentence defining $\mathcal{C}$. The *logical width* of $\mathcal{C}$, denoted by $W(\mathcal{C})$, is the minimum *variable width* of a sentence defining $\mathcal{C}$, i.e., the number of first-order variables occurring in the sentence where different occurrences of the same variable do not count.

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 *distinguishes* G and H, i.e., is true on one of the graphs and false on the other.

▶ **Lemma 1.** *$D(\mathcal{C}) = \max\{ D(G, H) : G \in \mathcal{C},\ H \notin \mathcal{C}\}$.*

Proof. In one direction, note that whenever $G \in \mathcal{C}$ and $H \notin \mathcal{C}$, we have $D(G, H) \leq D(\mathcal{C})$ because any sentence defining $\mathcal{C}$ distinguishes G and H. For the other direction, suppose that every such G and H are distinguished by a sentence $\Phi_{G,H}$ of quantifier depth at most d. Specifically, suppose that $\Phi_{G,H}$ is true on G and false on H. We have to show that $\mathcal{C}$ can be defined by a sentence of quantifier depth at most d. For a graph $G \in \mathcal{C}$, consider the sentence $\Phi_G \overset{\mathrm{def}}{=} \bigwedge_H \Phi_{G,H}$, where the conjunction is over all $H \notin \mathcal{C}$. This sentence distinguishes G from all $H \notin \mathcal{C}$ and has quantifier depth at most d. The only problem with it is that the conjunction over H is actually infinite. Luckily, there are only finitely many pairwise inequivalent first-order sentences about graphs of quantifier depth d; see, e.g., [23, Theorem 2.4]. Removing all but one formula $\Phi_{G,H}$ from each equivalence class, we make Φ_G a legitimate finite sentence. Now, consider $\Phi \overset{\mathrm{def}}{=} \bigvee_G \Phi_G$, where the disjunction is over all $G \in \mathcal{C}$. It can be made finite in the same way. The sentence Φ defines $\mathcal{C}$ and has quantifier depth at most d. ◀

Thus, Lemma 1 is a simple consequence of the fact that there are only finitely many pairwise inequivalent first-order statements of bounded quantifier depth. Note that the last fact does not hold true for the variable width. We define

$$W^*(\mathcal{C}) = \max\{\,W(G,H) : G \in \mathcal{C},\ H \notin \mathcal{C}\}\,.$$

Equivalently, $W^*(\mathcal{C})$ is equal to the minimum k such that $\mathcal{C}$ is definable in the infinitary logic $L^k_{\infty\omega}$; see, e.g., [19, Chapter 11]. Obviously, $W^*(\mathcal{C}) \le W(\mathcal{C})$, and we will see in Section 4.2 that this inequality can be strict. Summarizing, we have

$$W^*(\mathcal{C}) \le W(\mathcal{C}) \le D(\mathcal{C}). \tag{1}$$

The value of $W(\mathcal{C})$ admits the following characterization. If $W(G,H) \le k$, let $D^k(G,H)$ denote the minimum quantifier depth of a first-order k-variable sentence distinguishing G and H. We set $D^k(G,H) = \infty$ if $W(G,H) > k$. The following fact can be proved similarly to Lemma 1.

▶ **Lemma 2.** *$W(\mathcal{C}) > k$ if and only if there is a sequence of graph pairs (G_i, H_i) with $G_i \in \mathcal{C}$ and $H_i \notin \mathcal{C}$ such that $D^k(G_i, H_i) \to \infty$ as $i \to \infty$.*

Lemmas 1 and 2 reduce estimating the logical depth and width to estimating the parameters $D(G,H)$ and $D^k(G,H)$ over $G \in \mathcal{C}$ and $H \notin \mathcal{C}$. The first inequality in (1) can be used for obtaining lower bounds for $W(\mathcal{C})$ by estimating $W(G,H)$ over $G \in \mathcal{C}$ and $H \notin \mathcal{C}$. The parameters $D(G,H)$, $D^k(G,H)$, and $W(G,H)$ have a very useful combinatorial characterization.

In the *k-pebble Ehrenfeucht-Fraïssé game*, the board consists of two vertex-disjoint graphs G and H. Two players, *Spoiler* and *Duplicator* (or *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. Note that the pebbles can be reused and change their positions during the 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 [13]:

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

We are interested in the property of containing a specified induced subgraph. We write $F \sqsubset G$ to say that G contains an induced subgraph isomorphic to F. Thus, $\mathcal{I}(F) = \{G : F \sqsubset G\}$. Let $D[F] = D(\mathcal{I}(F))$ and, similarly, $W[F] = W(\mathcal{I}(F))$ and $W^*[F] = W^*(\mathcal{I}(F))$. Thus, $W^*[F]$ is the maximum $W(G,H)$ over all G containing an induced copy of F and all H not containing such a copy. As a particular case of (1), we have

$$W^*[F] \le W[F] \le D[F] \le \ell$$

for every F with ℓ vertices.

The vertex set of a graph G will be denoted by $V(G)$. Throughout the paper, we consider simple undirected graphs without loops. Let $\overline{G}$ denote the complement of G, that is, $V(\overline{G}) = V(G)$ and two vertices are adjacent in $\overline{G}$ exactly when they are not adjacent in G.

▶ **Lemma 3.** *$D[F] = D[\overline{F}]$, $W^*[F] = W^*[\overline{F}]$, and $W[F] = W[\overline{F}]$.*

Proof. The first equality follows from the equality $D(G,H) = D(\overline{G},\overline{H})$ by Lemma 1. Indeed, $F \sqsubset G$ iff $\overline{F} \sqsubset \overline{G}$. Therefore,

$$D[F] = \max\left\{ D(G,H) : G \sqsupset F,\ H \not\sqsupset F\right\} = \max\left\{ D(\overline{G},\overline{H}) : \overline{G} \sqsupset \overline{F},\ \overline{H} \not\sqsupset \overline{F}\right\}$$
$$= \max\left\{ D(G,H) : G \sqsupset \overline{F},\ H \not\sqsupset \overline{F}\right\} = D[\overline{F}].$$

The second equality follows similarly from the equality $W(G,H) = W(\overline{G},\overline{H})$ by the definition of $W^*[F]$. The third equality follows from the equality $D^k(G,H) = D^k(\overline{G},\overline{H})$ by Lemma 2. ◀

2.1 Further graph-theoretic definitions

A graph G is called F-free if $F \not\sqsubset G$. The vertex-disjoint union of graphs G and H will be denoted by $G + H$. Correspondingly, sG is the vertex-disjoint union of s copies of G. The *lexicographic product* $A \cdot B$ of two graphs A and B is defined as follows: $V(A \cdot B) = V(A) \times V(B)$, and (u,v) and (x,y) are adjacent in $A \cdot B$ if u and x are adjacent in A or if $u = x$ and v and y are adjacent in B. In other words, $A \cdot B$ is obtained from A by substituting each vertex u with an induced copy B_u of B and drawing all edges between B_u and B_x whenever u and x are adjacent.

A vertex is *isolated* if it has no adjacent vertex and *universal* if it is adjacent to all other vertices in the graph. Two vertices are called *twins* if they have the same adjacency to the rest of the graph.

Throughout the paper, $\log n$ means the logarithm base 2.

3 The extension index and a lower bound for $W^*[F]$

Let $k \geq 2$. By the *k-extension property* we mean the first-order sentence EA_k of quantifier depth k (also called the *kth extension axiom*) saying that, for every two disjoint sets $X, Y \subset V(G)$ with $|X \cup Y| < k$, there is a vertex $z \notin X \cup Y$ adjacent to all $x \in X$ and non-adjacent to all $y \in Y$. Note that EA_2 says exactly that a graph has neither isolated nor universal vertex. For convenience, we also set $EA_1 \overset{\text{def}}{=} \exists z(z = z)$.

Note that, if $G \models EA_k$ and F has at most k vertices, then $F \sqsubset G$. Suppose that F has more than 1 vertex. We define the *extension index* of F, denoted by $E[F]$, as the minimum k such that $H \models EA_k$ implies $F \sqsubset H$. Equivalently, $E[F]$ is the maximum k for which there is a graph H such that $H \models EA_{k-1}$ while $F \not\sqsubset H$. Note that $E[F] \leq \ell$ for any ℓ-vertex graph F.

▶ **Lemma 4.** *$W^*[F] \geq E[F]$.*

Proof. As easily seen, if both G and H have the $(k-1)$-extension property, then Duplicator has a winning strategy in the $(k-1)$-pebble Ehrenfeucht-Fraïssé game on graphs G and H and, hence, $W(G,H) \geq k$. Therefore, it suffices to show that there are G and H such that $F \sqsubset G$, $F \not\sqsubset H$, and both of them satisfy EA_{k-1} for $k = E[F]$. Such a graph H exists by the definition of the extension index. Such a graph G exists because, as very well known (see, e.g., [25]), for fixed k and ℓ a random graph $G(n, 1/2)$ has the k-extension property and contains every ℓ-vertex graph as an induced subgraph with probability approaching 1 as n increases. Recall that $G(n,p)$ refers to the probability distribution on graphs with vertex set $\{1,\dots,n\}$ where each two vertices are adjacent with probability p independently of the other pairs. ◀

▶ **Example 5.**

1. $E[P_3] = 3$ because $H = 2K_2$ is P_3-free and satisfies EA_2. By Lemma 4, $W^*[P_3] = W[P_3] = 3$.
2. $E[K_3] = 3$, as also certified by $H = 2K_2$ (or by $H = C_4$).

We can determine $E[K_\ell]$ for any ℓ using a relationship between $E[F]$ and the chromatic number of F.

▶ **Theorem 6.** *$E[F] \geq \chi(F)$.*

Proof. Let $k = \chi(F) - 1$. We have to show that there is a graph G having the kth extension property and containing no induced copy of F.

Let $T_{k,n}$ denote the k-partite Turán graph with kn vertices. The vertex set of $T_{k,n}$ is split into k vertex classes $V_1, \ldots, V_k$, each consisting of n vertices. Two vertices of $T_{k,n}$ are adjacent if and only if they belong to different vertex classes. Obviously, $\chi(T_{k,n}) = k$. Since $\chi(T_{k,n}) < \chi(F)$, the graph $T_{k,n}$ itself and any of its subgraphs do not contain an induced copy of F. Let $\mathbb{T}_{k,n}$ be a random subgraph of $T_{k,n}$, obtained from $T_{k,n}$ by deleting each edge with probability $1/2$, independently of the other edges. In order to prove the theorem, it suffices to show that $\mathbb{T}_{k,n}$ has the kth extension property with nonzero probability if n is chosen sufficiently large.

Consider two disjoint vertex sets X, Y in $\mathbb{T}_{k,n}$ such that $|X \cup Y| = k - 1$ and estimate the probability that they violate EA_k. Fix a vertex class V_m disjoint with $X \cup Y$. A particular vertex $z \in V_m$ is adjacent to all $x \in X$ and non-adjacent to all $y \in Y$ with probability 2^{-k+1}, and the converse happens with probability $1 - 2^{-k+1}$. The probability that none of the vertices in V_m has the "right" adjacency pattern to X and Y is equal to $(1 - 2^{-k+1})^n$. Using the inequality

$$1 - x \leq \mathrm{e}^{-x} \text{ for all reals } x, \tag{2}$$

we conclude that two sets X, Y violating EA_k exist with probability at most

$$\binom{kn}{k-1} 2^{k-1}(1-2^{-k+1})^n \leq \left(\frac{kn\,\mathrm{e}}{k-1}\right)^{k-1} 2^{k-1}\mathrm{e}^{-2^{-k+1}n} = \left(\frac{2k\,\mathrm{e}}{k-1}\right)^{k-1} \mathrm{e}^{(k-1)\ln n - 2^{-k+1}n},$$

which approaches 0 as n increases (since k is fixed). It follows that EA_k is violated by $\mathbb{T}_{k,n}$ with probability strictly less than 1 if n is chosen sufficiently large. ◀

▶ **Corollary 7.** *$E[K_\ell] = \ell$.*

It may seem plausible at first glance that $E[F] = \ell$ for every F with ℓ vertices. Nevertheless, in Section 4 we will see that this is not always the case as, for example, $E[F] = 3$ for F being the paw and the path on 4 vertices. The best general lower bound for $E[F]$ we can show is given by the following lemma.

▶ **Lemma 8.** *Let F be a graph with $\ell \geq 2$ vertices. Then*

$$E[F] \geq \lfloor \frac{1}{2}\ell - 2\log_2 \ell + 3 \rfloor.$$

Proof. The lemma is trivially true if $\ell \leq 15$ because in this case it just states that $E[F] \geq 2$. We, therefore, suppose that $\ell \geq 16$.

Denote $k = \lfloor \frac{1}{2}\ell - 2\log \ell + 2 \rfloor$. Suppose that ℓ is even and set $n = 2^{\ell/2-1}$. It suffices to show that the random graph $G(n, 1/2)$ with a non-zero probability has the k-extension property and simultaneously contains no induced copy of F.

The probability of $F \sqsubset G(n,1/2)$ is bounded from above by the value of

$$p(\ell,n) = n(n-1)(n-2)\cdots(n-\ell+1)2^{-\ell(\ell-1)/2}.$$

Since

$$2^{-\ell(\ell-1)/2} = (2n)^{-\ell+1} = 2^{-\ell+1}n^{-\ell+1} = \frac{1}{2}\,n^{-\ell-1},$$

we have

$$p(\ell,n) = \frac{n(n-1)(n-2)\cdots(n-\ell+1)}{2\,n^{\ell+1}} < \frac{1}{2n}.$$

It remains to prove that $G(n,1/2)$ has the k-extension property with probability at least $1/(2n)$.

The probability of $G(n,1/2) \not\models EA_k$ is bounded from above by the value of

$$q(n,k) = \binom{n}{k-1}2^{k-1}(1-2^{-k+1})^{n-k+1}.$$

In its turn,

$$q(n,k) < 4\,n^{k-1}(1-2^{-k+1})^{n}.$$

By (2), the last value is bounded from above by

$$q'(n,k) = 4\,\exp\left(\ln n(k-1) - n\,2^{-k+1}\right).$$

Denote $k' = \frac{1}{2}\,\ell - 2\log\ell + 2$. Since the function $f(x) = \ln n\,x - n\,2^{-x}$ is monotonically increasing,

$$\begin{aligned} q(n,k) < q'(n,k') &= 4\,n^{k'-1}\exp\left(-\frac{n}{2^{\ell/2-2\log\ell+1}}\right) = 4\,n^{k'-1}\exp\left(-\frac{\ell^2}{4}\right) \\ &= 4\,n^{k'-1}(2n)^{-\log \mathrm{e}\,\ell/2} = 4\,n^{k'-1}2^{-\log \mathrm{e}\,\ell/2}n^{-\log \mathrm{e}\,\ell/2} = 4\,n^{k'-1}(2n)^{-\log \mathrm{e}}n^{-\log \mathrm{e}\,\ell/2} \\ &= 4\,\mathrm{e}^{-1}n^{-\ell(\log \mathrm{e}-1)/2-2\log\ell-\log \mathrm{e}+1}. \end{aligned}$$

For $\ell \geq 16$, this gives us $q(n,k) < 2\,n^{-11}$. Therefore, $G(n,1/2)$ has the k-extension property with probability more than $1-2\,n^{-11}$. This is well more than $1/(2n)$, as desired.

If ℓ is odd, set $n = 2^{(\ell-3)/2}$ and proceed similarly to above. ◀

▶ Remark. The bound of Lemma 8 cannot be much improved as long as the argument is based on $G(n,1/2)$. Indeed, it is known [15] that there is a function $\ell_0(n) = 2\log n - 2\log\log n + \Theta(1)$ such that the clique number of $G(n,1/2)$ is equal to $\ell_0(n)$ or to $\ell_0(n)+1$ with probability $1-o(1)$. In [16] it is shown that there is a function $k(n) = \log n - 2\log\log n + \Theta(1)$ such that, with probability $1-o(1)$, $G(n,1/2)$ satisfies $EA_{k(n)}$ but does not satisfy $EA_{k(n)+6}$. It follows that, if n is chosen so that $G(n,1/2)$ does not contain a subgraph K_ℓ with high probability, then $G(n,1/2)$ satisfies EA_k with non-negligible probability for, at best, $k = \frac{1}{2}\ell - \log\ell + \Theta(1)$.

As an immediate consequence of Lemmas 4 and 8 we obtain the following result.

▶ **Theorem 9.** *Let F be a graph with $\ell \geq 2$ vertices. Then*

$$W^*[F] > \frac{1}{2}\,\ell - 2\log_2\ell + 2.$$

4 Four-vertex subgraphs

Our next goal is to determine the values of $D[F]$, $W[F]$, and $W^*[F]$ for all graphs F with at most 4 vertices. It is enough to consider connected F, as follows from Lemma 3 and the fact that the complement of a disconnected graph is connected. The two connected 3-vertex graphs are considered in Example 5, and we now focus on connected graphs with 4 vertices. Recall that $W^*[K_4] = 4$ by Corollary 7 (or just because $W(K_4, K_3) = 4$).

4.1 The paw subgraph ($K_3 + e$)

▶ **Lemma 10** (Olariu [21]). *A graph H is paw-free if and only if each connected component of H is triangle-free or complete multipartite.*

Note that a graph B is complete multipartite iff the complement of B is a vertex-disjoint union of complete graphs. The latter condition means exactly that $\overline{B}$ is P_3-free. Thus, B is complete multipartite iff it is $(K_2 + K_1)$-free, where $K_2 + K_1 = \overline{P_3}$.

▶ **Theorem 11.** $D[K_3 + e] = W[K_3 + e] = W^*[K_3 + e] = E[K_3 + e] = 3$.

Proof. We have $E[K_3 + e] > 2$ because, for example, C_4 satisfies the 2nd extension axiom and does not contain $K_3 + e$ as a subgraph.

In order to prove that $D[K_3 + e] \le 3$, we have to describe a winning strategy for Spoiler in the 3-round Ehrenfeucht-Fraïssé game on graphs $G \sqsupset K_3 + e$ and $H \not\sqsupset K_3 + e$. Let v_1, v_2, v_3, v_4 be vertices spanning a paw in G. We suppose that v_1 and v_2 have degree 2, v_3 has degree 3, and v_4 has degree 1 in this subgraph. In the first round Spoiler pebbles v_1. Suppose that Duplicator responds with a vertex u_1 in a connected component B of H. By Lemma 10, B is either K_3-free or a multipartite graph with at least three parts. In the former case Spoiler wins by pebbling v_2 and v_3. In the latter case Spoiler pebbles v_4 in the second round. The distance between v_1 and v_4 in G is 2. If Duplicator responds in a connected component of H other than B, then he loses in the next round. Therefore, Duplicator is forced in the second round to pebble a vertex u_2 in the same part of B that contains u_1. In this case, Spoiler wins by pebbling the vertex v_2. Indeed, this vertex is adjacent to v_1 and not adjacent to v_4, while u_1 and u_2 have the same adjacency to any other vertex in H. ◀

4.2 The path subgraph (P_4)

$F = P_4$ is a remarkable example showing that the parameters $W^*[F]$ and $W[F]$ can have different values. Specifically, we prove that $W^*[P_4] = 3$ and $W[P_4] = 4$.

▶ **Theorem 12.** $W^*[P_4] = E[P_4] = 3$.

The proof of Theorem 12 is based on a well-known characterization of the class of P_4-free graphs. A graph is called a *cograph* if it can be built from copies of the single-vertex graph K_1 by using disjoint unions and complementations. It is known [6] that a graph is P_4-free if and only if it is a cograph. For the proof of Theorem 12 we need the following definitions, that we borrow from [22].

We call G *complement-connected* if both G and $\overline{G}$ are connected. An inclusion-maximal complement-connected induced subgraph of G will be called a *complement-connected component* of G or, for brevity, a *cocomponent* of G. Cocomponents have no common vertices and their vertex sets form a partition of $V(G)$. Note that G is a cograph if and only if all cocomponents of G are single-vertex graphs.

The *decomposition* of G, denoted by $Dec\,G$, is the set of all connected components of G. Furthermore, given $i \geq 0$, we define the *depth i decomposition* $Dec_i\,G$ of G by

$$Dec_0\,G = Dec\,G \quad \text{and} \quad Dec_{i+1}\,G = \bigcup_{E \in Dec_i\,G} Dec\,\overline{E}.$$

Note that

$$\Pi_i = \{\, V(E) : E \in Dec_i\,G\} \tag{3}$$

is a partition of $V(G)$, and Π_{i+1} refines Π_i. Once the partition stabilizes, that is, $\Pi_{i+1} = \Pi_i$, it coincides with the partition of G into its cocomponents. The *depth i environment* of a vertex $v \in V(G)$, denoted by $Env_i(v)$, is the graph E in $Dec_i\,G$ containing v.

▶ **Lemma 13.** *Suppose that a graph G contains an induced copy of P_4 and let $U \subseteq V(G)$ be such that $G[U] \cong P_4$. Consider the 3-pebble Ehrenfeucht-Fraïssé game on G and another graph H. Let $x, x' \in V(G)$ and $y, y' \in V(H)$, and assume that the pairs x, y and x', y' were selected by the players in the same rounds. If $x, x' \in U$ and $Env_l(y) \neq Env_l(y')$, then Spoiler has a strategy allowing him to win in this position playing all the time in U and making no more than $2l + 2$ moves.*

Proof. We use induction on l. In the base case of $l = 0$, the vertices y and y' lie in different connected components of H, while the distance between x and x' in G is at most 3. Therefore, Spoiler is able to win with one extra pebble in 2 moves.

Let $l \geq 1$. Suppose that $Env_{l-1}(y) = Env_{l-1}(y') = E$ (for else Spoiler wins by the induction assumption). Note that $Env_l(y)$ and $Env_l(y')$ are connected components of $\overline{E}$. Since P_4 is self-complementary, $\overline{G[U]} \cong P_4$. Therefore, if Duplicator moves only in $V(E)$, Spoiler will win in at most 2 next moves. Once Duplicator makes one of these moves outside $V(E)$, this creates a position with two vertices $\tilde{x}$ and $\tilde{x}'$ pebbled by Spoiler in U such that $Env_{l-1}(\tilde{y}) \neq Env_{l-1}(\tilde{y}')$ for the corresponding vertices $\tilde{y}$ and $\tilde{y}'$ pebbled by Duplicator in H. The induction assumption applies. ◀

Proof of Theorem 12. Consider graphs $G \sqsupset P_4$ and $H \not\sqsupset P_4$. Since H is a cograph, $Dec_i\,H$ for some i consists of single-vertex graphs. By Lemma 13, this readily implies that $W(G, H) \leq 3$. Indeed, when Spoiler pebbles two vertices on an induced P_4 in G, then whatever Duplicator responds, this creates a position as in Lemma 13. Thus, $W^*[P_4] \leq 3$.

The lower bound $E[P_4] > 2$ is certified, for example, by C_4. ◀

Theorem 12 implies that the class of graphs containing an induced P_4 is definable in the infinitary logic $L^3_{\infty\omega}$. It turns out that this class is not definable in 3-variable first-order logic.

▶ **Theorem 14.** *$W[P_4] = 4$.*

Our proof of Theorem 14 is based on Lemma 2. It suffices to exhibit a sequence of graph pairs G_i, H_i such that G_i contains an induced copy of P_4, H_i does not, and $D^3(G_i, H_i) \to \infty$ as i increases.

Given a graph K, we define its ith power K^i by $K^1 = K$ and $K^{i+1} = \overline{K^i + K^i}$. Now, let $H_i = (K_1)^i$. This is a cograph and, therefore, $P_4 \not\sqsubset H_i$ (which is also easy to see directly, using induction and the fact that P_4 is self-complementary).

In order to construct G_i, we use the lexicographic product of graphs; see Section 2. Fix a graph A satisfying the 3rd extension axiom and containing P_4 as an induced subgraph (a large enough random graph has both of these properties with high probability). Now, let $G_i = H_i \cdot (A \cdot H_i)$. Obviously, $P_4 \sqsubset G_i$. Theorem 14 immediately follows by Lemma 2 from the following estimate.

▶ **Lemma 15.** $D^3(G_{4m+2}, H_{4m+2}) > m + 1$.

The proof of Lemma 15 can be found in the full version of the paper [27].

4.3 The claw ($K_{1,3}$) and the diamond ($K_4 \setminus e$) subgraphs

A *strongly regular graph* with parameters (n, k, λ, μ) is a regular graph with n vertices of degree k such that every two adjacent vertices have λ common neighbors and every two non-adjacent vertices have μ common neighbors. The simplest examples are sK_t (the vertex-disjoint union of s copies of the complete graph K_t) and their complements (complete s-partite graphs with each vertex class of size t). We call such strongly regular graphs *trivial*. A strongly regular graph is non-trivial exactly if $0 < \mu < k < n - 1$.

An example of a non-trivial strongly regular graph, that will be useful for us below, is the $m \times m$*-rook graph*. The vertex set of this graph is $\{\,(a, b) : 1 \le a, b \le m\}$, and two vertices (a_1, b_1) and (a_2, b_2) are adjacent if $a_1 = a_2$ or $b_1 = b_2$. In other words, each vertex represents a square of the $m \times m$ chess board, and two squares are adjacent if one is reachable from the other by a move of the rook. The $m \times m$-rook graph is strongly regular with parameters $(m^2, 2m - 2, m - 2, 2)$.

The condition $\lambda = 0$ means that a strongly regular graph is K_3-free. Every complete bipartite graph $K_{n,n} = \overline{2K_n}$ has this property and seven other triangle-free non-trivial graphs are known. It is open whether there is yet another such graph [12].

Suppose that H is a non-trivial non-triangle-free strongly regular graph with parameters (n, k, λ, μ). Thus, $\mu < k$ and it is also not hard to see that $\lambda < k - 1$ (otherwise every two adjacent vertices were twins and, by connectedness, every two vertices would be adjacent twins, implying that H is complete). These two inequalities readily imply that H satisfies the 3rd extension axiom.

▶ **Theorem 16.** *$W[K_{1,3}] = W^*[K_{1,3}] = E[K_{1,3}] = 4$ and $W[K_4 \setminus e] = W^*[K_4 \setminus e] = E[K_4 \setminus e] = 4$.*

Proof. We have to show that $E[K_{1,3}] > 3$ and $E[K_4 \setminus e] > 3$. By the discussion above, it suffices to exhibit a non-trivial non-triangle-free strongly regular graph that does not contain the claw and the diamond graphs as induced subgraphs. The 3×3-rook graph suits these needs. ◀

4.4 The cycle subgraph (C_4)

Let G be a connected graph. Given $u, v \in V(G)$, let $f^G_{i,j}(u, v)$ denote the number of vertices at distance i from u and at distance j from v. The graph G is called *distance-regular* if the number $f^G_{i,j}(u, v) = f^G_{i,j}(d)$ depends only on i, j, and the distance $d = d(u, v)$ between u and v. Note that such a graph is regular of degree $f^G_{1,1}(0)$. We call two distance-regular graphs G and H *similar* if

$$f^G_{i,j}(d) = 0 \iff f^H_{i,j}(d) = 0. \tag{4}$$

▶ **Lemma 17.** *If G and H are similar distance-regular graphs, then $W(G, H) > 3$.*

Proof. We show a strategy allowing Duplicator to win the 3-pebble game on G and H. In the first round she responds Spoiler's move arbitrarily. Let x and x' be the vertices pebbled in G and H respectively. Suppose that in the second round Spoiler pebbles a vertex y in G (the case that Spoiler plays in H is similar). Duplicator responds with a vertex y' in H such that $d(x',y')=d(x,y)$, which guarantees that she does not lose in this round. Such a choice of y' is possible because (4) implies that $f^H_{i,i}(0)>0$ for $i=d(x,y)$.

For any subsequent round, assume that $x,y\in V(G)$ and $x',y'\in V(H)$ are occupied by two pairs of pebbles and that $d(x',y')=d(x,y)$. Suppose that Spoiler puts the third pebble on a vertex z in G (the case that Spoiler plays in H is similar). It is enough to notice that Duplicator can pebble a vertex z' in H such that $d(z',x')=d(z,x)$ and $d(z',y')=d(z,y)$. Such a vertex exists because (4) implies that $f^H_{i,j}(d)>0$ for $i=d(z,x)$, $j=d(z,y)$, and $d=d(x,y)$. ◀

▶ **Theorem 18.** $W[C_4]=W^*[C_4]=4$.

Proof. By Lemma 17, it suffices to exhibit similar distance-regular graphs G and H such that G contains an induced copy of C_4 and H does not. We can take G to be the cubic graph (the skeleton of the 3-dimensional cube) and $H=C_6$. ◀

5 Lower bounds over highly connected graphs

As we discussed in Section 1, in the case of a connected pattern graph F it is algorithmically motivated to consider the parameter $W_\kappa[F]$, which is the minimum variable width of a sentence defining the graph class $\mathcal{I}(F)$ correctly only over graphs of *sufficiently large* connectedness. More precisely, $W_\kappa[F]$ is equal to the minimum k for which there is a k-variable sentence Φ and a number s such that $G\models\Phi$ iff $F\sqsubset G$ for all s-connected graphs G. Moreover, we define

$$W^*_\kappa[F]=\min_s\max\{\,W(G,H):\ F\sqsubset G,\ F\not\sqsubset H,\ \text{and both } G \text{ and } H \text{ are } s\text{-connected}\}\,.$$

This parameter is an analog of $W_\kappa[F]$ for the infinitary logic, and we have

$$W^*_\kappa[F]\le W_\kappa[F]\le W[F] \text{ and } W^*_\kappa[F]\le W^*[F]\le W[F].$$

The *join* of graphs A and B is denoted by $A*B$. Recall that this is the graph obtained from the disjoint union of A and B by adding all possible edges between a vertex of A and a vertex of B.

▶ **Lemma 19.** $W(A*B,A'*B)\ge W(A,A')$.

Proof. In the game on $A*B$ and $A'*B$, Duplicator can use her strategy for the game on A and A'. Each time that Spoiler moves in the B part of one graph, Duplicator just mirrors his move in the B part of the other graph. ◀

▶ **Lemma 20.** *Let F_0 be obtained from F by removing all universal vertices from this graph. Then $W^*_\kappa[F]\ge W^*[F_0]$.*

Proof. Let m denote the number of universal vertices in F. Suppose that $W^*[F_0]=W(G,H)$, where G contains an induced copy of F_0 and H does not. Let G' be obtained from G by adding new $s>m$ universal vertices, and let H' be defined similarly, i.e, $G'=G*K_s$ and $H'=H*K_s$. Note that G' contains an induced copy of F and H' still does not contain even an induced copy of F_0 (no new vertex of H' can appear in an induced copy of F_0 because it would be universal there). We have $W(G',H')\ge W(G,H)$ by Lemma 19. This proves the claim because G' and H' are s-connected and s can be chosen arbitrarily large. ◀

▶ **Theorem 21.**

1. $W^*_\kappa[F] = W^*[F]$ *whenever* F *has no universal vertex.*
2. $W^*_\kappa[F] = W^*[F]$ *whenever* $W^*[F] > 3$ *and* F *has no adjacent twins or no non-adjacent twins.*

Proof. Part 1 is an immediate consequence of Lemma 20. To establish Part 2, we have to prove that $W^*_\kappa[F] \geq W^*[F]$. Since with 3 pebbles Spoiler can force playing on connected components, the assumption $W^*[F] > 3$ implies that $W^*[F] = W(G, H)$ for some connected G and H such that $F \sqsubset G$ and $F \not\sqsubset H$. Assume that F has no adjacent twins. Consider $G' = G \cdot K_s$ and $H' = H \cdot K_s$ (recall that $\cdot$ denotes the lexicographic product of graphs). Note that G' and H' are s-connected and observe that $W(G', H') \geq W(G, H)$. Moreover, G' still contains an induced copy of F, and H' still does not (because if an induced subgraph of H' contains two vertices from the same K_s-part, they are adjacent twins in this subgraph). Since s can be chosen arbitrarily large, this implies that $W^*_\kappa[F] \geq W(G, H) = W^*[F]$, as required. If F has no non-adjacent twins, the same argument works with $G' = G \cdot \overline{K_s}$ and $H' = H \cdot \overline{K_s}$. ◀

An example of a graph to which Theorem 21 is non-applicable is the diamond. It has universal vertices and both adjacent and non-adjacent twins.

If an ℓ-vertex pattern graph F has no universal vertex, then $W^*_\kappa[F] = W^*[F]$ and, therefore, $W^*_\kappa[F] \geq (\frac{1}{2} - o(1))\ell$ by Theorem 9. Lemma 20 works well also for F with few universal vertices. For example, we have $W^*_\kappa[K_{1,\ell}] \geq W^*[\overline{K_\ell}] = W^*[K_\ell] = \ell$. However, if F has many universal vertices, then we need a different approach.

Similarly to $E[F]$, we define $E_\kappa[F]$ to be the maximum k such that, for each s, there is an s-connected graph H with $H \models EA_{k-1}$ while $F \not\sqsubset H$.

The following relations are easy to prove similarly to Lemma 4 and Theorem 6 (note that, for each fixed s, the random multipartite graph $\mathbb{T}_{k,n}$ in the proof of Theorem 6 is s-connected with high probability).

▶ **Lemma 22.** $W^*_\kappa[F] \geq E_\kappa[F] \geq \chi(F)$.

▶ **Theorem 23.** *If* F *has* ℓ *vertices, then* $W^*_\kappa[F] > \frac{1}{3}\ell - \frac{4}{3}\log_2 \ell$.

Proof. Denote the number of universal vertices in F by m, and let F_0 be obtained from F by removing all these vertices. By Lemma 20 and Theorem 9,

$$W^*_\kappa[F] \geq W^*[F_0] > \frac{1}{2}(\ell - m) - 2\log \ell. \tag{5}$$

By Lemma 22,

$$W^*_\kappa[F] \geq \chi(F) \geq m. \tag{6}$$

Combining the bounds (5) and (6), we obtain

$$3\, W^*_\kappa[F] > \ell - 4\log \ell,$$

which implies the bound stated in the theorem. ◀

Finally, we determine the values of $W^*_\kappa[F]$ and $W_\kappa[F]$ for small connected pattern graphs. As a particular case of Lemma 22, we have $W^*_\kappa[K_3] = 3$ and $W^*_\kappa[K_4] = 4$. According to the discussion in Section 1, we have $W_\kappa[P_3] \leq D_v[P_3] \leq 2$. On the other hand, $W^*_\kappa[P_3] \geq 2$ just because there are highly connected graphs with an induced copy of P_3 and without it, for example, $K_n \setminus e$ and K_n respectively.

▶ **Theorem 24.**
1. $W^*_\kappa[K_3 + e] = W[K_3 + e] = 3$;
2. $W^*_\kappa[P_4] = W^*[P_4] = 3$ *and* $W_\kappa[P_4] = W[P_4] = 4$;
3. $W^*_\kappa[F] = W[F] = 4$ *for all remaining connected* F *on 4 vertices.*

Proof.
1. In the trivial direction, $W^*_\kappa[K_3 + e] \le W[K_3 + e] = 3$, the equality being established in Theorem 11. On the other hand, we have
$$W^*_\kappa[K_3 + e] \ge W^*[K_2 + K_1] = W^*[\overline{K_2 + K_1}] = W^*[P_3] \ge E[P_3] = 3,$$
where we use Lemma 20, Lemma 3, and Part 1 of Example 5.
2. We have $W^*_\kappa[P_4] = W^*[P_4] = 3$ by Part 1 of Theorem 21 and by Theorem 12. Since Lemma 20 easily extends to the relation $W_\kappa[F] \ge W[F_0]$, we also have $W_\kappa[P_4] = W[P_4] = 4$.
3. We have $W^*_\kappa[C_4] = W^*[C_4] = 4$ by Part 1 of Theorem 21 and by Theorem 18.

We also have $W^*_\kappa[K_{1,3}] = E_\kappa[K_{1,3}] = 4$ and $W^*_\kappa[K_4 \setminus e] = E_\kappa[K_4 \setminus e] = 4$ by the first inequality in Lemma 22. Indeed, $E_\kappa[K_{1,3}] = E_\kappa[K_4 \setminus e] = 4$ because the $m \times m$ rook's graph is a strongly regular graph containing no induced copy of $K_{1,3}$ and no induced copy of $K_4 \setminus e$. Moreover, the $m \times m$ rook's graph is $(2m - 2)$-connected by the following general fact: The connectivity of a connected strongly regular graph equals its vertex degree [2]. ◀

6 Conclusion

- The equality $D[K_3 + e] = 3$ is currently the only example we know that shows that $D[F]$ and $W[F]$ can be strictly less than the number of vertices in F. Are there other such graphs? Are there infinitely many of them? On the other hand, we only know that $W[F] \ge (\frac{1}{2} - o(1))\ell$ for all F with ℓ vertices. This does not even exclude the possibility that the time bound $O(n^{W[F]})$ for Induced Graph Isomorphism can be better than the Nešetřil-Poljak bound $O(n^{(\omega/3)\ell + c})$ for infinitely many pattern graphs F.
- An example of $F = P_4$ shows that $W^*[F]$ can be strictly less than $W[F]$ but we do not know how far apart from each other $W^*[F]$ and $W[F]$ can generally be.
- Note that Lemmas 4 and 8 show the following hierarchy of graph parameters:
$$(1/2 - o(1))\,\ell \le E[F] \le W^*[F] \le W[F] \le D[F] \le \ell, \tag{7}$$
where ℓ is the number of vertices in F. It seems that we currently have no example separating the parameters $W[F]$ and $D[F]$ or the parameters $E[F]$ and $W^*[F]$. An important question is whether or not $E[F] = (1 - o(1))\ell$.
- It follows from (7) that $W[F] \le (2 + o(1))W^*[F]$. In other terms, in the context of Induced Subgraph Isomorphism, the infinitary logic cannot be much more succinct than the standard first-order logic with respect to the number of variables. More generally, is it true that $W(\mathcal{C}) = O(W^*(\mathcal{C}))$ for all first-order definable graph properties?
- We have checked that $D[F] = W[F] = \ell$ for all F with $\ell \le 4$ vertices excepting for the paw graph and its complement. Since it remains open if the equality holds true for all larger graphs, it seems reasonable to examine the pattern graphs on 5 vertices. If there is a 5-vertex F with $W[F] = 4$, the resulting decision procedure for $\mathcal{I}(F)$ would be competitive to (or, at least comparable with) the currently known algorithmic results for 5-vertex induced subgraphs [9, 28].
- We have a (rather trivial) example of $F = P_3$ showing that $W_\kappa[F]$ can be strictly smaller than $W[F]$. Are there other such graphs?

References

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

2 Andries E. Brouwer and Dale M. Mesner. The connectivity of strongly regular graphs. *Eur. J. Comb.*, 6(3):215–216, 1985. doi:10.1016/S0195-6698(85)80030-5.

3 Jianer Chen, Xiuzhen Huang, Iyad A. Kanj, and Ge Xia. Strong computational lower bounds via parameterized complexity. *J. Comput. Syst. Sci.*, 72(8):1346–1367, 2006. doi:10.1016/j.jcss.2006.04.007.

4 Yijia Chen and Jörg Flum. On parameterized path and chordless path problems. In *Proc. of the 22nd Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 250–263. IEEE Computer Society, 2007. doi:10.1109/CCC.2007.21.

5 Derek G. Corneil, Yehoshua Perl, and Lorna K. Stewart. A linear recognition algorithm for cographs. *SIAM J. Comput.*, 14(4):926–934, 1985. doi:10.1137/0214065.

6 D. G. Corneil, H. Lerchs, and L. Stewart Burlingham. Complement reducible graphs. *Discrete Appl. Math.*, 3(3):163–174, 1981. doi:10.1016/0166-218X(81)90013-5.

7 Bruno Courcelle. The monadic second-order logic of graphs I. Recognizable sets of finite graphs. *Inf. Comput.*, 85(1):12–75, 1990. doi:10.1016/0890-5401(90)90043-H.

8 Friedrich Eisenbrand and Fabrizio Grandoni. On the complexity of fixed parameter clique and dominating set. *Theor. Comput. Sci.*, 326(1-3):57–67, 2004. doi:10.1016/j.tcs.2004.05.009.

9 Peter Floderus, Miroslaw Kowaluk, Andrzej Lingas, and Eva-Marta Lundell. Detecting and counting small pattern graphs. *SIAM J. Discrete Math.*, 29(3):1322–1339, 2015. doi:10.1137/140978211.

10 Peter Floderus, Mirosław Kowaluk, Andrzej Lingas, and Eva-Marta Lundell. Induced subgraph isomorphism: Are some patterns substantially easier than others? *Theoretical Computer Science*, 605:119–128, 2015. doi:10.1016/j.tcs.2015.09.001.

11 François Le Gall. Powers of tensors and fast matrix multiplication. In *Proc. of the Int. Symposium on Symbolic and Algebraic Computation (ISSAC'14)*, pages 296–303. ACM, 2014. doi:10.1145/2608628.2608664.

12 Chris D. Godsil. Problems in algebraic combinatorics. *Electr. J. Comb.*, 2:F#1, 1995. URL: http://www.combinatorics.org/Volume_2/PDFFiles/v2i1f1.pdf.

13 Neil Immerman. *Descriptive complexity.* New York, NY: Springer, 1999. doi:10.1007/978-1-4612-0539-5.

14 Alon Itai and Michael Rodeh. Finding a minimum circuit in a graph. *SIAM J. Comput.*, 7(4):413–423, 1978. doi:10.1137/0207033.

15 Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random graphs.* New York, Berlin: Wiley, 2000.

16 Jeong Han Kim, Oleg Pikhurko, Joel H. Spencer, and Oleg Verbitsky. How complex are random graphs in first order logic? *Random Struct. Algorithms*, 26(1-2):119–145, 2005. doi:10.1002/rsa.20049.

17 Ton Kloks, Dieter Kratsch, and Haiko Müller. Finding and counting small induced subgraphs efficiently. *Inf. Process. Lett.*, 74(3-4):115–121, 2000. doi:10.1016/S0020-0190(00)00047-8.

18 Yuan Li, Alexander A. Razborov, and Benjamin Rossman. On the AC^0 complexity of Subgraph Isomorphism. In *Proc. of the 55th IEEE Ann. Symposium on Foundations of Computer Science (FOCS'14)*, pages 344–353. IEEE Computer Society, 2014. doi:10.1109/FOCS.2014.44.

19 Leonid Libkin. *Elements of Finite Model Theory.* Texts in Theoretical Computer Science. An EATCS Series. Springer, 2004. doi:10.1007/978-3-662-07003-1.

20 Jaroslav Nešetřil and Svatopluk Poljak. On the complexity of the subgraph problem. *Commentat. Math. Univ. Carol.*, 26:415–419, 1985.

21 Stephan Olariu. Paw-free graphs. *Inf. Process. Lett.*, 28:53–54, 1988. `doi:10.1016/0020-0190(88)90143-3`.

22 Oleg Pikhurko, Joel Spencer, and Oleg Verbitsky. Decomposable graphs and definitions with no quantifier alternation. *Eur. J. Comb.*, 28(8):2264–2283, 2007. `doi:10.1016/j.ejc.2007.04.016`.

23 Oleg Pikhurko and Oleg Verbitsky. Logical complexity of graphs: a survey. In Martin Grohe and Janos Makowsky, editors, *Model theoretic methods in finite combinatorics*, volume 558 of *Contemporary Mathematics*, pages 129–179. American Mathematical Society (AMS), Providence, RI, 2011.

24 Benjamin Rossman. On the constant-depth complexity of k-clique. In *Proc. of the 40th Ann. ACM Symposium on Theory of Computing (STOC'08)*, pages 721–730. ACM, 2008. `doi:10.1145/1374376.1374480`.

25 Joel H. Spencer. *The strange logic of random graphs.* Berlin: Springer, 2001. `doi:10.1007/978-3-662-04538-1`.

26 Oleg Verbitsky and Maksim Zhukovskii. The descriptive complexity of Subgraph Isomorphism without numerics. In *Proc. of the 12th Int. Computer Science Symposium in Russia (CSR'17)*, volume 10304 of *Lecture Notes in Computer Science*, pages 308–322. Springer, 2017. A full version: `http://arxiv.org/abs/1607.08067`. `doi:10.1007/978-3-319-58747-9_27`.

27 Oleg Verbitsky and Maksim Zhukovskii. On the first-order complexity of Induced Subgraph Isomorphism. E-print, 2017. URL: `http://arxiv.org/abs/1704.02237`.

28 Virginia Vassilevska Williams, Joshua R. Wang, Richard Ryan Williams, and Huacheng Yu. Finding four-node subgraphs in triangle time. In *Proc. of the 26th Ann. ACM-SIAM Symposium on Discrete Algorithms (SODA'15)*, pages 1671–1680. SIAM, 2015. `doi:10.1137/1.9781611973730.111`.