The First Order Definability of Graphs: Upper Bounds for Quantifier Depth

Oleg Pikhurko[*] Helmut Veith[†] Oleg Verbitsky[‡]

Abstract

Let $D(G)$ denote the minimum quantifier depth of a first order sentence that defines a graph G up to isomorphism in terms of the adjacency and equality relations. Call two vertices of G similar if they have the same adjacency to any other vertex and denote the maximum number of pairwise similar vertices in G by $\sigma(G)$. We prove that $\sigma(G)+1 \le D(G) \le \max\{\sigma(G)+2, (n+5)/2\}$, where n denotes the number of vertices of G. In particular, $D(G) \le (n+5)/2$ for every G with no transposition in the automorphism group. If G is connected and has maximum degree d, we prove that $D(G) \le c_d n + O(d^2)$ for a constant $c_d < 1/2$. A linear lower bound for graphs of maximum degree 3 with no transposition in the automorphism group follows from an earlier result by Cai, Fürer, and Immerman (*Combinatorica* 12(4):389–410, 1992). Our upper bounds for $D(G)$ hold true even if we allow only definitions with at most one alternation in any sequence of nested quantifiers.

In passing we establish an upper bound for a related number $D(G, G')$, the minimum quantifier depth of a first order sentence which is true on exactly one of graphs G and G'. If G and G' are non-isomorphic and both have n vertices, then $D(G, G') \le (n+3)/2$. This bound is tight up to an additive constant of 1. If we additionally require that a sentence distinguishing G and G' is existential, we prove only a slightly weaker bound $D(G, G') \le (n+5)/2$.

[*]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213-3890 Web: `http://www.math.cmu.edu/~pikhurko/`

[†]Institut für Informatik, TU München, Germany. Research done in part while affiliated with Institut für Informationssysteme, Technische Universität Wien, A-1040 Wien, Austria. Supported by the European Community Research Training Network “Games and Automata for Synthesis and Validation” (GAMES) and by the Austrian Science Fund Project Z29-INF. E-mail: `veith@in.tum.de`

[‡]Department of Mechanics & Mathematics, Kyiv University, Ukraine. Research was done in part while visiting Institut für Informationssysteme, Technische Universität Wien. Supported by the Austrian Science Fund Project Z29-INF. E-mail: `oleg@ov.litech.net`

1 Introduction

1.1 Statement of the problem and overview of our results

From the logical point of view, a graph G is a structure with a single irreflexive and symmetric binary relation capturing the vertex adjacency. We consider first order sentences about graphs in the laconic language consisting of two relation symbols, $\leftrightarrow$ for the adjacency and $=$ for the equality. *First order* means that quantification is over vertices (as opposed to *second order* that would permit to quantify also over sets of vertices). We say that a sentence Φ *defines* a graph G if Φ is true on G and false on every graph non-isomorphic to G.

A well-known basic principle of finite model theory says that every finite graph has a defining sentence. For example, a graph G with vertex set $V(G) = \{1, \ldots, n\}$ and edge set $E(G)$ is defined by

$$\begin{gathered}\exists x_1 \ldots \exists x_n \Big(\bigwedge_{i \neq j} \neg\,(x_i = x_j) \wedge \bigwedge_{\{i,j\} \in E(G)} x_i \leftrightarrow x_j \wedge \bigwedge_{\{i,j\} \notin E(G)} \neg\,(x_i \leftrightarrow x_j) \Big) \\ \wedge\, \forall x_1 \ldots \forall x_n \forall x_{n+1} \Big(\bigvee_{i \neq j} x_i = x_j \Big).\end{gathered} \tag{1}$$

The sentence (1) is an exhaustive description of G and seems rather wasteful. We want to know if there is a more succinct way of defining a graph on n vertices. Among a few natural succinctness measures for a first order sentence Φ, we focus on its *quantifier depth* (or *quantifier rank*), which is the maximum number of nested quantifiers in Φ. Let $D(G)$ denote the minimum quantifier depth of a sentence defining G. We will call this graph invariant the *logical depth* of G. By (1) we have

$$D(G) \leq n + 1. \tag{2}$$

Since this bound is attained by the complete and the empty graph, it cannot be improved to a better bound in n. However, this fact does not make a further analysis worthless. As a first step, (2) can actually be improved to $D(G) \leq n$ for all G on n vertices different from the complete and the empty graph. We aim at yet further refining this bound. Since the complete and the empty graphs are the two graphs with the richest automorphism group, it is reasonable to try to improve (2) in dependence on how many automorphisms G has.

We call two vertices u and v of G *similar* if the transposition of u and v is an automorphism of G. Similarity is an equivalence relation on the vertex set and every similarity class is a *homogeneous* set of vertices, i.e., either a clique or an independent set. Denote the maximum number of pairwise similar vertices in G by $\sigma(G)$. We prove that

$$\sigma(G) + 1 \leq D(G) \leq \max\left\{\sigma(G) + 2, \frac{n+5}{2}\right\}. \tag{3}$$

These bounds can be restated as a dichotomy result. If $\sigma(G) \leq (n+1)/2$, which is a rather weak restriction on the automorphism group of G, we have $D(G) \leq (n+5)/2$, which is twice better than (2). If $\sigma(G) > (n+1)/2$, we are no less lucky having $D(G) \in \{\sigma(G)+1, \sigma(G)+2\}$. Moreover, in the latter case we are able to determine which of the two values is right: $D(G) = \sigma(G)+1$ iff the largest similarity class is an inclusion-maximal homogeneous set.

As an immediate consequence of (3), we have an upper bound

$$D(G) \le \frac{n+5}{2}$$

for all G whose automorphism group does not contain transpositions. We do not know if the factor of $1/2$ here is best possible, but it should be stressed that no sublinear upper bound is possible for this class of graphs. This follows from an earlier linear lower bound of Cai, Fürer, and Immerman [1].

The factor of $1/2$ can be improved for graphs with bounded vertex degree: For each $d \ge 2$ there is a constant $c_d < 1/2$ such that $D(G) \le c_d\, n + O(d^2)$ for any graph G with no isolated vertices and edges whose maximum degree is d. We do not try to find the best possible c_d being content with a constant strictly less than $1/2$. Notice a simpler linear lower bound $D(G_n) \ge n/(d+1)$ for $n = m(d+1)$ and G_n being the vertex disjoint union of m copies of the complete graph on $d+1$ vertices. Moreover, the aforementioned linear lower bound of Cai, Fürer, and Immerman holds even for connected graphs with maximum degree 3.

We say that a sentence Φ *distinguishes* graphs G and G' if Φ is true on exactly one of these graphs. Let $D(G, G')$ denote the minimum quantifier depth of a such sentence. Our estimation of $D(G)$ is based on two facts. First, $D(G)$ is equal to the maximum $D(G, G')$ over all G' non-isomorphic to G. Second, $D(G, G')$ admits a purely combinatorial characterization. This number is equal to the length of the *Ehrenfeucht game* on G and G' under the condition that the players play optimally.

Estimating $D(G, G')$ is not only important for our analysis of the logical depth of a graph, but interesting in its own right because of a connection to the graph isomorphism problem to be discussed in Section 1.2 below. We prove that

$$D(G, G') \le \frac{n+3}{2} \quad \text{if } G \text{ and } G' \text{ are non-isomorphic and both have } n \text{ vertices.} \tag{4}$$

To our best knowledge, no upper bound better than the trivial bound of n has been observed in the literature so far. Simple examples of graphs with $D(G, G') \ge (n+1)/2$ (see Example 2.6) show that the bound (4) is best possible for even n veritably and for odd n up to an additive constant of 1.

Without affecting the quantifier depth, we can assume that all negations in a first order sentence Φ stay in front of relation symbols. Under this assumption the *alternation number* of Φ is equal to the maximum number of alternations of nested quantifiers in Φ, from $\exists$ to $\forall$ and vice versa. Note that the generic defining sentence (1) has alternation number 0 and, moreover, the existential part of (1) suffices to distinguish G from every non-isomorphic G' that has the same number of vertices. It is worth noting in this respect that our results do not require a much bigger alternation number: The upper bounds (3) and (4) are proved with alternation number 1. Moreover, if we relax (4) only by an additive constant of 1, we can manage solely with existential distinguishing sentences.

1.2 A relation to the Weisfeiler-Lehman algorithm

Previous work on the logical definability of graphs is motivated by the relevance of the subject to the Graph Isomorphism problem. An important role is here played by another succinctness measure of a first order sentence Φ. Let $W(G)$ (resp. $W(G, G')$) denote the minimum number of variables in a Φ defining G (resp. distinguishing G and G'), where different occurrences of the same variable do not count. The number $W(G)$ will be referred to as the *logical width* of a graph G.

Analyzing the computational complexity of Graph Isomorphism from the logical perspective, it is worthwhile to enrich first order logic with counting quantifiers. In *logic with counting* we allow expressions like $\exists^{\ge m} x \Psi$ to say that there are at least m vertices x for which the statement Ψ holds. Irrespective of m, each counting quantifier contributes 1 in the quantifier depth. Let ${}_cW(G)$ and ${}_cW(G, G')$ be analogs of $W(G)$ and $W(G, G')$ in logic with counting.

Cai, Fürer, and Immerman [1] discovered a close connection of the logical width of a graph in logic with counting to the *multidimensional Weisfeiler-Lehman algorithm* for Graph Isomorphism. We refer the reader to [1] for a detailed exposition of the algorithm. An important parameter, which occurs at the exponent of an upper bound for the running time, is *dimension* of the algorithm. As established in [1], the optimum dimension on input (G, G') is equal to ${}_cW(G, G') - 1$. Since ${}_cW(G, G') \le {}_cW(G)$, the Weisfeiler-Lehman algorithm recognizes Graph Isomorphism in polynomial time for every class of graphs whose width in logic with counting is bounded by a constant. Examples of such classes are graphs of bounded genus [5] and graphs of bounded treewidth [6]. On the other hand, Cai, Fürer, and Immerman came up with a remarkable construction of non-isomorphic G and G' with ${}_cW(G, G') = \Omega(n)$. A simple analysis of their argument shows that this lower bound holds even for graphs with bounded vertex degree and with no transposition in the automorphism group. Some additional efforts undertaken in the earlier version of our paper [12] show that the multiplicative constant hidden in the Ω-notation is at least 0.00465.

Since every sentence of quantifier depth k can be equivalently rewritten with at most k variables, we have ${}_cW(G, G') \le W(G, G') \le D(G, G')$ and hence our bound (4) implies that the optimum dimension of the Weisfeiler-Lehman algorithm on input (G, G') does not exceed $(n + 1)/2$. Though a linear upper bound for the dimension does not entail any good bound for the running time, at least it shows an interesting combinatorial property of the algorithm that was previously never observed. Restating the results again in logical terms, for the maximum ${}_cW(G)$ over G on n vertices we have an upper bound of $0.5\,n + 0.5$ (due to the present paper) and a lower bound of $0.00465\,n$ (due to [1]). It would be interesting to make the gap closer.

1.3 Subsequent work

Since the appearance of the preliminary version [12] of this paper, the subject has developed in the following directions: Extending our current analysis from graphs to more general structures, obtaining stronger upper bounds on $D(G)$ for particular classes of graphs, estimating $D(G_{n,p})$ for a random graph $G_{n,p}$ in the the evolutional Erdős-Rényi model, and investigating the minimum $D(G)$ over G on n vertices. Here we survey some of these results.

In the preliminary version of this paper [12], we show how without much additional effort, the bounds (3) and (4) carry over to directed graphs and, more generally, to arbitrary relational structures with maximum arity 2. The interested reader can make the necessary modifications in the proof or find details in [12]. As another application of our approach, in [12] we also treat k-uniform hypergraphs. We obtain analogs of (3) and (4) that read

$$\sigma(G) + 1 \le D(G) \le \max\left\{\sigma(G) + k, \left(1 - \frac{1}{k}\right) n + 2k - 1\right\}$$

and

$$D(G, G') \le \left(1 - \frac{1}{k}\right) n + 2k - 1$$

for k-hypergraphs G and G' on n vertices. It remains open if the latter bound is tight for $k \ge 3$ since the only lower bound we know for any k is $(n+1)/2$. The case of general relational structures, where we cannot reckon on any symmetry of relations, is much more subtle. We treat it in [13] and obtain similar results comprising a bound of $(1 - \frac{1}{2k})\, n + k^2 - k + 2$, where k denotes the maximum relation arity of a structure. Note that for graphs the latter bound reads $\frac{3}{4}\, n + 4$ and therefore it does not supersede the results of the present paper.

In [15] it is shown that $D(G) = O(\log n)$ if G is a tree of bounded degree or a biconnected outerplanar graph. This upper bound complements the popular lower bounds $D(P_n, P_{n+1}) > \log_2 n - 3$ (e.g. [14, Theorem 2.1.3]) and $D(C_n, C_{n+1}) > \log_2 n$ (e.g. [3, Example 2.3.8]), where P_n is the path and C_n is the cycle on n vertices.

A thorough average-case analysis is undertaken in [8]. If a graph is chosen on n vertices uniformly at random, its logical depth is determined with high precision: $|D(G_{n,1/2}) - \log_2 n| = O(\log \log n)$ with probability $1 - o(1)$.

In [11] the "best case" behavior of $D(G)$ is investigated. Namely, we define a *succinctness function* $q(n)$ whose value is equal to the minimum $D(G)$ over graphs on n vertices. It is not hard to see that $q(n) \to \infty$ as $n \to \infty$ but it is not so clear how fast or slowly $q(n)$ grows. In [11] it is proved that $q(n)$ can be so small if compared to n that the gap between the two numbers cannot be estimated by any computable function, namely, there is no general recursive function f such that $f(q(n)) \ge n$. In particular, $q(n)$ does not admit any non-constant monotone computable lower bound. However, if we "smooth" $q(n)$ by considering its least monotone upper bound $q^*(n) = \max_{m \le n} q(m)$, we have $q^*(n) = (1 + o(1)) \log^* n$, where $\log^* n$ is the smallest number of iterations of the binary logarithm needed to bring n below 1.

It is worth noting that, like in the present paper, all the upper bounds mentioned above hold true even under the assumption that the alternation number is bounded by a small constant (dependent on a particular result).

1.4 Organization of the paper

Section 2 serves to recall relevant definitions from graph theory and logic as well as to state basic facts on the Ehrenfeucht games. In Sections 3 and 4 we prove our main results (4) and (3) respectively. In Section 5 we establish a variant of (4) for the fragment of first order logic with no quantifier alternation. The logical depth of graphs with bounded vertex degree is estimated in Section 6.

2 Preliminaries

2.1 Graphs

Given a graph G, we denote its vertex set by $V(G)$ and its edge set by $E(G)$. The number of vertices of G is called the *order* of G. The *neighborhood* of a vertex v consists of all vertices adjacent to v and is denoted by $\Gamma(v)$.

The *complement* of G, denoted by $\overline{G}$, is the graph on the same vertex set $V(G)$ with all those edges that are not in $E(G)$. Given G and G' with disjoint vertex sets, we define the *disjoint union* $G \sqcup G'$ to be the graph with vertex set $V(G) \cup V(G')$ and edge set $E(G) \cup E(G')$.

A set $S \subseteq V(G)$ is called *independent* if it contains no pair of adjacent vertices. S is a *clique* if all vertices in S are pairwise adjacent. The *complete graph* of order n, denoted by K_n, is a graph of order n whose vertex set is a clique. The complement of K_n is the *empty graph* of order n. The *complete bipartite graph* with vertex classes V_1 and V_2, where $V_1 \cap V_2 = \emptyset$, is a graph with the vertex set $V_1 \cup V_2$ and the edge set consisting of all edges $\{v_1, v_2\}$ for $v_1 \in V_1$ and $v_2 \in V_2$.

If $X \subseteq V(G)$, then $G[X]$ denotes the subgraph *induced* by G on X. If $X, Y \subseteq V(G)$ are disjoint, then $G[X, Y]$ denotes the bipartite graph induced by G on vertex classes X and Y, that is, $V(G[X, Y]) = X \cup Y$ and $E(G[X, Y])$ contains exactly those edges of G which connect a vertex in X with a vertex in Y.

We call $X \subseteq V(G)$ *homogeneous* if it is a clique or an independent set. We call a pair of disjoint sets $X, Y \subseteq V(G)$ *homogeneous* if $G[X, Y]$ is a complete or an empty bipartite graph.

We write $G \cong H$ if graphs G and H are isomorphic. If $U \subseteq V(G)$ and $W \subseteq V(H)$, we call a one-to-one map $\phi : U \to W$ a *partial isomorphism from G to H* if it is an isomorphism from $G[U]$ to $H[W]$.

2.2 Logic

For a backgrounding material in formal logic we refer the reader to any of the books [3, 7, 9] and use this subsection just to fix the terminology. We consider first order formulas in the relational vocabulary consisting of two relation symbols, for the vertex adjacency and the vertex equality. All formulas are supposed to be over the set of connectives $\{\neg, \wedge, \vee\}$ and with negations occurring only in front of relation symbols. A *sentence* is a closed first order formula. Referring to *nested quantifiers* in a sentence Φ we mean a sequence of quantifiers in Φ in which every subsequent quantifier is in the scope of the preceding one. The *quantifier depth* of Φ is defined to be the maximum number of nested quantifiers in Φ and denoted by $D(\Phi)$. The *alternation number* of Φ, denoted by $A(\Phi)$, is one smaller than the maximum number of alternating nested quantifiers in Φ (cf. inductive definitions in [10] or [13]). The *width* of Φ, denoted by $W(\Phi)$, is equal to the number of variables in Φ, where different occurrences of the same variable do not count[1].

Definition 2.1. Let G and G' be non-isomorphic graphs. A sentence Φ *distinguishes G from G'* if Φ is true on G but false on G'. By $D(G, G')$ (resp. $W(G, G')$) we denote the minimum $D(\Phi)$ (resp. $W(\Phi)$) where Φ ranges over all first order sentences which distinguish G from G'. If $k \geq 0$, then $D_k(G, G')$ denotes a variant of $D(G, G')$ under the additional condition that $A(\Phi) \leq k$.

Obviously $D(G, G') = D(G', G)$ and the other numbers $D_k(G, G')$ and $W(G, G')$ are symmetric as well. Note also that $W(G, G') \leq D(G, G') \leq D_k(G, G') \leq D_{k-1}(G, G')$ for every $k \geq 1$. The first inequality here follows from the observation that every Φ has an equivalent sentence Ψ with $W(\Psi) \leq D(\Phi)$.

Definition 2.2. A sentence Φ *defines* a graph G (up to isomorphism) if Φ distinguishes G from every non-isomorphic graph G'. By $D(G)$ (resp. $W(G)$) we denote the minimum $D(\Phi)$ (resp. $W(\Phi)$) over Φ defining G. These graph parameters will be referred to as, respectively, the *logical depth* and *logical width* of G. If $k \geq 0$, then $D_k(G)$ denotes a variant of $D(G)$ under the additional condition that $A(\Phi) \leq k$.

Notice relations $W(G) \leq D(G) \leq D_k(G) \leq D_{k-1}(G)$.

Proposition 2.3.

$$\begin{aligned} D(G) &= \max\{\, D(G, G') : G' \not\cong G \}\,, \\ D_k(G) &= \max\{\, D_k(G, G') : G' \not\cong G \}\,, \\ W(G) &= \max\{\, W(G, G') : G' \not\cong G \}\,. \end{aligned}$$

The first two equalities in Proposition 2.3 easily follow from the well-known fact that over a fixed finite vocabulary there are only finitely many inequivalent first order sentences of bounded quantifier depth. The third equality is not so obvious but immediately follows from a result of Dawar, Lindell, and Weinstein in [2]. The next proposition can be easily either proved directly or deduced from Proposition 2.5 below.

Proposition 2.4. *Let G and G' be non-isomorphic graphs.*

1. $D(G, G') = D(\overline{G}, \overline{G'})$, $D_k(G, G') = D_k(\overline{G}, \overline{G'})$, *and* $W(G, G') = W(\overline{G}, \overline{G'})$.

2. $D(G) = D(\overline{G})$, $D_k(G) = D_k(\overline{G})$, *and* $W(G) = W(\overline{G})$.

[1] We slightly deviate from the standard notion of width, cf. [4].

2.3 Games

We use the Immerman-Poizat version of an Ehrenfeucht game. Let G and G' be graphs with disjoint vertex sets. The *r-round l-pebble Ehrenfeucht game on G and G'*, which will be denoted by $\textsc{Ehr}_r^l(G, G')$, is played by two players, Spoiler and Duplicator, with l pairwise distinct pebbles $p_1, \ldots, p_l$, each given in duplicate. Spoiler starts the game. A *round* consists of a move of Spoiler followed by a move of Duplicator. At each move Spoiler takes a pebble, say p_i, selects one of the graphs G or G', and places p_i on a vertex of this graph. In response Duplicator should place the other copy of p_i on a vertex of the other graph. It is allowed to move previously placed pebbles to another vertex and place more than one pebble on the same vertex.

After each round of the game, let x_i (resp. x_i') denote the vertex of G (resp. G') occupied by p_i, irrespectively of who of the players placed the pebble here. If p_i is off the board at this moment, x_i and x_i' are undefined. If after every of r rounds the component-wise correspondence $(x_1, \ldots, x_l)$ to $(x_1', \ldots, x_l')$ is a partial isomorphism from G to G', this is a win for Duplicator; Otherwise the winner is Spoiler.

Note that, if we prohibit moving pebbles from one vertex to another in $\textsc{Ehr}_r^r(G, G')$, this will not affect the outcome of the game. We denote this variant of $\textsc{Ehr}_r^r(G, G')$ by $\textsc{Ehr}_r(G, G')$.

The *k-alternation* Ehrenfeucht game on G and G' is a variant of the game in which Spoiler is allowed to switch from one graph to another at most k times during the game, i.e., in at most k rounds he can choose the graph other than that in the preceding round.

Proposition 2.5. *Let G and G' be non-isomorphic graphs.*

1. *$W(G, G')$ equals the minimum l such that Spoiler has a winning strategy in $\textsc{Ehr}_r^l(G, G')$ for some r.*

2. *$D(G, G')$ equals the minimum r such that Spoiler has a winning strategy in $\textsc{Ehr}_r(G, G')$.*

3. *$D_k(G, G')$ equals the minimum r such that Spoiler has a winning strategy in the k-alternation $\textsc{Ehr}_r(G, G')$.*

We refer the reader to [7, Theorem 6.10] for the proof of the first two claims and to [10] for the third claim.

The examples below are obtained by simple application of Proposition 2.5.

Example 2.6.

1. $W(K_m \sqcup \overline{K_m}, K_{m+1} \sqcup \overline{K_{m-1}}) = m$, $D(K_m \sqcup \overline{K_m}, K_{m+1} \sqcup \overline{K_{m-1}}) = m + 1$.

2. $W(K_{m+1} \sqcup \overline{K_m}, K_m \sqcup \overline{K_{m+1}}) = D(K_{m+1} \sqcup \overline{K_m}, K_m \sqcup \overline{K_{m+1}}) = m + 1$.

3 Distinguishing non-isomorphic graphs

Theorem 3.1. *If G and G' are non-isomorphic graphs of the same order n, then $D_1(G, G') \le (n+3)/2$.*

The proof takes up the rest of this section. It is based on the characterization of $D_1(G, G')$ as the length of the 1-alternation Ehrenfeucht game on G and G' given by Proposition 2.5.3. The most essential part of the proof is contained in Lemma 3.14 that gives a winning strategy for Spoiler. This lemma will be used also in subsequent sections to prove our other results.

3.1 Spoiler's preliminaries

We first introduce a couple of useful relations between vertices of a graph that will be intensively exploited in the course of the proofs.

Definition 3.2. We call vertices u and v of a graph G *similar* and write $u \sim v$ if the transposition (uv) is an automorphism of G. Let $[u]_G = \{v \in V(G) : v \sim u\}$, $\sigma_G(u) = |[u]_G|$, and $\sigma(G) = \max_{u \in V(G)} \sigma_G(u)$. If the graph is clear from the context, the subscript G may be omitted. We will call the numbers $\sigma(v)$ and $\sigma(G)$ the *similarity indices* of the vertex v and the graph G respectively.

We will say that t *separates* u and v if t is adjacent to exactly one of the two vertices. Thus, $u \sim v$ iff these vertices are inseparable by any $t \in V(G) \setminus \{u, v\}$.

Lemma 3.3.

1. *$\sim$ is an equivalence relation on $V(G)$.*
2. *Every equivalence class $[u]$ is a homogeneous set.*

Proof. The lemma is straightforward. The only care should be taken to check the transitivity. Given pairwise distinct u, v, and w, let us deduce from $u \sim v$ and $v \sim w$ that $u \sim w$. For every $t \ne u, w$, we need to show that u and t are adjacent iff so are w and t. If $t \ne v$, this is true because both adjacencies are equivalent to the adjacency of v and t. There remains the case that $t = v$. Then u and v are adjacent iff so are u and w (as $v \sim w$), which in turn holds iff so are w and v (as $u \sim v$). ∎

Definition 3.4. Given $X \subset V(G)$, we will denote its complement by $\overline{X} = V(G) \setminus X$. Let $u, v \in \overline{X}$. We write $u \equiv_X v$ if the identity map of X onto itself extends to an isomorphism from $G[X \cup \{u\}]$ to $G[X \cup \{v\}]$.

In other words, $u \equiv_X v$ if these vertices have the same adjacency pattern to X, i.e., $\Gamma(u) \cap X = \Gamma(v) \cap X$. Clearly, $\equiv_X$ is an equivalence relation on $\overline{X}$.

Definition 3.5. $\mathcal{C}(X)$ is the partition of $\overline{X}$ into $\equiv_X$-equivalence classes.

Let us notice a few straightforward properties of this partition. If $X_1 \subseteq X_2$, then $\mathcal{C}(X_2)$ is a refinement of $\mathcal{C}(X_1)$ on $\overline{X_2}$. For any X, the $\sim$-equivalence classes restricted to $\overline{X}$ refine the partition $\mathcal{C}(X)$.

Definition 3.6. Let $X \subset V(G)$. We say that X is $\mathcal{C}$-*maximal* if $|\mathcal{C}(X \cup \{u\})| \le |\mathcal{C}(X)|$ for any $u \in \overline{X}$.

Lemma 3.7. *Let $X \subset V(G)$ be $\mathcal{C}$-maximal. Then the partition $\mathcal{C}(X)$ has the following properties.*

1. *Every C in $\mathcal{C}(X)$ is a homogeneous set.*
2. *If C_1 and C_2 are distinct classes in $\mathcal{C}(X)$ and have at least two elements each, then the pair C_1, C_2 is homogeneous.*

Proof. **1.** Suppose, to the contrary, that C is neither a clique nor an independent set. Then C contains three vertices u, v, and w such that u and v are adjacent but u and w are not. However, if we move u to X, then C splits into two classes, one containing u and another containing w. Hence the number of equivalence classes increases at least by 1, a contradiction.

2. Suppose that this is not true, for example, $u \in C_1$ is adjacent to $v \in C_2$ but not to $w \in C_2$. If we move u to X, then C_2 splits into two non-empty classes and $C_1 \setminus \{u\}$ stays non-empty. Again the number of equivalence classes increases, a contradiction. ∎

Lemma 3.8. *In every graph G of order n, there exists a $\mathcal{C}$-maximal set of vertices X such that*

$$|\mathcal{C}(X)| \ge |X| + 1 \tag{5}$$

and hence

$$|X| \le \frac{n-1}{2}. \tag{6}$$

Proof. Such an X can be constructed, starting from $X = \emptyset$, by repeating the following procedure. As long as there exists $u \in \overline{X}$ such that $\mathcal{C}(X \cup \{u\}) > \mathcal{C}(X)$, we move u to X. As soon as there is no such u, we arrive at X which is $\mathcal{C}$-maximal. The relation (5) is true as it holds at the beginning and is preserved in each step. The bound (6) follows from the inequality $|X| + |\mathcal{C}(X)| \le n$. ∎

Definition 3.9. Let $X \subset V(G)$.
$Y(X)$ is the union of all single-element classes in $\mathcal{C}(X)$.
$Z(X) = V(G) \setminus (X \cup Y(X))$.
$\mathcal{D}(X)$ is the partition of $Z(X)$ defined by $\mathcal{D}(X) = \mathcal{C}(X \cup Y(X))$.

Clearly, $\mathcal{D}(X)$ refines the partition induced on $Z(X)$ by $\mathcal{C}(X)$.

Lemma 3.10. *If $X \subset V(G)$ is $\mathcal{C}$-maximal, then every class D in $\mathcal{D}(X)$ consists of pairwise similar vertices. Thus, $\mathcal{D}(X)$ coincides with the partition induced on $Z(X)$ by $\sim$-equivalence classes.*

Proof. Let u and v be distinct elements of the same class $D \in \mathcal{D}(X)$. These vertices cannot be separated by any vertex $t \in X \cup Y(X)$ by the definition of $\mathcal{D}(X)$. Assume that they are separated by a $t \in Z(X)$. Let C_1 be the class in $\mathcal{C}(X)$ including D and C_2 be the class in $\mathcal{C}(X)$ containing t. Since $t \notin Y(X)$, the class C_2 has at least one more element except t. If $C_1 \neq C_2$, moving t to X splits up C_1 and does not eliminate C_2. If $C_1 = C_2$, moving t to X splits up this class and splits up or does not affect the others. In either case $|\mathcal{C}(X)|$ increases, giving a contradiction. ∎

Definition 3.11. Let $\phi : X \to X'$ be a partial isomorphism from G to G'. Let $v \in \overline{X}$ and $v' \in \overline{X'}$. We call vertices v and v' *ϕ-similar* and write $v \equiv_\phi v'$ if ϕ extends to an isomorphism from $G[X \cup \{v\}]$ to $G'[X' \cup \{v'\}]$.

Note that, if $u \equiv_X v$ and $u' \equiv_{X'} v'$, then $u \equiv_\phi u'$ iff $v \equiv_\phi v'$. This makes the following definition correct.

Definition 3.12. Let $\phi : X \to X'$ be a partial isomorphism from G to G'. Let $C \in \mathcal{C}(X)$ and $C' \in \mathcal{C}(X')$. We call C and C' *ϕ-similar* and write $C \equiv_\phi C'$ if $v \equiv_\phi v'$ for some (equivalently, for all) $v \in C$ and $v' \in C'$.

Notice that, if $u \equiv_\phi u'$ and $v \equiv_\phi v'$, then the relations $u \equiv_X v$ and $u' \equiv_{X'} v'$ are true or false simultaneously. It follows that the ϕ-similarity is a matching between the classes in $\mathcal{C}(X)$ and the classes in $\mathcal{C}(X')$, i.e., no class can have more than one ϕ-similar counterpart in the other graph.

3.2 Spoiler's strategy

Definition 3.13. Let $l \geq 1$. If $v \in V(G)$, the notation $H = G \oplus_l v$ means that

- $\sigma_G(v) \geq 2$ and
- H is a graph obtained from G by adding new vertices $v_1, \ldots, v_l$ so that $[v]_H = [v]_G \cup \{v_1, \ldots, v_l\}$.

In other words, each v_i has the same adjacency pattern to $V(G) \setminus \{v\}$ as v and is adjacent or not to v (as well as to any other v_j) depending on if $[v]_G$ is a clique or an independent set.

Convention. In the sequel, writing $H = G \oplus_l v$ we will assume that H is an arbitrary isomorphic copy of $G \oplus_l v$. When considering the Ehrenfeucht game on G and H, the vertex sets of these graphs will be assumed disjoint.

Lemma 3.14. *If G and G' are non-isomorphic graphs of orders n and n' respectively and $n \le n'$, then*

$$D_1(G, G') \le (n+5)/2 \tag{7}$$

unless $G' = G \oplus_{n'-n} v$ for some $v \in V(G)$.

Proof. We will describe a strategy of Spoiler winning $\text{E}\textsc{hr}_r(G, G')$ for $r = \lfloor (n+5)/2 \rfloor$ unless $G' = G \oplus_{n'-n} v$. The strategy splits the game in two phases.

Phase 1

Spoiler selects a $\mathcal{C}$-maximal set of vertices $X \subset V(G)$ such that $|\mathcal{C}(X)| \ge |X|+1$, whose existence is guaranteed by Lemma 3.8. Denote $s = |X|$, the number of rounds in Phase 1, and $t = |\mathcal{C}(X)|$. The bounds (5) and (6) read

$$t \ge s+1 \tag{8}$$

and

$$s \le \frac{n-1}{2}. \tag{9}$$

If Duplicator loses in Phase 1, by (9) this happens within the claimed bound.

Let $X' \subseteq V(G')$ consist of the vertices selected in Phase 1 by Duplicator and $\phi : X \to X'$ be the bijection defined by the condition that x and $\phi(x)$ are selected by the players in the same round. We assume that Phase 1 finishes without Duplicator losing and hence ϕ is a partial isomorphism from G to G'. The following useful observation is straightforward from Definition 3.11.

Claim 3.14.1. Whenever after Phase 1 Spoiler selects a vertex v in $V(G) \setminus X$ or $V(G') \setminus X'$, Duplicator responds with a ϕ-similar vertex or otherwise immediately loses. □

Phase 2

Denote the classes of $\mathcal{C}(X)$ by $C_1, \ldots, C_t$ and the classes of $\mathcal{C}(X')$ by $C'_1, \ldots, C'_{t'}$. If there is a class C_i or C'_j without any ϕ-similar counterpart, respectively, in $\mathcal{C}(X')$ or in $\mathcal{C}(X)$, then Spoiler selects a vertex in this class and wins according to Claim 3.14.1, making at total $s+1 \le (n+1)/2$ moves and at most one alternation between the graphs. From now on, we therefore assume that the ϕ-similarity determines a perfect matching between $C_1, \ldots, C_t$ and $C'_1, \ldots, C'_{t'}$, where actually $t = t'$. For the notational convenience, we assume that $C_i \equiv_\phi C'_i$ for all $i \le t$.

Furthermore, if there is a singleton C_i or C'_j whose ϕ-similar counterpart has at least two vertices, Spoiler selects such two vertices and, again by Claim 3.14.1, wins in $s+2 \le (n+3)/2$ moves with at most one alternation. We will therefore assume that $|C_i| = 1$ iff $|C'_i| = 1$. Without loss of generality, assume that $|C_i| = |C'_i| = 1$ iff $i \le q$.

Denote $Y = Y(X)$ and $Y' = Y(X')$. Let $C_i = \{y_i\}$ and $C'_i = \{y'_i\}$ for $i \le q$. Thus, $Y = \{y_1, \ldots, y_q\}$ and $Y' = \{y'_1, \ldots, y'_q\}$. Define $\phi^* : X \cup Y \to X' \cup Y'$, an extension of ϕ, by $\phi^*(y_i) = y'_i$.

Claim 3.14.2. ϕ^* is a partial isomorphism from G to G', unless Spoiler wins in the next 2 moves with no alternation, having made at total $s+2 \le (n+3)/2$ moves.

Proof of Claim. For every $i \le q$, the restriction $\phi^* : X \cup C_i \to X' \cup C'_i$ is an isomorphism because C_i and C'_i are ϕ^*-similar. The restriction $\phi^* : Y \to Y'$ should be an isomorphism as well by the following reason. If there are $i, j \le q$ such that y_i and y_j are adjacent but y'_i and y'_j are not or vice versa, then Spoiler wins on the account of Claim 3.14.1 by selecting y_i and y_j. $\square$

We will therefore assume that ϕ^* is indeed a partial isomorphism from G to G'. Let $Z = Z(X)$ and $Z' = Z(X')$. Denote the classes of $\mathcal{D}(X)$ by $D_1, \ldots, D_p$ and the classes of $\mathcal{D}(X')$ by $D'_1, \ldots, D'_{p'}$. Note that

$$p \ge t - q \tag{10}$$

and

$$p = t - q \text{ iff } \mathcal{D}(X) = \{C_{q+1}, \ldots, C_t\}.$$

Claim 3.14.3. Whenever in Phase 2 Spoiler selects a vertex $v \in Z \cup Z'$, Duplicator responds with a ϕ^*-similar vertex or otherwise Spoiler wins in the next round at latest, with no alternation between G and G' in this round.

Proof of Claim. Let u be the vertex selected by Duplicator in response to v and assume that $u \not\equiv_{\phi^*} v$. Suppose that $v \in Z'$ (the case of $v \in Z$ is completely similar). If $u \notin Z$, then $u \not\equiv_{\phi} v$ and Duplicator has already lost by Claim 3.14.1. If $u \in Z$, then there exists a vertex $w \in X \cup Y$ such that u and w are adjacent but v and $\phi^*(w)$ are not or vice versa. If $w \in X$, again Duplicator has already lost. If $w \in Y$, then in the next round Spoiler selects $\phi^*(w)$ and wins. $\square$

Claim 3.14.3 implies that every class in $\mathcal{D}(X)$ or $\mathcal{D}(X')$ has a ϕ^*-similar counterpart in, respectively, $\mathcal{D}(X')$ or $\mathcal{D}(X)$ unless Spoiler wins making in Phase 2 two moves and at most one alternation between the graphs. We will therefore assume that this is true, that is, the ϕ^*-similarity determines a perfect matching between the classes $D_1, \ldots, D_p$ and $D'_1, \ldots, D'_{p'}$, where actually $p = p'$. For the notational convenience, we assume that $D_i \equiv_{\phi^*} D'_i$ for all $i \le p$.

Claim 3.14.4. Unless Spoiler is able to win making 2 moves and at most 1 alternation in Phase 2, the following conditions are met.

1. For every $i \le p$, D_i and D'_i are simultaneously cliques or independent sets.
2. For every pair of distinct $i, j \le p$, $G[D_i, D_j]$ and $G'[D'_i, D'_j]$ are simultaneously complete or empty bipartite graphs.

Proof of Claim. 1. Suppose that both D_i and D'_i have at least 2 vertices. Since D_i consists of $X \cup Y$-similar and hence X-similar vertices, by Lemma 3.7.1, D_i is either a clique or an independent set. This is actually true for the class $C \in \mathcal{C}(X)$ including D_i. If D'_i is not a clique or an independent set simultaneously with D_i, Spoiler wins in 2 moves with 1 alternation by selecting in D'_i two vertices which are non-adjacent in the former case and adjacent in the latter case. Indeed, if Duplicator responds with two vertices in C, those are in the opposite adjacency relation. If at least one Duplicator's response is not in C, he loses by Claim 3.14.1.

2. Assume first that D_i and D_j are included in the same class $C \in \mathcal{C}(X)$. Then D'_i and D'_j must be included in the same $C' \in \mathcal{C}(X')$. By Lemma 3.7.1, $G[C]$ is either complete or empty and hence so is $G[D_i, D_j]$. The graphs $G'[C']$ and $G'[D'_i, D'_j]$ must be complete or empty respectively unless Spoiler wins in 2 moves with 1 alternation similarly to Item 1.

Assume now that D_i and D_j are included in different classes of $\mathcal{C}(X)$, C^1 and C^2 respectively. Since both C^1 and C^2 have at least 2 vertices, $G[D_i, D_j]$ is either complete or empty according to Lemma 3.7.2. If $G'[D'_i, D'_j]$ is not, respectively, complete or empty, then Spoiler wins in 2 moves with 1 alternation by selecting, respectively, non-adjacent or adjacent vertices, one in D'_i and another in D'_j. Indeed, if Duplicator responds with one vertex in C^1 and another in C^2, those are in the opposite adjacency relation. Otherwise Duplicator loses by Claim 3.14.1. $\square$

Thus, in what follows we assume that the two conditions in Claim 3.14.4 are obeyed. Together with the fact that the D_i's and the D'_i's are classes of the partitions $\mathcal{C}(X \cup Y)$ and $\mathcal{C}(X' \cup Y')$, this implies that each D_i and each D'_i consists of pairwise similar vertices (in the sense of Definition 3.2). Moreover, since $D_i \equiv_{\phi^*} D'_i$ for every $i \le p$, where ϕ^* is an isomorphism from $G[X \cup Y]$ to $G'[X' \cup Y']$ and D_i and D'_i are simultaneously cliques or independent sets, the graphs G and G' would be isomorphic if $|D_i| = |D'_i|$ for every $i \le p$. Since G and G' are supposed to be non-isomorphic, there is D_i such that $|D_i| \ne |D'_i|$. We will call such a D_i *useful* (for Spoiler).

Claim 3.14.5. If D_i is useful and $p > t - q$, then Spoiler is able to win having made in Phase 2 at most $\min\{|D_i|, |D'_i|\} + 2$ moves and at most 1 alternation between G and G'. If $p = t - q$, then $\min\{|D_i|, |D'_i|\} + 1$ moves and 1 alternation suffice.

Proof of Claim. Spoiler selects $\min\{|D_i|, |D'_i|\}+1$ vertices in the larger of the classes D_i and D'_i. Duplicator is enforced to at least once reply with not a ϕ^*-similar vertex. Then, according to Claim 3.14.3, Spoiler wins in the next move at latest. If $p = t-q$ and hence the D-classes coincide with the C-classes, this extra move is not needed. This follows from Claim 3.14.1 because in this case violation of the ϕ^*-similarity causes violation of the ϕ-similarity. $\square$

Suppose that there are two useful classes, D_i and D_j. Observe that

$$\begin{aligned} &|D_i| + |D_j| = |Z| - \textstyle\sum_{l \ne i,j} |D_l| \le (n - s - q) - (p - 2) \\ &\le \begin{cases} (n - s - q) - (t - q - 1) \le n - 2s & \text{if } \ p > t - q, \\ (n - s - q) - (t - q - 2) \le n - 2s + 1 & \text{if } \ p = t - q, \end{cases} \end{aligned} \tag{11}$$

where we use (10) and (8). It follows that one of the useful classes has at most $(n-2s+1)/2$ vertices if $p = t-q$ and at most $(n-2s)/2$ vertices if $p > t-q$. Therefore, if $p = t - q$, Spoiler wins the game in at most $s + (n - 2s + 1)/2 + 1 = (n + 3)/2$ moves and, if $p > t - q$, in at most $s + (n - 2s)/2 + 2 = (n + 4)/2$ moves, which is within the required bound (7).

Finally, suppose that there is a unique useful class D_m. According to Claim 3.14.5, Spoiler is able to win in at most $|D_m| + 2$ moves, with the total number of moves $s + |D_m| + 2$ that is within the required bound (7) provided $|D_m| = 1$. Thus, we arrive at the conclusion that the bound (7) may not hold true in the only case that there is exactly one useful class D_m and $|D_m| \geq 2$. Note that we then have $n' > n$ and $|D'_m| = |D_m| + (n' - n)$. It remains to notice that, if we remove $n' - n$ vertices from D'_m, we obtain a graph isomorphic to G. It follows that $G' = G \oplus_{n'-n} v$ with $v \in D_m$. ∎

Note a direct consequence of Lemma 3.14 and Proposition 2.3, that will be significantly improved in the next section.

Corollary 3.15. *If $\sigma(G) = 1$, then $D_1(G) \leq (n+5)/2$.* ∎

Proof of Theorem 3.1. Lemma 3.14 immediately gives us an upper bound of $(n+5)/2$, which is a bit worse than we now claim. To improve it, we go trough lines of the proof of Lemma 3.14 but make use of the equality $n = n'$. The latter causes the following changes. Since $n' = n$, there must be at least two useful classes, D_i and D_j, such that $|D_i| < |D'_i|$ and $|D_j| > |D'_j|$. If $p = t - q$, the bound of $(n+3)/2$ has been actually proved, and we only need to tackle the case that $p > t - q$. Similarly to (11), we have

$$2|D_i| + 2|D'_j| + 2 \leq |D_i| + |D_j| + |D'_i| + |D'_j| \leq 2\left((n - s - q) - (p - 2)\right) \leq 2(n - 2s).$$

It follows that at least one of $|D_i|$ and $|D'_j|$ does not exceed $(n - 2s - 1)/2$. By Claim 3.14.5, Spoiler wins in totally at most $s + (n - 2s - 1)/2 + 2 = (n+3)/2$ moves. ∎

In the conclusion of this section, we state a lemma for further use in Section 6. This lemma is actually a corollary from the proof of Lemma 3.14. More precisely, it is a variant of Claim 3.14.5, where we take into account Lemma 3.10.

Lemma 3.16. *Let G and G' be arbitrary non-isomorphic graphs. Suppose that $X \subset V(G)$ is $\mathcal{C}$-maximal. Then Spoiler wins the 1-alternation Ehrenfeucht game on G and G' in at most $|X| + \max_{v \notin X} \sigma_G(v) + 2$ rounds.* ∎

4 Defining a graph

Theorem 4.1. *Let G be a graph of order n.*

1.
$$\sigma(G) + 1 \leq W(G) \leq \max\left\{\sigma(G) + 1, \frac{n+5}{2}\right\}.$$

Thus if, and only if, $\sigma(G) \leq (n+3)/2$, we have an upper bound $W(G) \leq (n+5)/2$, while if $\sigma(G) \geq (n+3)/2$, we explicitly know the exact value $W(G) = \sigma(G) + 1$.

2.
$$\sigma(G) + 1 \leq D(G) \leq D_1(G) \leq \max\left\{\sigma(G) + 2, \frac{n+5}{2}\right\}.$$

Thus if $\sigma(G) \leq (n+1)/2$, we have an upper bound $D_1(G) \leq (n+5)/2$, while if $\sigma(G) \geq (n+1)/2$, we know that $D(G) \in \{\sigma(G)+1, \sigma(G)+2\}$.

3. *Assume that $\sigma(G) \geq (n+2)/2$. Then $D(G) = D_1(G) = \sigma(G) + 1$ if the largest similarity class in G is an inclusion-maximal homogeneous set and $D(G) = D_1(G) = \sigma(G) + 2$ otherwise.*

4. *Given G, one can efficiently check whether or not $W(G) \le (n+5)/2$. If this is not true, one can efficiently compute the exact value of $W(G)$. All the same holds for $D(G)$.*

Item 4 is worth noting in view of the fact that algorithmic computability of the logical depth and width of a graph, even with no efficiency requirements, is unclear. A reason for this is that the question if a given first order sentence defines a graph is known to be undecidable [11].

The proof of the theorem is based on the following four lemmas.

Lemma 4.2. *Let x_i (resp. x'_i) denote the vertex of G (resp. G') selected in the i-th round of $\textsc{Ehr}_r(G, G')$. Then, as soon as a move of Duplicator violates the condition that $x_i \sim x_j$ iff $x'_i \sim x'_j$, Spoiler wins either immediately or in the next move possibly with one alternation between the graphs.*

Proof. Suppose, for example, that Duplicator selects x'_j so that $x'_i \not\sim x'_j$ while $x_i \sim x_j$ for some $i < j$. Suppose that the correspondence between the x_m's and the x'_m's, $1 \le m \le j$, is still a partial isomorphism. Then there is $y \in V(G')$ adjacent to exactly one of x'_i and x'_j. Note that such y could not be selected by the players previously. In the next move Spoiler selects y and wins, whatever the move of Duplicator is. ∎

Lemma 4.3. *Let G be a graph of order n, v be a vertex of G with $\sigma_G(v) = s \ge 2$, and $G' = G \oplus_l v$ for an arbitrary $l \ge 1$. Then*

$$s+1 \le W(G,G') \le D_1(G,G') \le s+1+\frac{n+1}{s+1} \le \begin{cases} (n+5)/2 & \text{for } 2 \le s \le (n-1)/2, \\ s+3-1/(n/2+1) & \text{for } s \ge n/2. \end{cases}$$

Proof. The lower bound is given by the following strategy for Duplicator in $\textsc{Ehr}_r^s(G, G')$. Whenever Spoiler selects a vertex outside $[v]$ in either graph, Duplicator selects its copy in the other graph. If Spoiler selects an unoccupied vertex similar to v, then Duplicator selects an arbitrary unoccupied vertex similar to v in the other graph. Clearly, this strategy preserves the isomorphism arbitrarily long, that is, is winning for every r.

The upper bound for $D_1(G, G')$ is ensured by the following Spoiler's strategy winning in the 1-alternation $\textsc{Ehr}_r(G, G')$ for $r = \lfloor s+1+\frac{n+1}{s+1} \rfloor$. In the first round Spoiler selects a vertex in $[v]_{G'}$. Suppose that Duplicator replies with a vertex in $[u]_G$.

Case 1: $|[u]_G| \le s$. Spoiler continues to select vertices in $[v]_{G'}$. In the $(s+1)$-th round at latest, Duplicator selects a vertex outside $[u]_G$. Spoiler wins in the next move by Lemma 4.2, having made at most $s+2$ moves and one alternation.

Case 2: $|[u]_G| \ge s+1$. Spoiler selects one vertex in each similarity class of G' containing at least $s+1$ vertices. Besides $[v]_{G'}$, there can be at most $\frac{n-s}{s+1}$ such classes. At latest in the $\lfloor \frac{n-s}{s+1} + 1 \rfloor$-th round Duplicator selects either another vertex in a class with an already selected vertex (then Spoiler wins in one extra move by Lemma 4.2) or a vertex in $[w]_G$ with $|[w]_G| \le s$. In the latter case Spoiler selects s

more vertices in the corresponding class of G'. Duplicator is forced to move outside $[w]_G$ and loses in the next move by Lemma 4.2. Altogether there are made at most $\lfloor \frac{n-s}{s+1} + 1 \rfloor + s + 1 \le s + 1 + \frac{n+1}{s+1}$ moves.

If $s \ge n/2$, the last inequality of the lemma is straightforward and, if $2 \le s \le (n-1)/2$, it follows from the fact that the function $f(x) = x + \frac{n+1}{x}$ attains its maximum on the segment $[3, (n+1)/2]$ at $x = (n+1)/2$. ∎

Using Lemma 4.3, Lemma 3.14 can now be refined.

Lemma 4.4. *If G and G' are non-isomorphic graphs of orders $n \le n'$, then*

$$D_1(G, G') \le (n+5)/2$$

unless

$$\sigma(G) \ge n/2 \text{ and } G' = G \oplus_{n'-n} v \text{ for some } v \in V(G) \text{ with } \sigma_G(v) = \sigma(G). \tag{12}$$

In the latter case we have

$$\sigma(G) + 1 \le W(G, G') \le D_1(G, G') \le \sigma(G) + 2. \tag{13}$$

Note that the condition (12) determines G' up to isomorphism with two exceptions if n is even. Namely, for $G = K_m \sqcup \overline{K_m}$ and $G = \overline{K_m \sqcup \overline{K_m}}$ there are two ways to extend G to G'.

The gap between the bounds (13) can be completely closed.

Lemma 4.5. *Let G and G' be graphs of orders $n \le n'$. Assume the condition (12). Then $W(G, G') = \sigma(G) + 1$ for all such G and G', $D_0(G, G') = \sigma(G) + 1$ if $[v]_G$ is an inclusion-maximal homogeneous set, and $D(G, G') = \sigma(G) + 2$ if $[v]_G$ is not.*

Proof. For simplicity we assume that $[v]_G$ is an independent set. Otherwise we can switch to $\overline{G}$ and $\overline{G'}$ by Proposition 2.4. Denote $s = \sigma(G) = |[v]_G|$.

Case 1: $[v]_G$ is maximal independent. We show the upper bound $D_0(G, G') \le s + 1$ by describing Spoiler's strategy winning the 0-alternation $\textsc{Ehr}_{s+1}(G, G')$. Spoiler selects $s + 1$ vertices in $[v]_{G'}$. Duplicator is forced to select at least one vertex $u_1 \in [v]_G$ and at least one vertex $u_2 \notin [v]_G$. Since $[v]_G$ is a maximal independent set, u_1 and u_2 are adjacent and this is Spoiler's win.

Case 2: $[v]_G$ is not maximal. We first show the bound $W(G, G') \le s + 1$ by describing Spoiler's strategy winning $\textsc{Ehr}^{s+1}_{s+2}(G, G')$. As in the preceding case, Spoiler selects $s + 1$ vertices in $[v]_{G'}$ and there are $u_1 \in [v]_G$ and $u_2 \notin [v]_G$ selected in response by Duplicator. Assume that u_1 and u_2 are not adjacent for otherwise Duplicator loses immediately. Since u_1 and u_2 are not similar, there is $u \in V(G) \setminus \{u_1, u_2\}$ adjacent to exactly one of u_1 and u_2. It follows that $u \notin [v]_G$. Note that u could not be selected by Duplicator in the first $s + 1$ rounds without immediately losing. Therefore, Duplicator has selected in $[v]_G$ at least two vertices, say, u_0 and u_1. In the $(s + 2)$-th round Spoiler removes the pebble from u_0 to u and wins because the counterparts of u_1 and u_2 in G' are similar and hence equally adjacent or non-adjacent to any counterpart of u.

We now show the bound $D(G,G') > s+1$ by describing Duplicator's strategy winning $\textsc{Ehr}_{s+1}(G,G')$. Whenever Spoiler selects a vertex of either graph, Duplicator selects its copy in the other graph, with the convention that the copy of a vertex in $[v]_{G'}$ is an arbitrary unselected vertex in $[v]_G$. This is impossible in the only case when Spoiler selects $s+1$ vertices all in $[v]_{G'}$. Then Duplicator, in addition to s vertices of $[v]_G$, selects one more vertex extending $[v]_G$ to a larger independent set. ∎

Proof of Theorem 4.1.
1-2. We start with proving the lower bound $D(G) \ge W(G) \ge \sigma(G)+1$. The case that $\sigma(G) = 1$ is trivial because $W(G) \ge 2$ for all G. Given G with $\sigma(G) \ge 2$, fix a vertex v with $\sigma_G(v) = \sigma(G)$ and let $G' = G \oplus_1 v$. By Lemma 4.3 we obtain $W(G) \ge W(G,G') \ge \sigma(G)+1$.

Let us now prove the upper bounds. By Proposition 2.3, we have to show that

$$W(G,G') \le \max\{\sigma(G)+1, (n+5)/2\} \tag{14}$$

and

$$D_1(G,G') \le \max\{\sigma(G)+2, (n+5)/2\} \tag{15}$$

for every G' non-isomorphic to G. Denote the order of G' by n'. If $n' \ge n$, then (15) follows directly from Lemma 4.4 and (14) follows from Lemmas 4.4 and 4.5.

If $n' < n$, we use Lemma 4.4 with G and G' interchanged (we will refer to (12) and (13) with G and G' as well as n and n' interchanged). If the condition (12) is false, then $D_1(G,G') \le (n'+5)/2 \le (n+4)/2$. If (12) is true, then $\sigma(G) > \sigma(G')$ and by (13) we have $D_1(G,G') \le \sigma(G')+2 \le \sigma(G)+1$. In either case

$$W(G,G') \le D_1(G,G') \le \max\{\sigma(G)+1, (n+4)/2\} \tag{16}$$

and we again come to (14) and (15).
3. Suppose that

$$\sigma(G) \ge (n+2)/2. \tag{17}$$

Let v be a representative of the largest similarity class in G. We already know that in any case $\sigma(G)+1 \le D(G) \le D_1(G) \le \sigma(G)+2$. If the homogeneous set $[v]_G$ is not inclusion-maximal, let $G' = G \oplus_1 v$. By Lemma 4.5 we have $D(G) \ge D(G,G') = \sigma(G)+2$ and hence $D(G) = D_1(G) = \sigma(G)+2$.

The case that $[v]_G$ is inclusion-maximal is a bit more complicated. We have to show that

$$D_1(G,G') \le \sigma(G)+1 \tag{18}$$

for every G' non-isomorphic to G. If $n' < n$, we have already established (16) which implies (18) on the account of (17). Assume that $n' \ge n$.

If the condition (12) is true, we have (18) because $D_0(G,G') = \sigma(G)+1$ by Lemma 4.5. If (12) is false, by Lemma 4.4 we have

$$D_1(G,G') \le (n+5)/2. \tag{19}$$

This proves (18) if $\sigma(G) \geq (n+3)/2$. In the only remaining case of $\sigma(G) = (n+2)/2$, note that n is even and hence (19) is equivalent to $D_1(G, G') \leq (n+4)/2 = \sigma(G)+1$.

4. By Item 1 of the theorem, $W(G) \leq (n+5)/2$ iff $\sigma(G) \leq (n+3)/2$ and, if $\sigma(G) > (n+3)/2$, then $W(G) = \sigma(G) + 1$. It remains to notice that, given G, its vertex set is easy to split into the similarity classes and therefore the similarity index $\sigma(G)$ is efficiently computable.

With the logical depth $D(G)$ we have to be a bit more careful. By Item 2 of the theorem, $D(G) \leq (n+5)/2$ whenever $\sigma(G) \leq (n+1)/2$. By Item 3, if $\sigma(G) > (n+1)/2$, we can determine $D(G)$ exactly and hence decide whether or not $D(G) \leq (n+5)/2$ (which still may happen if $\sigma(G) \leq (n+3)/2$). We can do it efficiently because the maximality of a homogeneous set with respect to inclusion is easy to check. ∎

Remark 4.6. Throughout the paper we tacitly assume that we consider only finite graphs. However, in a natural strengthening of Definition 2.2 we could require that a defining sentence for a graph G should distinguish G from every non-isomorphic G' of *any* cardinality, both finite and infinite. Denote the logical depth of G modified in this way by $D^*(G)$. We obviously have $D(G) \leq D^*(G)$ but it does not seem so obvious whether or not $D^*(G)$ can be strictly greater. Anyway the upper bounds stated in Theorems 4.1 and 6.2 for $D(G)$ hold true for $D^*(G)$ with virtually the same proof. In particular, Lemmas 3.14, 4.3, 4.4, and 4.5 remain true if G is finite and G' is infinite under an appropriate modification of Definition 3.13. Note also that the number $D(G, G')$ for non-isomorphic G and G' may not exist if both G and G' are infinite (see, e.g., [14, Theorem 3.3.2]) but always exists if at least G is finite. Under the latter condition Propositions 2.3 and 2.5 also remain true, the former with $D^*(G)$ in place of $D(G)$.

5 Distinguishing graphs with no quantifier alternation

Theorem 3.1 is proved in a rather strong form: The class of distinguishing sentences is restricted to those with alternation number 1. We now further restrict the alternation number to the smallest possible value of 0. Note that, if graphs are distinguished by a 0-alternation formula Φ, they are distinguished as well by a subformula of Φ containing either only existential or only universal quantifiers. Somewhat surprisingly, even under this restriction on the class of distinguishing sentences Theorem 3.1 holds true just with a little bit worse bound.

In terms of the Ehrenfeucht game, we restrict the ability of Spoiler to alternate between graphs during play (see Proposition 2.5.3). He only keeps the freedom to choose a graph in which he will move all the time.

Theorem 5.1. *If G and G' are non-isomorphic graphs of the same order n, then $D_0(G, G') \leq (n+5)/2$.*

Proof. We will describe a strategy for Spoiler winning the 0-alternation game $\textsc{Ehr}_{\lfloor (n+5)/2 \rfloor}(G, G')$. Given a set of vertices X in a graph H and a partial isomorphism $\phi : X \to X'$ to another graph, we will use the notions introduced in Section 3.1: the partitions $\mathcal{C}(X)$ and $\mathcal{D}(X)$, the set $Y(X)$, and the ϕ-similarity relation $\equiv_\phi$. We set the following notation:

$$\begin{aligned} s(X) &= |X|, \\ t(X) &= |\mathcal{C}(X)|, \\ c(X) &= \max\{\, |C| : C \in \mathcal{C}(X)\}\,, \\ d(X) &= \max\{\, |D| : D \in \mathcal{D}(X)\}\,. \end{aligned}$$

For brevity, we will not indicate the dependence on X, writing merely Y, s, t, c, and d.

At the start of the game Spoiler chooses $H \in \{G, G'\}$ and a set $X \subset V(H)$ with $t \geq s + 1$ according to the following criteria.

Criterion 1: First of all, he maximizes s.

Criterion 2: Then, if there is still some choice, he minimizes c.

Criterion 3: Finally, he minimizes d.

Let us assume $H = G$. As $s + t \leq n$, we have $s \leq (n-1)/2$.

Spoiler selects all vertices in X in any order. Denote the set of vertices of G' selected in response by Duplicator by X'. Let $t' = t(X')$ and $c' = c(X')$. Assume that Duplicator has not lost up to now, that is, has managed to maintain the partial isomorphism $\phi : X \to X'$. Let $C_1, \ldots, C_t$ (resp. $C'_1, \ldots, C'_{t'}$) be all classes in $\mathcal{C}(X)$ (resp. $\mathcal{C}(X')$).

If there is a class C_i without ϕ-similar counterpart in $\mathcal{C}(X')$, Spoiler wins in one move by selecting a vertex in C_i, having made $s + 1 \leq (n+1)/2$ moves at total. We therefore suppose that $t' \geq t$ and, for every $i \leq t$, the classes C_i and C'_i are ϕ-similar. Thus, $t' \geq s + 1 = s' + 1$ and, by Criterion 2 of the choice of (H, X), we conclude that there is C'_m such that $|C_i| \leq |C'_m|$ for all i.

Define $C_i = \emptyset$ for all $t < i \leq t'$. Suppose first that for some $i \leq t'$ we have $|C_i| \neq |C'_i|$ (this is necessary so if $t' > t$). As G and G' have the same order, there must be an $i \leq t'$ with $|C_i| > |C'_i|$. Spoiler wins by selecting $|C'_i| + 1$ vertices inside C_i. Observe that $|C'_i| < |C_i| \leq |C'_m|$ and that

$$2\,|C'_i| + 1 \leq |C'_i| + |C'_m| \leq (n - s') - (t' - 2) \leq n - 2s + 1.$$

The total number of Spoiler's moves is therefore at most $s + |C'_i| + 1 \leq n/2 + 1$, within the required bound.

Suppose from now on, that $t' = t$ and $|C_i| = |C'_i|$ for all $i \leq t$. In particular,

$$c = c'. \tag{20}$$

Without loss of generality, assume that $|C_i| = |C'_i| = 1$ precisely for $i \leq q$. Note that $Y = \bigcup_{i=1}^{q} C_i$ and $Y' = \bigcup_{i=1}^{q} C'_i$, where $Y' = Y(X')$. Similarly to the proof of Lemma

3.14, we extend ϕ to $\phi^* : X \cup Y \to X' \cup Y'$ by the condition that ϕ^* maps each C_i with $i \le q$ onto C'_i. Similarly to Claim 3.14.2, if ϕ^* is not an isomorphism from $G[X \cup Y]$ to $G'[X' \cup Y']$, then Spoiler wins by selecting 2 vertices in Y, having made altogether $s + 2 \le (n+3)/2$ moves. In the sequel we therefore suppose that ϕ^* is a partial isomorphism from G to G'. We will make use of the following observation, provable similarly to Claim 3.14.3. Let $Z = V(G) \setminus (X \cup Y)$.

Claim 5.1.1. From now on, whenever Spoiler selects a vertex $v \in Z$, Duplicator responds with a ϕ^*-similar vertex or otherwise loses in the next round at latest with no alternation. □

Let $D_1, \ldots, D_p$ (resp. $D'_1, \ldots, D'_{p'}$) be all classes in $\mathcal{D}(X)$ (resp. $\mathcal{D}(X')$). We now claim that every class D_i has a ϕ^*-similar counterpart in $\mathcal{D}(X')$ or otherwise Spoiler wins in at most 2 next moves with no alternation, having made altogether at most $s + 2 \le (n+3)/2$ moves. Indeed, if a D_i has no ϕ^*-similar counterpart, Spoiler selects a vertex in the D_i and wins either immediately or in the next move by Claim 5.1.1. We hence will assume that $p' \ge p$ and, for all $i \le p$, the classes D_i and D'_i are ϕ^*-similar. If $p' > p$, define $D_i = \emptyset$ for all $p < i \le p'$.

We now show that each class in $\mathcal{D}(X)$ or $\mathcal{D}(X')$ consists of pairwise similar vertices as defined in Definition 3.2. Suppose, to the contrary, that vertices u and v lie in the same D_i and some w is connected to one of u, v but not to the other. By the definition of D_i such w must lie in Z; but then moving w to X we increase t at least by one. Indeed, if w belongs to the same D_i, the class $C^1 \in \mathcal{C}(X)$ including D_i splits up into at least two subclasses, containing u and v respectively, while no class in $\mathcal{C}(X)$ disappears. If w belongs to another D_j, the class C^1 splits up as well, while the class $C^2 \in \mathcal{C}(X)$ including D_j still stays because it has at least two elements. Since the relation $t \ge s + 1$ is preserved, we get a contradiction with Criterion 1 in the choice of (H, X). The same argument applies for $\mathcal{D}(X')$.

It follows that, for any distinct $i, j \le p$, each of $G[D_i]$, $G'[D'_i]$, $G[D_i, D_j]$ and $G'[D'_i, D'_j]$ is either complete or empty. The same is true about every $G[D_i, \{v\}]$ and $G'[D'_i, \{v'\}]$ for $v \in X \cup Y$ and $v' \in X' \cup Y'$. We now claim that, for every $i \le p$, $j \le p$ such that $j \ne i$, and $v \in X \cup Y$,

1. $G[D_i]$ with at least 2 vertices is complete iff $G'[D'_i]$ is,
2. $G[D_i, D_j]$ is complete iff $G'[D'_i, D'_j]$ is, and
3. $G[D_i, \{v\}]$ is complete iff $G'[D'_i, \{\phi^*(v)\}]$ is

or otherwise Spoiler wins in at most 3 next moves with no alternation, having made altogether $s + 3 \le (n+5)/2$ moves. For example, consider the case that $G[D_i]$ has at least 2 vertices and is complete but $G'[D'_i]$ is empty. Then Spoiler selects two vertices in D_i. If both Duplicator's responses are in D'_i, he loses immediately. Otherwise Duplicator responds at least once with a vertex which is not ϕ^*-similar. Then Spoiler wins in the next move according to Claim 5.1.1.

We therefore suppose that the above three conditions are obeyed for all $i, j \le p$ and $v \in X \cup Y$. It follows that, if $|D_i| = |D'_i|$ for all $i \le p'$ and, in particular, $p' = p$, then G and G' should be isomorphic. Since this is not so, there is $l \le p'$ such that $|D_l| \neq |D'_l|$. As G and G' have the same order, we can assume that

$$|D_l| > |D'_l|. \tag{21}$$

Note that $p' > t - q$ for else the D'-classes are identical to the C'-classes, which contradicts (21). Thus

$$p' \ge s + 2 - q. \tag{22}$$

It follows from (20) and Criterion 3 of the choice of (H, X) that there exists $k \le p'$ such that $|D_i| \le |D'_k|$ for all i. We have $|D'_l| < |D_l| \le |D'_k|$, so

$$2\,|D'_l| + 1 \le |D'_l| + |D'_k| \le (n - s - q) - (p' - 2) \le n - 2s,$$

where the latter inequality follows from (22).

Now, Spoiler selects $|D'_l| + 1$ vertices inside D_l. Duplicator cannot reply to this with all moves in D'_l and hence replies at least once with a vertex which is not ϕ^*-similar. According to Claim 5.1.1, Spoiler wins either immediately or in the next round. The total number of moves is at most

$$s + |D'_l| + 1 + 1 \le s + \frac{n - 2s - 1}{2} + 2 = \frac{n + 3}{2},$$

as required. ∎

6 Defining graphs of bounded degree

The *degree* of a vertex v in a graph, denoted by $\deg(v)$, is the number of edges incident to v. The *maximum degree* of a graph G is defined by $\Delta(G) = \max_{v \in V(G)} \deg(v)$. The *distance* between vertices v and u in a graph, $\mathrm{dist}(v, u)$, is the smallest number of edges in a path from v to u. If $U \subseteq V(G)$, then $\mathrm{dist}(v, U) = \min_{u \in U} \mathrm{dist}(v, u)$. Recall that the similarity index $\sigma_G(v)$ of a vertex v is defined in Definition 3.2.

Lemma 6.1. *If v is a non-isolated vertex of a graph G, then*

$$\sigma_G(v) \le \Delta(G) + 1. \tag{23}$$

Proof. By Lemma 3.3, the similarity class $[v]_G$ is either a clique or an independent set. If it is a clique, then the bound (23) is clear. Otherwise there must exist a vertex $u \notin [v]_G$ adjacent to v. As u is adjacent to every vertex in $[v]_G$, we have $\sigma_G(v) \le \deg(u) \le \Delta(G)$ in this case. ∎

Theorem 6.2. *Let $d \ge 2$. If G is a graph of order n with $\Delta(G) = d$ that has no isolated vertex and no isolated edge, then*

$$D_1(G) < c_d\, n + d^2 + d + 4$$

for a constant $c_d = \frac{1}{2} - \frac{1}{4} d^{-2d-5}$.

The constant c_d as stated in the theorem is far from being best possible. We do not try to optimize it; Our goal is more moderate, just to show the existence of a c_d strictly less than 1/2. In the case of $d = 2$, which is simple and included just for uniformity, an optimal bound is $D_1(G) \le n/3 + O(1)$. Without the assumption that G has no isolated vertex and edge, the theorem holds true for no $c_d < 1/2$: A counterexample is provided by G which is the disjoint union of isolated edges and the complete graph on $d + 1$ vertices.

Proof. According to Proposition 2.3, we have to estimate $D_1(G, G')$ for any G' non-isomorphic to G. Referring to Proposition 2.5.3, we design a strategy for Spoiler winning the 1-alternation Ehrenfeucht game on G and G' in at most $c_d\, n + d^2 + d + 4$ rounds. Clearly, we may assume that $\Delta(G') \le d$ for otherwise Spoiler wins in at most $d + 2$ moves by selecting a star $K_{1,d+1}$ in G'. Another assumption we made on G' is the absence of isolated vertices and edges in this graph. It is clear that otherwise Spoiler wins in 3 moves with 1 alternation.

A *component* of a graph is a maximal connected induced subgraph. We call a component *small* if it has at most $d^2 + 1$ vertices. Denote the union of all small components of G by K and the remaining part of the graph by H. The similar notation K' and H' will be used for G'. Since $G = K \sqcup H$ and $G' = K' \sqcup H'$ are non-isomorphic, we have $K \not\cong K'$ or $H \not\cong H'$. Consider the former case first.

Spoiler enforces play on non-isomorphic K and K' as explained below. He starts in K if this graph has at least as many components as K' has and starts in K' otherwise. Without loss of generality assume the former.

Spoiler selects one vertex in each component of K. This takes at most $n/3$ moves as every component of G has at least 3 vertices. He keeps doing so until one of the following happens.

1. Duplicator moves outside K'. Denote the vertex of H' that he selects in this move by v'. Then Spoiler switches to G' and wins in at most d^2+1 extra moves by selecting a connected induced subgraph on $d^2 + 2$ vertices including v'.

2. Duplicator makes two moves in the same component of K'. Denote the two selected vertices by v' and u'. Then Spoiler switches to G' and wins in at most $d^2 - 1$ extra moves by selecting a path from v' to u'.

3. While Spoiler selects a vertex in a component C of K, Duplicator responds with a vertex in a component C' of K' such that $C' \not\cong C$. Then Spoiler wins in at most d^2 extra moves by selecting all vertices of C if $|C| \ge |C'|$ or all vertices of C' otherwise.

It is clear that one of the three situations must happen sooner or later. Thus, Spoiler wins in at most $n/3 + d^2 + 1$ moves with at most 1 alternation between G and G'.

Consider now the case that $H \not\cong H'$. Spoiler enforces play on H and H' moving all the time in these graphs. Once in response to Spoiler's move $v \in V(H) \cup V(H')$ Duplicator moves in K or K', Spoiler wins in at most d^2+1 extra moves by selecting a connected induced subgraph on d^2+2 vertices including v. We will therefore assume that Duplicator agrees to play on H and H'.

Denote the order of H by m. We split our description of Spoiler's strategy into three phases.

PHASE 1

Spoiler will make moves in pairs. Let $i \geq 1$. Denote the vertices selected by him in the $(2i-1)$-th and $2i$-th rounds of Phase 2 by x_i and y_i respectively. Suppose that Spoiler has already made $2(i-1)$ moves and selected a set $X_{i-1} = \{x_1, y_1, \ldots, x_{i-1}, y_{i-1}\} \subset V(H)$. Let us explain how x_i and y_i are now selected.

If there is a vertex $x \in V(H)$ such that

- $\mathrm{dist}(x, X_{i-1}) \geq 5$ and
- for all y with $\mathrm{dist}(x, y) \leq 2$ we have $\deg(y) \leq \deg(x)$,

then Spoiler selects this x for x_i.

Claim 6.2.1. Suppose that $x_i = x$ does exist. Then there are vertices u, y, v such that $\{x, u\}, \{u, y\}, \{y, v\} \in E(H)$ while $\{x, y\}, \{x, v\} \notin E(H)$.

Proof of Claim. Let C be the component of H containing x. It should contain a vertex v with $\mathrm{dist}(x, v) = 3$ for else every vertex of C would be at distance at most 2 from x and hence C would have at most $1 + d + d(d-1) = d^2 + 1$ vertices. Let (x, u, y, v) be an arbitrary path from x to v. The vertices u, y, v are as desired. □

If $x_i = x$ is selected, Spoiler chooses some u, y, v as in the claim and takes the y for y_i.

If no such x exists, Phase 1 ends. Suppose that this phase lasts $2r$ rounds. Given $X \subset V(H)$, let $\mathcal{C}(X)$ denote the partition of $V(H) \setminus X$ into $\equiv_X$-equivalence classes (see Definition 3.4).

Claim 6.2.2. $|\mathcal{C}(X_i)| \geq |\mathcal{C}(X_{i-1})| + 3$ if $i < r$ and $|\mathcal{C}(X_r)| \geq |\mathcal{C}(X_{r-1})| + 2$.

Proof of Claim. We will show that, if we extend X_{i-1} to X_i, one of the classes in $\mathcal{C}(X_{i-1})$ splits up into at least 4 parts if $i < r$ and into at least 3 parts if $i = r$.

By the choice of u, $y = y_i$, and v, we have $\mathrm{dist}(x_i, u) = 1$ and $\mathrm{dist}(x_i, v) = 3$. Since $\mathrm{dist}(x_i, X_{i-1}) \geq 5$, neither u and v is in X_{i-1}. Note that u is adjacent to both x_i and y_i, while v is adjacent to y_i but not to x_i.

Since $\deg(x_i) \geq \deg(y_i)$ and $\Gamma(y_i) \setminus \Gamma(x_i)$ contains at least one vertex v, there must be a vertex w adjacent to x_i but not to y_i. Like u and v, we have $w \notin X_{i-1}$.

Thus, u, v, and w belong to pairwise distinct classes of $\mathcal{C}(X_i)$. If we assume that $i < r$, we are able to find a vertex in yet another class. Indeed, consider $z = x_{i+1}$. Since $\mathrm{dist}(z, X_i) \geq 5$, this vertex is adjacent neither to x_i nor to y_i.

On the other hand, every of u, v, w, and z is at distance at least 2 from X_{i-1}. Therefore all of them are in the same class of $\mathcal{C}(X_{i-1})$. □

PHASE 2

As long as possible, Spoiler extends $X = X_r$ by one vertex so that $|\mathcal{C}(X)|$ increases at least by 1. Phase 2 ends as soon as Spoiler arrives at a $\mathcal{C}$-maximal set $X \subset V(H)$ (in the sense of Definition 3.6 with respect to H).

Suppose that Phase 2 lasts h rounds. At the end of this phase we therefore have

$$|X| = 2r + h$$

and

$$|\mathcal{C}(X)| \geq 1 + 3(r-1) + 2 + h = 3r + h.$$

It follows that $|\mathcal{C}(X)| \geq |X| + r$ and hence $m \geq |X| + |\mathcal{C}(X)| \geq 2|X| + r$. We conclude that

$$|X| \leq \frac{m-r}{2}.$$

Phase 3

Spoiler now plays precisely as in Phase 2 of the strategy designed in the proof of Lemma 3.14. By Lemma 3.16, with Lemma 6.1 taken into account, Spoiler wins the game on H and H' making totally at most

$$|X| + d + 3 \leq \frac{m-r}{2} + d + 3 \tag{24}$$

moves. It therefore remains to show that r is linearly related to m, that is, Phase 1 cannot be too short.

Claim 6.2.3. Let $V_k = \{ x \in V(H) : \mathrm{dist}(x, X_r) \geq 2k+3 \}$. Then $V_{d+1} = \emptyset$.

Proof of Claim. Provided $V_i \neq \emptyset$, denote $d_i = \max \{ \deg(x) : x \in V_i \}$ and consider an arbitrary $z \in V_i$ with $\deg(z_i) = d_i$. Since Phase 1 has ended in $2r$ rounds and no further x_{r+1} can be chosen any more, z does not satisfy the criteria for x_{r+1}. This means that there is y such that $\mathrm{dist}(y, z) \leq 2$ and $\deg(y) > \deg(z)$. The latter implies that $y \notin V_i$, i.e., $\mathrm{dist}(y, X_r) \leq 2i+2$. It follows that $\mathrm{dist}(z, X_r) \leq 2i+4$ and therefore $z \notin V_{i+1}$. We conclude that V_{i+1} contains no vertex of degree d_i and hence either $V_{i+1} = \emptyset$ or $d_{i+1} < d_i$. Since the chain $d_1 > d_2 > d_3 > \ldots$ can have length at most d, we must have $V_i = \emptyset$ no later than for $i = d+1$. $\square$

Thus, $|V(H) \setminus V_{d+1}| = m$. By the definition of V_{d+1} we have

$$|V(H) \setminus V_{d+1}| \leq |X_r|(1 + d + d(d-1) + d(d-1)^2 + \ldots + d(d-1)^{2d+3}) < |X_r| d^{2d+5}$$

and hence $r > m/(2d^{2d+5})$. Substituted in (24), this shows that Spoiler wins in less than

$$\left(\frac{1}{2} - \frac{1}{4d^{2d+5}} \right) m + d + 3 \leq c_d\, n + d + 3$$

moves. Recalling that Duplicator can at some time deviate from playing within H and H', we have to increase this bound by $d^2 + 1$. ∎

Acknowledgments

The third author is grateful to Georg Gottlob and members of his department at the Vienna University of Technology for their hospitality during this author's visit in the summer 2002. We thank Eric Rosen for useful references. We also thank Daniel Zeitlin, Roman Gershman, and anonymous referees for critical expository comments that led to a substantial reorganization of the original version of the paper.

References

[1] J.-Y. Cai, M. Fürer, N. Immerman. An optimal lower bound on the number of variables for graph identification. *Combinatorica* 12(4):389–410 (1992).

[2] A. Dawar, S. Lindell, S. Weinstein, Infinitary logic and inductive definability over finite structures. *Information and Computation* 119:160–175 (1995).

[3] H.-D. Ebbinghaus, J. Flum. *Finite model theory.* Springer Verlag, 2nd rev. ed. (1999).

[4] E. Grädel. Finite model theory and descriptive complexity. In: *Finite Model Theory and Its Applications.* Texts in Theoretical Computer Science, an EATCS Series. Springer (2005).

[5] M. Grohe. Isomorphism testing for embeddable graphs through definability. In: *Proc. of the 32nd ACM Ann. Symp. on Theory of Computing* 63–72 (2000).

[6] M. Grohe, J. Marino. Definability and descriptive complexity on databases of bounded tree-width. In: *Proc. of the 7th International Conference on Database Theory*, Lecture Notes in Computer Science 1540, Springer-Verlag, 70–82 (1999).

[7] N. Immerman. *Descriptive complexity.* Springer-Verlag (1999).

[8] J.-H. Kim, O. Pikhurko, J. Spencer, O. Verbitsky. How complex are random graphs in first order logic? *Random Structures and Algorithms* 26:119–145 (2005).

[9] L. Libkin. *Elements of finite model theory.* Berlin: Springer (2004).

[10] E. Pezzoli. Computational complexity of Ehrenfeucht-Fraïssé games on finite structures. In: *Proc. of the 12th Conference on Computer Science Logic 1998.* G. Gottlob, E. Grandjean, K. Seyr Eds. Lecture Notes in Computer Science 1584, Springer-Verlag, 159–170 (1999).

[11] O. Pikhurko, J. Spencer, O. Verbitsky. Succinct definitions in first order graph theory. *Annals of Pure and Applied Logic*, to appear. Published online 31 May 2005 at *http://www.sciencedirect.com.*

[12] O. Pikhurko, H. Veith, O. Verbitsky. First order definability of graphs: upper bounds on quantifier rank. E-print: *http://arxiv.org/abs/math.CO/0311041* (2003).

[13] O. Pikhurko, O. Verbitsky. Descriptive complexity of finite structures: saving the quantifier rank. *Journal of Symbolic Logic* 70(2):419–450 (2005).

[14] J. Spencer. *The strange logic of random graphs.* Springer Verlag (2001).

[15] O. Verbitsky. The first order definability of graphs with separators via the Ehrenfeucht game. *Theoretical Computer Science* 343(1-2):158–176 (2005).