Discrete Mathematics and Theoretical Computer Science DMTCS vol. **(subm.)**, by the authors, 0–1

Decomposable graphs and definitions with no quantifier alternation

Oleg Pikhurko[1] and Joel Spencer[2] and Oleg Verbitsky[3†]

[1] *Department of Math Sciences, Carnegie Mellon University, Pittsburgh, PA 15213; pikhurko@andrew.cmu.edu*
[2] *Courant Institute, New York University, New York, NY 10012; spencer@cs.nyu.edu*
[3] *Institut für Informatik, Humboldt Universität Berlin, Berlin, D-10099; verbitsky@informatik.hu-berlin.de*

received 4 May 2005, *revised* , *accepted* .

Let $D(G)$ be the minimum quantifier depth of a first order sentence Φ that defines a graph G up to isomorphism in terms of the adjacency and the equality relations. Let $D_0(G)$ be a variant of $D(G)$ where we do not allow quantifier alternations in Φ. Using large graphs decomposable in complement-connected components by a short sequence of serial and parallel decompositions, we show examples of G on n vertices with $D_0(G) \le 2\log^* n + O(1)$. On the other hand, we prove a lower bound $D_0(G) \ge \log^* n - \log^* \log^* n - O(1)$ for all G. Here $\log^* n$ is equal to the minimum number of iterations of the binary logarithm needed to down n to 1.

Keywords: descriptive complexity of graphs, first order logic, Ehrenfeucht game on graphs, graph decompositions

1 Introduction

Given a finite graph G, how succinctly can we describe it using first order logic and the laconic language consisting of the adjacency and the equality relations? A first order sentence Φ *defines* G if Φ is true precisely on graphs isomorphic to G. All natural succinctness measures of Φ are of interest: the *length* $L(\Phi)$ (a standard encoding of Φ is supposed), the *quantifier depth* $D(\Phi)$ which is the maximum number of nested quantifiers in Φ, and the *width* $W(\Phi)$ which is the number of variables used in Φ (different occurrences of the same variable are not counted). All the three characteristics inherently arise in the analysis of the computational problem of checking if a Φ is true on a given graph [3]. They give us a small hierarchy of descriptive complexity measures for graphs: $L(G)$ (resp. $D(G)$, $W(G)$) is the minimum $L(\Phi)$ (resp. $D(\Phi)$, $W(\Phi)$) of a Φ defining G. These graph invariants will be referred to as the *logical width, depth, and length* of G. We have $W(G) \le D(G) \le L(G)$. The former number is of relevance for graph isomorphism testing, see [2]. $W(G)$ and $D(G)$ admit a purely combinatorial characterization in terms of the Ehrenfeucht game, see [2, 8].

We here address the logical depth of a graph. We focus on the following general question: How do restrictions on logic affect the descriptive complexity of a graph? Call a first order sentence Φ to be

† Supported by an Alexander von Humboldt fellowship.

a*-alternation* if it contains negations only in front of relation symbols and every sequence of nested quantifiers in Φ has at most a quantifier alternations. Let $D_a(G)$ denote a variant of $D(G)$ for a-alternation defining sentences, so $D(G) \leq D_{a+1}(G) \leq D_a(G)$. The logic of 0-alternation sentences is most restrictive and provably weaker than the unbounded first order logic. Whereas the problem of deciding if a first order sentence is satisfiable by some graph is unsolvable, it becomes solvable if restricted to 0-alternation sentences (the latter due to Ramsey's logical work [7] founding the combinatorial Ramsey theory).

It is not hard to observe that $D_0(G) \leq n+1$ where n denotes the number of vertices in G. This bound is in general best possible as $D(K_n) = n+1$. Nevertheless, it admits a non-obvious improvement under a rather small restriction on the automorphism group of G. If the latter does not contain any transposition of two vertices, then $D_1(G) \leq (n+5)/2$ [6]. No sublinear improvement is possible because of the sequence of asymmetric graphs with $W(G) = \Omega(n)$ constructed in [2]. In [4] we prove that $D(G) = \log_2 n - \Theta(\log_2 n \log_2 n)$ and $D_0(G) \leq (2 + o(1)) \log_2 n$ for almost all G.

After obtaining these worst-case and average-case results, we undertake a "best-case" analysis in [5]. We define the *succinctness function* $q(n) = \min \{D(G) : G \text{ has order } n\}$ and show that its values may be superrecursively small if compared to n: $f(q(n)) \geq n$ for no recursive f. A weaker but still surprising succinctness result is also obtained for the fragment of first order logic with no quantifier alternation. Let $q_0(n) = \min \{D_0(G) : G \text{ has order } n\}$.

Theorem 1 $q_0(n) \leq 2 \log^* n + O(1)$ *for infinitely many* n.

In [5] this theorem is proved by considering G in a certain class of asymmetric trees and estimating $D_0(G)$ in terms of the radius of a tree. We here reprove this result by showing the same definability phenomenon in a different class of graphs. We consider G in a class of graphs with small complement-connected induced subgraphs and estimate $D_0(G)$ in terms of the number of the *serial* and *parallel decompositions* [1] decomposing G in the complement-connected components.

We also present a new result complementing Theorem 1.

Theorem 2 $q_0(n) \geq \log^* n - \log^* \log^* n - O(1)$ *for all* n.

As a consequence, $q_0(n) \leq f(q(n))$ for no recursive f, which also shows a superrecursive gap between the graph invariants $D(G)$ and $D_0(G)$.

2 Definitions

We use the following notation: $V(G)$ is the vertex set of a graph G; $diam\, G$ is the diameter of G; $\overline{G}$ is the complement of G; $G \sqcup H$ is the disjoint union of graphs G and H; $G \subset H$ means that G is isomorphic to an induced subgraph of H (we will say that G is *embeddable* in H); $G \sqsubset H$ means that G is isomorphic to the union of some of the connected components of H.

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, *cocomponent* of G. Cocomponents have no common vertices and partition $V(G)$.

The *decomposition* of G, denoted by $Dec\, G$, is the set of all connected components of G (this is a set of graphs, not just isomorphism types). Furthermore, given $i \geq 0$, we define the *depth* i *decomposition* of G by $Dec_0\, G = Dec\, G$ and $Dec_{i+1}\, G = \bigcup_{F \in Dec_i\, G} Dec\, \overline{F}$. Note that $P_i = \{V(F) : F \in Dec_i\, G\}$ is a partition of $V(G)$ and that P_{i+1} refines P_i. The *depth* i *environment* of a vertex $v \in V(G)$, denoted by $Env_i(v)$, is the F in $Dec_i\, G$ containing v.

We define the *rank* of a graph G, denoted by $rk\,G$, inductively as follows: (1) If G is complement-connected, then $rk\,G = 0$. (2) If G is connected but not complement-connected, then $rk\,G = rk\,\overline{G}$. (3) If G is disconnected, then $rk\,G = 1 + \max\{rk\,F \,:\, F \in Dec\,G\}$. In other terms, $rk\,G$ is the smallest k such that $P_{k+1} = P_k$ or such that P_k consists of $V(F)$ for all cocomponents F of G.

In the *Ehrenfeucht game* on two disjoint graphs G and H two players, Spoiler and Duplicator, alternatingly select vertices of the graphs, one vertex per move. Spoiler starts and is always free to move in any of G and H; Then Duplicator must choose a vertex in the other graph. Let $x_i \in V(G)$ and $y_i \in V(H)$ denote the vertices selected by the players in the i-th round. Duplicator wins the k-round game if the component-wise correspondence between $x_1, \ldots, x_k$ and $y_1, \ldots, y_k$ is a partial isomorphism from G to H; Otherwise the winner is Spoiler. In the *0-alternation game* Spoiler plays all the game in the same graph he selects in the first round.

Assume $G \not\cong H$. Let $D(G, H)$ (resp. $D_0(G, H)$) denote the minimum $D(\Phi)$ over (resp. 0-alternation) first order sentences Φ that are true on one of the graphs and false on the other. The Ehrenfeucht theorem relates $D(G, H)$ and the length of the Ehrenfeucht game on G and H. We will use the following version of the theorem: $D_0(G, H)$ is equal to the minimum k such that Spoiler has a winning strategy in the k-round 0-alternation Ehrenfeucht game on G and H. It is also useful to know that $D_0(G) = \max\{D_0(G, H) \,:\, H \not\cong G\}$.

We define the tower-function by $\mathit{Tower}\,(0) = 1$ and $\mathit{Tower}\,(i) = 2^{\mathit{Tower}\,(i-1)}$ for each subsequent i.

3 Upper bound: Proof of Theorem 1

Lemma 1 *Consider the Ehrenfeucht game on graphs G and H. Let $x, x' \in V(G)$, $y, y' \in V(H)$ and assume that the pairs x, y and x', y' are selected by the players in the same rounds. Furthermore, assume that $Env_l(x) \neq Env_l(x')$, $Env_l(y) = Env_l(y')$, and $diam\,Env_i(y) \le 2$ for every $i \le l$. Then Spoiler can win in at most $l + 1$ rounds (l rounds if G is connected), playing all the time in H.*

Proof: We proceed by induction on l. The base case is $l = 0$ if G is disconnected and $l = 1$ if G is connected. If y and y' are adjacent in $Env_l(y)$, Duplicator has already lost. Otherwise, Spoiler uses the fact that $diam\,Env_l(y) = 2$ and selects y'' adjacent in $Env_l(y)$ to both y and y'. Duplicator cannot do so with any x'' because x and x' are in different components of G if $l = 0$ or $\overline{G}$ if $l = 1$.

Assume that $l \ge 1$. Let $0 \le m \le l$ be the minimum number such that $x' \notin Env_m(x)$. If $m < l$, Spoiler wins in the next $m + 1 \le l$ moves by induction. If $m = l$, Spoiler uses the same trick as in the base case and enforces Duplicator to make a move x'' outside $Env_{l-1}(x)$. By the induction hypothesis, Spoiler needs l extra moves to win. □

As long as Duplicator avoids meeting the conditions of Lemma 1 (in particular, selects $x' \in Env_l(x)$ whenever Spoiler selects $y' \in Env_l(y)$), we will say that she *bewares of the environment threat.*

Let $rk\,G = k$. We call G *uniform* if $Dec_{k-1}\,G$ contains no complement-connected graph, that is, every cocomponent appears in $Dec_k\,G$ and no earlier. We call G *inclusion-free* if the following two conditions are true for every $i < k$: (1) For any $K \in Dec_i\,G$, $\overline{K}$ contains no isomorphic connected components. (2) If two elements K and M of $Dec_i\,G$ are non-isomorphic, then neither $\overline{K} \sqsubset \overline{M}$ nor $\overline{M} \sqsubset \overline{K}$.

Lemma 2 (Main Lemma) *Let G be a uniform inclusion-free graph. Suppose that every cocomponent of G has exactly c vertices. Then $D_0(G) \le 2\,rk\,G + c + 1$.*

Proof: Let $rk\,G = k$. Fix a graph $H \not\cong G$. We will design a strategy allowing Spoiler to win the 0-alternation Ehrenfeucht game on G and H in at most $2k + c + 1$ moves. Since $D_0(G) = D_0(\overline{G})$, without loss of generality we will assume that G is connected. Since the case of $k = 0$ is trivial, we will also assume that $k \geq 1$.

Case 1: H has a cocomponent C non-embeddable in any cocomponent of G. If C has no more than c vertices, Spoiler selects all C. Otherwise he selects $c + 1$ vertices spanning a complement-connected subgraph in C (it is not hard to show that this is always possible). If Duplicator's response A is within a cocomponent of G, then $C \ncong A$ by the assumption. Otherwise A is not complement-connected and Duplicator loses anyway.

In the sequel we will assume that Duplicator bewares of the environment threat during all game.

Case 2: $G \subset H$ or there are $l \leq k$ and $A \in Dec_l\, G$ properly embeddable in some $B \in Dec_l\, H$, and not Case 1. Spoiler plays in H. If $G \subset H$, set $A = G$, $B = H$, and $l = 0$. Let H_0 be a copy of A in B. At the first move Spoiler selects an arbitrary $y_0 \in V(B) \setminus V(H_0)$. Denote Duplicator's response in G by x_0 and set $G_0 = Env_l(x_0)$. From now on Spoiler plays in H_0. Since we are not in Case 1, B is not a cocomponent of H and hence $diam\, B \leq 2$. Since Duplicator is supposed to beware of the environment threat, from now on she is enforced to play in G_0.

Subcase 2.1: $G_0 \ncong H_0$. Assume that $l < k$ (the case of $l = k$ will be covered by the last phase of the strategy). Since G_0 and H_0 are non-isomorphic copies of elements of $Dec_l\, G$ and G is inclusion-free, Spoiler is able to make his next choice y_1 in some $H_1 \in Dec\, \overline{H_0}$ absent in $Dec\, \overline{G_0}$. Denote Duplicator's response in G_0 by x_1 and set $G_1 = Env_{l+1}(G)$. Note that G_1 and H_1 are non-isomorphic copies of elements of $Dec_{l+1}\, G$. Playing in the same fashion in the subsequent $k - l - 1$ rounds, Spoiler finally achieves the players' moves in some non-isomorphic $G_{k-l} \in Dec_k\, G$ and H_{k-l}, the latter being a copy of an element of $Dec_k\, G$. Both the graphs have c vertices. Now Spoiler selects the $c - 1$ remaining vertices of H_k and wins whatever Duplicator responds.

Subcase 2.2: $G_0 \cong H_0$. Though the graphs are isomorphic, the crucial fact is that G_0, unlike H_0, contains a selected vertex. By the definition of an inclusion-free graph, every automorphism of $A \cong G_0 \cong H_0$ takes each cocomponent onto itself. Therefore every isomorphism between G_0 and H_0 matches cocomponents of these graphs in the same way. Let Y be the counterpart of $Env_k(x_0)$ in H_0 with respect to this matching. In the second round Spoiler selects an arbitrary y_1 in Y. Denote Duplicator's answer by x_1. If $x_1 \in Env_k(x_0)$, Spoiler selects all Y and wins. Otherwise there is $m \leq rk\, A$ such that $Env_m(x_1)$ in G_0 and $Env_m(y_1)$ in H_0 are non-isomorphic. This allows Spoiler to apply the strategy of Subcase 2.1.

Case 3: Neither Case 1 nor Case 2. Spoiler plays in $G_0 = G$. His first move x_0 is arbitrary. Denote Duplicator's response in H by y_0 and set $H_0 = Env_0(y_0)$. Since we are not in Case 2, $G_0 \not\subset H_0$. As G_0 is inclusion-free, $\overline{G_0}$ has a connected component G_1 with no isomorphic copy in $\overline{H_0}$. Spoiler selects x_1 arbitrarily in G_1. Let Duplicator respond with y_1 somewhere in H_0 and denote $H_1 = Env_1(y_1)$. Thus $G_1 \ncong H_1$ and $G_1 \not\subset H_1$, the latter again because we are not in Case 2. In the next round Spoiler again selects a vertex in a component G_2 of $\overline{G_1}$ absent in $\overline{H_1}$. Continuing in the same fashion, Spoiler finally enforces playing the game on some $G_m \in Dec_m\, G_0$ and $H_m \in Dec_m\, H_0$ with $G_m \not\subset H_m$ under one of the two terminal conditions: (1) $m < k$ and H_m (or its complement) is a cocomponent of H. (2) $m = k$. In the first case note that, as we are not in Case 1, H_m is embeddable in some cocomponent of G (or its complement) and hence has at most c vertices. Therefore it suffices for Spoiler to select altogether $c + 1$ vertices in G_m to win (recall the assumption that Duplicator bewares of the environment threat and hence cannot move outside H_m). In the second case G_m is a cocomponent of G and hence has c vertices. Spoiler selects all G_m. Since Duplicator's response must be complement-connected, she is enforced to play within a cocomponent of H_m and hence loses.

Length of the game. The above strategy allows Spoiler to win in at most $k+c$ moves under the condition that Duplicator bewares of the environment threat. If Duplicator ignores this threat, Spoiler needs $k+1$ additional moves according to Lemma 1. $\square$

Let R_0 consist of all complement-connected graphs of order 5. Assume that R_{i-1} is already specified. Fix F_i to be the family of all $\lfloor |R_{i-1}|/2 \rfloor$-element subsets of R_{i-1}. Define R_i to be the set of the complements of $\bigsqcup_{G \in S} G$ for all S in F_i. Note that R_i consists of inclusion-free uniform graphs of rank i whose cocomponents all have 5 vertices. All graphs in R_i have the same order; Denote it by N_i. Let $M_i = |R_i|$. By the construction,

$$M_{i+1} = \binom{M_i}{\lfloor M_i/2 \rfloor} = \sqrt{\frac{2+o(1)}{\pi M_i}}\, 2^{M_i} \text{ and } N_{i+1} = \lfloor M_i/2 \rfloor N_i > M_i.$$

A simple estimation shows that $N_i \geq \textit{Tower}\,(i - O(1))$. To complete the proof of Theorem 1, choose G_i in R_i. Using Main Lemma, we obtain $q_0(N_i) \leq D_0(G_i) \leq 2i + 6 \leq 2 \log^* N_i + O(1)$.

4 Lower bound: Proof-sketch of Theorem 2

Let $L_a(G)$ denote the minimum length of an a-alternation sentence defining G.

Lemma 3 $L_a(G) \leq \textit{Tower}\,(D_a(G) + \log^* D_a(G) + O(1))$.

An analog of this lemma for $L(G)$ and $D(G)$ appears in [5] but its proof does not work under restrictions on the alternation number. The proof of Lemma 3 will appear in the full version.

Given n, denote $k = q_0(n)$ and fix a graph G on n vertices such that $D_0(G) = k$. By Lemma 3, G is definable by a 0-alternation Φ of length at most $\textit{Tower}\,(k + \log^* k + O(1))$. Using the standard reduction, we convert Φ to an equivalent prenex $\exists^*\forall^*$-sentence Ψ (i.e. existential quantifiers in Ψ all precede universal quantifiers). Since the reduction does not increase the total number of quantifiers, $D(\Psi) \leq L(\Phi)$. It is well known and easy to prove that, if a prenex $\exists^*\forall^*$-sentence Ψ is true on some structure, then it is true on some structure of order at most $D(\Psi)$. Since the Ψ is true only on G, we have $n \leq D(\Psi) \leq L(\Phi) \leq \textit{Tower}\,(k + \log^* k + O(1))$, which proves the theorem.

References

[1] A. Brandstädt, V. B. Le, J. P. Spinrad. *Graph classes: a survey.* SIAM Monographs (1999).

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

[3] 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).

[4] 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).

[5] O. Pikhurko, J. Spencer, O. Verbitsky. Succinct definitions in the first order theory of graphs. Accepted for publication in *Annals of Pure and Applied Logic*. (arxiv.org/abs/math.LO/0401307)

[6] O. Pikhurko, H. Veith, O. Verbitsky. First order definability of graphs: tight bounds on quantifier rank. *Submitted* (2004) (arxiv.org/abs/math.CO/0311041)

[7] F. Ramsey. On a problem of formal logic. *Proc. London Math. Soc. 2-nd series* 30:264–286 (1930).

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