УДК 519.95

O. Verbitsky

TWO CONSEQUENCES OF THE DICHOTOMY THEOREM ON FIRST ORDER DEFINABILITY OF GRAPHS

O. V. Verbitsky. *Two consequences of the dichotomy theorem on first order definability of graphs*, Matematychni Studii, **22** (2004) 3–9.

Given two graphs G and H, a vertex u of G and a vertex v of H such that there is no isomorphism from G to H taking u to v, let $V(G,u,H,v)$ denote the minimum number of variables in a first order formula $\Phi(x)$ that is true on (G,u) but false on (H,v). Let $Var(G)$ be the maximum of $V(G,u,H,v)$ over all vertices u of G and all possible pairs (H,v). Refining upon a result of Immerman and Lander, we prove that the class of graphs G on n vertices with $Var(G) \le (n+5)/2$ is efficiently recognizable and that, if $Var(G) > (n+5)/5$, then the exact value of $Var(G)$ is efficiently computable. We also solve a particular case of an open problem on the computational complexity of Ehrenfeucht games posed by Pezzoli.

О. В. Вербицкий. *Два следствия теоремы дихотомии об определимости первого порядка графов* // Математичні Студії. – 2004. – Т.22, №1. – С.3–9.

Для графов G, H и их вершин u, v таких, что никакой изоморфизм из G на H не переводит u в v, пусть $V(G,u,H,v)$ обозначает наименьшее число переменных в формуле первого порядка $\Phi(x)$, истинной на (G,u), но ложной на (H,v). Пусть $Var(G)$ — максимальное значение $V(G,u,H,v)$ по всевозможным вершинам u графа G и парам (H,v). Усиливая результат Иммермана и Ландера, мы доказываем, что класс графов G на n вершинах с $Var(G) \le (n+5)/2$ эффективно распознаваем и что, если $Var(G) > (n+5)/5$, то точное значение $Var(G)$ эффективно вычислимо. Мы также частично отвечаем на открытый вопрос Пеззоли о вычислительной сложности игр Эренфойхта.

1. Introduction. We consider a first order language of graph theory containing two binary relation symbols for adjacency and equality. A graph is considered a structure with a single anti-reflexive and symmetric binary relation. Thus, every closed first order formula Φ is either true or false on G. Given two non-isomorphic graphs G and G', we say that Φ *distinguishes G from G'* if Φ is true on G but false on G'. Let $V(G,G')$ denote the minimum number of variables in a formula distinguishing G from G' (different occurrences of the same variable are not counted). Notice the trivial upper bound

$$V(G,G') \le n \quad \text{for } G \text{ and } G' \text{ both on } n \text{ vertices} \tag{1}$$

(using one variable for each of n vertices, one can explicitly list all adjacencies and non-adjacencies of G).

2000 *Mathematics Subject Classification:* 03C13, 05C60.

The analog of $V(G,G')$ in logic with *counting quantifiers* proves closely related to the Weisfeiler-Lehman graph canonization algorithm that was studied since the seventies (see e.g. [1] for historical survey). An important combinatorial parameter, occurring in all bounds for the running time, is *dimension* of the algorithm. Immerman and Lander [4] came up with a logical characterization of the dimension, which was developed in [1]. This characterization implies that the optimum dimension of the Weisfeiler-Lehman algorithm sufficient to recognize non-isomorphism of two graphs G and G' does not exceed $V(G,G')$. This was a motivation of recent work by Pikhurko, Veith, and the author [7], where we improved the trivial bound (1) to

$$V(G,G') \le \frac{n+3}{2} \quad \text{for } G \text{ and } G' \text{ both on } n \text{ vertices.} \tag{2}$$

As simple examples show, this bound is tight up to an additive constant of 3/2.

We say that a first order formula Φ *defines* a graph G if Φ distinguishes G from all non-isomorphic graphs with any number of vertices. Let $V(G)$ be the minimum number of variables in a formula defining G. By a result of [2], we have

$$V(G) = \max\{\, V(G,G') : G' \not\cong G\}, \tag{3}$$

where $\cong$ denotes isomorphism of graphs. For G on n vertices, an obvious upper bound is

$$V(G) \le n+1 \tag{4}$$

(we need variables $x_1,\dots,x_n$ for the vertices and yet another variable x_{n+1} to say that any x_{n+1} is equal to one of $x_1,\dots,x_n$). Bound (4) is attained by the complete graph and therefore cannot be improved. Nevertheless, in [7] we prove the following result.

Call two vertices of a graph *similar* if they are either simultaneously adjacent or not to any other vertex. This is an equivalence relation and each equivalence class spans either a complete or an empty subgraph. Let $\sigma(G)$ denote the maximum number of pairwise similar vertices in G.

Proposition 1. (The Dichotomy Theorem, Pikhurko-Veith-Verbitsky [7]) *Let G be a graph on n vertices. If $\sigma(G) \le (n+3)/2$, then $V(G) \le (n+5)/2$. Otherwise $V(G) = \sigma(G)+1$.*

We call this result *Dichotomy Theorem* because it implies that, for any graph G, either $V(G) \le (n+5)/2$ or $V(G)$ is easily computable. Note in this respect that it seems plausible that generally $V(G)$ is an incomputable function of a graph. An evidence in favour of this hypothesis is provided by the classical Trakhtenbrot theorem which is based on simulation of a Turing machine computation by a first order sentence about finite graphs.

In the context of logical characterization of the Weisfeiler-Lehman algorithm, Immerman and Lander [4] consider a generalization of $V(G)$, which we will denote by $Var(G)$. This logical invariant of a graph generalizes $V(G)$ in the following two aspects.

First, we now consider a graph coupled with one of its vertices. We call two such pairs (G,u) and (H,v) *isomorphic* and write $(G,u) \cong (H,v)$ if there is an isomorphism from G to H taking u to v. We say that a formula $\Phi(x)$ with one free variable x *distinguishes* (G,u) from (H,v) if $\Phi(x)$ is true on G with x assigned the value u and false on H with x assigned the value v. Given non-isomorphic (G,u) and (H,v), let $V(G,u,H,v)$ denote the minimum number of variables in a formula $\Phi(x)$ distinguishing (G, u) from (H, v) (note that the variable x is counted).

Second, we now deal with an extended class of structures. A *colored graph* is a structure that, in addition to the anti-reflexive and symmetric binary relation, has countably many unary relations C_i, $i \geq 1$. The truth of $C_i(v)$ for a vertex v is interpreted as coloration of v in color i. An isomorphism of colored graphs preserves the adjacency relation and, moreover, matches a vertex of one graph to an equally colored vertex of the other graph. We consider finite colored graphs, whose vertices can have only finitely many colors.

We adapt the graph invariant $\sigma(G)$ for colored graphs by letting $\sigma(G)$ denote the maximum number of pairwise similar equally colored vertices in G. All notions we have introduced so far have a perfect sense for colored graphs. Moreover, both bound (2) and Proposition 1 are proved in [7] for colored graphs.

Definition 1. (Immerman-Lander [4]) Given a graph G, let $Var(G)$ be the maximum $V(\hat{G}, u, H, v)$ over all colorations $\hat{G}$ of G, colored graphs H, and vertices u of $\hat{G}$ and v of H.

Note that $Var(G)$ is equal to the minimum k such that any coloration of G with one designated vertex is definable in the infinitary logic $L^k_{\infty\omega}$. This number is of relevance to the 1-dimensional Weisfeiler-Lehman algorithm, known also as *vertex refinement* or *canonical labeling algorithm* (see [4]). We have $V(G) \leq Var(G)$ (this follows from equality (6) in Section 3).

Immerman and Lander [4, Proposition 1.4.3] strengthen (4) proving that

$$Var(G) \leq n + 1 \tag{5}$$

for all G with n vertices. The first result of the present paper refines upon the Immerman-Lander bound, being a version of the Dichotomy Theorem for $Var(G)$.

Theorem 1. *Let G be a graph on n vertices. If $\sigma(G) \leq (n+3)/2$, then $Var(G) \leq (n+5)/2$. Otherwise $Var(G) = \sigma(G) + 1$.*

The proof is based on Proposition 1. Note that Theorem 1 is not a straightforward modification of Proposition 1 as we have to tackle the aforementioned aspects in which $Var(G)$ differs from and is more complicated than $V(G)$. In particular, we here borrow no proof ideas from [7] making use only of the result itself. Another technique used in the proof of Theorem 1 is a characterization of $V(G, G')$ in terms of the Ehrenfeucht game on G and G' (more precisely, we use the version of the game suggested by Immerman and Poizat).

Pezzoli [6] studies the Ehrenfeuch game on its own right. She addresses a computational problem of determining, given two structures G, G' and a number r, who of the players, Spoiler or Duplicator, has a winning strategy in the r-round Ehrenfeucht game on G and G'. It is proved that, for structures over any fixed vocabulary containing at least one binary and one ternary relation, the problem is PSPACE-complete. Pezzoli asks if this result holds true for structures with a binary relation only (i.e. directed graphs). In the present paper we answer Pezzoli's question negatively in a particular case when r is large comparatively to the number of elements in the smaller of the structures G and G' (which, however, does not exlude an affirmative answer to the general question). It is well know that the isomorphism problem for structures over an arbitrary fixed vocabulary has the same complexity as the isomorphism problems for graphs. The graph isomorphism problem is in the class NP and is not NP-complete unless the polynomial time hierarchy collapses to its second level (see, e.g., [5]).

Theorem 2. *Let τ be a vocabulary containing only binary relation symbols. Let G and G' denote τ-structures and n denote the number of elements in G. The problem of determining, given G and G' and a number $r \geq (n+5)/2$, who of the players has a winning strategy in the r-round Ehrenfeucht game on G and G' is computationally equivalent to recognition if G and G' are isomorphic.*

Referring to the r-round Ehrenfeucht game on G and G', we mean the game $\textsc{Ehr}_r(G, G')$ defined in the next section. The proof of Theorem 2 is based on another version of the Dichotomy Theorem proved in [7].

2. Notation and definitions.

Graphs. Given a graph G, we denote its vertex set by $\mathcal{V}(G)$. The order of G will be sometimes denoted by $|G|$, that is, $|G| = |\mathcal{V}(G)|$. The neighborhood of a vertex v consists of all vertices adjacent to v and is denoted by $\Gamma(v)$. If $X \subseteq \mathcal{V}(G)$, then $G[X]$ denotes the subgraph induced by G on X. A one-to-one map $\phi\colon S \to S'$, where $S \subseteq \mathcal{V}(G)$ and $S' \subseteq \mathcal{V}(G')$, is a *partial isomorphism* from G to G' if ϕ is an isomorphism from $G[S]$ to $G'[S']$.

The Ehrenfeucht game. Let G and G' be colored graphs with disjoint vertex sets. The r-round l-pebble Ehrenfeucht game on G and G', denoted by $\textsc{Ehr}^l_r(G, G')$, 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 remove previously placed pebbles to another vertex and place more than one pebble on the same vertex.

After each round of the game, for $1 \leq i \leq l$ let u_i (resp. v_i) denote the vertex of G (resp. G') occupied by p_i, irrespectively of who of the players placed the pebble on this vertex. If p_i is off the board at this moment, u_i and v_i are undefined. If after every of r rounds it is true that

$$u_i = u_j \text{ iff } v_i = v_j \text{ for all } 1 \leq i < j \leq l,$$

and the component-wise correspondence $(u_1, \ldots, u_l)$ to $(v_1, \ldots, v_l)$ is a partial isomorphism from G to G', this is a win for Duplicator; otherwise the winner is Spoiler.

Proposition 2. (Immerman [3, Theorem 6.10]) *$V(G, G')$ equals the minimum l such that Spoiler has a winning strategy in $\textsc{Ehr}^l_r(G, G')$ for some r.*

If we prohibit removing 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 $\textsc{Ehr}_r(G, G')$.

3. Proof of Theorem 1 Let K and H be two colored graphs. If $K \not\cong H$, then

$$\max_{u,v} V(K, u, H, v) = V(K, H).$$

Indeed, the inequality $V(K, u, H, v) \leq V(K, H)$ for all u, v is straightforward. To show that for some u and v we have the equality, let $\Phi_{u,v}(x)$ be a formula with the smallest number of variables distinguishing (K, u) from (H, v). We can assume that each $\Phi_{u,v}(x)$ really contains the free variable x for else a closed $\Phi_{u,v}$ distinguishes K and H, and we are done immediately. Then, for an arbitrary $u \in \mathcal{V}(K)$, the formula $\exists x \bigwedge_{v \in \mathcal{V}(H)} \Phi_{u,v}(x)$ distinguishes G from H and has as many variables as some $\Phi_{u,v}(x)$ has.

If $K \cong H$ and ϕ is an isomorphism from H to K, then obviously $V(K,u,H,v) = V(K,u,\phi(H),\phi(v))$ and hence in this case $\max_{u,v} V(K,u,H,v) = \max_{u,v} V(K,u,K,v)$, where the latter maximum is over vertices u and v of K that cannot be taken to one another by any automorphism of K, that is, belong to different orbits of the automorphism group $\mathrm{Aut}(G)$.

Summarizing, we conclude that

$$Var(G) = \max_{\hat{G}} \Big\{ V(\hat{G}), \max_{u,v} V(\hat{G},u,\hat{G},v) \Big\}, \tag{6}$$

where $\hat{G}$ is a coloration of G, and u and v belong to different orbits of $\mathrm{Aut}(\hat{G})$. Our nearest aim is to prove for such u and v that

$$V(\hat{G},u,\hat{G},v) \le (n+4)/2, \tag{7}$$

where n is the order of $\hat{G}$.

By [4, Fact 1.6.1], $V(\hat{G},u,\hat{G},v)$ is equal to the smallest l such that, for some r, in the game $\mathrm{EHR}_r^l(\hat{G},\hat{G})$ on vertex-disjoint copies of $\hat{G}$ Spoiler has a winning strategy from the position with u and v matched by a pair of identical pebbles. It therefore suffices to show that $l \le (n+4)/2$ pebbles are enough for Spoiler to win. We assume that u and v are equally colored for otherwise $l = 1$ is enough. Take two colors that do not occur in $\hat{G}$, say, red and blue. Given $w \in \mathcal{V}(\hat{G})$, define $\hat{G}_w$ to be the graph $\hat{G}[\mathcal{V}(G) \setminus \{w\}]$ with additional coloring the vertices that were adjacent to w in red and the remaining vertices in blue. Since mapping u to v cannot be extended to an automorphism of $\hat{G}$, we have $\hat{G}_u \not\cong \hat{G}_v$. It follows that Spoiler can win by keeping one pair of pebbles on u and v and playing $\mathrm{EHR}_r^{l-1}(\hat{G}_u,\hat{G}_v)$ with $l-1 = \max\{V(G,G') : G \not\cong G', |G| = |G'| = n-1\}$ and r as large as needed. Using bound (2) (the generalized version for colored graphs), we conclude that Spoiler is able to win with $l \le 1 + ((n-1)+3)/2 = (n+4)/2$ pebbles thereby proving (7).

Let G be a graph on n vertices. We are now prepared to prove that $Var(G) \le (n+5)/2$ whenever $\sigma(G) \le (n+3)/2$. Consider an arbitrary coloration $\hat{G}$ of G. Taking into account (6) and (7), it is enough to estimate $V(\hat{G})$. Obviously, $\sigma(\hat{G}) \le \sigma(G)$. It follows that $\sigma(\hat{G}) \le (n+3)/2$ and, by Proposition 1 (the version for colored graphs), $V(\hat{G}) \le (n+5)/2$ as required.

We now prove that $Var(G) = \sigma(G) + 1$ whenever $\sigma(G) > (n+3)/2$. Again, consider an arbitrary coloration $\hat{G}$ of G. If $\sigma(\hat{G}) \le (n+3)/2$, then $V(\hat{G}) \le (n+5)/2$ as before. If $\sigma(\hat{G}) > (n+3)/2$, then, by Proposition 1, $V(\hat{G}) = \sigma(\hat{G}) + 1 \le \sigma(G) + 1$. Since $\sigma(G)+1 > (n+5)/2$, we have $V(\hat{G}) \le \sigma(G) + 1$ for all $\hat{G}$. Taking into account also (7), we obtain $Var(G) \le \sigma(G) + 1$. Since this bound is attained by $V(\hat{G})$ for $\hat{G} = G$, we conclude that $Var(G) = \sigma(G) + 1$.

4. Computational complexity of the Ehrenfeucht game. To facilitate the exposition, we prove Theorem 2 in the case of graphs with noting that for the general case of arbitrary binary structures the proof is virtually the same.

Given non-isomorphic graphs G and G', let $D(G,G')$ denote the minimum r such that Spoiler has winning strategy in $\mathrm{EHR}_r(G,G')$. Furthermore, let $D(G) = \max\{D(G,G') : G' \not\cong G\}$. It is known that $D(G)$ is equal to the minimum quantifier rank of a first order formula defining G up to isomorphism. We use results about $D(G)$ obtained in [7].

Proposition 3. (Pikhurko-Veith-Verbitsky [7]) *There is an efficient algorithm that, given a graph G on n vertices, determines whether or not $D(G) \le (n+5)/2$.*

The proof of this result in [7] is based on estimation of $D(G, G')$ for various types of the pair (G, G'). We will need some particular facts from this analysis.

If $v \in \mathcal{V}(G)$, let $[v]_G$ denote the equivalence class consisting of all vertices in G similar to v (recall that the similarity relation is introduced in Section 1). If $[v]_G$ has at least 2 elements, then the notation $G \oplus v$ stands for a graph H obtained from G by adding a new vertex v' so that $[v]_H = [v]_G \cup \{v'\}$. In other words, v' is similar to v and adjacent to v depending on if $[v]_G$ is a clique or an independent set. Furthermore, we define $G \oplus 0v = G$ and $G \oplus kv = (G \oplus (k-1)v) \oplus v$ for a positive integer k.

Proposition 4. (Pikhurko-Veith-Verbitsky [7])
1. If G and G' are graphs of orders $n \le n'$, then

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

unless

$$\sigma(G) \ge n/2 \text{ and } G' = G \oplus (n'-n)v \text{ for some } v \in \mathcal{V}(G) \text{ with } |[v]_G| = \sigma(G). \qquad (8)$$

2. If condition (8) is true, then $\sigma(G) + 1 \le D(G, G') \le \sigma(G) + 2$ and the exact value of $D(G, G')$ is an efficiently computable function of G.

Note that condition (8) determines G' up to isomorphism provided $\sigma(G) > n/2$. Propositions 3 and 4 are proved in [7] for any class of structures over a vocabulary with only binary relation symbols (with a natural extention of the similarity relation and the operation $G \oplus v$).

Proof of Theorem 2. Let $n = |G|$. The reduction of the graph isomorphism recognition to the winner recognition is obvious: G and G' are isomorphic iff G' has the same order n and Duplicator wins $\text{EHR}_n(G, G')$.

The reduction in the other direction proceeds as follows. First decide whether $G \cong G'$. If so, the winner is Duplicator. If not, decide whether or not $D(G) \le (n+5)/2$ by using Proposition 3. In the case $D(G) \le (n+5)/2$ the bound imposed on r implies $r \ge D(G)$ and, by the definition of $D(G)$, the winner is Spoiler. In the case $D(G) > (n+5)/2$, decide whether the pair G, G' satisfies condition (8), possibly with G and G' interchanged (here we again need the ability to test graph isomorphism). If (8) is false, the winner is Spoiler because $r \ge D(G, G')$ by the bound for $D(G, G')$ in Item 1 of Proposition 4 and the bound imposed on r. If (8) is true, compute $D(G, G')$ by using Item 2 of Proposition 4. The winner is Duplicator if $r < D(G, G')$ and Spoiler otherwise. □

REFERENCES

1. Cai J.-Y., Fürer M., Immerman N. *An optimal lower bound on the number of variables for graph identification,* Combinatorica **12** (1992), no. 4, 389–410.

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

3. Immerman N. Descriptive complexity, Springer-Verlag, 1999.

4. Immerman N., Lander E. *Describing graphs: a first-order approach to graph canonization,* In: Complexity Theory Retrospective, A. Selman Ed., Springer-Verlag, (1990), 59–81.

5. Köbler J., Schöning U., Torán J. *The graph isomorphism problem: its structural complexity,* Birkhäuser, 1993.

6. Pezzoli E. *Computational complexity of Ehrenfeucht-Fraïssé games on finite structures,* In: *Proc. of the CSL'98 Conf.*, G. Gottlob, K. Seyr Eds. Lecture Notes in Computer Science 1584, Springer-Verlag, 1999, 159–170.

7. Pikhurko O., Veith H., Verbitsky O. *First order definability of graphs: tight bounds on quantifier rank,* Submitted manuscript, 2003, 50 pp. Available at *http://arxiv.org/abs/math.CO/0311041*

Department of Mechanics and Mathematics,
Kyiv National University

Received 01.03.2004