УДК 519.95

O. V. Verbitsky

ON A DUAL VERSION OF THE WEISFEILER–LEHMAN ALGORITHM

We address a simplified variant of the Weisfeiler–Lehman graph canonization algorithm that corresponds to the fragment of first order logic with bounded number of variables precisely in the same way as the standard variant corresponds to this fragment enriched with counting quantifiers. We propose a natural dual version of the color refinement subroutine and prove that the dual algorithm has optimum dimension one greater than the optimum dimension of the standard algorithm.

1. Introduction. The Weisfeiler–Lehman algorithm for recognition of graph isomorphism was invented in the sixties and since then has been intensively studied for decades (see e.g. [1, 2, 3]). The most important complexity characteristic of the algorithm is its *dimension* (the algorithm description and relevant definitions are postponed to Section 3). Cai, Fürer, and Immerman [2] characterized the minimum dimension of the algorithm sufficient to detect (non)isomorphism of input graphs G and G' as the number one smaller than the minimum number of variables in a first order formula with counting quantifiers that is true on G but false on G'. This characterization turned out very useful for proving lower bounds for the algorithm dimension needed to process graphs on n vertices. For this purpose it was also involved the relationship between the first order expressibility and the Ehrenfeucht game on graphs G and G' [4].

We here discuss a reduced version of the Weisfeiler–Lehman algorithm that corresponds to the fragment of first order logic with bounded number of variables precisely in the same way as the standard version corresponds to this fragment enriched with counting quantifiers. The reduced version has a simplified color refinement subroutine where, in contrast to the standard version, multiplicities of equally colored adjoining configurations are not recorded. By this reason, the reduced version sometimes (e.g. even for trees) requires much higher dimension. Nevertheless, the best upper bound we know for the worst case dimension of the standard Weisfeiler–Lehman algorithm actually holds also for the reduced version, what makes investigation of the reduced version worthwhile.

We propose a natural dual version of the color refinement procedure and characterize the optimum dimension of the dual algorithm in terms of the length of the Ehrenfeucht game. As a result, we prove that the dual algorithm has optimum dimension one greater than the optimum dimension of the standard algorithm. This shows that the running time of the dual version is nearly as good as in the standard algorithm. It is not excluded that on some classes of inputs the dual algorithm may have better space complexity. It would be interesting to make both empiric and theoretical comparative analysis of the space complexity of both versions.

2. Notation and definitions. Given a graph G, we denote its vertex set by $V(G)$. Given an ordered k-tuple of vertices $\bar{u} = (u_1, \ldots, u_k) \in V(G)^k$, let $s = s(\bar{u})$ be the number of distinct components in $\bar{u}$ and define a function $F_{\bar{u}} : \{1, \ldots, k\} \to \{1, \ldots, s\}$ by $F_{\bar{u}}(i) = |\{u_1, \ldots, u_i\}|$. Furthermore, let $G_{\bar{u}}$ be the graph on the vertex set $\{1, \ldots, s\}$ with vertices a and b adjacent iff, for the smallest i and j such that $F_{\bar{u}}(i) = a$ and $F_{\bar{u}}(j) = b$, u_i and u_j are adjacent in G. The pair $(F_{\bar{u}}, G_{\bar{u}})$ is an *isomorphism type* of $\bar{u}$ and will be denoted by $[\bar{u}]$.

If $w \in V(G)$ and $i \le k$, we let $\bar{u}^{i,w}$ denote the result of substituting w in place of u_i in $\bar{u}$. The *order* of G is the number of its vertices. We write $G \cong G'$

to say that graphs G and G' are isomorphic. A *coloring* of $V(G)^k$ is a map from $V(G)^k$ to an arbitrary set, whose elements are called *colors*.

3. Description of the algorithm. We distinguish two modes of the algorithm. In the *canonization mode* the algorithm takes as an input a graph G and is purported to output its *canonic form* $W(G)$, that is, it is required that $W(G) = W(G')$ iff $G \cong G'$. In the *isomorphism testing mode* the algorithm takes as an input two graphs G and G' and should decide if $G \cong G'$. We split our description of the k-dimensional Weisfeiler–Lehman algorithm and versions thereof in three stages.

Initial coloring

The algorithm assigns each $\bar{u} \in V(G)^k$ color $W_G^{k,0}(\bar{u}) = [\bar{u}]$ (in a suitable encoding).

Color refinement step

In the r-th step each $\bar{u} \in V(G)^k$ is assigned color

$$W_G^{k,r}(\bar{u}) = \left(W_G^{k,r-1}(\bar{u}), \left\{ (W_G^{k,r-1}(\bar{u}^{1,w}), \ldots, W_G^{k,r-1}(\bar{u}^{k,w})) : w \in V(G)\right\}\right).$$

This description regards the reduced version that is discussed in the introduction and is considered in the paper. In the proper Weisfeiler–Lehman algorithm, the second component of $W_G^{k,r}(\bar{u})$ is a multiset rather than a set.

In parallel we start description of the dual version. We will denote the r-th dual coloring of $\bar{u} \in V(G)^l$ by $\tilde{W}_G^{l,r}(\bar{u})$. The initial dual and standard colorings coincide: $\tilde{W}_G^{l,0}(\bar{u}) = W_G^{k,0}(\bar{u})$.

Dual color refinement:

$$\tilde{W}_G^{l,r}(\bar{u}) =$$
$$\left(\tilde{W}_G^{l,r-1}(\bar{u}), \left\{ \tilde{W}_G^{l,r-1}(\bar{u}^{1,w}) : w \in V(G)\right\}, \ldots, \left\{ \tilde{W}_G^{l,r-1}(\bar{u}^{l,w}) : w \in V(G)\right\}\right).$$

Below we analize the dual version. As it will be easily seen, all the same holds true for the standard version. The following fact is straightforward.

Proposition 1. *If ϕ is an isomorphism from G to G', then for all k, r, and $\bar{u} \in V(G)^k$ it holds $\tilde{W}_G^{k,r}(\bar{u}) = \tilde{W}_G^{k,r}(\phi^k(\bar{u}))$.*

Proposition 2. *For every pair of graphs G and G' there is a number R such that for all $\bar{u} \in V(G)^k$, $\bar{v} \in V(G')^k$, and $r > R$*

$$\tilde{W}_G^{k,r}(\bar{u}) = W_{G'}^{k,r}(\bar{v}) \quad \textit{iff} \quad \tilde{W}_G^{k,R}(\bar{u}) = W_{G'}^{k,R}(\bar{v}).$$

Moreover, if $R_k(G, G')$ denotes the smallest such R, then $R_k(G, G') < |G|^k + |G'|^k$.

P r o o f. By Proposition 1 it suffices to prove the claim for arbitrary isomorphic copies of G and G' and we therefore can suppose that $V(G)$ and $V(G')$ are disjoint. Colorings $\tilde{W}_G^{k,r}$ and $\tilde{W}_{G'}^{k,r}$ determine a partition of the union $V(G)^k \cup V(G')^k$ into monochromatic classes. Denote this partition by Π^r. Since the $(r+1)$-th color incorporates the r-th color, Π^{r+1} is a subpartition of Π^r. It is clear that we eventually have $\Pi^{R+1} = \Pi^R$ and the smallest such R is less than $|V(G)|^k + |V(G')|^k$.

Computing an output

Isomorphism testing mode. The algorithm terminates color refinement as soon as the partition Π^r of $V(G)^k \cup V(G')^k$ coincides with Π^{r-1}, i.e., after performing $r = R_k(G, G') + 1$ refinement steps. The algorithm decides that $G \cong G'$ iff

$$\left\{ \tilde{W}_G^{k,r}(a^k) : a \in V(G)\right\} = \left\{ \tilde{W}_G^{k,r}(b^k) : b \in V(G')\right\}, \tag{1}$$

where w^k denotes a diagonal vector $(w_1, \dots, w_k)$ with all $w_i = w$.

Canonization mode. The algorithm performs $r = 2|G|^k - 1$ refinement steps and outputs the set $\left\{ \tilde{W}_G^{k,r}(u^k) : u \in V(G) \right\}$.

This completes description of (the dual version of) the algorithm. The dual version differs from the standard one only in the color refinement step. We will refer to the dual k–dimensional version as the DUAL k–WL ALGORITHM and to the standard k–dimensional version as the STANDARD k–WL ALGORITHM.

In the above description we skipped some important implementation details. Denote the minimum length of the code of $\tilde{W}_G^{k,r}(\bar{u})$ over all $\bar{u}$ by $L(r)$. As easily seen, for any natural encoding we should expect that $L(r) \geq (k+1)L(r-1)$. To prevent increasing $L(r)$ at the exponential rate, before every refinement step we arrange colors of all k–tuples in the lexicographic order and replace each color with its number. In the canonization mode we should keep substitution tables of all steps. In the isomorphism testing mode this is unnecessary but it should be stressed that color renaming must be common for both input graphs.

For each dimension k, making the decision in the isomorphism testing mode can be done in space bounded by a polynomial in n. This is so because checking the condition (1) reduces to deciding, given $\bar{u} \in V(G)^k$ and $\bar{v} \in V(G')^k$, if $\tilde{W}_G^{k,r}(\bar{u}) = \tilde{W}_{G'}^{k,r}(\bar{v})$. The latter reduces to checking equalities of type $\tilde{W}_G^{k,r-1}(\bar{u}) = \tilde{W}_{G'}^{k,r-1}(\bar{v})$ for at most kn^2+1 pairs $(\bar{u}, \bar{v})$, that can be done one by one with recording only the results of preceding checks.

4. Relation to the Ehrenfeucht game. Let G and G' be graphs with disjoint vertex sets. The r–round k–pebble Ehrenfeucht game on G and G', denoted by $\text{EHR}_r^k(G, G')$, is played by two players, Spoiler and Duplicator, with using k pairwise distinct pebbles $p_1, \dots, p_k$, 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 $1 \leq i \leq k$ let x_i (resp. y_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, x_i and y_i are undefined. If after every of r rounds it is true that

$$x_i = x_j \text{ iff } y_i = y_j \text{ for all } 1 \leq i < j \leq k,$$

and the component–wise correspondence $(x_1, \dots, x_k)$ to $(y_1, \dots, y_k)$ is a partial isomorphism from G to G', this is a win for Duplicator; Otherwise the winner is Spoiler.

Given k–tuples $\bar{u} \in V(G)^k$ and $\bar{v} \in V(G')^k$, we use notation $\text{EHR}_r^k(G, \bar{u}, G', \bar{v})$ to denote the r–round k–pebble Ehrenfeucht game on G and G' starting from the position with $(x_1, \dots, x_k) = \bar{u}$ and $(y_1, \dots, y_k) = \bar{v}$.

Let $\text{L}(G, G')$ denote the minimum k such that Spoiler has winning strategy in $\text{EHR}_r^k(G, G')$ for some r. As proved in [2], if G and G' are non–isomorphic graphs with the same number of vertices, then the STANDARD k–WL ALGORITHM recognizes G and G' as non-isomorphic iff $k \geq \text{L}(G, G') - 1$. We are able to obtain a similar result for the DUAL k–WL ALGORITHM.

Proposition 3. *For all $\bar{u} \in V(G)^k$ and $\bar{v} \in V(G')^k$ the equality*

$$\tilde{W}_G^{k,r}(\bar{u}) = \tilde{W}_{G'}^{k,r}(\bar{v}) \tag{2}$$

holds iff Duplicator has a winning strategy in $\text{EHR}_r^k(G, \bar{u}, G', \bar{v})$.

P r o o f. We proceed by induction on r. The base case $r = 0$ is straightforward by the definitions of the initial coloring and the game. Assume that the proposition is true for $r-1$ rounds.

Let x_i and y_i denote the vertices in G and G' respectively marked by the i-th pebble pair. Assume (2) and consider the Ehrenfeucht game on G, G' with initial configuration $(x_1,\ldots,x_k) = \bar{u}$ and $(y_1,\ldots,y_k) = \bar{v}$. First of all, this configuration is non–losing for Duplicator since (2) implies that $[\bar{u}] = [\bar{v}]$. Assume that in the first move Spoiler removes the j-th pebble, say in G, from u_j to another vertex $a \in V(G)$. By the definition of $\tilde{W}_G^{k,r}$, it holds

$$\left\{ \tilde{W}_G^{k,r-1}(\bar{u}^{j,w}) : w \in V(G) \right\} = \left\{ \tilde{W}_{G'}^{k,r-1}(\bar{v}^{j,w}) : w \in V(G') \right\}.$$

It follows that there is $b \in V(G')$ such that $\tilde{W}_G^{k,r-1}(\bar{u}^{j,a}) = \tilde{W}_{G'}^{k,r-1}(\bar{v}^{j,b})$. In his first move Duplicator selects such b. Note that Duplicator does not lose after the first round because $[\bar{u}^{j,a}] = [\bar{v}^{j,b}]$. Furthermore, Duplicator wins $\mathrm{EHR}_r^k(G,\bar{u},G',\bar{v})$ iff he wins $\mathrm{EHR}_{r-1}^k(G,\bar{u}^{j,a},G',\bar{v}^{j,b})$. The latter is true by the induction assumption.

Conversely, assume that (2) is false. It follows that either $\tilde{W}_G^{k,r-1}(\bar{u}) \neq \tilde{W}_{G'}^{k,r-1}(\bar{v})$ (then Spoiler wins in $r-1$ moves by the induction assumption) or there is $j \le k$ and a vertex a in one of the graphs, say in G, such that for every b in the other graph, resp. in G', $\tilde{W}_G^{k,r-1}(\bar{u}^{j,a}) \neq \tilde{W}_{G'}^{k,r-1}(\bar{u}^{j,b})$. In the latter case Spoiler in his first move removes the j-th pebble to a. Let b be the vertex that Duplicator in response marks by the other copy of the j-th pebble. Starting from the next round the players essentially play $\mathrm{EHR}_{r-1}^k(G,\bar{u}^{j,a},G',\bar{v}^{j,b})$. Spoiler has a winning strategy by the induction assumption. The proof is complete.

If $G \cong G'$, then any version of the Weisfeiler–Lehman algorithm recognizes G and G' as isomorphic for every dimension k. This follows from Proposition 1. If $G \not\cong G'$, then the algorithm may be wrong if k is chosen too small.

Proposition 4. *If G and G' are non–isomorphic graphs of the same order n, then the* DUAL k–WL ALGORITHM *recognizes G and G' as non–isomorphic iff $k \ge \mathrm{L}(G,G')$.*

P r o o f. Look at the decision criterion (1) and note that Duplicator has a winning strategy in $\mathrm{EHR}_r^k(G,G')$ iff for every $a \in V(G)$ (resp. $b \in V(G')$) there is $b \in V(G')$ (resp. $a \in V(G)$) such that Duplicator has a winning strategy in $\mathrm{EHR}_{r-1}^k(G,a,G',b)$ or, equivalently, in $\mathrm{EHR}_{r-1}^k(G,a^k,G',b^k)$. It follows by Propositions 2 and 3 that the k-dimensional algorithm decides that $G \cong G'$ iff Duplicator has a winning strategy in $\mathrm{EHR}_r^k(G,G')$ for all r. Recall that $\mathrm{L}(G,G')$ is equal to the smallest k such that Spoiler has a winning strategy in $\mathrm{EHR}_r^k(G,G')$ for some r. Therefore the decision of the k-dimensional algorithm is correct iff $k \ge \mathrm{L}(G,G')$.

We call the smallest dimension of the algorithm giving correct outputs on graphs of order n the *optimum dimension.* Let $\mathrm{L}(n)$ denote the maximum $\mathrm{L}(G,G')$ over all non–isomorphic G and G' both of order n. In these terms Proposition 4 is rephrased as follows.

Theorem. *The optimum dimension of the* DUAL WL ALGORITHM *is equal to* $\mathrm{L}(n)$.

In [5] we prove that $\mathrm{L}(n) \le (n+3)/2$.

Corollary. *The optimum dimension of the* DUAL WL ALGORITHM *does not exceed* $(n+3)/2$.

1. *Babai L.* Automorphism groups, isomorphism, reconstruction// Chapter 27 of the *Handbook of Combinatorics*, 1447–1540. R. L. Graham, M. Grötschel, L. Lovász Eds. Elsevier Publ., 1995.
2. *Cai J.-Y., Fürer M., Immerman N.* An optimal lower bound on the number of variables for graph identification// Combinatorica, 1992. – Vol. 12 – № . 4 – P. 389–410.
3. *Grohe M.* Isomorphism testing for embeddable graphs through definability// Proc. of the 32nd ACM Ann. Symp. on Theory of Computing, 2000. – P. 63–72.
4. *Ehrenfeucht A.* An application of games to the completeness problem for formalized theories// Fundam. Math., 1961. – Vol. 49 – P. 129–141.
5. *Pikhurko O., Veith H., Verbitsky O.* First order definability of graphs: tight bounds on quantifier rank// Submitted, 2003.

ПРО ДУАЛЬНУ ВЕРСІЮ АЛГОРИТМУ ВАЙСФАЙЛЕРА–ЛЕМАНА

Ми розглядаємо спрощений варіант алгоритму канонізації графів Вайсфайлера–Лемана, який відповідає фрагментові логіки першого порядку з обмеженою кількістю змінних таким же чином, як стандартний варіант відповідає цьому фрагментові збагаченому рахуючими кванторами. Ми пропонуємо природну дуальну версію процедури подрібнення кольорів і доводимо, що оптимальна розмірність дуального алгоритму на одиницю перевищує оптимальну розмірність стандартного алгоритму.

О ДУАЛЬНОЙ ВЕРСИИ АЛГОРИТМА ВАЙСФАЙЛЕРА–ЛЕМАНА

Мы рассматриваем упрощенный вариант алгоритма канонизации графов Вайсфайлера–Лемана, соответствующий фрагменту логики первого порядка с ограниченным числом переменных точно также, как стандартный вариант соответствует этому фрагменту обогащенному считающими кванторами. Мы предлагаем естественную дуальную версию процедуры уточнения цветов и доказываем, что оптимальная размерность дуального алгоритма на единицу больше, чем оптимальная размерность стандартного алгоритма.

Київський національний
університет ім. Т. Шевченка

Отримано
15.07.03