COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA **19** (3) (1999) 321–333

ISOMORPHISM OF COLOURED GRAPHS WITH SLOWLY INCREASING MULTIPLICITY OF JORDAN BLOCKS

SERGEI EVDOKIMOV* and ILIA PONOMARENKO†

Received November 29, 1994

We show that the isomorphism test for n-vertex edge coloured graphs with multiplicity of Jordan blocks bounded by k can be done in time $(k^k n)^{O(1)}$.

1. Introduction

The Graph Isomorphism Problem (ISO) is to recognize in an efficient way whether two given graphs are isomorphic, i.e., whether there is a bijection of their vertex sets preserving the adjacency of vertices. The computation complexity status of the ISO is unknown at present and the best general isomorphism test for n-vertex graphs runs in time $n^{O(\sqrt{n/\log n})}$ (see [2]).

The failure in the attempts to find a polynomial-time [1] algorithm for the ISO in the class of all graphs led to the investigation of the problem in some special classes of them. There is a great variety of such results, we mention only a few of them. There exist polynomial-time algorithms for graphs with bounded degree [10] and for graphs with bounded eigenvalue multiplicity [1]. We also mention an $n^{O(\log n)}$-algorithm for tournaments (directed graphs with exactly one arc between any two distinct vertices) [3].

Paper [1] contains an $n^{O(k)}$-isomorphism test for the class of undirected n-vertex graphs with eigenvalue multiplicity bounded by k. The question arises: can

Mathematics Subject Classification (1991): 05C60, 05C85, 05B30

* Research supported by the Volkswagen-Stiftung Program on Computational Complexity and by RFFI, grant 96-15-96060.

† Research supported by the Volkswagen-Stiftung Program on Computational Complexity and by RFFI, grant 96-15-96060.

[1] Below under "polynomial time" we always mean "polynomial time in n".

the complexity of the algorithm be improved by pulling k out of the exponent. The purpose of the paper is to give a positive answer to this question, which would provide a polynomial-time isomorphism test not only for small k but also for k slowly increasing with respect to n (one more algorithm of such a kind for another problem was described in [7]). To formulate the main result we make use of the following function $F(k)$ introduced in fact in [6]:

$$F(k) = \sup_G [G : \mathrm{sol}(G)] \tag{1}$$

where G runs over all transitive groups the degree of any irreducible constituent of the permutation representation of which is at most k, and $\mathrm{sol}(G)$ is the maximal solvable normal subgroup of G. It was proved in [6] that $F(k) < +\infty$ for all k.

Let Γ be an edge coloured graph and $A_i = A_i(\Gamma)$ be the adjacency matrix of the binary relation corresponding to the ith coloured class, $i = 1, \dots, s$. Set

$$m(\Gamma) = \min_i m(A_i)$$

where $m(A_i)$ is the maximum multiplicity of a Jordan block of the matrix A_i. Using the standard linear algebra technique one can find $m(\Gamma)$ in polynomial time. We prove the following result.

Theorem 1. *The isomorphism test for n-vertex edge coloured graphs Γ with $m(\Gamma) \le k$ can be done in time*

$$f(k) \cdot n^{O(1)} \quad \text{with} \quad f(k) = k^{k-1} F(k)^2$$

where $F(k)$ is defined by (1).

It follows from [6] that in fact $F(k) \le J(k)^{\log_2 k}$ where $J(k)$ is the Jordan function (M. Isaacs [9] improved this upper bound to $F(k) \le J(k)$). Furthermore, $J(k) \le k^{O(k^2/\log^2 k)}$ (see [4]) and $J(k) \le (k!)^{O(1)}$ under the Classification of Finite Simple Groups (CFSG) (see [12]). So

$$f(k) \le e^{O(k^2/\log k)} \quad \text{and} \quad f(k) \le e^{O(k \log k)} \ \text{(CFSG)}.$$

Thus the isomorphism test of Theorem 1 is polynomial-time for

$$k \le O((\log n \log\log n)^{1/2})$$

uncoditionally and for $k \le O(\log n / \log\log n)$ under the CFSG respectively.

As a corollary of Theorem 1 we have the following answer to the above question.

Theorem 2. *The isomorphism test for n-vertex graphs with eigenvalue multiplicity bounded by k can be done in time $f(k)n^{O(1)}$ where $f(k)$ is as in Theorem 1.*

We prove Theorem 1 by using one of the oldest approaches to the ISO developed by B. Weisfeiler and A. Lehman (see [13]). With each coloured graph Γ they associate an algebra $W(\Gamma)$ (called the cellular algebra of Γ) which is the smallest matrix

algebra over $\mathbf{C}$ containing the adjacency matrices $A_i(\Gamma)$, the identity matrix and the all-one matrix, and closed under the Hermitian conjugation and the Hadamard (componentwise) multiplication. They also showed that $\mathrm{Aut}(\Gamma) = \mathrm{Aut}(W(\Gamma))$ where the latter group consists by definition of all permutation matrices commuting with any matrix of $W(\Gamma)$. Given a coloured graph Γ the cellular algebra $W(\Gamma)$ can be constructed in polynomial time.

Let Γ be a graph satisfying the hypothesis of Theorem 1. The cellular algebra $W(\Gamma)$ is a semisimple algebra over $\mathbf{C}$. So the standard matrix representation of $W(\Gamma)$ is a sum of irreducible representations. A straightforward check shows that the multiplicity of each of them is at most k. Thus Theorem 1 can be deduced from the following statement (as to isomorphisms and canonical labelings of cellular algebras see Section 3).

Theorem 3 (MAIN THEOREM). *A canonical labeling and the automorphism group of a cellular algebra on n points can be found in time $f(k)n^{O(1)}$ where k is its maximum irreducible representation multiplicity and $f(k)$ is as in Theorem 1.*

It follows from [5] that the number k in Theorem 3 can be found in polynomial time.

The proof of the MAIN THEOREM for a primitive cellular algebra W is contained in Subsection 3.4. To reduce the general case to the primitive one we use for cellular algebras an interpretation of the standard permutation group technique. To get the required upper bound we need to control the groups arising throughout the algorithm. To do this we observe that the maximum degree of an irreducible representation of the group $\mathrm{Aut}(W)$ entering its standard permutation representation is bounded by k. This implies by the definition of $F(k)$ that $[G : \mathrm{sol}(G)] \le F(k)$ for each transitive constituent G of $\mathrm{Aut}(W)$. Thus the problem is reduced to permutation group computation with solvable groups developed in [3].

The paper consists of four sections. Section 2 contains the definition of a cellular algebra and related concepts as well as the basic notations used along the paper. It also contains some statements concerning representation properties of cellular algebras. In Section 3 we consider the canonization problem for cellular algebras. In Section 4 we prove the MAIN THEOREM and deduce Theorem 1 from it.

Notation. As usual by $\mathbf{C}$ we denote the field of complex numbers. If L is a linear space over $\mathbf{C}$, then the set of all linear operators on L is denoted by $\mathrm{End}(L)$.

Throughout the paper V denotes a finite set with $n = |V|$ elements. A subset of $V \times V$ is called a relation on V. If E is an equivalence (i.e. reflexive, symmetric and transitive relation) on V, then V/E denotes the set of all equivalence classes modulo E.

The algebra of all complex matrices whose rows and columns are indexed by the elements of V is denoted by Mat_V, its unity element (the identity matrix) by I_V and the all-one matrix by J_V. For $U \subset V$ the algebra Mat_U is considered as a subalgebra of Mat_V.

Each bijection $g: V \to V'$ defines a natural algebra isomorphism from Mat_V onto $\mathrm{Mat}_{V'}$. The image of a matrix A under it will be denoted by A^g.

If G is a group, then $H \le G$ means that H is a subgroup of G. The index of H in G is denoted by $[G:H]$.

The group of all permutations of V is denoted by $\mathrm{Sym}(V)$. In all our algorithms a permutation group $G \le \mathrm{Sym}(V)$ will be given by a set of at most n^2 generators (as to this fact and the standard permutation group algorithms see ([11])).

For integers l, m the set $\{l, l+1, \dots, m\}$ is denoted by $[l, m]$. We write $[m]$, $\mathrm{Sym}(m)$ and Mat_m instead of $[1, m]$, $\mathrm{Sym}([m])$ and $\mathrm{Mat}_{[m]}$ respectively.

2. Cellular algebras and their representations

2.1. By a *cellular algebra* on V we mean a subalgebra W of Mat_V containing the identity matrix I_V and the all-one matrix J_V, and closed under the Hermitian conjugation and the Hadamard (componentwise) multiplication denoted by $\circ$ below. The elements of V are called the *points*, the set V is called the *point set* of W. It easily follows from the definition that W is a semisimple algebra over $\mathbf{C}$.

Since W is closed under the Hadamard multiplication, it has a uniquely determined linear base $\mathcal{R} = \mathcal{R}(W)$ consisting of $\{0,1\}$-matrices such that

$$\sum_{R \in \mathcal{R}} R = J_V \quad \text{and} \quad R \in \mathcal{R} \ \Leftrightarrow \ R^T \in \mathcal{R} \tag{2}$$

where R^T is the transpose of R. The linear base $\mathcal{R}$ is called the *standard basis* of W and its elements the *basis matrices*. Set $\mathrm{Cel}(W) = \{U \subset V : \ I_U \in \mathcal{R}\}$. Each element of $\mathrm{Cel}(W)$ is called a *cell* of W. Obviously,

$$V = \bigcup_{U \in \mathrm{Cel}(W)} U \qquad \text{(disjoint union)}.$$

For two cells U, U' the number of 1's in the uth row (resp. vth column) of a basis matrix R does not depend on the choice of $u \in U$ (resp. $v \in U'$). If $|\mathrm{Cel}(W)| = 1$, these numbers (for rows and columns) coincide and their common value is called the *degree* of R.

Each matrix $R \in \mathcal{R}$ being a $\{0,1\}$-matrix is the adjacency matrix of some binary relation on V called a *basis relation* of W. By (2) the set of all of them form a partition of $V \times V$. We use all the notations introduced for basis matrices also for basis relations.

The set of all cellular algebras on V is ordered by inclusion. The algebra Mat_V is obviously the largest element of the set. We write $W \le W'$ if W is a subalgebra of W'. If $A_1, \dots, A_m \in \mathrm{Mat}_V$, then the intersection of all cellular

algebras on V containing $W, A_1, \dots, A_m$ is also a cellular algebra on V. It is denoted by $W[A_1, \dots, A_m]$.

Let $W \le \mathrm{Mat}_V$ and $W' \le \mathrm{Mat}_{V'}$ be cellular algebras and $g : V \to V'$ be a bijection such that $W' = W^g$. Then W and W' are called *isomorphic* and g is called an *isomorphism* from W to W'. Clearly, g induces a bijection between the sets $\mathcal{R}(W)$ and $\mathcal{R}(W')$. The group of all isomorphisms from W to itself contains a normal subgroup

$$\mathrm{Aut}(W) = \{g \in \mathrm{Sym}(V) : \ A^g = A, \ A \in W\}.$$

called the *automorphism group* of W. We stress that each $g \in \mathrm{Aut}(W)$ induces the identical map of W.

2.2. A cellular algebra W on V acts on the linear space $\mathbf{C}^V$ spanned by the set V. Below we consider the induced actions of some special subalgebras of W on subspaces of $\mathbf{C}^V$.

Let U be a union of cells of W. The subalgebra $I_U W I_U \subset W$ invariantly acts on the subspace $I_U \mathbf{C}^V$ of $\mathbf{C}^V$ identified with $\mathbf{C}^U$. So it can be viewed as a subalgebra of Mat_U. Clearly, it is closed under the Hermitian conjugation and the Hadamard multiplication and contains I_U and J_U. Thus it is a cellular algebra on U called the *restriction* of W to U and denoted by W_U.

Let E be an *equivalence of* W, i.e. that on V for which the matrix $I_E = \sum_{U \in V/E} J_U/|U|$ belongs to W. The subalgebra $I_E W I_E \subset W$ invariantly acts on the subspace $I_E \mathbf{C}^V$ of $\mathbf{C}^V$ identified with $\mathbf{C}^{V/E}$ via the mapping $U \mapsto \sum_{v \in U} v$, $U \in V/E$. So it can be viewed as a subalgebra of $\mathrm{Mat}_{V/E}$. Denote it by W/E. Clearly, W/E contains $I_{V/E}$, $J_{V/E}$ and is closed under the Hermitian conjugation.

Lemma 2.1. *The algebra W/E is closed with respect to the Hadamard multiplication in* $\mathrm{Mat}_{V/E}$.

Proof. For two basis matrices $R, S \in \mathcal{R}$ we write $R \overset{E}{\sim} S$ if S enters the decomposition of QRQ in the standard basis of W where $Q = I_E$ (we make use of the fact that $Q \in W$). This relation is obviously reflexive and transitive. Since $R \overset{E}{\sim} S$ iff $QRQ \circ QSQ \ne 0$, it is also symmetric. Thus it is an equivalence relation and its class containing R coincides with the set $\{S \in \mathcal{R} : c_S^R \ne 0\}$ where c_S^R are defined by $QRQ = \sum_{S \in \mathcal{R}} c_S^R S$. Besides, given $u, v \in V$ we have $(QRQ)_{u,v} = d/(|U||U'|)$ where U, U' are the classes of E containing u and v respectively and d is the number of $1's$ of the matrix R in $U \times U'$. So $c_S^R = c_R^R$ for all S with $c_S^R \ne 0$. Thus given $R_1, R_2 \in \mathcal{R}$ we have $QR_1Q = (c_{R_1}^{R_1}/c_{R_2}^{R_2})QR_2Q$ whenever $QR_1Q \circ QR_2Q \ne 0$. It follows that the algebra QWQ (and hence W/E) is closed under the Hadamard multiplication. ∎

The lemma implies that W/E is a cellular algebra on V/E called the *cellular factoralgebra* of W modulo E [2]. Clearly, given W and E the standard basis of W/E can be constructed in polynomial time.

2.3. Let $\Delta: W \to \mathrm{End}(L)$ be a representation of a semisimple algebra W over $\mathbf{C}$ on a linear space L. Denote by $\mathrm{Spec}(W)$ the set of all primitive central idempotents of the algebra W. For each $P \in \mathrm{Spec}(W)$ the restriction of Δ to the subspace $PL \subset L$ is a multiple of an irreducible representation of W. Denote its multiplicity by $m(P,\Delta)$ and set

$$m(\Delta) = \max_{P \in \mathrm{Spec}(W)} m(P,\Delta).$$

If W is a cellular algebra, set $m(W) = m(\Delta)$ where Δ is the standard representation of W, i.e. a faithful linear representation of W on $\mathbf{C}^V$ induced by the action of Mat_V on $\mathbf{C}^V$. We call $m(W)$ the *multiplicity* of W.

Proposition 2.2. *Let $W \le \mathrm{Mat}_V$ be a cellular algebra. Then*

(1) *if $W' \ge W$, then $m(W') \le m(W)$,*
(2) *if U is a union of cells of W, then $m(W_U) \le m(W)$,*
(3) *if E is an equivalence of W, then $m(W/E) \le m(W)$.*

Proof. Since the standard representation of W is equivalent to the restriction of that of W', statement (1) is clear. Further, the standard representation of the algebra W_U (resp. W/E) is obviously equivalent to the representation

$$\Delta_Q : QWQ \to \mathrm{End}(Q\mathbf{C}^V) \quad \text{with} \quad Q = I_U \ (\text{resp. } Q = I_E)$$

obtained from the standard representation of W by restriction. So statements (2) and (3) are consequences of the following lemma.

Lemma 2.3. *Let $\Delta : W \to \mathrm{End}(L)$ be a linear representation of a semisimple algebra W over $\mathbf{C}$, Q be an idempotent of W and $\Delta_Q : QWQ \to \mathrm{End}(QL)$ be the representation of the algebra QWQ obtained from Δ by restriction. Then*

$$m(\Delta_Q) \le m(\Delta).$$

Moreover, the mapping $P \mapsto PQ$ defines a bijection between the sets $\{P \in \mathrm{Spec}(W) : PQ \ne 0\}$ and $\mathrm{Spec}(QWQ)$.

Proof. Let $P \in \mathrm{Spec}(W)$ and $PQ \ne 0$. Since the idempotent P is primitive, the algebra PW is isomorphic to $\mathrm{End}(\mathbf{C}^r)$ for some positive integer r. Then the image of PQ with respect to this isomorphism is a nontrivial idempotent T of $\mathrm{End}(\mathbf{C}^r)$. So $PQWQ$ is isomorphic to $\mathrm{End}(T\mathbf{C}^r)$. Thus $PQ \in \mathrm{Spec}(QWQ)$ and $m(PQ,\Delta_Q) = m(P,\Delta)$. It follows that if $1 = \sum_{P \in \mathrm{Spec}(W)} P$ is the decomposition of unity of W, then $Q = \sum_{P,PQ \ne 0} PQ$ is the decomposition of unity of QWQ and

$$m(\Delta_Q) = \max_{P,PQ \ne 0} m(PQ,\Delta_Q) \le \max_P m(P,\Delta) = m(\Delta). \blacksquare$$

[2] A special case of this construction was considered in [13, Section I].

3. Canonical labeling of cellular algebras

3.1. Below by a cellular algebra W we mean one with a linear order on the set of its basis relations. This order induces a linear order on the set $\mathrm{Cel}(W)$ and lexicographic orders on the sets of all relations and all equivalences of W (recall that any such relation is a union of basis ones). It also induces linear orders on the sets of basis relations of the algebras W/E and W_U where E is an equivalence of W and U is a union of cells of W respectively. By isomorphisms of cellular algebras here we mean those preserving the order of basis relations. We say that two cellular algebras on the same set are equal if the identity map of this set is an isomorphism from one to the other. If $g: V \to V'$ is a bijection, then by definition the order of the basis relations of the algebra $W^g \le \mathrm{Mat}_{V'}$ is induced by that of W. Thus, g is an isomorphism from W to W' iff $W^g = W'$.

Given a cellular algebra W on V and $A \in \mathrm{Mat}_V$, we put in order the set of the basis relations of the algebra $W[A]$ according to the Weisfeiler–Lehman canonical algorithm, so that the following holds (see [13, Section M]):

(W-L) $(W[A])^g = W^g[A^g]$ for any bijection g with domain V.

The algebra $W[A]$ (with the corresponding order) can be constructed in polynomial time.

3.2. Our approach to the canonization problem goes back to [3] and is similar to that of [7]. By a *coset* on a finite set V of cardinality n we mean a set $C = Gg$ where $g: V \to [n]$ is a bijection and $G \le \mathrm{Sym}(V)$ is a group (equivalently, $C = gG$ with $G \le \mathrm{Sym}(n)$). A *V-pair* is by definition a pair $P = (W, C)$ where W is a cellular algebra on V and C is a coset on V. It is isomorphic to a V'-pair P', $P \cong P'$, if there exists a bijection $g: V \to V'$ such that $P^g = P'$ where $P^g = (W^g, g^{-1}C)$.

Let $\mathcal{P}$ be a class of pairs P closed under isomorphisms of pairs and $\mathcal{P}_0$ be its subclass consisting of all $[n]$-pairs over all positive integer n. A mapping $\mathrm{CF}: \mathcal{P} \to \mathcal{P}_0$ is called a *canonical form* for $\mathcal{P}$ if

(C1) $\forall P \in \mathcal{P}: \quad \mathrm{CF}(P) \cong P$,

(C2) $\forall P, P' \in \mathcal{P}: \quad P \cong P' \Leftrightarrow \mathrm{CF}(P) = \mathrm{CF}(P')$.

Any bijection h for which $P^h = \mathrm{CF}(P)$ is called a *canonical labeling* of P. All of them form a coset $\mathrm{CL}(P) = \mathrm{Aut}(P)h = (\mathrm{Aut}(W) \cap G)h$ where $P = (W, Gg)$. It is called the *canonical labeling coset* of P with respect to CF. Obviously,

$$\mathrm{CL}(P^g) = g^{-1}\,\mathrm{CL}(P) \tag{3}$$

for any bijection g with domain V. Conversely, let $P \mapsto \mathrm{CL}(P)$ be an arbitrary mapping taking a V-pair $P \in \mathcal{P}$ to a nonempty set $\mathrm{CL}(P)$ of bijections from V to $[n]$ such that P^h does not depend on the choice of $h \in \mathrm{CL}(P)$ and also (3) is satisfied. Then the mapping $P \mapsto P^h$, $h \in \mathrm{CL}(P)$, is a canonical form for $\mathcal{P}$ and $\mathrm{CL}(P)$ is the canonical labeling coset of P with respect to it.

If $C=\mathrm{Sym}(V)g$ for all $P=(W,C)\in\mathcal{P}$, then we do not refer to C and speak about a canonical form, a canonical labeling and a canonical labeling coset of W. Thus, $\mathrm{CL}(W)=\mathrm{Aut}(W)h=h\,\mathrm{Aut}(\mathrm{CF}(W))$ where h is any canonical labeling of W.

In our algorithms a coset $C=Gg$ will be given by g and a generating set of G.

3.3. Let G be a finite group. Denote by $\mathrm{sol}(G)$ the maximal normal solvable subgroup of G. Clearly, if $G=G_1\times G_2$, then $[G:\mathrm{sol}(G)]=[G_1:\mathrm{sol}(G_1)]\cdot[G_2:\mathrm{sol}(G_2)]$, and if H is a subgroup or a homomorphic image of G, then $[H:\mathrm{sol}(H)]\le[G:\mathrm{sol}(G)]$.

Proposition 3.1. *A canonical labeling coset of a pair (W,gG), $G\le\mathrm{Sym}(n)$, can be found in time $t^2n^{O(1)}$ where $t=t(G)=\max\limits_H[H:\mathrm{sol}(H)]$ with H running over all transitive constituents of G.*

Proof. It is easy to see that $(W_1,C_1)\cong(W_2,C_2)$ iff $C_1=g_1G$, $C_2=g_2G$ for $G\le\mathrm{Sym}(n)$ and $W_1^{g_1}$ is G-isomorphic to $W_2^{g_2}$. So the canonization problem for pairs is reduced to that for cellular algebras on the set $V=[n]$ with respect to the group G in sense of [3]. To solve the last problem we associate with $W\le\mathrm{Mat}_V$ a string $s=s(W)$ on $V\times V$ which is a mapping taking a 2-tuple to the index number of the basis relation containing it. So the problem is reduced to finding a canonical placement coset $\mathrm{CP}(s,G)$ of the string s with respect to the induced action of G on $V\times V$ (see [3]). We also observe that if U_1 and U_2 are orbits of G on V, then $U_1\times U_2$ is an invariant set of G on $V\times V$. Thus each transitive constituent H of G on $V\times V$ is a homomorphic image of a subgroup of the group $H_1\times H_2$ where H_1 and H_2 are some transitive constituents of G on V. By the definition of t we conclude that $[H:\mathrm{sol}(H)]\le t^2$.

Following [3] without loss of generality we assume that G is transitive on $V\times V$. In this case the above argument shows that $[G:H]\le t^2$ where $H=\mathrm{sol}(G)$. Let $G=\cup_{i=1}^r g_iH$ be a disjoint union of the cosets of G by H. Set

$$\mathrm{CP}(s,G)=\bigcup_{i\in T}\mathrm{CP}(s,g_iH)$$

where $\mathrm{CP}(s,g_iH)$ is the canonical placement coset of s with respect to g_iH found by the algorithm of [3, p.3] and $T=\{i\in[r]:\ \mathrm{CF}(s,g_iH)=\mathrm{CF}(s,g_{i_0}H)\}$ with i_0 providing a lexicographic maximum of $\mathrm{CF}(s,g_iH)$, $i\in[r]$. As in [3] one can easily prove that $\mathrm{CP}(s,G)$ is a canonical placement coset of s with respect to G.

To estimate the running time of the algorithm we observe that the group H and the permutations g_i, $i\in[r]$, can be found in time $rn^{O(1)}$ (see [11]). Since H is a solvable group, finding $\mathrm{CP}(s,g_iH)$ can be done in time $n^{O(1)}$. Besides, by the definition of t we have $r\le t^2$. Thus $\mathrm{CP}(s,G)$ can be constructed in time $t^2n^{O(1)}$. ∎

We will apply Proposition 3.1 to a pair (W,C) with C being the direct sum of cosets $C_i=\mathrm{Aut}(W_i)g_i$ where $W_i\le\mathrm{Mat}_{V_i}$ is a cellular algebra with $m(W_i)\le k$ (see Section 4). In this case the multiplicities of irreducible representations of W_i in its

standard representation coincide with the degrees of irreducible representations of the subalgebra of Mat_{V_i} centralizing W_i (see [14]). This implies that the degree of each irreducible representation of the group $\mathrm{Aut}(W_i)$ entering the permutation representation is at most k. So $t(G) \le \max_i t(\mathrm{Aut}(W_i)) \le F(k)$ where F is the function defined by (1).

3.4. In this subsection we consider the canonization problem for *primitive* cellular algebras, i.e. those having exactly one cell and exactly two equivalences the adjacency matrices of which coincide with the identity matrix and the all-one matrix respectively. We stress that a cellular algebra on a one-point set is not primitive according to this definition.

Following [8] a tuple $(v_1,\dots,v_s)\in V^s$ is called an *irredundant base* of a cellular algebra W on V if

$$W_{v_1,\dots,v_s} = \mathrm{Mat}_V \quad \text{and} \quad \{v_i\} \notin \mathrm{Cel}(W_{v_1,\dots,v_{i-1}}) \ \text{ for all } i \in [s]$$

where $W_{v_1,\dots,v_s} = W[I_{v_1},\dots,I_{v_s}]$ with $I_{v_i} = I_{\{v_i\}}$ (we successively add the matrices $I_{v_1},\dots,I_{v_s}$ to W applying at each step the Weisfeiler–Lehman canonical algorithm).

Theorem 3.2. *In the class of all primitive cellular algebras a canonical labeling coset* $\mathrm{CL}(W)$ *(and consequently the automorphism group* $\mathrm{Aut}(W)$*) of an algebra* W *on* n *points can be found elementwise in time* $d^{b-1}n^{O(1)}$ *where* d *is the minimum degree of a nonreflexive basis relation of* W *and* b *is the maximum size of its irredundant base. Moreover*

$$|\mathrm{CL}(W)| \le d^{b-1}n. \tag{4}$$

Proof. Let R be a nonreflexive basis relation of W of minimal degree with minimal index number. For $t=(v_1,\dots,v_s)\in V^s$ set

$$X(t,R) = \{u \in X :\ R(u) \not\subset X\}$$

where $R(u) = \{v \in V :\ (u,v) \in R\}$ and $X = \{v \in V :\ \{v\} \in \mathrm{Cel}(W_t)\}$ with $W_t = W_{v_1,\dots,v_s}$. Notice that the set X and hence $X(t,R)$ is linearly ordered according to the ordering of basis matrices of W_t. It follows from the primitivity of W that R is strongly connected (see [13, p.55]) and so

$$X(t,R) = \emptyset \ \Rightarrow\ W_t = \mathrm{Mat}_V\,. \tag{5}$$

Denote by $T=T(R)$ the set of all irredundant bases t of W such that

(i) $X(t_i,R)\ne\emptyset, \quad i\in[s-1]$,

(ii) $(v_i^*, v_{i+1})\in R, \quad i\in[s-1]$

where $t_i=(v_1,\dots,v_i)$ and v_i^* is the maximal element of the set $X(t_i,R)$. It follows from (i) and (5) that $T\ne\emptyset$ whereas (ii) implies that

$$|T| \le nd^{b-1}. \tag{6}$$

Each $t \in T$ defines by (W-L) a uniquely determined bijection $h_t : V \to [n]$ corresponding to the ordering of diagonal matrix units of the algebra $W_t = \mathrm{Mat}_V$. Let $W^{h_{t_0}}$ be the lexicographic leader over all W^{h_t} with $t \in T$. Set

$$\mathrm{CL}(W) = \{h_t : \ t \in T, \ W^{h_t} = W^{h_{t_0}}\}.$$

It follows from (W-L) that if $g : V \to V'$ is a bijection, then $R' = R^g$, $T' = T^g$ and $\mathrm{CL}(W') = g^{-1}\mathrm{CL}(W)$ where $W' = W^g$. Since W^h obviously does not depend on the choice of $h \in \mathrm{CL}(W)$, we conclude that $\mathrm{CL}(W)$ is a canonical labeling coset of W (see Subsection 3.2). Inequality (4) follows from (6) after taking into account that $|\mathrm{CL}(W)| \le |T|$.

The complexity of listing $\mathrm{CL}(W)$ is estimated by that of constructing the set T. The last problem is reduced by (6) to at most bnd^{b-1} calls of the Weisfeiler–Lehman canonical algorithm. ∎

When applying Theorem 3.2 in Section 4 we use the following estimates of the numbers b and d which are easily deduced from Theorem 4.3 and equality (16) of [8]:

$$b \le m(W), \qquad d \le m(W) \tag{7}$$

where $m(W)$ is the multiplicity of W (see Subsection 2.3).

4. Proofs of Theorems

In this section we prove Theorem 3 and deduce Theorem 1 from it.

Proof of Theorem 3. Denote by SOLV and PRIM the algorithms of Proposition 3.1 and Theorem 3.2 respectively. Thus $\mathrm{SOLV}(W,C)$ is a canonical labeling coset of a pair (W,C). Similarly, $\mathrm{PRIM}(W)$ is a canonical labeling coset of a primitive cellular algebra W.

Let $C_i = G_i g_i$ be a coset on V_i, $i \in [s]$. Denote by g the bijection from the disjoint union V of V_i's to $[n]$ where $n = \sum_{i=1}^{s} n_i$ with $n_i = |V_i|$, such that $v^g = v^{g_i} + \sum_{j=1}^{i-1} n_j$ whenever $v \in V_i$. It is easy to see that the coset $C = Gg$ on V with $G = \prod_{i=1}^{s} G_i$ does not depend on the choice of $g_i \in C_i$. We denote it by $\coprod_{i=1}^{s} C_i$ and call the *direct sum* of C_i.

Algorithm
Input: a cellular algebra W on V.
Output: a canonical labeling coset $A(W)$ of W.

Step 1. If $|V| = 1$, then output $\{h\}$ where h is the unique mapping from V to $[1]$.

Step 2. If $|\mathrm{Cel}(W)| > 1$, then for each $U \in \mathrm{Cel}(W)$ find recursively $C_U = A(W_U)$. Output $\mathrm{SOLV}(W,C)$ where $C = \coprod_{i=1}^{s} C_{U_i}$ with $s = |\mathrm{Cel}(W)|$ and U_i being the ith cell of W.

Step 3. Let E be the equivalence of W with minimal index number such that W/E is primitive. For each $U \in V/E$ find recursively $C_U = A(W_{[U]})$ where $W_{[U]} = W[I_U]_U$. Find $\widetilde{C} = \mathrm{PRIM}(W/E)$ and for each $h \in \widetilde{C}$ find $D_h = \mathrm{SOLV}(W_h, C_h)$ where $W_h = W[\sum_{i=1}^{s} iI_{h^{-1}(i)}]$ and $C_h = \coprod_{i=1}^{s} C_{h^{-1}(i)}$ with $s = |V/E|$. Output $\cup_{h\in S} D_h$ where $S = \{h \in \widetilde{C}\colon \mathrm{CF}(W_h, C_h) = \mathrm{CF}(W_{h_0}, C_{h_0})\}$ with h_0 providing a lexicographic maximum of $\mathrm{CF}(W_h, C_h)$, $h \in \widetilde{C}$. ∎

To prove the correctness of the algorithm it suffices to verify that W^h does not depend on $h \in A(W)$ and also that $A(W^g) = g^{-1}A(W)$ for any bijection $g\colon V \to V'$ (see Subsection 3.2). However, the former condition is an immediate consequence of the definition of $A(W)$. So we check only the latter one using for this the induction on $|V|$. If $|V| = 1$, then the algorithm terminates at Step 1 and we are done. Otherwise, set $W' = W^g$ and consider two cases.

If the algorithm terminates at Step 2, then $U_i' = U_i^g$ and $W'_{U_i'} = (W_{U_i})^{g_i}$ for all $i \in [s]$ where $g_i : U_i \to U_i'$ is the bijection induced by g. By the induction hypothesis $C'_{U_i'} = g_i^{-1} C_{U_i}$ for all i. So $C' = g^{-1}C$ and (3) implies that

$$A(W') = \mathrm{SOLV}(W^g, g^{-1}C) = g^{-1}\,\mathrm{SOLV}(W, C) = g^{-1}A(W).$$

If the algorithm terminates at Step 3, then $E' = E^g$ and $W'/E' = (W/E)^{\widetilde{g}}$ where $\widetilde{g}\colon V/E \to V'/E'$ is the bijection induced by g. So (3) implies that

$$\widetilde{C}' = \mathrm{PRIM}(W'/E') = \widetilde{g}^{-1}\,\mathrm{PRIM}(W/E) = \widetilde{g}^{-1}\widetilde{C}. \tag{8}$$

Let $h \in \widetilde{C}$. Then by (W-L) we have $W'_{h'} = (W_h)^g$ where $h' = \widetilde{g}^{-1}h$ and hence $W'_{[U_i']} = (W_{[U_i]})^{g_i}$ for all $i \in [s]$ where $U_i = h^{-1}(i)$, $U_i' = h'^{-1}(i)$ and $g_i : U_i \to U_i'$ is the bijection induced by g. So by the induction hypothesis $C'_{U_i'} = g_i^{-1} C_{U_i}$ for all i and hence $C'_{h'} = g^{-1}C_h$. Thus $(W'_{h'}, C'_{h'}) = (W_h, C_h)^g$ whence $D'_{h'} = g^{-1}D_h$ by (3). This implies by (8) that if h runs over S, then h' runs over S'. Thus

$$A(W') = \bigcup_{h\in S} D_{h'} = \bigcup_{h\in S} g^{-1}D_h = g^{-1}A(W)$$

which completes the proof of the correctness.

Let us estimate the running time $t(W)$ of the algorithm applied to a cellular algebra W. Denote by $t(k,n)$ the maximum of $t(W)$ taken over all algebras W on n points with $m(W) \le k$. We will prove by induction on n that $t(k,n) \le k^{k-1}F(k)^2 n^{O(1)}$. Let the algorithm terminate at Step 2. It follows from Proposition 2.2 that $m(W_U) \le m(W) \le k$ for all $U \in \mathrm{Cel}(W)$. So the

coset $C_U = \mathrm{Aut}(W_U)h_U$ can be found in time $t(k,|U|)$. By the definition of F we have $[H : \mathrm{sol}(H)] \le F(k)$ for each transitive constituent H of $\mathrm{Aut}(W_U)$ (see end of Subsection 3.3). So by Proposition 3.1 the coset $A(W)$ can be found in time $F(k)^2 n^{c_2}$ for some constant c_2. Therefore

$$t(W) \le \sum_{U \in \mathrm{Cel}(W)} t(k,|U|) + F(k)^2 n^{c_2} + n^{O(1)} \tag{9}$$

Let the algorithm terminate at Step 3. By Proposition 2.2 $m(W/E) \le m(W) \le k$ and $m(W_{[U]}) \le m(W) \le k$ for all $U \in \mathrm{Cel}(W)$. So the coset C_U can be found in time $t(k,n/s)$. The coset $\widetilde{C}$ can be found in time $k^{k-1}s^{c_1}$ by Theorem 3.2 and inequalities (7). By Proposition 2.2, Proposition 3.1 and the definition of F the coset D_h, $h \in \widetilde{C}$, can be found in time $F(k)^2 n^{c_2}$ where c_2 is as above. Besides, it follows from Theorem 3.2 and inequalities (7) that $|\widetilde{C}| \le k^{k-1}s$. Therefore,

$$t(W) \le k^{k-1}s^{c_1} + s \cdot t(k,n/s) + k^{k-1}s \cdot F(k)^2 n^{c_2} + n^{O(1)}. \tag{10}$$

It follows from (9) and (10) by induction that there exists a constant c for which

$$t(k,n) \le k^{k-1} F(k)^2 n^c.$$

Theorem 3 is completely proved. ∎

Proof of Theorem 1. Denote by $W(\Gamma)$ the cellular algebra generated by all the matrices A_i. Then obviously $m(W(\Gamma)) \le m(A_i)$ for all i. Thus by the hypothesis of the theorem

$$m(W(\Gamma)) \le \min_i m(A_i) = m(\Gamma) \le k.$$

Define a linear order on the standard basis of $W(\Gamma) = W[A]$ where $A = \sum_{i=1}^{s} iA_i$ according to the Weisfeiler–Lehman canonical algorithm. Then by (W-L) two graphs Γ and Γ' are isomorphic iff $W(\Gamma)$ is isomorphic to $W(\Gamma')$ and the corresponding coefficients in the decompositions of the matrices A and A' with respect to the standard bases coincide. Since the latter condition can easily be tested, Theorem 1 follows from Theorem 3. ∎

References

[1] L. Babai, D. Y. Grigoriev, D. M. Mount: Isomorphism of graphs with bounded eigenvalue multiplicity, *Proc. 14th ACM STOC*, (1982), 310–324.

[2] L. Babai, W. M. Kantor, E. M. Luks: Computation complexity and the classification of finite simple groups, *Proc. 24th FOCS*, (1983), 162–171.

[3] L. Babai, E. M. Luks: Canonical labeling of graphs, *Proc. 15th ACM STOC*, (1983), 1–15.

[4] C. W. Curtis, I. Reiner: *Representation theory of finite groups and associative algebras*, New York and London, Interscience Publishers, 1962.

[5] W. Eberly: Decomposition of algebras over **R** and **C**, *Computational Complexity,* **1** (1991), 211–234.

[6] S. Evdokimov, I. Ponomarenko: Transitive permutation groups with representations of bounded degree, *Zapiski Nauchnykh Seminarov POMI,* **223** (1995), 108–119.

[7] S. Evdokimov, I. Ponomarenko: On geometric graph isomorphism problem, *Journal of Pure and Applied Algebra,* **117 & 118** (1997), 253–276.

[8] S. Evdokimov, I. Ponomarenko: On primitive cellular algebras, Research Report No. 85168-CS, University of Bonn (February 1997) (to appear in Zapiski Nauchnykh Seminarov POMI 256, (1999)).

[9] I. M. Isaacs: Constituents of permutation characters, 1997 (to appear).

[10] E. M. Luks: Isomorphism of graphs of bounded valence can be tested in polynomial time, *J. Comp. Sys. Sci.,* **25** (1982), 42–65.

[11] W. M. Kantor, E. M. Luks: Computing in quotient groups, *Proc. 22nd ACM STOC*, (1990), 524–534.

[12] L. Pyber: How abelian is a finite group?, in: *The Mathematics of Paul Erdös,* Vol. I. (R. L. Graham et al. eds.), Algorithms and Combinatorics **13**, Springer-Verlag, Berlin, 1997, 372–384.

[13] B. Ju. Weisfeiler (editor), *On construction and identification of graphs*, Springer Lecture Notes, **558** (1976).

[14] H. Weyl: *The classical groups, their invariants and representations*, London, Humphrey Milford Oxford University Press. XII, 1939.

Sergei Evdokimov

St. Petersburg Institute for Informatics and Automation Academy of Sciences of Russia
evdokim@pdmi.ras.ru

Ilia Ponomarenko

St. Petersburg Department of Mathematical Institute Academy of Sciences of Russia
inp@pdmi.ras.ru