CANONICAL LABELING OF GRAPHS

László Babai
Dept. of Algebra & Number Theory
Eötvös University
H-1088 Budapest

Eugene M. Luks[1]
Dept. of Mathematics
Bucknell University
Lewisburg, PA 17837

ABSTRACT. We announce an algebraic approach to the problem of assigning *canonical forms* to graphs. We compute canonical forms and the associated canonical labelings (or renumberings) in polynomial time for graphs of bounded valence, in moderately exponential, $\exp(n^{\frac{1}{2}+o(1)})$, time for general graphs, in subexponential, $n^{\log n}$, time for tournaments and for 2-(v,k,λ) block designs with k,λ bounded and $n^{\log\log n}$ time for λ-planes (symmetric designs) with λ bounded. We prove some related problems NP-hard and indicate some open problems.

1. Introduction.

The computational complexity of finding canonical representatives for the isomorphism classes of finite algebraic and combinatorial structures is a long-standing unresolved question in the theory of computation. As such structures can be canonically represented by polynomial-time computable graphs [HP], [Mi1], it would suffice to find canonical forms for graphs.

It would appear that the canonical form problem for graphs is closely related to the problem of testing isomorphism; the second task can be performed at least as fast as the first and, in most instances, an isomorphism test for a class of graphs either consisted of a procedure for canonizing or else had an analogue for that problem (cf. remarks in [Lip], [Mi1]). In some recent studies, however, the gap between these problems seemed wider. In [Ba1] the "tower of groups" approach was introduced and used in a polynomial-time Las Vegas isomorphism test of colored graphs with bounded color classes. The same method yields a polynomial-time isomorphism test for graphs with bounded multiplicities of eigenvalues [BGM]. In [FHL], these Las Vegas algorithms were replaced by deterministic versions. In [Lu1], deeper group-theoretic techniques were described that yield a polynomial-time test for graphs of bounded valence. In the same paper, subexponential isomorphism tests for tournaments and for symmetric (v,k,λ) block designs were announced. An ingenious valence reduction procedure led Zemlyachenko [ZKT], [Ba3] to a moderately exponential ($\exp(n^{1-c})$) test for general graphs via the techniques of [Lu1]. Subsequent improvements of the bounded valence algorithm have brought this bound down to $\exp(c\sqrt{n}\log n)$ [Lu2] (We use the letter c to denote a positive absolute constant throughout, but possibly a different one each time). In contrast to algorithms with a combinatorial flavor, [HT] [FM] [Mi2] [Mi3] [Ba2], none of these group-theoretic isomorphism tests appeared to have implications for canonical forms. Indeed both [Ba1] and [Lu1] explicitly ask whether the methods can be modified to perform this other, potentially more useful, job. For graphs with bounded color-classes, this was soon done [KL] by a naive "lexicographic leader" idea. However, a similar approach leads to NP-hard problems even in the context of trivalent graphs (cf. §3.1). Thus, despite the fact that trivalent graph isomorphism had been brought down to $O(n^3\log n)$ [GHLSW] the fastest canonical form algorithm for this class was apparently $n^{c\sqrt{n}}$ (applying the bounded color class result of [KL] via reductions of [Ba1]). Remarks on this discrepency appear also in [CG], where a combinatorial technique of canonization of general graphs in c^n is given.

If the difference between the problems appears subtle, we offer the following (naive) observation. The algebraic methods for testing isomorphism involve the determination of generators for Aut(X), the automorphism group of a graph X. In fact, the ability to do so is both necessary and sufficient for isomorphism testing [Ma]. Does knowledge of Aut(X) lead to a canonical form? In the canonical form problem the objective is to select, wisely, from the various representations. If, as is almost always the case, Aut(X) is trivial, the number of such representations is n!. How do we select?

The main purpose, then, of the present paper is to close the remaining complexity gap. The approach begins with an algebraization of the

[1] Research supported by NSF grant MCS 81-02856

canonical form problem. For this purpose, we propose the string placement problem with respect to a permutation group action. General graph canonization is easily reducible to general string placement, but, much more significantly, special cases of graph canonization are reducible to special cases of string placement in which crucial properties of the groups are forced. We offer a string canonization algorithm and show how its timing can be tied into the group structure.

In the applications, our CF-algorithms reach the time bounds known for isomorphism. At the end we briefly mention an alternative, mainly combinatorial, method to obtain an $\exp(n^{2/3+o(1)})$ canonization of general graphs. Finally, we list some outstanding open problems related to isomorphism and canonization. One of these (no. 5) suggests that, the present work notwithstanding, there might be a greater gap between the problems than has been suspected.

Some preliminaries: For a graph $X = \langle V,E \rangle$, $\mathcal{V}(X)$ refers to the vertex set, V. The group of permutations of an n element set is denoted S_n, or, if the set A requires explication Sym(A). The subgroup generated by a set Φ is indicated by $\langle \Phi \rangle$. In all algorithms, input and output groups are assumed to be specified by a generating set (see §1.2 of [Lu1]).

2. Canonical forms

2.1 Canonical labeling-cosets. For a graph X with $\mathcal{V}(X) = V$ and $\sigma \in \mathrm{Sym}(V)$ we note by X^σ the graph obtained by joining u and v whenever $u^{\sigma^{-1}}$ and $v^{\sigma^{-1}}$ are adjacent in X. The graphs X and Y are *isomorphic*, denoted $X \cong Y$ iff $Y = X^\sigma$ for some σ. Let $\mathcal{K}$ denote a class of graphs, closed under isomorphisms, on a linearly ordered vertex set V (e.g. $V = \{1,2,\ldots,n\}$). A function $CF: \mathcal{K} \to \mathcal{K}$ is a *canonical form* for $\mathcal{K}$ if

(i) For X in $\mathcal{K}$, $CF(X) \cong X$

(ii) For X,Y in $\mathcal{K}$, $X \cong Y$ (if and only if)

$$CF(X) = CF(Y)$$

The definition can be extended to digraphs, possibly with colored vertices/edges and, more generally, to finite structures with any number of relations/operations.

It is convenient to generalize the notion of canonical forms as follows. Let G be a group acting on V. We say that X is G-*isomorphic* to Y, denoted $X \cong_G Y$, if $Y = X^\sigma$ for some $\sigma \in G$. Let $\mathcal{K}$ be a class of structures on V closed under G-isomorphisms, i.e.

(0) If $X \in \mathcal{K}$ and $\sigma \in G$ then $X^\sigma \in \mathcal{K}$.

We call a function $CF: \mathcal{K} \to \mathcal{K}$ a *canonical form with respect to* G if

(i) For X in $\mathcal{K}$, $CF(X) \cong_G X$.

(ii) For X,Y in $\mathcal{K}$, $X \cong_G Y$ (if and only if)

$$CF(X) = CF(Y)$$

We shall usually write CF(X,G) for such a form, reserving the notation CF(X) for CF(X, Sym(V)). Given some CF(-,G), there is a natural extension to cosets of G. If $\mathcal{K}$ is closed under $\langle \sigma, G \rangle$ where $\sigma \in \mathrm{Sym}(V)$, we define $CF(X,\sigma G)$, *the canonical form of* X *w.r.t.* σG, to be $CF(X^\sigma,G)$ It is important to observe that $CF(X,\sigma G)$ depends only on the coset σG and not on the choice of σ, for if $\sigma G = \tau G$ then $\tau^{-1}\sigma \in G$ so that $X^\sigma = X^{\tau(\tau^{-1}\sigma)} \cong_G X^\tau$, whence $CF(X^\sigma,G) = CF(X^\tau,G)$. A canonical form corresponds to a set of *labelings*, namely the renumberings of V which put X in canonical form

$$CL(X,\sigma G) = \{\tau \in \sigma G \mid X^\tau = CF(X,\sigma G)\}.$$

Clearly (I) $CL(X,\sigma G) = \sigma CL(X^\sigma,G)$

(II) $CL(X,\sigma G) = \tau \mathrm{Aut}_G(X^\tau)$ for any $\tau \in CL(X,\sigma G)$.

Here $\mathrm{Aut}_G(Y)$ denotes the group of G-automorphisms of Y. In particular, $CL(X,\sigma G)$ is a subcoset of σG which we call a *canonical labeling-coset of* X *w.r.t.* σG.

For purposes of recursion, it will be useful to have algorithms which return the full coset CL (although, for the structures we study, an oracle for CF, in fact an oracle for any complete set of invariants, i.e. for a certificate, could be used to construct CL.). We observe that properties I and II characterize canonical labeling-cosets. To be precise, let $\overline{G}$ be a subgroup of Sym(V) and suppose $\mathcal{K}$ is closed under $\overline{G}$-isomorphism. Denote by $\mathcal{C}$ the set of subcosets of $\overline{G}$. Then

Lemma 2.1 *Let* $CL: \mathcal{K} \times \mathcal{C} \to \mathcal{C}$ *be a function such that, if* X *is in* $\mathcal{K}$ *and* $\sigma \in \sigma G \in \mathcal{C}$, *then* $CL(X,\sigma G) \subseteq \sigma G$ *and* (I) and (II) *hold. Then*

$CF(X,\sigma G) = X^\tau$ *for any* $\tau \in CL(X,\sigma G)$ *defines a canonical form on* $\mathcal{K}$ *w.r.t. the subcoset* σG *and* CL *is the corresponding canonical labeling-coset.*

Proof: First, X^τ does not depend on the choice of τ. For, if $\tau, \mu \in CL(X,\sigma G)$ then $\sigma G = \tau G = \mu G$ so that (II) yields $\tau Aut_G(X^\tau) = \mu Aut_G(X^\mu)$; it follows that $\tau^{-1}\mu \in Aut_G(X^\tau)$ so that $X^\tau = (X^\tau)^{\tau^{-1}\mu} = X^\mu$. To show that the map $X \mapsto CF(X,G)$ is a canonical form w.r.t. G we need only verify $CF(X,G) = CF(Y,G)$ if, for $\rho \in G$, $Y = X^\rho$. But by (I), $CL(X,G) = CL(X,\rho G) = \rho CL(Y,G)$. Hence, if $\tau \in CL(Y,G)$ then $\rho\tau \in CL(X,G)$ and $CF(X,G) = X^{\rho\tau} = Y^\tau = CF(Y,G)$. Finally, the fact that $CF(X,\sigma G) = CF(X^\sigma,G)$ is immediate from (I). ▨

2.2 Strings and graphs. Let V be a linearly ordered set and Σ an alphabet. A *Σ-string* on V is a function $x : V \longrightarrow \Sigma$. The set of all Σ-strings on V is denoted by Σ^V. Strings can be regarded as particular structures with only unary relations. Thus, for $\sigma \in Sym(V)$ the string x^σ satisfies $x^\sigma(v) = x(v^{\sigma^{-1}})$. Canonical forms for strings are defined as above. Since we have found 'labeling' awkward in this setting, we shall refer to the subcoset of σG which maps x to $CF(x,\sigma G)$ as a *canonical placement-coset*, denoted $CP(x,\sigma G)$.

Whereas for graphs, etc., our basic problem is to find canonical forms w.r.t. $Sym(V)$, this problem becomes trivial for strings: the lexicographically first string obtained by reordering V will do it. However, the canonical form problem for graphs is easily reduced to the canonical form problem for strings with respect to a particular group. The adjacency matrix of an n-vertex graph (digraph, colored graph, etc.) is a string of length n^2 (indices ordered lexicographically) over a suitable alphabet. Then $\sigma \in Sym(V)$ acts on such strings via $a^\sigma(i,j) = a(i^{\sigma^{-1}}, j^{\sigma^{-1}})$.

Observation 2.2. *A canonical form for graphs w.r.t. $\sigma G \subseteq Sym(V)$ is precisely a canonical form for strings of length n^2 w.r.t. the induced action of σG.*

Switching back and forth between graphs and strings will enable us to combine geometric and algebraic ideas, each in their natural setting. Our basic tool is the string placement algorithm of §3.2, the canonicity of which is rigorously proved. When referring to combinatorial properties of graphs, we shall use procedures whose canonicity is intuitively clear and in fact has been used in the literature (cf. [CG] [We]). There are two sorts of particularly notable examples.

(1) Refinement procedures. The simplest example is the classification of the vertices of a graph X by their valence. Let V_i denote the set of vertices of valence i. Let the permutation $\sigma \in Sym(V)$ reorder the vertices by their valences: $i^\sigma \leq j^\sigma$ iff $deg(i) \leq deg(j)$. Let $H = Sym(V_0^\sigma) \times \cdots \times Sym(V_n^\sigma)$. Now we have reduced the canonization problem w.r.t. G to one w.r.t. H by setting

$$CL(X,G) = CL_1(X,\sigma H),$$

where CL_1 refers to any applicable canonical labeling. Note that, while σ is not canonical, the coset σH is.

(2) Individualization. Canonization with respect to G can be broken into a set of canonization problems with respect to cosets of the stabilizer subgroups G_v ($v \in V$). One can then take the lexicographically first of the resulting n canonical forms and recover the canonical coset as the union of those corresponding to the same canonical form.

While our references herein to the canonicity of such procedures will be informal, they can be justified directly from our definitions. Details of an algebraic machinery which encompasses such proofs will appear in the final paper.

3. String placement

3.1 Lexicographic placement. Suppose $G \subseteq Sym(A)$ and $x \in \Sigma^A$. A natural candidate for $CF(x,G)$ is the *lexicographic leader* in the G-orbit of x. As we remarked, this is easily computable when $G = Sym(A)$. Thus it is worth observing that the general problem of finding such lexicographic leaders is NP-hard. To see this, consider the interpretation of adjacency matrices as $\{0,1\}$-strings on $A(n) = \{(i,j) \mid 1 \leq i, j \leq n\}$. Order $A(n)$ so that $A(m)$, for $m < n$, (the upper left square) always precedes its complement. Then knowledge of the lexicographic leader w.r.t. the natural action of S_n would reveal the size of the largest independent set in the corresponding graph. (We remark that similar observations have been made by D. Corneil [Co] and G. Miller [Mi5]).

Nevertheless, the NP-hardness of the general problem is not, in itself, a deterrent to investigations for restricted groups. By way of analogy, we do not know how to test general string isomorphism subexponentially but the problem is in P for groups that turn up in the study of graphs of bounded valence [Lu1]. Thus, it is worth pointing out that the Lexicographic Leader problem remains NP-hard even for those groups, indeed even for a very restricted subclass.

Proposition 3.1 *The problem of finding the lexicographic leader in the* G*-orbit of* x *is* NP*-hard even if* G *is restricted to be an elementary abelian* 2*-group (every element has order* 2*).*

Proof: We reduce from 3-Dimensional Matching (see, e.g., [GJ]). Thus let $M \subseteq U \times V \times W$ be an instance of 3DM, that is, $|U| = |V| = |W|$ and we ask whether a subset of M projects bijectively onto U,V and W. Form the set, Q, consisting of unordered pairs, $\{m,m'\}$, of elements of M which overlap (i.e. have a common coordinate) and the set, P of ordered pairs, (m,m'), of elements of M which overlap. Fix any orderings of U,V,W,Q,P and form the ordered set A by taking, in sequence, $U_1V_1W_1Q_1PQ_2U_2V_2W_2$, wherein the subscript indicates an ordered copy of the respective set. For each $m = (u,v,w)$ in M, form the involution σ_m so that

(i) σ_m switches u_1 and u_2, v_1 and v_2, w_1 and w_2 (u_i is the image of u in U_i, etc.)

(ii) for each m' which overlaps m, σ_m switches $\{m,m'\}_1$ in Q_1 with (m,m') in P and switches (m',m) in P with $\{m,m'\}_2$ in Q_2.

Set $G = \langle \{\sigma_m\}_{m \in M} \rangle$. Finally, let x be the $\{0,1\}$-string on A which takes the value 1 on U_1,V_1,W_1,Q_2 and the value 0 on U_2,V_2, W_2,Q_1,P. Then, one checks that a matching exists in M if and only if the lexicographic leader in the G-orbit of x takes the value 0 on U_1,V_1, W_1, Q_1. ▨

Remarks. 1. Another proof, our initial one, employs the 2-group actions constructed by A. Lubiw [Lub] from instances of 3-SAT.

2. If the orbits on A are restricted to have length ≤ 2 (which, itself, forces the group to be an elementary abelian 2-group) there is a straightforward polynomial-time solution to the Lexicographic Leader problem. However, the above proof shows the problem becomes NP-hard for 2-groups if the orbits are allowed to have length ≤ 4. Another avenue of generalization appears difficult as well. For every $p > 2$, the problem is NP-hard for elementary abelian p-groups with orbits of length p.

3.2 A string canonization algorithm. We shall give an algorithm which computes canonical placement for Σ^A, with respect to any G. After proving it 'works' we discuss the timing for special G.

Some additional preliminaries: Both Σ^A and the collection of subsets of A inherit lexicographic orders. In particular, if $G \subseteq \mathrm{Sym}(A)$, it makes sense to refer to the *first* orbit of G. If G acts transitively on A there is also a 'first' minimal G-block system determined as follows. If G acts primitively the system is A itself. Otherwise, let a denote the first element in A. Find the first b in A, $b \neq a$ such that the G-invariant equivalent relation generated by $a \sim b$ is non-trivial. The induced partition is ordered and so the process may be repeated until the block system is minimal.

We now present the algorithm. To allow for recursion, we compute canonical placement-cosets w.r.t. σG for substrings induced on any G-invariant subset B of A, denoted, for convenience, $CP_B^X(\sigma G)$. Thus $CP(x,\sigma G)$ is $CP_A^X(\sigma G)$. We denote by x_B the restriction of x to B.

The algorithm -

INPUT: $x \in \Sigma^A$; a coset σG in Sym(A); a G-stable subset B of A.

OUTPUT: $CP_B^X(\sigma G)$, a subcoset of σG.

METHOD:

(1) If $|B| = 1$ then $CP_B^X(\sigma G) = \sigma G$

(2) If G is intransitive on B, let C be the first G-orbit in B, $B = C \cup D$. Then

$$CP_B^X(\sigma G) = CP_D^X\, CP_C^X(\sigma G)$$

(3) If G is transitive on B and $|B| > 1$, let H be the stabilizer of the blocks in the first minimal G-block system in B. Decompose σG as

$$\sigma G = \bigcup_{i=1}^{r} \sigma_i H$$

Say

$$CP_H^x(\sigma_i H) = \rho_i H_i$$

Reorder these cosets so that

$$(x^{\rho_1})_B = (x^{\rho_2})_B = \cdots = (x^{\rho_s})_B < (x^{\rho_{s+1}})_B \leq \cdots$$

(Choices of ρ_i's and of reordering will have to be justified). (Note that we only look at x^{ρ_i} on B). Then

$$CP_B^x(\sigma G) = \rho_1 \langle H_1, \{\rho_1^{-1}\rho_i\}_{1 \leq i \leq s} \rangle$$ ▨

Recalling Lemma 2.1, we prove

Lemma 3.2

(i) CP_B^x *is well defined*

(ii) *If* $(x^{\sigma})_B = y_B$ *then* $CP_B^x(\sigma G) = \sigma CP_B^y(G)$

(iii) *If* $\tau \in CP_B^x(\sigma G)$ *then*

$CP_B^x(\sigma G) = \{\mu \in \sigma G \mid (x^{\mu})_B = (x^{\tau})_B\}$;

equivalently, $CP_B^x(\sigma G) = \tau\, \mathrm{Aut}_G((x^{\tau})_B)$.

(*Note* $\mathrm{Aut}_G((x^{\tau})_B)$ *is a group since* B *is* G*-stable).*

Proof: We prove (i), (ii), (iii) simultaneously via a double induction on $|B|$, $|G|$. If either $|B|$ or $|G|$ is 1 then $CP_B^x(\sigma G) = \sigma G$ and (ii), (iii) are easy; (i) is no problem since we do not enter (3) in the algorithm. Suppose then that $|G| > 1$ and $|B| > 1$ and that (i) (ii), (iii) hold if either the subset is smaller or the subset is the same and the group is smaller.

Proof of (i): Choices are made only in (3). By (iii) for (H,B) the string $(x^{\rho_i})_B$ is unaffected by choice of ρ_i in $\rho_i H_i$ so the collection $\{\rho_i H_i\}_{i \leq s}$ is well defined. Also by (iii), $H_1 = \mathrm{Aut}_H((x^{\rho_i})_B) = H_i$ for $i \leq s$. But, since the groups $H_1, \ldots, H_s$ are identical, the output in (iii) is precisely the smallest subcoset of σG containing $\bigcup_{i \leq s} \rho_i H_i$ (Actually, a consequence of the lemma is that the output equals this union).

Proof of (ii): Assume $(x^{\sigma})_B = y_B$. Suppose first that the pair (G,B) sends us into case (2). Then, by induction for C,D, $CP_B^x(\sigma G) = CP_C^x CP_D^x(\sigma G) = CP_C^x(\sigma CP_D^y(G)) = \sigma CP_C^y CP_D^y(G) = \sigma CP_B^y(G)$ (note that $CP_D^y(G)$ is a group by (iii) for D). Suppose next that (G,B) sends us into case (3). Let $G = \bigcup \tau_i H$ so that $\sigma G = \bigcup \sigma\tau_i H$ and say $CP_B^y(\tau_i H) = \rho_i H_i$. Let $z_i = x^{\sigma\tau_i}$. Since $\tau_i \in G \subseteq \mathrm{Stab}(B)$, $(z_i)_B = (y^{\tau_i})_B$ so, as $|H| < |G|$,

$$CP_B^x(\sigma\tau_i H) = \sigma\tau_i CP_B^{z_i}(H) = \sigma CP_B^y(\tau_i H) = \sigma\rho_i H_i.$$

Since (i) has been established for (G,B) we may assume the strings $\{(y^{\rho_i})_B\}$ and $\{(x^{\sigma\rho_i})_B\}$ were those considered in processing $CP_B^y(G)$, $CP_B^x(\sigma G)$, respectively. But, as $\rho_i \in G \subseteq \mathrm{Stab}(B)$, $(x^{\sigma\rho_i})_B = (y^{\rho_i})_B$ so these are identical collections. Thus, $CP_B^y(G)$ and $CP_B^x(\sigma G)$ are the smallest subcosets of $G, \sigma G$ containing $\bigcup_{i \leq s} \rho_i H$ and $\bigcup_{i \leq s} \sigma\rho_i H_i$, respectively. Hence

$$CP_B^x(\sigma G) = \sigma CP_B^y(G).$$

Proof of (iii): The inclusion

$$\{\mu \in \sigma G \mid (x^{\mu})_B = (x^{\tau})_B\} \subseteq CP_B^x(\sigma G)$$

is derivable from (ii) (now established for (G,B)). To see this, set $y = x^{\tau}$ so that, as $\tau G = \sigma G$, $CP_B^x(\sigma G) = \tau CP_B^y(G)$. If also $(x^{\mu})_B = y_B$ for $\mu \in \sigma G$ then $CP_B^x(\sigma G) = \mu CP_B^y(G)$ and so $\mu = \mu\tau^{-1}\tau \in \mu\tau^{-1} CP_B^x(\sigma G) = CP_B^x(\sigma G)$. For the reverse inclusion, suppose first that (G,B) sends us into case 2. Let $\mu \in CP_B^x(\sigma G)$. We must show $(x^{\mu})_B = (x^{\tau})_B$. Since

$$\mu, \tau \in CP_B^x(\sigma G) = CP_D^x(CP_C^x(\sigma G)) \subseteq CP_C^x(\sigma G),$$

the induction hypothesis and (iii) for C yields $(x^{\mu})_C = (x^{\tau})_C$. Since $\mu, \tau \in CP_D^x(-)$, $(x^{\mu})_D = (x^{\tau})_D$ (induction for D). The result follows since $B = C \cup D$.

So suppose finally that (G,B) sends us to case 3. We need to show

$$\langle H_1, \{\rho_1^{-1}\rho_i\}_{i \leq s} \rangle \subseteq \mathrm{Aut}_G((x^{\rho_1})_B)$$

By induction on H

$$H_1 = \mathrm{Aut}_H((x^{\rho_1})_B) \subseteq \mathrm{Aut}_G((x^{\rho_1})_B)$$

So we need only worry about the elements $\rho_1^{-1}\rho_i$, $i \le s$. But the fact that $\rho_1^{-1}\rho_i \in G \subseteq \mathrm{Stab}(B)$ and $(x^{\rho_1})_B = (x^{\rho_i})_B$ puts $\rho_1^{-1}\rho_i$ in $\mathrm{Aut}_G((x^{\rho_1})_B)$. ▨

Hence, by Lemma 2.1, we conclude

Theorem 3.3 *The map* $x \longrightarrow x^\tau$, *where* $\tau \in CP_A^x(\sigma G)$, *is a canonical form for* Σ^A *w.r.t.* σG *and* $CP_A^x(\sigma G)$ *is the corresponding canonical placement-coset.*

3.3 Comment on timing. Good groups. Composition width. The group operations (including finding orbits, first minimal block system, stabilizer of block decomposition) require only polynomial (in $|A|$) time. Ignoring these, the decomposition in case (2) leads to a recurrence

$$t(|B|) \le t(|C|) + t(|B| - |C|)$$

for the timing. The bottleneck is in the passage in case (3) from a problem for (G,B) to [G:H] problems for (H,B). However, each of the latter problems decomposes into problems on disjoint orbits, each of size $\le |B|/m$ where m is the number of blocks in the first minimal block decomposition.

For a group G, let the *composition width* of G, denoted $cw(G)$, be the smallest positive integer d such that every nonabelian composition factor of G embeds in the symmetric group S_d. (For solvable groups, $cw(G) = 1$). Standard arguments show $cw(H) \le cw(G)$ if H is a subgroup or a homomorphic image of G. For a reason to become obvious soon, we call a class of groups *good* if the composition widths of its members are bounded.

The following result shows how the timing of our algorithm is controlled by $cw(G)$.

Theorem 3.4 (Babai, Cameron, Pálfy [BCP]). *If* G *is a primitive permutation group of degree* n *and* $cw(G) \le d$ *then* $|G| \le n^{\omega(d)}$. *(See (*) below)*

For $d = 1$ (solvable groups), Pálfy [Pa] proves $\omega(d) < 3.4$.

It is implicit in [BCP , p. 162, ℓ. 9-11] that $\omega(d) < 2 + \log(da(d))$ where $a(d) = \max\{|\mathrm{Aut}\ H| \mid H$ is a simple subgroup of $\mathrm{Sym}(d)\}$. Using consequences of the classification of finite simple groups [Ca2] we obtain $a(d) \le d!$ for sufficiently large d, hence

$$(*) \qquad \omega(d) < d \log d + c.$$

One can avoid use of the classification, invoking more elementary group theory [Ba4], [Ba5] to prove $\omega(d) < cd \log^4 d$ (cf. [Ba3]).

By Theorem 3.4, case (3) of the algorithm yields a recurrence of the form

$$t(|B|) \le m^{\omega(d)+1}\ t(|B|/m).$$

We have then

Theorem 3.5 *The canonical placement algorithm for* Σ^A *w.r.t.* G *runs in time* $O(n^{\omega(d)+c})$ *where* $n = |A|$ *and* $d = cw(G)$. *In particular, the algorithm runs in polynomial time if we consider good groups (bounded* cw).

An immediate application (cf. Observation 2.2) is

Corollary 3.6 *If* $\mathcal{K}$ *is a class of (possibly colored, directed) graphs on a vertex set* V, *closed under isomorphisms by* $G \subseteq \mathrm{Sym}(V)$, *then a canonical form and corresponding canonical labeling-coset w.r.t.* G *for* X *in* $\mathcal{K}$ *can be found in* $n^{\omega(d)+c}$ *time.*

3.4 Lexicographic placement revisited. In view of Proposition 3.1, it is worth noting that there is a sense in which efficient lexicographic placement is now available for good groups. To be precise, Z. Galil [Ga] pointed to an interpretation of the canonization algorithm as lexicographic placement relative to an easily determined reordering. Galil's suggestion develops into a striking counterpoint to Proposition 3.1.

Proposition 3.7 *Let* A, Σ *be linearly ordered,* $G \subseteq \mathrm{Sym}(A)$. *There is a canonical reordering of* A *relative to which the lexicographic leader problem for every* x *in* Σ^A *w.r.t.* G *is solvable in* $|A|^{\omega(d)+c}$ *time,* $d = cw(G)$. *Furthermore, the reordering can be determined in polynomial time.*

Outline of proof: The essential idea is that one can create, in a canonical fashion from G and A, a tree, $T = \mathrm{TREE}(G,A)$, of subsets through which the recursion will always descend, so that

(i) The leaf set is A

(ii) For any mode B, the stabilizer of B in G acts trivially or primitively on the sons.

The tree is then laid out so that the sons of any mode appear, left to right, in increasing order. The reordering is obtained by numbering the entire leaf set, left to right.

In the earlier algorithm, the decomposition of a subset B was guided by the action of the subgroup at hand, denoted now by G^*. Now, in computing

$CP^x_B(\sigma G)$ at a node B, we replace (2) and (3) of the algorithm by

(2') Let H be the stabilizer in G^* of the sons $B_1, \ldots, B_m$ (listed in order). Let $\sigma G^* = \bigcup \tau_i H$ and determine, for each i,

$$\rho_i H_i = CP^x_{B_m} \cdots CP^x_{B_1}(\tau_i H)$$

Proceed with $\{\rho_i H_i\}$ as in (3).

The output for $CP^x_A(G)$ this time is always the lexicographic leader in the G-orbit of x (relative to the reordered A). The idea now is that, for $i < j$, all the points in B_i precede all the points in B_j; thus lexicographic placement of x on B is achievable by lexicographically placing on B_1, then B_2, etc.

For the timing, the crucial observation is that, either $H = G^*$, so there's only one τ_i or else the B_i all have size $|B|/m$ and $|G^*/H| \leq$ the order of a primitive group acting on $\{B_1, \ldots, B_m\} \leq m^{\omega(d)}$.

There are several reasonable choices for $T = TREE(G,A)$. One such is analogous to the 'structure tree' construction of [GHLSW, Theorem 1] (and is useful in extending the tricks of that paper to speed up trivalent canonization). If G acts intransitively on A, $TREE(G,A)$ is the union, joined to a new root, of $\{TREE(G,A_i)\}$ where $\{A_i\}$ is the set of orbits. If G acts transitively, let $\{A_i\}$ be the first minimal block system, A_1 the first block and let $T' = TREE(G_1,A_1)$, where G_1 is the stabilizer in G of A_1; choose any $\{\sigma_i\}$ so that $\sigma_i(A_1) = A_i$; then $TREE(G,A)$ is the union of $\{\sigma_i(T')\}$.

4. Applications to graphs.

4.1 Tournaments. It is convenient to use the language of round-robin tournaments with no draws. The players are the vertices of the tournament. Each pair of players play exactly once. An arrow from v to w indicates that v beat w. We show

Theorem 4.1 *Canonical forms for tournaments, T, can be computed in $n^{c \log n}$ time, where $n = |V(T)|$ and $c = \frac{1}{2} + o(1)$ (logarithms are taken base 2).*

It is well known that the automorphism groups of tournaments have odd order (an involution would reverse an arrow). Thus, by the Odd Order Theorem of Feit and Thompson [FT], tournaments have solvable automorphism groups. However, it requires some effort to force the appearance of these groups in a string setting.

Let T be a tournament on the vertex set V. We seek a canonical labeling-coset for T w.r.t. $Sym(V)$. If the tournament is not regular (i.e. if there are vertices with different out-valences) we can reduce the group to $\bigtimes_{i \geq 0} Sym(V_i)$, where V_i is the set of vertices of out-valence i. Then, denoting the induced tournament on V_i by T_i, we find, recursively,

$$CL_{tour}(T_i, Sym(V_i)) = \rho_i H_i \quad \text{for } i \geq 0.$$

So we may let

$$CL_{tour}(T, Sym(V)) = CL_{Cor.3.6}(T, \bigtimes_i \rho_i H_i).$$

Note that we are dealing here with a coset of the good group $\bigtimes_i Aut(T_i)$.

Suppose, then, the tournament is regular so that each vertex has out-valence $(n-1)/2$. Here we use the individualization process, fixing v and finding the canonical labeling-cosets w.r.t. the cosets of $Sym(V')$, where $V' = V - \{v\}$. Viewing $CL(T,\sigma Sym(V'))$ as $\sigma\, CL(T^\sigma, Sym(V'))$, we have n problems for regular tournaments T' on V w.r.t. $Sym(V')$ (which fixes v). In such a case, V' immediately splits in half, $V' = V'_1 \cup V'_2$, vertices being assigned to V'_1 or V'_2 according to whether they beat or are beaten by v. Thus, again we can replace the group $Sym(V')$ by $Sym(V'_1) \times Sym(V'_2)$, find, recursively

$$CL_{tour}(T'_i, Sym(V'_i)) = \rho_i H_i \quad \text{for } i = 1,2$$

where T'_i is the induced tournament on V'_i, and let

$$CL_{tour}(T', Sym(V')) = CL_{Cor\ 3.6}(T', \rho_1 H_1 \times \rho_2 H_2).$$

The non-regular case leads to a timing inequality

$$t(n) \leq \Sigma_i t(n_i) + n^c \text{ where } n_i = |V_i|$$

and the regular case to

$$t(n) \leq n(2t(\tfrac{n-1}{2}) + n^c).$$

The proposition follows.

4.2 Bipartite graphs. We consider the problem of finding canonical forms for a bipartite graph with respect to a group action on one of the sides. We describe an algorithm whose complexity is sensitive to the valence on that side and to the composition width of the group. The algorithm will serve as a subroutine, in the next sections, for extending canonical labeling-cosets through a nested sequence of subgraphs. Thus, we have a set $A = B \dot\cup C$, a coset σG acting on B, the symmetric group $Sym(C)$ acting on C and a bipartite $X = (A,E)$ graph with edge set $E \subseteq B \times C$. Let d_{out} denote the maximum valence of vertices of B and d_{in} the maximum valence of vertices in C. (We think of the edges being oriented from B to C).

In order to find some $CL(X,\sigma G \times Sym(C))$, the first naive approach is to adopt Luks' idea [Lu1,§3.1] to represent the vertices of C by their neighborhoods in B. Let $[B]^{d_{in}}$ denote the set of subsets of B of size $\le d_{in}$. The ordering of B induces an ordering of $[B]^{d_{in}}$. Let $f : [B]^{d_{in}} \to \{0,1,\ldots,d_{out}\}$ associate with each $Y \in [B]^{d_{in}}$ the number $f(Y)$ of those vertices in C whose neighborhood is precisely Y. Now f is a string which we have to canonically place with respect to the induced σG-action on $[B]^{d_{in}}$. The subcoset $CP(f,\sigma G) = \overline{\sigma}\,\overline{G}$ then easily extends to a subcoset $CL(X,\sigma G \times Sym(C)) = \Psi H$ where $\Psi H\big|_B = \overline{\sigma}\,\overline{G}$.

The timing of this procedure depends on that for CP. It runs in polynomial time if $cw(G)$ and d_{in} are bounded. In particular, it suffices for polynomial-time canonization of graphs of bounded valence (next subsection). An unsatisfying feature, however, is the blow-up in the problem size, which multiples the exponent in the running time by a factor of d_{in}.

We are aware, at present, of at least four tricks which avoid this blow-up. Each was originally designed to improve the running time of Luks' isomorphism test from essentially n^{cd^2} to n^{cd} where d is the valence. Each is sufficient to improve the Zemlyachenko-Luks bound, $\exp(n^{2/3})$, for general graph isomorphism [ZKT],[Ba3] to $\exp(n^{\frac{1}{2}})$. We give a brief account here of these ideas, leaving the details for the full paper.

1. The first trick involves a modification of the string placement algorithm to capitalize on the sparceness of the strings arising in the above construction (one letter predominates). It is an analogue of the Schnorr-Weber [SW] (see also [GHLSW]) speedup of the string isomorphism (≡ color isomorphism) algorithm. First one expands the terminating case (1) to

(1') If x^σ is constant on B, $CP^x_B(\sigma G) = \sigma G$.

Secondly, in the transitive case, one first places the blocks themselves according to the vectors which indicate the number of occurences of each α in Σ. Thus one only comes to the original case (3) when these numbers are the same for each B_i. The timing can be expressed as $m^{\omega(d)}|A|^c$ where m is the number of occurences in x of the second most frequent letter.

2. The second trick, due to Luks, does not reduce the problem to string placement immediately but adapts the ideas therein to split the set B directly. The base case ($|B| = 1$) is then a placement problem for a subset of size d_{out}, which has a naive solution in $n^{d_{out}+c}$ steps (this can be improved to $4^{d_{out}} n^c$). Thus one only requires an additive d_{out} term in the running time. Details will appear in [Lu2]. We remark that this algorithm finds, more generally, $CL(X,\sigma H)$, for X as above and $H \subseteq G \times Sym(C)$. The timing involves only $cw(G)$ and d_{out}.

3. A third trick removes the dependence on d_{out} in computing $CL(X,\sigma G \times Sym(C))$. In [Mi4], G. Miller succeeded in determining (in our notation) $Aut(X) \cap (\sigma G \times Sym(C))$ in polynomial time for G of bounded composition width, irrespective of the valences in X. His method is adaptable to produce a suitable CL and his results suggest broader applications.

4. We have chosen, for its simplicity, to describe the details of a fourth trick. Due to Babai, it retains the dependence on d_{out} (it enters in our applications anyway since it affects the composition width of the output).

We may assume X has no isolated vertices in C. First we consider the string

$x : B \longrightarrow \{0,1,\ldots,d_{out}\}$ where $x(v)$ is the valence of v. Compute $CP(x,\sigma G) = \rho H$ and let $CL(X,\sigma G \times Sym(C)) = CL(X,\rho H \times Sym(C))$. Note that H preserves valences of X^{ρ} in B so we can extend the action of ρH to E allowing arbitrary permutations of edges with common orgin. The coset obtained, $\hat{\rho}\hat{H}$, is the largest subcoset of $Sym(B \times C)$ which maps E to E^{ρ}, respects the blocks of edges having common origin and restricts to ρH on these origins. The kernel of the epimorphism $\hat{H} \longrightarrow H$ (projection) is the direct product of groups $S_{x(b)}$ for $b \in B$. Hence $cw(\hat{H}) \leq \max(d_{out}, cw(G))$.

Our next step is to consider the string $y : E^{\rho} \times E^{\rho} \longrightarrow \{0,1\}$ where $y(e_1,e_2) = 1$ iff e_1 and e_2 terminate at the same vertex in C. We order $E^{\rho} \subseteq B \times C$ lexicographically, and obtain $CP(y,\hat{H}) = \tau K$ with respect to the lexicographic order of $E^{\rho} \times E^{\rho}$ and the induced $\hat{H}$ action. Now τK is a subcoset of $\hat{H}$. Let $\psi : C \to C$ denote the permutation defined by $u^{\psi} < v^{\psi}$ iff $F(u) < F(v)$ where $F(u) = \min\{e \mid e \in E^{\rho\tau}, e \text{ is incident with } u\}$. Let ϕ denote the permutation induced by τ on B; let $\pi = (\rho\phi,\psi)$ act on $B \cup C$. The group K can be viewed as acting on $B \cup C$; clearly $K = Aut(X^{\pi}) \cap (G \times Sym(C))$. Setting $CL(X,\sigma G \times Sym(C)) = \pi K$ we obtain the desired canonical labeling.

Since the composition widths of G and $\hat{H}$ are bounded by $d = \max(d_{out}, cw(G))$, the total running time is $O(n^{\omega(d)+c})$, $n = |A|$.

4.3 Graphs of valence $\leq d$. We show

Theorem 4.2. *Canonical forms for graphs, X, can be computed in* $O(n^{\omega(d-1)+c})$ *steps where* $n = |v(x)|$, $d =$ valence (X).

It suffices to canonize connected graphs, for $CF(X)$ can be the canonized components taken in lexicographic order of adjacency matrices and $CL(X)$ is easily constructible from the canonical labeling-cosets of the components. We observe, next, that it suffices to canonize connected graphs, X, with an edge e individualized, denoted $X(e)$. The motivation for edge individualization is the significant effect it has on the automorphism group. The complete group, even in the trivalent case, is unrestricted. On the other hand, the following was proved in [Lu1].

Lemma 4.3. *Let* X *be a connected graph of valence* $\leq d$ *and let* e *be an edge in* X. *Then the composition factors of* $Aut(X(e))$ *are subgroups of* S_{d-1}. *In particular,* $cw(Aut(X(e))) \leq d-1$.

The proof of Lemma 4.3 depended upon the observation that the kernels of the homomorphisms

$$\pi_r : Aut(X_{r+1}(e)) \longrightarrow Aut(X_r(e))$$

are direct products of symmetric groups, S_t for $t \leq d-1$; herein X_r is the subgraph consisting of all vertices and edges lying on paths of length $\leq r$ through e and π_r is induced by restriction. We note that this property of kernel (π_r) is a consequence of the boundedness of the <u>outvalence</u> in the induced bipartite graph on $V_r \times V_{r+1}$ where $V_r = \mathcal{V}(X_r) \setminus \mathcal{V}(X_{r-1})$; neither the invalence nor the valences within the induced graph on V_r are involved. With this in mind, we define the <u>outvalence</u> of $X(e)$ to be the maximum over r of the induced outvalences on $V_r \times V_{r+1}$ and note

Lemma 4.4 *The conclusion of* Lemma 4.3 *holds under the assumption of outvalence* $(X(e)) \leq d-1$.

Henceforth, we weaken the valence assumption on the connected graph X to: there exists at least one edge e such that outvalence $(X(e)) \leq d-1$.

In the canonization of $X(e)$ we may cut immediately from the group $Sym(\mathcal{V}(X))$ to K_{n-1} where $K_r = Sym(V_1) \times Sym(V_2) \times \cdots \times Sym(V_r)$.

We proceed, inductively, through the X_r. Assume we have defined $CF(X_r,K_r)$ and determined the corresponding $CL(X_r,K_r) = \sigma G$. Then $CL(X_{r+1},K_{r+1})$ is determined in two steps. First, let Y_r be the induced subgraph of X on $\mathcal{V}(X_r)$ $(= X_r$ together with the edges between vertices in $V_r)$. We canonize Y_r w.r.t. σG by taking

$$CL_{cor\ 3.6}(Y_r,\sigma G) = \rho H.$$

Next, let Z_r denote the bipartite graph induced between $\mathcal{V}(X_r)$ and V_{r+1}. Using the 'bipartite' algorithm of subsection 4.2, we let

$$CL(X_{r+1},K_{r+1}) = CL_{bip}(Z_r, \rho H \times Sym(V_{r+1})).$$

The resulting coset may be viewed as a subcoset of $\mathrm{Sym}(V(X_{r+1}))$.

Since G, and therefore H, has composition width $\leq d-1$, the total time to compute $CL(X_{n-1}, K_{n-1})$ is $O(n^{\omega(d-1)+c})$.

We remark that the above can be improved to $O(n^{cd/\log d})$ using other techniques introduced in [Lu2].

4.4. General graphs. Zemlyachenko's trick. We obtain

Theorem 4.3 *Canonical forms for graphs, X, can be computed in* $\exp(n^{\frac{1}{2}+o(1)})$ *time, where* $n = |\mathcal{V}(X)|$.

The link to general graph isomorphism is the remarkable Valence Reduction Lemma of Zemlyachenko. To set the stage, let $X = (V,E)$ be a vertex colored graph, that is, there is a coloring function f from V into an initial segment of $\{1,2,3,\ldots\}$. We denote the color class $f^{-1}(i)$ by C_i. We say that X has *color valence* $\leq d$ *w.r.t.* f if, for every v and i, either the the number of neighbors or the number of non-neighbors of v in C_i is $\leq d$. For the purposes of isomorphism-testing or canonization of vertex-colored graphs, it is often useful to recolor the vertices according to the familiar naive refinement procedure ([CG], [Ba3]), so that the number of neighbors in C_i of a vertex in C_j is a function of i and j alone. The Valence Reduction Lemma states

Proposition 4.4 (Zemlyachenko [ZKT], [Ba3]) *Let X be a vertex colored graph with* $|\mathcal{V}(X)| = n$. *and suppose* $d \leq n$. *Then there is a sequence of* $k < 4n/d$ *vertices such that the assignment of* k *new colors to these (individualization) followed by naive refinement results in a graph with color-valence* $\leq d$.

To this we add an extension of Theorem 4.2

Proposition 4.5 *Canonical forms for vertex-colored graphs, X, can be computed in* $O(n^{\omega(d)+c})$ *steps where* $n = |\mathcal{V}(X)|$ *if color-valence* $(X) \leq d$.

Given these results, canonical forms for general graphs can be obtained by individualizing all sequences of 4n/d vertices, canonizing the resulting graphs with color-valence $\leq d$, then taking those with lexicographically least adjacency matrix. So, if $d = \sqrt{n}$, we perform an $\exp(n^{\frac{1}{2}+o(1)})$ step procedure, $\exp(n^{\frac{1}{2}+o(1)})$ times.

We comment briefly on the proof of Proposition 4.5. It would be a straightforward extension of the results of subsection 4.3 if the bound actually involved valences, not valence-or-covalence; if so, one forms the nested sequence $\{X_r\}$ of subgraphs in which each successive level adds accessible vertices from just one color class. The final trick, then, involves reduction to this situation. For this, one simply switches edges and non-edges between C_i and C_j if it was the covalence which was small. This brings the valences down and may have the (harmless) side effect of disconnecting the graph. A CL for the modified graph, X', will be a CL for X. Note that non-isomorphic X may yield isomorphic modified graphs, X', at this stage which would have the same CL(X'). However, the corresponding canonical forms, $X^{CL(X')}$, would not be identical.

4.5 Designs.

We first consider balanced incomplete block designs [Ry] with parameters (v,k,λ): v is the number of vertices, k the size of each block and λ the number of blocks common to each pair of vertices. (The other commonly used parameters, b and r, are functions of these). We assume $3 \leq k < v$ and $\lambda \geq 1$, thereby excluding the trivial cases. We show

Theorem 4.6 *Canonical forms for block-designs with parameters* (v,k,λ) *can be computed in* $v^{f(k,\lambda)} + \log v$ *time.*

Our estimate for $f(k,\lambda)$ is $c + \omega(\max(k-2,\lambda))$.

It is known that the isomorphism problem for block designs is isomorphism complete, even for triple systems (k = 3) [CC]. On the other hand, Miller [Mi2] has shown that isomorphism testing, in fact canonical labeling, can be done in $n^{\log n}$ for Steiner triple systems ($k = 3$, $\lambda = 1$). The reason is that a Steiner triple system can be viewed as a quasigroup and therefore has a set of $\leq 1 + \log n$ generators. Having individualized these, one can canonically order the remaining vertices in polynomial time. The choice of generators has to be repeated at most $n^{\log n}$ times and the lexicographically least of the resulting multiplication tables is selected to be canonical.

We combine this idea with information about the automorphism groups. To establish the existence of a 'small' generating set one shows that any *subdesign* $Y = (W,\mathcal{C})$ of $X = (V,\mathcal{B})$ (i.e. $W \subseteq V$, $\mathcal{C} \subseteq \mathcal{B}$ and Y is a block design with parameters $(|W|, k, \lambda)$) satisfies $|W| \leq (v-1)/(k-1)$. Since the set of subdesigns is closed under intersection, any subset 'generates' a subdesign. So

Lemma 4.7 X *has a generating set of size* $\leq 1 + \log v/\log(k-1)$.

Unfortunately, unlike the Steiner triple system case, the stabilizer of a set S of generators of X in $\mathrm{Aut}(X)$ is not necessarily the identity. However, one shows

Lemma 4.8 *The composition factors of* $\mathrm{Aut}_S(X)$ *are subgroups of* S_d *where* $d = \max(\lambda, k-2)$.

We employ this in an extension process analogous to the one in §4.3. (The log v term in the exponent is due to the number of choices of S). For a sequence $S = (u_1,\ldots,u_s)$ we build a chain $\{W_i\}$ of subsets of V by: $W_1 = \{u_1\}$ and while $W_i \neq V$, if W_i induces a subdesign then $W_{i+1} = W_i \cup \{\text{first } u_j \text{ not in } W_i\}$ else $W_{i+1} = W_i \cup \{B \in \mathcal{B} \mid |B \cap W_i| \geq 2\}$. Then the nested graphs $\{X_j\}$ are taken to be bipartite, X_{2i-1} and X_{2i} both have the set W_i on one side, the vertices on the other side represent those blocks entirely in W_i (for X_{2i-1}) or those in W_{i+1} (for X_{2i}). Edges correspond to incidence. The extension of $CL(X_j)$ to $CL(X_{j+1})$ works as in §4.3 for j even. For j odd we do not have a bound on d_{out} (this wouldn't bother us if we had developed Miller's trick in §4.2) but we get around this by considering another bipartite graph Y_j, having the set of unordered pairs of elements of $\mathcal{V}(X_j)$ on the left and $\mathcal{V}(X_{j+1}) \setminus \mathcal{V}(X_j)$ on the right. A pair $\{x,x'\}$ will be adjacent to a block B on the right if $x,x' \in B$. Now the vertices on the left side of Y_j have degree $\leq k-2$, justifying the timing. ▨

We turn next to *symmetric designs*, i.e. we suppose the number of points equals the number of blocks. If $\lambda = 1$, these are the *projective planes*. Miller [Mi2] showed that canonical forms for projective planes can be computed in $n^{\log \log n}$ time. Using ideas somewhat similar to the above (although note that *subdesigns* now refer to *subplanes* $(W,\mathcal{B}_W) \subseteq (V,\mathcal{B})$ where

$$\mathcal{B}_W = \{B \cap W \mid B \in \mathcal{B},\ |B \cap W| \geq 2\}$$

we establish

Theorem 4.9 *Canonical forms for symmetric* (v,k,λ)*-designs can be found in* $v^{\omega(\lambda)+\log\log v + c}$ *time.*

Remark. As far as we know, no infinite family of such designs is known for any $\lambda \geq 2$.

We remark finally that similar ideas can be used to find canonical forms for strongly regular graphs (cf [Ca1]) with parameters (v,k,λ,μ) in $v^{c \log v + \omega(\max(\lambda,\mu))}$ time. Again, the applicability may be limited because it appears to be an open question whether there exist an infinite number of connected strongly regular graphs with bounded λ,μ. It is conceivable, however, that for small k it might improve Babai's bound $\exp(cv \log^2 v/k)$ for $k \leq n/2$ [Ba2] (Note: $k \geq \sqrt{v}$).

5. An alternative moderately exponential graph canonization.

There is now an $\exp(n^{2/3+o(1)})$ graph canonization algorithm available which does not use any group theory except for the "tower of groups" algorithm [Ba1],[FHL]. The method starts with a Zemlyachenko valence reduction to valence $\leq n^{1/3}$ at the cost of individualizing $\leq 4n^{2/3}$ vertices.

The next step uses the following result:

Theorem 5.1 *Let* X *be a connected graph on* n *vertices of valence* $\leq cn^{2/3}$. *Then there exists a set* S *of* $O(n^{2/3}\log n)$ *vertices such that by individualizing the vertices in* S *and applying the Weisfeiler-Lehman edge-refinement* [We] *the vertex set breaks into color-classes of size* $O(n^{2/3})$.

The proof rests on estimates for distinguishing sets in coherent configurations in the spirit of [Ba4].

The concluding step is the [KL] version of [Ba1]: canonical forms for graphs with bounded color-classes. The cost of this third step can actually be reduced to $\exp(n^{1/3 + o(1)})$, using [Ba6], leaving the entire algorithm with only

two $\exp(n^{2/3})$ bottlenecks.

This result appears to indicate that coherent configurations and other combinatorial techniques might be relevant in the search for improved complexity estimates.

6. Problems and Comments

1. Can hypergraph isomorphism (respectively, canonization) be determined in simply exponential, c^n, time where n is the size of the vertex set? Note that the input itself can be exponential in n, so we can not expect any better. If the hypergraphs have bounded rank (= the maximum cardinality of an edge) then isomorphism is decidable in c^n time ([Lu2]) and the technique extends to canonization. (The result makes essential use of the simple groups classification). Are there moderately exponential methods for this class? We observe that it is possible to reduce 3-hypergraph, and even 4-hypergraph, isomorphism (respectively, canonization) to graph isomorphism (respectively, canonization) on an n^2 element set. Hence, a moderately exponential algorithm for 4-hypergraph isomorphism is a necessary condition for the reduction of graph isomorphism to $\exp(n^{\frac{1}{2}-\varepsilon})$ for some $\varepsilon > 0$.

2. Subset stabilizers for arbitrary permutation groups can be computed in $\exp(n^{\frac{1}{2}} + o(1))$ time [Ba6]. They can also be computed in $4^d n^c$ time, where d is the size of the subset, c an absolute constant [Lu2]. Both of these results have canonical placement analogues with the same time bounds (for the latter, see [BKL]). Is there a common generalization?

3. We indicated that, with respect to a certain natural ordering of the indices, the problem of finding the lexicographic leader among the possible adjacency matrices of a graph is NP-hard. Is this the case with respect to the usual lexicographic ordering of the indices? We conjecture that the problem is NP-hard with respect to any predetermined ordering of the indices. (In this regard, however, compare Proposition 3.7). We further conjecture that the problem remains NP-hard even for special classes of graphs, e.g. trivalent graphs, trees. We are split over a prediction for binary trees.

4. Blass and Gurevich [BG] constructed a polynomial-time recognizable equivalence over strings for which determining the kth digit of the lexicographic leader of a class is Δ_2^p-complete. Recall, (Δ_2^p is the class of languages recognizable in polynomial time using an oracle for an NP-set). Is this still the case for the equivalence defined by a permutation group action? By a 2-group action?

5. We point out a situation where a significant complexity gap between isomorphism testing and canonization remains. Consider a class $\mathcal{C}$ of 'good' graphs on a vertex set V, e.g. graphs of bounded valence, tournaments, graphs of bounded eigenvalue multiplicity, etc. and an arbitrary group $G \subseteq \mathrm{Sym}(V)$. The group intersection algorithm of [Lu1,§4] shows that the isomorphisms from X to $X' \in \mathcal{C}$ lying in G can be computed in essentially the time currently required for testing isomorphism (i.e. in $\mathrm{Sym}(V)$). However, the methods do not yet seem to extend to finding canonical forms for 'good' graphs, $CF(X,G)$, with respect to arbitrary groups, G. We do not know such a CF even for the class of binary trees. (Cf Corollary 3.6 where we show an answer for good groups in arbitrary graphs).

6. The significance of a canonical form in mathematics is, very often, its simplicity and transparent structure. Although that is not the motivation for studies in the computational complexity of graph canonization, the question remains whether the canonized graphs constructed herein have any noteworthy combinatorial structure. One of us thinks it would be worthwhile to investigate the matter.

REFERENCES

[Ba1] L. Babai, *Monte-Carlo algorithms in graph isomorphism testing*, preprint, Univ. Montréal (1979).

[Ba2] L. Babai, *On the complexity of canonical labeling of strongly regular graphs*, SIAM J. Comp. 9 (1980), 212-216.

[Ba3] L. Babai, *Moderately exponential bound for graph isomorphism*, Proc. Conf. FCT '81 Szeged, Lecture Notes in Computer Science 117, Springer 1981, 34-50.

[Ba4] L. Babai, *On the order of uniprimitive permutation groups*, Ann. of Math 113 (1981), 553-568.

[Ba5] L. Babai, *On the order of doubly transitive permutation groups,* Inventiones Math 65, 473-484.

[Ba6] L. Babai, *Permutation group intersection in* $\exp(n^{\frac{1}{2}+o(1)})$ *time,* to appear.

[BCP] L. Babai, P. Cameron, P. Pàlfy, *On the orders of primitive groups with restricted nonabelian composition factors,* J. Alg., 79 (1982), 161-168.

[BGM] L. Babai, D. Grigoriev, D. Mount, *Isomorphism of directed graphs with bounded eigenvalue multiplicity,* Proc. 14th ACM Symp Thy Comp (1982), 310-324.

[BG] A. Blass and Y. Gurevich, *Equivalence relations, invariants and normal forms,* to appear.

[BKL] L. Babai, P. Klingsberg, E. Luks, *Canonical labeling for vertex colored graphs,* to appear.

[Ca1] P. Cameron, Strongly regular graphs, Selected Topics in Graph Theory, ed. L. Beineke, R. Wilson, Academic Press, (1979).

[Ca2] P. Cameron, *Finite permutation groups and finite simple groups,* Bull. London Math. Soc. 13 (1981), 1-22.

[Co] D. Corneil, Private communication.

[CG] D. Corneil and M. Golberg, *On graph certificates,* Congressus Num. 35 (1982), 181-190.

[CC] C. Colbourne and M. Colbourne, *The complexity of combinatorial isomorphism problems,* Proc. Can.-France Comb. Coll., Montreal (1979).

[FT] W. Feit and J. Thompson, *Solvability of groups of odd order,* Pac. J. Math. 13 (1983), 775-1029.

[FM] S. Filotti and J. Mayer, *A polynomial-time algorithm for determining the isomorphism of graphs of fixed genus,* Proc. 12th ACM Symp Thy Comp (1980), 236-243.

[FHL] M. Furst, J. Hopcroft, E. Luks, *Polynomial-time algorithms for permutation groups,* 21st IEEE Symp. Found. Comp. Sci. (1980), 36-41.

[Ga] Z. Galil, private communication.

[GHLSW] Z. Galil, C. Hoffman, E. Luks, C. Schnorr, A. Weber, *An* $O(n^3 \log n)$ *deterministic an* $O(n^3)$ *probabilistic isomorphism test for trivalent graphs,* 23rd IEEE Symp. Found. Comp. Sci. (1982), 118-125.

[GJ] M. Garey and D. Johnson, Computers and Intractibility: A guide to the theory of NP-completeness, Freeman, San Francisco (1979).

[HP] Z. Hedrlin and P. Pultr, *On full embeddings of categories of algebras,* Ill. J. Math. 10(1966), 392-406.

[HT] J. Hopcroft and R. Tarjan, *Isomorphism of planar graphs* (working paper), *Complexity of Computer Computations,* Plenum (1972), 131-152.

[KL] P. Klingsberg, E. Luks, *Succinct certificates for a class of graphs,* St. Joseph's Univ. preprint (1981) (See [BKL]).

[Lip] R. Lipton, *The beacon set approach to graph isomorphism,* Yale Dept. Comp Sci. preprint #135 (1978).

[Lu1] E. Luks, *Isomorphism of graphs of bounded valence can be tested in polynomial time,* J. Comp. Sys. Sci. 25 (1982), 42-65.

[Lu2] E. Luks, *On the complexity of fixed valence groph isomorphism and the implications for general graph isomorphism,* to appear.

[Lub] A. Lubiw, *Some NP-complete problems similar to graph isomorphism,* SIAM J. Comp, 10 (1980), 11-21.

[Ma] R. Mathon, *A note on the graph isomorphism counting problem,* Inf. Proc. Let. 8 (1979), 131-132.

[Mi1] G. Miller, *Graph isomorphism, general remarks,* J. Comp. Sys. Sci., 18 (1979), 128-142.

[Mi2] G. Miller, *On the* $n^{\log n}$ *isomorphism technique,* Proc. 10th ACM Symp Thy. Conp. (1978), 51-58.

[Mi3] G. Miller, *Isomorphism testing for graphs of bounded genus,* Proc. 12th ACM Symp Thy Comp. (1980), 225-235.

[Mi4] G. Miller, *Isomorphism of graphs which are pairwise* k*-separable,* to appear.

[Mi5] G. Miller, private communication.

[Pa] P. Pálfy, *A polynomial bound for the orders of primitive solvable groups,* J. Alg., (1982), 127-137.

[Ry] H. Ryser, Combinatorial Mathematics, MAA 1963.

[SW] C. Schnorr and A. Weber, *Constructing the automorphism group* $\mathrm{Aut}_e(X)$ *for trivalent graphs in time* $O(n^3 \log n)$, Tech. Rep., U. Frankfurt (1981).

[We] B. Weisfeiler, ed., *On Construction and Identification of Graphs,* Lecture Notes in Math 558, Springer, 1976.

[ZKT] V. Zemlyachenko, N. Kornienko, R. Tyshkevich, *Graph isomorphism problem* (Russian) The Theory of Computation I, Notes Sci. Sem LOMI 118 (1982).