GRAPH ISOMORPHISM, GENERAL REMARKS\*

Gary L. Miller Departments of Mathematics and Computer Science The University of Rochester

An open question is the computational complexity of recognizing when two graphs are isomorphic. In an attempt to answer this question we shall analyze the relative computational complexity of generalizations and restrictions of the graph isomorphism problem. In the first Section we show graph isomorphism of regular undirected graphs is complete over isomorphism of explicitly given structures (say Tarski models from logic). Then we show that valence seems to be important. Finally we analyze symmetric cubic graphs.

\*This research was in part supported by the National Research Council,  ${\sf Grant\ \#NRC-A5549}$ .

A graph G is a set of edges over a set of nodes, nodes being denoted by N(G). The graph is undirected unless otherwise noted. The number of edges associated with a node is the valence of the node. The valence of a graph is equal to the maximum over the valences of the nodes. A graph is said to be regular if all nodes have the same valence.

We shall need the computational notations:

- P(NP) is all sets recognizable in (non)deterministic polynomial time; and
- A ≤ pB denote that A is polynomial time reducible to B.

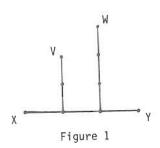
## I. Completeness of Graph Isomorphism over Isomorphism

The main result of this Section is
Theorem 2, which states that isomorphism of undirected graphs is complete over the general
isomorphism problem. We first state and prove a
special case which contains most of the ideas
and techniques to be used to prove the general
case.

 $\frac{\text{Theorem 1:}}{\text{rected graph isomorphism}} \stackrel{\leq}{=} p \text{ undirected graph isomorphism.}$ 

<u>Proof</u>: Suppose that G and G' are two directed graphs on n nodes. We define a map or procedure, say  $\alpha$ , from directed graphs to undirected graphs, such that  $G \not\approx G'$  iff  $\alpha(G) \not\approx \alpha(G')$ . Given G we construct  $\alpha(G)$  as follows:

- 1) For each node of G construct a node for  $\alpha(G)$ .
- 2) For each directed arc of G (say (X→Y)) construct a "gadget" using 7 new nodes and connect it to X and Y as in Figure 1.



By the construction it should be clear that if q is an isomorphism of G onto G' then the natural extension of g to  $\alpha(G)$  is also an isomorphism of  $\alpha(G)$  onto  $\alpha(G')$ . Thus, to complete the proof of the theorem it suffices to prove the following lemma:

Lemma: If g is an isomorphism from  $\alpha(G)$  onto  $\alpha(G')$  then g restricted to the nodes of G is an isomorphism of G onto G'.

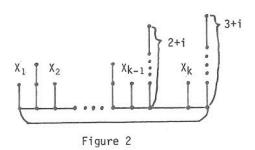
Proof: The spectrum of a node X in a graph on n nodes is a sequence of natural numbers  $S(X) = \langle a_1, \ldots, a_n \rangle$  such that  $a_i$  is the number of nodes whose minimum distance to X is i. Note that the spectrum is an invariant under isomorphism. Now, the spectra of V and W of Figure 1 have the form  $\langle 1,1,2,\ldots \rangle$  and  $\langle 1,1,1,2,\ldots \rangle$  respectively. If  $X \in N(G)$  and X is of valence  $\ell$  in G then the spectrum of X in  $\alpha(G)$  is of the form  $\langle \ell,2\ell,\ldots \rangle$ . Thus, the nodes V (or W) from gadgets in  $\alpha(G)$  are invariant under isomorphism. Therefore gadgets are invariants. The function g maps N(G) onto N(G'). Finally, g restricted to N(G) is an isomorphism of G onto G', since  $X \rightarrow Y$  iff X is connected to Y by a gadget in  $\alpha(G)$ .

A <u>structure</u> is a set A with relations  $R_1,\ldots,R_m$ , where  $R_1 = A^A$ , which we will denote by  $< A,R_1,\ldots,R_m >$ . We will say  $< A,R_1,\ldots,R_m >$  is <u>homeomorphic</u> to  $< A^1,R^1,\ldots,R^1_m >$  if there exists a map g from A to A' such that  $< X_1,\ldots,X_k > \epsilon R_i$  implies  $< g(X_1),\ldots,g(X_k) > \epsilon R^1_i$ ,  $1 \le i \le m$ , and they are isomorphic if g is one-to-one.

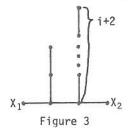
To prove that undirected graph isomorphism is complete over isomorphism of structures, using the techniques developed in the proof of the last theorem, we will need to define a general construct  $\alpha.$ 

Given a structure  ${}^<\!A,R_1,\dots,R_m{}^>$  we defined  $\alpha({}^<\!A,\dots{}^>)$  as follows:

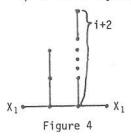
- 1) for each element of A construct a node for  $\alpha(\langle A,R_1,\ldots,R_m\rangle);$
- 2) a) for each ordered sequence  $<X_1,...,X_k>\epsilon R_1$ ,  $k\ge 3$ , construct a  $R_1$ -gadget;



b) for each  $<X_1,X_2>\epsilon R_i$ ;



c) for each  $\langle X_1 \rangle \in R_i$ ;



By argument similar to those previously used we see the spectrum of the "leaves" are unique hence invariant under isomorphism. Thus,  $R_{\dot{1}}\text{-gadgets}$  are invariant which implies A is an invariant. Finally, any isomorphism of  $\alpha(A)$  onto  $\alpha(A')$  induces an isomorphism of A onto A'. This proves the following theorem.

 $\label{eq:theorem 2: Isomorphism of structures $\le p$ graph isomorphism.}$ 

By a group we shall mean a multiplication or Caley table. Since a group can be viewed as a trinary relation over a set, namely  $\langle X,Y,Z \rangle$  iff  $X \cdot Y = Z$ , we get the following theorem:

Theorem 3: (Miller, Monk) Group isomorphism

The best-known upper bound is  $O(n^{\log n + 3})$  due to Tarjan. For a discussion of this result and generalization to latin squares and some graphs derived from latin squares, see [10].

The next result says that when we consider graphs of valence  $\alpha$  where  $\alpha$  is odd we need only consider the subcase of regular graphs of valence  $\alpha.$ 

Theorem 4: Isomorphism of graphs of valence  $\alpha$   $\leq$  p isomorphism of regular graphs of valence  $\alpha$ , when  $\alpha$  is odd.

 $\begin{array}{lll} \underline{Proof}\colon & \text{Consider a $T_{\alpha,n}$ gadget, with nodes} \\ \{X,a_{i,j},b_{i,j} | 1 \leq i \leq \alpha-1, \ 1 \leq j \leq n \} & \text{with connections:} \\ & \{ < X,a_{i,1} > | 1 \leq i \leq \alpha-1 \} \\ & \{ < a_{i,j},b_{k,j} > | 1 \leq i, \ k \leq \alpha-1 \ \text{and} \ 1 \leq j \leq n \} \\ & \{ < b_{i,j},a_{i,j+1} > | 1 \leq j < n \} \\ & \{ < b_{i,n},b_{i+1,n} > | 1 < i < \alpha-1, \ i \ \text{odd} \} \end{array}$ 

For example,  $T_{3,2}$  is given in Figure 5.-

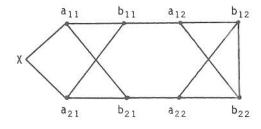


Figure 5

Given a graph G of valence  $\alpha$ ,  $\alpha$  odd, we can pick n large enough so that  $T_{\alpha,n}$  never occurs in G. Now by simply attaching copies of  $T_{\alpha,n}$  to nodes of G we can increase the valence of any node to  $\alpha$ . Thus Theorem 4 is proved.

This gadget has the property that all nodes have valence  $\alpha$  except one which has valence  $\alpha$ -1. Any other gadget it would seem also needs to have this property. But if  $\alpha$  is even and K is a gadget with the above property then K is a graph with an odd number of nodes of odd valence. By a simple counting argument of Euler's we see this is impossible. Thus to extend the theorem

to the even case seems to require we tie together collections of nodes of odd valence, which seems difficult in light of Section II.

Up to an increase in valence of at most one we can assume our graphs are regular.

Corollary: Graph isomorphism < p regular graph isomorphism.

## II. Bounded Valence

All the constructions of Section I preserve valence in the sense that the valence of G equals the valence of  $\alpha(G)$ . Our goal in this Section is to analyze the importance of valence in the isomorphism problem. A natural formalization of this problem is the following open question.

Open Question: Graph isomorphism < p isomorphism over graphs of bounded valence.

Let us first consider bounding the valence to 3. One way of constructing a cubic graph from an arbitrary graph is to replace nodes of valence n (n>3) by n-gon. This procedure is not well defined, as the following example shows. Consider graph A from Figure 6A. There are two ways to replace X by a 4-gon, giving graphs B and B' (see Figures 6B and 6C). These two graphs are not isomorphic, in fact, B is planar while B' is not planar. Thus it seems replacing nodes of higher valence by polygons fails because the polygons induce an orientation on the arcs attached to them. If we could find a polynomial time procedure which uniquely replaces nodes by polygons independent of how the graphs are presented we would be close to producing a polynomial time algorithm for doing the general isomorphism problem.

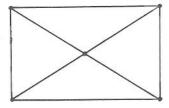


Figure 6A

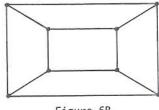


Figure 6B

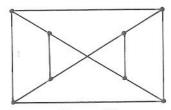


Figure 6C

Since the n-gon is only one of an infinite number of possible graphs that might work we now formalize the properties we seem to need of such a graph and then proceed to show that no such graph can exist.

Definition: An isomorphism gadget (n-gadget) is a connected graph G with valence n together with n+1 distinguished nodes of valence n-1, say Γ, such that the group of automorphisms which stabilize  $\Gamma$  induces all permutations of  $\Gamma$ , i.e., induces  $S_{n+1}$  on  $\Gamma$ .

The main theorem is:

Theorem 5: For n#4 no gadget exists.

For the proof of the general case we refer the reader to [11]. Consider the special case when n=3. In this case we use the following theorem.

Theorem 6: (Babai, Lovasz) If G is a connected graph of valence 3 and H is a group of automorphisms of G which leaves some edge of G fixed then H is a 2-group (H is of order 2<sup>m</sup>, for some m.).

<u>Proof</u>: Suppose the theorem is false. Let H be as in the hypothesis of the theorem and let P divide the order of H, p an odd prime. Further, let  $< X_0, X_1 >$  be the edge fixed by H and  $X_2$  and  $X_3$  be the other two possible neighbors of  $X_1$ . If H' is the subgroup of H which also fixes  $X_2$  and  $X_3$  then  $[H:H'] \le 2$ . By our assumption that P divides |H| and the fact that  $[H:H'] \le 2$  we have P divides |H'|. Using induction and the fact that G is connected, Theorem 6 is proved.

Suppose G is a 3-gadget and  $X_1, X_2, X_3, X_4$  are distinguished nodes of G. Let H be the fixer of  $X_1$ . By attaching an extra edge to  $X_1$ , H satisfies Theorem 6. But, H induces  $S_3$  on  $\{X_2, X_3, X_4\}$  by definition, therefore H is not a 2-group. This contradicts Theorem 6. Thus, 3-gadgets do not exist.

## III. Short Proofs of Nonisomorphism

It is often stated that efficient graph isomorphism algorithms are useful to Chemistry since molecules can be viewed as a graph where the nodes are the atoms and edges are the bonds. A problem which arises is classifying molecules, namely, we have a very large table of molecules and we are given some new molecule and asked whether or not it is already in the list. Since the number of molecules is potentially exponential in the number of atoms per molecule, even a linear time isomorphism algorithm naively produces a potentially exponential search. We now attempt to characterize a feasible solution to the Chemist's problem.

A function f from a class of objects A to the natural numbers is called a <u>certificate</u> with respect to some equivalence relation  $\equiv$  if for all G, G' in A G  $\equiv$  G' iff f(G) = f(G'). In the case that A is incidence matrices and  $\equiv$  is isomorphism then a computable f exists. We shall say that f is a <u>deterministic certificate</u> if f is a certificate and it is computable in polynomial time.

If graph isomorphism has deterministic certificates then graph isomorphism is in P.
Thus deterministic certificates seems too strong a condition. If f is a certificate which is

computable in nondeterministic polynomial time then f is called a <u>succinct certificate</u>. The definition of nondeterministically computable function is given in [9] for completeness. We define it for partial functions.

<u>Definition</u>: A function f over a domain A is said to be computable in nondeterministic polynomial time if there exists a nondeterministic machine M such that on all input  $X \in A$  some path halts and all halting paths must output f(X) in polynomial time.

The existence of a succinct certificate for graphs under isomorphism seems to formally characterize what [7] calls a complete set of invariants for graphs.

Open Question: What is the relation between the following four properties, where  $\equiv$  is an equivalence relation over a set A, other than  $1) \Rightarrow 2) \Rightarrow 4$  and  $1) \Rightarrow 3) \Rightarrow 4$ :

- <A,≡> has deterministic certificates;
- 2) equivalence of A over ≡ is in P;
- <A,≡> has succinct certificates;
- 4) equivalence of A over ≡ is in NP NP?

It is not known if graph isomorphism satisfies any of the above four conditions.

Since polynomial time reducibility oreserves all of the conditions a positive solution for graph isomorphism would imply a positive solution for structures. In particular, group isomorphism is not known to satisfy any of the four conditions. It seems we need to find a tractable restriction of the class of graphs so as to solve the molecular classification problem.

Since molecules have bounded valence and Theorem 6 gives us reason to believe graphs of bounded valence may be easier, we restrict our attention to these graphs. Valence 3 graphs are the first interesting case and by Theorem 4 we need only consider graphs of uniform valence 3, cubic graphs.

Open Question: Is cubic graph nonisomorphism in
NP?

There are many ways of partitioning nodes of a graph into classes invariant under the automorphism group, with the goal of either finding an isomorphism or eliminating possible isomorphisms. If the automorphism group is transitive on vertices then the only invariance partition is the trivial one. Thus, the vertex transitive graph seems like an interesting subcase to consider. Now with one further restriction, namely, that not only is the automorphism group transitive on vertices but also transitive on arcs (transitive over paths of length 1), we are able to say something interesting:

Theorem 7: Arc transitive cubic graph nonisomorphism is in NP.

In fact, a stronger fact is true:

<u>Theorem 8</u>: Arc transitive cubic graphs have succinct certificates.

Since an incidence matrix can easily be viewed as a natural number, we need only construct unique incidence matrices, i.e., enumerations of the vertices of the graph which produce identical incidence matrices. Now arc transitive implies vertex transitive; thus, we can start our enumeration of the vertices of the graph from any vertex. Given a vertex x of a graph G and a sequence of automorphism  $a_1, \ldots, a_k$  such that  $a_1, \ldots, a_k$  generates a vertex transitive group, then there exists a natural enumeration of the vertices associated with  $x, a_1, a_2, \ldots, a_k$ ; namely,

<u>Definition</u>: Given a graph G, vertex x, and sequence of automorphisms  $a_1, \ldots, a_k$  as above we inductively assign an automorphism and ordered successors to every vertex y as follows:

- 1) assign the identity to x;
- 2) given (f,f(x)) we let  $fa_i(x)$  be a successor of f(x) with automorphism  $fa_i$  when  $fa_i(x)$  has no

automorphism assigned to it, for 1 < i < k.

We shall call this assignment the ordered tree associated with  $x,a_1,\ldots,a_k$ .

In the case when the graph is connected and cubic arc transitive there exist in fact two conical automorphisms. Following [12], we introduce the following notations and definition.

An s-arc is a path  $X_0,\ldots,X_S$  and a 1-arc is simply an arc. A graph is s-arc transitive if the automorphism group is transitive on s-arcs. A group acting on a graph is s-regular if it acts regularly on s-arcs (uniquely maps s-arcs to s-arcs). Now, Tutte proved that if a cubic graph is arc transitive then it is s-regular for some s<5. Tutte also proved that there exist cubic graphs which are s-regular for  $1 \le \le 5$ .

Suppose G is an s-regular graph and S is some s-arc, say  $X_0,\ldots,X_S$ , and the other two neighbors to  $X_S$  are X and Y. Now, S has two unique successors,  $X_1,\ldots,X_S,X$  and  $X_1,\ldots,X_S,Y$  which we will denote by  $S_1$  and  $S_2$ . Let  $a_1$  and  $a_2$  be the unique automorphisms of S which send S to  $S_1$  and  $S_2$  respectively. Automorphisms which push arcs forward are called shuntings. Tutte also proved that  $a_1$  and  $a_2$  in a very natural way generate the automorphism group of G.

So the tree associated with  $X_S$ ,  $a_1$ ,  $a_2$ , say  $T(X_S,a_1,a_2)$  is in fact a subgraph of G which is a rooted ordered spanning tree of G. Let  $M(G,a_1,a_2)$  denote the incidence matrix induced by some fixed standard traversal of the spanning tree  $T(X_S,a_1,a_2)$ . Since G is s-transitive the matrices  $M(G,a_1,a_2)$  and  $M(G,a_2,a_1)$  are independent of our choice of S.

Since matrices are linearly ordered, we can choose the minimum of the two, say M(G). Therefore we have defined a certificate for arc transitive cubic graphs, namely, f(G) = M(G). But it is not clear that f is computable in non-deterministic polynomial time. In nondeterministic polynomial time we can guess the shuntings  $a_1$  and  $a_2$ , but we need to also recognize that G is at most s-transitive. Thus, we need to show that the set of s-regular cubic graphs is in NP for each s. A stronger fact is provable. First

we formally define shuntings.

<u>Definition</u>: A shunting in G is an ordered pair (x,a) where x is a vertex and an automorphism of G such that a(x) is adjacent to x and  $a^2(x) \neq x$ . If G is finite then  $a^1(x)$ ,  $i \in \mathbb{Z}$ , determines a simple closed path which is rotated by a. Two shuntings (x,a), (x,b) have <u>overlap s</u> if  $a^{-k}(x) = b^{-k}(x)$  for  $0 \le k \le s$  and  $a(x) \ne b(x)$ ,  $a^{-(s+1)}(x) = b^{-(s+1)}(x)$ . Finally, (x,a) is <u>conjugate</u> to (y,b) is there exists an automorphism a such that  $(x,b) = (ax,aaa^{-1})$ . Using this notation we can show:

<u>Theorem 9</u>: Given two shuntings of overlap  $t \ge 1$  for some cubic graph G then in polynomial time one can find the automorphism group of G.

Proof: Since the automorphism group of G contains 3.25-1.n elements where s is the transivity and n is the number of vertices, the size of the group is only linear in the number of nodes. Using the shuntings  $(x,a_1)$  and  $(x,a_2)$  we can construct the subgroup generated by a1,a2, denoted  $\langle a_1, a_2 \rangle$ . Now,  $\langle a_1, a_2 \rangle$  is t'-regular for some t'<5, by Tuttes' result. If the overlap of  $(x,a_1)$  and  $(x,a_2)$ , t, is strictly less than t' we can find new shuntings with overlap t' in <a1,a2>. Without loss of generality, we can assume that the overlap is in fact t=t'. Our theorem can be restated: given a t-regular subgroup of an s-regular group, for a cubic graph, find the s-regular group. Certain of the pairs (t,s) can not exist by the following theorem:

<u>Theorem 10</u>: If a group of automorphisms for a cubic graph is 4- or 5-regular then it cannot contain a 2- or 3-regular subgroup.

Proof: (See [6]).

We next consider the cases when s=t+1, that is, H is the t-regular subgroup of a t+1 regular group A. We show how to construct A from H. By our counting argument, the index of H in A is 2. H is a normal subgroup of A. Now there exists a unique element w in A of order 2 which

fixes S. By the normality of H and the uniqueness of  $a_1$  and  $a_2$  in H, we have  $wa_1w=a_2$ , i.e.,  $(x,a_1)$  and  $(x,a_2)$  are conjugate. We can rewrite this as  $wa_1=a_2w$  and  $wa_2=a_1w$ . The automorphism w is uniquely defined by

$$w(a_{i_1},...,a_{i_k}(X)) = a_{\gamma(i_1)},...,a_{\gamma(i_k)}(X)$$

where  $i_i \in \{1,2\}$  and  $\gamma(1)=2$ ,  $\gamma(2)=1$ .

All this boils down to  $M(G,a_1,a_2)$  is identical with  $M(G,a_2,a_1)$  if we assume the tree traversal used in constructing M is reasonable.

Thus if w exists we can quickly find it; in fact, it is not hard to show that  $wa_1$  and  $a_2$  are two shunting functions of overlap t+1.

The cases t=1, S=3 and t=1, S=5 can be handled by the following theorem:

<u>Theorem 11</u>: If H,A are 1 and 3(5) regular groups respectively acting on some cubic graph then there exists a 2(4) regular group B such that  $H \le B \le A$ .

Thus we need only deal with the case t=1 and S=4. The smallest 4-regular cubic graph is Heawood's graph on 14 nodes; its automorphism group contains 1-regular subgroups. We shall show that all graphs which have both a 1-regular subgroup and a 4-regular subgroup "look like" Heawood's graph. Let G be a cubic graph which is 4-transitive and let H be a 1-regular group over G. Then H contains shuntings of overlap one, say  $(x,a_1)$  and  $(x,a_2)$ . Using this notation we have the following:

Theorem 12: (Djokovic, Miller) Given G, H,  $a_1$ , and  $a_2$  as above then there exists a 1-regular subgroup of Heawood's graph with shuntings  $(y,b_1)$  and  $(y,b_2)$  with overlap 1, such that the map

$$f(a_{i_1}...a_{i_k}x) = b_{i_1}...b_{i_k}y$$

is a well-defined covering of G over Heawood's graph,  $g(a_{i_1} \dots a_{i_k}) = b_{i_1} \dots b_{i_k}$  is a well-defined homomorphism from H to  $\langle b_1, b_2 \rangle$  and finally, (f,g) form a covering morphism. This covering morphism allows, in a natural way, the lifting of the

full automorphism group of Heawood's graph to G.

Proof: (See [6]).

1111

Summing up, the lattice of possible regular subgroups is (see [6]):

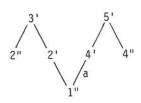


Figure 7

Each inclusion is of index 2 except a. Thus we can climb up the lattice using the normality trick except for inclusion a. For inclusion a we rely on the fact that the graph is a covering of Heawood's graph.

## REFERENCES

- Babai, L. and Lavasz, L. "Permutation Groups and Almost Regular Graphs," <u>Studia Scien-tiarum Mathematicarum Hungarica</u> 8 (1973), 141-150.
- Biggs, N. <u>Algebraic Graph Theory</u>, Cambridge University Press, 1974.
- 3. Cook, S.A. "The Complexity of Theoremproving Procedures," <u>Conference Record of</u> <u>3rd ACM Symposium on the Theory of Comput-</u> ing (1970), 151-158.
- Djokovic, D. "Automorphisms of Graph and Coverings," <u>Journal of Combinatorial</u> <u>Theory B</u>, 16 (1974), 243-247.
- Djokovic, D. "On Regular Graphs V," to appear in <u>J Comb. Theory B</u>.
- Djokovic, D. and Miller, G.L. "Regular Groups of Automorphisms for Cubic Graphs," to appear.
- 7. Harary, F. <u>Graph Theory</u>, Addison-Wesley Publishing Co., Reading, Mass., 1969.
- Karp, R.M. "Reducibility Among Combinatorial Problems," <u>Complexity of Computer Computa-</u> <u>tions</u>, Miller, R.E. and Thatcher, J.W. (Eds.), Plenum Press, NY (1972), 85-103.
- Miller, G.L. "Riemann's Hypothesis and Test for Primality," <u>JCSS</u> 13 (1976).
- 10. Miller, G.L. "On the nlog n Isomorphism Technique," Univ. of Rochester Computer Science Dept., TR17 (1976).
- 11. Miller, G.L. "Graph Isomorphism and Isomorphism of Explicitly Given Structures,"
  U of Roch. CS Dept., TR18 (1976).
- 12. Tutte, W.T. "A Family of Cubical Graphs,"

  Proc. Cambridge Phil. Soc. 43 (1947), 459474.
- Tutte, W.T. "On the Symmetry of Cubic Graphs," Can. J Math. 11 (1959), 621-624.
- 14. Tutte, W.T. <u>Connectivity in Graphs</u>, Univ. of Toronto Press (1966).