Electronic Colloquium on Computational Complexity, Report No. 40 (2008)

Property Testing of Equivalence under a Permutation Group Action

László Babai and Sourav Chakraborty
University of Chicago
{laci, sourav}@cs.uchicago.edu

Abstract

For a permutation group G acting on the set Ω we say that two strings $x, y : \Omega \to \{0,1\}$ are G-*isomorphic* if they are equivalent under the action of G, i. e., if for some $\pi \in G$ we have $x(i^\pi) = y(i)$ for all $i \in \Omega$. Cyclic Shift, Graph Isomorphism and Hypergraph Isomorphism are special cases, and subcases corresponding to certain classes of groups have been central to the design of efficient isomorphism testing for subclasses of graphs (Luks 1982).

We study the complexity of G-isomorphism in the context of property testing: we want to find the randomized decision tree complexity of distinguishing the cases when x and y are G-isomorphic from the cases when they are at least δ-far from being G-isomorphic (in normalized Hamming distance). Error can be 1-sided or 2-sided. In each case we consider two models. In the query-1 model we assume y is known and only x needs to be queried. In the query-2 model we have to query both x and y.

We give various upper and lower bounds for the four combinations of models considered in terms of $n = |\Omega|$ and $|G|$. In many cases, substantial gaps remain between the upper and lower bounds. However, for *primitive permutation groups*, we obtain a tight (up to polylog(n) factors) bound of $\widetilde{\Theta}(\sqrt{n \log |G|})$ for the 1-sided error query complexity in the query-2 model and a tight bound of $\widetilde{\Theta}(\log |G|)$ for the 1-sided error query complexity in the query-1 model. This result extends results of Fischer and Matsliah (2006) on Graph Isomorphism to a surprisingly general class of groups which also includes isomorphism of uniform hypergraphs of any rank. Besides the fact that they include Graph Isomorphism, primitive permutation groups are significant because they form the "building blocks" of all permutations groups, providing the base cases of a natural divide-and-conquer approach successfully exploited in algorithm design (Luks, 1982).

While all our bounds are in terms of the order of G, it seems likely that tighter bounds will depend on the finer structure of G; our result on primitive groups is a first step in this direction.

1 Introduction

"Property testing" is a branch of decision tree complexity (query complexity) theory: with a small number of randomized queries to the unknown input string, we want to have a good chance of distinguishing the cases when the input has a given property from the cases when the input is "far" from any string having the property.

This concept was introduced in the context of program checking by Blum, Luby and Rubinfeld [14] who showed that *linearity* of a function over a vector space can be tested with a *constant* number of queries. A central ingredient in the proof of the MIP=NEXP theorem [11] was the proof that *mulitinearity* can be tested with a *polylogarithmic* number of queries. These two papers were among the roots of the technical developments culminating in the PCP Theorem [8, 7].

Rubinfeld and Sudan [23] formally defined property testing in the context of algebraic properties. Subsequently, the interest in property testing was extended to graph properties, with applications to learning and

ISSN 1433-8092

approximation [20]. In recent years the field of combinatorial property testing has enjoyed a rapid growth (see, e. g., [2, 3, 4, 5, 1], cf. [22, 16]).

Notably, Alon and Shapira [1] show that graph properties that are invariant under vertex removal (i. e., are inherited by induced subgraphs) are testable by a constant number of queries. Isomorphism to a given graph is an important example of a graph property that is not hereditary.

The immediate motivation of our work in this chapter comes from papers by Fischer [17] and Fischer and Matsliah [18] who consider the Graph Isomorphism problem in the property testing model. Here two graphs are given as inputs and we have to test whether they are isomorphic or "far" from being isomorphic.

In this chapter we consider a generalization of graph isomorphism. Let us fix a permutation group G acting on the set Ω. Given two input strings $x, y : \Omega \to \{0, 1\}$, we say x is "G-isomorphic" to y if y is a a π-shift of x for some $\pi \in G$. We want to test the property "x is G-isomorphic to y," that is, we want to distinguish the case when x and y are G-isomorphic from the case when every string that is G-isomorphic to y is far from x. [Formal definitions are given in Section 2.]

Graph Isomorphism is a special case: we need to choose Ω to be the set of unordered pairs of the set V of vertices; and $G = \mathrm{Sym}^{(2)}(V)$ the induced action on Ω of $\mathrm{Sym}(V)$, the symmetric group acting on V (so $n = \binom{|V|}{2}$). We note that the induced symmetric group action on pairs is primitive (does not admit nontrivial invariant partitions of the permutation domain). This fact defines the direction in which we extend results on Graph Isomorphism. We note that by considering the induced symmetric group action on k-tuples, another primitive action, we also cover the case of k-uniform hypergraphs. Here k need not be a constant. Various finite geometries also correspond to primitive groups, so G-isomorphism includes equivalence under geometric transformations (projective, orthogonal, symplectic, etc.).

Besides the fact that the case of primitive groups includes Graph Isomorphism and its immediate generalizations (hypergraphs) as well as geometric equivalence, primitive permutation groups are significant because they form the "building blocks" of all permutations groups in the sense that a "structure tree" can be built of wich the leaves constitute the permutation domain and the action of G extends to the tree in such a way that the action of the stabilizer of any node in the tree is primitive on the children of the node (cf. [13]). This structure tree formalizes the natural divide-and-conquer approach successfully exploited in algorithm design [12, 13, 21].

In "property testing" we want to output 1 if the inputs are G-isomorphic and 0 if they are "far" from being G-isomorphic. The complexity is the number of queries made to the input. We consider two models depending on whether we have to query both x and y or we have to query only one of them (the other is known). We call the models query-2 and query-1, respectively. A property test can have 1-sided or 2-sided-error.

In this paper we focus mainly on property testing of G-isomorphism when the group is primitive. Our main results are the tight bounds on the query complexity when we are allowed only 1-sided error, that is, the algorithm has to output 1 with probability 1 when the two inputs are G-isomorphic and we have to output 0 with high probability when the inputs are "far" from being isomorphic. The main results are the following.

Theorem 1.1. *[Tight bounds for primitive groups] If G is a primitive group then*

1. *The 1-sided-error query complexity for testing G-isomorphism in the query-2 model is $\widetilde{\Theta}(\sqrt{n \log |G|})$.*
2. *The 1-sided-error query complexity for testing G-isomorphism in the query-1 model is $\widetilde{\Theta}(\log |G|)$.*

Theorem 1.1 generalizes a result of Fischer and Matsliah [18] on Graph Isomorphism. The lower bound parts of this result is the main technical contribution of this paper and is proved in Section 3. For the lower

	Query-1 Complexity	Query-2 Complexity
1-sided-error testing	$\widetilde{\Theta}(\log \lvert G\rvert)^{\ddagger}$, $\Omega(\log n)^{\dagger}$	$\widetilde{\Theta}(\sqrt{n \log \lvert G\rvert})^{\ddagger}$
2-sided-error testing	$O(\log \lvert G\rvert)$, $\Omega(\log n)^{\dagger}$	$O(\sqrt{n \log \lvert G\rvert})$

† The lower bound holds when G is transitive and $|G| = 2^{O(n^{1-\epsilon})}$.
‡ The lower bound is for primitive G and the upper bound has no tilde.

Table 1: Bounds on the query complexity of Testing of Equivalence under G-isomorphism.

bound proofs we crucially use a classification of primitive groups based on the O'Nan–Scott Theorem (see [15]).

We also prove some upper and lower bounds for the other cases. But in most of these cases, a significant gap remains between the upper and lower bounds. We present these results in the appendix. The following is the list of results we prove in the appendix. The tilde in the asymptotic notation indicates polylog(n) factors.

Proposition 1.2. *[Upper bound]*

1. *The query-1 complexity of 1-sided and 2-sided error G-isomorphism testing is $O(1 + \log |G|)$.*
2. *The query-2 complexity of 1-sided and 2-sided error G-isomorphism testing is $O(\sqrt{n(1 + \log |G|)})$.*

In Table 1, we abbreviated the expression $1 + \log |G|$ to $\log |G|$ for better typography. The only case where this makes a difference is when $|G| = 1$ so the results as stated in the Table 1 assume $|G| \geq 2$.

Theorem 1.3. *[Lower bound] Let G be a transitive group of order $2^{O(n^{1-\epsilon})}$. Then the 2-sided-error query-1 complexity of the property testing of G-isomorphism is $\Omega(\log n)$.*

Note that we have tighter lower bound for the same case when G is primitive.

In Section 2 we give the formal definitions. In Sections 3, 4 and 5 we give the proofs of the above three results. In Section 6 we state further nearly tight bounds that follow from our results (in addition to Theorem 1.1).

Table 1 summarizes our results on G-isomorphism. Table 2 gives the results of Fischer and Matsliah on Graph Isomorphism. In Table 3 we specialize our results to the case of Graph Isomorphism for comparison with the results of Fischer and Matsliah.

2 Preliminaries

2.1 Definitions

Let Ω be a set of size n. The permutations of Ω form the **symmetric group** $\mathrm{Sym}(\Omega)$ of order $n!$. We write the action of $\pi \in \mathrm{Sym}(\Omega)$ as $i \mapsto i^{\pi}$. For a subset $S \subseteq \Omega$ we set $S^{\pi} = \{i^{\pi} : i \in S\}$.

A subgroup G of $\mathrm{Sym}(\Omega)$ is a **permutation group**; Ω is the **permutation domain** on which G acts. G has **order** $|G|$ and **degree** n.

G is **transitive** if $(\forall i, j \in \Omega)(\exists \pi \in G)(i^\pi = j)$. A partition $\Omega = \Omega_1 \dot\cup \ldots \dot\cup \Omega_m$ is **invariant** under $\pi \in \mathrm{Sym}(\Omega)$ if $(\forall i)(\exists j)(\Omega_i^\pi = \Omega_j)$. The partition is invariant under G if it is invariant under every $\pi \in G$. The trivial partitions correspond to $m = 1$ or $m = n$; these are always invariant. If G is transitive and does not admit any nontrivial invariant partition then G is **primitive**. The largest primitive permutation groups of degree n other than the symmetric and the alternating groups (groups of even permutations) have order $\exp(O(\sqrt{n}\log^2 n))$ ([9, 10]) so except for the two classes of "giants" of order $n!$ and $n!/2$, resp., $\log(|G|) = \widetilde{O}(\sqrt{n})$ for all primitive groups of degree n.

We use the notation $[n] = \{1, 2, 3, ..., n\}$. Most often we take $\Omega = [n]$ and write S_n for $\mathrm{Sym}([n])$.

Definition 2.1. A *partial assignment* is a function $p : S \to \{0, 1\}$ where $S \subseteq [n]$. We call S the support of this partial assignment and often denote $|S|$ as $|p|$. We call x a (full) assignment if $x : [n] \to \{0, 1\}$. (Note than a string $x \in \{0, 1\}^n$ can be thought of as a full assignment.) We say $p \subseteq x$ if x is an extension of p, *i. e.*, if $p = x|_S$ (the restriction of x to S).

$\mathrm{Ham}(x, y)$ will denote the Hamming distance of the strings (full assignments) x and y.

Definition 2.2. Let $T \subseteq [n]$ and let $\pi \in S_n$.

Let G be a permutation group acting on $[n]$. Then the sets T^π, where $\pi \in G$, are called the *G-shifts* of T. If $p : T \to \{0, 1\}$ is a partial assignment then we define $p^\pi : T^\pi \to \{0, 1\}$ as $p^\pi(i) = p(i^{\pi^{-1}})$.

Given two full assignments x and y and a permutation group G we denote by $d_G(x, y)$ the minimum distance between the G-shifts of x and y. That is,

$$d_G(x, y) = \min_{\pi_1, \pi_2 \in G} \mathrm{Ham}(x^{\pi_1}, y^{\pi_2}). \tag{1}$$

Since G is a group, we have

$$d_G(x, y) = \min_{\pi \in G} \mathrm{Ham}(x, y^\pi) = \min_{\pi \in G} \mathrm{Ham}(x^\pi, y). \tag{2}$$

If $d_G(x, y) = 0$ then we say "x is G-**isomorphic** to y."

A **2-sided property tester** for G-isomorphism is a probabilistic decision tree, say $\mathcal{A}$, such that given $x, y \in \{0, 1\}^n$

$$\text{if } d_G(x, y) = 0 \text{ then with probability} > \tfrac{2}{3} \text{ we have } \mathcal{A}(x, y) = 1, \text{ and,}$$

$$\text{if } d_G(x, y) \geq \delta n \text{ then with probability} > \tfrac{2}{3} \text{ we have } \mathcal{A}(x, y) = 0.$$

An **1-sided error property tester** is one which makes no mistake if $d_G(x, y) = 0$.

The complexity of a property tester is the maximum (over all possible inputs) of the minimum number of bits that need to be queried. If neither x nor y is given (so both need to be queried) then we speak of a *query-2 tester* and correspondingly of **query-2 complexity**. If one of them is given (we always assume y is given) and only the other (that is x) needs to be queried then we speak of a *query-1 tester* and **query-1 complexity**.

The trivial upper bound on the complexity of query-2 testers is $2n$ and of query-1 testers is n.

All our upper bound results hold for any permutation group G. But for our lower bound results we need some more structure on G. In Theorem 1.3 we assume that the group is transitive while Theorem 1.1 holds for primitive groups. Our main tool for primitive groups is the O'Nan–Scott Theorem (see Section 3).

	Query-1 Complexity	Query-2 Complexity
1-sided-error testing	$\widetilde{\Theta}(\|V\|)$	$\widetilde{\Theta}(\|V\|^{3/2})$
2-sided-error testing	$\widetilde{\Theta}(\sqrt{\|V\|})$	$\Omega(\|V\|)$, $\widetilde{O}(\|V\|^{5/4})$

Table 2: The results of Fischer and Matsliah for Graph Isomorphism.

2.2 Previous Results

The query complexity of the property testing version of graph isomorphism has been well studied. Fischer and Matsliah [18] gave some tight bounds. In case of graph isomorphism the group that acts is $S^{(2)}_{|V|}$, where V is the vertex set of the graph. Hence the order of the group is $|V|!$. Table 2 shows the main results of [18].

2.3 Chernoff bounds

We shall repeatedly use the following version of the Chernoff bounds, as presented by N. Alon and J. Spencer [6, Corollary A.14].

Let $X_1, X_2, \ldots, X_k$ be mutually independent indicator random variables and $Y = \sum_{i=1}^{k} X_i$. Let the expected value of Y be $\mu = E[Y]$. For all $\alpha > 0$,

$$\Pr[|Y - \mu| > \alpha\mu] < 2e^{-c_\alpha \mu},$$

where $c_\alpha > 0$ depends only on α.

3 Query Complexity for 1-sided-error Testing of Equivalence under some Primitive Group Action

3.1 Structure of Primitive Groups

Definition 3.1. Let G be a permutation group acting on a set A and H a permutation group acting on a set B. The *wreath product* $G \wr H$ is the split extension of the base group G^B (the cartesian product of $|B|$ copies of G) by H, where H acts on G^B by permuting the factors as it does the elements of B. Identifying G^B with the set of functions $f : B \to G$ we have $h^{-1}fh(b) = f(h^{-1}(b))$ for $h \in H$, $b \in B$.

There are two natural actions of $G \wr H$.

1. The imprimitive action on $A \times B$. The base group acts in the first coordinate by the rule $f(a, b) = (f(b)(a), b)$ and H acts on the second coordinate in the usual way.

2. The product action on the set A^B of $B \to A$ functions, where the base group acts coordinatewise (that is, if $p \in A^B, f \in G^B$, then $(fp)(b) = f(b)(p(b))$ and H acts by permuting the coordinates $((hp)(b) = p(h^{-1}(b))$ for $g \in G^B, h \in H$).

Note that these are two permutation representations of the same group. Note also that $G \wr H$ has $G^{|B|}$ as a normal subgroup with H as the quotient.

The structure of primitive permutation groups is described by the O'Nan–Scott Theorem. A useful consequence of that theorem is given by Cameron.

Theorem 3.2. *[15] There is a (computable) constant c with the property that, if G is a primitive permutation group of degree n, then one of the following holds:*

1. $|G| \leq n^{c \log n}$.
2. *G is a subgroup of $Aut(A_m^{(k)}) \wr S_\ell$ (product action) containing $(A_m^{(k)})^\ell$, where $A_m^{(k)}$ is the alternating group A_m acting on k-element subsets. [We can assume without loss of generality that $1 \leq k \leq \frac{m}{2}$].*

So in the case $|G| > n^{c \log n}$ the degree of G is given by

$$n = \binom{m}{k}^\ell \text{ and therefore } n \geq m^\ell. \tag{3}$$

It follows that $\ell \leq \log_2 n$. Also since we can assume $k \leq \frac{m}{2}$, so

$$\binom{m}{k} \geq \left(\frac{m}{k}\right)^k \geq 2^k \text{ and therefore } k \leq \log_2 n. \tag{4}$$

In fact if $|G| > n^{c \log n}$ then we obtain the bound on the size of G as

$$|G| \leq (m!)^\ell (\ell!) < m^{m\ell} \ell^\ell \leq n^m \ell^\ell \text{ [From Equation 3]} \tag{5}$$

Since $\ell \leq \log_2 n$ we have from Equation 5,

$$c(\log n)^2 < \log(|G|) < (m \log n + \ell \log \ell) \sim m \log n. \tag{6}$$

The last asymptotic equality holds because $\ell < \log n$ and therefore $\ell \log \ell = o(\log^2 n)$.

Therefore,

$$\log |G| \lesssim m \log n \text{ and } m \gtrsim c \log n. \tag{7}$$

It follows in particular that $m \geq 7$ (for sufficiently large n). The significance of this is in the known fact that for $m \geq 7$ we have

$$Aut(A_m) = S_m, \tag{8}$$

and therefore $Aut(A_m^{(k)}) = S_m^{(k)}$.

Observation 3.3. *If $k = O(\sqrt{m})$ then*

$$\binom{m}{k} = \Theta\left(\frac{m^k}{k!}\right).$$

Corollary 3.4. *Either $\sqrt{n \log |G|} = \widetilde{O}(\sqrt{n})$ or*

$$m \binom{m}{k}^{\ell/2} = \widetilde{O}(\sqrt{n \log |G|})$$

.

Proof. Let $k > \sqrt{m}$. Then

$$n = \binom{m}{k} > \left(\frac{m}{k}\right)^k > 2^k > 2^{\sqrt{m}}.$$

Therefore $m = O((\log n)^2)$ which implies from Equation 6 $\log |G| < (\log n)^3$. Hence if $k > \sqrt{m}$ we have $\sqrt{n \log |G|} = \widetilde{O}(\sqrt{n})$. The corollary now follows from Observation 3.3. □

Definition 3.5. Let $A, B \subset [n]$ and $p : A \to \{0,1\}$ and $q : B \to \{0,1\}$ be two partial assignments. Let G be a permutation group on $[n]$. Then p and q are said to be G-**agreeable** if there exists a full assignment x on $[n]$ and two elements $\pi_1, \pi_2 \in G$ such that x is an extension of both p^{π_1} and q^{π_2}. Since G is a group this is same as saying p and q are G-agreeable if there exists an element $\pi \in G$ and a full assignment x such that x is an extension of both p^π and q. We say that p and q are agreeable through π.

We say that the partial assignments p and q are compatible if there is a full assignment x on $[n]$ which is an extension of both p and q.

Definition 3.6. Let G be a permutation group on $[n]$. Let x and y be two full assignments on $[n]$. Then x and y are called k-G-agreeable if for any sets $A, B \subset [n]$ with $|A|, |B| \leq k$, the partial assignments $x|_A$ and $y|_B$ are G-agreeable.

3.2 G-Agreeability Lemma for G Primitive

The following proposition is folklore.

Proposition 3.7. *Let G be a transitive group on $[n]$. Let us fix $A, B \subset [n]$ and let us select $\pi \in G$ uniformly at random. Then*

$$E(|A^\pi \cap B|) = \frac{|A||B|}{n}. \tag{9}$$

Proof. By G-symmetry, for each $b \in B$ we have $\Pr(b \in A^\pi) = \frac{|A|}{n}$. Now the linearity of expectation yields the result. □

Corollary 3.8. *Let G be a transitive group on $[n]$. Let $A, B \subset [n]$ with $|A|, |B| \leq \epsilon\sqrt{n}$. Then,*

$$\Pr_{\pi \in G}[A^\pi \cap B = \phi] > (1 - \epsilon^2)$$

In particular if A and B are the support of the partial functions p and q, respectively, then p and q are G-agreeable.

Proof. Immediate from Proposition 3.7 by Markov's inequality. □

A simple consequence of Corollary 3.8 is that if G is a transitive group then any two full assignments x and y on $[n]$ are $\sqrt{n}$-G-agreeable.

Next we state the most technical lemma of this chapter - the G-Agreeability Lemma for primitive groups.

Lemma 3.9 (G-Agreeability Lemma)**.** *Let G be a primitive group. Then there exist two full assignments x and y on $[n]$ such that $d_G(x, y) \geq n/6$ and x and y are $\widetilde{O}(\sqrt{n \log |G|})$-$G$-agreeable.*

3.3 Lower Bounds for 1-sided error Testing

Proof of Part 1 of Theorem 1.1

Let $\mathcal{A}$ be a 1-sided-error query-2 property tester for G-isomorphism. Let the inputs be x and y. After the queries are made we get two partial functions $x|_{Q_x}$ and $y|_{Q_y}$. Now if $x|_{Q_x}$ and $y|_{Q_y}$ are G-agreeable then we have no proof that $d_G(x,y) \neq 0$. Since $\mathcal{A}$ is a 1-sided-error tester, it has to output 1. So by Lemma 3.9 we see that there exists x and y such that $d_G(x,y) \geq \frac{1}{6}n$ and $\mathcal{A}(x,y)$ has to be 1 if the query size is $\widetilde{O}(\sqrt{n \log |G|})$. So the result follows from the lemma.

Proof of Part 2 of Theorem 1.1

We recall the example for lower bound of 1-sided query-1 complexity of graph isomorphism given by Fischer and Matsliah [18]. The unknown graph is the complete graph on n vertices while the known graph is the union of $n/2$ isolated vertices and a complete graph on $n/2$ vertices. Note that without querying at least $n/4$ pairs of vertices it is impossible to give a certificate of non-isomorphism. This gives the lower bound of $n/4$ for the graph isomorphism case.

A similar example can be given in case of isomorphism under primitive group action. First of all we assume that the primitive group is of size more than $n^{c \log n}$ where the c is same as in Cameron's Theorem 3.2. Now we use the structure of the primitive group given by Cameron. We continue with the same notation as in Section 3.4. We partition V_1 into three disjoint parts, namely $V_a, V_b,$ and V_c, where $|V_a| = |V_b| = |V_c| = \frac{m}{3}$. The known input is

$$x(W) = 1 \text{ iff } W \in \binom{V_a, V_c, V_2 \ldots, V_\ell}{1, k-1, k, \ldots, k}$$

The unknown input is

$$y(W) = 1 \text{ iff } W \in \binom{(V_a \cup V_b), V_c, V_2, \ldots, V_\ell}{1, k-1, k, \ldots, k}$$

Note that one need to make at least $m/6$ queries to give a certificate of non-isomorphism between the two inputs. Now from Equation 7 we get a lower bound of $\Omega(\frac{\log(|G|)}{\log n})$.

3.4 Proof of the G-Agreeability Lemma for Primitive Groups

Proof of Lemma 3.9. If $|G| \leq n^{c \log n}$ then $\sqrt{n \log |G|} = \widetilde{O}(\sqrt{n})$ and the result follows from Corollary 3.8. Therefore from Theorem 3.2 and Corollary 3.4 we may assume that G is a subgroup of $S_m^{(k)}) \wr S_\ell$ (product action) containing $(A_m^{(k)})^\ell$, and $k < \sqrt{m}$. Hence in rest of the proof we will use from Lemma 3.3 that

$$\binom{m}{k} = \Theta\left(\frac{m^k}{k!}\right)$$

where the impied constant is absolute.

If $\ell = 1$ and $G = S_m^{(2)}$ then G is the group of automorphisms of the complete graph on m vertices. This case was settled by Fischer and Matsliah [18]. We generalize their technique.

For our convenience we have the following definition.

Definition 3.10. *Let $T_1, T_2, \ldots, T_s$ be disjoint sets and $r_1, r_2, \ldots, r_s$ be positive integers satisfying $\sum_{i=1}^{s} r_i = R$. Then by $\binom{T_1, T_2, \ldots, T_s}{r_1, r_2, \ldots, r_s}$ we mean the set of R-tuples formed by r_i distinct elements from the set T_i for all $1 \leq i \leq s$. That is,*

$$\binom{T_1, T_2, \ldots, T_s}{r_1, r_2, \ldots, r_s} = \left\{ \bigcup_{i=1}^{s} S_i \Big| S_i \subseteq T_i, |S_i| = r_i \right\}$$

G is a subgroup of $S_m^{(k)} \wr S_\ell$ (product action) containing $(A_m^{(k)})^\ell$. G is naturally isomorphic to a subgroup of $S_m \wr S_\ell$, acting in its imprimitive action on $\mathcal{V} = \cup_{i=1}^{\ell} V_i$, where $|V_i| = m$ and the V_i are all disjoint. Then any full assignment is a function from the set $\binom{V_1, \ldots, V_\ell}{k, k, \ldots, k}$ to $\{0, 1\}$.

We will first have to define two full assignments, x and y, on n bits. The group G is a map from $\mathcal{V}$ to $\mathcal{V}$. The rest of our proof has the following two parts:

- Define the full assignments x and y and prove that $d_G(x, y) > \delta n$ for some constant δ.
- Let Q_x and Q_y be two query sets for x and y, respectively such that both $|Q_x|$ and $|Q_y|$ is $\widetilde{O}(\sqrt{n \log |G|})$. Then we prove that there exist a permutation $\pi = \pi_1 \times \pi_2 \times \cdots \times \pi_\ell \in (A_m^{(k)})^\ell$ such that Q_x^π and Q_y are compatible.

We start with defining x.

Definition of the full assignments x and y

We partition V_1 into three disjoints parts U_1, U_2 and U_3 such that

$$|U_3| = m\left(1 - \frac{1}{k}\right),\ |U_1| = m\left(\frac{1}{2k} + \epsilon\right) \text{ and } |U_2| = m\left(\frac{1}{2k} - \epsilon\right)$$

We define x and y as

$$x(W) = 1 \text{ iff } W \in \binom{U_1, U_3, V_2, \ldots, V_\ell}{1, k-1, k, k, \ldots, k}$$

$$y(W) = 1 \text{ iff } W \in \binom{U_2, U_3, V_2, \ldots, V_\ell}{1, k-1, k, k, \ldots, k}$$

Note that a map from $\mathcal{V}$ to $\mathcal{V}$ gives a reordering of the bits is x.

Now note that number of 1s in x and y is $m\binom{m(1-\frac{1}{k})}{k-1}(\frac{1}{2k} + \epsilon)\binom{m}{k}^{\ell-1}$ and $m\binom{m(1-\frac{1}{k})}{k-1}(\frac{1}{2k} - \epsilon)\binom{m}{k}^{\ell-1}$ respectively. So from the difference in number of 1s in x and y we see that

$$d_G(x, y) \geq 2\epsilon m \binom{m(1-\frac{1}{k})}{k-1} \binom{m}{k}^{\ell-1}$$

For $k = 1$ the right-hand side is $2\epsilon m^\ell = 2\epsilon n$. If $k \neq 1$ then from Lemma 3.3 and the fact that $\left(1 - \frac{1}{k}\right)^{k-1} \geq \frac{1}{e}$ we obtain

$$2\epsilon m \binom{m(1-\frac{1}{k})}{k-1} \sim 2\epsilon \frac{m^k (1-\frac{1}{k}))^{k-1}}{(k-1)!} \geq \epsilon k \frac{2m^k}{ek!} = \Theta\left(\epsilon k \binom{m}{k}\right).$$

So if we choose $\epsilon = \frac{1}{6ck}$ where c is the constant implied in the Θ notation then we get that

$$d_G(x, y) \geq \frac{1}{6}\binom{m}{k}^\ell = \frac{1}{6}n.$$

Now we give the second part of the proof. Let Q_x and Q_y be query sets for x and y, respectively, such that $|Q_x|, |Q_y| \leq M$ where $M = \frac{m}{18k}\sqrt{\binom{m(1-\frac{1}{k})}{k-1}\binom{m}{k}^{\ell-1}}$.

To prove that x and y are M-G-agreeable, we have to give a $\pi \in (A_m^{(k)})^\ell \subseteq G$ that maps $\mathcal{V}$ to $\mathcal{V}$ such that Q_x^π and Q_y agrees.

If $a \in U_1$ then we define

$$q_x(a) = \left\{ w \in \binom{U_1, U_3, V_2, \ldots, V_\ell}{1, k-1, k, \ldots, k} \middle| w \in Q_x \text{ and } a \in w \right\}$$

Similarly if $b \in U_2$ let

$$q_y(b) = \left\{ w \in \binom{U_1, U_3, V_2, \ldots, V_\ell}{1, k-1, k, \ldots, k} \middle| w \in Q_y \text{ and } b \in w \right\}$$

Now by an averaging argument there exist sets $A \subset U_1$ and $B \subset U_2$ such that $|A|, |B| > \frac{2m}{9k}$ and for all $a \in A$ and $b \in B$ we have

$$|q_x(a)|, |q_y(b)| \leq \frac{9k}{m} M.$$

Let $H = A_{m(1-\frac{1}{k})}^{(k-1)} \times (A_m^{(k)})^{\ell-1}$ acting on the set $\binom{U_3, V_2, \ldots, V_\ell}{k-1, k, k, \ldots, k}$. Pick a random element $\pi' \in H$. Note that H acts transitively on the set $\binom{U_3, V_2, \ldots, V_\ell}{k-1, k, k, \ldots, k}$.

Fix an arbitrary even bijection from A to B, i. e., an even permutation of $[n]$ which maps A to B. Let $a \in A$ be mapped to $b \in B$. We call a pair (a, b) acceptable if $q_x(a)^{\pi'} \cap q_y(b) = \phi$. We want to calculate the probability of a pair (a, b) being acceptable.

Note that q_a and q_b are two subsets of $\binom{U_3, V_2, \ldots, V_\ell}{k-1, k, \ldots, k}$. So from Lemma 3.8 we get that probability that a and b are compatible is more that $\frac{3}{4}$.

So the expected number of (a, b) pairs that are acceptable is $\geq \frac{3}{4}\frac{2m}{9k} = \frac{m}{6k} = \epsilon m$. So there exist a permutation $\pi' \in H$ such that ϵm of the (a, b) pairs are acceptable. These acceptable pairs along with the permutation π' give a map from a set $A' \subset A \subset U_1$ to $B' \subset B \subset U_2$ such that Q_x and Q_y are compatible. Now we have

$$|U_1 \backslash A'| = |U_1|$$

and

$$|U_2| = |U_2 \backslash B'|.$$

Hence π' and the map from the acceptable pairs can be extended to a mapping π from $\mathcal{V}$ to $\mathcal{V}$ by mapping $U_1 \backslash A'$ and U_2 to U_1 and $U_2 \backslash B'$ respectively, such that Q_x^π and Q_y are compatible.

Finally from Corollary 3.4 we have $M = \widetilde{O}(\sqrt{n \log |G|})$. □

4 Upper bounds for Transitive groups

Proof of Proposition 1.2

Definition 4.1. We define **query sequence** as the sequence of elements of $[n]$ consisting of the positions of the bits of the input that will be queried. Repetition is permitted.

The proofs of both parts of Proposition 1.2 are rather simple applications of the Chernoff bound; we describe the proofs for completeness.

Proof of Part 1 of Proposition 1.2. In this part we only have to query bits of x. Let us choose a real number p, $0 < p < 1$, appropriately (see below). The length of the query sequence Q will be $m = pn$. We say that two partial functions p, q *contradict* at i if both $p(i)$ and $q(i)$ are defined and $p(i) \neq q(i)$. The following is the test:

1. Construct the query sequence $Q = (a_1, \ldots, a_m)$ by choosing pn elements of $[n]$ independently at random. (So there is a small chance that the same element is chosen twice.)
2. Query the bits of x corresponding to Q. So we obtain the partial function $x|_Q$.
3. If for some $\pi \in G$ the partial function $x|_Q^\pi$ and y contradict in fewer than $\delta pn/2$ places then output 1. Otherwise output 0.

Now to prove that the above test works we have to show that the test outputs the correct answer with probability at least $\frac{2}{3}$.

Given a permutation $\pi \in G$, we say that the ith bit queried contradicts y along π if $x(a_i) \neq y(a_i^\pi)$. We define the $(0,1)$-variable X_i^π by

$$X_i^\pi = 1 \text{ if the } i\text{th bit queried contradicts } y \text{ along } \pi.$$

$X^\pi = \sum X_i^\pi$ is the number of places the partial information of the two strings contradicts along π. Since the members of the query sequence are chosen independently, the X_i^π are mutually independent indicator random variables ($i = 1, \ldots, m$; π is fixed). So we can use the Chernoff bound to estimate the value of X^π.

Suppose that one of the following conditions holds for a given $\pi \in G$:

(i) x^π and y agrees completely;

(ii) x^π and y differ in more than δn places.

It follows that the expected value of X^π is less than 0 in Case (i) and greater than $pn\delta$ in Case (ii). Let $\eta = \delta/2$.

So, using the Chernoff bound we obtain,

$$\Pr\left[|X_\pi - E(X_\pi)| > \eta np\right] \leq 2\exp\left(-c_{\eta/\delta} np\right). \tag{10}$$

If there exists $\pi \in G$ satisfying condition (i) (so the correct answer is 1), the probability we err is less than the right-hand side of this inequality.

If every $\pi \in G$ satisfies (ii) (so the correct answer is 0), the probability we err is less than $|G|$ times the right-hand side by the union bound. So in any case, the probability of error is less than $|G| \exp\left(-c_{\eta/\delta} np\right)$.

If we take $p = \frac{2+\log(|G|)}{c_{\eta/\delta} n}$ then the probability of error is less than $\frac{1}{3}$.

Note that this is an 1-sided error algorithm. So the query-1 complexity for the 1-sided error G-isomorphism testing is $O(pn) = O(1 + \log|G|)$.

□

Proof of Part 2 of Proposition 1.2. In this part we have to query both x and y. Again we choose a real number p, $0 < p < 1$, appropriately (see below). The total length of the query sequence will be $2m = 2pn$. The following is the test:

1. Construct two query sequences $Q_1 = (a_1, \ldots, a_m)$ and $Q_2 = (b_1, \ldots, b_m)$, by choosing these $2m$ elements of $[n]$ independently at random.
2. Query the bits of x and y corresponding to Q_1 and Q_2 respectively. So we obtain the partial functions $x|_{Q_1}$ and $y|_{Q_2}$.
3. If for some group element $\pi \in G$, the partial function $x|^{\pi}_{Q_1}$ contradicts the partial function $y|_{Q_2}$ in fewer than $p^2 n\delta/2$ places then output 1. Otherwise output 0.

To prove that the above test works we have to show that the test outputs the correct answer with probability at least $\frac{2}{3}$.

Given a permutation $\pi \in G$, we say that the ith bit queried in x contradicts $y|Q_2$ along π if $a_i^{\pi} \in Q_2$ and $x(a_i) \neq y(a_i^{\pi})$. We define the $(0,1)$-random variable X_i^{π} by

$$X_i^{\pi} = 1 \text{ if the } i\text{th bits queried contradict along } \pi.$$

$X^{\pi} = \sum X_i^{\pi}$ is the number of places the queried information about the two strings contradicts along π. Since the members of the query sequences are chosen independently, the X_i^{π} are mutually independent indicator random variables ($i = 1, \ldots, m$; π is fixed). So we can use the Chernoff bound to estimate the value of X^{π}.

For any group element π, let D_{π} be the set of positions of the bits of x^{π} that differ from y. The expected number number of bits in D_{π} that are queried is $p|D_{\pi}|$. Now for X_i^{π} to be 1 we must also have $a_i^{\pi} \in Q_2$. Now the expected number of bits in D_{π} that are queried in both x and y is $p^2|D_{\pi}|$. So $E[X^{\pi}] = p^2|D_{\pi}|$.

Suppose that one of the following conditions holds for a given $\pi \in G$:

(i) x^{π} and y agrees completely;

(ii) x^{π} and y differ in more than δn places.

If condition (i) holds then $|D_{\pi}|$ is less than 0 and hence $E[X^{\pi}]$ is less than $\epsilon p^2 n$. In case of condition (ii), $|D_{\pi}|$ is greater than δn and hence $E[X^{\pi}]$ is greater than $\delta p^2 n$. Let $\eta = \delta/2$. From the Chernoff bound we get,

$$\Pr\left[|X^{\pi} - E[X^{\pi}]| > \eta p^2 n\right] \leq 2\exp\left(-c_{\eta/\delta} p^2 n\right) \tag{11}$$

If there exists $\pi \in G$ satisfying condition (i) (so the correct answer is 1), the probability we err is less than the right-hand side of this inequality.

If for every π condition (ii) is satisfied then the probability we err is less than $|G|$ times the right-hand side by the union bound.

If we take $p = \sqrt{\frac{2+\log|G|}{c_{\eta/\delta} n}}$ the error is less than $\frac{1}{3}$.

Again note that this algorithm is also 1-sided. So the query-2 complexity of 1-sided error G-isomorphism testing is $O(\sqrt{n(1+\log|G|)})$. □

5 Lower bounds for Transitive Groups

Proof of Theorem 1.3

We will use the following lemma and the Theorem.

Lemma 5.1. *If G is a transitive permutation group and S is a subset of $[n]$, $|S| = k$ then there exist at least $\frac{n}{k^2}$ pairwise disjoint G-shifts of S.*

Theorem 5.2. ([16, 19]) *Let $x \in \{0,1\}^n$. Suppose that there exists a distribution D_P on inputs $y \in \{0,1\}^n$ such that $f(x,y) = 1$, and a distribution D_N on inputs $z \in \{0,1\}^n$ such that x and z are ϵ-far from satisfying the f. Suppose further that for any $\mathcal{Q} \subset [n]$ of size q, and any $g : \mathcal{Q} \to \{0,1\}$, we have $\frac{2}{3}\Pr_{D_P|\mathcal{Q}}(g) < \Pr_{D_N|\mathcal{Q}}(g)$. Then any 2-sided-error property test for f requires at least q queries.*

Proof of Theorem 1.3. Let x be a full assignment. For any subset $P \subseteq [n]$ of size k and any $Q \in \{0,1\}^k$ let $p_Q = |\{\pi \in G : x^\pi|_P = Q\}|$. We call x "almost universal" if for all $Q \in \{0,1\}^k$ and for all subset P of size k, we have $|p_Q - \frac{n}{2^k}| \leq \frac{n}{5(2^k)}$. Note that this means that if we pick $\pi \in G$ at random then for all $Q \in \{0,1\}^k$ and for all subset P we have

$$|\Pr[x^\pi|_P = Q] - \mu| \leq \mu/5$$

where $\mu = 1/2^k$.

We prove the existence of an almost universal string using the probabilistic method. Pick a random full assignment x. Fix a subset $P \subset [n]$ of size k and queries the bits of x corresponding to the indices in P. For a fixed $Q \in \{0,1\}^k$ we will estimate p_Q. Using Lemma 5.1 we can place $\frac{n}{k^2}$ disjoint G-copies of the subset P in $[n]$. Let $\mathcal{S}$ denote the set of disjoint copies of P. Let $v_i^Q(x)$ be the $(0,1)$-indicator variable indicating whether the i-th G-copy of P in $\mathcal{S}$ is same as Q. Since x is chosen randomly these random variables are independent. Let $v_Q(x) = \sum v_i^Q(x)$. So $v_Q(x)$ be the number of times Q occurs in $\mathcal{S}$. The expected value of $v_Q(x)$ is $\frac{n}{k^2 2^k}$. So using the Chernoff bound

$$\Pr\left[\left|v_Q(x) - \frac{n}{k^2 2^k}\right| > \frac{n}{5k^2 2^k}\right] \leq 2\exp\left(-\frac{c_{1/5}n}{k^2 2^k}\right).$$

So using the union bound we get

$$\Pr\left[\forall \pi \in G, \forall Q \in \{0,1\}^k, \forall P, \left|v_Q(x^\pi) - \tfrac{n}{k^2 2^k}\right| \leq \tfrac{n}{5k^2 2^k}\right]$$
$$\geq 1 - 2\exp\left(-\tfrac{c_{1/5}n}{k^2 2^k}\right)|G|\tbinom{n}{k}2^k.$$

If $|G| = 2^{O(n^{1-\epsilon})}$ for any positive ϵ and $k \leq (1-\gamma)(\log n)$ (where $\gamma > 0$), this probability is non-zero. Now since we had exactly (n/k^2) times of disjoint copies of P, so there is a string x such that

$$\forall \pi \in G, \forall Q \in \{0,1\}^n, \forall P, |\Pr[x^\pi|_P = Q] - \mu| \leq \mu/5$$

where $\mu = 1/2^k$.

Similarly we can show that existence of a full assignment such that it is $\frac{1}{3}$-far from x and still "almost universal." The argument is similar. Now probability that a random string is $\frac{1}{3}$-close to x is less than $\frac{1}{2^{o(n)}}$. Using the same argument as above we can say that the probability that a random string is $\frac{1}{3}$-far from x and is an "almost universal" string is more than $\left(1 - \frac{1}{2^{o(n)}} - 2\exp\left(-\frac{c_{1/5}n}{k^2 2^k}\right)|G|\binom{n}{k}2^k\right)$. This is also positive

for $k \leq \frac{\log n}{2}$ if $|G| = 2^{o(n)}$. Hence there exists a full assignment $y \in \{0,1\}^n$ which is $\frac{1}{3}$-far from x and is "almost universal."

Now let x be the string that to which we have full access. The unknown string is chosen from the following two distributions.

- D_P: Uniform random G-shift of x.
- D_N: Uniform random G-shift of y.

Now we now that x and y are $\frac{1}{3}$-far. And since both are "almost universal," for all subset $P \subset [n]$ of size $(1-\gamma)\log n$ and all $Q \in \{0,1\}^k$,

$$2/3 \Pr_{\pi \in G}[x^\pi|_P = Q] \leq \Pr_{\pi}[y^\pi|_P = Q].$$

Now by Theorem 3.5 we can say that it will be impossible to test G-isomorphism with less than $(1-\gamma)\log n$ queries. So the query-1 complexity of any property tester of G-isomorphism is $\Omega(\log n)$.

□

6 Tight bounds and comparisons

In addition to our main result, Theorem 1.1, we obtain tight bounds for polynomial-sized groups. Note that these include all linear groups of bounded dimension. These in turn include most finite simple groups: all the classical finite simple groups of bounded dimension (linear, symplectic, orthogonal, and unitary groups) and all exceptional simple groups of Lie type, not only in their "natural" representations but in any representation (cf. [15]).

Corollary 6.1. *Let G be a transitive permutation group and $|G| = n^{O(1)}$. Then the query-1 complexity of 1-sided-error and 2-sided-error property testing of G-isomorphism is $\Theta(\log n)$.*

The next corollary gives an essentially tight bound for all groups of order $\exp(\text{polylog}(n))$. This includes all linear groups, and also includes all finite simple groups in any representation, except for the alternating groups.

Corollary 6.2. *Let G be a permutation group and assume $\log(|G|) = (\log n)^{O(1)}$. Then the query-2 complexity of 1-sided error property testing of G-isomorphism is $\Theta(\sqrt{n})$.*

For comparison with the results of Fischer amd Matsliah, we also include a table of corollaries of our results when specialized to Graph Isomorphism. In this case we take $n = \binom{|V|}{2}$ and define G to be $G = S_{|V|}^{(2)}$ (the induced symmetric group action) and hence $\log(|G|) = \log(|V|!) \sim \sqrt{n/2}\log n$. In the case of 1-sided error, query-2 complexity, the special cases of our general bounds (for primitive groups) match the Fischer–Matsliah bounds.

	Query-1 Complexity	**Query-2 Complexity**
1-sided-error testing	$\widetilde{\Theta}(\|V\|)^\dagger$	$\widetilde{\Theta}(\|V\|^{3/2})^\dagger$
2-sided-error testing	$\widetilde{O}(\|V\|)$, $\Omega(\log(\|V\|))$	$\widetilde{O}(\|V\|^{3/2})$

† Matches the Fischer–Matsliah bounds.
‡ New results.

Table 3: Corollaries of our results to Graph Isomorphism.

7 Future Work

We obtain tight bounds for the 1-sided error query complexity when the group is primitive. Obtaining tight bound for the 2-sided error query complexity would be the obvious next step. Also we want to obtain tight bounds in the case when the group is transitive and not just primitive. In the tight bounds for primitive groups that we obtain, we use the classification of primitive groups. But is the special structure of primitive groups essential or are there similar bounds for the general transitive groups?

A test case would be the automorphism group of a complete binary tree in its action on the leaves. We have reason to believe that a solution to this case would bring us close to solving the general case of transitive groups. Let T_N denote the complete binary tree with $N = 2^h$ leaves. Let G be the action of the automorphism group of the tree on the leaves. Given a string x of length N we place the bits of x on the leaves of the tree T_N. Then G permutes the bits of x. For this particular transitive group, the query-1 and query-2 complexities of testing G-isomorphism are wide open both in the 1-sided error and 2-sided error cases.

References

[1] N. Alon and A. Shapira. A characterization of the (natural) graph properties testabale with one-sided error. *Proceedings of the 46th IEEE FOCS*, pages 429–438, 2005.

[2] Noga Alon, Eldar Fischer, M. Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20:451–476, 2000.

[3] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: It's all about regularity. *Proceedings of the 38th ACM STOC*, pages 251–260, 2006.

[4] Noga Alon and Asaf Shapira. Testing subgraphs in directed graphs. *STOC*, pages 700–709, 2003.

[5] Noga Alon and Asaf Shapira. Linear equations, arithmetic progressions and hypergraph property testing. *SODA*, pages 708–717, 2005.

[6] Noga Alon and J.H. Spencer. The probabilistic method. *Wiley-Interscience (John Wiley & Sons), New York*, 1992.

[7] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and hardness of approximation problems. *J. ACM*, 45:501–555, 1998.

[8] Sanjeev Arora and S. Safra. Probabilistic checking of proofs: a new characterization of np. *J. ACM*, 45:70–122, 1998.

[9] László Babai. On the order of uniprimitive permutation groups. *Annals of Mathematics*, 113:553–568, 1981.

[10] László Babai. On the order of doubly transitive permutation groups. *Inventiones Math*, 65:473–484, 1982.

[11] László Babai, Lance Fortnow, and Carsten Lund. Nondeterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1:3–40, 1991.

[12] László Babai and E. M. Luks. Canonical labeling of graphs. *Proc. 15th STOC, ACM Press*, pages 171–183, 1983.

[13] László Babai, E.M. Luks, and Á. Seress. Permutation groups in nc. *Proc. 19th STOC, ACM Press*, pages 409–420, 1987.

[14] Manuel Blum, M. Luby, and Ronnit Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47:549–595, 1993.

[15] Peter J. Cameron. Finite permutation groups and finite simple groups. *Bull. London Math. Soc.*, 13:1–22, 1981.

[16] Eldar Fischer. The art of uninformed decisions: A primer to property testing. *Current Trends in Theoretical Computer Science: The Challenge of the New Century, G. Paun, G. Rozenberg and A. Salomaa (editors), World Scientific Publishing*, I:229–264, 2004.

[17] Eldar Fischer. The difficulty of testing for isomorphism against a graph that is given in advance. *SIAM Journal of Computing*, 34:1147–1158, 2005.

[18] Eldar Fischer and Arie Matsliah. Testing graph isomorphism. *SODA*, 2006.

[19] Eldar Fischer, Ilan Newman, and J. Sgall. Functions that have read-twice constant width branching programs are not necessarily testable. *Random Structures and Algorithms*, 24:175–193, 2004.

[20] Oded Goldreich, Saffi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *JACM*, 45(4):653–750, 1998.

[21] E. M. Luks. Isomorphism of graphs of bounded valence can be tested in polynomial time. *J.C.S.S.*, 25:42–65, 1982.

[22] Dana Ron. Property testing (a tutorial). *Handbook of Randomized Computing, (editors: S. Rajasekaran, P. M. Pardalos, J. H. Reif and J. D. P. Rolim eds), Kluwer Press*, II, 2001.

[23] Ronnit Rubinfeld and Madhu Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, 25:252–271, 1996.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/