Diameters of finite simple groups: sharp bounds and applications

By Martin W. Liebeck and Aner Shalev*

Abstract

Let $G$ be a finite simple group and let $S$ be a normal subset of $G$. We determine the diameter of the Cayley graph $\Gamma(G, S)$ associated with $G$ and $S$, up to a multiplicative constant. Many applications follow. For example, we deduce that there is a constant $c$ such that every element of $G$ is a product of $c$ involutions (and we generalize this to elements of arbitrary order). We also show that for any word $w = w(x_1, \ldots, x_d)$, there is a constant $c = c(w)$ such that for any simple group $G$ on which $w$ does not vanish, every element of $G$ is a product of $c$ values of $w$. From this we deduce that every verbal subgroup of a semisimple profinite group is closed. Other applications concern covering numbers, expanders, and random walks on finite simple groups.

1. Introduction

The purpose of this paper is to obtain sharp estimates on the diameters of Cayley graphs of finite simple groups with respect to certain generating sets. The generating sets with which we are concerned here are conjugacy classes, and more generally, normal subsets. (Recall that a normal subset of a group is a subset which is invariant under conjugation — that is, a union of conjugacy classes.) A normal subset $S$ of a finite simple group $G$ is said to be nontrivial if $S \not\subseteq \{1\}$. We denote by $S^m$ the set of all products of length $m$ of elements of $S$. Our main result is as follows.

Theorem 1.1. There exists a constant $c$ such that if $G$ is a finite simple group and $S \subseteq G$ is a nontrivial normal subset, then $S^m = G$ for any $m \geq c \log |G|/ \log |S|$. 

*1991 Mathematics Subject Classification: Primary 20D06; Secondary 20P05.

The second author is supported by the Israel Science Foundation.
Note that since $|S^m| \leq |S|^m$ we have $S^m \neq G$ for $m < \log |G|/ \log |S|$, and so the estimate in the theorem is best possible up to a multiplicative constant.

We shall obtain quite a range of applications of this theorem. The first application concerns Cayley graphs. Let $\Gamma(G,S)$ denote the Cayley graph of the group $G$ with respect to a generating subset $S$. Thus $\Gamma(G,S)$ is a directed graph, with vertex set $G$ and edges $(g,gs)$, $(g \in G, s \in S)$. The diameter $\text{diam} \Gamma$ of a directed graph $\Gamma$ is the maximal directed distance between two points in $\Gamma$. Theorem 1.1 immediately implies

**Corollary 1.2.** There is a constant $c$ such that whenever $S$ is a non-trivial normal subset of a finite simple group $G$,

$$\text{diam} \Gamma(G,S) \leq c \log |G|/ \log |S|.$$ 

In particular, the diameter is bounded when $|S| \geq |G|^\varepsilon$ where $\varepsilon > 0$ is bounded away from zero. Now recall that a finite undirected graph $\Gamma$ is defined to be a $\delta$-expander (where $\delta > 0$), if for all sets $A$ of vertices with $|A| \leq |\Gamma|/2$, we have $|\text{bdry}(A)| > \delta |A|$, where $\text{bdry}(A)$ is the boundary of $A$, that is, the set of vertices in $\Gamma \setminus A$ which are joined to some vertex in $A$. For definitions and background on expanders; see [26]. By a result of Babai and Szegedy [3] (see also [2]), any finite vertex-transitive graph $\Gamma$ is a $\delta$-expander with $\delta = 1/(\text{diam} \Gamma + 1/2)$. Combining this with Corollary 1.2 we obtain

**Corollary 1.3.** For any $\varepsilon > 0$ there exists $\delta > 0$ such that if $S$ is a normal subset of size at least $|G|^\varepsilon$ of a finite simple group $G$, with $S = S^{-1}$ (i.e. $s \in S \Rightarrow s^{-1} \in S$), then the Cayley graph $\Gamma(G,S)$ is a $\delta$-expander.

Of course, the expanders produced in this way have unbounded valency. Theorem 1.1 implies that, whenever $S$ is a normal subset of $G$ which is large in the sense that $|S| \geq |G|^\varepsilon$ ($\varepsilon > 0$), then $S^c = G$ for some constant $c$ depending only on $\varepsilon$. This general recipe produces various interesting results. For example, if $S$ is the set of involutions in $G$, then $|S| \geq c|G|^{1/2}$ by [24, 4.2 and 4.3]. We therefore obtain

**Corollary 1.4.** There exists a constant $c$ such that any element of any finite simple group $G$ is a product of $c$ involutions in $G$.

In fact a much more general result along these lines can be obtained by the same method (see §8 for the proof):

**Corollary 1.5.** Let $k \geq 2$ be an integer. There exists a constant $c$ (independent of $k$) such that, if $G$ is any finite simple group which contains an element of order $k$, then every element of $G$ is a product of $c$ elements of order $k$, all of which may be chosen from the same conjugacy class of $G$. 
In a similar vein, several results have been proved, showing that any element of a finite simple group $G$ is a short product of elements which are images of certain words: for example, in [32], Wilson proves that every element is a product of a bounded number of commutators, while in [30], [28] it is shown that for simple groups $G$, not of exponent dividing $n$, every element is a product of at most $f(n)$ elements which are $n^{th}$ powers.

We generalize these results to arbitrary words in the following theorem, which could be viewed as a noncommutative Waring-type theorem. Denote by $F_d$ the free group of rank $d$, with generators $x_1, \ldots, x_d$.

**Theorem 1.6.** For any word $1 \neq w(x_1, \ldots, x_d) \in F_d$ there exists a constant $c = c(w)$ with the following property. Let $G$ be a finite simple group such that $w$ is not identically $1$ in $G$. Then every element $g \in G$ can be written as a product of length $c$ of values of $w$.

Passing to profinite groups, we obtain the following.

**Corollary 1.7.** Let $G$ be a Cartesian product of infinitely many finite simple groups. Then any verbal subgroup of $G$ is closed.

Indeed, this follows since Theorem 1.6 shows that the verbal subgroup is a finite product of closed subsets of $G$. Up to now only a few instances of Corollary 1.7 had been established, dealing with the verbal subgroups $G'$ and $G^n$ (see [32], [30], [28]).

We now turn to some other types of applications of Theorem 1.1. Of particular interest is the case when the normal subset $S$ is equal to $C$, a conjugacy class in $G$. The covering number $\text{cn}(C, G)$ of $C$ in $G$ is defined to be the minimal $m$ with $C^m = G$ (see [1]). Theorem 1.1 yields a sharp bound on the covering number of an arbitrary class in an arbitrary finite simple group.

**Corollary 1.8.** There exists a constant $c$ such that if $G$ is a finite simple group and $C \subset G$ is a nontrivial conjugacy class, then

$$\text{cn}(C, G) \leq c \log |G| / \log |C|.$$
elements. Then $|G| \leq q^{8r^2}$, while $|C| \geq q^r$ for every nontrivial class $C$ in $G$ (these facts are well known, and can for example easily be deduced from [18, Tables 5.1 and Theorem 5.2.2]). This yields $\log |G|/ \log |C| \leq 8r$, so applying 1.8 we obtain

**Corollary 1.9.** There exists a constant $c$ such that if $G$ is a finite simple group of Lie type of Lie rank $r$ then

$$cn(G) \leq cr.$$ 

This is actually proved with an explicit constant in [8], [20].

The next application concerns generation properties. Define the *generation number* $gn(C; G)$ of a conjugacy class $C$ in a finite simple group $G$ to be the minimal $d$ such that $G$ can be generated by $d$ elements from $C$. Let the generation number $g(G)$ of $G$ be the maximum of $gn(C; G)$ over all nontrivial classes $C$ of $G$. Observe that

$$gn(G; C) \geq \log |G| / \log |C|,$$

since the index of the intersection of the centralizers of $d$ elements from $C$ is at most $|C|^d$. Using Corollary 1.8 it is easy to obtain the following.

**Corollary 1.10.** There exists a constant $c$ such that for any nontrivial conjugacy class $C$ in a finite simple group $G$ we have

$$gn(C; G) \leq c \log |G| / \log |C|.$$

Indeed, it is well known that any finite simple group can be generated by two elements. Let $x, y$ be generators for $G$. Write $x = x_1 \ldots x_m$ and $y = y_1 \ldots y_m$ where $x_i, y_i \in C$ and $m \leq c \log |G| / \log |C|$. Then $G$ is generated by the $2m$ elements $x_1, \ldots, x_m, y_1, \ldots, y_m$. In fact this argument can be slightly improved. By [11] every nontrivial element of a finite simple group $G$ is contained in a generating pair. Forming a generating pair with the first generator belonging to $C$ we see that, in the previous notation, $gn(C; G) \leq m + 1$.

If $G$ is a group of Lie type then maximizing over $C$ yields

**Corollary 1.11.** There exists a constant $c$ such that if $G$ is a rank $r$ simple group of Lie type then

$$g(G) \leq cr.$$ 

Results of this type – with explicit constants – are proved in [15], [14] and have found many applications.

Let us now state another by-product of the main result.
Theorem 1.12. There exists a constant $c$ such that if $G$ is a finite simple group, and $g \in G$ is chosen at random, then the probability that every element of $G$ is a product of $c$ conjugates of $g$ tends to 1 as $|G| \to \infty$.

The case of alternating groups is already known (see [4]), so our contribution is for groups of Lie type. Theorem 1.12 joins a growing number of recent results on random generation of finite simple groups (see for instance [17], [23], [24], [25], [12], [11], [13]).

Finally, our results and methods have some applications to random walks on finite simple groups. Recall that with the Cayley graph $\Gamma(G, S)$ one can associate a random walk as follows. Start at the identity, and when at vertex $g \in G$, choose at random a generator $s \in S$ and go on to the vertex $gs$. If $P^t(g)$ denotes the probability of reaching the vertex $g$ after $t$ steps, one is then interested in finding the minimal $t$ such that $P^t$ is close to the uniform distribution on $G$. Such a number $t$ is called the mixing time of $\Gamma(G, S)$; see Section 8 below for the precise definition, as well as Diaconis’ monograph [5] for background and details.

The case where $S = C$, a conjugacy class of a simple group $G$, is of particular interest, and has been studied in a number of papers, for example [7], [6], [29], [10], [27]. Two of the most general results available on the mixing time of Cayley graphs of type $\Gamma(G, C)$ are those of Roichman [29, 6.1] for alternating groups, and Gluck [10, 6.4] for groups of Lie type. Nevertheless, much remains to be done on this problem.

Since the diameter of $\Gamma(G, C)$ is roughly $\log |G|/\log |C|$ (up to a constant), it is natural to ask whether a similar estimate holds for the mixing time. However this is not the case: for instance, if $G = A_n$ and $C$ is a class of $n/2$-cycles then $\log |G|/\log |C|$ is constant, whereas by [29, 6.1] the respective mixing time is at least $c \log n$. Similar examples occur for groups of Lie type: for example, letting $C$ be the class of an $n/2$-dimensional Singer cycle in $G = \text{PSL}_n(q)$ when $n \to \infty$, we see that corresponding mixing time is unbounded (as a random bounded product of conjugates of $x$ acts trivially on a subspace of large dimension), whereas $\log |G|/\log |C|$ is bounded (it is about 4).

Since, by our main result, the diameter of $\Gamma(G, C)$ is bounded above by a constant, this example provides a negative solution to [27, Problem 4.13].

However, for particularly large classes $C$, we do obtain a satisfactory result.

Theorem 1.13. Let $G$ be a simple group of Lie type of rank $n$ over $\mathbb{F}_q$. Suppose $x \in G$ satisfies $|C_G(x)| \leq q^{(2-\varepsilon)n}$ for some $\varepsilon > 0$ and let $C = x^G$. Then the mixing time of the random walk on $\Gamma(G, C)$ is bounded above in terms of $\varepsilon$ alone.
The condition $|C_G(x)| \leq q^{(2-\varepsilon)n}$ holds for some interesting classes of elements, for example classes of regular elements; when $G$ is of type PSL, these include all cyclic matrices in $G$.

For a general class $C$ in a finite simple group $G$ we can still derive upper bounds on the mixing time, but these do not strike us as particularly strong. Indeed, working in continuous time Diaconis and Saloff-Coste [6] show that the mixing time of $\Gamma(G, C)$ is at most $cd^2\log |G|$, where $c$ is a constant and $d$ is the diameter of $\Gamma(G, C)$ (this can be deduced from [6, 2.3, 5.4]). Using Corollary 1.2 we deduce the following.

**Corollary 1.14.** For a finite simple group $G$ and a conjugacy class $C \subset G$, the mixing time of $\Gamma(G, C)$ is at most $c\log^3 |G|/\log^2 |C|$.

Our results rely on the classification of finite simple groups in a weak sense, namely, the existence of finitely many sporadic groups. However, for each of the results stated above, given that $G$ is a known simple group our proofs do not invoke the classification.

Let us now describe the structure of this paper. Sections 2–7 are devoted to the proof of Theorem 1.1. In Section 2 we reduce the theorem to the case where the set $S$ is a single conjugacy class. This case is then resolved in Section 3 for alternating groups and in Sections 4–7 for groups of Lie type. The case of alternating groups is rather easy, and the case of groups of Lie type of bounded rank is known. Hence most of the work is done for classical groups of large rank. The main idea in this case is to use a certain trick of merging short Jordan blocks to larger ones, until an element with a very small centralizer is produced. At this stage we apply character ratios with estimates on the number of conjugacy classes [22] to finish the proof. The applications are discussed in Section 8, where Corollary 1.5 and Theorems 1.6, 1.12 and 1.13 are proved.

We are grateful to Persi Diaconis and Bob Guralnick for useful discussions.

### 2. Reduction to conjugacy classes

In this section we reduce the proof of Theorem 1.1 to the case where the normal subset $S$ is a single conjugacy class of $G$. We begin with two lemmas for use in the proof for alternating groups.

**Lemma 2.1.** If $x \in S_n$ is a fixed-point-free permutation then $|C_{S_n}(x)| \leq (n!)^{1/2^{1+\varepsilon(n)}}$, where $\varepsilon(n) \to 0$ as $n \to \infty$.  


Proof. Using the well-known formula for the order of a centralizer in $S_n$ in terms of the cycle lengths of the permutation one easily sees (assuming $n$ is even) that the largest centralizer of a fixed-point-free permutation is attained by an involution, whose centralizer in $S_n$ has order $2^{n/2}(n/2)! \leq (n!)^{1/2+\varepsilon}$. The case of odd $n$ is similar and is left to the reader.

In the next lemma, for $x \in A_n$ we call the set of points in $\{1, 2, \ldots, n\}$ which are moved by $x$ the support of $x$.

**Lemma 2.2.** Let $C$ be a conjugacy class in $A_n$ and suppose the elements of $C$ have support of size $k$. Then for $\delta > 0$ we have

\[
n^{(1/3-\delta)k} \leq |C| \leq n^k,
\]

where the inequality on the left holds for $n$ large enough.

**Proof.** The inequality on the right is trivial, since

\[
|C| \leq \binom{n}{k} k! \leq n^k.
\]

To prove the inequality on the left let $x \in C$. We may suppose that $x$ moves $1, \ldots, k$ and leaves the rest of the letters fixed. Regarding $x$ also as an element of $S_k$ we obtain

\[C_{A_n}(x) \leq C_{S_k}(x) \times S_{n-k}.\]

Since $x$ is a fixed-point-free element of $S_k$ we obtain (using Lemma 2.1)

\[|C_{A_n}(x)| \leq (k!)^{1/2+\varepsilon}(n-k)!,\]

where $\varepsilon = \varepsilon(k) \to 0$ as $k \to \infty$. This yields

\[|C| = \frac{n!}{2|C_{A_n}(x)|} \geq \frac{1}{2} \binom{n}{k} (k!)^{1/2-\varepsilon}.
\]

Suppose $\delta > 0$ is given and that $n$ is large enough. If $k \leq n^{2/3}$ we obtain

\[
\binom{n}{k} = \frac{n(n-1) \cdots (n-k+1)}{k(k-1) \cdots 1} \geq \{n(n-1) \cdots (n-k+1)\}^{1/3} \geq 2n^{(1/3-\delta)k}.
\]

If $k > n^{2/3}$ we obtain

\[(k!)^{1/2-\varepsilon} \geq 2n^{(1/3-\delta)k}.
\]

In both cases the result follows.

**Lemma 2.3.** A finite simple group $G$ has at most $t^{3+o(1)}$ classes of size at most $t$. 

Proof. Denote by \( N(t) \) the number of classes of size at most \( t \) in \( G \). Suppose first that \( G = A_n \) (\( n \) large). Let \( C \subset G \) be a class of size at most \( t \), and let the elements of \( C \) have support of size \( k \). Assuming \( n \) is large and using the previous lemma we obtain
\[
n^{k/4} \leq |C| \leq t, \text{ so } k \leq 4 \log t/\log n.
\]
However, the number of classes in \( A_n \) of support \( k \) is at most \( 2^{p(k)} \), where \( p \) is the partition function. It is known that \( p(k) \leq c^{\sqrt{k}} \) for some constant \( c \). It follows that
\[
N(t) \leq \sum_{k \leq 4 \log t/\log n} c^{\sqrt{k}} \leq (4 \log t/\log n) \cdot c^{2 \sqrt{\log t/\log n}} = t^{o(1)}.
\]
Now assume \( G \) is a simple group of Lie type, and let \( q, r \) denote the field size and the Lie rank of \( G \) respectively. It is well known that the minimal size of a nontrivial class in \( G \) is at least \( q^r \) (this can easily be deduced from [18, 5.2.2], for example). Hence we may assume \( t \geq q^r \). On the other hand, by [22, Theorem 1], the number of classes \( k(G) \) in \( G \) is at most \( q^{3r} \). Hence we have \( N(t) \leq t^3 \).

**Corollary 2.4.** There exists \( \varepsilon > 0 \) such that any normal subset \( S \) of a finite simple group \( G \) contains a conjugacy class \( C \) such that \( |C| \geq |S|^\varepsilon \).

**Proof.** Let \( t \) be the maximal size of a class in \( S \). We have to show that \( |S| \leq t^c \) for some absolute constant \( c \). It suffices to show that the union of all classes in \( G \) of size at most \( t \) is of size at most \( t^c \). According to the previous lemma there are at most \( t^b \) such classes (for some fixed \( b \)), and so their union has size at most \( t^c \) for \( c = b + 1 \).

Theorem 1.1 is now easily reduced to the case where \( S \) is a conjugacy class. Indeed, given \( S \), let \( C \) be a class contained in it which satisfies \( |C| \geq |S|^\varepsilon \). Theorem 1.1 for the case of conjugacy classes yields \( C^m = G \) for all \( m \geq \epsilon \log |G|/\log |C| \). But \( \log |C| \geq \epsilon \log |S| \), so \( c \log |G|/\log |C| \leq c' \log |G|/\log |S| \) where \( c' = c\varepsilon^{-1} \). It follows that \( S^m = G \) for all \( m \geq c' \log |G|/\log |S| \), which is what we wanted.

### 3. Alternating groups

In Sections 3–7 we prove Theorem 1.1 in the case \( S = C \), a conjugacy class. The proof for alternating groups follows rather easily from results of Brenner [4].

**Lemma 3.1.** Let \( C \) be a fixed-point-free class in \( A_n \) (\( n \geq 9 \)). Then \( C^8 = A_n \).
Proof. A class \( C \) in \( A_n \) is said to be exceptional if it is not a class in \( S_n \). In this case the elements of \( C \) have distinct cycle lengths which are all odd.

Suppose first that \( C \) is nonexceptional. Then it follows from Theorem 3.05 of [4] that \( C^4 = A_n \). Now assume \( C \) is exceptional and let \( x \in C \). Since \( n \geq 9 \) and \( x \) is fixed-point-free by hypothesis, one of the cycles in \( x \) has odd length \( l \geq 7 \). So suppose \( x = yz \) where \( y = (12\ldots l) \) and \( z \) moves only \( l+1,\ldots,n \).

By Lemma 2.05 of [4], \( (y^h)^2 = A_l \setminus \{1\} \). In particular we can find \( w \in A_l \) such that the class of \( yy^w \) in \( A_l \) is both fixed-point-free and nonexceptional. This implies that the class of \( xx^w \) in \( A_n \) is fixed-point-free and nonexceptional. Thus \( C^2 \) contains some fixed-point-free nonexceptional class \( D \), so \( C^8 \) contains \( D^4 = A_n \).

Lemma 3.2. Let \( C \) be a class in \( A_n \) with elements having support of size \( k > 0 \). Then \( \text{cn}(C, A_n) \leq 16n/k \).

Proof. Write \( n = qk + r \) where \( 0 \leq r < k \). Let \( x \in C \) and suppose its support is \( \{1,\ldots,k\} \). Then for \( i = 1,\ldots,q \) there is \( x_i \in C \) whose support is \( \{(i-1)k+1,\ldots,(i-1)k+k\} \). Now the element \( y = x_1 \cdots x_q \in C^q \) can be viewed as a fixed-point-free permutation in \( A_{qk} \). Using the previous lemma (assuming \( qk \geq 9 \) as we may) we see that \( C^{8q} \) contains a subgroup \( A_{qk} \) of \( A_n \) embedded naturally. This means that \( C^{8q} \) contains all permutations of support at most \( qk \) in \( A_n \). Note that \( qk > n/2 \). Using this it is easy to write any element of \( A_n \) as a product of two elements of \( A_n \) of support \( \leq qk \): this can be done by splitting a length \( l \) cycle as a product of two cycles of lengths \( a,b \) whose supports overlap in one point (so that \( l = a + b - 1 \)). It now follows that \( C^{16q} = A_n \), yielding the result.

Theorem 3.3. Let \( C \) be a nontrivial class in \( A_n \) and let \( \varepsilon > 0 \). Then
\[
\text{cn}(C, A_n) \leq (16 + \varepsilon) \log |A_n|/\log |C|,
\]
provided \( n \) is large enough.

Proof. Let \( k \) be the support size of \( C \). By Lemma 2.2 we have \( |C| \leq n^k \). Hence
\[
\log |A_n|/\log |C| \geq (1 - o(1)) \frac{n \log n}{k \log n} = (1 - o(1)) \frac{n}{k}.
\]
The conclusion now follows using the previous lemma.

4. Preliminaries on groups of Lie type

Throughout, let \( G = G_n(q) \) be a quasisimple group of Lie type, of untwisted rank \( n \) over the field \( \mathbb{F}_q \). (The untwisted rank is the Lie rank of the corresponding simple algebraic group.)
Lemma 4.1. Theorem 1.1 holds if $G$ is of bounded rank.

Proof. Assume $G = G_n(q)$ is simple, with $n$ bounded, and let $1 \neq x \in G$ and $C = x^G$. By [1, Chap. 1, 10.1], for $g \in G$, the number of ways of writing $g$ as a product of $m$ conjugates of $x$ is

$$N = \frac{|x^G|^m}{|G|} \sum_{\chi \in \text{Irr}(G)} \frac{\chi(x)^m \chi(g^{-1})}{\chi(1)^{m-1}} \geq \frac{|x^G|^m}{|G|}(1 - \sum_{\chi \neq 1} (\frac{|\chi(x)|}{\chi(1)})^m \chi(1)^2).$$

By [22, Th. 1], $|\text{Irr}(G)| < (6q)^n$, and by [9], $\frac{|\chi(x)|}{\chi(1)} < c/q^{1/2}$ for $1 \neq \chi \in \text{Irr}(G)$. Combining this with the trivial bound $\chi(1)^2 \leq |G| \leq q^{8n^2}$, we see that the bracketed term above is at least

$$1 - (6q)^n (c/q^{1/2})^m q^{8n^2},$$

which, as $n$ is bounded, is positive for sufficiently large $m$. \qed

In view of Lemma 4.1, we now concentrate on classical groups of large rank.

Lemma 4.2. Let $G = G_n(q)$ be a perfect classical group $SL_n^\varepsilon(q)$, $Sp_{2n}(q)$, $SO_{2n}^\varepsilon(q)$ or $SO_{2n+1}(q)$, and let $G_1$ be the corresponding full isometry group $GL_n^\varepsilon(q)$, $Sp_{2n}(q)$, $GO_{2n}^\varepsilon(q)$ or $GO_{2n+1}(q)$. Let $1 \neq x \in G_1$, and suppose either that $n$ is bounded, or that $|C_{G_1}(x)| < q^{(2-\varepsilon)n}$ (where $\varepsilon > 0$). Then there is a constant $m_0$ (depending on $\varepsilon$ in the latter case) such that for all $m \geq m_0$,

$$(x^G)^m = x^m G.$$

Proof. Set $H = G \langle x \rangle$, and assume first that $|C_{G_1}(x)| < q^{(2-\varepsilon)n}$ and $n$ is large. Observe that for $\chi \in \text{Irr}(H)$ nonlinear, $|\chi(x)| < |C_H(x)|^{1/2} < q^{(1-\varepsilon/2)n}$; moreover, $\chi(1) \geq (q^n - 1)/2$ by [19], and $|\text{Irr}(H)| < q.(6q)^n$ by [22, Theorem 1]. Hence

$$\sum_{\chi \in \text{Irr}(H), \chi \text{ nonlinear}} \frac{|\chi(x)|^m}{\chi(1)^{m-2}} \leq q.(6q)^n q^{(1-\varepsilon/2)nm} \left(\frac{1}{2} (q^n - 1)^{m-2}\right),$$

which is less than 1 for sufficiently large $m, n$.

Now let $h \in G$, and $g = x^m h$. By [1, Chap. 1, 10.1], the number of ways of writing $g$ as a product of $m H$-conjugates of $x$ is

$$N = \frac{|x^H|^m}{|H|} \sum_{\chi \in \text{Irr}(H)} \frac{\chi(x)^m \chi(g^{-1})}{\chi(1)^{m-1}}.$$
Since \( g^{-1}x^m \in G \), the contribution of the linear characters to the above sum is \( |H : G| \). As shown above, the contribution of the nonlinear characters is less than 1; hence \( N > 0 \). Since \( g \) is an arbitrary element of \( x^m G \), it follows that \( x^m G = (x^H)^m = (x^G)^m \), as required.

In the case where \( n \) is bounded, we argue similarly, using [9] as in the previous lemma to bound the sum over nonlinear characters.

Lemma 4.3. Let \( T_k \in T_k, T_l \in T_l, D_{k-1} \in D_{k-1} \) and \( D_{l-1} \in D_{l-1} \). Then

\[
\begin{pmatrix}
D_{k-1} & T_k \\
T_l & D_{l-1}
\end{pmatrix}
= T_{k+l-1} \in T_{k+l-1}.
\]

Proof. The proof is trivial. \( \square \)

Lemma 4.4. If \( T_k \in T_k \) then \( |C_{\text{GL}_k(q)}(T_k)| < q^k \).

Proof. Let \( V = V_k(q) \) be the underlying vector space of dimension \( k \) over \( \mathbb{F}_q \). As \( T = T_k \) is a cyclic matrix, there is a vector \( v \in V \) such that \( V = \langle v, vT, vT^2, \ldots, vT^{k-1} \rangle \). If \( C \in C_{\text{GL}_k(q)}(T) \) then \( C \) is uniquely determined by the vector \( w = vC \), since for any \( i \), \( vT^iC = vCT^i = wT^i \). Therefore

\[
|C_{\text{GL}_k(q)}(T_k)| \leq |V - \{0\}| = q^k - 1.
\]

5. Proof of Theorem 1.1 for \( L_n(q) \)

In this section, let \( G = \text{SL}_n(q) \). For a subset \( Y \) of \( \text{GL}_n(q) \), define \( Y^G = \bigcup_{y \in Y} y^G \).

Lemma 5.1. Let \( r \geq 2 \), \( X = \text{SL}_r(q) \times I_{n-r} \leq G \) and let \( t \in \text{GL}_r(q) \times D_{n-r} \). Then there is a constant \( c \) such that for any \( m \geq cn/r \),

\[
((tX)^G)^m = t^m G.
\]
Proof. Let \( t = \text{diag}(t_1, D) \in \text{GL}_r(q) \times D_{n-r} \), and let \( \lambda = \det t_1 \). Define \( J_r(\lambda) \) to be the \( r \times r \) matrix
\[
J_r(\lambda) = \begin{pmatrix}
\lambda & \lambda & \cdots \\
1 & 1 & \cdots \\
& \ddots & \ddots & \ddots \\
& & 1 & 1 & 1
\end{pmatrix}
\] (with 0 in all blank space). Note that \( \text{diag}(J_r(\lambda), D) \in tX \).

Write \( D = \text{diag}(\alpha_1, \ldots, \alpha_{n-r}) \), and for \( i < j \leq n-r \) set \( D_{i,j} = \text{diag}(\alpha_i, \ldots, \alpha_j) \).

Let \( n = a(r-1) + s \) with \( 0 \leq s < r-1 \). For \( 0 \leq k \leq a-1 \), define
\[
u_k = \begin{pmatrix}
D_{1,k(r-1)} & J_r(\lambda) \\
& D_{k(r-1)+1,n-r}
\end{pmatrix}
\]
and set
\[
u_a = \begin{pmatrix}
D \\
& I_{r-s} \\
& & J_a(\lambda)
\end{pmatrix}.
\]

Then each \( u_i \in (tX)^G \). Moreover, by Lemma 4.3 we have
\[
u_a u_{a-1} \cdots u_1 u_0 = T \in T_n.
\]
It follows that \((tX)^G\) contains \( T \). By Lemma 4.4, \(|C_G(T)| \leq q^n\). Hence Lemma 4.2 implies that there is a constant \( c \) such that for \( m \geq c \), \((T^G)^m = T^m G\). The conclusion follows.

**Lemma 5.2.** Suppose \( n_1, \ldots, n_k \) are integers such that each \( n_i \geq 2 \) and \( r = \sum n_i \leq n \). Let
\[
Y = \text{SL}_{n_1}(q) \times \cdots \times \text{SL}_{n_k}(q) \times I_{n-r} \leq G,
\]
and let
\[
t \in \text{GL}_{n_1}(q) \times \cdots \times \text{GL}_{n_k}(q) \times D_{n-r} \leq \text{GL}_{n}(q).
\]

Then there is a constant \( c \) such that for any \( m \geq cn/r \),
\[
((tY)^G)^m = t^m G.
\]
Proof. Write 
\[ t = (t_1, \ldots, t_k, D) \in \text{GL}_{n_1}(q) \times \cdots \times \text{GL}_{n_k}(q) \times D_{n-r} \]
and let \( \lambda_i = \det t_i \). For \( 3 \leq i \leq k \), define
\[
v_i = \left( \begin{array}{c} I_{n_i-1}^{(n_i-1)}(n_i-1) \\ \vdots \\ J_{n_i}(\lambda_i) \\ I_{n-(1+\sum_i(n_i-1))} \end{array} \right).
\]
Further, if \( k = 2m \) is even, set
\[
v_1 = \text{diag}(J_{n_1}(\lambda_1), I_{r-n_1-m}, \lambda_2, \ldots, \lambda_{2m}, D),
v_2 = \text{diag}(I_{n_1-1}, J_{n_2}(\lambda_2), I_{r-n_1-n_2-m+1}, \lambda_1, \lambda_3, \ldots, \lambda_{2m-1}, D).
\]
And if \( k = 2m-1 \) is odd, define \( v_1, v_2 \) in the same way, but omitting \( \lambda_{2m} \) from \( v_1 \) (and having \( I_{r-n_1-m+1} \) in the middle).

Observe that \([v_i, v_j] = 1\) when \( |i - j| \geq 2\). Suppose \( k = 2m \). Let
\[
u = v_1 v_3 \cdots v_{2m-1}, \quad v = v_2 v_4 \cdots v_{2m}.
\]
Then both \( u \) and \( v \) lie in \((tY)^G\). Observe that
\[
uv = v_1(v_3v_2)(v_5v_4) \cdots (v_{2m-1}v_{2m-2})v_{2m}.
\]
By Lemma 4.3,
\[
v_3v_2 = \text{diag}(D_{n_1-1}, T_{n_2-1+n_3}, D_{n_1-n_2-n_3+2}),
v_5v_4 = \text{diag}(D_{n_1-1+n_2-1+n_3}, T_{n_4-1+n_5}, D_{n-(n_1+\cdots+n_5)+3}),
\]
and so on, where \( D_i \in D_i, T_i \in T_i \). Using this it is straightforward to check that \( uv \) is conjugate to \( \text{diag}(T_s, D_{n-s}) \), where \( s = 1 + \sum_i(n_i - 1) \) and \( T_s \in T_s, D_s \in D_s \). The same holds if \( k = 2m - 1 \), omitting the \( v_{2m} \) term.

Thus \( ((tY)^G)^2 \) contains \( w = \text{diag}(T_s, D_{n-s}) \). Put \( X = \text{SL}_s(q) \times I_{n-s} < G \).

By Lemma 4.4, \( |C_{\text{GL}_s(q)}(T_s)| < q^s \), and hence by Lemma 4.2, there is a constant \( d' \) such that for \( d \geq d' \) we have \( (w^X)^d = w^d X \). Then \( (tY)^G)^{2d} \) contains \( t^{2d} X \).

By Lemma 5.1, there is a constant \( e \) such that for \( m \geq en/s \), we have
\[
((t^{2d}X)^G)^m = t^{2dm} G.
\]
The conclusion follows. \( \square \)

We now discuss the general form of an element of \( G = \text{SL}_n(q) \). Denote by \( J_k \) the unipotent \( k \times k \) Jordan block matrix
\[
\begin{pmatrix}
1 & 1 & \cdots \\
1 & 1 & \cdots \\
& & \cdots \\
1 & 1 \\
1
\end{pmatrix}.
\]
For \( x \in G \), we have \( x = su \), a uniquely determined commuting product of a semisimple element \( s \) and a unipotent element \( u \), and \( C_G(x) = C_G(s) \cap C_G(u) \). The element \( s \) is conjugate to a matrix of the form

\[
\text{diag}(\alpha_1 I_{k_1}, \ldots, \alpha_r I_{k_r}, \lambda_1, \ldots, \lambda_1, \ldots, \lambda_t, \ldots, \lambda_t),
\]

where each \( \alpha_j \in \mathbb{F}_q^* \), each \( \lambda_j \) is an irreducible \( a_j \times a_j \) matrix \((a_j \geq 2)\), occurring \( b_j \) times in the above, \( \sum k_i + \sum a_j b_j = n \); moreover

\[
C_{\text{GL}_m(q)}(s) \cong \prod_{j=1}^r \text{GL}_{k_j}(q) \times \prod_{j=1}^t \text{GL}_{b_j}(q^{a_j}).
\]

Then \( u \) is a unipotent element in this centralizer: say the projection of \( u \) in \( \text{GL}_{k_j}(q) \) is conjugate to \((J_i^{m_{ij}})\) (a matrix with each Jordan block \( J_i \) occurring \( m_{ij} \) times, where \( \sum_i m_{ij} = k_j \), and the projection in \( \text{GL}_{b_j}(q^{a_j}) \) is \((J_i^{n_{ij}})\), where \( \sum_i n_{ij} = b_j \). Thus \( x = su \) is conjugate to

\[
(\dagger) \quad \text{diag}((\alpha_1 J_i^{m_{i1}}, \ldots, \alpha_r J_i^{m_{ir}}), \lambda_1 \otimes (J_i^{n_{i1}}), \ldots, \lambda_t \otimes (J_i^{n_{it}})).
\]

We may take \( m_{i1} \geq m_{i2} \geq \ldots \geq m_{ir} \).

**Lemma 5.3.** There is a constant \( c \) such that if \( x \in G \) is as in \((\dagger)\) above, then

\[
\frac{\log |G|}{\log |x^G|} \geq \frac{cn^2}{n^2 - m_{i1}^2}.
\]

**Proof.** Obviously \( C_G(x) \) contains a subgroup \( \text{SL}_{m_{i1}}(q) \times I_{n-m_{i1}} \), so \( |x^G| \leq c n^2 - m_{i1}^2 \), and the result follows. \( \square \)

**Lemma 5.4.** There is a constant \( c' \) such that if \( 1 \neq x \in G \) is as in \((\dagger)\) above, and \( m \geq c'n/(n - m_{i1}) \), then

\[
(x^G)^m = G.
\]

**Proof.** Write

\[
x = \text{diag}(\alpha_1 I_{m_{i1}}, \ldots, \alpha_r I_{m_{ir}}, B_1, \ldots, B_t, C_1, \ldots, C_p),
\]

where the \( B_i \) are all the \( 2 \times 2 \) blocks in \((\dagger)\) (i.e. the blocks \( \alpha_j J_2 \) and the blocks of the form \( \lambda_j \otimes J_1 \) with \( a_j = 2 \)), and the \( C_i \) are the remaining blocks of the form \( \alpha_j J_k \) \((k \geq 3)\) and \( \lambda_j \otimes J_k \) \((a_j k \geq 3)\).

Say \( C_i \) is \( c_i \times c_i \) (so \( c_i \geq 3 \)). If \( C_i = \alpha_j J_k \) then \( |C_{\text{GL}_{c_i}(q)}(C_i)| = (q-1)q^{c_i-1} \); and if \( C_i = \lambda_j \otimes J_k \) then \( |C_{\text{GL}_{c_i}(q)}(C_i)| = (q^{a_j} - 1)q^{a_j-1} \leq q^{a_j} \). Hence in any case, Lemma 4.2 gives

\[
(1) \quad (C_i^{\text{SL}_{c_i}(q)})^m = C_i^m \text{SL}_{c_i}(q)
\]

for all \( m \geq c \), a constant.
Now define \( D_i = \text{diag}(B_{2l-1}, B_{2i}) \in \text{GL}_4(q) \) \((1 \leq i \leq [l/2])\), and let \( E_i \) \((1 \leq i \leq v)\) be 3×3 nonscalar diagonal matrices with eigenvalues among the \( \alpha_j \), the number \( v \) of \( E_i \) being maximal such that each \( \alpha_j \) appears as an eigenvalue of \( \text{diag}(E_1, \ldots, E_v) \) with multiplicity at most \( m_{1j} \). Easily, \( w = \sum^r_i m_{1j} - 3v \leq m_{111} \). Moreover, by Lemma 4.2, there is a constant \( c' \) such that for \( m \geq c' \),

\[(D^3_{i\text{SL}}(q))^{m} = D^m_{i\text{SL}}(q), \quad (E^3_{i\text{SL}}(q))^{m} = E^m_{i\text{SL}}(q).\]

For \( l \) even, let

\[Y = \prod^m_i \text{SL}_{c_i}(q) \times \prod^{l/2}_i \text{SL}_4(q) \times \prod^v_i \text{SL}_3(q) \times I_w,\]

chosen so that \( x \in N_{G}(Y) \) in the obvious way. (If \( l \) is odd take the middle sum to have upper limit \((l - 1)/2\) and the right-hand term to be \( I_{w+2} \).) From (1) and (2) we conclude that for \( m \geq \max(c, c') \), \( (x^Y)^m \) contains \( x^mY \). Since \( w \leq m_{111} \), the conclusion now follows from Lemma 5.2.

Theorem 1.1 for \( L_n(q) \) follows from the previous two lemmas, since

\[
\frac{c'n}{n-m_{11}} = \frac{c'n(n+m_{11})}{n^2-m^2} < \frac{2c'n^2}{n^2-m^2} < c'' \log |G|/\log |x^G|.
\]

6. Proof of Theorem 1.1 for symplectic and orthogonal groups

In this section let \( G = \text{Sp}(V) \) or \( \text{SO}(V)' \), a symplectic or orthogonal group on \( V \) over \( \mathbb{F}_q \) of BN-rank \( l \); thus \( G \cong \text{Sp}_{2l}(q), \text{SO}_{2l+1}(q)' \), \( \text{SO}_{2l+1}(q)' \) or \( \text{SO}_{2l+2}(q)' \). Let \( I(V) \) denote the corresponding full isometry group \( \text{Sp}(V) \) or \( \text{GO}(V) \), and write \( n = \dim V \).

If \( W, W' \) are totally singular subspaces of \( V \) of equal dimension \( r \) such that \( W \oplus W' \) is nondegenerate, then \( \text{Sp}(V), \text{SO}(V) \) contain a subgroup \( \text{GL}(W, W') \cong \text{GL}_r(q) \) inducing \( \text{GL}_r(q) \) on \( W \) and \( W' \), and acting trivially on \( (W \oplus W')' \). Let \( S(W, W') \) be the subgroup \( \text{SL}_r(q) \) of this, so \( S(W, W') < G \). For \( g \in \text{GL}_r(q) \), write \( g(W, W') \) for the corresponding element of \( \text{GL}(W, W') \).

**Lemma 6.1.** Let \( \dim W = \dim W' = r \geq 2 \), \((r, q) \neq (2, 2), (2, 3)\), and \( X = S(W, W') \) as above. Let \( t \in \text{GL}(W, W') \). There is a constant \( c \) such that for any \( m \geq cn/r \),

\[
((tX)^G)^m = t^m G.
\]

**Proof.** Choose a basis of \( V \) containing vectors \( e_i, f_i \) \((1 \leq i \leq l)\), where \( E = \langle e_i : 1 \leq i \leq l \rangle \) and \( E' = \langle f_i : 1 \leq i \leq l \rangle \) are maximal totally singular subspaces of \( V \), and \((e_i, f_i) = 1\) for all \( i \). Write \( l = a(r-1)+s \) with \( 0 \leq s \leq r-1 \), and for \( 0 \leq k \leq a-1 \) define
\[ W_k = \langle e_i : k(r - 1) + 1 \leq i \leq (k + 1)r \rangle, \]
\[ W'_k = \langle f_i : k(r - 1) + 1 \leq i \leq (k + 1)r \rangle. \]

Set
\[ W_a = \langle e_i : a(r - 1) + 1 \leq i \leq n \rangle, \]
\[ W'_a = \langle f_i : a(r - 1) + 1 \leq i \leq n \rangle. \]

Let \( \lambda \) be the determinant of \( t \) as an element of \( \text{GL}_r(q) \), and define \( J_c(\lambda) \) as in the previous section. For \( 0 \leq k \leq a - 1 \), set \( u_k = J_r(\lambda)(W_k, W'_k) \), and set \( u_a = J_a(\lambda)(W_a, W'_a) \). As in the proof of Lemma 5.1 we have
\[ u_a u_{a-1} \ldots u_0 = T_l(E, E'), \]
where \( T_l \in T_l \) (as an element of \( \text{GL}_d(q) \)). Thus \( ((tX)^G)^{a+1} \) contains \( T_l(E, E') \).

Hence by Lemma 4.2, there is a constant \( c \) such that for \( m \geq c(a+1) \), \( ((tX)^G)^m \) contains \( t^mS(E, E') \).

For a fixed such value of \( m \), pick an irreducible element \( y \in \text{GL}_d(q) \) of determinant \( \lambda^m \), such that \( C_{\text{GL}_d(q)}(y) \leq \text{GL}_1(q^l) \). Then \( y(E, E') \in t^mS(E, E') \), and moreover, \( |C_G(y(E, E'))| \leq |\text{GL}_1(q^l)|(q + 1) \). Hence by Lemma 4.2, for \( m \geq c'c(a+1) \), we have
\[ ((tX)^G)^m = t^m G, \]
as required.

**Lemma 6.2.** Suppose \( W_1 \oplus W'_1, \ldots, W_k \oplus W'_k \) are mutually orthogonal non-degenerate subspaces, with all \( W_i, W'_i \) totally singular, and \( \dim W_i = \dim W'_i = n_i \geq 3 \). Write \( r = \sum{n_i} \). Let
\[ Y = \prod_{i=1}^{k} S(W_i, W'_i) \leq G, \]
(so that \( Y \cong \prod \text{SL}_{n_i}(q) \)), and let \( t \in \prod \text{GL}(W_i, W'_i) \). There is a constant \( c \) such that for \( m \geq cn/r \),
\[ ((tY)^G)^m = t^m G. \]

**Proof.** Let \( t \) map to \( (t_1, \ldots, t_k) \in \prod \text{GL}_{n_i}(q) \), and write \( \lambda_i = \det t_i \) for \( 1 \leq i \leq k \). We work with a basis containing vectors \( e_i, f_i \) (\( 1 \leq i \leq l \)) as in the previous proof. Define \( r_i = \sum_{j=1}^{i-1}(n_j - 1) \), and set
\[ X_i = \langle e_j : r_i + 1 \leq j \leq r_i + n_i \rangle, \quad X'_i = \langle e_j : r_i + 1 \leq j \leq r_i + n_i \rangle \]
for all \( i \). Write \( s = 1 + \sum(n_i - 1) \), and put \( W = \sum X_i = \langle e_i : 1 \leq i \leq s \rangle, W' = \sum X'_i = \langle f_i : 1 \leq i \leq s \rangle. \)
Now define \( v_1, \ldots, v_k \) in similar fashion to the proof of Lemma 5.2. Thus for \( i \geq 3 \),
\[
v_i = J_{n_i}(\lambda_i)(X_i, X'_i),
\]
while (for \( k = 2m \))
\[
v_1 = \text{diag}(J_{n_1}(\lambda_1), I_{s-n_1-m}, \lambda_2, \lambda_4, \ldots, \lambda_{2m}, I_{l-s})(W, W'),
\]
\[
v_2 = \text{diag}(\lambda_1, J_{n_1-2}, J_{n_2}(\lambda_2), I_{s-n_1-n_2-m+2}, \lambda_3, \lambda_5, \ldots, \lambda_{2m-1}, I_{l-s})(W, W')
\]

(there is no \( \lambda_{2m} \) term for \( k = 2m - 1 \)).

Take \( k = 2m \) or \( 2m - 1 \), and let \( u = v_1v_3 \ldots v_{2m-1} \), \( v = v_2v_4 \ldots v_{2m-2}v_{2m} \)
(no \( v_{2m} \) term if \( k = 2m - 1 \)). Then \( u, v \in (tY)^G \), and we see as in the proof of
Lemma 5.2 that uv is conjugate to \( T_s(W, W') \) (where \( T_s \in T_s \) as a member of
\( \text{GL}_s(q) \)). Therefore by 4.2, \((tY)^G\) contains \( t^jS(W, W') \) for \( j \geq c \). Now the
result follows from Lemma 6.1.

We now discuss the general form of elements of \( G = \text{Sp}(V) \) or \( \text{SO}(V)' \). For
\( x \in G \) we have \( x = su \), a unique commuting product of a semisimple element
\( s \) and a unipotent element \( u \). We have
\[
V \downarrow \langle s \rangle = V_1 \downarrow V_- \downarrow \sum(Y_i \oplus Y'_i) \downarrow \sum Z_i,
\]
where \( V_1 = C_V(s), V_- = C_V(-s), Y_i, Y'_i \) are totally singular with \( Y_i \oplus Y'_i \)
nondegenerate, \( Z_i \) are nondegenerate, and \( Y_i, Y'_i, Z_i \) are all irreducible \( \langle s \rangle \)-modules. Collect the \( Y_i, Y'_i, Z_i \) into homogeneous components \( W_j, W'_j, X_j \), and say
\[
s^{W_j} = \text{diag}(\lambda_1, \ldots, \lambda_j) \ (\lambda_j \text{ irreducible } a_j \times a_j, \text{ and } b_j \text{ copies}),
\]
\[
s^{X_j} = \text{diag}(\mu_1, \ldots, \mu_j) \ (\mu_j \text{ irreducible } c_j \times c_j, \text{ and } d_j \text{ copies}).
\]

Then
\[
C_I(V)(s) = I(V_1) \times I(V_-) \times \prod \text{GL}_{b_j}(q^{a_j}) \times \prod \text{GU}_{d_j}(q^{c_j/2}).
\]
Thus \( u \) is a unipotent element in this centralizer, and \( x = su \) is conjugate to a
matrix of the form
\[
\text{diag}(u^{V_1}, -u^{V_-}, (\lambda_1 \otimes (J_{l^{m_{11}}})))(W_1, W'_1), \ldots, (\lambda_k \otimes (J_{l^{m_{kk}}})))(W_k, W'_k),
\]
\[
\mu_1 \otimes (J_{l^{m_{11}}}), \ldots, \mu_1 \otimes (J_{l^{m_{kk}}}).
\]

In the symplectic case, by [31, pp. 34–38], \( u^{V_1} \) and \( u^{V_-} \) contain Jordan blocks
\( J_i (i \text{ odd}) \) with even multiplicity, and blocks \( J_i (i \text{ even}) \) act on nondegenerate
subspaces; thus \( u^{V_{1\pm1}} \) is an orthogonal direct sum
\[
\bigoplus_{i \text{ even}} J_i \downarrow \bigoplus_{i \text{ odd}} J_i(E_i, E'_i),
\]
where $E_i, E'_i$ are totally singular subspaces and $E_i \oplus E'_i$ is nondegenerate. A similar discussion applies in the orthogonal case, interchanging the words “even” and “odd” when $p \neq 2$.

Let $m_\varepsilon$ be the number of $J_1$ blocks in $u^\varepsilon$, ($\varepsilon = \pm 1$), and let $m = \max(m_1, m_{-1})$.

**Lemma 6.3.** For $x = su$ as above, we have
\[
\frac{\log |G|}{\log |x^G|} \geq \frac{cn^2}{n^2 - m^2}.
\]

**Proof.** In the symplectic case, $C_G(x)$ contains $\Sp_m(q)$, so
\[
|x^G| \leq cq^{n(n+1)/2-m(m+1)/2};
\]
whence
\[
\frac{\log |G|}{\log |x^G|} \geq \frac{c'n(n+1)}{n(n+1) - m(m+1)},
\]
and the conclusion follows. The same argument applies in the orthogonal case. \(\Box\)

**Lemma 6.4.** For $x = su$ as above, there is a constant $c$ such that for $d \geq cn/(n - m)$, we have
\[
(x^G)^d = G.
\]

**Proof.** Observe that
\[
|C_{\GL_n(q)}(J_i)| < cq^i, \quad |C_{\GL_k(q)}(\lambda_j \otimes J_i)| < cq^k (a_ji = k),
\]
and for $X = \Sp_r(q)$ or $\SO_r(q)$ with $r = 2k$ or $2k + 1$,
\[
|C_X(J_i)| < cq^k, \quad |C_X(\mu_j \otimes J_i)| < cq^k (c_ji = r).
\]
Using Lemma 4.2, and arguing as in the proof of Lemma 5.4, we deduce that there is a constant $c$, and a subgroup
\[
Y = \prod I_{n_i}(q) \times \prod \GL(W_i, W'_i) \times \{\pm I_w\},
\]
such that $x \in Y$ and $(x^G)^d$ contains $x^dY'$ for $d \geq c$, where each $I_{n_i}(q) = \Sp_{n_i}(q)$ or $\SO_{n_i}(q)$, each $n_i \geq 6$, each $\dim W_i = \dim W'_i \geq 3$, $I_w$ is the $w \times w$ identity and $w = \dim V - \sum n_i - 2 \sum \dim W_i \leq m$. Now the conclusion follows from Lemma 6.2. \(\Box\)

As deduced at the end of the previous section, Theorem 1.1 for symplectic and orthogonal groups follows immediately the previous two lemmas.
7. Proof of Theorem 1.1 for unitary groups

The proof of Theorem 1.1 for unitary groups is very similar to that for the classical groups already handled. Let \( G = \text{SU}_n(q) \), with natural module \( V = \text{V}_n(q^2) \). If \( W, W' \) are totally singular subspaces of \( V \) of equal dimension \( r \) such that \( W \oplus W' \) is nondegenerate, write \( \text{GL}(W, W') \) for the subgroup of \( G \) inducing \( \text{GL}_r(q^2) \) on \( W \) and \( W' \), and acting trivially on \( (W \oplus W')^\perp \). Let \( S(W, W') \) be the subgroup \( \text{SL}_r(q^2) \) of \( \text{GL}(W, W') \). For \( g \in \text{GL}_r(q^2) \) write \( g(W, W') \) for the corresponding element of \( \text{GL}(W, W') \). The proofs of Lemmas 6.1 and 6.2 give

**Lemma 7.1.** Suppose \( W_1 \oplus W_1', \ldots, W_k \oplus W_k' \) are mutually orthogonal non-degenerate subspaces, with all \( W_i, W_i' \) totally singular, and \( \dim W_i = \dim W_i' = n_i \geq 3 \). Write \( r = \sum n_i \). Let

\[
Y = \prod_{i=1}^{k} S(W_i, W_i') \leq G,
\]

and let \( t \in \prod \text{GL}(W_i, W_i') \). Then there is a constant \( c \) such that for \( m \geq cn/r \),

\[
((tY)^G)^m = t^m G.
\]

Now let \( x = su \in G \), a commuting product of a semisimple element \( s \) and a unipotent element \( u \). Then

\[
V \downarrow \langle s \rangle = V_{\alpha_1} \downarrow \ldots \downarrow V_{\alpha_k} \downarrow \sum (Y_i \oplus Y_i') \downarrow \sum Z_i,
\]

where each \( V_{\alpha_i} \) is the nondegenerate eigenspace for eigenvalue \( \alpha_i (\alpha_i^{q+1} = 1) \), \( Y_i, Y_i' \) are totally singular with \( Y_i \oplus Y_i' \) nondegenerate, \( Z_i \) are nondegenerate with \( \dim Z_i \geq 2 \), and \( Y_i, Y_i', Z_i \) are all irreducible \( \langle s \rangle \)-modules. Collect the \( Y_i, Y_i', Z_i \) into homogeneous components \( W_j, W_j', X_j \), and say

\[
s^{W_j} = \text{diag}(\lambda_j, \ldots, \lambda_j) \quad (\lambda_j \text{ irreducible } a_j \times a_j, \text{ and } b_j \text{ copies}),
\]

\[
s^{X_j} = \text{diag}(\mu_j, \ldots, \mu_j) \quad (\mu_j \text{ irreducible } c_j \times c_j, \text{ and } d_j \text{ copies}).
\]

Then

\[
C_{\text{GU}(V)}(s) = \prod \text{GU}(V_{\alpha_i}) \times \prod \text{GL}_{b_j}(q^{2a_j}) \times \prod \text{GU}_{d_j}(q^{c_j}).
\]

Thus \( u \) is a unipotent element in this centralizer, and \( x = su \) is conjugate to a matrix of the form

\[
diag((\alpha_1(J_{p_{11}}^{(i)}) \ldots, \alpha_k(J_{p_{1k}}^{(i)})), (\lambda_1 \otimes (J_{p_{21}}^{(i)}))(W_1, W_1'), \ldots, (\lambda_k \otimes (J_{p_{2k}}^{(i)}))(W_k, W_k'),
\]

\[
\mu_1 \otimes (J_{p_{31}}^{(i)}), \ldots, \mu_l \otimes (J_{p_{3l}}^{(i)})).
\]

Take \( p_{11} \geq p_{12} \geq \ldots \geq p_{1l} \).
Since $C_G(x)$ contains $SU_{p_{11}}(q)$, we see that
$$\frac{\log |G|}{\log |x^G|} \geq \frac{cn^2}{n^2 - p_{11}^2}.$$ And arguing as in the proof of Lemma 6.4, we see that there is a constant $c'$ such that for $m \geq c'n/(n - p_{11})$, we have
$$(x^G)^m = G.$$ The result follows in the usual way.

8. Applications

Most of the consequences of the main results have already been derived in the introduction. It remains to prove Corollary 1.5 and Theorems 1.6, 1.12 and 1.13.

Proof of Corollary 1.5. For this it suffices by Theorem 1.1 to prove the following.

**Lemma 8.1.** Let $k \geq 2$ be an integer. There exists $\varepsilon > 0$ (independent of $k$) such that if $G$ is a finite simple group containing an element of order $k$, then $G$ contains a conjugacy class $C$ of elements of order $k$, with $|C| > |G|^\varepsilon$.

**Proof.** For $G = A_n$ choose an element $x \in G$ of order $k$ with support $\Delta$ of maximal possible size. Then $|\Delta| > n/2$, since otherwise we could multiply $x$ by an element $y$ acting like $x$ on the complement of $\Delta$, and then $xy$ would be an element of order $k$ with larger support than $x$. Now the result follows from Lemma 2.2.

Now consider $G = G(q)$, a simple group of Lie type over $\mathbb{F}_q$. Since the size of any conjugacy class of $G$ is at least $cq$ ($c$ a constant), we may take $G$ to be of unbounded rank, hence to be a classical group of large dimension. Let $V$ be the natural module for $G$, and let $n = \dim V$. As in [25], for an element $h \in G$, define $\nu(h)$ as follows: let $\hat{V} = V \otimes \mathbb{F}_q$, let $\hat{h}$ be an element of $\text{SL}(V)$ which maps to $h$ modulo scalars, and set
$$\nu(h) = \min \{\dim[\hat{V}, \lambda \hat{h}] : \lambda \in \mathbb{F}_q\}.$$ Pick an element $x \in G$ of order $k$ with $\nu(x)$ a large as possible. We claim that $\nu(x) > n/8$. For suppose $\nu(x) \leq n/8$. Then there exists $\lambda \in \mathbb{F}_q$ such that $\dim[\hat{V}, \lambda \hat{x}] \leq n/8$. Thus $\dim C_V(\lambda \hat{x}) \geq 7n/8$. It follows that $\lambda \in \mathbb{F}_q$ (if $G$ is unitary), and also $\lambda = \pm 1$ if $G$ is symplectic or orthogonal. Moreover, there is a subspace $W$ (nondegenerate if $G \neq L_n(q)$) such that $\dim W \geq n/4$, and $V = W \oplus W'$ with $\hat{x}$ fixing $W, W'$ and acting as $\lambda^{-1}I$ on $W$. We can find a
nontrivial element \( y \in G \) of order dividing \( k \) such that \( y \) fixes \( W \) and \( W' \) and acts trivially on \( W' \). Then \( xy \) has order \( k \) but \( \nu(xy) > \nu(x) \), contradicting the choice of \( x \).

Thus \( \nu(x) > n/8 \), as claimed. Write \( x = su \), a product of commuting semisimple and unipotent elements \( s, u \in G \). Then either \( \nu(s) > n/16 \) or \( \nu(u) > n/16 \). Now the result follows from the proof of [25, 3.4].

**Proof of Theorem 1.6.** Given a word \( w = w(x_1, \ldots, x_d) \) in the free group \( F_d \) on \( x_1, \ldots, x_d \) and a group \( G \), denote

\[
w(G) = \{ w(g_1, \ldots, g_d) : g_1, \ldots, g_d \in G \}.
\]

By a result of Jones [16], every proper variety of groups contains only finitely many finite simple groups. This means that, if \( w \neq 1 \), then \( w(G) \neq \{ 1 \} \) for all large enough finite simple groups \( G \).

**Lemma 8.2.** For every nontrivial word \( w \in F_d \) there exists \( \varepsilon = \varepsilon(w) > 0 \), such that if \( G \) is a finite simple group, then either \( w(G) = \{ 1 \} \) or \( |w(G)| \geq |G|^\varepsilon \).

**Proof.** Let \( G \) be a finite simple group such that \( w(G) \neq \{ 1 \} \). Then \( |w(G)| \geq 2 \) and so to prove \( |w(G)| \geq |G|^\varepsilon \) we may assume that \( G \) is large.

Suppose \( G = A_n \). Let \( k = k(w) \) be minimal with \( w(A_k) \neq \{ 1 \} \). Write \( n = qk + r \) where \( 0 \leq r < k \). Then \( A_n \) has a naturally embedded subgroup \( H \) of type \( (A_k)^r \). Consider the set \( w(H) \). In each factor \( A_k \) there is a nontrivial value of \( w \), whose support is necessarily at least 3. This implies that \( w(G) \) contains a permutation \( x \) of support \( \geq 3q \). Let \( C \subseteq w(G) \) be the class of \( x \). Applying Lemma 2.2 we obtain

\[
|C| \geq n^{(1/3-o(1))3q} = n^{(1-o(1))q} \geq n^{(1-o(1))n(2k)^{-1}}.
\]

It follows that

\[
|w(G)| \geq |C| \geq |A_n|^{(2k)^{-1-0(1)}},
\]

as required.

Suppose now that \( G \) is a group of Lie type over \( \mathbb{F}_q \). We may assume that \( G \) has large Lie rank, otherwise the result is trivial (as \( \text{cn}(G) \) is bounded). So we may take \( G = \text{Cl}_n(q) \) to be a classical group with natural module \( V \) of large dimension \( n \). Here a similar argument works. Let \( k = k(w) \) be minimal such that \( w \) does not vanish on \( \text{Cl}_k(q) \) (a group of the same type as \( G \), with natural module of dimension \( k \)). Take \( n > 2k \), and write \( n = mk + r \), where \( m, r \in \mathbb{Z} \) and \( k \leq r < 2k \). Then \( G \) has a naturally embedded subgroup \( H \) of type \( (\text{Cl}_k(q))^{\alpha} \).
Consider $w(H)$. In each factor $\text{Cl}_k(q)$ there is a nontrivial value of $w$, and multiplying these together, we see that $w(H)$ contains an element which has a power $h$ of prime order, such that $\nu(h) \geq m$ (where $\nu(h)$ is as defined above in the proof of 8.1). Then by [25, 3.4], we have

$$|h^G| > cq^{mn/2} > cq^{n^2/6k} > c|G|^{1/6k},$$

and the result follows.

Theorem 1.6 now follows from results 1.1 and 8.2.

Proof of Theorem 1.12. In view of Theorem 1.1, to prove Theorem 1.12 it suffices to prove the following.

**Lemma 8.3.** There exists $\varepsilon > 0$ such that if $G$ is a finite simple group, and $x \in G$ is randomly chosen, then $|x^G| \geq |G|^\varepsilon$ with probability tending to 1 as $|G| \to \infty$.

**Proof.** Fix $\delta > 0$ and set

$$G_\delta = \{ x \in G : |x^G| = \delta \frac{|G|}{k(G)} \},$$

where $k(G)$ is the number of classes in $G$. Then

$$|G_\delta| \leq k(G) \delta \frac{|G|}{k(G)} = \delta |G|.$$

It follows that if $x \in G$ is randomly chosen, then with probability at least $1 - \delta$, we have $|x^G| \geq \delta |G|/k(G)$.

Now for $G = A_n$ we have $k(G) \leq P(n) = |G|^{o(1)}$ (where $P(n)$ is the partition function), and by [22, Theorem 1], for $G$ of Lie type we have $k(G) \leq |G|^{1/3 + o(1)}$. Hence $|G|/k(G) \geq |G|^{2/3 - o(1)}$, and the conclusion follows by choosing (say) $\delta = |G|^{-1/6}$.

**Proof of Theorem 1.13.** We need some notation. Let $P^t$ denote the induced distribution on $\Gamma(G, C)$ after $t$ steps of the random walk, and let $U$ be the uniform distribution on $G$. For distributions $f, g$ on $G$ we define the total variation distance between $f$ and $g$ by

$$||f - g|| = \frac{1}{2} \sum_{x \in G} |f(x) - g(x)|.$$ 

The mixing time $t = t(G, C)$ is defined to be the minimal positive integer $t$ for which $||P^t - U|| \leq \frac{1}{2e}$. (Then for $n \geq t$, $||P^t - U|| \leq e^{-n/t}$.)

Now, by the upper bound lemma of Diaconis and Shahshahani [7] we have

$$||P^t - U|| \leq \frac{1}{4} \sum_{\chi \neq 1} |\chi(x)/\chi(1)|^2 \chi(1)^2,$$
where the sum is over the nontrivial irreducible characters of $G$. Arguing as in the proofs of Lemmas 4.1 and 4.2 above, we see that the sum above is bounded above by each of the expressions

$$(6q)^n(c/q^{1/2})^{2t}q^{8n^2},$$

and

$$(6q)^n q^{(1-\varepsilon/2)2nt} \left(\frac{1}{2}(q^n - 1)\right)^{2t-2}.$$  

For $t > t(n, \varepsilon)$ the first expression is arbitrarily close to zero; if $n > n(\varepsilon)$ and $t > t(\varepsilon)$, then the second expression is arbitrarily close to zero. It follows that, if $t$ is larger than some function of $\varepsilon$, then $||P^t - U||$ is arbitrarily small. This completes the proof.

Imperial College of Science, Technology, and Medicine, London, England, UK  
E-mail address: miliebeck@ic.ac.uk  

The Hebrew University, Jerusalem, Israel  
E-mail address: shalev@math.huji.ac.il  

References  

(Received October 28, 1999)