

CHARACTERS OF SYMMETRIC GROUPS: SHARP BOUNDS AND APPLICATIONS

MICHAEL LARSEN AND ANER SHALEV

ABSTRACT. We provide new estimates on character values of symmetric groups which hold for all characters and which are in some sense best possible. It follows from our general bound that if a permutation $\sigma \in S_n$ has at most $n^{o(1)}$ cycles of length $< m$, then $|\chi(\sigma)| \leq \chi(1)^{1/m+o(1)}$ for all irreducible characters χ of S_n . This is a far reaching generalization of a result of Fomin and Lulov.

We then use our various character bounds to solve a wide range of open problems regarding mixing times of random walks, covering by powers of conjugacy classes, as well as probabilistic and combinatorial properties of word maps.

In particular we prove a conjecture of Rudvalis and of Vishne on covering numbers, and a conjecture of Lulov and Pak on mixing times of certain random walks on S_n .

Our character-theoretic methods also yield best possible solutions to Waring type problems for alternating groups A_n , showing that if w is a non-trivial word, and $n \gg 0$, then every element of A_n is a product of two values of w .

Contents:

1. Introduction
2. Virtual degrees
3. Character values
4. Random walks
5. Covering numbers
6. Normal subsets
7. Word maps

Michael Larsen was partially supported by NSF grant DMS-0354772. Aner Shalev was partially supported by an Israel Science Foundation Grant. Both authors were partially supported by a Bi-National Science Foundation United States-Israel Grant 2004052.

1. INTRODUCTION

The basic representation theory of symmetric groups was developed a century ago, yet many questions on character values, essential for various applications, remained largely open. Such estimates depend on a choice of Young diagram λ (corresponding to the character) and a choice of conjugacy class σ^{S_n} (at which the character is evaluated). There has been considerable progress in the combinatorics of Young diagrams, their asymptotic shape, as well as in bounding character values in the regime where λ is, in a suitable sense, well-behaved (see, e.g., [Ro], [Bi], [RS] and the references therein). However, for many applications it is essential to obtain good bounds which hold for *all* characters, which is our goal here.

There has been particular interest in showing that character ratios $|\chi(\sigma)|/|\chi(1)|$ in S_n are very small (under suitable assumptions on the character χ and the permutation σ). See for instance Roichman's celebrated paper [Ro], showing that certain character ratios are exponentially small. It turns out that such ratios are usually super-exponentially small, namely bounded above by $(n!)^{-\epsilon}$ for some $\epsilon > 0$.

We look for upper bounds on the character value $|\chi(\sigma)|$ in terms of the character degree $\chi(1)$. Particularly useful are bounds of the form

$$|\chi(\sigma)| \leq \chi(1)^\alpha \text{ for all } \chi \in \text{Irr } S_n,$$

where α depends on σ and is as small as possible. A breakthrough in this direction was obtained by Fomin and Lulov [FL] in 1996 for very special permutations σ , called *homogeneous*. Let $(a_1^{f_1} a_2^{f_2} \cdots a_k^{f_k})$ denote the conjugacy class of S_n consisting of the permutations which decompose into f_i a_i -cycles ($i = 1, \dots, k$). It follows from [FL] that if $\sigma \in (m^{n/m})$, then

$$|\chi(\sigma)| \leq cn^{1/2(1-1/m)} \chi(1)^{1/m} \text{ for all } \chi \in \text{Irr } S_n,$$

where c is an absolute constant. Now, most characters of S_n have super-polynomial degree, in the sense that $\log \chi(1)/\log n \rightarrow \infty$ as $n \rightarrow \infty$. For such characters χ it follows that

$$(1) \quad |\chi(\sigma)| \leq \chi(1)^{1/m+o(1)}.$$

Here and throughout this paper $o(1)$ denotes a real number depending on n which tends to 0 as $n \rightarrow \infty$.

Our first results provide bounds of this type for arbitrary permutations σ and arbitrary characters χ . We need some notation.

Let σ be a permutation in S_n . Let $\text{fix}(\sigma)$ denote the number of fixed points of σ , and let $\text{cyc}(\sigma)$ be the number of cycles (including 1-cycles) of σ . Given $k \geq 1$ let $f_\sigma(k)$ denote the number of cycles of length k in

the cycle decomposition of σ , and $F_\sigma(k) := \sum_{i=1}^k f_\sigma(i)$ the number of cycles of length at most k in that decomposition.

Next, let Σ_k denote the union of all σ -orbits of length at most k , so that

$$|\Sigma_k| = \sum_{i=1}^k i f_\sigma(i).$$

We define the *orbit growth sequence* of σ to be the sequence e_1, e_2, \dots such that for every positive integer k ,

$$n^{e_1 + \dots + e_k} = \max(|\Sigma_k|, 1).$$

Thus any orbit growth sequence can be regarded as a probability measure on the positive integers. Let

$$E(\sigma) := \sum_{k \geq 1} \frac{e_k}{k} = \frac{1}{\log n} \sum_{k \geq 1} \frac{\log^+ |\Sigma_k| - \log^+ |\Sigma_{k-1}|}{k}$$

denote the expected value of $\frac{1}{n}$ for this measure. (Here $\log^+ x = \max(\log x, 0)$.)

We also define the *cycle growth sequence* of σ to be the sequence b_1, b_2, \dots such that for every positive integer k ,

$$n^{b_k} = \max(F_\sigma(k), 1).$$

Define

$$B(\sigma) := \sum_{k \geq 1} \frac{b_k}{k(k+1)}.$$

It is easy to see that $B(\sigma) \leq E(\sigma) \leq 1$. Moreover, we have $E(\sigma) = 1$ if and only if $B(\sigma) = 1$ if and only if $\sigma = 1$.

We can now state our main character-theoretic result.

Theorem 1.1. *For all $\epsilon > 0$ there exists N such that for all integers $n \geq N$, all permutations $\sigma \in S_n$, and all irreducible characters χ of S_n we have*

$$\begin{aligned} (i) \quad & |\chi(\sigma)| \leq \chi(1)^{E(\sigma) + \epsilon}; \\ (ii) \quad & |\chi(\sigma)| \leq \chi(1)^{B(\sigma) + \epsilon}. \end{aligned}$$

The proof of part (i) is rather complicated, and is based on a new concept of the *virtual degree* of a character developed here. We then deduce part (ii) by showing that $E(\sigma) \leq B(\sigma) + o(1)$.

Roughly speaking, Theorem 1.1 shows that the only obstruction to small character values is the existence of many short cycles in the underlying permutation σ .

Note that the orbit growth sequence (e_k) of $\sigma \in (m^{n/m})$ satisfies $e_m = 1$ and $e_k = 0$ for $k \neq m$. Thus $E(\sigma) = 1/m$ and Theorem 1.1 implies the bound (1) above for *all* characters χ (not just those of super-polynomial degree).

Theorem 1.1 has a whole range of applications to various fields. We start by stating the character-theoretic ones.

Theorem 1.2. *Let $\sigma \in S_n$ and $\chi \in \text{Irr } S_n$.*

(i) *If σ is fixed-point-free, or has $n^{o(1)}$ fixed points, then*

$$|\chi(\sigma)| \leq \chi(1)^{1/2+o(1)}.$$

(ii) *If m is a positive integer and σ has at most $n^{o(1)}$ cycles of length less than m , then*

$$|\chi(\sigma)| \leq \chi(1)^{1/m+o(1)}.$$

A result very similar to part (i) above was conjectured by Lulov and Pak. Indeed, Conjecture 4.3 in [LP] states that, if σ is fixed-point-free, then $|\chi(\sigma)| \leq cn^{1/4}\chi(1)^{1/2}$ for all characters χ . Part (i) above shows that this is true with a small error term (tending to 0).

Note that if $\sigma \in (2^{n/2})$ and χ is a character whose degree is quadratic in n , then $|\chi(\sigma)|$ is roughly $\chi(1)^{1/2}$. Thus part (i) above is best possible.

Part (ii) of Theorem 1.2 also generalizes the Fomin-Lulov inequality (1), and can be shown to be best possible too.

In the next result we bound all character values of σ in terms of the number of fixed points of σ . A recent result of Müller and Puchta [MS] states that, if $f = \text{fix}(\sigma)$, then $|\chi(\sigma)| \leq \chi(1)^{1-\delta}$, where

$$\delta = ((1 - 1/\log n)^{-1} \frac{12 \log n}{\log(n/f)} + 18)^{-1}.$$

Note that we always have $\delta < 1/30$, even when σ is fixed-point-free. Therefore the bound above is not strong enough to imply Theorem 1.2. Here we prove a stronger bound as follows.

Theorem 1.3. *Let $\sigma \in S_n$ and let $f = \max(\text{fix}(\sigma), 1)$. Then for all $\chi \in \text{Irr}(S_n)$ we have*

$$|\chi(\sigma)| \leq \chi(1)^{1 - \frac{\log(n/f)}{2 \log n} + o(1)}.$$

For example, if $f = n^{1-\alpha}$, then Theorem 1.3 shows that $|\chi(\sigma)| \leq \chi(1)^{1-\alpha/2+o(1)}$. This result has an intriguing consequence, showing that one particular character value has strong implications for all character values. Indeed, let $\chi_0 \in \text{Irr}(S_n)$ be the irreducible character of degree $n-1$. Then $\chi_0(\sigma) = \text{fix}(\sigma) - 1$, and so

$$|\chi_0(\sigma)| \leq \chi_0(1)^{1-\alpha} \text{ implies } |\chi(\sigma)| \leq \chi(1)^{1-\alpha/2+o(1)} \text{ for all } \chi \in \text{Irr}(S_n).$$

Our final character estimate uses the number of cycles as the main parameter.

Theorem 1.4. *Fix $\alpha \leq 1$ and let $\sigma \in S_n$ be a permutation with at most n^α cycles. Then for all $\chi \in \text{Irr}(S_n)$ we have*

$$|\chi(\sigma)| \leq \chi(1)^{\alpha+o(1)}.$$

This improves a technical result from [LaSh], where it is shown that $\text{cyc}(\sigma) \leq n^{1/128}$ implies $|\chi(\sigma)| \leq \chi(1)^{1/63}$.

Theorem 1.4 is essentially best possible. For example let $k = n^\alpha$ and $\sigma \in (1^{k-1}(n-k+1))$, and let χ_0 be the degree $n-1$ character of S_n as above. Then $\text{cyc}(\sigma) = n^\alpha$ and $\chi_0(\sigma) = n^\alpha - 2$.

Aside from its intrinsic interest, the study of character values of symmetric groups has important applications to various fields; these include random walks, covering numbers, subgroup growth, Fuchsian groups, Riemann surfaces and other fields. See for instance [LiSh2] for more details.

Random walks on symmetric groups with respect to a conjugacy class C as a generating set have been studied extensively in the past decades. See Diaconis and Shahshahani [DS] for transpositions, Lulov [Lu] and Vishne [V] for homogeneous classes, Lulov and Pak [LP] and Müller and Schlage-Puchta [MS]. A main problem investigated is determining the mixing time $T(C)$ of the random walk, namely the time t required until we reach an almost uniform distribution on A_n or $S_n \setminus A_n$ (note that the sign after t steps of the walk is fixed). See also Diaconis [D1], [D2] for more background, and Section 4 below for the exact definition of the mixing time $T(C)$. Here we obtain very general results, and in particular determine the mixing time up to a surprisingly small multiplicative constant.

We start with our main result on mixing time, which follows from Theorem 1.1 above.

Theorem 1.5. *For every positive integer t and $\epsilon > 0$ there exists N such that if $n \geq N$ and $\sigma \in S_n$ satisfies*

$$E(\sigma) \leq 1 - 1/t - \epsilon$$

then

$$T(\sigma^{S_n}) \leq t.$$

The same result holds with $B(\sigma)$ replacing $E(\sigma)$.

We now present various consequences of this theorem. The first one solves the natural asymptotic question, for which classes C the mixing time is bounded. It turns out that the answer depends only on the

number of fixed points of the permutations in C , which we denote by $\text{fix}(C)$.

Theorem 1.6. *A series of conjugacy classes $C_n \subset S_n$ has bounded mixing time if and only if there exists $\epsilon > 0$ such that*

$$\text{fix}(C_n) \leq n^{1-\epsilon}$$

for all large enough n .

We note that the necessity of this condition was already observed in [V], and the sufficiency is the main part.

The next theorem determines the mixing time essentially up to a factor of 2.

Theorem 1.7. *For any $\epsilon > 0$ there exists N such that if $n \geq N$ and $\sigma \in S_n$ satisfies $\text{fix}(\sigma) \leq n^{1-\epsilon}$, then*

$$\frac{\log n}{\log(n/f)} \leq T(\sigma^{S_n}) \leq 2 \frac{\log n}{\log(n/f)} + 1$$

where $f = \max(\text{fix}(\sigma), 1)$.

The case where $\text{fix}(\sigma) = 0$ received particular attention. In the homogeneous case, where $C = (m^{n/m})$, a breakthrough was made in Lulov's thesis [Lu]. It is shown there that $T(C) = 3$ if $m = 2$ and $T(C) = 2$ if $m \geq 3$.

It remained an open problem whether similar results hold for all fixed-point-free classes. See Lubotzky's survey [Lub] (where this problem is mentioned, and stated – too optimistically? – as solved). In [LP, 4.1] the following is stated as a main conjecture.

Lulov-Pak Conjecture: Let C_n be a sequence of conjugacy classes in S_n with no fixed points. Then (for $n \gg 0$) the mixing time $T(C_n)$ is either 2 or 3.

Here we prove this conjecture, even allowing some fixed points.

Theorem 1.8. *Let $C = \sigma^{S_n}$.*

- (i) *If σ is fixed-point-free, or has at most $n^{o(1)}$ fixed points, then $T(C) \leq 3$ for all large n .*
- (ii) *If σ has at most $n^{o(1)}$ cycles of length 1 and 2 then $T(C) = 2$ for all large n .*

We now turn to applications of our character theoretic results to covering problems.

Given a conjugacy class $C \neq \{1\}$ in a finite simple group G , there exist integers k such that $C^k = G$. Define the *covering number* $cn(C, G)$

of C in G to be the minimal such k . Substantial work has been devoted to the study of these covering numbers (see e.g. [AH, EGH, LL, LiSh1]). In [LiSh1] they are determined for all finite simple groups G and classes C up to an (unspecified) multiplicative constant.

A particular challenge is to show that $C^2 = G$ in certain cases. Indeed a conjecture of Thompson states that every finite simple group G has a class C with this property. In spite of considerable progress this is still open in general.

Now let $C = \sigma^{S_n}$ be a conjugacy class in S_n . When can we say that $C^2 = A_n$? This problem has quite a long history. Gleason [Hu] seems to have been the first to observe that $C^2 = A_n$ if σ is an n -cycle. See also Bertram [Be]. Later this was generalized by Brenner [Br] to the case where σ consists of two cycles (or more generally two non-trivial cycles with few additional fixed points). The case of permutations with more general cycle structure remained wide open for a long time.

In the recent work [LaSh] we show that if $\sigma \in S_n$ has at most $n^{1/128}$ cycles, and n is sufficiently large, then $(\sigma^{S_n})^2 = A_n$. Here we improve it as follows.

Theorem 1.9. *For every $\epsilon > 0$ there exists N such that if $n \geq N$ and $\sigma \in S_n$ is a permutation with at most $n^{1/4-\epsilon}$ cycles, then $(\sigma^{S_n})^2 = A_n$.*

Moreover, if σ has no fixed points, then $\text{cyc}(\sigma) \leq n^{1-\epsilon}$ already implies $(\sigma^{S_n})^2 = A_n$

It turns out that a much weaker assumption can be made on the number of cycles of σ , provided σ doesn't have many cycles of length 1 and 2. Indeed we have

Theorem 1.10. *For all $\epsilon > 0$ there exists N such that for all $n \geq N$ and all $\sigma \in S_n$, if*

$$(2) \quad \text{cyc}(\sigma) < \left(\frac{1}{4} - \epsilon\right) n$$

and any of the following conditions is satisfied, then $(\sigma^{S_n})^2 = A_n$:

$$(3) \quad E(\sigma) < \frac{1}{4} - \epsilon,$$

$$(4) \quad \text{fix}(\sigma^2) < n^{1/4-\epsilon},$$

$$(5) \quad \text{fix}(\sigma) = 0 \text{ and } \text{fix}(\sigma^2) < n^{1-\epsilon}.$$

The above theorem has various consequences. One, which turns out to be quite useful in the study of word maps, is the following.

Corollary 1.11. *For all $\epsilon > 0$ there exists N such that for all $n \geq N$ and all $\sigma \in S_n$,*

$$E(\sigma) \leq 1/4 - \epsilon \text{ implies } (\sigma^{S_n})^2 = A_n.$$

The precise covering number of classes in S_n was largely unknown, even in the homogeneous case. The following conjecture is posed in [BR] (attributed to Rudvalis) and more recently in [V].

Rudvalis-Vishne Conjecture: If $m \geq 4$, then $(m^{n/m})^2 = A_n$ for all n divisible by m .

The case $m = 4$ is settled in [BR]. Using Theorem 1.10 we settle the general conjecture for all large n .

Theorem 1.12. *There exists N such that, if $m \geq 4$ and $n \geq N$ is a multiple of m , then $(m^{n/m})^2 = A_n$.*

For $m \geq 5$ we also prove an extended version of the Rudvalis-Vishne conjecture, where we allow some fixed points. For instance we show that $(1^a m^b)^2 = A_n$ if a is bounded and $n \gg 0$ (in fact $a \leq n^{1/4-\epsilon}$ suffices). This result has nice applications to Waring-type problems for word maps. See the end of Section 7 for details.

The results above have a somewhat unexpected consequence. While up to now results of the form $C^2 = A_n$ were proved for very few and special classes C , it follows from our results that for almost all $\sigma \in S_n$ we have $(\sigma^{S_n})^2 = A_n$. Indeed, by the Erdős-Turán theory [ET] a random permutation in S_n has about $\log n$ cycles, and so by Theorem 1.9 the square of its class is A_n .

Moreover, using Corollary 1.11 and other tools, we can show that the probability that $(\sigma^{S_n})^2 = A_n$ tends to 1 rather fast.

Theorem 1.13. *For every $\epsilon > 0$ there exists N such that, if $n \geq N$ and $\sigma \in S_n$ is randomly chosen, then $(\sigma^{S_n})^2 = A_n$ with probability $\geq 1 - \exp(-n^{1/4-\epsilon})$.*

In general we define the covering number $cn(C, S_n)$ of a non-trivial class C in S_n to be the minimal integer $k \geq 2$ such that C^k is a coset of A_n in S_n . Our methods have applications also to classes with higher covering numbers. We show the following.

Theorem 1.14. *There exists N such that for all $n \geq N$ and all $\sigma \in S_n$ with $\text{fix}(\sigma) \leq n/5$ we have*

$$(\sigma^{S_n})^4 = A_n.$$

The exponent 4 here is best possible. Indeed, there are even fixed-point-free classes C whose covering number is 4, for instance $C = (2^{n/2})$ (see [V]).

One of the motivations behind the above result is a somewhat surprising consequence regarding connections between covering and mixing. In general it is known that bounded covering number does not imply bounded mixing time. Also if C is a class of mixing time t then the covering number of C is not necessarily bounded by t (though it can be bounded by $2t$). However, we show below that bounded mixing time implies covering number at most 4.

Corollary 1.15. *Let $C_n \subset S_n$ ($n > 1$) be a series of conjugacy classes such that the mixing times $T(C_n)$ are bounded. Then we have $C_n^4 = A_n$ for all sufficiently large n .*

This result is again best possible. Indeed, the class $C = (2^{n/2})$ has mixing time 3 and covering number 4.

Our final result on covering numbers deals with arbitrary classes. Recall that the *support* of a permutation $\sigma \in S_n$ is $n - \text{fix}(\sigma)$. It is easy to see that for any class $C \subset S_n$ of permutations of support s we have

$$cn(C, S_n) \geq \lceil n/s \rceil.$$

Indeed, if $k < n/s$ then all the permutations in C^k have support $\leq ks < n$.

Using Theorem 1.14 we show the following.

Theorem 1.16. *Let $1 \neq \sigma \in A_n$ be a permutation of support s and let $C = \sigma^{S_n}$. Then*

$$cn(C, A_n) \leq 4\lceil n/s \rceil.$$

The constant 4 here is best possible, as shown by the case $C = (2^{n/2})$. In fact our proof shows that if $s \leq n/5$ or $s \geq 4n/5$ then $cn(C, A_n) \leq 4\lceil n/s \rceil$.

The final applications of our study of character values are related to *word maps*. By a word we mean an element $w = w(x_1, \dots, x_d)$ of the free group on x_1, \dots, x_d . Given a group G , the word w gives rise to a map $f_w : G^d \rightarrow G$ induced by substitution. We let $w(G) \subseteq G$ denote the image of f_w , namely the set of values of w in G .

In recent years there has been growing interest in word maps and their images, especially in connection to algebraic groups, free groups, profinite groups, and Waring type problems in simple groups. See Borel [Bo] for algebraic groups, Nikolov and Segal [NS1], [NS2] for profinite groups, as well as [LiSh1], [DPSSh], [L], [Sh] and [LaSh].

Recall that Waring problem, solved by Hilbert, states that every positive integer is a sum of $g(k)$ k th powers, for some suitable function g . See for instance [Na] for details and extensive background.

Group theoretic versions of this problem have been studied lately, where one attempts to express group elements as short product of values of a word w . It is shown in [Sh] that for every word $w \neq 1$ there is a number $N = N_w$ such that if G is a finite (non-abelian) simple group of order at least N , then

$$w(G)^3 = G.$$

The character estimates and methods developed in this paper enable us to obtain more refined information on the behavior of word maps on symmetric and alternating groups.

Theorem 1.17. *Let w be a non-identity word. Choose $\sigma_1, \sigma_2 \in w(A_n)$ at random (with uniform distribution on $w(A_n)$). Then, as $n \rightarrow \infty$, the product $\sigma_1\sigma_2$ is almost uniformly distributed on A_n*

This means that, if, for $\pi \in A_n$, $P(\pi)$ is the probability that $\pi = \sigma_1\sigma_2$, and U is the uniform distribution on A_n , then $\|P - U\| \rightarrow 0$ as $n \rightarrow \infty$, where the norm is the L_1 -norm. Thus a random walk on A_n with the generating set $w(A_n)$ has mixing time 2.

In fact the same method establishes a similar result for $\sigma_i \in w_i(A_n)$ ($i = 1, 2$), where w_1, w_2 are two non-identity words.

There is another sense in which a product of two non-identity words behaves randomly.

Theorem 1.18. *Let w be a non-identity word in d variables. Consider the map $f : A_n^{2d} \rightarrow A_n$ defined by*

$$f(\sigma_1, \dots, \sigma_d, \tau_1, \dots, \tau_d) = w(\sigma_1, \dots, \sigma_d)w(\tau_1, \dots, \tau_d).$$

(i) *If $\sigma_1, \dots, \sigma_d, \tau_1, \dots, \tau_d \in A_n$ are randomly chosen, then $f(\sigma_1, \dots, \sigma_d, \tau_1, \dots, \tau_d)$ is almost uniformly distributed on A_n .*

(ii) *The map f is almost measure preserving, namely, if $Y \subseteq A_n$ then*

$$|f^{-1}(Y)|/|A_n^{2d}| = |Y|/|A_n| + o(1).$$

(iii) *If $X \subseteq A_n^{2d}$ then*

$$|f(X)|/|A_n| \geq |X|/|A_n^{2d}| - o(1).$$

The results on measure preservation can be compared with recent results from [GSh] on measure-preservation of commutator maps, and of the word map associated to $x_1^2x_2^2$.

Our proof of Theorem 1.18 combines the character bounds proved here with an interesting result of Nica [Ni] on the typical cycle structure

of $w(\sigma_1, \dots, \sigma_d)$. Again, a similar result holds for the map sending $\sigma_1, \dots, \sigma_d, \tau_1, \dots, \tau_e$ to $w_1(\sigma_1, \dots, \sigma_d)w_2(\tau_1, \dots, \tau_e)$, where $w_1 \in F_d$ and $w_2 \in F_e$ are any two non-identity words.

Note that Theorems 1.17 and 1.18 imply $|w(A_n)^2|/|A_n| \rightarrow 1$ but they do not imply $w(A_n)^2 = A_n$ for large n . However, by combining results from [Ni] with our present results on covering numbers we show the following.

Theorem 1.19. *Let w be a non-identity word in d variables, and let $\sigma_1, \dots, \sigma_d \in A_n$ be d randomly chosen elements. Then*

$$(w(\sigma_1, \dots, \sigma_d)^{S_n})^2 = A_n$$

with probability tending to 1 as $n \rightarrow \infty$.

In particular $w(A_n)^2 = A_n$ for all large n .

An alternative proof of the second assertion above will appear in [LaSh]; it uses highly non-elementary methods, involving algebraic geometry, analytic number theory and groups of Lie type. Our proof here uses other tools (such as free probability through [Ni] and more delicate character bounds) and gives a bit more, namely that the class of a random w -value already has covering number 2.

Note that $w(A_n)$ (being characteristic in A_n) is a normal subset of S_n , and by results from [LaSh] it is very large, namely

$$|w(A_n)| \geq |A_n|n^{-29/9-o(1)}.$$

Can we deduce that $w(A_n)^2 = A_n$ just from these facts?

It turns out that the answer is positive. In fact our final result shows that $W^2 = A_n$ even for much smaller normal subsets W .

Theorem 1.20. *Let $W \subseteq A_n$ be a normal subset of S_n , and suppose*

$$|W| \geq |A_n| \exp(-n^{1/4-\epsilon})$$

for some fixed $\epsilon > 0$. Then there exists N depending only on ϵ such that if $n \geq N$ then

$$W^2 = A_n.$$

In particular, if $w \neq 1$ is a word, then $w(A_n)^2 = A_n$ for all large n .

In [NP] Nikolov and Pyber use a method of Gowers [G] to prove results of the form $W^3 = G$ where W is a large subset of a finite group G . More specifically, $|W|$ should be at least $|G|/k^{1/3}$ where k is the minimal degree of a character $1 \neq \chi \in \text{Irr}(G)$.

While this method provides strong results for simple groups of Lie type, the result for A_n is rather weak, since $k = n - 1$ in this case. Obviously, taking $W = A_{n-1} \subset A_n$ we see that $|W| \geq |A_n|n^{-1}$ does not imply $W^2 = A_n$ for arbitrary subsets.

This shows that the condition in Theorem 1.20 that W is a normal subset is essential. With this assumption this theorem proves covering in two (and not three) steps, which Gowers' method cannot provide.

Our notation is rather standard. If P is a real function on the finite group G , then $\|P\| = \sum_{g \in G} |P(g)|$ denotes the L_1 -norm of P . We shall use this norm to measure the distance between various probability distributions and the uniform distribution U on G . The so called *Witten zeta function* ζ^G of G is defined by

$$\zeta^G(s) = \sum_{\chi \in \text{Irr}(G)} \chi(1)^{-s},$$

where $\text{Irr}(G)$ is the set of complex irreducible characters of G . The zeta function of S_n plays a major role in this paper (see [LiSh2] for some of its properties). The idea is to use Theorem 1.1 and its consequences to reduce important formulae involving character degrees and values to certain values of ζ^{S_n} involving character degrees alone, and then apply known properties of ζ^{S_n} .

Finally, some words on the structure of the paper.

In Section 2 we define the notion of the *virtual degree* $D(\lambda)$ of a Young diagram λ . We show that it is a good approximation for the degree $\chi_\lambda(1)$ of the character associated with λ . Using this notion we then study character values in Section 3, where Theorems 1.1–1.4 are proved.

Section 4 is devoted to random walks on S_n and to the proof of Theorems 1.5–1.9. We combine our estimates on character values with the Diaconis-Shahshahani upper bound lemma [DS] to bound the distance from the uniform distribution by a certain value of ζ^{S_n} , thereby bounding the mixing time. In Section 5 we study covering numbers of conjugacy classes C in S_n and prove results 1.10–1.12 and 1.14–1.16. Here too character bounds play a key role, but they only show that certain permutations σ (with $E(\sigma)$ not too large) lie in the suitable power C^k ; to show that the remaining permutations also lie there we develop a *cancellation method*. Consequently, the arguments in Section 5 are partly probabilistic and character-theoretic, and partly combinatorial, based on explicit constructions.

Section 6 is devoted to more general mixing and covering problems, where conjugacy classes are replaced with normal subsets, and are allowed to change over time. This is where results 1.13 and 1.20 are proved. The generalizations to normal subsets are essential in the study of various properties of word maps in Section 7, where we also apply [Ni] and prove theorems 1.17–1.19. We end this paper with yet another

proof of the $w(A_n)^2 = A_n$ theorem, which is based on the extended Rudvalis-Vishne conjecture established here.

2. VIRTUAL DEGREES

In this section and the next we give upper bound estimates for the characters of irreducible representations of symmetric groups. We assume the reader is familiar with the basic representation theory of symmetric groups, see for instance James [Ja] for background.

Let n be a positive integer and λ a Young diagram with $n = |\lambda|$ boxes, with rows of length λ_i . Let

$$(a_1, \dots, a_m | b_1, \dots, b_m)$$

be the Frobenius notation for λ , i.e., $a_i := \lambda_i - i$ and $b_i := \lambda_i^t - i$. If λ' is another Young diagram, we write $\lambda' \leq \lambda$ if $\lambda'_i \leq \lambda_i$ for all i . Let χ_λ denote the character of the irreducible representation of S_n associated to λ . We write $d(\lambda) := \chi_\lambda(1)$ and define

$$D(\lambda) := \frac{(n-1)!}{\prod_{i=1}^m a_i! b_i!}.$$

We call $D(\lambda)$ the *virtual degree* of χ_λ . The idea, developed in this section, is that the virtual degree can be viewed as an approximate substitute for the actual degree, one which is easier to work with when using the Murnaghan–Nakayama rule.

Recall that by the Hook Formula [Ja, 20.1],

$$d(\lambda) = \chi_\lambda(1) = \frac{n!}{\prod h_{ij}},$$

where $h_{ij} = \lambda_i + \lambda_j^t + 1 - i - j$ are the hook lengths of λ .

Throughout this section we set $a := a_1$, $b := b_1$, $c := n - (a + b + 1)$, $s := a + b$. We write $h(\lambda)$ for the product of hook lengths associated with the a boxes in the first row, excluding the upper left corner. Likewise $v(\lambda)$ is the product of hook lengths for the b boxes in the first column excluding the upper left corner.

Lemma 2.1. *For all Young diagrams λ we have*

$$\begin{aligned} a! &\leq h(\lambda) \leq (a+1)^{c/a} a! \\ b! &\leq v(\lambda) \leq (b+1)^{c/b} b! \end{aligned}$$

Proof. We recall the inequality of Hardy–Littlewood–Pólya [HLP, Section 3.17]. If $x_1 \geq x_2 \geq \cdots \geq x_a$, $y_1 \geq y_2 \geq \cdots \geq y_a$, and for $1 \leq i \leq a$,

$$x_1 + x_2 + \cdots + x_i \geq y_1 + y_2 + \cdots + y_i,$$

with equality when $i = a$, then we say that the sequence x_i *majorizes* the sequence y_i . If x_i majorizes y_i and $f(x)$ is convex, then

$$(6) \quad f(x_1) + f(x_2) + \cdots + f(x_a) \geq f(y_1) + f(y_2) + \cdots + f(y_a).$$

If $x_1, \dots, x_a \geq 0$ is a decreasing sequence of integers with sum c , and

$$y_1 - 1 = \cdots = y_r - 1 = y_{r+1} = \cdots = y_a = \lfloor c/a \rfloor,$$

where $r = c - \lfloor c/a \rfloor x_a$, then the sequence x_i majorizes the sequence y_i . (Note that y_i is the unique non-increasing a -term sequence of integers with sum c .) Applying (6) with $f(x) = -\log x$, we deduce that

$$(x_1 + a) \cdots (x_{a-1} + 2)(x_a + 1)$$

is maximized when $x_i = y_i$ for all i . Therefore,

$$a! \leq h(\lambda) \leq a! \prod_{i=1}^{x_a} \frac{a+i}{i} \prod_{j=1}^r \frac{a+x_a+2-j}{a+x_a+1-j}.$$

As the fraction $\frac{a+x_a+2-j}{a+x_a+1-j}$ is increasing in j and

$$\prod_{j=1}^a \frac{a+x_a+2-j}{a+x_a+1-j} = \frac{a+x_a+1}{x_a+1},$$

we have

$$h(\lambda) \leq a!(a+1)^{x_a} \left(\frac{a+x_a+1}{x_a+1} \right)^{r/a} \leq a!(a+1)^{c/a}.$$

The argument for columns is the same. \square

We now turn to the main result of this section.

Theorem 2.2. *We have*

$$\lim_{|\lambda| \rightarrow \infty} \frac{\log D(\lambda)}{\log d(\lambda)} = 1.$$

Proof. For any Young diagram λ , let λ^i denote the diagram obtained from it by omitting the first i rows and the first i columns. Let

$$A(\lambda) := \frac{D(\lambda)}{D(\lambda^1)} = \frac{(n-1)!}{a!b!(n-s-2)!} \quad (\text{where } (-1)! = 1 \text{ by our convention})$$

and

$$a(\lambda) := \frac{d(\lambda)}{d(\lambda^1)} = \frac{n!}{(s+1)h(\lambda)v(\lambda)(n-s-1)!},$$

where the last inequality follows from the Hook Formula.

We claim that for all $\epsilon > 0$ there exists N_ϵ such that

$$C_{|\lambda|}^{-1}a(\lambda)^{1-\epsilon} \leq A(\lambda) \leq C_{|\lambda|}a(\lambda)^{1+\epsilon},$$

where $C_i := N_\epsilon$ if $i < N_\epsilon$ and $C_i := 1$ if $i \geq N_\epsilon$. This claim implies the theorem since

$$D(\lambda) = \prod_{i=0}^{\infty} \frac{D(\lambda^i)}{D(\lambda^{i+1})} = \prod_{i=0}^{\infty} A(\lambda^i),$$

$$d(\lambda) = \prod_{i=0}^{\infty} \frac{d(\lambda^i)}{d(\lambda^{i+1})} = \prod_{i=0}^{\infty} a(\lambda^i),$$

and

$$\prod_{i=0}^{\infty} C_{|\lambda^i|} < N_\epsilon^{N_\epsilon}.$$

To prove the claim it is enough to show that

$$\lim_{|\lambda| \rightarrow \infty} \frac{\log A(\lambda)}{\log a(\lambda)} = 1.$$

When $a = 0$ or $b = 0$ we have $a(\lambda) = A(\lambda)$, so we assume that a and b are positive and therefore that $A(\lambda) \rightarrow \infty$. We fix an integer $k > 0$ and assume that $s > k^2$. We now consider the following cases:

- I. $c \leq k^2$;
- II. $c > k^2$ and $\min(a, b) \leq k$;
- III. $a, b > k, k^2 < c \leq \sqrt{ks}$;
- IV. $a, b > k, c > \sqrt{ks}$.

In every case we have

$$\frac{A(\lambda)}{a(\lambda)} = \frac{h(\lambda)v(\lambda)(s+1)\max(c, 1)}{a!b!n} \geq \frac{h(\lambda)v(\lambda)}{2a!b!} \geq \frac{1}{2}.$$

I. By Lemma 2.1,

$$\frac{A(\lambda)}{a(\lambda)} < (a+1)^{k^2/a}(b+1)^{k^2/b}k^2 < 4^{k^2}k^2.$$

As $a(\lambda) \ll A(\lambda) \ll a(\lambda)$ and $A(\lambda) \rightarrow \infty$,

$$(7) \quad \lim_{|\lambda| \rightarrow \infty} \frac{\log A(\lambda)}{\log a(\lambda)} = 1.$$

II. The hook length of the box $i \geq 0$ units from the right end of the first row is $\leq b+1+i$ and that of the box $j \geq 0$ units from the bottom in the first column is $\leq a+1+j$. Therefore,

$$\frac{A(\lambda)}{a(\lambda)} < \binom{a+b}{a} \binom{a+b}{b} (s+1) < s^{2k}(s+1)$$

and

$$A(\lambda) = \frac{(n-1)!}{a!b!(c-1)!} > \frac{(n-1)!}{s!(c-1)!} > \frac{(s+k^2+1)!}{s!(k^2)!} > \frac{s^{k^2+1}}{(k^2)!}.$$

Thus

$$(8) \quad 1 + o(1) < \frac{\log A(\lambda)}{\log a(\lambda)} < \frac{k^2 + 1}{k^2 - 2k} + o(1).$$

III. By Lemma 2.1,

$$\frac{A(\lambda)}{a(\lambda)} < [(a+1)^{1/a}(b+1)^{1/b}]^c c < ((k+1)^{2/k})^c c < (1+k^{-1/2})^{ck^{-1/4}} c$$

if k sufficiently large. Moreover,

$$A(\lambda) > \frac{(n-1)!}{s!(c-1)!} = \frac{s+1}{1} \frac{s+2}{2} \dots \frac{s+c-1}{c-1} \frac{s+c}{c} > (1+k^{-1/2})^c.$$

Thus,

$$(9) \quad 1 + o(1) < \frac{\log A(\lambda)}{\log a(\lambda)} < \frac{1}{1 - k^{-1/4}} + o(1).$$

IV. As in paragraph II,

$$\frac{A(\lambda)}{a(\lambda)} < \binom{a+b}{a} \binom{a+b}{b} (s+1) < 4^{a+b} (s+1) = 4^s (s+1)$$

and

$$\begin{aligned} A(\lambda) &> \frac{(n-2)!}{s!(c-1)!} \geq \binom{s+s\sqrt{k}-1}{s} = \left(\frac{s\sqrt{k}}{1} \dots \frac{s\sqrt{k}+s-1}{s} \right) \\ &> \sqrt{k}^s \end{aligned}$$

so

$$(10) \quad 1 + o(1) < \frac{\log A(\lambda)}{\log a(\lambda)} < \frac{\log k}{\log k - \log 16} + o(1).$$

Now, sending $k \rightarrow \infty$, equations (7)–(10) imply the theorem. \square

We also need the following result, relating the virtual degrees of a Young diagram and its sub-diagram.

Lemma 2.3. *If $\lambda' \leq \lambda$, then*

$$\frac{\log D(\lambda')}{\log D(\lambda)} \leq \frac{\log |\lambda'|}{\log |\lambda|}.$$

Proof. For fixed λ' , assuming without loss of generality $a'_1 \geq b'_1$, and fixed $|\lambda|$, the value of $\log D(\lambda)$ is minimized when $a_1 = a'_1 + |\lambda| - |\lambda'|$, $a_i = a'_i$ for $i \geq 2$, and $b_i = b'_i$ for all i . It follows that there exists $K \geq 1$ such that

$$D(\lambda) = \frac{(|\lambda| - 1)!}{a_1! K},$$

$$D(\lambda') = \frac{(|\lambda'| - 1)!}{a'_1! K},$$

and so

$$\begin{aligned} \frac{\log D(\lambda')}{\log D(\lambda)} &= \frac{\log(|\lambda'| - 1) + \log(|\lambda'| - 2) + \cdots + \log(a'_1 + 1) - \log K}{\log(|\lambda| - 1) + \log(|\lambda| - 2) + \cdots + \log(a_1 + 1) - \log K} \\ &\leq \frac{\log(|\lambda'| - 1) + \log(|\lambda'| - 2) + \cdots + \log(a'_1 + 1)}{\log(|\lambda| - 1) + \log(|\lambda| - 2) + \cdots + \log(a_1 + 1)} \\ &\leq \frac{\log |\lambda'|}{\log |\lambda|}. \end{aligned}$$

□

By the *index* of a box in a Young diagram λ , we mean the number of squares to the left of it minus the number of squares above it. Thus $D(\lambda)$ equals $(|\lambda| - 1)!$ divided by the product of the absolute values of the indices of all boxes of λ of nonzero index. We recall that a *rim hook* μ of λ is the union of a sequence of squares in λ such that each box in the sequence is directly to the right of or directly above the previous box and $\lambda \setminus \mu$ is a Young diagram. The indices of boxes in a rim hook form a set of consecutive integers. The largest of these integers corresponds to a box on the right of its row; if the integer is positive, the row is among the first m . Likewise the smallest term corresponds to a box on the bottom of its column and if it is negative, then the column is among the first m .

If μ is a rim k -hook, i.e., a rim hook of length k , whose indices $c_1 < \cdots < c_k$ are non-negative, its highest index box lies in the i th row, $i \leq m$, and

$$(11) \quad \frac{D(\lambda \setminus \mu)^{1/k}}{D(\lambda)^{1/k}} \leq \frac{a_i}{n-1}.$$

Indeed,

$$\frac{D(\lambda \setminus \mu)}{D(\lambda)} = \frac{c_k}{n-1} \frac{c_k-1}{n-2} \cdots \frac{c_k-k+1}{n-k},$$

and

$$\frac{c_k-j}{n-1-j} \leq \frac{c_k}{n-1} = \frac{a_i}{n-1}.$$

If $c_1 < \dots < c_k \leq 0$, the lowest index box lies in the j th column, $j \leq m$, and

$$(12) \quad \frac{D(\lambda \setminus \mu)^{1/k}}{D(\lambda)^{1/k}} \leq \frac{b_j}{n-1}.$$

If $c_1 < 0$ and $c_k > 0$, then the highest index box lies at the end of the i th row and the lowest index box lies at the bottom of the j th column, $1 \leq i, j \leq m$, $c_1 = -b_j$, and $c_k = a_i$. Then

$$(13) \quad \begin{aligned} \frac{D(\lambda \setminus \mu)^{1/k}}{D(\lambda)^{1/k}} &\leq \left[\frac{b_j! \cdot 1 \cdot a_i!}{(n-1)(n-2) \cdots (n-a_i-b_j-1)} \right]^{1/k} \\ &\leq \left[\frac{(a_i+b_j)!}{(n-1)(n-2) \cdots (n-a_i-b_j)} \right]^{1/k} \leq \frac{a_i+b_j}{n-1}. \end{aligned}$$

Combining equations (11)–(13), we deduce

$$(14) \quad \sum_{\mu} \left(\frac{D(\lambda \setminus \mu)}{D(\lambda)} \right)^{1/k} \leq \frac{1}{n-1} \sum_{i=1}^m a_i + \frac{1}{n-1} \sum_{j=1}^m b_j < 1,$$

where μ in the left hand summation ranges over all rim k -hooks.

3. CHARACTER VALUES

In this section we use the estimates of the previous section, together with the Murnaghan–Nakayama rule [Ja, 21.1], to prove Theorem 1.1.

Lemma 3.1. *Let λ be a Young diagram with n boxes, and let σ be a permutation in the class $(1^{c_1} 2^{c_2} \cdots k^{c_k})$ of S_n . Let σ' be a permutation in $(1^{c_1} \cdots (k-1)^{c_{k-1}}) \subset S_{n'}$. Then*

$$\frac{|\chi_{\lambda}(\sigma)|}{D(\lambda)^{1/k}} \leq \sup_{\lambda'} \frac{|\chi_{\lambda'}(\sigma')|}{D(\lambda')^{1/k}}$$

where λ' ranges over all Young diagrams with n' boxes which are contained in λ .

Proof. If $\sigma = \rho\tau$, where ρ is a k -cycle and τ consists of the remaining cycles in σ , then the Murnaghan–Nakayama rule gives an equality of the form

$$\chi_{\lambda}(\sigma) = \sum_{\mu} (-1)^{l(\mu)} \chi_{\lambda \setminus \mu}(\tau)$$

and therefore an inequality

$$(15) \quad |\chi_{\lambda}(\sigma)| \leq \sum_{\mu} |\chi_{\lambda \setminus \mu}(\tau)|,$$

where the sum is taken over all rim k -hooks of λ . Clearly,

$$(16) \quad |\chi_{\lambda \setminus \mu}(\tau)| \leq D(\lambda \setminus \mu)^{1/k} \sup_{\lambda'} \frac{|\chi_{\lambda'}(\tau)|}{D(\lambda')^{1/k}},$$

where λ' ranges over all Young subdiagrams of λ containing $n - k$ boxes. By (15), (16), and (14),

$$\begin{aligned} \frac{|\chi_{\lambda}(\sigma)|}{D(\lambda)^{1/k}} &\leq \sum_{\mu} \frac{|\chi_{\lambda \setminus \mu}(\tau)|}{D(\lambda)^{1/k}} \\ &\leq \sum_{\mu} \frac{D(\lambda \setminus \mu)^{1/k}}{D(\lambda)^{1/k}} \sup_{\lambda'} \frac{|\chi_{\lambda'}(\tau)|}{D(\lambda')^{1/k}} \\ &< \sup_{\lambda'} \frac{|\chi_{\lambda'}(\tau)|}{D(\lambda')^{1/k}}. \end{aligned}$$

The lemma now follows by induction on k . □

The above lemma for $k = 1$ yields the following result of independent interest, relating the degree and the virtual degree.

Corollary 3.2. *For every Young diagram λ we have*

$$d(\lambda) \leq D(\lambda).$$

We now prove Theorem 1.1.

Proof. We first prove part (i) of the theorem and then deduce part (ii).

Every orbit growth sequence is supported on $\{1, 2, \dots, k\}$ for some k . We prove the theorem by induction on k . The induction hypothesis is

$$(17) \quad |\chi_{\lambda}(\sigma)| \leq D(\lambda)^{e_1 + e_2/2 + e_3/3 + \dots + e_k/k}$$

if $e_1 + \dots + e_k = 1$. The case $k = 1$ is the inequality $d(\lambda) \leq D(\lambda)$ given in the corollary above. Assume this hypothesis for $k - 1$. Let $\sigma \in 1^{c_1} 2^{c_2} \dots k^{c_k}$ and $\sigma' \in 1^{c_1} 2^{c_2} \dots (k - 1)^{c_{k-1}}$, and let λ' denote an n' -box subdiagram of λ for which the supremum in Lemma 3.1 is achieved. Let e'_1, e'_2, \dots denote the orbit growth sequence of σ' ; thus,

$$e'_i = \begin{cases} \frac{e_i}{e_1 + \dots + e_{k-1}} & \text{if } i < k \\ 0 & \text{if } i \geq k. \end{cases}$$

By Lemma 2.3,

$$D(\lambda') \leq D(\lambda)^{\frac{\log n'}{\log n}} = D(\lambda)^{e_1 + \dots + e_{k-1}}.$$

By the induction hypothesis,

$$\begin{aligned}
|\chi_{\lambda'}(\sigma)| &\leq D(\lambda')^{e'_1 + \dots + \frac{e'_{k-1}}{k-1}} \\
&= D(\lambda')^{\frac{e_1 + \dots + \frac{e_{k-1}}{k-1}}{e_1 + \dots + e_{k-1}} - \frac{1}{k}} D(\lambda')^{1/k} \\
&\leq D(\lambda)^{e_1 + \dots + \frac{e_{k-1}}{k-1} - \frac{1}{k}(e_1 + \dots + e_{k-1})} D(\lambda')^{1/k}.
\end{aligned}$$

By Lemma 3.1,

$$\frac{|\chi_{\lambda}(\sigma)|}{D(\lambda)^{1/k}} \leq D(\lambda)^{e_1 + \dots + (e_k/k - 1/k)}.$$

Equation (17) follows by induction, and part (i) of the theorem follows from Theorem 2.2.

To deduce part (ii) it suffices to show that

$$E(\sigma) \leq B(\sigma) + o(1).$$

Let (b_k) be the cycle growth sequence of σ . Clearly $b_1 = e_1$ and in general, since $|\Sigma_i| \leq iF_{\sigma}(i)$ we have

$$\sum_{j=1}^i e_j \leq b_i + \frac{\log i}{\log n}.$$

Therefore, for every positive integer k ,

$$\begin{aligned}
E(\sigma) &= \sum_{i=1}^{\infty} \frac{e_i}{i} = \sum_{i=1}^{\infty} \frac{1}{i} \sum_{j=1}^i e_j - \sum_{i=2}^{\infty} \frac{1}{i} \sum_{j=1}^{i-1} e_j \\
&= \sum_{i=1}^{\infty} \frac{1}{i} \sum_{j=1}^i e_j - \sum_{i=1}^{\infty} \frac{1}{i+1} \sum_{j=1}^i e_j \\
&= \sum_{i=1}^{\infty} \left(\frac{1}{i} - \frac{1}{i+1} \right) \sum_{j=1}^i e_j \\
&\leq \sum_{i=1}^k \left(\frac{1}{i} - \frac{1}{i+1} \right) \sum_{j=1}^i e_j + \sum_{i=k+1}^{\infty} \left(\frac{1}{i} - \frac{1}{i+1} \right) \\
&\leq \sum_{i=1}^k \left(\frac{1}{i} - \frac{1}{i+1} \right) \left(b_i + \frac{\log i}{\log n} \right) + \frac{1}{k+1} \\
&\leq \sum_{i=1}^k \frac{b_i}{i(i+1)} + \frac{\log k}{\log n} + \frac{1}{k+1}.
\end{aligned}$$

Given $\epsilon > 0$ choose k such that $1/(k+1) < \epsilon/2$. Then for $n \gg 0$ we have $\frac{\log k}{\log n} < \epsilon/2$ and so

$$E(\sigma) < B(\sigma) + \epsilon.$$

This proves part (ii). □

We can now prove Theorems 1.2–1.4.

Proof. Setting $f = \max(\text{fix}(\sigma), 1)$, we have

$$e_1 = \frac{\log f}{\log n},$$

so

$$E(\sigma) \leq e_1 + \frac{1 - e_1}{2} \leq \frac{1 + e_1}{2} \leq \frac{\log n + \log f}{2 \log n} = 1 - \frac{\log(n/f)}{2 \log n},$$

and Theorem 1.1 implies Theorem 1.3. This in turn implies part (i) of Theorem 1.2. For part (ii), we use the fact that if σ has less than n^ϵ cycles of length less than m , then $b_1, b_2, \dots, b_{m-1} \leq \epsilon$, so

$$B(\sigma) = \sum_{i=1}^{\infty} \frac{b_i}{i(i+1)} = \sum_{i=1}^{m-1} \frac{b_i}{i(i+1)} + \sum_{i=m}^{\infty} \frac{b_i}{i(i+1)} < \epsilon + \frac{1}{m}.$$

Part (ii) of Theorem 1.2 now follows from part (ii) of Theorem 1.1.

For Theorem 1.4 we note that $\text{cyc}(\sigma) \leq n^\alpha$ implies $b_k \leq \alpha$ for all k . Hence

$$B(\sigma) \leq \sum_{k=1}^{\infty} \frac{\alpha}{k(k+1)} = \alpha.$$

The result now follows from part (ii) of 1.1. □

4. RANDOM WALKS

For a subset C of a finite group G , let P_C denote the uniform distribution on C , and set

$$P_C^t = \underbrace{P_C * \dots * P_C}_{t \text{ times}},$$

where $*$ denotes convolution over G . Consider the random walk on G starting with 1, and performing successive multiplications by randomly chosen elements of C . Then the probability to reach $g \in G$ after t steps of this random walk on G based on C is $P_C^t(g)$.

In this section C will be a conjugacy class of G , but we shall also deal with general normal subsets in Section 6.

Let $U = U_G$ be the uniform distribution on G . Consider the L_1 -distance $\|P_C^t(g) - U\|$ between the two distributions defined above. The *mixing time* $T(C)$ of the random walk is defined as the minimal time t such that

$$\|P_C^t(g) - U\| < 1/e.$$

In our asymptotic contexts (G, C) usually ranges over an infinite sequence (G_n, C_n) , and when we say that $T(C_n) \leq t$ we usually prove a somewhat stronger statement, that $\|P_{C_n}^t - U_{G_n}\| \rightarrow 0$ as $n \rightarrow \infty$.

We start with the useful upper bound lemma of Diaconis and Shahshahani [DS].

Lemma 4.1. *With the above notation we have*

$$\|P_C^t - U_G\| \leq \sum_{1 \neq \chi \in \text{Irr } G} \frac{\chi(C)^{2t}}{\chi(1)^{2t-2}}.$$

When we study random walks on S_n based on a conjugacy class C , the sign after t steps of the walk is fixed. If we then obtain an almost uniform distribution on a coset of A_n (in the sense of L_1 -distance $\leq 1/e$, or tending to 0) we say that the mixing time is at most t .

For simplicity of notation we shall mostly discuss S_n -classes C which are contained in A_n , the general case being very similar. If $C = \sigma$ is such a class, then standard Fourier techniques show that

$$(18) \quad \|P_C^t - U_{A_n}\| \leq \sum_{\chi \in \text{Irr } S_n, \chi(1) > 1} \frac{\chi(C)^{2t}}{\chi(1)^{2t-2}}.$$

If $C \subset S_n \setminus A_n$ then the same holds with the uniform distribution $U_{\sigma^t A_n}$ on $\sigma^t A_n$, where $\sigma \in C$.

We also need a result from [LiSh2] concerning the Witten zeta function of S_n , defined by

$$\zeta^{S_n}(s) = \sum_{\chi \in \text{Irr } S_n} \chi(1)^{-s}.$$

Lemma 4.2. *Fix a real number $s > 0$. Then we have*

$$\zeta^{S_n}(s) = 2 + O(n^{-s}).$$

In particular

$$\sum_{\chi \in \text{Irr } S_n, \chi(1) > 1} \chi(1)^{-s} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

We can now obtain the following

Proposition 4.3. *Let C be a conjugacy class of S_n and suppose $C \subset A_n$. Let $\gamma > 0$ be a real number satisfying*

$$|\chi(C)| \leq \chi(1)^{1-\gamma} \text{ for all } \chi \in \text{Irr } S_n.$$

Then for an integer $t > \gamma^{-1}$ we have

$$\|P_C^t - U_{A_n}\| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Consequently, if n is large enough (given γ), then the mixing time $T(C)$ satisfies

$$T(C) \leq \lceil \gamma^{-1} \rceil + 1.$$

Proof. Combining the bound (18) with the inequality $|\chi(C)| \leq \chi(1)^{1-\gamma}$ it follows that

$$\|P_C^t - U_{A_n}\| \leq \sum_{\chi \in \text{Irr } S_n, \chi(1) > 1} \chi(1)^{-s} = \zeta^{S_n}(s) - 2,$$

where $s = 2t\gamma - 2$. Since $t > \gamma^{-1}$ we have $s > 0$, so applying Lemma 4.2 above we obtain

$$\|P_C^t - U_{A_n}\| = O(n^{-s}),$$

which implies the first assertion. The second follows immediately from the first. \square

Proof of Theorem 1.5.

We have $E(\sigma) \leq 1 - 1/t - \epsilon$, so applying Theorem 1.1, there exists N (depending on ϵ) such that if $n \geq N$ we have

$$|\chi(\sigma)| \leq \chi(1)^{1-1/t-\epsilon/2} \text{ for all } \chi \in \text{Irr}(S_n).$$

We now apply the above proposition with $\gamma = 1/t + \epsilon/2$. Since $\gamma^{-1} < t$ we obtain

$$T(\sigma^{S_n}) \leq \lceil \gamma^{-1} \rceil + 1 \leq t$$

for all n large enough. \square

Proof of Theorem 1.6.

If $T(C)$ is bounded then a result of Vishne [V] shows that for $\sigma \in C$ we have $\text{fix}(\sigma) \leq n^{1-\epsilon}$ for some $\epsilon > 0$. It remains to prove the converse. Suppose $\sigma \in C$ and $f = \text{fix}(\sigma) \leq n^{1-\epsilon}$ where $\epsilon > 0$. Then we have $E(\sigma) \leq (1 - \epsilon) + \epsilon/2 = 1 - \epsilon/2$. Choose a minimal integer $t > 2/\epsilon$ and let $\delta = \epsilon/2 - 1/t$. Then $E(\sigma) = 1 - 1/t - \delta$, and so by Theorem 1.5 we have

$$T(\sigma^{S_n}) \leq t = \lceil 2\epsilon^{-1} \rceil + 1.$$

for all large n . The result follows. \square

Proof of Theorem 1.7.

The upper bound on $T(C)$ follows from the inequality above.

The lower bound follows from the fact that if $f = \text{fix}(\sigma) = n^{1-\epsilon}$ then the intersection of the sets of fixed points of t random conjugates of σ has expected cardinality $n^{1-t\epsilon}$. Hence, if $t\epsilon < 1$, a random product of t elements of C has too many fixed points and there is no mixing yet. \square

Proof of Theorem 1.8.

Let $\sigma \in C$.

If $\text{fix}(\sigma) \leq n^{o(1)}$ then

$$E(\sigma) \leq 1/2 + o(1) \leq 1 - 1/3 - \epsilon,$$

for some $\epsilon > 0$ and all $n \gg 0$. Applying Theorem 1.5 we conclude that $T(C) \leq 3$. This proves part (i) of the theorem and the Lulov-Pak conjecture.

For part (ii), note that $E(\sigma) \leq 1/3 + o(1)$ so $E(\sigma) \leq 1 - 1/2 - \epsilon$ for some $\epsilon > 0$ and all $n \gg 0$. It follows from Theorem 1.5 that $T(C) = 2$. \square

In fact our arguments give rise to the following quantitative solution to the Lulov-Pak conjecture.

Corollary 4.4. *Let $C_n \subset S_n$ be a sequence of conjugacy classes with no fixed points. Let $\sigma_n \in C_n$. Then for any fixed $\epsilon > 0$ we have*

$$\|P_{C_n}^3 - U_{\sigma_n A_n}\| = O(n^{-1+\epsilon}).$$

Proof. As above we have $\chi(C_n) \leq \chi(1)^{1/2+o(1)}$. Substituting this in (18) gives $\|P_{C_n}^3 - U_{\sigma_n A_n}\| \leq \zeta^{S_n}(1 - \delta) - 2$ for any fixed $\delta > 0$ and $n \gg 0$. The result now follows from 4.2. \square

5. COVERING NUMBERS

The motivating problem of this section is to find, for each conjugacy class C in a symmetric group S_n , the minimum constant k such that C^k is a coset of A_n . (Of course, C^k is always *contained* in such a coset, so this means C^k is as large as possible.) In particular, we prove that for most C , in a suitable sense, $C^2 = A_n$.

Our basic estimate is the following consequence of Theorem 1.1.

Proposition 5.1. *For all $\epsilon > 0$ there exists N such that for all $n \geq N$ and all $\sigma \in S_n$ and $\tau \in A_n$ with*

$$2E(\sigma) + E(\tau) < 1 - \epsilon,$$

we have $\tau \in (\sigma^{S_n})^2$.

Proof. Let p denote the probability that $xy = \tau$, where x and y are independent, uniformly distributed random elements of σ^{S_n} .

By a formula of Frobenius we have

$$(19) \quad p = \frac{1}{n!} \sum_{\chi \in \text{Irr}(S_n)} \frac{\chi(\sigma)^2 \bar{\chi}(\tau)}{\chi(1)}.$$

Note that the contribution of the two linear characters of S_n to the right hand side is $2/n!$, so

$$\left| p - \frac{2}{n!} \right| \leq \frac{1}{n!} \sum_{\chi \in \text{Irr}(S_n), \chi(1) > 1} \frac{|\chi(\sigma)|^2 |\chi(\tau)|}{\chi(1)}.$$

By taking N sufficiently large, we can guarantee that $|\chi(\sigma)| \leq \chi(1)^{E(\sigma) + \epsilon/4}$ and $|\chi(\tau)| \leq \chi(1)^{E(\tau) + \epsilon/4}$. This yields

$$\left| p - \frac{2}{n!} \right| \leq \frac{1}{n!} \sum_{\chi \in \text{Irr}(S_n), \chi(1) > 1} \frac{\chi(1)^{2E(\sigma) + E(\tau) + 3\epsilon/4}}{\chi(1)} = \frac{1}{n!} (\zeta^{S_n}(s_0) - 2),$$

where

$$s_0 := 1 - 2E(\sigma) - E(\tau) - 3\epsilon/4.$$

Since $2E(\sigma) + E(\tau) < 1 - \epsilon$ by our hypothesis it follows that $s_0 > \epsilon/4$.

Now, applying Lemma 4.2 above we have $\zeta^{S_n}(s_0) - 2 \rightarrow 0$ as $n \rightarrow \infty$. Therefore $p = (2 + o(1))/n!$, and so $p > 0$ when n is sufficiently large. The result follows. \square

We remark that in spite of its strength and novelty, this proposition alone is never enough to prove $(\sigma^{S_n})^2 = A_n$, since $E(\tau)$ can be arbitrarily close to 1. We must combine it with another technique which works well when $E(\tau)$ is large, namely the cancellation method, which we shall now develop.

For any permutation σ , define $f_\sigma(k)$ to be the number of k -cycles in σ . We write $f_\tau \prec f_\sigma * f_\sigma$ if and only if $\tau \in \sigma^{S_n} \sigma^{S_n}$. Using the natural embedding $S_{n_1} \times S_{n_2} \hookrightarrow S_{n_1+n_2}$, we see that if $f_1 \prec g_1 * g_1$ and $f_2 \prec g_2 * g_2$, then $(f_1 + f_2) \prec (g_1 + g_2) * (g_1 + g_2)$. Turning this around, $f_3 \prec g_3 * g_3$ and $f_2 \prec g_2 * g_2$ imply $f_1 \prec g_1 * g_1$ as long as $f_1 := f_3 - f_2$ and $g_1 := g_3 - g_2$ take only nonnegative values. We can therefore use a collection of simple expressions of the form $f_2 \prec g_2 * g_2$ to perform systematic reduction of this kind. The following three lemmas will be used in this way.

Lemma 5.2. *Let $n \geq 4$. If $(2^{c_2} \dots r^{c_r})$ is any even conjugacy class of permutations, there exist c_1 and m such that*

$$(1^{c_1} \dots r^{c_r}) \subset (n^m)(n^m).$$

Proof. Let the sequence $a_1, a_2, \dots, a_k \geq 2$ consist of c_i repetitions of i for $2 \leq i \leq r$. Let $m = \frac{1}{2} \sum (a_i - 1)$. We claim that there exist $\sigma, \tau \in n^m \subset S_{nm}$ such that $\sigma\tau$ has exactly k non-trivial orbits X_1, \dots, X_k , $|X_i| = a_i$, and for all i with $a_i \geq 3$, there exists $j \in \{1, 2, \dots, mn\}$ such that $\{j, \tau(j)\} \subset X_i$.

It is enough to check this for $k = 1$, in which case a_1 is odd, and for $k = 2$, when a_1 and a_2 are both even. We proceed by induction: if the claim is true for a_1, \dots, a_k with $a_k \geq 3$, then it is true for $a_1, \dots, a_{k-1}, a_k + 2$. Indeed, if $\sigma, \tau \in (m^n)$, $\sigma\tau$ has k non-trivial orbits X_i , $|X_i| = a_i$, and $\{j, \tau(j)\} \subset X_k$, let

$$\sigma' := \sigma((mn+1) (mn+2) \cdots (mn+n))$$

and

$$\tau' := (\tau(j) (mn+1))[\tau((mn+n) \cdots (mn+2) (mn+1))](\tau(j) (mn+1)).$$

The fixed points of $\sigma'\tau'$ are the fixed points of $\sigma\tau$, together with $mn+3, \dots, mn+n$. The $\sigma\tau$ -orbits X_1, \dots, X_{k-1} are $\sigma'\tau'$ -orbits, as is

$$X_k \cup \{mn+1, mn+2\}.$$

There are four base cases:

- $a_1 = 3$

$$\sigma = (1 \ 2 \ \cdots \ n), \quad \tau = (n \ n-1 \ \cdots \ 3 \ 1 \ 2);$$

- $a_1 = a_2 = 2$

$$\sigma = (1 \ 2 \ \cdots \ n), \quad \tau = (n \ n-1 \ \cdots \ 4 \ 1 \ 2 \ 3);$$

- $a_1 = 2, a_2 = 4$

$$\sigma = (1 \ 2 \ \cdots \ n)(1' \ 2' \ \cdots \ n'),$$

$$\tau = (n \ (n-1) \ \cdots \ 3 \ 2 \ 1')(n' \ (n-1)' \ \cdots \ 4' \ 2' \ 3' \ 1);$$

- $a_1 = 4, a_2 = 4$

$$\sigma = (1 \ 2 \ \cdots \ n)(1' \ 2' \ \cdots \ n')(1'' \ 2'' \ \cdots \ n''),$$

$$\tau = (n \ n-1 \ \cdots \ 3 \ 2 \ 1')(n' \ (n-1)' \ \cdots \ 3' \ 1 \ 2'')(n'' \ (n-1)'' \ \cdots \ 3'' \ 1'' \ 2'').$$

□

Lemma 5.3. *For all integers $a \geq 2$ and $b \geq 1$, there exist integers u and v such that $au = bv$ and*

$$(b^v) \subset (a^u)^2.$$

Proof. This follows from Proposition 5.1 as long as $2/a + 1/b < 1$. So we must consider the cases $a = 2, b = 1$, and $(a, b) \in \{(3, 2), (3, 3), (4, 2)\}$. For these, it suffices to find some finite group G and elements $x, y \in G$ of order a whose product is of order b . We then realize x, y as permutations via the regular representation of G . The $a = 2$ case can be handled with G a dihedral group, the $b = 1$ case with G cyclic, and the remaining cases with $G = S_4$. \square

Lemma 5.4. *For all positive integers a and b with $2/a + 1/b < 1$, there exist positive integers p, q and r such that*

$$(b^r) \subset (1^p a^q)(1^p a^q).$$

Proof. Letting $p := b$, taking q divisible by b and sufficiently large, and defining $r := \frac{p+aq}{b}$, we can make sure that

$$2E(1^p a^q) + E(b^r) < 1,$$

so the lemma follows from Proposition 5.1. \square

We now prove Theorem 1.10 (and the weaker version, Theorem 1.9).

Proof. We can replace (2) by the equivalent condition that the average cycle length exceeds $4 + \epsilon'$ for some fixed $\epsilon' > 0$.

Let τ be an element of A_n . We would like to show that if n is sufficiently large, τ can be written as the product of two conjugates of σ . We proceed by a sequence of reduction steps, mainly using the cancellation method, until we reduce to a case which can be handled by the character method.

Step 1. For any fixed integer N , we may assume without loss of generality that the support of τ is $\geq N$. Indeed, if $f_\sigma(k) > 0$ for some $k \geq N$, then $\tau \in (\sigma^{S_n})^2$ since every element of A_k is the product of two k -cycles by a theorem of Gleason [Hu]. On the other hand if every cycle of σ has length less than N , then for n sufficiently large, we can use Lemma 5.2 to represent τ .

Step 2. We may assume that if $f_k(\sigma) > 0$, then $f_1(\tau) < k$, i.e., the minimum orbit length k in σ is greater than the number of fixed points of τ . Indeed, otherwise, we can cancel a k -cycle in σ against k 1-cycles in τ . To justify this, we must check that we can perform the reduction step repeatedly without sacrificing any of the hypotheses (2)–(5).

Let σ' denote a permutation in $S_{n'} = S_{n-k}$ obtained by removing a k -cycle from σ . The average cycle length of σ' is at least as great as that of σ , so (2) holds for σ' . If $k \geq 3$, then (4) holds for σ' unconditionally. If $k \leq 2$ and (4) holds for σ , then

$$\text{fix}(\sigma'^2) = \text{fix}(\sigma^2) - k < n^{1/4-\epsilon} - k < (n-k)^{1/4-\epsilon} = n'^{1/4-\epsilon}.$$

If (5) holds for σ , then $k \geq 2$. If $k = 2$, then

$$\text{fix}(\sigma'^2) = \text{fix}(\sigma^2) - 2 < n^{1-\epsilon} - 2 < (n-2)^{1-\epsilon} = n'^{1-\epsilon};$$

if $k > 2$, then $\text{fix}(\sigma'^2) = \text{fix}(\sigma^2) = 0$. In either case, (5) holds for σ' . If (3) holds, then

$$\begin{aligned} E(\sigma') &= \frac{1}{\log n'} \sum_{i=k}^{\infty} \frac{\log^+ |\Sigma'_i| - \log^+ |\Sigma'_{i-1}|}{i} \\ &\leq \frac{1}{\log n'} \sum_{i=k}^{\infty} \frac{\log^+ |\Sigma_i| - \log^+ |\Sigma_{i-1}|}{i} \\ &= \frac{\log n}{\log n'} E(\sigma) \\ &< \frac{\log n}{\log n'} \left(\frac{1}{4} - \epsilon \right). \end{aligned}$$

Iterating this process, we obtain new permutations, which we denote σ'' and τ'' in $S_{n''}$, such that $f_k(\sigma'') > 0$ implies $f_1(\tau'') < k$, for which we must show $\tau'' \in (\sigma''^{S_{n''}})^2$. By Step 1, we know that $n'' \geq N$, where we may take N as large as we wish. Conditions (2), (4), and (5) hold for σ'' if they hold for σ . For (3), if the minimal cycle length in σ'' is ≥ 5 , replacing ϵ by $\epsilon'' < 1/20$, we have $E(\sigma'') < 1/4 - \epsilon''$. Otherwise, as we have removed only 1-cycles, 2-cycles, 3-cycles, and 4-cycles, we have a lower bound for n'' linear in n , with a positive constant depending only on ϵ . Setting $\epsilon'' = \epsilon/2$,

$$E(\sigma'') < \frac{\log n}{\log n''} (1/4 - \epsilon) < 1/4 - \epsilon'',$$

if n is sufficiently large. Replacing the original σ , τ , and ϵ (if necessary) by σ'' , τ'' , and ϵ'' (if necessary), we may therefore assume, without sacrificing the original hypotheses, that $f_k(\sigma) > 0$ implies $f_1(\tau) < k$ as claimed.

Step 3. We may assume that $f_\tau(1) < n^{2/3}$. Otherwise, since all the cycles in σ have length greater than $f_\tau(1)$, there can be at most $n^{1/3}$ cycles in σ . When n is sufficiently large the number of fixed points in τ is more than 7 times the number of cycles in σ , and this implies τ is a product of two conjugates of σ by [LaSh, Proposition 6.1].

Step 4. We may assume that $f_\tau(1) < 12$. Indeed, let k be the length of the smallest orbit of σ . Otherwise, $k > 12$, so $E(\sigma) \leq 1/13$. If e_1, e_2, \dots is the orbit growth sequence of τ , then $e_1 \leq 2/3$, so $E(\tau) \leq 5/6$. For n sufficiently large, $\tau \in \sigma^{S_n} \sigma^{S_n}$ by Proposition 5.1.

Step 5. At this point we can complete the proof if (3) holds. Indeed, if n is sufficiently large, $f_\tau(1) < 12$ implies $E(\tau) < \frac{1}{2} + \epsilon$. The theorem now follows from Proposition 5.1.

Step 6. For any integer r there exists s such that we may assume that either σ has less than s a -cycles for all $2 \leq a \leq 4$, or τ has less than s b -cycles for all $b \leq r$. Indeed we may cancel sets u of a -cycles against sets of v b -cycles by Lemma 5.3 with $au = bv \leq s$. If $\sigma' \in S_{n'} = S_{n-au}$ is obtained by removing u a -cycles from $\sigma \in S_n$, the average cycle length of σ' is greater than that of σ by (2). It is not always the case that if σ satisfies (4) or (5), the same is true for σ' . However, if this process is iterated and ultimately produces elements $\sigma'', \tau'' \in S_{n''}$, then $n - n'' < 4(1/4 - \epsilon)n$, so $n'' > 4\epsilon n$. If we set $\epsilon'' = \epsilon/2$, then if n is sufficiently large, (4) and (5) hold with σ , n , and ϵ replaced by σ'' , n'' , and ϵ'' respectively. Replacing the original σ , n , and ϵ by σ'' , n'' , and ϵ'' respectively, the reduction step is justified.

Step 7. For any integer r there exists s such that we may assume one of the following conditions holds:

- (a) $f_\sigma(i) < s$ for $1 \leq i \leq 4$.
- (b) $f_\sigma(i) < s$ for $2 \leq i \leq r$.
- (c) $f_\tau(i) < s$ for $1 \leq i \leq r$.

Indeed, we are already assuming that σ has a bounded number of 2-cycles, 3-cycles, and 4-cycles. We use Lemma 5.4 to cancel combinations of 1-cycles and a -cycles, $5 \leq a \leq r$, in σ against b -cycles, $b \leq r$, in τ . Each such combination has at least one 1-cycle; as $f_1(\sigma) < n^{1/4 - \epsilon}$, if n is sufficiently large we may assume that the length n'' of the resulting permutation is at least $n - n^{1/4}$. Replacing ϵ by $\epsilon/2$, therefore, we may assume that after all such cancellations are completed, the original hypotheses are satisfied with the new σ , τ , n , and ϵ .

Step 8. We may assume that $2E(\sigma) + E(\tau) < 1 - \frac{\epsilon}{20}$. In case (a), $E(\sigma) < 1/5 + o(1)$, and $E(\tau) = 1/2 + o(1)$. In case (b), if r is sufficiently large

$$E(\sigma) < 1/4 - \epsilon + \frac{3/4 + \epsilon}{r + 1} + o(1) < 1/4 - \epsilon/2 + o(1),$$

and $E(\tau) = 1/2 + o(1)$. In case (c),

$$E(\sigma) < \max\left(1/4 - \epsilon + \frac{3/4 + \epsilon}{3}, \frac{1 - \epsilon}{2} + \frac{\epsilon}{3}\right) + o(1) = \frac{1}{2} - \frac{\epsilon}{6} + o(1),$$

and taking r large enough, we may assume $E(\tau) < \frac{\epsilon}{12}$.

The theorem now follows from Proposition 5.1. □

We can now easily deduce Corollary 1.11.

Proof. Fix $0 < \epsilon < 1/20$. In view of Theorem 1.10 it suffices to show that there exists N such that, if $n \geq N$ and $\sigma \in S_n$ satisfies $E(\sigma) \leq 1/4 - \epsilon$ then $\text{cyc}(\sigma) < (1/4 - \epsilon)n$.

Let (e_k) be the orbit growth sequence of σ . Then we have

$$e_1 + e_2 + e_3 + e_4 \leq 4(e_1 + e_2/2 + e_3/3 + e_4/4) \leq 4E(\sigma).$$

Assuming $E(\sigma) \leq 1/4 - \epsilon$ we obtain

$$e_1 + e_2 + e_3 + e_4 \leq 1 - 4\epsilon.$$

This shows that $|\Sigma_4| \leq n^{1-4\epsilon}$, namely at most $n^{1-4\epsilon}$ points lie in σ -cycles of length at most 4. Clearly σ has at most $n/5$ cycles of length exceeding 4, and so

$$\text{cyc}(\sigma) \leq n^{1-4\epsilon} + n/5 = (1/5 + n^{-4\epsilon})n.$$

The right hand side is clearly $< (1/4 - \epsilon)n$ for n sufficiently large. Corollary 1.11 is proved. \square

We deduce Theorem 1.12 immediately.

Proof. The case $m = 4$ is in [BR], so suppose $m \geq 5$. For any $\epsilon < 1/20$ conditions (2) and (5) are satisfied. Hence Theorem 1.10 yields the result. \square

Note that Theorem 1.10 also implies the stronger version of the Rudvalis-Vishne conjecture for $n \gg 0$, where some fixed points are allowed (see the paragraph following Theorem 1.12).

We now prove Theorem 1.14.

Proof. We claim that there exists $\tau \in (\sigma^{S_n})^2$ such that the number of cycles in τ of length at most 4 is bounded. Applying Theorem 1.10 to τ we then obtain the required conclusion.

We prove the claim by the method of cancellation, τ being defined precisely to make the needed cancellation work. The equation

$$(20) \quad [(12)(34)][(23)(45)] = (12453)$$

shows that two 2-cycles and a 1-cycle in σ can be cancelled against a 5-cycle in τ . We can repeat this process as needed without sacrificing the property

$$(21) \quad \text{fix}(\sigma) \leq n/5.$$

We may therefore assume either $f_\sigma(1) = 0$ or $f_\sigma(2) \leq 1$.

If $0 \leq k < m$ and $k + m$ is odd, then

$$(1 \ 2 \ \cdots \ m)(k + 1 \ k + 2 \ \cdots \ k + m)$$

is a $(k + m)$ -cycle. If $f_\sigma(1) \geq 2$, $f_\sigma(m) > 0$ for some $m \geq 3$, and (21) is satisfied, then cancelling an m -cycle and k 1-cycles in σ against an

$(m + k)$ -cycle in τ for a suitable value of k with $m + k \geq 5$, we can arrange that (21) is still satisfied, or $f_\sigma(1) \leq 1$ after the cancellation has been carried out. Indeed we may take k to be the largest integer $\leq \max(f_\sigma(1), m - 1)$ of the same parity as $m - 1$. Iterating, we may assume that $f_\sigma(1) \leq 1$, since the alternative, $f_\sigma(m) = 0$ for all $m \geq 3$, $f_\sigma(2) \leq 1$, $f_\sigma(1) > 1$, and (21) still satisfied, is impossible.

For $2 \leq r \leq 8$ we use Lemma 5.3 to cancel sets of r -cycles in σ against sets of 5-cycles in τ . Thus $f_\sigma(r)$ can be assumed bounded for $1 \leq r \leq 8$.

For $r \geq 9$, a single r -cycle in σ can be cancelled against one r -cycle in τ or two $r/2$ -cycles in τ depending on whether r is odd or even. This implies that $\tau \in (\sigma^{S_n})^2$ can be chosen with a bounded number of cycles of length at most 4, proving the claim. □

Proof of Corollary 1.15.

This follows immediately from Theorem 1.5 and Theorem 1.14. Indeed bounded mixing time of σ^{S_n} implies $\text{fix}(\sigma) \leq n^{1-\epsilon}$ for all large n where $\epsilon > 0$, which yields $\text{fix}(\sigma) \leq n/5$ for all large n , so $(\sigma^{S_n})^4 = A_n$. □

Before proving Theorem 1.16 we need the following.

Lemma 5.5. *Let $\sigma \in S_n$ be a permutation of support s and suppose $n/2 \leq s < n$. Then there exists a conjugate σ' of σ in S_n such that $\sigma\sigma'$ is fixed-point-free.*

Proof. We use the cancellation method, using the fact that the following products have no fixed points:

$$(22) \quad [(1\ 2 \cdots k)][(n - k + 1 \cdots m - 1\ m)], \quad k \geq m/2, \quad m > 2,$$

$$(23) \quad [(1\ 2)(3\ 4 \cdots k + 2)][(k + 2\ 1)(2\ 3 \cdots k + 1)], \quad k \geq 1.$$

The only point which requires care is that after each cancellation, the number of remaining fixed points is still at least half the length of the permutation.

Step 1. Using (22), with $m - k = \min(k, f_\sigma(1))$, we may assume that if $f_\sigma(2) = 0$ or $f_\sigma(2) \geq 2$, then $f_\sigma(k) = 0$ for all $k \geq 3$.

Step 2. Using (22) with $m - k = \min(k, f_\sigma(1) - 1)$, we may assume that if $f_\sigma(2) = 1$ and $f_\sigma(1) \geq 1$, then $f_\sigma(k) = 0$ for all $k \geq 3$.

Step 3. Using (22) with $m = k$, we may assume that if $f_\sigma(1) = 0$ and $f_\sigma(2) = 1$, then σ has exactly two cycles, one of length 2 and one of length strictly greater than 2.

Step 4. Using (23), we may assume that σ belongs to $(1^i 2^j)$ with $i \leq 2j$, and $j = 1$ implies $i > 0$.

Step 5. Using (22) with $k = 2$ and $m = \min(2, i)$, we may assume $j \leq 2$

Step 6. We have reduced to the cases $(1^i 2^1)$, $1 \leq i \leq 2$; and $(1^i 2^2)$, $0 \leq i \leq 4$. The $j = 1$ cases follow from (22) and this takes care of the $j = 2$ cases with $i \geq 2$. The cases (2^2) and $(1^1 2^2)$ follow from (23) and (20) respectively. \square

We now turn to the proof of Theorem 1.16.

Proof. Write $n = ks + r$ where $0 \leq r < s$ and $k = \lceil n/s \rceil$. Let $\sigma_1, \dots, \sigma_k$ be k conjugates of σ with disjoint support, and set $\tau = \sigma_1 \dots \sigma_k$. Then τ has support ks and so $\text{fix}(\tau) = n - ks < s$.

Case 1. $s \leq n/5$.

Then $(\tau^{S_n})^4 = A_n$ by Theorem 1.14. Since $\tau \in C^k$ it follows that $C^{4k} = A_n$. Thus

$$cn(C, S_n) \leq 4k = 4\lceil n/s \rceil$$

in this case.

Case 2. $s > n/5$.

If $r \leq n/5$ argument above still works, yielding the same conclusion. So we may assume $r > n/5$ and in particular $r > 0$, so $k + 1 = \lceil n/s \rceil$.

We claim that there exists a conjugate σ_{k+1} of σ satisfying the following conditions:

- (i) The support of σ_{k+1} is disjoint from the support of $\sigma_1 \dots \sigma_{k-1}$.
- (ii) $\sigma_k \sigma_{k+1}$ have support of size $s + r$.

Indeed, this claim follows immediately from the lemma above, applied for the symmetric group of degree $s + r$ (acting on the fixed points of $\sigma_1 \dots \sigma_{k+1}$).

Note that conditions (i) and (ii) imply that $\tau = \sigma_1 \dots \sigma_k \sigma_{k+1}$ is fixed-point-free, and so $(\tau^{S_n})^4 = A_n$. This yields

$$cn(C, S_n) \leq 4(k + 1) = 4\lceil n/s \rceil,$$

completing the proof. \square

6. NORMAL SUBSETS

In this section we prove some more general results related to random walks and covering. Here we allow the generating set of the walk to change with time, and instead of conjugacy classes we deal with normal subsets (namely unions of classes). Thus we are interested in sequences W_1, \dots, W_k where each $W_i \subseteq S_n$ is a normal subset, and in the random variable $x = x_1 \cdots x_k$ where $x_i \in W_i$ is randomly chosen.

We shall study the behavior of such random variables, and then apply our results to word maps in the next section.

For simplicity of notation we shall assume our normal subsets are contained in A_n and let U be the uniform distribution on A_n .

One of the main results of this section is the following.

Theorem 6.1. *Let $W_1, \dots, W_k \subseteq A_n$ be normal subsets of A_n , where $k \geq 2$. Suppose $\alpha_1, \dots, \alpha_k$ are fixed real numbers in the interval $[0, 1]$ such that*

- (i) $|A_n|/|W_i| \leq \exp(n^{\alpha_i})$ for all $i = 1, \dots, k$.
- (ii) $\alpha_1 + \dots + \alpha_k < k - 1$.

Let $P_i = P_{W_i}$, the uniform distribution on W_i .

Then for any $\gamma < (k - 1) - \sum_{i=1}^k \alpha_i$ there exists a constant c (depending only on γ) such that

$$\|P_1 * \cdots * P_k - U\| \leq cn^{-2\gamma}.$$

In particular, if $x_i \in W_i$ are randomly chosen, then $x_1 \cdots x_k$ is almost uniformly distributed on A_n .

Before proving the theorem we make several remarks.

First, condition (i) above can be replaced by the weaker condition $|A_n|/|W_i| \leq n^{cn^{\alpha_i}}$ for all $i = 1, \dots, k$, where c is some fixed constant. This is because $n^{cn^{\alpha_i}} \leq \exp(n^{\alpha_i + \epsilon})$ for any $\epsilon > 0$ and large enough n .

Secondly, we show that this result is essentially best possible. More precisely, for any $\alpha_1, \dots, \alpha_k \geq 0$ with $\alpha_1 + \dots + \alpha_k \geq k - 1$ we construct normal subsets W_1, \dots, W_k satisfying $|A_n|/|W_i| \leq n^{cn^{\alpha_i}}$ for all $i = 1, \dots, k$ such that $\|P_1 * \cdots * P_k - U\|$ is bounded away from zero.

Indeed, choose an integer $c > 1$ and let W_i be the conjugacy class $(1^{cn^{\alpha_i}}(n - cn^{\alpha_i}))$ of S_n . We may assume $W_i \subset A_n$ (adding 1 to the length of the long cycle if needed). Let F_i be the set of fixed points of a random element of W_i . Then the expected value E of $|F_1 \cap \dots \cap F_k|$ is $np_1 \cdots p_k$ where $p_i = |F_i|/n = cn^{\alpha_i - 1}$. Therefore

$$E = c^k n^{\alpha_1 + \dots + \alpha_k - (k-1)} \geq c^k \geq 2^2 = 4.$$

It follows that, if $x_i \in W_i$ is randomly chosen, and $\sigma = x_1 \cdots x_k$, then the expected value of $\text{fix}(\sigma)$ is at least c^k . This implies that σ is not almost uniform and $\|P_1 * \cdots * P_k - U\|$ is bounded away from zero.

We now prove Theorem 6.1.

We need some preparations.

Lemma 6.2. *The probability that a randomly chosen permutation $\sigma \in S_n$ has at least f cycles of length i is at most $(f!i^f)^{-1}$.*

Proof. Let $p(i, f, n)$ denote the above probability. It is well known that the conjugacy class (i^f) in S_{fi} has size $(fi)!/f!i^f$. Now, the probability that $\sigma \in S_n$ acts in any given way on a subset X of $\{1, \dots, n\}$ of size fi is $(n-fi)!/n!$. Hence the probability that $\sigma(X) = X$ and σ induces on X a permutation of type (i^f) is $(fi)!(n-fi)!/(n!f!i^f)$. Summing up over the $\binom{n}{fi}$ subsets X of size fi we obtain

$$p(i, f, n) \leq \binom{n}{fi} (fi)!(n-fi)!/(n!f!i^f) = (f!i^f)^{-1}.$$

□

The next result sheds light on the distribution of the random variable $E(\sigma)$ occurring in Theorem 1.1.

Proposition 6.3. *Fix a real number $0 < \alpha < 1$. For every $\epsilon > 0$ there exists N such that, if $n \geq N$, then the probability that a randomly chosen permutation $\sigma \in S_n$ satisfies $E(\sigma) \leq \alpha$ is at least $1 - \exp(-n^{\alpha-\epsilon})$.*

Proof. Given α and $0 < \epsilon < \alpha$ fix an integer $c > 2/\epsilon, 1/\alpha$, and define

$$f = c^{-2}n^{\alpha-1/c}.$$

Note that for large n we have $f > n^{\alpha-\epsilon/2}$.

Let T be the set of permutations in S_n having less than f i -cycles for all $i = 1, \dots, c-1$. By the previous lemma we have

$$1 - |T|/|S_n| \leq \sum_{i=1}^{c-1} (f!i^f)^{-1} \leq (c-1)/f!.$$

Our choice of f implies $(c-1)/f! < \exp(-n^{\alpha-\epsilon})$ for large n , and so

$$|T|/|S_n| > 1 - \exp(-n^{\alpha-\epsilon}).$$

To prove the proposition it therefore remains to show that if $\sigma \in T$ then $E(\sigma) \leq \alpha$.

Indeed, let $\sigma \in T$ and let $1 \leq i < c$. Then the union of the σ -orbits of length at most i has size at most $(1+2+\dots+i)(f-1) < c^2f \leq n^{\alpha-1/c}$. Define a sequence (e_k) satisfying

$$e_1 := \alpha - 1/c, e_c := 1 - (\alpha - 1/c), e_k := 0 \text{ for } k \neq 1, c.$$

Then it is easy to see that

$$E(\sigma) \leq \sum_{k \geq 1} e_k/k = e_1 + e_c/c < \alpha - 1/c + 1/c = \alpha.$$

This completes the proof. \square

Proof of Theorem 1.13.

We apply Proposition 6.3 with $\alpha = 1/4 - \epsilon/2$ (and $\epsilon/2$ replacing ϵ) to deduce that $E(\sigma) \leq 1/4 - \epsilon/2$ with probability $\geq 1 - \exp(-n^{1/4-\epsilon})$. Now, by Corollary 1.11, $E(\sigma) \leq 1/4 - \epsilon/2$ implies $(\sigma^{S_n})^2 = A_n$ for all $n \gg 0$. The result follows. \square

Combining the above proposition with Theorem 1.1 we obtain a result of some intrinsic interest.

Proposition 6.4. *Fix $0 < \alpha < 1$. For every $\epsilon > 0$ there exists N such that if $n \geq N$ and $\sigma \in S_n$ is randomly chosen, then the probability that*

$$|\chi(\sigma)| \leq \chi(1)^\alpha \text{ for all } \chi \in \text{Irr } S_n$$

is at least $1 - \exp(-n^{\alpha-\epsilon})$.

Proof. Fix $0 < \epsilon < \alpha$. The above proposition shows that there exists N_1 such that for all $n \geq N_1$, the probability that $\sigma \in S_n$ satisfies $E(\sigma) \leq \alpha - \epsilon/2$ is at least $1 - \exp(-n^{\alpha-\epsilon})$.

By Theorem 1.1 there exists $N \geq N_1$ such that if $n \geq N$ and $E(\sigma) \leq \alpha - \epsilon/2$ then

$$|\chi(\sigma)| \leq \chi(1)^\alpha \text{ for all } \chi \in \text{Irr } S_n.$$

The result follows. \square

In the next result we study the typical behavior of $E(\sigma)$ and $|\chi(\sigma)|$ where σ is chosen from some large subset W of A_n .

Corollary 6.5. *Fix $0 < \alpha < 1$ and let $W \subseteq A_n$ be a subset satisfying $|A_n|/|W| \leq \exp(n^\alpha)$, and let $\sigma \in W$ be randomly chosen.*

(i) For any $\epsilon > 0$ there exists N depending only on ϵ such that if $n \geq N$ then the probability that $E(\sigma) \leq \alpha + \epsilon$ is at least $1 - \exp(-n^\alpha)$.

(ii) For any $\epsilon > 0$ there exists N depending only on ϵ such that if $n \geq N$ then the probability that

$$|\chi(\sigma)| \leq \chi(1)^{\alpha+\epsilon} \text{ for all } \chi \in \text{Irr}(S_n)$$

is at least $1 - \exp(-n^\alpha)$. In particular this probability tends to 1 as $n \rightarrow \infty$.

Proof. For each real number β let

$$C(\beta) := \{\sigma \in A_n : E(\sigma) \leq \beta\}.$$

By Proposition 6.4 we have

$$U(A_n \setminus C(\alpha + \epsilon)) \leq \exp(-n^{\alpha+\epsilon/2})$$

for all sufficiently large n . Therefore

$$P_W(W \setminus C(\alpha + \epsilon)) \leq \exp(-n^{\alpha+\epsilon/2}) \cdot |A_n|/|W| \leq \exp(-n^{\alpha+\epsilon/2}) \exp(n^\alpha).$$

Since $\exp(-n^{\alpha+\epsilon/2} + n^\alpha) \leq \exp(-n^\alpha)$, we see that

$$P_W(W \setminus C(\alpha + \epsilon)) \leq \exp(-n^\alpha),$$

which proves part (i). Part (ii) follows from part (i) (applied say with $\epsilon/2$) using Theorem 1.1. \square

The following is an extension of the Diaconis-Shahshahani upper bound lemma to (possibly) different conjugacy classes. The proof is very similar and so we omit it.

Lemma 6.6. *Let $C_1, \dots, C_k \subset A_n$ be conjugacy classes of S_n , and let $P_i = P_{C_i}$. Then*

$$\|P_1 * \dots * P_k - U\| \leq \sum_{\chi \in \text{Irr}(S_n), \chi(1) > 1} \frac{|\chi(C_1)^2 \dots \chi(C_k)^2|}{\chi(1)^{2k-2}}.$$

The next result can be regarded as an extension of Theorem 1.5 for the case of random walks with different conjugacy classes.

Proposition 6.7. *Let $C_1, \dots, C_k \subset A_n$ be conjugacy classes of S_n , let $\sigma_i \in C_i$, and let $\alpha_i = E(\sigma_i)$, $P_i = P_{C_i}$. Suppose $\alpha_1 + \dots + \alpha_k < k - 1 - \gamma$ for some fixed $\gamma > 0$. Then there exists constants c, N (depending only on γ) such that if $n \geq N$ we have*

$$\|P_1 * \dots * P_k - U\| \leq cn^{-2\gamma}.$$

*In particular $P_1 * \dots * P_k$ is almost uniform as $n \rightarrow \infty$.*

Proof. Since $(k - 1) - \sum_{i=1}^k \alpha_i > \gamma$ by our assumption we can choose $\epsilon > 0$ such that

$$(k - 1) - \sum_{i=1}^k \alpha_i - k\epsilon > \gamma.$$

By Theorem 1.1 there is $N > 0$ such that for all χ and $i = 1, \dots, k$ we have

$$|\chi(C_i)| \leq \chi(1)^{\alpha_i + \epsilon}.$$

Plugging this into the upper bound of the previous lemma we obtain

$$\|P_1 * \cdots * P_k - U\| \leq \sum_{\chi \in \text{Irr}(S_n), \chi(1) > 1} \frac{|\chi(1)|^{2(k\epsilon + \sum_{i=1}^k \alpha_i)}}{\chi(1)^{2k-2}} = \zeta^{S_n}(s) - 2,$$

where $s = 2((k-1) - \sum_{i=1}^k \alpha_i - k\epsilon) > 2\gamma$. Therefore, by Lemma 4.2 there is a constant c (depending on γ) such

$$\zeta^{S_n}(s) - 2 \leq \zeta^{S_n}(2\gamma) - 2 \leq cn^{-2\gamma}.$$

We conclude that for all $n \geq N$ we have

$$\|P_1 * \cdots * P_k - U\| \leq cn^{-2\gamma}.$$

This yields the first conclusion, and the second one follows too. \square

Proof of Theorem 6.1.

We adopt the notation of the theorem. We have $\gamma < (k-1) - \sum_{i=1}^k \alpha_i$. We can therefore choose $\epsilon > 0$ such that

$$\gamma + k\epsilon < (k-1) - \sum_{i=1}^k \alpha_i.$$

Now define, for $i = 1, \dots, k$, $\alpha'_i := \alpha_i + \epsilon$, and

$$T_i := \{\sigma \in W_i : E(\sigma) \leq \alpha'_i\}.$$

Then by Corollary 6.5 we have

$$|T_i|/|W_i| \geq 1 - \exp(-n^{\alpha_i}).$$

Since W_i is a normal subset, so is T_i . Therefore each T_i is a union of conjugacy classes C_{ij} , so the distribution P_{T_i} can be written as $\sum_j a_{ij} P_{C_{ij}}$, where $a_{ij} > 0$ and $\sum_j a_{ij} = 1$. This shows that

$$\|P_{T_1} * \cdots * P_{T_k} - U\| \leq \sum_{j_1, \dots, j_k} a_{1j_1} \cdots a_{kj_k} \|P_{C_{1j_1}} * \cdots * P_{C_{kj_k}} - U\|.$$

Let γ be as in Theorem 6.1. It follows from Proposition 6.7 (applied with α'_i in the role of α_i) that there is a constant c_1 such that for all $n \gg 0$ and j_1, \dots, j_k we have

$$\|P_{C_{1j_1}} * \cdots * P_{C_{kj_k}} - U\| \leq c_1 n^{-2\gamma}.$$

This easily implies

$$\|P_{T_1} * \cdots * P_{T_k} - U\| \leq c_1 n^{-2\gamma}.$$

It also follows that

$$\|P_{W_i} - P_{T_i}\| \leq c_2 \exp(-n^{\alpha_i}).$$

Using this and the fact that $\|P * Q\| = \|P\| \cdot \|Q\|$ we easily deduce that

$$\|P_{W_1} * \cdots * P_{W_k} - U\| \leq cn^{-2\gamma}$$

for some fixed $c > c_1$ and large enough n . Theorem 6.1 is proved. \square

The case $W_1 = \dots = W_k = W$ in Theorem 6.1 leads to the following mixing time theorem for random walks based on a large normal subset.

Theorem 6.8. *Let $W \subseteq A_n$ be a normal subset of S_n , and let $k \geq 2$. Suppose $|A_n|/|W| \leq \exp(n^\alpha)$ where $\alpha < 1 - 1/k$. Then for all $n \gg 0$ the mixing time $T(W)$ is at most k .*

The case $k = 2$ is of particular interest.

Corollary 6.9. *Let $W \subseteq A_n$ be a normal subset of S_n . Suppose*

$$|A_n|/|W| \leq \exp(n^{1/2-\epsilon})$$

for some fixed $\epsilon > 0$. Then for all $n \gg 0$ we have $T(W) = 2$.

The condition on $|W|$ is essentially best possible. For example take W to be the conjugacy class of an l -cycle where $l = n - 2n^{1/2}$ say. See more details in the discussion following Theorem 6.1.

We conclude this section proving Theorem 1.20 on covering properties of normal subsets. We need the following.

Proposition 6.10. *Fix $0 < \alpha < 1/4$ and let $W \subseteq A_n$ be a normal subset of S_n satisfying*

$$|A_n|/|W| \leq \exp(n^\alpha).$$

Then there exists N depending only on α such that, if $\sigma \in W$ is randomly chosen, then the probability that $(\sigma^{S_n})^2 = A_n$ is at least $1 - \exp(-n^\alpha)$.

Proof. Let $\epsilon = (1/4 - \alpha)/2 > 0$. Applying Corollary 6.5 we conclude that there exists N_1 depending on α such that, if $n \geq N_1$ and $\sigma \in W$ is randomly chosen, the probability that

$$E(\sigma) \leq \alpha + \epsilon = 1/4 - \epsilon$$

is at least $1 - \exp(-n^\alpha)$.

Now, by Corollary 1.11, there exists $N \geq N_1$ depending on ϵ such that, if $n \geq N$ then

$$E(\sigma) < 1/4 - \epsilon \text{ implies } (\sigma^{S_n})^2 = A_n.$$

The result follows. \square

We can now prove Theorem 1.20.

Proof. This follows immediately from Proposition 6.10, since for $n \gg 0$ the probability that $(\sigma^{S_n})^2 = A_n$ is positive, and $W \supseteq \sigma^{S_n}$. \square

Results 6.9 and 1.20 constitute useful tools in the next section, dealing with applications to word maps.

7. WORD MAPS

In this section we apply our results and methods to the study of word maps. Our aim is two-fold: to analyze probabilistic properties of word maps on A_n , and to provide best-possible solutions to related Waring-type problems. In particular we give three different proofs that for any non-identity word w we have $w(A_n)^2 = A_n$ for all large n .

We start by proving Theorem 1.17, showing that a random walk on A_n based on $w(A_n)$ has mixing time two.

Proof. Let $w \neq 1$ be a word. By [LaSh], if $W = w(A_n)$ then

$$|A_n|/|W| \leq n^{29/9+o(1)} \leq \exp(n^{o(1)}).$$

Applying Corollary 6.9 we immediately conclude that

$$T(w(A_n)) = 2$$

for all $n \gg 0$. \square

We also obtain the following extended and more refined version of Theorem 1.17.

Corollary 7.1. *Let w_1, w_2 be any non-identity words. Then for any $\epsilon > 0$ we have*

$$\|P_{w_1(A_n)} * P_{w_2(A_n)} - U\| = O(n^{-2+\epsilon}).$$

*In particular $P_{w_1(A_n)} * P_{w_2(A_n)}$ is almost uniform as $n \rightarrow \infty$.*

Proof. This follows from Theorem 6.1 with $k = 2, W_i = w_i(A_n)$ and $\alpha_1 = \alpha_2 = \epsilon/5$. \square

We now discuss an important result of Nica [Ni] which will be applied in the proofs below. We need some notation. A word $1 \neq w = w(x_1, \dots, x_d) \in F_d$ is called *primitive* if it is not a proper power of another word. Given a word $w \neq 1$ we can write $w = v^e$ where v is a primitive word and e is a positive integer. We refer to e as the *exponent* of w . Given w and positive integers n, i, j we let $P_{n,i,j}(w)$ denote the probability that, as $\sigma_1, \dots, \sigma_d \in S_n$ are randomly chosen, the permutation $w(\sigma_1, \dots, \sigma_d)$ has at least j cycles of length i .

We can now state Nica's theorem.

Theorem 7.2. *Let $w \neq 1$ be a word of exponent e in F_d . Fix positive integers i, j . Then we have*

$$|P_{n,i,j}(w) - P_{n,i,j}(x_1^e)| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

For example, if w is primitive then we have $P_{n,i,j}(w) = P_{n,i,j}(x_1) + o(1)$, so the limit distribution of the number of i -cycles in $w(\sigma_1, \dots, \sigma_d)$ is the same as that of the number of i -cycles in a random permutation σ_1 , which is well known to be given by a Poisson distribution.

In order to prove Theorem 1.18 we need some preparations. We first show, using Theorem 7.1 above as a main tool, that the invariant $E(\sigma)$ is very small on random word values.

Proposition 7.3. *Let $w = w(x_1, \dots, x_d)$ be a non-trivial word and fix $\epsilon > 0$. Choose $\sigma_1, \dots, \sigma_d \in S_n$ at random. Then the probability that*

$$E(w(\sigma_1, \dots, \sigma_d)) < \epsilon$$

tends to 1 as $n \rightarrow \infty$.

Proof. Since $E(\sigma) \leq B(\sigma) + o(1)$ it suffices to prove a similar result for the invariant B instead of E . Hence it suffices to show that for every $\delta > 0$ there exists N such that for all $n \geq N$, as $\sigma_1, \dots, \sigma_d \in S_n$ are chosen at random, the probability that

$$B(w(\sigma_1, \dots, \sigma_d)) < \delta$$

is at least $1 - \delta$.

Let e be the exponent of w . Given $\delta > 0$ choose a positive integer i such that $1/(i+1) < \delta$.

We claim that there is an absolute constant c such that, for any integer $f \geq 2$, the probability that σ_1^e has at least e^2if cycles of length at most i is bounded above by $c/f!$.

To show this note that the number of cycles of length at most i in σ_1^e is at most e times the number of cycles of length at most ei in σ_1 . Thus, if σ_1^e has at least e^2if cycles of length $\leq i$, then σ_1 has at least EIF cycles of length $\leq ei$. This implies that, for some $k \leq ei$, σ_1 has at least f cycles of length i . By Lemma 6.2 the latter event holds with probability at most

$$\sum_{k=1}^{ei} (f!k^f)^{-1} < c(f!)^{-1},$$

where $c = \sum_{k \geq 1} k^{-2} = \pi^2/6$. This proves the claim.

Now, choose $f \geq 2$ so that $c/f! < \delta$. It follows from the claim above that the probability that σ_1^e has at least e^2if cycles of length at most i is less than δ .

By Nica's theorem above, assuming $n \gg 0$, the same holds for the probability that $\sigma = w(\sigma_1, \dots, \sigma_d)$ has at least e^2if cycles of length at most i . This shows that with probability $> 1 - \delta$ the cycle growth sequence (b_k) of σ satisfies

$$b_k < \frac{\log(e^2if)}{\log n} \text{ for all } k \leq i.$$

It follows that, with probability $> 1 - \delta$ we have

$$B(\sigma) = \sum_{k \geq 1} \frac{b_k}{k(k+1)} < \frac{\log(e^2if)}{\log n} + \sum_{k > i} \frac{1}{k(k+1)} = \frac{\log(e^2if)}{\log n} + \frac{1}{i+1}.$$

Since $\frac{1}{i+1} < \delta$ we can choose N such that for all $n \geq N$ we have $\log e^2if / \log n + \frac{1}{i+1} < \delta$. It follows that for such n we have $B(\sigma) < \delta$ with probability $> 1 - \delta$ as required.

This completes the proof. \square

We can now show that character values of random word values are usually very small.

Theorem 7.4. *Let $w = w(x_1, \dots, x_d)$ be a non-trivial word and fix $\epsilon > 0$. Choose $\sigma_1, \dots, \sigma_d \in S_n$ at random. Then the probability that*

$$|\chi(w(\sigma_1, \dots, \sigma_d))| \leq \chi(1)^\epsilon \text{ for all } \chi \in \text{Irr}(S_n)$$

tends to 1 as $n \rightarrow \infty$.

Proof. This follows immediately combining the above result with Theorem 1.1. \square

Proof of Theorem 1.18.

We first prove part (i) of the theorem.

Fix $0 < \epsilon < 1/3$. By the theorem above, for $n \gg 0$ the probability that

$$|\chi(w(\sigma_1, \dots, \sigma_d))| \leq \chi(1)^\epsilon \text{ for all } \chi \in \text{Irr}(S_n)$$

is at least $1 - \epsilon$. The same holds for the probability that randomly chosen $\pi \in A_n$ satisfies

$$|\chi(\pi)| \leq \chi(1)^\epsilon \text{ for all } \chi \in \text{Irr}(S_n).$$

Suppose now that $w(\sigma_1, \dots, \sigma_d)$, $w(\tau_1, \dots, \tau_d)$, and π satisfy the above character inequalities. Let C and D denote the S_n -conjugacy classes of $w(\sigma_1, \dots, \sigma_d)$ and $w(\tau_1, \dots, \tau_d)$. Let $N(C, D, \pi)$ denote the number of ways to write $\pi = \sigma\tau$ where $\sigma \in C$ and $\tau \in D$. Then we have

$$N(C, D, \pi) = \frac{|C||D|}{n!} \sum_{\chi \in \text{Irr}(S_n)} \frac{\chi(C)\chi(D)\bar{\chi}(\pi)}{\chi(1)}.$$

Now,

$$\sum_{\chi \in \text{Irr}(S_n)} \frac{\chi(C)\chi(D)\bar{\chi}(\pi)}{\chi(1)} = 2 + \sum_{\chi \in \text{Irr}(S_n), \chi(1) > 1} \frac{\chi(C)\chi(D)\bar{\chi}(\pi)}{\chi(1)}.$$

By our assumptions $|\chi(C)| \leq \chi(1)^\epsilon$, and the same holds for $|\chi(D)|$ and $|\chi(\pi)|$. Therefore

$$\left| \sum_{\chi \in \text{Irr}(S_n), \chi(1) > 1} \frac{\chi(C)\chi(D)\bar{\chi}(\pi)}{\chi(1)} \right| \leq \sum_{\chi \in \text{Irr}(S_n), \chi(1) > 1} \chi(1)^{3\epsilon-1} = \zeta^{S_n}(1-3\epsilon) - 2.$$

Since $1 - 3\epsilon > 0$ we have $\zeta^{S_n}(1 - 3\epsilon) - 2 = o(1)$ by Lemma 4.2. We conclude that

$$N(C, D, \pi) = (2 + o(1)) \frac{|C||D|}{n!} = (1 + o(1)) \frac{|C||D|}{|A_n|}.$$

In particular, for any $\delta > 0$ there exists N such that if $n \geq N$ then π can be written in at least $(1 - \delta)|C||D|/|A_n|$ ways as a product of an element of C with an element of D .

Letting C, D, π vary (assuming the same bounds on their character values) it follows that, for $n \geq N$, at least $(1 - \epsilon)|A_n|$ permutations $\pi \in A_n$ can be written as

$$\pi = w(\sigma_1, \dots, \sigma_d)w(\tau_1, \dots, \tau_d)$$

in at least $(1 - \epsilon)^2(1 - \delta)|A_n|^{2d-1}$ ways. For $\pi \in A_n$ let $P(\pi)$ be the probability that $\pi = w(\sigma_1, \dots, \sigma_d)w(\tau_1, \dots, \tau_d)$ as $\sigma_i, \tau_j \in A_n$ are randomly chosen. Then P is a probability distribution on A_n , and for at least $(1 - \epsilon)|A_n|$ elements $\pi \in A_n$ we have $P(\pi) \geq (1 - \epsilon)^2(1 - \delta)|A_n|^{-1}$. Letting $\epsilon, \delta \rightarrow 0$ this easily implies

$$\|P - U\| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Part (i) of Theorem 1.18 is proved.

Parts (ii) and (iii) are elementary consequences of part (i). Indeed, a function which is almost uniformly distributed in the L_1 -norm is almost measure preserving. See for instance [GSh] for more details. \square

Theorem 1.18 has interesting consequences. For example, applying part (iii) of the theorem for the set X of generating pairs of A_n , whose size is $(1 - o(1))|A_n^2|$ (see Dixon [Di]), we obtain

Corollary 7.5. *Fix an integer $k > 0$. Then, as $n \rightarrow \infty$, almost all permutations $\pi \in A_n$ can be written as $\pi = \sigma^k \tau^k$ where σ, τ generate A_n*

Finally we prove Theorem 1.19.

Proof. Fix $0 < \epsilon < 1/4$. By Proposition 7.3, if $\sigma_1, \dots, \sigma_d \in S_n$ are randomly chosen, then

$$E(w(\sigma_1, \dots, \sigma_d)) < \epsilon$$

with probability tending to 1 as $n \rightarrow \infty$. It now follows using Corollary 1.11 that $(w(\sigma_1, \dots, \sigma_d)^{S_n})^2 = A_n$ with probability tending to 1.

Note that this implies a similar result when $\sigma_1, \dots, \sigma_d$ are chosen randomly from A_n . This proves the first assertion.

In particular it follows that, if n is large enough given w , then $(\sigma^{S_n})^2 = A_n$ for some $\sigma \in w(A_n)$. Since $w(A_n) \supseteq \sigma^{S_n}$ we conclude that $w(A_n)^2 = A_n$ for all $n \gg 0$. Theorem 1.19 is proved. \square

We conclude this paper with a third proof of the $w(A_n)^2 = A_n$ theorem, based on the extended Rudvalis-Vishne conjecture established here. The idea is to use almost semiregular embeddings of suitable p -groups in A_n , giving rise to almost homogeneous classes in $w(A_n)$. We need the following.

Proposition 7.6. *For any word $w \neq 1$ there exist positive integers $m \geq 5$ and c such that, for every positive integer n we have $w(A_n) \supseteq (1^a m^b)$ for some $a, b \geq 0$ with $a \leq c$.*

Proof. Fix a prime $p \geq 5$. Let $w \in F_d$. Since the free group F_d is residually- p , there exists a finite p -group H such that $w(H) \neq \{1\}$. Choose such a p -group of minimal order (depending on w alone). Choose an element $1 \neq h \in w(H)$.

Set $c = |H| - 1$. We may assume $n \geq c$ (otherwise the result holds trivially with $a = n, b = 0$). Write $n = q|H| + a$ where q, a are integers, $q \geq 1$ and $0 \leq a \leq c$. Let ϕ be an embedding of H into S_n with q regular orbits and a orbits of size 1. Note that $\phi(H) \subseteq A_n$ since p is odd.

Finally, let $\sigma = \phi(h)$, and let m be the order of h . Then $m \geq p \geq 5$, and $\sigma^{S_n} = (1^a m^b)$ for some $b \geq 1$. Clearly $\sigma \in \phi(w(H)) = w(\phi(H)) \subseteq w(A_n)$, and so $w(A_n) \supseteq (1^a m^b)$ as required. \square

Now, by the extended Rudvalis-Vishne conjecture, stated after Theorem 1.12, we have $(1^a m^b)^2 = A_n$ for $n \gg 0$. Combining this with Proposition 7.6 we conclude that $w(A_n)^2 = A_n$ for $n \gg 0$.

REFERENCES

- [AH] Z. Arad and M. Herzog (Eds), *Products of Conjugacy Classes in Groups*, Springer Lecture Notes **1112**, Springer-Verlag, Berlin, 1985.
- [Be] E. Bertram, Even permutations as a product of two conjugate cycles, *J. Comb. Th. Ser. A* **12** (1972), 368–380.
- [Bi] P. Biane, Representations of symmetric groups and free probability. *Adv. Math.* **138** (1998), no. 1, 126–181.
- [Bo] A. Borel, On free subgroups of semisimple groups, *Enseign. Math.* **29** (1983), 151–164.
- [Br] J.L. Brenner, Covering theorems for finite nonabelian simple groups. IX. How the square of a class with two nontrivial orbits in S_n covers A_n , *Ars Combinatorica* **4** (1977), 151–176.
- [BR] J.L. Brenner and J. Riddell, Covering theorems for finite nonabelian simple groups. VII. Asymptotics in the alternating groups, *Ars Combinatorica* **1** (1976), 77–108.
- [D1] P. Diaconis, *Group Representations in Probability and Statistics*, Institute of Mathematical Statistics Lecture Notes - Monograph Series, Vol. 11, 1988.
- [D2] P. Diaconis, Random walks on groups: characters and geometry, in *Groups St Andrews 2001 in Oxford, Vol. I*, 120–142, London Math. Soc. Lecture Note Series **304**, Cambridge Univ. Press, Cambridge, 2003.
- [DS] P. Diaconis, M. Shahshahani, Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** (1981), no. 2, 159–179.
- [Di] J. Dixon, The probability of generating the symmetric group. *Math. Z.* **110** (1969), 199–205.
- [DPSSh] J.D. Dixon, L. Pyber, Á. Seress, A. Shalev, Residual properties of free groups and probabilistic methods, *J. reine angew. Math. (Crelle's)* **556** (2003), 159–172.
- [EGH] E.W. Ellers, N. Gordeev and M. Herzog, Covering numbers for Chevalley groups, *Israel J. Math.* **111** (1999), 339–372.
- [ET] P. Erdős and P. Turán, On some problems of a statistical group theory. I, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **4** (1965), 175–186.
- [FL] S. Fomin and N. Lulov, On the number of rim hook tableaux. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **223** (1995), Teor. Predstav. Din. Sistemy, Kombin. i Algoritm. Metody. I, 219–226, 340; translation in *J. Math. Sci. (New York)* **87** (1997), no. 6, 4118–4123.
- [GSh] S. Garion and A. Shalev, Commutator maps, measure preservation, and T -systems, to appear in *Trans. Amer. Math. Soc.*
- [G] W.T. Gowers, Quasirandom groups, Preprint, 2007, to appear.
- [HLP] G.H. Hardy, J.E. Littlewood, and G. Pólya, *Inequalities*. Second Edition. Cambridge, at the University Press, 1952.
- [Hu] D. Husemoller, Ramified coverings of Riemann surfaces. *Duke Math. J.* **29** (1962), 167–174.
- [Ja] G.D. James, *The representation theory of the symmetric groups*, Lecture Notes in Math. **682**, Springer-Verlag, Berlin, 1978.
- [L] M. Larsen, Word maps have large image, *Israel J. Math.* **139** (2004), 149–156.
- [L2] M. Larsen, How often is a partition an n 'th power?, arXiv: math.CO/9712223.
- [LaSh] M. Larsen and A. Shalev, Word maps and Waring type problems, Preprint, 2007.

- [LL] R. Lawther and M.W. Liebeck, On the diameter of a Cayley graph of a simple group of Lie type based on a conjugacy class, *J. Comb. Theory, Ser. A* **83** (1998), 118–137.
- [LiSh1] M.W. Liebeck and A. Shalev, Diameter of simple groups: sharp bounds and applications, *Annals of Math.* **154** (2001), 383–406.
- [LiSh2] M.W. Liebeck and A. Shalev, Fuchsian groups, coverings of Riemann surfaces, subgroup growth, random quotients and random walks, *J. Algebra* **276** (2004), 552–601.
- [Lish3] M.W. Liebeck and A. Shalev, Fuchsian groups, finite simple groups, and representation varieties, *Invent. Math.* **159** (2005), 317–367.
- [Lub] A. Lubotzky, Cayley graphs: eigenvalues, expanders and random walks, in *Surveys in combinatorics, Stirling 1995*, London Math. Soc. Lecture Note Series **218**, pp. 155–189, Cambridge Univ. Press, 1995.
- [Lu] N. Lulov, Random walks on the symmetric groups, Ph. D. Thesis, Harvard University, 1996.
- [LP] N. Lulov and I. Pak, Rapidly mixing random walks and bounds on characters of the symmetric groups, *J. Algebraic Combin.* **16** (2002), 151–163.
- [MS] T.W. Müller and J-C. Schlage-Puchta, Character theory of symmetric groups, subgroup growth of Fuchsian groups, and random walks, *Adv. in Math.* **213** (2007), 919–982.
- [Na] M.B. Nathanson, *Additive Number Theory: the classical bases*, Graduate Texts in Mathematics **164**, Springer, 1996.
- [Ni] A. Nica, On the number of cycles of given length of a free word in several random permutations, *Random Structures Algorithms* **5** (1994), no. 5, 703–730.
- [NP] N. Nikolov and L. Pyber, Product decompositions of quasirandom groups and a Jordan type theorem, Preprint, 2007.
- [NS1] N. Nikolov and D. Segal, On finitely generated profinite groups. I. strong completeness and uniform bounds, *Annals of Math.* **165** (2007), 171–238.
- [NS2] N. Nikolov and D. Segal, On finitely generated profinite groups. II. Products in quasisimple groups, *Annals of Math.* **165** (2007), 239–273.
- [RS] A. Rattan and P. Sniady, Upper bound on the characters of the symmetric groups for balanced Young diagrams and a generalized Frobenius formula, Preprint, arXiv: math.RT/0610540.
- [Ro] Y. Roichman, Upper bound on the characters of the symmetric groups, *Invent. Math.* **125** (1996), no. 3, 451–485.
- [Sh] A. Shalev, Word maps, conjugacy classes, and a non-commutative Waring-type theorem, to appear in *Annals of Math.*
- [V] U. Vishne, Mixing and covering in the symmetric groups, *J. Algebra* **205** (1998), 119–140.

E-mail address: larsen@math.indiana.edu

DEPARTMENT OF MATHEMATICS, INDIANA UNIVERSITY, BLOOMINGTON, IN 47405, U.S.A.

E-mail address: shalev@math.huji.ac.il

EINSTEIN INSTITUTE OF MATHEMATICS, HEBREW UNIVERSITY, GIVAT RAM, JERUSALEM 91904, ISRAEL