On the diameter of permutation groups

By Harald A. Helfgott and Ákos Seress

Abstract

Given a finite group $G$ and a set $A$ of generators, the diameter $\text{diam}(\Gamma(G, A))$ of the Cayley graph $\Gamma(G, A)$ is the smallest $\ell$ such that every element of $G$ can be expressed as a word of length at most $\ell$ in $A \cup A^{-1}$. We are concerned with bounding $\text{diam}(G) := \max_A \text{diam}(\Gamma(G, A))$.

It has long been conjectured that the diameter of the symmetric group of degree $n$ is polynomially bounded in $n$, but the best previously known upper bound was exponential in $\sqrt{n \log n}$. We give a quasipolynomial upper bound, namely,

$$\text{diam}(G) = \exp\left( O((\log n)^3 \log \log n) \right) = \exp\left( (\log \log |G|)^{O(1)} \right)$$

for $G = \text{Sym}(n)$ or $G = \text{Alt}(n)$, where the implied constants are absolute. This addresses a key open case of Babai’s conjecture on diameters of simple groups. By a result of Babai and Seress (1992), our bound also implies a quasipolynomial upper bound on the diameter of all transitive permutation groups of degree $n$.

1. Introduction

1.1. Groups and their diameters. Let $A$ be a set of generators for a group $G$. The (undirected) Cayley graph $\Gamma(G, A)$ is the graph whose set of vertices is $V = G$ and whose set of edges is $E = \{ \{g, ga\} : g \in G, a \in A \}$. The diameter $\text{diam}(\Gamma)$ of a graph $\Gamma(V, E)$ is defined by

$$\text{diam}(\Gamma) = \max_{v_1, v_2 \in V} \min_{P \text{ a path from } v_1 \text{ to } v_2} \text{length}(P).$$

In particular, the diameter of a Cayley graph $\Gamma(G, A)$ is the maximum, for $g \in G$, of the length $\ell$ of the shortest expression $g = a_1^{\varepsilon_1} a_2^{\varepsilon_2} \cdots a_\ell^{\varepsilon_\ell}$ with $a_i \in A$ and $\varepsilon_i \in \{-1, 1\}$ for each $i = 1, \ldots, \ell$. We may define the diameter $\text{diam}(G)$ of a finite group to be the maximal diameter of the Cayley graphs $\Gamma(G, A)$ for all generating sets $A$ of $G$.

Much recent work on group diameters has been motivated by the following conjecture:
Conjecture 1 (Babai, published as [BS92, Conj. 1.7]). For all finite simple groups $G$,
\[ \text{diam}(G) \leq (\log |G|)^{O(1)}, \]
where the implied constant is absolute.

Here and henceforth, $|S|$ denotes the number of elements of a set $S$.

The first class of finite simple groups for which Conjecture 1 was established was $\text{PSL}_2(\mathbb{Z}/p\mathbb{Z})$ with $p$ prime, by Helfgott [Hel08]. The paper [Hel08] initiated a period of intense activity [BG08a], [BG08b], [Din11], [BGS10], [Hel11], [GH], [Var12], [BGS11], [PS10], [BGT11], [GH10], [GV12] on the diameter problem and the related problem of expansion properties of Cayley graphs.

As far as work in this vein on the diameter of finite simple groups is concerned, the best results to date are those of Pyber, Szabó [PS10] and Breuillard, Green, Tao [BGT11]. Their wide-ranging generalisation covers all simple groups of Lie type, but (just like [GH]) the diameter estimates retain a strong dependence on the rank; thus, they prove Conjecture 1 only for groups of bounded rank. The problem for the alternating groups remained wide open.

These two issues are arguably related: product theorems (of the type $|A \cdot A \cdot A| \gg |A|^{1+\delta}$ familiar since [Hel08]) are false both in the unbounded-rank case and in the case of alternating groups, and the counterexamples described in both situations in [PPSS12], [PS10] are based on similar principles.

In the present paper we address the case of alternating (and symmetric) groups. We expect that some of the combinatorial difficulties we overcome will also arise in the context of linear groups of large rank.

For $G = \text{Alt}(n)$, Conjecture 1 stipulates that $\text{diam}(\text{Alt}(n)) = n^{O(1)}$; [BS92] refers to this special case of Conjecture 1 as a “folklore” conjecture. Indeed, this has long been a problem of interest in computer science (see [KMS84], [McK84], [BHK+90], [BBS04], [BH05]). On a more playful level, bounds on the diameter of permutation groups are relevant to every permutation puzzle (e.g., Rubik’s cube).

The best previously known upper bound on $\text{diam}(G)$ for $G = \text{Alt}(n)$ or $G = \text{Sym}(n)$ was more than two decades old:

\[ \text{diam}(G) \leq \exp((1 + o(1))\sqrt{n \log n}) = \exp((1 + o(1))\sqrt{\log |G|}), \]
due to Babai and Seress [BS88]. (We write $\exp(x)$ for $e^x$.)

---

1This list is not meant to be exhaustive.

2See, e.g., I. Pak’s remarks (made already before [PS10], [BGT11]) on the relative difficulty of the work remaining to do in the linear case (to be finished “in the next 10 years”) and of the problem on $\text{Alt}(n)$, for which there was “much less hope” [Pak].
1.2. Statement of results. Recall that a function $f(n)$ is called quasipolynomial if $\log(f(n))$ is a polynomial function of $\log n$. Our main result establishes a quasipolynomial upper bound for $\text{diam}(\text{Alt}(n))$ and $\text{diam}(\text{Sym}(n))$.

**Main Theorem.** Let $G = \text{Sym}(n)$ or $\text{Alt}(n)$. Then

$$\text{diam}(G) \leq \exp\left( O((\log n)^4 \log \log n) \right),$$

where the implied constant is absolute.

The quasipolynomial bound extends to a much broader class of permutation groups. Recall that a permutation group $G$ acting on a set $\Omega$ is called transitive if

$$\forall \alpha, \beta \in \Omega, \, \exists g \in G \text{ such that } g \text{ takes } \alpha \text{ to } \beta.$$  

The size $|\Omega|$ of the permutation domain is called the degree of $G$.

Kornhauser et al. [KMS84] and McKenzie [McK84] raised the question of what classes of permutation groups may have polynomial diameter bound in their degree. A weaker, quasipolynomial bound for all transitive groups was formally conjectured in [BS92]:

**Conjecture 2 ([BS92, Conj. 1.6]).** If $G$ is a transitive permutation group of degree $n$, then $\text{diam}(G) \leq \exp((\log n)^{O(1)})$.

Babai and Seress [BS92] linked Conjecture 2 to the diameter of alternating groups:

**Theorem 1.1 ([BS92, Thm. 1.4]).** If $G$ is a transitive permutation group of degree $n$, then

$$\text{diam}(G) \leq \exp\left( O((\log n)^3) \text{diam}(\text{Alt}(k)) \right),$$

where $\text{Alt}(k)$ is the largest alternating composition factor of $G$.

Combining our Main Theorem with Theorem 1.1, we immediately obtain

**Corollary 1.2.** Conjecture 2 is true; indeed the diameter of any transitive permutation group $G$ of degree $n$ is

$$\text{diam}(G) \leq \exp\left( O((\log n)^4 \log \log n) \right).$$

We note that Theorem 1.1 is not only used to prove Corollary 1.2 — it also comes into play as an inductive tool in the proof of the Main Theorem (see Lemma 6.3). Since Theorem 1.1 relies on the Classification of Finite Simple Groups, so does the Main Theorem.

It is well known that, for any finite group $G$ and any set $A$ of generators of $G$, the eigenvalues $\lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \cdots$ of the adjacency matrix of $\Gamma(G, A)$ satisfy

$$\lambda_0 - \lambda_1 \geq \frac{1}{\text{diam}(\Gamma(G, A))^2}.$$  

(1.3)
(See [DSC93, Cor. 1] or the references [Ald87], [Bab91], [Gan91], [Moh91] therein.) Because of (1.3), we obtain immediately that
\[ \lambda_0 - \lambda_1 \geq \exp(-O((\log n)^d \log \log n)), \]
with consequences on expansion and the mixing rate. (See, for example, [Lov96], [HLW06].)

Finally, the Main Theorem and Corollary 1.2 extend to directed graphs. Given \( G = \langle A \rangle \), the directed Cayley graph \( \vec{\Gamma}(G, A) \) is the graph with vertex set \( G \) and edge set \( \{(g, ga) : g \in G, a \in A\} \). The diameter of \( \vec{\Gamma}(G, A) \) is defined by (1.1), where “path” should be read as “directed path”; \( \text{diam}(G) \) is the maximum of \( \text{diam}(\vec{\Gamma}(G, A)) \) taken as \( A \) varies over all generating sets \( A \) of \( G \). Thanks to Babai’s bound \( \text{diam}(G) = O(\text{diam}(G) \cdot (\log |G|)^2) \) [Bab06, Cor. 2.3], valid for all groups \( G \), we obtain immediately from Corollary 1.2 that

**Corollary 1.3.** Let \( G \) be a transitive group on \( n \) elements. Then
\[ \text{diam}(G) \leq \exp(O((\log n)^d \log \log n)). \]

1.3. **General approach.** An analogy underlies recent work on growth in groups: much of basic group theory carries over when, instead of subgroups, we study sets that grow slowly (\( |A \cdot A \cdot A| \leq |A|^{1+\varepsilon} \)). This realisation is clearer in [Hel11] than in [Hel08] and has become current since then. (The term “approximate group” [Tao08] actually first arose in a different context, namely, the generalisation of some arguments in classical additive combinatorics to the non-abelian case; see also [Hel08, §2.3], [SSV05, Lemma 4.2]. The analogy between subgroups and slowly growing sets was also explored in a model-theoretic setting in later work by Hrushovski [Hru12].)

This analogy is more important than whether one works with approximate subgroups in Helfgott’s sense (\( |A \cdot A \cdot A| \leq |A|^{1+\varepsilon} \), or more generally \( |A \cdot A \cdot A| \leq f(|A|) \) for some specified \( f \)) or Tao’s sense [Tao08, Def. 3.7]; the two definitions are essentially equivalent, and we will actually work with neither. We could phrase part of our argument in terms of statements of the form \( |A^k| \leq |A|^{1+\varepsilon} \), but \( k \) would sometimes be larger than \( n \); applying the tripling lemma ([RT85], [Hel08, Lemma 2.2], [Tao08, Lemma 3.4]) to such statements would weaken them fatally.

There is another issue worth emphasising: the study of growth needs to be relative. We should not think simply in terms of a group acting on itself

\[ \text{diam}(G) \leq \exp(O((\log n)^d \log \log n)). \]
by multiplication — even if, in the last analysis, this is the only operation available to us. Rather, growth statements often need to be thought of in terms of the action of a group $G$ on a set $X$ and the effect of this action on subsets $A \subseteq G, B \subseteq X$. (Here $X$ may or may not be endowed with a structure of its own.) This was already clear in [Hel11, Prop. 3.1] and [GH10] and is crucial here: a key step will involve the action of a normaliser $N_G(H)$ on a subgroup $H \leq G$ by conjugation.

1.4. Relation to previous work. Our debt to previous work on permutation groups is manifold. It is worthwhile to point out that some of our main techniques are adaptations to sets of classification-free arguments on the properties of subgroups of $\text{Sym}(n)$ by Babai [Bab82], Pyber [Pyb93], Bochert [Boc89], and Liebeck [Lie83]. Of particular importance is Babai and Pyber’s work on the order of 2-transitive groups [Bab82], [Pyb93].

We shall also utilise existing diameter bounds. Besides Theorem 1.1, we shall use the main idea from [BS88] (see Lemma 3.19) and the following theorem by Babai, Beals, and Seress. For a permutation $g$ of a set $\Omega$, the support $\text{supp}(g)$ is the subset of elements of $\Omega$ that are displaced by $g$.

**Theorem 1.4.** ([BBS04]) For every $\varepsilon < 1/3$, there exists $K(\varepsilon)$ such that, if $G = \text{Alt}(n)$ or $\text{Sym}(n)$ and $A$ is a set of generators of $G$ containing an element $x \in A$ with $1 < |\text{supp}(x)| \leq \varepsilon n$, then

$$\text{diam}(\Gamma(G, A)) \leq K(\varepsilon)n^8.$$ 

We will use this theorem repeatedly in Section 6. As we shall make clear in Section 4, we also apply — crucially — one of the main methods involved in the proof of Theorem 1.4, namely, the use of short random walks to mimic a uniform distribution.

We note that until recently Theorem 1.4 gave the largest known explicit class of Cayley graphs of $\text{Sym}(n)$ or $\text{Alt}(n)$ that has polynomially bounded diameter. In late 2010, partly based on ideas from [BBS04], Bamberg et al. [BGH+12] proved that if a set of generators of $\text{Sym}(n)$ or $\text{Alt}(n)$ contains an element of support size at most $0.63n$, then the diameter of the Cayley graph is bounded by a polynomial of $n$.

1.5. Outline. Let us begin in medias res, focusing on a crucial moment at which growth is achieved. Classical reasons aside, this will allow us to emphasise the link to [Hel08], [Hel11], [BGT11], [PS10], and [GH10], while

---

4Cf. the role of [LP11] (especially Theorems 4.2 and 6.2) which, in order to provide alternatives to the Classification of Finite Simple Groups, did (both more and less generally) for subgroups what [Hel11, §5] did for sets and was later translated back to sets for use in [BGT11].
repeating one of the main motifs: growth results from the action of a group on a set, often, as is the case here, by conjugation.

The setup for the crucial step will involve a set $A \subset \text{Sym}([n])$ with $A = A^{-1}$ and a fairly large set $\Sigma \subset [n]$ ($[n] := \{1, 2, \ldots, n\}$) such that the pointwise stabiliser\footnote{Defined as in (2.1). The notation here follows Dixon and Mortimer [DM96] and Seress [Ser03] rather than Wielandt [Wie64]. Wielandt writes $A_\Sigma$ for the pointwise stabiliser, which we denote by $A_{(\Sigma)}$; we write $A_\Sigma$ for the setwise stabiliser.} $A_{(\Sigma)}$ generates a group $\langle A_{(\Sigma)} \rangle$ with a large orbit $\Gamma \subset [n] \setminus \Sigma$. (Say, for concreteness, that $|\Sigma| \geq (\log n)^2$ and $|\Gamma| > 0.95n$.) The setwise stabiliser $\langle A_{(\Sigma)} \rangle$ acts on the pointwise stabiliser $\langle A_{(\Sigma)} \rangle$ by conjugation.

We can assume that $\langle A_{(\Sigma)} \rangle$ acts as the alternating or symmetric group on $\Gamma$, as otherwise we are done by a different argument (called descent in Section 6; we will discuss it later). It follows that we can find a set $S$ of at most six elements of $\langle A_{(\Sigma)} \rangle^\ell$, $\ell$ fairly small, such that $\langle S \rangle$ is doubly transitive on $\Gamma$. (This implication is far from trivial; we prove a general result of this kind (Corollary 4.7) showing that, if a set $A'$ generates $\text{Sym}([m])$ or $\text{Alt}([m])$, then there is a small set $S \subset (A')^\ell$, $\ell$ fairly small, such that $\langle S \rangle$ is $k$-transitive.)

Consider the action of the elements of $A_\Sigma$ on the elements of $S$ by conjugation. By an orbit-stabiliser principle, either (a) an element $g \neq e$ of $A_\Sigma$ fixes (i.e., commutes with) every element of $S$, or (b) the orbit $\{gsg^{-1} : g \in A_\Sigma\}$ of some $s \in S$ is of size $|A_\Sigma|^{1/6}$. In case (a), since $\langle S \rangle$ is doubly transitive, $g$ fixes every point of $\Gamma$. We have thus constructed a nonidentity element $g \in A$ with small support, and we are done by Theorem 1.4. In case (b), we have constructed many ($\geq |A_\Sigma|^{1/6}$) distinct elements $gsg^{-1}$ in the pointwise stabiliser $\langle A_\Sigma \rangle$. This is what we call creation in Section 6.

The questions are now — how do we get to the point at which we began our narrative? And how do we use the conclusion we have just shown, namely, the creation of many elements in the pointwise stabiliser?

Let us start with the first question. For the conclusion to be strong, $A_{(\Sigma)}$ should be large — for instance, large in comparison to $A_{(\Sigma)}$ or $\langle A_{(\Sigma)} \rangle$. Now, $A_{(\Sigma)}$ can be much larger than $\langle A_{(\Sigma)} \rangle$ only if $A$ occupies a large number $R$ of cosets of $\text{Sym}([n])_{(\Sigma)}$ in $\text{Sym}([n])$. (By pigeonhole, $|\langle A_{(\Sigma)} \rangle| \geq |A|/R$.) Our aim will be to find a large $\Sigma$ such that $R$ is larger than $(dn)^{\lfloor 2 \rfloor}$, where $d > 1/2$ is a constant.

This is also an intermediate aim in [Pyb93] (which treats subgroups, not sets). Much as there, we use this as follows: $R$ is larger than $(dn)^{\lfloor 2 \rfloor}$, and so $AA^{-1}$ intersects at least $d^{\lfloor 2 \rfloor} |\Sigma|!$ cosets of $(\text{Sym}([n]))_{(\Sigma)}$ within $(\text{Sym}([n]))_{(\Sigma)}$ (by pigeonhole); this means that the projection (by restriction) of $AA^{-1}$ to $\text{Sym}(\Sigma)$ has size at least $d^{\lfloor 2 \rfloor} |\Sigma|!$. At this point Pyber uses the fact (due to Liebeck [Lie83] and based on Bochert [Boc89]) that, if a subgroup $H$ of
Sym(Σ) is of size at least $s = d^{\Sigma|\Sigma|!}$, where $d > 1/2$, then there must be a large orbit $\Delta \subset \Sigma$ of $H$ such that the restriction of $H$ to $\Delta$ equals Alt($\Delta$) or Sym($\Delta$). We will show (Proposition 3.15) that, even if $H \subset \text{Sym}(\Sigma)$ is just a set, not a subgroup, the assumption that $H$ is of size at least $s$ implies that the restriction of $H^\ell$ to $\Delta$ equals all of Alt($\Delta$) or Sym($\Delta$), where $\ell$ is relatively small. (This works because the proof of Bochert’s nineteenth-century result is algorithmic.) The fact that we obtain all of Alt($\Delta$) or Sym($\Delta$) is particularly important for what we called a “descent argument” (as in “infinite descent”) algorithmic. The key here is that the outcome of a random walk of moderate length takes a pair $(x, y)$ to any other pair $(x', y')$ with almost uniform probability.

Now, as we said, we must find a large $\Sigma$ such that $A$ (or $A^\ell$, $\ell$ moderate) occupies a large number of cosets of Sym([n]), i.e., sends (Σ) to many different tuples. Pyber shows this (for $A$ a subgroup) by constructing $\Sigma = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ so that

$$|\alpha_i A^{\langle \alpha_1, \ldots, \alpha_{i-1}\rangle}| \geq dn$$

for every $1 \leq i \leq m$. (The use of stabiliser chains $A > A^{\langle \alpha_1\rangle} > A^{\langle \alpha_1, \alpha_2\rangle} > \cdots$ goes back to the algorithmic work of Sims [Sim70], [Sim71], as does the use of the size of the orbits in (1.4); see [Ser03, §4.1].) This step also works when $A$ is a subset (Lemma 3.17). The difficult part, of course, is to show that elements $\alpha_1, \alpha_2, \ldots, \alpha_m$ satisfying (1.4) exist.

Here [Pyb93] uses Babai’s splitting lemma [Bab82], which states that, if $H < \text{Sym}(\text{[n]})$ is a doubly transitive permutation group and $\Sigma \subset \text{[n]}$ is such that $H^{\langle \Sigma\rangle}$ has no orbits of size $>(1-\epsilon)n$, then there is a set $\Sigma' \subset \text{[n]}$ with $|\Sigma'| \ll (\log n)|\Sigma|$ such that $H^{\langle \Sigma'\rangle}$ consists only of the identity. In fact, $\Sigma' = \Sigma^S = \{x^s : x \in \Sigma, s \in S\}$, where $S$ is a subset of $H$ of size $|S| \ll \log n$. Babai constructs $S$ by choosing $O(\log n)$ elements randomly from $H$ with the uniform distribution. A random element of $H$ takes a pair $(x, y)$ of distinct elements of $\text{[n]}$ to any other such pair $(x', y')$ with the same probability $\left((n(n-1)/2)^{-1}\right)$ no matter what $(x', y')$ is. Now, given any distinct $x, y \in \text{[n]}$, it is almost certain that they will be taken to elements $x^g, y^g$ of different orbits of $H^{\langle \Sigma\rangle}$ by some $g \in S \subset H$, simply because a positive proportion of all pairs $(x', y')$ lie in different orbits (by the fact that there is no orbit of size $>(1-\epsilon)n$). Then, $x$ and $y$ belong to different orbits of $gH_Sg^{-1} = H_{\Sigma^S_{g^{-1}}}$ and thus to different orbits of $H_{\Sigma^S}$. Summing probabilities over all $x$ and $y$, we obtain that, with positive probability, every two distinct $x, y \in \text{[n]}$ belong to different orbits of $H_{\Sigma^S}$. This implies that $H_{\Sigma^S}$ is trivial.

We adapt this entire argument so as to hold for a set $A \subset \text{Sym}(\text{[n]})$ instead of a subgroup $H < \text{Sym}(\text{[n]})$; as usual, sometimes $H$ is replaced by $A$ and sometimes by $AA^{-1}$ or $A^\ell$, where $\ell$ is moderate ($\ell \ll n^{O(1)}$). The key here is that the outcome of a random walk of moderate length takes a pair $(x, y)$ to any other pair $(x', y')$ with almost uniform probability.
We apply the resulting generalisation of the splitting lemma (Proposition 5.2) and point out that \((\AA^{-1})_{\Sigma'} = \{e\}\) implies \(|\Sigma'| \gg \log_n |\AA|\) (by pigeonhole) and so \(|\Sigma| \gg (\log |\AA|)/(\log n)^2\). In other words, we are guaranteed to be able to construct a stabiliser chain with long orbits as in (1.4) (for any \(d < 1\)) until \(m\) gets to size proportional to \((\log |\AA|)/(\log n)^2\). We call this the organising step.

Now that we have the stabiliser chain, and thus the proper setup for the creation step, how do we use the outcome of the creation step? In [Hel08] and the work that followed, the main intermediate result stated that a generating set \(\AA\) always grew in size \((|\AA^3| \geq |\AA|^{1+\delta}\) [Hel08, Key Proposition]); to prove that the diameter \(\Gamma(G, \AA)\) was small, one just had to apply this key proposition over and over \((|\AA^3| \geq |\AA|^{1+\delta}, |\AA^9| \geq |\AA^3|^{1+\delta} \geq |\AA|^{(1+\delta)^2}, \ldots)\). Here we will also prove our diameter bound by iteration; however, the quantity whose growth we will keep track of during iteration will not be the size of \(\AA^\ell\), but rather the length of the sequence \(\alpha_1, \alpha_2, \ldots\) we have constructed satisfying (1.4) (for \(\AA^\ell\) instead of \(\AA\)).

The iteration is conducted as follows. We actually construct the first \((\log n)^2\) elements of \(\alpha_1, \alpha_2, \ldots\) by brute force, by raising \(\AA\) to an \(n^{O((\log n)^2)}\)-th power. (This works by Lemma 3.9.) Now we get to the main step that gets repeated (Proposition 6.4): given a sequence \(\alpha_1, \ldots, \alpha_m\) satisfying (1.4) (for \(\AA^\ell\) instead of \(\AA\)), we use the creation step to construct at least \((m!)^{1/6}\) elements of \((\AA^\ell')_{\alpha_1, \ldots, \alpha_m}\), where \(\ell' \leq n^{O(\log n)\ell}\); then we use the organising step to construct new elements \(\alpha_{m+1}, \ldots, \alpha_{m'}\) \((m' \geq m + cm \log m)/(\log n)^2\) so that (1.4) is satisfied for all \(i = 1, 2, \ldots, m'\) (with \(\AA^{\ell'}\) instead of \(\AA^\ell\)). (We actually repeat the organising step several times after each creation step; this helps us save a log in the final exponent.) Repeating this, we keep on lengthening the sequence \(\alpha_1, \alpha_2, \ldots\) until it gets to be of length almost \(n\), and then we are done easily.

Needless to say, in the above outline, we have left out details that will be treated in full in the body of the text. Let us discuss one more thing now — namely, what we have called the descent step. We reach it when we have constructed a set \(\Sigma = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}\) such that (a) the restriction of \(A_{\Sigma}\) to \(\Sigma\) acts as \(\text{Alt}(\Delta)\) or \(\text{Sym}(\Delta)\) on a large subset \(\Delta \subset \Sigma\), (b) the restriction of \(\langle A_{\Sigma} \rangle\) to \([n] \setminus \Sigma\) does not act like \(\text{Alt}\) or \(\text{Sym}\) on any subset of \([n] \setminus \Sigma\) larger than 0.95n (say).

Now we can use Theorem 1.1 (Babai-Seress) and obtain from (b) that the diameter of \(\langle A_{\Sigma} \rangle\) is bounded in terms of the diameter of \(\text{Alt}(k), k = [0.95n]\). (It is here, and only here, that the Classification Theorem is needed, since Theorem 1.1 is based on it.) Now we can use, inductively, our own main theorem on the diameter of \(\text{Alt}(n)\), with \(k\) instead of \(n\). This gives a bound on the diameter of \(\langle A_{\Sigma} \rangle\). At this point we use Lemma 3.19 (which is [BS87,
Lemma 3]; see also [BLS87]). This shows that (a) implies that \( \langle A_\Sigma \rangle \) contains a nonidentity element \( g \) of small support. We can now apply Theorem 1.4 (Babai-Beals-Seress) to bound the diameter of our group \( G = \text{Alt}(n) \) or \( G = \text{Sym}(n) \) with respect to \( A \). Note that [BS87, Lemma 3] would be prohibitively expensive if used as a constructive result; here we are using it to show the existence of an element, which we know can be constructed as a relatively short word thanks to the bound on the diameter of \( \langle A_\Sigma \rangle \) we obtained through Theorem 1.1.

1.6. Acknowledgements. We are deeply grateful to both Pablo Spiga and Nick Gill for stimulating discussions and for their constant help. Gordon Royle organised the first author’s visit to Australia; if it were not for him, our collaboration might not have happened. Thanks are also due to László Babai, Martin Kassabov, Igor Pak, Peter Sarnak, and Andrzej Żuk for their advice. Detailed comments by two anonymous referees have certainly helped improve the paper.

Ákos Seress was supported by the NSF and by ARC Grant DP1096525. Travel was supported in part by H. A. Helfgott’s Philip Leverhulme prize. We benefited from the kind hospitality of the University of Western Australia and the École Normale Supérieure during our visits to each other’s institutions.

2. Notation

We write \([n] = \{1, 2, \ldots, n\}\). For a set \( \Omega \), \( \text{Sym}(\Omega) \) and \( \text{Alt}(\Omega) \) are the symmetric and alternating groups acting on \( \Omega \). As is customary, we often write \( \text{Alt}(n) \) and \( \text{Sym}(n) \) for \( \text{Alt}([n]) \) and \( \text{Sym}([n]) \) — particularly when we are thinking of these groups as abstract groups as opposed to their actions.

We write \( H \leq G \) to mean that \( H \) is a subgroup of \( G \) and \( H \triangleleft G \) to mean that \( H \) is a normal subgroup. We say that a group \( S \) is a section of a group \( G \) if there exist subgroups \( H \) and \( K \) of \( G \) with \( K \triangleleft H \) and \( H/K \cong S \). We denote the identity element of a group by \( e \).

Let \( A \) be a subset of a group \( G \). We write \( A^{-1} = \{a^{-1} : a \in A\} \), \( A^k = \{a_1a_2 \cdots a_k : a_1, \ldots, a_k \in A\} \). In [Hel08], [Hel11], the first author wrote \( A_\ell \) to mean \( (A \cup A^{-1} \cup \{e\})^\ell \); this does not seem to have become standard and would also not do here due to the potential confusion with alternating groups. (Recall that \( A_n \) is in common usage as a synonym for \( \text{Alt}(n) \).) We will often include \( A = A^{-1}, e \in A \) explicitly in our assumptions so as to simplify notation. A set \( A \) with \( A = A^{-1} \) is said to be symmetric.

We write \( |A| \) for the number of elements of a set \( A \). (All of our sets and groups are finite.) Given a group \( G \) and a subgroup \( H \leq G \), we write \( [G : H] \) for the index of \( H \) in \( G \).

Let a group \( G \) act on a set \( X \). As is customary in the study of permutation groups, given \( g \in G \) and \( \alpha \in X \), we write \( \alpha^g \) for the image of \( \alpha \) under the
action of $g$. We speak of the orbit $\alpha^A = \{\alpha^g : g \in A\}$ of a point $\alpha$ under the action of a set $A$ of permutations. Our actions are right actions by default: $(\alpha^g)^h = \alpha^{gh}$. In consequence, we also use right cosets by default, i.e., cosets $Hg$ (and so $G/H$ is the set of all such cosets). Clearly $|G/H| = [G:H]$.

We define the commutator $[g,h]$ by $[g,h] = g^{-1}h^{-1}gh$. Again, this choice is customary for permutation groups.

Define

$$A\Sigma = \{g \in A : \Sigma^g = \Sigma\} \quad \text{the setwise stabiliser},$$

$$A_{\langle \Sigma \rangle} = \{g \in A : \forall \alpha \in \Sigma (\alpha^g = \alpha)\} \quad \text{the pointwise stabiliser}.$$  \hspace{1cm} (2.1)

If $\Sigma = \{g_1, \ldots, g_m\}$, the setwise stabiliser is denoted by $A_{\{g_1, \ldots, g_m\}}$ and the pointwise stabiliser by $A_{\langle g_1, \ldots, g_m \rangle}$.

Given a permutation $g \in \text{Sym}(\Omega)$, we define its support $\text{supp}(g)$ to be the set of elements of $\Omega$ moved by $g$: $\text{supp}(g) = \{\alpha \in \Omega : \alpha^g \neq \alpha\}$. If a subset $\Delta \subseteq \Omega$ is invariant under $g$, i.e., $\Delta$ is a union of cycles of $g$, then we define $g|_\Delta \in \text{Sym}(\Delta)$ as the restriction (natural projection) of $g$ to $\Delta$: the permutation $g|_\Delta$ acts on $\Delta$ as $g$ does. If $\Delta$ is invariant under some $D \subseteq \text{Sym}(\Omega)$, then $D|_\Delta = \{g|_\Delta : g \in D\}$.

A partition $\mathcal{B} = \{\Omega_1, \Omega_2, \ldots, \Omega_k\}$ of a set $\Omega$ ($\Omega_i$ nonempty) is called a system of imprimitivity for a transitive group $G \leq \text{Sym}(\Omega)$ if $G$ permutes the sets $\Omega_i$ for $1 \leq i \leq k$. For $|\Omega| \geq 2$, a transitive group $G \leq \text{Sym}(\Omega)$ is called primitive if there are only the two trivial systems of imprimitivity for $G$: the partition into one-element sets, and the partition consisting of one part $\Omega_1 = \Omega$.

We say that a graph (or a multigraph) is regular with degree or valency $d$ if there are $d$ edges adjoining every vertex; that is, “degree” and “valency” of a vertex mean the same thing. In a directed graph, the out-degree of a vertex $x$ is the number of edges starting at $x$ while the in-degree is the number of edges terminating at $x$. A directed graph is called strongly connected if for any two vertices $x, y$, there is a directed path from $x$ to $y$.

By $f(n) \ll g(n)$, $g(n) \gg f(n)$, and $f(n) = O(g(n))$ we mean one and the same thing, namely, that there are $N > 0$, $C > 0$ such that $|f(n)| \leq C \cdot g(n)$ for all $n \geq N$.

We write $\log_2 x$ to mean the logarithm base 2 of $x$ (and not to mean $\log \log x$).

3. Preliminaries on sets, groups and growth

3.1. Orbits and stabilisers. The orbit-stabiliser theorem from elementary group theory carries over to sets. This is a fact whose importance to the area is difficult to overemphasise. It underlies already [Hel08] at a key point (Proposition 4.1); the action at stake there is that of a group $G$ on itself by conjugation.
ON THE DIAMETER OF PERMUATION GROUPS

The setting for the theorem is the action of a group $G$ on a set $X$. The stabiliser $G_x$ of a point $x \in X$ is the set $\{g \in G : x^g = x\}$.

**Lemma 3.1** (Orbit-stabiliser theorem for sets). Let $G$ be a group acting on a set $X$. Let $x \in X$, and let $A \subseteq G$ be nonempty. Then

\[ |AA^{-1} \cap G_x| \geq \frac{|A|}{|x^A|}. \]  

Moreover, for every $B \subseteq G$,

\[ |AB| \geq |A \cap G_x||x^B|. \]

The usual orbit-stabiliser theorem is the special case $A = B = H$, $H$ a subgroup of $G$.

**Proof.** By the pigeonhole principle, there exists an image $x' \in x^A$ such that the set $S = \{a \in A : x^a = x'\}$ has at least $|A|/|x^A|$ elements. For any $a, a' \in S$, $x^{a(a')^{-1}} = (x')^{(a')^{-1}} = x$. Hence

\[ |AA^{-1} \cap G_x| \geq |SS^{-1}| \geq |S| \geq \frac{|A|}{|x^A|}. \]

Let $b_1, b_2, \ldots, b_\ell \in B$, $\ell = |x^B|$, be elements with $x^{b_i} \neq x^{b_j}$ for $i \neq j$. Consider all products of the form $ab_i$, $a \in A \cap G_x$, $1 \leq i \leq \ell$. If two such products $ab_i, a'b_i'$ are equal, then $x^{b_i} = x^{a'b_i} = x^{a'b_i'} = x^{b_i'}$. This implies $b_i = b_i'$. Since $ab_i = a'b_i'$, we conclude that $a = a'$. We have thus shown that all products $ab_i$, $a \in A \cap G_x$, $1 \leq i \leq \ell$, are in fact distinct. Hence

\[ |AB| \geq |(A \cap G_x) \cdot \{b_i : 1 \leq i \leq \ell\}| \]

\[ = |A \cap G_x| \cdot \ell = |A \cap G_x| \cdot |x^B|. \quad \Box \]

As the following corollaries show, the relation between the size of $A$, on the one hand, and the size of orbits and stabilisers, on the other, implies that growth in the size of either orbits or stabilisers induces growth in the size of $A$ itself.

**Corollary 3.2.** Let $G$ be a group acting on a set $X$. Let $x \in X$. Let $A \subseteq G$ be a nonempty set with $A = A^{-1}$. Then, for any $k > 0$,

\[ |A^{k+1}| \geq \frac{|A^k \cap G_x|}{|A^2 \cap G_x|} |A|. \]

**Proof.** By (3.2),

\[ |A^{k+1}| \geq |A^k \cap G_x||x^A| \geq \frac{|A^k \cap G_x|}{|A^2 \cap G_x|} |A^2 \cap G_x||x^A|. \]

Since $|A^2 \cap G_x||x^A| \geq |A|$ (by (3.1)), we obtain (3.3). \qed
**Corollary 3.3.** Let $G$ be a group acting on a set $X$. Let $x \in X$. Let $A \subseteq G$ be a nonempty set with $A = A^{-1}$. Then, for any $k > 0$,

\[(3.4) \quad |A^{k+2}| \geq \frac{|x^{A^k}|}{|x^A|} |A|.
\]

*Proof.* By (3.2) and (3.1),

\[|A^{k+2}| \geq |A^2 \cap G_x||x^{A^k}| \geq \frac{|A|}{|x^A|} |x^{A^k}| = \frac{|x^{A^k}|}{|x^A|} |A|.
\]

\[\square\]

3.2. **Lemmas on subgroups and quotients.** We start by recapitulating some of the simple material in [Hel11, §7.1]. The first lemma guarantees that we can always find many elements of $AA^{-1}$ in any subgroup of small enough index.

**Lemma 3.4** ([Hel11, Lemma 7.2]). Let $G$ be a group and $H$ a subgroup thereof. Let $A \subseteq G$ be a nonempty set. Then

\[(3.5) \quad |AA^{-1} \cap H| \geq \frac{|A|}{r},
\]

where $r$ is the number of cosets of $H$ intersecting $A$. In particular,

\[|AA^{-1} \cap H| \geq \frac{|A|}{[G : H]}.
\]

*Proof.* By the orbit-stabiliser principle (3.1) applied to the natural action of $G$ on $G/H$ by multiplication on the right.\(^6\) (Set $x = He = H$.) \[\square\]

The following two lemmas should be read as follows: growth in a subgroup gives growth in the group; growth in a quotient gives growth in the group.

**Lemma 3.5** (essentially [Hel11, Lemma 7.3]). Let $G$ be a group and $H$ a subgroup thereof. Let $A \subseteq G$ be a nonempty set with $A = A^{-1}$. Then, for any $k > 0$,

\[(3.6) \quad |A^{k+1}| \geq \frac{|A^k \cap H|}{|A^2 \cap H|} |A|.
\]

*Proof.* By Corollary 3.2 applied to the action of $G$ on $G/H$ by multiplication on the right (with $x = He = H$). \[\square\]

For a group $G$ and a subgroup $H \leq G$, we define the coset map $\pi_{G/H} : G \to G/H$ that maps each $g \in G$ to the right coset $Hg$ containing $g$.

---

\(^6\)Recall that we are following the convention that $G/H$ is the set of right cosets $Hg$. 
Lemma 3.6 (essentially [Hel11, Lemma 7.4]). Let $A \subseteq G$ be a nonempty set with $A = A^{-1}$. Then, for any $k > 0$,

$$|A^{k+2}| \geq \frac{|\pi_{G/H}(A^k)|}{|\pi_{G/H}(A)|} |A|.$$ 

Proof. By Corollary 3.3, applied with $G$ acting on $X := G/H$ by multiplication on the right and with $x := H$ seen as an element of $G/H$. □

The following lemma is a generalisation of Lemma 3.4.

Lemma 3.7. Let $G$ be a group, let $H,K$ be subgroups of $G$ with $H \leq K$, and let $A \subseteq G$ be a nonempty set. Then

$$|\pi_{K/H}(AA^{-1} \cap K)| \geq \frac{|\pi_{G/H}(A)|}{|\pi_{G/K}(A)|} \geq \frac{|\pi_{G/H}(A)|}{|G : K|}.$$ 

In other words: if $A$ intersects $r|G : H|$ cosets of $H$ in $G$, then $AA^{-1}$ intersects at least $r|G : H|/|G : K| = r|K : H|$ cosets of $H$ in $K$. (As usual, all our cosets are right cosets.)

Proof. Since $A$ intersects $|\pi_{G/H}(A)|$ cosets of $H$ in $G$ and $|\pi_{G/K}(A)|$ cosets of $K$ in $G$, and every coset of $K$ in $G$ is a disjoint union of cosets of $H$ in $G$, the pigeonhole principle implies that there exists a coset $Kg$ of $K$ such that $A$ intersects at least $k = |\pi_{G/H}(A)|/|\pi_{G/K}(A)|$ cosets $Ha \subseteq Kg$. Let $a_1, \ldots, a_k$ be elements of $A$ in distinct cosets of $H$ in $Kg$. Then $a_1a_1^{-1} \in AA^{-1} \cap K$ for each $i = 1, \ldots, k$. Finally, note that $Ha_1a_1^{-1}, \ldots, Ha_k a_k^{-1}$ are $k$ distinct cosets of $H$. □

The above lemmas fall into two types: either (a) they reduce the problem of proving growth in $G$ to that of proving growth in a smaller structure (a subgroup in Lemma 3.5, a quotient in Lemma 3.6), or (b) they produce many elements in a smaller structure (a group in Lemma 3.4, a quotient in Lemma 3.7).

Lastly, we present a result of a somewhat different nature. It is a version of Schreier’s lemma (rewritten slightly as in [GH10, Lemma 2.10]). Usually, if a set $A$ generates a group $G$, that does not mean that, for $H$ a subgroup of $G$, the intersection $A \cap H$ will generate $H$. However, Lemma 3.8 tells us, if $A$ projects onto $G/H$, then $A^3 \cap H$ does generate $H$. We will use Lemma 3.8 in the proof of Lemma 6.2 (for $G$ a setwise stabiliser ($\text{Sym}(n)$)$_\Delta$ and $H$ the corresponding pointwise stabiliser ($\text{Sym}(n)$)$_{\langle\Delta\rangle}$).

Lemma 3.8 (Schreier). Let $G$ be a group and $H$ a subgroup thereof. Let $A \subseteq G$ with $A = A^{-1}$ and $e \in A$. Suppose $A$ intersects each coset of $H$ in $G$. Then $A^3 \cap H$ generates $\langle A \rangle \cap H$. Moreover, $\langle A \rangle = \langle A^3 \cap H \rangle A$. 

Proof. Let \( C \subseteq A \) be a full set of right coset representatives of \( H \), with \( e \in C \). We wish to show that \( \langle A \rangle = \langle A^3 \cap H \rangle C \). (This immediately implies both \( \langle A \rangle = \langle A^3 \cap H \rangle A \) and \( \langle A \rangle \cap H = \langle A^3 \cap H \rangle \).)

Clearly \( e \in \langle A^3 \cap H \rangle C \). It is thus enough to show that, if \( g = hc \), where \( h \in \langle A^3 \cap H \rangle \) and \( c \in C \), and \( a' \in A \), then \( ga' \) still lies in \( \langle A^3 \cap H \rangle C \). This is easily seen: since \( C \) is a full set of coset representatives, there is a \( c' \in C \) with \( c' = h'ca' \) for some \( h' \in H \), and thus
\[
ga' = hca' = h((h')^{-1})h'ca' = h((h')^{-1})c' \in \langle A^3 \cap H \rangle \langle A^3 \cap H \rangle C = \langle A^3 \cap H \rangle C,
\]
where we use the fact that \( h' = c'(a')^{-1}c^{-1} \in A^3 \).

3.3. Actions and generators. The proofs of the next two lemmas share a rather simple idea. Indeed, both lemmas can be seen as consequences of the well-known fact that every connected graph has a spanning tree.\(^7\) The graph would be the union of the permutation graphs (with \( X \) as the vertex set) induced by the elements of the set \( A \).

We give two brief proofs without graphs.

Lemma 3.9. Let \( G \) be a group acting transitively on a finite set \( X \). Let \( A \subseteq G \) with \( A = A^{-1} \), \( e \in A \), and \( G = \langle A \rangle \). Then, for any \( x \in X \),
\[
x^{A^\ell} = X,
\]
where \( \ell = |X| - 1 \).

Proof. Consider the orbits \( \{ x \} \subseteq x^A \subseteq x^{A^2} \subseteq \cdots \). Let \( \ell' \) be the smallest integer with \( x^{A^{\ell'+1}} = x^{A^\ell} \). As \( x^{A^{\ell'+2}} = (x^{A^{\ell'+1}})^A = (x^{A^\ell})^A = x^{A^{\ell+1}} = x^{A^\ell} \), we have \( x^{A^\ell} = x^{(A)} = x^G = X \). Since
\[
\{ x \} \subseteq x^A \subseteq x^{A^2} \subseteq \cdots \subseteq x^{A^{\ell'}} = X,
\]
we have \( \ell' \leq |X| - 1 \). \(\)\(\)

Lemma 3.10. Let \( G \) be a group acting transitively on a finite set \( X \). Let \( A \subseteq G \) with \( A = A^{-1} \) and \( G = \langle A \rangle \). Then there is a subset \( A' \subseteq A \), \( |A'| < |X| \), such that \( \langle A' \rangle \) acts transitively on \( X \).

Proof. Let \( x \in X \). Let \( A_1 = \{ g \} \), where \( g \) is any element of \( A \) such that \( x^g \neq x \). For each \( i \geq 1 \), let \( A_{i+1} = A_i \cup \{ g_i \} \), where \( g_i \) is an element of \( A \) such that \( x^{(A_i \cup \{ g_i \})} \supseteq x^{(A_i)} \). If no such element \( g_i \) exists, we can conclude that \( x^{(A_i)} \) is taken to itself by every \( g_i \in A \). This implies that \( x^{(A_i)} \) is taken to itself by every product of elements of \( A \), and thus \( (x^{(A_i)}) \langle A \rangle = x \langle A \rangle \) equals \( x^{(A_i)} \).

Hence, we have a chain
\[
\{ x \} \subseteq x^{(A_1)} \subseteq x^{(A_2)} \subseteq \cdots \subseteq x^{(A_i)} = x^{(A)} = X.
\]
Clearly \( i \leq |X| - 1 \), and so \( |A_i| \leq |X| - 1 \). Let \( A' = A_i \). \(\)\(\)

\(^7\)We thank an anonymous referee for this comment.
3.4. Large subsets of $\text{Sym}(n)$. Let us first prove a result on large subgroups of $\text{Sym}(n)$.

**Lemma 3.11.** Let $n \geq 84$. Let $G \leq \text{Sym}(n)$ be transitive, with a section isomorphic to $\text{Alt}(k)$ for some $k > n/2$. Then $G$ is either $\text{Alt}(n)$ or $\text{Sym}(n)$.

**Proof.** Since $k \geq 5$, the group $\text{Alt}(k)$ is simple. Hence some composition factor of $G$ has a section isomorphic to $\text{Alt}(k)$. Assume that $G$ is imprimitive, and let $B$ be a nontrivial system of imprimitivity for $G$. Write $b = |B|$ and $m = n/b$, and let $K$ be the kernel of the action of $G$ on $B$. Since $G/K$ is isomorphic to a subgroup of $\text{Sym}(b)$, $K$ is isomorphic to a subgroup of $\text{Sym}(m)^b$ and $b, m < k$, we obtain that $G$ has no section isomorphic to $\text{Alt}(k)$, a contradiction. This shows that $G$ is primitive.

From [PS80], we obtain that either $G \geq \text{Alt}(n)$ or $|G| \leq 4^n$. Since $|G| \geq |\text{Alt}(k)| = k!/2 \geq [n/2]!/2$, a direct computation shows that the latter case arises only for $n < 84$. □

Our aim for the rest of this subsection will be to show that, if $A \subseteq \text{Sym}(n)$ is very large, then $A^{|\Omega|}$ contains a copy of $\text{Alt}(\Delta)$, $|\Delta| > n/2$. The next lemma generalises Bochert’s theorem [Boc89], [DM96, Thm. 3.3B] to subsets. Recall that, for $g \in \text{Sym}(\Omega)$, we define the support of $g$ by $\text{supp}(g) = \{\alpha \in \Omega : \alpha^g \neq \alpha\}$.

**Lemma 3.12.** Let $n \geq 5$. Let $A \subseteq \text{Sym}([n])$ with $A = A^{-1}$, $e \in A$. If $\langle A \rangle$ is a primitive permutation group and $|A| > n!/(\lfloor n/2 \rfloor)!$, then $A^n$ is either $\text{Alt}([n])$ or $\text{Sym}([n])$.

This is an example of how one can sometimes modify a proof of a result about subgroups to give a result about sets: the proof follows the lines of Bochert’s essentially algorithmic proof plus some bookkeeping.

**Proof.** Given $A \subseteq \text{Sym}([n])$ as in the statement of the lemma, let $k$ be the smallest integer such that there exists $\Delta \subseteq [n]$ with $|\Delta| = k$ and $(A^2)_{(\Delta)} = \{e\}$. Let $\Delta$ be one such set.

Suppose that $k \leq n/2$. Then $\text{Sym}([n])_{(\Delta)}$ has $n!/(n-k)! < |A|$ cosets in $\text{Sym}([n])$. Thus, by the pigeonhole principle, there exist two distinct elements $a$ and $b$ of $A$ in the same coset. Hence $ab^{-1} \in \text{Sym}([n])_{(\Delta)}$; that is, $ab^{-1} \in (A^2)_{(\Delta)}$. This contradicts the definition of $k$. We conclude that $k > n/2$.

The set $\Omega = [n] \setminus \Delta$ has cardinality less than $k$, so by definition there exists $g \in (A^2)_{(\Omega)}$ with $g \neq e$. Let $\delta \in \Delta$ with $\delta^g \neq \delta$. As the set $\Delta \setminus \{\delta\}$ has cardinality less then $k$, by the definition of $k$, there exists $h \in (A^2)_{(\Delta \setminus \{\delta\})}$ with $h \neq e$. Then $\text{supp}(h) \subset \Omega \cup \{\delta\}$. Necessarily, $\delta \in \text{supp}(h)$, otherwise $(A^2)_{(\Delta)}$ contains the nonidentity element $h$. Hence $\text{supp}(g) \cap \text{supp}(h) = \{\delta\}$ and so the commutator $x = [g, h]$ is a 3-cycle. Note that $[g, h] \in A^8$. 

ON THE DIAMETER OF PERMUTATION GROUPS

15
Now, since \(\langle A\rangle\) is primitive and contains a 3-cycle, by Jordan’s theorem [DM96, Thm. 3.3A] we obtain that \(\langle A\rangle \geq \text{Alt}(\{n\})\). In particular, \(\langle A\rangle\) is 3-transitive, and thus its action by conjugation on the set \(X\) of all 3-cycles is transitive. By Lemma 3.9,

\[ x^{A^\ell} = X, \]

where \(\ell = |X| = n(n-1)(n-2)/3\) and \(A^\ell\) acts on \(x\) by conjugation. Thus

\[ A^{n(n-1)(n-2)/3} = A^{n(n-1)(n-2)/3} \]

contains all 3-cycles in \(\text{Alt}([n])\).

Since any element of \(\text{Alt}([n])\) can be written as a product of at most \(\lfloor n/2 \rfloor\) 3-cycles, we obtain that \(A^{n^3-1}\) contains \(\text{Alt}([n])\). Also, if \(A\) contains an odd permutation, then \(A = \text{Sym}([n])\). □

What happens, however, if \(\langle A\rangle\) is not transitive, let alone primitive? We shall see first that, if \(A\) is large, then \(\langle A\rangle\) must have at least a large orbit. In the following two lemmas, we use the inequalities

\[ \left(\frac{n}{\text{e}}\right)^n < n! < 3\sqrt{n} \left(\frac{n}{\text{e}}\right)^n. \]  

**Lemma 3.13.** Let \(H < \text{Sym}(n)\) with \(|H| \geq d^n n!\) for some number \(d\) with \(0.5 < d < 1\). If \(n\) is greater than a bound depending only on \(d\), then \(H\) has an orbit of length at least \(dn\).

**Proof.** Let \(k := \lfloor dn \rfloor\). Suppose that the longest orbit length of \(H\) is less than \(dn\). Then, as is well known, \(|H| \leq k!(n-k)!\). (The size of a direct product of symmetric groups \(\text{Sym}(\Omega_i)\) only goes up if we pass elements from the smaller sets \(\Omega_i\), \(i \geq 2\), to the largest set \(\Omega_1\).)

Now, by (3.7), we have the following inequalities:

\[ k^n \left(\frac{n}{\text{e}}\right)^n < \left(\frac{k}{n}\right)^n n! \leq d^n n! \leq |A| \leq |\langle A\rangle| \leq k!(n-k)! \]

\[ < 9\sqrt{k(n-k)} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{\text{e}}\right)^{n-k} \leq 9\left(\frac{k}{n}\right)^{n-k} \right(\frac{n-k}{\text{e}}\right)^{n-k}. \]

Simplifying the left-hand side together with the right-hand side, we obtain

\[ k^{n-k} < \frac{9}{2} n(n-k)^{n-k}; \]  

that is, \(\left(\frac{k}{n-k}\right)^{n-k} < \frac{9}{2} n\).

We define \(c := \left(\frac{d}{1-d}\right)^{1-d}\). As

\[ \lim_{n \to \infty} \left(\frac{k}{n-k}\right)^{n-k} = c > 1, \]

for large enough \(n\), depending only on \(d\), we have \(\left(\frac{k}{n-k}\right)^{n-k} > \left(\frac{1+c}{2}\right)^n\). However, \(\left(\frac{1+c}{2}\right)^n < \frac{9}{2} n\) is false if \(n\) is greater than a bound depending only on \(d\), proving our claim.

□
Using Bochert’s theorem [Boc89], Liebeck derived a result ([Lie83, Lemma 1.1]; see [Jor57, pp. 68–75] for a classical result of the same kind) on large subgroups of Sym(n). It does not assume transitivity or primitivity. We will generalise it to sets (Proposition 3.15). In a somewhat strengthened version [DM96, Thm. 5.2B], the result from [Lie83] states the following, among other things: if H is a subgroup of Sym(n), n ≥ 9, and

\[(3.10) \quad \text{Alt}(n)_{[n]\Delta} \leq H \leq \text{Sym}(n)_{[n]\Delta}.\]

Here, of course, Alt(n)_{[n]\Delta} ~ Alt(\Delta) and Sym(n)_{[n]\Delta} = Sym(n)\Delta; in particular, (3.10) implies that \(\Delta\) is an orbit of \([n]\). It is easy to see that, if \(|H| \geq d^n n!\), 0.5 < d < 1, then (3.9) is fulfilled for \(m = [dn]\), provided that n is larger than a constant depending only on d: by Stirling’s formula,

\[(3.11) \quad \left(\frac{n}{[dn]}\right) \gg \frac{1}{\sqrt{n}} \left(\frac{n}{[dn]}\right)^n \gg \frac{1}{n^{3/2}} \left(\frac{1}{d^d(1-d)^{1-d}}\right)^n\]

and, since \(d^d(1-d)^{1-d} < d\) for \(d \in (1/2, 1)\), this is certainly greater than \((1/d)^n\) for \(n\) large enough. The inequality \(\frac{1}{2} (\frac{n}{[n/2]} \gg 2^n/\sqrt{n} \Rightarrow \frac{1}{2} (\frac{n}{[n/2]} > (1/d)^n\)

immediately for all large \(n\). Thus (3.10) holds for some \(\Delta\) with \(|\Delta| > dn\).

We will show an analogue of (3.10) holds for a set \(A\) instead of a subgroup \(H\) (Proposition 3.15). This can be shown in two ways: we can use Liebeck’s result (3.10) for groups, or we can give an elementary proof using only counting arguments. (Both [Lie83] and [DM96] do a detailed examination of the subgroup structure of Sym(n) in order to give a result valid for small \(n\).

Let us first give an elementary proof of a somewhat weaker statement.

**Lemma 3.14.** Let \(d\) be a number with 0.5 < \(d\) < 1. If \(A \subseteq \text{Sym}(\{n\})\) (with \(A = A^{-1}\)) has cardinality \(|A| \geq d^n n!\) and \(n\) is larger than a bound depending only on \(d\), then there exists an orbit \(\Delta \subseteq [n]\) of \(\langle A\rangle\) such that \(|\Delta| \geq dn\) and \((A^{n})_{\Delta} = \text{Alt}(\Delta)\) or \(\text{Sym}(\Delta)\).

**Proof.** By Lemma 3.13, for large enough \(n\), the group \(\langle A\rangle\) has an orbit \(\Delta\) of length \(k \geq dn\). Write \(\rho = k/n\), and note that \(d \leq \rho \leq 1\). The group \(G = B|_{\Delta}\) has order at least \(d^n n!/ (n-k)!\), so estimating \(k!(n-k)!\) from above as in (3.8) and estimating \(n!\) from below by (3.7), we obtain

\[(3.12) \quad [\text{Sym}(\Delta) : G] \leq \frac{k!(n-k)!}{d^n n!} \leq \frac{9}{4} \frac{k^k(n-k)^{n-k}}{d^n n^n} = \frac{9}{4} \left(\frac{n\rho(1-\rho)^{1-\rho}}{d}\right)^n \leq \frac{9}{4} \left(\frac{n}{d}\right)^n.\]
Next, we show that for large values of $n$, the transitive group $G$ cannot be imprimitive. Indeed, if $G$ is imprimitive, then using (3.7), we have

\begin{equation}
[\text{Sym}(\Delta) : G] \geq \frac{1}{2} \left( \frac{k}{|k/2|} \right) > \frac{1}{2} \frac{(\frac{k}{2e})^k}{9k(\frac{k}{2e})^k} > \frac{1}{18n} 2^{\rho n}.
\end{equation}

A direct computation shows that the function $f(\rho) = 2^{1/\rho} \rho (1 - \rho)^{(1 - \rho)/\rho}$ is monotone increasing in the interval $[1/2, 1]$ with supremum 2. Hence, comparing the upper and lower bounds for $[\text{Sym}(\Delta) : G]$ deduced in (3.12) and (3.13), we obtain

\begin{equation}
\frac{9}{4} n 2^{\rho n} \left( \frac{1}{2d} \right)^n > \frac{1}{18n} 2^{\rho n}.
\end{equation}

As $d > 1/2$, for large enough $n$, we have $(2d)^n > (18n)(\frac{9}{4} n)$, and therefore (3.14) cannot hold.

Hence $G$ is primitive and $A|\Delta$ is a set of size at least $d^n n!/(n - k)! \geq d^n k! > k!/(\lfloor k/2 \rfloor)!$ (where the last inequality holds for $n$ greater than a lower bound depending only on $d$). Therefore, by Lemma 3.12, $(A|\Delta)^n$ is either $\text{Alt}(\Delta)$ or $\text{Sym}(\Delta)$, and hence so is $(A^n)^{|\Delta} = (A|\Delta)^n$.

Now we get the full analogue of (3.10).

**Proposition 3.15.** Let $d$ be a number with $0.5 < d < 1$. Let $A \subseteq \text{Sym}(n)$ with $A = A^{-1}$ and $e \in A$. If $|A| \geq d^n n!$ and $n$ is larger than a bound depending only on $d$, then there exists an orbit $\Delta \subseteq [n]$ of $\langle A \rangle$ such that $|\Delta| \geq dn$ and $(A^n)^{|\Delta \setminus \Delta)} \Delta$ contains $\text{Alt}(\Delta)$.

**Proof.** By Lemma 3.14, there is an orbit $\Delta$ of $\langle A \rangle$ such that $|\Delta| \geq dn$ and $(A^n)^{|\Delta \setminus \Delta)} \Delta$ is $\text{Alt}(\Delta)$ or $\text{Sym}(\Delta)$. Let $A' = A^n$.

It is clear that $|A'| \geq |\text{Alt}(\Delta)| > |\text{Sym}([n] \setminus \Delta)|$. Thus, by the pigeonhole principle, there are $h_1, h_2 \in A'$, $h_1 \neq h_2$, such that $h_1|n \setminus \Delta = h_2|n \setminus \Delta$, and so $g = h_1 h_2^{-1}$ fixes $[n] \setminus \Delta$ pointwise.

We show that $((A')^{14})([n] \setminus \Delta)$ contains an element $g'$ such that $g'|\Delta$ is a 3-cycle. If $g|\Delta$ has at least two fixed points, then there exists an element $h \in A'$ such that $h|\Delta$ is a 3-cycle, with $\text{supp}(h|\Delta)$ intersecting $\text{supp}(g|\Delta)$ in exactly one point. Then $g' = [g, h] \in (A')^{2+1+2+1} = (A')^6$ fixes $[n] \setminus \Delta$ pointwise and $g'|\Delta$ is a 3-cycle. If $g$ contains a cycle $(\alpha \beta \gamma \delta \cdots)$ of length at least 4, then we choose an element $h \in A'$ with $h|\Delta = (\alpha \beta \gamma)$ and let $g' = [g, h] \in (A')^6$. Then $g'$ fixes $[n] \setminus \Delta$ pointwise and $g'|\Delta$ is the 3-cycle $(\alpha \beta \delta)$.

In all other cases, $|\text{supp}(g|\Delta)| \geq |\Delta| - 1 \geq 6$ (assuming $n \geq 13$, which implies $|\Delta| \geq 7$) and all nontrivial cycles of $g$ have length 2 or 3. Hence $g|\Delta$ contains at least two 3-cycles or at least two 2-cycles.
If \( g|_{\Delta} \) contains the cycles \((\alpha\beta\gamma)\) and \((\delta\eta\nu)\), then we choose an element \( h \in A' \) with \( h|_{\Delta} = (\alpha\eta)(\beta\delta\gamma\nu) \). A little computation shows that \( g' = [g, h] \) fixes \([n] \setminus \Delta\) pointwise and \( g'|_{\Delta} \) is the 3-cycle \((\delta\eta\nu)\).

Finally, suppose \( g \) contains the 2-cycles \((\alpha\beta)\) and \((\gamma\delta)\). We again choose an element \( h \in A' \) with \( h|_{\Delta} = (\alpha\beta\gamma) \); then \( \text{supp}([g, h]) = \{\alpha, \beta, \gamma, \delta\} \) and \( [g, h] \) fixes \([n] \setminus \Delta\) pointwise. Since \( [g, h] \in (A')^6 \) also fixes at least two points of \( \Delta \), we deduce as in the very first case of our analysis that the commutator \( g' = [[g, h], h'] \) with an appropriate \( h' \in A' \) is a 3-cycle. Note that \( g' \in (A')^{6+1+6+1} = (A')^{14} \).

Given any 3-cycle \( s \in \text{Sym}(\Delta) \), we can conjugate \( g' \) by an appropriate element of \( A' \) to get an element of \(((A')^{16})([n]\setminus\Delta)\) whose restriction to \( \Delta \) equals \( s \). Now, every element of \( \text{Alt}(\Delta) \) is the product of at most \(|\Delta|/2\) 3-cycles. Hence \(((A')^{16}|_{[n]/\Delta})|_{\Delta} \) contains \( \text{Alt}(\Delta) \). \( \square \)

An anonymous referee kindly provides the following argument, showing that Proposition 3.15, which is a generalisation of (3.10), can be proven using (3.10).

**Second proof of Proposition 3.15.** (This proof gives Proposition 3.15 with \( A^{2(n+1)}n^4 \) instead of \( A^{8n^5} \).) By (3.10) applied to \( H = \langle A \rangle \), there is a set \( \Delta \) with \(|\Delta| > dn \) such that (a) \( H \) is contained in \( \text{Sym}(n)_{\Delta} \), and (b) \( H \) contains the subgroup \( D = \text{Alt}(n)_{[n]\setminus\Delta} \); i.e., \( H|_{\Delta} \) contains \( \text{Alt}(\Delta) \). Let \( B = A^2 \cap D \). By Lemma 3.4,

\[
|B| = |A^2 \cap D| \geq \frac{|A|}{[\text{Sym}(n)_{\Delta} : D]} \geq \frac{d^n n!}{2(n - |\Delta|)!} > \frac{d^n|\Delta|!}{2^{n+1}} > \frac{|\Delta|!}{2^{2|\Delta|+1}}.
\]

For \( n \) sufficiently large (and hence \(|\Delta| \) sufficiently large), \( 2^{2|\Delta|+1} < |\Delta|/2 \), and so we obtain that \( |B| > |\Delta|/|\Delta|/2 \).

Since \( \langle A|_{\Delta} \rangle = H|_{\Delta} \) contains \( \text{Alt}(\Delta) \), \( \langle (A \cup B)|_{\Delta} \rangle \) is \( \text{Alt}(\Delta) \) or \( \text{Sym}(\Delta) \) — and, in particular, it is primitive. Hence, a first application of Lemma 3.12 (with \( \Delta \) instead of \([n]\)) implies that \( \langle (A \cup B)|_{\Delta} \rangle n^4 \) is \( \text{Alt}(\Delta) \) or \( \text{Sym}(\Delta) \).

The set \( S = \{ gbg^{-1} : g \in (A \cup B)^n, b \in B \} \) is in \( D \); moreover, \( \langle S|_{\Delta} \rangle \) is normal in \( \text{Alt}(\Delta) \). Since \( S|_{\Delta} \) is nontrivial (by \(|B| > 1\)), we conclude that \( \langle S|_{\Delta} \rangle = \text{Alt}(\Delta) \). Now we apply Lemma 3.12 (again with \( \Delta \) instead of \([n]\)) and obtain that \( \langle S|_{\Delta} \rangle n^4 = \text{Alt}(\Delta) \). Since \( S^4 \subset A^{2(n^4+2)n^4} \), we are done. \( \square \)

**3.5. Bases and stabiliser chains.** Given a permutation group \( G \) on a set \( \Omega \), a subset \( \Sigma \) of \( \Omega \) is called a base if \( G_{(\Sigma)} = \{ e \} \). This definition goes back to Sims [Sim70]. If, instead of \( G \), we consider a subset \( A \) of \( \text{Sym}(\Omega) \), then, as the following lemma suggests, it makes sense to see whether \( (AA^{-1})_{(\Sigma)} \) (rather than \( A_{(\Sigma)} \)) equals \( \{ e \} \).
Lemma 3.16. Let $A \subseteq \text{Sym}(\Omega)$, \(|\Omega| = n\). If $\Sigma \subseteq \Omega$ is such that \((AA^{-1})_{(\Sigma)} = \{e\}\), then \(|\Sigma| \geq \log_n |A|\).

Proof. Notice first that \([\text{Sym}(\Omega) : (\text{Sym}(\Omega))_{(\Sigma)}] \leq n|\Sigma|\). By the pigeonhole principle, if \(|A| > n|\Sigma|\), then there exists a right coset of \((\text{Sym}(\Omega))_{(\Sigma)}\) containing more than one element of $A$, and thus
\[
|AA^{-1}_{(\Sigma)}| = |AA^{-1} \cap (\text{Sym}(\Omega))_{(\Sigma)}| > 1.
\]

Hence, if \((AA^{-1})_{(\Sigma)} = \{e\}\), then we have \(|A| \leq n|\Sigma|\); i.e., \(|\Sigma| \geq \log_n |A|\). \(\Box\)

The use of stabiliser chains $H > H_\alpha > H_{(\alpha_1, \alpha_2)} > \cdots$ is very common in computational group theory (starting, again, with the work of Sims; see references in [Ser03, §4.1]). We may study a similar chain $A > A_\alpha > A_{(\alpha_1, \alpha_2)} > \cdots$ when $A$ is merely a set.

Lemma 3.17. Let $\Sigma = \{\alpha_1, \ldots, \alpha_m\} \subseteq [n]$ and $A \subseteq \text{Sym}([n])$. Suppose that
\[
|A_{(\alpha_1, \ldots, \alpha_i-1)}| \geq r_i
\]
for all $i = 1, 2, \ldots, m$. Then $A^m$ intersects at least $\prod_{i=1}^m r_i$ cosets of $(\text{Sym}([n]))_{(\Sigma)}$.

Proof. For each $1 \leq i \leq m$, write $\Delta_i = \alpha_i^A_{(\alpha_1, \ldots, \alpha_i-1)}$; thus \(|\Delta_i| \geq r_i\). For each $\delta \in \Delta_i$, pick $g_\delta \in A_{(\alpha_1, \ldots, \alpha_i-1)}$ with $\alpha_i^g = \delta$ and write $S_i = \{g_\delta : \delta \in \Delta_i\}$. Clearly, \(|S_i| = |\Delta_i|\) and $S_i \subseteq A$. We show that for every two distinct tuples
\[
(s_1, s_2, \ldots, s_m), (s'_1, s'_2, \ldots, s'_m) \in S_1 \times \cdots \times S_m,
\]
the products $P = s_m s_{m-1} \cdots s_1$ and $P' = s'_m s'_{m-1} \cdots s'_1$ belong to two distinct cosets of $(\text{Sym}([n]))_{(\Sigma)}$. From this it follows that $A^m$ intersects at least $|S_1| \cdots |S_m| = |\Delta_1| \cdots |\Delta_m| \geq \prod_{i=1}^m r_i$ cosets of $(\text{Sym}([n]))_{(\Sigma)}$.

We argue by contradiction; i.e., we assume that $P$ and $P'$ map \((\alpha_1, \ldots, \alpha_m)\) to the same $m$-tuple. Let $j$ be the smallest index such that $s_j \neq s'_j$. Then $Q = P s_1^{-1} \cdots s_{j-1}^{-1}$ and $Q' = P' s'_1^{-1} \cdots s'_{j-1}^{-1}$ also map \((\alpha_1, \ldots, \alpha_m)\) to the same $m$-tuple. Note that for all $k \leq m$, $s_k$ and $s'_k$ fix \((\alpha_1, \ldots, \alpha_{k-1})\) pointwise. Thus
\[
\alpha_j^Q = \alpha_j^{s_j} \neq \alpha_j^{s'_j} = \alpha_j^{Q'},
\]
contradicting our assumption. \(\Box\)

We thus see that, if we choose $\alpha_1, \alpha_2, \ldots$ so that the orbits $\alpha_i^A_{(\alpha_1, \ldots, \alpha_i-1)}$ are large, we get to occupy many cosets of $(\text{Sym}([n]))_{(\Sigma)}$. By Lemma 3.7, this will enable us to occupy many cosets of $(\text{Sym}([n]))_{(\Sigma)}$ in the setwise stabiliser $(\text{Sym}([n]))_{\Sigma}$. We will then be able to apply Proposition 3.15 to build a large alternating group within $\text{Sym}(\Sigma) \cong (\text{Sym}([n]))_{\Sigma}/(\text{Sym}([n]))_{(\Sigma)}$. This procedure is already implicit in [Pyb03, Lemma 3]; indeed, what amounts to this is
signalled by Pyber as the main new element in his refinement [Pyb93, Thm. A] of Babai’s theorem on the order of doubly transitive groups [Bab82]. The main difference is that we have to work, of course, with sets rather than groups; we also obtain a somewhat stronger conclusion due to our using Proposition 3.15 rather than invoking Liebeck’s lemma directly.

**Lemma 3.18.** Let \( A \subseteq \text{Sym}(\{n]\) with \( A = A^{-1} \) and \( e \in A \). Let \( \Sigma = \{\alpha_1, \ldots, \alpha_m\} \subseteq [n] \) be such that

\[
(3.15) \quad \left| \alpha_i^{A_{\{\alpha_1, \ldots, \alpha_{i-1}\}}} \right| \geq dn
\]

for all \( i = 1, 2, \ldots, m \), where \( d > 0.5 \). Then, provided that \( m \) is larger than a bound \( C(d) \) depending only on \( d \), there exists \( \Delta \subseteq \Sigma \) with \( |\Delta| \geq d|\Sigma| \) and

\[
\text{Alt}(\Delta) \subseteq ((A^{16m^6})_\Sigma)(\Sigma\setminus\Delta)/\Delta.
\]

**Proof.** By (3.15) and Lemma 3.17, \( A^m \) intersects at least \( (dn)^m \) cosets of \( \text{Sym}(\{n]\) in \( \text{Sym}(\{n]\). Since

\[
[\text{Sym}(\{n]\) : \text{Sym}(\{n]\)\_\Sigma] = \frac{[\text{Sym}(\{n]\) : \text{Sym}(\{n]\)\_\Sigma]}{[\text{Sym}(\{n]\)\_\Sigma : \text{Sym}(\{n]\)\_\Sigma]} \leq \frac{n^m}{m!},
\]

Lemma 3.7 implies (with \( G = \text{Sym}(\{n]\), \( K = \text{Sym}(\{n]\)\_\Sigma, \( H = \text{Sym}(\{n]\)\_\Sigma\), and \( A^m \) instead of \( A \)) that

\[
|\pi_{K/H}(A^{2m} \cap K)| \geq \frac{|\pi_{G/H}(A^m)|}{\nu^m/m!} \geq \frac{(dn)^m}{n^m/m!} = d^m m!.
\]

Note that \(|\pi_{K/H}(A^{2m} \cap K)| = |(A^{2m})\_\Sigma/\Sigma|\). We can thus apply Proposition 3.15 (with \( m \) instead of \( n \), and \( A' = (A^{2m})\_\Sigma/\Sigma \) instead of \( A \)) and obtain that there is a set \( \Delta \subseteq \Sigma \) such that \( |\Delta| \geq dm \) and \( ((A')^{8m^5})_{\Sigma\setminus\Delta}/\Delta \) contains \( \text{Alt}(\Delta) \). \( \square \)

### 3.6. Existence of elements of small support

The following lemma is essentially [BS87, Lemma 3] (or [BS88, Lemma 1]; see also [BLS87]).

**Lemma 3.19.** Let \( \Delta \subseteq [n], |\Delta| \geq c(\log n)^2, c > 0 \). Let \( H \leq (\text{Sym}(n)\_\Delta \). Assume \( H\_\Delta \) is \( \text{Alt}(\Delta) \) or \( \text{Sym}(\Delta) \).

Let \( \Gamma \) be any orbit of \( H \). Then, if \( n \) is larger than a bound depending only on \( c \), \( H \) contains an element \( g \) with \( g|_{\Delta} \neq 1 \) and \( |\text{supp}(g|_{\Gamma})| < |\Gamma|/4 \).

**Proof.** Let \( p_1 = 2, p_2 = 3, \ldots, p_k \) be the sequence of the first \( k \) primes, where \( k \) is the least integer such that \( p_1 p_2 \cdots p_k > n^4 \). Much as in [BS87], we remark that, by elementary bounds towards the prime number theorem,

\[
2p_1 + p_2 + \cdots + p_k < c(\log n)^2,
\]

provided that \( n \) be larger than a bound depending only on \( c \). Thus \( H \) contains an element \( h \) such that \( h|_{\Delta} \) consists of \( |\Delta| - (2p_1 + p_2 + \cdots + p_k) \) fixed points
and cycles of length $p_1, p_1, p_2, p_3, \ldots, p_k$. (We need two cycles of length $p_1 = 2$ because we want an even permutation on $\Delta$.)

We can now reason as in [BS87, Lemma 3] or [BS88, Lemma 1]. For every $\gamma \in \Gamma$, denote by $\kappa_\gamma$ the length (possibly 1) of the cycle of $h$ containing $\gamma$ and for $i \leq k$, define $\Gamma_i := \{ \gamma \in \Gamma : p_i | \kappa_\gamma \}$. Then

$$\sum_{\gamma \in \Gamma} \sum_{p_i | \kappa_\gamma} \log p_i < |\Gamma| \log n$$

because $\kappa_\gamma < n$ implies that for all $\gamma$, the inner sum is less than $\log n$. Exchanging the order of summation,

$$\sum_{\gamma \in \Gamma} \sum_{p_i | \kappa_\gamma} \log p_i = \sum_{i=1}^{k} |\Gamma_i| \log p_i.$$

If $|\Gamma_i| \geq |\Gamma|/4$ for all $i \leq k$, then

$$\sum_{i=1}^{k} |\Gamma_i| \log p_i \geq \frac{|\Gamma|}{4} \log \left( \prod_{i=1}^{k} p_i \right) > \frac{|\Gamma|}{4} \log(n^k) = |\Gamma| \log n,$$

contradicting (3.17). Hence there is a prime $p \leq p_k$ such that $p | \kappa_\gamma$ for fewer than $|\Gamma|/4$ elements $\gamma$ of $\Gamma$. Denoting the order of $h$ by $|h|$, we define $g = h^\ell$ for $\ell := |h|/p$. We obtain that $|\text{supp}(g)\Delta| < |\Gamma|/4$. We also have that $g$ is nontrivial, since $g|\Delta$ contains a $p$-cycle. Clearly $g \in H$, and so we are done. \(\square\)

4. Random walks and generation

4.1. Random walks. The aim of this subsection is to present some basic material on random walks. As stated in the outline, our later use of random walks to mimic the uniform distribution in combinatorial arguments is clearly influenced by [BBS04]; indeed, this subsection is very close to the first two thirds of [BBS04, §2].

Let $\Gamma$ be a strongly connected directed multigraph with vertex set $V = V(\Gamma)$. For $x \in V(\Gamma)$, we denote by $\Gamma(x)$ the multiset of endpoints of the edges starting at $x$ (counted with multiplicities in case of multiple edges). We are interested in the special case when $\Gamma$ is regular of valency $d$ (i.e., $|\Gamma(x)| = d$, for each $x \in V(\Gamma)$) and $\Gamma$ is also symmetric in the sense that for all vertices $x, y \in V(\Gamma)$, the number of edges connecting $x$ to $y$ is the same as the number of edges connecting $y$ to $x$. These two conditions imply that the adjacency matrix $A$ of $\Gamma$ is symmetric and all row and column sums are equal to $d$.

\(\text{Since we need only the existence of } g \text{ for the moment, we are not concerned by the fact that } l \text{ is very large. Compare this to the situation in [BS88], where the use of a large } l \text{ causes diameter bounds much weaker than those in the present paper.} \)
A lazy random walk on $\Gamma$ is a stochastic process where a particle moves from vertex to vertex; if the particle is at vertex $x$ such that $\Gamma(x) = \{y_1, \ldots, y_d\}$, then the particle

- stays at $x$ with probability $\frac{1}{2}$;
- moves to vertex $y_i$ with probability $\frac{1}{2d}$ for all $i = 1, \ldots, d$.

Here we are concerned with the asymptotic rate of convergence for the probability distribution of a particle in a lazy random walk on $\Gamma$. For $x, y \in V(\Gamma)$, write $p_k(x, y)$ for the probability that the particle is at vertex $y$ after $k$ steps of a lazy random walk starting at $x$. For a fixed $\varepsilon > 0$, the $\ell_\infty$-mixing time for $\varepsilon$ is the minimum value of $k$ such that

$$\frac{1}{|V(\Gamma)|} (1 - \varepsilon) \leq p_k(x, y) \leq \frac{1}{|V(\Gamma)|} (1 + \varepsilon)$$

for all $x, y \in V(\Gamma)$.

We can give a crude (and well known; see, e.g., [BBS04, Fact 2.1]) upper bound on the $\ell_\infty$ mixing time for regular symmetric multigraphs in terms of $N = |V(\Gamma)|$, $\varepsilon$, and the valency $d$ alone.

**Lemma 4.1.** Let $\Gamma$ be a connected, regular, and symmetric multigraph of valency $d$ and with $N$ vertices. Then the $\ell_\infty$ mixing time for $\varepsilon$ is at most $N^2 d \log(N/\varepsilon)$.

**Proof.** Let $A$ be the adjacency matrix of $\Gamma$. Since $A$ is symmetric, the eigenvalues of $A$ are real; moreover, their modulus is clearly no more than $d$ in magnitude. Let

$$d = \mu_1 \geq \mu_2 \geq \cdots \geq \mu_N \geq -d$$

be the eigenvalues of $A$, and write $P = I/2 + A/2d$, where $I$ is the $N \times N$-identity matrix. The matrix $P$ is the probability transition matrix for the Markov process described by a lazy random walk on $\Gamma$.

The sum of every row or column of $P$ is 1; i.e., $P$ is a doubly stochastic matrix. The eigenvalues of $P$ are

$$1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0,$$

with $\lambda_i = 1/2 + \mu_i/2d$ for each $i = 1, \ldots, N$. It is well known that the asymptotic rate of convergence to the uniform distribution of a lazy random walk is determined by $\lambda_2$: since $P$ is symmetric, there is a basis of $\mathbb{R}^N$ consisting of orthogonal eigenvectors $v_1, v_2, \ldots, v_n$ of $P$ with eigenvalues $\lambda_1, \ldots, \lambda_N$, where every eigenvector $v_i$ has $\ell_2$-norm 1 with respect to (say) the counting measure; writing $e_x$ for the probability distribution having value 1 at $x$ and 0 elsewhere,
we see that (by Cauchy-Schwarz and Plancherel)
\[
\sum_{j=1}^N |\langle e_x, v_j \rangle| |\langle v_j, e_y \rangle| \leq \sqrt{\sum_{j=1}^N |\langle e_x, v_j \rangle|^2} \sqrt{\sum_{j=1}^N |\langle v_j, e_y \rangle|^2} \leq |e_x|_2 \cdot |e_y|_2 = 1
\]
and, since
\[
p_k(x, y) = \langle P^k e_x, e_y \rangle = \left( \sum_{j=1}^N \langle e_x, v_j \rangle \cdot P^k v_j, e_y \right) = \sum_{j=1}^N \langle e_x, v_j \rangle \cdot \lambda_j^k \langle v_j, e_y \rangle
\]
\[
= \frac{1}{\sqrt{N}} \cdot 1^k \cdot \frac{1}{\sqrt{N}} + \sum_{j=2}^N \langle e_x, v_j \rangle \cdot \lambda_j^k \langle v_j, e_y \rangle,
\]
we see that
\[
|p_k(x, y) - \frac{1}{N}| \leq \frac{\lambda_2^k}{N} \sum_{j=2}^N |\langle e_x, v_j \rangle| |\langle v_j, e_y \rangle| \leq \lambda_2^k.
\]

By [Fie72, Lemma 2.4 and Thm. 3.4], we have
\[
\lambda_2 \leq 1 - 2(1 - \cos(\pi/N))\mu(P),
\]
where \(\mu(P) = \min_{\emptyset \neq M \subseteq V} \sum_{i \in M \cap \bar{M}} p_{ij}\). As \(\Gamma\) is a connected regular graph of valency \(d\), we have \(\mu(P) \geq 1/2d\). Using the Taylor series for \(\cos(x)\), we see that \((1 - \cos(\pi/N)) \geq 1/N^2\). Thence \(|p_k(x, y) - 1/N| \leq (1 - 1/(N^2d))^k\). Since \(1 - x \leq e^{-x}\) for all \(x\), we obtain \(|p_k(x, y) - 1/N| \leq \varepsilon/N\) for \(k \geq N^2d \log(N/\varepsilon)\), as desired. \(\square\)

We will generally study regular symmetric multigraphs of the following type. (The following argument is already present in [BBS04, §2]; indeed, the only difference between Lemma 4.2 here and corresponding material in [BBS04, §2] is that Lemma 4.2 applies to ordered as opposed to unordered \(k\)-tuples.) Let \(G\) be a group and \(A\) be a subset of \(G\) with \(A = A^{-1}\) and \(e \in A\). Let \(G\) act on a set \(X\). We take the elements of \(X\) as the vertices of our multigraph and draw one edge from \(x \in X\) to \(x' \in X\) for every \(a \in A\) such that \(x^a = x'\). A walk on the graph then corresponds to the action of an element of \(A^\ell\) on an element \(x\) of \(X\), where \(\ell\) is the length of the walk and \(x\) is the starting point of the walk.

Lemma 4.1 then gives us a lower bound on how large \(\ell\) has to be for the action of \(A^\ell\) on \(X\) to have a rather strong randomising effect.

**Lemma 4.2.** Let \(H\) be a \(k\)-transitive subgroup of \(\text{Sym}([n])\). Let \(A\) be a set of generators of \(H\) with \(A = A^{-1}\) and \(e \in A\). Then there is a subset \(A' \subseteq A\) with \(A' = (A')^{-1}\) such that, for every \(\varepsilon > 0\), for any \(\ell \geq 2n^3k \log(n^3k/\varepsilon)\), and for any \(k\)-tuples \(x = (x_1, \ldots, x_k), y = (y_1, \ldots, y_k)\) of distinct elements of \([n]\), the probability of the event
\[
\overline{y} = \overline{x}^{g_1g_2 \cdots g_\ell}
\]
for \( g_1, \ldots, g_r \in A' \) (chosen independently, with uniform distribution on \( A' \setminus \{e\} \) and with the identity being assigned probability 1/2) is at least \((1 - \varepsilon)\frac{(n-k)!}{n!}\) and at most \((1 + \varepsilon)\frac{(n-k)!}{n!}\).

Proof. Let \( \Delta \) be the set of \( k \)-tuples of distinct elements of \([n]\). Since \( H \) acts transitively on \( \Delta \) and since \( \langle A \rangle = H \), Lemma 3.10 gives us a subset \( A' \) of \( A \) with \( \langle A' \rangle \) transitive on \( \Delta \) and with \(|A'| < |\Delta|\). Set \( A_0 = A' \cup A'^{-1} \). Let \( \Gamma \) be the multigraph with vertex set \( \Delta \) and with \( \Gamma(\pi) = \{\pi^a \mid a \in A_0\} \) as the multiset of neighbours of \( \pi \) for each \( \pi \in \Delta \). Clearly, \( \Gamma \) is a regular graph of valency \(|A_0| \leq 2|\Delta|\) and with \(|\Delta| \leq n^k\) vertices. Now the statement follows from Lemma 4.1 applied to \( \Gamma \). \qed

4.2. Generators. Given \( A \subseteq \text{Sym}([n]) \) such that \( \langle A \rangle = \text{Alt}([n]) \) or \( \text{Sym}([n]) \), how long can it take to construct a small set of generators for a transitive subgroup of \( \langle A \rangle \)? This subsection is devoted to answering that question. We start by proving two auxiliary lemmas.

Lemma 4.3. Let \( A \subseteq \text{Sym}([n]) \), \( e \in A \). Assume \( \langle A \rangle \) is transitive. Then there is a \( g \in A^n \) such that \(|\text{supp}(g)| \geq n/2\).

Proof. For each \( i \in [n] \), let \( g_i \) be an element of \( A \) moving \( i \). (If no such element existed, then \( \langle A \rangle \) could not be transitive.) Let \( g = g_1^{r_1}g_2^{r_2} \cdots g_n^{r_n} \), where \( r_1, r_2, \ldots, r_n \in \{0, 1\} \) are independent random variables taking the values 0 and 1 with equal probability.

Let \( \alpha \in [n] \) be arbitrary. Let \( j \) be the largest integer such that \( g_j \) moves \( \alpha \). Then \( g \) moves \( \alpha \) if and only if \( g' = g_1^{r_1} \cdots g_j^{r_j} \) moves \( \alpha \). Take \( r_1, r_2, \ldots, r_{j-1} \) as given. If \( \beta = g_1^{r_1} \cdots g_{j-1}^{r_{j-1}} \) equals \( \alpha \), then \( g' \) moves \( \alpha \) if and only if \( r_j = 1 \); this happens with probability 1/2. If \( \beta \neq \alpha \), then \( g' \) certainly moves \( \alpha \) if \( r_j = 0 \), and thus moves \( \alpha \) with probability at least 1/2. Thus \( g \) moves \( \alpha \) with probability at least 1/2. \qed

Summing over all \( \alpha \), we see that the expected value of the number of elements of \([n]\) moved by \( g \) is at least \( n/2 \). In particular, there is a \( g \in A^n \) moving at least \( n/2 \) elements of \([n]\). \(\)

\(^9\)Such an element \( g \) is called a random subproduct of the sequence \((g_i)\). This notion was introduced by [BLS88] in the context of the analysis of algorithms on permutation groups. See, e.g., [Ser03, §2.3] for other applications.

\(^{10}\)This argument essentially appears in [BLS88, §6.2] (without proof). It appears again, with proof and in a much more general context, in [BCF+91]. Indeed, Lemma 4.3 here follows immediately from [BCF+91, Lemma 2.2] (with \( K \) equal to a point stabiliser), and the idea of the proof of Lemma 4.3 given here is exactly the same as that of [BCF+91, Lemma 2.2]. We thank an anonymous referee for this remark.
The following is the simplest sphere-packing lower bound, applied to the Hamming distance. (The Hamming distance on \( \mathbb{0,1}^k \) is \( d(\vec{x}, \vec{y}) = |\{1 \leq j \leq k : x_j \neq y_j\}| \).)

**Lemma 4.4.** Let \( n > 0, k \geq 4.404 \log_2 n, \rho > 1. \) Let \( U = \{0,1\}^k \) be the set of \( \{0,1\} \)-sequences of length \( k \). Then there exists \( V \subseteq U, |V| > n \) such that any two sequences in \( V \) differ in more than \( \log_2 n \) coordinates.

**Proof.** In general, for \( U \) a metric space and \( V \subseteq U \) maximal with respect to the property that the distance between any two points of \( V \) is greater than \( r \), the closed balls of radius \( r \) around the points of \( V \) clearly cover \( U \); hence, if the notion of volume is well defined, \( |V| \) is at most \( \text{Vol}(U) \) divided by the volume of a closed ball of radius \( r \). Applying this to the Hamming distance, we obtain that, for \( V \subseteq U \) maximal,

\[
|V| \geq \frac{2^k}{\sum_{j=0}^{\lfloor r \rfloor} \binom{k}{j}}.
\]

By, e.g., [MS77, §10.11, Lemma 8],

\[
\sum_{j=0}^{\lfloor r \rfloor} \binom{k}{j} \leq 2^{kH(\lfloor r \rfloor/k)} \leq 2^{kH(r/k)}
\]

for \( 0 \leq r \leq k/2 \), where \( H(x) = -x \log_2 x - (1 - x) \log_2 (1 - x) \) is the binary entropy function. Let \( r = \log_2 n \). It is easy to check that, for \( 0 \leq \rho \leq 1/4.404 \), \( 1 - H(\rho) > \rho \). Hence

\[
|V| \geq 2^{k(1-H(r/k))} > 2^{k\cdot r/k} = 2^r = n.
\]

The following lemma is the main step toward answering the question raised at the beginning of the subsection. Most of the proof goes to show that, for some \( g \in A^n, h \in A^\ell \), and a random \( \beta \in [n] \), the orbit of \( \beta \) under \( \langle g, h \rangle \) is rather large. The following is a brief sketch. If \( \beta \) were being acted upon by many random elements of \( \text{Sym}([n]) \) in succession, it would indeed traverse many points. Now think of this obvious remark as being strengthened twice. First, let \( g \) have large support and let \( h \) be a random element of \( \text{Sym}([n]) \). If we let \( h \) act on \( \beta \) and then let \( g \) act (or not) on \( \beta^h \), and we let this happen over and over, the effect is a great deal as if \( \beta \) were being acted upon by random elements in succession: if \( \beta \) has arrived at a point \( x \) where it has not been before, then the random element \( h \) acts on it in a way that, as far as we are concerned, is essentially random, in that it is almost independent of any of the parts of \( h \) we have seen so far. This makes the action of the fixed element \( g \) on \( x^h \) itself random. Here comes the second strengthening: it is actually enough for \( h \) to be the outcome of a random walk of moderate length \( \ell \leq n^{O(\log n)} \); as we know (Lemma 4.2), such an \( h \) pretends to be a random element of \( \text{Sym}([n]) \).
very ably as far as its action on $k$ tuples, $k \ll \log n$, is concerned. These are all the tuples that we have to deal with, since the above argument gives us large orbits after $O(\log n)$ steps.

The proof below is just a detailed and rigorous version of this sketch.

**Lemma 4.5.** Let $A \subseteq \text{Sym}([n])$ with $A = A^{-1}$, $e \in A$, and $\langle A \rangle = \text{Sym}([n])$ or $\text{Alt}([n])$. Then there are $g \in A^n$, $h \in A^{[n^{27\log n}]}$ such that the action of $(g,h)$ on $[n]$ has at most $175(\log n)^2$ orbits, provided that $n$ is larger than an absolute constant.

**Proof.** We will show that, for some $g \in A^n$, for $h \in A^\ell$ ($\ell \leq [n^{27\log n}]$) taken randomly in a sense we will specify, and for any $\beta \in [n]$, the expected value of $1/|\beta(g,h)|$ is at most $175(\log n)^2/n$. (Here $\beta(g,h)$ denotes the orbit of $\beta$ under the action of $(g,h) \leq \text{Sym}([n])$.) Now, $\sum_{\beta \in [n]} 1/|\beta(g,h)|$ is just the number of orbits of $(g,h)$ (since each such orbit contributes $|\beta(g,h)|^{-1} = 1/n$ to the sum). Hence, by the additivity of expected values,

$$
\mathbb{E}(\text{number of orbits of } (g,h)) = \sum_{\beta \in [n]} \mathbb{E}\left(\frac{1}{|\beta(g,h)|}\right) \leq 175(\log n)^2.
$$

In particular, this will imply that there exists an $h \in A^\ell$ such that the number of orbits of $(g,h)$ is at most $175(\log n)^2$, and so we will be done.

Let $k = \lceil 4.404 \log_2 n \rceil$. By Lemma 4.3, there is an element $g \in A^n$ with

$$
|\text{supp}(g)| = \alpha n \geq n/2.
$$

Let $\varepsilon = 1/n$ and $\ell = [2n^{6k}\log(n^{2k}/\varepsilon)]$. (It is easy to check that, for $n$ larger than an absolute constant, $\ell \leq [n^{27\log n}]$.) Let $h \in A^\ell$ be the outcome of a random walk of length $\ell$ as in Lemma 4.2.

Consider all words of the form

$$
f(\bar{a}) = hg^{a_1}hg^{a_2} \cdots hg^{a_k},
$$

where $\bar{a} = (a_i : 1 \leq i \leq k)$ runs through all sequences in $U = \{0,1\}^k$. For $\beta \in [n]$, we wish to estimate $|\beta(g,h)|$ from below by counting the number of different images $f_\beta(\bar{a}) := \beta_{f(\bar{a})}$ for $\bar{a} \in U$.

To this end, for fixed elements $\bar{a} = (a_1, \ldots, a_k)$ and $\bar{a}' = (a'_1, \ldots, a'_k)$ in $U$ and $\beta \in [n]$, we wish to bound from above the probability that $f_\beta(\bar{a}) = f_\beta(\bar{a}')$. We will do this by examining all possible trajectories $(\beta_1, \ldots, \beta_k), (\beta'_1, \ldots, \beta'_k)$, where

$$
\beta_1 = \beta h g^{a_1}, \quad \beta_2 = \beta_1 h g^{a_2}, \ldots, \beta_k = \beta_{k-1} h g^{a_k} \quad \text{and} \quad \beta'_1 = \beta h g^{a'_1}, \ldots, \beta'_k = \beta_{k-1} h g^{a'_k},
$$

counting how many satisfy $\beta_k \neq \beta'_k$, and then estimating the probability (for $h$ chosen randomly in the manner described above) that such a pair of trajectories be traversed following $f(\bar{a})$ and $f(\bar{a}')$. 
Let $R = \{1 \leq i \leq k : a_i \neq a'_i\}$; let the elements of $R$ be $k_1 < k_2 < \cdots < k_r$, where $r = |R|$. Let $r_0 \leq r$ be fixed. Let $k' = k_{r_0}$. Consider all tuples $(\beta_1, \beta_2, \ldots, \beta_k, \beta'_{k'}, \ldots, \beta'_{k'}) \in [n]^{(2k-k') + 1}$ such that

(a) $\beta_1, \beta_2, \ldots, \beta_k, \beta'_{k'}, \ldots, \beta'_{k'}$ are distinct from each other and from $\beta$,
(b) $\beta_1^{g_{a_1}}, \beta_2^{g_{a_2}}, \ldots, \beta_k^{g_{a_k}}; (\beta'_{k'+1})^{g_{a'_{k'+1}}}, \ldots, (\beta'_{k'})^{g_{a'_{k'}}}$ are distinct from each other;
(c) $\beta_{kj} \notin \text{supp}(g)$ for every $j < r_0$, but $\beta_{k'} \in \text{supp}(g)$;
(d) $(\beta'_{k'})^{g_{a'_{k'}}} = (\beta'_{k'})^{g_{a'_{k'}}}$.

The number of such tuples is at least

$$
(1) \quad \left( \prod_{j=1}^{r_0-1} (n - |\text{supp}(g)| - j) \right) \cdot (|\text{supp}(g)| - 1) \cdot \prod_{j=(r_0+1)}^{2k-k'} (n - (2j - 1)),
$$

where we count tuples by choosing first $\beta_{kj} \in [n] \setminus \text{supp}(g)$ for $1 \leq j < r_0$, then $\beta_{k'} \in \text{supp}(g)$, then the other $\beta_i$ and $\beta'_i$. To justify the estimate on the number of choices at each stage, notice that at the $j$th choice with $j \leq r_0 - 1$, we have to make selections from $[n] \setminus \text{supp}(g)$ so as to satisfy (c) while keeping them different from previous selections and from $\beta$ (to satisfy (a)). Then $\beta_{k'}$ can be chosen as an arbitrary element of $\text{supp}(g)$ different from $\beta$. At this point, (b) is still satisfied automatically. At later choices, if $\beta_i$ or $\beta'_i$ is selected at stage $j$, then enforcing (a) eliminates $j$ possibilities and enforcing (b) eliminates $j - 1$, not necessarily different, possibilities. Note that (4.1) also gives a valid lower estimate (namely, 0) in the case when $r_0 - 1 \geq n - |\text{supp}(g)| > 0$. (The negative terms in the first product in (4.1) are made harmless by a term equal to 0.)

By Lemma 4.2 (with $2k - k'$ instead of $k$, and with properties (a), (b) as inputs), the probability that a random $h \in A^k$ satisfies

$$
(2) \quad (\beta, \beta_1, \ldots, \beta_{k-1}, \beta'_{k'}, \ldots, \beta'_{k-1})^h
$$

is at least $(1 - \varepsilon) \frac{(n-(2k-k'))!(n-2k-k')}{n^k}$. If $h$ satisfies (4.2), then $h^{g_{a_1}} = \beta_1$, $h^{g_{a_2}} = \beta_2, \ldots, h^{g_{a_k}} = \beta_k$. By properties (c) and (d), we also have $h^{g_{a'_1}} = \beta_1$, $h^{g_{a'_2}} = \beta_2, \ldots, h^{g_{a'_{k'}}} = \beta'_{k'}$. By properties (c) and (d), we also have $h^{g_{a'_{k'+1}}} = \beta'_{k'+1}, \ldots, h^{g_{a'_{k}}} = \beta'_{k'}.$. Thus, in particular, any two distinct tuples

$$
(\beta_1, \beta_2, \ldots, \beta_k, \beta'_{k'}, \ldots, \beta'_{k})
$$

give us mutually exclusive events, even for different values of $r_0$. Note also that, by property (a) and what we have just said, $f_\beta(\bar{a}) = \beta_k \neq \beta_k = f_\beta(\bar{a'}).$
Hence the probability $P$ that $f_\beta(\vec{a}) \neq f_\beta(\vec{a}')$ is at least

\[(4.3) \quad P \geq \sum_{r_0=1}^{r} \frac{1 - \varepsilon}{n^{2k-k_0}} \cdot \left( \prod_{j=1}^{r_0-1} (n - \alpha n - j) \right) \cdot (\alpha n - 1) \cdot \prod_{j=(r_0+1)}^{2k-k_0} (n - (2j - 1))
\]

\[> \sum_{r_0=1}^{r} \left( 1 - \frac{1}{n} \right) \left( 1 - \frac{4k}{n} \right)^{2k} \left( \alpha - \frac{1}{n} \right) \cdot \prod_{j=1}^{r_0-1} \left( 1 - \frac{\alpha - j}{n} \right).
\]

If $\alpha n = |\text{supp}(g)| \geq n - k$, then we estimate $P$ from below by the summand $r_0 = 1$ in (4.3), yielding

\[P > \left( 1 - \frac{1}{n} \right) \left( 1 - \frac{4k}{n} \right)^{2k} \left( \alpha - \frac{1}{n} \right) > 1 - \frac{1}{n} - \frac{8k^2}{n} - \frac{k + 1}{n} \geq 1 - \frac{9k^2}{n},
\]

with the last inequality valid for $n \geq 2$.

If $\alpha n = |\text{supp}(g)| < n - k$ then, estimating the terms $(1 - \alpha - j/n)$ in the last product in (4.3) from below by $(1 - \alpha - k/n)$, we obtain

\[(4.4) \quad P > \left( 1 - \frac{1}{n} \right) \left( 1 - \frac{4k}{n} \right)^{2k+1} \left( \alpha - \frac{1}{n} \right) \left( 1 - \frac{1 - \alpha - (k/n))}{1 - (1 - \alpha - (k/n))} \right)^r
\]

\[= \left( 1 - \frac{4k}{n} \right)^{2k+1} \left( \frac{\alpha - (1/n)}{\alpha + (k/n)} \right) \left( 1 - (1 - \alpha - (k/n)) \right)^r.
\]

Since $\alpha \geq 1/2$, we have $\frac{\alpha - (1/n)}{\alpha + (k/n)} \geq 1 - \frac{2(k+1)}{n}$ and $(1 - \alpha - (k/n))^r < (1/2)^r$, implying

\[P > 1 - \frac{4k(2k + 1)}{n} - \frac{2(k + 1)}{n} - \frac{1}{2^r} > 1 - \frac{9k^2}{n} - \frac{1}{2^r}
\]

if $n \geq 3$ (since then $k \geq 7$).

We conclude that, for any two nonidentical tuples,

\[\vec{a} = (a_1, \ldots, a_k) \in \{0, 1\}^k, \quad \vec{a}' = (a'_1, \ldots, a'_k) \in \{0, 1\}^k
\]

and for any $\beta \in [n]$,

\[\text{Prob}(\beta^{h_{g_1}h_{g_2}^1h_{g_3}^2\ldots h_{g_{\alpha k}}^k} = \beta^{h_{g_1'^1h_{g_2'^2}\ldots h_{g_{\alpha k}^k}'}}) < \frac{9k^2}{n} + \frac{1}{2d(\vec{a}, \vec{a}')}.
\]

where $d(\vec{a}, \vec{a}')$ is the Hamming distance between $\vec{a}$ and $\vec{a}'$, i.e., the number of indices $1 \leq j \leq k$ for which $a_j \neq a'_j$.

By Lemma 4.4, there exists a set $V$ of more than $n$ tuples so that any two tuples differ in more than $\log_2 n$ coordinates. For fixed $\beta \in [n]$, writing
\begin{align*}
f_\beta(\bar{a}) &= \beta h \bar{a} \gamma_1 h \bar{a}^2 \cdots h \bar{a}^k, \quad \bar{a} \in V \text{ for the random variable } \beta \mapsto f_\beta(\bar{a}) \text{ defined using a random } h \in A^t, \text{ we obtain that} \\
\mathbb{E}(\|\{ (\bar{a}, \bar{a}') \in V^2 : f_\beta(\bar{a}) = f_\beta(\bar{a}') \} \|) &= \sum_{\bar{a}, \bar{a}' \in V} \mathbb{P}(f_\beta(\bar{a}) = f_\beta(\bar{a}')) \\
&\leq |V| + \left( \frac{9k^2}{n} + \frac{1}{2d(\bar{a}, \bar{a}')} \right) |V| (|V| - 1) < \frac{|V|^2}{n} + \left( \frac{9k^2}{n} + \frac{1}{n} \right) |V|^2 \\
&< (9k^2 + 2) \frac{|V|^2}{n} < 175(\log n)^2 \frac{|V|^2}{n}
\end{align*}

for \( n \) larger than an absolute constant.

Concerning the length of the orbit \( \beta^{(g, h)} \), we have

\[
\mathbb{E} \left( \frac{1}{|\beta^{(g, h)}|} \right) \leq \mathbb{E} \left( \frac{1}{|\{ f_\beta(\bar{a}) : \bar{a} \in V \}|} \right) \leq \mathbb{E} \left( \frac{|\{ (\bar{a}, \bar{a}') \in V^2 : f_\beta(\bar{a}) = f_\beta(\bar{a}') \}|}{|V|^2} \right) \leq \frac{175(\log n)^2}{n},
\]

where we use Cauchy-Schwarz in the second step for the numbers \( m_i \) that measure how many times a particular value \( \gamma_i \) occurs among the \( f_\beta(\bar{a}) \), for some \( \bar{a} \in V \).

\textbf{Proposition 4.6.} Let \( A \subseteq \text{Sym}(\mathbb{N}) \) with \( A = A^{-1}, e \in A, \) and \( \langle A \rangle = \text{Sym}(\mathbb{N}) \) or \( \text{Alt}(\mathbb{N}) \). If \( n \) is larger than an absolute constant, then there are \( g_1, g_2, g_3 \in A^{\lfloor n^{2\log n} \rfloor} \) such that \( \langle g_1, g_2, g_3 \rangle \) is transitive.

\textbf{Proof.} Let \( g, h \) be as in \textbf{Lemma 4.5}. Let \( \varepsilon = 1/n^2, \ell = [2n^6 \log(n^2/\varepsilon)] \).

Let \( g' \in A^t \) be the outcome of a random walk of length \( \ell \) as in \textbf{Lemma 4.2}.

Note that \( \ell \leq [n^{2\log n}] \) for \( n \) larger than an absolute constant.

Let \( \Delta \) be the union of orbits of \( \langle g, h \rangle \) of length less than \( \sqrt{n} \). Since, by \textbf{Lemma 4.5}, there are at most \( 175(\log n)^2 \) orbits of \( \langle g, h \rangle \), we have \( |\Delta| < 175\sqrt{n}(\log n)^2 \). Let \( S \) be a set consisting of one element \( \alpha \) of each orbit of length less than \( \sqrt{n} \). Then, for each \( \alpha \in S \), \textbf{Lemma 4.2} implies that

\[
\mathbb{P}(\alpha^\ell \in \Delta) \leq (1 + \varepsilon) \frac{|\Delta|}{n} < \left( 1 + \frac{1}{n^2} \right) \frac{175(\log n)^2}{\sqrt{n}}
\]

and so

\begin{equation}
\mathbb{P} \left( (\exists \alpha \in S) (\alpha^\ell \in \Delta) \right) < \left( 1 + \frac{1}{n^2} \right) \frac{175^2(\log n)^4}{\sqrt{n}}.
\end{equation}

Let \( \kappa \) be an orbit of \( \langle g, h \rangle \) contained in \( n \setminus \Delta \); by definition, \( |\kappa| \geq \sqrt{n} \).

Let \( \kappa_0 \) be the largest orbit; by the pigeonhole principle, \( |\kappa_0| > n/(175(\log n)^2) \).

Then

\[
\mathbb{E}(|\kappa^\ell \cap \kappa_0|) = \sum_{\alpha \in \kappa} \mathbb{P}(\alpha^\ell \in \kappa_0) \geq \sum_{\alpha \in \kappa} (1 - \varepsilon) \frac{|\kappa_0|}{n} = (1 - \varepsilon) \frac{|\kappa||\kappa_0|}{n},
\]
whereas
\[
\mathbb{E} \left( |\kappa^g' \cap \kappa_0|^2 \right) = \sum_{\alpha, \beta \in \kappa} \text{Prob} \left( \alpha^g' \in \kappa_0 \land \beta^g' \in \kappa_0 \right)
\]
\[
= \sum_{\alpha \in \kappa} \text{Prob}(\alpha^g' \in \kappa_0) + \sum_{\alpha, \beta \in \kappa, \alpha' \neq \beta, \beta' \in \kappa_0} \text{Prob}((\alpha, \beta)^g' = (\alpha', \beta'))
\]
\[
\leq \sum_{\alpha \in \kappa} (1 + \varepsilon) \frac{|\kappa_0|}{n} + \sum_{\alpha, \beta \in \kappa, \alpha' \neq \beta} (1 + \varepsilon) \frac{|\kappa_0|(|\kappa_0| - 1)}{n(n - 1)}
\]
\[
\leq (1 + \varepsilon) \left( \frac{|\kappa_0||\kappa|}{n} + \frac{|\kappa|^2|\kappa_0|^2}{n^2} \right).
\]

Thus
\[
\text{Var}(\kappa^g' \cap \kappa_0) = \mathbb{E}(\kappa^g' \cap \kappa_0)^2 - \mathbb{E}(\kappa^g' \cap \kappa_0)^2
\]
\[
\leq (1 + \varepsilon) \left( \frac{|\kappa_0||\kappa|}{n} + \frac{|\kappa|^2|\kappa_0|^2}{n^2} \right) - (1 - \varepsilon)^2 \frac{|\kappa_0|^2|\kappa|^2}{n^2}
\]
\[
\leq 3\varepsilon \frac{|\kappa|^2|\kappa_0|^2}{n^2} + (1 + \varepsilon) \frac{|\kappa_0||\kappa|}{n} < \left( 1 + \frac{4}{n} \right) \frac{|\kappa_0||\kappa|}{n}.
\]

By Chebyshev’s inequality,
\[
\text{Prob}(\kappa^g' \cap \kappa_0 = \emptyset) \leq \frac{\text{Var}(\kappa^g' \cap \kappa_0)}{\mathbb{E}(\kappa^g' \cap \kappa_0)^2}
\]
\[
\leq \frac{|\kappa||\kappa_0|/n(1 + 4/n)}{(1 - \varepsilon)^2 \frac{|\kappa|^2|\kappa_0|^2}{n^2}} \leq \frac{12n}{|\kappa||\kappa_0|} < \frac{12 \cdot 175(\log n)^2}{\sqrt{n}}.
\]

Hence
\[
(4.6) \quad \text{Prob} \left( (\exists \kappa \subseteq ([n] \setminus \Delta)) (\kappa^g' \cap \kappa_0 = \emptyset) \right) < \frac{12 \cdot 175^2(\log n)^4}{\sqrt{n}}.
\]

Now, for \( n \) larger than a constant,
\[
\left( 1 + \frac{1}{n^2} \right) \frac{175^2(\log n)^4}{\sqrt{n}} + \frac{12 \cdot 175^2(\log n)^4}{\sqrt{n}} < 1.
\]

Therefore, (4.5) and (4.6) imply that with positive probability, (a) \( \kappa^g' \) intersects \([n] \setminus \Delta\) for every orbit \( \kappa \) not contained in \([n] \setminus \Delta\), and (b) \( \kappa^g' \) intersects \( \kappa_0 \) for every orbit \( \kappa \) contained in \([n] \setminus \Delta\). In particular, this happens for some \( g' \in A^\ell \). Properties (a) and (b) imply that \( \langle g, h, g' \rangle \) is transitive. We set \( g_1 = g, g_2 = h, g_3 = g' \) and are done. \( \square \)
We will later use\footnote{If we wished to, we could use it to obtain a set \( S \) of generators of \( \operatorname{Alt}([n]) \) or \( \operatorname{Sym}([n]) \) simply by setting \( k = 6 \): the Classification of Finite Simple Groups implies that a 6-transitive group must be either alternating or symmetric.} the following corollary with \( k = 2 \).

**Corollary 4.7.** Let \( A \subseteq \operatorname{Sym}([n]) \) with \( A = A^{-1} \), \( e \in A \), and \( \langle A \rangle = \operatorname{Sym}([n]) \) or \( \operatorname{Alt}([n]) \). Let \( k \geq 1 \). If \( n \) is larger than a constant depending only on \( k \), then there is a set \( S \subseteq A^{\lfloor n^{28 \log n} \rfloor} \) of size at most \( 3k \) such that \( \langle S \rangle \) is \( k \)-transitive.

**Proof.** Let \( \alpha_1 \in [n] \) be arbitrary. Since \( \langle A \rangle \) is transitive, Lemma 3.9 implies that \( \alpha_1^{A^n} = [n] \). Let \( G = \operatorname{Sym}([n]) \), \( H = G_{\alpha_1} \), \( A' = A^n \). Since \( \alpha_1^{A^n} = [n] \), \( A' \) intersects every coset of \( H \) in \( G \). By Schreier’s Lemma (Lemma 3.8), it follows that \( (A')^3 \cap H \) generates \( \langle A \rangle \cap H \), which is either \( \operatorname{Sym}([n] \setminus \{\alpha_1\}) \) or \( \operatorname{Alt}([n] \setminus \{\alpha_1\}) \). Let \( A_1 = (A')^3 \cap H \).

Iterating, we obtain a sequence of sets \( A_0 = A, A_1, A_2, \ldots, A_{k-1} \subseteq \operatorname{Sym}([n]) \) and a sequence of elements \( \alpha_1, \alpha_2, \ldots, \alpha_{k-1} \in [n] \) such that \( A_i \subseteq A_i^{2^{27 \log n}} \) and \( \langle A_i \rangle \) is \( \operatorname{Sym}([n] \setminus \{\alpha_1, \ldots, \alpha_i\}) \) or \( \operatorname{Alt}([n] \setminus \{\alpha_1, \ldots, \alpha_i\}) \).

Let \((g_1)_i, (g_2)_i, (g_3)_i \) be as in Proposition 4.6, applied with \( A_i \) instead of \( A \). Then \((g_1)_i, (g_2)_i, (g_3)_i \in A_i^{2^{27 \log n}} \) and \( \langle (g_1)_i, (g_2)_i, (g_3)_i \rangle \subseteq \operatorname{Sym}([n] \setminus \{\alpha_1, \ldots, \alpha_i\}) \) is transitive on \( [n] \setminus \{\alpha_1, \ldots, \alpha_i\} \) for \( 0 \leq i \leq k - 1 \). Thus, for \( S = \bigcup_{i=0}^{k-1} A_i \), \( \langle S \rangle \) is \( k \)-transitive on \( [n] \). \( \square \)

5. The splitting lemma and its consequences

We will prove what is in effect an adaptation of Babai’s splitting lemma (proven for groups in [Bab82, Lemma 3.1]) to the case of sets. This is a key point in this paper: the splitting lemma will allow us to construct long stabiliser chains with large orbits.

The following easy lemma will make an “unfolding” step possible.

**Lemma 5.1.** Let \( A \subseteq \operatorname{Sym}([n]) \), \( \Sigma \subseteq [n] \), and \( g \in \operatorname{Sym}([n]) \). Then

\[
gA_{\langle \Sigma^g \rangle}g^{-1} = (gAg^{-1})_{\langle \Sigma \rangle}.
\]

**Proof.** We have \( \operatorname{Sym}([n])_{\langle \Sigma^g \rangle} = g^{-1} \operatorname{Sym}([n])_{\langle \Sigma \rangle}g \). Therefore,

\[
A_{\langle \Sigma^g \rangle} = A \cap \operatorname{Sym}([n])_{\langle \Sigma^g \rangle} = A \cap g^{-1} \operatorname{Sym}([n])_{\langle \Sigma \rangle}g = g^{-1}(gAg^{-1} \cap \operatorname{Sym}([n])_{\langle \Sigma \rangle})g = g^{-1}(gAg^{-1})_{\langle \Sigma \rangle}g.
\]

Notice a feature of the following statement — there is a high power of \( A \) in the assumptions, not just in the conclusion. We will “unfold” the high power of \( A \) in the course of the proof. (By \( \Sigma^S \) we mean the set \( \Sigma^S = \{\alpha^S : \alpha \in \Sigma, g \in S\} \).)
Proposition 5.2 (Splitting Lemma). Let $A \subseteq \text{Sym}([n])$ with $A = A^{-1}$, $e \in A$, and $(A)$ 2-transitive. Let $\Sigma \subseteq [n]$. Assume that there are at least $pn(n-1)$ ordered pairs $(\alpha, \beta)$ of distinct elements of $[n]$ such that there is no $g \in (A^{\lfloor 9n^6 \log n \rfloor})_{\Sigma}$ with $\alpha^g = \beta$. Then there is a subset $S$ of $A^{\lfloor 5n^6 \log n \rfloor}$ with

$$(AA^{-1})_{\Sigma S} = \{e\}$$

and

$$|S| \leq \left\lceil \frac{2}{\log(3/(3-2\rho))} \cdot \log n \right\rceil.$$

Proof. Set $\ell = \lceil 2n^6 \log(n^2/(1/3)) \rceil$; note that $\ell \leq \lfloor 5n^6 \log n \rfloor$ and $2\ell + 2 \leq \lfloor 9n^6 \log n \rfloor$ for $n \geq 5$. (For $n < 5$, the statement is trivial.) By Lemma 4.2 applied with $k = 2$ and $\varepsilon = 1/3$, we obtain that given any two distinct elements $\alpha, \beta \in [n]$ and $g \in A^\ell$, the pair $(\alpha^g, \beta^g)$ adopts any possible value $(\alpha', \beta')$ with probability at least $(1 - 1/3)/(n(n-1))$, where we choose $g \in A^\ell$ with the distribution in Lemma 4.2 ($g = g_1 g_2 \cdots g_\ell$, $g_i$ chosen independently from $A' \cup \{e\}$, where $A'$ is a symmetric subset of $A$). Since this distribution is symmetric, this is the same as saying that $(\alpha^{g^{-1}}, \beta^{g^{-1}})$ adopts any possible value $(\alpha', \beta')$ with probability at least $(1 - 1/3)/(n(n-1))$.

Now, given $(\alpha, \beta)$ and $g \in A^\ell$, we have $h \in (AA^{-1})_{\Sigma g}$ and $\alpha^h = \beta$ if and only if $ghg^{-1} \in g(AA^{-1})_{\Sigma g} g^{-1}$ and $(\alpha^{g^{-1}}) g^{ghg^{-1}} = \beta^{g^{-1}}$. By Lemma 5.1 applied to $AA^{-1}$, we have that $ghg^{-1} \in (gAA^{-1})_{\Sigma g} g^{-1}$ only if $ghg^{-1} \in (gAA^{-1})_{\Sigma g} g^{-1}$, which in turn can happen only if $ghg^{-1} \in (A^{2\ell+2})_{\Sigma g}$. Thus, if there is no element $j \in (A^{2\ell+2})_{\Sigma g}$ with $\alpha^{g^{-1}j} = \beta^{g^{-1}}$, then there is no element $h \in (AA^{-1})_{\Sigma g}$ with $\alpha^h = \beta$. (This is the “unfolding” step we referred to before.)

Since by hypothesis there are at least $pn(n-1)$ ordered pairs $(\alpha', \beta')$ such that there is no element $j \in (A^{2\ell+2})_{\Sigma g}$ with $\alpha^{gj} = \beta'$, and since $(\alpha^{g^{-1}}, \beta^{g^{-1}})$ equals any such pair with probability at least $(2/3)/(n(n-1))$, we see that the probability that there is no element $h \in (AA^{-1})_{\Sigma g}$ with $\alpha^h = \beta$ is at least $2\rho/3$.

Let $S$ be a set of $r$ random $g \in A^\ell$ (chosen independently, with the distribution as above). The probability that for every $g \in S$ there is an element $h \in (AA^{-1})_{\Sigma g}$ with $\alpha^h = \beta$ is at most $(1 - 2\rho/3)^r$. This must happen if there is an element $h \in (AA^{-1})_{\Sigma S}$ such that $\alpha^h = \beta$. Thus, the probability that there is such an $h$ is at most $(1 - 2\rho/3)^r$, and the probability that there is such an $h$ for at least one of the $n(n-1)$ pairs $(\alpha, \beta)$ is at most $n(n-1)(1 - 2\rho/3)^r$.

Setting $r = \lfloor (\log n^2)/(\log 3/(3-2\rho)) \rfloor$, we obtain that the probability that there is such an $h$ for at least one pair is less than 1. Hence there is a set $S \subseteq A^\ell$ with at most $r$ elements such that, for every pair $(\alpha, \beta)$ of distinct elements of
[n], there is no h ∈ (AA\(^{-1}\))(Σ) with α\(^h\) = β. This implies immediately that the only element of (AA\(^{-1}\))(Σ) is the identity.

**Corollary 5.3.** Let A ⊆ Sym([n]) with A = A\(^{-1}\), e ∈ A, and (A) 2-transitive. Let A′ = A\([^{6n^6\log n}]\). Let Σ ⊆ [n] be such that

\[|\alpha_{A(\Sigma)}^n| < (1 - \rho)n\]

for every α ∈ [n], where ρ ∈ (0, 1). Then

\[|\Sigma| > \frac{\log |A|}{\log(3/(3 - 2\rho)) \cdot \log n} \cdot \log n.\]

In particular, if ρ = 0.05, then |Σ| > (log |A|)/(60(\log n)^2).

**Proof.** Since |\alpha_{A(\Sigma)}^n| < (1 - \rho)n for every α ∈ [n], there are at least \(\rho n(n - 1)\) pairs (α, β) such that there is no g ∈ A′(Σ) with α\(^g\) = β. By Proposition 5.2, there is a set S ⊆ Sym([n]) such that (AA\(^{-1}\))(Σ) = {e} and |S| ≤ \(\frac{2}{\log(3/(3 - 2\rho))} \cdot \log n\). Since (AA\(^{-1}\))(Σ) = {e}, we know, by Lemma 3.16, that |Σ| ≥ log \(|A|\). Clearly |Σ| = |S||Σ| ≤ |S||Σ|. Hence

\[|\Sigma| ≥ \frac{\log |A|}{|S|} ≥ \frac{\log |A|}{\frac{2}{\log(3/(3 - 2\rho))} \cdot \log n} \cdot \log n.\]

A key idea in the proof of the Main Theorem is the following. For A ⊆ Sym([n]), we can construct A′ = A\([^{5n^6\log n}]\) and a set Σ = {α₁, α₂, ...} ⊆ [n] starting with an empty set and taking at each step α\(_i\) to be an element such that |\alpha_{A(\Sigma)}^i(α_1, ..., α_{i-1})| ≥ (1 - \rho)n (say); if no such element exists, we stop the procedure. By Corollary 5.3, |Σ| must be large.

An application of Lemma 3.18 will give that, for A″ = (A′)\([^{16n^6}]\), the set (A″)Σ contains a copy of Alt(Δ), where Δ ⊆ Σ and |Δ| ≥ (1 - \rho)|Σ|. Such a large alternating group certainly looks like a valuable tool.

### 6. Proof of the Main Theorem

The core of this section is Proposition 6.4. It is a growth result, but not quite of type |A · A · A| ≥ |A|\(^{1+\epsilon}\) or |A\(^k\)| ≥ |A|\(^{1+\epsilon}\). What will grow by a factor at each step is not the number of elements |A| of A, but rather the length m of a sequence α₁, ..., α\(_m\) such that the orbits

\[\alpha_1^A, \alpha_2^A(\alpha_1), \alpha_3^{A(\alpha_1, \alpha_2)}, ..., \alpha_m^{A(\alpha_1, \alpha_2, ..., \alpha_{m-1})}\]

are all large.

This growth result (Proposition 6.4) will be applied iteratively. There are two ways for the iteration to stop: (a) an element we construct could fix a large set pointwise (we call this the case of exit), or (b) a group we work with
could fail to have a large alternating composition factor. In case (a), we obtain all of $G = \text{Alt}(|n|)$ in a few steps by Theorem 1.4. In case (b), we can descend to the problem of proving small diameter for $n'$ smaller than $n$ by a constant factor. (Here, as in “infinite descent,” the term “descent” means the same as induction, seen backwards.)

* * *

Let us sketch briefly the proof of Proposition 6.4. First, we use (6.4) to construct many elements in the setwise stabiliser $G_{\Sigma}$, where $\Sigma = \{\alpha_1, \ldots, \alpha_m\}$; in fact we get an entire copy of a large alternating group in $(G_{\Sigma})_|\Gamma|$ (Lemma 3.18). This is the setup. Then comes the creation step: we use the action by conjugation of $G_{\Sigma}$ on the pointwise stabiliser $G_{(\Sigma)}$ to construct many elements of $G_{(\Sigma)}$ (Lemma 6.1). We organise these new elements (all in a power $A'$ of $A$) as follows: we apply Corollary 5.3 (a consequence of the splitting lemma) to lengthen our stabiliser chain $A' \supseteq A'_{\alpha_1} \supseteq \cdots \supseteq A'_{(\alpha_1, \ldots, \alpha_m)} \supseteq \cdots$ up to $A'_{(\alpha_1, \ldots, \alpha_{m+1})}$ in such a way that the orbits (defined as in (6.1)) are still large. We repeat the organiser step about $\gg (\log n)/(\log m)$ times. There are only two ways for this procedure to stop prematurely, namely, exit and descent (cases (a) and (b) discussed above).

* * *

We start by proving the lemma containing the creation step: we give a way to construct many elements in a subgroup $H^-$ of a group $G$. The basic idea is the application of the orbit-stabiliser principle to the action by conjugation of a subgroup $H^+ \leq N_G(H^-)$ on $H^-$, where $N_G(H^-)$ is the normaliser of $H^-$. 

**Lemma 6.1.** Let $G = \text{Sym}(|n|)$ or $\text{Alt}(|n|)$, $H^- \leq G$, $H^+ \leq N_G(H^-)$, $\Gamma$ an orbit of both $H^-$ and $H^+$. Let $Y = \{y_1, \ldots, y_r\} \subseteq H^-$ be such that $\langle Y \rangle |\Gamma$ is 2-transitive on $\Gamma$. Let $B \subseteq H^+$. Then either

(a) there is a $b \in BB^{-1} \setminus \{e\}$ fixing $\Gamma$ pointwise, or
(b) $|B^{-1}YB \cap H^-| \geq |B|^{1/r}$.

**Proof.** Consider the action of $B$ on $\bar{y} = (y_1, \ldots, y_r)$ by conjugation: for $b \in B$, we define $\bar{y}^b := (y_1^b, \ldots, y_r^b)$, where $y_i^b = b^{-1}yb$. Assume first that there are two distinct elements $b_1, b_2 \in B$ such that $\bar{y}^{b_1} |\Gamma = \bar{y}^{b_2} |\Gamma$. Then $b_1b_2^{-1} |\Gamma$ centralises $\bar{y} |\Gamma$, implying that $b_1b_2^{-1} |\Gamma \in C(\langle Y \rangle |\Gamma) = \{e\}$. (As is well known and can be easily seen, the centraliser of a doubly transitive group, such as $\langle Y \rangle |\Gamma < \text{Sym}(\Gamma)$, is trivial.) Hence $b_1b_2^{-1} \in B$ fixes $\Gamma$ pointwise without being the identity, i.e., conclusion (a) holds.

Assume now that the restrictions $\bar{y}^b |\Gamma$ are all distinct. Hence, by the pigeonhole principle, there exists an index $j \in \{1, \ldots, r\}$ such that the set $W$ of conjugates of $y_j$ by $B$ satisfies $|W |\Gamma| \geq |B|^{1/r}$. Observe that all elements of
W are in \( H^- \), as \( Y \subset H^- \) and \( B \subset N_G(H^-) \). Hence \(|B^{-1}YB \cap H^-| \geq |W| \geq |B|^{1/r}\).

The following useful lemma is in part an easy application of Schreier’s lemma and in part a consequence of a trick based on the following trivial fact: one clearly cannot have two disjoint copies within \([n]\) of an orbit of size greater than \(n/2\).

**Lemma 6.2.** Let \( \Delta \subseteq [n] \). Let \( B^+ \subseteq \langle \text{Sym}(\Delta) \rangle \) with \( B^+ = (B^+)^{-1} \), \( e \in B^+ \). Assume \( B^+|_\Delta \) is \( \text{Alt}(\Delta) \) or \( \text{Sym}(\Delta) \). Let \( B^- = ((B^+)^3|_\Delta) \).

Then \( \langle B^- \rangle = \langle B^+ \rangle|_\Delta \triangleleft \langle B^+ \rangle \). Furthermore, if \( \langle B^- \rangle \) has an orbit \( \Gamma \) of length greater than \( n/2 \), then \( \Gamma \) is also an orbit of \( \langle B^+ \rangle \).

**Proof.** Since \( B^+|_\Delta \) is a group \( \langle \text{Alt}(\Delta) \rangle \) or \( \langle \text{Sym}(\Delta) \rangle \), \( B^+|_\Delta = \langle B^+ \rangle|_\Delta \). Thus \( B^+ \) contains an element from every coset of \( \langle B^+ \rangle|_\Delta \) in \( \langle B^+ \rangle \) and so, by Lemma 3.8, \( B^- \) contains a set of generators of \( \langle B^+ \rangle|_\Delta \). Hence \( \langle B^- \rangle = \langle B^+ \rangle|_\Delta \). In particular, \( \langle B^- \rangle \triangleleft \langle B^+ \rangle \), as \( \langle B^+ \rangle|_\Delta \) is the kernel of the action of \( \langle B^+ \rangle \) on \( \Delta \).

The orbits of the normal subgroup \( \langle B^- \rangle \) of \( \langle B^+ \rangle \) are blocks of imprimitivity for \( \langle B^+ \rangle \). Since one cannot have two blocks of length greater than \( n/2 \), \( \langle B^+ \rangle \) leaves \( \Gamma \) invariant as a set, and so \( \Gamma \) is also an orbit of \( \langle B^+ \rangle \). \( \square \)

The following lemma is also crucial to the descent step. In the proof of the lemma, we use Lemma 3.19 to guarantee the existence of an element that we then construct by other means.

**Lemma 6.3.** Let \( G = \text{Sym}([n]) \) or \( \text{Alt}([n]) \). Let \( \Delta \subseteq [n] \), \( |\Delta| \geq (\log n)^2 \). Let \( A \subseteq G \) with \( A = A^{-1} \), \( e \in A \) and \( \langle A \rangle = G \). Let \( B^+ \subseteq (A^l|_\Delta \), \( l \geq 1 \), with \( B^+ = (B^+)^{-1} \), \( e \in B^+ \). Assume \( B^+|_\Delta \) is \( \text{Alt}(\Delta) \) or \( \text{Sym}(\Delta) \). Let \( B^- = ((B^+)^3|_\Delta) \). Assume \( \langle B^- \rangle \) has an orbit \( \Gamma \) of length at least \( \rho n \), for some \( \rho > 8/9 \).

If all alternating composition factors \( \text{Alt}(k) \) of \( \langle B^- \rangle \) satisfy \( k \leq \delta n \), where \( \delta > 0 \), and

\[
(6.2) \quad \max_{k \leq \delta n} \text{diam}(\text{Alt}(k)) \leq D_\delta,
\]

for some \( D_\delta > 0 \), and \( n \) is larger than an absolute constant, then

\[
A|_{le^{c(\log n)^3}\cdot D_\delta} \supseteq \text{Alt}([n]),
\]

where \( c = c(\rho) \) depends only on \( \rho \).

**Proof.** The group \( U := \langle B^- \rangle|_\Gamma \) is transitive. It is also isomorphic to a quotient of \( \langle B^- \rangle \), so \( U \) also has no alternating composition factors \( \text{Alt}(k) \) with \( k > \delta n \). By Theorem 1.1 and by (6.2), there exists an absolute constant \( C_1 \).
such that for
\begin{equation}
(6.3) \quad u := |e^{C_1(\log n)^3} \cdot D_3|, \quad (B^-)^u |_{\Gamma} = U.
\end{equation}
Let $H = \langle B^+ \rangle$. By Lemma 6.2, $\Gamma$ is an orbit of $H$. If $n$ is large enough that Lemma 3.19 applies then there exists a nonidentity element $g \in H$ of support less than $|\Gamma|/4$ on $\Gamma$. Take $h \in B^+$ with $h|_\Delta = g|_\Delta$. Then $gh^{-1} \in \langle B^+ \rangle(\Delta) = \langle B^- \rangle$ and so, by (6.3), there exists $b \in (B^-)^u$ with $gh^{-1}|_{\Gamma} = b|_{\Gamma}$. Therefore, $bh \in (B^+)^{3u+1}$ satisfies $bh|_{\Gamma} = g|_{\Gamma}$. Since $g$ fixes at least $(3/4)|\Gamma| \geq (3/4) \cdot pn > (2/3)n$ points in $\Gamma$, we have $|\text{supp}(bh)| \leq (1 - (3/4)\rho)n < n/3$. By Theorem 1.4, $(A \cup \{bh, (bh)^{-1}\})^{Kn^8}$ contains $\text{Alt}([n])$, where $K = K(\varepsilon)$ ($\varepsilon = 1 - (3/4)\rho < 1/3$) is the number defined in Theorem 1.4. Since $A \cup \{bh, (bh)^{-1}\} \subseteq A^{3u+1}$, we are done.

We come to the key results in the paper. They will be given as two separate propositions, proved by a back-and-forth inductive process. For the sake of clarity, we will state them in terms of functions $F_1, F_2 : \mathbb{R}^+ \to \mathbb{R}^+$ obeying certain relations; we will later specify functions satisfying these relations.

**Proposition 6.4.** Let $G = \text{Sym}([n])$ or $\text{Alt}([n])$. Let $A \subseteq G$ with $A = A^{-1}$, $e \in A$, and $\langle A \rangle = G$. Let $\alpha_1, \alpha_2, \ldots, \alpha_{m+1} \in [n]$ be such that
\begin{equation}
(6.4) \quad \left| A_{\alpha_1, \ldots, \alpha_{m+1}} \right| \geq \frac{9}{10} n
\end{equation}
for every $i = 1, 2, \ldots, m+1$, where $m \geq (\log n)^2$.

There are absolute constants $n_0 \in \mathbb{Z}^+$ and $K, c_1, c_2, c_3 > 0$ such that the following holds. Assume $n \geq n_0$. Assume also that Proposition 6.5 holds for all smaller values of $n$ with respect to some increasing function $F_2 : \mathbb{R}^+ \to \mathbb{R}^+$. Let $F_1 : \mathbb{R}^+ \to \mathbb{R}^+$ be such that, for all $n \in \mathbb{Z}^+$,
\begin{equation}
(6.5) \quad F_1(n) \geq \max \left( n^{c_3 \log n} e^{c_1(\log n)^3} F_2(0.95n), 2Kn^{c_3 \log n + 8} \right).
\end{equation}
Then either
\begin{equation}
(6.6) \quad A_{F_1(n)} \supseteq \text{Alt}([n]),
\end{equation}
or there are $\alpha_{m+2}, \alpha_{m+3}, \ldots, \alpha_{m+l+1} \in [n]$, $l \geq c_2(m \log m)/(\log n)$, such that
\begin{equation}
(6.7) \quad \left| A_{\alpha_1, \ldots, \alpha_l, \alpha_{m+1}} \right| \geq \frac{9}{10} n
\end{equation}
for $A' = A^{n^{c_3 \log n}}$ and every $i = 1, 2, \ldots, m + l + 1$.

An easy application of Proposition 6.4 proves Proposition 6.5 (which is equivalent to our Main Theorem). Conversely, in order to prove Proposition 6.4, we will use Proposition 6.5 for smaller values of $n$ in an inductive process. In the proofs of Propositions 6.4 and 6.5, we assume that $n$ is greater than a well-defined (but not explicitly computed) absolute constant $n_0$; we take
n_0 to be large enough to satisfy the assumptions made in the course of both proofs. In the statement of Proposition 6.4, the assumption is made explicitly; in the statement of Proposition 6.5, the assumption is allowed by (6.8), which implies that, when n \leq n_0, the bound diam(\Gamma(G,Y)) \leq F_2(n) is trivial and there is nothing to prove.

Proposition 6.5. Let G = Sym([n]) or Alt([n]). Let Y \subseteq G with Y = Y^{-1}, e \in Y and G = \langle Y \rangle. Assume Proposition 6.4 holds for n with respect to some function F_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+. Let c_2 and c_3 be the absolute constants in the statement of Proposition 6.4; let n_0 be at least as large as in Proposition 6.4. Let F_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+ be such that (6.8)

\[ F_2(n) \geq \max \left( e^{(\log n)^3 + 2 \log n + c'_3 \log \log n} F_1(n) + 2, n_0! \right) \]

for some c' > c_2 and all n \in \mathbb{Z}^+. Then

\[ \text{diam}(\Gamma(G,Y)) \leq F_2(n), \]

provided that n_0 is larger than a constant depending only on c_2 and c'.

The proof consists just of a repeated use of Proposition 6.4, plus some accounting.

Proof. We can assume that n is large enough that m_0 \leq 0.1n \leq n - 3 for m_0 = \lfloor (\log n)^2 \rfloor + 1 and so G acts transitively on the set X of all (m_0 + 1)-tuples. Hence, by Lemma 3.9, the set A_0 := Y^{nm_0+1} \supseteq Y^{X} acts transitively on the set of all (m_0 + 1)-tuples. Thus (6.4) holds with A_0 instead of A, m_0 instead of m, and \alpha_i = i for i = 1, 2, \ldots, m_0 + 1. We apply Proposition 6.4 with these parameters, assuming n \geq n_0, where n_0 is the absolute constant in the statement of Proposition 6.4. We obtain either (6.6) or (6.7).

In the latter case, we set \ell_0 = \ell, m_1 = m_0 + \ell_0, and iterate: we apply Proposition 6.4 to

\[ A_1 = A_0^r, \quad A_2 = A_1^r = A_0^{r^2}, \quad A_3 = A_2^r = A_0^{r^3}, \ldots \]

where \( r = \lfloor n^{c_3 \log n} \rfloor \). (After each step, we “save” the output \ell to \ell_i and set \( m_{i+1} = m_i + \ell_i \).) We stop when we obtain (6.6); say this happens when we apply Proposition 6.4 with A = A_k = A_0^{r^k}.

It remains to estimate k. By Proposition 6.4,

\[ m_{i+1} \geq (1 + (c_2 \log m_i) / (\log n)) \cdot m_i. \]

We want to compute how many times we have to iterate (6.9) before we run into a contradiction with \( m_i \leq n \).

For 1 \leq j \leq \log n, let \( t_j \) be the largest index i between 0 and k such that \( m_i < e^j \); if no such index exists, set \( t_j = 1 \). We have \( m_0 \geq 3 \) and so
t_1 = 1. By (6.9) and \((1 + c_2 j / (\log n))^{(\log n) / (c_2 j)} + 2 > e\), we have \(t_{j+1} \leq t_j + [(\log n) / (c_2 j)] + 3\). Thus

\[
t_{[\log n]} + 1 \leq t_1 + 1 + \sum_{j=1}^{[\log n]} (t_{j+1} - t_j) \\
\leq 2 + \sum_{1 \leq j \leq \log n} \left( \frac{\log n}{c_2 j} + 3 \right) \leq c' \log n \log \log n
\]

for any \(c' > 1/c_2\), with the last inequality valid if \(n\) is larger than a constant depending only on \(c\) and \(c'\). Since \(t_{[\log n]} + 2 > k\) (because \(m_k \leq n\), we get that \(k \leq c' \log n \log \log n\).

Thus

\[
A_k = A_0^k \subseteq Y_{n^{(\log n)^2 j + 2, \rho}} \subseteq Y_{n^{(\log n)^3 + 2 \log n + c' \cdot 3 \cdot \log \log n}}.
\]

Then, by (6.6) (valid for \(A = A_k\)), we obtain

\[
\text{Alt}([n]) \subseteq (Y_{e^{(\log n)^3 + 2 \log n + c' \cdot 3 \cdot \log \log n}})^{[F_1(n)]} \subseteq Y^{[F_2(n)]} - 1
\]

for \(n\) larger than a constant. If \(Y \subseteq \text{Alt}([n])\), then \(Y^{[F_2(n)]} - 1 = \text{Alt}([n])\). If \(Y\) contains an odd permutation, then \(Y^{[F_2(n)]} = \text{Sym}([n])\). \(\Box\)

We finally turn to the proof of Proposition 6.4.

Proof of Proposition 6.4. We can assume that \(n\) is large enough that \(m \geq (\log n)^2 > C(0.9)\), where \(C(0.9)\) is as in Lemma 3.18. Apply Lemma 3.18 with \(d = 0.9\) and \(\Sigma = \{\alpha_1, \ldots, \alpha_m\}\). We obtain a set \(\Delta \subseteq \Sigma\) such that \(|\Delta| \geq 0.9|\Sigma|\) and \((A^{16m^6})^{\langle |\Sigma| \Delta \rangle} \Delta\) contains \(\text{Alt}(\Delta)\). Let

\[
B^+ = \left\{ g \in \left( A^{16m^6} \right)^{\langle |\Sigma| \Delta \rangle} : g|\Delta \in \text{Alt}(\Delta) \right\}, \quad B^- = \left( (B^+)^3 \right)^{\langle |\Sigma| \Delta \rangle}.
\]

This is our initial setup: we have a large set \(B^+\) in the setwise stabiliser \(G_{\Sigma}\); furthermore, we have constructed a large subset \(\Delta \subseteq \Sigma\) such that \(B^+ \subseteq (G_{\Sigma})^{\langle |\Sigma| \Delta \rangle}\) and \(B^+|\Delta = \text{Alt}(\Delta)\). We also have a set \(B^-\) in the pointwise stabiliser \(G_{\Sigma}\). By (6.4) with \(i = m + 1\), \(a_{m+1}^B \geq \frac{m}{3^n} n\), and so \(\langle B^- \rangle\) has an orbit \(\Gamma\) of length at least \(0.9n\). By Lemma 6.2, \(\Gamma\) is also an orbit of \(\langle B^+ \rangle\).

We would like \(\langle B^- \rangle\) to act as an alternating or symmetric group on \(\Gamma\); let us show that, if this is not the case, we obtain descent. We are assuming that Proposition 6.5 holds for \(n' < n\) (inductive hypothesis). Hence, if \(\langle B^- \rangle\) has no composition factor \(\text{Alt}(k)\) with \(k > 0.95n\), then Lemma 6.3 (descent) gives us

\[
A^{16m^6e_1(\log n)^3 \cdot F_2(0.95n)} \supseteq \text{Alt}([n]),
\]

for \(n\) larger than an absolute constant, where \(c_1 = c(0.9)\) is from Lemma 6.3. By (6.5), we conclude that (6.6) holds and we are done. (We are assuming that \(n\) is larger than a constant, so that \(16n^6 \leq c_3 \log n\), where \(c_3 > 0\) will be set later.)
Thus, we can suppose from now on that \( \langle B^- \rangle \) does have a composition factor \( \text{Alt}(k) \) for some \( k > 0.95n \). The only orbit of \( \langle B^- \rangle \) that can be of length at least \( k \) is \( \Gamma \), so \( \langle B^- \rangle |_\Gamma = \langle B^- |_\Gamma \rangle \) must contain \( \text{Alt}(k) \) as a section. Hence, by Lemma 3.11, \( \langle B^- |_\Gamma \rangle \geq \text{Alt}(\Gamma) \). (We can assume \( 0.95n > 84 \), and thus Lemma 3.11 does apply.) Note we also get that \( |\Gamma| > 0.95n \).

Now that we know that \( \langle B^- |_\Gamma \rangle \geq \text{Alt}(\Gamma) \). Corollary 4.7 gives us a small set of elements \( Y = \{y_1, y_2, \ldots, y_6\} \subseteq (B^-) |_{[\log 2n, \log n]} \) such that \( \langle Y \rangle |_\Gamma \) is 2-transitive on \( \Gamma \). We apply Lemma 6.1 (creation) with \( H^- = \langle B^- \rangle \), \( H^+ = \langle B^+ \rangle \), \( B = B^+ \), and \( r = 6 \). (The condition \( H^- < H^+ \) is fulfilled thanks to Lemma 6.2.)

If conclusion (a) in Lemma 6.1 holds, then there is a \( b \in B^+(B^+)^{-1} \setminus \{e\} \) with \( \text{supp}(b) \leq 0.05n \). Theorem 1.4 thus gives us that \( (A \cup \{b\})^{Kn^8} \supset \text{Alt}([n]) \), where \( K = K(0.1) \geq K(0.05) \) is an absolute constant. (We set \( K = K(0.1) \), instead of \( K = K(0.05) \), because we are planning to use the same constant later.) By (6.5),

\[
2 \cdot 48m^6 \cdot Kn^8 < 96Kn^{14} \leq F_1(n),
\]

and so (provided that \( n \) is larger than a constant) (6.6) holds and we are done. (This is what we call an exit from the procedure.)

We can thus assume that conclusion (b) in Lemma 6.1 holds; i.e., we have created a set \( W = (B^+)^{-1}YB^+ \cap \langle B^- \rangle \) with \( |W| \geq |B^+|^{1/6} \). Note that \( (B^+)^{-1}YB^+ \subseteq A |_{[\log 2n, \log n]} \) (for \( n \) larger than a constant) and \( |B^+| \geq |\text{Alt}(\Delta)| = (1/2)|\Delta!| \geq m^{0.899m} \) (for \( m \) larger than a constant; recall that \( |\Delta| \geq 0.9m \)). Hence

(6.10)

\[
|A |_{[\log 2n, \log n]} \cap \langle B^- \rangle | \geq m^{0.149m}.
\]

Now that we have created many elements in the pointwise stabiliser of \( \Sigma \), it is our task to organise them: we wish to produce \( \alpha_m, \ldots, \alpha_{m+\ell+1} \) satisfying (6.7).

This can be done in two ways. One is short and simple, gives a bound of \( l \gg m(\log m)/(\log n)^2 \), and results in a bound of \( O((\log n)^8(\log \log n)) \) in the exponent of the final result. The other is longer, but gives the stronger bound of \( l \gg m(\log m)/(\log n) \) promised in the statement of the proposition, and results in a bound of \( O((\log n)^4 \log \log n) \) in the exponent of the final result. Let us go through both arguments for the sake of clarity.

In the first argument, we simply apply Corollary 5.3 with \( \text{Sym}(\Gamma) \) instead of \( \text{Sym}([n]) \) and \( A |_{[\log 2n, \log n]} \cap \langle B^- \rangle \supset B^- \) instead of \( A \). We obtain that any maximal sequence of elements \( \alpha_m, \ldots, \alpha_{m+\ell+1} \) satisfying (6.7) must be of length

\[
\gg (\log |A |_{[\log 2n, \log n]} \cap \langle B^- \rangle |)/(\log n)^2 \gg \frac{\log m^{0.149m}}{(\log n)^2} \gg \frac{m(\log m)}{(\log n)^2}.
\]

Thus \( \ell \gg m(\log m)/(\log n)^2 \).
Let us now carry out the second argument in detail. The basic idea is that
the creation step has given us enough elements that we can apply the organiser
step several times in succession.

For \( i \geq 0 \), we define recursively \( A_i, B_i \subseteq (A) \) and a sequence \( \Sigma_i \) of points
in \([n]\). Let \( A_0 = A^{\lceil n^{29 \log n} \rceil} \), \( m_0 = m \), \( \Sigma_0 = (\alpha_1, \ldots, \alpha_{m_0+1}) \), and \( B_0 =
(\alpha_0)\Sigma_0\{\alpha_{m_0+1}\}) \).

If \( A_i, \Sigma_i, B_i \) are already defined then let \( A_{i+1} = A_i^{\lceil n^{6 \log n} \rceil} \) and let \( \Sigma_{i+1} \)
be a maximal extension \( \Sigma_{i+1} = (\alpha_1, \ldots, \alpha_{m_{i+1}+1}) \) of \( \Sigma_i = (\alpha_1, \ldots, \alpha_{m_i+1}) \) such that

\[
\left| \alpha_j^{(A_{i+1})_{\{\alpha_1, \ldots, \alpha_{j-1}\}}} \right| \geq 0.9n
\]

for all \( j = 1, 2, \ldots, m_{i+1} + 1 \). Finally, let

\[
A_{i+1} = (A_{i+1})^{29n^6} \quad \text{and} \quad B_{i+1} = (A_{i+1})\Sigma_{i+1}\{\alpha_{m_{i+1}+1}\}).
\]

Note that for all \( i \geq 0 \), \( (B_i) \) has an orbit \( \Gamma_i \) of length at least \( 0.9n \) because
\( |\alpha_{m_i+1}| \geq 0.9n \). We went up to \( i = m + 1 \) in condition (6.4) and up to
\( i = m + l + 1 \) in conclusion (6.7) (rather than \( i = m \) and \( i = m + l \), respectively)
so that we could do this useful trick!

We stop the recursion, and set \( w := i \) for the last \( i \) for which \( A_i \) is defined,
if either

(a) \( |B_i|_{\Gamma_i} \leq |B_i| \), i.e., there are two elements \( b_1, b_2 \in B_i \) such that \( b_1 b_2^{-1} \)
fixes \( \Gamma_i \) pointwise; or

(b) \( |\Gamma_i| \leq 0.95n \) or \( (B_i|_{\Gamma_i}) \not\supset \Alt(\Gamma_i) \); or

(c) \( n^{m_i-m_0} > m^{0.149n} \).

By (6.10), we have \( |B_0| \geq m^{0.149n} \).

First, we estimate the differences \( m_{i+1} - m_i \). If the recursion did not stop
after the definition of \( A_i, B_i, \) and \( \Sigma_i \) then, in particular, the stopping criterion
(c) is not fulfilled at step \( i \). Lemma 3.4, applied with \( (B_0) \) as \( G, G(\Sigma_i\{\alpha_{m_i+1}\}) \)
as \( H \), and \( B_0 \) as \( A \), then implies that

\[
|B_i| \geq \left| B_i \right| H \geq \frac{|B_0|}{n^{m_i-m_0}} \geq m^{0.149n}.
\]

Also, by the criteria (a) and (b), we have \( |B_i|_{\Gamma_i} = |B_i| \) and \( (B_i|_{\Gamma_i}) \) acts as
\( \Alt(\Gamma_i) \) or \( \Sym(\Gamma_i) \) on \( \Gamma_i \), where \( |\Gamma_i| > 0.95n \).

Since \( 0.9n < 0.95 \cdot 0.95n \leq 0.95|\Gamma_i| \), we can apply Corollary 5.3 with \( \rho =
0.05, B_i|_{\Gamma_i} \) instead of \( A \), and \( \Gamma_i \) instead of \([n]\), and obtain that, for \( 1 \leq i < w \),

\[
m_{i+1} - m_i > \frac{\log |B_i|}{60(\log n)^2} \geq \frac{c_2 m \log m}{60(\log n)^2},
\]
where we define $c_2 := 0.149/2 = 0.0745$. (This is what we have called an *organiser* step. It is ultimately based on the splitting lemma (Proposition 5.2), of which Corollary 5.3 is a corollary.)

At the same time, $n^{m_{w-1}-m_0} \leq \sqrt{m^{0.149m}}$ implies

$$m_{w-1} - m_0 \leq \frac{c_2 \log m}{\log n} m.$$ 

Since $m_{w-1} - m_0 = \sum_{i=1}^{w-1} (m_i - m_{i-1})$, from (6.12) it follows that

$$\frac{c_2 \log m}{\log n} m > (w - 1) \frac{c_2 \log m}{60(\log n)^2},$$

and we conclude that $w - 1 < 60 \log n$. Hence

$$A_w = A_{0}^{\lfloor 9m^6 \log n \lfloor w (48n)^w \subseteq A_n^{\lfloor 29 \log n \lfloor - \lfloor 432n^{12} \log n \lfloor w \subseteq A_n^{\lfloor c_3 \log n \lfloor}$$

for $c_3 := 750 > 29 + 12 \cdot 60$, provided that $n$ is larger than an absolute constant.

If $n^{m_{w-1}-m_0} > \sqrt{m^{0.149m}}$ (stopping condition (c)), then

$$m_{w} - m_0 \geq \frac{c_2 \log m}{\log n} m$$

and so, setting $\ell = m_w - m_0$, we obtain (6.7).

(In other words, as long as our *organising* has consumed less than the square-root of the material we created, we are organising rapidly; if our organising has consumed at least the square-root of the said material, then we have already organised plenty.)

If we stopped because condition (a) holds then $A^2_w$ contains a nontrivial element $b_1 b_2^{-1}$ with support less than 0.1$n$. By Theorem 1.4, $(A \cup \{b_1 b_2^{-1}\}) K_n^{\delta} \supseteq \text{Alt}([n])$, where $K = K(0.1)$ is an absolute constant. By (6.5),

$$2 \cdot \lfloor n^c \log n \rfloor \cdot K n^{\delta} \leq F_1(n),$$

and so we obtain (6.6). (This is an exit case.)

Finally, suppose we stopped in case (b), i.e., $\langle B_w \mid \Gamma_w \rangle \not\supset \text{Alt}(\Gamma_w)$ or $|\Gamma_w| \leq 0.95n$. As $|\Sigma|_w \geq m > C(0.9)$, we can apply Lemma 3.18 with $\Sigma_w \setminus \alpha_{m_{w}+1}$ as $\Sigma$ and $A'$ as $A$ to obtain $\Delta_w \subseteq \Sigma_w \setminus \alpha_{m_{w}+1}$, $|\Delta_w| \geq 0.9|\Sigma_w \setminus \alpha_{m_{w}+1}|$ such that

$$B^+_w = \langle ((A'_w)^{16n^6})_{\Sigma_w \setminus \{\alpha_{m_{w}+1}\}}(\Sigma_w \setminus \{\alpha_{m_{w}+1}\}) \cup \Delta_w \rangle$$

satisfies $(B^+_w)|_{\Delta_w} = \text{Alt}(\Delta_w)$. (This is a fresh setup.) Also, by Lemma 6.2, $B^-_w = ((B^+_w)^3|_{\Delta_w})$ generates $\langle B^+_w \rangle |_{\Delta_w} \cup \langle B^+_w \rangle$. Note that $B^-_w \subseteq B_w$ and $\langle B^-_w \rangle$ has an orbit of length at least 0.9$n$, simply because $B^-_w$ contains $(A'_w)_{\Sigma_w \setminus \{\alpha_{m_{w}+1}\}}$, and the orbit of $\alpha_{m_{w}+1}$ under $(A'_w)_{\Sigma_w \setminus \{\alpha_{m_{w}+1}\}}$ is of length $\geq 0.9n$ by (6.11).

We are ready for another descent. The group $\langle B^-_w \rangle$ has no composition factor $\text{Alt}(k)$ with $k > 0.95n$, because such a factor would be a section of $\langle B_w \rangle$ and Lemma 3.11 would imply that $\langle B_w \mid \Gamma_w \rangle$ is an alternating group on $> 0.95n$ elements, in contradiction with condition (b). Thus the hypotheses of
Lemma 6.3 are satisfied with $\delta = 0.95$ and $\rho = 0.9$ and, by the assumption that Proposition 6.5 holds for $n' \leq 0.95n < n$ (inductive hypothesis), Lemma 6.3 gives us that

$$A[n^{c_3 \log n} e^{(\log n)^3} F_2(0.95n)] = \text{Alt}(\{n\}),$$

where $c = c(0.9)$. We apply (6.5) and conclude that (6.6) holds.$\square$

We now use Proposition 6.5 to prove both the Main Theorem and Corollary 1.3 (for Sym$(n)$ and Alt$(n)$).

**Theorem 6.6.** Let $G = \text{Sym}(n)$ or $\text{Alt}(n)$. Then

$$\text{diam}(G) = O(e^{c(\log n)^4 \log \log n}),$$

(6.13)

$$\overset{\text{diam}(G)}{\rightarrow} = O(e^{(c+1)(\log n)^4 \log \log n}),$$

for an absolute constant $c > 0$.

As we shall see, $c_1 = 49071$ is valid (and by no means optimal).

**Proof.** We must find functions $F_1, F_2$ satisfying (6.5) and (6.8). We can set

$$F_2(n) = e^{(\log n)^3 + 2 \log n + c' c_3 (\log n)^3 \log \log n} F_1(n) + 2$$

for $c' > c_2$ arbitrary. Now we must make sure that

(6.14) $$F_1(n) \geq n^{c_3 \log n} e^{c_1 (\log n)^3} \cdot (e^{c' c_3 (\log 0.95n)^3 \log \log 0.95n + (\log 0.95n)^3 + 2 \log 0.95n} F_1(0.95n) + 2).$$

(Here we can assume $n > 1$ so that $\log \log n$ is well defined.) Choose $c_4 > c' c_3$. Then, for $n$ larger than a constant $n_0'$ depending only on $c_1, c_3, c'$, and $c_4$, (6.14) will hold provided that

(6.15) $$F_1(n) \geq e^{c_4 (\log n)^3 \log \log n} \max(F_1(0.95n), 1).$$

For any $c > c_4/(4 \log 0.95)$ and any $C \geq 1$, (6.15) is fulfilled by

$$F_1(n) = C e^{c (\log n)^4 \log \log n},$$

provided that $n$ is larger than a constant $n_0''$ depending only on $c$ and $c_4$. We set $C = n_0''!$, where $n_0'' = \max(n_0, n_0', n_0'', 2K)$. Then (6.5) holds for all $n \geq n_0''$, and (6.8) holds with $n_0!$ replaced by $n_0''!$. We now apply Proposition 6.5 for our $n$, with $n_0$ replaced by $n_0''$; it uses Proposition 6.4, which in turn uses Proposition 6.5 for smaller $n$, and so on. The recursion ends when $n \leq \max(n_0'', 1)$, as then Proposition 6.5 is trivially true (due to the bound $F_2(n) \geq n_0''!$ in (6.8)).

We obtain that

(6.16) $$\text{diam}(\Gamma(G, Y)) \leq C e^{c (\log n)^4 \log \log n}$$
for any set $Y$ of generators of $G$ with $Y = Y^{-1}$, $e \in Y$. A quick calculation shows that, since $c_2 = 0.0745$ and $c_3 = 750$ (see the proof of Proposition 6.4), we can set $c' = 13.423 > 1/0.0745$, $c_4 = 10068 > c'c_3$ and

$$c = \left\lfloor \frac{c_4}{4\log 0.95} \right\rfloor = 49071.$$

Let $A$ be an arbitrary set of generators of $G$. Let $Y = A \cup A^{-1} \cup \{e\}$. The undirected Cayley graph $\Gamma(G, Y)$ is just the undirected Cayley graph $\Gamma(G, A)$ with a loop at every vertex; their diameters are the same. Thus, by (6.16),

$$\text{diam}(\Gamma(G, A)) = \text{diam}(\Gamma(G, Y)) \leq Ce^{c\log n} - 4\log \log n.$$

By [Bab06, Cor. 2.3],

$$\text{diam}(\Gamma(G, A)) \leq O \left( \text{diam}(G)(n \log n)^2 \right) \leq O \left( e^{(c+1)(\log n)^4 \log \log n} \right). \quad \square$$

References


ON THE DIAMETER OF PERMUTATION GROUPS


HARALD A. HELFGOTT and ÁKOS SERESS


(Received: September 27, 2011)
(Revised: April 22, 2013)

ÉCOLE NORMALE SUPÉRIEURE, PARIS, FRANCE

E-mail: harald.helfgott@ens.fr

THE UNIVERSITY OF WESTERN AUSTRALIA, CRAWLEY, AUSTRALIA and

THE OHIO STATE UNIVERSITY, COLUMBUS, OH