JOURNAL OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 18, Number 1, Pages 157–192
S 0894-0347(04)00467-9
Article electronically published on September 2, 2004

DIMENSION AND RANDOMNESS IN GROUPS ACTING ON ROOTED TREES

MIKLÓS ABÉRT AND BÁLINT VIRÁG

1. Introduction

Let $T = T(d)$ denote the infinite rooted d-ary tree and let $H \subseteq \mathrm{Sym}(d)$ be a permutation group. Let $W(H)$ denote the infinite iterated wreath product of H acting on T with respect to H. For example, $W(\mathrm{Sym}(d))$ is the full automorphism group of $T(d)$. Let $W_n(H)$ denote the n-fold wreath product of H, acting on T_n, the d-ary tree of depth n.

The case $H = C_p$, the cyclic group of order p, is of particular interest. The pro-p group $W(p) = W(C_p)$ obtained this way is called the group of p-adic automorphisms. The group $W_n(p)$ is called the *symmetric p-group* of depth n, as it can also be obtained as the Sylow p-subgroup of the symmetric group $\mathrm{Sym}(p^n)$.

The first goal of this paper is to analyze the elements of $W(H)$. We answer a question of Turán (see Pálfy and Szalay [1983]), which asks for an analogy of the famous theorem of Erdős and Turán [1965] about the distribution of orders of random elements in $\mathrm{Sym}(n)$.

Let p^{K_n} denote the order of a random element of the symmetric p-group $W_n(p)$ and let α_p be the solution of the equation $\alpha(1-\alpha)^{1/\alpha-1} = 1-1/p$ in the open unit interval.

Theorem 1. *We have $K_n/n \to \alpha_p$ in probability.*

A similar result holds for general H (see Section 2). The proof of this theorem depends on a new measure-preserving bijection between conjugacy classes of random elements and Galton-Watson trees, bringing the theory of stochastic processes to bear upon the subject. Theorem 1 was conjectured by Pálfy and Szalay [1983]; they proved the upper bound and that the variance of K_n remains bounded as $n \to \infty$. Puchta [2001] showed that the limit in Theorem 1 exists. In a related recent paper, Evans [2002] studies the random measure given by the eigenvalues of the natural representation of $W_n(H)$.

Received by the editors February 16, 2003.

2000 *Mathematics Subject Classification.* Primary 20E08, 60J80, 37C20; Secondary 20F69, 20E18, 20B27, 28A78.

Key words and phrases. Groups acting on rooted trees, Galton-Watson trees, Hausdorff dimension, pro p-groups, generic subgroups, symmetric p-group.

The first author's research was partially supported by OTKA grant T38059 and NSF grant #DMS-0401006.

The second author's research was partially supported by NSF grant #DMS-0206781 and the Canada Research Chair program.

The next goal is to understand the subgroup structure of $W(p)$. Since every countably based pro-p group can be embedded into $W(p)$, we investigate subgroups according to their Hausdorff dimension. We show that the Hausdorff spectrum for the closure of subgroups generated by 3 elements is the whole unit interval. More precisely,

Theorem 2. *For each $\lambda \in [0,1]$ there exists a subgroup $G \subseteq W(p)$ generated by 3 elements such that the closure of G is λ-dimensional.*

This result leads to the solution of a problem of Shalev [2000], who asked whether a finitely generated pro-p group can contain finitely generated subgroups of irrational Hausdorff dimension.

In the context of pro-finite groups Hausdorff dimension was introduced by Abercrombie [1994]. It was deeply analyzed by Barnea and Shalev [1997] (see also Shalev [2000] and Barnea et al. [1998]). They show that for closed subgroups, Hausdorff dimension agrees with the lower Minkowski (box) dimension. In the case of $W(H)$, this can be computed from the size of the congruence quotients $G_n = GW^{[n]}/W^{[n]}$, where $W^{[n]}$ denotes the stabilizer of level n in W. In fact, $\dim_{\mathrm{H}} G$ is given by the lim inf of the *density sequence*

$$\gamma_n(G) = \log |G_n| / \log |W_n|. \tag{1}$$

For instance, the closure of Grigorchuk's group, a subgroup of $W(2)$, has dimension $5/8$.

Theorem 2 is obtained using probabilistic methods. These methods give the following analogue of the theorem of Dixon [1969] saying that two random elements of $\mathrm{Sym}(n)$ generate $\mathrm{Sym}(n)$ or the alternating group $\mathrm{Alt}(n)$ with probability tending to 1. Perhaps surprisingly, no deterministic construction is known for large subgroups of $W_n(p)$ with a bounded number of generators as in the following theorem.

Theorem 3. *Let $\varepsilon > 0$. Let G_n be the subgroup generated by three random elements of the symmetric p-group $W_n(p)$. Then $\mathbf{P}\left(|G_n| > |W_n(p)|^{1-\varepsilon} \right) \to 1$ as $n \to \infty$.*

For the infinite group $W(p)$, this takes the following form: with probability 1, three random elements generate a subgroup with 1-dimensional closure (see Theorem 7.2). We conjecture that the same holds for two elements instead of three. Since three random elements generate a spherically transitive subgroup with positive probability, and $W(p)$ is not topologically finitely generated, the above result shows the existence of a rich family of topologically finitely generated transitive full-dimensional subgroups of $W(p)$.

The analysis of randomly generated subgroups allows us to answer a question of Sidki [2001]. He asked whether the binary odometer (a.k.a. adding machine), an element that acts on every level of $T(2)$ as a full cycle, can be embedded into a free subgroup of $W(2)$. Call a subgroup $G \subseteq W(H)$ strongly free if it is free and every nontrivial element fixes only finitely many vertices. For instance, both the trivial group and the cyclic subgroup generated by the odometer have this property.

Theorem 4. *Let $G \subseteq W(H)$ be a strongly free subgroup and let $g \in W(H)$ be a random element. Then the group $\langle G, g \rangle$ is strongly free with probability* 1.

Next, we explore a general dimension theory of groups acting on the rooted trees $T(p)$. The first result in this direction shows that closed subgroups of $W(p)$ are 'perfect' in the sense of Hausdorff dimension. Recall that G' denotes the derived subgroup of the group G.

Theorem 5. *Let $G \subseteq W(p)$ be a closed subgroup. Then* $\lim(\gamma_n(G) - \gamma_n(G')) = 0$. *In particular,* $\dim_{\mathrm{H}} G = \dim_{\mathrm{H}} G'$.

As a corollary, solvable subgroups are zero-dimensional, and positive-dimensional closed subgroups cannot be *abstractly* generated by countably many solvable subgroups.

Unlike the infinite group $W(p)$, the finite symmetric p-groups $W_n(p)$ do have large Abelian subgroups of density $1 - 1/p$. However, they cannot be glued to form a large subgroup of $W(p)$. The following theorem explains the background of this phenomenon.

Theorem 6. *Let $G \subseteq W(p)$ be a solvable subgroup with solvable length d. Then we have $\sum_{n=0}^{\infty} \gamma_n(G) \leq Cd$, where C is a constant depending on p only.*

One motivation to study dimension of subgroups in $W(p)$ comes from the theory of just infinite pro-p groups, which are regarded as the simple groups in the pro-p category. A group G is just infinite if every proper quotient of G is finite. By the characterization of Grigorchuk [2000] (which is based on Wilson [1971]), groups with this property fall into two classes: one is the so-called branch groups, which have a natural action on $T(p)$ (see Section 9 for details). Boston [2000] suggests a direct connection between Grigorchuk's classes and Hausdorff dimension; he conjectures that a just infinite pro-p group is branch if and only if it has an embedding into $W(p)$ with positive-dimensional image.

In the following we contrast known results about branch groups with new results on 1-dimensional subgroups. We first show the following.

Theorem 7. *Let $G \subseteq W(p)$ be a spherically transitive 1-dimensional closed subgroup. Then every nontrivial normal subgroup of G is 1-dimensional.*

It would be interesting to see whether a similar result holds for spherically transitive positive-dimensional closed subgroups.

Wilson [2000] conjectured that every just infinite pro-p branch group contains a nonabelian free pro-p subgroup. An affirmative result is known if the group is not virtually torsion-free (see Grigorchuk et al. [2000a]). Another beautiful result in this vein is given by Zelmanov [2000] on groups satisfying the Golod-Shafarevich condition. In light of Boston's conjecture it is natural to formulate the following.

Conjecture 8. *Let $G \subseteq W(p)$ be a positive-dimensional closed subgroup. Then G contains a nonabelian free pro-p subgroup.*

This might be more attackable than Wilson's original conjecture. A step in this direction is the following.

Theorem 9. *Let $G \subseteq W(p)$ be a spherically transitive 1-dimensional closed subgroup. Then G contains a nonabelian free pro-p subgroup.*

Another measure of largeness for profinite groups is whether they contain dense free subgroups (see Pyber and Shalev [2001] and Soifer and Venkataramana [2000]). Wilson [2000] showed that just infinite pro-p branch groups contain dense free subgroups. Let $d(G)$ denote the minimal number of topological generators for G, which may be infinite.

Theorem 10. *Let $G \subseteq W(p)$ be a closed subgroup of dimension 1 and let $k \geq d(G)$. Then G contains a dense free subgroup of rank k.*

In other words, the free group F_k is residually S, where S denotes the set of congruence quotients of G.

A common tool throughout the paper is the so-called *orbit tree* of a subgroup $G \subseteq W(H)$, the quotient graph of T modulo the orbits of G. Orbit trees reflect the conjugacy relation (see Gawron et al. [2001]) and can be used to describe the structure of Abelian subgroups in $W(p)$ (see Section 8). The key observation leading to Theorem 1 is the fact that the orbit tree of a random element is a Galton-Watson tree. A crucial step in Theorems 3 and 4 is that the orbit tree of a randomly generated subgroup has finitely many rays with probability 1, i.e., the closure has finitely many orbits on the boundary of T (Proposition 3.10). Theorem 5 also traces back to estimates on orbit trees.

Another tool used in the paper is word maps between the spaces $W(p)^k$. A nontrivial word $w \in F_k$ can be thought of as a map $W(H)^k \to W(H)$ via evaluation. Bhattacharjee [1995] proved that random substitution into a word map gives 1 with probability 0, or equivalently, the kernel of the map has Haar measure zero. As a result, random elements of $W(H)$ generate a free subgroup with probability 1. The key result leading to Theorem 10 is the following generalization.

Theorem 11. *The kernel of a word map is not full dimensional in $W(H)^k$.*

Another generalization leading to Theorem 4 is that random evaluation of a word map yields an element of $W(H)$ which fixes only finitely many vertices with probability 1.

The structure of the paper is as follows. Section 2 contains the proof of Theorem 1. It also introduces notation and discusses conjugacy classes and random elements in $W(H)$. Section 3 discusses the orbit structure of random subgroups. In Section 4 we explore word maps and prove Theorems 4 and 11. Sections 5 and 6 introduce the technical tools needed for proving Theorem 3, which is done in Section 7. Section 8 discusses small subgroups, and covers Theorems 5 and 6. Section 9 is about high-dimensional subgroups, and contains the proofs of Theorems 7, 8, 9, and 10.

2. Elements

This section studies the statistical properties of elements of $W(p)$. Random elements are described via the family tree of a Galton-Watson branching process. This allows us to answer Turán's question (see Theorem 1).

We first describe our notation for (iterated) **wreath products**. Let (H_i, X_i), $1 \le \ell \le n$, be a sequence of permutation groups where n might be infinite. Let $d_i = |X_i|$. For $0 \le \ell \le n$, define the **level** ℓ as

$$\partial T_\ell := X_1 \times \ldots \times X_\ell,$$

and the tree T_n which has vertex set $V(T_n)$ given by elements of the disjoint union of all levels. The sequence $w \in V(T_n)$ is a child of $v \in V(T_n)$ (or v is the parent of w) if w begins as v and has one extra element. The edge set E is the collection of $\{v, w\}$ for all such v, w.

Consider a map g that for each $\ell < n$ maps

$$\partial T_\ell \to H_{\ell+1}. \tag{2}$$

Let W be the set of all such maps. An element $g \in W$ acts on T_n bijectively via the rule

$$(x_1, \ldots, x_\ell)^g := (x_1^{g(())}, x_2^{g((x_1))}, \ldots, x_\ell^{g((x_1,\ldots,x_{\ell-1}))}), \tag{3}$$

and $W = H_n \wr \ldots \wr H_1$ is easily checked to be a subgroup of $\mathrm{Aut}(T_n)$. This construction also works in the case $n = \infty$, and the notation $W = \ldots \wr H_2 \wr H_1$ is compatible with the right actions we are considering.

If n is finite, then the group structure of W_n does not depend on the permutation representation of the last group H_n on $\mathrm{Sym}(n)$, only on the abstract structure of H_n. This defines $K \wr H$, where K is an abstract group. When we talk about $g \in H_2 \wr H_1$, the wreath product of two groups, we will often use the shorthand $g(x)$ for $g((x))$, where $x \in X_1$. The rule (3) yields the multiplication rule

$$(gh)(x) = g(x)h(x^g).$$

We will use the notation $W_n(H)$ for the wreath product of n copies of a permutation group (H, X), and the shorthand $W(H) = W_\infty(H)$. The most important cases are the **symmetric p-group** $W_n(C_p)$ and the full automorphism group $W_n(\mathrm{Sym}(d))$ of T_n.

If G is a group, and $S \subset G$, then $\langle S \rangle$ denotes the group **abstractly generated** by S, that is, the set of elements that can be obtained from S via repeated group operations. If G is a topological group, then we call the closure $\overline{\langle S \rangle}$ the group **topologically generated** by S.

Let G be a group acting on a tree T.

Definition 2.1. The **orbit tree of the group** G is the quotient graph T_G of T modulo the set of orbits of G.

It is easy to check that orbit trees are in fact trees.

For the purpose of Proposition 2.3 the orbit trees of group elements need to contain more information about the local structure of automorphisms. Let v be a vertex of a tree $T = X^\infty$, and let k denote the length of the orbit of v under an automorphism g. Then g^k acts on the child edges of v, which are labeled naturally by elements of X. Let $\ell_g(e) \subseteq X$ denote the orbit of the child edge e. For a set (e.g. orbit) $\mathcal{E}$ of edges at the same level, let $\ell_g(\mathcal{E}) = \ell_g(e)$ for the lexicographically smallest $e \in \mathcal{E}$.

Definition 2.2. The **orbit tree of a tree automorphism** g is the quotient graph T_g of T modulo the set of orbits of g, together with the labeling ℓ_g of the edges of T_g.

The orbit tree of $g \in W(H)$ is an H-labeled tree, defined as follows.

An **H-labeled tree** is a rooted tree together with an edge-labeling ℓ so that for $v \in V$, ℓ is a bijection between the child edges of v and the cycles of some $h_v \in H$.

Two H-labeled trees are called **equivalent** if (1) there exists a graph isomorphism $v \mapsto v'$, $e \mapsto e'$ between them, and (2) for each v, there exists an $h_v \in H$ so that $\ell(e') = \ell(e)^{h_v}$ for all child edges e' of v'.

The following extension of a theorem of Sushchansky [1984] is not difficult to check, and motivates the notion of orbit trees.

Proposition 2.3. *Two elements of $W(H)$ are conjugates if and only if their orbit trees are equivalent.*

In most important cases, equivalence is only a little more than graph isomorphism. We say that two labellings of a tree are equivalent if they define equivalent H-labeled trees. In the case $H = \mathrm{Sym}(d)$, two labellings are equivalent iff the length of the edge label cycles agree. In this case, it suffices to label edges by the

cycle lengths. If $H = C_2$, then any two labellings are equivalent. If $H = C_p$ for p prime, then each vertex has either 1 or p children, and labeling is the same as cyclicly ordering the children whenever there are p of them.

We now turn to the description of the orbit trees of random elements. Our convention is that unless otherwise specified, the term **random elements** denotes independent random elements chosen according to uniform (in the infinite case Haar) measure.

Definition 2.4. Labeled Galton-Watson (GW) trees. Let $\bar{L}$ be the set of finite sequences with elements from a set of labels L, and let ν be a probability distribution on $\bar{L}$. We define the probability distribution $\mathsf{GW}(\nu)$ on infinite trees with edges labeled by L by the following inductive construction. In the first generation, we have 1 individual. Given the individuals at level n, each of them has a sequence of child edges picked from the distribution ν independently. The other endpoints of the child edges form generation $n+1$.

Note that the unlabeled tree is just a classical Galton-Watson process; once the unlabeled tree has been laid down, each family is labelled independently at random using a mechanism that only depends on the family size.

Let (H, X) be a permutation group, and let ν_H be the distribution of the sequence of orbits of a uniform random element.

Proposition 2.5 (Random elements and GW trees)**.** *If $g \in W(H)$ is a Haar random element, then the distribution of the orbit tree T_g is $\mathsf{GW}(\nu_H)$.*

Proof. We prove this by induction. The inductive hypothesis is that the first n levels of T_g have the same distribution as the first n levels of a $\mathsf{GW}(\nu_H)$ tree.

Indeed, suppose this is true for n. Let $w = (v, \dots, v^{g^{k-1}}) \in T_g$ be an orbit of an element on level n of T. Then the child edges of w in T_g correspond to, and are labeled by, orbits of $g^k(v)$. But $g^k(v) = g(v)g(v^g)\dots g(v^{g^{k-1}})$, and the factors on the right-hand side are chosen independently uniformly from H, which implies that their product also has uniform distribution. □

The simplest consequence of this is the following asymptotic law. Let $r(H)$ denote the permutation rank of H, i.e. the number of orbits of the action of H on pairs of elements of X (e.g. $r(H) = 2$ if H is 2-transitive on X).

Corollary 2.6. *Let the group H be transitive on X, and let $g_n \in W_n(H)$ be random. Then as $n \to \infty$,*

$$n\mathbf{P}(g_n \textit{ fixes a vertex on level } n) \to 2/(r(H) - 1).$$

Proof. Let N be the number of fixed points of a random element of H on X; then it is well known that $\mathbf{E}N = 1$, $\mathbf{E}N^2 = r$, so $\operatorname{Var} N = r - 1$.

A vertex at level n of T_{g_n} corresponds to a fixed point if and only if all of its ancestor edges are labeled by a fixed point. This happens if it lies in the subtree of T_{g_n} gained by removing all descendant subtrees which have an ancestor edge not labeled by a fixed point. Such a subtree has Galton-Watson distribution with branching structure given by the number of fixed points N of a random element in H. Since $\mathbf{E}N = 1$, this is a critical tree, so by a theorem in branching processes (see Athreya and Ney [1972]) the probability that it survives until generation n is asymptotic to $2/(n \operatorname{Var} N)$, proving the corollary. □

The following corollary is immediate.

Corollary 2.7. *Let H be transitive on X, and let $g \in W(H)$ be chosen according to Haar measure. Then g fixes only finitely many vertices of T with probability* 1.

Let for an integer n, let $p(n)$ denote the highest power of p dividing n. Let $h \in H$ be a uniform random element, and let p be a prime. Let $\mu_p(k) = \mu_{H,p}(k)$ denote the expected number of orbits v of h with $p(|v|) = k$. Let $\hat{\mu}_p$ be the generating function for μ_p:

$$\hat{\mu}_p(z) = \sum_{k=0}^{\infty} \mu_p(k) z^k$$

and note that the sum has only finitely many nonzero terms.

Proposition 2.8. *Let $g_n \in W_n$ be a random element, and let*

$$\alpha_p = \min_{\lambda>0} \frac{\log \hat{\mu}_p(e^\lambda)}{\lambda}. \tag{4}$$

Then $\mathbf{E}\, p(|g_n|)/n \to \alpha_p$ and there exist $c_1, c_2 > 0$ so that for all n, k we have

$$\mathbf{P}[|p(|g_n|) - \mathbf{E}\, p(|g_n|)| \geq k] \leq c_1 e^{-c_2 k},$$

in particular, $\mathrm{Var}(p(|g_n|))$ *remains bounded.*

The bounded variance statement is somewhat surprising, since in most limit theorems variance increases with n (in most cases it is on the order of n). The asymptotics of the order of a typical element is immediate from this result.

Corollary 2.9. *As $n \to \infty$, we have*

$$\log |g_n|/n \to \sum \alpha_p \log p$$

in probability, where the sum ranges over primes p dividing $|H|$.

As a special case, we get an answer to Turán's question, which we restate here. Let p^{K_n} denote the order of a random element of the symmetric p-group $W_n(p)$ and let α_p be the solution of the equation

$$\alpha(1-\alpha)^{1/\alpha - 1} = 1 - 1/p \tag{5}$$

in $(0, 1)$.

Theorem 1. *We have $K_n/n \to \alpha_p$ in probability.*

Turán's goal was to find a natural analogue of the theorem of Erdős and Turán [1965] about the asymptotics of the order of a random element in $\mathrm{Sym}(n)$. Their paper started the area of statistical group theory. Theorem 1 was conjectured by Pálfy and Szalay [1983]; they proved the upper bound and a linear lower bound with a different constant. Puchta [2001] shows that the limit exists. He also studies uniformly randomly chosen *conjugacy classes* in Puchta [2003]. In a related recent paper motivated by random matrix theory, Evans [2002] studies the random measure given by the eigenvalues of the natural representation of $W_n(H)$.

Proof of Theorem 1. H is the cyclic group of order p, so

$$\hat{\mu}_p(z) = 1 + z(p-1)/p.$$

Let $f(\lambda) = \log \hat{\mu}_p(e^\lambda)$. Then (4) equals the minimum of $f(\lambda)/\lambda$, and setting the derivative to 0 we get

$$f(\lambda)/\lambda = f'(\lambda). \tag{6}$$

At equality, this expression gives α. The right-hand side is the logarithmic derivative of $\hat{\mu}_p(e^\lambda)$, so it equals $\alpha = (e^\lambda(p-1)/p)/(1+e^\lambda(p-1)/p)$. Using this, we easily express λ and then $f(\lambda)$ from α. Substitution into (6) and algebraic manipulations give $\mathbf{E}K_n/n \to \alpha_p$, and convergence in probability follows since the variance of K_n is bounded. □

For the proof of Proposition 2.8, we need to understand the order of an element in terms of its orbit tree. A simple inductive argument shows that if $v \in T_g$ is an orbit, then its length is given by the product of the lengths of labels on the simple path $\pi((), v)$ from the root () to v in T_g:

$$|v| = \prod_{e \in \pi((),v)} |\ell(e)|, \tag{7}$$

and therefore

$$p(|v|) = \sum_{e \in \pi((),v)} p(|\ell(e)|). \tag{8}$$

Proposition 2.11 below shows that all properties about the length of orbits of W_n can be understood in terms of a branching random walk.

Let $\{p_i\}_{1 \le i \le d}$ be a sequence of primes. Consider an orbit o of an action group H on a set X, and let $\mathrm{ex}(o, X)$ denote the point in $\mathbb{Z}^d$ whose ith coordinate is the exponent of the highest power of p_i dividing $|o|$. Let $\mathrm{ex}(H, X)$ denote the multiset $\{\mathrm{ex}(o) \mid o \text{ orbit of } H\}$.

Definition 2.10. Let ν be a probability distribution on the space S of finite multisets of elements of $\mathbb{Z}^d$. A **branching random walk** is a probability measure on infinite sequences $\{\Xi_n;\, n \ge 0\}$ with elements from S. It is constructed recursively as follows. At time $n = 0$ we have $\Xi_0 = \{0\}$, the set containing the origin.

If Ξ_n is already constructed, then each $x \in \Xi_n$ has "offspring" $x+Y_1, \dots, x+Y_N$, where N is random. Here $(Y_1, \dots, Y_N)$ is picked from ν independently from the past and from the offspring of other individuals at time n. Ξ_{n+1} is the multi-set union of all the offspring of $x \in \Xi_n$.

We call the measure

$$\mu(x) = \mathbf{E}\left(\sum_{i=1}^{N} \mathbf{1}(Y_i = x)\right)$$

the **occupation measure** of the BRW. Let $\hat{\mu}$ denote the generating function for the measure μ.

In our case, the relative offspring positions are given by the sequence

$$\{\, (p_1(|o|), \dots, p_d(|o|)) \,:\, o \text{ orbit of } h \,\},$$

where h is a uniform random element of H.

Proposition 2.11 (Orbit lengths and BRW)**.** *Let $g \in W(H)$ be a Haar random element, and let $\Xi_n = ex(\langle g \rangle, \partial T_n)$. Then $\{\Xi_n;\, n \ge 0\}$ is a branching random walk on $\mathbb{Z}^d$.*

Proof. This is immediate from the formula (8) and the proof of Proposition 2.5. □

Since $p(|g|)$ equals the maximum of $p(|o|)$ for the orbits of g, we need to understand the position of the highest element of Ξ_n. The corresponding questions for the branching random walk have been answered by Hammersley [1974], Biggins [1977], and Dekking and Host [1991]. The part of their results that is relevant to our setting is stated in the following theorem.

Theorem 2.12. *Let X_n denote the position of the greatest individual of a 1-dimensional BRW whose occupation measure μ has finite support. Let*

$$\alpha = \inf_{\lambda>0} \frac{\log \hat{\mu}(e^{\lambda})}{\lambda}.$$

Then

$$\mathbf{E}X_n/n \to \alpha,$$

and there exist $c_1, c_2 > 0$ so that for all n,

$$\mathbf{P}[|X_n - \mathbf{E}X_n| \geq k] \leq c_1 e^{-c_2 n}.$$

Here we present a sketch of the proof of the first statement in a simple case.

Proof. Let $\mu^{*n}(k)$ denote the expected number of individuals at position k at time n. If we find the highest k for which this quantity is about 1, that means that on average there will be 1 individual at position k, and less than one individual at positions higher than k. Intuitively, this k seems to be a good guess for the position of the highest individual. There are technical arguments to make this intuition rigorous.

The next step is to solve $\mu^{*n}(k) = k$. By considering one step in the branching random walk one discovers the convolution formula

$$\mu^{*n}(k) = \sum_{j=0}^{k} \mu^{*n-1}(j)\mu(k-j)$$

but convolution turns into product for generating functions: $\widehat{\mu^{*n}} = \hat{\mu}^n$. For example, for the symmetric p-group,

$$\hat{\mu}(z)^n = (1 + z(p-1)/p)^n,$$

and we are left with finding the power k of z for which the coefficient in the above expression is about 1. This gives the approximate equation

$$\binom{n}{k} = \left(\frac{p}{p-1}\right)^k, \quad \alpha = k/n, \tag{9}$$

and by the Stirling formula the asymptotic version of this equation is (5). For the general case, the theory of large deviations is a standard tool for approximating $\mu^{*n}(\alpha n)$ for large n. □

3. Orbit trees of random subgroups

Just as for single elements, there is a natural graph structure on the orbits of a subgroup $G \subseteq W(H)$; see Definition 2.1. In this section, we prove that the orbit tree of random subgroup (i.e. subgroups generated by random elements) is a multi-type Galton-Watson tree. As a result, we show that the closure of a random subgroup has finitely many orbits on ∂T with probability 1, and compute the probability that it is transitive.

The following definition is a generalization of ordinary GW trees to the case when the branching mechanism may depend on the type of the individual.

Definition 3.1. Multi-type GW trees. Let Υ be a finite or countable set called **types**. Let $\bar{\Upsilon}$ be a set of finite sequences of Υ. For each $y \in \Upsilon$ let μ_y be a probability measure on $\bar{\Upsilon}$. A multi-type Galton-Watson tree $\mathsf{GW}(\mu_\cdot)$ is a random tree constructed as follows. The first generation consists of an individual of a given type $y \in \Upsilon$. Individuals in generation n have children according to the measure μ_y, where y is the type of the individual. The children of a generation are picked independently and they form the next generation.

Let $j \geq 1$, $k \geq 0$, and let (H, X) be a permutation group. Consider the action of the subgroup generated by $(j-1)k+1$ random elements of H on X. Consider the list of the length of the orbits, and replace each element ℓ of this list by ℓk. The distribution of the new list yields a probability measure $\mu_{H,j,k}$ on finite sequences of integers.

Proposition 3.2. *Let G be the subgroup generated by j random elements of $W(H)$. Then T_G has multi-type Galton-Watson distribution $\mathsf{GW}(\mu_{H,j,\cdot})$, where the types correspond to the length of the orbits.*

First we introduce some notation. Let $G = H \wr K$ be a permutation group, let G_x denote the stabilizer of x, and let G_{x+} denote the subgroup of elements $g \in G_x$ for which $g(x) = id$. The group G_x/G_{x+} "ignores" the action everywhere except at x; it is naturally isomorphic to K.

The following observation is immediate.

Fact 3.3. *Let $G \subseteq W(H)$, and let $T_{G,v}$ be the descendant subtree of the T_G at the orbit w of $v \in T$. Then $T_{G,v}$ is isomorphic to $T_{G_v/G_{v+}}$ with the labels multiplied by the length of w.*

Lemma 3.4. *Let (H, X) be a permutation group and let K be a group. Consider the subgroup G generated by j Haar random elements of $H \wr K$. Conditioned on the orbit V of x, G_x/G_{x+} has the same distribution as a subgroup generated by $|V|(j-1)+1$ random elements of G.*

The following definition is needed for the proof.

Definition 3.5. Let K be a group acting on a set X, and let S be a subset of K. The **Schreier graph** $\mathcal{G}(H, X, S)$ is a directed multi-graph whose vertex set is X, and the edge set is $X \times S$, where the edge (v, g) connects v to v^g.

Proof. Let g_i denote the randomly chosen generators of H. Consider the Schreier graph $\mathcal{G}(G, V, \{g_1, \ldots, g_j\})$, and let E denote its edge set. Let π_1 denote the fundamental group of the graph. It consists of equivalence classes of cycles starting from x. The cycles may contain an edge in either direction. The equivalence relation is generated by adding or removing immediately retraced steps.

It is well known that for a connected graph, π_1 is a free group generated by $|E| - |V| + 1$ elements. Here is one way to specify a set of generators. Take a spanning tree of (V, E). For each edge e in the set E_0 of edges outside the spanning tree, the path w_e follows the path in the tree from x to e_-, moves along e and then back in the tree to x. Clearly, $|E_0| = |E| - |V| + 1$.

Each path from x is described by a word in the generators g_i and their inverses. If two paths are equivalent, then the corresponding words w have evaluate at x to

the same element of K. Moreover, a word is in G_x if and only if it corresponds to a cycle.

It follows that G_x is generated by the words corresponding to generators of π_1. It suffices to show that for the generators discussed above, $\{w_e/G_{x+}\}_{e\in E_0}$ are independent random elements of G_x/G_{x+}.

For an edge $e = (v, g^i) \in E$, let $g(e) = g_i(v)$. Respectively, let $g(e) = g_i(v)^{-1}$ for the reversed edge. If $w = e_1 \dots e_k$ is a cycle starting from x, then

$$w/G_{x+} = g(e_1) \dots g(e_k).$$

Now fix the values of $g(e)$ corresponding to $e \in E \setminus E_0$. Then for $e \in E_0$, w_e/G_{x+} equals $g(e)$ multiplied on the left and on the right by fixed elements. Since the $\{g(e)\}_{e\in E_0}$ are uniform random and independent, so are the $\{w_e/G_{x+}\}_{e\in E_0}$, as required. □

Corollary 3.6. *For $v \in T$ of orbit length k the subgroup G_v/G_{v+} has the same distribution as $G_{(j-1)k+1}$.*

The following fact is also immediate.

Fact 3.7. *Let $v_1, \dots, v_k$ be vertices of T. Let A denote the event that none of the vertices $w(v_i)$ of T_H equals another or an ancestor of another. Conditioned on A, the subgroups G_{v_i}/G_{v_i+} are independent.*

Proof of Proposition 3.2. Proposition 3.2 easily follows from Corollary 3.6, Fact 3.3, and Fact 3.7. □

Let $q(j)$ denote the probability that j random elements generate a transitive subgroup G of H. Looking at the probability that T_G does not branch we get the following

Corollary 3.8.

$$\mathbf{P}(G \text{ is transitive on level } n) = \prod_{\ell=0}^{n-1} q\left(1 + (j-1)|X|^\ell\right). \tag{10}$$

A subgroup of $W(H)$ is called **spherically transitive** if it acts transitively on every level of the tree. As an example of Corollary 3.8, consider two random elements of the automorphism group of the binary tree. In this case $j = 2$, $H = C_2$, we have $q(n) = 1 - 2^{-n}$, and

$$\mathbf{P}(G \text{ is spherically transitive}) = \prod_{\ell=0}^{\infty}(1 - 2^{-2^\ell - 1}) \sim 0.63.$$

In the general case, we have

Corollary 3.9. *Let $j > 1$, and assume that H has a transitive subgroup generated by j elements. Then with positive probability, j independent Haar random elements of $W(H)$ generate a spherically transitive subgroup. Moreover, this probability tends to 1 as $j \to \infty$.*

Proof. Clearly, $1 - q(nj) \le (1 - q_j)^n$. This means that $1 - q(1 + (j-1)|X|^\ell)$ is summable, so for $n = \infty$ the product (10) is positive. □

Proposition 3.10. *Let G be the subgroup generated by two independent Haar random elements of $W(H)$. Then with probability 1, T_G has only finitely many rays. Equivalently, the number of orbits at level ℓ remains bounded as $\ell \to \infty$.*

Proof. Consider the vertices of T_G that are orbits of length 1 in T. These induce a subtree with a subcritical GW distribution (the number of fixed points of just one random element has expectation 1 by Pólya theory).

This implies that with probability 1 (1) there exist finitely many fixed vertices of T, (2) there exist finitely many vertices of T with orbits of length at most k for each k.

On the basis of Corollary 3.9 let k be so large that the chance that $(|X|-1)k+1$ random elements generate a transitive subgroup of $W(H)$ is at least $1-1/|X|$.

Now consider the tree built to a level where each orbit has at least k elements. Then each such orbit will stop splitting forever with probability at least $1-1/|X|$, and if it does not, then it splits into at most $|X|$ offspring with probability at most $1/|X|$. Thus the orbit process is dominated by the subcritical GW tree with 0 or $|X|$ offspring with probability at least $1-1/|X|$ and at most $1/|X|$, respectively. Hence all orbits stop splitting with probability 1. □

4. Word maps

Let G be a group with finite Haar measure. A **randomly evaluated word** in G is a G-valued random variable defined by a word w in the free group F_k by substituting k independent Haar random elements of G into the generators of F_k.

Randomly evaluated words are perhaps the simplest possible maps $G^k \to G$. Bhattacharjee [1995] showed that with probability 1, a nontrivial word randomly evaluated in $W(H)$ does not give the identity; moreover, it does not stabilize a fixed ray in the tree. In this section we strengthen this result in two directions, and give a sufficient condition for a list of randomly evaluated words to have independent uniform distribution.

Our first result is key in answering a question of Sidki [2001].

Proposition 4.1. *A nontrivial randomly evaluated word in $W(H)$ fixes only finitely many vertices of T with probability 1.*

Proof. Let $w = w_1 \dots w_\ell$, where each w_i is one of the random generators or its inverse, and w is in reduced form. We use induction on the length ℓ of w. The $\ell = 1$ case is exactly Corollary 2.7.

Let $\overline{w}$ denote the random evaluation of the word w. Let $v \in T$ be a vertex. Let $w_i' = w_1 \dots w_i$, $v_i = v^{\overline{w}_i'}$. Make the assumption A that $v^{\overline{w}} = v_\ell = v$, but $v_i \neq v_j$ for $0 \le i < j < \ell$. We use the notation $g(v)$ for the action of $g \in W(H)$ on the descendant subtree of v. Then

$$\overline{w}(v) = \prod_{i=1}^{\ell} \overline{w}_i'(v^{\overline{w}_{i-1}'}).$$

Given the event A, the factors in the product are independent Haar random elements of $W(H)$, and therefore $\overline{w}(v)$ is also a Haar random element of $W(H)$. This means that with probability one, by Corollary 2.7, only finitely many of the descendants of v are fixed by $\overline{w}$.

Now we are ready to prove the proposition. By the inductive hypothesis, there are only finitely many vertices v which are fixed by the evaluation of any of the

proper sub-words $w_i \dots w_j$, $i < j$, of w. Let L be the greatest level at which there is such a vertex. Then at level $L+1$ event A holds for all vertices v that are fixed by $\overline{w}$. In particular, all vertices at this level have finitely many offspring that are fixed by $\overline{w}$, as required. □

This proof generalizes to the infinite wreath product of arbitrary (not necessarily identical) finite transitive permutation groups. From probabilistic point of view, this property is interesting because it describes a critical phenomenon in the statistical physics sense. One manifestation of this is that the probability that a word is satisfied on levels 1 to n decays polynomially in n, as opposed to exponentially (see Corollary 2.6).

The following corollary is immediate.

Corollary 4.2. *Let G be a subgroup abstractly generated by k random elements of $W(H)$. Then G is a free group acting freely on ∂T.*

In fact, we have that all elements of the countable group G have only finitely many fixed vertices in T. We call a countable free subgroup G of $W(H)$ with this property **strongly free**. We are ready to prove Theorem 4, which we restate here.

Theorem 4. *Let $G \subseteq \Gamma(H)$ be a strongly free subgroup and let $g \in \Gamma(H)$ be a random element. Then the groupd $\langle G, g\rangle$ is strongly free withd probability* 1.

Proof. The proof of Proposition 4.1 did not use the full power of randomness of all generators; it was sufficient to have one random generator.

In fact, we have the following corollary to the proof of Proposition 4.1. Let g_1 be random and $g_2, \dots, g_k$ fixed elements of $W(H)$. Let w be a reduced word in the g_i, containing g_1 or its inverse. If the evaluation all proper sub-words of w fixes finitely many vertices of T with probability 1, then the same is true for $\overline{w}$.

Theorem 4 follows. □

The **odometer** s is an element of $W(p)$ that acts as a full cycle on $W(p)$, or equivalently, one that has an orbit tree given by a single ray. More precisely, it is a specific such element, defined by the following identity. For a vertex $v = (v_1, \dots, v_n)$ at level n let $N(v) = v_1 + v_2 p + \dots + v_n p^{n-1}$. Then for every n and vertex v at level n we have

$$N(v^s) = N(v) + 1 \qquad (\mathrm{mod}\, p^n). \tag{11}$$

The odometer s is one of the simplest examples of elements of $W(p)$ that are given by finite automata. Groups generated by such elements include Grigorchuk's group, and have a rich theory; see Grigorchuk et al. [2000b]. Sidki [2000] showed that s and another element given by a finite automaton cannot generate a free subgroup of $W(p)$. He asked (Sidki [2001]) whether the odometer and *any* other element can generate a free group. Since the odometer generates a strongly free cyclic subgroup, we have

Corollary 4.3 (Answer to a question of Sidki)**.** *The odometer and a random element abstractly generate a free subgroup of $W(p)$ with probability* 1.

Consider the following norm $\|\cdot\|$ and distance dist on the group $W(H)$. Let n be the greatest number so that g fixes all vertices at level n, and let

$$\|g\| = |W_n|^{-1}, \quad \mathrm{dist}(g,h) := \|gh^{-1}\|, \tag{12}$$

where $|W_n|$ is just the order of W_n. The L_1-metric on the product space $W(H)^k$ gives rise to a notion of Hausdorff dimension in the usual way. If $S \subseteq W(H)^k$, then it is easy to check that upper Minkowski dimension, which dominates Hausdorff dimension, satisfies

$$\dim_{\overline{M}} S = \limsup_{n\to\infty} \left(\log |S/(W^{[n]})^k| \,/\, \log |W_n| \right). \tag{13}$$

Simple maps $\mathbb{R}^k \to \mathbb{R}$ tend to have the property that the pre-images of points are at most $k-1$ dimensional. A word in k letters evaluated in $W(H)$ is a simple map $w : W(H)^k \to W(H)$. It is natural to ask whether the analogous statement is true here. The answer is no, for $k = 1$ a counterexample is the map $x \mapsto x^{p^k}$ in $W(H)$. If $g \in W(H)$ acts as a full cycle $(v_1, \dots, v_{p^\ell})$ on level ℓ of the tree, and the action $g(v_i)$ on the descendant subtrees satisfy $g(v_1) \dots g(v_{p^\ell}) = 1$, then g is in the kernel of this map. The set of such elements has dimension $1 - p^{-\ell}$.

The 1-dimensional example above easily generalizes to show that the kernel of a k-letter word map can have Hausdorff dimension arbitrarily close to k. However, the kernel of a map cannot be full dimensional, as the following restatement of Theorem 11 shows.

Theorem 4.4. *Let w be a nontrivial word in the letters $g_1, \dots, g_k$, and let*

$$\mathcal{K} = \{(g_1, \dots, g_k) \in W(H)^k \,:\, w = 1\}.$$

Then $\dim_{\mathrm{H}}(\mathcal{K}) \le \dim_{\overline{M}}(\mathcal{K}) < k$.

Proof. The first inequality holds in every metric space. We first make the assumption (i) that there exists elements in H for which $w \neq 1$. Let ℓ denote the length of the word (i.e., the sum of the absolute values of the exponents), and let $b = |X|$. Fix the action $A_n = (g_i/W^{[n-1]};\ 1 \le i \le k)$ of the variables on T_{n-1}, and let $s(A_n)$ denote the number of liftings of these elements to W_n for which $w = 1$. Let $h = |H|$, and let $t_n = h^{kb^n}$ denote the total number of liftings. Let $s_n = \max_{A_n} s(A_n)$. Then we have

$$|\mathcal{K}/(W^{[n]})^k| \le s_1 \dots s_n,$$

and therefore by (13)

$$\dim_{\overline{M}} \mathcal{K} \le k \limsup \frac{\log(s_1 \dots s_n)}{\log(t_1 \dots t_n)} \le k \limsup \frac{\log s_n}{\log t_n}.$$

The last inequality is simple arithmetic.

Since the action on T_{n-1} is fixed, for $v \in \partial T_{n-1}$ we have that $w(v)$ is a word in the letters $g_i(u)$, $u \in \partial T_{n-1}$. More precisely, we get $w(v)$ by replacing each occurrence of g_i in w by some $g_i(u)$, where the u depend on the action A_n and need not be the same for different occurrences of g_i. Together with our assumption (i), this implies that there is a counterexample to $w(v)$ in H.

We claim that there exists a large subset $S \subseteq \partial T_{n-1}$ so that if $u, v \in S$, then the words $w(u)$, $w(v)$ have no letter in common. Indeed, each word uses at most ℓ letters, and conversely, each letter appears in at most ℓ words (more precisely, the number of words using $g_i(u)$ is at most the total number of times g_i is used in w). This means that the graph in which u, v are neighbors iff $w(u)$, $w(v)$ share a letter has maximal degree at most $d = \ell(\ell - 1)$ (ignoring loops). Such a graph has an independent set S of size at least $\lceil |\partial T_{n-1}|/(d+1) \rceil$ (using the greedy algorithm).

Let $v \in S$ and consider the set of letters L that contribute to $w(v)$. The number of assignments of values in H to the letters L is $h^{|L|}$; the number of assignments for which $w(v) = 1$ is at most $h^{|L|} - 1$. Since these sets of letters are disjoint for different v, we have that

$$s(A_n) \leq t_n((h^\ell - 1)/h^\ell)^{|S|}$$

and therefore with $\varepsilon = \log(h^\ell/(h^\ell - 1))$, we get

$$\log s_n \leq \log t_n - \varepsilon b^n/(d+1) \leq (1 - \varepsilon') \log t_n,$$

where $\varepsilon' > 0$ does not depend on n. This implies the claim of the proposition when assumption (i) holds. Otherwise, for some finite j the group $H' = W_j(H)$ has a counterexample, since $W(H)$ contains a free subgroup of rank k (see Bhattacharjee [1995] and Corollary 4.2 here). Since the natural isomorphism between $W(H)^k$ and $W(H')^k$ preserves full-dimensionality, the proposition applied to H' concludes the proof. □

Remark 4.5. We have shown that the pre-image of the element $1 \in W(H)$ is not full dimensional. It is easy to modify this proof to get that the pre-image of any point $g \in W(H)$ has dimension at most $1 - \varepsilon'(w)$, where $\varepsilon'(w)$ does not depend on g.

Corollary 4.6. *Let $G \subseteq W(H)$ be a subgroup of Hausdorff dimension* 1*. Then k random elements generate a free subgroup with probability* 1*.*

Proof. It suffices to show that for each word w in k letters, the probability that $\overline{w} = 1$ in G is zero. This probability equals the measure of $\mathcal{K}_w$ in G^k. But G^k is a full-dimensional subgroup of $W(H)^k$, so every subset of G^k with positive measure has full dimension in $W(H)^k$. □

Let G be a finite p-group. Our study of random generation requires a sufficient condition for a set of randomly evaluated words to have independent uniform distribution on G. For a word w in the variables g_i define the **exponent sum vector** of w as the vector whose coordinate i is the sum of the exponents of g_i in w. Let $\overline{w}$ denote the evaluation of the word w. We have not been able to locate the proof of the following simple lemma in the literature.

Lemma 4.7 (Linear and probabilistic independence)**.** *Let $w_1, \ldots, w_k$ be words in the letters $g_1, \ldots, g_{n+m}$, so that the exponent sum vectors restricted to the first n coordinates are linearly independent mod p. Fix $g_{n+1}, \ldots, g_{n+m} \in G$. Then the substitution map $G^n \to G^k$,*

$$(g_1, \ldots, g_n) \mapsto (\overline{w}_1, \ldots, \overline{w}_k),$$

covers G^k evenly, i.e. it is $|G|^{n-k}$-to-one and onto.

A special case of Lemma 4.7 says that if w is a word in which the exponent sum of some letter is relatively prime to the order of a finite p-group G, then the evaluation map covers G evenly. We state without proof that this also holds for nilpotent groups but not for arbitrary finite groups.

Proof. Assume that the $g_1, \ldots, g_n$ are chosen independent uniformly and at random (i.u.). It suffices to show that the evaluations $\overline{w}_j(g)$ are i.u. We prove this by induction; it is clearly true for $G = 1$. Let $Z \lhd G$ be a subgroup of order p of the center of G, and our inductive hypothesis is that the claim holds for G/Z. Let R

be a set of coset representatives of Z, and let $r : G \to R$ be a coset representative map; we will also think of r as a map $G/Z \to R$.

Write $g_i = z_i r(g_i)$. Then $z_i \in Z$, $r(g_i) \in R$ are jointly i.u. for $i \leq n$. Thus $z_i \in Z$, $r(g_i)/Z \in G/Z$ are jointly i.u. for $i \leq n$. Because the z_i are in the center, we have

$$\overline{w}_i(g) = \overline{w}_i(z)\overline{w}_i(r(g)) = \overline{w}_i(z)b_i r(\overline{w}_i(g)) = \overline{w}_i(z)b_i r(\overline{w}_i(g/Z)),$$

where $b_i \in Z$ and the random vector b depends on the vector $r(g)$, but not on z. By the inductive hypothesis, the vector $r(\overline{w}_i(g/Z))$ is i.u. in R^n. The assumption implies that $\overline{w}_i(z)$ are i.u., and therefore given b, $\overline{w}_i(z)b_i$ are i.u. Thus $\overline{w}_i$ are products of i.u. elements of Z and i.u. coset representatives, and these $2n$ random variables are also jointly independent. The claim follows. □

By considering finite quotients, we see that the assumption of Lemma 4.7 implies that for any pro-p group G the substitution map is a measure-preserving surjection.

5. Schreier graphs and homology

The goal of this section is to establish some tools for proving Theorem 3 about random generation.

Let K be a p-group acting on a set X, and let $\mathcal{G} = \mathcal{G}(K, X, S)$ be the Schreier graph as in Definition 3.5.

Let $\tau\mathcal{G}$ denote the vector space of mod p **1-chains**, that is, the set of abstract linear combinations of the edges of $\mathcal{G}$ with coefficients in $\mathbb{F}_p$.

If w is a word in S, i.e., an element of the free group indexed by S, then for each $v \in X$, $w(v)$ can be thought of as a path $(v, v^{\overline{w}_1}, v^{\overline{w}_n})$, where the $w_1, \dots, w_n$ are the initial sub-words of w and $\overline{w}$ denotes evaluation in K.

For a path π, let $\tau\pi \in \tau\mathcal{G}$ denote the **1-chain of** π, where each edge appears with a coefficient given by the number of times it is used in π.

Recall that the **exponent sum vector** of w is the vector whose coordinate i is the sum of the exponents of g_i in w.

Lemma 5.1 (Independent 1-chains)**.** *Let Υ be a set of words in S whose exponent sum vectors are linearly independent. Then $\{\tau w(v) \ : \ w \in \Upsilon, v \in X\}$ are linearly independent.*

More generally, let Υ be a set of words in S whose exponent sum vectors restricted to $S' \subseteq S$ are linearly independent. Then $\{\tau w(v) \ : \ w \in \Upsilon, v \in X\}$ restricted to S'-edges are linearly independent.

Proof. We prove the more general version. Let $S = \{g_1, \dots, g_\ell\}$, let $n \leq \ell$, and let $S' = \{g_1, \dots, g_n\}$. Without loss of generality we may extend the set Υ so that $|\Upsilon| = n$, and the exponent sum vectors are still linearly independent. Let $w_1, \dots, w_n$ denote the elements of Υ. Consider the wreath product $D = C_p \wr K$. For $i > n$ let $d_i := ((0, \dots, 0), g_i) \in C_p \wr K$. Consider the map that evaluates the words w_i at elements $d_j \in D$ (d_j is a variable for $j \leq n$ and is fixed for $j > n$):

$$\begin{aligned} \varphi : D^n &\to D^n, \\ \varphi_i : (d_1, \dots, d_n) &\mapsto \overline{w}_i. \end{aligned}$$

This map φ is a bijection because of the linear independence assumption and Lemma 4.7. Now restrict φ to all possible extensions of $(g_1, \dots, g_n) \in K^n$, and let

$$(\nu_i', g_i') = \varphi_i\left((\nu_1, g_1), \dots, (\nu_n, g_n)\right).$$

Then the g_i' are determined by the g_i, and φ defines an isomorphism of vector spaces

$$\begin{aligned} f : \mathbb{F}_p^{|X|n} &\to \mathbb{F}_p^{|X|n}, \\ (\nu_1, \dots, \nu_n) &\mapsto (\nu_1', \dots, \nu_n'). \end{aligned}$$

If the reduced form $w \in \Upsilon$ is $a_1 \cdots a_m$, then for $v \in X$ the evaluation in D satisfies

$$\overline{w}(v) = \overline{a}_1(v)\overline{a}_2(v^{\overline{a}_1}) \cdots \overline{a}_m(v^{\overline{a}_1 \cdots \overline{a}_{m-1}}). \tag{14}$$

Here the notation $g(v)$ means the value of g at v as in the definition of the wreath product (2). Since the edges of the subgraph $\mathcal{G}(X, S')$ are of the form

$$\{(v, g_i) : v \in X, 1 \le i \le n\},$$

which agree with the coordinates of $\nu = (\nu_1, \dots, \nu_n)$, we may think of ν as a 1-chain for $\mathcal{G}(X, S')$. Then (14) implies that for $w \in \Upsilon$, $v \in X$ we have

$$f(\nu)(w, v) = \overline{w}(v) = \tau' w(v) \cdot \nu,$$

where $\tau' w(v)$ is the restriction of $\tau w(v)$ to S'-edges, and $\cdot$ is the natural $\mathbb{F}_p$ scalar product of 1-chains in $\mathcal{G}(X, S')$. Since f is bijective, the statement of the lemma follows. □

The subspace of $\tau\mathcal{G}$ generated by $\tau\pi$ for cycles π is called the (first) **homology group** $\mathcal{H}\mathcal{G}$ of $\mathcal{G}$. For a word w and $v \in X$ let $w(\circ v)$ denote the path $(w^\ell)(v)$, where ℓ is the smallest positive power so that $\overline{w}^\ell$ fixes v. Note that the path $w(\circ v)$ is a cycle, and so $\tau w(\circ v) \in \mathcal{H}\mathcal{G}$.

Lemma 5.4 is important for our study of random generation. We will need to show that the homology of certain cycles in a Schreier graph of a p-group K with 3 variables $S = \{g_1, g_2, g_3\}$ is rich enough. The generators g_1 and g_2 provide a skeleton, and g_3 is needed for extra flesh. We first need to make some definitions.

Definition 5.2. For $v \in X$, let $\kappa(v)$ denote the smallest positive power of p so that there exists a word w in $\{g_1, g_2, g_* = g_3^\kappa\}$ so that $v^{\overline{w}} = v$ and the exponent sum of g_* in w is not divisible by p. Since $g_3^{|K|} = 1$ fixes v, we have $\kappa(v) \le |K|$. Let v_X denote a vertex for which κ is minimal, let w_X denote the corresponding word, and let $\kappa(X) = \kappa(v_X)$.

Assume that

$$g_* = g_3^{\kappa(X)} \text{ fixes no point in } X. \tag{15}$$

Let $\tau_3\pi$ denote the 1-chain of a path π restricted to g_3-edges of $\mathcal{G}(K, X, \{g_1, g_2, g_3\})$. For a $\{g_1, g_2, g_*\}$-path π, let $\tau_*\pi$ denote the 1-chain in $\mathcal{G}(K, X, \{g_1, g_2, g_*\})$ restricted to g_*-edges.

If w is a word in which the exponent sum of g_3 is not divisible by p, then it follows from Lemma 5.1 that $\{\tau_3 w(u) : u \in X\}$ are linearly independent. For the proof of Lemma 5.4, the ideal situation is when there is such a word satisfying $v^{\overline{w}} = v$. But it can happen that there is no such word, so we need the following extra lemma.

Lemma 5.3 (Independence for added noise)**.** *Assume* (15)*, and let* $u \in X \setminus \{v_X\}$*. The vectors* $\tau_3(w_X(v_X))$*,* $\tau_3(w_X(\circ u))$ *are linearly independent.*

Proof. Let $\kappa = \kappa(X)$, $v = v_X$, $w = w_X$, and let $g_* = g_3^\kappa$. Consider the partition of X into $\{g_1, g_2, g_*\}$-orbits $\mathcal{Q}$. For $x \in X$, let $Q_x \in \mathcal{Q}$ denote the orbit of x.

Let $Q \in \mathcal{Q}$. We claim that the sets Q, Q^{g_3}, $\dots$, $Q^{g_3^{\kappa-1}}$ are disjoint. Suppose the contrary; then for $x, z \in Q$, we have $x^{g_3^i} = z^{g_3^j}$ with $0 \le i < j < \kappa$. Then $x = z^{g_3^{j-i}}$. Let w_{xz} be the word in $\{g_1, g_2, g_*\}$ so that $x^{\overline{w}_{xz}} = z$. Then the evaluation of the word $w_{xz} g_3^{j-i}$ fixes z and its g_3-exponent sum is not divisible by κ. But this contradicts the minimality in the definition of κ.

Consider the homomorphism φ of 1-chains defined by

$$\begin{array}{rcl} \varphi \,:\, \tau\mathcal{G}(K, Q, \{g_*\}) & \longrightarrow & \tau\mathcal{G}(K, X, \{g_3\}), \\ \tau_* g_*(x) & \mapsto & \tau_3 g_*(x) \end{array}$$

which maps the 1-chain of a g_*-edge to the 1-chain of its expansion, a path of length κ. By the previous paragraph, for different $x, z \in Q$ the paths $g_*(x)$, $g_*(z)$ in $\mathcal{G}(X, \{g_3\})$ have disjoint edges. It follows that φ is injective.

Since the exponent sum of g_* in w is not divisible by p, we can apply Lemma 5.1 to the Schreier graph $\mathcal{G}(K, Q, \{g_1, g_2, g_*\})$ and $S' = \{g_*\}$. We get that the 1-chains $\{\tau_*(w(x)) \,:\, x \in Q\}$ are linearly independent. Since φ is an injective homomorphism, $\{\tau_3(w(x)) : x \in Q\}$ are linearly independent. In particular, $\tau_3(w(v)) \neq 0$.

Let $x \in X$ and let q be the minimal positive integer so that $x^{\overline{w}^q} = x$. Then

$$\tau_3(w(\circ x)) = \sum_{\ell=0}^{q-1} \tau_3(w(x^{\overline{w}^\ell})) \neq 0 \tag{16}$$

since $x, x^{\overline{w}}, \dots, x^{\overline{w}^{q-1}}$ lie in the same $\{g_1, g_2, g_*\}$-orbit Q_x and so the terms are linearly independent. This implies that if $u \in Q_v \setminus \{v\}$, then the claim of the lemma holds.

Now assume that $u \notin Q_v$, i.e., Q_u and Q_v are disjoint. Recall the definition of the boundary operator ∂. It is a linear map from the 1-chains $\tau\mathcal{G}$ of a graph $\mathcal{G}$ to 0-chains i.e., abstract linear combinations of vertices of $\mathcal{G}$. It is defined by $(x, z) \mapsto z - x$ for each directed edge (x, z). The kernel of ∂ is the homology group $\mathcal{H}\mathcal{G}$.

For $x \in X$, $\tau_3 w(\circ x)$ is a linear combination of 1-chains of paths with length κ that start and end in Q_x. Thus $\partial\tau_3 w(\circ x)$ is supported on Q_x. In particular, $\partial\tau_3 w(v)$ and $\partial\tau_3 w(\circ u)$ are supported on Q_v and Q_u, respectively. Therefore it suffices to show that $\partial\tau_3 w(v) \neq 0$.

We first claim that $\partial\tau_* w(v) \neq 0$, in other words $\tau_* w(v)$ is not a linear combination of 1-chains of g_*-cycles. Indeed, all such cycles are of p-power length, and by assumption (15) there is no cycle of length 1. But $\tau_* w(v)$ cannot be a linear combination of longer cycles, since the exponent sum of g_* in w is not divisible by p.

If t is a one-chain in $\tau\mathcal{G}(K, Q, \{g_*\})$, then $\partial t = \partial\varphi(t)$; this is clear for generators (one-chains supported on single edges), and the general case follows by linearity. In particular, $\partial\tau_3 w(v) = \partial\varphi(\tau_3 w(v)) = \partial\tau_* w(v) \neq 0$, completing the proof. □

For a set of words Υ, and an element f of the free group $\mathcal{F}_\Upsilon$ indexed by Υ, let w_f denote the concatenation of words $w \in \Upsilon$ according to f.

Lemma 5.4 (Words with special homology)**.** *Assume* (15)*. There exists a set* Υ *of* $|v_X^{\langle g_1, g_2\rangle}| + 1$ *words in* $\{g_1, g_2\}$ *fixing* v_X*, so that for each fixed* $f \in \mathcal{F}_\Upsilon$ *and fixed*

$u \in X \setminus \{v_X\}$ the 1-chains

$$\tau(w_X w_f)(\circ u), \quad \{\tau w(v_X) \,:\, w \in \Upsilon \cup \{w_X\}\}$$

are linearly independent.

Proof. Let Υ be the generating set for the fundamental group of the Schreier graph $\mathcal{G}_{12} = \mathcal{G}(K, v_X^{\langle g_1, g_2\rangle}, \{g_1, g_2\})$ defined in the proof of Lemma 3.4, i.e., pick a spanning tree rooted at v_X, and for each edge (v_w, g_{i_w}) outside the tree consider the path $w(v_X)$ that moves from v_X to v_w in the tree, then along the edge, and finally back to v_X in the tree.

We need to check that if the linear combination

$$a_u \tau(w_X w_f)(\circ u) + a_X \tau(w_X(v_X)) + \sum_{w \in \Upsilon} a_w \tau(w(v_X)) = 0 \qquad (\mathrm{mod}\, p),$$

then all the coefficients a_i equal 0 $(\mathrm{mod}\, p)$. Indeed, by Lemma 5.3, the first two terms are linearly independent when restricted to g_3 edges, and the terms in the last sum vanish on g_3-edges, thus we have $a_+ = a_u = 0$. It is standard that $\tau\Upsilon$ are independent: they form a basis for the homology group $\mathcal{H}\mathcal{G}_{12}$. In particular, the edge (v_w, g_{i_w}) defined in the previous paragraph appears in exactly one of $\tau(v_X, w)$, so $a_w \equiv 0$. □

6. Slices of random subgroups

The goal of this section is to show that slices of random subgroups of $W(p)$ are large; in the next section we will show that these large slices generate a large subgroup.

The setup is as follows. Let K be a finite p-group acting on a set X.

Let $H = W_n(p)$. The group $H \wr K$ acts on the disjoint union of $|X|$ copies of T_n. For a subgroup $G_* \subseteq (H \wr K)$, let $\sigma G_* \subseteq G_*$ denote the **boundary slice**, that is, the pointwise stabilizer of the vertices on level $n-1$ of the trees. Note that σG_* is naturally an $\mathbb{F}_p$-vector space.

We first consider the case $|X| = |K| = 1$. The following lemma shows that it is possible to generate large slices by two elements, even if one is required to be a high power.

Lemma 6.1. *For $0 \le k < n$ there exists elements $s, g \in W_n(p)$ so that*

$$\dim(\sigma\langle s, g^{p^k}\rangle) = p^{n-1} - p^k + 1.$$

The following proof uses a technique introduced in Abért and Virág [In prep], but is self-contained.

Proof. Each element $v \in \partial T_{n-1}$ is represented by a sequence $(a_1, \ldots, a_{n-1})$ of elements of $\mathbb{F}_p$. We assign to each such element a number

$$N(v) = a_1 + a_2 p + \ldots + a_n p^{n-2}.$$

Recall (from the definition before (11)) that there exists an element, the odometer $s \in W(p)$, that acts as a full cycle on ∂T_{n-1} so that $N(v^s) = N(v) + 1 \ (\mathrm{mod}\, p^{n-1})$.

Each $g \in \sigma W_n$ we identify with the polynomial in $F_n := \mathbb{F}_p[x]/(x^{p^{n-1}} - 1)$ given by

$$\sum_{v \in \partial T_{n-1}} g(v) x^{N(v)}.$$

This is an isomorphism of vector spaces, and the action of s on σW_n corresponds to multiplication by x in F_n. Define the **weight** $\nu(f)$ of an element $f \in F_n$ as the highest ℓ so that $y^\ell = (x-1)^\ell$ divides f. It is clear that elements of F_n of different weight are linearly independent. Also $\nu(xf - f) = \nu(yf) = \nu(f) + 1$, unless $\nu(f) = p^{n-1} - 1$ (since $y^{p^{n-1}} = x^{p^{n-1}} - 1 = 0$). This implies that if an s-invariant subspace M of σW_n contains an element of weight ν, then it also contains an element of weight $\nu + 1$, then one of $\nu + 2$, up to weight $p^{n-1} - 1$. Hence $\dim M \geq p^{n-1} - \nu$. (Note that this gives a description of the so-called uniserial action of s on σW_n.)

Consider the element $g \in W_n$ with the following properties. The action of g agrees with the action s on ∂T_k. Also, for vertices v at levels greater than k, $g(v) = 0$, except that for the vertex $v \in \partial T_{n-1}$ with $N(v) = 0$ we have $g(v) = 1$.

Let $q = p^k$. One easily checks that $g^q \in \sigma W_n$. Moreover, for $v \in \partial T_{n-1}$ we have that $g^q(v) = 1$ if $N(v) < q$ and $g^q(v) = 0$ otherwise. Thus g^q corresponds to the polynomial

$$f = 1 + x + \ldots + x^{q-1} = y^{q-1}$$

of weight $q-1$ (the last equality can be checked via binomial expansion). Therefore the s-invariant subspace containing g^q has elements of all weights in $q-1 \ldots p^{n-1}-1$. The claim of the lemma follows. □

Now assume that K is generated by three elements $\tilde{g}_1, \tilde{g}_2, \tilde{g}_3$. Define κ, v_X, w_X as in Definition 5.2 (with tildes removed from g's), and assume (15). Consider the subgroup $G \subseteq H \wr K$ generated by uniform random liftings g_i of the generators $\tilde{g}_i$. Let

$$\varepsilon_* = \mathbf{P}(|v_X^{\langle \tilde{g}_1, \tilde{g}_2 \rangle}| + 1 \text{ independent random elements do not generate } H). \tag{17}$$

Lemma 6.2 (Controlled randomness in wreath products)**.** *Assume* (15). *Let $h \in H$. There exists a word $w_{(h)}$ in the g_i whose composition depends on the g_i so that $v_X^{\overline{w}_{(h)}} = v_X$, $\mathbf{P}(\overline{w}_{(h)}(v_X) = h) \geq 1 - \varepsilon_*$, and for $u \in X \setminus \{v_X\}$ the distribution of $\overline{w}_{(h)}(\circ u)$ is uniform on H.*

Remark 6.3. In the last step of the proof we will need the following simple fact. Let $Z, \{Y_i,\, i \in I\}$ be random variables so that for each $i \in I$, the two variables Z, Y_i are independent and Y_i has distribution μ. Let $f(z)$ be an I-valued deterministic function. Then $Z, Y_{f(Z)}$ are also independent and $Y_{f(Z)}$ has distribution μ.

Proof of Lemma 6.2. We will use the notation of Lemma 5.4. The linear independence statement of Lemma 5.4 translates to probabilistic independence by Lemma 4.7: for each fixed $f \in \mathcal{F}_\Upsilon$, and $u \neq v_X \in X$, the random variables

$$\{\overline{w}(v_X) : w \in \Upsilon \cup \{w_X\}\} \text{ and } \overline{w}_X \overline{w}_f(\circ u) \text{ are independent and uniform on } H. \tag{18}$$

Consider the group generated by $\{\overline{w}(v_X) : w \in \Upsilon\}$; let A denote the event that this is the full group H. Then $\mathbf{P}(A) = 1 - \varepsilon_*$.

Order the elements of the free group $\mathcal{F}_\Upsilon$ in an arbitrary fixed way. On the event A, let $f \in \mathcal{F}_\Upsilon$ be the smallest element so that $\overline{w}_X \overline{w}_f(v_X) = h$. On the complement A^c let $f = 1$.

Let $w_{(h)} = w_X w_f$. The claim (18) also holds for our random choice of f as it only depends on the values $\{\overline{w}(v_X) : w \in \Upsilon \cup \{w_X\}\}$. Following Remark 6.3, we

set $Z = \{\overline{w}(v_X) : w \in \Upsilon \cup \{w_X\}\}$, $I = F_\Upsilon$, $Y_f = \overline{w}_X \overline{w}_f(\circ u)$, and $f = f(Z)$ only depends on Z. □

The **rigid vertex stabilizer** $\mathcal{R}_v G \subseteq G$ of a vertex v (respectively, $\mathcal{R}_V G$ for a vertex set V) is the pointwise stabilizer of the vertices not descendant to v (respectively, vertices not descendant to some $v \in V$).

Lemma 6.4 (Slices for a descendant tree of a vertex)**.** *Assume condition* (15). *Let* $\theta = |v_X^{\langle \tilde{g}_1, \tilde{g}_2 \rangle}|$. *There exists* $c_0, c_1, c_2, c_3 > 0$ *depending on* p *only so that if* $|X| \le p^{c_0 n}$, *then the event*

$$\dim \sigma \mathcal{R}_{v_X} G \ge (1 - p^{-c_1 n + 1}) \dim \sigma W_n \tag{19}$$

has probability at least $1 - e^{-c_2 n}|X| - e^{c_3(n-\theta)}$. *Moreover,* (19) *implies that for the orbit* v_X^G *we have*

$$\dim \sigma \mathcal{R}_{v_X^G} G \ge (1 - p^{-c_1 n + 1})|v_X^G| \dim \sigma W_n. \tag{20}$$

Proof. By Proposition 2.8 there exists $c_4 < 1$ and $c_2 > 0$ so that for all n a uniformly chosen random $h_0 \in H = W_n(p)$ has $\mathbf{P}(h_0^{p^{\lfloor c_4 n \rfloor}} \ne 1) \le e^{-c_2 n}$. Now let $c_0 \in (0, 1 - c_4)$, let $c_1 = c_4 + c_0 < 1$, and let $k = \lfloor c_1 n \rfloor$.

Let $v = v_X$, let $G_v \subseteq G$ denote the stabilizer of v, and consider the projection

$$\begin{aligned} \pi : G_v &\to H, \\ g &\mapsto g(v). \end{aligned}$$

Consider the two elements g, s used in Lemma 6.1 for generating a large slice. Recall the definition of ε_* in (17). Two applications of Lemma 6.2 show that there exist words $w_{(g)}, w_{(s)}$ so that

$$\mathbf{P}(\overline{w}_{(g)}(v) = g,\ \overline{w}_{(s)}(v) = s) \ge 1 - 2\varepsilon_*, \tag{21}$$

and for $u \in X \setminus \{v\}$ the element $\overline{w}_{(g)}(\circ u) \in H$ is uniform random. Let p^r be the length of the orbit of u under the action of $\overline{w}_{(g)}$; clearly, $p^r \le |X| \le p^{c_0 n}$. Let $h = \overline{w}_{(g)}^{p^k}$. Then

$$h(u) = (\overline{w}_{(g)})^{p^k}(u) = (\overline{w}_{(g)}(\circ u))^{p^{k-r}},$$

which is the p^{k-r}-th power of a uniform random element of H, where

$$k - r \ge \lfloor c_1 n \rfloor - \lfloor c_0 n \rfloor = \lfloor c_1 n - \lfloor c_0 n \rfloor \rfloor \ge \lfloor c_4 n \rfloor.$$

Then we have

$$\mathbf{P}\left(h(u) = 1 \text{ for all } u \in X \setminus \{v\}\right) \ge 1 - |X| e^{-c_2 n}. \tag{22}$$

Assume that the events in (21), (22) hold, in particular $h \in \mathcal{R}_v G$. By Lemma 6.1, we have

$$\pi h = h(v) = (\overline{w}_g(v))^{p^k} = g^{p^k} \in \sigma W_n$$

and the smallest s-invariant subspace M of σW_n containing πh has

$$\dim M \ge p^{n-1} - p^k + 1. \tag{23}$$

The element h is in $\mathcal{R}_v G$, and thus so are its $\overline{w}_{(s)}$-conjugates. Therefore the subspace M has an isomorphic π-pre-image $M_v \subseteq \mathcal{R}_v G$. The inequality (23) implies our claim (19).

By definition, the value of ε_* in (21) is bounded above by the chance q that $\theta + 1$ random elements do not generate W_n (or, equivalently, its Frattini quotient

$\mathbb{F}_p^n$). It is known that $q \leq e^{c_3(n-\theta)}$, where c_3 is an absolute constant. Since we use Lemma 6.2 twice, we can increase c_3 to include a factor of 2. The claimed probability bound follows from (21) and (22). Since G is transitive on v_X^G, (20) follows by conjugation. □

7. Dimension of random subgroups

The notion of Hausdorff dimension was introduced by Abercrombie [1994] to the setting of compact groups. Recall the norm $\|\cdot\|$ and distance dist defined in (12). This turns $W(H)$ into a compact metric space, and Hausdorff dimension is defined in the usual way. Translated to our setting the result of Barnea and Shalev [1997] says that for closed subgroups $G \subseteq W(H)$ Hausdorff dimension agrees with lower Minkowski (box) dimension. Since balls in $W(H)$ have simple structure, Minkowski dimension is easily checked to coincide with the lim inf of the **density sequence** γ_ℓ defined in the Introduction (1). For technical purposes, we will now introduce γ_ℓ in a slightly more general setting, for the wreath product of $W(p)$ and a finite p-group.

Let K_0 be a p-group acting on a set X_0. Let $W_{K_0} = W(p) \wr K_0$ acting on $|X_0|$ copies of the infinite p-ary tree T_p. Let $W_+^{[\ell]}$ denote the subgroup that stabilizes all vertices at level ℓ of all trees. Given a subgroup $G \subseteq W_+$, define the density sequence

$$\gamma_\ell(G) = \frac{\log |G/W_+^{[\ell]}|}{\log |W_+/W_+^{[\ell]}|}.$$

The density sequence γ_ℓ defined in (1) covers the case when K_0 is the trivial group acting on a singe-element set. In general, we have

$$\dim_{\mathrm{H}}(G) = \liminf_{\ell\to\infty} \gamma_\ell(G).$$

Proposition 7.1. *Let G be the subgroup generated by independent uniform random liftings of three fixed elements of K_0 to $W_+ = W(p) \wr K_0$. With probability 1, the density sequence $\gamma_\ell(G)$ satisfies*

$$\gamma_\ell(G) > 1 - e^{-c\ell}$$

for $c = c(p) > 0$ fixed and all sufficiently large ℓ. In particular, $\dim_{\mathrm{H}}(G) = 1$.

The most important case is when K_0 acts trivially on the single-element set X_0.

Theorem 7.2. *The subgroup of $W(p)$ generated by three random elements has* 1*-dimensional closure with probability* 1*.*

These random subgroups are the first known examples of finitely generated subgroups of $W(p)$ with 1-dimensional closure. We propose

Conjecture 7.3. *Theorem 7.2 holds even for two random elements.*

Another immediate corollary of Proposition 7.1 is a strengthening of Theorem 3:

Theorem 7.4. *Let $\varepsilon > 0$. Let G_n be the subgroup generated by three random elements of the symmetric p-group $W_n(p)$. Then $\mathbf{P}\left(|G_n| > |W_n(p)|^{1-\varepsilon}\right) \to 1$ as $n \to \infty$. More precisely, for some $q < p$ we have $\mathbf{P}(\log[W_n : G_n] < q^n) \to 1$.*

This theorem is the p-group analogy of the famous result of Dixon [1969] saying that two random elements of $\mathrm{Sym}(n)$ generate $\mathrm{Sym}(n)$ or the alternating group $\mathrm{Alt}(n)$ with probability tending to 1. In fact, for finite simple groups G, the probability that two elements generate G tends to 1 as $|G| \to \infty$; see Shalev [1999]. The methods used here are fundamentally different from the usual subgroup counting technique.

Another interesting consequence of Proposition 7.1 is the following strengthening of Theorem 2.

Theorem 7.5. *For each $d \in [0,1]$ there exists a topologically finitely generated free pro-p subgroup of $W(p)$ of Hausdorff dimension d.*

This result leads to the solution of a problem of Shalev [2000], who asked whether a topologically finitely generated pro-p group can contain topologically finitely generated subgroups of irrational Hausdorff dimension. Indeed, we may take three elements generating a full-dimensional subgroup $G_1 \subseteq W(p)$, and three more elements that generate a d-dimensional subgroup $G_d \subseteq W(p)$. It is straightforward to check that the dimension of G_d in $\langle G_1, G_d \rangle$ with respect to the filtration inherited from $W(p)$ is d.

Let T' be a subtree of T which has the same root as T and each vertex has 0 or p offspring. Let $W_{T'} \subseteq W(p)$ be the pointwise stabilizer of T'. The measure $\mu(T')$ of T' can be defined as $\lim |T'_n|/|T_n|$, where T'_n means the vertices at level n. The sequence is nonincreasing, so the limit always exists. It is easy to check that

$$\dim_{\mathrm{H}}(W_{T'}) = 1 - \mu(T').$$

Let V_0 denote the set of vertices of T' that have no children. Since $\mu(T')$ can take any value in $[0,1)$ (even if we assume that V_0 is infinite), $\dim_{\mathrm{H}} W_{T'}$ can take any value in $(0,1]$. In the following we show that random subgroups G of $W_{T'}$ are free and $\dim_{\mathrm{H}}(G) = \dim_{\mathrm{H}}(W_{T'})$ with probability 1. This proves Theorem 7.5 for $d > 0$; the easy case $d = 0$ is left to the reader.

Proposition 7.6. *The subgroup G topologically generated by $k \geq 3$ random elements of $W_{T'}$ satisfies*

$$\dim_{\mathrm{H}}(G) = \dim_{\mathrm{H}}(W_{T'}) = 1 - \mu(T')$$

with probability 1. *If V_0 is infinite, then G is a free pro-p group with probability* 1.

Proof of Proposition 7.6. Let V_{0+} denote the vertices V_0 and their descendants in T. Clearly, for $g \in W_{T'}$ and $v \in V \setminus V_{0+}$ we have $g(v) = 0$. Moreover, Haar measure on $W_{T'}$ agrees with picking $g(v)$ uniformly, independently at random for each $v \in V_{0+}$.

Let g_i denote the random elements, let $\varepsilon > 0$, and let ℓ be so that $|\partial_\ell T'|/|\partial_\ell T| < \mu(T') + \varepsilon$. Condition on the action K_0 of g_i on $X_0 = \partial T_\ell \setminus \partial T'_\ell$. By the first paragraph, given this information, the g_i are uniform random liftings. Proposition 7.1 applied to this action, with dimension appropriately rescaled, implies

$$\dim_{\mathrm{H}} G \geq |\partial_\ell T \setminus \partial_\ell T'|/|\partial_\ell T| > 1 - \mu(T') - \varepsilon.$$

With probability 1, this is true for all rational ε, so $\dim_{\mathrm{H}}(G) \geq \dim_{\mathrm{H}}(W_{T'})$, and the other inequality is trivial.

If P is a finite p-group of size p^d generated by $h_1, \ldots, h_k$, then P can be embedded into $W_d(p)$. This implies that if $g_1, \ldots, g_k$ are uniform independent random

elements in $W(p)$, then with positive probability the map $f : g_i \to h_i$ extends to a continuous homomorphism from $\overline{\langle g_1, \dots, g_k \rangle}$ onto P. For the vertices $v \in V_0$ the action of G on the descendant subtrees of v are independent, generated by k independent random elements. So if V_0 is infinite, then with probability 1 for any P and $h_1, \dots, h_k$ as above, there is a homomorphism $f : G \to P$ such that $f(g_i) = h_i$ $(1 \le i \le k)$. This implies that G is free pro-p with probability 1. □

Before we proceed to the proof of the main proposition we pose a conjecture and a related question.

Conjecture 7.7. *Let G be a closed subgroup of $W(p)$. Then G contains a topologically finitely generated subgroup of the same Hausdorff dimension.*

It may be that such subgroups are abundant. An affirmative answer to the following question would imply Conjectures 7.7 and 7.3.

Question 7.8. Let G be a closed subgroup of $W(p)$. Let g_1, g_2 be random elements chosen according to Haar measure on G. Is it always true that $\dim_{\mathrm{H}}(\langle g_1, g_2 \rangle) = \dim_{\mathrm{H}}(G)$ with probability 1?

Proof of Proposition 7.1. We now proceed to prove the key proposition. The proof uses almost all theorems proved so far, in particular, it depends crucially on the results of Sections 5 and 6. The proof is split into a deterministic part (the main proof) and a probabilistic part (Lemmas 7.9 and 7.10).

Let X_m denote the vertices at level m of the trees. For a G-orbit R of X_m, let $\kappa(R) = \kappa(v_R)$ as in Definition 5.2, with $K = G$ and $X = X_m$ there. Let $R_{m,1}, \dots, R_{m,r_m}$ denote the G-orbits of X_m, listed so that $\kappa(R_{m,i})$ is nondecreasing in i. Orbits with equal κ are listed according to their smallest vertex; recall that vertices are p-ary strings, which have a natural ordering as integers in p-ary notation.

Let $G_{m,i}$ denote the action of G on $X_{m,i} = R_{m,i} \cup \ldots \cup R_{m,r_m}$ and its descendants and ancestors. So $G_{m,i}$ is the quotient of G by the kernel of this action.

For a group G_* acting on disjoint union of rooted trees, let $\iota_\ell G_*$ denote the action on all levels $\ell' \le \ell$ of the trees; so $\iota_\ell G_*$ is a quotient of G_* by the kernel of this action.

Let $L_{m,i}$ denote the pointwise stabilizer in $G_{m,i}$ of all vertices in $X_{m,i+1}$ and their descendants. We claim that this decomposition freezes in the following sense.

Lemma 7.9. *There exits m_* so that with probability 1 for $m \ge m_*$ and any G-orbit $R \subseteq X_m$, the children R' of vertices in R form a G-orbit as well, and we have $\kappa(R') = \kappa(R)$.*

As a consequence, for all $m \ge m_*$ we have $r_m = r_{m_*}$, for all i we have

$$R_{m,i} = (\textit{descendants of } R_{m_*,i}) \cap X_m$$

as well as $G_{m,i} = G_{m_*,i}$ and $L_{m,i} = L_{m_*,i}$.

It is easy to see that for all $\ell \ge m \ge 1$ we have

$$|\iota_\ell G| = |\iota_\ell L_{m,1}| \cdots |\iota_\ell L_{m,r_m}|. \tag{24}$$

We now decompose the $\iota_\ell L_{m,i}$ further. Let $\sigma_\ell L_i \subseteq \iota_\ell L_i$ denote the pointwise stabilizer of level $\ell - 1$ of the trees. Then $\sigma_\ell L_i$ is naturally an $\mathbb{F}_p$-vector space. For $\ell > \ell_0$ we easily get

$$|\iota_\ell L_{m,i}| = |\iota_{\ell_0} L_{m,i}|\, |\sigma_{\ell_0+1} L_{m,i}| \cdots |\sigma_\ell L_{m,i}|. \tag{25}$$

Lemma 7.10. *There exists $c = c(p) > 0$ and a random variable ℓ_0 so that for all $\ell \geq \ell_0$ and $1 \leq i \leq r_{m_*}$ we have $m(\ell) := \lfloor (\log \ell)^2 \rfloor \geq m_*$ and*

$$\dim \sigma_\ell L_{m(\ell),i} \geq (1 - p^{-c\ell})|R_{m(\ell),i}| \dim \sigma W_{\ell - m(\ell)}.$$

We first prove the proposition assuming the lemmas, then proceed to prove the lemmas.

Proof of Proposition 7.1. The decompositions (24) and (25) give

$$|\iota_\ell G| \geq \prod_{i=1}^{r_{m_*}} \prod_{k=\ell_0}^{\ell} |\sigma_k L_{m_*,i}|.$$

By Lemma 7.9 we have $L_{m_*,i} = L_{m(k),i}$. By Lemma 7.10 we get

$$\begin{aligned} \log_p |\iota_\ell G| &\geq \sum_{i=1}^{r_{m_*}} \sum_{k=\ell_0}^{\ell} (1 - p^{-ck})|R_{m(k),i}| \dim \sigma W_{k-m(k)} \\ &= \sum_{k=\ell_0}^{\ell} (1 - p^{-ck})|X| p^{m(k)} p^{k-m(k)-1} \\ &= \log_p |\iota_\ell W_+| - \log_p |\iota_{\ell_0} W_+| - |X| \sum_{k=\ell_0}^{\ell} p^{(1-c)k-1}. \end{aligned}$$

Note that the first error term is a constant depending on ℓ_0, and the second one is bounded by $c_2 p^{-c\ell} \log_p |\iota_\ell W_+|$. For large enough ℓ this gives the bound

$$\log_p |\iota_\ell G| \geq (1 - e^{-c'\ell}) \log_p |\iota_\ell W_+|,$$

as required. □

Proof of Lemmas 7.9 and 7.10. For $b \geq 1$, let $A_{b,m}$ denote the following event:

(i) $r_m \leq b$,
(ii) all orbits of the action of $\langle g_1, g_2 \rangle$ on X_m have size at least p^m/b, and
(iii) the action of $g_3^{\kappa(X_m)}$ on X_m has no fixed point.

By Proposition 3.10, there exists a random variable m_0 so that for $m \geq m_0$ the children of any G-orbit of X_m form a G-orbit of X_{m+1} with probability 1. Similarly, there exists $m_* \geq m_0$ so that for $m \geq m_*$ the children of any $\langle g_1, g_2 \rangle$-orbit of X_m form a $\langle g_1, g_2 \rangle$-orbit of X_{m+1}. Therefore

(a) There exists a random b^* so that for $b \geq b_*$ and all m (i)-(ii) hold.
(b) If w is a word in the generators fixing $v \in X_m$, $m > m_*$, then for some $\{g_1, g_2\}$-word w' all children of v are fixed by ww'. This implies that $\kappa(X_{m+1}) = \kappa(X_m)$ and that the claim of Lemma 7.9 holds.

Part (b) implies that there exists a random b' so that $\kappa(X_m) < b'$ for all m with probability 1. Every random word in $W(p)$ fixes only finitely many vertices of T (Proposition 4.1). Therefore there exists a random variable $m_+ \geq m_*$ so that (iii) holds for all $m \geq m_+$ with probability 1. Thus with probability 1

$$\text{for all } b \geq b_*, m \geq m_+ \text{ the event } A_{b,m} \text{ holds.} \tag{26}$$

For $\ell \geq 1$ let $m = m(\ell) := \lfloor (\log \ell)^2 \rfloor$, and let $n = n(\ell) := \ell - m$. The event $A_{b,m}$ depends only on the information contained in the $g_i' := g_i / W_+^{[m]} \in W_m \wr K_0$. The

conditional distribution of the $g_i/W^{[\ell]}$ given this information is just independent uniform random liftings of the g_i' to $W_n \wr (W_m \wr K_0) = W_\ell \wr K_0$.

Let B_ℓ denote the event that for all $1 \le i \le r_m$ we have

$$\dim \sigma_\ell L_{m,i} \ge (1 - p^{-c_1 n(\ell)+1})|R_{m,i}| \dim \sigma W_n.$$

We apply Lemma 6.4 (20) for $i = 1, \dots, r_m$, with $K = G_{m,i}$ and $X = X_{m,i}$. For $b \ge 1$ we get

$$\begin{aligned} \mathbf{P}(B_\ell^c | A_{b,m}) &\le r_m(e^{-c_2 n}|X_m| - e^{c_3(n-p^m/b)}) \\ &\le b(e^{-c_2 n}|K_0|p^m - e^{c_3(n-p^m/b)}) \\ &\le e^{-c_4 \ell + c_5} \end{aligned}$$

for some deterministic constants c_4, $c_5 > 0$ depending only on b, p and $|K_0|$. In particular,

$$\mathbf{P}(B_\ell^c \cap A_{b,m(\ell)}) \le \mathbf{P}(B_\ell^c | A_{b,m(\ell)}) \le e^{-c_4 \ell + c_5}.$$

Since the right-hand side is summable, the Borel-Cantelli lemma implies that there exists a random variable $\ell_*(b) < \infty$ so that with probability one

$$\text{for all } \ell \ge \ell_*(b) \text{ the event } B_\ell \text{ or } A_{b,m(\ell)}^c \text{ holds.} \tag{27}$$

By (26) and (27), the event B_ℓ holds whenever $b \ge b_*$, $\ell \ge \ell_*(b)$ and $m(\ell) \ge m_+$. The claim of Lemma 7.10 follows by comparing ℓ and $n(\ell)$. □

8. Small subgroups

In this section we state general results on small subgroups of $W(p)$.

First we give a description on the Abelian subgroups of $W(p)$ using orbit trees. A rooted tree is a 1-p tree if all the vertices have 1 or p children. We call a vertex **solo**, if it has 1 child. For an arbitrary rooted tree R let $S(R)$ denote the number of solo vertices of R. We say that a group $A \subseteq W_n(p)$ **belongs to** R if $R = T_n/A$.

Lemma 8.1. *Let R be a 1-p tree of depth n and let $A \subseteq W_n(p)$ be an Abelian subgroup which belongs to R. Then $\log_p |A| \le S(R)$. Equality holds if and only if A is a maximal Abelian subgroup which belongs to R.*

Proof. We use induction on n. For $n = 1$ the proposition is trivial. Suppose it holds for $n - 1$.

Let $B \subseteq A$ be the stabilizer of level 1 of T_n. Then B has index 1 or p in A. Let $\Lambda_1, \dots, \Lambda_p$ denote the subtrees of T_n starting from a vertex of level 1. Let B_i denote the restriction of the action of B to Λ_i and let $|B_i|$.

If $A = B$ (i.e., if the root of R is not solo), then A maps each Λ_i onto itself so we have

$$|A| \le \prod_{i=1}^{p} |B_i| = p^{\sum_{i=1}^p k_i}.$$

Using induction we get

$$\sum_{i=1}^{p} k_i \le \sum_{i=1}^{p} S(\Lambda_i/B_i) = S(R).$$

If $|A : B| = p$ (i.e., the root is solo), let $g \in A \setminus B$ be an arbitrary element. Then $g^p \in B$, so the action of g on T_n/B has order p. This means that g induces graph isomorphisms between the trees Λ_i/B_i, so $S(T_n/B) = pS(\Lambda_1/B_1)$ and also $S(R) = S(\Lambda_1/B_1) + 1$. Now let $b \in B$ be any element which lies in the kernel of the

restriction of B to Λ_1, i.e., fixes Λ_1 pointwise. Since g permutes the Λ_i transitively and commutes with b, it means that b fixes all the Λ_i pointwise, that is, b is the identity. Thus B is isomorphic to B_1. So using induction

$$|A| = p\,|B_1| \le p^{S(\Lambda_1/B_1)+1} = p^{S(R)}.$$

If the equality $|A| = p^{S(R)}$ holds, then A is trivially maximal with the property $R = T_n/A$. The other direction follows by induction in the same way as the inequality above. □

As a corollary, it is possible to show that Abelian subgroups of $W(p)$ are zero dimensional. Theorem 6 gives a more general result in this direction. Moreover, Theorem 5 shows that Abelian groups are also small dimensional as sections: the derived subgroup of a closed subgroup $G \subseteq W(p)$ has the same dimension as G.

For Theorem 5 we need a general asymptotic result on rooted trees. Let $d \ge 2$ be an integer. A rooted tree T is defined to be a d-bounded tree if every vertex has at most d children. It is a 1-d tree if every vertex has either 1 or d children. Let U be a subset of a d-bounded tree T. We define the *density sequence* of U by

$$\delta_n(U) = \frac{r_n(U)}{d^n},$$

where $r_n(U)$ is the number of vertices in U of level n. We define the **density**

$$\delta(U) = \lim_{n\to\infty} \delta_n(U)$$

if this limit exists. In particular, $\delta(T)$ always exists, since $\delta_n(T)$ is monotone decreasing.

The following straightforward lemma will be useful.

Lemma 8.2. *Let T be a d-bounded tree and let U be the set of vertices of degree less than d. Then $\sum_{i=0}^{\infty} \delta_i(U) \le d(1-\delta(T))$. In particular, $\delta(U) = 0$. If in addition T is a 1-d tree, then*

$$\sum_{i=0}^{\infty} \delta_i(U) = \frac{d}{d-1}(1-\delta(T)).$$

Proof. For the first claim, we may assume that all the vertices have degree d or $d-1$. Then $r_i(U) = dr_i(T) - r_{i+1}(T)$. From this, using $r_0(T) = 1$ we have

$$\sum_{i=0}^{n} \delta_i(U) = \sum_{i=0}^{n} d\left(\frac{r_i(T)}{d^i} - \frac{r_{i+1}(T)}{d^{i+1}}\right) = d\left(1 - \frac{r_{n+1}(T)}{d^{n+1}}\right)$$

which gives $\sum_{i=1}^{\infty} \delta_i(U) = d(1-\delta(T))$. If every vertex has degree 1 or d, then we have $r_i(U) = (dr_i(T) - r_{i+1}(T))/(d-1)$ and the last statement of the lemma follows. □

Now we are ready to prove the "perfectness" theorem, restated here.

Theorem 5. *Let $G \subseteq W(p)$ be a closed subgroup. Then $\lim(\gamma_n(G) - \gamma_n(G')) = 0$. In particular, $\dim_{\mathrm{H}} G = \dim_{\mathrm{H}} G'$.*

Proof. Let $F = T(p)/G$ be the orbit tree of the action of G on $T(p)$. We treat F and $T(p)$ as 1-p trees. Fix positive integers d and k.

We partition the vertices of F into three classes as follows. Let A be the set of orbits which have exactly p^k k-th cousins in F including themselves. Let B be the set of orbits of size at least p^d and let C be the set of remaining vertices.

We claim that $\delta(B \cup C) = 0$. To see this, let us define a new graph F' on the vertices on F by putting an edge between u and v if u is the k-th grandson of v. Then F' is the union of k disjoint p^k-bounded trees. Now let D be the set of F'-parents of elements in $B \cup C$. Then D is the set of vertices in F' which has less than p^k children, so by Lemma 8.2 D has zero density in each of the components of F'. But then $B \cup C$ has zero density in F.

For an arbitrary p-group P acting on a finite set X of size m let us define the **commutator density** by $c(P) = \log_p |P : P'|$. We use the following facts on $c(P)$:

(a) Comparing the p-group P to the Sylow p-subgroup we have
$$c(P) \leq \log_p \left|\mathrm{Syl}_p(\mathrm{Sym}(m))\right| \leq m.$$
(b) If P is transitive on X, then $c(P) \leq Km/\sqrt{\log m}$, where K is an absolute constant. This is a result of Kovács and Newman [1988].
(c) If P is intransitive and $X = \bigcup_{i=1}^{j} X_i$ such that each X_i is P-invariant, then we have $c(P) \leq \sum_{i=1}^{j} c(P_i)$, where P_i is the action of P on X_i.
(d) If P acts diagonally on the X_i, i.e., when elements of P fixing any of the X_i pointwise must fix the whole X, then $c(P) = c(P_1)$.

Now let n be an arbitrary level and let A_n, B_n and C_n be the set of vertices of T which belong to A, B and C, respectively. Let G_n be the action of G on the n-th level of T and let G_a, G_b and G_c be the action of G_n on A_n, B_n and C_n.

The orbits lying in A can be partitioned into sets of size p^k such that the action of G_a on them is diagonal. Then (d), (c) and (a) imply
$$c(G_a) \leq |A_n| / p^k \leq p^{n-k}.$$
Using (b) and (c) we have
$$c(G_b) \leq K |B_n| / \sqrt{d} \leq K p^n / \sqrt{d},$$
since all the orbits lying in B have size at least p^d. Lastly (a) implies
$$c(G_c) \leq |C_n| \leq \delta_n(C) p^{n+d-1},$$
since all the orbits in C_n have size at most p^{d-1}. Combining the above, (c) on A_n, B_n and C_n implies
$$c(G_n)p^{-n} \leq p^{-k} + K/\sqrt{d} + \delta_n(C)p^{d-1}$$
and letting $n \to \infty$ we get
$$\limsup_{n\to\infty} c(G_n)p^{-n} \leq p^{-k} + K/\sqrt{d} + \delta(C)p^{d-1}, \tag{28}$$
where $\delta(C) \leq \delta(B \cup C) = 0$. The inequality (28) holds for every positive k and d, so $c(G_n)p^{-n} \to 0$, and the proof is complete. □

As an immediate corollary, solvable subgroups are zero dimensional. Using a Baire category argument we obtain the following.

Corollary 8.3. *Let $G \subseteq W(p)$ be a positive-dimensional subgroup. Then G cannot be abstractly generated by countably many solvable subgroups.*

Proof. Suppose that $H_1, H_2, \ldots \subseteq G$ are solvable subgroups generating G as an abstract group. Since closure preserves solvability, we can assume that the H_i are closed. Then
$$G = \bigcup_{k, n_i \in \mathbb{N}} H_{n_1} H_{n_2} \cdots H_{n_k}$$

is the countable union of the finite products formed from the subgroups H_i. The sets $H_{n_1}H_{n_2}\cdots H_{n_k}$ are closed, so by the Baire category theorem one of them has to contain an open set U. Then U contains a coset of an open subgroup in G, so

$$\dim_{\mathrm{H}} H_{n_1}H_{n_2}\cdots H_{n_k} \geq \dim_{\mathrm{H}} U = \dim_{\mathrm{H}} G > 0.$$

By Theorem 5 H_{n_i} are zero dimensional, moreover, $\lim \gamma_n(H_{n_i}) = 0$. It is easy to see that

$$\gamma_n(H_{n_1}H_{n_2}\cdots H_{n_k}) \leq \sum_{i=1}^{k} \gamma_n(H_{n_i})$$

for all n. This implies $\dim_{\mathrm{H}} H_{n_1}H_{n_2}\cdots H_{n_k} = 0$, a contradiction. □

Note that the full $W(p)$ can be easily generated *topologically* by an element and an Abelian subgroup.

Since solvable subgroups have dimension zero, we need to find a more refined way to measure how large they are. This can be done by *summing* the density sequence rather than taking its limit, as Theorem 6 shows.

Theorem 6. *Let $G \subseteq W(p)$ be a closed subgroup. Then* $\lim(\gamma_n(G) - \gamma_n(G')) = 0$. *In particular,* $\dim_{\mathrm{H}} G = \dim_{\mathrm{H}} G'$.

The example $W_n \subset W(p)$ shows that this is the best possible bound up to the constant factor.

Proof. We can directly deduce this theorem from Lemma 8.1 and Lemma 8.2 in the case when G is Abelian.

Instead of $\gamma_n(G)$, it is more convenient to first estimate a slightly different density. Let

$$\overline{\gamma}_n(G) = (p-1)\frac{\log|G_n|}{p^n}, \qquad s_n(G) = \sum_{i=0}^{n} \overline{\gamma}_i(G).$$

Let $m_{n,d} = \max\{s_n(H) \mid H \subseteq W_n(p) \text{ with solvable length } d\}$.

Let $\Lambda_1, \ldots, \Lambda_p$ denote the subtrees of W_n starting from level 1. Let $K \subseteq H$ denote the stabilizer of level 1. Either $K = H$ or $|H : K| = p$. Let K_j denote the action of K on Λ_j. The groups K_j are d-solvable so $s_{n-1}(K_j) \leq m_{n-1,d}$ for all j.

If $H = K$, then we have

$$s_n(H) \leq \sum_{j=1}^{p} s_{n-1}(K_j)/p \leq m_{n-1,d}. \tag{29}$$

If $|H : K| = p$, then let A denote the restriction of K to the set $\Lambda_2 \cup \ldots \cup \Lambda_n$ and let $N \subseteq K_1$ be the kernel of this restriction, i.e., the set of elements which act trivially outside Λ_1. Let $h \in H \backslash K$ be an element moving the first level and let $D = [N, h]$. Then h moves all the vertices of the first level so the restriction of D to Λ_1 is again N. Hence from $D \subseteq G'$ we see that N is $d-1$-solvable. Now $|H| = p\,|N|\,|A|$ and $|A| \leq |K_2| \cdots |K_p|$ so we have

$$\overline{\gamma}_i(H) \leq \frac{1}{p^i} + \frac{1}{p}\left(\overline{\gamma}_{i-1}(N) + \sum_{j=2}^{p} \overline{\gamma}_{i-1}(K_j)\right),$$

from which we get

$$s_n(H) \leq \frac{p}{p-1} + \frac{1}{p}\left(m_{n-1,d-1} + (p-1)m_{n-1,d}\right). \tag{30}$$

Summarizing, we have $m_{n,0} = 0$ and a recursive upper bound on $m_{n,d}$ given by the maximum of the right-hand side of (29) and (30). From this the bound

$$m_{n,d} \leq dp^2/(p-1)$$

easily follows. On the other hand, trivially $|\gamma_n(G) - \overline{\gamma}_n(G)| \leq 1/p^n$ which yields

$$\sum_{n=0}^{\infty} \gamma_n(G) \leq Cd$$

with $C = (p^2+p)/(p-1)$. □

A possible generalization of the result on the zero dimensionality of solvable groups would be that every subgroup of $W(p)$ which satisfies a nontrivial identity is zero dimensional. Note that Conjecture 8 would imply this even for pro-p identities. Another possible direction arises from the observation that linear groups tend to be zero dimensional in $W(p)$.

Conjecture 8.4. *Let R be a commutative pro-p ring and let $G \subseteq GL_n(R)$ be a linear group over R. Then for every embedding of G into $W(p)$ the image of G has zero dimension.*

It is easy to see that the conjecture holds for $R = \mathbb{Z}_p$; in this case G is p-adic analytic and so is the product of finitely many procyclic subgroups (Dixon et al. [1991]). This question is also related to Conjecture 8 in the sense that by Barnea and Larsen [1999] nonabelian free pro-p groups are not linear over local fields, so Conjecture 8 implies Conjecture 8.4 in the case when R is a local field.

9. Large subgroups

In this section we analyze 1-dimensional subgroups of $W(p)$. The main results are that spherically transitive 1-dimensional groups have normal spectra $\{0, 1\}$ and that they contain nonabelian free pro-p subgroups as well as dense free subgroups. Recall that the normal spectra of a group G is the set of possible Hausdorff dimensions of normal subgroups of G.

One of the natural sources of interesting subgroups in $W(p)$ are just infinite pro-p groups. A group G is just infinite if every proper quotient of G is finite; G is hereditarily just infinite if all subgroups of G of finite index are just infinite. Just infinite pro-p groups are regarded as the simple groups in the pro-p category. Many questions on pro-p groups can be reduced to the just infinite case, since (unlike in the profinite setting Zalesskii [2002]) every finitely generated pro-p group possesses just infinite pro-p quotients.

Building on the work of Wilson [1971], Grigorchuk [2000] showed that just infinite pro-p groups fall into the following two classes:

(i) groups containing an open normal subgroup which is the direct power of a hereditarily just infinite pro-p group;
(ii) pro-p branch groups.

Groups from class (ii) have a natural action on $T(p)$. In fact, one can define them as spherically transitive subgroups $G \subseteq W(p)$ such that the rigid level stabilizers of G have finite index in G. The first example for such a group was the closure of the first Grigorchuk group, a subgroup of $W(2)$ with many remarkable properties. Bartholdi and Grigorchuk [2000] calculated that it has Hausdorff dimension 5/8.

In du Sautoy et al. [2000], Boston suggested a direct connection between Grigorchuk's classes and Hausdorff dimension: he conjectured that a just infinite pro-p group is branch if and only if it can be obtained as a positive-dimensional subgroup of $W(p)$. In this section we contrast known properties of branch groups with new results on 1-dimensional groups. The first result concerns the normal subgroup structure of such groups.

In general, the normal spectra of 1-dimensional groups can easily be the full interval $[0,1]$. The simplest example for this is the stabilizer of an infinite ray in $W(p)$; it is the full countable direct power of $W(p)$ with weighted dimensions $1/p^n$. However, if we assume that the group is spherically transitive, i.e., it acts transitively on every level, we get a completely different picture. It turns out that these groups, just as the free pro-p group (see Newman et al. [2000]), are "simple" in terms of Hausdorff dimension.

Theorem 9.1. *Let $G \subseteq W(p)$ be a spherically transitive 1-dimensional closed subgroup. Then for every $1 \neq N \lhd G$ we have $\dim_{\mathrm{H}} N = 1$.*

We need some technical lemmas. The first one is an asymptotic version of Theorem 7. As the existence of large Abelian subgroups in $W_n(p)$ shows, we have to assume that the density is close to 1.

Lemma 9.2. *There exist a constant C depending only on p such that for every subgroup $G \subseteq W_n(p)$ with $\gamma_n(G) \geq 1-\epsilon$ we have $\gamma_n(G') \geq 1-\epsilon-\delta$, where $\delta = C \min\{n, -\log_p \epsilon\}^{-1/2}$.*

Proof. We consider the action of G on level n of T_n. Let r_i be the number of orbits of size p^i $(0 \leq i \leq n)$, and let r denote the total number of orbits. Then

$$\sum_{i=0}^{n} r_i p^i = p^n \tag{31}$$

and

$$\log_p |G| \leq \sum_{i=0}^{n} r_i \log_p |W_i(p)| = \sum_{i=0}^{n} r_i \frac{p^i-1}{p-1} = \frac{p^n - r}{p-1}.$$

From this and the density assumption

$$\log_p |G| \geq (1-\epsilon)\frac{p^n-1}{p-1}$$

we get the inequality $r \leq \epsilon(p^n-1)+1$. Using the theorem of Kovács and Newman [1988] with (b) and (c) from the proof of Theorem 5 we see that

$$\log_p |G:G'| \leq K \sum_{i=0}^{n} r_i \frac{p^i}{\sqrt{i}},$$

where K is an absolute constant. Separating the sum at an arbitrary $m \leq n$ and using (31) we get

$$\sum_{i=0}^{n} \frac{r_i p^i}{\sqrt{i}} \leq \frac{r p^m}{\sqrt{m}} + \frac{p^n}{\sqrt{m+1}}.$$

Therefore with $m_0 = \min\left\{n, -\log_p \epsilon\right\}$ and $C = 4K(p-1)$ we get

$$\log_p |G:G'| \leq K(p^n-1)\frac{3+\epsilon p^m}{\sqrt{m}} \leq C \log_p |W_n(p)| / \sqrt{m_0},$$

and this implies the stated bound on the density $\gamma_n(G')$. □

Consider a subgroup $G \subseteq W_n(p)$, where n may be infinite. We define the **rigid stabilizer** $\mathcal{R}_k G$ of level k as follows. Recall that the rigid vertex stabilizer $\mathcal{R}_v G$ of a vertex v is the pointwise stabilizer of the vertices not descendant to v. Then $\mathcal{R}_k G$ is defined as the group generated by the $\mathcal{R}_v G$ of vertices v at level k.

We need a slight reformulation of Theorem 4 of Grigorchuk [2000]. This is the key theorem to prove that the first Grigorchuk group, or more generally, torsion branch groups, are just infinite.

Lemma 9.3. *Let $G \subseteq W_n(p)$ be transitive and let $g \in G$ be an element which moves a vertex at level k. Then the normal subgroup generated by g in G contains $(\mathcal{R}_{k+1}G)'$, the derived subgroup of the rigid stabilizer of level $k+1$.*

The proof is exactly the same as in Grigorchuk [2000].

We also need a lemma which measures how close subdirect products of high density are to direct products in size.

Lemma 9.4. *Let $G \subseteq W_n(p)^d$ satisfying $|G| \geq |W_n|^{d-\epsilon}$. Then G contains a direct product $H = H_1 \times \cdots \times H_d$ with $|H| \geq |W_n|^{d(1-\epsilon)}$.*

Proof. Let $i \leq d$ be a coordinate. Project G to the coordinates not equal to i and let H_i be the kernel of this projection. Then the image has size at most $|W_n|^{d-1}$ so $|H_i| \geq |W_n|^{1-\epsilon}$. Since the full direct product $H = H_1 \times \cdots \times H_d$ lies in G, the statement of the lemma follows. □

Now we are ready to prove Theorem 7, restated here.

Theorem 7. *Let $G \subseteq W(p)$ be a closed subgroup. Then* $\lim(\gamma_n(G) - \gamma_n(G')) = 0$. *In particular,* $\dim_{\mathrm{H}} G = \dim_{\mathrm{H}} G'$.

Proof. Let $N \lhd G$ be a nontrivial normal subgroup, and fix k so that N moves a vertex at level $k-1$. For $n \geq k$ let $G_n = GW^{[n]}/W^{[n]}$, and let $N_n = NW^{[n]}/W^{[n]}$. Then N_n is a nontrivial normal subgroup of $G_n \subseteq W_n$. Let $K \subseteq G_n$ denote the stabilizer of level k. Then $|G_n : K| \leq |W_k|$ so by the 1-dimensionality of G

$$\log|K| = \log|W_n|\,(1-\epsilon_n),$$

where $\epsilon_n \to 0$ as $n \to \infty$. Now K acts on the subtrees starting at level k so it is naturally a subgroup of $W_{n-k}(p)^{p^k}$. Lemma 9.4 implies that K contains a large direct product; in particular, the rigid stabilizer $\mathcal{R}_k = \mathcal{R}_k G$ satisfies

$$\log|\mathcal{R}_k| \geq p^k \log|W_{n-k}|\,(1-p^k\epsilon_n) \geq \log|W_n|\,(1-\epsilon_n')$$

with $\epsilon_n' \to 0$. By Lemma 9.2 the commutator R_k' satisfies

$$\log|N_n| \geq \log|\mathcal{R}_k'| \geq \log|W_n|\,(1-\epsilon_n'-\delta_n),$$

where $\delta_n = C\min\{n, -\log_p \epsilon'\}^{-1/2}$. Letting $n \to \infty$ we get $\gamma_n(N) \to 1$, as required. □

We are ready to prove Conjecture 8 in the 1-dimensional spherically transitive case.

Theorem 9. *Let $G \subseteq W(p)$ be a spherically transitive 1-dimensional closed subgroup. Then G contains a nonabelian free pro-p subgroup.*

In particular, G involves every finite p-group. Recall that a group G **involves** a group H if it can be obtained as a quotient of a finite index subgroup of G. Our first lemma confirms this for every 1-dimensional subgroup.

Lemma 9.5. *Let $G \subseteq W(p)$ be a closed subgroup of dimension 1. Then G involves every finite p-group.*

Proof. It suffices to prove that G involves W_n for every n. We define the **n-sample** S_v of G at the vertex v of depth k as follows. Take the stabilizer of level k. Project this group onto the descendant subtree of v and cut it at level $k+n$. Let $S_v \subseteq W_n$ denote the action on this subtree.

Now for any ℓ it is easy to see that $G_{n\ell} = GW^{[n\ell]}/W^{[n\ell]}$ satisfies

$$|G_{n\ell}| \leq \prod_v |S_v|\,,$$

where the product is taken over all vertices in $T_{n\ell}$ at levels divisible by n. Then

$$\gamma_{n\ell}(G) \leq \frac{\sum_v \log|S_v|}{\sum_v \log|W_n|} \leq \max_v \frac{\log|S_v|}{\log|W_n|}.$$

As $\ell \to \infty$, the left-hand side converges to 1, so there exists a v such that $S_v = W_n$. It is easy to see that G involves all its samples. □

As a corollary, 1-dimensional groups cannot satisfy any pro-p identity. It seems plausible that the lemma holds for arbitrary positive-dimensional closed subgroups.

Let $G \subseteq W(p)$ be a closed subgroup and let r be an infinite ray with vertex r_n at level n. There are two possible notions of the n-th stabilizer of r in G: the stabilizer subgroup $F_n \subseteq G_n = GW^{[n]}/W^{[n]}$ of r_n, and the congruence quotient $K_n = KW^{[n]}/W^{[n]}$ of the stabilizer $K \subseteq G$ of r. In general, $K_n \subseteq F_n$, but they need not coincide. However, equality holds for spherically transitive groups of high enough dimension.

Lemma 9.6. *Let $G \subseteq W(p)$ be a spherically transitive closed subgroup satisfying $\dim_{\mathrm{H}} G > 1/p$ and let r be an infinite ray. Then there exists n_0, such that for all $n > n_0$ we have $K_n = F_n$. In particular, $\dim_{\mathrm{H}} K = \dim_{\mathrm{H}} G$.*

Proof. Let $V_n = (G \cap W_n)W_{n+1}/W_{n+1}$, the level n subspace of G. Then we have $\dim V_n \leq p^n$ and $\log_p |G_n| = \dim V_0 + \ldots + \dim V_{n-1}$. Our assumption and a straightforward computation implies that there exists n_0 such that for all $n > n_0$ the subspace V_n is nontrivial.

Now let $n > n_0$ and let $g_n \in F_n$. We have to show that g_n can be extended to K, that is, there exists $g \in K$ so that $g/W^{[n]} = g_n$. In fact, using induction, it suffices to show that there exists an extension g_{n+1} to F_{n+1}, the stabilizer of the vertex v of r at level $n+1$.

Let S be the set of possible extensions of g to G_{n+1} Then S forms a coset of V_n in G_{n+1}. Now G is spherically transitive and V_n is a nontrivial G_n-invariant subspace, thus the projection of V_n to the coordinate $v \in r$ is the entire $\mathbb{F}_p$. Since S stabilizes v, the projection of S to v also equals $\mathbb{F}_p$. So there exist $g_{n+1} \in S$ which stabilizes the children of v, as required.

For $n > n_0$ we have $|G_n : K_n| = |G_n : F_n| = p^n$, so $\dim_{\mathrm{H}} K = \dim_{\mathrm{H}} G$. □

Proof of Theorem 9. Let r be the leftmost infinite ray, that is, the set of vertices indexed by sequences of zeros. Let F_k denote the stabilizer of r in the finite group

$W_k(p)$. Let K be the stabilizer of r in G. We claim that K has a closed normal subgroup such that the quotient group is isomorphic to the full direct product $D = \prod_k F_k$.

Choose n_0 according to Lemma 9.6. From Lemma 9.5 it follows that for every k there exists infinitely many vertex v such that the k-sample $S_v = W_k$. Since G is spherically transitive, samples at the same level are isomorphic, so we can choose a strictly increasing sequence n_k such that $S_{v_k} = W_k$, where the vertex v_k is at level n_k on the ray r. We can also assume that $n_{k+1} - n_k > k$ for all k. Now K naturally projects into $\prod_k F_k$ by the following function π. For $g \in K$ let the k-th coordinate $\pi_k(g)$ be the action of g on the subtree of length k starting at the vertex $v_k \in r$. As g stabilizes r, the image $\pi_k(g)$ will lie in F_k.

We show that the projection π is surjective, that is, for every $h \in \prod_k F_k$ we can find $g \in K$ such that $\pi(g) = h$. We construct g by a series of level extensions. From level n_k to level $n_k + k$ we use the following argument. Since g stabilizes v_k, the set of possible extensions of g under v_k is the union of cosets of the sample $S_{v_k}(G)$. But n_k was chosen in a way such that the k-sample $S_{v_k}(G)$ is the full $W_k(p)$. This implies that up to level $n_k + k$ we can prescribe g in an arbitrary way, so we can assume $\pi_k(g) = h_k$. For levels between $n_k + k$ and n_{k+1} we use Lemma 9.6 to obtain an extension which stabilizes v_{k+1}. Since G is closed, our series of extensions produces an element of G, so the claim holds and K projects onto D.

Since $W_{k-1}(p)$ is a subgroup of F_k for all k, every finite p-group can be embedded into F_k for large enough k. This implies that the full product D has a nonabelian free pro-p subgroup topologically generated, say, by the elements x, y. Then any two preimages $\tilde{x}$, $\tilde{y}$ in G generate a nonabelian free pro-p subgroup, since pro-p words satisfied by $\tilde{x}$ and $\tilde{y}$ are also satisfied by x and y. The proof is complete. □

Wilson [2000] showed that just infinite pro-p branch groups contain dense free subgroups (on the existence of dense free subgroups in profinite group; see Pyber and Shalev [2001] and Soifer and Venkataramana [2000]). The same holds for 1-dimensional subgroups. Let $d(G)$ denote the minimal number of topological generators for G (this might be infinite). An abstract subgroup of a topological group G is **dense** if its closure equals G.

Theorem 10. *Let $G \subseteq W(p)$ be a closed subgroup of dimension 1 and let $k \geq d(G)$. Then G contains a dense free subgroup of rank k.*

In other words, the free group F_k is residually S, where S denotes the set of congruence quotients of G.

Proof. By Corollary 4.6 the subgroup generated by k random elements is free of rank k with probability 1.

If $k = \infty$, then even the set of k random elements is dense in G with probability 1, since $G \subseteq W(p)$ has a countable base. If $d(G) \leq k < \infty$, then topological generation depends only on the Frattini quotient $C_p^{d(G)}$ of G. This is generated by $k \geq d(G)$ random elements with positive probability. □

Acknowledgments

We thank Yuval Peres for recommending the relevant branching random walk references, Péter Pál Pálfy for comments and bringing Turán's problem to our attention, and László Pyber for asking a question leading to Corollary 8.3, which

was the starting point of this project. We thank Russell Lyons and an anonymous referee for their useful remarks about previous versions of this paper.

References

[1994] A. G. Abercrombie. Subgroups and subrings of profinite rings. *Math. Proc. Cambridge Philos. Soc.*, 116(2):209–222. MR1281541 (95h:11078)

[In prep] M. Abért and B. Virág. Polynomials of p-trees. In preparation.

[1972] K. B. Athreya and P. E. Ney. *Branching processes*. Springer-Verlag, New York. Die Grundlehren der mathematischen Wissenschaften, Band 196. MR0373040 (51:9242)

[1999] Y. Barnea and M. Larsen. A non-abelian free pro-p group is not linear over a local field. *J. Algebra*, 214(1):338–341. MR1684856 (2000d:20031)

[1997] Y. Barnea and A. Shalev. Hausdorff dimension, pro-p groups, and Kac-Moody algebras. *Trans. Amer. Math. Soc.*, 349(12):5073–5091. MR1422889 (98b:20041)

[1998] Y. Barnea, A. Shalev, and E. I. Zelmanov. Graded subalgebras of affine Kac-Moody algebras. *Israel J. Math.*, 104:321–334. MR1622319 (99d:17025)

[2000] L. Bartholdi and R. I. Grigorchuk. Lie methods in growth of groups and groups of finite width. In *Computational and geometric aspects of modern algebra (Edinburgh, 1998)*, volume 275 of *London Math. Soc. Lecture Note Ser.*, pages 1–27. Cambridge Univ. Press, Cambridge. MR1776763 (2001h:20046)

[1995] M. Bhattacharjee. The ubiquity of free subgroups in certain inverse limits of groups. *J. Algebra*, 172(1):134–146. MR1320624 (96c:20044)

[1977] J. D. Biggins. Chernoff's theorem in the branching random walk. *J. Appl. Probability*, 14(3):630–636. MR0464415 (57:4345)

[2000] N. Boston. p-adic Galois representations and pro-p Galois groups. In *New horizons in pro-p groups*, volume 184 of *Progr. Math.*, pages 329–348. Birkhäuser, Boston, MA. MR1765126 (2001h:11073)

[1991] F. M. Dekking and B. Host. Limit distributions for minimal displacement of branching random walks. *Probab. Theory Related Fields*, 90(3):403–426. MR1133373 (93b:60189)

[1969] J. D. Dixon. The probability of generating the symmetric group. *Math. Z.*, 110:199–205. MR0251758 (40:4985)

[1991] J. D. Dixon, M. P. F. du Sautoy, A. Mann, and D. Segal. *Analytic pro-p-groups*, volume 157 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge. ISBN 0-521-39580-1. MR1152800 (94e:20037)

[2000] M. du Sautoy, D. Segal, and A. Shalev, editors. *New horizons in pro-p groups*, volume 184 of *Progress in Mathematics*. Birkhäuser, Boston, MA. ISBN 0-8176-4171-8. MR1765115 (2001b:20001)

[1965] P. Erdős and P. Turán. On some problems of a statistical group-theory. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 4:175–186. MR0184994 (32:2465)

[2002] S. N. Evans. Eigenvalues of random wreath products. *Electron. J. Probab.*, 7:no. 9, 15 pp. (electronic). MR1902842 (2003f:60015)

[2001] P. W. Gawron, V. V. Nekrashevych, and V. I. Sushchansky. Conjugation in tree automorphism groups. *Internat. J. Algebra Comput.*, 11(5):529–547. MR1869230 (2002k:20046)

[2000] R. I. Grigorchuk. Just infinite branch groups. In *New horizons in pro-p groups*, volume 184 of *Progr. Math.*, pages 121–179. Birkhäuser, Boston, MA. MR1765119 (2002f:20044)

[2000a] R. I. Grigorchuk, W. N. Herfort, and P. A. Zalesskii. The profinite completion of certain torsion p-groups. In *Algebra (Moscow, 1998)*, pages 113–123. de Gruyter, Berlin. MR1754662 (2001i:20058)

[2000b] R. I. Grigorchuk, V. V. Nekrashevich, and V. I. Sushchanskiĭ. Automata, dynamical systems, and groups. *Tr. Mat. Inst. Steklova*, 231(Din. Sist., Avtom. i Beskon. Gruppy):134–214. MR1841755 (2002m:37016)

[1974] J. M. Hammersley. Postulates for subadditive processes. *Ann. Probability*, 2:652–680. MR0370721 (51:6947)

[1988] L. G. Kovács and M. F. Newman. Generating transitive permutation groups. *Quart. J. Math. Oxford Ser. (2)*, 39(155):361–372. MR0957277 (89i:20008)

[2000] M. F. Newman, C. Schneider, and A. Shalev. The entropy of graded algebras. *J. Algebra*, 223(1):85–100. MR1738253 (2001c:16078)

[1983] P. P. Pálfy and M. Szalay. On a problem of P. Turán concerning Sylow subgroups. In *Studies in pure mathematics*, pages 531–542. Birkhäuser, Basel. MR0820249 (87d:11073)

[2001] J.-C. Puchta. Unpublished.

[2003] J.-C. Puchta. The order of elements of p-sylow subgroups of the symmetric group. Preprint.

[2001] L. Pyber and A. Shalev. Residual properties of groups and probabilistic methods. *C. R. Acad. Sci. Paris Sér. I Math.*, 333(4):275–278. MR1854764 (2002g:20114)

[1999] A. Shalev. Probabilistic group theory. In *Groups St. Andrews 1997 in Bath, II*, volume 261 of *London Math. Soc. Lecture Note Ser.*, pages 648–678. Cambridge Univ. Press, Cambridge. MR1676661 (2001b:20117)

[2000] A. Shalev. Lie methods in the theory of pro-p groups. In *New horizons in pro-p groups*, volume 184 of *Progr. Math.*, pages 1–54. Birkhäuser, Boston, MA. MR1765116 (2001d:20026)

[2000] S. Sidki. Automorphisms of one-rooted trees: growth, circuit structure, and acyclicity. *J. Math. Sci. (New York)*, 100(1):1925–1943. Algebra, 12. MR1774362 (2002g:05100)

[2001] S. Sidki. Oxford University Algebra Seminar Talk, October 30, 2001.

[2000] G. A. Soifer and T. N. Venkataramana. Finitely generated profinitely dense free groups in higher rank semi-simple groups. *Transform. Groups*, 5(1):93–100. MR1745714 (2001h:22017)

[1984] V. Sushchansky. Isometry groups of the Baire p-spaces. *Dop. AN URSR*, 8:27–30. (in Ukranian).

[1971] J. S. Wilson. Groups with every proper quotient finite. *Proc. Cambridge Philos. Soc.*, 69:373–391. MR0274575 (43:338)

[2000] J. S. Wilson. On just infinite abstract and profinite groups. In *New horizons in pro-p groups*, volume 184 of *Progr. Math.*, pages 181–203. Birkhäuser, Boston, MA. MR1765120 (2001i:20060)

[2002] P. A. Zalesskii. Profinite groups admitting just infinite quotients. *Monatsh. Math.*, 135(2):167–171. MR1894095 (2003b:20043)

[2000] E. Zelmanov. On groups satisfying the Golod-Shafarevich condition. In *New horizons in pro-p groups*, volume 184 of *Progr. Math.*, pages 223–232. Birkhäuser, Boston, MA. MR1765122 (2002f:20037)

Department of Mathematics, University of Chicago, 5734 University Ave., Chicago, Illinois 60637

E-mail address: abert@math.uchicago.edu

Department of Mathematics, University of Toronto, 100 St George St., Toronto, Ontario, Canada M5S 3G3

E-mail address: balint@math.toronto.edu