TREES, HOMOLOGY, AND AUTOMORPHISM GROUPS OF RIGHT-ANGLED ARTIN GROUPS

JAVIER ARAMAYONA, JOSÉ L. FERNÁNDEZ, PABLO FERNÁNDEZ AND
CONCHITA MARTÍNEZ-PÉREZ

ABSTRACT. We study the homology of an explicit finite-index subgroup of the automorphism group of a partially commutative group, in the case when its defining graph is a tree. More concretely, we give a lower bound on the first Betti number of this subgroup, based on the number and degree of a certain type of vertices, which we call *deep*. We then use combinatorial methods to analyze the average value of this Betti number, in terms of the size of the defining tree.

1. INTRODUCTION

Let K be an (undirected) finite graph, and write $V(K)$ and $E(K)$ for its set of nodes and edges, respectively. The *right-angled Artin group* (RAAG, for short) defined by K is the group A_K given by the presentation

$$A_K = \langle a \in V(K) \mid [a,b] = 1 \iff ab \in E(K) \rangle,$$

where ab denotes the edge joining a and b, and $[a,b] = aba^{-1}b^{-1}$.

Observe that if K is a complete graph with n nodes, then $A_K \cong \mathbb{Z}^n$; at the other end of the spectrum, if K has no edges then $A_K \cong F_n$, the free group on n letters. For a fixed number n of nodes, the groups A_K interpolate between these two extremal cases of $\mathbb{Z}^n$ and F_n. For instance, for a complete bipartite graph K, A_K is a direct product of two free groups, while for a disconnected graph K, A_K is the free product of the RAAGs defined by the connected components of K.

In this paper we will study the automorphism group $\mathrm{Aut}(A_K)$ of A_K which, by the discussion of the paragraph above, interpolates between the important cases of $\mathrm{Aut}(F_n)$ and $\mathrm{Aut}(\mathbb{Z}^n) = \mathrm{GL}(n, \mathbb{Z})$. More concretely, we will restrict our attention to the case when the defining graph K is a tree.

1.1. **Abelianization of finite-index subgroups.** Let G be a finitely generated group. Recall that the *abelianization* of a group G is the quotient $G^{\mathrm{ab}} := G/[G,G]$, where $[G,G]$ is the commutator subgroup of G. By definition, G^{ab} is abelian, and has a further incarnation as the first homology group $H_1(G, \mathbb{Z})$ of G. Observe that, since G^{ab} is abelian, it may be decomposed as $B \oplus \mathbb{Z}^N$, where B is a finite abelian group. The number N is called the *rank* of G^{ab}, also known as the *first Betti number* of G, denoted $b_1(G)$.

A celebrated theorem of Kazhdan [10] implies that if $G < \mathrm{Aut}(\mathbb{Z}^n) = \mathrm{GL}(n, \mathbb{Z})$ has finite index ($n \geq 3$), then $b_1(G) = 0$. Motivated by this, a

well-known open question asks whether the same holds true for finite index subgroups of $\mathrm{Aut}(F_n)$, where $n \geq 4$. We remark that the condition $n \geq 4$ is necessary, for Grunewald–Lubotzky [9] have constructed an explicit finite-index subgroup of $\mathrm{Aut}(F_3)$ with positive first Betti number. On the other hand, a recent result of Kaluba–Nowak–Ozawa [13] asserts that $\mathrm{Aut}(F_5)$ has Kazhdan's Property (T), and therefore all of its finite-index subgroups have finite abelianization.

We may consider the analogous problem for automorphism groups of arbitrary RAAGs, although one needs to be slightly careful about how to formulate it. Indeed, it is often the case that $\mathrm{Aut}(A_K)$ contains A_K as a subgroup of finite index (see Charney–Farber [4] and Day [5] for explicit results in this direction), and $b_1(A_K) \geq 1$ so long as K has at least one node. With this in mind, one may still search for combinatorial conditions on K that guarantee the existence of finite index subgroups of $\mathrm{Aut}(A_K)$ with positive first Betti number, and which apply to graphs K for which A_K has infinite index in $\mathrm{Aut}(A_K)$ (this is the case, for instance, when K is a tree). In this direction, Bregman–Fullarton [3] have identified a class of graphs K_f, which they call *focused*, such that A_{K_f} has infinite index in $\mathrm{Aut}(A_{K_f})$, and $\mathrm{Out}(A_{K_f})$ is virtually abelian [3, Proposition 3.1]. In particular, for this class of graphs $\mathrm{Aut}(A_{K_f})$ has a finite-index subgroup with infinite abelianization.

The discussion of these types of conditions is the objective of this paper. Before proceeding any further, we introduce some definitions and notations about graphs and, in particular, trees.

1.2. **Graphs and trees.** Given a graph K, the set of neighbors of a vertex $v \in V(K)$ will be called the *link* of v:

$$\mathrm{lk}(v) = \{w \in V(K) \mid vw \in E(K)\}.$$

The *degree* of v is the cardinality of $\mathrm{lk}(v)$. We define the *star* of v as $\mathrm{st}(v) := \mathrm{lk}(v) \cup \{v\}$. For simplicity, by $\mathrm{st}(v)$ we also mean the subgraph with these vertices and the edges from v in K.

If K_2 is a subgraph of a graph K_1, we denote by $K_1 \smallsetminus K_2$ the full subgraph of K_1 induced by the vertices which belong to K_1 but not to K_2.

We endow $V(K)$ with its usual combinatorial distance d, namely, given $u, v \in V(K)$, we define $d(u, v)$ as the minimal n for which there exist nodes $u = w_0, w_1, \ldots, w_n = v$ in $V(K)$ with $w_j w_{j+1} \in E(K)$ for all j.

1.2.1. *Trees.* A graph T is a *tree* if it is connected and every edge separates T into two connected components, or, alternatively, if it is connected and has no *cycles*. From now on, we will use T to refer to a tree.

A node of a tree T is a *leaf* if its degree is 1. The *boundary* ∂T of T is the set of leaves of T and, for a node v of T, we write ∂_v to abbreviate $d(v, \partial T)$, the distance from v to ∂T.

A node v of a tree T is called *deep* if $\partial_v \geq 3$. The subset of deep nodes of a tree T is denoted by $D(T)$. A tree is termed *shallow* if it has no deep nodes, i.e., all its nodes are at distance at most 2 from the boundary.

We say that a tree T is *rooted* if it has one distinguished node, called the *root* of T. A tree T with n vertices is called *labeled* if its set of vertices is $\{1, \ldots, n\}$. An *unlabeled* tree is an equivalence class of labeled trees under graph isomorphism.

1.3. **Automorphisms of RAAGs defined by trees.** In [1], the first and fourth named authors identified two properties of a graph K, each of which implies the existence of finite-index subgroups of $\mathrm{Aut}(K)$ with positive first Betti number; see Corollary 1.4 and Theorem 1.6 of [1]. In this paper we will study one of these conditions in the particular case when T is a tree. At this point we remark that, apart from forming a natural subclass, RAAGs defined by trees are also interesting from a topological viewpoint, as they are examples of fundamental groups of certain three-dimensional manifolds called *graph manifolds*.

We shall denote by $\mathrm{Aut}^\star(A_K)$ the finite-index subgroup of $\mathrm{Aut}(A_K)$ generated by *transvections*, *partial conjugations*, and *thin inversions*; see Section 2 for an expanded definition.

The following theorem is Proposition 5.3 in [1], which is simply a restatement of Theorem 1.6 in [1] in the particular case when the given graph is a tree.

Theorem ([1])**.** *If the tree T is not shallow, then $b_1(\mathrm{Aut}^\star(A_T)) \geq 1$.*

In this note we refine the methods of [1] in order to give a better lower bound on this rank, again in the particular case when T is a tree.

We introduce, for any tree T, the graph invariant

$$\Upsilon(T) := \sum_{v \in D(T)} \sum_{w \in \mathrm{lk}(v)} (\deg(w) - 1).$$

As we shall see in Section 2, the invariant $\Upsilon(T)$ counts precisely the number of the so-called partial conjugations of A_T defined by deep nodes.

Observe that for a deep node v in a tree, $\sum_{w \in \mathrm{lk}(v)} (\deg(w) - 1) \geq 2$, and thus for any tree T,

$$(1.1) \qquad\qquad \Upsilon(T) \geq 2|D(T)|.$$

The first result of this paper is the following lower bound of the Betti number of $\mathrm{Aut}^\star(A_T)$.

Theorem A. *For any tree T, the bound $b_1(\mathrm{Aut}^\star(A_T)) \geq \Upsilon(T)$ holds.*

This result implies that if T is a tree with at least one deep node, then $b_1(\mathrm{Aut}^\star(A_T)) \geq 1$, as asserted in the result from [1] stated above.

1.4. **Combinatorics of deep nodes.** Next we turn our attention to the combinatorial analysis of deep nodes and shallow trees, and to the study of the "typical size" of the combinatorial invariant $\Upsilon(T)$.

We will carry out this study in terms of *labeled* trees. This corresponds to considering RAAGs with labeled generators. Of course, isomorphism classes of *unlabeled* trees correspond to isomorphism classes of RAAGs.

(The combinatorics of unlabeled trees is much more involved than the one for labeled trees. Compare, for instance, Otter's formula for the number of unrooted unlabeled trees with n vertices, see page 481 in [8], with Cayley's formula for the labeled case.)

Let $\mathcal{U}_n$ denote the set of trees with n nodes labeled with $\{1, \ldots, n\}$. Cayley's formula says that the cardinality of this set $\mathcal{U}_n$ is exactly n^{n-2} for $n \geq 1$.

As we will see below (Theorem B), for a typical tree T in $\mathcal{U}_n$, $b_1(\mathrm{Aut}^\star(A_T))$ is quite large; although we point out that for every n there are trees T

with n nodes such that $b_1(\mathrm{Aut}^\star(A_T)) = 0$; see Lemma 2.11. It seems that the proportion of trees in $\mathcal{U}_n$ for which $b_1(\mathrm{Aut}^\star(A_T)) = 0$ is asymptotically negligible as $n \to \infty$. It would be nice to ascertain this, and to establish the precise speed of convergence to 0.

Theorem 3.7 below asserts that

$$\lim_{n\to\infty} \frac{1}{|\mathcal{U}_n|} \sum_{T\in\mathcal{U}_n} \frac{|D(T)|}{n} = c_3 \,.$$

The constant c_3 is about 0.3522, and thus we can say that for n large a typical labeled tree with n nodes has about 35.22% of deep nodes.

Concerning the invariant $\Upsilon(T)$, as we will see in Theorem 3.8, we have that

$$\lim_{n\to\infty} \frac{1}{|\mathcal{U}_n|} \sum_{T\in\mathcal{U}_n} \frac{|\Upsilon(T)|}{n} = d_3 \,;$$

the value of the constant d_3 is ≈ 2.070. In particular, we may say that for n large and a typical tree $T \in \mathcal{U}_n$, the invariant $\Upsilon(T)$ is about $d_3\, n$.

In other words, we will get:

Theorem B. *For n large and a typical tree $T \in U_n$, we have*

$$b_1(\mathrm{Aut}^\star(A_T)) \geq d_3\, n.$$

More concretely,

$$\liminf_{n\to\infty} \frac{1}{|U_n|} \sum_{T\in U_n} \frac{b_1(\mathrm{Aut}^\star(A_T))}{n} \geq d_3.$$

For the sake of completeness at this point, we remark that the explicit values of the constants c_3 and d_3 are

$$c_3 = e^{-1-2/e+e^{-1/e}} \quad \text{and} \quad d_3 = 2 - \frac{1}{e} + \left(1 - \frac{1}{e}\right) e^{-1/e} \,.$$

1.5. **Plan of the paper.** Section 2 is devoted to the proof of Theorem A. Section 3 contains the combinatorial analysis which leads to the proof of Theorem B.

2. RAAGs AND THEIR AUTOMORPHISMS

Recall from the introduction that, given a finite graph K, the *right-angled Artin group* (RAAG, for short) defined by K is the group A_K with presentation

$$A_K = \langle v \in V(K) \mid [v, w] = 1 \iff vw \in E(K)\rangle.$$

In order to relax notation, we will blur the distinction between nodes of K and generators of A_K. For instance, given a vertex $v \in V(K)$ we will write v^{-1} for the inverse of the generator corresponding to v in A_K. In addition, we will write $V(K)^{-1}$ for the set of inverses of elements of $V(K)$, when viewed as generators of A_K.

Here we will focus on the automorphism group $\mathrm{Aut}(A_K)$ of A_K. Our first aim is to describe a standard generating set for $\mathrm{Aut}(A_K)$, introduced by Laurence [11] and Servatius [15]. Before doing so, we will need to introduce a certain partial order on the set of vertices of K.

2.1. **A partial order on the set of nodes.** There is a standard partial order $\leq$ on the set of nodes of K, whereby $v \leq w$ if $\mathrm{lk}(v) \subset \mathrm{st}(w)$. We will write $v \sim w$ to mean $v \leq w$ and $w \leq v$; it is easy to see that $\sim$ is an equivalence relation. We will say that a node $v \in V(K)$ is *thin* if its equivalence class, with respect to $\sim$, has exactly one element.

We record the following observation for future use:

Lemma 2.1. *Let T be a tree with at least three nodes, and let v and w be distinct vertices. Then*

- (i) *$v \leq w$ if and only if $v \in \partial T$ and $d(v, w) \leq 2$;*
- (ii) *$v \sim w$ if and only if $v, w \in \partial T$ and there exists $u \in V(T)$ with $v, w \in \mathrm{lk}(u)$.*

Note that (ii) above implies that the $\sim$-equivalence classes of nodes with $k \geq 2$ elements consist precisely of sets of k leaves which are neighbors of a same node. A further consequence is that the thin nodes of a tree are either nodes that are not leaves, or leaves whose only neighbor is not connected to other leaves.

2.2. **Laurence–Servatius generators.** We distinguish the following four types of automorphisms of $\mathrm{Aut}(A_K)$:

- (i) *Graphic automorphisms.* Every automorphism of K induces an element of $\mathrm{Aut}(A_K)$, which we call *graphic*.
- (ii) *Inversions.* Given $v \in V(K)$, the *inversion* on v is the automorphism that sends v to v^{-1}, and fixes the rest of generators.
- (iii) *Transvections.* Given $u, v \in V(K)$, the *transvection* t_{uv} sends u to uv, and fixes the rest of generators. It is not difficult to see that $t_{uv} \in \mathrm{Aut}(A_K)$ if and only if $u \leq v$.
- (iv) *Partial conjugations.* Let $u \in V(K)$, and let Y be a connected component of $K \smallsetminus \mathrm{st}(u)$. The *partial conjugation* $c_{Y,u}$ is the automorphism given by $c_{Y,u}(v) = u^{-1}vu$ if $v \in Y$, and $c_{Y,u}(v) = v$ otherwise.

Laurence [11] and Servatius [15] proved that these four types of automorphisms suffice to generate $\mathrm{Aut}(A_K)$:

Theorem 2.2 ([11], [15])**.** *Let K be any graph. Then $\mathrm{Aut}(A_K)$ is generated by the sets of graphic automorphisms, inversions, transvections, and partial conjugations.*

2.3. **Day's presentation of** $\text{Aut}(A_K)$**.** More recently, building on work of McCool [12], Day [6] gave an explicit finite presentation of $\text{Aut}(A_K)$ in terms of *Whitehead automorphisms*, which we now briefly recall.

Let $L = V(K) \cup V(K)^{-1} \subset A_K$, and consider the obvious extension to L of the partial order $\leq$. A *type* (1) *Whitehead automorphism* is an automorphism of A_K which is induced by a permutation of L. A *type* (2) *Whitehead automorphism* is determined by a subset $A \subset L$, plus an $a \in L$ with $a \in A$ but $a^{-1} \notin A$. Then we set $(A, a)(a) = a$ and, for $c \neq a$,

$$(A, a)(c) = \begin{cases} c, & \text{if } c \notin A \text{ and } c^{-1} \notin A, \\ ca, & \text{if } c \in A \text{ and } c^{-1} \notin A, \\ a^{-1}c, & \text{if } c \notin A \text{ and } c^{-1} \in A, \\ a^{-1}ca, & \text{if } c \in A \text{ and } c^{-1} \in A. \end{cases}$$

As remarked by Day [6], not every choice of $A \subset L$ and $a \in L$ gives rise to an automorphism of A_K. Using a slight abuse of notation, by the *link* (resp. *star*) of $a \in L \setminus K$ we mean the link (resp. star) in K of a^{-1}. In [6], Lemma 2.5, Day proved the following result.

Lemma 2.3 ([6], Lemma 2.5)**.** *Let* $A \subset L$, *and* $a \in L$ *with* $a \in A$ *but* $a^{-1} \notin A$. *Then* $(A, a) \in \text{Aut}(A_K)$ *if and only if*

(1) *the set* $K \cap A \cap A^{-1} \setminus \text{lk}(a)$ *is a union of connected components of* $K \setminus \text{st}(a)$,
(2) *for each* $x \in A - A^{-1}$ *we have* $x \leq a$.

Remark 2.4. Observe that every Laurence–Servatius generator of $\text{Aut}(A_K)$ may be expressed in terms of Whitehead automorphisms. This is clear for graphic automorphisms and inversions, which are type (1) automorphisms.

In the case of partial conjugations, if Y is a union of connected components of $K \setminus \text{st}(a)$,

$$c_{Y,a} = (Y \cup Y^{-1} \cup a, a)$$

and in particular

$$c_a := c_{K-a,a} = (L - a^{-1}, a)$$

Finally, if τ_{ba} is a transvection (so, in particular, $b \leq a$) then

$$\tau_{ba} = (\{a, b\}, a).$$

In [6], Day proved the following.

Theorem 2.5 ([6])**.** $\text{Aut}(A_K)$ *is the group generated by the set of all White-head automorphisms, subject to the following relations:*

(R1) $(A, a)^{-1} = (A - a \cup a^{-1}, a^{-1})$,
(R2) $(A, a)(B, a) = (A \cup B, a)$ *whenever* $A \cap B = \{a\}$,
(R3) $(B, b)(A, a)(B, b)^{-1} = (A, a)$, *whenever* $\{a, a^{-1}\} \cap B = \varnothing$, $\{b, b^{-1}\} \cap A = \varnothing$, *and at least one of* $A \cap B = \varnothing$ *or* $b \in \text{lk}(a)$ *holds,*
(R4) $(B, b)(A, a)(B, b)^{-1} = (A, a)(B - b \cup a, a)$, *whenever* $\{a, a^{-1}\} \cap B = \varnothing$, $b \notin A$, $b^{-1} \in A$, *and at least one of* $A \cap B = \varnothing$ *or* $b \in \text{lk}(a)$ *holds,*
(R5) $(A - a \cup a^{-1}, b)(A, a) = (A - b \cup b^{-1}, a)\sigma_{a,b}$, *where* $b \in A$, $b^{-1} \notin A$, $b \neq a$ *but* $b \sim a$, *and where* $\sigma_{a,b}$ *is the type* (1) *automorphism such that* $\sigma_{a,b}(a) = b^{-1}$, $\sigma_{a,b}(b) = a$, *fixing the rest of generators.*

(R6) $\sigma(A,a)\sigma^{-1} = (\sigma(A),\sigma(a))$, *for every σ of type* (1).
(R7) *All the relations among type* (1) *Whitehead automorphisms.*
(R9) $(A,a)(L-b^{-1},b)(A,a)^{-1} = (L-b^{-1},b)$, *whenever* $\{b,b^{-1}\}\cap A = \varnothing$, *and*
(R10) $(A,a)(L-b^{-1},b)(A,a)^{-1} = (L-a^{-1},a)(L-b^{-1},b)$, *whenever* $b \in A$ *and* $b^{-1} \notin A$.

Remark 2.6. In Day's list of relations [6] there is an extra type of relator, which Day calls (R8); however, as he mentions in [6], Remark 2.9, this relation is redundant and therefore we omit it from the list above.

2.4. **The group** $\mathrm{Aut}^\star(A_K)$**.** From now on we will restrict our attention to an explicit finite-index subgroup of $\mathrm{Aut}(A_K)$, which we will denote by $\mathrm{Aut}^\star(A_K)$. Before introducing this subgroup, we need a definition. Recall that a node is said to be thin if its equivalence class, with respect to the relation $\sim$, has only one element. Consequently, we call an inversion *thin* if it fixes every thin node; in other words, it is the inversion about a node that is not thin.

Now, let $\mathrm{Aut}^\star(A_K)$ be the subgroup of $\mathrm{Aut}(A_K)$ generated by transvections, partial conjugations, and thin inversions. Observe that $\mathrm{Aut}^\star(A_K)$ has finite index in $\mathrm{Aut}(A_K)$.

In [1], the first and fourth named authors proved that $\mathrm{Aut}(A_K)$ has a finite-index subgroup that surjects onto $\mathbb{Z}$. In that paper, it was claimed that one such finite-index subgroup is the one generated by transvections, partial conjugations, and *all* inversions, which was denoted $\mathrm{Aut}^0(A_K)$. However, the proof given in [1] is not correct; this issue was fixed in an updated version of [1] (see [2]), where it was proved that $\mathrm{Aut}^\star(A_K)$ surjects to $\mathbb{Z}$. In order to do so, one needs to prove that Day's presentation can be restricted in the obvious way to a presentation for $\mathrm{Aut}^\star(A_K)$. For completeness, we include a proof here. Write $\mathrm{Sym}^1(A_K)$ for the subgroup of $\mathrm{Aut}(A_K)$ consisting of thin inversions, plus those graphic automorphisms that preserve setwise the equivalence classes for $\sim$ and fix every node of K that is thin. One has:

Proposition 2.7. *The group* $\mathrm{Aut}^\star(A_K)$ *has a finite presentation with generators the set S of type* (2) *Whitehead automorphisms and* $\mathrm{Sym}^1(A_K)$, *and relators* (R1), (R2), (R3), (R4), (R5), (R9), (R10) *above together with*

(R6)' $\sigma(A,a)\sigma^{-1} = (\sigma(A),\sigma(a))$, *for every $\sigma \in \mathrm{Sym}^1(A_K)$.*
(R7)' *All the relations among automorphisms in* $\mathrm{Sym}^1(A_K)$.

Proof. First, it follows directly from the definition that $\mathrm{Aut}^\star(K)$ is generated by all the type (2) Whitehead automorphisms, and every thin inversion. Thanks to relator (R5), we may add the elements of $\mathrm{Sym}^1(A_K)$ to this list of generators.

Let R^1 be the list (R1)–(R10) of Day's relators, except that (R6) and (R7) are substituted by (R6)' and (R7)'. Observe that every relation in R^1 is indeed a relation in $\mathrm{Aut}^\star(A_K)$. Therefore, it remains to justify why these form a complete set of relations in $\mathrm{Aut}^\star(A_K)$.

By Theorem B of [6], every automorphism $\alpha \in \mathrm{Aut}(A_K)$ may be written as a product $\alpha = \beta\delta$, where β lies in the subgroup of $\mathrm{Aut}(A_K)$ generated by *short-range* automorphisms, and δ is in the subgroup generated by *long-range* automorphisms. Here, we say that $\gamma \in \mathrm{Aut}(A_K)$ is *long-range* if either

it is a type (1) Whitehead automorphism, or it is a type (2) Whitehead automorphism specified by a subset (A, v) such that γ fixes all the elements adjacent to v in K. Similarly, we say that $\gamma \in A_K$ is *short-range* if it is a type (2) Whitehead automorphism specified by a subset (A, v) and γ fixes all the elements of K not adjacent to v. Following Day, we denote by Ω_l (resp. Ω_s) the set of all long-range (resp. short-range) automorphisms.

Consider now $\alpha \in \text{Aut}^\star(A_K)$, and observe that all short-range automorphisms are in $\text{Aut}^\star(A_K)$. The proof of the splitting in Theorem B of [6] is based in the so called *sorting substitutions* in [6], Definition 3.2. Of these, only substitution (3.1) involves an element possibly not in $\text{Aut}^\star(A_K)$, and this element is just moved along, meaning that if our initial string consists solely of elements in the generating set S, then so does the final string. Moreover, observe that the relators needed for these moves all lie in R^1 (an explicit list of the relators needed, case by case, can be found in Lemma 3.4 of [6]). All this implies that up to conjugates of relators in R^1, we may write $\alpha = \beta\delta$, with β in the subgroup of $\text{Aut}^\star(A_K)$ generated by Ω_s, and δ in the subgroup generated by $\Omega_l^1 = \Omega_l \cap \text{Aut}^\star(A_K)$.

By Proposition 5.5 of [6], the subgroup of $\text{Aut}^\star(A_K)$ generated by Ω_s has a presentation where every generator is a short-range automorphism or an element of $\text{Sym}^1(A_K)$, and where every relator lies in R^1. Indeed, in the proof of Proposition 5.5 in [6], the generators that we need to add to Ω_s to get the desired presentation are precisely the elements of the form σ_{ab} of (R5), which belong to $\text{Sym}^1(A_K)$.

In addition, the subgroup $\text{Aut}^\star(A_K)$ generated by Ω_l^1 has a presentation where every relator is in R^1. To see that this is indeed the case, first recall from Proposition 5.4 of [6] that the subgroup of $\text{Aut}(A_K)$ generated by Ω_l admits a presentation in which every relation (also in the list (R1)–(R10) of Theorem 2.5) is written in terms of Ω_l. In order to prove this, Day uses a certain inductive argument called the *peak reduction algorithm*. However, by Remark 3.22 of [6], every element of $\text{Aut}^\star(A_K)$ may be peak-reduced using elements of $\text{Aut}^\star(A_K)$ *only*. Indeed, the only subcase of Remark 3.22 in [6] that is problematic in this setting is the use of subcase (3c) of Lemma 3.18 in [6]. But the relator used in that subcase is precisely (R5), where the type (1) Whitehead automorphism is σ_{ab}, and thus lies in $\text{Sym}^1(A_K)$.

Moreover, the process of peak reduction needs relators in R^1 only; this is a consequence of the fact, observed already in Remark 3.22 of [6], that type (1) Whitehead automorphisms are only moved around when lowering peaks, and if they lie in Ω_l^1 then the needed relator is precisely (R5), where the type (1) Whitehead automorphism is σ_{ab} and thus lies in $\text{Sym}^1(A_K)$.   □

Remark 2.8. Let H be any subgroup of type (1) automorphisms which contains $\sigma_{a,b}$ for all a, b with $a \sim b$. As one of the referees pointed out to us, the obvious adaptation of the argument above may be used to give a finite presentation of the subgroup $\text{Aut}^H(A_K)$ of $\text{Aut}(A_K)$ generated by H and all type (2) automorphisms.

2.5. **Proof of Theorem A.** In what follows we will assume that T is a tree with at least 3 nodes. Recall that ∂T denotes the set of leaves of T, that is, the set of nodes of degree one. Before embarking in the proof of Theorem A, we make some preliminary observations.

First, an immediate consequence of Lemma 2.1 is that if a is a deep node of T, then there is no transvection of the form τ_{ca}. Furthermore, recall that the same lemma implies that the $\sim$-equivalence classes with more than one element consist precisely of sets of $k \geq 2$ leaves adjacent to a same node. Thus the subgroup of $\mathrm{Aut}(A_T)$ whose elements are those graphic automorphisms which (setwise) preserve these classes is generated by the graphic automorphisms that fix the whole T, apart from two leaves adjacent to a same node, which are possibly interchanged by an involution.

Let $a \in V(T)$. Observe that the number of partial conjugations of the form $c_{Y,a}$ coincides with the number of connected components of $T \smallsetminus \mathrm{st}(a)$. Moreover, this number can be computed as

$$\sum_{c \in \mathrm{lk}(a)} (\mathrm{degree}(c) - 1).$$

Set

$$\Omega = \bigcup \{c_{Y,a} \mid d(a, \partial T) \geq 3, Y \text{ connected component of } T \smallsetminus \mathrm{st}(a)\}\}.$$

Finally, in order to relax notation, we will simply write H_1 instead of $H_1(\mathrm{Aut}^\star(A_T), \mathbb{Z})$. After all this notation, Theorem A will be a consequence of the following stronger result.

Theorem 2.9. *Let* $\pi \colon \mathrm{Aut}^\star(A_T) \to H_1$ *be the abelianization map. Then* $\pi(\Omega)$ *is a linearly independent set in* H_1.

Accepting momentarily the validity of Theorem 2.9, we now explain how to deduce Theorem A from it:

Proof of Theorem A. In the light of the discussion before Theorem 2.9, we have that the cardinality of Ω is equal to

$$\Upsilon(T) = \sum_{v \in D(T)} \sum_{w \in \mathrm{lk}(v)} (\mathrm{deg}(w) - 1),$$

where again $D(T)$ denotes the set of deep nodes of T. Thus the result follows from Theorem 2.9. $\square$

Finally, we prove Theorem 2.9.

Proof of Theorem 2.9. Let n be the cardinality of Ω and consider the map

$$\varphi : \Omega \to \bigoplus_{c \in \Omega} \mathbb{Z}$$
$$c \mapsto 1_c.$$

We claim that this map can be extended to a well defined epimorphism $\mathrm{Aut}^\star(A_T) \twoheadrightarrow \mathbb{Z}^{|\Omega|}$. To show this, we will first extend φ to the set of Whitehead automorphisms that generate $\mathrm{Aut}^\star(A_T)$ and then check that the Day relators are preserved. In order to do so, we map all automorphisms in $\mathrm{Sym}^1(A_T)$ to 0.

Consider an arbitrary type (2) Whitehead automorphism (A, a). If there is some leaf b such that $d(a, b) \leq 2$, then we map $(A, a) \mapsto 0$. Otherwise, assume first that a is a node of T. Using Lemma 2.3 and relators (R2) we may write (A, a) as a product of partial conjugations $c_{Y,a} \in \Omega$ (observe that there is no element $b \leq a$) and the set of possible $Y's$ appearing in this

expression is uniquely determined from A. We define the image of (A, a) in the obvious way using this expression; note that the last observation implies that this is well defined. Finally, in the case when $a^{-1} \in T$, set $\varphi(A, a) = -\varphi(A - a \cup a^{-1}, a^{-1})$. Now we have an extended map which we also denote φ and claim that it respects Day relators. We do not have to worry about (R1) and (R2) because of the way φ is defined. About (R3) and (R9), they are preserved because $\mathbb{Z}^n$ is abelian. For (R6)' and (R7)' we only have to consider elements in $\mathrm{Sym}^1(A_T)$. Relator (R7)' is not an issue either, because all the terms therein vanish. So we are left with (R4), (R5) and (R10) and (R6)'. About (R4), as $\mathbb{Z}^n$ is abelian we only have to check that φ maps $(B - b \cup a, a)$ to 0, but this is obvious because the facts that $b \notin A$, $b^{-1} \in A$ and that (A, a) is well defined imply $b \leq a$, hence b is a leaf and $d(a, b) \leq 2$. Exactly the same argument works for (R5) and (R10): in the case of (R5) we have $a \sim b$, thus both are leaves and everything is mapped to 0. And in the case of (R10), we know that $b \in A$, $b^{-1} \notin A$, and (A, a) is well defined; thus $b \leq a$, and we argue as before to conclude that $(L - a^{-1}, a)$ is mapped to 0.

At this point, we only have to consider (R6)'. We claim that if a is a deep node, and (A, a) is well defined, then $(\sigma(A), \sigma(a)) = (A, a)$ for any $\sigma \in \mathrm{Sym}^1(A_T)$; note that this will imply that φ preserves (R6)'. In fact, it suffices to show the claim for $A = Y \cup Y^{-1} \cup a$ and Y a connected component of $T \smallsetminus \mathrm{st}(a)$. As T is a tree, such a Y must have more than one element and must itself be a tree with a vertex in $\mathrm{lk}(a)$, which we may regard as the root of Y. Moreover, if $c \sim b$ are leaves in T and one of them happens to be in Y then so is the other. Therefore $\sigma(Y \cup Y^{-1}) = Y \cup Y^{-1}$. On the other hand, since a is thin we have that $\sigma(a) = a$, by the definition of $\mathrm{Sym}^1(A_T)$, so the claim follows. $\qquad\square$

2.6. **A remark on the bound given by Theorem A.** Before continuing, we stress that the lower bound given by Theorem A is most definitely not sharp. On the other hand, not every element of $\mathrm{Aut}^\star(A_T)$ projects to a non-trivial element of H_1. In this direction, we have:

Lemma 2.10. *Let T be a tree with at least three vertices, and let $\mathrm{Aut}^*(A_T) \to H_1$ be the abelianization map. The following elements have trivial image:*

  i) *Every transvection $t_{da} = (\{d, a\}, a)$ satisfying that:*
     – *either a is a leaf, and there is a third leaf $b \notin \{a, d\}$ such that a, b, d have a common neighbor.*
     – *d is adjacent to a, and there is a leaf $b \neq d$ adjacent to a.*
  ii) *Partial conjugations $c_{Y,a}$ where a is a leaf and there is a second leaf $b \neq a$ such that a, b have a common neighbor.*

In order to prove the lemma, we will mainly use relators (R4) and (R10). It will be useful to reformulate them as follows (we emphasize that this reformulation does not make use of the hypothesis that T is a tree).

  (R4) Let $B_1 \subseteq L$ be such that (B_1, a) is well defined. Assume that there is some $b \in L$ with $b \leq a$ and $b, b^{-1} \notin B_1$ such that $(B_1 - a \cup b, b)$ is well defined and that for some $A \subseteq L$ we have (A, a) well defined, $b \notin A$, $b^{-1} \in A$, and at least one of $A \cap B_1 = \{a\}$ or $b \in \mathrm{lk}(a)$ holds,

Then
$$(B_1, a) \text{ vanishes in } H_1.$$

(R10) Let $b, a \in L$ such that $b \leq a$. Then
$$c_a \text{ vanishes in } H_1,$$
where c_a denotes conjugation (of every node of T) by a.

We are now ready to prove Lemma 2.10:

Proof of Lemma 2.10. First, note that since t_{da} is defined, then d is necessarily a leaf by Lemma 2.1. Moreover, in both cases we have $b \leq a$, and thus the element (A, a), with $A = \{b^{-1}, a\}$, is well defined. Now, in case i) let $B_1 = \{d, a\}$ so $t_{da} = (B_1, a)$. As the hypothesis implies $d \leq b$, we see that $(B_1 - a \cup b, b) = (\{d, b\}, b)$ is well defined, thus using (R4) we deduce that t_{da} vanishes in H_1.

Consider now case ii). Let
$$T - \mathrm{st}(a) = \{b\} \sqcup Y_1 \sqcup \cdots \sqcup Y_t$$

be the partition of $T \smallsetminus \mathrm{st}(a)$ into connected components. Observe that the connected components of $T \smallsetminus \mathrm{st}(b)$ are precisely
$$T - \mathrm{st}(b) = \{a\} \sqcup Y_1 \sqcup \cdots \sqcup Y_t$$

also. For any i, set $B_1 = Y_i \cup Y_i^{-1} \cup a$ and as before $A = \{b^{-1}, a\}$. Using (R4) we deduce that $c_{Y_i,a}$ vanishes in H_1. Moreover, the fact that $b \leq a$ implies by (R10) that c_a also vanishes in H_1, and as an iterated use of (R2) implies
$$c_a = c_{b,a} \prod c_{Y_i,a},$$
we see that the same happens for $c_{b,a}$. $\qquad\square$

As a consequence, we may easily exhibit a class of trees T for which the first Betti number of $\mathrm{Aut}^\star(A_T)$ vanishes.

Lemma 2.11. *Let T be a tree such that every node is either a leaf or it has at least three leaves as neighbors. Then $b_1(\mathrm{Aut}^\star(A_T)) = 0$.*

Proof. Observe first that the result is true if T is just an edge. Otherwise, recall that a consequence of Day's presentation is that $\mathrm{Aut}^\star(A_T)$ is generated by certain type (1) Whitehead automorphisms, all of which have finite order, and the following two kinds of type (2) Whitehead automorphisms:

  i) Transvections $\tau_{y,a} = (\{y, a\}, a)$ with $y \leq a$,
  ii) Partial conjugations $c_{Y,a} = (Y \cup Y^{-1} \cup a, a)$ with Y a connected component of $T - \mathrm{st}_T(a)$.

Therefore it suffices to check that both types of elements i) and ii) vanish in H_1. In case i) this follows from the hypothesis and Lemma 2.10. The same happens in case ii) unless a is not a leaf. In this latter case, take the vertex $z \in \mathrm{st}(a)$ connects a to Y and a leaf b that is adjacent to z. Observe that in a similar way as we did in Lemma 2.10, putting $A = \{b^{-1}, a\}$ and $B_1 = Y \cup Y^{-1} \cup a$, relator (R4) implies that $c_{Y,a}$ vanishes in H_1. $\qquad\square$

Before closing this section, we briefly discuss an example of a type of tree T such that $\mathrm{Aut}(A_T)$ has infinitely many finite-index subgroups with zero Betti number. Specifically, suppose T contains a vertex with degree n, and n leaves as neighbors. Then $A_T = \mathbb{Z} \times F_n$, where the $\mathbb{Z}$-factor is generated by the vertex of degree n. This group satisfies properties (B1) and (B2) in [1], and thus, by Theorem 1.1 in that paper, we obtain that $b_1(H) = 0$ for any $H \leq \mathrm{Aut}(A_T)$ of finite index containing the Torelli subgroup.

In the light of these results, a natural question is:

**Question 2.12.** *Let T be a tree. What is the exact value of $b_1(\mathrm{Aut}^\star(A_T))$?*

More generally, one may ask:

Question 2.13. *Let K be an arbitrary graph, and $\mathrm{Aut}^1(A_K)$ the subgroup of $\mathrm{Aut}(A_K)$ generated by transvections and partial conjugations. Calculate $b_1(\mathrm{Aut}^1(A_K))$ in terms of the combinatorics of K.*

3. Deep nodes and shallow trees

Recall that a node v of a tree T is called deep if $\partial_v \geq 3$, that the collection of deep nodes of T is denoted $D(T)$, and that a tree T with no deep nodes is termed shallow.

Some examples of shallow trees and trees with deep nodes follow; colors indicate distance to the boundary.
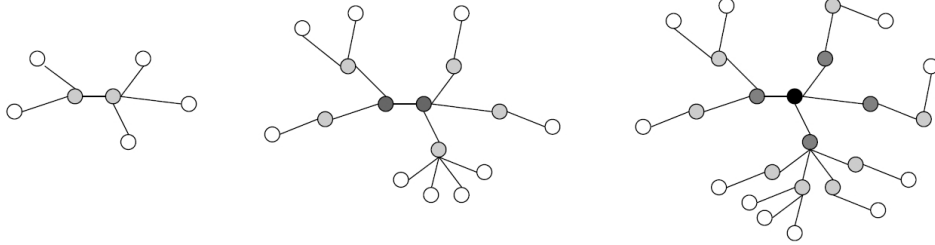
The class of rooted labeled trees is denoted by $\mathcal{T}$, while the class of general (unrooted) labeled trees is denoted by $\mathcal{U}$. The respective subclasses of trees with nodes labeled with $\{1, \ldots, n\}$ are denoted $\mathcal{T}_n$ and $\mathcal{U}_n$, for each $n \geq 1$. Cayley's formula (see [8], page 127) tells us that

$$t_n := |\mathcal{T}_n| = n^{n-1}, \quad \text{for } n \geq 1,$$

and that

$$u_n := |\mathcal{U}_n| = n^{n-2}, \quad \text{for } n \geq 1.$$

We endow $\mathcal{U}_n$ with the uniform probability distribution; claiming that a certain property occurs with probability p in $\mathcal{U}_n$ is tantamount to claiming that the proportion of trees in $\mathcal{U}_n$ satisfying that property is p.

3.1. **Notation and some basic results.** The "symbolic method" readily translates questions about counting trees into analytic questions about power series. We introduce here some basic notation and results, and refer to the comprehensive treatise of Flajolet and Sedgewick [8] for details.

Given a sequence $(a_n)_{n=0}^{\infty}$, its (*ordinary*) *generating function* (ogf, for short) is the power series $f(z)$ given by

$$f(z) = \sum_{n=0}^{\infty} a_n \, z^n$$

for all $z \in \mathbb{D}(0, \varepsilon)$, for some $\varepsilon > 0$. We will write $a_n = \text{COEFF}_n(f(z))$.

The function $g(z)$ is the *exponential generating function* (for short, egf) of the sequence (a_n) if

$$g(z) = \sum_{n=0}^{\infty} \frac{a_n}{n!} \, z^n$$

for all $z \in \mathbb{D}(0, \varepsilon)$, for some $\varepsilon > 0$.

A basic tool for handling combinatorial questions about trees is the Lagrange inversion formula (see, for instance, Section A.6, page 732, in [8]).

Lemma 3.1 (Lagrange inversion formula)**.** *Let $h(z)$ and $f(z)$ be two holomorphic functions on some neighborhood of $z = 0$, say $\mathbb{D}(0, \varepsilon)$, such that $f(0) \neq 0$, and*

$$h(z) = z f(h(z))$$

in $\mathbb{D}(0, \delta)$. Then, for any function g holomorphic at 0,

$$\text{COEFF}_n \left[g(h(z)) \right] = \text{COEFF}_{n-1} \left[g'(z) \frac{f(z)^n}{n} \right], \qquad \textit{for each } n \geq 1 \,.$$

Note that $h(0) = 0$.

Trees and generating functions. We let $T(z)$ denote the egf of the sequence (t_n), namely:

$$(3.1) \qquad\qquad T(z) = \sum_{n=0}^{\infty} \frac{t_n}{n!} \, z^n, \quad |z| < \frac{1}{e}.$$

The function $T(z)$ satisfies the following implicit equation:

$$(3.2) \qquad\qquad\qquad T(z) = z \, e^{T(z)}.$$

Here, the factor z accounts for the root, while $e^{T(z)}$ represents the (possibly empty) set of trees stemming from the root. See the details in [8], page 127.

As an immediate corollary of Lemma 3.1 we state:

Corollary 3.2. *For $n, k \geq 1$,*

$$\text{COEFF}_n \left[T(z)^k \right] = \frac{k}{n} \, \frac{n^{n-k}}{(n-k)!}.$$

Stirling numbers. For $n \geq 0$ and $k \geq 0$, the symbol $S(n, k)$ stands for the (double) sequence of Stirling numbers of the second kind, that count the number of partitions of a set of size n into k non-empty (unlabeled) subsets, with the conventions $S(n, 0) = 0 = S(0, k)$ for $n \geq 1$ and $k \geq 1$, but $S(0, 0) = 1$.

We shall use the following identities (see [8], page 736). For $k \geq 0$,

$$(3.3) \qquad\qquad \sum_{n=0}^{\infty} \frac{S(n, k)}{n!} \, x^n = \frac{1}{k!} \, (e^x - 1)^k.$$

Also,

$$(3.4) \qquad \sum_{k,n \geq 0} S(n,k)\, \frac{x^n}{n!}\, y^k = e^{y(e^x - 1)}.$$

Notice that, writing the Taylor series of $(e^x - 1)^k$, and identifying coefficients, formula (3.3) gives

$$(3.5) \qquad \frac{S(n,k)}{n!} = \frac{1}{k!} \sum_{\substack{q_1,\ldots,q_k \geq 1 \\ q_1+\cdots+q_k=n}} \frac{1}{q_1! \cdots q_k!}.$$

Taking a derivative with respect to x in (3.4), we get

$$(3.6) \qquad \sum_{k \geq 0, n \geq 1} S(n,k)\, \frac{x^{n-1}}{(n-1)!}\, y^k = y\, e^x\, e^{y(e^x - 1)};$$

and multiplying by x and differentiating again with respect to x,

$$(3.7) \qquad \sum_{k \geq 0, n \geq 1} S(n,k)\, n\, \frac{x^{n-1}}{(n-1)!}\, y^k = y\, e^x\, e^{y(e^x - 1)}\, (1 + x + xy\, e^x).$$

3.2. Deep nodes.
Our objective now is to study how abundant deep nodes are in a typical labeled tree with n nodes, as $n \to \infty$. Our argument starts analyzing rooted trees (Sections 3.2.1 and 3.2.2), and then settles (Section 3.2.3) the same question about unrooted trees, which is the more relevant case for our purposes.

3.2.1. *Proportion of rooted trees with the root at distance ≥ 3 to the border.* Recall that $\mathcal{T}$ denotes the class of all rooted trees and that $\mathcal{T}_n$ denotes the subclass of rooted trees labeled with $\{1,\ldots,n\}$. Again, we endow $\mathcal{T}_n$ with the uniform probability distribution. Probabilities and expectations, denoted by $\mathbf{P}_n$ and $\mathbf{E}_n$, refer to this probability space. Recall that $t_n = |\mathcal{T}_n| = n^{n-1}$.

Call $\mathcal{T}^{(3)}$ the subclass of rooted trees whose root is a deep node, $\partial_{\mathrm{root}} \geq 3$. In such trees, the root has, say, $k \geq 1$ descendants, which in turn have $q_1,\ldots,q_k \geq 1$ descendants, none of which is a leaf (this guarantees distance ≥ 3 from the root to the leaves). Call $N = q_1 + \cdots + q_k$.

Consider the following subclasses of $\mathcal{T}$:

$$\mathcal{T}^{(3)}_{k,q_1,\ldots,q_k} = \left\{ \begin{array}{c} k \text{ descendants of the root, with } q_1,\ldots,q_k \geq 1 \\ \text{descendants, respectively, and } \partial_{\mathrm{root}} \geq 3 \end{array} \right\}$$

$$\mathcal{T}^{(3)}_{k,N} = \left\{ \begin{array}{c} k \text{ descendants of the root, } N \text{ nodes} \\ \text{in the second generation, } \partial_{\mathrm{root}} \geq 3 \end{array} \right\}$$

$$= \bigcup_{\substack{q_1,\ldots,q_k \geq 1 \\ q_1+\cdots+q_k=N}} \mathcal{T}^{(3)}_{k,q_1,\ldots,q_k}.$$

Observe that

$$\mathcal{T}^{(3)} = \bigcup_{k \geq 1} \bigcup_{N \geq k} \mathcal{T}^{(3)}_{k,N}.$$

In all cases, an extra subindex n would indicate the corresponding subclass of trees with nodes labeled with $\{1,\ldots,n\}$.

We have the following asymptotic result.
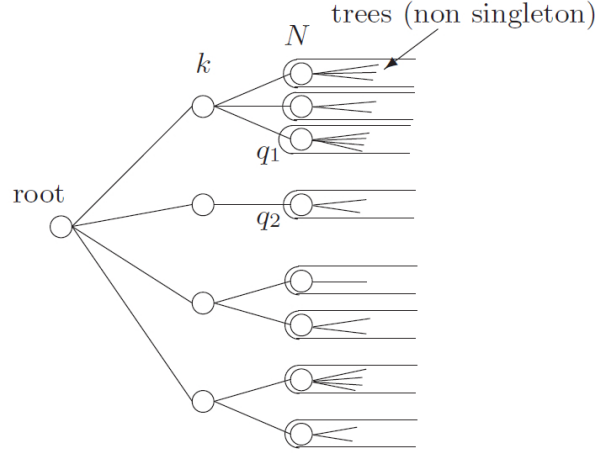
FIGURE 1. Trees with $\partial_{\text{root}} \geq 3$.

Theorem 3.3.

$$\lim_{n \to \infty} \mathbf{P}_n(\mathcal{T}_n^{(3)}) = e^{-1-2/e+e^{-1/e}} =: c_3.$$

Proof. Let $k, q_1, \ldots, q_k$ be fixed. The symbolic method (for labeled structures) of Flajolet and Sedgewick ([8]) yields that the egf of the class $\mathcal{T}_{k,q_1,\ldots,q_k}^{(3)}$ is given by

$$(3.8) \qquad \sum_{n=0}^{\infty} \frac{|\mathcal{T}_{n;k,q_1,\ldots,q_k}^{(3)}|}{n!} z^n = z \frac{z^k}{k!} \frac{(T(z)-z)^{q_1}}{q_1!} \cdots \frac{(T(z)-z)^{q_k}}{q_k!}.$$

Here, the factor z accounts for the root, the factor z^k for its k neighbours, and each factor $(T(z)-z)^{q_i}$, for the tree (with more than one vertex) springing from each one of those neighbours. The factorials in the denominators take into account the absence of order among first neighbours, and among the subsequent trees.

We can rewrite (3.8) as

$$\sum_{n=0}^{\infty} \frac{|\mathcal{T}_{n;k,q_1,\ldots,q_k}^{(3)}|}{n!} z^n = \frac{1}{k!} \frac{1}{q_1! \cdots q_k!} z^{k+1} (T(z)-z)^N,$$

from where we get that

$$\mathbf{P}_n(\mathcal{T}_{n;k,q_1,\ldots,q_k}^{(3)}) = \frac{|\mathcal{T}_{n;k,q_1,\ldots,q_k}^{(3)}|}{|\mathcal{T}_n|}$$

$$= \frac{n!}{n^{n-1}} \frac{1}{q_1! \cdots q_k!} \frac{1}{k!} \text{COEFF}_n[z^{k+1} (T(z)-z)^N].$$

Now, writing $a_j = (k+1) + (N-j)$ and using Corollary 3.2,

$$\text{COEFF}_n\left[z^{k+1}\left(T(z)-z\right)^N\right] = \text{COEFF}_{n-k-1}\left[\sum_{j=0}^{N}\binom{N}{j}T(z)^j(-1)^{N-j}z^{N-j}\right]$$

$$= \sum_{j=0}^{N}\binom{N}{j}(-1)^{N-j}\,\text{COEFF}_{n-a_j}\left[T(z)^j\right]$$

$$= \sum_{j=0}^{N}\binom{N}{j}(-1)^{N-j}\frac{j}{n-a_j}\frac{(n-a_j)^{n-a_j-j}}{(n-a_j-j)!}.$$

This yields

$$\mathbf{P}_n(\mathcal{T}_{n;k,q_1,\ldots,q_k}^{(3)})$$

$$(3.9)\qquad = \frac{n!}{n^{n-1}}\frac{1}{q_1!\cdots q_k!}\frac{1}{k!}\sum_{j=0}^{N}\binom{N}{j}(-1)^{N-j}\frac{j}{n-a_j}\frac{(n-a_j)^{n-a_j-j}}{(n-a_j-j)!}.$$

Notice that

$$\frac{n!}{n^{n-1}}\frac{1}{n-a_j}\frac{(n-a_j)^{n-a_j-j}}{(n-a_j-j)!}$$

$$(3.10)\qquad = \frac{n}{n-a_j}\frac{n(n-1)\cdots(n-a_j-j+1)}{n^{a_j+j}}\left(1-\frac{a_j}{n}\right)^{n-a_j-j},$$

which tends to e^{-a_j} when $n\to\infty$.

This gives, recalling that $a_j = (k+1) + (N-j)$, that

$$\lim_{n\to\infty}\mathbf{P}_n(\mathcal{T}_{n;k,q_1,\ldots,q_k}^{(3)}) = \frac{1}{q_1!\cdots q_k!}\frac{1}{k!}\sum_{j=0}^{N}\binom{N}{j}(-1)^{N-j}j\,e^{-a_j}$$

$$= \frac{1}{q_1!\cdots q_k!}\frac{1}{k!}e^{-N-(k+1)}\sum_{j=0}^{N}\binom{N}{j}(-1)^{N-j}j\,e^j$$

$$= \frac{1}{q_1!\cdots q_k!}\frac{1}{k!}e^{-(k+1)}N\left(1-\frac{1}{e}\right)^{N-1},$$

where in the last step we have used the binomial theorem.

Now, summing in all tuples $q_1,\ldots,q_k\geq 1$ with sum N, we get

$$\lim_{n\to\infty}\mathbf{P}_n(\mathcal{T}_{n;k,N}^{(3)}) = \frac{1}{k!}e^{-(k+1)}N\left(1-\frac{1}{e}\right)^{N-1}\sum_{\substack{q_1,\ldots,q_k\geq 1\\ q_1+\cdots+q_k=N}}\frac{1}{q_1!\cdots q_k!}$$

$$(3.11)\qquad = e^{-(k+1)}S(N,k)\frac{1}{(N-1)!}\left(1-\frac{1}{e}\right)^{N-1},$$

using (3.5).

Finally, summing over k and N, we get

$$\lim_{n\to\infty}\mathbf{P}_n(\mathcal{T}_n^{(3)}) \overset{(\star)}{=} \sum_{k,N}e^{-(k+1)}S(N,k)\frac{1}{(N-1)!}\left(1-\frac{1}{e}\right)^{N-1}$$

$$(3.12)\qquad = \frac{1}{e}\sum_{k,N}S(N,k)\frac{(1-1/e)^{N-1}}{(N-1)!}\left(\frac{1}{e}\right)^k = \frac{1}{e}\frac{1}{e}e^{1-1/e}e^{(e^{1-1/e}-1)/e}$$

(for the last identity, use (3.6) with $x = 1 - 1/e$ and $y = 1/e$).

To justify the interchange of limit and (the double) sum in $(\star)$, we observe, from (3.9) and (3.10), that

$$|\mathbf{P}_n(\mathcal{T}^{(3)}_{n;k,q_1,\ldots,q_k})| \leq \frac{1}{q_1! \cdots q_k!} \frac{1}{k!} \sum_{j=0}^{N} \binom{N}{j} j = \frac{1}{q_1! \cdots q_k!} \frac{1}{k!} \, N \, 2^{N-1},$$

and so

$$|\mathbf{P}_n(\mathcal{T}^{(3)}_{n;k,N})| \leq \frac{N}{k!} \, 2^{N-1} \sum_{\substack{q_1,\ldots,q_k \geq 1 \\ q_1 + \cdots + q_k = N}} \frac{1}{q_1! \cdots q_k!}$$

$$\leq \frac{N}{k!} \, 2^{N-1} \, \mathrm{COEFF}_N[e^{kz}] = \frac{1}{2} \frac{1}{k!} \frac{(2k)^N}{(N-1)!}.$$

As

$$\sum_{k \geq 1} \sum_{N \geq k} \frac{1}{k!} \frac{(2k)^N}{(N-1)!} < +\infty,$$

dominated convergence justifies $(\star)$. $\qquad\qquad\square$

Remark 3.4 (Rooted labeled trees with the root farther away from the leaves)**.** For $k \geq 0$, denote by $\mathcal{T}^{(k)}$ the subclass of rooted trees in which the root is, at least, k units away from the boundary ($\partial_{\mathrm{root}} \geq k$). Write $\Psi_k(z)$ for its egf.

For $k = 0$, $\mathcal{T}^{(0)} = \mathcal{T}$, and the corresponding egf is just the Cayley's function, $\Psi_0(z) = T(z)$.

The symbolic method (see [8]) gives that the sequence $(\Psi_k(z))$ of egfs satisfies the recurrence relation

$$(3.13) \qquad \Psi_0(z) = T(z), \quad \Psi_k(z) = z \left(e^{\Psi_{k-1}(z)} - 1\right), \quad k \geq 1.$$

To see this, take a tree in $\mathcal{T}^{(k)}$, delete its root (and the edges departing from it), and observe that we are left with a non-empty set of rooted trees in $\mathcal{T}^{(k-1)}$.

In particular, using Cayley's formula (3.2), we get

$$(3.14) \qquad \Psi_1(z) = z \left(e^{\Psi_0(z)} - 1\right) = z \left(e^{T(z)} - 1\right) = T(z) - z.$$

and

$$(3.15) \qquad \Psi_2(z) = z \left(e^{\Psi_1(z)} - 1\right) = z \left(e^{T(z)-z} - 1\right) = T(z) \, e^{-z} - z.$$

In the latter case, the particular structure of $\Psi_2(z)$ allows to obtain the asymptotic behaviour of its coefficients in a direct manner (avoiding a combinatorial argument similar to that used in the proof of Theorem 3.3), using a trick of Schur and Szász (see [8], Theorem VI.12, p. 434). The result in this case is that

$$\lim_{n \to \infty} \mathbf{P}_n(\mathcal{T}^{(2)}_n) = e^{-1/e} \approx 0.6922.$$

For $k = 3$, instead,

$$(3.16) \qquad \Psi_3(z) = z \left(e^{\Psi_2(z)} - 1\right) = z \, e^{T(z)e^{-z}} \, e^{-z} - z,$$

and the simple approach sketched above for $k = 2$ does not work. That is why we had to go through the combinatorial argument of the proof of Theorem 3.3, to obtain

$$\lim_{n\to\infty} \mathbf{P}_n(\mathcal{T}_n^{(3)}) = \frac{1}{e}\, e^{-1/e}\, e^{(e^{1-1/e}-1)/e} \approx 0.3522.$$

Notice that the height of a rooted tree is the *maximum* distance from the root to the leaves, while the distance ∂_{root} is the *minimum* distance from the root to the leaves. The egfs $\Phi_k(z)$ of trees of height $\leq k$ satisfy

(3.17) $$\Phi_0(z) = z, \quad \Phi_k(z) = z\, e^{\Phi_{k-1}(z)}, \quad k \geq 1.$$

The asymptotics of the proportion that rooted trees of height k occupy in $\mathcal{T}_n$ is well known, starting with the Rényi–Szekeres analysis of (3.17) (see [14]).

It would be nice to have a general analogous analysis of the recurrence (3.13) that could lead to an answer to:

Question 3.5. *For $k \geq 4$, and as $n \to \infty$, what is the proportion that trees in $\mathcal{T}_n^{(k)}$ do occupy in $\mathcal{T}_n$?*

3.2.2. *Mean of the sum of degrees of descendants of the root.* In the same probability space $\mathcal{T}_n$ (rooted trees labeled with $\{1,\ldots,n\}$, with uniform distribution), consider the random variable

$$Y_n(T) = \mathbf{1}_{\{\partial_{\text{root}}\geq 3\}} \cdot N_n(T) = \begin{cases} N_n(T) & \text{if } \partial_{\text{root}} \geq 3, \\ 0 & \text{otherwise,} \end{cases}$$

where $N_n(T)$ is the number of nodes in the second generation of the graph T counted from the root (see Figure 1). Observe that

$$N_n(T) = \sum_{v\in \text{lk(root of } T)} (\deg(v) - 1).$$

The following asymptotic result holds.

Theorem 3.6.
$$\lim_{n\to\infty} \mathbf{E}_n(Y_n) = 2 - \frac{1}{e} + \left(1 - \frac{1}{e}\right) e^{-1/e} =: d_3.$$

The numerical value of d_3 is ≈ 2.070.

Proof. Recalling (3.11) and (3.12), we observe that

$$\lim_{n\to\infty} \mathbf{E}_n(Y_n) = \frac{1}{c_3} \sum_{k,N} N \cdot S(N,k)\, \frac{(1-1/e)^{N-1}}{(N-1)!}\, e^{-(k+1)}$$

$$= \frac{1}{c_3}\frac{1}{e} \sum_{k,N} N \cdot S(N,k)\, \frac{(1-1/e)^{N-1}}{(N-1)!}\, \left(\frac{1}{e}\right)^k$$

$$= \frac{1}{c_3}\frac{1}{e}\frac{1}{e}\, e^{1-1/e}\, e^{(e^{1-1/e}-1)/e} \left[2 - \frac{1}{e} + \frac{1}{e}\left(1 - \frac{1}{e}\right) e^{1-1/e}\right]$$

$$= \left[2 - \frac{1}{e} + \frac{1}{e}\left(1 - \frac{1}{e}\right) e^{1-1/e}\right]$$

where we have used (3.7) with $x = 1 - 1/e$ and $y = 1/e$, and the value of c_3 from Theorem 3.3. The interchange of limit and double sum can be justified by dominated convergence, along the same lines as in the proof of Theorem 3.3. $\square$

3.2.3. *From rooted to unrooted trees.* Theorems 3.3 and 3.6 can be readily reinterpreted in the context of unrooted trees.

Fix n and consider the collection $\mathcal{U}_n$ of the n^{n-2} trees labeled with $\{1, \ldots, n\}$ endowed with the uniform probability. For the sake of clarity, we denote probability and expectation in $\mathcal{U}_n$ with $\mathbf{P}'_n$ and $\mathbf{E}'_n$, respectively.

Let $X_n(T)$ denote the random variable in $\mathcal{U}_n$ that counts the number of deep nodes of T:

$$X_n(T) = \sum_{v \in V(T)} \mathbf{1}_{\{\partial_v \geq 3\}} = |D(T)|$$

Consider now a 0-1 matrix M, of dimensions $n \times n^{n-2}$, with columns labeled with $T_1, T_2, \ldots$, the collection of trees in $\mathcal{U}_n$, and with rows labeled with the nodes $\{1, \ldots, n\}$, where in the (j, T_i)-entry we place a 1 if the node j of T_i is deep, and we place a 0 otherwise.

Summing the entries of the matrix M and dividing by n^{n-2}, we obtain the mean value of X_n:

$$\mathbf{E}'_n(X_n) = \frac{1}{n^{n-2}} \sum_{T \in \mathcal{U}_n} X_n(T).$$

Each (unrooted) tree T_i leads to n different rooted trees $T_i^{(1)}, \ldots, T_i^{(n)}$ by choosing any of its nodes as the root; here, $T_i^{(j)}$ means that node j has been selected as the root in the tree T_i.

Now build a 0-1 matrix M' of dimensions $n \times n^{n-1}$: rows are labeled with the n nodes, and the columns with the collection of rooted trees in the following order: first $T_1^{(1)}, \ldots, T_1^{(n)}$, then $T_2^{(1)}, \ldots, T_2^{(n)}$, etc. The value of the entry $(v_i, T_j^{(k)})$ is 1 if $i = k$ and the node i (the root of the tree $T_j^{(i)}$) is at distance ≥ 3 to the boundary; and it is 0 otherwise.

The sum of the entries of M', divided by n^{n-1}, gives the probability that in a rooted labeled tree, the root is at distance ≥ 3 to its boundary. As the sum of the entries of M' equals the sum of the entries of M, recalling Theorem 3.3, we deduce the following.

Theorem 3.7. *As $n \to \infty$, the expectation of the proportion of nodes in a labeled tree on n nodes that are at distance ≥ 3 from any leaf tends to c_3, i.e.,*

$$\lim_{n \to \infty} \frac{1}{n} \mathbf{E}'_n(X) = c_3.$$

Next, in the probability space $\mathcal{U}_n$ of unrooted trees T labeled with $\{1, \ldots, n\}$ and endowed with uniform probability, consider the random variable

$$\Upsilon(T) = \sum_{v \in V(T)} \mathbf{1}_{\{\partial_v \geq 3\}} \cdot N_v = \sum_{v \in \mathcal{D}(T)} N_v$$

where N_v is the number of nodes two units away from v. Observe that

$$N_v(T) = \sum_{w \in \mathrm{lk}(v)} (\deg(w) - 1).$$

An analogous argument as above using Theorem 3.6 instead of Theorem 3.3 yields:

Theorem 3.8.

$$\lim_{n \to \infty} \frac{1}{n} \mathbf{E}'_n(\Upsilon) = d_3.$$

References

[1] Aramayona, J. and Martínez-Pérez, C.: On the first cohomology of automorphism groups of graph groups. *J. Algebra* **452** (2016), 17–41.

[2] Aramayona, J. and Martínez-Pérez, C.: On the first cohomology of automorphism groups of graph groups. Preprint, arXiv:`1504.07449, v4`, Nov. 2015.

[3] Bregman, C. and Fullarton, N.: Infinite groups acting faithfully on the outer automorphism group of a right-angled Artin group. *Michigan Math. J.* **66** (2017), no. 3, 569–580.

[4] Charney, R. and Farber, M.: Random groups arising as graph products. *Algebr. Geom. Topol.* **12** (2012), no. 2, 979–995.

[5] Day, M. B.: Finiteness of outer automorphism groups of random right-angled Artin groups. *Algebr. Geom. Topol.* **12** (2012), no. 3, 1553–1583.

[6] Day, M. B.: Peak reduction and finite presentations for automorphism groups of right angled Artin groups. *Geom. Topol.* **13** (2009), no. 2, 817–855.

[7] Droms, C.: Isomorphisms of graph groups. *Proc. Amer. Math. Soc.* **100** (1987), no. 3, 407–408.

[8] Flajolet, P. and Sedgewick, R.: *Analytic Combinatorics.* Cambridge University Press, 2009.

[9] Grunewald, P. and Lubotzky, A.: Linear representations of the automorphism group of free groups. *Geom. Funct. Anal.* **18** (2009), no. 5, 1564–1608.

[10] Každan, D.: On the connection of the dual space of a group with the structure of its closed subgroups. (Russian). *Funkcional. Anal. i Priložen.* **1** (1967), 71–74.

[11] Laurence, M. R.: A generating set for the automorphism group of a graph group. *J. London Math. Soc. (2)* **52** (1995), no. 2, 318–334.

[12] McCool, J.: A faithful polynomial presentation of $\mathrm{Out}(F_3)$. *Math. Proc. Camb. Phil. Soc.* 106 (1989), no. 2, 207–213.

[13] Kaluba, M., Nowak, P., and Ozawa, N.: $\mathrm{Aut}(F_5)$ has property (T). Preprint, arXiv: `1712.07167`, 2017.

[14] Rényi, A. and Szekeres, G.: On the height of trees. *J. Austral. Math. Soc.* **7** (1967), 497–507.

[15] Servatius, H.: Automorphisms of graph groups. *J. Algebra* **126** (1989), no. 1, 34–60.

Javier Aramayona: Department of Mathematics, Universidad Autónoma de Madrid, 28049 Madrid, Spain; and Instituto de Ciencias Matemáticas (ICMAT).
  E-mail address: `javier.aramayona@uam.es`

José L. Fernández: Department of Mathematics, Universidad Autónoma de Madrid, 28049 Madrid, Spain.
  E-mail address: `joseluis.fernandez@uam.es`

Pablo Fernández: Department of Mathematics, Universidad Autónoma de Madrid, 28049 Madrid, Spain.
  E-mail address: `pablo.fernandez@uam.es`

Conchita Martínez-Pérez: Department of Mathematics, Universidad de Zaragoza, Pedro Cerbuna s/n, 50009 Zaragoza, Spain.
  E-mail address: `conmar@unizar.es`