String Powers in Trees

Tomasz Kociumaka(✉), Jakub Radoszewski, Wojciech Rytter, and Tomasz Waleń

Faculty of Mathematics, Informatics and Mechanics,
University of Warsaw, Warsaw, Poland
{kociumaka,jrad,rytter,walen}@mimuw.edu.pl

Abstract. We investigate the asymptotic growth of the maximal number $\mathsf{powers}_\alpha(n)$ of different α-powers (strings w with a period $|w|/\alpha$) in an edge-labeled unrooted tree of size n. The number of different powers in trees behaves much unlike in strings. In a previous work (CPM, 2012) it was proved that the number of different squares in a tree is $\mathsf{powers}_2(n) = \Theta(n^{4/3})$. We extend this result and analyze other powers. We show that there are phase-transition thresholds:

1. $\mathsf{powers}_\alpha(n) = \Theta(n^2)$ for $\alpha < 2$;
2. $\mathsf{powers}_\alpha(n) = \Theta(n^{4/3})$ for $2 \le \alpha < 3$;
3. $\mathsf{powers}_\alpha(n) = \mathcal{O}(n \log n)$ for $3 \le \alpha < 4$;
4. $\mathsf{powers}_\alpha(n) = \Theta(n)$ for $4 \le \alpha$.

The difficult case is the third point, which follows from the fact that the number of different cubes in a rooted tree is linear (in this case, only cubes passing through the root are counted).

1 Introduction

Repetitions are a fundamental notion in combinatorics on words. For the first time they were studied more than a century ago by Thue [14] in the context of square-free strings, that is, strings that do not contain substrings of the form $W^2 = WW$. Since then, α-free strings, avoiding string powers of exponent α (of the form W^α), have been studied in many different contexts; see [13]. Another line of research is related to strings that are rich in string powers. It has been shown that the number of different squares in a string of length n does not exceed $2n - \Theta(\log n)$ (see [5,7,8]); stronger bounds are known for cubes [12].

Repetitions are also considered in labeled trees and graphs. In this model, a repetition corresponds to a sequence of labels of edges (or nodes) on a simple path. The origin of this study comes from a generalization of square-free

This work was supported by the Polish National Science Center, grant no 2014/13/B/ST6/00770.

T. Kociumaka—Supported by Polish budget funds for science in 2013–2017 as a research project under the 'Diamond Grant' program.

J. Radoszewski—Supported by the Polish Ministry of Science and Higher Education under the 'Iuventus Plus' program in 2015–2016 grant no 0392/IP3/2015/73. Receives financial support of Foundation for Polish Science.

F. Cicalese et al. (Eds.): CPM 2015, LNCS 9133, pp. 284–294, 2015.
DOI: 10.1007/978-3-319-19929-0_24

strings and α-free strings, called non-repetitive colorings of graphs. A survey by Grytczuk [6] presents several results of this kind. In particular, non-repetitive colorings of labeled trees were considered [2]. Strings related to paths in graphs have also been studied in the context of hypertexts [1].

Enumeration of squares in labeled trees has already been considered from both combinatorial [4] and algorithmic point of view [9]. Our study is a continuation of the results of [4], where it has been proved that the maximum number of different squares in a labeled tree with n nodes is of the order $\Theta(n^{4/3})$. As our main result we show a *phase transition* property: for every exponent $2 < \alpha < 3$, a tree of n nodes may contain $\Omega(n^{4/3})$ string α-powers, whereas it may only have $\mathcal{O}(n \log n)$ powers of exponent $\alpha \geq 3$.

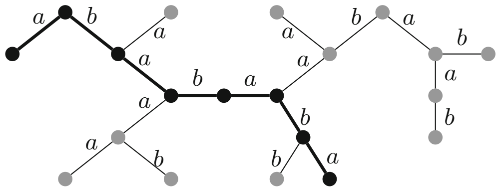

Fig. 1. There are 5 different cubic substrings in this tree: a^3, $(ab)^3$, $(ba)^3$, $(aab)^3$, $(baa)^3$. Hence, $\mathsf{powers}_3(T) = 5$. Note that the cube $(ab)^3$ occurs twice; also a^3 has multiple occurrences. The most repetitive substring, a 3.5-power $(ab)^{3.5}$, is marked in the figure.

Let T be a tree whose edges are labeled with symbols from an alphabet Σ. We denote the size of the tree, that is, the number of nodes, by $|T|$. A *substring* of T is the sequence of labels of edges on any simple path in T. We define $\mathsf{powers}_\alpha(T)$ as the number of different substrings of T which are powers of (possibly fractional) exponent α; see Fig. 1. We denote $\mathsf{powers}_\alpha(n) = \max_{|T|=n} \mathsf{powers}_\alpha(T)$. The bound $\mathsf{powers}_2(n) = \Theta(n^{4/3})$ has been shown in [4]. Here, we prove the following asymptotic bounds:

$\alpha \in (1, 2)$	$\mathsf{powers}_\alpha(n) = \Theta(n^2)$
$\alpha \in [2, 3)$	$\mathsf{powers}_\alpha(n) = \Theta(n^{4/3})$
$\alpha \in [3, 4)$	$\mathsf{powers}_\alpha(n) = \mathcal{O}(n \log n)$
$\alpha \geq 4$	$\mathsf{powers}_\alpha(n) = \Theta(n)$

2 Preliminaries

2.1 Combinatorics of Strings

Let V be a string over an alphabet Σ. We denote its letters by $V_1, \ldots, V_m$ and its length m by $|V|$. By V^R we denote the *reverse* string $V_m \ldots V_1$. For

$1 \le i \le j \le m$ a string $V[i..j] = V_i \ldots V_j$ is a *substring* of V. We say that a positive integer q is a *period* of V if $V_i = V_{i+q}$ holds for $1 \le i \le m - q$. In this case we also say that the prefix of V of length q is a period of V.

For an integer i, $1 \le i \le m$, a substring $V[1..i]$ is called a *prefix* of V, and $V[i..m]$ is called a *suffix* of V. A string U is a *border* of V if it is both a prefix and a suffix of V. It is well known that a string of length m has a border of length b if and only if it has a period $m - b$.

Fact 1 ([11]). *Let B_1, B_2 be borders of a string V. If $|B_1| < |B_2| \le 2|B_1|$, then B_1 and B_2 have the same shortest period p, which is a divisor of $|B_2| - |B_1|$.*

We say that a string V is an *α-power* (a power of exponent α) of a string U, denoted as $V = U^\alpha$, if $|V| = \alpha|U|$ and U is a period of V. Here, $\alpha \ge 1$ may otherwise be an arbitrary rational number. Powers of exponent $\alpha = 2$ are called *squares*, and powers of exponent $\alpha = 3$ are called *cubes*. By U^* we denote the set of all integer powers of U. A string V is called *non-primitive* if $V = U^k$ for some string U and an integer $k \ge 2$. Otherwise, V is called *primitive*. Primitive strings enjoy several useful properties; see [3,13].

Fact 2 (Synchronization Property). *If P is a primitive string, then it occurs exactly twice as a substring of P^2.*

Fact 3. *Let p be a period of a string X and P be any substring of X of length p. If p is the shortest period of X, then P is primitive. Conversely, if P is primitive and $p \le \frac{1}{2}|X|$, then p is the shortest period of X.*

Fact 4. *Let X be a string. Suppose that an integer p is a period of a prefix Y of X and of a suffix Z of X. If $|X| \le |Y| + |Z| - p$, then p is a period of X.*

2.2 Labeled Trees

Let T be a labeled tree. If u and v are two nodes of T, then by $val(u, v)$ we denote the sequence of labels of edges on the path from u to v. We call $val(u, v)$ a *substring* of T and (u, v) an *occurrence* of the string $val(u, v)$ in T. A *rooted tree* is a tree T with one of its nodes r designated as a root. For any two nodes u, v, by $lca(u, v)$ we denote their lowest common ancestor in T. A substring of a rooted tree is *anchored* at r if it corresponds to a path passing through r, i.e., if it has an occurrence (u, v) such that $lca(u, v) = r$. A *directed tree* T_r is a rooted tree with all its edges directed towards its root r. Every substring of a directed tree corresponds to a directed path in the tree. The following fact is a simple generalization of the upper bound of $2n$ on the number of squares in a string of length n; see [5,7].

Lemma 5. *A directed tree with n nodes contains at most $2n$ different square substrings.*

Proof. It suffices to note that there are at most two topmost occurrences of different squares starting at each node of the tree; see [5,7,10]. ◻

3 Cubes in Rooted Trees

In this section, we show that a rooted tree T with n nodes contains $\mathcal{O}(n)$ different cubes anchored at its root r.

3.1 Cube Decompositions

For a non-empty string X, (U, V) is a *cube decomposition* of X^3 if $UV = X^3$ and there exist nodes u and v in T such that $lca(u, v) = r$, $val(u, r) = U$ and $val(r, v) = V$. A cube decomposition is called *leftist* if $|U| \geq |V|$ and *rightist* if $|U| \leq |V|$. Due to the following lemma, it suffices to consider cubes with a leftist cube decomposition.

Lemma 6. *In a rooted tree the numbers of different cubes with a leftist decomposition and with a rightist decomposition are equal.*

Proof. (U, V) is a leftist cube decomposition of a cube X^3 if and only if (V^R, U^R) is a rightist cube decomposition of a cube Y^3 where $Y = X^R$. □

If $|U|, |V| < 2|X|$, then (U, V) is called a *balanced* cube decomposition. Otherwise, it is *unbalanced*. It turns out that the number of cubes with an unbalanced decomposition is simpler to bound.

Lemma 7. *A rooted tree with n nodes contains at most $2n$ different cubes with a leftist unbalanced cube decomposition.*

Proof. Let T be a tree rooted in r and let T_r be the corresponding directed tree. If (U, V) is an unbalanced leftist decomposition of a cube X^3, then $|U| \geq 2|X|$ and thus X^2 occurs as a square substring in T_r. By Lemma 5 there are at most $2n$ such different squares. □

A cube X^3 is called a *p-cube* if X is primitive. Otherwise it is called an *np-cube*. A bound on the number of np-cubes also follows from Lemma 5.

Lemma 8. *A rooted tree with n nodes contains at most $4n$ different np-cubes with a leftist cube decomposition.*

Proof. Let X^3 be an np-cube with a leftist decomposition (U, V) in a tree T rooted at r. We have $X = Y^k$ for a primitive string Y and an integer $k \geq 2$. Let $\ell = \left\lfloor \frac{3k}{4} \right\rfloor$. Note that $Y^{2\ell}$ is a proper prefix of U and thus a square in the directed tree T_r. Consider an assignment $Y^{3k} \mapsto Y^{2\ell}$. Observe that a single square can be assigned this way at most two cubes: $Y^{2\ell}$ can be assigned to $Y^{4\ell}, Y^{4\ell+1}, Y^{4\ell+2}$, or $Y^{4\ell+3}$, but no more than two of these exponents may be divisible by 3.

By Lemma 5 there are at most $2n$ different squares in the directed tree T_r. Therefore the number of different np-cubes with a leftist cube decomposition is bounded by $4n$. □

3.2 Essential Cube Decompositions

Thanks to Lemmas 6–8, from now on we only consider p-cubes in T which have a balanced leftist cube decomposition. We call such a decomposition an *essential* cube decomposition. In this section, we classify such decompositions into two types and provide a separate bound for either type.

Observation 9. *Let (U, V) be an essential cube decomposition of a p-cube X^3. Then $U = XB$ for a non-empty string B which is a border of U (and a prefix of X) and satisfies $\frac{1}{3}|U| \le |B| < \frac{1}{2}|U|$.*

Motivated by the observation, for a string U we define

$$\mathcal{B}(U) = \left\{B : B \text{ is a border of } U \text{ and } \tfrac{1}{3}|U| \le |B| < \tfrac{1}{2}|U|\right\}.$$

Moreover, by $\mathcal{B}'(U)$ we denote a set formed by the two longest strings in $\mathcal{B}(U)$ (we assume $\mathcal{B}'(U) = \mathcal{B}(U)$ if $|\mathcal{B}(U)| \le 2$).

Definition 10. Let (U, V) be an essential cube decomposition of X^3 and let $U = XB$. This decomposition is said to be of *type 1* if $B \in \mathcal{B}'(U)$ and of *type 2* otherwise.

Note that the string U and its border B uniquely determine the cube X^3. Since $|\mathcal{B}'(U)| \le 2$, the following observation follows directly from the definition above.

Observation 11 (Type-1 Reconstruction). *For every string U there are at most two strings V such that (U, V) is an essential decomposition of type 1 of some cube $X^3 = UV$.*

Below we prove a similar property of type-2 decompositions. Before that, we need to characterize them more carefully. The following lemma lists several properties of type-2 decompositions; see also Fig. 2.

Lemma 12. *Let (U, V) be a type-2 essential decomposition of a p-cube X^3. Then there exists a primitive string P such that:*

(a) $|P| \le \frac{1}{6}|X|$,
(b) *X has a prefix of the form P^* of length at least* $2|X| - |V| + |P|$,
(c) *X has P as a suffix, but does not have a suffix of the form P^* of length* $|V| - |X|$ *or more.*

Proof. Let $\mathcal{B}(U) = \{B_0, \ldots, B_\ell\}$ with $|B_0| < \ldots < |B_\ell|$. Since (U, V) is a type-2 decomposition of X, we have $U = XB_k$ for some k satisfying $0 \le k \le \ell - 2$. In particular, this implies $\ell \ge 2$.

By Fact 1, all borders in $\mathcal{B}(U)$ share a common shortest period, whose length in particular divides $|B_{i+1}| - |B_i|$ for any i $(0 \le i < \ell)$. We denote this period by P. By Fact 3, P is primitive. Let $p = |P|$ and let $p' = |B_0| \bmod p$. Moreover, let P' be the prefix of P of length p'. Observe that $B_0 = P^j P'$ for some integer j, and in general $B_i = P^{j+i} P'$.

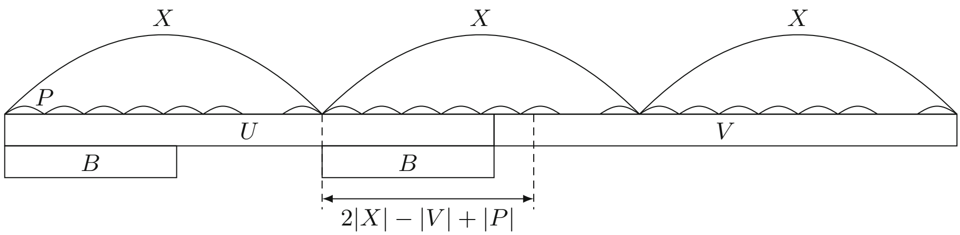

Fig. 2. Type 2 essential cube decomposition (U, V) of a cube X. Here, B is a border of U. Note that P is a period of B, but not a period of X or U.

(a) We have $\frac{1}{3}|U| \leq |B_0| < |B_\ell| < \frac{1}{2}|U|$ and $|B_\ell| - |B_0| = \ell \cdot p \geq 2p$. Thus, $p \leq \frac{1}{2}\left(\frac{1}{2} - \frac{1}{3}\right)|U| = \frac{1}{12}|U|$. Moreover, $|U| \leq 2|X|$, and as a consequence we get $|P| = p \leq \frac{2}{12}|X| = \frac{1}{6}|X|$.

(b, c) Note that $U = XB_k$ has B_ℓ as a suffix, and $B_\ell = P^{\ell-k}B_k$. Thus $P^{\ell-k}$ and, in particular, P is a suffix of X. Moreover, B_ℓ is a prefix of U, so U has $P^{j+\ell}$ as a prefix and, in particular, P is a prefix of X. Therefore, P is a border of X. Observe that P is not a period of X. Otherwise, due to synchronization property of primitive strings (Fact 2), X would be a power of P, which is a contradiction with X^3 being a p-cube.

Consequently, $|P^{j+\ell}| < |X|$, so $P^{j+\ell}$ is a prefix X. Moreover, we have $|P^{j+\ell}| \geq |B_{\ell-1}| \geq |B_k| + |P|$ since $k \leq \ell - 2$, and $|B_k| = |U| - |X| = 3|X| - |V| - |X| = 2|X| - |V|$. Thus, X indeed has a prefix Y of the form P^* whose length is at least $2|X| - |V| + |P|$. Now, suppose that X has a suffix Z of the form P^* whose length is at least $|V| - |X|$. We would have $|X| \leq |Y| + |Z| - |P|$, so Fact 4 would imply that P is a period of X, which we have already proved impossible. □

Lemma 13. (Type-2 Reconstruction). *For every string V there is at most one string U such that (U, V) is an essential cube decomposition of type 2 of some cube $X^3 = UV$.*

Proof. Suppose there is at least one string U which satisfies the assumption of the lemma. We shall prove that U can be uniquely determined from V. Let $UV = X^3$ and let P be the primitive string obtained through Lemma 12. Our goal is to recover P and then X from V.

Recall that $|X| < |V| \leq \frac{3}{2}|X|$ by the definition of essential cube decomposition. We have $X = V[i..|V|]$ for $i = |V| - |X| + 1$. Additionally, let $j = |X|$. Note that $j - i + 1 = 2|X| - |V|$, so Lemma 12(b) implies that $V[i..j'] = P^k$ for a position $j' \geq j$ and an integer exponent k. Observe that

$$i = |V| - |X| + 1 \leq \tfrac{1}{3}|V| + 1 \quad \text{and} \quad j = |X| \geq \tfrac{2}{3}|V|,$$

so $p = |P|$ is a period of $V' = V[\lfloor\frac{1}{3}|V|\rfloor + 1..\lceil\frac{2}{3}|V|\rceil]$. By Lemma 12(a), $|P| \leq \frac{1}{6}|X| \leq \frac{1}{6}|V| \leq \frac{1}{2}|V'|$ and P is primitive. Thus, by Fact 3, p can be uniquely determined as the shortest period of V'.

Once we know p, we can easily determine P: by Lemma 12(c), P is a suffix of X and thus a suffix of V. Hence, $P = V[|V| - p + 1..|V|]$.

Next, we determine the smallest position $i' > \frac{1}{3}|V|$ where P occurs in V. This occurrence must lie within $V[i..j]$, so $i \equiv i' \pmod{p}$ by the synchronization property of primitive strings (Fact 2). Let ℓ be the largest integer such that P^ℓ is a suffix of X. Then ℓ is simultaneously the largest integer such that P^ℓ is a suffix of V and the largest integer such that P^ℓ is a suffix of $V[1..i-1]$ (since $\ell p < |V| - |X|$ by Lemma 12(c)). The former lets us uniquely determine ℓ. The latter implies that $\ell' := \ell + \frac{i'-i}{p}$ is the largest integer such that $P^{\ell'}$ is a suffix of $V[1..i'-1]$. Since ℓ' is uniquely determined by V, so is i, and thus also $X = V[i..|V|]$. This concludes the proof that the string U can be uniquely determined from V. In particular, at most one such string exists. □

3.3 The Upper Bound

Theorem 14. *A rooted tree with n nodes contains $\mathcal{O}(n)$ cubes anchored at its root.*

Proof. Let T be a tree with n nodes rooted in r. The whole proof reduces to proofs of the following two claims.

Claim. There are $\mathcal{O}(n)$ different cubes in T having a non-essential cube decomposition.

Proof. A non-essential decomposition of a cube is rightist, leftist unbalanced or a leftist decomposition of an np-cube. In each case, by Lemmas 6–8, there are $\mathcal{O}(n)$ different cubes with such a decomposition. □

Claim. There are $\mathcal{O}(n)$ different p-cubes in T having an essential cube decomposition.

Proof. For each p-cube X^3 with an essential decomposition let us fix a single such decomposition UV and a single pair of nodes (u, v) that gives this decomposition.

If UV is a type-1 decomposition, we *charge* one token to the node u, otherwise we charge one token to v. By Observation 11 and Lemma 13, each node receives at most 3 tokens. □

This concludes the proof of the theorem. □

4 Powers in Trees

In this section we prove the announced bounds for powers_α for $\alpha > 1$.

Let S_m be a string $\mathtt{a}^m\mathtt{b}\mathtt{a}^m$. Note that S_m can be seen as a tree with a linear structure. Though the following fact can be treated as a folklore result, we provide its proof for completeness.

Theorem 15. *For every rational $\alpha \in (1, 2)$, we have $\mathsf{powers}_\alpha(S_m) = \Omega(|S_m|^2)$.*

Proof. Let $\alpha = 1 + \frac{x}{y}$ where x, y are coprime positive integers. For every positive integer $c \le \frac{m}{y}$, we construct $c(y-x)$ different powers of exponent α and length $cy\alpha$ that occur in S_m:

$$\mathtt{a}^i \mathtt{b} \mathtt{a}^{cy-1-i} \mathtt{a}^{cx} \quad \text{for } cx \le i < cy.$$

Note that $i < cy \le m$ and $cy - 1 - i + cx < cy \le m$, so they indeed occur as substrings of S_m. In total we obtain

$$\sum_{1 \le c \le \frac{m}{y}} c(y-x) = \Theta\left(\frac{m^2(y-x)}{y^2}\right) = \Theta(m^2)$$

different α-powers. Moreover, $|S_m| = \Theta(m)$, so this implies $\mathsf{powers}_\alpha(S_m) = \Omega(|S_m|^2)$. □

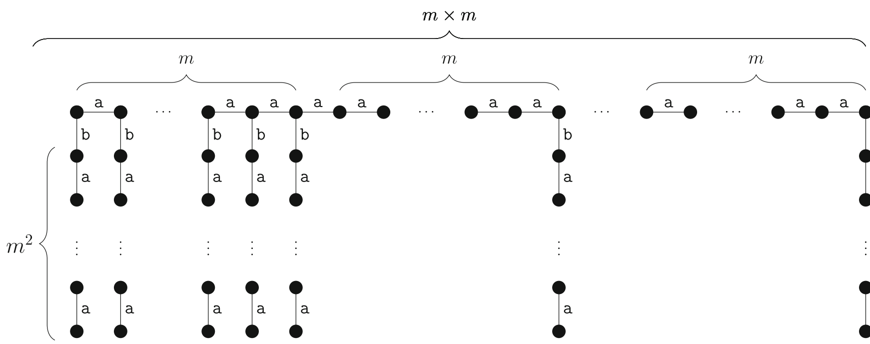

Fig. 3. Lower bound example T_m for powers of exponent $\alpha \in (2, 3)$.

Recall that for $\alpha = 2$, it has been shown that $\mathsf{powers}_2(n) = \Theta(n^{4/3})$ [4]. It turns out that the same bound applies for any $\alpha \in (2, 3)$. Moreover, the lower bound on $\mathsf{powers}_\alpha(n)$ is realized by the same family of trees called *combs*; see Fig. 3. A comb T_m consists of a path of length m^2 called the *spine*, with at most one *branch* attached to each node of the spine. Branches are located at positions $\{0, 1, 2, \ldots, m-1, m, 2m, 3m, \ldots, m^2\}$ of the spine. All edges of the spine are labeled with letters a. Each branch is a path starting with a letter b, followed by m^2 edges labeled with letters a.

Theorem 16. *For every rational* $\alpha \in (2, 3)$*, we have* $\mathsf{powers}_\alpha(T_m) = \Omega(|T_m|^{4/3})$.

Proof. Let $\alpha = 2 + \frac{x}{y}$ where x, y are coprime positive integers. For every positive integer $c \le \frac{m^2}{y}$, we construct $c(y-x)$ different α-powers of length $cy\alpha$ that occur in T_m:

$$(\mathtt{a}^i \mathtt{b} \mathtt{a}^{cy-1-i})^2 \mathtt{a}^{cx} \quad \text{for } cx \le i < cy.$$

Let us prove that these powers indeed occur in T_m. In [4] it was shown that for every $0 < j < m^2$ there are two branches whose starting nodes (on the spine) satisfy $distance(u, v) = j$. We apply this fact for $j = cy - 1$ and align letters $\mathtt{b}$ at the edges incident to u and v. Each branch contains m^2 edges labeled with $\mathtt{a}$. Since $i < cy \le m^2$ and $cy - 1 - i + cx < cy \le m^2$, this is enough to extend an occurrence of $\mathtt{ba}^{cy-1}\mathtt{b}$ to an occurrence of $(\mathtt{a}^i\mathtt{ba}^{cy-1-i})^2\mathtt{a}^{cx}$. Altogether this gives $\Theta(m^4)$ different α-powers. Since $|T_m| = \Theta(m^3)$, the number of the considered powers in T_m is $\Omega(|T_m|^{4/3})$. □

The upper bound for cubes and, consequently, for powers of rational exponent $\alpha \in (3, 4)$, is a consequence of the main result of the previous section.

Theorem 17. *For every rational $\alpha \ge 3$, we have $\mathsf{powers}_\alpha(n) = \mathcal{O}(n \log n)$.*

Proof. Recall that a *centroid* of a tree T is a node r such that each connected component of $T \setminus \{r\}$ is a tree with at most $\frac{n}{2}$ nodes. It is a well-known fact that every tree has a centroid.

We have already shown (Theorem 14) that the number of cubes in the tree T passing through a fixed node r is $\mathcal{O}(n)$. Now we need to count the remaining cubes in T. After removing the node r, the tree is partitioned into components $T_1, \ldots, T_k$. Hence, the number of cubes in T can be written as:

$$\mathsf{powers}_3(T) \le \mathcal{O}(|T|) + \sum_i \mathsf{powers}_3(T_i).$$

The components satisfy $\sum_i |T_i| = n - 1$ and $|T_i| \le \frac{n}{2}$, so a solution to this recurrence yields $\mathsf{powers}_3(n) = \mathcal{O}(n \log n)$. For every $\alpha \ge 3$, each power U^α of exponent α induces a cube U^3, so $\mathsf{powers}_\alpha(n) = \mathcal{O}(n \log n)$. □

The final result related to the powers function may be interpreted as a generalization of the $2n$ upper bound on the number of different squares in a string.

Theorem 18. *For every $\alpha \ge 4$, $\mathsf{powers}_\alpha(n) = \Theta(n)$.*

Proof. For a string $\mathtt{a}^n$, we have $\Theta(n/\alpha) = \Theta(n)$ distinct α-powers. For the proof of a linear upper bound, let T be a tree with n nodes and let r be any of its nodes. Let T_r be a directed tree obtained from T by selecting r as its root. Then any power U^α in T of exponent $\alpha \ge 4$ corresponds to square U^2 or $(U^R)^2$ in T_r. Thus, the conclusion follows from Lemma 5. □

5 Final Remarks

We have presented an almost complete asymptotic characterization of the function powers_α specifying the maximum number of different powers of exponent α in a tree of given size. What remains is an exact asymptotic bound for powers_α, $\alpha \in [3, 4)$, for which we have shown an $\mathcal{O}(n \log n)$ upper bound.

It can be shown (see Fact 19) that a tree with n nodes contains $\mathcal{O}(n)$ different cubes of the form $(\mathtt{a}^i\mathtt{ba}^j)^3$. In comparison, the lower bound constructions for $\alpha < 3$ rely on counting powers of the form $(\mathtt{a}^i\mathtt{ba}^j)^\alpha$.

Fact 19. *A tree with n nodes with edges labeled with* $\{\mathtt{a}, \mathtt{b}\}$ *contains* $\mathcal{O}(n)$ *cubes of the form* $(\mathtt{a}^i\mathtt{b}\mathtt{a}^j)^3$.

Proof. Let T be a tree with n nodes. Suppose that T is rooted at an arbitrary node r. Nevertheless, we bound the number of all cubes of the form $(\mathtt{a}^i\mathtt{b}\mathtt{a}^j)^3$ in T, including those which are not anchored at r. We shall assign each such cube to a single node of T so that each node of T is assigned at most two cubes. For a particular occurrence of a cube $X^3 = (\mathtt{a}^i\mathtt{b}\mathtt{a}^j)^3$ which starts in node u and ends in node v with $q = lca(u, v)$, we define the assignment as follows:

(A) if the string $val(u, q)$ contains at least two characters $\mathtt{b}$, then the cube is assigned to node u,
(B) otherwise (in that case $val(q, v)$ contains at least two characters $\mathtt{b}$) the cube is assigned to node v.

Let us prove that such procedure assigns at most one cube of type (A) and at most one cube of type (B) to a single node. If we fix the node and type of the assignment, we shall be able to uniquely recover the cube X^3 by going towards the root until we encounter the second edge labeled with $\mathtt{b}$. Indeed, suppose u is a fixed node and consider the assignment of type (A). Let X_1 be the shortest prefix of $val(u, r)$ that contains exactly one character $\mathtt{b}$ and let X_2 be the shortest prefix of $val(u, r)$ that contains exactly two characters $\mathtt{b}$. Then $X = \mathtt{a}^{|X_1|-1}\mathtt{b}\mathtt{a}^{|X_2|-|X_1|-1}$. For the assignment of type (B), we use a symmetric procedure. □

We conclude with the following conjectures.

Conjecture 20 (Weak conjecture). $\mathsf{powers}_\alpha(n) = \Theta(n)$ for every $\alpha > 3$.

Conjecture 21 (Strong conjecture). $\mathsf{powers}_3(n) = \Theta(n)$.

References

1. Amir, A., Lewenstein, M., Lewenstein, N.: Pattern matching in hypertext. J. Algorithms **35**(1), 82–99 (2000)
2. Brešar, B., Grytczuk, J., Klavžar, S., Niwczyk, S., Peterin, I.: Nonrepetitive colorings of trees. Discrete Math. **307**(2), 163–172 (2007)
3. Crochemore, M., Hancart, C., Lecroq, T.: Algorithms on Strings. Cambridge University Press, Cambridge (2007)
4. Crochemore, M., Iliopoulos, C.S., Kociumaka, T., Kubica, M., Radoszewski, J., Rytter, W., Tyczyński, W., Waleń, T.: The maximum number of squares in a tree. In: Kärkkäinen, J., Stoye, J. (eds.) CPM 2012. LNCS, vol. 7354, pp. 27–40. Springer, Heidelberg (2012)
5. Fraenkel, A.S., Simpson, J.: How many squares can a string contain? J. Comb. Theory Ser. A **82**(1), 112–120 (1998)
6. Grytczuk, J.: Thue type problems for graphs, points, and numbers. Discrete Math. **308**(19), 4419–4429 (2008)
7. Ilie, L.: A simple proof that a word of length n has at most $2n$ distinct squares. J. Comb. Theory Ser. A **112**(1), 163–164 (2005)

8. Ilie, L.: A note on the number of squares in a word. Theor. Comput. Sci. **380**(3), 373–376 (2007)
9. Kociumaka, T., Pachocki, J., Radoszewski, J., Rytter, W., Waleń, T.: Efficient counting of square substrings in a tree. Theor. Comput. Sci. **544**, 60–73 (2014)
10. Kociumaka, T., Radoszewski, J., Rytter, W., Waleń, T.: Maximum number of distinct and nonequivalent nonstandard squares in a word. In: Shur, A.M., Volkov, M.V. (eds.) DLT 2014. LNCS, vol. 8633, pp. 215–226. Springer, Heidelberg (2014)
11. Kociumaka, T., Radoszewski, J., Rytter, W., Waleń, T.: Internal pattern matching queries in a text and applications. In: Indyk, P. (ed.) 26th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, pp. 532–551. SIAM (2015)
12. Kubica, M., Radoszewski, J., Rytter, W., Waleń, T.: On the maximum number of cubic subwords in a word. Eur. J. Comb. **34**(1), 27–37 (2013)
13. Lothaire, M.: Combinatorics on Words. Cambridge Mathematical Library, second edn. Cambridge University Press, New York (1997)
14. Thue, A.: Über unendliche Zeichenreihen. Norske Videnskabers Selskabs Skrifter Mat.-Nat. Kl. **7**, 1–22 (1906)