ENUMERATION OF HIGHLY BALANCED TREES

STEPHAN WAGNER

ABSTRACT. Bereg and Wang defined a new class of highly balanced d-ary trees which they call k-trees; these trees have the interesting property that the internal path length and thus the Wiener index can be calculated quite easily. A k-tree is characterized by the property that all levels, except for the last k levels, are completely filled. Bereg and Wang claim that the number of k-trees is exponentially increasing, but do not give an asymptotic formula for it. In this paper, we study the number of d-ary k-trees and the number of mutually non-isomorphic d-ary k-trees, making use of a technique due to Flajolet and Odlyzko.

1. INTRODUCTION AND PRELIMINARIES

Bereg and Wang [2] study a class of highly balanced trees which they call k-trees:

Definition 1. *A rooted binary tree is called a k-tree if every node of depth less than $h - k$ has exactly two children, where h is the depth of the tree. In other words, there are exactly 2^j nodes of depth j for $0 \leq j \leq h - k$.*

This tree family has interesting properties with respect to the Wiener index (i.e. the sum of all distances between pairs of vertices), which is the reason why it was studied in [2]. Obviously, 0-trees are necessarily complete binary trees, which implies that the number of vertices is $2^{h-1} - 1$. On the other hand, if $k \geq 1$, Bereg and Wang claim that the number of k-trees increases exponentially without further specifying the growth rate. In this paper, we will be concerned with the enumeration of k-trees, where the definition is further extended in the obvious way to d-ary trees.

We note that many important classes of trees satisfy balance conditions of a similar type. For instance, Kemp [6] studies trees with the property that all root-to-leaf paths have the same length and obtains exact and asymptotic enumeration results for this class of trees. Another, very well-known example are 2-3-trees (see for instance [7, 9]), which are indeed completely balanced. They play an important role in computer science, and so do the so-called AVL-trees, i.e. binary trees with the property that the heights of the two subtrees stemming from each vertex differ by at most 1 [7, 8]. Indeed, we will see that the enumeration of 2-3-trees is quite similar to the enumeration of k-trees, even though there are differences due to the fact that 2-3-trees allow 2 and 3 as potential outdegrees of a node. This becomes particularly clear if one considers 1-trees: note that there is exactly one 1-tree of size $2^k - 1$ for each positive integer k (namely the complete binary tree). Thus, even though the number of 1-trees of size $\leq n$ grows exponentially, this is not true for 1-trees of size equal n. This phenomenon does not occur for 2-3-trees.

Let us first discuss the trivial case, where we consider k-trees as ordered trees, and isomorphisms are not taken into account. In this case, if $p_r(z)$ denotes the generating function for d-ary k-trees with the property that the $(r-k)$-th level is full (i.e. it has d^{r-k} vertices of depth $r-k$) and the height is at most r, we have

$$p_r(z) = zp_{r-1}(z)^d$$

for $r \geq k$ (simply note that all d subtrees attached to the root have to be k-trees again), while the initial value p_{k-1} is just the generating function for d-ary trees of height $\leq k-1$ (including the empty tree!). The latter is given by $p_{k-1}(z) = q_{k-1}(z)$, where $q_0(z) = 1 + z$ and

$$q_h(z) = 1 + zq_{h-1}(z)^d$$

for $1 \leq h \leq k-1$. Now a simple induction shows that

$$p_r(z) = z^{\frac{d^{r-k+1}-1}{d-1}} p_{k-1}(z)^{d^{r-k+1}},$$

and there are strong results available for the coefficients of powers of a given polynomial (equivalently, one can regard it is the iterated convolution of a discrete distribution), see for instance [10]. In the particularly simple case $k = 1$, one obtains

$$p_r(z) = z^{\frac{d^r-1}{d-1}}(1+z)^{d^r},$$

and one has the following result:

Proposition 1. *The number of d-ary* 1*-trees of size n is exactly*

$$\binom{d^r}{n - \frac{d^r-1}{d-1}},$$

where r is chosen in such a way that $d^r - 1 \leq (d-1)n \leq d^{r+1} - 1$.

A much more interesting problem arises if one is interested in the number of *non-isomorphic* k-trees. In this case, the above relations have to be modified accordingly, so that we have

$$p_r(z) = zZ(S_d, p_{r-1}(z))$$

and

$$q_h(z) = 1 + zZ(S_d, q_{h-1}(z)),$$

where $Z(S_d)$ denotes the cycle index of the symmetric group, see [4]. Explicitly, this can be written as

$$p_r(z) = z \sum_{j_1+2j_2+\ldots+dj_d=d} \prod_{\ell=1}^{d} \frac{p_{r-1}(z^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!}$$

and analogously for q_h. These recursions are not polynomial any longer, which complicates the asymptotic analysis.

For polynomial iterations of the type $p_r(z) = P(z, p_{r-1})$, Flajolet and Odlyzko [3] provide a very powerful result for the asymptotics of the coefficients of p_r under some technical assumptions. This cannot be applied directly to our situation, but the techniques used in the aforementioned paper [3] can be used again, and wide parts are even completely identical. The main difficulty lies in an estimate for the values of our polynomials at the powers of a complex number z. We will obtain the following main theorem:

Theorem 2. *The number of non-isomorphic k-trees with n vertices and the property that the $(r-k)$-th level is full and the height is at most r is given by*

$$[z^n]p_r(z) = d!^{1/(d-1)} \rho^{-m} d^{-r/2} \left(2\pi \left(\rho^2 \psi''(\rho) + \rho\psi'(\rho)\right)\right)^{-1/2} \exp(d^r \psi(\rho))(1 + O(r^3 d^{-r/2}))$$

uniformly in n if $\lambda_1 d^r \le m = n - \frac{d^{r-k+1}-1}{d-1} \le \lambda_2 d^r$ for fixed real numbers with

$$0 < \lambda_1 < \lambda_2 < \frac{d^k - 1}{d^k - d^{k-1}},$$

where

- *ρ is chosen in such a way that $\rho\psi'(\rho) = \frac{m}{d^r}$,*
- *ψ is an analytic function on the set of positive reals.*

From this we can draw the following corollary:

Corollary 3. *The coefficients of $p_r(z)$ asymptotically follow a normal distribution with mean*

$$\mu_r \sim \left(\psi'(1) + \frac{d^{1-k}}{d-1}\right) d^r$$

and variance

$$\sigma_r^2 \sim (\psi'(1) + \psi''(1))\, d^r.$$

In the following section, we provide some auxiliary results that are needed for the proof of the main theorem, which is given in Section 3. The special case $d = 2$, $k = 1$, that is of particular interest in [2] as well, is treated in Section 4, followed by some concluding remarks.

2. Auxiliary results

In the following, it will be advantageous to set

$$y_r(z) = z^{-\frac{d^{r-k+1}-1}{d-1}} p_r(z)$$

for $r \geq k-1$. Note that since

$$Z(S_d, z^a f(z)) = z^{ad} Z(S_d, f(z)),$$

this implies

$$(1) \qquad y_r(z) = Z(S_d, y_{r-1}(z)) = \sum_{j_1+2j_2+\ldots+dj_d=d} \prod_{\ell=1}^{d} \frac{y_{r-1}(z^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!}.$$

Clearly, $[z^n]y_r(z)$ is now the number of non-isomorphic d-ary k-trees with $n + \frac{d^{r-k+1}-1}{d-1}$ vertices, whose $(r-k)$-th level is full and whose height is at most r. In particular, $[z^0]y_r(z) = [z^1]y_r(z) = 1$: note that the first $r-k$ levels contain $\frac{d^{r-k+1}-1}{d-1}$ vertices, and there is only one way (up to isomorphism) to add no vertex or one more vertex. Hence, the polynomial $y_r(z)$ starts with

$$y_r(z) = 1 + z + \cdots.$$

It is not difficult to show that the coefficient of z^2 already tends to ∞ as r does. Indeed, one has

$$[z^2]y_r(z) = \begin{cases} r & k = 1, \\ r-k+2 & k > 1, \end{cases}$$

since a tree with the prescribed properties and $\frac{d^{r-k+1}-1}{d-1} + 2$ vertices is uniquely determined by the level of the first common ancestor of the two vertices at level $r-k+1$ (or possibly $r-k+2$, if $k > 1$).

As a simple consequence, if follows immediately that

$$\lim_{r\to\infty} y_r(z) = \infty$$

for $z > 0$. This helps us to prove the following simple lemma:

Lemma 4. *For any $\delta > 0$, there exist positive constants $A_1 = A_1(\delta)$, $B_1 = B_1(\delta)$ depending only on δ such that*

$$y_r(z) \geq A_1 \exp(B_1 d^r)$$

whenever $z \geq \delta$.

Proof. Since y_r is increasing (its coefficients are non-negative), it obviously suffices to consider $z = \delta$. Because of the fact that $y_r(\delta)$ tends to ∞, we can choose $r_0 \geq k$ large enough to have

$$y_{r_0}(\delta) > d!^{1/(d-1)}.$$

Now, for $r > r_0$, we obtain

$$y_r(\delta) = Z(S_d, y_{r-1}(\delta)) \geq \frac{1}{d!} y_{r-1}(\delta)^d$$

and therefore

$$\frac{y_r(\delta)}{d!^{1/(d-1)}} \geq \frac{1}{d!^{d/(d-1)}} y_{r-1}(\delta)^d = \left(\frac{y_{r-1}(\delta)}{d!^{1/(d-1)}} \right)^d .$$

Induction now shows that

$$y_r(\delta) \geq d!^{1/(d-1)} \left(\frac{y_{r_0}(\delta)}{d!^{1/(d-1)}} \right)^{d^{r-r_0}} ,$$

proving the lemma (for $r < r_0$, the inequality is trivial if A_1, B_1 are sufficiently small). ∎

In order to obtain more precise asymptotic information, we need to show that $\frac{1}{d!} y_{r-1}(z)^d$ is indeed the dominant term in $Z(S_d, y_{r-1}(z))$. This motivates the following lemma:

Lemma 5. *For any $\delta > 0$, there exist positive constants $A_2 = A_2(\delta)$, $B_2 = B_2(\delta)$ depending only on δ such that*

$$\frac{y_r(z^\ell)}{y_r(z)^\ell} \leq A_2 \exp(-B_2 d^r)$$

whenever $2 \leq \ell \leq d$ and $\delta \leq z \leq \delta^{-1}$.

Proof. If $z \leq 1$, simply note that

$$\frac{y_r(z^\ell)}{y_r(z)^\ell} \leq y_r(z)^{1-\ell}$$

since y_r is increasing, and apply Lemma 4. If $z \geq 1$, we work with the "mirrored" function

$$\tilde{y}_r(z) = z^{\frac{d^{r+1}-d^{r-k+1}}{d-1}} y_r\left(\frac{1}{z}\right)$$

(note that $\frac{d^{r+1}-d^{r-k+1}}{d-1}$ is the degree of y_r), which satisfies the same recursion. Therefore,

$$\frac{y_r(z^\ell)}{y_r(z)^\ell} = \frac{\tilde{y}_r(1/z^\ell)}{\tilde{y}_r(1/z)^\ell}$$

tends to 0 at a doubly exponential rate by the same arguments. This completes the proof of the lemma. ∎

Now we extend this to a strip in the complex plane:

Lemma 6. *For any $\delta > 0$, there exist $\eta > 0$ and positive constants $A_3 = A_3(\delta)$, $B_3 = B_3(\delta)$ depending only on δ such that*

$$\left|\frac{y_r(z^\ell)}{y_r(z)^\ell}\right| \le A_3 \exp(-B_3 d^r)$$

for any $2 \le \ell \le d$ and any complex z in the region

$$R(\delta) = \left\{z \in \mathbb{C} : |\operatorname{Arg} z| \le \eta,\ \delta \le \operatorname{Re} z \le \delta^{-1}\right\}.$$

Proof. Set $\alpha = \frac{A_2}{4}$, $\beta = \frac{2B_2}{3}$, $\gamma = 2$ and $\delta = \frac{B_2}{3d}$, with A_2 and B_2 as in the previous lemma. Furthermore, choose r_0 sufficiently large and η sufficiently small so that

$$\left|\frac{y_r(z^\ell)}{y_r(z)^\ell}\right| \le \alpha \exp(-\beta d^r) \tag{2}$$

holds for $r = r_0$ and all $z \in R(\delta)$ and $2 \le \ell \le d$ (which is possible by Lemma 5 since $\beta < B_2$) as well as

$$\left|\frac{y_r(z)}{y_r(|z|)}\right| \ge \gamma \exp(-\delta d^r) \tag{3}$$

(again for $r = r_0$ and all $z \in R(\delta)$; note that the right-hand side is < 1 for sufficiently large r, so that validity of the inequality can be ensured by choosing η sufficiently small). Furthermore, let r_0 be large enough so that

$$\frac{\gamma^d \exp(-\delta d^{r+1}) - \alpha d! \exp(-\beta d^r)}{1 + \alpha d! \exp(-\beta d^r)} \ge \gamma \exp(-\delta d^{r+1}) \tag{4}$$

for all $r \ge r_0$. Since $\gamma > 1$ and $\beta > \delta d$, this is also possible.

We will prove by induction that the inequalities (2) and (3) remain true for $r > r_0$. To this end, we apply the induction hypothesis together with the recursion (1) to obtain

$$
\begin{aligned}
|y_{r+1}(z)| &= \left| \sum_{j_1+2j_2+\ldots+dj_d=d} \prod_{\ell=1}^{d} \frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!} \right| \\
&= \left| \frac{1}{d!} y_r(z)^d + \sum_{\substack{j_1+2j_2+\ldots+dj_d=d \\ j_1<d}} \prod_{\ell=1}^{d} \frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!} \right| \\
&\geq \left| \frac{1}{d!} y_r(z)^d \right| - \sum_{\substack{j_1+2j_2+\ldots+dj_d=d \\ j_1<d}} \prod_{\ell=1}^{d} \frac{|y_r(z^\ell)|^{j_\ell}}{\ell^{j_\ell} j_\ell!} \\
&\geq \frac{1}{d!} |y_r(z)|^d - \sum_{\substack{j_1+2j_2+\ldots+dj_d=d \\ j_1<d}} \prod_{\ell=1}^{d} \frac{y_r(|z|^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!} \\
&\geq \frac{1}{d!} y_r(|z|)^d \left(\gamma \exp(-\delta d^r)\right)^d - \frac{d!-1}{d!} \alpha \exp(-\beta d^r) y_r(|z|)^d \\
&\geq \left(\frac{\gamma^d}{d!} \exp(-\delta d^{r+1}) - \alpha \exp(-\beta d^r) \right) y_r(|z|)^d.
\end{aligned}
$$

and analogously

$$
y_{r+1}(|z|) \leq \left(\frac{1}{d!} + \alpha \exp(-\beta d^r) \right) y_r(|z|)^d.
$$

Together with (4), this implies

$$
\left| \frac{y_{r+1}(z)}{y_{r+1}(|z|)} \right| \geq \gamma \exp(-\delta d^{r+1}),
$$

as required. Furthermore, we have

$$
\begin{aligned}
\left| \frac{y_{r+1}(z^\ell)}{y_{r+1}(z)^\ell} \right| &\leq \frac{y_{r+1}(|z|^\ell)}{|y_{r+1}(z)^\ell|} \leq \frac{y_{r+1}(|z|^\ell)}{y_{r+1}(|z|)^\ell} \cdot \frac{1}{\gamma^\ell} \exp(\delta \ell d^{r+1}) \\
&\leq \frac{A_2}{\gamma^\ell} \exp(-B_2 d^{r+1}) \exp(\delta \ell d^{r+1}) \\
&\leq \frac{A_2}{\gamma^2} \exp((\delta d - B_2) d^{r+1}) = \alpha \exp(-\beta d^{r+1})
\end{aligned}
$$

for $2 \leq \ell \leq d$ by Lemma 5, which completes the induction. Therefore, if we choose $A_3 \geq \alpha$ and $B_3 \leq \beta$ in such a way that the inequality holds for $r < r_0$ as well, the lemma follows. ∎

This allows us to determine a uniform estimate for $y_r(z)$ within the region $R(\delta)$:

Proposition 7. *For any $\delta > 0$ there is a positive constant $C = C(\delta)$ depending only on δ such that*

$$y_r(z) = \exp\left(\psi(z)d^r + \frac{\log d!}{d-1}\right)\left(1 + O\left(\exp(-Cd^r)\right)\right)$$

for all $z \in R(\delta)$, where

$$\psi(z) = d^{1-k}\log y_{k-1}(z) - \frac{d^{1-k}}{d-1}\log d! + \sum_{j=k-1}^{\infty} d^{-j-1}\log\left(\frac{d!y_{j+1}(z)}{y_j(z)^d}\right)$$

is analytic in $R(\delta)$.

Proof. We use the classical approach for sequences with a doubly exponential growth, compare [1] and [3]. Taking logarithms in (1) yields

$$\begin{aligned}\log y_{r+1}(z) &= \log \sum_{j_1+2j_2+\ldots+dj_d=d}\prod_{\ell=1}^{d}\frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell}j_\ell!} \\ &= \log\left(\frac{1}{d!}y_r(z)^d + \sum_{\substack{j_1+2j_2+\ldots+dj_d=d\\ j_1<d}}\prod_{\ell=1}^{d}\frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell}j_\ell!}\right) \\ &= d\log y_r(z) - \log d! + \log\left(1 + d!\sum_{\substack{j_1+2j_2+\ldots+dj_d=d\\ j_1<d}}\prod_{\ell=1}^{d}\frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell}j_\ell!y_r(z)^{\ell j_\ell}}\right).\end{aligned}$$

Here, the principal value of the logarithm is used. Note that $y_r(z) \neq 0$ for $z \in R(d)$ by the previous lemma. We write $v_r(z)$ as an abbreviation for $\log y_r(z)$ and $q_r(z)$ for the last summand in the above equation. Then,

$$v_{r+1}(z) = dv_r(z) - \log d! + q_r(z)$$

for $r \geq k-1$. Iterating this equation yields

$$v_r(z) = d^{r-k+1} v_{k-1}(z) - \frac{d^{r-k+1}-1}{d-1} \log d! + \sum_{j=k-1}^{r-1} d^{r-j-1} q_j(z)$$
$$= \frac{\log d!}{d-1} + d^r \left(d^{1-k} v_{k-1}(z) - \frac{d^{1-k}}{d-1} \log d! + \sum_{j=k-1}^{r-1} d^{-j-1} q_j(z) \right)$$
$$= \frac{\log d!}{d-1} + d^r \left(d^{1-k} v_{k-1}(z) - \frac{d^{1-k}}{d-1} \log d! + \sum_{j=k-1}^{\infty} d^{-j-1} q_j(z) \right.$$
$$\left. - \sum_{j=r}^{\infty} d^{-j-1} q_j(z) \right)$$
$$= \frac{\log d!}{d-1} + d^r \psi(z) - \sum_{j=r}^{\infty} d^{r-j-1} q_j(z).$$

Note that the infinite sum converges uniformly since $q_j(z)$ is bounded as a consequence of Lemma 6, thus implying analyticity of $\psi(z)$. The same lemma shows that the remainder term satisfies

$$\sum_{j=r}^{\infty} d^{r-j-1} q_j(z) = O\left(\exp(-B_3 d^r)\right),$$

since $q_r(z) = O\left(\exp(-B_3 d^r)\right)$. This concludes the proof. ∎

Our next lemma concerns the behavior outside the region $R(\delta)$:

Lemma 8. *For any $\delta > 0$, there exist positive constants $A_4 = A_4(\delta)$, $B_4 = B_4(\delta)$ depending only on δ such that*

$$y_r(z) \leq A_4 \exp(-B_4 d^r) y_r(|z|)$$

whenever $\delta \leq |z| \leq \delta^{-1}$ and $z \notin R(\delta)$.

Proof. Since $y_r(z) = 1 + z + \ldots$, the triangle inequality implies $|y_r(z)| \leq y_r(|z|)$ with equality only for $z = |z|$. Hence, for any fixed r we can choose $\alpha < 1$ in such a way that

$$|y_r(z)| \leq \alpha y_r(|z|)$$

whenever $z \notin R(\delta)$. Now, note that

$$y_{r+1}(z) \leq \frac{1}{d!} |y_r(z)|^d + \sum_{\substack{j_1+2j_2+\ldots+dj_d=d \\ j_1<d}} \prod_{\ell=1}^{d} \frac{y_r(|z|^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!}$$
$$\leq \frac{\alpha^d}{d!} y_r(|z|)^d + O\left(y_r(|z|)^d \exp(-B_3 d^r)\right)$$

in this case as well as

$$y_{r+1}(|z|) = \frac{1}{d!} y_r(|z|)^d + O\left(y_r(|z|)^d \exp(-B_3 d^r)\right).$$

If K is the implied constant in the error terms and $A_4 < 1$ and B_4 are chosen appropriately to ensure that

$$|y_r(z)| \le A_4 \exp(-B_4 d^r) y_r(|z|)$$

for $r = r_0$ and

$$\frac{A_4^d}{d!} \exp(-B_4 d^{r+1}) + K \exp(-B_3 d^r) \le A_4 \exp(-B_4 d^{r+1}) \left(\frac{1}{d!} + K \exp(-B_3 d^r)\right)$$

for $r \ge r_0$ (which is the case if r_0 is large enough and B_4 small enough, since we assume $A_4 < 1$), the desired inequality follows inductively for $r \ge r_0$. Again, it holds trivially for $r \le r_0$ if the parameters are chosen appropriately. ■

Before we can go on to prove our main theorem, we still need some more information on the function ψ that is summarized in the following lemma:

Lemma 9. *The function ψ satisfies*

$$(z\psi'(z))' > 0$$

for all $z > 0$, and one has

$$\lim_{z\to 0} z\psi'(z) = 0 \quad \text{and} \quad \lim_{z\to\infty} z\psi'(z) = \frac{d^k - 1}{d^k - d^{k-1}}.$$

Proof. Note first that we have

$$\psi(z) = \lim_{r\to\infty} d^{-r} \log y_r(z),$$

and since the convergence is uniform on compact subsets of $(0, \infty)$, it follows that

$$z\psi'(z) = \lim_{r\to\infty} d^{-r} \frac{z y_r'(z)}{y_r(z)}.$$

and

$$(z\psi'(z))' = \lim_{r\to\infty} d^{-r} \left(\frac{z y_r'(z)}{y_r(z)}\right)'.$$

Now, note that for a polynomial $a(z) = \sum_n a_n z^n$ with positive coefficients,

$$z\left(\frac{za'(z)}{a(z)}\right)' = \frac{z^2 a''(z)}{a(z)} + \frac{za'(z)}{a(z)} - \left(\frac{za'(z)}{a(z)}\right)^2 = \frac{\sum_n n^2 a_n z^n}{\sum_n a_n z^n} - \left(\frac{\sum_n n a_n z^n}{\sum_n a_n z^n}\right)^2$$

is the variance of an associated random variable A with $\mathbb{P}(A = n) = \frac{a_n z^n}{a(z)}$ and is thus always non-negative. Furthermore, if $a(z)$ is not a monomial, then the variance is strictly positive. Furthermore, note that if $a(z) = a_1(z) + a_2(z)$ and A, A_1, A_2 are the associated random variables, then A is a mixture of A_1 and A_2 (with weights $\lambda = \frac{a_1(z)}{a(z)}$ and $1 - \lambda = \frac{a_2(z)}{a(z)}$, respectively, i.e. A_1 is chosen with probability λ, and A_2 with probability $1 - \lambda$). It follows that

$$\begin{aligned}\mathbb{V}(A) &= \mathbb{E}(A^2) - \mathbb{E}(A)^2 \\ &= \lambda\,\mathbb{E}(A_1^2) + (1-\lambda)\,\mathbb{E}(A_2^2) - (\lambda\,\mathbb{E}(A_1) + (1-\lambda)\,\mathbb{E}(A_2))^2 \\ &= \lambda\,\mathbb{V}(A_1) + (1-\lambda)\,\mathbb{V}(A_2) + \lambda(1-\lambda)\,(\mathbb{E}(A_1) - \mathbb{E}(A_2))^2 \geq \lambda\,\mathbb{V}(A_1)\end{aligned}$$

and thus

$$z\left(\frac{za'(z)}{a(z)}\right)' \geq \frac{a_1(z)}{a(z)} \cdot z\left(\frac{za_1'(z)}{a_1(z)}\right)'.$$

We apply this to the recurrence

$$y_{r+1}(z) = \frac{y_r(z)^d}{d!} + \dots$$

to obtain

$$z\left(\frac{zy_{r+1}'(z)}{y_{r+1}(z)}\right)' \geq d \cdot \frac{d!y_{r+1}(z)}{y_r(z)^d} \cdot z\left(\frac{zy_r'(z)}{y_r(z)}\right)'.$$

Iteration yields

$$z\left(\frac{zy_r'(z)}{y_r(z)}\right)' \geq d^{r-k+1} \cdot z\left(\frac{zy_{k-1}'(z)}{y_{k-1}(z)}\right)' \prod_{j=k}^{r}\left(\frac{d!y_j(z)}{y_{j-1}(z)^d}\right). \tag{5}$$

By Proposition 7,

$$\frac{d!y_j(z)}{y_{j-1}(z)^d} = 1 + O\left(\exp(-Cd^{j-1})\right),$$

which implies that the product

$$\prod_{j=k}^{r}\left(\frac{d!y_j(z)}{y_{j-1}(z)^d}\right)$$

converges to a positive number. Since $y_{k-1}(z)$ is certainly not a monomial by its definition,

$$z\left(\frac{zy_{k-1}'(z)}{y_{k-1}(z)}\right)' > 0,$$

and so (5) shows that

$$(z\psi'(z))' = \lim_{r\to\infty} d^{-r}\left(\frac{zy_r'(z)}{y_r(z)}\right)' > 0.$$

For the remaining part of the lemma, note first that

$$y'_{r+1}(z) = \sum_{j_1+2j_2+\ldots+dj_d=d} \frac{d}{dz} \prod_{\ell=1}^{d} \frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!}$$

$$= \sum_{j_1+2j_2+\ldots+dj_d=d} \left(\sum_{\ell=1}^{d} \frac{\ell j_\ell z^{\ell-1} y'_r(z^\ell)}{y_r(z^\ell)} \right) \prod_{\ell=1}^{d} \frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!}$$

and thus

$$\frac{zy'_{r+1}(z)}{y_{r+1}(z)} = \frac{\sum_{j_1+2j_2+\ldots+dj_d=d} \left(\sum_{\ell=1}^{d} \frac{\ell j_\ell z^\ell y'_r(z^\ell)}{y_r(z^\ell)} \right) \prod_{\ell=1}^{d} \frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!}}{\sum_{j_1+2j_2+\ldots+dj_d=d} \prod_{\ell=1}^{d} \frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!}}.$$

We already know that $\left(\frac{zy'_r(z)}{y_r(z)}\right)' > 0$, i.e. this quotient is increasing. Therefore, for $z \leq 1$, we have the following upper estimate:

$$\frac{zy'_{r+1}(z)}{y_{r+1}(z)} \leq \frac{\sum_{j_1+2j_2+\ldots+dj_d=d} \left(\frac{zy'_r(z)}{y_r(z)} \sum_{\ell=1}^{d} \ell j_\ell \right) \prod_{\ell=1}^{d} \frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!}}{\sum_{j_1+2j_2+\ldots+dj_d=d} \prod_{\ell=1}^{d} \frac{y_r(z^\ell)^{j_\ell}}{\ell^{j_\ell} j_\ell!}} = d \cdot \frac{zy'_r(z)}{y_r(z)}.$$

Hence, $d^{-r} \cdot \frac{zy'_r(z)}{y_r(z)}$ decreases for $r \to \infty$. Noting that $y_r(0) = 1$ for all r, we have

$$\lim_{z \to 0} \frac{zy'_r(z)}{y_r(z)} = 0$$

for all r, which implies, together with the monotonicity, that

$$\lim_{z \to 0} \lim_{r \to \infty} \frac{zy'_r(z)}{y_r(z)} = 0,$$

as required. To prove the second limit formula, we can use the same trick as in the proof of Lemma 5: if

$$\tilde{y}_r(z) = z^{\frac{d^{r+1}-d^{r-k+1}}{d-1}} y_r\left(\frac{1}{z}\right)$$

is the mirrored function again, then it is easy to check that

$$\frac{zy'_r(z)}{y_r(z)} = \frac{d^{r+1} - d^{r-k+1}}{d-1} - \frac{1/z\tilde{y}'_r(1/z)}{\tilde{y}_r(1/z)}.$$

Dividing by d^r and passing to the limit, we obtain

$$\lim_{z \to \infty} z\psi'(z) = \frac{d^k - 1}{d^k - d^{k-1}} - \lim_{z \to 0} \lim_{r \to \infty} \frac{z\tilde{y}'_r(z)}{\tilde{y}_r(z)},$$

and the limit on the right hand side can be treated in the same way as the analogous limit involving y_r. This completes the proof. ■

3. Proof of the main theorem

Proposition 7, together with Hwang's Quasi-power Theorem [5], shows that the coefficients of $p_r(z)$ follow a normal distribution. In order to obtain more detailed information, we apply the saddle point method. The rest of the proof can be literally taken from [3]. We start with the integral representation

$$[z^n]p_r(z) = \left[z^{n-\frac{d^{r-k+1}-1}{d-1}}\right] y_r(z) = \frac{1}{2\pi i}\int_\Gamma y_r(z) z^{-\left(n-\frac{d^{r-k+1}-1}{d-1}\right)-1}\,dz$$

for any simple closed curve Γ around 0. Write $m = n - \frac{d^{r-k+1}-1}{d-1}$, and set $\lambda = \frac{m}{d^r}$. Then, $\lambda_1 \le \lambda \le \lambda_2$ by our condition on n. Hence, in view of Lemma 9, there is a unique ρ such that

$$\rho\psi'(\rho) = \lambda,$$

and $\rho_1 \le \rho \le \rho_2$, where ρ_1 and ρ_2 are such that $\rho_1\psi'(\rho_1) = \lambda_1$ and $\rho_2\psi'(\rho_2) = \lambda_2$. Now, we choose Γ to be the circle of radius ρ around the origin. By Proposition 7 there is a constant $\theta_0 > 0$ such that ψ is analytic in the region determined by

$$\rho_1 \le |z| \le \rho_2 \quad \text{and} \quad |\operatorname{Arg} z| \le \theta_0.$$

Within that region, we have the expansion

$$\operatorname{Re}\psi(\rho e^{i\theta}) = \psi(\rho) - \frac{\theta^2}{2}\left(\rho^2\psi''(\rho) + \rho\psi'(\rho)\right) + O(\theta^4), \tag{6}$$

and by taking θ_0 small enough, we can ensure that

$$\operatorname{Re}\psi(\rho e^{i\theta}) \le \psi(\rho) - \frac{\theta^2}{4}\left(\rho^2\psi''(\rho) + \rho\psi'(\rho)\right). \tag{7}$$

If Γ_1 denotes the part of the circle Γ with $|\operatorname{Arg} z| \ge \theta_0$, then by Proposition 7 and Lemma 8, we have

$$\frac{1}{2\pi i}\int_{\Gamma_1} y_r(z) z^{-m-1}\,dz = O\left(\rho^{-m}\exp\left(d^r(\psi(\rho) - C_1)\right)\right).$$

Furthermore, if $\Gamma_2 = \Gamma \setminus \Gamma_1$, then

$$\frac{1}{2\pi i}\int_{\Gamma_2} y_r(z) z^{-m-1}\,dz = \frac{1}{2\pi i}\int_{\Gamma_2} z^{-m-1}\exp\left(\psi(z)d^r + \frac{\log d!}{d-1}\right)dz + O\left(\rho^{-m}\exp\left(d^r(\psi(\rho) - C_2)\right)\right)$$

by Proposition 7. Here, the positive constants C_1 and C_2 only depend on ρ_1, ρ_2 and θ_0. It remains to determine the integral on the right hand side. To this end, we split Γ_2 further into

$$\Gamma_3 = \left\{z : |z| = \rho, |\operatorname{Arg} z| \le \theta_1 = rd^{-r/2}\right\}$$

and $\Gamma_4 = \Gamma_2 \setminus \Gamma_3$. Then, by (7),

$$\operatorname{Re}\psi(\rho e^{i\theta}) \le \psi(\rho) - C_3 r^2 d^{-r}$$

holds on Γ_4 for some positive constant C_3, and thus

$$\frac{1}{2\pi i}\int_{\Gamma_4} z^{-m-1}\exp\left(\psi(z)d^r + \frac{\log d!}{d-1}\right) dz = O\left(\rho^{-m}\exp\left(d^r\psi(\rho) - C_3 r^2\right)\right).$$

Finally, we are left with the integral

$$\begin{aligned} J &= \frac{1}{2\pi i}\int_{\Gamma_3} z^{-m-1}\exp\left(\psi(z)d^r + \frac{\log d!}{d-1}\right) dz \\ &= \frac{1}{2\pi}\int_{-\theta_1}^{\theta_1} \exp\left(\psi(\rho e^{i\theta})d^r + \frac{\log d!}{d-1} - m\log\rho - im\theta\right) d\theta. \end{aligned}$$

By the choice of ρ,

$$d^r\psi(\rho e^{i\theta}) - im\theta = d^r\psi(\rho) - \frac{\theta^2}{2}d^r\left(\rho^2\psi''(\rho) + \rho\psi'(\rho)\right) + O\left(d^r|\theta|^3\right),$$

and so we obtain

$$\begin{aligned} J = {} & \frac{d!^{1/(d-1)}}{2\pi}\rho^{-m}\exp(d^r\psi(\rho)) \\ & \int_{-\theta_1}^{\theta_1}\exp\left(-\frac{\theta^2}{2}d^r\left(\rho^2\psi''(\rho)+\rho\psi'(\rho)\right)\right)\left(1 + O\left(d^r|\theta|^3\right)\right) d\theta \end{aligned}$$

and finally

$$\begin{aligned} J = {} & d!^{1/(d-1)}\rho^{-m}d^{-r/2}\left(2\pi\left(\rho^2\psi''(\rho)+\rho\psi'(\rho)\right)\right)^{-1/2} \\ & \exp(d^r\psi(\rho))(1 + O(r^3 d^{-r/2})). \end{aligned}$$

Putting everything together yields Theorem 2.

In order to prove Corollary 3, we first note that

$$\begin{aligned} \log\left([z^n]p_r(z)\right) = {} & \left(\psi(\rho) - \frac{m}{d^r}\log\rho\right)d^r - \frac{r}{2}\log d + \frac{1}{d-1}\log d! \\ & - \frac{1}{2}\log\left(2\pi\left(\rho^2\psi''(\rho)+\rho\psi'(\rho)\right)\right) + O\left(r^3 d^{-r/2}\right) \end{aligned}$$

by Theorem 2. By the choice of ρ, the principal term can be rewritten as

$$\left(\psi(\rho) - \rho\psi'(\rho)\log\rho\right)d^r.$$

The derivative of the function $f(z) = \psi(z) - z\psi'(z)\log z$ is given by

$$f'(z) = -\left(\psi'(z) + z\psi''(z)\right)\log z,$$

and since we already know that $\psi'(z)+z\psi''(z)$ is strictly positive (Lemma 9), the function has a unique maximum at $z = 1$. The second derivative at $z = 1$ is given by

$$f''(1) = -\left(\psi'(1) + \psi''(1)\right).$$

If we define m_0 by $\psi'(1) = m_0 d^{-r}$, we have

$$d^{-r}(m - m_0) = \rho\psi'(\rho) - \psi'(1) = (\psi'(1) + \psi''(1))(\rho - 1) + O\left((\rho-1)^2\right).$$

Hence,

$$\begin{aligned}
\log\left([z^n]\frac{p_r(z)}{p_r(1)}\right) &= \frac{f''(1)}{2}(\rho-1)^2 d^r + O\left((\rho-1)^3 d^r + (\rho-1) + r^3 d^{-r/2}\right) \\
&= -\frac{\psi'(1)+\psi''(1)}{2}\left(\frac{d^{-r}(m-m_0)}{\psi'(1)+\psi''(1)}\right)^2 d^r \\
&\quad + O\left(d^{-2r}(m-m_0)^3 + d^{-r}(m-m_0) + r^3 d^{-r/2}\right) \\
&= -\frac{1}{2(\psi'(1)+\psi''(1))} d^{-r}(m-m_0)^2 \\
&\quad + O\left(d^{-2r}(m-m_0)^3 + d^{-r}(m-m_0) + r^3 d^{-r/2}\right),
\end{aligned}$$

and so we finally obtain

$$[z^n]\frac{p_r(z)}{p_r(1)} \sim \exp\left(-\frac{1}{2(\psi'(1)+\psi''(1))d^r}(m-m_0)^2\right)$$

if m is "close" to the peak m_0, i.e. $m - m_0 = o(d^{2r/3})$. This proves the corollary.

4. A special case

As an illustrative example, let us consider the case $d = 2$, $k = 1$, i.e. binary trees with the property that all levels except for the last one are completely filled. We have

$$p_r(z) = zZ(S_2, p_{r-1}(z)) = \frac{z}{2}\left(p_{r-1}(z)^2 + p_{r-1}(z^2)\right)$$

with initial value $p_0(z) = 1 + z$. Hence the iterates are

$$\begin{aligned}
p_1(z) &= z + z^2 + z^3, \\
p_2(z) &= z^3 + z^4 + 2z^5 + z^6 + z^7, \\
p_3(z) &= z^7 + z^8 + 3z^9 + 3z^{10} + 5z^{11} + 3z^{12} + 3z^{13} + z^{14} + z^{15},
\end{aligned}$$

and so on. It is obvious that the coefficients have to be symmetric at every step. The total number $p_r(1)$ satisfies the recurrence

$$p_r(1) = \frac{1}{2}\left(p_{r-1}(1)^2 + p_{r-1}(1)\right),$$

which leads to the sequence $2, 3, 6, 21, 231, 26796, 359026206, \ldots$ (sequence A007501 in Sloane's encyclopedia, see [11]; it is also noted there that the sequence counts nonisomorphic complete binary trees with leaves colored using two colors, which is clearly equivalent to our construction).

Now, the function $\psi(z)$ can be computed numerically (see Figure 1). The aforementioned sequence is asymptotically equal to

$$p_r(z) \sim 2 \cdot e^{2^r \psi(1)} \approx 2 \cdot 1.34576817^{2^r}.$$

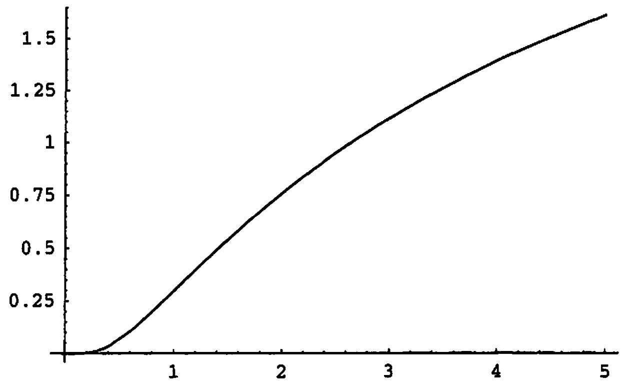

FIGURE 1. Plot of the function $\psi(z)$

For a binary 1-tree with a given number n of vertices, the height is unique. Hence, the number of nonisomorphic binary 1-trees is given immediately by Theorem 2: if b_n is this number and $(1+\epsilon_1)2^r < n < (1-\epsilon_2)2^{r+1}$ for fixed $\epsilon_1, \epsilon_2 > 0$ and a positive integer r, then

$$(8) \qquad b_n \sim 2^{(1-r)/2} \rho^{-m} \left(\pi \left(\rho^2 \psi''(\rho) + \rho\psi'(\rho)\right)\right)^{-1/2} \exp\left(2^r \psi(\rho)\right),$$

uniformly in n, where $m = n - 2^r + 1$ and ρ is chosen in such a way that $\rho\psi'(\rho) = 2^{-r}m$.

Furthermore, Corollary 3 shows that the coefficients of p_r asymptotically follow a normal distribution, where mean and variance can be given explicitly: the mean is exactly $3 \cdot 2^{r-1} - 1$ (which follows immediately from the symmetry of the coefficients), the variance is given by $2^{r-1}(1 - p_r(1)^{-1}) \sim 2^{r-1}$. The latter can be derived from the recurrence: it suffices to note that

$$v_r = \frac{p_r''(1)}{p_r(1)} + \frac{p_r'(1)}{p_r(1)} - \left(\frac{p_r'(1)}{p_r(1)}\right)^2$$

satisfies

$$v_r = 2v_{r-1}\left(1 + \frac{1}{p_{r-1}(1) + 1}\right),$$

from which the formula for v_r follows easily by induction. Around the peak $m_0 = 2^{r-1}$ (equivalently, $n_0 = 3 \cdot 2^{r-1} - 1$), (8) reduces to

$$b_n \sim 2^{1-r/2} \pi^{-1/2} \exp\left(\psi(1)2^r - \frac{1}{2^r}(m - m_0)^2\right),$$

as shown in the previous section.

Finally, the following plot (Figure 2) shows the number of binary 1-trees with n vertices (on a logarithmic scale).

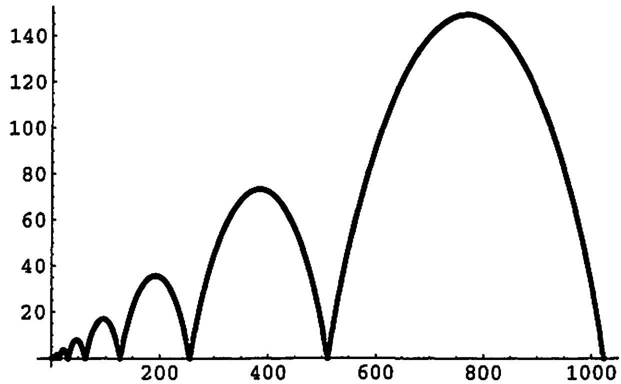

FIGURE 2. The number of binary 1-trees of given size

5. CONCLUSION

Of course, the presented method is not restricted to the specific recurrences that arise from the study of k-trees. Generally, similar results can be expected if a sequence $p_r(z)$ of polynomials is described recursively by an equation of the form

$$p_r(z) = F(p_{r-1}(z), p_{r-1}(z^2), \ldots, p_{r-1}(z^d)),$$

where F is a multivariate polynomial with nonnegative coefficients. The presented case, where cycle indices occur in the recurrences, is typical for combinatorial applications.

REFERENCES

[1] A. V. Aho and N. J. A. Sloane. Some doubly exponential sequences. *Fibonacci Quart.*, 11(4):429–437, 1973.

[2] S. Bereg and H. Wang. Wiener indices of balanced binary trees. *Discrete Appl. Math.*, 155(4):457–467, 2007.

[3] P. Flajolet and A. M. Odlyzko. Limit distributions for coefficients of iterates of polynomials with applications to combinatorial enumerations. *Math. Proc. Cambridge Philos. Soc.*, 96(2):237–253, 1984.

[4] F. Harary and E. M. Palmer. *Graphical enumeration.* Academic Press, New York, 1973.

[5] H.-K. Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European J. Combin.*, 19(3):329–343, 1998.

[6] R. Kemp. Balanced ordered trees. In *Proceedings of the Fifth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science (Poznań, 1991)*, volume 5, pages 99–121, 1994.
[7] D. E. Knuth. *The art of computer programming. Volume 3.* Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973. Sorting and searching, Addison-Wesley Series in Computer Science and Information Processing.
[8] D. E. Knuth. *The art of computer programming.* Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, second edition, 1975. Volume 1: Fundamental algorithms, Addison-Wesley Series in Computer Science and Information Processing.
[9] A. M. Odlyzko. Periodic oscillations of coefficients of power series that satisfy functional equations. *Adv. in Math.*, 44(2):180–205, 1982.
[10] A. M. Odlyzko and L. B. Richmond. On the unimodality of high convolutions of discrete distributions. *Ann. Probab.*, 13(1):299–306, 1985.
[11] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences. published electronically at `http://www.research.att.com/~njas/sequences`.

Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa
E-mail address: `swagner@sun.ac.za`