J. Austral. Math. Soc. (*Series A*) **58** (1995), 298–311

THE NUMBER OF TREES WITH LARGE DIAMETER

TOMASZ ŁUCZAK

(Received 13 February 1992)

Communicated by L. Caccetta

Abstract

In the paper we study the asymptotic behaviour of the number of trees with n vertices and diameter $k = k(n)$, where $k/\sqrt{n} \to \infty$ as $n \to \infty$ but $k = o(n)$.

1991 *Mathematics subject classification* (*Amer. Math. Soc.*): 05C30.

1. Introduction

The *diameter* of a connected graph G is the largest distance between its vertices, where the distance between two vertices is defined as the number of edges in the shortest path connecting them. Let $t(n, k)$ denote the number of labelled trees with n vertices and diameter equal to k. The asymptotic value of $t(n, k)$ for k which is near $\sqrt{n}$ was established by Szekeres [3] by a delicate analysis of the generating function. The purpose of this work is to present a simple combinatorial argument by which one can extrapolate Szekeres' result to all values of k such that $k/\sqrt{n} \to \infty$ but $k/n \to 0$ as $n \to \infty$.

2. The number of trees with large height—a crude upper bound

In this section we study the behaviour of $h(n, k)$, the number of labelled rooted trees on n having height k, where by the *height* we mean the maximum distance from a fixed vertex v_0, called the *root*, to any other vertex of a graph. (Here and below we shall always assume that v_0 is the lexicographically first vertex.)

Research supported by KBN grant 2 1087 91 01.

Our starting point is the following result of Rényi and Szekeres [2], which determines the limit value of $h(n,k)$ when k is of order $\sqrt{n}$.

THEOREM 1. *Let n, k be natural numbers and $\beta = 2n/k^2$. Then*

$$(1)\quad p_n(k) = \frac{h(n,k)}{n^{n-2}} = (2+o(1))\sqrt{\frac{2\pi}{n}}\beta^2 \sum_{i=1}^{\infty}\left(2i^4\pi^4\beta - 3i^2\pi^2\right)\exp(-\beta\pi^2 i^2),$$

uniformly for every $0 < c \le |\beta| \le C$ and any positive constants c and C.

In particular, for n large enough and for every $1 \le k \le n-1$, we have $p_n(k) < 100/\sqrt{n}$.

Let us note that, since c in Theorem 1 could be chosen arbitrarily small, there exists a function $\gamma(n)$ which tends to infinity as $n \to \infty$ such that (1) holds uniformly for every $1 \le |1/\beta| \le \gamma(n)$. Throughout the paper we shall always assume that this function $\gamma(n)$ is non-decreasing, $\gamma(1) > 10^{10}$ and, for n large enough, $\gamma(n) < \log\log\log n$.

The formula for $p_n(k)$, given in (1), can be transformed (for example, using Poisson's formula) to the form

$$2\sqrt{\frac{2\pi}{n}}\sum_{i=1}^{\infty}\left(\frac{2i^4}{\sqrt{\pi}\beta^{3/2}} - \frac{3i^2}{\sqrt{\pi\beta}}\right)\exp\left(-\frac{i^2}{\beta}\right) = \sum_{i=1}^{\infty}\left(\frac{2i^4k^3}{n^2} - \frac{6i^2k}{n}\right)\exp\left(-\frac{i^2k^2}{2n}\right).$$

Thus, for every function $\gamma'(n) \le \gamma(n)$ such that $\gamma'(n) \to \infty$ as $n \to \infty$, uniformly for every $k = k(n)$ such that $\gamma'(n) \le k^2/n \le \gamma(n)$ we have,

$$(2)\qquad p_n(k) = (1+o(1))\frac{2k^3}{n^2}\exp\left(-\frac{k^2}{2n}\right).$$

It turns out that the left hand side of (2), slightly adjusted, can easily be shown to be an upper bound for $p_n(k)$, for all k of the order larger than $\sqrt{n}$.

LEMMA 1. *Let*

$$f(n) = \max_{k \ge \sqrt{n\log\log\gamma(n)}}\left\{p_n(k)\Big/\frac{2k^3}{n^2}\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}\right)\right\}.$$

Then

$$\limsup_{n\to\infty} f(n) \le 1.$$

PROOF. Note first that

$$h(n,k) \le \binom{n-1}{k} k!\,k(n-1)^{n-k-2} = (n-1)_k k(n-1)^{n-k-2}$$

$$(3)\qquad \le n^{n-2}k\exp\left(-\frac{k^2}{2n} - \frac{k^3}{6n^2}\right),$$

so, for $k \geq n^{0.67}$,

$$p_n(k) \Big/ \frac{2k^3}{n^2} \exp\left(-\frac{k^2}{2n} + \frac{k^3}{3n^2}\right) \leq \frac{n^2}{k^4} \exp\left(-\frac{k^3}{2n^2}\right) \leq 0.5.$$

(Here and below we claim that all inequalities are valid only for n large enough.)

Suppose that the assertion of Lemma 1 does not hold. Then, for some constant $\epsilon > 0$, there exist an absolute constant C and a function $z(n)$ such that $z(n) > 1 + \epsilon$ and for every n_0, one can find $n \geq n_0$ such that

$$p_n(k) \geq z(n) \frac{2k^3}{n^2} \exp\left(-\frac{k^2}{2n} + \frac{k^3}{3n^2}\right) \tag{4}$$

for some $\sqrt{n \log\log \gamma(n)} \leq k \leq n^{0.67}$, whereas for every $m \leq n$ we have

$$f(m) \leq C f(n) \leq 2Cz(n).$$

We shall show that (4) leads to a contradiction.

Let us define an (n, k, l)-*structure* as a triple (T', P, T''), where T' is a rooted tree of $|T'| \leq n - l$ vertices, $P = v_0 v_1 \ldots v_{k-l}$ is a path of length l contained in T' which starts at the root, and T'' is a rooted tree with $n - |T'|$ vertices with height equal to $l - 1$. Suppose that a rooted tree T has height k and path $P' = v_0 v_1 v_2 \ldots v_k$ joining the root of T to the highest leaf of T. (If there are many such leaves, take as v_k the lexicographically first one.) Then one may obtain from T an (n, k, l)-structure by setting $P = v_0 v_1 \ldots v_{k-l}$, and picking as T' and T'' trees obtained from T by deleting edge $v_{k-l} v_{k-l+1}$, where vertex v_{k-l+1} serves as the root of T'. Thus, the number $a(n, k, l)$ of (n, k, l)-structures is a rather natural upper bound for $h(n, k)$. In fact, we shall prove later that for suitably chosen l, $h(n, k) = (1 + o(1)) a(n, k, l)$.

Clearly, for $a(n, k, l)$, we have

$$\begin{aligned} \frac{a(n,k,l)}{n^{n-2}} &= \sum_{m=l}^{n-1-k+l} \binom{n-1-k+l}{m} (m+1)^{m-1} p_m(l)(k-l) \frac{(n-m-1)^{n-m-k+l-2}}{n^{n-2}} \\ &\quad \times \binom{n-1}{k-l} (k-l)! \\ &= \sum_m \frac{n!}{n^{n-1}} \frac{(m+1)^m}{(m+1)!} \frac{(n-m-1)^{n-m-k+l-2}}{(n-1-k+l-m)!} (k-l) p_m(l). \end{aligned}$$

Hence, using Stirling's formula, we get

$$\frac{a(n,k,l)}{n^{n-2}} = \frac{n^{n+1/2}}{n^{n-1}} \sum_m \frac{(m+1)^{m-1}}{(m+1)!} \frac{(n-m-1)^{n-m-k+l-2}}{(n-m-1-k+l)^{n-m-k+l-1/2}}$$

$$\times(k-l)p_m(l)\exp\left(-k+l+O\left(\frac{1}{m}+\frac{1}{n-m-k+l}\right)\right)$$

$$=\frac{1}{\sqrt{2\pi}}\sum_m \frac{k-l}{m^{3/2}}\left(1+\frac{k-l}{n-m-1-k+l}\right)^{n-m-1-k+l}$$

$$\times p_m(l)\exp\left(-k+l+O\left(\frac{1}{m}+\frac{1}{n-m-k+l}+\frac{m}{n}\right)\right), \tag{5}$$

where all constants hidden in $O(\cdot)$ can be bounded from above uniformly for all m.

If $\sqrt{n\log\log\gamma(n)}\le k\le n^{0.67}$ then

$$\left(1+\frac{k-l}{n-m-1-k+l}\right)^{n-m-1-k+l}\le\exp\left(k-l-\frac{(k-l)^2}{2(n-m)}+\frac{k^3}{3n^2}\right),$$

so, from (4),

$$\frac{a(n,k,l)}{n^{n-2}}\le\left\{\sum_{m=l}^{n-1-k+l}\frac{k}{m^{3/2}}\exp\left(\frac{(k-l)^2}{2(n-m)}+O\left(\frac{1}{m}+\frac{1}{n-m-k+l}+\frac{m}{n}\right)\right)p_m(l)\right\}$$
$$\times\frac{1}{\sqrt{2\pi}}\exp\left(\frac{k^3}{3n^2}\right).$$

We shall estimate the above expression for $l=(n/2k)\log\gamma(n)$. Let us consider first the case when $m\le m_-$, where $m_-=n^2/(20k^2)\log\gamma(n)<l^2/\log\gamma(m)$. Then, due to our assumption,

$$p_m(l)\le f(m)\frac{2l^3}{m^2}\exp\left(-\frac{l^2}{2m}+\frac{l^3}{3m^2}\right)\le 2Cz(n)\frac{2l^3}{m^2}\exp\left(-\frac{l^2}{2m}+\frac{l^3}{3m^2}\right)$$

and, for n large enough,

$$\frac{1}{\sqrt{2\pi}}\exp\left(\frac{k^3}{3n^2}\right)\sum_{m=l}^{m_-}\frac{k}{m^{3/2}}\exp\left(-\frac{(k-l)^2}{2(n-m)}+O\left(\frac{1}{m}+\frac{1}{n-m-k+l}+\frac{m}{n}\right)\right)p_m(l)$$
$$\le 2Cz(n)\exp\left(\frac{k^3}{3n^2}\right)\sum_{m=l}^{m_-}\frac{kl^3}{m^{7/2}}\exp\left(-\frac{l^2}{2m}+\frac{(k-l)^2}{2(n-m)}+\frac{l^3}{3m^2}\right)$$
$$\le 2Cz(n)\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}\right)\sum_{m=l}^{m_-}\frac{kl^3}{m^{7/2}}\exp\left(-\frac{l^2}{2m}+\frac{kl}{n-m}+\frac{l^3}{3m^2}\right)$$
$$\le 2Cz(n)\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}\right)\sum_{m=l}^{l^2/\log\gamma(n)}\frac{kl^3}{m^{7/2}}\exp\left(-\frac{l^2}{10m}\right)$$
$$\le 2Cz(n)\frac{k(\log\gamma(n))^2}{l^2}\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}-\log\gamma(n)/10\right)$$
$$\le\frac{2Cz(n)}{\log\gamma(n)}\frac{k^3}{n^2}\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}\right).$$

Now set $m_+ = (4n^2/k^2)\log\gamma(n)$ and consider the case when $m_- \le m \le m_+$. For such m we have $0.1\log\gamma(m) \le l^2/2m \le \log\gamma(m)$ so we can approximate $p_m(l)$ using (2). Thus

$$\frac{1}{\sqrt{2\pi}}\exp\left(\frac{k^3}{3n^2}\right)\sum_{m=m_-}^{m_+}\frac{k}{m^{3/2}}\exp\left(\frac{(k-l)^2}{2(n-m)}+O\left(\frac{1}{m}+\frac{m}{n}\right)\right)p_m(l)$$

$$= \frac{1+o(1)}{\sqrt{2\pi}}\exp\left(\frac{k^3}{3n^2}\right)\sum_{m=m_-}^{m_+}\frac{2kl^3}{m^{7/2}}\exp\left(-\frac{l^2}{2m}-\frac{(k-l)^2}{2(n-m)}\right). \tag{6}$$

The function $g(x) = a^2/x + b^2/(c-x)$ attains the maximum for $x = ac/(a+b)$. Set $m_0 = ln/k$ and $\Delta m = m - m_0$. Then (6) becomes

$$\frac{2+o(1)}{\sqrt{2\pi}}\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}\right)\sum_{m=m_--m_0}^{m_+-m_0}\frac{kl^3}{(m_0+\Delta m)^{7/2}}\exp\left(-\frac{(\Delta m)^2}{2}\frac{l^2}{m_0^3}\right)$$

$$= (2+o(1))\frac{k^3}{n^2}\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}\right).$$

Finally, note that if $m \ge m_+$ then

$$\frac{(k-l)^2}{2(n-m)} \ge \frac{k^2}{2n} - \frac{kl}{n-m} + \frac{m(k-l)^2}{n^2} \ge \frac{k^2}{2n}.$$

Thus, since from Theorem 1 $\max_l\{p_m(l)\} \le O(1/\sqrt{m})$, we arrive at

$$\frac{1}{\sqrt{2\pi}}\exp\left(\frac{k^3}{3n^2}\right)\sum_{m=m_+}^{n-k+l-1}\frac{k}{m^{3/2}}\exp\left(\frac{(k-l)^2}{2(n-m)}+O\left(\frac{1}{m}+\frac{1}{n-m-k+l}+\frac{m}{n}\right)\right)p_m(l)$$

$$\le O(k)\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}\right)\sum_{m=m_+}^{n-k+l-1}\frac{1}{m^2}$$

$$\le \frac{O(k)}{m_+}\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}\right) \le \frac{O(k^3)}{n^2\log\gamma(n)}\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}\right).$$

Hence

$$p_n(k) \le \frac{a(n,k,(n/2k)\log\gamma(n))}{n^{n-2}}$$

$$\le \left(\frac{Cz(n)+O(1)}{\log\gamma(n)}+1+o(1)\right)\left(\frac{2k^3}{n^2}\exp\left(-\frac{k^2}{2n}+\frac{k^3}{3n^2}\right)\right)$$

contradicting (4).

3. The number of trees with large height—the asymptotic behaviour

In this part of the paper, using the upper bound for $p_n(k)$ provided by Lemma 1, we repeat the argument from the previous section to get the limit value for $h(n,k)$ when $k^2/n \to \infty$ but $k = o(n)$. However, in order to do it we should know that, for suitably chosen l, $a(n,k,l) = (1+o(1))h(n,k)$.

Let $F(n,k)$ denote a forest chosen uniformly from all forests with the vertex set $\{1,2,\dots,n\}$ and $n-k$ edges, such that vertices $1,2,\dots,k$ belong to different trees. Moreover define $H(n,k)$ as the result of adding edges $\{1,2\},\{2,3\},\dots,\{k-1,k\}$ to $F(n,k)$. Now, in order to show that $a(n,k,l) = (1+o(1))h(n,k)$ it is enough to prove that *almost surely* (that is, with probability tending to 1 as $n \to \infty$) the graph $H(n,k)$ contains no paths starting at vertex 1 longer than $k+l-2$.

LEMMA 2. *Let $k^2/n \to \infty$, $k = o(n)$ and $\omega(n)$ be any function which tends to infinity with n. Then almost surely each path contained in $H(n,k)$ which starts at vertex 1 is shorter than $k + \omega(n)n/k$.*

PROOF. Let T_i, for $i = 1,2,\dots,k$, denote the tree of $F(n,k)$ which contains vertex i. We shall show first that almost surely every T_i contains less than $\hat{m}(i) = (k - i + \sqrt{\omega(n)}n/k)^2$ vertices. Indeed, since it is well known that almost surely the maximum size of a tree in the random forest $F(n,k)$ is less than $(4n^2/k^2)\log n$ (see Pavlov [1]), the size of T_i is less than $\hat{m}(i)$ for every $i \le k - 3(n/k)\log n$. On the other hand, for the expected number of trees T_i such that $i > i_0 = k - 3(n/k)\log n$, and with T_i having more than $\hat{m}(i)$ vertices, we have

$$\begin{aligned}
&\sum_{i>i_0}\sum_{m>\hat{m}(i)} \binom{n-k}{m}(m+1)^{m-1}\frac{(k-1)(n-m-1)^{n-m-k-1}}{kn^{n-k-1}} \\
&\qquad \le \sum_{i>i_0}\sum_{m>\hat{m}(i)} \frac{1}{m^{3/2}}\frac{(n-k)^{n-k+1/2}(n-m)^{n-m-k-1}}{(n-k-m)^{n-k-m+1/2}n^{n-k-1}} \\
&\qquad \le \sum_{i<k-i_0}\sum_{m>m(k-i)} \frac{1}{m^{3/2}}\exp\left(-\frac{k^2m}{3n^2}\right) \qquad (7)\\
&\qquad \le \sum_{i=1}^{3(n/k)\log n} \frac{12}{i+\sqrt{\omega(n)}n/k}\exp\left(-\frac{k^2(i+\sqrt{\omega(n)}n/k)^2}{3n^2}\right) \\
&\qquad \le 40\exp(-\omega(n)/3) \to 0.
\end{aligned}$$

Let X be the random variable which counts all trees T_i with less than $\hat{m}(i)$ vertices and with height at least $\hat{h}(i) = k - i + \omega(n)n/k$. Since $\hat{h}^2(i)/\hat{m}(i) \to 0$, the probability that the height of T_i is larger than $\hat{h}(i)$ provided that it has $m \le \hat{m}(i)$

vertices is, due to Lemma 1, bounded from above by

$$(1+o(1)) \sum_{k \geq \hat{h}(i)} \frac{2k^3}{m^2} \exp\left(-\frac{k^2}{2m} + \frac{k^3}{3m^2}\right) \leq (1+o(1)) \frac{4\hat{h}^2(i)}{m} \exp\left(-\frac{\hat{h}^2(i)}{2m} + \frac{\hat{h}^3(i)}{3m^2}\right).$$

Thus, calculations similar to that from (7) lead to the following formula for the expectation of X

$$\begin{aligned} EX &\leq \sum_{i=1}^{k} \sum_{m \leq \hat{m}(i)} \frac{4\hat{h}^2(i)}{m^{5/2}} \exp\left(-\frac{k^2 m}{3n^2} - \frac{\hat{h}^2(i)}{2m} + \frac{\hat{h}^3(i)}{3m^2}\right) \\ &\leq \sum_{i=1}^{k} \sum_{m \leq \hat{m}(i)} \frac{4\hat{h}^2(i)}{m^{5/2}} \exp\left(-\frac{k^2 m}{6n^2} - \frac{\hat{h}^2(i)}{6m}\right) \\ &\leq 40 \sum_{i=1}^{k} \sqrt{\frac{k^3 \hat{h}(i)}{n^2}} \exp\left(-\frac{k^2 \hat{h}^2(i)}{6n^2}\right) \\ &\leq \omega(n) \exp(-\omega(n)) \to 0. \end{aligned}$$

Thus, almost surely $H(n, k)$ contains no trees T_i with height at least $\hat{h}(i) = k - i + \omega(n)n/k$ and the assertion follows.

THEOREM 2. *Let $k = k(n)$ be a function of n such that $k/\sqrt{n} \to \infty$ as $n \to \infty$ but $k = o(n)$. Then*

$$h(n, k) = (1 + o(1)) \frac{2n! k^3 n^{n-k-4}}{(n-k)!}. \tag{8}$$

PROOF. Since, for $k \leq \sqrt{n\gamma(n)}$, (8) follows from (2) and Stirling's formula it is enough to prove Theorem 2 for $k \geq \sqrt{n\gamma(n)}$. Due to Lemma 2, $h(n, k) = (1+o(1))a(n, k, l)$ whenever $lk/n \to \infty$ as $n \to \infty$. Let us set $l = (n/k) \log \gamma(n/k)$. Then (5) becomes

$$\begin{aligned} \frac{a(n,k,l)}{n^{n-k-2}} \frac{(n-k)!}{n!} &= \frac{1}{\sqrt{2\pi}} \sum_m \frac{k-l}{m^{3/2}} \left(1 + \frac{k-l}{n-m-1-k+l}\right)^{n-m-1-k+l} \\ &\quad \times \frac{n^k (n-k)!}{n!} p_m(l) \exp\left(-k + l + O\left(\frac{1}{m} + \frac{1}{n-m-k+l} + \frac{m}{n}\right)\right). \end{aligned} \tag{9}$$

Set $m_- = (n^2/50k^2) \log \gamma(n/k)$ and $m_+ = (n^2/k^2) \log \gamma(n/k) \log\log \gamma(n/k)$. As in the proof of Lemma 2 we shall split the sum in (9) into three parts and estimate each of them separately.

Note first that, by elementary calculations,

$$\left(1+\frac{k-l}{n-m-1-k+l}\right)^{n-m-1-k+l}$$
$$= (1+o(1)\left(1+\frac{k-l}{n-k+l-1}\right)^{n-1-k+l}\exp\left(\frac{m(k-l)^2}{n^2}+O\left(\frac{mk^3}{n^3}+\frac{m^2k^2}{n^3}\right)\right).$$

Thus, since $k^2/n \geq \gamma(n)$, for $m \leq m_-$ we get

$$\left(1+\frac{k-l}{n-m-1-k+l}\right)^{n-m-1-k+l}$$
$$(10) \qquad = (1+o(1))\left(1+\frac{k-l}{n-k+l-1}\right)^{n-1-k+l}\exp\left(\frac{mk^2}{2n^2}\right).$$

Moreover,

$$(1+o(1))\left(1+\frac{k-l}{n-k+l-1}\right)^{n-1-k+l} = (1+o(1))\left(1+\frac{k}{n-k}\right)^{n-k}\exp\left(-l+\frac{kl}{n}\right)$$
$$(11) \qquad = (1+o(1))\frac{n!}{(n-k)!}\frac{1}{n^k}\exp\left(k-l+\frac{kl}{n}\right).$$

Hence, for $m \leq m_-$, using Lemma 1 we get

$$\frac{1}{\sqrt{2\pi}}\frac{n^k(n-k)!}{n!}\sum_{m\leq m_-}\frac{k}{m^{3/2}}p_m(l)\left(1+\frac{k-l}{n-m-1-k+l}\right)^{n-m-1-k+l}$$
$$\times \exp\left(-k+l+O\left(\frac{1}{m}+\frac{m}{n}\right)\right)$$
$$(12) \qquad \leq (1+o(1))\sum_{m\leq m_-}\frac{kl^3}{m^{7/2}}\exp\left(\frac{kl}{n}+\frac{mk^2}{2n^2}+\frac{l^3}{3m^2}-\frac{l^2}{2m}+O\left(\frac{1}{m}+\frac{m}{n}\right)\right).$$

But for $m \leq m_-$ we have

$$\frac{kl}{n}+\frac{mk^2}{2n^2}+\frac{l^3}{3m^2}-\frac{l^2}{2m} < -\frac{l^2}{20m},$$

so the left hand side of (12) can be bounded from above by

$$(13) \quad (1+o(1))\sum_{m\leq m_-}\frac{kl^3}{m^{7/2}}\exp\left(-\frac{l^2}{20m}\right) \leq \frac{50kl}{m_-^{3/2}}\exp\left(-\frac{l^2}{20m_-}\right) \leq \frac{k^3}{n^2\log\gamma(n/k)}.$$

Similarly as in the proof of Lemma 1, using (2), we get

$$\frac{1}{\sqrt{2\pi}}\frac{n^k(n-k)!}{n!}\sum_{m=m_-}^{m_+}\frac{k}{m^{3/2}}p_m(l)\left(1+\frac{k-l}{n-m-1-k+l}\right)^{n-m-1-k+l} \times \exp\left(-k+l+O\left(\frac{1}{m}+\frac{m}{n}\right)\right)$$
$$\le \frac{1+o(1)}{\sqrt{2\pi}}\sum_{m=m_-}^{m_+}\frac{2kl^3}{m^{7/2}}\exp\left(\frac{kl}{n}+\frac{mk^2}{2n^2}-\frac{l^2}{2m}\right).$$

and setting $m_0 = ln/k$, $\Delta m = m - m_0$ leads to

$$\frac{1+o(1)}{\sqrt{2\pi}}\sum_{m=m_-}^{m_+}\frac{2kl^3}{m^{7/2}}\exp\left(\frac{kl}{n}+\frac{mk^2}{2n^2}-\frac{l^2}{2m}\right)$$
$$= \frac{1+o(1)}{\sqrt{2\pi}}\sum_{\Delta m=m_--m_0}^{m_+-m_0}\frac{2kl^3}{(m_0+\Delta m)^{7/2}}\exp\left(-\frac{(\Delta m)^2}{2}\frac{l^2}{m_0^3}\right)$$
$$= \frac{1+o(1)}{\sqrt{2\pi}}\frac{2kl^2}{m_0^2}\int_{-\infty}^{\infty}e^{-x^2/2}dx = (1+o(1))2k^3/n^2. \tag{14}$$

In order to deal with large values of m note that for every $x \in (0, 1/2)$ and $y \in (0, 1)$

$$(1+x/(1-y))^{1-y} \le (1+x)\exp(-0.1x^2y^2).$$

Thus, for $m \ge m_+$, we have

$$\left(1+\frac{k-l}{n-m-1-k+l}\right)^{n-m-1-k+l} \le \left(1+\frac{k-l}{n-k+l-1}\right)^{n-1-k+l}\exp\left(-\frac{m^2(k-l)^2}{10n^2}\right)$$
$$\le \left(1+\frac{k-l}{n-k+l-1}\right)^{n-1-k+l}\exp\left(-\frac{kl}{n}\right).$$

and (11) together with the fact that $p_m(l) \le 100/\sqrt{m}$ implies that

$$\frac{1+o(1)}{\sqrt{2\pi}}\sum_{m\ge m_+}\frac{k}{m^{3/2}}\left(1+\frac{k-l}{n-m-1-k+l}\right)^{n-m-1-k+l}$$
$$\times\frac{n^k(n-k)!}{n!}p_m(l)\exp\left(-k+l+O\left(\frac{1}{m}+\frac{1}{n-m-k+l}+\frac{m}{n}\right)\right)$$
$$\le 50\sum_{m\ge m_+}\frac{k}{m^2} \le \frac{50k}{m_+} \le \frac{50}{\log\log\gamma(n/k)}\frac{k^3}{n^2}. \tag{15}$$

Thus, the assertion follows from (9), (13), (14) and (15).

As a simple consequence of Theorem 2 we get a new upper bound for $h(n,k)$, which, for large k, is much better than the one given in Lemma 1.

COROLLARY 1. *There exists an absolute constant A such that for every n and every* $k \geq \sqrt{n}$

$$h(n,k) \leq An!k^3n^{n-k-4}/(n-k)!. \tag{16}$$

PROOF. Let us suppose that the assertion does not hold. Then we may find a sequence $\{n_i\}_{i=1}^{\infty}$ and a function $k(n)$ such that $k(n) \geq \sqrt{n}$ and

$$\lim_{i\to\infty} \frac{h(n_i, k(n_i))(n_i-k)!}{(n_i)!k^3 n_i^{n_i-k-4}} = \infty. \tag{17}$$

Due to Theorems 1 and prefthm:3.1 the function $k(n)$ could be chosen in such a way that $n/k(n) \leq C$ for some constant C. However, in such a case, from the trivial upper bound given in (3) we get

$$\frac{h(n_i, k(n_i))(n_i-k)!}{(n_i)!k^3 n_i^{n_i-k-4}} \leq \frac{n_i^2}{k^2(n_i)} \leq C^2$$

contradicting (17).

REMARK. After some more work it can be shown that if $k(n)/n \to a$, where $0 < a \leq 1$, then for some constant $\alpha(a) > 0$

$$h(n,k) = (1+o(1))\alpha(a)n!k^3n^{n-k-4}/(n-k)!. \tag{18}$$

Theorem 2 states that $\alpha(a) \to 2$ as $a \to 0$ and one could easily check that $\alpha(a) \to 1$ as $a \to 1$. However, to determine the exact value of $\alpha(a)$ for $0 < a < 1$ one probably needs more sophisticated tools than the elementary combinatorial approach presented in this paper.

4. Trees with large diameter

The asymptotic behaviour of the number $t(n,k)$ of trees with n vertices and diameter k was considered by Szekeres in [3], who found the limiting value of $t(n,k)$ for $k \sim \sqrt{n}$.

THEOREM 3. *Let n, k be natural numbers and $\bar{\beta} = n/(2k^2)$. Then*

$$\begin{aligned}\frac{t(n,k)}{n^{n-2}} = \frac{1+o(1)}{3}\sqrt{\frac{2\pi}{n}} \sum_{i=1}^{\infty} \Big[& 4\pi^8 i^8 \bar{\beta}^6 - 36\pi^6 i^6 \bar{\beta}^5 + 75\pi^4 i^4 \bar{\beta}^4 \\ & - 30\pi^2 i^2 \bar{\beta}^3 + 4\pi^6 i^6 \bar{\beta}^4 - 10\pi^4 i^4 \bar{\beta}^2 \Big] \exp(-\bar{\beta}\pi^2 i^2),\end{aligned} \tag{19}$$

uniformly for every $0 < c < |\bar{\beta}| \leq C$ and any positive constants c and C.

The main result of this section is stated in the following theorem.

THEOREM 4. *Let $k = k(n)$ be a function of n such that $k = o(n)$ but $k/\sqrt{n} \to \infty$ as $n \to \infty$. Then*

$$t(n,k) = (1+o(1))\frac{2n!k^5n^{n-k-5}}{(n-k)!}. \tag{20}$$

REMARK. Note that $t(n,k) = (2+o(1))(k/n)^5\exp(-k^2/2n + O(k^3/n^2))$. Thus, if we transform (19) using Poisson's formula, in the resulting sum the polynomial coefficient of $\exp(-1/\bar{\beta})$ disappears.

PROOF. Let us consider first the case when k is odd. Each tree with diameter $k = 2r+1$ could be, in a natural way, decomposed into two rooted trees, each having height r, so

$$t(n, 2r+1) = \frac{1}{2}\sum_{m=h+1}^{n-h+1}\binom{n}{m} mh(m,r)(n-m)h(n-m,r),$$

where the factor 1/2 appears since we count each tree twice. If m is contained between $n/2$ and $3n/4$ then we could use Theorem 2 to estimate $h(m,r)$ and $h(n-m,r)$, so, using Stirling's formula, we get

$$\begin{aligned}
&\frac{1}{2}\sum_{m=n/2}^{3n/4}\binom{n}{m} mh(m,r)(n-m)h(n-m,r)\\
&= \frac{1+o(1)}{2}\sum_m \frac{n!}{m!(n-m)!}\,\frac{2r^3m!m^{m-r-3}}{(m-r)!}\,\frac{2r^3(n-m)!(n-m)^{n-m-r-3}}{(n-m-r)!}\\
&= \frac{1+o(1)}{\sqrt{2\pi}}\,\frac{2r^6n!n^{n-2r+1/2}}{(n-2r)!}\sum_m \frac{1}{m^{7/2}(n-m)^{7/2}}\,\frac{(n-2r)^{n-2r}m^{m-r}(n-m)^{n-m-r}}{n^{n-2r}(m-r)^{m-r}(n-m-r)^{n-m-r}}.
\end{aligned}$$

Set $m = n/2 + \Delta m$. Then

$$\begin{aligned}
&\sum_{m=n/2}^{3n/4}\frac{1}{m^{7/2}(n-m)^{7/2}}\,\frac{(n-2r)^{n-2r}m^{m-r}(n-m)^{n-m-r}}{n^{n-2r}(m-r)^{m-r}(n-m-r)^{n-m-r}}\\
&= \sum_{\Delta m=-n/2}^{n/2}\frac{2^7}{(n^2-4(\Delta m)^2)^{7/2}}\,\frac{(n-2r)^{n-2r}(n+2\Delta m)^{n/2+\Delta m-r}(n-2\Delta m)^{n/2-\Delta m-r}}{n^{n-2r}(n+2\Delta m-r)^{n/2+\Delta m-r}(n-2\Delta m-r)^{n/2-\Delta m-r}}\\
&= \sum_{\Delta m=-n/2}^{n/2}\frac{2^7}{(n^2-4(\Delta m)^2)^{7/2}}\left(1-\frac{16r(n-r)(\Delta m)^2}{(n-2r)^2(n^2-4(\Delta m)^2)}\right)^{r-n/2}\\
&\qquad\qquad\left(1-\frac{8r\Delta m}{(n-2r+2\Delta m)(n-2\Delta m)}\right)^{\Delta m}
\end{aligned}$$

$$= \sum_{\Delta m=-n/2}^{n/2} \frac{2^7}{(n^2-4(\Delta m)^2)^{7/2}} \exp\left(-\frac{8r^2(\Delta m)^2}{n^3} + O\left(\frac{r(\Delta m)^2}{n^{5/2}} + \frac{r^2(\Delta m)^3}{n^4}\right)\right)$$

$$= (1+o(1))\frac{2^5\sqrt{2\pi}}{rn^{11/2}}.$$

Hence

$$\frac{1}{2}\sum_{m=n/2}^{3n/4}\binom{n}{m} mh(m,r)(n-m)h(n-m,r) = (1+o(1))\frac{2^6 n! r^5 n^{n-2r-5}}{(n-2r)!}$$

$$= (1+o(1))\frac{2n!(2r+1)^5 n^{n-(2r+1)-5}}{(n-(2r+1))!}.$$

Moreover, one can use (16) and repeat the above calculations to show that the sum of all terms for which either $m \le n/2$ or $m \ge 3n/4$ is $o(n!r^5 n^{n-(2r+1)-5}/(n-2r-1)!)$. Thus, (20) holds for odd k.

Now let $k = 2r$. Then, similarly as in the odd case, each tree T having diameter $k = 2r$ can be viewed as two rooted trees T' and T'' whose roots are joined by an edge, where the heights of T' and T'' equals r and $r-1$ respectively. Furthermore, for each tree T, such a decomposition could be done in at least two ways. Hence, since $h(n,k) = (1+o(1))h(n,k+1)$, as an upper bound for $t(n,2r)$ we get

$$\frac{1}{2}\sum_{m=1}^{r}\binom{n}{m} mh(m,r-1)(n-m)h(n-m,r)$$

$$= \frac{1+o(1)}{2}\sum_{m=1}^{r}\binom{n}{m} mh(m,r)(n-m)h(n-m,r)$$

$$= (1+o(1))t(n,2r+1). \tag{21}$$

However, the number of decomposition of a tree T with even diameter into T' and T'' can be larger than 2. Indeed, it might happen that deleting from T the common midpoint w of all paths of length $2r$ results in a forest of rooted trees, among which more than two have height $r-1$. (By the root of a tree in such a forest we mean the vertex previously joined to w.) Thus, in order to show that the number of trees for which this happens is a negligible fraction of the number of all trees with diameter k, it is enough to check that a 'typical' tree T'' having height r contains no two edge-disjoint paths of length r which start at the root.

Note first the following simple fact.

FACT 1. *Let $F(n,k)$ be a forest chosen at random from all forests with vertex set $\{1,2,\dots,n\}$, which consists of k components each of them containing precisely one from vertices $\{1,2,\dots,k\}$ and $\omega(n)\to\infty$ as $n\to\infty$. Then the probability that* *the component of* $F(n,k)$ *containing* 1 *has more than* $n\omega(n)/k$ *vertices tends to 0 as* $n \to \infty$.

PROOF. By symmetry, the expected size of each component of $F(n,k)$ is n/k. Thus the assertion follows from Markov's inequality.

From the proof of Theorem 2 it follows that to build a rooted tree T with n vertices and height k one should set l slightly larger than n/k, choose $k-l$ vertices, build a path $P = v_0v_1v_2\ldots v_{k-l}$ starting at the root, choose roughly m other vertices, build on these vertices a rooted tree having height $k-l-1$, join the root of this tree to the last vertex v_{k-l} of P and finally, on the remaining $n-m$ vertices, build a forest F such that each of its components contains precisely one of vertices $v_0, v_1, v_2, \ldots, v_{k-l}$. Thus, by the fact above, if T is chosen at random from all rooted trees with n vertices and height k then the component of F containing root v_0 almost surely has size less than $\sqrt{n} \le k$, provided $k/\sqrt{n} \to \infty$ as $n \to \infty$. Hence, the number of rooted trees of n vertices which have height r and contain two or more paths of length r starting at the root is negligible when compared with the number of all rooted trees of size n and height r, provided $r/\sqrt{n} \to 0$ as $n \to \infty$. Consequently, (21) gives the correct value of $t(n, 2r)$ and the assertion follows.

REMARK. Similarly as in the case of the height of rooted trees one can also prove that

$$t(n,k) \le A'n!k^5n^{n-k-5}/(n-k)!,$$

for every $k \ge \sqrt{n}$ and some absolute constant A'.

Moreover, if $k(n)/n \to a$, where $a < 0 \le 1$, then

$$t(n,k) = (1+o(1))\frac{\alpha^2(a)n!k^5n^{n-k-5}}{2(n-k)!},$$

where the constant $\alpha(a)$ is defined by (18)).

5. Acknowledgement

I would like to thank Adam Neugebauer for an interesting and stimulating discussion.

References

[1] Yu. L. Pavlov, 'A case of limit distribution of the maximal volume on a tree in a random forest', *Math. Notes* **25** (1979), 387–392.

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

[3] G. Szekeres, 'Distribution of labelled trees by diameter', in: *Combinatorial Mathematics X, Proc. Conf. Adelaide, Australia 1982*, Lecture Notes in Math. 1036 (Springer, Berlin, 1983) pp. 392–397.

Department of Discrete Mathematics
Adam Mickiewicz University
Poznań
Poland