Journal of Combinatorial Theory, Series B 100 (2010) 161–170

Contents lists available at ScienceDirect

Journal of Combinatorial Theory, Series B

www.elsevier.com/locate/jctb

The average order of a subtree of a tree

Andrew Vince [a], Hua Wang [b,1]

[a] Department of Mathematics, University of Florida, 358 Little Hall, PO Box 118105, Gainesville, FL 32611-8105, United States
[b] Department of Mathematical Sciences, Georgia Southern University, PO Box 8093, Statesboro, GA 30460, United States

ARTICLE INFO

Article history:
Received 17 March 2008
Available online 11 June 2009

Keywords:
Average order
Subtree

ABSTRACT

Let T be a tree all of whose internal vertices have degree at least three. In 1983 Jamison conjectured in JCT B that the average order of a subtree of T is at least half the order of T. In this paper a proof is provided. In addition, it is proved that the average order of a subtree of T is at most three quarters the order of T. Several open questions are stated.

1. Introduction

The *order* of a tree is the number of its vertices. The following notation will be used throughout the paper. For a tree T, its order is denoted $n := n(T)$ and the number of leaves $n' := n'(T)$. This paper concerns the average order of a subtree of a tree T. If T has N subtrees (not including the empty tree) of orders $n_1, n_2, \dots, n_N$, then let

$$\mu_T := \frac{1}{N}\sum_{i=1}^{N} n_i$$

denote the *average order of subtrees* of T. If T has order n, call

$$D_T := \frac{\mu_T}{n}$$

the *density* of T. The density allows us to compare average subtree order of trees of different sizes. The density is also the probability that a vertex chosen at random from T will belong to a randomly chosen subtree of T. In [1] Jamison conjectured that if T is a tree whose internal vertices have degree at least three, then the average order of a subtree is at least half the order of T. In other words, for

E-mail addresses: avince@math.ufl.edu (A. Vince), hwang@georgiasouthern.edu (H. Wang).
[1] Fax: +1 912 478 0654.

0095-8956/$ – see front matter
doi:10.1016/j.jctb.2009.05.006

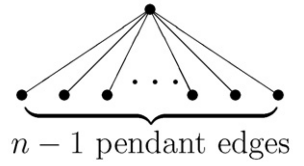

Fig. 1. Star S_n.

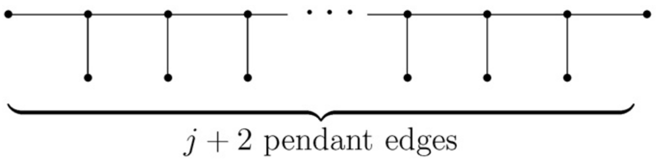

Fig. 2. Caterpillar C_j with $j+2$ leaves.

such a tree, $D_T \geqslant \frac{1}{2}$. The bound given below is the main result of this paper. The proof of the lower bound appears in Section 2 and the proof of the upper bound in Section 3.

Theorem 1. *If T is a tree all of whose internal vertices have degree at least three, then*

$$\frac{1}{2} \leqslant D_T < \frac{3}{4}.$$

Both bounds are best possible in the sense that there exists an infinite sequence $\{S_n\}$ of trees such that $\lim_{n\to\infty} D_{S_n} = 1/2$ and a sequence $\{C_j\}$ of trees such that $\lim_{j\to\infty} D_{C_j} = 3/4$. The sequence S_n of stars of order n, shown in Fig. 1, is an example in the first case. The sequence $\{C_j\}$ of caterpillars with $j+2$ leaves shown in Fig. 2 is an example in the second case. In fact, the exact formula for the density

$$D_{C_j} = \frac{27j2^{j-1} - 3\cdot 2^{j+2} + 4j + 16}{(2j+2)(9\cdot 2^j - j - 6)}$$

can be derived using the recursions in Lemma 1. (The somewhat complicated derivation involves solving recurrences giving the parameters of C_{j+1} in terms of the parameters of C_j. The recurrences are obtained by taking the root in Lemma 1 as a vertex of degree 3 at one of the two ends of the caterpillar.)

This paper concerns exclusively those trees whose internal vertices have degree at least 3. They are also referred to as *homeomorphically irreducible trees* and *series reduced trees*. Let this family of trees be denoted $\mathcal{T}_3$. It is known [1] that, for trees in general (no restriction on the degrees), the tree of order n that minimizes the average order is the path P_n, in which case $\mu_{P_n} = (n+2)/3$. Therefore, for trees in general, $D_T \geqslant 1/3$, and this is best possible. Examples appear in [1] of sequences $\{T_j\}$ of trees such that $D_{T_j} \to 1$ as $j \to \infty$. So, in the general case, the trivial upper bound $D_T \leqslant 1$ is best possible.

Concerning trees whose internal vertices have degree at least 3, several questions remain open. The first question concerns the upper bound of 3/4 on the density. Consider the ratio $diam(T)/n'(T)$ of the diameter to the number of leaves of a tree T in $\mathcal{T}_3$. The ratio $diam(T)/n'(T)$ is close to 1 if and only if both $n'(T)/n(T)$ is close to 1/2 and $diam(T)/n(T)$ is close to 1/2. Moreover, $n'(T)/n(T)$ is close to 1/2 if and only if the average degree of an internal vertex is close to 3, and $diam(T)/n(T)$ is close to 1/2 if and only if T is "stretched out." This motivates the following question.

Question 1. Is it the case that a sequence $\{T_j\}$ of distinct trees in $\mathcal{T}_3$ satisfies

$$\lim_{j\to\infty} D_{T_j} = \frac{3}{4}$$

if and only if $diam(T_j)/n'(T_j) \to 1$?

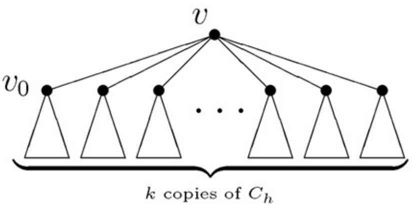

Fig. 3. Joining k copies of C_h.

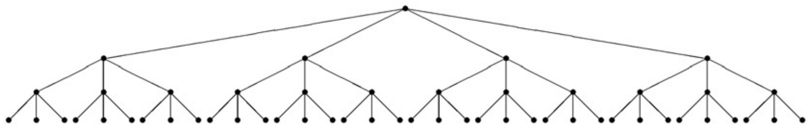

Fig. 4. Complete k-ary tree of height h ($k = 4$ and $h = 3$).

The next two questions concern the lower bound of $1/2$ on the density.

Question 2. For a sequence $\{T_j\}$ of distinct trees in $\mathcal{T}_3$, what are necessary and sufficient conditions for $D_{T_j} \to 1/2$?

Consider the following two conditions on a sequence $\{T_j\}$ of distinct trees in $\mathcal{T}_3$.

1. *Bounded diameter.* There is a number B such that $diam(T) \leqslant B$ for all $T \in \{T_j\}$.
2. *Unbounded degree.* For any number B there is a $T \in \{T_j\}$ and a vertex v of T such that $deg(v) \geqslant B$.

Clearly the bounded diameter condition implies the unbounded degree condition. However, even the bounded diameter condition is not a sufficient condition for $D_{T_j} \to 1/2$ as $j \to \infty$. An example is as follows. Fix integer h, and let C_h be the caterpillar defined above, with v_0 an arbitrarily chosen vertex of C_h. Using k disjoint copies of C_h, let T_k be the tree formed by joining the k copies of the vertex v_0 in the caterpillars to an isolated vertex v. See Fig. 3. It can be shown using Lemma 1 that, for h sufficiently large, $D_{T_k} \nrightarrow 1/2$ as $k \to \infty$. But clearly $\{T_k\}$ has bounded diameter.

There are certainly sequences $\{T_j\}$ of bounded diameter for which $D_{T_j} \to 1/2$ as $j \to \infty$. For example, let h be fixed and let T_k be the complete k-ary tree of height h (see Fig. 4) or simply consider the sequence $\{S_j\}$ of stars on j vertices. Then calculation using Lemma 1 shows that $D_{T_k} \to 1/2$ as $k \to \infty$. This motivates the following question.

Question 3. For a sequence $\{T_j\}$ of trees in $\mathcal{T}_3$, is either the bounded diameter or the unbounded degree condition a necessary condition for $D_{T_j} \to 1/2$?

Moon and Meir [2] determined the average density over all trees of order n to be $1 - e^{-1} \approx .6321$ as $n \to \infty$.

Question 4. What is the limit of the average density over all trees in $\mathcal{T}_3$ of order n as $n \to \infty$?

2. Proof of the lower bound

A tree with no internal vertices of degree 2 satisfies the first of the following inequalities, and if there is at most one internal node of degree 2, then the second

$$2n' \geqslant n + 2, \qquad 2n' \geqslant n + 1. \tag{1}$$

A *root* of a tree is a designated vertex, its *children* the adjacent vertices.

($*$) *It is assumed throughout this section that each internal vertex, with the possible exception of the root, has degree at least* 3. *The root is assumed to have degree at least* 2.

For a tree rooted at vertex v, the number of subtrees containing the root, and the number not containing the root, are denoted $N(v)$ and $\overline{N}(v)$, respectively. Each subset L of leaves of a tree rooted at v determines a unique subtree containing v, i.e., the subtree whose vertices are those with a descendant in L. Thus there is an injection from the set of subsets of leaves into the set of rooted subtrees, giving $N(v) \geqslant 2^{n'}$. Clearly, if the tree is not a star rooted at the center, then the inequality is strict

$$N(v) > 2^{n'}. \tag{2}$$

The average order of a subtree containing the root, and not containing the root, are denoted $\mu(v)$ and $\overline{\mu}(v)$, respectively. The following formula relates the parameters of the unrooted tree to those of the rooted tree

$$\mu_T = \frac{\mu(v)N(v) + \overline{\mu}(v)\overline{N}(v)}{N(v) + \overline{N}(v)}.$$

Letting $D(v) = \mu(v)/n$ and $\overline{D}(v) = \overline{\mu}(v)/n$ denote the densities for subtrees containing and not containing the root, respectively, the above equation becomes

$$D_T = \frac{D(v)N(v) + \overline{D}(v)\overline{N}(v)}{N(v) + \overline{N}(v)}. \tag{3}$$

If $v_1, v_2, \ldots, v_k$ are the children of the root v, let $T_1, T_2, \ldots, T_k$ denote the connected components of $T - v$, considered as trees rooted at $v_1, v_2, \ldots, v_k$, respectively. Let $N_i, \overline{N}_i, \mu_i, \overline{\mu}_i$ denote the number of subtrees of T_i containing the root v_i, not containing the root, the average order of those containing the root, and not containing the root, respectively.

Lemma 1. *The following recursive formulas are valid for a rooted tree*:

$$N(v) = \prod_{i=1}^{k}(N_i + 1), \qquad \mu(v) = 1 + \sum_{i=1}^{k} \mu_i \frac{N_i}{N_i + 1},$$
$$\overline{N}(v) = \sum_{i=1}^{k}(N_i + \overline{N}_i), \qquad \overline{\mu}(v) = \frac{\sum_{i=1}^{k}(\mu_i N_i + \overline{\mu}_i \overline{N}_i)}{\sum_{i=1}^{k}(N_i + \overline{N}_i)}.$$

Proof. The recursion for $\overline{N}(v)$ is clear. The recursion for $N(v)$ comes from the fact that there is an obvious bijection between the set of subtrees rooted at v and the set of k-element subsets $\{t_1, t_2, \ldots, t_k\}$ where t_i is a rooted subtree of T_i – including the empty tree.

Concerning the recursion for $\mu(v)$, let S be the sum of the orders of all rooted subtrees of T. As explained above, each rooted subtree t_i of T_i is contained in exactly $\prod_{j \neq i}(N_j + 1)$ rooted subtrees of T. The subtree t_i therefore contributes $n(t_i)\prod_{j \neq i}(N_j + 1)$ toward the sum S. Hence all the subtrees of T_i combined contribute $\mu_i N_i \prod_{j \neq i}(N_j + 1)$. Summing over all subtrees $T_1, T_2, \ldots, T_k$ gives

$$S = \sum_{i=1}^{k} \mu_i N_i \prod_{j \neq i}(N_j + 1) + N(v),$$

the term $N(v)$ being the contribution from the root v. Hence, using the recursion for $N(v)$,

$$\mu(v) = \frac{\sum_{i=1}^{k} \mu_i N_i \prod_{j \neq i}(N_j + 1) + N(v)}{N(v)} = 1 + \sum_{i=1}^{k} \mu_i \frac{N_i}{N_i + 1}.$$

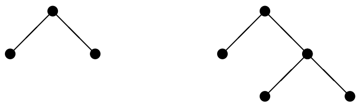

Fig. 5. Figure a (on the left) and b (on the right).

The recursion for $\overline{\mu}(v)$ is clear by splitting the subtrees of the T_i into those that contain and do not contain their root, respectively. $\square$

Lemma 2. *If T is a tree rooted at vertex v, then $N(v) > \overline{N}(v)$.*

Proof. It is easy to verify by induction that, for any positive integers N_i, it is the case that $\prod_{i=1}^{k}(N_i+1) \geqslant 2\sum_{i=1}^{k} N_i$ if $k \geqslant 2$. The first case $k=2$ is equivalent to $(N_1-1)(N_2-1) \geqslant 0$. The induction step from $k-1$ to k is equivalent to $N_k \geqslant 2$.

The statement of the lemma is again proved by induction, on the order of T. It is trivially true for the tree of order 1 (a single vertex). Applying this inequality above, the induction hypothesis, and Lemma 1 to any tree T satisfying our assumption $(*)$ gives

$$N(v) = \prod_{i=1}^{k}(N_i+1) \geqslant 2\sum_{i=1}^{k} N_i > \sum_{i=1}^{k}(N_i+\overline{N}_i) = \overline{N}(v). \qquad \square$$

Lemma 3. *If T is a tree rooted at vertex v then, except for the graph in Fig.* 5b, *we have*

$$D(v) \geqslant \frac{1}{2}\left(1+\frac{1}{n'+1}\right).$$

Proof. Let $T_1,\dots,T_k$ be the subtrees as defined earlier. Let k_1 be the number of such subtrees of order 1, k_2 the number isomorphic to the rooted tree in Fig. 5a, and k_3 the number isomorphic to the rooted tree in Fig. 5b. Let $T_1,\dots,T_{k'}$ be the remaining subtrees. The proof proceeds by induction on the order of the tree T. It is easy to check that the statement is true for the rooted tree of order 1 and the rooted tree of order 3. (The rooted tree of order 2 is not relevant because the root is assumed to have at least 2 children.) Applying the induction hypothesis to the recursion for $\mu(v)$ in Lemma 1 yields the following. The induction hypothesis can be used in the second inequality below because each tree T_i, $i=1,2\dots,k'$, is not the tree in Fig. 5b. Inequality (2) implies $N_i \geqslant 2^{n_i'}+1 \geqslant 2n_i'+2$ for subtrees T_i with $n_i' \geqslant 3$, which is also used

$$\begin{aligned}
\mu(v) &= 1+\sum_{i=1}^{k}\mu_i\frac{N_i}{N_i+1} = 1+\frac{1}{2}k_1+\frac{8}{5}k_2+\frac{31}{11}k_3+\sum_{i=1}^{k'}\mu_i\frac{N_i}{N_i+1}\\
&\geqslant 1+\frac{1}{2}k_1+\frac{8}{5}k_2+\frac{31}{11}k_3+\frac{1}{2}\sum_{i=1}^{k'}\left(1+\frac{1}{n_i'+1}\right)n_i\frac{N_i}{N_i+1}\\
&\geqslant 1+\frac{1}{2}k_1+\frac{8}{5}k_2+\frac{31}{11}k_3+\frac{1}{2}\sum_{i=1}^{k'}\left(1+\frac{1}{n_i'+1}\right)\left(\frac{2n_i'+2}{2n_i'+3}\right)n_i\\
&\geqslant 1+\frac{1}{2}k_1+\frac{8}{5}k_2+\frac{31}{11}k_3+\sum_{i=1}^{k'}\frac{n_i'+2}{2n_i'+3}n_i.
\end{aligned}$$

To prove that $\mu(v) \geqslant \frac{1}{2}(1+\frac{1}{n'+1})n$ it now suffices to prove

$$1+\frac{1}{2}k_1+\frac{8}{5}k_2+\frac{31}{11}k_3+\sum_{i=1}^{k'}\frac{n_i'+2}{2n_i'+3}n_i$$
$$\geqslant \frac{1}{2}\left(1+k_1+3k_2+5k_3+\sum_{i=1}^{k'}n_i+\frac{1+k_1+3k_2+5k_3+\sum_{i=1}^{k'}n_i}{1+k_1+2k_2+3k_3+\sum_{i=1}^{k'}n_i'}\right),$$

which simplifies to

$$1+\frac{1}{5}k_2+\frac{7}{11}k_3+\sum_{i=1}^{k'}\frac{n_i}{2n_i'+3}\geqslant \frac{1+k_1+3k_2+5k_3+\sum_{i=1}^{k'}n_i}{1+k_1+2k_2+3k_3+\sum_{i=1}^{k'}n_i'}. \tag{4}$$

To prove the inequality above, first consider the case $k'=0$. In this case it is sufficient to show that

$$f(k_1,k_2,k_3):=\left(1+\frac{1}{5}k_2+\frac{7}{11}k_3\right)(1+k_1+2k_2+3k_3)-(1+k_1+3k_2+5k_3)\geqslant 0.$$

Since we are assuming throughout this section that the root has degree at least 2, we have $k_1+k_2+k_3\geqslant 2$. It is easy to check that $f(k_1,k_2,k_3)\geqslant 0$ for all values where $k_1+k_2+k_3=2$. Taking partial derivatives reveals that $f(k_1,k_2,k_3)$ is non-decreasing with respect to k_1 for all $k_1,k_2,k_3\geqslant 0$ and non-decreasing with respect to both k_2 and k_3 if either $k_2\geqslant 1$ or $k_3\geqslant 1$. This implies that $f(k_1,k_2,k_3)\geqslant 0$ for all values where $k_1+k_2+k_3\geqslant 2$.

Next consider the case $k'>0$. From the paragraph above, either $f(k_1,k_2,k_3)\geqslant 0$ or $k_1=k_3=0$, $k_2=1$. In either case, a little elementary algebra shows that inequality (4) holds if

$$1+\sum_{i=1}^{k'}\frac{n_i}{2n_i'+3}\geqslant \frac{1+\sum_{i=1}^{k'}n_i}{1+\sum_{i=1}^{k'}n_i'}.$$

Since $n_i\geqslant n_i'+1$, we have $\frac{n_i}{2n_i'+3}>\frac{1}{3}$. Also, since $2n_i'\geqslant n_i+1$ from inequality (1), we have $2>\frac{1+\sum_{i=1}^{k'}n_i}{1+\sum_{i=1}^{k'}n_i'}$. Therefore, if $k'\geqslant 3$, then

$$1+\sum_{i=1}^{k'}\frac{n_i}{2n_i'+3}>1+3(1/3)=2>\frac{1+\sum_{i=1}^{k'}n_i}{1+\sum_{i=1}^{k'}n_i'}.$$

This only leaves the case $k'\leqslant 2$, in which case we need to show

$$1+\frac{n_1}{2n_1'+3}\geqslant\frac{1+n_1}{1+n_1'}\quad\text{and}\quad 1+\frac{n_1}{2n_1'+3}+\frac{n_2}{2n_2'+3}\geqslant\frac{1+n_1+n_2}{1+n_1'+n_2'}.$$

The case $k'=1$ (on the left) simplifies to $n_1'(2n_1'+3)\geqslant n_1(n_1'+2)$, which in turn follows from the fact that $2n_1'\geqslant n_1+1$ from inequality (1). Although algebraically tedious, the case $k'=2$ can similarly be verified using the fact that $2n_i'\geqslant n_i+1$. □

In the tree T, let v_0 be a vertex that maximizes $N(v)$, i.e., $N(v_0)\geqslant N(v)$ for all vertices v. Call v_0 a *maximizing vertex*.

Lemma 4. *If T is a tree rooted at a maximizing vertex v_0, then*

$$N_1\leqslant\prod_{i\neq 1}(N_i+1).$$

Proof. Let $v_1, v_2, \dots, v_k$ be the children of the root v_0. Let $v_0, u_1, u_2, \dots, u_s$ be the vertices adjacent to v_1, and let $N(u_i)$ denote the number of subtrees of $T - v_1$ rooted at u_i. Let N_0 denote the number of subtrees of $T - v_1$ rooted at v_0. By the recursions for $N(v_0)$ in Lemma 1,

$$N(v_0) = (N_1 + 1)\prod_{i\neq 1}(N_i + 1) = N_1\prod_{i\neq 1}(N_i + 1) + \prod_{i\neq 1}(N_i + 1),$$

$$N(v_1) = (N_0 + 1)\prod_{i=1}^{s}\bigl(N(u_i) + 1\bigr) = \Bigl(1 + \prod_{i\neq 1}(N_i + 1)\Bigr)N_1 = N_1\prod_{i\neq 1}(N_i + 1) + N_1.$$

The result now follows from the assumption that $N(v_0) \geqslant N(v_1)$. $\square$

Lemma 5. *If T is a tree rooted at a maximizing vertex v_0 with $deg(v_0) \geqslant 3$, then*

$$\prod_{i=1}^{k}(N_i + 1) > \frac{4}{9}\left(\sum_{i=1}^{k} N_i\right)^2.$$

Proof. Order the children of v_0 so that $N_1 \geqslant N_2 \geqslant \cdots \geqslant N_k$. Consider

$$f(N_1) = \frac{\prod_{i=1}^{k}(N_i + 1)}{(\sum_{i=1}^{k} N_i)^2}$$

as a function of variable N_1 (the other N_i considered as constants). Taking the derivative reveals that the function is increasing if $N_1 \leqslant (\sum_{i\neq 1} N_i) - 2$ and decreasing if $N_1 \geqslant (\sum_{i\neq 1} N_i) - 2$. By the shape of the graph of $f(N_1)$ and by Lemma 4, the function f can attain a minimum only if $N_1 = N_2$ or $N_1 = \prod_{i\neq 1}(N_i + 1)$. It can be checked that $f(N_2) \geqslant f(\prod_{i\neq 1}(N_i + 1))$, so the minimum is attained for $N_1 = \prod_{i\neq 1}(N_i + 1)$. (The inequality is easy to check when $k = 2$. For $k \geqslant 3$, let $P = \prod_{i=3}^{k} N_i$ and $S = \sum_{i=3}^{k} N_i$, and express the inequality in terms of P, S, and $q := N_2$, regarding these variable as real numbers. Holding P fixed, it follows from some algebra that the inequality is true if it is true when all the factors of P and summands of S are equal to q, i.e., $P = q^r$, $S = rq$, where r is not necessarily an integer. Once this substitution is made, additional algebra suffices to verify the inequality.) Therefore

$$\frac{\prod_{i=1}^{k}(N_i + 1)}{(\sum_{i=1}^{k} N_i)^2} = f(N_1) \geqslant f\left(\prod_{i\neq 1}(N_i + 1)\right) = \frac{(\prod_{i\neq 1}(N_i + 1))(\prod_{i\neq 1}(N_i + 1) + 1)}{(\sum_{i\neq 1} N_i + \prod_{i\neq 1}(N_i + 1))^2}.$$

Letting $a = \prod_{i\neq 1}(N_i + 1)$ and $b = \sum_{i\neq 1} N_i$, it is sufficient to show that

$$\frac{a(a+1)}{(a+b)^2} > \frac{4}{9}.$$

By the inequality used in the proof of Lemma 2 we have, in the case $deg(v_0) \geqslant 3$,

$$\frac{b}{a} = \frac{\sum_{i\neq 1} N_i}{\prod_{i\neq 1}(N_i + 1)} \leqslant \frac{1}{2}.$$

Therefore

$$\frac{a(a+1)}{(a+b)^2} \geqslant \frac{a(a+1)}{(a+a/2)^2} = \frac{4}{9}\left(1 + \frac{1}{a}\right) > \frac{4}{9}. \qquad \square$$

Lemma 6. *If T is a tree rooted at a maximizing vertex v_0 with $deg(v_0) \geqslant 3$, then*

$$N(v_0) > \frac{1}{9}\overline{N}(v_0)^2.$$

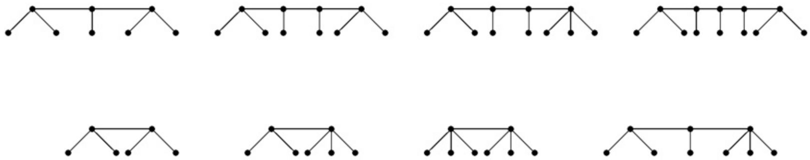

Fig. 6. The exceptions.

Proof. Using Lemmas 1, 2, and 5, we have

$$N(v_0) = \prod_{i=1}^{k}(N_i + 1) > \frac{4}{9}\left(\sum_{i=1}^{k} N_i\right)^2 = \frac{1}{9}\left(\sum_{i=1}^{k}(N_i + N_i)\right)^2$$
$$> \frac{1}{9}\left(\sum_{i=1}^{k}(N_i + \overline{N}_i)\right)^2 = \frac{1}{9}\overline{N}(v_0)^2. \qquad \square$$

Lemma 7. *For a tree T all of whose internal vertices have degree at least three and a maximizing vertex v_0, we have $D_T > \frac{1}{2}$ if any one of the following conditions hold.*

(a) $N(v_0) \geqslant n'\overline{N}(v_0)$;
(b) $\overline{N}(v_0) \geqslant 9n'$;
(c) $N(v_0) \geqslant 9n'^2$.

Proof. By formula (3), $D_T > \frac{1}{2}$ if and only if

$$\big(2D(v_0) - 1\big)N(v_0) > \big(1 - 2\overline{D}(v_0)\big)\overline{N}(v_0).$$

Clearly $\overline{D}(v_0) \geqslant 1/n$. This and Lemma 3 imply that the above inequality holds if

$$N(v_0) > (n' + 1)\left(1 - \frac{2}{n}\right)\overline{N}(v_0).$$

Inequality (1), namely $2n' \geqslant n + 2$, implies that this is the case if

$$N(v_0) \geqslant n'\overline{N}(v_0), \tag{5}$$

which is condition (a) in the statement of the lemma. It is routine to check that a leaf of T cannot be the maximizing vertex. Hence v_0 has degree at least 3 and by applying Lemma 6, the above inequality holds if $\overline{N}(v_0) \geqslant 9n'$, which is statement (b) of the lemma. If, on the other hand, $\overline{N}(v_0) < 9n'$, then condition (a) holds if $N(v_0) \geqslant 9n'^2$, which is condition (c) in the statement of the lemma. $\square$

Proof of the lower bound in Theorem 1. By (2) we have $N(v_0) \geqslant 2^{n'} \geqslant 9n'^2$ for $n' \geqslant 10$. Therefore, by condition (c) in Lemma 7, it is sufficient to verify conditions (a), (b) or (c) for all trees with 9 or fewer leaves. There are relatively few such trees. Moreover, as we systematically constructed the trees according to the number subtrees of order one in $T - v_0$, many possibilities could be quickly eliminated by using Lemmas 2 and 4. Only the star S_4 and the 8 in Fig. 6 failed to satisfy any of the conditions of Lemma 7. The density of each of these was computed and found to be at least $\frac{1}{2}$. $\square$

3. Proof of the upper bound

Again, it is assumed throughout this section that the degrees of all internal vertices of rooted trees are at least 3, except possibly the root.

Proof of the upper bound in Theorem 1. It will be shown that, for any tree T rooted at vertex v, all of whose internal vertices have degree at least three,

(i) $D(v) < \frac{3}{4}$ and
(ii) $\overline{D}(v) < \frac{3}{4}$.

The upper bound then follows immediately from Eq. (3). Statements (i) and (ii) are proved in the next two lemmas. □

Lemma 8. *If T is a tree rooted at v with $deg(v) \geqslant 2$, then*

$$D(v) \leqslant \frac{3}{4}\left(1 - \frac{1}{3n}\right).$$

Proof. Let $T_1, \dots, T_k$ be the subtrees as defined earlier. Let k_1 be the number of such subtrees of order 1, and let $T_1, \dots, T_{k'}$ be the remaining subtrees. The proof proceeds by induction on the order of the tree T. It is easy to check that the statement is true for th rooted tree of order 3. Consider any tree satisfying assumption (∗). Applying the induction hypothesis to the recursion for $\mu(v)$ in Lemma 1 yields the following

$$\begin{aligned}
\mu(v) &= 1 + \sum_{i=1}^{k} \mu_i \frac{N_i}{N_i+1} = 1 + \frac{1}{2}k_1 + \sum_{i=1}^{k'} \mu_i \frac{N_i}{N_i+1} \\
&\leqslant 1 + \frac{1}{2}k_1 + \frac{3}{4}\sum_{i=1}^{k'}\left(1 - \frac{1}{3n_i}\right) n_i \frac{N_i}{N_i+1} \\
&\leqslant 1 + \frac{1}{2}k_1 + \frac{3}{4}\sum_{i=1}^{k'}\left(1 - \frac{1}{3n_i}\right) n_i \\
&= 1 + \frac{1}{2}k_1 + \frac{3}{4}\sum_{i=1}^{k'} n_i - \frac{k'}{4}.
\end{aligned}$$

To prove that $\mu(v) \leqslant \frac{3}{4}(1 - \frac{1}{3n})n = \frac{3}{4}(1 + k_1 + \sum_{i=1}^{k'} n_i) - \frac{1}{4}$ it now suffices to prove

$$1 + \frac{1}{2}k_1 + \frac{3}{4}\sum_{i=1}^{k'} n_i - \frac{k'}{4} \leqslant \frac{3}{4}\left(1 + k_1 + \sum_{i=1}^{k'} n_i\right) - \frac{1}{4},$$

which simplifies to $2 \leqslant k_1 + k'$, which is true because $k_1 + k' = deg(v) \geqslant 2$. □

Lemma 9. *If T is a tree rooted at v with $deg(v) \geqslant 2$, then*

$$\overline{D}(v) < \frac{3}{4}.$$

Proof. Let $T_1, \dots, T_k$ be the subtrees as previously defined. Let k_1 be the number of such subtrees of order 1, and let $T_1, \dots, T_{k'}$ be the remaining subtrees. The proof proceeds by induction on the order of the tree T. It is easy to check that the statement is true for the rooted tree of order 3. Applying Lemma 8 and the induction hypothesis to the recursion for $\overline{\mu}(v)$ in Lemma 1 yields

$$\overline{\mu}(v) = \frac{\sum_{i=1}^{k}(\mu_i N_i + \overline{\mu}_i \overline{N}_i)}{\sum_{i=1}^{k}(N_i + \overline{N}_i)} \quad \Longrightarrow$$

$$\overline{D}(v)=\frac{\sum_{i=1}^{k}(D_in_iN_i+\overline{D}_in_i\overline{N}_i)}{n\sum_{i=1}^{k}(N_i+\overline{N}_i)}=\frac{k_1+\sum_{i=1}^{k'}(D_in_iN_i+\overline{D}_in_i\overline{N}_i)}{(1+k_1+\sum_{i=1}^{k'}n_i)(k_1+\sum_{i=1}^{k'}(N_i+\overline{N}_i))}$$
$$<\frac{k_1+\frac{3}{4}\sum_{i=1}^{k'}n_i(N_i+\overline{N}_i)}{(k_1+\sum_{i=1}^{k'}n_i)(k_1+\sum_{i=1}^{k'}(N_i+\overline{N}_i))}.$$

To prove that $\overline{D}(v)<3/4$ it now suffices to prove

$$\frac{3}{4}\left(\left(k_1+\sum_{i=1}^{k'}n_i\right)\left(k_1+\sum_{i=1}^{k'}(N_i+\overline{N}_i)\right)-\sum_{i=1}^{k'}n_i(N_i+\overline{N}_i)\right)\geqslant k_1.$$

Because

$$\sum_{i=1}^{k'}n_i\sum_{i=1}^{k'}(N_i+\overline{N}_i)\geqslant\sum_{i=1}^{k'}n_i(N_i+\overline{N}_i),$$

it suffices to prove

$$\frac{3}{4}k_1\left(k_1+\sum_{i=1}^{k'}(N_i+\overline{N}_i)\right)\geqslant k_1,$$

which is true because

$$\frac{3}{4}k_1\left(k_1+\sum_{i=1}^{k'}(N_i+\overline{N}_i)\right)\geqslant\frac{3}{4}k_1(k_1+k')=\frac{3}{4}k_1\,deg(v)\geqslant\frac{3}{2}k_1. \qquad \square$$

Acknowledgments

We thank Professor R. Jamison for bringing the conjecture on average subtree order to the attention of one of the authors.

References

[1] R. Jamison, On the average number of nodes in a subtree of a tree, J. Combin. Theory Ser. B 35 (1983) 207–223.
[2] A. Meir, J.W. Moon, On subtrees of certain families of rooted trees, Ars Combin. 16 (1983) 305–318.