On Trees Containing All Small Trees

F.R.K. CHUNG **R.L. GRAHAM***
D. COPPERSMITH

ABSTRACT

In this paper we investigate trees (i.e., connected, acyclic graphs) which contain all trees on n vertices as subgraphs. Let us denote by $u(n)$ the minimum number of vertices such a "universal" tree can have. We prove that:

$$u(n) = n^{(\log n - 2 \log \log n + O(1))/2 \log 2},$$

resolving an earlier conjecture of the authors.

1. Introduction.

A number of papers have appeared recently (e.g., [1], [3], [4], [5], [6]) which investigate questions of the following type: For a given class of graphs $\mathcal{G}$, how large must a graph $U(\mathcal{G})$ be which contains all $G \in \mathcal{G}$ subgraphs? Here, size can be measured by number of vertices or number of edges, subgraphs may or may not be required to be induced, graphs may or may not be directed, etc. (for undefined graph theory terminology, see [7]).

In this note we study this problem in the case that $\mathcal{G}$ is the class of all (unrooted) trees on n vertices and $U(\mathcal{G})$ itself is required to be a tree. Let us denote by $u(n)$ the minimum number of vertices such a "universal" tree can have. Our main result is the following rather sharp estimate for $u(n)$:

$$u(n) = n^{(\log n - 2 \log \log n + O(1))/2 \log 2}. \tag{1}$$

In fact, (1) holds even if $U(n)$ is only required to contain all *binary* (i.e., maximum degree three) trees on n vertices.

This completes the work begun by M. Goldberg and E. Lifshitz [6], who showed that if $r(n)$ is the size of the smallest *rooted* tree containing all rooted trees on n vertices as (rooted) subtrees then $r(n)$ satisfies:

$$r(1) = 1, \quad r(n+1) = 1 + \sum_{k=1}^{n} r\left(\left[\frac{n}{k}\right]\right).$$

In particular, $r(n)$ also satisfies (1).

2. Preliminaries.

By a *uniform* rooted tree we mean a rooted tree in which the number of descendants of each vertex depends only on its level (i.e., distance from the root). We denote by $T(b_0, b_1, \ldots, b_{k-1})$ a tree whose root has b_0 descendants, each vertex at the jth level has b_j descendants, $0 \le j < k$, (where we assume $b_j > 0$) and the leaves are precisely the vertices at the kth level. Such a tree has *height* k. By convention, a tree with one vertex is a uniform tree of height 0, denoted by $T(0)$. Also, we let $v(T)$ denote the number of vertices of T.

The following result is due to Goldberg and Lifshitz [6]. We supply a short proof for ease of reference.

Lemma [*6*]. For a given rooted tree T, the number of nonisomorphic uniform trees contained in T is equal to the number of vertices of T.

Proof. It is enough to prove that for each k, the number of nonisomorphic uniform trees occurring as subtrees of T with height k is equal to the number of vertices of T at level k.

The proof is by induction on k. For $k = 0$, the root corresponds to the unique uniform tree with height 0 to be found as a subtree of T, namely $T(0)$. Assume $k > 0$. For each $j > 0$ let $A(j)$ denote the set of vertices of T at level $k-1$

having at least j descendants each. Let $T'(j)$ be the minimum rooted subtree of T containing $A(j)$. By induction, there are $|A(j)|$ nonisomorphic uniform rooted trees of height $k-1$ occurring as subtrees of $T'(j)$. To each such tree $T(b_0, b_1, \ldots, b_{k-1})$ attach j of the descendants to each vertex at level $k-1$, thus forming a uniform tree $T(b_0, b_1, \ldots, b_{k-1}, j)$ in T. All uniform trees of height k in T can be obtained this way, and each of these occurs exactly once. Thus the number of uniform trees of height k is just $\sum |A(j)|$. But this is the number of descendants of vertices at level $k-1$, i.e., the number of vertices of T at level k and the lemma is proved. ∎

3. The Main Results.

Let us call a rooted tree T a *binary* tree if vertices of T have degree at most 3, except for the root which has degree at most 2.

Lemma 1. The number $b(n)$ of uniform (rooted) binary trees with at most n vertices satisfies

$$b(n) = n^{(\log n - 2 \log \log n + O(1))/2 \log 2} . \tag{4}$$

Proof. Let B be a uniform binary tree having $2k+1$ vertices. Then B is either formed from two identical uniform binary trees, each with k vertices and joined to a common root, or a single uniform binary tree with $2k$ vertices joined to B's root. On the other hand, if B has $2k$ vertices then it can only be formed from a uniform binary tree with $2k-1$ vertices. Thus $b(n)$ satisfies

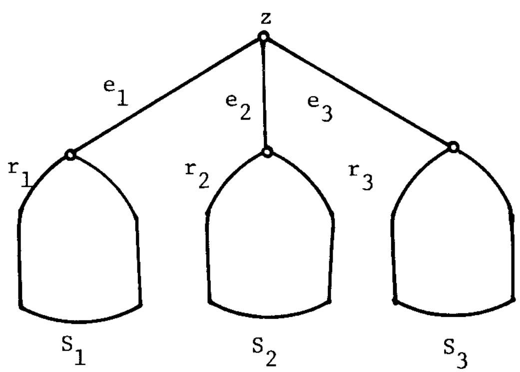
$$\begin{aligned} b(1) &= 1, \\ b(2k+1) &= b(2k) + b(k), \\ b(2k) &= b(2k-1) \qquad k \geq 1. \end{aligned} \tag{5}$$

However, straightforward calculations now show that $b(n)$ satisfies (4). In fact, from the work of Mahler, de Bruijn and others (see [8], [2], [10]) much more is known about the behavior of $b(n)$ and other similarly defined functions. ∎

Theorem 1. Let $U = U(n)$ be an unrooted tree containing as subtrees all unrooted trees on n vertices. Then the number of vertices $v(U)$ of U satisfies

$$v(U) \geq n^{(\log n - 2 \log\log n + O(1))/2 \log 2} . \tag{6}$$

Proof. Among the unrooted trees occurring as subtrees of U are those shown in Figure 1.



A Triple of Uniform Trees (z, S_1, S_2, S_3)
Figure 1.

In (z, S_1, S_2, S_3), each S_i is a rooted uniform binary tree connected at its root r_i to the edge e_i, the other end of which is z. All the S_i are assumed to be different. We order the edges of z so that (z, S_1, S_2, S_3) is distinct from (z, S_2, S_1, S_3) as triples of uniform trees, even though they are identical as unrooted trees; this is a minor point which causes no difficulty.

Let us say that three vertices x_1, x_2, x_3 of T are *noncollinear* if no x_i lies on the path connecting the other two. For noncollinear vertices x_1, x_2, x_3, let $z(x_1, x_2, x_3)$ be the unique vertex belonging to all three paths joining the three pairs of x_i. The number of ordered triples of noncollinear vertices (x_1, x_2, x_3) is less than $v(U)^3$. Thus, letting $P(z,x)$ denote the path connecting z and x, we have

$$
\begin{aligned}
v(U)^3 &\geq \sum_{\substack{(x_1,x_2,x_3)\\ \text{noncollinear}}} 1 = \sum_{z} \sum_{\substack{(e_1,e_2,e_3)\\ \text{edges from } z}} \sum_{\substack{(x_1,x_2,x_3)\\ e_i \in P(z,x_i)}} 1 \\
&= \sum_{z} \sum_{(e_1,e_2,e_3)} v(T_1)\, v(T_2)\, v(T_3)
\end{aligned}
\tag{7}
$$

where T_i is the subtree of T consisting of all t such that $e_i \in P(t,z)$.

Let us consider T_i as a rooted tree whose root is the vertex which together with z forms the edge e_i. Let $R(T_i)$ be the number of nonisomorphic uniform rooted trees S occurring as rooted subtrees of T_i with $v(S) \leq (n-1)/3$. By the Lemma, $R(T_i) \leq v(T_i)$. Thus, from (7)

$$
\begin{aligned}
(v(U))^3 &\geq \sum_{z} \sum_{e_i} \prod_{i} v(T_i) \\
&= \sum_{z} \sum_{e_i} \prod_{i} R(T_i) \\
&\geq \sum_{z} \sum_{e_i} \prod_{i} B(T_i)
\end{aligned}
\tag{8}
$$

where $B(T_i)$ denotes the number of nonisomorphic uniform rooted binary subtrees in T_i.

Now, each triple of nonisomorphic uniform binary trees (z, S_1, S_2, S_3) with $v(S_i) \leq (n-1)/3$ is contained in U and thus, is counted in this sum at least once. (Also the triple (z, S_2, S_1, S_3) will be counted.) The number of such triples is

$$\beta(\beta-1)(\beta-2)$$

where $\beta = b\left(\left[\frac{n-1}{3}\right]\right)$. Furthermore, since the S_i are nonisomorphic, no unrooted tree can arise from more than one triple (z, S_1, S_2, S_3).

Therefore,

$$v(U)^3 \geq \sum_z \sum_{e_i} \prod_i B(T_i) \geq \left(b\left(\left[\frac{n-1}{3}\right]\right) - 2\right)^3 \tag{9}$$

Lemma 1 applied to (9) now implies (6) and the Theorem is proved.

Theorem 2. There exist trees $U^*(n)$ containing all trees on n vertices as subtrees with

$$v(U^*(n)) = n^{(\log n - 2\log\log n + O(1))/2\log 2} .$$

Proof. The construction needed for the proof appears in an earlier paper of Goldberg and Lifshitz [6] and two of the authors [5]. Basically, what is done is the following. Let $G(k)$ denote a (rooted) universal tree for rooted trees with at most k vertices. Then $G(n+1)$ can be constructed from $G\left(\left[\frac{n}{k}\right]\right)$ by attaching the roots of each $G\left(\left[\frac{n}{k}\right]\right)$ to a common "super-root". This leads to the recurrence

$$v(G(n+1)) = 1 + \sum_{k=1} v\left(G\left(\left[\frac{n}{k}\right]\right)\right) \tag{10}$$

which can be shown to have a solution of the form

$$v(G(n)) = n^{(\log n - 2\log\log n + O(1))} . \tag{11}$$

Since $G(n)$ also contains all unrooted trees on n vertices then the theorem is proved. ■

Theorems 1 and 2 can be combined to yield the previous estimate claimed for $u(n)$, namely

$$u(n) = n^{(\log n - 2 \log \log n + O(1))/2 \log 2} .$$

We note that in the same way we can construct a tree $U_d(n)$ having maximum degree $d \geq 3$ which contains every maximum degree d tree on n vertices as a subgraph and which also has $n^{(\log n - 2 \log \log n + O(1))/2 \log 2}$ vertices.

4. Concluding Remarks.

Although the recurrence (10) was previously known to hold for this construction of universal trees, the solution for $v(G(n))$ was only given in the crude form

$$v(G(n)) = n^{(1+O(1))\log n/\log 4}$$

since it was not expected at that time that this would be particularly close to the truth. When the prior work of Goldberg and Lifshitz on the rooted case came to the attention of the authors, it became apparent that it might indeed be closer than we had suspected. In fact, as it turns out, all of the bounds have the same form, differing only by powers of n (coming from the $O(1)$ error terms).

By rather more complicated arguments it can in fact be shown that

$$u(n) = n^{-3/2+o(1)} r(n) .$$

The complete details of the proof of this assertion problem will appear in a later paper. This shows (as one would expect) that rooted universal trees really have to be substantially larger than unrooted ones.

A nice question of this type which very recently has been almost completely resolved concerns so-called "caterpillars", i.e., trees in which the vertices of degree exceeding one induce a path.

If $c(n)$ denotes the minimum number of vertices a caterpillar can have which contains all n-vertex caterpillars as subgraphs, we now know

$$\frac{c_1 n^2}{\log n} < c(n) < c_2 n^2 .$$

Earlier work of this type on caterpillars appears in this *Proceedings* by Kimble and Schwenk [9].

REFERENCES

1. L. Babai, F.R.K. Chung, P. Erdös and R.L. Graham, On graphs which contain all sparse graphs, to appear.

2. N.G. de Bruijn, On Mahler's partition problem, *Nederl. Akad. Wetensch., Proc.*,51 (1948), 659-669 and *Indagationes Math.* 10 (1948), 210-220.

3. F.R.K. Chung and R.L. Graham, On graphs which contain all small trees, *J. Comb. Th.*,24 (1978), 14-23.

4. F.R.K. Chung and R.L. Graham, On universal graphs, *Ann. N.Y. Acad. Sci.*,319 (1979), 136-140.

5. F.R.K. Chung, R.L. Graham and N. Pippenger, On graphs which contain all small trees, II, *Coll. Math. Soc. János Bolyai,* 18 *Combinatorics,* Keszthely (1976), 213-223.

6. M. Goldberg and E. Lifshitz, On minimal universal trees, *Mat. Zametki,* 4 (1968), 371-378 (Russian).

7. F. Hararh, *Graph Theory,* Addison-Wesley, New York (1969).

8. K. Mahler, On a special functional equation, *J. London Math. Soc.,* 15 (1940), 115-122.

9. R.J. Kimble and A.J. Schwenk, On universal caterpillars and trees, this *Proceedings.*

10. A.M. Odlyzko, Periodic Oscillations of coefficients of power series that satisfy functional equations, to appear.

Bell Laboratories and IBM T.J. Watson Research Center