1-1-1976

On Graphs Which Contain All Small Trees, II

F. R. K. Chang
University of California - San Diego

R. L. Graham
University of California - San Diego

Nicholas Pippenger
Harvey Mudd College

COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
18. COMBINATORICS, KESZTHELY (HUNGARY), 1976.

ON GRAPHS WHICH CONTAIN ALL SMALL TREES, II.

F.R.K. CHUNG — R.L. GRAHAM — N. PIPPENGER

INTRODUCTION

Let $\mathscr{T}_n$ denote the class of all trees* with n edges and denote by $s(\mathscr{T}_n)$ the minimum number of edges a graph G can have which contains all $T \in \mathscr{T}_n$ as subgraphs. In a previous paper [2], two of the authors established the following bounds on $s(\mathscr{T}_n)$:

$$(1) \qquad \frac{1}{2} n \log n < s(\mathscr{T}_n) < n^{1 + \frac{1}{\log \log n}}$$

where n is taken sufficiently large. In this note, we strengthen the upper bound on $s(\mathscr{T}_n)$ considerably. In addition we also consider the same problem in the case that G is restricted to be a tree, with $s_{\mathscr{T}}(\mathscr{T}_n)$ denoting the corresponding minimum number of edges. Surprisingly, we show that $s_{\mathscr{T}}(\mathscr{T}_n)$ does *not* grow exponentially in n, answering a question in [2]. It is annoying, however, that at present we cannot even show that $s_{\mathscr{T}}(\mathscr{T}_n)$ must exceed $n^{2+\epsilon}$ for large n.

*The reader may consult [1] or [3] for any undefined graph-theoretic terminology.

W-SUBTREES OF A TREE

Before establishing new bounds on $s(\mathcal{T}_n)$ and $s_{\mathcal{T}}(\mathcal{T}_n)$, we first require a result concerning the decomposition of trees.

Let W be a nonempty set of vertices of a tree T. By a *W-subtree of T*, we mean a subtree T' of T consisting of one of the components C formed from T by the removal of all the vertices of W, except for those vertices of W adjacent to some vertex of C (and the edges joining them).

Example.

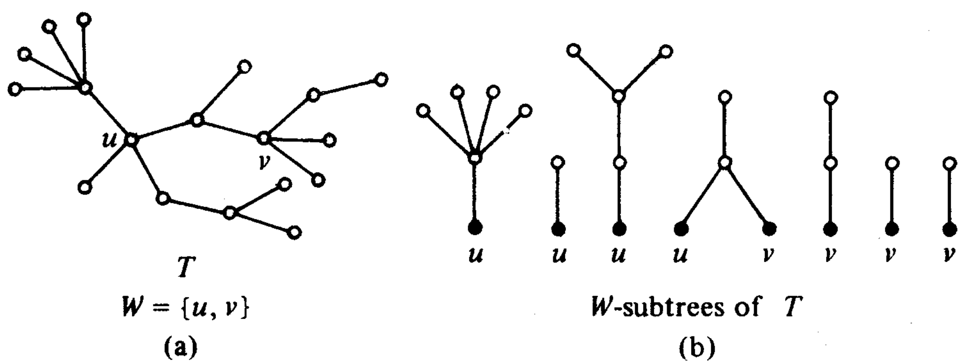

Fig. 1

As usual, we let $\|G\|$ denote the number of edges of a graph G.

Lemma. *Let w be a nonnegative integer. Then if α is sufficiently large, any tree T with at least $\alpha + 1$ edges has a subset of vertices W with $|W| \leqslant w + 1$ so that for some set $\mathscr{C}$ of W-subtrees of T we have*

$$\alpha < \sum_{T' \in \mathscr{C}} \|T'\| \leqslant \left(1 + \left(\tfrac{2}{3}\right)^w\right) \alpha. \tag{2}$$

Proof. For $w = 0$, this is a result in [2]. Assume $w = 1$. We know that if α is large enough then for some vertex u there is a set $\mathscr{C}(u)$ of $\{u\}$-subtrees of T such that

(3) $$\alpha < \sum_{T' \in \mathscr{C}(u)} \| T' \| \leq 2\alpha.$$

If

$$\sum_{T' \in \mathscr{C}(u)} \| T' \| \leq \frac{5}{3} \alpha,$$

then the lemma holds for $w = 1$. Hence, we may assume

$$\frac{5}{3} \alpha < \sum_{T' \in \mathscr{C}(u)} \| T' \| \leq 2\alpha.$$

Let T_1 be the subtree of T formed by taking the union of all $T' \in \mathscr{C}(u)$. Again, for α sufficiently large, there exists a vertex v of T_1 so that for some set $\mathscr{C}(v)$ of (v)-subtrees of T_1, we have

$$\frac{\alpha}{3} \sum_{T'' \in \mathscr{C}(v)} \| T'' \| \leq \frac{2\alpha}{3}.$$

Consider the set $\mathscr{C}'(v)$ all of $\{v\}$-subtrees of T_1 which are *not* in $\mathscr{C}(v)$. Then

$$\alpha < \sum_{T' \in \mathscr{C}'(v)} \| T' \| \leq \frac{5}{3} \alpha.$$

However, a $\{v\}$-subtree of T_1 is a $\{u, v\}$-subtree of T. This proves the lemma for the case $w = 1$. The inductive proof of (2) for general w follows very similar lines and will not be given.∎

AN UPPER BOUND ON $s(\mathscr{T}_n)$

Theorem 1.

$$s(\mathscr{T}_n) = O(n \log n \,(\log \log n)^2).$$

Proof. For $p \geq 0$, let us define the graph $G_{w,p}$ as follows. $G_{w,0} = K_{w+1}$, the complete graph on $w + 1$ vertices. For $p > 0$, $G_{w,p}$ will denote the graph formed from K_{w+1} and two disjoint copies of $G_{w,p-1}$, by placing an edge between each vertex of K_{w+1} and each vertex of each of the copies of $G_{w,p-1}$ (see Figure 2).

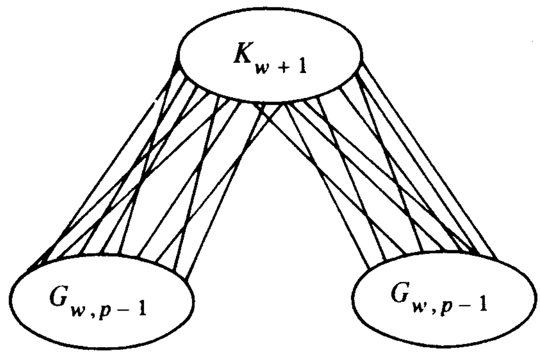

Fig. 2

Simple inductive arguments show that $|G_{w,p}| = O(w2^p)$ and $\|G_{w,p}\| = O(w^2 p 2^p)$ (where $|G|$ denotes the number of vertices in G). It is also not difficult to see that $G_{w,p}$ contains all trees with at most

$$\left(\frac{2+\left(\frac{2}{3}\right)^w}{1+\left(\frac{2}{3}\right)^w}\right)^p$$

edges. For $p = 1$, the expression is less than 2 and the claim is trivial. For $p > 1$, application of the preceding Lemma with

$$\alpha = \frac{1}{1+\left(\frac{2}{3}\right)^w}\left(\frac{2+\left(\frac{2}{3}\right)^w}{1+\left(\frac{2}{3}\right)^w}\right)^{p-1},$$

guarantees a set W of $w+1$ vertices (which may be assigned to the vertices of K_{w+1} in $G_{w,p}$) and a decomposition of the W-subtrees into two classes, each having at most

$$\left(\frac{2+\left(\frac{2}{3}\right)^w}{1+\left(\frac{2}{3}\right)^w}\right)^{p-1}$$

edges (which may be assigned to the two copies of $G_{w,p-1}$ in $G_{w,p}$).

If we now choose $q = \left\lceil \frac{\log 2n}{\log 2} \right\rceil$ and $w = \left\lceil \frac{\log q}{\log \frac{3}{2}} \right\rceil$ we find that

$$\| G_{w,q} \| = O(n \log n \, (\log \log n)^2).$$

Furthermore, a simple calculation shows that

$$\left(\frac{2 + \left(\frac{2}{3}\right)^w}{1 + \left(\frac{2}{3}\right)^w} \right)^q \geqslant 2^q \left(1 - \frac{1}{2} \left(\frac{2}{3}\right)^w \right)^q \geqslant 2^{q-1} \geqslant n,$$

so that $G_{w,q}$ contains as subgraphs all trees with at most n edges. ∎

TREES CONTAINING ALL SMALL TREES

We next turn our attention to the case in which G is restricted to be a tree. As mentioned in the introduction, it was asked in [2] whether or not $s_{\mathscr{T}}(\mathscr{T}_n)$, the corresponding minimum number of edges in this case, must grow exponentially in n. This is settled by Theorem 2.

Before presenting this result, we first list the values of $s\ (\quad_n)$ for $n \leqslant 7$. We also show trees which produce these values (see Fig. 3). The corresponding proofs for these results are straightforward (using degree sequence considerations) and are omitted.

n	$s_{\mathscr{T}}(\mathscr{T}_n)$
1	1
2	2
3	4
4	6
5	9
6	13
7	17

Table 1

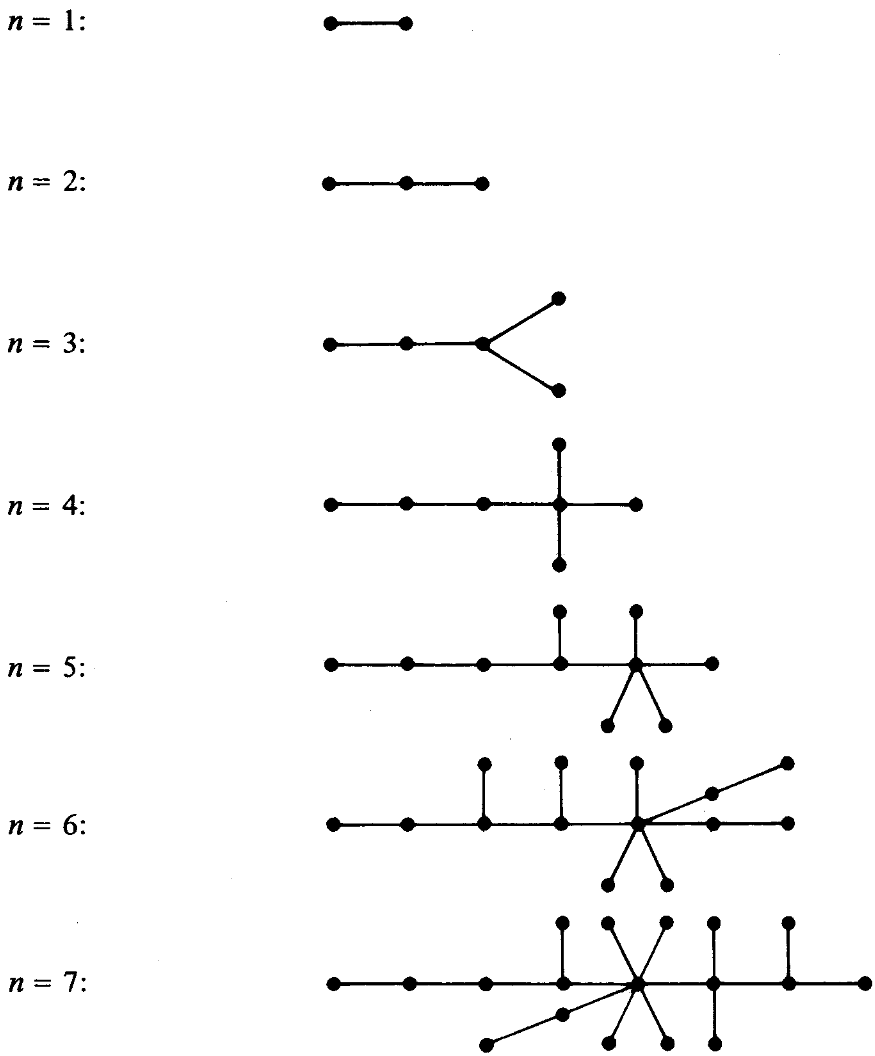

Fig. 3

Theorem 2.

$$s_{\mathscr{T}}(\mathscr{T}_n) \leqslant \frac{2\sqrt{2}}{n} \exp \frac{\log^2 n}{2 \log 2}$$

for n sufficiently large.

Proof. Let us consider a family of rooted trees $\bar{G}(x)$ with a root at some vertex of degree 1 which contains as subgraphs all rooted trees on at most x edges which have a root at some vertex of degree 1. For $1 \leq k < n$, let $\bar{G}\,\frac{n-1}{k}$ have as its root r_k. Form the graph $\bar{G}(n)$ (as shown in Fig. 4) by identifying all the r_k as a single vertex r^* and adjoining a root r_n of degree 1 to r^*. We note that $\bar{G}(x) = \bar{G}(n)$ where n is the integral part of x.

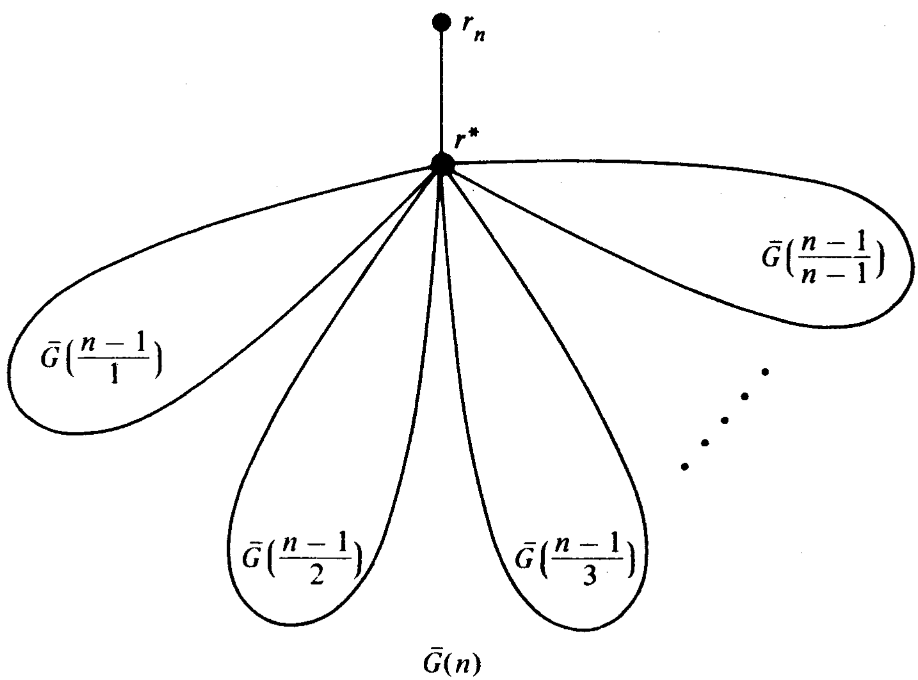

Fig. 4

It is easy to see that if $\bar{f}$ satisfies

(4) $$\bar{f}(x) \geq \sum_{k=1}^{[x]} \bar{f}\left(\frac{x-1}{k}\right),$$

for sufficiently large x then

(5) $$\| \bar{G}(n) \| \leq \bar{f}(n).$$

We claim that it will suffice to have $\bar{f}$ satisfy

$$(6) \qquad \bar{f}(x) \geqslant \bar{f}(x-1) + 2\bar{f}\left(\frac{x+1}{2}\right)$$

in order for (4) to hold. For (6) implies

$$\begin{aligned} \bar{f}(x) &\geqslant \bar{f}(x-1) + 2\bar{f}\left(\frac{x-1}{2}\right) \geqslant \\ &\geqslant \bar{f}(x-1) + 2\bar{f}\left(\frac{x-1}{2}\right) + 4\bar{f}\left(\frac{x-3}{4}\right) \geqslant \\ &\geqslant \bar{f}(x-1) + 2\bar{f}\left(\frac{x-1}{2}\right) + 4\bar{f}\left(\frac{x-1}{4}\right) + 8\bar{f}\left(\frac{x+7}{8}\right) \geqslant \\ &\vdots \\ &\geqslant \sum_{2^k < x} 2^{k-1}\bar{f}\left(\frac{x-1}{2^k}\right) \geqslant \sum_{k=1}^{[x]} \bar{f}\left(\frac{x-1}{k}\right). \end{aligned}$$

A straightforward computation now shows that the choice

$$\bar{f}(x) = e^{\frac{\log^2 x}{2\log 2}}$$

satisfies (6) for x sufficiently large.

Let $G(x)$ be a graph as shown in Figure 5.

It is immediate that $G(x)$ contains all $T \in \mathscr{T}_n$ as subgraphs and we have

$$s_{\mathscr{T}}(\mathscr{T}_n) \leqslant \| G(x) \| \leqslant \frac{2\sqrt{2}}{n} \cdot \exp\left(\frac{(\log n)^2}{2\log 2}\right).$$

This proves the theorem.∎

Let $s^*_{\mathscr{T}}(\mathscr{T}_n)$ be the minimum number of edges a *rooted* tree can have which contains all *rooted* trees of n edges as subgraphs. Of course, the inequality

$$s_{\mathscr{T}}(\mathscr{T}_n) \leqslant s^*_{\mathscr{T}}(\mathscr{T}_n)$$

is immediate. In fact, we now show that if $s_{\mathscr{T}}(\mathscr{T}_n)$ grows polynomially in n, then so does $s^*_{\mathscr{T}}(\mathscr{T}_n)$.

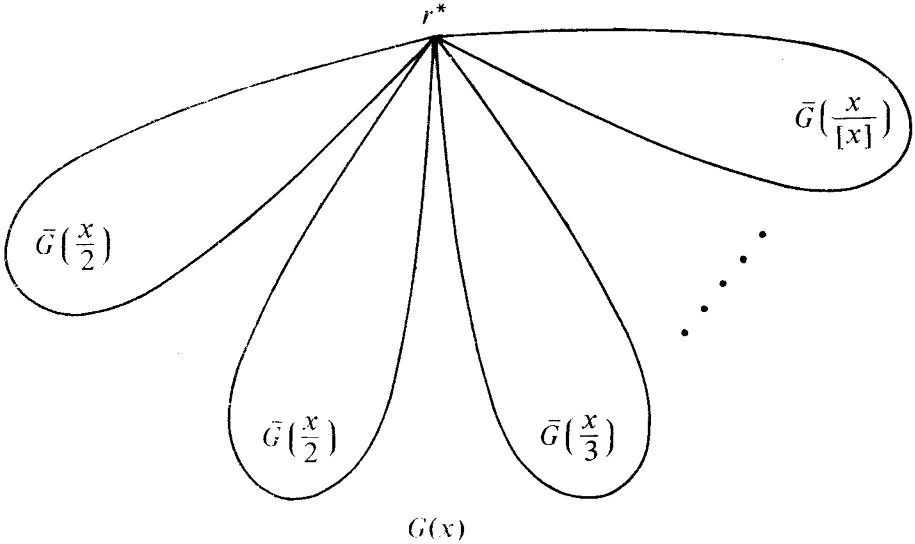

Fig. 5

Theorem 3.

$$s_{\mathscr{T}}(\mathscr{T}_n) \leqslant s_{\mathscr{T}}^*(\mathscr{T}_n) \leqslant s_{\mathscr{T}}(\mathscr{T}_n) \cdot (s_{\mathscr{T}}(\mathscr{T}_n) + 1).$$

Proof. Let G_n be a tree with $s_{\mathscr{T}}(\mathscr{T}_n)$ edges which contains all $T \in \mathscr{T}_n$ as subgraphs. Let $G_n(v)$, $v \in G_n$, be a rooted tree which has the same structure as G_n and which has v as its root. Now, form the rooted tree H_n (as shown in Fig. 6) by identifying all the roots v in $G_n(v)$ for $v \in G_n$.

It is easily verified that H_n contains all rooted trees with n edges and satisfies

$$s_{\mathscr{T}}^*(\mathscr{T}_n) \leqslant \| H_n \| \leqslant s_{\mathscr{T}}(\mathscr{T}_n)(s_{\mathscr{T}}(\mathscr{T}_n) + 1).$$

This proves the theorem.∎

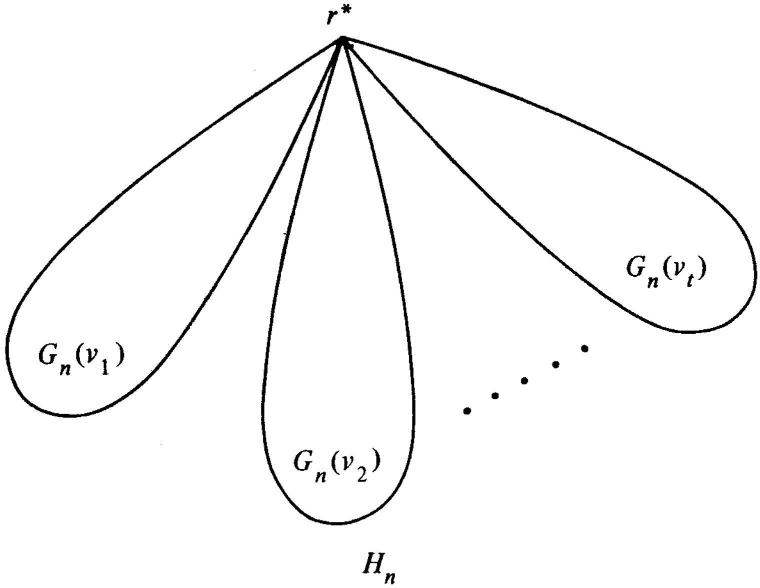

Fig. 6

CONCLUDING REMARKS

As remarked earlier, the best known lower bound for $s(\mathcal{T}_n)$ is $\frac{1}{2}\, n \log n$ which is not too far from the upper bound of $O(n \log n\, (\log\log n)^2)$ of Theorem 1. Perhaps the lower bound is the correct order of magnitude. Unfortunately, the only lower bound presently known for $s_{\mathcal{T}}(\mathcal{T}_n)$ is rather weak. By considering the possible locations of the vertices of degree 1 of the $T \in \mathcal{T}_n$, it can be argued that

$$s_{\mathcal{T}}(\mathcal{T}_n) > cn^2$$

for some $c > 0$. It seems likely that

$$\frac{s_{\mathcal{T}}(\mathcal{T}_n)}{n^k} \to \infty$$

for any fixed k.

REFERENCES

[1] C. Berge, *Graphs and hypergraphs,* North-Holland, Amsterdam, 1973.

[2] F.R.K. Chung – R.L. Graham, On graphs which contain all small trees (to appear in *Jour. Comb. Th.*, (B)).

[3] F. Harary, *Graph theory,* Addison-Wesley, New York, 1969.

F.R.K. Chung – R.L. Graham

Bell Laboratories, Murray Hill, New Jersey.

N. Pippenger

IBM Thomas J. Watson Research Center, Yorktown Heights, New York.