Reprinted from JOURNAL OF COMBINATORIAL THEORY, Series B
All Rights Reserved by Academic Press, New York and London

Vol. 24, No. 1, February 1978
Printed in Belgium

On Graphs which Contain All Small Trees

F. R. K. CHUNG AND R. L. GRAHAM

Bell Laboratories, Murray Hill, New Jersey 07974

Communicated by the Managing Editors

We investigate those graphs G_n with the property that any tree on n vertices occurs as subgraph of G_n. In particular, we consider the problem of estimating the minimum number of edges such a graph can have. We show that this number is bounded below and above by $\frac{1}{2}n \log n$ and $n^{1+1/\log \log n}$, respectively.

INTRODUCTION

A typical question in extremal graph theory[1] is one which asks for the maximum number of edges a graph G on n vertices can have so that G does not contain some given graph H (or class of graphs $\mathscr{H}$) as a subgraph. Perhaps the most well-known result of this type is the theorem of Turán [9, 10] which asserts that if $H = K_m$, the complete graph on m vertices, then this maximum number is just

$$\frac{(m-2)}{2(m-1)}(n^2 - r^2) + \binom{r}{2}$$

where r is the unique integer satisfying

$$r \equiv n (\mathrm{mod}\ m-1) \qquad \text{and} \qquad 1 \leqslant r \leqslant m-1.$$

In general, let $t(H; n)$ (or $t(\mathscr{H}; n)$) denote this maximum number of edges when the forbidden subgraph is the graph H (or class of graphs $\mathscr{H}$). The preceding result can be stated as

$$t(K_m; n) = \frac{m-2}{2(m-1)}(n^2 - r^2) + \binom{r}{2}, \tag{1}$$

where r is defined as before.

[1] For any undefined terminology, see [6].

0095-8956/78/0241-0014$02.00/0

Numerous other results along these lines are available, although usually only estimates for $t(H; n)$ are known, as opposed to exact values. For example:

(i) If C_4 denotes the cycle of length 4, then it has been shown [4, 7] that

$$t(C_4 ; n) \sim \tfrac{1}{2} n^{3/2}.$$

(ii) If $K_{3,3}$ denotes the complete bipartite graph with vertex sets of sizes 3 and 3, then Brown [3] and others [7] have proved

$$cn^{5/3} < t(K_{3,3} ; n) \leqslant (2^{1/3} n^{5/3} + 3n)/2.$$

(where $c, c_1, c_2, \ldots,$ will hereafter denote suitable positive constants).

(iii) If C_{2m} denotes the cycle of length $2m$, then it has been shown by Erdös [5] and Bondy and Simonovits [2] that

$$\frac{cn \log n}{\log \log n} < t(C_{2m} ; n) < c_m n^{1+1/m}.$$

An interesting old (and apparently difficult) conjecture of Erdös and Sós [5] asserts that for $\mathscr{T}_m$, the class of all trees with m edges:

Conjecture. $t(\mathscr{T}_m ; n) = [(m-1)\, n/2]$.

In this paper we consider the *complementary* extremal problem. That is, for a given class $\mathscr{H}$, what is the *least* number $s(\mathscr{H}; n)$ of edges a graph G on n vertices can have so that *all* $H \in \mathscr{H}$ are subgraphs of G? Also of interest to us will be the quantity $s(\mathscr{H})$, defined to be the least number of edges *any* graph G (with no restriction on its number of vertices) can have so that all $H \in \mathscr{H}$ are subgraphs of G.

In contrast to the situation for $t(\mathscr{H}; n)$, very few results for $s(\mathscr{H}; n)$ or $s(\mathscr{H})$ are known. It has been shown by Bondy [1] that

$$n - 1 + \frac{\log(n-1)}{\log 2} < s(\mathscr{C}_n ; n) < n + \frac{\log n}{\log 2} + H(n) + O(1),$$

where $H(n)$ denotes $\min\{k\colon \overbrace{\log \log \cdots \log}^{k} n < 2\}$ and $\mathscr{C}_n$ denotes the set of all cycles of length at most n.

Our results deal almost exclusively with the case in which $\mathscr{H}$ is $\mathscr{T}_n$, the set of all trees with n edges. In particular, we show for all sufficiently large n,

$$\tfrac{1}{2} n \log n < s(\mathscr{T}_n) < n^{1+1/\log \log n}. \tag{2}$$

From the definitions it is clear that

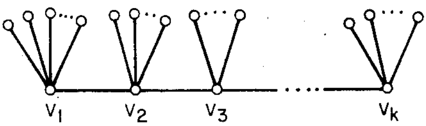
$$s(\mathscr{T}_n) \leqslant s(\mathscr{T}_n ; n+1). \tag{3}$$

It is not known whether, in fact, (3) always holds with *equality*.

A Lower Bound

By the *degree sequence* of a graph G, denoted by ds(G), we mean the *nonincreasing* sequence $(d_1, d_2, \ldots, d_t)$ formed from the set of *degrees* of the vertices of G (where, as usual, the degree of a vertex v is the number of edges incident to v). It is not difficult to see that if H is a subgraph of G then for any j, the jth component of ds(H) is less than or equal to the jth component of ds(G).

THEOREM 1. $s(\mathscr{T}_n) > \frac{1}{2} n \log n$.



T(k)

FIGURE 1

Proof. Let G be a graph containing as subgraphs all trees T in n edges. Now, for each k, $1 \leqslant k \leqslant n+1$, there exists a tree $T(k) \in \mathscr{T}_n$ so that

$$\text{ds}(T(k)) = (d_1^{(k)}, d_2^{(k)}, \ldots, d_{n+1}^{(k)})$$

with $d_k^{(k)} \geqslant n/k$. To see this, consider the tree $T(k)$ shown in Fig. 1.

If $d(j)$ denotes the degree of vertex v_j then

$$\sum_{j=1}^{k} d(j) = n + k - 1.$$

Hence, by distributing the edges as uniformly as possible we can guarantee that

$$d(j) \geqslant \left[\frac{n+k-1}{k}\right] \geqslant \frac{n}{k} \qquad \text{for} \quad 1 \leqslant j \leqslant k.$$

Thus, the kth term of ds($T(k)$) must be as large as n/k. If ds(G) $= (d_1, d_2, \ldots, d_t)$, then by the previous observation,

$$d_k \geqslant n/k \qquad \text{for} \quad 1 \leqslant k \leqslant n+1. \tag{4}$$

However, the number of edges of G, denoted by $\| G \|$, satisfies

$$\| G \| = \tfrac{1}{2} \sum_{k=1}^{t} d_k$$
$$\geqslant \tfrac{1}{2} \sum_{k=1}^{n+1} n/k = \tfrac{1}{2} n \log n + \gamma n + O(1)$$

as $n \to \infty$, where γ denotes Euler's constant. This proves the theorem. ∎

An Upper Bound

In this section we establish the upper bound of (2). Before doing so, we first require a preliminary result. For a vertex v of a tree T, a subtree T' of T consisting of one of the components C formed from T by the removal of v, together with v and the edge joining it to C, is called a *v-subtree* of T. For example, if T is the tree shown in Fig. 2a, then the v-subtrees of T are shown in Fig. 2b.

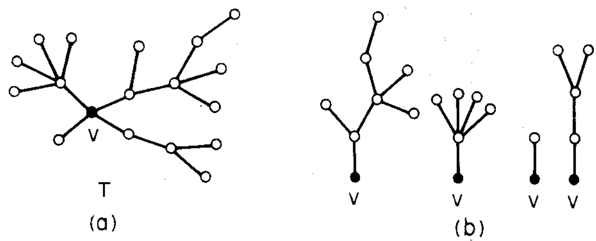


FIG. 2. Example of v-subtrees.

LEMMA. *Suppose T is a tree with at least $k+1$ edges. Then for some vertex v of T, there is a set $C(v)$ of v-subtrees of T so that*

$$k + 1 \leqslant \sum_{T' \in C(v)} \| T' \| \leqslant 2k. \tag{5}$$

Proof. If $\| T \| = k + 1$, the result is immediate. Hence we may assume $\| T \| \geqslant k + 2$. Choose a vertex v_0 of degree 1 and let $\{v_0, v_1\}$ denote the edge incident to v_0. Consider the set $\mathscr{C}(v_1)$ of v_1-subtrees of T not containing v_0. Thus, we have

$$\sum_{T' \in \mathscr{C}(v_2)} \| T' \| \geqslant k + 1.$$

If all $T' \in \mathscr{C}(v_1)$ satisfy $\| T' \| \leqslant k$, then by taking increasing unions of the elements of $\mathscr{C}(v_1)$, it is readily seen that a set $C(v) = C(v_1)$ can be formed which satisfies (5). Thus, we may assume that $\| T_1 \| \geqslant k + 1$ for some $T_1 \in \mathscr{C}(v_1)$. If $\| T_1 \| = k + 1$, then the result is immediate. Hence we may further assume $\| T_1 \| \geqslant k + 2$. Let $\{v_1, v_2\}$ be the edge of T_1 containing v_1 and let $\mathscr{C}(v_2)$ denote the set of v_2-subtrees of T not containing v_1. Thus,

$$\sum_{T' \in \mathscr{C}(v_2)} \| T' \| \geqslant k + 1.$$

As before, if all $T' \in \mathscr{C}(v_2)$ satisfy $\| T' \| \leqslant k$ then it is easily seen that (5) can be satisfied. On the other hand, if some $T' \in \mathscr{C}(v_2)$ has $\| T' \| \geqslant k + 1$, then we let $\{v_2, v_3\}$ denote the edge of T' incident to v_3 and we consider the

v_3-subtrees of T not containing v_2, etc. By continuing in this manner, the lemma follows by induction. ▮

THEOREM 2.

$$s(\mathscr{T}_n) < n^{1+1/\log\log n} \tag{6}$$

for all sufficiently large n.

Proof. Let $\mathscr{S}(\mathscr{T}_n)$ denote the class of all graphs which contain all $T \in \mathscr{T}_n$ as subgraphs. The key to the proof of (6) is the following construction. Suppose $G_1 \in \mathscr{S}(\mathscr{T}_{2k-1})$ and $G_2 \in \mathscr{S}(\mathscr{T}_{n-k-2})$ for some k. Form the graph G by joining a new vertex x to all the vertices in G_1 and G_2 (which we assume are disjoint). It follows at once from the lemma that $G \in \mathscr{S}(\mathscr{T}_n)$.

If $|H|$ denotes the number of vertices of a graph H then we have in the above construction

$$\begin{aligned} |G| &= |G_1| + |G_2| + 1, \\ \|G\| &= \|G_1\| + |G_1| + \|G_2\| + |G_2|. \end{aligned} \tag{7}$$

The general plan is to construct graphs $G(n)$ in $\mathscr{S}(\mathscr{T}_n)$ for large n by applying the preceding construction recursively. However, at each stage the choice of an appropriate k must be made. On one hand, if k is chosen too large, e.g., k is always chosen to be a fixed proportion α of n, then we find that $\|G(n)\|$ will grow faster than n^β, where $\beta = \beta(\alpha) > 1$ and $\beta(\alpha) \to 1$ as $\alpha \to 0$. On the other hand, if k is chosen too small, e.g., so that k is constant, then the corresponding $G(n)$ will have $\|G(n)\|$ growing like $c'n^2$. What we use is something in between, namely, we choose $k = k(n)$ to be about $n/\log n$.

For a large fixed constant c (to be specified later), let $P(x)$ denote the following assertion:

There exists a graph $G(x) \in \mathscr{S}(\mathscr{T}_x)$ satisfying:

(i) $|G(x)| < cx^{1+0.8/\log\log x} = v(x)$;

(ii) $\|G(x)\| < cx^{1+0.9/\log\log x} = f(x)$.

In order to prove (6), it will suffice to show that $P(x)$ holds for all $x \geq 2$. This, however, will follow from the inequalities

$$v(2x/\log x) + v(x - x/\log x) \leq v(x); \tag{8}$$

$$v(x) + f(2x/\log x) + f(x - x/\log x) \leq f(x). \tag{9}$$

For, if (8) and (9) hold, then by (7), the monotonicity of $v(x)$ and $f(x)$ and the fact that $v(x) \geq x$, we obtain by the preceding construction a graph $G(x) \in \mathscr{S}(\mathscr{T}_x)$ satisfying (i) and (ii).

Rather than grind out the rather straightforward proofs of (8) and (9) in detail, we limit ourselves to sketching (8) ((9) is similar).

In the following sequence of inequalities, each one is implied by the following one (always for sufficiently large x).

$$x^{1+0.8/\log\log x} - \left(x - \frac{x}{\log x}\right)^{1+0.8/\log\log(x-x/\log x)}$$

$$\geqslant \left(\frac{2x}{\log x}\right)^{1+0.8/\log\log(2x/\log x)}, \tag{10}$$

$$\exp\left(\frac{0.8\log x}{\log\log x}\right) - \left(1 - \frac{1}{\log x}\right)\exp\left(\frac{0.8\log(x - x/\log x)}{\log\log(x - x/\log x)}\right)$$

$$\geqslant \frac{2}{\log x}\exp\left(\frac{0.8\log(2x/\log x)}{\log\log(2x/\log x)}\right),$$

$$\exp\left(\frac{0.8\log(x - x/\log x)}{\log\log(x - x/\log x)}\right) \geqslant 2\exp\left(\frac{0.8\log(2x/\log x)}{\log\log(2x/\log x)}\right), \tag{11}$$

$$\frac{0.8(\log x + \log(1 - 1/\log x))}{\log\log x}$$

$$\geqslant \log 2 + \frac{0.8(\log 2 + \log x - \log\log x)}{\log\log(2x/\log x)}.$$

But

$$\log x\left(\frac{1}{\log\log(2x/\log x)} - \frac{1}{\log\log x}\right) \sim \frac{1}{\log\log x}$$

as $x \to \infty$ so that (11) is valid provided we have

$$\frac{\log(1 - 1/\log x)}{\log\log x} \geqslant \frac{\log 2}{0.8} + \frac{2 + \log 2 - \log\log x}{1.1\log\log x},$$

$$\left(1 - \frac{1.1\log 2}{0.8}\right)\log\log x \geqslant 2 + \log 2 - 1.1\log\left(1 - \frac{1}{\log x}\right),$$

which clearly holds for large x since the coefficient of $\log\log x$ is positive.

For a suitable fixed m (chosen large enough so that the preceding approximations are valid), $P(x)$ is clearly valid for all x satisfying $2 \leqslant x \leqslant m$ (by an appropriate choice of c). Thus, by induction $P(x)$ holds for all $x \geqslant 2$. This proves the theorem. ∎

In exactly the same way, one can show that for any $\epsilon > 0$,

$$s(\mathscr{T}_n) < n^{1+\log(2+\epsilon)/\log\log n} \tag{12}$$

for all sufficiently large n.

Some Exact Values

In Table I, we list the values of $s(\mathscr{T}_n)$ for some small values of n. We also list the best bound $\| G^*(n)\|$ on $s(\mathscr{T}_n)$ which we can obtain using the construction of the preceding section (i.e., we optimize the values of k in the

TABLE I

Some Exact Values of $s(\mathscr{T}_n)$

n	$s(\mathscr{T}_n)$	$\| G^*(n) \|$
0	0	0
1	1	1
2	2	2
3	4	4
4	6	6
5	8	8
6	11	11
7	13	14
8	?	17
9	?	21
10	?	25

construction of a suitable $G^*(n)$). The values for $s(\mathscr{T}_0)$ and $s(\mathscr{T}_1)$ are obvious. To establish the values of $s(\mathscr{T}_n)$, $2 \leqslant n \leqslant 5$, it suffices to note that any graph $G \in \mathscr{S}(\mathscr{T}_n)$ must contain a path of length n and a vertex of degree n. This forces $\| G \| \geqslant 2n - 2$ and this lower bound is achieved by the graphs $G^*(n)$ shown in Fig. 3.

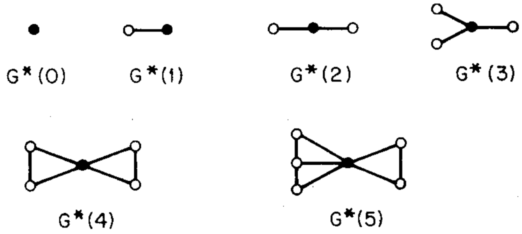


FIG. 3. Optimal elements of $\mathscr{S}(\mathscr{T}_n)$, $n \leqslant 5$.

For $n = 6$, suppose $G \in \mathscr{S}(\mathscr{T}_6)$ with $\| G \| = 10$. Since G contains a path of length 6 and a vertex v^* of degree 6, there are only three possibilities, shown in Fig. 4. However, the two trees T_1 and T_2 shown in Fig. 5 must also be subgraphs of G. But the only possibility for T_1 in Fig. 4 is (*c*) (shown in Fig. 6) and this case is impossible for T_2 ! Thus, $\| G \| \geqslant 11$ for $G \in \mathscr{S}(\mathscr{T}_6)$.

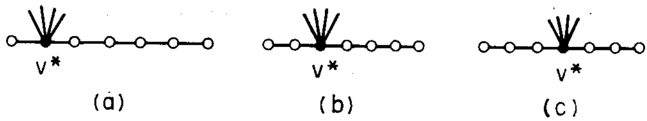


FIG. 4. Possibilities for $G \in \mathcal{S}(\mathcal{T}_6)$ with $\| G \| = 10$.

FIGURE 5

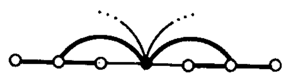


FIGURE 6

For $n = 7$, the obvious lower bound of 12 can be increased to 13 by noting that any $G \in \mathcal{S}(\mathcal{T}_7)$ must contain the tree with degree sequence (4, 4, 1, 1, 1, 1, 1, 1). Two graphs which achieve these lower bounds are shown in Fig. 7. Note that all examples given thus far for optimal elements of $\mathcal{S}(\mathcal{T}_n)$ have

FIG. 7. Optimal elements of $\mathcal{S}(\mathcal{T}_n)$, $n = 6, 7$.

just $n + 1$ vertices, the minimum number possible. Whether this always happens is not known.

SOME RELATED QUESTIONS

The preceding results suggest a number of related problems, several of which we now mention.

1. Is $s(\mathcal{T}_n ; n + 1) = s(\mathcal{T}_n)$?

2. Is $s(\mathcal{T}_n) = O(n \log n)$?

3. What can be said about $s(\mathcal{H})$ for other classes $\mathcal{H}$? For example, if $\mathcal{H} = \mathcal{G}_n$, the class of *all* graphs with n edges, is $s(\mathcal{G}_n) = o(n^{1+\epsilon})$ for all $\epsilon > 0$? Of course, the same questions make sense when we replace the word "edges" by "vertices," both in the definitions of $s(\mathcal{H})$ and/or in $\mathcal{G}_n$.

4. We could define $s_{\mathscr{K}}(\mathscr{H})$ for classes of graphs $\mathscr{K}$ and $\mathscr{H}$ by

$$s_{\mathscr{K}}(\mathscr{H}) = \min\{\| K \|: K \in \mathscr{S}(\mathscr{H}) \cap \mathscr{K}\}.$$

An interesting example of this is the case $\mathscr{K} = \mathscr{T}$, the set of *all* finite trees. Does $s_{\mathscr{T}}(\mathscr{T}_n)$ grow faster than any power of n?

5. Suppose we define $\mathscr{S}^*(\mathscr{H})$ by

$$\mathscr{S}^*(\mathscr{H}) = \{G\text{: Each } H \in \mathscr{H} \text{ occurs as an \textit{induced} subgraph of } G\}.$$

What is the behavior of $s^*(\mathscr{H}) = \min\{\| G \|: G \in \mathscr{S}^*(\mathscr{H})\}$? One would generally expect $s^*(\mathscr{H})$ to be much larger than $s(\mathscr{H})$ as $\mathscr{H}$ becomes large. It has been shown by Moon [8] that for $\mathscr{H} = \mathscr{G}_n'$, the class of all graphs with n vertices,

$$2^{1/2(n-1)} \leqslant s^*(\mathscr{G}_n') \leqslant n \cdot 2^{(n-1)/2} \qquad \text{for } n \text{ odd},$$

$$\leqslant \frac{3}{2\sqrt{2}} n \cdot 2^{(n-1)/2} \qquad \text{for } n \text{ even}.$$

6. All of the preceding questions can be asked more generally for *hypergraphs*. One suspects that in this case results might be significantly more difficult to obtain than for the case of ordinary graphs. For example, even the analog of Turán's theorem for 3-uniform hypergraphs is not currently known.

Note added in proof. It has very recently been shown by N. Pippenger and the authors that (6) can be strengthened to $s(\mathscr{T}_n) = O(n \log n(\log \log n)^2)$.

References

1. J. A. Bondy, Pancyclic graphs, I, *J. Combinatorial Theory Ser. B* **11** (1971), 80–84.
2. J. A. Bondy and M. Simonovits, Cycles of even length in graphs, Research Report CORR 73-2, Univ. of Waterloo, 1973.
3. W. G. Brown, On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.* **9** (1966), 281–85.
4. P. Erdös, On extremal problems of graphs and generalized graphs, *Isr. J. Math.* **2** (1964), 183–190.
5. P. Erdös, Extremal problems in graph theory, *in* "Theory of Graphs and Its Applications," Proc. Sympos. Smolenice, pp. 29–36, Publ. Czechoslovak Acad., Prague, 1964.
6. F. Harary, "Graph Theory," Addison–Wesley, New York, 1969.
7. T. Kövari, V. T. Sós, and P. Turán, On a problem of K. Zarankiewicz, *Coll. Math.* **3** (1955), 50–57.

8. J. W. MOON, On minimal n-universal graphs, *Proc. Glasgow Math. Soc.* 7 (1965), 31–33.

9. P. TURÁN, Egy gráfelméletei szélsöértékfeladatról, *Matem. Physikai Lapok* **48** (1941), 436–452.

10. P. TURÁN, On the theory of graphs, *Coll. Math.* **3** (1954), 19–30.

Printed by the St Catherine Press Ltd., Tempelhof 37, Bruges, Belgium.