JOURNAL OF COMBINATORIAL THEORY, Series B 31, 348–355 (1981)

Note

Universal Caterpillars

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

Bell Laboratories, Murray Hill, New Jersey 07974

AND

J. SHEARER*

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Communicated by the Editors

Received March 24, 1981

For a class $\mathscr{C}$ of graphs, denote by $u(\mathscr{C})$ the least value of m so that for some graph U on m vertices, every $G \in \mathscr{C}$ occurs as a subgraph of U. In this note we obtain rather sharp bounds on $u(\mathscr{C})$ when $\mathscr{C}$ is the class of caterpillars on n vertices, i.e., tree with property that the vertices of degree exceeding one induce a path.

INTRODUCTION

Recently several of the authors have investigated graphs $U(\mathscr{C})$ which are "universal" with respect to various classes $\mathscr{C}$ of graphs. By this we mean that every graph $G \in \mathscr{C}$ occurs as a subgraph of $U(\mathscr{C})$. The usual goal has been to estimate $u(\mathscr{C})$, the minimum number of edges such a universal graph $U(\mathscr{C})$ can have. Typical examples of known results are:

(i) $\mathscr{C}_1 = \{\text{trees on } n \text{ vertices}\}$,

$$(\tfrac{1}{2} + o(1))\, n \log n < u(\mathscr{C}_1) < \left(\frac{5}{\log 4} + o(1)\right) n \log n; \tag{1}$$

* The work by this author was done while he was a consultant at Bell Laboratories.

0095-8956/81/060348–08$02.00/0

(ii) $\mathscr{C}_2 = \{\text{graphs with } n \text{ edges}\}$,

$$\frac{cn^2}{\log^2 n} < u(\mathscr{C}_2) < (1 + o(1)) \frac{n^2 \log \log n}{\log n}; \tag{2}$$

(iii) $\mathscr{C}_3 = \{\text{trees on } n \text{ vertices}\}$, $u^*(\mathscr{C}_3)$ defined as the minimum number of edges in a universal *tree*,

$$u^*(\mathscr{C}_3) = n^{(1 + o(1)) \log n / \log 4}. \tag{3}$$

Proofs of these and other results can be found in [1–7, 10, 11].

In this note we take up the same question for a special class of trees known as caterpillars (in general, we will use the graph theoretic terminology of [8]). Specifically, a *caterpillar* is a tree with the property that its vertices of degree greater than one induce a path (see [9] or [12] for many other characterizations of caterpillars).

Define c_n to be the minimum number of edges a *caterpillar* can have that is universal for all caterpillars with n vertices. Estimates for c_n have been given by Kimble and Schwenk in [9]. In particular, they show

$$\frac{n^2}{4e \log n} < c_n < \frac{3n^2 \log \log n}{\log n} \tag{4}$$

for n sufficiently large.

Our main result will be the improvement of the upper bound in (4) to

$$c_n < \frac{cn^2}{\log n} \tag{4'}$$

for a suitable constant c, which is therefore the best possible up to a constant factor.

COVERING FUNCTIONS ON $\mathbb{Z}_n$

We now shift the scene of our discussion from graphs to functions defined on the ring $\mathbb{Z}_n$ of integers modulo n. It will be easy to see the relevance of results obtained here to the estimation of c_n.

To begin with, for a fixed integer n and functions $f: \mathbb{Z}_n \to \mathbb{R}^+$, the set of nonnegative reals, and $g: \mathbb{Z}_n \to \mathbb{Z}^+$, the set of nonnegative integers, we say that f *covers* g if for some $a \in \mathbb{Z}_n$

$$f(x) \geqslant g(x + a) \qquad \text{for all } x \in \mathbb{Z}_n. \tag{5}$$

Further, call f $\mathbb{Z}_n$-*covering* if f covers every $g: \mathbb{Z}_n \to \mathbb{Z}^+$ with

$$w(g) := \sum_{x \in \mathbb{Z}_n} g(x) = n.$$

Finally, define $\lambda(n)$ by

$$\lambda(n) = \min\{w(f): f \text{ is } \mathbb{Z}_n\text{-covering}\}.$$

THEOREM. *For appropriate positive constants* c_1, c_2,

$$\frac{c_1 n^2}{\log n} < \lambda(n) < \frac{c_2 n^2}{\log n}. \tag{6}$$

Proof. We first show the lower bound. The argument is similar to one occurring in [9]. For a number t (which will be specified later; it will be about $\log n$), we consider for each t-set $T \subseteq \mathbb{Z}_n$ the function $g_T: \mathbb{Z}_n \to \mathbb{Z}^+$ by

$$\begin{aligned} g_T(x) &= \lfloor n/t \rfloor \qquad \text{if} \quad x \in T, \\ &= 0 \qquad\qquad \text{otherwise,} \end{aligned}$$

where $\lfloor x \rfloor$ denotes the integer part of x. Suppose f covers g_T for every such $T \subseteq \mathbb{Z}_n$. Let S denote $\{x: f(x) \geqslant n/t\}$ and $s = |S|$. Up to cyclic equivalence these are at least $\binom{n}{t} \cdot (t/n)$ such g_T's. Since there are just $\binom{s}{t}$ different t-subsets of S then we must have

$$\binom{s}{t} \geqslant \binom{n}{t} \frac{t}{n}. \tag{7}$$

Thus,

$$s \geqslant n \left(\frac{t}{n}\right)^{1/t}$$

and

$$w(f) = \sum_{x \in \mathbb{Z}_n} f(x) \geqslant \frac{s \cdot n}{t} = \frac{n^{2-1/t}}{t^{1-1/t}}. \tag{8}$$

Choosing $t \sim \log n$ gives

$$w(f) \geqslant \frac{n^2}{\log n} \left(\frac{\log n}{n}\right)^{1/\log n} = (1 + o(1))\, e^{-1} \frac{n^2}{\log n} \tag{9}$$

as required.

The proof of the upper bound of (6) will use the so-called probability method. Define d to be the integer satisfying

$$100 \leqslant \log_d n < e^{100}, \tag{10}$$

where we will use the abbreviation

$$\log_i x = \log(\log(\cdots (\log x) \cdots)),$$

the i-fold iterated (natural) logarithm. For $1 \leqslant i \leqslant d$, define

$$s_i = n/\log_i^2 n, \qquad k_i = (\log n)/(3 \log_{i+1} n)$$

and $k_0 = 1$. Note that

$$k_0 < k_1 < k_2 < \cdots < k_d < \log n$$

and

$$k_d > \frac{\log n}{3 \log 100}$$

For a fixed $g: \mathbb{Z}_n \to \mathbb{Z}^+$ with $w(g) = n$, define G_i to be the set $\{x \in \mathbb{Z}_n : n/k_i < g(x) \leqslant n/k_{i-1}\}$ for $1 \leqslant i \leqslant d$ and let g_i denote $|G_i|$. Note that since $w(g) = n$ then

$$\sum_{i=1}^{d} \frac{g_i}{k_i} \leqslant 1. \tag{11}$$

From this it follows that

$$\sum_{i=1}^{d} g_i \leqslant k_d < \log n. \tag{12}$$

We next defne a *random* function $f: \mathbb{Z}_n \to \mathbb{R}^+$ with the following structure. For $1 \leqslant i \leqslant d$, a random subset S_i of $\mathbb{Z}_n$ with $|S_i| = s_i$ is selected. For each $x \in S_i, f(x)$ will be defined to be n/k_{i-1}. In addition, every $x \notin \bigcup_{i=1}^{d} S_i$ will have $f(x) = n/k_d$. (Of course, what we are really doing is assigning a uniform probability measure to each of the possible functions of this form).

For such an f, we have

$$\begin{aligned} w(f) &\leqslant \sum_{i=1}^{d} s_i \frac{n}{k_{i-1}} + \frac{n \cdot n}{k_d} \\ &\leqslant \sum_{i=1}^{d} \frac{n}{\log_i^2 n} \cdot \frac{3n \log_i n}{\log n} + \frac{300\, n^2}{\log n} \\ &\leqslant \frac{n^2}{\log n} \left(3 \sum_{i=1}^{d} \frac{1}{\log_i n} + 300 \right) < \frac{cn^2}{\log n} \end{aligned}$$

for a suitable c.

We next must show there is such an f which is $\mathbb{Z}_n$–covering. To do this, we first estimate the probability that f does not cover a fixed translate of g, say $g(x + a)$. Let $G_i(a)$ denote the corresponding set G_i for this translate of g. We are actually going to require f to cover $g(x + a)$ in a special way if it is to be counted as covering $g(x+a)$. We will say that f *sharply covers* $g(x+a)$ if $G_i(a) \subseteq S_i$, $1 \leqslant i \leqslant d$.

Since there are $\binom{n}{s_i}$ ways of choosing S_i, of which $\binom{n-g_i}{s_i-g_i}$ contain $G_i(a)$ then the pobability that $G_i(a) \subseteq S_i$ is

$$\binom{n-g_i}{s_i-g_i} \Big/ \binom{n}{s_i}.$$

Since the d events $\{G_i(a) \subseteq S_i\}$, $1 \leqslant i \leqslant d$, are independent then the intersection

$$E(a) := \{G_i(a) \subseteq S_i : 1 \leqslant i \leqslant d\}$$

satisfies

$$\Pr\{E(a)\} = \prod_{i=1}^{d} \binom{n-g_i}{s_i-g_i} \Big/ \binom{n}{s_i}. \tag{13}$$

Next, observe that for translates $g(x+a)$ and $g(x+b)$ for which

$$G_i(a) \cap G_j(b) = \varnothing \qquad \text{for all } i, j, \tag{14}$$

we have

$$\Pr\{E(a) \mid E(b)\} \leqslant \Pr\{E(a)\}, \tag{15}$$

i.e.,

$$\Pr\{E(a) \cap E(b)\} \leqslant \Pr\{E(a)\} \Pr\{E(b)\}.$$

Thus, if $\bar{E}$ denotes the complement of the event E,

$$\Pr\{\bar{E}(a) \cap \bar{E}(b)\} \leqslant \Pr\{\bar{E}(a)\} \Pr\{\bar{E}(b)\} \tag{16}$$

and more generally, if $g(x+a_1), \ldots, g(x+a_u)$ are "disjoint" translates of g; i.e., $G_i(a_j) \cap G_k(a_l) = \varnothing$ for all i, j, k, l, then

$$\Pr\left\{\bigcap_{i=1}^{u} \bar{E}(a_i)\right\} \leqslant \prod_{i=1}^{u} \Pr\{\bar{E}(a_i)\}. \tag{17}$$

Thus, the probability that f does not sharply cover any of the translates $g(x+a_1), \ldots, g(x+a_u)$ is at most $(1 - \Pr\{E(0)\})^u$, since $\Pr\{E(a)\} = \Pr\{E(0)\}$ for all $a \in \mathbb{Z}_n$.

At this point it will be useful to find a lower bound on u, the number of disjoint translates of g we can find. For any $y \in \mathbb{Z}_n$ there are exactly $\sum_{i=1}^{d} g_i$ translates of g which hit y, i.e., such that $y \in \bigcup_{i=1}^{d} G_i(a)$. Thus, by (12) each translate of g rules out fewer than $\log^2 n$ other translates and so, we can certainly find $n/\log^2 n$ disjoint translates of g, i.e., we can take

$$u \geqslant n/\log^2 n. \tag{18}$$

Next, we need an upper bound on the number of different g's there are. For each choice of g_i, $1 \leqslant i \leqslant d$, there are at most $\binom{n}{g_i}$ ways to select the sets G_i. For each $x \in G_i$ there are at most $1 + n/k_{i-1}$ ways to assign a value of g to it. The locations of the $x \in \mathbb{Z}_n$ for which $g(x) \leqslant n/k_d$ are irrelevant, since $f(x)$ is always at least n/k_d for every $x \in \mathbb{Z}_n$.

Thus, a crude upper bound on the total number of g's with $w(g) = n$ is

$$\prod_{i=1}^{d} (1 + k_i) \max_{\substack{0 \leqslant g_i \leqslant k_i \\ 1 \leqslant i \leqslant d}} \prod_{i=1}^{d} \binom{n}{g_i} \prod_{i=1}^{d} \left(1 + \frac{n}{k_{i-1}}\right)^{g_i} \leqslant n^{3 \log n}.$$

for n sufficiently large. Since for each one of them, the fraction of f's which do not (sharply) cover it is at most $(1 - \Pr\{E(0)\})^u$ then there must exist *some* f which covers all g's provided

$$n^{3 \log n}(1 - \Pr\{E(0)\})^u < 1. \tag{19}$$

Taking logarithms, by (18) it is enough that

$$3 \log^2 n + \frac{2}{\log^2 n} \log(1 - \Pr\{E(0)\}) < 0. \tag{20}$$

Using (13), the inequality

$$\binom{n-g}{s-g} \Big/ \binom{n}{s} \geqslant \left(\frac{s-g}{n-g}\right)^g$$

and the inequality $-x \geqslant \log(1-x)$ for $x < 1$, it follows that it is enough that for n sufficiently large

$$\sum_{i=1}^{d} g_i \log \left(\frac{s_i - g_i}{n - g_i}\right) > \log((3 \log^4 n)/n),$$

or, since $\log(1-x) \geqslant -x - x^2$ for $0 \leqslant x < \frac{1}{2}$,

$$\sum_{i=1}^{d} g_i \log \frac{s_i}{n} > 6 \log_2 n - \log n. \tag{21}$$

But

$$\log \frac{n}{s_i} = 2 \log_{i+1} n = \frac{2 \log n}{3k_i}$$

so that it is enough that

$$\frac{2}{3}\log n \sum_{i=1}^{d} \frac{g_i}{k_i} < \log n - 6 \log_2 n.$$

However, by (11) this easily holds for n sufficiently large.

Consequently, there must exist an f of the required form covering all the g's. By the previous calculation, such an f has $w(f) < cn^2/\log n$ for some fixed c. This proves the theorem. ∎

The application of the Theorem to the estimate for c_n is immediate. Simply observe that a universal caterpillar for n-vertex caterpillars can be formed by placing $f(x)$ edges at the "vertex" $x \in \mathbb{Z}_n$, "opening up" the cycle $\mathbb{Z}_n$ to form a caterpillar and joining two copies of this graph together. It seems certain that for some c^*

$$c_n \sim c^* \frac{n^2}{\log n}.$$

It would be interesting to determine the exact value of c^* in this case.

We also note that analogues to the Theorem can be proved in the more general setting in which our functions are defined on an n-set S on which some permutation group G acts. We can say that f covers g in this case if $f(x) \geq g(x^\sigma)$ for some $\sigma \in G$ and all $x \in S$. In general, one can ask for estimates of the minimum weight a function can have which covers all $g: S \to \mathbb{R}^+$ with $w(g) = m$. However, we will not pursue this here.

References

1. L. Babai, F. R. K. Chung, P. Erdös, and R. L. Graham, On graphs which contain all sparse graphs, to appear.
2. J. A. Bondy, Pancyclic graphs, I. *J. Combin. Theory Ser. B* **11** (1971), 80–84.
3. F. R. K. Chung and R. L. Graham, On graphs which contain all small trees, *J. Combin. Theory Ser. B* **24** (1978), 14–23.
4. F. R. K. Chung and R. L. Graham, On universal graphs, *in* "Proceedings, Second Int. Conf. on Combin. Math." (A. Gewirtz and L. Quintas, Eds.); *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, "Proc., 1976 Hungarian Colloq. on Combinatorics," pp. 213–223, North-Holland, Amsterdam, 1978.
6. F. R. K. Chung, R. L. Graham, and D. Coppersmith, On graphs containing all small trees, *in* "The Theory and Applications of Graphs," (G. Chartrand, Ed.), pp. 255–264, Wiley, New York, 1981.
7. M. K. Goldberg and E. M. Lefschitz, On minimal universal trees, *Mat. Zametki* **4** (1968), 371–379.

8. F. HARARY, "Graph Theory," Addison–Wesley, Reading, Mass., 1969.
9. R. J. KIMBLE AND A. J. SCHWENK, On universal caterpillars, to appear.
10. J. W. MOON, On minimal n-universal graphs, *Proc. Glasgow Math. Soc.* **7** (1965), 32–33.
11. L. NEBESKÝ, On tree-complete graphs, *Časopis Pěst. Mat.* **100** (1975), 334–338.
12. B. ZELINKA, Caterpillars, *Časopis Pěst. Mat.* **102** (1977), 179–185.