JOURNAL OF COMBINATORIAL THEORY, Series B 41, 61–79 (1986)

Maximum Induced Trees in Graphs

PAUL ERDÖS

Hungarian Academy of Sciences, Budapest, Hungary

MICHAEL SAKS*

Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903

AND

VERA T. SÓS†

Bell Laboratories, Murray Hill, New Jersey

Let $t(G)$ be the maximum size of a subset of vertices of a graph G that induces a tree. We investigate the relationship of $t(G)$ to other parameters associated with G: the number of vertices and edges, the radius, the independence number, maximum clique size and connectivity. The central result is a set of upper and lower bounds for the function $f(n, \rho)$, defined to be the minimum of $t(G)$ over all connected graphs with n vertices and $n-1+\rho$ edges. The bounds obtained yield an asymptotic characterization of the function correct to leading order in almost all ranges. The results show that $f(n, \rho)$ is surprisingly small; in particular $f(n, cn) = 2 \log \log n + O(\log \log \log n)$ for any constant $c > 0$, and $f(n, n^{1+\gamma}) = 2 \log(1 + 1/\gamma) \pm 4$ for $0 < \gamma < 1$ and n sufficiently large. Bounds on $t(G)$ are obtained in terms of the size of the largest clique. These are used to formulate bounds for a Ramsey-type function, $N(k, t)$, the smallest integer so that every connected graph on $N(k, t)$ vertices has either a clique of size k or an induced tree of size t. Tight bounds for $t(G)$ from the independence number $\alpha(G)$ are also proved. It is shown that every connected graph with radius r has an induced path, and hence an induced tree, on $2r-1$ vertices.

I. INTRODUCTION, DEFINITIONS AND MAIN RESULTS

Let $t(G)$ be the maximum size of a subset of vertices of a graph G that induces a tree. We investigate the relationship of $t(G)$ to other parameters associated with G: the number of vertices and edges, the radius, independence number, maximum clique size and connectivity.

* The work of this author was supported by the NSF under Grant MCS 81-02448.

† Permanent address: Eotvos University of Budapest, Hungary.

0095-8956/86 $3.00

By a graph we will mean an undirected graph without loops or multiple edges. Unless otherwise stated, all graphs are assumed to be connected. If G is a graph, $V(G)$ and $E(G)$ denote its vertex and edge set. An edge between v and w is denoted by $\langle v, w \rangle$. We write $n(G)$, $e(G)$, and $\alpha(G)$, respectively, for the number of vertices of G, the number of edges of G and the independence number of G. The *acyclotomic number*, $\rho(G)$, is defined for connected graphs to be $e(G) - n(G) + 1$.

If $W \subseteq V(G)$, the graph induced on W is written G_W. $G - W$ denotes the graph induced on $V(G) - W$.

The function $p(G)$ is defined to be the maximum size of a subset of vertices that induces a path; trivially, $p(G) \leqslant t(G)$. P_n denotes the graph consisting of a path of n vertices.

For any vertices $v, w \in V(G)$, $d_G(v, w)$ is the distance from v to w, if there is no ambiguity we write simply $d(v, w)$. We will often use the triangle inequality, $d(v, w) \leqslant d(v, x) + d(x, w)$. The *diameter* of G is defined by $\operatorname{diam}(G) = \max\{d(v, w)\colon v, w \in V(G)\}$. The *centrality* of a vertex v, written $c_G(v)$, is the maximum of $d(v, w)$ over all vertices $w \in V(G)$. The *radius* of G is given by $\operatorname{rad}(G) = \min\{c(v)\colon v \in V(G)\}$. A vertex v for which $c(v) = \operatorname{rad}(G)$ is a *center* of G.

A vertex of G whose neighbors form a clique is called a *simplicial vertex* of G.

If $v \in V(G)$ and k is a positive integer, *multiplying v by k* means replacing v by a clique on k vertices, each inheriting the edges that v had. A graph G' is a multiple of G if it can be obtained from G by multiplying vertices.

The Ramsey number $R(a, b)$ for positive integers a and b is the smallest integer n so that every graph on n vertices has either a clique of size a or an independent set of size b.

All logarithms are to the base 2.

The main results of this paper are upper and lower bounds on $t(G)$ (or $p(G)$) in terms of certain parameters or structural properties of G. (Throughout, inequalities involving n may hold only for n sufficiently large.) In Section 2 we relate $t(G)$ to the radius by establishing:

THEOREM 2.1. *$p(G) \geqslant 2\operatorname{rad}(G) - 1$, for any connected graph G.*

Sections 3 through 7 are concerned with bounds on $t(G)$ which can be stated in terms of $e(G)$ and $n(G)$. Since a connected graph has at least $n(G) - 1$ edges it is convenient to state our results in terms of the acyclotomic number $\rho(G)$. For $n \geqslant 1$ and $0 \leqslant \rho \leqslant \binom{n-1}{2}$ let $\mathscr{G}(n, \rho)$ denote the class of connected graphs G with $n(G) = n$ and $\rho(G) = \rho$. We define $u(n, \rho) = \max\{t(G)\colon G \in \mathscr{G}(n, \rho)\}$ and $f(n, \rho) = \min\{t(G)\colon G \in \mathscr{G}(n, \rho)\}$. In Section 3 we prove the following easy result.

THEOREM 3.1. *$u(n, \rho) = \min\{t\colon \binom{t}{2} > \binom{n-1}{2} - \rho\}$.*

In Section 4 we begin our study of $f(n, \rho)$, which is considerably more complicated than $u(n, \rho)$. We state several upper and lower bounds which are then proved in Sections 5, 6, and 7. These bounds provide a nearly exact asymptotic description of $f(n, \rho)$. The following theorem gives this for three important ranges:

THEREOREM 4.1. (i) $f(n, \rho) = 2n/(\rho + 2) + o(n/(\rho + 2))$, *for* $\rho = o(n/\log \log n)$.

(ii) $f(n, cn) = 2 \log \log n + O(\log \log \log n)$, *for any constant* $c > 0$.

(iii) *For* n *sufficiently large and* $0 < \gamma < 1$, $f(n, n^{1+\gamma}) = 2 \log(1 + 1/\gamma) + \varepsilon$, *where* $-2 \leqslant \varepsilon \leqslant 4$.

These results indicate that $t(G)$ can be surprisingly small for relatively sparse graphs. The constructions which realize the value of $f(n, \rho)$ are line graphs of trees, and are described in Sections 5 and 6. Section 8 establishes the following relationship between $t(G)$ and the independence number of G.

THEOREM 8.2. *For any connected graph G with n vertices and any integer* $1 \leqslant m \leqslant (n-1)/2$:

$$\alpha(G) > \frac{(m-1)\,n}{m} + 1 \qquad \textit{implies} \quad t(G) \geqslant 2m + 1$$

$$\alpha(G) > \frac{(m-1)\,n + 1}{m} + 1 \qquad \textit{implies} \quad t(G) \geqslant 2m + 2$$

and these bounds are best possible; for any $1 \leqslant m \leqslant (n-1)/2$ *there exist graphs* $G_1(m, n)$ *and* $G_2(m, n)$ *on* n *vertices with* $\alpha(G_1) > (m-1)\,n/m$ *and* $t(G_1) = 2m$, *and* $\alpha(G_2) > ((m-1)\,n + 1)/m$ *and* $t(G_2) = 2m + 1$.

In Section 9 we relate $t(G)$ to the maximum clique size of G. The function $c(n, k)$ is defined to be the minimum of $t(G)$ over all graphs with n vertices and no clique of size k. We obtain the following bounds on $c(n, k)$:

THEOREM 9.1. (i) $c(n, k) \geqslant 2 \log n/((k-2) \log \log n) - 3$ *for* $k \geqslant 3$, $n \geqslant 4$.

(ii) $c(n, 3) \leqslant c_1 \sqrt{n} \log n$ *for some constant* c_1.

(iii) $c(n, k) \leqslant 2 \lceil \log(n-1)/\log(k-2) \rceil + 2$ *for* $k \geqslant 4$.

Theorem 9.1 can be restated as a "Ramsey" type theorem.

THEOREM 9.1′. *For any positive integers* $k \geqslant 3$ *and* $t \geqslant 2$ *there exists a minimum integer* $N(k, t)$ *such that every connected graph on at least* $N(k, t)$ *vertices has either a clique of size k or an induced tree of size t. Moreover*

(i) $N(k, t) < 2R(k-1, t)^{(t+1)/2}$ *for* $k \geq 3$, $t \geq 2$.

(ii) *There exists a constant* C_1 *so that*

$$N(3, t) \geq R\left(3, \frac{t}{2}\right) \geq C_1 \frac{t^2}{(\log t)^2}.$$

(iii) $N(k, t) \geq (k-2)^{t/2-2} + 1$ *for* $k \geq 4$.

Section 10 contains some open problems. It is worth noting that computing $t(G)$ for an arbitrary connected graph G is difficult.

PROPOSITION 1.1. *The problem "given a connected graph G and integer t, is* $t(G) > t$*?" is NP complete.*

Proof. The problem "given a graph H and integer k does H have an independent set of size k?" is a well-known NP-complete problem [4]. Given H and k, let n be the number of vertices of H and let G be the graph obtained by adjoining a path on n vertices to H, one endpoint of which is joined by an edge to every vertex in H. The problem of whether H has an independent set of size k is easily seen to be equivalent to whether G has an induced tree of size $n+k$. ▮

II. BOUNDS ON $p(G)$ FROM THE RADIUS

If v and w are two vertices of maximum distance, then since the shortest path between them is an induced path, we have

PROPOSITION 2.1. $p(G) \geq \operatorname{diam}(G) + 1$.

A related fact is

THEOREM 2.2. $p(G) \geq 2\operatorname{rad}(G) - 1$.

Note that since $\operatorname{rad}(G)$ can be as large as $\operatorname{diam}(G)$, this theorem may give as much as an extra factor of 2 over the bound of Proposition 2.1. The following proof of Theorem 2.2 was provided by Fan Chung, replacing a cumbersome proof given in a previous version of the paper.

Proof. Let $G = (V, E)$ have radius r. We can assume, by induction on $|V|$, that no connected induced subgraph of G has radius r. Let v_r be a vertex that is not a cutpoint. Since the graph induced on $V - v_r$ is connected, it has radius less than r; let v_0 be a center of this graph. Then $d(v_0, w) \leq r-1$ for $w \neq v_r$ and so $d(v_0, v_r)$ must equal r. Let $v_0, v_1, \ldots, v_r$ be a shortest path from v_0 to v_r. There exists a vertex w with $d(v_2, w) \geq r$. Therefore

$d(v_0, w) \geqslant r-2$ and also $d(v_0, w) \leqslant r-1$ since $w \neq v_r$. Let P be a shortest path from v_0 to w. If any vertex u in P is adjacent to v_j for some $j \geqslant 2$, then

$$\begin{aligned} d(v_0, w) &= d(v_0, u) + d(u, w) \geqslant d(v_0, v_j) - 1 + d(v_j, w) - 1 \\ &\geqslant d(v_0, v_j) - 2 + d(v_2, w) - d(v_2, v_j) \geqslant r, \end{aligned}$$

a contradiction. Hence $v_r, v_{r-1}, \ldots, v_1, v_0$ followed by P is a path of $2r$-1 or $2r$ vertices that fails to be induced only if v_1 is adjacent to some vertex of P. If P has $r-2$ vertices this is impossible since $d(v_1, w) \geqslant r-1$. If P has $r-1$ vertices then v_1 may be adjacent to the first vertex of P. In that case, deleting v_0 yields the desired path. ∎

III. An Upper Bound on $t(G)$ from the Number of Edges

In the next few sections we compute bounds on $t(G)$ from the number of vertices and edges of G. Recall from the introduction that $\mathscr{G}(n, \rho)$ is the class of connected graphs with n vertices and $n-1+\rho$ edges and $u(n, \rho)$ is the maximum of $t(G)$ over all graphs in $\mathscr{G}(n, \rho)$.

THEOREM 3.1. $u(n, \rho) = \min\{t: \binom{t}{2} > \binom{n-1}{2} - \rho\}$.

Proof. Fix n and ρ and let $t^* = \min\{t: \binom{t}{2} > \binom{n-1}{2} - \rho\}$. If $G \in \mathscr{G}(n, \rho)$ and T is an induced tree in G, then since T has $|T|-1$ out of a possible $\binom{|T|}{2}$ arcs, we have $\rho \leqslant \binom{n-1}{2} - \binom{|T|}{2} + |T| - 1$ or $\binom{|T|-1}{2} \leqslant \binom{n-1}{2} - \rho$, hence $|T| \leqslant t^*$.

Conversely, we can construct a graph G in $\mathscr{G}(n, \rho)$ with $t(G) = t^*$ by taking a tree T on t^* vertices, a clique C on $n - t^*$ vertices and adding the required number of edges between T and C. ∎

IV. The Function $f(n, \rho)$

Next we tackle the problem of describing the function $f(n, \rho)$, defined in the introduction to be the minimum of $t(G)$ over graphs with n vertices and $n+\rho-1$ edges. Unlike $u(n, \rho)$, the computation of $f(n, p)$ is in general difficult. We first state a theorem which describes $f(n, \rho)$ in three important ranges.

THEOREM 4.1. (i) *$f(n, \rho) = 2n/(p+2) + o(n/(\rho+2))$ for $\rho = o(n/\log\log n)$.*

(ii) *$f(n, cn) = 2\log\log n + O(\log\log\log n)$ for any constant $c > 0$.*

(iii) *For* $0<\gamma<1$ *and* n *sufficiently large*, $f(n, n^{1+\gamma}) = 2\log(1+1/\gamma)+\varepsilon$, *where* $-2 \leqslant \varepsilon \leqslant 4$.

Theorem 4.1 can be deduced from five theorems which we state here and prove in the next three sections. The first three give upper bounds on $f(n, \rho)$ and the other two give lower bounds.

THEOREM 4.2. *For* $\rho \leqslant (n-1)/2$,

$$f(n, \rho) \leqslant 2\left(\frac{n+3}{\rho+2} + \log\lceil(\rho+2)\rceil - 1\right).$$

THEOREM 4.3. *For* $\rho \geqslant 8n$,

$$f(n, \rho) \leqslant 2\log(\log\rho/(\lfloor\log(\rho/n)\rfloor - 2)) + 3.$$

THEOREM 4.4. *For* $32 \leqslant \rho \leqslant 8n$,

$$f(n, \rho) \leqslant 2\log\log\rho + \frac{2n}{\rho} + \frac{c_1 n}{\rho\log\log\rho} + c_2\log\log\log\rho$$

for some constants c_1 *and* c_2.

THEOREM 4.5. $f(n, \rho) \geqslant 2n/(\rho+2)$.

THEOREM 4.6. *For* $n \geqslant 4$ *and* $\rho \geqslant n/\log\log n$

$$f(n, \rho) \geqslant 2\log\frac{\log\rho + \log\log\log n + 2}{\log\rho - \log n + 5\log\log\log n + 4} - 1.$$

These theorems can be combined to give a nearly exact asymptotic description of $f(n, \rho)$. The only gap in this description is when $\rho = c\,n/\log\log n$ for some constant c in which case we know that $f(n, \rho) = c_1\log\log n + o(\log\log n)$, where c_1 is between $\max(2, 2/c)$ and $2+2/c$.

The proofs of the upper bounds in Theorems 4.2, 4.3, and 4.4 are obtained by construction. The graphs used in the construction are line graphs of trees. The lower bounds in Theorems 4.5 and 4.6 are proved in Section 7.

Before proceeding with these proofs we investigate some properties of $f(n, \rho)$ which will be needed later.

It is clear that $f(n, \binom{n-1}{2}) = 2$ and $f(n, \binom{n-1}{2} - 1) = 3$. We consider first the question: how small can ρ be so that $f(n, \rho) = 3$? The following proposition characterizes graphs for which $t(G) = 3$.

PROPOSITION 4.7. *If G is connected then $t(G) \leqslant 3$ if and only if the complement of G is a union of disjoint complete bipartite graphs and isolated vertices.*

Proof. If $t(G) \leqslant 3$ then G has no independent set of size 3. This follows from Theorem 8.1 but can be proved directly. For if x, y, and z are an independent set then $d(x, y) = d(y, z) = 2$, since $\operatorname{diam}(G) \leqslant t(G) - 1 = 2$. Thus x and y have a common neighbor v and y and z have a common neighbor u. If $v = u$, G has an induced 4-star and otherwise G has an induced 4-path.

Therefore the complement of G is bipartite; let C be a connected component of $\bar{G}$ with bipartition X and Y. If C has an induced path on four vertices then so does G, thus $\operatorname{diam}(C) = 2$. If $x \in X$ and $y \in Y$ then $d(x, y)$ is odd, but then $d(x, y)$ must be one. Therefore C is a complete bipartite graph.

On the other hand, if $\bar{G}$ is a union of disjoint complete bipartite graphs and isolated vertices then it contains no induced triangle or 4-path, so G contains no induced 4-star or 4-path and $t(G) \leqslant 3$. ∎

As a simple consequence of this proposition we obtain:

PROPOSITION 4.8. *For any $n \geqslant 3$, the smallest value of ρ such that $f(n, \rho) = 3$ is $\rho = (n-2)^2/4$. Moreover, for any ρ with $(n-2)^2/4 \leqslant \rho < \binom{n-1}{2}$, $f(n, \rho)$ is either 3 or 4.*

Proof. By Proposition 3.1 we know that $t(G) = 3$ only if $\bar{G}$ consists of disjoint complete bipartite graphs, and since G is connected, $\bar{G}$ has at least two components. The number of edges in $\bar{G}$ (when $n = n(G)$ is fixed) is thus maximized if $\bar{G}$ consists of the complete bipartite graph on $(n-1)/2$ and $(n-1)/2$ vertices together with one isolated vertex. In this case, $\rho(G) = (n-2)^2/4$.

If $\rho > (n-2)^2/4$, we can add additional edges to the graph constructed above to obtain a graph in $\mathscr{G}(n, \rho)$ whose largest induced tree has at most 4 vertices. ∎

It seems natural to suppose that, like $u(n, \rho)$, $f(n, \rho)$ should be a decreasing function of ρ, i.e., the fewer the number of edges, the larger the size of the tree one is forced to have. However, if we take $\rho = (n-2)^2/4 + 2$ with $n \geqslant 9$, we find that there is no G in $\mathscr{G}(n, \rho)$ satisfying the conditions of Proposition 4.7. Therefore, $f(n, (n-2)^2/4 + 2) > f(n, (n-2)^2/4)$ and $f(n, \rho)$ can decrease with ρ.

It is the case, however, that $f(n, \rho)$ is "almost" a decreasing function of ρ, in the sense that increasing ρ cannot increase $f(n, \rho)$ by very much.

LEMMA 4.9. *Suppose $G \in \mathscr{G}(n_1, \rho_1)$, $n_2 \leqslant n_1$, and $\rho_2 \geqslant \rho_1$. Then*

(i) *there exists a graph* $G' \in \mathscr{G}(n_2, \rho_2)$ *so that* $t(G') \leqslant t(G)+2$

(ii) *if G has a simplicial vertex then there exists a graph* $G \in \mathscr{G}(n_2, \rho_2)$ *so that* $t(G') \leqslant t(G)+1$.

Proof. We prove (ii) first. Let w be a simplicial vertex of G and let H be a connected induced subgraph of G with n_2 vertices including w; then $\rho(H) \leqslant \rho_1$ and $t(H) \leqslant t(G)$. Let C be the clique in H consisting of w and its neighbors. List the vertices of $H-C$ as $v_1, v_2, \ldots, v_k$ so that $H-\{v_1, v_2, \ldots, v_j\}$ is connected for each $1 \leqslant j \leqslant k$ and let H_j denote the graph obtained from $H-\{v_1, v_2, \ldots, v_j\}$ by multiplying w by $j+1$. Since H_k is the complete graph and $H_0=H$, there exists an index i such that $\rho(H_i) \geqslant \rho_2 > \rho(H_{i-1})$. Let G' be the graph obtained from H_i by deleting $\rho(H_i)-\rho_2$ edges incident to w, which is possible since w has degree at least $\rho(H_i)-\rho(H_{i-1})$ in H_i. Now $p(G')=\rho_2$ and $\rho(G') \leqslant t(H_i)+1 \leqslant t(H)+1 \leqslant t(G)+1$. This completes the proof of (ii).

To prove (i), let w be a non-cutpoint of G and delete all edges incident on w except one. The resulting graph H has $t(H) \leqslant t(G)+1$ and is in $\mathscr{G}(n_1, \rho_1')$, where $\rho_1' \leqslant \rho_1$ and w is a simplicial vertex. Now apply (ii) to the graph H. ∎

An immediate consequence of Lemma 4.9 is

THEOREM 4.10. *For* $n_1 \geqslant n_2$ *and* $\rho_1 \leqslant \rho_2$, $f(n_1, \rho_1)+2 \geqslant f(n_2, \rho_2)$.

V. LINE GRAPHS OF TREES

Recall that the line graph $G=L(H)$ of a graph H is the graph whose vertex set is the edge set of H, with two vertices in G joined by an arc if their corresponding edges are incident on a common vertex in H.

PROPOSITION 5.1. (i) *If H is any graph and G is its line graph then*

$$t(G)=\rho(G) \geqslant p(H)-1.$$

(ii) *If H is a tree then* $t(G)=d(H)$.

Proof. Line graphs have no induced $K_{1,3}$ so every induced tree of G is a path and $t(G)=p(G)$. If $v_1, v_2, \ldots, v_k$ induces a path in H then $e_1, e_2, \ldots, e_{k-1}$ is an induced path in G, where $e_i=\langle v_i, v_{i+1}\rangle$; so $p(G) \geqslant p(H)-1$.

On the other hand, an induced path $e_1, \ldots, e_q$ in G corresponds to the (not necessarily induced) path $v_1, \ldots, v_{q+1}$ in H, where $e_i=\langle v_i, v_{i+1}\rangle$. If H is a tree then every path is induced and therefore $p(G)=p(H)-1$. Moreover $d(H)=p(H)-1$ for a tree, hence (ii). ∎

The line graph construction will be used to obtain graphs with given n and ρ for which $t(G)$ is small. These examples will be built from two special classes of trees.

A *balanced regular tree*, $B(c, k)$, is a rooted tree in which k vertices have degree c and the remainder are leaves, with the depth (or equivalently, height) of any two leaves differing by at most one.

If $s_0, s_1, s_2, \ldots, s_k$ are integers, the *layered tree* $T(s_0, s_1, \ldots, s_k)$ is a rooted tree of depth $k+1$ whose root has s_k sons and each vertex at distance i from the root has s_{k-i} sons.

In Section 7, we use these classes of trees to bound $f(n, \rho)$. For ρ small relative to $n(\rho = o(n/\log n))$ we use the line graphs of $B(3, k)$ with a path attached to each leaf. For $\rho > 8n$ we use the line graphs of layered trees in which the numbers s_i satisfy $s_i = S_{i-1}^2/2$. For intermediate values of ρ we use the line graph of a tree consisting of a layered tree with a copy of $B(3, k)$ attached to each leaf.

In order to describe the precise constructions and prove the bounds we will need a clear statement of the relationship between a tree and its line graph. The following can be easily verified by the reader.

LEMMA 5.2. *Let T be a tree on m vertices with degrees $(d_1, d_2, \ldots, d_m)$, and let G be its line graph. Then*

(i) $n(G) = m - 1$

(ii) $\rho(G) = \sum_{i=1}^{m} \binom{d_i - 1}{2}$

(iii) *the number of simplicial vertices of G equals the number of leaves of T.* ∎

VI. PROOFS OF THE UPPER BOUNDS ON $f(n, \rho)$

In this section we prove Theorems 4.2, 4.3, and 4.4.

Proof of Theorem 4.2. We first state the following lemma, which follows from Proposition 5.1 (ii) and Lemma 5.2.

LEMMA 6.1. *The line graph L of $B(c, k)$ satisfies:*

(i) $n(L) = (c-1)k + 1$

(ii) $\rho(L) = \binom{c-1}{2} k$

(iii) *L has $(c-2)k + 2$ simplicial vertices*

(iv) $t(L) \leqslant 2\lceil \log k/\log(c-1) \rceil + 2$.

Given n and $\rho \leqslant (n-1)/2$, construct the line graph of $B(3, \rho)$. It has

$2\rho+1$ vertices, $\rho+2$ simplicial vertices, and acyclotomic number ρ. To each of its free vertices attach a path on either $\lceil(n-2\rho-1/(\rho+2)\rceil$ or $\lfloor(n-2\rho-1)/(\rho+2)\rfloor$ vertices to obtain a graph with n vertices, acyclotomic number ρ, and longest induced path of length $2(\log(\rho+2)+(n+3)/(\rho+2)-1)$ (See Fig. 6.1). This proves Theorem 4.2. ∎

Proof of Theorem 4.3. For positive integers $s_0, s_1, ..., s_r$ we will write T_j to denote the layered tree $T(s_0, s_1, ..., s_j)$, L_j to denote the line graph of T_j and v_j, m_j, and ρ_j to denote, respectively, the number of vertices, the number of simplicial vertices, and the acyclotomic number of L_j.

LEMMA 6.2. *For any sequence* $s_0, s_1, ..., s_r$ *of positive integers and* $0 \leq j \leq r$:

(i) $m_j = s_0 s_1 s_2 \cdots s_j$

(ii) $v_j = m_j(1 + 1/m_0 + 1/m_1 + \cdots + 1/(m_{j-1}))$

(iii) $\rho_j = (m_j/2)((s_0-1) + (s_1-1)/m_0 + (s_2-1)/m_1 + \cdots + (s_j-1)/m_{j-1})) - s_j + 1$

(iv) $t(L_j) = 2j+2$.

Proof. Let C_j be the clique in L_j corresponding to the s_j edges incident on the root of T_j. The graphs L_j with distinguished cliques C_j can be constructed inductively as follows: $L_0 = C_0$ is a clique on s_0 vertices. Given L_j and distinguished clique C_j let L_j' be the graph obtained by adding an additional vertex that is connected to every vertex in C_j. L_{j+1} consists of S_{j+1} copies of L_j' in which the s_{j+1} added vertices are joined as the clique C_{j+1}.

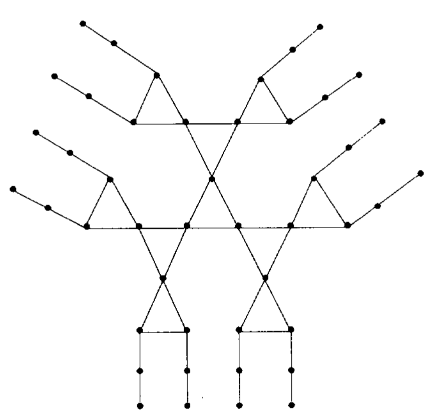

FIG. 6.1. The construction for Theorem 4.2 with $n=45$ and $\rho=10$ consisting of $B(3, 10)$ with 12 paths of 2 vertices adjoined. Note $t(G)=p(G)=10$.

Viewed in this way it is straightforward to obtain the recurrences:

$$m_{j+1} = m_j s_{j+1}, \qquad m_0 = s_0;$$

$$v_{j+1} = (v_j + 1)\, s_{j+1}, \qquad v_0 = s_0;$$

$$\rho_{j+1} = s_{j+1}(\rho_j + s_j - 1) + \binom{s_{j+1} - 1}{2}, \qquad \rho_0 = \binom{s_{0-1}}{2};$$

$$t(L_j) = t(L_j) + 2, \qquad t(L_0) = 2,$$

which are solved as given by the lemma. ∎

Our aim is to construct a tree with small diameter whose line graph has a prescribed number of vertices and acyclotomic number. In order to do this it is useful to consider the question: for a fixed number of vertices and radius r how should $s_0, s_1, \dots, s_r$ be chosen so as to minimize the acyclotomic number of $L(s_0, \dots, s_r)$. To get a rough answer we approximate the total number of vertices in $L(s_0, \dots, s_r)$ by $s_0 s_1 \cdots s_r$, and minimize ρ_r subject to this fixed. Consider how ρ_r changes if we permit only s_j and s_{j+1} to vary.

By Lemma 6.2 (iii) we have that ρ_r is approximately

$$\frac{m_r}{2}\left(s_0 + \frac{s_1}{m_0} + \frac{s_2}{m_1} + \cdots + \frac{s_j}{m_{j+1}} + \frac{s_{j+1}}{m_j} + \cdots + \frac{s_r}{m_{r-1}}\right).$$

Note that the only terms which vary if s_j and s_{j+1} change (but their product remains constant) are the terms

$$\frac{s_j}{m_{j-1}} + \frac{s_{j+1}}{m_j} = \frac{1}{m_{j-1}}\left(s_j + \frac{s_{j+1}}{s_j}\right).$$

Minimizing $s_j + s_{j+1}/s_j$ subject to $s_j s_{j+1}$ a constant yields $s_{j+1} = s_j^2/2$.

Thus we conclude that a reasonable choice for the sequence $s_0, s_1, \dots, s_r$ is to choose one satisfying the above recurrence. Our aim now is, given v and ρ, to construct such a sequence.

Lemma 6.3. *If $s_0, s_1, \dots, s_r$ is a sequence of integers satisfying $s_{i+1} \leqslant s_i^2/2$ then $\rho_r/m_r \leqslant s_0$.*

Proof. By Lemma 6.2 (iii), $\rho_r \leqslant (m_r/2)(s_0 + s_1/m_0 + \cdots + s_r/m_{r-1})$. An easy induction shows that, assuming $s_{i+1} \leqslant s_i^2/2$ for all i, $s_{i+1}/m_i \leqslant s_0/2^{i+1}$. Thus $\rho \leqslant (m_r/2)(s_0 + s_0/2 + s_0/2^2 + \cdots + s_0/2^r) \leqslant (mr)(s_0)$. ∎

For any integer $q \geqslant 2$, define the sequence $\sigma_{q,0}, \sigma_{q,1}, \sigma_{q,2}, \dots$ by the recurrence $\sigma_{q,0} = 2^q$, $\sigma_{q,i} = (\sigma_{q,i-1})^2/2$. Solving this yields $\sigma_{q,i} = 2 \cdot 2^{2^i(q-1)}$.

LEMMA 6.4. *The graph $L(\sigma_{q,0}, \sigma_{q,1},..., \sigma_{q,r})$ satisfies*

(i) $m_j = 2^{j+1} 2^{(q-1)(2^{j+1}-1)}$

(ii) $\rho_j \leqslant 2^q m_j$

for each $j \geqslant 0$.

Proof. (i) Follows from Lemma 6.2 (i) and (ii) from Lemma 6.3. ∎

Now given n and ρ, set $q = \lfloor \log(\rho/n) \rfloor - 1$ and $r = \lceil \log(\log \rho / (\lfloor \log(\rho/n) \rfloor - 2)) \rceil$ and consider the graph $L(\sigma_{q,0}, \sigma_{q,1},..., \sigma_{q,r})$. By Lemma 6.4 (i), $m_r = 2^{r+1} 2^{(q-1)(2^{r+1}-1)} \geqslant 2^{(q-1)(2^{r+1}-1)}$. Now we have,

$$(q-1)(2^{r+1}-1) = \left(\left\lfloor \log \frac{\rho}{n} \right\rfloor - 2\right)\left(\frac{\log \rho}{\lfloor \log(\rho/n) \rfloor - 2} - 1\right)$$

$$= \log \rho - \log \frac{\rho}{n} + 2 \geqslant \log n + 2$$

so $m_r \geqslant 2^{\log n + 2} > n$.

Let k be the largest index so that L_k has fewer than n vertices; we must have $k < r$. Choose s_{k+1} as small as possible so that $G = L(\sigma_{q,0}, \sigma_{q,1},..., \sigma_{q,k}, s_{k+1})$ has at least n vertices. We have $t(G) \leqslant 2(k+1) + 2 \leqslant 2r + 2$. We claim that $\rho(G) \leqslant \rho$, in which case by Lemma 4.9 (ii) there exist G' in $\mathscr{G}(n, \rho)$ with $t(G') \leqslant t(G) + 1 \leqslant 2r + 3 = 2 \log(\log \rho / (\lfloor \log(\rho/n) \rfloor - 2)) + 3$ as required to prove Theorem 4.3.

Now by Lemma 6.4, $\rho(G) \leqslant 2^q m(G) = 2^{\log(\rho/n) - 1} m(G) \leqslant \rho m(G)/2n$. Therefore it suffices to show that $m(G)/2n \leqslant 1$. By the choice of s_{k+1}, we know that $H = L(\sigma_{q,0}, \sigma_{q,1},..., \sigma_{q,k}, s_{k+1} - 1)$ has fewer than n vertices (assuming $s_{k+1} \geqslant 2$; the case $s_{k+1} = 1$ is trivial). Now by Lemma 6.2 (ii) we can see that $n(G) \leqslant 2n(H)$ so we have $m(G) \leqslant n(G) < 2n(H) < 2n$ as required to finish the proof of the theorem. ∎

Proof of Theorem 4.4. Given n and ρ with $32 \leqslant \rho \leqslant 8n$, let $n_1 = \lceil \rho/8 \log \log \rho \rceil$ and $\rho_1 = 8n_1$. By the construction of the previous theorem, there is a tree T_1 whose line graph L_1 satisfies $2n_1 \geqslant m(L_1) \geqslant n_1$, $\rho(L_1) \leqslant \rho_1$ and

$$t(L_1) \leqslant 2 \log(\log \rho_1 / \lfloor \log \rho_1/n_1 \rfloor - 2) + 3 \leqslant 2 \log \log \rho + O(1).$$

Now let $n_2 = \lceil (n - n(L_1))/m(L_1) \rceil$ and $\rho_2 = \lceil (\rho - \rho(L_1))/m(L_1) \rceil - 2$. By the construction for Theorem 4.2 there exists a tree T_2 whose line graph L_2 has n_2 vertices, acyclotomic number ρ_2, and

$$t(G_2) \leqslant 2\left(\frac{\lceil n_2+3 \rceil}{\rho_2+2} + \lceil \log(\rho_2+2) \rceil - 1\right)$$
$$\leqslant \frac{2n}{\rho} + \frac{c_1 n}{\rho \log\log\rho} + c_2 \log\log\log\rho$$

for some constants c_1 and c_2. Now let T_3 be the tree obtained by taking T_1 and $m(L_1)$ copies of T_2 and joining each leaf of T_1 to the root of some copy of T_2, and let $L_3 = L(T_3)$. This yields a graph with at least n vertices, acyclotomic number at most ρ_1, and

$$t(L_3) \leqslant t(L_1) + t(L_2) + 2 \leqslant 2 \log\log\rho + 2\frac{n}{\rho}$$
$$+ \frac{c_1 n}{\rho \log\log\rho} + c_2 \log\log\log\rho + O(1).$$

Finally applying Theorem 4.10 and adjusting the constants give the desired bounds. ▌

Remark. Graphs with high connectivity: The constructions used in this section produce graphs with a large number of cutpoints (every vertex of the line graph of a tree is either a simplicial vertex or a cutpoint). Nevertheless, the results obtained are not very sensitive to vertex connectivity, except for small ρ (less than a certain constant times n).

By multiplying each vertex of a graph G by the integer κ we obtain a κ-vertex connected graph G' with $t(G') = t(G)$, $n(G') = \kappa \cdot n(G)$ and $\rho(G') \leqslant \kappa^2(\rho(G) + n(G))$. Using this construction we can obtain, for example,

THEOREM 6.5. *For any integer $\kappa \geqslant 2$ there exists a constant $c(\kappa)$ so that for $c > c(\kappa)$ and sufficiently large n, there exists a graph G in $\mathscr{G}(n, cn)$ with vertex connectivity κ so that $t(G) = f(n, cn) + o(f(n, cn))$.*

Similar results can be formulated in cases where ρ grows faster than a constant times n.

VII. PROOFS OF THE LOWER BOUNDS ON $f(n, \rho)$

Proof of Theorem 4.5. Fix f; we show by induction on $\rho(G)$ that any connected graph G with $f \geqslant 2n(G)/(\rho(G)+2)$ has an induced tree on f vertices. The result is trivial if $\rho(G)=0$. Assume $\rho(G)>0$ and that the result holds for all G' with $\rho(G') < \rho(G)$. Let C be the subset of vertices that are contained in some cycle of G and let v and w be vertices in C whose distance from each other is maximum. Let $S(v)$ (resp. $S(w)$) denote the graph obtained from G by deleting the component of $G-v$ (resp. $G-w$) which contains w (resp. v).

Every vertex in $S(v)$ is farther from w than v is so $S(v) \cap C = \{v\}$, and, similarly $S(w) \cap C = \{w\}$, therefore $S(v)$ and $S(w)$ must induce trees in G. If $|S(v)| + |S(w)| \geqslant f$ then $S(v) \cup S(w) \cup \{$a shortest path from w to $v\}$ induces a tree in G of size at least f. So assume $|S(v)| + |S(w)| \leqslant f-1$ and, without loss of generality, $|S(v)| = b \leqslant (f-1)/2$. Note that $S(v)$ spans a graph with $b-1$ edges and v has at least 2 edges in $G - S(v)$, since $v \in C$. Thus $H = G - S(v)$ has $n-b$ vertices and $\rho(H) \leqslant \rho(G) - 1$. It is easy to verify then that $2n(H)/(\rho(H)+2) \geqslant f$ so, by induction, H, and therefore G, has an induced tree on f vertices. ∎

Proof of Theorem 4.6. Let $G = (V, E)$ with $n = n(G)$, $\rho = \rho(G)$, and $t = t(G)$. We show that t is bounded below by the quantity given in the theorem. It is routine to verify that this quantity is less than or equal to $2 \log \log n$, so it suffices to prove the bound under the assumption that $t \leqslant 2 \log \log n$. Let v^* be a center of G, let $r = \operatorname{rad}(G)$ and partition V into sets $V_0 = \{v^*\}$, $V_1, V_2, \ldots, V_r$, where $V_i = \{w \mid d(v^*, w) = i\}$. An edge between two vertices in the same block V_i is said to be *internal* to V_i. Let $n_i = |V_i|$ and let a_i be the number of edges internal to V_i.

If w is any vertex not equal to v^* and $w \in V_i$ then w is joined to some vertex in V_{i-1}. Thus at least $n-1$ edges of G are not internal so the total number of internal edges is at most ρ. In particular, we have

LEMMA 7.1. *For* $1 \leqslant i \leqslant r$, $a_i \leqslant \rho$.

LEMMA 7.2. *For* $1 \leqslant i \leqslant r$, V_i *contains no independent set of size* $n_{i-1}(t-1)+1$.

Proof. If $I \subseteq V_i$ is independent then since every vertex in I is joined to some vertex in V_{i-1} at least $|I|/(n_{i-1})$ of these vertices have a common neighbor in V_{i-1}, which induces a star on $|I|/(n_{i-1})+1$ vertices which can not be bigger than t, so $|I| \leqslant n_{i-1}(t-1)$. ∎

LEMMA 7.3. *For* $1 \leqslant i \leqslant r$, $a_i \geqslant (n_i/2)(n_i/n_{i-1}(t-1)-1)$.

Proof. The bound follows from Lemma 7.2 and the complementary form of Turan's theorem [6], which says that a graph on m vertices with no independent set of size $\alpha+1$ has at least $(m/2)(m/\alpha - 1)$ edges. ∎

We can now use these results to bound the size of V_k.

LEMMA 7.4. *If* $k \leqslant \log((\log \rho + \log\log\log n + 2)/(\log \rho - \log n + 5\log\log\log n + 4))$ *then*

$$n_k \leqslant (2t\rho)^{1-2^{-k}} \leqslant \frac{n}{4(\log\log n)^4}.$$

Proof. The second inequality is an unpleasant but simple calculation. (The inequality obtained by taking the log of both sides follows from the assumption $t \leqslant 2\log\log n$ and the fact that for k as hypothesized, $1-2^{-k} \leqslant (\log n - 4\log\log n - 2)/(\log \rho + \log\log n + 2))$. The first inequality is proved by induction on k. If $k=0$, it is trivial. Now suppose $k>0$ and $n_{k-1} \leqslant (2t\rho)^{1-2^{-k-1}}$. We show that $n_k \leqslant \sqrt{2n_{k-1}\rho t}$ which suffices to prove the induction step.

Case 1. $n_k \geqslant t^2 n_{k-1}$. By Lemmas 7.1 and 7.2, $\rho \geqslant a_k \geqslant (n_{k-1}/2)(n_k/(n_{k-1}(t-1))-1)$. Under the case assumption this is greater than $n_k^2/(2tn_{k-1})$ so $n_k \leqslant \sqrt{2t\rho n_{k-1}}$ as required.

Case 2. $n_k < t^2 n_{k-1}$. By the induction hypothesis and the second inequality, $n_{k-1} \leqslant n/4(\log\log n)^4$. Since ρ is at least $n/\log\log n$ and $t \leqslant 2\log\log n$, we have $n_{k-1} \leqslant 2\rho\log\log n/8(\log\log n)^4 \leqslant 2\rho/t^3$. Thus $t^4 n_{k-1}^2 \leqslant 2\rho n_{k-1} t$ so by the case assumption $n_k \leqslant \sqrt{2\rho n_{k-1} t}$. ∎

LEMMA 7.5. $r(G) \geqslant 1 + \log((\log \rho + \log\log\log n + 2)/(\log \rho - \log n + 5\log\log n + 4))$.

Proof. At least one of the V_i's has size at least $(n-1)/r$, since $V_1, \dots, V_r$ partition $V - v^*$. By Theorem 2.1, $(n-1)/r \geqslant (n-1)/(t+1)/2 > (2n-2)/(2\log\log n + 1)$. By Lemma 7.4, if $k < \log((\log \rho + \log\log\log n + 2)/(\log \rho - \log n + 5\log\log\log n + 4))$ then $n_k \leqslant n/4(\log\log n)^4$ which for $n \geqslant 4$ is at most $(2n-2)/(2\log\log n + 1)$. Thus, for k bounded as above no level has size $(n-1)/r$, so the radius must be at least $\log((\log \rho + \log\log\log n + 2)/(\log \rho - \log n + 5\log\log\log n + 4))$. ∎

Applying Theorem 2.1 and this corollary yields the desired bound on t and proves the theorem. ∎

It should be noted that the above proof shows that the graph G not only has a tree of the required size, but that it has such a tree that is either a path or a star.

VIII. BOUNDS ON $t(G)$ FROM THE INDEPENDENCE NUMBER

In this section we derive bounds on $t(G)$ in terms of the independence number $\alpha(G)$. We begin with a simple upper bound on $t(G)$.

PROPOSITION 8.1. *$t(G) \leqslant 2\alpha(G)$ and this bound is best possible.*

Proof. If T is a maximum induced tree of G then an independent set in T is independent in G. Since trees are bipartite we have $\alpha(G) \geqslant \alpha(T) \geqslant |T|/2 = t(G)/2$. For any α, the graph $G = P_{2\alpha}$ attains the bound. ▮

Next we consider the problem of using $\alpha(G)$ to get a lower bound on $t(G)$. Trivially, for $\alpha(G) = 1$, $t(G) = 2$ and for $\alpha(G) = 2$, $t(G) \geqslant 3$. In the proof of Proposition 4.7 it was shown that if $\alpha(G) = 3$, $t(G) \geqslant 4$. How large an independence number is needed to guarantee an induced tree of size 5? Surprisingly, no constant α is sufficient. In fact, as we will see there exist graphs on n vertices with $\alpha > n/2$ for which $t(G) = 4$.

THEOREM 8.2. *For any connected graph G with n vertices and any integer $1 \leqslant m \leqslant (n-1)/2$,*

$$\alpha(G) > \frac{(m-1)\,n}{m} + 1 \qquad \text{implies} \qquad t(G) \geqslant 2m + 1$$

$$\alpha(G) > \frac{(m-1)\,n + 1}{n} + 1 \text{ implies} \qquad t(G) \geqslant 2m + 2$$

and these bounds are best possible: for any $1 \leqslant m \leqslant (n-1)/2$, there exist graphs $G_1(m, n)$ and $G_2(m, n)$ on n vertices with $\alpha(G_1) > (m-1)\,n/m$ and $t(G_1) = 2m$ and $\alpha(G_2) > ((m-1)\,n + 1)/m$ and $t(G_2) \leqslant 2m + 1$.

Proof. We first construct graphs $G_1(m, n)$ and $G_2(m, n)$ to show the bounds are best possible. Write $n = qm + r$, where $1 < r \leqslant m$. The graph $G_1(m, n)$ consists of q stars with m vertices whose centers are connected as a clique C. The remaining r vertices are each connected to each vertex in the clique (See Fig. 8.1). All the vertices not in C are independent so $\alpha(G_1) = q(m-1) + r > ((m-1)/m)\,n$. The maximal induced trees of G consist of one vertex from C and all its neighbors outside C ($m + r$ total vertices) or two vertices of C together with their stars ($2m$ vertices).

If $m \equiv 1 \bmod n$ ($r \neq 1$ above) then the graph $G_1(m, n)$ also serves as $G_2(m, n)$. If, however, $m \equiv 1 \bmod n$ then $\alpha(G_1(m, n)) = ((m-1)\,n + 1)/m + 1$ so $G_1(m, n)$ does not meet the requirements for $G_2(m, n)$. In this case let $G_2(m, n)$ consist of $(q-1)$ stars of size m with centers connected as a clique with the remaining $m + 1$ vertices connected to the clique. This graph has $\alpha(G_2(m, n)) = ((m-1)\,n + 1)/m + 1$ and $t(G) = 2m + 1$ as required.

Next we show that the lower bounds on $t(G)$ hold for all graphs. Fix m; let G be a connected graph with vertex set V of size n and let I be an independent set in G of size $\alpha > (m-1)(n)/m + 1$. We show by induction on

n that $t(G) \geqslant 2m+1$ (a nearly indentical argument shows that $t(G) \geqslant 2m+2$ if $\alpha > ((m-1)n+1)/m+1$). If $n \leqslant 2m$ then the assumption about α gives $\alpha \geqslant n$ which is impossible. For $n = 2m+1$, then $\alpha \geqslant 2m$ and G must be a star of size $2m+1$. So suppose $n \leqslant 2m+1$.

Let v and w be vertices in $V - I$ such that the distance between them is maximum. Let $S(v)$ (resp. $S(w)$) denote the graph obtained by deleting from G the component of $G - v$ (resp. $G - w$) which contains w (resp. v). Then every vertex in $S(v) - v$ and $S(w) - w$ is in I (otherwise we could find a pair of vertices in $V - I$ with the distance between them greater than $d(v, w)$), so $S(v)$ and $S(w)$ are stars. If $|S(v)| + |S(w)| \geqslant 2m+1$ then $S(v) \cup S(w)$ together with a shortest path from w to v is the desired tree. Assume, therefore, that $|S(v)| + |S(w)| \leqslant 2m$ and, without loss of generality, $q = |S(v)| \leqslant m$. The graph $G' = G - S(w)$ is connected has $n' = n - q$ vertices and an independent set of size $\alpha' = \alpha + 1 - q$. We have

$$\alpha' = \alpha + 1 \frac{q}{m} \geqslant \alpha - q + \frac{q}{m} > \frac{(m-1)\,n}{m} + 1 - q + \frac{q}{m}$$

$$= \frac{(m-1)(n-q)}{m} + 1 = \frac{(m-1)\,n'}{m} + 1.$$

So G' has an independent set of the required size and by induction has a tree of size $2m+1$. ∎

IX. $t(G)$ for Graphs with no Large Cliques

The construction given in Section VII which produced sparse graphs on n vertices with largest induced tree of size $2 \log \log n$, prescribed graphs with very large cliques. It is natural to ask how small $t(G)$ can be for connected graphs with no large cliques. Define $c(n, k)$ to be the minimum of $t(G)$ over all graphs with n vertices having no complete subgraph on k vertices. We will prove:

Theorem 9.1. (i) $c(n, k) \geqslant 2 \log n/((k-2) \log \log n) - 3$ for $k \geqslant 3$, $n \geqslant 4$.

(ii) $c(n, 3) \leqslant c_1 \sqrt{n} \log n$ for some constant c_1.

(iii) $c(n, k) \leqslant 2 \log(n-1)/\log(k-2) + 2$ for $k \geqslant 4$.

This theorem can be restated as a "Ramsey"-type theorem:

Theorem 9.1′. For any positive integers $k \geqslant 3$ and $t \geqslant 2$ there exists a minimum integer $N(k, t)$ such that every connected graph on at least

$N(k, t)$ vertices has either a clique of size k or an induced tree of size t. Moreover,

(i) $N(k, t) < 2R(k-1, t)^{t+1/2}$, for $k \geqslant 3$ $t \geqslant 2$.

(ii) There exists a constant C_1 such that

$$N(3, t) \geqslant R\left(3, \frac{t}{2}\right) \geqslant C_1\,(t^2/(\log t)^2).$$

(iii) $N(k, t) \geqslant (k-2)^{(t/2-1)} + 1$ for $k \geqslant 4$.

Proof. The lower bound on $c(n, k)$ (and the upper bound on $N(k, t)$) is obtained by an argument similar to that used to prove Theorem 4.6. Let G be a graph on n vertices with no k clique and let $t = t(G)$. Let v^* be a center, let $r = r(G)$ and partition V into sets $V_0, V_1, ..., V_r$, where $V_i = \{v \mid d(v^*, v) = i\}$.

LEMMA 9.2. *For* $1 \leqslant i \leqslant r$, $n_i/(n_{i-1})$ *is less than the Ramsey number* $R(k-1, t)$.

Proof. Every vertex in V_i has a neighbor in V_{i-1} so some vertex in V_{i-1} has at least $n_i/(n_{i-1})$ neighbors in V_i. These $n_i/(n_{i-1})$ vertices cannot contain a clique of size $k-1$ (otherwise G has a k clique) nor can it have an independent set of size t or else G has a star on $t+1$ vertices. Thus $n_i/(n_{i-1}) < R(k-1, t)$. ∎

From this Lemma we get that $n_i \leqslant R(k-1, t))^i$ and $n = \sum_{i=0}^{r} n_i \leqslant \sum_{j=0}^{r} R(k-1, t)^j$. By Theorem 2.1 we have $r \leqslant (t+1)/2$ so $n \leqslant \sum_{j=0}^{(t+1)/2} R(k-1, t)^j \leqslant 2R(k-1, t)^{(t+1)/2}$, proving (i) in Theorem 9.1′.

Using the result of Erdös and Szekeres [3] that $R(a, b) \leqslant \binom{a+b-2}{a-1} \leqslant (b-1)^{a-1}$ we get $n \leqslant 2(t-1)^{(k-2)(t+1)/2}$. We want to show that t must be at least $2 \log n/((k-2) \log\log n) - 3$. We can assume that $t-1 \leqslant \log n$ otherwise the claim is immediate. Thus $n \leqslant 2(\log n)^{(k-2)(t+1)/2}$. Solving for t yields the desired lower bound of $(2 \log n)/((k-2) \log\log n) - 3$.

The upper bound on $c(n, 3)$ and the lower bound on $N(3, t)$ follow from Proposition 8.1. Since $t(G) \leqslant 2\alpha(G)$, if G has no independent set of size $t/2$ then G has no tree of size t. Thus $N(3, t) \geqslant R(3, t/2)$. Erdös [2] showed that there is a constant C_0 so that $R(3, s) \geqslant C_0(s^2/(\log s)^2)$, so there is a constant C_1 such that $N(3, t) \geqslant C_1(t^2/(\log t)^2)$ and a constant c_1 so that $c(n, 3) \leqslant c_1\sqrt{n} \log n$.

The upper bound on $c(n, k)$ (and the lower bound on $N(k, t)$) for $k \geqslant 3$ is obtained by using the line graph of the regular tree

$B(k-1, (n-1)/(k-2))$, with some leaves deleted if necessary to make the number of vertices equal to n. Lemma 6.1 bounds the size of its largest induced tree at $2(\log(n-1)/(\log(k-2))+2$. ∎

X. Some Open Problems

The upper and lower bounds proved for $c(n, k)$ in the last section are not tight, particularly in the case $k=3$. Here the bounds were $O(\log n/\log\log n) \leqslant c(n, 3) \leqslant O(\sqrt{n}\log n)$. It would be interesting to improve these bounds.

Although we gave an example where $f(n, \rho)$ can decrease when ρ is increased, this occurs at a value of ρ where $f(n, \rho)=3$. If $f(n, \rho_1) \geqslant 4$ and $\rho_2 < \rho_1$, can $f(n, \rho_2)$ be greater than $f(n, \rho_1)$? Also, if $f(n, \rho) \geqslant 4$, is $f(n, \rho)$ always minimized by the line graph of a tree?

Finally, for esthetic reasons, it would be nice to fill the last little gap in our description of $f(n, \rho)$ and determine the exact leading order behavior of $f(n, cn/\log\log n)$.

References

1. B. Bollobas, "Extremal Graph Theory," Academic Press, London, 1978.
2. P. Erdös, Graph theory and probability, II, *Canad. J. Math.* (1961), 346–352.
3. P. Erdös and G. Szekeres, A combinatorial problem in geometry, *Compositio Math.* **2** (1935), 463–470.
4. M. R. Garey and D. J. Johnson, "Computers and Intractibility: A Guide to the Theory of *NP*-Completeness," Freeman, San Francisco, 1979.
5. R. L. Graham, B. Rothchild, and J. Spencer, "Ramsey Theory," Wiley, New York, 1980.
6. P. Turan, On an extermal problem in graph theory, *Mat. Fiz. Lapok* **48** (1941), 436–452.