Reprinted from Canadian Mathematical Bulletin, Volume 13, Number 3, September 1970

BULLETIN **13** (3), 1970

ON SUBTREES OF DIRECTED GRAPHS WITH NO PATH OF LENGTH EXCEEDING ONE

BY
R. L. GRAHAM

The following theorem was conjectured to hold by P. Erdös [1]:

THEOREM 1. *For each finite directed tree T with no directed path of length 2, there exists a constant $c(T)$ such that if G is any directed graph with n vertices and at least $c(T)n$ edges and n is sufficiently large, then T is a subgraph of G.*

In this note we give a proof of this conjecture. In order to prove Theorem 1, we first need to establish the following weaker result.

THEOREM 2. *For each finite directed tree T with no directed path of length 2, there exists a constant $c'(T)$ such that if G is any directed graph with no directed path of length 2, n vertices and at least $c'(T)$ edges, and n is sufficiently large, then T is a subgraph of G.*

Proof of Theorem 2. First note that if G has no directed path of length 2, then each vertex of G is either a *source* (all edges directed out), a *sink* (all edges directed in), or *isolated.*

Define the graph $A(d, k)$ for $d \geq 2$, $k \geq 0$, as follows:

$A(d, 0)$ consists of a single isolated vertex p.

$A(d, k)$ is formed from $A(d, k-1)$ by adjoining to each vertex of degree 1, d new edges and vertices so that the resulting graph still has no path of length 2, where for $k=1$ we take p to be a *source.*

Thus, $A(d, k)$ consists of the vertex p surrounded by k alternating layers of sinks and sources (cf. Figure 1).

The jth layers of $A(d, k)$ consists of d^j vertices. We note the immediate

Fact. If T is a directed tree with no directed path of length 2, if the longest *undirected* path in T has length m, and if the maximal degree of a vertex of T is d, then T is a subgraph of $A(d, m+1)$.

We now prove by induction on k that Theorem 2 holds for $T=A(d, k)$. By the preceding fact, this is sufficient to establish Theorem 2 for general T.

For $k=0$, this is immediate. Assume the result holds for a fixed $k \geq 0$ and all d. Let D denote $1+d+d^2+\cdots+d^k$, the total number of vertices of $A(d, k)$ and let $M=D+d$. Let C denote $c'(A(d, k))+d^kM$ which exists by the induction hypothesis. Suppose G is a graph with no directed path of length 2, n vertices and at least Cn edges, where n is a large integer to be specified later. Assume further that k is *even*

Received by the editors November 11, 1969.

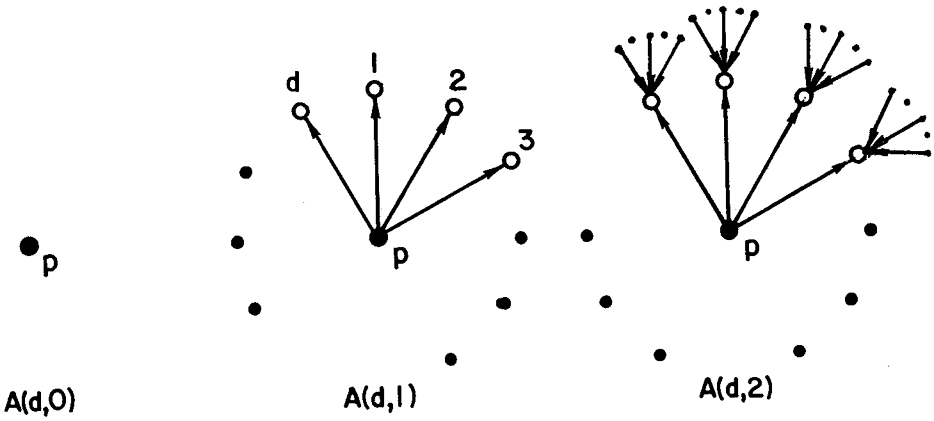

FIG. 1

(the case of k odd is similar and will be omitted). Form the subgraph G' of G by deleting from G all *source* vertices of degree $\leq d^k M$, of which there are, say, u of these, and their incident edges. Note that this operation does not decrease the degree of any vertex of G of degree $>d^k M$. By construction, in G' all source vertices have degree $>d^k M$. By the choice of C, we have $u<n$. Since we have removed at most $ud^k M$ edges from G in forming G', then G' has $n-u$ vertices and at least

$$\begin{aligned} Cn-ud^kM &\geq c'(A(d,k))n+(n-u)d^kM \\ &\geq c'(A(d,k))n \\ &\geq c'(A(d,k))(n-u) \end{aligned}$$

edges. Since G' has less than $(n-u)^2$ edges then

$$(n-u)^2 > c'(A(d,k))n$$

and

$$n-u > \sqrt{c'(A(d,k))n}.$$

For n sufficiently large, $n-u$ becomes arbitrarily large and we may apply the induction hypothesis to G'. This implies that G' contains a copy of $A(d,k)$ as a subgraph. Let us examine the outside layer of vertices of this subgraph $A(d,k)$, i.e., the vertices of degree 1. Since k is even (by assumption), these vertices are sources. Denote them by $v_1, v_2, \ldots, v_{d^k}$. With each v_i, we associate the set S_i of vertices of G' which are adjacent to v_i. That is, $s \in S_i$ if and only if (v_i, s) is an edge of G'. By the construction of G', $|S_i|>d^kM$. It is not difficult to see that this implies that we can extract a *system* of disjoint *representative subsets* R_i, $1\leq i\leq d^k$, i.e., a set of subsets such that:

(i) $R_i \cap R_j = \varnothing$ for $i \neq j$,

(ii) $R_i \subseteq S_i$, $1 \leq i \leq d^k$,

(iii) $|R_i| = M$, $1 \leq i \leq d^k$.

Finally, form R_i' from R_i by deleting all vertices which lie in the subgraph $A(d, k) \subseteq G'$. Thus, $|R_i'| \geq M - D = d$ for $1 \leq i \leq d^k$. By reconnecting the vertices of the R_i' to the subgraph $A(d, k)$ so that they are sinks, we see that we have $A(d, k+1) \subseteq G' \subseteq G$. The case for odd k is similar. This completes the induction step and Theorem 2 is proved.

Proof of Theorem 1. Let G be a directed graph with n vertices and at least $2c'(A(D+d, k))n$ edges. We shall show that for n sufficiently large, $A(d, k)$ is a subgraph of G. By choosing $c(A(d, k)) = 2c'(A(D+d, k))$, Theorem 1 will then be established for $T = A(d, k)$, and by a previous remark, this suffices to prove it for general T.

We can assume G has no isolated vertices (for otherwise they may be deleted without harm). Form the graph G^* from G by the following operation: Replace each vertex v of G by a *pair* of vertices v', v'' such that all directed edges going into v now go into v', and all directed edges going away from v now go away from v'' (cf. Figure 2). The vertices v' and v'' will be called *mates* of one another.

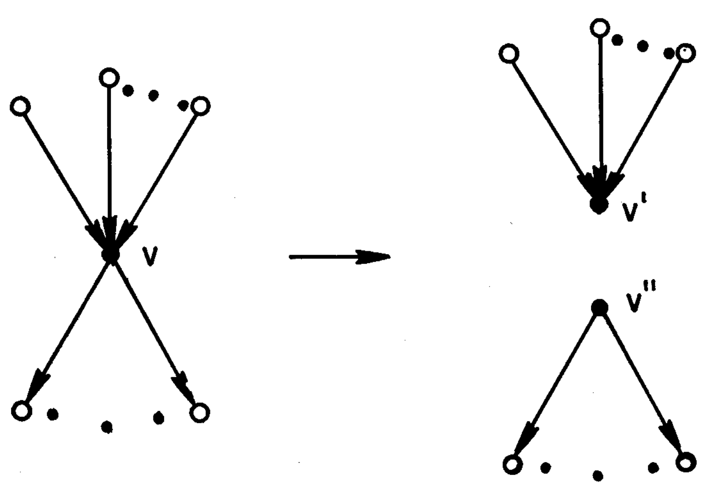

FIG. 2

G^* has the property that it has no path of length 2, it has $n^* \leq 2n$ vertices and at least

$$2c'(A(D+d, k))n \geq c'(A(D+d, k))n^*$$

edges. Hence, for n sufficiently large, we may apply Theorem 2 to G^*. This implies that G^* contains the subgraph $A(D+d, k)$.

We next recursively delete certain vertices and edges from G^* as follows:

(1) Delete from $A(D+d, k) \subseteq G^*$ the mate $m(p)$ of p (the central vertex of $A(D+d, k)$), all edges incident to $m(p)$ and all other vertices and edges of $A(D+d, k)$ which are not connected to p after the deletion of $m(p)$.

(2) Next select d of the remaining first level vertices of $A(D+d, k)$, say, $u_1, u_2, \ldots, u_d$, and delete all the other first level vertices, incident edges and new components formed by these deletions.

(3) For each of the u_i, $1 \leq i \leq d$ (which are sinks) delete from what is currently left of $A(D+d, k)$ the *mates* $m(u_i)$ of the u_i, all incident edges and all newly formed components (i.e., vertices and edges not connected to p). Since each u_i is originally adjacent to $D+d \geq 1+d+d$ vertices in the second level, then after this deletion each u_i is now still adjacent to at least d vertices on the second level.

(4) For each u_i, select d of the second level vertices to which it is adjacent, say, $u_{i1}, u_{i2}, \ldots, u_{id}$, and delete all remaining second level vertices, incident edges and new components.

(ω) We can continue this construction since $D = 1+d+\cdots+d^k$ until we have finally constructed by selective deletions a copy of $A(d, k)$ with the important property that this $A(d, k)$ *does not contain both a vertex and its mate*. This, however, is sufficient to guarantee that $A(d, k)$ is a subgraph of the original graph G. This completes the proof of Theorem 1.

Reference

1. P. Erdös, (personal communication).

Bell Telephone Laboratories, Inc.
Murray Hill, New Jersey