An Extremal Result for Paths [a]

PAUL ERDŐS,[b] R. J. FAUDREE,[c] R. H. SCHELP,[c,d] AND
M. SIMONOVITS[e]

[b]*AT&T Bell Laboratories*
Murray Hill, New Jersey 07974
and
[e]*Mathematical Institute*
Hungarian Academy of Sciences
Budapest, Hungary

[c]*Department of Mathematical Sciences*
Memphis State University
Memphis, Tennessee 38152

INTRODUCTION

One of the best known extremal results involving paths is the following one proved more than 25 years ago.

THEOREM [2]: A graph G_m on m vertices with at least $[m(k-1)+1]/2$ edges contains a path P_{k+1} on $k+1$ vertices. Furthermore, when $m = kt$ the graph tK_k contains the maximal number of edges in an m vertex graph with no P_{k+1} and is the unique such graph.

There are many other results in the literature that use that a graph with many edges or with high-degree vertices has a long path. Several such results are given in the references [1–4].

The problem we address here is of a similar nature. Let m, n, and k be fixed positive integers with $m > n \geq k$. We wish to determine the minimum value l such that each graph on m vertices with l vertices of degree at least n contains a P_{k+1}.

A plausible minimum value for l is suggested by the following graph. Let $m = t(n+1)+r$, $0 \leq r < n+1$, with $k < 2n+1$. Then the graph consisting of t vertex disjoint copies of $H = \overline{K}_{n+1-\lfloor (k-1)/2 \rfloor} + K_{\lfloor (k-1)/2 \rfloor}$ contains $t\lfloor (k-1)/2 \rfloor$ vertices of degree n and no P_{k+1}. When k is even and $r + \lfloor (k-1)/2 \rfloor \geq n$, the number of vertices of degree $\geq n$ in this graph can be increased by 1 to $t\lfloor (k-1)/2 \rfloor + 1$ without forcing the graphs to contain a P_{k+1}. Simply take one of the vertices of degree $\lfloor (k-1)/2 \rfloor$ and make it adjacent to the r vertices in no copy of H.

Thus we have the following conjecture.

CONJECTURE: Let m, n, and k be fixed positive integers with $m > n \geq k$ and set $\delta = 2$ when k is even and $\delta = 1$ when k is odd. If G_m is a graph on m vertices and at least $l = \lfloor (k-1)/2 \rfloor \lfloor m/(n+1) \rfloor + \delta$ vertices of degree $\geq n$, then G_m contains a P_{k+1}.

[a]The research of one of the authors (R.J.F.) was supported by Office of Naval Research Grant N00014-88-K-0080. The research of another author (R.H.S.) was partially supported by National Science Foundation Grant DMS-8603717.

[d]To whom correspondence should be addressed.

Although we do not prove the conjectured result, we do show that the value of l given in the conjecture is "essentially" correct. Much attention is given to the special case when $n + 1 \le m \le 2n + 1$. In this case we show that approximately $k/2$ vertices of degree $\ge n$ is enough to guarantee that G_m contains a P_{k+1}. Unfortunately, even for this interval of values we are not able to prove the exact statement of the conjecture.

It should be mentioned that the problem considered is not of much interest when $k \ge n + 1$. In fact if m is such that there exist $l_1, l_2, \ldots, l_s$ with $n + 1 \le l_i \le k$ for each i and $\Sigma_{i=1}^{s} l_i = m$, then $K_{l_1} \cup K_{l_2} \cup \cdots \cup K_{l_s}$ is a graph that has all its vertices of degree $\ge n$, yet contains no P_{k+1}. This can always be done if m is large enough.

RESULTS

Before presenting the results we introduce some nonstandard notation. The symbol G_m will always represent the graph of interest, which is assumed to have m vertices. The vertices of G_m will be partitioned into two classes, those of degree $\ge n$, which will be called *high-degree* vertices, and the remaining vertices, which will be called *low-degree* vertices. A path is called a *high–low* path if it begins and ends with high-degree vertices and alternates between high- and low-degree vertices as one moves from one end of the path to the other. Further a path is called a *high-end* path if it is simply a path beginning and ending with high-degree vertices; nothing is assumed about the degree of the internal vertices of the path. Also degenerate paths are allowed, so that a single high-degree vertex is thought of as both a high-end path and a high–low path. Whenever E is affixed as a superscript on the usual symbol for a path, it is assumed the path is high-end. Thus P_i^E will denote a high-end path on i vertices.

A principal result of the paper establishes that the conjecture is approximately correct when $m \le 2n + 1$. In order to prove this we need two lemmas.

LEMMA 1: Let $P^{(1)}, P^{(2)}, \ldots, P^{(t)}$ be a vertex disjoint collection of high-end paths in G_m, $m \le 2n + 1$. Then there exists a high-end path P containing each of the paths $P^{(1)}$, $P^{(2)}, \ldots, P^{(t)}$.

COROLLARY 1: There exists a high-end path P in G_m, $m \le 2n + 1$, containing all high-degree vertices of G_m.

A bit more notation is needed to state the second lemma. Consider a vertex disjoint family $P^{(1)}, P^{(2)}, \ldots, P^{(t)}$ of high–low paths in G_m. Partition the vertices of these paths into two sets H and L, where H consists of all the high-degree vertices of $\bigcup_{i=1}^{t} P^{(i)}$ and L all low-degree vertices. Thinking of the vertices of each path $P^{(i)}$ as numbered from left to right, let R denote the set of right-hand end-vertices of the t paths. Let H' be those high-degree vertices that are not right-hand end-vertices of some $P^{(i)}$, that is, let $H' = H - R$. For each $h' \in H'$ let l' be that low-degree vertex in L that follows h' on some path $P^{(i)}$. A vertex h of H is *good* if $h \in R$ or if $h = h' \in H'$ and l' is adjacent to at least three vertices of R. As usual $V(G)$ will denote the vertex set of the graph G.

LEMMA 2: Let $P^{(1)}, P^{(2)}, \ldots, P^{(t)}$ be a vertex-disjoint family of high–low paths in G_m, and let $h_1, h_2 \in H$ be a pair of good vertices. Then there exist t vertex-disjoint high–low paths $Q^{(1)}, Q^{(2)}, \ldots, Q^{(t)}$ such that $\bigcup_{i=1}^{t} V(Q^{(i)}) = \bigcup_{i=1}^{t} V(P^{(i)})$ with h_1 and h_2 end-vertices of some $Q^{(i)}$ and $Q^{(j)}$, respectively, $i \ne j$.

THEOREM 1: Let G_m contain at least k high-degree vertices, with $m \leq 2n + 1$. Then G_m contains

$$\begin{cases} P^E_{2k-7}, & \text{when } k \leq n/2 + 3 \\ P^E_{n-1}, & \text{when } (n+1)/2 + 3 \leq k \leq n \\ P^E_{k}, & \text{when } k \geq n + 1. \end{cases}$$

COROLLARY 2: Under the conditions of the theorem, G_m contains a

$$\begin{cases} P_{2k-5}, & \text{when } k \leq n/2 + 3 \\ P_{n+1}, & \text{when } (n+1)/2 + 3 \leq k \leq n \\ P_{k}, & \text{when } k \geq n + 1. \end{cases}$$

COROLLARY 3: Let G_m contain at least k high-degree vertices, $m \leq 2n$. Then G_m contains a C_l where

$$\begin{cases} l \geq 2k - 7, & \text{when } k \leq n/2 + 3 \\ l \geq n - 1 & \text{when } (n+1)/2 + 3 \leq k \leq n \\ l \geq k & \text{when } n + 1 \leq k. \end{cases}$$

The graph $(K_k + \overline{K}_{n+1+k}) \cup \overline{K}_{m-n-1}$ shows (when $k \leq n$) that these results are close to the best possible. For $k \geq n + 1$ the graph K_k shows P_k is the longest path possible.

As more evidence that the conjecture is correct we prove the following theorem.

THEOREM 2: Let k be a positive integer. Then there exists a constant c such that if n is large enough with respect to k, each graph G_m, $m > n$, with at least $[m/(n+1)]$ $[(k-1)/2] + c$ vertices of degree $\geq n$ contains a P_{k+1}.

In the proof of the theorem we in fact show $c \geq 4$ works when $n \geq k^2 - 3k + 1$.

PROOFS

Proof of Lemma 1: Let $P^{(1)}, P^{(2)}, \ldots, P^{(l)}$ be a vertex disjoint family of high-end paths in G_m. Consider a vertex disjoint family $Q^{(1)}, Q^{(2)}, \ldots, Q^{(j)}$ of high-end paths whose vertex set includes all vertices of the family $P^{(1)}, P^{(2)}, \ldots, P^{(l)}$ and is chosen such that j is minimal. We need only prove $j = 1$.

Suppose $j > 1$ and let

$$Q^{(1)} - x_1, x_2, \ldots, x_{l_1} = u_1 \qquad \text{and} \qquad Q^{(2)} - y_1, y_2, \ldots, y_{l_2} = u_2.$$

Let U_1 and U_2 be the set of vertices adjacent to u_1 and u_2, respectively, and set

$$\begin{aligned} U_2' = \{z \mid\, & z = x_i \text{ if } x_{i+1} \in Q^{(1)} \cap U_2, \\ & z = y_{i+1} \text{ if } y_i \in Q^{(2)} \cap U_2, \text{ or} \\ & z \in U_2 - (Q^{(1)} \cup Q^{(2)})\}. \end{aligned}$$

Since j is minimal, it is easy to check that $u_1, y_1 \notin U_1 \cup U_2'$, $|U_2| = |U_2'|$, and $U_1 \cap U_2' = \phi$. Therefore, $|U_1 \cup U_2'| = |U_1| + |U_2| \leq m - 2 \leq 2n - 1$, contradicting that $|U_1|, |U_2| \geq n$. Hence, $j = 1$ and the proof is complete. □

Note that Corollary 1 is an immediate consequence of Lemma 1, since single high-degree vertices are considered to be high–low paths.

Proof of Lemma 2: Let $P^{(1)}, P^{(2)}, \ldots, P^{(t)}$ be a vertex disjoint family of high–low paths in G_m and let $h_1, h_2 \in H$ be a pair of good vertices. There are several cases to consider.

Case 1: $h_1, h_2 \in H'$ and lie on the same path.

Without loss of generality assume that h_1 and h_2 are vertices of $P^{(1)}$ and that l_1 and l_2 (the successors of h_1 and h_2, respectively) are adjacent to the right-hand end-vertices of $P^{(2)}$ and $P^{(3)}$, respectively. (Note that a good vertex in H' requires $t \geq 3$.) Set

$$P^{(1)} = x_1, x_2, \ldots, x_i = h_1, x_{i+1} = l_1, x_{i+2}, \ldots, x_j = h_2, x_{j+1} = l_2, x_{j+2}, \ldots, x_{r_1},$$
$$P^{(2)} = y_1, y_2, \ldots, y_{r_2},$$
$$P^{(3)} = z_1, z_2, \ldots, z_{r_3}.$$

Replace these three paths by

$$Q^{(1)} = x_1, x_2, \ldots, h_1,$$
$$Q^{(2)} = y_1, y_2, \ldots, y_{r_2}, l_1 = x_{i+1}, x_{i+2}, \ldots, x_j = h_2,$$
$$Q^{(3)} = z_1, z_2, \ldots, z_{r_3}, l_2 = x_{j+1}, x_{j+2}, \ldots, x_{r_1}.$$

These three paths together with $P^{(4)}, \ldots, P^{(t)}$ give the required family.

Case 2: $h_1, h_2 \in H'$ and lie on different paths.

Without loss of generality assume h_1 is on path $P^{(1)}$, h_2 is on path $P^{(2)}$, and let l_1 and l_2 be the successors of h_1 and h_2, respectively. If l_1 and l_2 are adjacent to different right-hand end-vertices among the paths $P^{(3)}, P^{(4)}, \ldots, P^{(t)}$, four new paths are found in a fashion similar to that described in Case 1. The only other possibility is that both l_1 and l_2 are adjacent to the right-hand end-vertices of the first three paths and to no others. Thus in this case we can assume that

$$P^{(1)} = x_1, x_2, \ldots, x_i = h_1, x_{i+1} = l_1, x_{i+2}, \ldots, x_{r_1},$$
$$P^{(2)} = y_1, y_2, \ldots, y_j = h_2, y_{j+1} = l_2, y_{j+2}, \ldots, y_{r_2},$$
$$P^{(3)} = z_1, z_2, \ldots, z_{r_3}.$$

Then replace these paths by

$$Q^{(1)} = x_1, x_2, \ldots, x_j = h_1,$$
$$Q^{(2)} = y_1, y_2, \ldots, y_j = h_2,$$
$$Q^{(3)} = x_{r_1}, \ldots, x_{i+1} = l_1, y_{r_2}, \ldots, y_{j+1} = l_2, z_{r_3}, z_{r_3-1}, \ldots, z_1.$$

These paths together with $P^{(4)}, \ldots, P^{(t)}$ again give the required family.

Case 3: $h_1 \in H'$ and $h_2 \in R$ and lie on the same path.

Without loss of generality assume

$$P^{(1)} = x_1, x_2, \ldots, x_i = h_1, x_{i+1} = l_1, x_{i+2}, \ldots, x_{r_1} = h_2,$$
$$P^{(2)} = y_1, y_2, \ldots, y_{r_2}$$

with l_1 adjacent to y_{r_2}. Replace these paths by

$$Q^{(1)} = x_1, x_2, \ldots, x_i = h_1$$
$$Q^{(2)} = y_1, y_2, \ldots, y_{r_2}, l_1 = x_{i+1}, x_{i+2}, \ldots, x_{r_1} = h_2$$

to obtain the desired family.

Case 4: $h_1 \in H'$ and $h_2 \in R$ and lie on different paths.

We may assume

$$P^{(1)} = x_1, x_2, \ldots, x_i = h_1, x_{i+1} = l_1, x_{i+2}, \ldots, x_{r_1},$$
$$P^{(2)} = y_1, y_2, \ldots, y_{r_2} = h_2,$$
$$P^{(3)} = z_1, z_2, \ldots, z_{r_3}$$

with l_1 adjacent to z_{r_3}. Then replace $P^{(1)}$ and $P^{(3)}$ by

$$Q^{(1)} = x_1, x_2, \ldots, x_i = h_1,$$
$$Q^{(3)} = z_1, z_2, \ldots, z_{r_3}, l_1 = x_{i+1}, x_{i+2}, \ldots, x_{r_1}$$

to obtain an appropriate family.

Case 5: $h_1, h_2 \in R$.

There is nothing to prove, since the original set of paths $P^{(1)}, P^{(2)}, \ldots, P^{(t)}$ fulfill the required conditions. □

Proof of Theorem 1: Partition the set of vertices of G_m into two sets H and L where H denotes the set of high-degree vertices and L the low-degree vertices of the graph. Form a minimal family (minimal number of paths) of disjoint high–low paths $P^{(1)}$, $P^{(2)}, \ldots, P^{(t)}$ such that all vertices of H are included in this family of paths. Recall that this is possible, since single vertices are allowed as high–low paths. Let L_1 be the subset of L of low-degree vertices used in the family of paths and let $L_2 = L - L_1$ be the remaining low-degree vertices.

Observe that $|L_1| + t = |H|$ and by Lemma 1 that there exists a high-end path P on $|H| + |L_1|$ vertices. Since $|H| + |L_1| = 2|H| - t$ and $|H| \geq t$, it follows that P has at least k vertices. Hence the theorem holds when $k \geq n + 1$.

We assume for the remainder of the proof that $5 \leq k \leq n$. Note that when $k \leq 4$ there is nothing to prove. Choose k_1 and x such that $k_1 + x = k = |H|$ with $k_1 = (n + 1)/2$ when $k \geq (n + 1)/2 + 3$, and $x = 3$ when $k \leq n/2 + 3$.

Since P has $|H| + |L_1| = 2k - t = 2k_1 + 2x - t$ vertices, the proof is complete if

$$2k_1 + 2x - t \geq \begin{cases} n - 1, & \text{when } k \geq (n + 1)/2 + 3 \\ 2k - 7, & \text{when } k \leq n/2 + 3. \end{cases} \tag{1}$$

This means that we can assume this inequality fails so that $t \geq 2x + 3$ when $k \geq (n + 1)/2 + 3$, and $t \geq 8 - 2x + 2$ when $k \leq n/2 + 3$.

To continue the proof we make some additional observations. Let s be the number of good vertices, in the sense of Lemma 2, in the paths $P^{(1)}, P^{(2)}, \ldots, P^{(t)}$. From the minimality of t it follows from Lemma 2 that each vertex of L_2 is adjacent to at most one of the s good vertices. Further, since the good vertices are of high-degree, each of these vertices has at least $n - |L_1| - |H| + 1$ adjacencies to vertices of L_2. Hence

$$s(n - |L_1| - |H| + 1) \leq |L_2| - m - 2k + t. \tag{2}$$

Also for each vertex h of H that is not good the vertex that follows it on its path $P^{(i)}$ is adjacent to at most two of the t right-hand end-vertices of the paths $P^{(1)}, P^{(2)}, \ldots, P^{(t)}$. But then these t right-hand end-vertices have a total of at least $(|H| - s)(t - 2) + t(n - |L_1| - |H| + 1)$ adjacencies to the vertices of L_2. From the minimality of t we obtain

$$(|H| - s)(t - 2) + t(n - |L_1| - |H| + 1) \leq |L_2| - m - 2k + t. \tag{3}$$

To complete the proof we need only show that under the assumed conditions, $t \geq 2x + 3$ when $k \geq (n + 1)/2 + 3$ and $t \geq 8 - 2x + 2$ when $k \leq n/2 + 3$, either inequality (2) or (3) fails to hold. Checking that this is the case amounts to looking at the number of good vertices in H. It is straightforward to show that inequality (2) fails when there are at least $(2k + t)/3$ good vertices, while inequality (3) fails when there are less than $(2k + t)/3$ good vertices. This completes the proof of the theorem. □

Corollary 2 is an immediate consequence of Theorem 1, since for $k \leq n$ each of the end-vertices of the existing high-end path has an additional adjacency off the path.

Proof of Corollary 3: Consider the high-end path P_i^E,

$$i = \begin{cases} 2k - 7, & \text{when } k \leq n/2 + 3 \\ n - 1, & \text{when } (n + 1)/2 + 3 \leq k \leq n, \text{ guaranteed by Theorem 1} \\ k, & \text{when } k \geq n + 1. \end{cases}$$

Let the vertices of this path be $u = x_1, x_2, \ldots, x_i = v$ with U the set of neighbors of u and V the set of neighbors of v. Set $V' = \{z \mid z = x_{j+1} \text{ if } x_j \in V \text{ or } z \in V - P_i^E\}$. It is easy to see if $V' \cap U \neq \phi$, then G_m contains a C_l, $l \geq i$. Also, $|V'| = |V|, |U| \geq n$ with $u \notin V' \cup U$. But then $|V' \cup U| \leq m - 1 \leq 2n - 1$, so that $V' \cap U \neq \phi$ and G_m contains the required cycle. □

Proof of Theorem 2: Assume n is considerably larger than k. In fact, we see in what follows that $n \geq k^2 - 3k + 1$ will suffice. Corollary 2 implies the result of this theorem when $m \leq 2n + 1$ and $c \geq 4$. Hence we assume $m > 2n + 1$ and that the theorem holds when the graph has less than m vertices.

Choose a maximal-length high-end path P on l vertices in G_m. Let this path P be $u = x_1, x_2, \ldots, x_l = v$. We suppose G_m contains no path on $k + 1$ vertices and reach a contradiction. Since P is a high-end path, this means $l \leq k - 2$.

Let $N(u)$ and $N(v)$ be the set of neighbors of the high-degree vertices u and v, respectively. Observe that neither $N(u) - P$ nor $N(v) - P$ contains high-degree vertices, since P is a high-end path of maximal length. Further

$$|N(u) - P|, |N(v) - P| \geq n - k + 3.$$

There are two possibilities to consider.

Case 1: $(N(u) - P) \cap (N(v) - P) = \phi$. Let G' be the graph obtained from G_m by deleting the vertices of $(N(u) - P) \cup (N(v) - P)$. Note that no vertex outside P of high-degree in G_m has adjacencies into the set $(N(u) - P) \cup (N(v) - P)$. Thus G' has at most $m - 2n + 2k - 6$ vertices and at least $[m/(n + 1)]\ [(k - 1)/2] + c - (k - 2)$ high-degree vertices. Since $n \geq k^2 - 3k + 1$,

$$\left[\frac{m - 2n + 2k - 6}{n + 1}\right]\left[\frac{k - 1}{2}\right] + c \leq \left[\frac{m}{n + 1}\right]\left[\frac{k - 1}{2}\right] + c - (k - 2),$$

and G' contains a P_{k+1}, a contradiction to the supposition that G_m contains no P_{k+1}.

Case 2: $(N(u) - P) \cap (N(u) - P) \neq \phi$. Let $w \in (N(u) - P) \cap (N(v) - P)$. Then $w, u = x_1, x_2, \ldots, x_l = v, w$ is a C_{l+1} in G_m. Clearly no two consecutive vertices on C_{l+1} are of high-degree; otherwise, G_m contains a high-end path on $l + 1$ vertices. Thus C_{l+1} contains at most $(l + 1)/2 \leq (k - 1)/2$ high-degree vertices. Also as in Case 1 no vertex outside P of high-degree has an adjacency into the set $(N(u) - P) \cup (N(v) - P)$. If C_{l+1} has fewer than $(k - 1)/2$ vertices of high-degree, then let G' be the graph obtained from G_m by deleting the vertices of $N(u) - P$, while if C_{l+1} has precisely $(k - 1)/2$ vertices of high-degree, then let G' be graph obtained by deleting the vertices of $N(u) \cup P$. In each case the number of vertices of high-degree in G' is at least $[m/(n + 1)]\ [(k - 1)/2] + c - z$ where z is the number of high-degree vertices on C_{l+1}. This is true since when C_{l+1} has exactly $(k - 1)/2$ vertices of high-degree, each high-degree vertex outside P has no adjacency to vertices of P. But in each of these two cases

$$\left[\frac{|G'|}{n + 1}\right]\left[\frac{k - 1}{2}\right] + c \leq \left[\frac{m}{n + 1}\right]\left[\frac{k - 1}{2}\right] + c - z$$

so that G' contains a P_{k+1}, a contradiction. This contradiction completes the proof of the theorem. □

QUESTIONS

A natural question concerns the extension of the result of Corollary 3 for cycles. In fact, does the graph obtained by identifying appropriate vertices from $\lfloor (m - 1)/n \rfloor$ copies of $H = \overline{K}_{n+1-\lfloor (k-1)/2 \rfloor} + K_{\lfloor (k-1)/2 \rfloor}$, one vertex from each copy of H, suggest the magnitude of the number of vertices of high-degree necessary for a graph G_m to contain a C_l, $l \geq k$? It is possible that the following holds. If $k \leq n$, and G_m contains no C_l ($l \geq k$), then G_m has at most $\lfloor (k-1)/2 \rfloor \lfloor (m-1)/n \rfloor + 1$ vertices of degree $\geq n$.

Another question related to the original conjecture occurs when the graph G_m is assumed to be connected. The graph, consisting of $\lfloor (m/(n+1) \rfloor$ copies of $H = \overline{K}_{n+1-\lfloor (k+1)/4 \rfloor} + K_{\lfloor (k+1)/4 \rfloor}$ identified at a fixed vertex of each $K_{\lfloor (k+1)/4 \rfloor}$, contains no P_{k+1} but does contain $\lfloor m/(n+1) \rfloor$ $\lfloor (k+1)/4 \rfloor$ high-degree vertices. This is approximately half of the number of high-degree vertices in the original conjecture. Is there a better extremal example, or does connectivity lower the extremal number of the conjecture by a factor of 2?

REFERENCES

1. Dirac, G. H. 1963. Extensions of Turan's theorem on graphs, Acta Math. Acad. Sci. Hung. **14:** 417–422.
2. Erdős, P. & T. Gallai. 1959. On maximal paths and circuits in graphs. Acta Math. Acad. Sci. Hung. **10:** 337–356.
3. Faudree, R. J. & R. H. Schelp. 1975. Path-path numbers in multicolorings. J. Comb. Theory B. **19:** 150–160.
4. Pósa, L. 1976. Hamiltonian circuits in random graphs. Discrete Math. **14:** 359–364.