SPANNING TREES WITH MANY LEAVES

Guoli Ding[*]

Department of Mathematics
Louisiana State University
Baton Rouge, LA 70803

Thor Johnson

and

Paul Seymour[†]

Department of Mathematics
Princeton University
Princeton, NJ 08544

October 1997
Revised April 1999

---

[*] Partially supported by NSF Grant DMS-9700623 and by DIMACS and Princeton University.

[†] Supported by ONR Grant N0014-97-1-0512 and NSF Grant DMS97-01598.

ABSTRACT

We show that if G is a simple connected graph with

$$|E(G)| \geq |V(G)| + \frac{1}{2}t(t-1)$$

and $|V(G)| \neq t+2$, then G has a spanning tree with $> t$ leaves, and this is best possible.

Given a connected graph G, determining the maximum number of leaves in a spanning tree of G is known to be NP-hard [1]. (All graphs in this paper are simple and finite. A *leaf* is a vertex with degree 1.) But we may ask for lower bounds on this number.

The first result of this type seems to be due to Storer [2], who stated (without proof) that every connected cubic graph with n vertices has a spanning tree with at least $n/4+2$ leaves, and this is tight. Linial conjectured that, more generally, every connected graph with n vertices and with minimum degree k has a spanning tree with at least $((k-2)/(k+1))n+c_k$ leaves, where c_k is a constant depending only on k. Some progress has been made toward this conjecture; Kleitman and West [3] proved it for $k=3$ with $c_3=2$, and Griggs and Wu [4] proved it for $k=4,5$ with $c_4=8/5$ and $c_5=2$. (All these bounds have been shown to be tight.)

Our approach is different. We do not impose any condition on the individual degrees in the graph; instead, for all n and t we determine the smallest $F(n,t)$ such that every connected graph with n vertices and at least $F(n,t)$ edges must have a spanning tree with more than t leaves. It turns out that $F(n,t)$ separates; we will show that if $t<n-2$, $F(n,t)-n$ is a function of t alone.

Incidentally, it was shown in about 1981 by Neil Robertson and the third author (unpublished) that for any fixed integer $t\geq 2$, every connected graph with no $K_{1,t+1}$ minor has a bounded number of vertices of degree $\neq 2$. An easy consequence of this is that there is a function $f(t)$ so that every connected graph G with $|E(G)|>|V(G)|+f(t)$ has a spanning tree with $>t$ leaves. But now we wish to sharpen this, to make it best possible.

For any integer $n>t\geq 2$, define

$$f(n,t)=\begin{cases} n+\frac{1}{2}(t^2-4) & \text{if } n=t+2 \text{ and } t \text{ is even}\\ n+\frac{1}{2}(t^2-5) & \text{if } n=t+2 \text{ and } t \text{ is odd}\\ n+\frac{1}{2}(t^2-t-2) & \text{if } n=t+1 \text{ or } n\geq t+3 \end{cases}$$

We shall prove the following two results (the first is easy, and the second is the main result of the paper).

Theorem 1. *For all integers $n>t\geq 2$ there is a connected graph with n vertices and $f(n,t)$ edges in which every spanning tree has $\leq t$ leaves.*

Theorem 2. *For all integers $n > t \geq 2$, every connected graph with n vertices and $> f(n,t)$ edges has a spanning tree with $> t$ leaves.*

Some notation: we use $G \setminus X$ to denote the graph obtained by deleting X (here X can be a vertex or an edge, or a set of vertices or edges); and if $X \subseteq V(G)$, $G|X$ denotes the restriction of G to X, that is, $G \setminus (V(G) - X)$.

Proof of Theorem 1. If $n = t + 2$, let H be a graph with n vertices and $\lceil \frac{1}{2} n \rceil$ edges, in which every vertex has degree ≥ 1. Let G be its complement; then G is connected, $|E(G)| = f(n,t)$, and no spanning tree of G has $\geq t + 1$ leaves, since no vertex of G has degree $n - 1$.

If $n = t + 1$ or $n \geq t + 3$, let G be obtained from a complete graph K_{t+1} by replacing some edge e by a path of $n - t$ edges between the ends of e. Then again it is easy to check that G is connected, $|E(G)| = f(n,t)$ and no spanning tree of G has $\geq t + 1$ leaves. ■

We prove Theorem 2 in two steps. First, we show:

Lemma. *Theorem 2 is true for all n and t such that $2 \leq t < n \leq t + 3$.*

Proof: If $n = t + 1$, no graph with n vertices has $> f(n,t)$ edges, so the claim is vacuous. Now suppose that $n = t + 2$. If $|V(G)| = n$ and $|E(G)| > f(n,t)$, then G has a vertex of degree $n - 1 = t + 1$, and hence the claim holds.

Finally, let $n = t + 3$, and let G be a connected graph with n vertices and $> f(n,t)$ edges. We suppose that every spanning tree of G has $\leq t$ leaves. Let H be the complement of G. Then $|V(H)| = n$; $|E(H)| \leq 2n - 6$ (since $|E(G)| > f(n,t)$); no vertex of H has degree $n - 1$ (since G is connected); every vertex of H has degree ≥ 2 (since G has maximum degree $\leq t$); and every two vertices of H have distance ≤ 2 (for otherwise there are adjacent vertices u, v of G such that every other vertex of G is adjacent to ≥ 1 of them, and then G has a spanning tree with $\geqslant n - 2 = t + 1$ leaves, a contradiction).

(1) Every vertex of H has degree $\leq n - 4$.

Subproof. Suppose that some v has degree $n-3$ or $n-2$ ($n-1$ is impossible by hypothesis). Then $|E(H \setminus v)| \leq 2n - 6 - (n - 3) < |V(H \setminus v)| - 1$ and so $H \setminus v$ is not connected. Let $x \neq v$ be a vertex not adjacent to v in H, and let y be a vertex in another component of $H \setminus v$. Then every xy path has length ≥ 3, a contradiction. This proves (1).

Now since $|E(H)| \le 2n-6$, there is a vertex x of degree ≤ 3; let S be its set of neighbours, and let $s = |S|$. Then $s \le 3$.

Let $R = V(H) - (S \cup \{x\})$. Since every vertex has distance ≤ 2 from x it follows that every vertex in R has ≥ 1 neighbour in S. For $i = 1, 2$, let R_i be the set of all $v \in R$ with exactly i neighbours in S and let $R_3 = R - (R_1 \cup R_2)$. Let there be p edges with both ends in R, and q with both ends in S. Thus,

$$2n - 6 \ge |E(H)| \ge p + q + s + |R_1| + 2|R_2| + 3|R_3|$$

and since $1 + s + |R_1| + |R_2| + |R_3| = n$, it follows that

$$|R_1| - p \ge q + |R_3| - s + 4 \ .$$

Let the restriction $H|R$ of H to R have k components. Then

$$k \ge |R| - p \ge q + |R_2| + 2|R_3| + 4 - s > |R_2 \cup R_3| \ .$$

Hence there is a component C of $H|R$ with $V(C) \subseteq R_1$. Choose $v \in V(C)$, and let $u \in S$ be its unique neighbour in S. Now for every $w \in R - V(C)$, since the distance between v and w is ≤ 2 it follows that w is adjacent to u. Since u has degree $\le n-4$ by (1), and $s \le 3$, there is a vertex $v' \in R$ not adjacent to u, and consequently $v' \in V(C) \subseteq R_1$. Let u' be the unique neighbour of v' in S, then $u \neq u'$. By the same argument, every vertex in $R - V(C)$ is adjacent to u', and so belongs to $R_2 \cup R_3$. Hence $R_1 = V(C)$, and so C is unique. Consequently, $k \le |R_2| + |R_3| + 1$.

Since $k \ge q + |R_2| + 2|R_3| + 4 - s$ and $s \le 3$, it follows that $q = 0$, $s = 3$ and $R_3 = \emptyset$. Let $S = \{u, u', u''\}$. Since u'' has degree ≥ 2, $q = 0$, $R_3 = \emptyset$ and every vertex in R_2 is adjacent only to u and u', it follows that u'' has a neighbour $v'' \in V(C) \subseteq R_1$, and consequently $v'' \neq v, v'$. As before, every vertex in $R - V(C)$ is adjacent to u'', and so belongs to R_3 $(= \emptyset)$; and hence it follows that $R = V(C)$. Moreover, since $k = |R| - p$ it follows that C is a tree.

Let v_0 be a vertex of C with degree 1 in C, and let v_1 be its neighbour in C. Let u_0, u_1 be their respective (unique) neighbours in S, and choose $u_2 \in S - \{u_0, u_1\}$. Then the distance between v_0 and u_2 is ≥ 3, a contradiction, as required. ■

Proof of Theorem 2.

We suppose for a contradiction that the theorem is false. Choose a connected graph G with $|V(G)|+|E(G)|$ minimum and an integer $t \geq 2$ with $|V(G)| > t$, so that $|E(G)| > f(|V(G)|, t)$ and every spanning tree in G has $\leq t$ leaves. Let $n = |V(G)|$. From the Lemma and the minimality of G it follows that

(1) $n \geq t+4$, $f(n,t) = n + \frac{1}{2}t(t-1) - 1$, *and* $|E(G)| = n + \frac{1}{2}t(t-1)$.

We claim

(2) *If* $u, v \in V(G)$ *are adjacent, there is a vertex adjacent to them both.*

Subproof. By (1) and the definition of f, $f(n-1,t) = f(n,t) - 1$, and so if u, v have no common neighbour we can produce a smaller counterexample by contracting the edge uv. This proves (2).

(3) *For every vertex* x, $G \setminus x$ *is connected.*

Subproof. Suppose not; then there are connected subgraphs G_1, G_2 of G with $G_1 \cup G_2 = G$, $V(G_1 \cap G_2) = \{x\}$ and hence $E(G_1 \cap G_2) = \emptyset$, with $|V(G_1)|$, $|V(G_2)| < |V(G)|$. For $i = 1, 2$, let $n_i = |V(G_i)|$, and let T_i be a spanning tree of G_i chosen with the maximum number of leaves, t_i say; and furthermore choose T_i so that, if possible, x is not a leaf of it. Now $n_i \neq 2$ by (2), and so $n_i > t_i \geq 2$. From the minimality of G it follows that

$$|E(G_i)| \leq f(n_i, t_i)$$

and so

$$n + \frac{1}{2}t(t-1) = |E(G)| = |E(G_1)| + |E(G_2)| \leq f(n_1, t_1) + f(n_2, t_2) \ .$$

But $n = n_1 + n_2 - 1$, and $f(n_i, t_i) \leq n_i + \frac{1}{2}(t_i^2 - 4)$, with strict inequality unless either $t_i = 2$, or $n_i = t_i + 2$ and t_i is even. Consequently

$$\frac{1}{2}t(t-1) \leq \frac{1}{2}(t_1^2 + t_2^2) - 3 \ .$$

Now $T_1 \cup T_2$ is a spanning tree of G with $\geq t_1 + t_2 - 2 + \varepsilon$ leaves, where $\varepsilon = 0$ if x is a leaf of both T_1, T_2 and $\varepsilon = 1$ otherwise. Consequently $t_1 + t_2 - 2 + \varepsilon \leq t$, and so

$$\frac{1}{2}(t_1 + t_2 - 2 + \varepsilon)(t_1 + t_2 - 3 + \varepsilon) \leq \frac{1}{2}(t_1^2 + t_2^2) - 3 \ ,$$

that is

$$\left(t_1 - \frac{5}{2} + \varepsilon\right)\left(t_2 - \frac{5}{2} + \varepsilon\right) + \frac{5}{2}\varepsilon - \frac{1}{2}\varepsilon^2 \le \frac{1}{4}\ .$$

If $\varepsilon = 1$ then $\left(t_1 - \frac{3}{2}\right)\left(t_2 - \frac{3}{2}\right) + 2 \le \frac{1}{4}$, which is false since $t_1, t_2 \ge 2$; and so $\varepsilon = 0$ and

$$\left(t_1 - \frac{5}{2}\right)\left(t_2 - \frac{5}{2}\right) \le \frac{1}{4}\ .$$

If $t_1 = t_2 = 3$ we must have equality throughout; but if $t_1 = 3$ then $f(n_1, t_1) < n_1 + \frac{1}{2}(t_1^2 - 4)$, a contradiction. We may therefore assume that $t_1 = 2$. Hence G_1 is a path or circuit, and by (2) G_1 is a 3-vertex circuit. But then T_1 can be chosen so that x is not a leaf of it, contradicting that $\varepsilon = 0$. This proves (3).

Let x be a vertex of G with maximum degree s say; and let S be the set of all its neighbours. In addition, choose x so that either some vertex in S has degree $< s$, or every vertex of G has degree s. (This is possible since G is connected.)

(4) $t \ge s \ge 3$.

Subproof. By (1) $|E(G)| > |V(G)|$, and so $s \ge 3$. Since every spanning tree of G has $\le t$ leaves and hence every tree subgraph of G also has $\le t$ leaves, it follows that $s \le t$. This proves (4).

Let $R = V(G) - (S \cup \{x\})$. A graph is *non-null* if it has at least one vertex.

(5) *$G|R$ is non-null and connected.*

Subproof. By (3) $G \setminus x$ is connected, and so by (1) and (4),

$$|E(G \setminus x)| = n + \frac{1}{2}t(t-1) - s \ge n + \frac{1}{2}((t-1)^2 - (t-1) - 2) > f(n-1, t-1)\ .$$

From the minimality of G, $G \setminus x$ has a spanning tree T with $\ge t$ leaves. Every vertex v in S is a leaf for T, for otherwise we could add the edge xv to T to obtain a tree with $\ge t+1$ leaves, a contradiction. Since T has $\ge t \ge 3$ leaves, it follows that $T \setminus S$ is non-null and connected, and consequently so is $G|R$. This proves (5).

Let there be e_1 edges with both ends in S, e_2 edges with both ends in R, and e_0 edges with one end in S and the other in R.

(6) $e_0 + 2e_2 \ge 2n + t(t-1) - s(s+1)$, *with equality only if every vertex of G has degree s.*

Subproof. Since every vertex has degree $\leq s$, by summing the degrees in S we obtain

$$s + e_0 + 2e_1 \leq s^2$$

with equality only if every vertex in S has degree s (and hence every vertex in G has degree s, from the choice of x). But

$$e_0 + e_1 + e_2 + s = |E(G)| = n + \frac{1}{2}t(t-1)$$

and the result follows on eliminating e_1. This proves (6).

For each vertex $y \in S$, let $G_y = G|(R \cup \{y\})$, and let $d(y)$ be the degree of y in G_y, that is, the number of neighbours of y in R. Let $d = \max\{d_y \colon\ y \in S\}$, and let $t_2 = t + 2 - s \geq 2$.

(7) *$sd \geq e_0$, $s \geq d + 2$, $t_2 \geq d + 1$, and $d \geq 1$.*

Subproof. $e_0 = \sum_{y \in S} d(y) \leq ds$, so the first claim follows. Choose $y \in S$ with $d(y) = d$. By (2) applied to x, y, it follows that y has a neighbour in S, and so its degree is at least $d + 2$; and hence $d + 2 \leq s$, proving the second claim. For the third, the set of edges incident with x in G, together with those incident with y in G_y, form a subtree of G with $s - 1 + d$ leaves, and so

$$s - 1 + d \leq t = s + t_2 - 2$$

and the third claim follows. Finally $d \geq 1$ since $R \neq \emptyset$ and G is connected.

(8) *If $y \in S$ satisfies $d(y) = d$, then G_y is connected, and every spanning tree of G_y has $\leq t_2$ leaves. In particular any neighbour in R of y has $\leq t_2 - 1$ neighbours in R, and every other vertex of R has $\leq t_2$ neighbours in R.*

Subproof. By (5), G_y is connected, with ≥ 4 vertices since $n \geq s + 4$ by (1) and (4). Let T be a spanning tree of G_y; then by adding the edges incident with x to it we obtain a spanning tree of G with $\geq s - 2$ more leaves. Consequently T has $\leq t - s + 2 = t_2$ leaves, and so every subtree of G_y has $\leq t_2$ leaves, and (8) follows.

$$\text{Define } \varepsilon = \begin{cases} \frac{1}{2}t_2 - 1 & \text{if } n - s = t_2 + 2 \text{ and } t_2 \text{ is even} \\ \frac{1}{2}t_2 - \frac{3}{2} & \text{if } n - s = t_2 + 2 \text{ and } t_2 \text{ is odd} \\ 0 & \text{otherwise .} \end{cases}$$

(9) $(s-2)\left(t_2-2-\frac{1}{2}d\right)\le\varepsilon$.

Subproof. Choose $y\in S$ with $d(y)=d$. Now $|V(G_y)|=n-s\ge t_2+2$, since $n\ge t+4$ by (1). From (8) and the minimality of G,

$$d+e_2=|E(G_y)|\le f(n-s,t_2)\ .$$

From (7), $e_0\le sd$, and so, substituting for e_0 and e_2 in (6), we obtain

$$sd+2(f(n-s,t_2)-d)\ge 2n+t(t-1)-s(s+1)$$

and since $t=s+t_2-2$, it follows that

$$f(n-s,t_2)\ge n+\frac{1}{2}t_2^2+st_2-3s-\frac{5}{2}t_2+3-\frac{1}{2}(s-2)d\ .$$

But from its definition,

$$f(n-s,t_2)=n-s+\frac{1}{2}t_2(t_2-1)-1+\varepsilon\ ,$$

and on substituting the claim follows.

(10) $t_2\le 3$ *and* $\varepsilon=0$.

Subproof. Suppose $t_2\ge 4$. Since $\varepsilon\le\frac{1}{2}t_2-1$, we deduce from (9) that

$$(s-2)\left(t_2-2-\frac{1}{2}d\right)\le\frac{1}{2}t_2-1$$

which can be rewritten as

$$\left(\frac{1}{2}d+\frac{1}{2}(d-1)+(s-d-2)\right)\left(\frac{1}{2}+\frac{1}{2}(t_2-4)+\frac{1}{2}(t_2-d-1)\right)\le\frac{1}{4}d\ .$$

But this is impossible, since $d-1\ge 0$, $s-d-2\ge 0$, $t_2-4\ge 0$, $t_2-d-1\ge 0$, and equality cannot hold in all four inequalities simultaneously. Thus $t_2\le 3$, and so $\varepsilon=0$ by the definition of ε. This proves (10).

(11) $t_2=3$ *and* $d=2$, *and* $s=t-1$.

Subproof. From (9) and (10) we deduce that $(s-2)\left(t_2-2-\frac{1}{2}d\right)\leq 0$. Since $s\geq 3$ by (4), it folows that $t_2\leq 2+\frac{1}{2}d$. Since $t_2\geq d+1$ and $d\geq 1$ by (7), we deduce that either $t_2=2$ and $d=1$, or $t_2=3$ and $d=2$.

If $t_2=2$ and $d=1$, then $s=t$. Since $n\geq t+3$ and $G|R$ is connected, there are two adjacent vertices u, v in R. By (2) they have a common neighbour w. Now $w\notin S$ since $d=1$. Choose a minimal path of G from S to $\{u,v,w\}$; and by adding this and two of uv, uw, vw to the edges incident with x, we obtain a tree in G with $\geq s+1>t$ leaves, a contradiction. This proves (11).

By (11), it follows that we have equality in (9), and hence $e_0=sd$, and we have equality in (6). Consequently every vertex of G has degree s. Since $e_0=sd$, it follows that $d(y)=d$ for every vertex $y\in S$, and so by (8), every $v\in R$ has $\leq t_2$ neighbours in R, with strict inequality if it has a neighbour not in R. Since every vertex has degree $s>t_2$ (because $s\geq d+2=4$ and $t_2=3$), it follows that every vertex in R has $\geq s-t_2+1=s-2$ neighbours in S. Consequently

$$2s=ds=e_0\geq(n-s-1)(s-2)\ .$$

On the other hand,

$$\frac{1}{2}sn=|E(G)|=n+\frac{1}{2}t(t-1)=n+\frac{1}{2}(s+1)s$$

and so

$$2s+sn\geq(n-s-1)(s-2)+2n+(s+1)s$$

which is impossible. This completes the proof. ■

It is easy to see that the proof just given can be converted into a polynomial time algorithm to find the tree, given a graph satisfying the hypotheses of the theorem. The algorithm would begin by checking that the graph is 2-connected (and if not, winning by looking at the blocks separately); checking that every edge is in a triangle (if not, winning by contracting the edge); then choosing a vertex of maximum degree, x say; checking that the graph stays connected when x and its neighbours are all deleted (and if not, winning by deleting x); then choosing a neighbour y of x as in the proof, and winning by deleting x and all neighbours except y (the proof shows that this graph must have enough edges for the algorithm to be applicable to it, except in one degenerate case which is easily treated separately). We omit the details.

REFERENCES

1. P. Lemke, “The maximum leaf spanning tree problem for cubic graphs is NP-complete”, Institute for Mathematics and its Application, preprint, 1988.
2. J.A. Storer, “Constructing full spanning trees for cubic graphs”, *Inform. Process. Lett.* 13 (1981) 8-11.
3. D.J. Kleitman and D.B. West, “Spanning trees with many leaves”, lecture at SIAM Discrete Math. Conf. San Francisco, 1988.
4. J.R. Griggs and M. Wu, “Spanning trees in graphs of minimum degree 4 or 5”, *Discrete Math.* 104 (1992) 167-183.