JOURNAL OF COMBINATORIAL THEORY, Series B 61, 63–70 (1994)

# Subtrees and Subforests of Graphs*

STEPHAN BRANDT

*Fachbereich Mathematik, Freie Universität Berlin, Graduiertenkolleg "Algorithmische Diskrete Mathematik," Arnimallee 2–6, D-14195 Berlin, Germany*

Received August 28, 1992

In this paper we present sufficient edge-number and degree conditions for a graph to contain all forests of given size. The edge-number bound answers in the affirmative a conjecture due to Erdős and Sós. Furthermore, we will give improved bounds for specified spanning subtrees of graphs. 

## 1. Introduction and Notation

A lot of graph-theoretical research has been performed on matchings in graphs and considerably less work has been spent on subtrees of graphs. The results are very deep in the former case and rather weak in the latter case. Almost no attention has been paid to the connecting concept of both: subforests of graphs. Set

$$f(k, n) = \max\left\{\binom{2k-1}{2}, \binom{k-1}{2} + (k-1)(n-k+1)\right\}. \qquad (1)$$

In 1963, Erdős and Sós (see [2]) stated the following conjecture:

*Conjecture* 1. Let $G$ be a graph with $n$ vertices and more than $f(k, n)$ edges. Then $G$ contains every forest with $k$ edges and without isolated vertices as a subgraph.

The main object of this paper is to prove this conjecture and the related degree bound. Recall the following well-known result that was attributed by Chvátal [1] to graph-theoretical folklore.

LEMMA 1. *Suppose $G$ is a graph with minimum degree at least $k$. Then $G$ contains every tree with $k$ edges.*

We will show that such graphs $G$ even contain all forests $F$ with $k$ edges and at most as many vertices as $G$. Furthermore, we will show that both

* This research was supported by Deutsche Forschungsgemeinschaft, Grant We 1265/2-3.



582b/61/1-5

0095-8956/94 $6.00

bounds cannot be improved. They were known to hold for trees as well as for matchings. The proofs are obtained by combining methods from both areas.

In the final section we will improve and generalize two results from [3, 6], one of them based on the degree bound for forests. The results give sufficient conditions for graphs to contain specified spanning subtrees.

Before stating the results let us first fix the notation needed. All graphs considered are simple, finite, and loopless. For a graph $G=(V(G), E(G))$ we denote its order by $|G|$, its size by $e(G)$ and its complement by $\bar{G}$. The neighborhood of a vertex $v \in V(G)$ is $N(v)$ and its cardinality $d(v)$. The maximum degree (minimum degree) of $G$ is denoted by $\Delta(G)$ (resp., $\delta(G)$). Let $S$ be a subset of $V(G)$. The number of neighbors of $v$ in $S$ will be $d(v, S)$ and the set of vertices adjacent to at least one vertex is $S$ is $N(S)$. The subgraph induced by $S$ is indicated by $\langle S \rangle$ and $\langle V(G) \backslash S \rangle$ is denoted by $G-S$. As usual we will write $G-v$ instead of $G-\{v\}$.

We will call an injection $\sigma : V(H) \rightarrow V(G)$ an embedding of $H$ in $G$, if $vw \in E(H)$ implies $\sigma(v)\,\sigma(w) \in E(G)$. The brief notation for the existence of such an embedding is $H \subseteq G$, in words: $G$ contains $H$ (as a subgraph). For two graphs $G$ and $H$ the graph $G+H$ is obtained by taking vertex disjoint copies of $G$ and $H$ and joining each vertex of $G$ to each vertex of $H$.

A forest is an acyclic graph, a tree is a connected forest, and a matching is a forest with all components of order two. Vertices of degree $\leqslant 1$ in forests will be called leaves.

## 2. Subforests of Graphs

Let us begin with the extension of Lemma 1 for forests. At the same time this is a generalization of a result due to Erdős and Pósa [5], who proved this bound for matchings.

THEOREM 1. *Suppose $F$ is a forest with $e(F)=k$, $G$ is a graph with $\delta(G) \geqslant k$ and $|F| \leqslant |G|$. Then $F \subseteq G$.*

This bound cannot be improved. Clearly $(k-1)$-regular graphs do not contain the star $K_{1,k}$ on $k$ edges and the graph $K_{k-1} + \bar{K}_{n-k+1}$ does not contain a matching on $k$ edges. So Theorem 1 is best possible in both the tree and the matching case.

Theorem 1 will be an important tool in the proof of the edge-number bound for forests that we will state now. Recall the function $f(k, n)$ from (1).

THEOREM 2. *Suppose $G$ is a graph of size $e(G) > f(k, n)$. Then $G$ contains every forest on $k$ edges without isolated vertices.*

This generalizes a result due to Erdős and Gallai [4] who proved this bound for matchings with $k$ edges. The graphs

$$K_{2k-1} \quad \text{and} \quad K_{k-1} + \bar{K}_{n-k+1}$$

contain no matching with $k$ edges. Hence the bound given in Theorem 2 is best possible. In case of forests that are more tree-like than matching-like (i.e., many edges, few components) the bound of Theorem 2 seems to become quite weak. So Erdős and Sós (see [2]) posed the following conjecture for trees.

*Conjecture* 2. Every graph $G$ of order $n$ and size $e(G) > \frac{1}{2}n(k-1)$ contains every tree with $k$ edges.

If this conjecture holds the bound is best possible since no $(k-1)$-regular graph contains the star $K_{1,k}$ on $k$ edges, and, in particular, the complete graph $K_k$ contains no tree with $k$ edges. Ajtai, Komlós, and Szemerédi [8] have proved an approximate version of Conjecture 2.

Before we give the proof of Theorem 1, let us first state an easy extension of Lemma 1 in the way that we will need it. In the sequel we will often deal with a special type of induced subforests of a forest. Suppose there is a sequence of forests $F = F^{(0)} \supseteq F^{(1)} \supseteq \cdots \supseteq F^{(r)} = F'$ and a sequence of vertices $(x_0, x_1, ..., x_{r-1})$ of $F$, such that $x_i$ is a leaf of $F^{(i)}$ and $F^{(i+1)} = F^{(i)} - x_i$ $(0 \leqslant i < r)$. We will call $F'$ a **reduced subforest** of $F$. So, informally, $F'$ is obtained from $F$ by successively deleting leaves.

LEMMA 2. *Suppose $G$ is a graph with $\delta(G) \geqslant k$. Let $F$ be a forest of order at most $k+1$ and let $F'$ be a reduced subforest of $F$. Then any embedding $\sigma$ of $F'$ in $G$ can be extended to an embedding of $F$ in $G$.*

Clearly Lemma 2 implies Lemma 1 since every subtree of a tree $T$ is a reduced subforest of $T$. In particular every single vertex is a reduced subforest.

*Proof.* Suppose there is an embedding $\sigma$ of $F'$ in $G$ that does not extend to an embedding of $F$ in $G$. Then there is an index $i$ $(0 \leqslant i < r)$, such that $\sigma$ can be extended to an embedding of $F^{(i+1)} = F^{(i)} - x_i$ in $G$ but not to an embedding of $F^{(i)}$ in $G$.

Note that $|G| \geqslant k+1$, so $x_i$ is not an isolated vertex in $F^{(i)}$. Fix an extension of $\sigma$ for $F^{(i+1)}$. Let $y_i$ be the neighbor of $x_i$ in $F^{(i)}$. Since $|N(\sigma(y_i)) \cap \sigma(F^{(i+1)})| \leqslant k-1$ but $d(\sigma(y_i)) \geqslant \delta(G) \geqslant k$ there is a neighbor $v \notin \sigma(F^{(i+1)})$ of $\sigma(y_i)$, and $\sigma(x_i) = v$ extends $\sigma$ to an embedding of $F^{(i)}$ in $G$, a contradiction. ∎

*Proof of Theorem* 1. Construct a subforest $F' \subseteq F$ by deleting one leaf from every component of $F$ and call the neighbor of this leaf (if it exists) the root. Let $R$ be the set of roots. Clearly $R$ is a reduced subforest of $F'$. Map the roots by an injection $\sigma$ on vertices of $G$, such that $e(\langle \sigma(R) \rangle)$ is minimal. Since $|F'| = k$, by Lemma 2 we can extend $\sigma$ to an embedding of $F'$ in $G$.

Set $S = V(G) \backslash \sigma(F')$ and suppose that $\sigma$ cannot be extended to an embedding of $F$ in $G$. Then by Hall's theorem [7] there is a subset $R' \subseteq \sigma(R)$ such that $|R'| > |S'|$, where $S' = N(R') \cap S$. For any vertex $v \in R'$ this yields $d(v, S) \leqslant |S'| < |R'|$, so $d(v, \sigma(R)) \geqslant |R| - |R'| + 1$, since $|F'| = k$. On the other hand, we have $d(w, \sigma(R)) \leqslant |R| - |R'|$ for every $w \in S \backslash S'$ and since $vw \notin E(G)$ this yields $e(\langle \sigma(R) \backslash \{v\} \cup \{w\} \rangle) \leqslant e(\langle \sigma(R) \rangle) - 1$, contradicting the minimality of $e(\langle \sigma(R) \rangle)$. ∎

Let us now turn to the related edge-number bound in Theorem 2. It can be shown by a simple proof that graphs with more than $f(k, n)$ edges contain an induced subgraph with minimum degree $k$ (for a proof, see, e.g., [9, p. 159]). The idea is that it is impossible to delete a graph with more than $f(k, n)$ edges by successively deleting vertices of degree less than $k$.

LEMMA 3. *Suppose $G$ is a graph on $n$ vertices with*

$$e(G) > \binom{k-1}{2} + (k-1)(n-k+1).$$

*Then $G$ has an induced subgraph $H$ with $\delta(H) \geqslant k$.*

Note that this need not be sufficient to contain all forests $F$ with $k$ edges, since the condition $|F| \leqslant |H|$ from Theorem 1 might be violated.

A big step towards the proof of Theorem 2 will be done in the following proposition, where we demonstrate that the edge-number bound above is sufficient for forest of size $k$ without isolated vertices or edges.

PROPOSITION 1. *If $e(G) > \binom{k-1}{2} + (k-1)(n-k+1)$, then $G$ contains every forest with $e(F) = k \leqslant \frac{1}{2}n$ having no components of order $\leqslant 2$.*

*Proof.* The Proposition holds for $k \leqslant 3$ by Lemma 1 and Lemma 3, since $F$ is a tree. Assume that $k \geqslant 4$ is the least value for which this result does not hold. Let $F$ be a forest with $c$ components of size $k$ and let $G$ be a least-order graph that does not contain $F$. We may assume that $n = 2k$, since otherwise either $\delta(G) \geqslant k$ or $e(G - v) > \binom{k-1}{2} + (k-1)(n-1-k+1)$ for a vertex $v$ with $d(v) \leqslant k-1$, both implying $F \subseteq G$.

Now consider an induced subgraph $H \subseteq G$ with $\delta(H) \geqslant k$ that is guaranteed by Lemma 3. Since $|F| = k + c \leqslant \frac{3}{4}n$ we may assume $|H| < \frac{3}{4}n$ (by Theorem 1). Suppose that $H$ has a subset $S \subseteq V(H)$ of $c$ vertices with $d_G(v) \geqslant \frac{3}{4}n - 1$ for all $v \in S$; then remove a leaf from every component of $F$ to obtain a subforest $F'$ of order $k$. Map the set $R$ of neighbors of the removed vertices by an injection $\sigma$ on $S$. Since $R$ is a reduced subforest of $F'$, $\sigma$ can be extended to an embedding of $F'$ in $H$. By Hall's theorem [7] there is a matching between $S$ and $V(G) \backslash \sigma(F')$ saturating $S$. So we obtain an embedding of $F$ in $G$, contradicting our assumption.

Thus there must be two adjacent vertices $v$ and $w$ in $H$ with $d_G(v) \leqslant d_G(w) < \frac{3}{4}n - 1$. Map an edge of one component $C$ of $F$ with $k'$ edges by $\sigma$ on $vw$ and extend $\sigma$ to an embedding of $C$ in $H$. Then

$$
\begin{aligned}
e(G - \sigma(C)) &> e(G) - \left(\frac{3}{2}n - 3\right) - \sum_{i=3}^{k'+1} (n - i) \\
&> \binom{k-1}{2} + (k-1)(n-k+1) - (k-1) - \sum_{i=2}^{k'+1} (n-i) \\
&= \binom{k-1}{2} + (k-1)(n-k) - k'(n-k) - \sum_{i=2}^{k'+1} (k-i) \\
&= \binom{k-k'-1}{2} + (k-k'-1)(n-k) \\
&= \binom{k-k'-1}{2} + (k-k'-1)(n-k'-1-(k-k')+1),
\end{aligned}
$$

and by induction using $k - k' \leqslant \frac{1}{2}(n - k' - 1)$, we obtain $F - C \subseteq G - \sigma(C)$, a contradiction. ∎

We are now prepared to prove the main theorem, Theorem 2, of this paper. Note that

$$\binom{2k-1}{2} \geqslant \binom{k-1}{2} + (k-1)(n-k+1),$$

only if $n \leqslant \frac{5}{2}k - 1$.

*Proof of Theorem* 2. First we may assume that the graph $G$ has no isolated vertex $v$, since then the bound holds for $G - v$. Furthermore, if $\delta(G) \geqslant k$ the result is true by Theorem 1 (note that $|G| \geqslant 2k$). Thus we may assume that $\delta(G) \leqslant k - 1$. Moreover, by Proposition 1 we may restrict our attention to forests with isolated edges.

In the rest of the proof we will again proceed by induction. Note that the theorem holds for $k=1$, since $G$ is non-empty. For $k \geqslant 2$ remove a vertex $v$ of minimum degree and one of its neighbors $w$ from $G$. Since

$$\begin{aligned} e(G-v-w) &> \binom{k-1}{2} + (k-1)(n-k+1) - (n+k-3) \\ &= \binom{k-2}{2} + (k-2)(n-2-(k-1)+1) \end{aligned}$$

and for $n-2 \leqslant \frac{5}{2}(k-1)-1$,

$$\begin{aligned} e(G-v-w) &> \binom{2k-1}{2} - (n+k-3) \\ &\geqslant \binom{2k-1}{2} - \left(\frac{7}{2}k - \frac{9}{2}\right) \geqslant \binom{2(k-1)-1}{2}, \end{aligned}$$

the graph $G-v-w$ contains $F-x-y$ for an isolated edge $xy \in E(F)$. Thus $F \subseteq G$. ∎

## 3. Sufficient Conditions for Graphs with Specified Spanning Subtrees

Let us now turn to another application of Theorem 1. A graph $G$ of order $n$ is called **panarboreal**, if it contains all trees of order $n$. Clearly panarboreal graphs must have a vertex joined to all other vertices in order to contain the star $K_{1,n-1}$. Now it can be shown that graphs with this property are panarboreal, provided that their minimum degree is sufficiently large.

Theorem 3. *Let $r \geqslant 1$ be an integer. Suppose $G$ is a graph of order $n \geqslant r^2 - r + 1$ with $\Delta(G) = n-1$ and $\delta(G) \geqslant n-r$. Then $G$ is panarboreal.*

This improves a result due to Faudree *et al.* [6] who proved this statement for $n \geqslant 3r^2 - 9r + 8$. It can be derived from Theorem 3 that every $G$ with $\Delta(G) = n-1$ and $\delta(G) \geqslant n - \sqrt{n}$ is panarboreal but our result yields no improvement concerning the following question posed in [3]:

> Is there a constant $c$ ($\frac{1}{2} < c < 1$) such that every graph $G$ with $\Delta(G) = n-1$ and $\delta(G) \geqslant cn$ is panarboreal?

This might even be true for every fixed $c > \frac{1}{2}$ if $n$ is sufficiently large. For the proof of Theorem 3 we need another result besides Theorem 1, which is a slight improvement of a result in [3, Theorem 2]. It was first used implicitly in [6]. It gives a minimum degree condition for graphs, assuring that they contain every spanning subtree of small maximum degree. In fact, we will prove a stronger statement, reflecting the structure of the tree in more detail, and derive the indicated result as a corollary. Instead of the maximum degree of the tree we will consider the initial part of its degree-sequence.

THEOREM 4. *Suppose* $D_T = (d(x_1), d(x_2), ..., d(x_n))$, $d(x_i) \geqslant d(x_{i+1})$ *for* $1 \leqslant i < n$, *is the degree-sequence of a tree* $T$ *of order* $n$ *and* $d_k = \sum_{i=1}^{k} d(x_i) \leqslant n-k-1$. *Then* $T \subseteq G$ *for every graph* $G$ *of order* $n$ *with* $\delta(G) \geqslant n-k-1$.

*Proof.* Assume this is false. Then there is a tree $T' \subseteq T$ with $T' \nsubseteq G$ but $T'-x \subseteq G$ for every leaf $x$. Fix an embedding $\sigma$ of $T'-x$ in $G$ and let $y$ be the neighbor of $x$ in $T'$. Then $\sigma(y)$ is not joined to any vertex $v \in V(G) \backslash \sigma(T'-x)$. Now consider such a vertex $v$. Let $Z$ be the set of vertices $z \in V(T'-x)$ for which the injection $\sigma_z$ mapping $z$ on $v$ and leaving $\sigma$ unchanged on the other vertices of $T'-x$ is an embedding. Clearly $\sigma_z$ is an embedding iff $v$ is adjacent to the image of every neighbor of $z$. The vertex $v$ is not adjacent to at most $k$ vertices in $\sigma(T'-x)$. Since $D_{T'-x}$ is lexicographically smaller than $D_T$ and since $v$ is not joined to $\sigma(y)$ we have $|Z| \geqslant |T'| - 1 - (d_k - 1)$. On the other hand, $\sigma(y)$ has at least $n-k-1$ neighbors in $\sigma(T'-x)$. Since $|T'| - 1 < |Z| + d(\sigma(y))$ there is a vertex in $\sigma(Z)$ joined to $\sigma(y)$, and $\sigma_z(x) = \sigma(z)$ is an embedding of $T'$ in $G$, contradicting the assumption. ∎

For the proof of Theorem 3 we only need the following immediate consequence.

COROLLARY 1. *Let* $G$ *be a graph and* $T$ *a tree both of order* $n$ *and suppose that* $\Delta(T) = \Delta$. *If* $\delta(G) \geqslant (\Delta/(\Delta+1))(n-1)$ *then* $T \subseteq G$.

*Proof.* Set $k = \lfloor (n-1)/(\Delta+1) \rfloor$. Then $\delta(G) \geqslant n-k-1$ and $d_k \leqslant \Delta k \leqslant \lfloor (\Delta/(\Delta+1))(n-1) \rfloor \leqslant n-k-1$. Thus by Theorem 4 we obtain $T \subseteq G$. ∎

What is the improvement of Theorem 4 compared with Corollary 1? We can easily derive results like the following, relating the minimum degree of a graph to the number of leaves of its spanning subtrees. The sum over the initial part of the degree-sequence of a tree $T$ bounds the number of leaves $\ell(T)$ of the tree from below. Recall that $\ell(T)$ can be computed by

$$\ell(T) = 2 + \sum_{v \in I} (d(v) - 2),$$

where the sum ranges over the set $I$ of the inner vertices of $T$.

COROLLARY 2. *Suppose $G$ is a graph of order $n$ and $T$ is a tree of order $n$ with*

$$\ell(T) \leqslant 3\delta(G) - 2n + 4.$$

*Then $T \subseteq G$.*

*Proof.* Suppose $T \not\subseteq G$. Then for $k = n - \delta(G) - 1$ we have $d_k > n - k - 1$. Thus,

$$\ell(T) \geqslant 2 + d_k - 2k > n - 3k + 1 = 3\delta(G) - 2n + 4. \quad \blacksquare$$

So, for example, graphs $G$ with $\delta(G) \geqslant (5n-8)/6$ contain every spanning tree with at most $\frac{1}{2}n$ leaves. The proof of Theorem 3 is an immediate consequence of Corollary 1 and Theorem 1.

*Proof of Theorem* 3. Let $T$ be a tree of order $n$. Suppose $\Delta(T) = \Delta \leqslant r-1$. Then $\delta(G) \geqslant n - r \geqslant (n-1)(r-1)/r \geqslant (\Delta/(\Delta+1))(n-1)$, since $n \geqslant r^2 - r + 1$. So Corollary 1 yields $T \subseteq G$.

Now suppose that $\Delta \geqslant r$. Let $x$ be a vertex of $T$ with $d(x) = \Delta(T)$ and let $v$ be a vertex of degree $n-1$ in $G$. Then $T - x$ is a forest satisfying $e(T-x) = n - 1 - \Delta \leqslant n - r - 1$ and $\delta(G - v) \geqslant n - r - 1$. From Theorem 1 we derive $T - x \subseteq G - v$ and thus $T \subseteq G$. $\blacksquare$

## REFERENCES


1. V. CHVÁTAL, Tree—Complete graph Ramsey numbers, *J. Graph Theory* **1** (1977), 93.
2. P. ERDŐS, Extremal problems in graph theory, *in* "Theory of Graphs and its Applications" (M. Fiedler, Ed.), pp. 29–36, Academic Press, New York, 1964.
3. P. ERDŐS, R. J. FAUDREE, C. C. ROUSSEAU, AND R. H. SCHELP, Graphs with certain families of spanning trees, *J. Combin. Theory Ser. B* **32** (1982), 162–170.
4. P. ERDŐS AND T. GALLAI, On maximal paths and circuits of a graph, *Acta Math. Acad. Sci. Hungar.* **10** (1959), 337–356.
5. P. ERDŐS AND L. PÓSA, On the maximal number of disjoint circuits of a graph, *Publ. Math. Debrecen* **9** (1962), 3–12.
6. R. J. FAUDREE, C. C. ROUSSEAU, R. H. SCHELP, AND S. SCHUSTER, Panarboreal graphs, *Israel J. Math.* **35** (1980), 177–185.
7. P. HALL, On representatives of subsets, *J. London Math. Soc.* **10** (1935), 26–30.
8. E. SZEMERÉDI, personal communication, 1992.
9. H. P. YAP, "Some Topics in Graph Theory," Cambridge Univ. Press, Cambridge, UK, 1986.