ON SUBGRAPH NUMBER INDEPENDENCE IN TREES

by

R. L. Graham
E. Szemerédi

STAN-CS-75-484
MARCH 1975

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

ON SUBGRAPH NUMBER INDEPENDENCE IN TREES

R. L. Graham
Stanford University
and
Bell Laboratories, Murray Hill, New Jersey

and

E. Szemerédi
Stanford University
and
The Hungarian Academy of Sciences

Abstract

For finite graphs F and G, let $N_F(G)$ denote the number of occurrences of F in G, i.e., the number of subgraphs of G which are isomorphic to F. If $\mathcal{F}$ and $\mathcal{G}$ are families of graphs, it is natural to ask them whether or not the quantities $N_F(G)$, $F \in \mathcal{F}$, are linearly independent when G is restricted to $\mathcal{G}$. For example, if $\mathcal{F} = \{K_1, K_2\}$ (where K_n denotes the complete graph on n vertices) and $\mathcal{G}$ is the family of all (finite) trees then of course $N_{K_1}(T) - N_{K_2}(T) = 1$ for all $T \in \mathcal{G}$. Slightly less trivially, if $\mathcal{F} = \{S_n : n = 1,2,3,\ldots\}$ (where S_n denotes the star on n edges) and $\mathcal{G}$ again is the family of all trees then

$$\sum_{n-1}^{\infty} (-1)^{n+1} N_{S_n}(T) = 1 \quad \text{for all } T \in \mathcal{G} .$$

It will be proved that such a linear dependence can never occur if $\mathcal{F}$ is finite, no $F \in \mathcal{F}$ has an isolated point and $\mathcal{G}$ contains all trees. This result has important applications in recent work of L. Lovász and one of the authors [2].

This research was supported in part by National Science Foundation grant GJ 36473X and by the Office of Naval Research contract NR 044-402.

ON SUBGRAPH NUMBER INDEPENDENCE IN TREES

by

R. L. Graham
Stanford University
Stanford, California

and

Bell Laboratories
Murray Hill, New Jersey

and

E. Szemerédi
Stanford University
Stanford, California

and

The Hungarian Academy of Sciences
Budapest, Hungary

INTRODUCTION

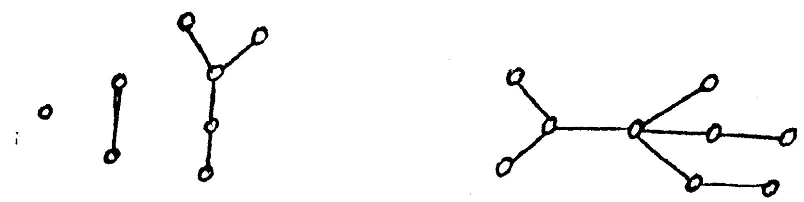
It is a trivial observation (in fact, almost a definition) that in any finite tree T, the number of vertices of T always exceeds the number of edges of T by exactly 3. In [1], it was asked to what extent this can happen for graphs in general. That is, given a finite family $\mathfrak{F}$ of graphs G, when can there be a fixed linear dependence between the number of occurrences of the $G \in \mathfrak{F}$ as subgraphs of a tree T which is valid for all finite* trees T. In this paper, we answer this question. In particular, this can never happen if none of the $G \in \mathfrak{F}$ have isolated points;

---

* All graphs considered in this paper will be finite. For terminology see [3].

SOME NOTATION

For a graph G, we let V(G) and E(G) denote the sets of vertices and edges of G, respectively. If H is a labelled graph (i.e., with distinguishable vertices) and G is an unlabelled graph, we define $N_G(H)$ to be the number of occurrences of G in H, i.e., the number of ways a subset of $|E(G)|$ edges can be selected from E(H) together with i vertices from V(H) if G has i isolated vertices, so that the resulting subgraph of H is isomorphic to G. Of course, the product of $N_G(H)$ and the order of the automorphism group of G is just $E_G(H)$, the number of ways of embedding G into H (considering G as labelled graph). For example, if G and H are as shown in Fig. 1 then $N_G(H) = 28$ and $E_G(H) = 112$.



G H

Fig. 1

Note that if the isolated point is removed from G to form G′ then $N_{G'}(H) = 14 = \frac{1}{2} N_G(H)$. Of course, in general, if G is formed from a graph G′ having no isolated points by adjoining i isolated points then

$$(1) \qquad N_G(H) = \binom{|V(H)| - |V(G')|}{i} N_{G'}(H)$$

THE MAIN RESULT

The primary result of this paper can be stated as follows.

Theorem. Let $\mathfrak{F}$ be a finite family of forests,* each having no isolated points, and suppose there exist real numbers A_F, $F \in \mathfrak{F}$, and A_0 such that the equation

$$\sum_{F \in \mathfrak{F}} A_F N_F(T) = A_0 \tag{2}$$

is valid for all trees T. Then $A_F = 0$ for all $F \in \mathfrak{F}$.

Remark. Since any subgraph of a tree is a forest then there is no loss of generality in assuming $\mathfrak{F}$ is a family of forests.

Proof: We may assume without loss of generality that among all families for which an equation of the form (2) is possible, $\mathfrak{F}$ has the least number of elements. The basic idea of the proof will be to construct a very large tree W^* for which one of the quantities $N_F(W^*)$ is much larger than all the others,' thereby forcing its coefficient A_F to be 0. However, this contradicts the minimality of $|\mathfrak{F}|$.

If T is a tree with a distinguished vertex v, we let $T^{(k)}$ denote the tree formed from T by adjoining k disjoint paths of length k to v. (See Fig. 2).

---

* i.e., acyclic graphs,

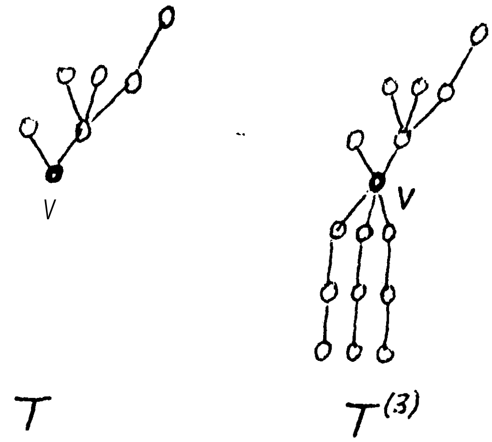


Fig. 2

Similarly, if F is a forest with components $T_1, \ldots, T_n$ having distinguished vertices $v_1, \ldots, v_n$, respectively, then $F^{(k)}$ denotes the forest with components $T_1^{(k)}, \ldots, T_n^{(k)}$.

We now define a (possibly empty) forest $W = W(\mathfrak{F})$ with components W_i and distinguished vertices $w_i \in V(W_i)$, $1 \leq i \leq t$, as follows:

(i) Some $F \in \mathfrak{F}$ occurs as a subgraph of $W(k)$ for some k.

(ii) $|E(W)|$ is minimal among all W satisfying (i).

Note that by (ii) no paths leave w_i in W_i. Define $\mathfrak{F}'$ to be the set $\{F \in \mathfrak{F} : F \subseteq W^{(k)}$ for some $k\}$.

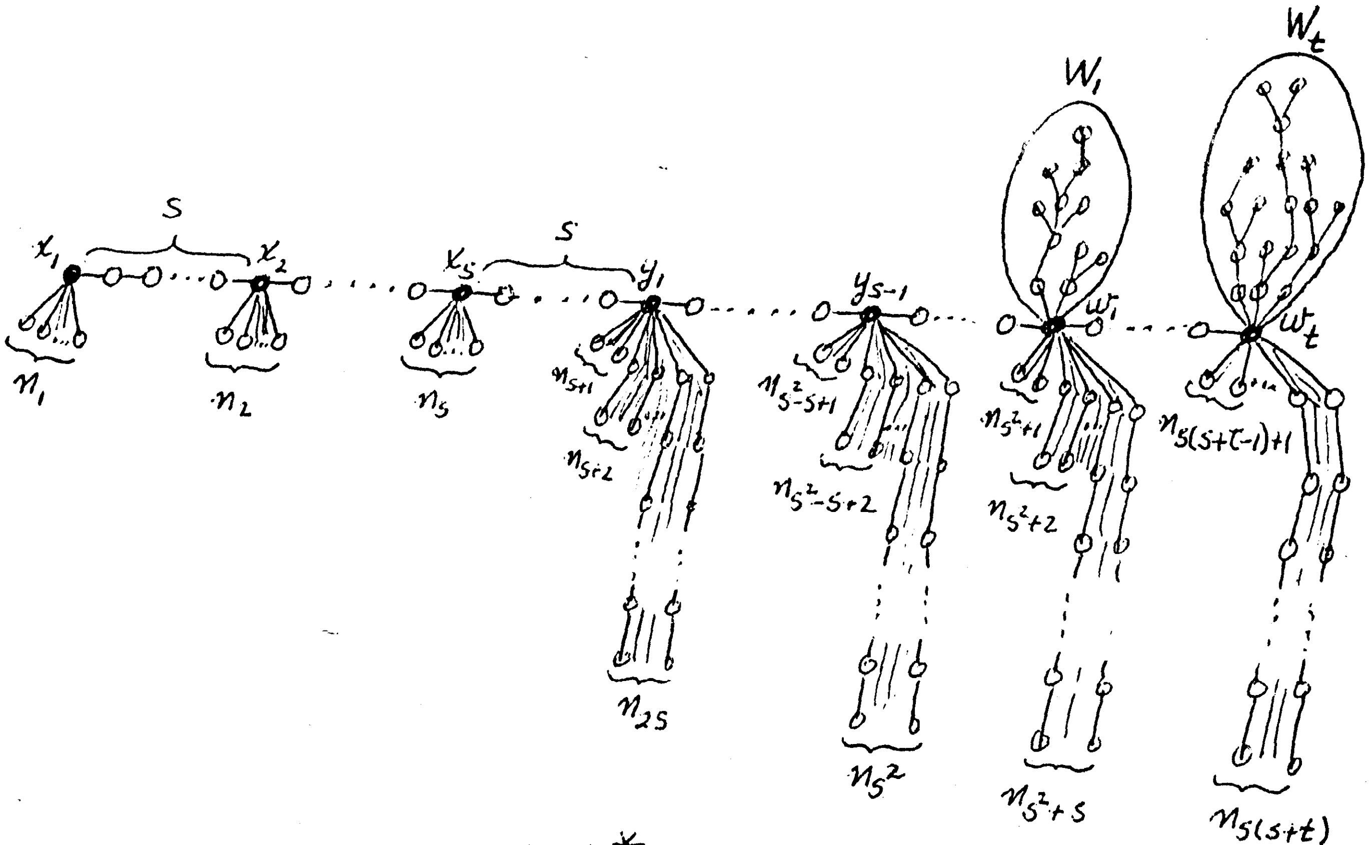
Next, we choose s to be a large fixed integer, depending only on $\mathfrak{F}$, to be determined later. For (large) integers n, define n_k by

$$n_k = [n^{1+s^{-k}}], \; 1 \leq k \leq s(s+t).$$

We are finally ready to define the tree $W^* = W^*(n)$.

1. W^* will have a subset of $2s+t-1$ vertices, called <u>special</u> vertices, denoted by $X = \{x_1, \ldots, x_s\}$, $Y = \{y_1, \ldots, y_{s-1}\}$ and $\{w_1, \ldots, w_t\}$.
2. For $1 \leq k \leq s$, x_k has n_k paths of length 1 attached to it.
3. For $1 \leq k \leq s-1$, y_k has n_{ks+j} paths of length j attached to it for $1 \leq j \leq s$.
4. For $1 \leq k \leq t$, w_k has $n_{s(s+k-1)+j}$ paths of length j attached to it for $1 \leq j \leq s$.
5. Also attached to w_k is a copy of W_k with w_k being the distinguished vertex of W_k.
6. The special vertices are joined sequentially by paths of length s, i.e., between adjacent vertices in the sequence $(x_1, \ldots, x_s, y_3, \ldots, y_{s-1}, w_1, \ldots, w_t)$ are placed paths of length s.

This completes the construction of W^*. In Fig. 3 we illustrate the structure of W^*.



W^*

Fig. 3

By hypothesis, we have

$$\sum_{F\in\mathfrak{F}} A_F N_F(W^*(n)) = A_0$$

for all n. However, since by the definition of $\mathfrak{F}'$, no $F \in \mathfrak{F} - \mathfrak{F}'$ occurs as a subgraph of $W(k)$ for any k, then it is not difficult to see that $N_F(W^*(n)) = 0$ for these F, provided we have chosen s and n sufficiently large. Hence, we have

$$\sum_{F\in\mathfrak{F}'} A_F N_F(W^*(n)) = A_0 \tag{3}$$

for all sufficiently large n. It is important to note that by the minimality assumptions we have made, any embedding of any $F \in \mathfrak{F}'$ into W^* must use <u>all</u> the edges of <u>all</u> the W_i, $1 \le i \le t$, in W^*, again, provided s and n are suffic

<u>Fact</u>. For any distinct F, $F' \in \mathfrak{F}'$, either

$$N_F(W^*)/N_{F'}(W^*) > n^{s^{-s^3}}$$

or

$$N_{F'}(W^*)/N_F(W^*) > n^{s^{-s^3}}$$

for n sufficiently large.

For suppose the Fact holds. Since we must have $|\mathfrak{F}'| > 1$, then there is some element $F^* \in \mathfrak{F}'$ such that

$$N_{F^*}(W^*)/N_F(W^*) > n^{s^{-s^3}}$$

for all $F \in \mathfrak{F}' - \{F^*\}$. By (3) we have

$$(4) \qquad A_{F^*} + \sum_{F \in \mathfrak{F}' - \{F^*\}} A_F\left(\frac{N_F(W^*)}{N_{F^*}(W^*)}\right) = \frac{A_0}{N_{F^*}(W^*)} .$$

But as $n \to \infty$, all terms in (4) tend to zero except A_{F^*} which is nonzero 'by hypothesis. This contradiction would then prove the theorem.

<u>Proof of Fact</u>: Let F and F′ be two distinct elements of $\mathfrak{F}'$. Partition the components of F into three classes: F_1, the set of <u>stars</u>, i.e., trees with at most one vertex of degree > 2; F_2, the non-stars which are <u>star-like</u>, i.e., trees with at most one vertex of degree ≥ 3; and F_3, the non-star-like trees, i.e., those having at least two vertices of degree ≥ 3.

Define F_1', F_2' and F_3' in an analogous way for F'. As we have noted earlier, F_3 must consist of t trees $T_1, \ldots, T_t$ where T_k is formed from W_k by adjoining a (nonempty!) set of paths to w_k (with a similar remark applying to F_3').

We need one more concept. A <u>weak attachment</u> <u>α of F to W*</u> is formed as follows. A vertex u_i is selected from each component C_i of F. These u_i are mapped by an injection α into the set of special vertices of W* with the restrictions that:

$$\alpha(u_i) = \begin{cases} x_j \text{ for some } j \text{ if } C_i \in F_1, \\ y_j \text{ for some } j \text{ if } C_i \in F_2, \\ w_j \text{ for some } j \text{ if } C_i \in F_3. \end{cases}$$

A weak attachment α of F to W* is said to be <u>proper</u> if α can be extended to an embedding of F into W^*. We let $|\alpha|$ denote the number of ways α can be extended to an embedding of F into W^*. Note that in a proper weak attachment α of F to W^*, u_i must be a vertex of C_i of maximal degree if $C_i \in F_1 \cup F_2$. Define the sequence $\tau(\alpha) = (\tau_1, \tau_2, \ldots, \tau_{s(s+t)})$ as follows:

$$\tau_k = \begin{cases} \text{number of paths of length 1 leaving } u_i \text{ for} \\ \alpha(u_i) = x_k,\ 1 < k \leq s, \\ \\ \text{number of paths of length } j \text{ leaving } u_i \text{ for } \alpha(u_i) = y_\ell \\ \text{where } k = \ell s + j \text{ for } 1 \leq j \leq s,\ 1 < \ell \leq s-1, \\ \\ \text{number of paths of length } j \text{ leaving } u_i \text{ for } \alpha(u_i) = w_m \\ \text{where } k = s^2 + (m-1)s + j \text{ for } 1 < j \ldots \end{cases}$$

It is then clear that

$$|\alpha| < K_0 \prod_{k=1}^{s(s+t)} n_k^{\tau_k}$$

where $K_0, K_1, \ldots,$ will denote constants depending on s and not on n. The sequences $\tau(\alpha)$ can be linearly ordered as follows.

For $\tau(\alpha) = (\tau_1, \tau_2, \ldots, \tau_{s(s+t)})$ and $\tau(\alpha') = (\tau'_1, \tau'_2, \ldots, \tau'_{s(s+t)})$, we define $\tau(\alpha') > \tau(\alpha)$ if either:

(i) $\displaystyle\sum_{k=1}^{s(s+t)} \tau'_k > \sum_{k=1}^{s(s+t)} \tau_k$; or

(ii) $\displaystyle\sum_{k=1}^{s(s+t)} \tau'_k = \sum_{k=1}^{s(s+t)} \tau_k$ and $\tau(\alpha')$ is lexicographically greater than $\tau(\alpha)$, i.e., for some m, $\tau'_k = \tau_k$ for $1 \leq k < m$ and $\tau'_m > \tau_m$.

We let $\tau^{(F)} = (\tau_1^{(F)}, \ldots, \tau_{s(s+t)}^{(F)})$ denote a <u>maximal</u> sequence $\tau(\alpha)$ in this ordering as α ranges over all proper weak attachments of F to W^*. The proof of the Fact will depend on the following assertion.

<u>Claim</u>: If $\tau^{(F')} > \tau^{(F)}$ then $N_{F'}(W^*)/N_F(W^*) > n^{s^{-s^3}}$ for n sufficiently large,

<u>Proof of Claim</u>: Suppose $\tau^{(F')} > \tau^{(F)}$. It is easily seen that

$$N_{F'}(W^*) \geq \prod_{k=1}^{s(s+t)} \binom{n_k}{\tau_k^{(F')}} > K_1 \prod_{k=1}^{s(s+t)} n_k^{\tau_k^{(F')}} \tag{6}$$

On the other hand, it is not hard to show that

$$(7) \qquad N_F(W^*) < K_2 \prod_{k=1}^{s(s+t)} n_k^{\tau_k(F)} .$$

To see this, we consider F as a labelled forest and we show that

$$E_F(W^*) < K_3 \prod_{k=1}^{s(s+t)} n_k^{\tau_k(F)}$$

for a suitable constant $K_3 = K_3(s)$.

First, the non-star-like trees in F_3 can only be embedded into the W_i parts of W* and, since the total number of proper weak attachments of F_3 to W* is bounded by a function of s, then the embedding of the non-star-like trees of $\mathcal{F}'$ contributes a factor of at most $K_4 \prod_{k=s^2+1}^{s(s+t)} n_k^{\tau'_k}$ where $\tau'(\beta) = (\tau'_{s^2+1}, \ldots, \tau'_{s(s+t)})$ is a (maximal) sequence derived from some proper weak attachment β of F_3 to W*.

Next, consider an embedding of a star-like tree $T \in F_2$ which is not a star. Suppose T is formed by adjoining m_k paths of length k, $1 < k < s$, to the "center" vertex u. Although it may be possible to embed T into W* by mapping u onto some $x_i \in X$ (e.g., when at most two of the m_k, $k \geq 2$, are nonzero), when this is done we must use edges in one of the paths of length s connecting x_i to adjacent special vertices of W*, and so, there are at most

$K_5 \, n_1^{-1} \, {}^{-\sum_{k=1}^{s} m_k}$ such embeddings, However, this factor is negligible compared to the corresponding factor of

$K_6 \, n_{s^2}^{\sum_{k=1}^{s} m_k}$ which we obtain if we em'bed T by mapping u onto some $y_i \in Y$ since

$$n_{s^2}^{m} / n_1^{m-1} > K_7 \, n^{m(1+s^{-s^2})} / n^{m(1+s^{-1})-1}$$

$$> K_8 \, n^{1/2}$$

provided s has been chosen sufficiently large for $\mathfrak{F}$ and n is sufficiently large.

Finally, we consider a star $S \in F_1$, say, consisting of m paths of length 1 adjoined to a vertex u. If $m \geq 3$ then in any embedding of F into W^*, u must be mapped onto some vertex in $X \cup Y$ since these are the only available vertices of degree ≥ 3. However, since $n_k/n_{k+1} \to \infty$ as $n \to \infty$ then the dominant contribution will certainly come from the embeddings which map u onto some $x_i \in X$ (in fact, the smaller the index i, the better). If $m \leq 2$ then there are many ways of' embedding S into W^*, for example, so that u does not map onto a special vertex of W^*. Again, however, the dominant term clearly comes from those embeddings which take u onto some special vertex $x_i \in X$.

Thus, all except a negligible fraction of the embeddings of F into W^* arc extensions of proper weak attachments α of F to W*. Note that if α and α' are proper weak attachments of F to W* and $\tau(\alpha') > \tau(\alpha)$ then by definition, either

$$\sum_{k=1}^{s(s+t)} \tau'_k > \sum_{k=1}^{s(s+t)} \tau_k$$

or

$$\sum_{k=1}^{s(s+t)} \tau'_k = \sum_{k=1}^{s(s+t)} \tau_k \text{ and for some } m \leq s(s+t),$$

$\tau'_k = \tau k$ for $1 \leq m < k$, and $\tau'_m > \tau_m$.

In the first case,

$$\prod_{k=1}^{s(s+t)} n_k^{\tau'_k} > K_9 \prod_{k=1}^{s(s+t)} n^{\tau'_k (1+s^{-k})}$$

$$= K_9 \cdot n^{\Sigma_{k=1}^{s(s+t)} \tau'_k} \cdot n^{\Sigma_{k=1}^{s(s+t)} \tau'_k s^{-k}}$$

$$\geq K_9 n^{1+\Sigma_{k=1}^{s(s+t)} \tau_k}$$

$$> K_{10} n^{1/2s} \prod_{k=1}^{s(s+t)}$$

for s and n sufficien y large. In the cond case

$$\prod_{k=1}^{s(s+t)} n_k^{\tau'_k} > K_9\, n^{\Sigma_{k=1}^{s(s+t)} \tau'_k} \cdot n^{\Sigma_{k=1}^{s(s+t)} \tau'_k s^{-k}}$$

$$= K_9\, n^{\Sigma_{k=1}^{s(s+t)} \tau_k} \cdot n^{\Sigma_{k=1}^{m-1} \tau_k s^{-k}} \bullet n^{\Sigma_{k=m}^{s(s+t)} \tau'_k s^{-k}}$$

But $\sum_{k=m}^{s(s+t)} \tau'_k s^{-k} \geq (\tau_m+1)s^{-m} = \tau_m s^{-m} + s^{-m}$

and $\sum_{k=m}^{s(s+t)} \tau_k s^{-k} \leq \tau_m s^{-m} + \sum_{k=m+1}^{s(s+t)} s^{1/2} \cdot s^{-k}$

$$\leq \tau_m s^{-m} + 2s^{-m-1/2}.$$

Hence, in either case,

$$(8) \qquad \prod_{k=1}^{s(s+t)} n_k^{\tau'_k} \Big/ \prod_{k=1}^{s(s+t)} n_k^{\tau_k} > K_{11}\, n^{s^{-m}-2s^{-m-1/2}} > K_{11}\, n^{s^{-2s^2}}$$

$$> K_{11}\, n^{1/s^{s^3}}$$

But since there are at most $K_{12} = K_{12}(s)$ proper weak attachments of F to W^* then by (5), (8), and the definition of $\tau^{(F)}$ we have

$$(9) \qquad E_F(W^*) < K_{13} \prod_{k=1}^{s(s+t)} n_k^{\tau_k^{(F)}}$$

Hence, from (7) and (9), we have

$$N_{F'}(W^*)/N_F(W^*) \geq N_{F'}(W^*)/E_F(W^*)$$

$$> K_{14} \prod_{k=1}^{s(s+t)} n_k^{\tau_k^{(F')}} \Big/ \prod_{k=1}^{s(s+t)} n_k^{\tau_k^{(F)}} > n^{1/s\, s^3}$$

for n sufficiently large and the Claim is proved.

From the preceding discussion it is not difficult to see that if $\tau^{(F)} = \tau^{(F')}$ then F and F' are isomorphic which contradicts the hypothesis that they are distinct elements of $\mathfrak{F}'$. Therefore, we must have $\tau^{(F)} \neq \tau^{(F')}$ and so the Fact always holds, provided s is sufficiently large. This completes the proof of the theorem. ∎

CONCLUDING REMARKS

As we have seen in Eq. (1), when F has isolated points then NF(T) can be written as

(9) $$N_F(T) = P(n)N_{F'}(T)$$

where P(n) is a polynomial (depending on F) in $n = |V(T)|$ and F' has no isolated points. However, such an expression, valid for all trees T, can always be written in the form

$$(10) \qquad P(n)N_{F'}(T) = \sum_{F \in \mathfrak{F}_{F'}(d)} A_F N_F(T)$$

where $\mathfrak{F}_{F'}(d)$ consists of all those forests which can be formed by adjoining exactly $d = \deg P(n)$ additional edges to F'. This follows by the observation that

$$(11) \qquad \binom{n-1-|E(F')|}{d} N_{F'}(T) = \sum_{F \in \mathfrak{F}_{F'}(d)} N_{F'}(F) N_F(T)$$

since the-left-hand side of (11) can be interpreted as counting the number of ways of selecting a copy of F' in T together with d additional edges of T. For example, if F' is the forest shown in Fig. 4(a) then

$$(12) \quad (n-4)N_{F'}(T) = 2N_{F_1}(T) + 4N_{F_2}(T) + 2N_{F_3}(T) + 3N_{F_4}(T)$$

where the F_i are given in Fig. 4(b).

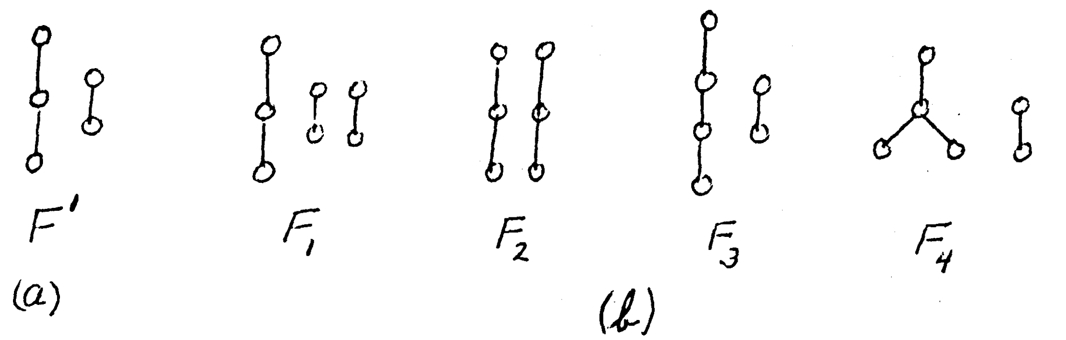


Fig. 4

We remark that if $\mathfrak{F}$ is allowed to be infinite then nontrivial linear dependences among the $N_F(T)$, $F \in \mathfrak{F}$, can exist. For example, if S_k denotes the star with k edges, then for $\mathfrak{F} = \{S_k : k=1,2,\ldots\}$ we have

(13) $$\sum_{k=1}^{\infty} (-1)^{k+1} N_{S_k}(T) = 1$$

for all trees T.

REFERENCES

[1] M. Edelberg, M. R. Garey, and R. L. Graham, On the distance matrix of a tree (to appear).

[2] Distance matrices of trees, R. L. Graham and L. Lovász (to appear).

[3] F. Harary, Graph Theory, Addison-Wesley, New York (1969).