Studies in Pure Mathematics
To the Memory of Paul Turán

On complete bipartite subgraphs contained in spanning tree complements

by

B. BOLLOBÁS (Cambridge), F. R. K. CHUNG (Murray Hill)
and R. L. GRAHAM (Murray Hill)

Abstract

The celebrated theorem of TURÁN answers the following question exactly:

How many edges can a graph with n vertices have without containing the complete graph K_r as a subgraph?

In a recent paper, ERDŐS, FAUDREE, ROUSSEAU and SCHELP investigate the analogous question for the complete bipartite graph $K_{a,b}$. In particular, they study the following problem: What is the largest number $f(n, k)$ such that no matter how $f(n, k)$ edges are deleted from K_n, the resulting graph always contains $K_{a,b}$ as a subgraph for all a and b satisfying $a+b\leqq n-k$. ERDŐS et al. show that for $\varepsilon<e^{-4}$, it is always possible to remove slightly more than $n/2$ edges from K_n and thereby prevent the occurrence of some $K_{a,b}$ in the remaining graph for $a+b\leqq n-[\varepsilon n]$.

The same authors also raise the question of estimating $f(n, k)$ when the removed edges form a *spanning subtree* of K_n. In this note we show that it is possible to obtain surprisingly sharp estimates for this problem.

Introduction

The celebrated theorem of TURÁN [3] answers the following question exactly:

How many edges can a graph with n vertices have without containing the complete graph K_r as a subgraph?

In a recent paper [2], ERDŐS, FAUDREE, ROUSSEAU and SCHELP investigate the analogous question for the complete bipartite graph $K_{a,b}$. In particular, they study the following problem: What is the largest number $f(n, k)$ such that no matter how $f(n, k)$ edges are deleted from K_n, the resulting graph always contains $K_{a,b}$ as a subgraph for all a and b satisfying $a+b\leqq n-k$.

They find that a rather abrupt change occurs as k increases from 1 to 2 by proving:

(i) $f(n,0)=\lceil n/2\rceil-1,\ n\geqq 2$;

(ii) $f(n,1)=\left\lceil \frac{n+1}{2}\right\rceil,\quad n\geqq 3$;

(iii) For fixed $k\geqq 2$ there exist positive constants A and B (depending on k) such that for n sufficiently large

$$n/2+A\sqrt{n}<f(n,k)<n/2+B\sqrt{n}.$$

(iv) For $0<\varepsilon<e^{-4}$, there is a $\delta=\delta(\varepsilon)>0$ so that for n sufficiently large

$$f(n, [\varepsilon n]) < \left(\frac{1}{2} + \delta\right)n.$$

The last result shows that it is always possible to remove slightly more than $n/2$ edges and thereby prevent the occurrence of some $K_{a,b}$ in the remaining graph for $a+b \leqq n-[\varepsilon n]$.

In [2] ERDŐS et al. raise the question of estimating $f(n, k)$ when the removed edges form a *spanning subtree* of K_n. In this case, a relatively large number $(n-1)$ of edges are removed, but they are required to be rather well-behaved. In this note we show that it is possible to obtain remarkably sharp estimates for this problem.

The main result

Let $f(n)$ denote the least integer with the property that for any spanning tree T of K_n and any $a+b \leqq n-f(n)$, $K_{a,b} \subseteq K_n - T$.

Theorem.

$$f(n) = (1+o(1))\frac{\log n}{\log 3}.$$

Proof. Let $\varepsilon > 0$ be fixed. We first show that

(1)
$$f(n) > (1-\varepsilon)\frac{\log n}{\log 3}$$

for infinitely many n. Suppose (1) fails to hold for all sufficiently large n. Take n to be of the form

$$n = (1+3+\ldots+3^{2r-1}) = \frac{1}{2}(3^{2r}-1)$$

for a large r to be specified later. For the spanning tree T of K_n we choose T_{2r}, the complete *ternary* tree with $2r$ levels (see Fig. 1)

Finally, we take a to be $\frac{1}{4}(3^{2r}-1)$.

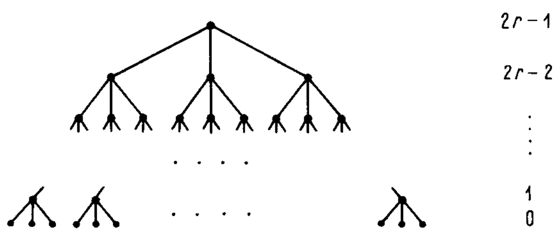

Fig.1. T_{2r}, the Complete Ternary Tree with $2r$ Levels

Assume now that for a fixed $\varepsilon>0$ there is a number b with

$$a+b>n-(1-\varepsilon)\frac{\log n}{\log 3}$$

such that $K_{a,b}\subseteq K_n-T_{2r}$. Note that expressed to the base 3,

$$n=\overbrace{11\ldots11}^{2r},\quad a=\overbrace{202\ldots02}^{r\,2\text{'s}}.$$

Also,

$$\frac{\log n}{\log 3}<2r-\frac{\log 2}{\log 3}<2r.$$

Thus,

$$b>n-a-(1-\varepsilon)\frac{\log n}{\log 3}>n-a-2(1-\varepsilon)r. \tag{2}$$

The number of vertices of K_n which do not belong to $K_{a,b}$ is just $n-(a+b)$ which is at most

$$(1-\varepsilon)\frac{\log n}{\log 3}<2(1-\varepsilon)r.$$

Their removal splits T_{2r} up into components C_i. The two vertex sets A and B of $K_{a,b}$ (where $|A|=a$, $|B|=b$) must each be contained in disjoint unions of these C_i. The plan is to show that any union of C_i's which contains at least a vertices must in fact contain more than $a+2(1-\varepsilon)r$ vertices, leaving fewer than $n-(a+2(1-\varepsilon)r)<b$ vertices from which to form B, which of course is a contradiction.

Let the number $n-(a+b)$ of vertices which do not belong to $K_{a,b}$ be denoted by αr. Thus

$$\alpha<2(1-\varepsilon).$$

We must make a careful analysis of the sizes and relationships of the components C_i into which T_{2r} is partitioned by the removal of the αr unused vertices.

To begin with, we will always remove lower level vertices before higher level vertices, i.e., vertices closer to the bottom are removed first. Thus, if vertex v in level m is removed then the top (parent) tree loses $\frac{1}{2}(3^{m+1}-1)$ vertices and three new complete ternary trees T_m are formed, each with $\frac{1}{2}(3^m-1)$ vertices (see Fig. 2). We say that the resulting four quantities

$$-\frac{1}{2}(3^{m+1}-1),\quad \frac{1}{2}(3^m-1),\quad \frac{1}{2}(3^m-1),\quad \frac{1}{2}(3^m-1)$$

form a *family*.

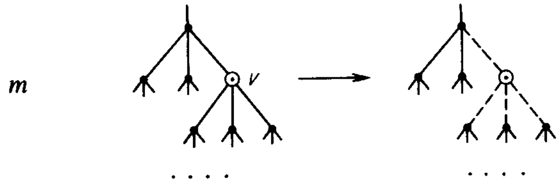

Fig. 2

After all αr points have been deleted, $\alpha r+1$ families will be formed. Let us examine what happens if some of the C_i are combined to form a set C containing A. The cardinality $|C|$ of C is a sum of $|C_i|$. For each family of the form $-\frac{1}{2}(3^{m+1}-1), \frac{1}{2}(3^m-1), \frac{1}{2}(3^m-1), \frac{1}{2}(3^m-1)$ there are just 8 possible sums it can contribute to $|C|$, depending on which members of the family occur in C. These sums are:

$$-\frac{1}{2}(3\cdot 3^m-1),\ -\frac{1}{2}(2\cdot 3^m),\ -\frac{1}{2}(3^m+1),\ -1,$$

$$\frac{1}{2}(3^m-1), \frac{1}{2}(2\cdot 3^m-2), \frac{1}{2}(3\cdot 3^m-3) \quad \text{and} \quad 0.$$

Hence, each family contributes an amount of the form $\frac{1}{2}(\delta\cdot 3^m+\gamma)$ to $|C|$ where $\delta=\pm 3, \pm 2, \pm 1$ or 0 and $|\gamma|\leqq 3$.

Since $A\subseteq C$ then

(3) $$\frac{1}{4}(3^{2r}-1)=a=|A|\leqq|C|=\sum_{i=0}^{\alpha r}\frac{1}{2}(\delta_i\cdot 3^{m_i}+\gamma_i)$$

Thus, we have

(4) $$\frac{1}{2}(3^{2r}-1)\leqq\sum_{i=0}^{\alpha r}\delta_i 3^{m_i}+\sum_{i=0}^{\alpha r}\gamma_i\leqq\sum_{i=0}^{\alpha r}\delta_i 3^{m_i}+3(\alpha r+1).$$

Since $3(\alpha r+1)$ has at most

$$\left\lceil\frac{\log 3(\alpha r+1)}{\log 3}\right\rceil\leqq\frac{2\log r}{\log 3}$$

nonzero digits base 3 (for large r) then $\frac{1}{2}(3^{2r}-1)-\sum_{i=0}^{\alpha r}\gamma_i$ has the form

$$\overbrace{111\ldots11}^{2r-v}x_1\ldots x_v \quad \text{(base 3)}$$

where $v<\dfrac{2\log r}{\log 3}$.

So, by (4) we have

$$\sum_{i=0}^{\alpha r} \delta_i 3^{m_i} \geqq \overbrace{111\ldots11}^{2r-v} x_1 \ldots x_v \geqq \sum_{j=1}^{2r-v} 3^{v+j-1} \tag{5}$$

where $v < \dfrac{2\log r}{\log 3}$, $\alpha < 2(1-\varepsilon)$ and $\delta_i = \pm 3,\ \pm 2,\ \pm 1$ or 0.

We now proceed to normalize the sum $\sum_{i=0}^{\alpha r} \delta_i 3^{m_i}$. To begin with, we may assume that $|\delta_i| \leqq 2$ since otherwise, we simply increase m_i by one. Next, observe that if $m_i = m_j$ for some $i \neq j$ then by writing

$$\delta_i + \delta_j = 3\delta_{i'} + \delta_{j'}$$

with $|\delta_{i'}| \leqq 1$, $|\delta_{j'}| \leqq 2$, we have

$$\delta_i 3^{m_i} + \delta_j 3^{m_j} = \delta_{i'} 3^{m_i+1} + \delta_{j'} 3^{m_i}$$

where either at least one more δ is 0 or $\sum_i |\delta_i|$ has decreased. Hence, we can continue this process until all the m_i in (5) are distinct. We can therefore rewrite

$$\sum_{i=0}^{\alpha r} \delta_i 3^{m_i} \quad \text{as} \quad \sum_{i=0}^{2r} \beta_i 3^i$$

where $\beta_i = \pm 2,\ \pm 1$ or 0 and *at most* $\alpha r + 1$ β's *are nonzero*. Equation (5) becomes

$$\sum_{i=0}^{2r} \beta_i 3^i \geqq \sum_{j=1}^{2r-v} 3^{v+j-1}\,. \tag{5'}$$

Next, suppose for some i we have

$$\beta_{i+1} = 2, \quad \beta_i = -2\,.$$

Since

$$2 \cdot 3^{i+1} - 2 \cdot 3^i = 1 \cdot 3^{i+1} + 1 \cdot 3^i$$

we can replace these β's by the new values $\beta_{i+1} = 1$, $\beta_i = 1$, thereby decreasing $\sum_i |\beta_i|$. In the same way $(\beta_{i+1}, \beta_i) = (1, -2), (2, -1)$ and $(1, -1)$ can be replaced by $(\beta_{i+1}, \beta_i) = (0, 1), (1, 2)$ and $(0, 2)$, respectively, thereby either decreasing $\sum_i |\beta_i|$ or keeping $\sum_i |\beta_i|$ constant and decreasing the number of negative β's. Hence, we may assume that in the sequence $\bar{\beta} = (\beta_{2r}, \beta_{2r-1}, \ldots, \beta_0)$, no positive β_{i+1} is followed by a negative β_i.

Let i_0 be the largest index such that $\beta_{i_0} \neq 1$, $i_0 < 2r$. We know that $i_0 \geqq (2-\alpha)r$ since at most $\alpha r + 1$ β's are nonzero. There are several possibilities:

(a) Suppose $\beta_{2r} \neq 0$. Then by the construction of the C_i, β_{2r} is positive and by the normalization process, $\beta_{2r-1} \geqq 0$ so that

$$\sum_{i=0}^{2r} \beta_i 3^i \geqq 3^{2r} - 2\sum_{i=0}^{2r-2} 3^i > 3^{2r} - 3^{2r-1} = 2\cdot 3^{2r-1}. \tag{6}$$

(b) Suppose $\beta_{2r}=0$ and $\beta_{i_0}=0$. Thus, $\bar\beta$ looks like

$$\bar\beta = (0, 1, 1, \dots, 1, 0, \beta_{i_0-1}, \dots, \beta_0)$$

and

$$\sum_{i=0}^{2r} \beta_i 3^i \leqq \sum_{i=i_0+1}^{2r-1} 3^i + 2\cdot\sum_{i=0}^{i_0-1} 3^i < \sum_{i=i_0}^{2r-1} 3^i$$

which contradicts (5′) for large r since

$$v < \frac{2\log r}{\log 3} \quad \text{and} \quad i_0 \geqq (2-\alpha)r > 2\varepsilon r.$$

(c) Suppose $\beta_{2r}=0$ and $\beta_{i_0}=2$. Thus, $\bar\beta$ looks like

$$\bar\beta = (0, 1, 1, \dots, 1, 2, \beta_{i_0-1}, \dots, \beta_0)$$

where $\beta_{i_0-1} \geqq 0$ by the normalization process. Hence

$$\sum_{i=0}^{2r} \beta_i 3^i \geqq \sum_{i=i_0+1}^{2r-1} 3^i + 2\cdot 3^{i_0} - 2\cdot\sum_{i=0}^{i_0-2} 3^i > \sum_{i=i_0}^{2r-1} 3^i + 2\cdot 3^{i_0-1}. \tag{7}$$

Note that the preceding arguments apply even if $i_0 = 2r-1$. Therefore, if (5′) holds then by (6) and (7)

$$\begin{aligned} |C| &= \frac{1}{2}\sum_{i=0}^{\alpha r} (\delta_i 3^{m_i} + \gamma_i) \geqq \frac{1}{2}\sum_{i=0}^{2r} \beta_i 3^i - \frac{3}{2}(\alpha r+1) > \\ &> \frac{1}{2}\sum_{i=i_0}^{2r-1} 3^i + 3^{i_0-1} - \frac{3}{2}(\alpha r+1). \end{aligned} \tag{8}$$

Consequently, the number of vertices of K_n which can be used for B is at most

$$\begin{aligned} n - \alpha r - |C| &< n - \frac{1}{2}\sum_{i=i_0}^{2r-1} 3^i - 3^{i_0-1} + \frac{3}{2}(\alpha r+1) - \alpha r < \\ &< \frac{1}{2}\sum_{i=i_0}^{2r-1} 3^i + \sum_{i=0}^{i_0-2} 3^i + \alpha r < \frac{1}{2}\sum_{i=i_0}^{2r-1} 3^i + \frac{1}{6}\cdot 3^{i_0} + \alpha r. \end{aligned} \tag{9}$$

On the other hand, by (2)

$$|B| = b > n - a - 2(1-\varepsilon)r$$

$$(10)\qquad \begin{aligned} &> \frac{1}{2}\sum_{i=0}^{2r-1} 3^i - 2(1-\varepsilon)r = \frac{1}{2}\sum_{i=i_0}^{2r-1} 3^i + \frac{1}{2}\sum_{i=0}^{i_0-1} 3^i - 2(1-\varepsilon)r > \\ &> \frac{1}{2}\sum_{i=i_0}^{2r-1} 3^i + \frac{1}{4}\cdot 3^{i_0} - 2r\,. \end{aligned}$$

Since $\alpha < 2$ and $i_0 > 2\varepsilon r$ then (9) and (10) are contradictory for large r. This shows that (1) must hold for infinitely many n.

To show that in fact (1) holds for all sufficiently large n, a very similar but somewhat more detailed calculation must be made. In place of a complete ternary tree we now use for T a "balanced" ternary tree $T(n)$ on n points. These trees are structurally very similar to complete ternary trees. However, they have the property that when a vertex at level m is removed from $T(n)$, the three new trees formed from vertices at levels below m are all balanced ternary trees on sets of vertices which now can differ by at most one in cardinality. It is not hard to show that such trees exist on any number of vertices. By keeping careful track of the perturbations produced by these trees in the preceding argument, a similar contradiction results. In fact, this argument shows that

$$f(n) > \frac{\log n}{\log 3} - \frac{(1+\varepsilon)\log n}{\log\log n}$$

for any fixed $\varepsilon > 0$ and all sufficiently large n.

We next must show that

$$(11)\qquad f(n) < (1+\varepsilon)\frac{\log n}{\log 3}$$

for all sufficiently large n. The basic fact we need for the proof of (11) is the following result.

Lemma. (F. R. K. Chung and R. L. Graham [1]). *Let T be a tree with at least α vertices. Then for some vertex p of T, the components $T_1, \ldots, T_d$ of T formed by the removal of p (and all incident edges) contain a subset $T_{i_1}, \ldots, T_{i_k}$ such that*

$$\alpha \leqq \sum_{j=1}^{k} |T_{i_j}| \leqq 2\alpha\,.$$

It follows from repeated application of the lemma that any number $x < |T| - \left[\frac{\log|T|}{\log 3}\right]$ can be written as a sum of component sizes formed by the deletion of at most $\left[\frac{\log|T|}{\log 3}\right]$ vertices from T. For, each time we delete a vertex, we can guarantee that with the new components formed, the distance from the target value a to a sum of component sizes we can achieve decreases by a factor of at most 1/3, since if a correction term Δ is needed at stage j, the by generating a component sum in $\left(\frac{2}{3}\Delta, \frac{4}{3}\Delta\right)$ the new correction term is at most $\frac{1}{3}\Delta$. (The reader should have no difficulty filling in the details.) This proves (11) and the theorem is proved.

It would be interesting to know to what extent results like this hold for almost all a, b with $a+b$ small enough, and, more generally, for almost all spanning trees of K_n.

References

[1] F. R. K. CHUNG and R. L. GRAHAM, On Graphs Which Contain all Small Trees. *J. Combinatorial Theory* (*B*), **24** (1978), 14–23.

[2] P. ERDOS, R. J. FAUDREE, C. C. ROUSSEAU and R. H. SCHELP, *On the Extremal Problem for Complete Bipartite Graphs*, to appear

[3] P. TURÁN, On an Extremal Problem in Graph Theory (Hungarian), *Mat. Fiz. Lapok*, **48** (1941) 436–452.

CAMBRIDGE UNIVERSITY
UNITED KINGDOM

BELL LABORATORIES
MURRAY HILL, NJ 07974
U.S.A.