Studies in Pure Mathematics
To the Memory of Paul Turán

On the number of complete subgraphs of a graph II

by

L. LOVÁSZ and M. SIMONOVITS (Budapest)

Abstract

Generalizing some results of P. ERDŐS and some of L. MOSER and J. W. MOON we give lower bounds on the number of complete p-graphs K_p of graphs in terms of the numbers of vertices and edges. Further, for some values of n and E we give a complete characterization of the extremal graphs, i.e. the graphs S of n vertices and E edges having minimum number of K_p's. Our results contain the proof of the longstanding conjecture of P. ERDŐS that a graph G^n with $[n^2/4]+k$ edges contains at least $k\left[\frac{n}{2}\right]$ triangles if $k<n/2$.

0. Notation

The graphs in this paper will be denoted by capital letters. We shall exclude loops and multiple edges, and all graphs will be non-oriented.

Let G be a graph: $e(G)$ will denote the number of edges of G, $v(G)=n$ the number of vertices. If x is a vertex, $st(x)$ will denote the set of neighbors of x, that is the set of vertices joined to x. $\sigma(x)$ will denote the cardinality of $st(x)$, that is, the degree of x and if we consider more graphs on the same set of vertices, $st_G(x)$, $\sigma_G(x)$ will denote the star and the degree in G. If G is a graph and A is a set of vertices of G, then $G(A)$ will denote the subgraph spanned by A. For given $n_1, \ldots, n_d$ $K_d(n_1, \ldots, n_d)$ is the complete d-partite graph with n_i vertices in its ith class. $K_d := K_d(1, \ldots, 1)$ is the complete d-graph and $k_d(G)$ denotes the number of complete K_d's of G. If A is a set of vertices and edges of G, $G-A$ denotes the graph obtained by deleting the vertices and edges of A from G and deleting all the edges incident to a vertex in A. If (x, y) does not belong to G, $G+(x, y)$ is the graph obtained by adding the edge (x, y) to G.

1. Introduction

Let $f_p(n, E) = \min\{k_p(G): e(G)=E,\ v(G)=n\}$.

Problem 1. *Determine the function $f_p(n, E)$.*

Problem 2. *Characterize the extremal graphs for given n and E, i.e. those graphs S for which $v(S)=n$, $e(S)=E$ and $k_p(S)=f_p(n, E)$.*

The history of Problem 1 is the following.

In 1941 TURÁN [7] proved that if $n \equiv r \pmod{p-1}$ and $0 \leqq r \leqq p-2$ and if

$$m(n, p) = \frac{p-2}{2(p-1)}(n^2 - r^2) + \binom{r}{2},$$

then every G on n vertices having at least $m(n, p)+1$ edges contains at least one K_p. For $E = m(n, p)$ there exists exactly one graph $T^{n,p-1}$ having n vertices and E edges and containing no K_p. This $T^{n,p-1}$ is a $K_{p-1}(n_1, \ldots, n_{p-1})$ where $\sum n_i = n$ and $|n_i - n/d| < 1$. RADEMACHER proved (unpublished) that any G with n vertices and $\geqq m(n, 3)+1$ edges contains not only one but at least $\left[\frac{n}{2}\right]$ K_3's. ERDŐS [2, 3], (first only for $p=3$, then for any $p \geqq 3$) proved the following.

Theorem A. *Let U_k^n denote a graph obtained from $T^{n,p-1}$ by adding k edges to it so that the new edges belong to the same class having maximum number of vertices (i.e. $[n/d]+1$ if n/d is not an integer, n/d otherwise) and the new edges do not form triangles, if this is possible. Then there exists a constant $c_p > 0$ such that for $k < c_p n$, U_k^n is an extremal graph of Problem 1; i.e. if*

$$v(G) = n \quad e(G) = e(U_k^n) = m(n, p-1) + k,$$

then

$$k_p(G) \geqq k_p(U_k^n) = k \prod_{0 \leqq i \leqq p-3} \left[\frac{n+i}{p-1}\right].$$

Problem 3. (ERDŐS). *How large can c_p be in the theorem above?*

Remark 1. If we add $k+1$ or more edges to the first class of $G = K_{p-1}(k+1, k, k, \ldots, k, k-1)$, then each new edge will be contained only in $(k-1)k^{p-3}$ K_p's and it is easy to see that this construction is better than $U_k^{k(p-1)}$. Thus Theorem A does not hold for $c_p > \frac{1}{p-1}$.

This paper contains an improvement of Theorem A (see Theorem 4 below) which yields that in Problem 3 the answer is $c = 1/(p-1)$. For $p=3$ the proof of this was given in [5]. The result will follow from a much more general theorem characterizing the extremal graphs of Problem 1 for many values of n and E. Before stating our results we introduce some notation.

Let p, n and E be integers such that $p \geqq 3$ and $m(n, p) \leqq E \leqq \binom{n}{2}$. We write E in the form

$$E = \left(1 - \frac{1}{t}\right)\frac{n^2}{2}$$

and set $d = \lfloor t \rfloor$. Thus

$$m(n, d+1) \leqq E < m(n, d+2).$$

We set $k=E-m(n,d+1)$. The numbers p and d will be considered fixed and n large relative to them.

The first theorem we state was proved for $p=3$ by GOODMAN [4] and it readily follows from results of MOON and MOSER [6]. We shall give a self-contained proof because some steps in the proof will be used later.

Theorem 1. *Let $v(G)=n$, $e(G)=E$, then*

$$k_p(G) \geqq \binom{t}{p}\left(\frac{n}{t}\right)^p. \tag{1}$$

Theorem 2. Let C be an arbitrary constant. There exist positive constants δ and C' such that if $0<k<\delta n^2$ and G is a graph on n vertices for which

$$k_p(G) \leqq \binom{t}{p}\left(\frac{n}{t}\right)^p + Ckn^{p-2} \tag{2}$$

then there exists a $K_d(n_1, \ldots, n_d)$ such that $\sum n_i = n$, $\left|n_i - \frac{n}{d}\right| < C'\sqrt{k}$, and G can be obtained from this $K_d(n_1, \ldots, n_d)$ by adding less than $C'k$ edges to it and then deleting less than $C'k$ edges from it.

Remark 2. *Theorem 2 is a "stability theorem" in the following sense: Let U_k^n be the graph obtained from $T^{n,d}$ by adding k edges to it (see Theorem A), then the k "extra edges" are contained in (approximately) $k\binom{d-1}{p-2}\left(\frac{n}{d}\right)^{p-2}$ K_p's and the graph $T^{n,d}$ has $\approx \binom{t}{p}\left(\frac{n}{t}\right)^p$ K_p's. Thus (2) means that G does not have much more K_p's than an extremal graph. Theorem 2 asserts that in this case G^n is very similar to $T^{n,d}$. This theorem is interesting only if k/n^2 is sufficiently small.*

Remark 3. Theorem 2 is sharp: $C'\sqrt{k}$ cannot be replaced by $o(\sqrt{k})$, $C'k$ cannot be replaced by $o(k)$. Indeed, if we add $3k$ edges to and delete k edges from $K_d\left(\frac{n}{d}+\sqrt{k}, \frac{n}{d}-\sqrt{k}, \frac{n}{d}, \ldots, \frac{n}{d}\right)$, then for the resulting graph G

$$k_p(G) \leqq \left(\frac{n}{d}\right)^p\binom{d}{p} + k\binom{d-1}{p-2}\left(\frac{n}{d}\right)^{p-2} < \left(\frac{n}{t}\right)^p\binom{t}{p} + k\binom{d-1}{p-2}\left(\frac{n}{d}\right)^{p-2}$$

while

$$e(G) = m(n, d+1) + k.$$

To formulate our main result we need to describe some classes of graphs.

Definition 1. Let $U_0(n, E)$ denote the class of those graphs with n points and E edges which arise from a complete d-partite graph S_0 by adding edges so that these new edges form no triangles. Let $U_1(n, E)$ denote the subclass where all new edges are contained in the same colorclass of S_0.

Definition 2. Let $U_2(n, E)$ denote the class of those graphs S with n points and E edges which have a set W of independent points such that $S-W$ is complete d-partite, and every point in W is connected to all points of all but one color-classes of $S-W$.

Theorem 3. *There exists a positive constant $\delta=\delta(p, d)$ such that if $0 \leqq k < \delta n^2$ then every extremal graph for Problem 1 is in the class $U_1(n, E)$ if $p \geqq 4$ and is in the class $U_0(n, E) \cup U_2(n, E)$ if $p=3$. In this latter case there exists at least one extremal graph in $U_1(n, E)$.*

We regard this theorem as a complete solution of Problem 1 for the values of n and E under consideration. However, this interpretation requires some explanation, since not all graphs in the classes U_0, U_1 or U_2 have the same number of K_p's and hence, not all of them are extremal. But once we know that our graph is in U_0, U_1 or U_2, its structure is simple enough to determine the best choice by simple arithmetic. Some remarks are in order here:

Proposition 1. *Let $S \in U_1(n, E)$ be extremal. Let $S_0 = K_d(n_1, \ldots, n_d)$, $n_1 \geqq \ldots \geqq n_d$. Then all edges in $E(S)-E(S_0)$ are spanned by the largest class. Furthermore, $|n_i - n_j| \leqq 1$ for $i, j \geqq 2$.*

Given a sequence $n_1 \geqq \ldots \geqq n_d$, all graphs with the above structure have the same number of K_p's. Hence their structure is completely determined if we know the value of n_1. This can be done by simple arithmetic which is not discussed here. We remark that it turns out that

$$n_1 = \frac{n}{d} + \frac{d-1}{d}\frac{k}{n} + o\left(\frac{k}{n}\right), \quad n_i = \frac{n}{d} - \frac{1}{d}\frac{k}{n} + o\left(\frac{k}{n}\right). \tag{3}$$

Proposition 2. *If $S \in U_0(n, E)$ is an extremal graph, then (for $k \leqq \delta n^2$) by moving all edges of $E(S)-E(S_0)$ to the largest color-class we can construct an $\ddot{S} \in U_1(n, E)$ for which $k_p(\ddot{S}) \leqq k_p(S)$. If we moved edges from a smaller class to a larger one, or if $p \geqq 4$, then $k_p(S) > k_p(\ddot{S})$, which contradicts that S is extremal. Thus if $S \notin U_1(n, E)$, then $p=3$ and all the edges of $E(S)-E(S_0)$ belong to color-classes of maximum size in S.*

Proposition 3. *Let $S \in U_2(n, E)$ be an extremal graph. Then every $x \in W$ is connected to all points of all but a possibly smallest color-class of $S-W$. Let B_0 be a smallest color-class of $S-W$. Then, if we change the graph S by connecting every $x \in W$ to all points of $S-W-B_0$ and an appropriate number of points in B_0, we get another extremal graph S'. This graph S' is in $U_1(n, E)$.*

These remarks make the following conjecture plausible:

Conjecture: *For every n and E* ($n \geqq n_0(p)$) *there is an extremal graph in* $U_1(n, E)$.

Let us consider the case when $p=d+1$ and $k < \left[\frac{n}{d}\right]$. Let S be an extremal graph in $U_1(n, E)$. Let $S_0=K_d(n_1, \ldots, n_d)$ and $n_1 \geqq \ldots \geqq n_d$. If the choice of S is not unique, choose one with n_1 minimal. We claim that $n_1 \leqq n_d+1$, i.e. $S_0=T^{n,d}$. Suppose that $n_1 \geqq n_d+2$. Let r denote the number of edges in $E(S)-E(S_0)$. Then simple computation and (3) yield that

$$r \leqq k + \sum_{i=1}^{d} \left(\frac{n}{d} - n_i\right)^2 \leqq \frac{n}{d} + O(1). \tag{4}$$

We have

$$k_p(S)=r \cdot n_2 \ldots n_d,$$

but if we add $r+n_d-n_1+1$ edges to $K_d(n_1-1, n_2, \ldots, n_{d-1}, n_d+1)$, then we get a graph S' with the same number of edges but, by the extremality of S and n_1, with $k_p(S') \geqq k_p(S)$. Hence

$$r \cdot n_2 \ldots n_d \leqq (r+n_d-n_1+1) n_2 \ldots (n_d+1),$$

or

$$r \geqq (n_1-n_d-1)(n_d+1). \tag{5}$$

Now, either $n_1 \leqq n_d+1$ and hence $S_0=T^{n,d}$, which we wish to prove, or by (3), $n_d = \frac{n}{d} + O(1)$, by (4) and (5)

$$(n_1-n_d-1)(n_d+1) \leqq \frac{n}{d} + O(1),$$

and therefore $n_1=n_d+2$, if n is sufficiently large. By Proposition 1 $n_i \leqq n_d+1=n_1-1$ for $i \geqq 2$. Thus, if S_0' is the complete d-partite graph of S', then $S_0'=T^{n,d}$ and so $k=r+n_d-$ $-n_1+1=r-1$. By (5),

$$k \geqq n_d+1 \geqq \left[\frac{n}{d}\right]$$

a contradiction.

Thus, assuming Theorem 3, we have proved

Theorem 4. *If* $E=m(n, p-1)+k$, *where* $k < \left[\frac{n}{p-1}\right]$, *then for* $p>3$ *the only, for* $p=3$ *one possible graph with n points and E edges, containing the least number of* K_p*'s is obtained by adding k edges to a largest class of* $T^{n,d}$.

Theorem 4 is clearly a sharpening of Erdős's Theorem 1.

We investigate one more special case. Let $0<x<1$ and $E\approx x\cdot\binom{n}{2}$, $n\to\infty$. Let S be a graph in $U_1(n, E)$ with minimum number of K_p's. Then

$$k_p(S)\approx f(x)\binom{n}{p},$$

where $f(x)$ can be determined as follows. If $1-\frac{1}{d}\leqq x\leqq 1-\frac{1}{d+1}$ and S is obtained from $S_0=K_d(n_1, n_2, \dots, n_d)$, then we put $n_1=(1-\alpha)n$ and for $i=2, \dots, d$, by $|n_i-n_j|<1 n_i\approx\frac{\alpha}{d-1}n$. Clearly,

$$(*)\qquad k_p(S)\approx\left\{x\binom{n}{2}-\alpha(1-\alpha)n^2-\binom{d-1}{2}\left(\frac{\alpha}{d-1}\right)^2 n^2\right\}\left(\frac{\alpha n}{d-1}\right)^{p-2}\binom{d-1}{p-2}+$$
$$+(1-\alpha)n\binom{d-1}{p-1}\left(\frac{\alpha n}{d-1}\right)^{p-1}+\binom{d-1}{p}\left(\frac{\alpha n}{d-1}\right)^{p}.$$

(Here $\{\dots\}$ is the number of edges in the first class of $K_d(n_1, \dots, n_d)$, $\{\dots\}\cdot\left(\frac{\alpha n}{d-1}\right)^{p-2}\binom{d-1}{p-2}$ is the number of K_p's containing such an edge. The next two terms are the numbers of K_p's containing 1 or 0 vertices from the first class.) Thus

$$\frac{k_p(S)}{\binom{n}{p}}\approx\{A\alpha^2+B\alpha+C\}\cdot\alpha^{p-2}=F(\alpha, x)$$

where $A=A(x,p,d)$, $B=B(x,p,d)$, $C=C(x,p,d)$ are constants easily calculated. $\frac{d}{d\alpha}F(\alpha, x)=0$ yields a quadratic equation, from which the optimal α can easily be determined. Substituting this α in (*) we obtain $f(x)$.

Define

$$g(x)=\liminf\left\{\frac{k_3(G^n)}{\binom{n}{3}}: e(G)\geqq x\binom{n}{2}\right\}.$$

Figure 1 shows what we know about the function $g(x)$. The dotted line shows the Goodman bound. This is equal to $g(x)$ when $x=1-\frac{1}{d}$, d integer. The broken line shows

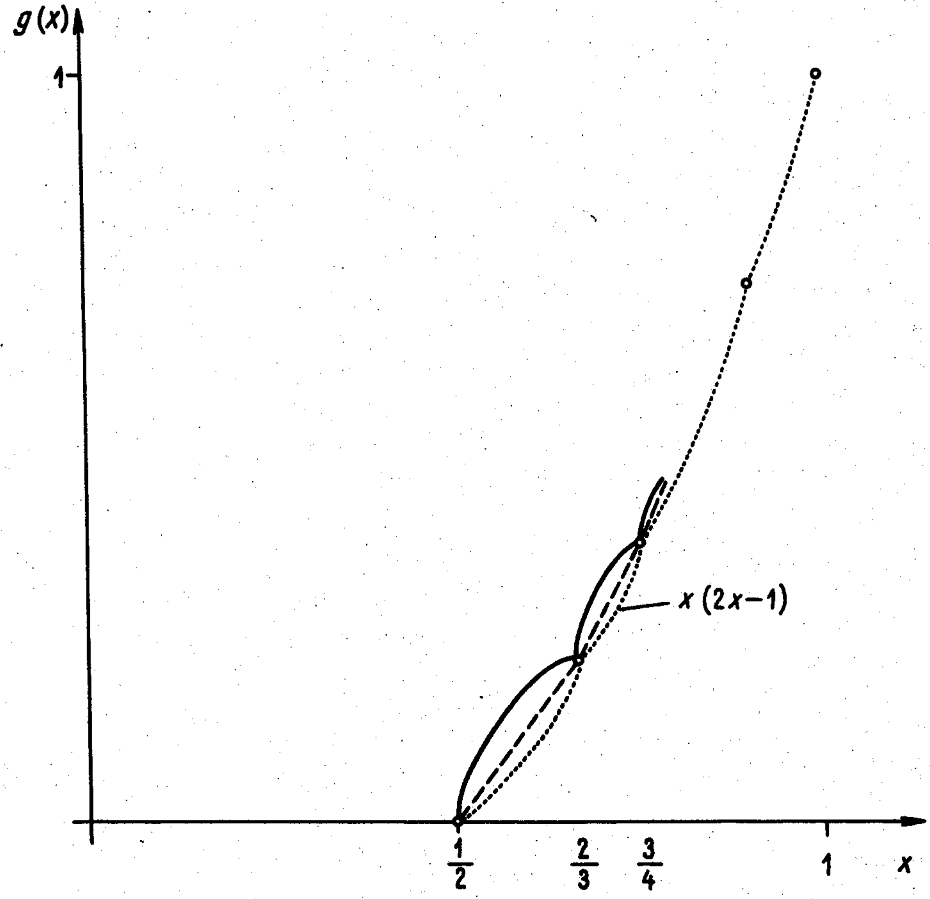

Fig. 1

the improvement given by Bollobás [1]. This proves that between these points $g(x)$ is above the chords. Finally, the continuous line shows the function $f(x)$. This is concave between the points $x=1-\frac{1}{d}$. If the conjecture formulated above is true, it follows that $g(x)=f(x)$. Clearly $g(x)\leqq f(x)$ and Theorem 3 implies that for each d there exists an $\varepsilon_d>0$ such that if $1-\frac{1}{d}\leqq x\leqq 1-\frac{1}{d}+\varepsilon_d$ then $f(x)=g(x)$. Unfortunately, ε_d is so small in our proof that we did not even dare to estimate ε_2.

2. Preliminaries: an inequality for the number of complete subgraphs

Let G be a graph with n points and E edges. Set $k_i=k_i(G)$. For each complete $(p-1)$-subgraph U, let $t_{i,U}$ denote the number of points connected to exactly $p-i-1$ points of U. Let t_i denote the number of induced subgraphs which consist of a K_{p-1} and a point joined to exactly $p-i-1$ points of this K_{p-1}. Clearly, for every U

$$\sum_{i=0}^{p-1} t_{i,U}=n-p+1$$

and

$$\sum_U t_{0,U} = p \cdot t_0 = pk_p, \quad \sum_U t_{1,U} = 2t_1,$$

$$\sum_U t_{i,U} = t_i \quad \text{for} \quad i \geqq 2.$$

So

$$k_{p-1} \cdot (n-p+1) = pt_0 + 2t_1 + t_2 + \ldots + t_{p-1}. \tag{6}$$

Denote, for each complete $(p-2)$-graph V, by r_V the number of complete $(p-1)$-graphs containing V. Then

$$\sum_V r_V = (p-1)k_{p-1}, \tag{7}$$

since each K_{p-1} contains exactly $p-1$ K_{p-2}'s.

Moreover

$$\sum_V \binom{r_V}{2} = t_1 + \binom{p}{2} k_p, \tag{8}$$

since any two K_{p-1}'s containing a given V yield a graph counted in $t_0 = k_p$ or t_1 depending on whether or not they are joined or not. Those subgraphs counted in t arise this way uniquely, and those counted in t_0 arise this way $\binom{p}{2}$ times.

Introducing the "deviation from average"

$$q_V = \frac{k_{p-1}}{k_{p-2}}(p-1) - r_V,$$

we have by (7)

$$\sum_V q_V = k_{p-1} \cdot (p-1) - \sum r_V = 0,$$

and hence

$$2\sum_V \binom{r_V}{2} = \sum_V 2\binom{\frac{k_{p-1}}{k_{p-2}}(p-1) - q_V}{2} = \frac{k_{p-1}^2}{k_{p-2}}(p-1)^2 + \sum q_V^2 - (p-1)k_{p-1}.$$

This, $t_0 = k_p$, (6), and (8) yield that

$$nk_{p-1} = pk_p + 2\left(\sum \binom{r_V}{2} - \binom{p}{2} t_0\right) + t_2 + \ldots + t_{p-1} + (p-1)k_{p-1} =$$

$$= pk_p + \frac{k_{p-1}^2}{k_{p-2}}(p-1)^2 + \sum q_V^2 - p(p-1)k_p + (t_2 + \ldots + t_{p-1})$$

whence

$$p(p-2)k_p = \frac{k_{p-1}^2(p-1)^2}{k_{p-2}} - nk_{p-1} + \sum q_v^2 + (t_2 + \dots + t_{p-1}).$$

Thus

$$\frac{k_p}{k_{p-1}} = \frac{1}{p(p-2)} \left\{ \frac{k_{p-1}}{k_{p-2}} (p-1)^2 - n \right\} + R \tag{9}$$

where

$$R = \frac{1}{p(p-2)k_{p-1}} \left\{ \sum q_v^2 + (t_2 + \dots + t_{p-1}) \right\}. \tag{9*}$$

In particular,

$$\frac{k_p}{k_{p-1}} \geqq \frac{1}{p(p-2)} \left\{ \frac{k_{p-1}}{k_{p-2}} (p-1)^2 - n \right\}. \tag{10}$$

This formula was remarked by Moon and Moser [6].

3. Proof of Theorem 1

First we give a lower bound on k_j/k_{j-1}. We shall prove that

$$k_j/k_{j-1} \geqq \frac{t-j+1}{j} \frac{n}{t}. \tag{11}$$

For $j=2$

$$k_2/k_1 = E/n = \left(1 - \frac{1}{t}\right).$$

By induction on j we obtain that

$$k_{j+1}/k_j \geqq \frac{1}{(j+1)(j-1)} \left\{ \frac{t-j+1}{j} \frac{n}{t} j^2 - n \right\} = \frac{t-j+2}{j+1} \frac{n}{t};$$

(we have used (10) for $p=j$ here). This proves (11). Since

$$k_p = k_1 (k_2/k_1)(k_3/k_2) \dots (k_p/k_{p-1}), \quad (k_1 = n),$$

we have, by (11),

$$k_p \geqq \left(\frac{n}{t}\right)^{p-1} \frac{(t-p+1)(t-p+2)\dots(t-1)}{p(p-1)(p-2)\dots 2.1} n = \left(\frac{n}{t}\right)^p \binom{t}{p}.$$

Thus Theorem 1 is proved.

4. Proof of Theorem 2

The basic inequality we shall use to prove Theorem 2 is (under the conditions of the theorem and with the notation of the previous proof

(12) $$\sum q_V^2+(t_2+\ldots+t_{p-1})=O(kn^{p-2}).$$

To establish (12) we shall carry out the proof of Theorem 1 a little more carefully. By Theorem 1 we know that

(13) $$k_{p-1}\geqq\binom{t}{p-1}\left(\frac{n}{t}\right)^{p-1}$$

By (9), (9*), (11) (applied with $j=p-1$) and (13) we obtain that

(14) $$k_p\geqq\frac{1}{p(p-2)}\{k_{p-1}\cdot((k_{p-1}/k_{p-2})(p-1)^2-n)+\sum q_V^2+(t_2+\ldots+t_{p-1})\}\geqq$$
$$\geqq\frac{1}{p(p-2)}\left\{\binom{t}{p-1}\left(\frac{n}{t}\right)^{p-1}\left(\frac{t-(p-1)+1}{p-1}\frac{n}{t}(p-1)^2-n\right)+\right.$$
$$\left.+\sum q_V^2+(t_2+\ldots+t_{p-1})\right\}=\binom{t}{p}\left(\frac{n}{t}\right)^p+\frac{1}{p(p-2)}\{\sum q_V^2+(t_2+\ldots+t_{p-1})\}.$$

This proves (12).

The method we shall use is the following. By an averaging process we show that there must be a complete d-graph K_d in G such that

(i) almost all the vertices of $G-K_d$ are joined to exactly $d-1$ vertices of K_d;

(ii) dividing the vertices of $G-K_d$ into the classes $C_0,\ldots,C_d$, where C_i contains the vertices joined to each vertex of K_d except the ith one ($i=1,\ldots,d$) and C_0 contains the remaining ones almost all the pairs (x,y) ($x\in C_i$, $y\in C_j$, $i\neq j$) belong to G.

It is convenient to reduce the proof first to the case $p=3$. If $p'<p$ and we know that (2) holds for p, then by (11)

$$k_p/k_{p'}=(k_p/k_{p-1})(k_{p-1}/k_{p-2})\ldots(k_{p'+1}/k_{p'})\geqq$$
$$\geqq\frac{(t-p+1)(t-p+2)\ldots(t-p')}{p(p-1)\ldots(p'+1)}\left(\frac{n}{t}\right)^{p-p'},$$

and hence (by (2))

$$k_{p'}\leqq\left(\frac{n}{t}\right)^{p'}\binom{t}{p'}+C''kn^{p'-2}.$$

In particular,

(15) $$k_3(G)\leqq\binom{t}{3}\left(\frac{n}{3}\right)^3+C''kn.$$

On the other hand, by (14),

$$k_3 \geqq \binom{t}{3}\binom{n}{3}^3 + \frac{1}{3}\{\sum q_V^2 + t_2\}, \tag{16}$$

where (14) is applied with $p=3$, V is a vertex of G, r_V (of (7)) reduces to the degree of V, and t_2 is the number of (3,1)-graphs: of subgraphs of 3 vertices with 1 edges. Finally, $q_V=(p-1)\frac{k_2}{k_1} - r_V = \frac{2E}{n} - \sigma(V)$ measures how near is the valence of the vertex V is to the average valence. By (15) and (16)

$$\sum q_V^2 = O(kn), \quad t_2 = O(kn).$$

Let W be a complete d-graph of G and let A_W denote the number of vertices joined to at most $d-2$ vertices of W. If z is a vertex joined to at most $d-2$ vertices of A_W, then there is an edge (x, y) in W forming a (3,1)-graph with z. A given (3,1)-graph is counted only $O(n^{d-2})$ times in $\sum A_W$, hence

$$\sum_W A_W = O(kn)\cdot O(n^{d-2}) = O(kn^{d-1}). \tag{17}$$

Let B_W be the number of pairs $(x, y) \notin E(G)$ such that either both x and y are joined to exactly $d-1$ vertices of W but these $d-1$ vertices are different for x and y, or x is joined to all vertices of W and y is joined to exactly $d-1$ ones. We can find a $z \in W$ joined to x but not joined to y and this triple (x, y, z) is a (3,1)-graph. For a given (3,1) graph we can find only $O(n^{d-1})$ W from which it can be obtained in the way given above. Hence

$$\sum_W B_W = O(kn)O(n^{d-1}) = O(kn^d). \tag{18}$$

Let $Q_W =: \sum_{V\in W} q_V^2$. (Here V is a vertex!) Trivially,

$$\sum_W \cdot Q_W = O(kn)\cdot O(n^{d-1}) = O(kn^d). \tag{19}$$

By (17), (18) and (19)

$$\sum_W (nA_W + B_W + Q_W) = O(kn^d).$$

By Theorem 1 applied with $p=d=\lfloor t\rfloor$ we know that the number of summands on the left, $k_d(G) \geqq c_1 n^d$ for some positive constant c_1. Therefore the average of $(nA_W + B_W + Q_W)$ is $O(k)$. Thus there exists a W in G for which

$$A_W = O(k/n), \quad B_W = O(k), \quad \text{and} \quad q_V = O(\sqrt{k}) \quad \text{if} \quad V\in W. \tag{20}$$

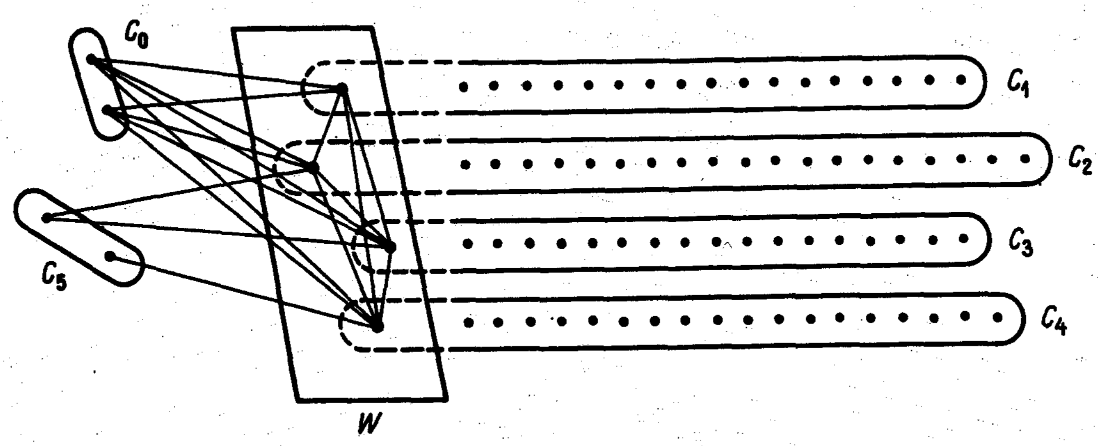

Fig. 2

Let C_i ($i=1, \ldots, d$) be the set of vertices of G joined to all the vertices of W but to the ith one denoted by V_i. Let C_0 be the set of vertices joined to W completely and C_{d+1} be the set of vertices joined to at most $d-2$ vertices of W. By (20), $|C_{d+1}|=A_W=$ $=O\left(\frac{k}{n}\right)=O(\sqrt{k})$, and for every $V_j \in W$ $\sigma(V_j)=r_{V_j}=\frac{2E}{n}-q_{V_i}=\left(1-\frac{1}{t}\right)n+O(\sqrt{k})$. Thus, for $j=1, 2, \ldots, d$,

$$|C_i|=|\bigcap_{j\neq i} st(V_j)| \geqq \frac{n}{d}+O(\sqrt{k})$$

and therefore (by $\sum|C_i|\leqq n$)

$$|C_i|=\frac{n}{d}+O(\sqrt{k}), \quad \text{and} \quad |C_0|=O(\sqrt{k}).$$

A short computation gives that if $n_i=n/d+O(\sqrt{k})$, then $e(K_d(n_1, \ldots, n_d))=$ $=m(n, d+1)+O(k)$. Let us consider the following classification of the vertices of G: C_i is the ith class for $i=2, 3, \ldots, d$ and $C_0 \cup C_1 \cup C_{d+1}$ is the first one, n_i is the number of vertices in the ith class, $i=1, 2, \ldots, d$.

By (20), more precisely, by $B_W=O(k)$ and $|C_{d+1}|=O(k/n)$, the number of pairs (x, y) not belonging to G where x and y belong to different classes is only $O(k)+$ $+O(k/n)O(n)=O(k)$. Since

$$e(K_d(n_1, \ldots, n_d))=m(n, d+1)+O(k),$$

(i.e. it is not too small!), by (3) the number of edges of G the end vertices of which belong to the same class is at most

$$E-(e(K_d(n_1, \ldots, n_d))-B_W-n|C_{d+1}|)=O(k).$$

This completes the proof.

5. Proof of Theorem 3

The proof is rather long and subdivided into steps (A)–(U). Occasionally we shall insert some remarks telling our plans for the next few steps. In steps (A) and (B) we approximate the extremal graphs with complete d-partite graphs and introduce some notation. In (C) we show that if $K_d(n_1, \ldots, n_d)$ is the graph approximating our extremal graph S, then $n_i - n_j$ is small.

All the inequalities below are stated only for the sufficiently large values of n.

(A) Let S be an extremal graph for Problem 1 for some n, E, and let

$$d = \max\{t : m(n, t+1) \leqq E\},$$

while $k = E - m(n, d+1)$.

It is clear that we may assume that $k = o(n^2)$. Indeed, if the theorem is true for all possible functions $k = k(n)$ such that $k = o(n)$ then there exists an $\varepsilon > 0$ such that the theorem is true for $k < \varepsilon n^2$ (p and d are fixed throughout).

(B) We can apply Theorem 2 to S. Let Z be a graph obtained from Turán's graph $T^{n,d}$ by adding k edges to it. Then $e(Z) = E$ and so by the extremality of S we have

$$k_p(S) \leqq k_p(Z) = \binom{d}{p}\left(\frac{n}{d}\right)^p + O(kn^{p-2}).$$

Thus Theorem 2 applies and we conclude that there is a constant c_1 such that S can be obtained from a $K_d(n_1, \ldots, n_d)$ by deleting and adding at most $c_0 k$ edges. The construction of S this way is not unique. Let us choose the graph $K_d(n_1, \ldots, n_d)$ in such a way that the number of edges to add is minimal. Let $A_1, \ldots, A_d$ denote the classes of $K_d(n_1, \ldots, n_d)$, $|A_i| = n_i$. Call the edges to be added to $K_d(n_1, \ldots, n_d)$ *horizontal* edges; the edges to be deleted from $K_d(n_1, \ldots, n_d)$ *missing* edges; the edges which occur in both S and $K_d(n_1, \ldots, n_d)$ *vertical edges.*

Let h and m denote the number of horizontal and missing edges, respectively. Clearly, $h \leqq c_0 k$ and $m \leqq h \leqq c_0 k$. Moreover, $m \leqq h - k$:

$$k = E - m(n, d+1) = \{e(K_d(n_1, \ldots, n_d)) + h - m\} - m(n, d+1) \leqq h - m.$$

Set

$$\sigma_i^+(x) = |A_i \cap st\, x|$$

$$\sigma_i^-(x) = |A_i - st\, x|$$

If $x \in A_j$ then let

$$\sigma^+(x) = \sigma_j^+(x)$$

and

$$\sigma^-(x) = \sum_{i \neq j} \sigma_i^-(x).$$

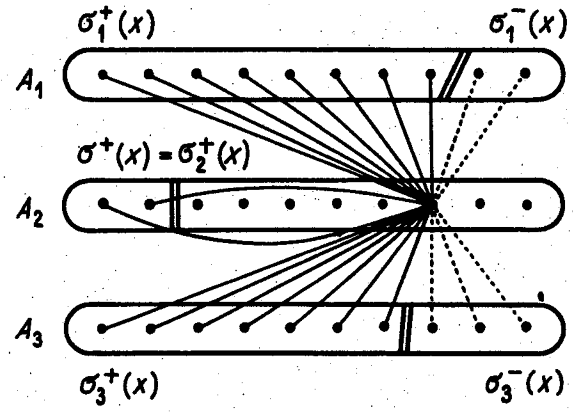

Fig. 3

Thus $\sigma^+(x)$ and $\sigma^-(x)$ denote the numbers of horizontal and missing edges adjacent to x, respectively.

Finally, set

$$\sigma^+ = \max_x \sigma^+(x),$$

$$\sigma^- = \max_x \sigma^-(x).$$

Note that the choice of the partition $\{A_1, \ldots, A_d\}$ implies that

(21) $$\sigma_i^+(x) \geqq \sigma^+(x) \quad (i=1, \ldots, d).$$

Hence

$$\sigma^+(x) < \frac{n}{d}$$

for all x, so

(22) $$\sigma^+ < \frac{n}{d}.$$

Introduce the numbers

$$R_1 = \binom{d-1}{p-2}\left(\frac{n}{d}\right)^{p-2}, \quad R_i = \binom{d-2}{p-i}\left(\frac{n}{d}\right)^{p-1} \quad (i \geqq 2).$$

These will occur frequently in various approximations.

Let

$$S_d^p(n_1, \ldots, n_d) = \sum_{i_1 \leqq \ldots \leqq i_p} \prod_{j=1}^{p} n_{i_j}.$$

If $n_1, \ldots, n_d$ are integers, clearly,

$$S_d^p(n_1, \ldots, n_d) = k_p(K_d(n_1, \ldots, n_d)).$$

(C) We show that $n_i = \frac{n}{d} + O(\sqrt{k})$. For let e.g. $n_1 = \max(n_1, \ldots, n_d)$, $n_2 = \min(n_1, \ldots, n_d)$. Then

$$e(S) \leqq e(K_d(n_1, \ldots, n_d)) + c_1 k = S_d^2(n_1, \ldots, n_d) + c_1 k =$$

$$= S_d^2\left(\frac{n_1+n_2}{2}, \frac{n_1+n_2}{2}, n_3, \ldots, n_d\right) + c_1 k - \frac{1}{4}(n_1-n_2)^2 \leqq$$

$$\leqq S_d^2\left(\frac{n}{d}, \ldots, \frac{n}{d}\right) + c_1 k - \frac{1}{4}(n_1-n_2)^2 .$$

On the other hand,

$$e(S) = S_d^2\left(\frac{n}{d}, \ldots, \frac{n}{d}\right) + k + O(1),$$

which yields that $n_1 - n_2 = O(\sqrt{k})$.

(D) Let $u, v \in V(G)$. We denote by $a_S(u, v) = a(u, v)$ the number of K_p's in $S + (u, v)$ containing the edge (u, v). We can obtain quite accurate estimations on these numbers.

... The first part of the proof consists of steps (A)–(M). In steps (D)–(M) we obtain step by step more and more information, sharper and sharper inequalities for quantities like

(i) $a(x, y)$, when (x, y) is an edge, in particular, a horizontal one
(ii) $a(u, v)$, where (u, v) is a missing edge
(iii) $\sigma^+ = \max \sigma^+(x)$
(iv) $t = t(x) =: \min(\sigma^+(x), \sigma^-(x))$
(v) $\sigma^+(x) + \sigma^+(y)$ for the edges (x, y) and for the missing edges (x, y)...

Let first (u, v) be a horizontal edge. Then

$$a(u, v) \geqq R_1 - [\sigma^-(u) + \sigma^-(v)] R_3 + O(\sqrt{k} \cdot n^{p-3}) . \tag{23}$$

Indeed, let us count the K_p's containing (u, v), as follows. Let e.g. $u, v \in A_1$ and $\ddot{S}$ denote the graph obtained from S by filling in all the missing edges. The number of K_p's in $\ddot{S}$ containing (u, v) but no other horizontal edge is

$$S_{d-1}^{p-2}(n_2, \ldots, n_d) = S_{d-1}^{p-2}\left(\frac{n}{d}, \ldots, \frac{n}{d}\right) + O(\sqrt{k}\, n^{p-3}) = R_1 + O(\sqrt{k}\, n^{p-3}) .$$

Let us delete now the missing edges which we have filled in. A missing edge disjoint from (u, v) destroys at most $O(n^{p-4})$ K_p's and since there are only $O(k)$ missing edges, this way we destroy only $O(k) \cdot O(n^{p-4}) < O(\sqrt{k}\, n^{p-3})$ K_p's. If we delete now a missing edge incident with u or v, say one connecting u to a point $w \in A_2$, then we destroy at most

$$S_{d-2}^{p-3}(n_3, \ldots, n_d) = R_3 + O(\sqrt{k}\, n^{p-4})$$

K_p's counted above. So deleting all such missing edges we destroy at most

$$(\sigma^-(u)+\sigma^-(v))\cdot R_3 + (\sigma^-(u)+\sigma^-(v))O(\sqrt{k}\, n^{p-4})) =$$

$$= (\sigma^-(u)+\sigma^-(v))\cdot R_3 + O(\sqrt{k}\, n^{p-3})$$

K_p's counted above. This proves (23).

Similar computation yields that if (u, v) is a missing edge then

$$a(u, v) \leqq R_2 + [\sigma^+(u)+\sigma^+(v)]R_3 + \sigma^+(u)\sigma^+(v)R_4 + O(\sqrt{k}\, n^{p-3}). \tag{24}$$

(E) The extremality of S implies that if $(x, y) \in E(S)$ but $(u, v) \notin E(S)$ then

$$a(x, y) \leqq a(u, v). \tag{25}$$

Indeed filling in (u, v) creates $a(u, v)$ K_p's, and then deleting (x, y) destroys *at least* $a(x, y)$ of them: filling in (u, v) may create K_p's containing (x, y), this is why the deletion of (x, y) may destroy more than $a(x, y)$ K_p's. By the extremality of S

$$k_p(S) \leqq k_p(S + (u, v) - (x, y)) \leqq k_p(S) + a(u, v) - a(x, y),$$

proving (25).

Now (25) will be applied in the following way: knowing more and more about the structure of the graph we shall be able to obtain always better and better bounds on $a(x, y)$ and $a(u, v)$; then (25) in turn gives more information on the graph. Another inequality, similar to (24) and (25) is that

$$a(x, y) \leqq R_1 \tag{26}$$

if $(x, y) \in E(S)$. For using induction on k, we know that

$$k_p(S) = k_p(S - (x, y)) + a(x, y) \geqq k_p(G) + a(x, y)$$

for some $G \in U_1(n, E-1)$. Let $G' \in U_1(n, E)$ be obtained from the same $K_d(n_1, \ldots, n_d)$ as G. Let $n_1 \geqq n_i$. Then

$$k_p(G') = k_p(G) + S_{d-1}^{p-2}(n_2, \ldots, n_d) \leqq k_p(G) + R_1$$

and hence

$$k_p(S) \geqq k_p(G) + a(x, y) \geqq k_p(G') + a(x, y) - R_1.$$

(F) Let (x, y) be a horizontal edge and (u, v) a missing edge. Then (23), (24) and (25) imply that

$$R_1 - R_2 \leqq [\sigma^-(x)+\sigma^-(y)+\sigma^+(u)+\sigma^+(v)] \cdot R_3 + \\ + \sigma^+(u)\sigma^+(v) \cdot R_4 + O(\sqrt{k}\, n^{p-3})$$

or, dividing by R_3,

$$\frac{n}{d} \leqq \sigma^-(x)+\sigma^-(y)+\sigma^+(u)+\sigma^+(v)+ \\ + \frac{d}{n}\frac{p-3}{d-p+2}\sigma^+(u)\cdot\sigma^+(v)+O(\sqrt{k}). \tag{27}$$

(G) The previous important inequality is used first to bound the number σ^+ from below. Using that

$$\sigma^+(u), \sigma^+(v) \leqq \sigma^+ < \frac{n}{d},$$

we obtain for each horizontal edge (x, y) that

$$\frac{n}{d} \leqq \sigma^-(x)+\sigma^-(y)+2\sigma^+ + \frac{p-3}{d-p+2}\sigma^+ + O(\sqrt{k}).$$

Summing for all horizontal edges (x, y), we get

$$h \cdot \frac{n}{d} \leqq \sum_x \sigma^+(x)\sigma^-(x) + h\sigma^+ \cdot \frac{2d-p+1}{d-p+2} + O(\sqrt{k}\, h) \leqq \\ \leqq \sigma^+ \sum_x \sigma^-(x) + h\sigma^+ \frac{2d-p+1}{d-p+1} + O(\sqrt{k}\, h) \leqq \\ \leqq \sigma^+ h\left(2+\frac{2d-p+1}{d-p+2}\right) + O(\sqrt{k}\, h),$$

since $\sum_x \sigma^-(x) = 2m \leqq 2h$. Thus

$$\sigma^+ \geqq \frac{d-p+2}{4d-3p+5}\frac{n}{d} + O(\sqrt{k}) \geqq \frac{n}{4d^2} + O(\sqrt{k}). \tag{28}$$

(H) Our next aim is to show that for every x, one of $\sigma^+(x)$, $\sigma^-(x)$ must be small. More precisely, let

$$t = t_x = \min(\sigma^+(x), \sigma^-(x)).$$

We want to show that

(29) $$t=o(n).$$

Set $\sigma_j^+(x)=\sigma_j$. By the choice of the partition, more precisely, by (21),

$$\sigma_j \geqq \sigma^+(x) \geqq t \quad (j=1, \ldots, d).$$

The number of K_p's containing x is

$$k_{p-1}(K_d(\sigma_1, \ldots, \sigma_d))+O(kn^{p-3})=S_d^{p-1}(\sigma_1, \ldots, \sigma_d)+O(kn^{p-3})$$

where the second term accounts for the K_p's containing a horizontal edge not adjacent to x and also for those p-tuples consisting of x and $p-1$ neighbors of it which span a missing edge (since $h, m \leqq c_1 k$, see (B)).

Suppose that e.g. $x \in A_1$. One of the numbers $\sigma_2^-(x), \ldots, \sigma_d^-(x)$, say $\sigma_2^-(x)$, is at least t/d.

Replace $s=\lfloor t/d \rfloor$ edges connecting x to A_1 by $\lfloor t/d \rfloor$ edges connecting x to A_2. Then the K_p's not containing x remain the same while the number of K_p's containing x becomes

$$k_{p-1}(K_d(\sigma_1-s, \sigma_2+s, \sigma_3, \ldots, \sigma_d)+O(k\, n^{p-3})=$$
$$=S_d^{p-1}(\sigma_1-s, \sigma_2+s, \sigma_3, \ldots, \sigma_d)+O(k\, n^{p-3}).$$

The number of K_p's cannot decrease by this operation, hence

$$S_d^{p-1}(\sigma_1, \ldots, \sigma_d)-S_d^{p-1}(\sigma_1-s, \sigma_2+s, \sigma_3, \ldots, \sigma_d) \leqq O(k\, n^{p-3}).$$

But the left hand side is

(30) $$s(\sigma_2-\sigma_1+s)S_{d-2}^{p-3}(\sigma_3, \ldots, \sigma_d)>s^2 t^{p-3}>\frac{1}{2d^2}\, t^{p-1}$$

whence

$$t=O(k^{\frac{1}{p-1}} \cdot n^{\frac{p-3}{p-1}})=o(n).$$

(I) Let $x_0 \in A_i$ be a point with

$$\sigma^+(x_0)=\sigma^+.$$

Then by (28) and (29),

$$\sigma^-(x_0)=o(n).$$

Clearly x_0 has a neighbor $y_0 \in A_i$ with

$$\sigma^-(y_0) \leqq m/\sigma^+=O(k/n)=o(n).$$

Hence, by (23),

$$a(x_0, y_0) \geqq R_1 + o(n^{p-2}).$$

So, by (25), for every pair $(u, v) \notin E(G)$ we have

(31) $$a(u, v) \geqq R_1 + o(n^{p-2}).$$

Applying (27) to the horizontal edge (x_0, y_0) and any missing edge (u, v) we obtain that

(32) $$\sigma^+(u) + \sigma^+(v) + \frac{d}{n}\frac{p-3}{d-p+2}\sigma^+(u)\cdot\sigma^+(v) \geqq \frac{n}{d} + o(n).$$

(J) Now we can easily show that $\sigma^- = O(\sqrt{k})$. First we prove the weaker

(33) $$\sigma^- = o(n).$$

Indeed, let v be a point with

$$\sigma^-(v) = \sigma^-.$$

For $c = \dfrac{1}{4d(p-3)}$ either $\sigma^+(v) \geqq cn$, and therefore (33) follows from (29), or $\sigma^+(v) \leqq cn$. In the second case for every missing edge (u, v) $\left(\text{by } \sigma^+(u) \leqq \dfrac{n}{d}\right)$

$$\frac{d}{n}\frac{p-3}{d-p+2}\cdot\sigma^+(u)\cdot\sigma^+(v) \geqq \frac{n}{4d}.$$

By (32)

$$\sigma^+(u) \geqq \frac{n}{d} + o(n) - \frac{n}{4d} - cn \geqq \frac{n}{6d}.$$

Therefore the number of such points u (i.e. $\sigma^-(v)$) is at most

$$h\Big/\frac{n}{6d} = O\left(\frac{k}{n}\right) = o(\sqrt{k}).$$

Now we improve (33). It implies that in (30) (in (H)) $\sigma_j = \sigma_j^+(x) = \dfrac{n}{d} + o(n)$, hence s^2t^{p-3} can be replaced by $s^2\cdot\left(\dfrac{n}{2d}\right)^{p-3}$. Hence in (H) we can improve $t = o(n)$ to $t = O(\sqrt{k})$, in (29); thus $\sigma^-(x_0) = O(\sqrt{k})$, which, in turn, yields that

(34) $$\sigma^- = O(\sqrt{k}).$$

An important consequence of (34) is that for any vertex $x \in V(S)$

(35) $$\sigma(x)=\left(1-\frac{1}{d}\right)n+\sigma^{+}(x)+O(\sqrt{k}).$$

$\left(\text{Here we use } n_i=\frac{n}{d}+O(\sqrt{k}), \text{ too.}\right)$ Another consequence is that if $u \in A_i$ $v \in A_j$ and $i \neq j$, then

(36) $$a(u, v)=R_2+[\sigma^{+}(u)+\sigma^{+}(v)]R_3+\sigma^{+}(u)\cdot\sigma^{+}(v)\cdot R_4+O(\sqrt{k}\,n^{p-3}).$$

Indeed, if we fill in all the missing edges adjacent to u or v, by (34), we create only $O(\sqrt{k}\,n^{p-3})$ K_p's containing (u, v). In the resulting graph an argument, similar to the proof of (23) works.

(K) Let $(x, y)\in E(S)$ (where (x, y) may be a horizontal or a vertical edge). We claim that

(37) $$\sigma^{+}(x)+\sigma^{+}(y)\leqq\frac{n}{d}+O(\sqrt{k}).$$

By (35), an equivalent form of (37), independent of the partition is

(37*) $$\sigma(x)+\sigma(y)\leqq\left(2-\frac{1}{d}\right)\cdot n+O(\sqrt{k}).$$

For let us assume first that x, y are in different classes. Then, by (26) and (36)

$$R_1\geqq a(x, y)\geqq R_2+[\sigma^{+}(x)+\sigma^{+}(y)]R_3+O(\sqrt{k}\,n^{p-3}),$$

proving (37). (Here we use that $R_1-R_2=R_3\cdot\frac{n}{d}$.) If $x, y \in A_1$, (say) then they have at least $\sigma^{+}(x)+\sigma^{+}(y)-|A_1|$ neighbors in A_1 in common and this yields

$$R_1\geqq a(x, y)\geqq R_1+O(\sqrt{k}\cdot n^{p-3})+(\sigma^{+}(x)+\sigma^{+}(y)-|A_1|)\cdot$$

$$\cdot\left(\binom{d-1}{p-3}\left(\frac{n}{d}\right)^{p-3}+O(\sqrt{kf}\,n^{p-4})\right).$$

This proves (37), for horizontal edges, too.

(L) An important consequence of (35), (36) and (37) *is that there exists a* $c_1>0$ *such that if* $\sigma(x)\geqq\left(1-\frac{1}{2d}\right)n+c_1\sqrt{k}$, *then the neighbors of* x *span no missing edge.* For suppose that (u, v) is a missing edge whose endpoints are adjacent to x. Let

$$\sigma(x)=\left(1-\frac{1}{2d}\right)n+r, \quad r>0.$$

Let $x \in A_i$ and $u \notin A_i$. By (25) and (36)

$$0 \leqq a(u, v)-a(u, x)=[\sigma^+(v)-\sigma^+(x)]R_3+$$

$$+\sigma^+(u)[\sigma^+(v)-\sigma^+(x)]R_4+O(\sqrt{k}\cdot n^{p-3})=$$

$$=[\sigma^+(v)-\sigma^+(x)+O(\sqrt{k})]\,[R_3+\sigma^+(u)R_4].$$

Therefore

$$\sigma^+(v) \geqq \sigma^+(x)+O(\sqrt{k}).$$

This and (35) yield that

$$\sigma^+(v)+\sigma^+(x) \geqq \frac{n}{d}+2r+O(\sqrt{k}).$$

Since $(u, v) \in E(S)$, by (37), applied with $y=v$,

$$r=O(\sqrt{k}).$$

(M) Let us fix a $c_2>c_1$. Set

$$V=\left\{x \in V(G)\colon \sigma^+(x)>\frac{n}{2d}+c_2\sqrt{k}\right\},$$

$$B_i=A_i-V,$$

$$b_i=|B_i|.$$

Let, further, h_i denote the number of horizontal edges spanned by B_i and m_{ij} the number of missing edges between B_i and B_j.

Note that if (u, v) is a missing edge, $u \in B_i$, $v \in B_j$ then, by $u \notin V$,

$$\sigma^+(u) \leqq \frac{n}{2d}+c_2\sqrt{k}$$

and hence (32) implies that there is a constant $c_3>0$ such that

$$\sigma^+(v)>c_3 n.$$

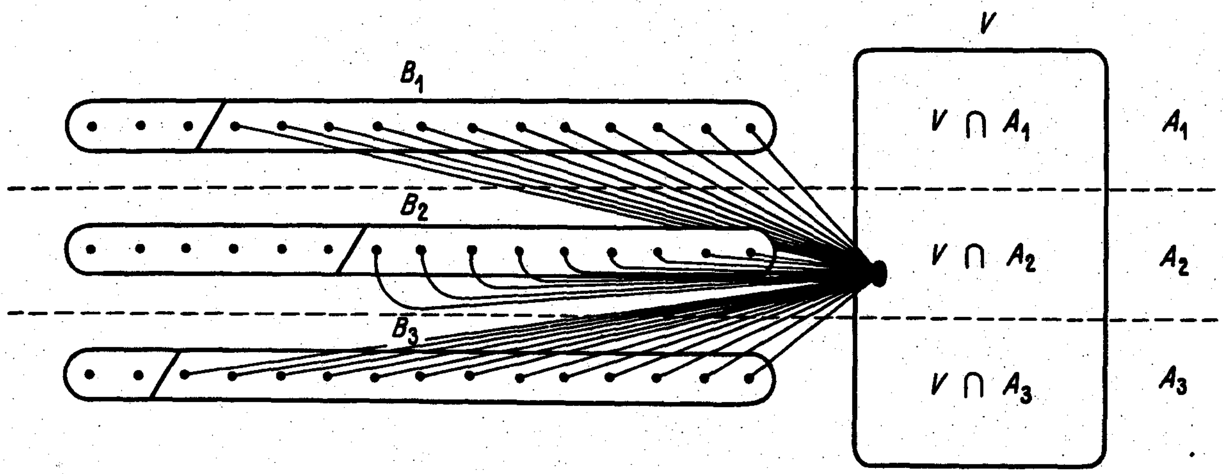

Fig. 4

Hence there are at most $h_j/c_3n = O(h_j/n)$ vertices in B_j incident with missing edges of $S-V$. This in particular implies that

(38) $$m_{ij} = O\left(\frac{h_ih_j}{n^2}\right).$$

(N) *We shall carry out now a number of transformations which finally lead to a graph Q with $v(Q)=v(S)$, $e(Q)=e(S)$ and $k_p(Q)<k_p(S)$ unless S is of a very simple structure. (By the extremality of S the second one must be the case.)*

(i) First construct $S-V=S'$.

(ii) Second, fill in the missing edges in S', to get S''.

(iii) Third, rearrange the horizontal edges in S'' as follows. Let B_i span h_i horizontal edges. Find the least number t_i such that $t_i(|B_i|-t_i)\geqq h_i$. Clearly, $t_i = O\left(\frac{h_i}{n}+1\right)=o(n)$. Further, $h_i \leqq t_in$. Let $F_i\subseteq B_i$, $|F_i|=t_i$ and $D_i=B_i-F_i$. Connect t_i-1 points of F_i to all points of D_i, and the remaining point u_i of F_i to $h_i-(t_i-1)(|B_i|-t_i)$ points of D_i. This yields the graph S'''. (See Fig. 5.)

(iv) Delete m_{ij} edges spanned by $B_i\cup B_j$. The precise way of selecting these edges depends on the values of m_{ij}, t_i, t_j, h_i and h_j and will be given below, when these cases will be distinguished. To be able to start the general discussion, first we assume only the following.

Condition (*).
If $v\in B_i$ and $t_i>1$, then we delete at most

$$\left\lceil \frac{m_{ij}}{t_i-1} \right\rceil$$

edges (v, w), $w\in B_j$.

The resulting graph is S^{IV}. (See Fig. 5.)

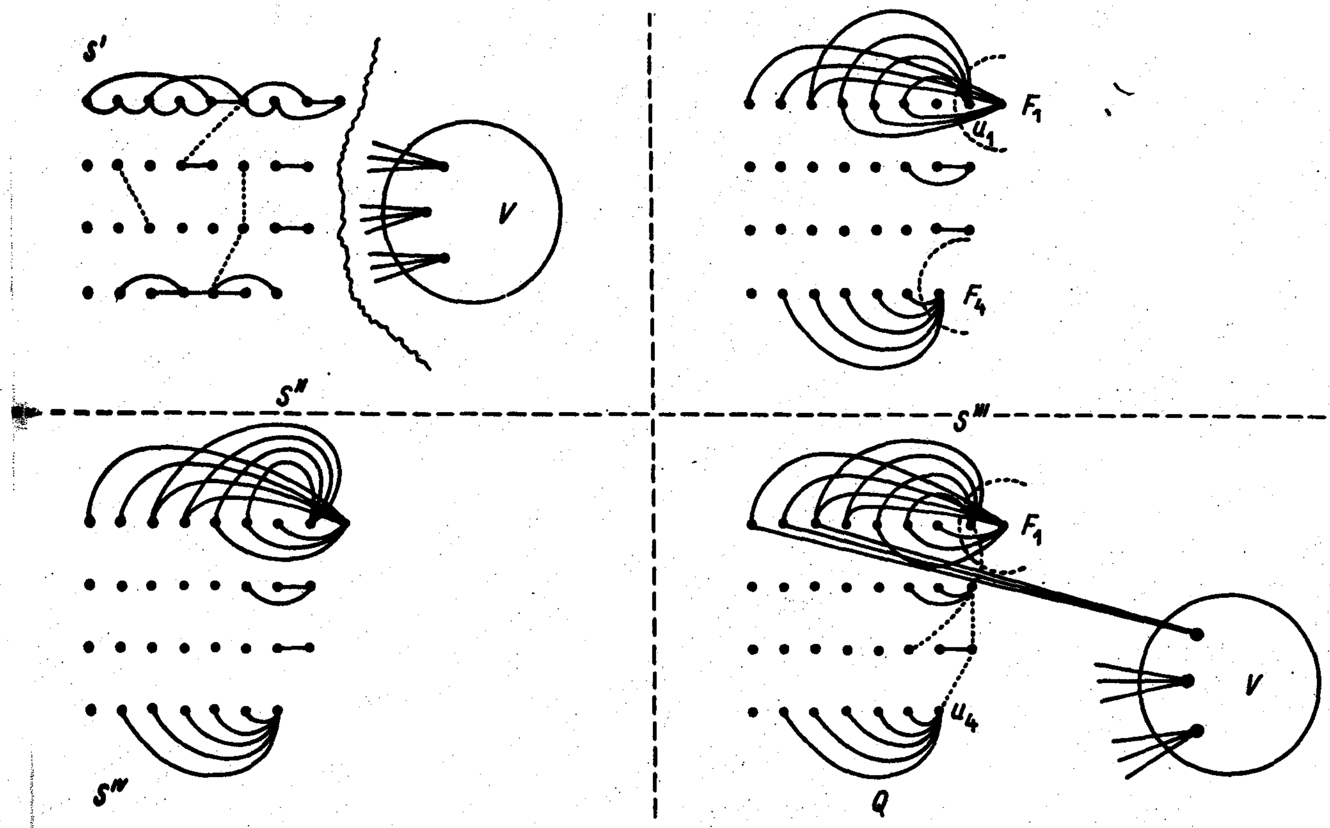

Fig. 5

(v) Connect each $x \in V$ to $\sigma_i^+(x)$ points of B_i which span the fewest edges in S^{IV}. The resulting graph is Q. Clearly $v(Q)=v(S)$ and $e(Q)=e(S)$.

(O) We first analyse the effect of (iii). Call a K_p *regular*, if it contains at most two points of each B_i. Clearly, every K_p in S''' is regular. On the other hand, it is easily seen that *S'' and S''' have the same number of regular K_p's*. Thus $k_p(S'') \geq k_p(S''')$.

Another property of S''' we need is that *for every r $(0 \leq r \leq b_i)$ the minimum number of edges spanned by a set X of r points of B_i is for S''' less than or equal to that of S''*. This is clear for $r \leq b_i - t_i$, since then $X \subseteq B_i$ yielding the minimum is an independent set in S'''. If $r \geq b_i - t_i + 1$, then X spans

$$h_i - (b_i - r)(b_i - t_i) \tag{39}$$

edges of S''': we take all points of B_i but $b_i - r$ ones from $F_i - u_i$. If $|X| = r$, $X \subseteq B_i$, then X spans at least

$$h_i - (b_i - r)\left(\frac{n}{2d} + c_2\sqrt{k}\right)$$

edges of S'', since $B_i - X$ represents at most $|B_i - X|\left(\frac{n}{2d} + c_2\sqrt{k}\right)$ edges. By $b_i - t_i = \frac{n}{d} + o(n)$ the minimum is smaller for S'''.

(P) We use the previous considerations to show that

$$k_p(Q)-k_p(S^{IV})\leqq k_p(S)-k_p(S'). \tag{40}$$

The left-hand side is the number of K_p's in Q containing any point in V. The right-hand side is the number of such K_p's in S. (Thus the meaning of (40) is that the transformations do not increase the number of K_p's meeting V.) It suffices to prove that each $x\in V$ is contained in no more K_p's of Q than of S. Let $x\in V$ and set $X=st_S x$. Without loss of generality we may assume that $st_Q x=X$ (this only means relabelling of the points). Let S_X and Q_X be the subgraphs of Q and S respectively, induced by X. What we want to show is that

$$k_{p-1}(S_X)\geqq k_{p-1}(Q_X). \tag{41}$$

Set $C_i=B_i\cap X$, and let γ_i and δ_i denote the numbers of horizontal edges induced by C_i in S and Q, respectively. Let us compare the numbers of K_{p-1}'s in S and Q, containing one horizontal edge from, say, each of $C_1, \dots, C_v$ and no other horizontal edge.

By (L) and the definition of V, X spans no missing edges in S. Thus, in S, the number of these K_{p-1}'s is exactly

$$\gamma_1, \dots, \gamma_v\, S_{d-v}^{p-2v-1}(|C_{v+1}|, \dots, |C_d|). \tag{42}$$

The corrresponding number in Q is at most

$$\delta_1 \dots \delta_v\, S_{d-v}^{p-2v-1}(|C_{v+1}|, \dots, |C_d|). \tag{43}$$

Since $\delta_i\leqq\gamma_i$ by (O) and the construction, and furthermore, every K_{p-1} in Q_X, being regular, is taken into consideration in the terms (43), the inequality (41) follows. (S may contain K_{p-1}'s not counted in the terms (42), namely those containing three or more points of a C_i.)

(Q) The previous section and the extremality of S imply that

$$k_p(S^{IV})\geqq k_p(S'). \tag{44}$$

Since every K_p in S^{IV} is regular, the number of regular K_p's in S^{IV} is at least as large as the number of regular K_p's in S'. Since step (iii) did not change the number of regular K_p's, it follows that the number of regular K_p's created in step (ii) is at least as large as the number of regular K_p's destroyed in step (iv).

Let Φ^{ij} denote the number of regular K_p's created when the missing edges between B_i and B_j are filled in; let Ψ^{ij} denote the number of regular K_p's destroyed when the m_{ij} edges corresponding to the missing edges between B_i and B_j are deleted in step (iv). Note that Φ^{ij} and Ψ^{ij} depend on the order in which the edges are filled in and deleted, so such an order must be fixed. However, this order will have no importance.

The following (unfortunately, rather tedious) analysis will show that *there is a* $c^*>0$ *such that*

(45) $$\Psi^{ij} \geqq \Phi^{ij} - c^* \cdot \sqrt{k}\, n^{p-3}$$

and the stronger inequality

(46) $$\Psi^{ij} \geqq \Phi^{ij} + d^2 \cdot c^* \sqrt{k}\, n^{p-3}.$$

holds, unless either $m_{ij}=0$ *or* $t_i=t_j=m_{ij}=1$.

The assertion above that "the number of regular K_p's created in step (ii) is at least as large as "the number of K_p's destroyed in step (iv)" means that

$$\sum_{i,j} \Psi^{ij} \leqq \sum_{i,j} \Phi^{ij}.$$

Therefore, by (45) and (46),

$$m_{ij}=0 \quad \text{or} \quad t_i=t_j=m_{ij}=1$$

for every i and j.

So let $i \neq j$ be given such that $m_{ij} \neq 0$ (if $m_{ij}=0$ we have nothing to do). Let us call a regular K_p to be of type ($\mu=0, 1, 2$) if it meets both B_i and B_j and contains μ horizontal edges in $B_i \cup B_j$. Let Φ_μ denote the type μ K_p's created in step (ii) and let Ψ_μ denote the type μ K_p's destroyed in step (iv).

Below *we shall first establish some upper bounds on* Φ_μ *and (lower) bounds on* Ψ_μ. Then, using some case distinction, we shall specify, how to delete the m_{ij} edges in step (iv) of (N) and show that in each case

$$\Psi^{ij} - \Phi^{ij} = (\Psi_0 + \Psi_1 + \Psi_2) - (\Phi_0 + \Phi_1 + \Phi_2)$$

is "too large", proving (46) or (45). What is an annoying but natural feature of our case distinction that *we shall have the most trouble with the cases, when* t_i *and* t_j *are very small* (1 *or* 2!).

When an edge between B_i and B_j is filled in, the number of type 0 K_p's created is at most

$$R_2 + O(\sqrt{k} \cdot n^{p-3}).$$

The corresponding numbers of type 1 and type 2 K_p's are

$$[\sigma_S^+(u) + \sigma_S^+(v)] R_3 + O(\sqrt{k} \cdot n^{p-3}) \leqq \frac{n}{d} R_3 + O(\sqrt{k} \cdot n^{p-3})$$

and

$$\sigma^+(u)\sigma^+(v) R_4 + O(\sqrt{k} \cdot n^{p-3}) \leqq \frac{n^2}{4d^2} R_4 + O(\sqrt{k} \cdot n^{p-3}),$$

(we have used the definition of V). So

(47) $$\Phi_0 \leqq m_{ij} R_2 + O(\sqrt{k} \cdot n^{p-3} m_{ij})$$

(48) $$\Phi_1 \leqq m_{ij} \frac{n}{d} R_3 + O(\sqrt{k} \cdot n^{p-3} m_{ij})$$

(49) $$\Phi_2 \leqq m_{ij} \frac{n^2}{4d} R_4 + O(\sqrt{k} \cdot n^{p-3} m_{ij}).$$

On the other hand, the numbers of K_p's of types 0, 1 and 2 destroyed by deleting an edge (u, v) in step (iv) are

(50) $$R_2 + O(\sqrt{k} \cdot n^{p-3})$$

(51) $$[\sigma^+_{S'''}(u) + \sigma^+_{S'''}(v)] R_3 + O(\sqrt{k} \cdot n^{p-3})$$

and

(52) $$\sigma^+_{S'''}(u) \cdot \sigma^+_{S'''}(v) R_4 + O(\sqrt{k} \cdot n^{p-3}),$$

respectively. This would be trivial if we counted the K_p's in S''' containing (u, v). However, we fixed an order of deleting the edges between the classes B_i and B_j. More precisely, we fixed an order on the pairs (i, j), and if (i^*, j^*) preceeds (i, j), then we should not count here the K_p's containing (u, v) but at the same time containing a (u^*, v^*), $u^* \in B_{i^*}$, $u^* \in B_{j^*}$, already deleted. The number of such K_p's is $O(m \cdot n^{p-4}) = O(k \cdot n^{p-4})$, for the edges (u^*, v^*) disjoint from (u, v). Let us estimate the number of those destroyed by the removal of an edge (u, w). If $t_i = 1$, i.e. $F_i = \{u\}$ then $h_i < b_i$ and so, by (38),

$$m_{il} = O\left(\frac{h_i h_l}{n^2}\right) = O\left(\frac{n \cdot k}{n^2}\right) = O\left(\frac{k}{n}\right)$$

for every l. Thus only $O\left(\frac{k}{n}\right)$ edges adjacent to u are removed at most and so the number of K_p's containing (u, v) and an edge adjacent to u and removed previously is at most $O\left(\frac{k}{n} \cdot n^{p-3}\right) = O(k \cdot n^{p-4})$. If $t_i \geqq 2$ then, by Condition (*) of (N)/(iv), the number of edges adjacent to u and removed previously (by (38) and $h_l \leqq t_l \cdot n$) is at most

$$\sum_l O\left(\frac{m_{ij}}{t_l}\right) \leqq \sum_l \left(\frac{h_i h_l}{t_l n^2}\right) = O\left(\frac{h_i}{n}\right) = O\left(\frac{k}{n}\right),$$

and we conclude as before. Thus (50), (51) and (52) are proved.

Now we need some case distinction. In the cases below we can always satisfy Condition (*) of step (iv) in (N).

Case (Q1). $m_{ij} \leqq (t_i - 1)(t_j - 1)$. Then we can remove m_{ij} edges connecting $F_i - u_i$ to $F_j - u_j$. For such an edge (u, v)

$$\sigma^+(u), \sigma^+(v) = \frac{n}{d} + O(\sqrt{k}),$$

whence by (50), (51) and (52),

$$\Psi_0 \geqq m_{ij} R_2 + O(m_{ij}\sqrt{k} \cdot n^{p-3})$$

$$\Psi_1 \geqq m_{ij} \frac{2n}{d} R_3 + O(m_{ij}\sqrt{k} \cdot n^{p-3})$$

$$\Psi_2 \geqq m_{ij} \frac{n^2}{d^2} R_4 + O(m_{ij}\sqrt{k} \cdot n^{p-3}).$$

Comparing with (47), (48) and (49) it follows that

$$\Psi^{ij} - \Phi^{ij} \geqq c_4 n^{p-2},$$

proving (46) and therefore (45), too.

Case (Q2). $(t_i - 1)(t_j - 1) < m_{ij} \leqq 4(t_i - 1)(t_j - 1)$. Then we can remove all edges between $F_i - u_i$ and $F_j - u_j$ and $m_{ij} - (t_i - 1)(t_j - 1)$ edges between $F_i - u_i$ and $B_j - F_j$. For the first $(t_i - 1)(t_j - 1)$ edges

$$\sigma^+(u), \sigma^+(v) \geqq \frac{n}{d} + O(\sqrt{k}),$$

for the rest still

$$\sigma^+(u) \geqq \frac{n}{d} + O(\sqrt{k}).$$

Hence, as before,

$$\Psi_0 - \Phi_0 \geqq O(\sqrt{k} \cdot n^{p-3} m_{ij}),$$

$$\Psi_1 - \Phi_1 \geqq (t_i - 1)(t_j - 1) \frac{n}{d} R_3 + O(\sqrt{k} \cdot n^{p-3} m_{ij}),$$

and

$$\Psi_2 - \Phi_2 \geqq \frac{4(t_i - 1)(t_j - 1) - m_{ij}}{4} \frac{n^2}{d^2} R_4 + O(\sqrt{k} \cdot n^{p-3} m_{ij}) \geqq O(\sqrt{k} \cdot n^{p-3} m_{ij}).$$

By $m_{ij} \leqq 4(t_i-1)(t_j-1)$ we have

$$\Psi^{ij}-\Phi^{ij} \geqq (t_i-1)(t_j-1)\frac{n}{d}R_3+O(\sqrt{k}\cdot n^{p-3}m_{ij})=$$
$$=(t_i-1)(t_j-1)\left[\frac{n}{d}R_3+O(\sqrt{k}\cdot n^{p-3})\right] \geqq c_5 n^{p-2},$$

proving (46) (and also (45)).

Case (Q3). $m_{ij} \geqq t_i t_j$, $t_i \geqq 2$, $t_j \geqq 3$. In this case remove the m_{ij} edges so that all edges between F_i and F_j are removed. Then no type 2 K_p's remain and hence

$$\Psi_2-\Phi_2 \geqq 0.$$

We have, similarly as before,

$$\Psi_0-\Phi_0 \geqq O(\sqrt{k}\cdot n^{p-3}m_{ij}),$$

and since $(t_i-1)(t_j-1)$ of the removed edges satisfy $\sigma^+_{S''}(u), \sigma^+_{S''}(v) \geqq n/d+O(\sqrt{k})$, and all but at most one of the rest has at least one endpoint u with $\sigma^+_{S''}(u) \geqq \frac{n}{d}+O(\sqrt{k})$, we have

$$\Psi_1-\Phi_1 \geqq ((t_i-1)(t_j-1)-1)\frac{n}{d}R_3+O(\sqrt{k}\cdot n^{p-3}m_{ij}).$$

By (38),

$$m_{ij}=O\left(\frac{h_i h_j}{n^2}\right) = O(t_i t_j)=O((t_i-1)(t_j-1)-1).$$

We conclude as before.

Case (Q4). $t_i=t_j=2$, $m_{ij} \geqq 4$. By (38), $m_{ij}=O(1)$. First we try the same construction as in case (Q3). As before we have

$$\Psi_2-\Phi_2 \geqq 0,$$
$$\Psi_0-\Phi_0 \geqq O(\sqrt{k}\cdot n^{p-3})$$

and looking also at the edges connecting u_i to F_j-u_j and u_j to F_i-u_i we get, similarly as above,

$$\Psi_1-\Phi_1 \geqq [\sigma^+_{S''}(u_i)+\sigma^+_{S''}(u_j)]\cdot R_3+O(\sqrt{k}\cdot n^{p-3}).$$

Now we are home, unless

$$\sigma^+_{S''}(u_i)+\sigma^+_{S''}(u_j) \leqq \sqrt[4]{k}\cdot n^{p-3+\frac{1}{2}}.$$

(Here and below we shall use $\sqrt[4]{kn^2}$ as a quantity which is $o(n)$ and for which $\sqrt{k}=o(\sqrt[4]{kn^2})$.) In the latter case we modify the rule used to delete the edges in step (iv). We do not delete (u_i, u_j), but delete an edge between F_i-u_i and B_j-F_j, instead. Putting (u_i, u_j) back creates at most

$$[\sigma_{S''}^{+}(u_i)+\sigma_{S''}^{+}(u_j)]R_3+\sigma_{S''}^{+}(u_i)\sigma_{S''}^{+}(u_j)R_4+O(\sqrt{k}\cdot n^{p-3})=o(n^{p-2})$$

K_p's, while deleting the edge between F_i-u_i and B_j-F_j destroys at least

$$\frac{n}{d}R_3+O(\sqrt{k}\cdot n^{p-3})$$

K_p's. Thus

$$\Psi^{ij}\geqq\Phi^{ij}+c_6n^{p-2}$$

proving (46). So we are finished again. The cases treated so far cover all cases with $t_i>1$ and $t_j>1$.

Case (Q5). $t_i=1$, $t_j\geqq 2$, $m_{ij}\leqq t_j-1$. The argument is basically the same as in case (Q1). However, we have to improve (48) and (49). Now $\sigma_S^{+}(u)\leqq\min\left(\frac{n}{2d}+c_2\sqrt{k},\, h_i\right)$ for any $u\in B_i$. Thus for any missing edge (u, v) $(u\in B_i,\ v\in B_j)$

$$\sigma_S^{+}(u)+\sigma_S^{+}(v)\leqq\frac{n}{2d}+\min\left(\frac{n}{2d}, h_i\right)+O(\sqrt{k}).$$

Hence

$$\Phi_1\leqq m_{ij}\left(\frac{n}{2d}+\min\left(\frac{n}{2d}, h_i\right)\right)R_3+O(\sqrt{k}\cdot n^{p-3}m_{ij}), \tag{53}$$

and

$$\Phi_2\leqq m_{ij}\cdot\frac{n}{2d}\cdot\min\left(\frac{n}{2d}, h_i\right)R_4+O(\sqrt{k}\cdot n^{p-3}m_{ij}). \tag{54}$$

On the other hand, deleting m_{ij} edges connecting u_i to F_j-u_j we obtain that

$$\Psi_1\geqq m_{ij}\left(\frac{n}{d}+h_i\right)R_3+O(\sqrt{k}\cdot n^{p-3})$$

and

$$\Psi_2\geqq m_{ij}\cdot\frac{n}{d}\cdot h_iR_4+O(\sqrt{k}\cdot n^{p-3}).$$

By (47), (50), (53) and (54),

$$\Psi^{ij}\geqq\Phi^{ij}+c_7n^{p-2}.$$

We are home.

Case (Q6). $t_i=1, m_{ij}\geqq t_j\geqq 2$. By (38) $m_{ij}=O(1)$, again. Then we delete all lines between u_i and F_j and $m_{ij}-t_j$ other lines between F_j-u_j and B_i-u_i. Then, as above, we have

$$\Psi_1\geqq\left[(t_j-1)\left(\frac{n}{d}+h_i\right)+(\sigma^+_{S''}(u_j)+h_i)+(m_{ij}-t_j)\frac{n}{d}\right]R_3+O(\sqrt{k}\cdot n^{p-3})$$

and by the same argument as in case (Q3), $\Psi_2\geqq\Phi_2$. Hence we get in the case $h_i\geqq\frac{n}{2d}$, using (53),

$$\Psi^{ij}-\Phi^{ij}\geqq\left[t_jh_i+\sigma^+_{S''}(u_j)-\frac{n}{d}\right]R_3+O(\sqrt{k}\cdot n^{p-3})\geqq$$

$$\geqq\left[\left(\frac{t_j}{2}-1\right)\frac{n}{d}+\sigma^+_{S''}(u_j)+t_j\left(h_i-\frac{n}{2d}\right)\right]R_3+O(\sqrt{k}\cdot n^{p-3}),$$

and in case $h_i\leqq\frac{n}{2d}$,

$$\Psi^{ij}-\Phi^{ij}\geqq\left[m_{ij}\frac{n}{2d}-(m_{ij}-t_j)h_i+\sigma^+_{S''}(u_j)-\frac{n}{d}\right]R_3+O(\sqrt{k}\cdot n^{p-3})\geqq$$

$$\geqq\left[\left(\frac{t_j}{2}-1\right)\frac{n}{d}+\sigma^+_{S''}(u_j)\right]R_3+O(\sqrt{k}\cdot n^{p-3}).$$

Hence (46) is proved, unless $t_j=2$, $\sigma^+_{S''}(u_j)\leqq\sqrt[4]{kn^2}$, $h_i\leqq\frac{n}{2d}+\sqrt[4]{kn^2}$. Even in this latter case

$$\Psi^{ij}-\Phi^{ij}\geqq O(\sqrt{k}\cdot n^{p-3}).$$

Put the edge (u_i, u_j) back and delete a line between F_j-u_j and B_i-u_i instead. This way we destroy at least

$$\frac{n}{2d}R_3+o(n^{p-2}).$$

more K_p's than before. This settles this case.

Case (Q7). $t_i=t_j=1$, $m_{ij}\geqq 2$, and e.g. $h_i\geqq h_j$. Again, $m_{ij}=O(1)$. Now we delete (u_i, u_j) and $m_{ij}-1$ horizontal lines. As before,

$$\Psi_2\geqq\Phi_2.$$

Note that in this case we deleted only one vertical edge. Thus

$$\Psi_0=R_2+O(\sqrt{k}\cdot n^{p-3}).$$

For Φ_0 we of course still have (47). Moreover,

$$\Psi_1 \geqq (h_i+h_j)R_3+(m_{ij}-1)R_1+O(\sqrt{k}\cdot n^{p-3})=$$
$$=\left[h_i+h_j+(m_{ij}-1)\frac{n}{d}\right]R_3+(m_{ij}-1)R_2+O(\sqrt{k}\cdot n^{p-3}).$$

We have to estimate Φ_1 a little more carefully than before. Consider two missing lines (u, v) and (w, t) in S', $u,w \in B_i$, $v,t \in B_j$, where, say, $u \neq w$ (we allow $v=t$). Then

$$[\sigma_S^+(u)+\sigma_S^+(v)]+[\sigma_S^+(w)+\sigma_S^+(t)] \leqq$$
$$\leqq (h_i+1)+\sigma_S^+(v)+\sigma_S^+(t) \leqq h_i+h_j+\frac{n}{2d}+O(\sqrt{k}).$$

Hence

$$\Phi_1 \leqq \left(h_i+h_j+\frac{n}{2d}+(m_{ij}-2)\frac{n}{d}\right)R_3+O(\sqrt{k}\cdot n^{p-3}).$$

Thus

$$\Psi^{ij}-\Phi^{ij} \geqq \frac{n}{2d}R_3+O(\sqrt{k}\cdot n^{p-3}),$$

proving (46) again.

Observe that the cases (Q1)–(Q7) prove (46) unless either $m_{ij}=0$ or $m_{ij}=t_i=t_j=1$. Thus we have proved that for every (i, j) $m_{ij}=0$ or $m_{ij}=t_i=t_j=1$. Let us consider the latter case.

Case (Q8). $m_{ij}=t_i=t_j=1$. Of course, we remove (u_i, u_j). Denoting the missing edge of S' between B_i and B_j by (v_i, v_j) we have, by the same type calculations as above,

$$\Psi^{ij}-\Phi^{ij} \geqq [(h_i-\sigma_S^+(v_i))+(h_j-\sigma_S^+(v_j))]R_3+O(\sqrt{k}\cdot n^{p-3}).$$

Hence indeed $\Psi-\Phi \geqq (\sqrt{k}\cdot n^{p-3})$ and it also follows that

$$\sigma_S^+(v_i) \geqq h_i-\sqrt[4]{kn^2} \qquad \sigma_S^+(v_j) \geqq h_j-\sqrt[4]{kn^2}$$

As we have seen at the end of (M), $\sigma^+(v_i) \geqq c_3 n$. Thus $h_i \approx \sigma^+(v_i)$, which implies that v_i is the unique point in B_i with the largest horizontal degree. Hence we may assume that $v_i=u_i$ and $v_j=u_j$. Hence S' and S^{IV} have the same missing edges. Therefore step (ii) and (iv) can be ignored: Q is obtained from S by steps (i), (iii) and (v).

Let us consider step (iii) again. If step (iii) is applied to S' (instead of S''), then the number of regular K_p's remains the same, if B_i meets no missing edge (v_i, v_j), then the number of regular K_p's decreases when a horizontal edge in B_i non-adjacent to v_i is replaced by a horizontal edge adjacent to v_i. (40) is not influenced by omitting steps (ii) and (iv). Thus we get that $k_p(Q) \geqq k_p(S)$. By the extremality of S $k_p(Q) = k_p(S)$ and if (v_i, v_j) is a missing edge ($v_i \in B_i$, $v_j \in B_j$), then all the horizontal edges of B_i are adjacent to v_i. It also follows that no B_i contains a triangle in S. Hence, if $x \in V$, we must have equality in (41), which implies that then *either x is joined to all points of B_i or the neighbors of x in B_i are independent in S.*

(R) We study now Q. Since Q is another extremal graph, it follows that there is at most one i such that $t_i \geqq 2$. Indeed, if t_i, $t_j \geqq 2$ $(i \neq j)$ then, by (Q), $m_{ij} = 0$. Considering an edge connecting $F_i - u_i$ to $F_j - u_j$ we would get a contradiction with (37). So suppose that $t_2, \ldots, t_d \leqq 1$ and thus $F_i = \{u_i\}$ or $F_i = \emptyset$ for $i \geqq 2$.

Consider now a pair of points $x \in B_i$, $y \in B_j$. By $x, y \notin V$

(55) $$\sigma_i^+(x) < b_i - t_i$$

and

(56) $$\sigma_i^+(y) < b_j - t_j .$$

If $\sigma_i^+(x)$, $\sigma_i^+(y) > 0$, we define the *shifting* of edges from x to y as follows. Replace t horizontal edges of form (x, u) by t horizontal edges of form (y, v). Clearly, the number of K_p's of the resulting graph $Q(t)$ is a quadratic function of $t: At^2 + Bt + C$, where $A \leqq 0$. Therefore either $Q(1)$ or $Q(-1)$ has less K_p's than Q, unless $A = 0$, which means that

(57) either $p = 3$ or (x, y) is a missing edge.

In both cases no K_p is containing horizontal edges of type (x, u) and (y, v) at the same time. Now $k_p(Q(t))$ is linear:

$$k_p(Q(t)) = k_p(Q) - (a(x, u) - a(y, v))t ,$$

(where $a(x, u)$ and $a(y, v)$ are independent of the choices of u and v).

Since $k_p(Q) \leqq \min(k_p(Q(-1)), k_p(Q(1)))$, thus

(58) $$a(x, u) = a(y, v)$$

and taking t as large as possible we obtain a $Q' = Q(t)$ for which either

(59) $$st_{Q'}(x) \cap B_i = \emptyset$$

or

(60) $$st_{Q'}(y) \cap B_j - F_j .$$

This operation is called shifting of edges from x to y. We shall use it to prove that *there is at most one missing edge in $Q - V$ (i.e. in $S - V$).*

First we prove that

(61) $$\sigma_Q^+(x)=O(\sqrt{k}),$$

(62) $$\sigma_Q^+(y)=\frac{n}{d}+O(\sqrt{k}),$$

and

(63) $$\sigma_S^+(x)=\sigma_Q^+(x)=\frac{n}{2d}+O(\sqrt{k}).$$

Finally,

(64) x, y are not adjacent to any point in V.

Indeed, if (x, y) is a missing edge, then (Q8) describes the situation: $x=u_i$, $y=u_j$ and all the horizontal edges of B_i are adjacent to x in Q, and in S. Hence

(65) $$\sigma_Q^+(x)=\sigma_S^+(x)\leqq\frac{n}{2d}+O(\sqrt{k})\quad\text{and}\quad\sigma_Q^+y)=\sigma_S^+(y)\leqq\frac{n}{2d}+O(\sqrt{k}).$$

Thus (60) implies

(66) $$_Q^+(x)=\sigma_Q^+(x)+\sigma_Q^+(y)-\sigma_Q^+(y)\leqq\frac{n}{d}-\sigma_Q^+(y)+O(\sqrt{k})\leqq O(\sqrt{k}).$$

This proves (61) in the second case and it is trivial, when (59) holds. (62) is trivial, when (60) holds. If we know only (59), then we apply (32) more precisely (to have $O(\sqrt{k})$), (34) *and* (27), *yielding* $\sigma^+(x)+\sigma^+(y)+\frac{d}{n}\cdot\frac{p-3}{d-p+2}\sigma^+(x), \sigma^+(y)<\frac{n}{d}+O(\sqrt{k})$, obtaining (62) again. (63) follows from (65) and (66), where we have equality.

Finally, if $w\in V$, then, by (62), $\sigma_Q^+(w)+\sigma_Q^+(y)\geqq\frac{3n}{2d}+O(\sqrt{k})$, therefore, applying (37) to (w, y) in Q' we obtain that w and y are not adjacent in Q, proving (64).

Let now (x, y) and (x', y') be two missing edges and assume that y and y' are in B_j and $B_{j'}$, where $j\neq j'$. By shifting the edges into y and y' we can achieve that $\sigma^+(y)+\sigma^+(y')=\frac{2n}{d}+o(n)$ in the obtained Q''.

By (35) and (37*), $(y, y')\notin E(Q'')$, hence (y, y') is a missing edge in Q and S as well. (Since the optimal partition may change during shifting the edges, we used (37*).) Now we shift the edges from y to x in Q but leaving $c_8\sqrt{k}$ edges at y, where c_8 is a sufficiently large constant. This will ensure that the arguments used to establish (32) in S work in Q as well. However, the missing edge (y, y') contradicts (32): $\sigma^+(y)=O(\sqrt{k})$ and

$\sigma^+(y') = \frac{n}{2d} + O(\sqrt{k})$. Thus we have proved that $Q-V$ and $S-V$ contain at most one missing edge.

(S) ... Below, in (S), (T) and (U) we complete the proof. In (S) we investigate the case, when $Q-V$ has no missing edges and at least two B_i's contain (horizontal) edges. In (U) we observe, that the remaining case, when $Q-V$ has (exactly) one missing edge, can be reduced to the cases (S), (T). Case (S) will be subdivided into (S1)–(S4) according to the distribution of the horizontal edges in B_i's. The basic method is to shift the edges so that the resulting graph contains an edge contradicting (37) of (K). Most of the difficulties occur when $p=3$.

First we prove the

Saturation principle. Every $x \in V$ is joined to all vertices of all but one sets $B_i - F_i$, where for $h_i = 0$ $F_i = :\emptyset$.

Indeed, if there are a $B_i - F_i$ and a $B_j - F_j$ not all the vertices of which are joined to x, then delete an edge (x, u), $u \in B_j - F_j$ and add an edge (x, v), $v \in B_i - F_i$. One can easily check that the number of K_p's of the resulting graph Q' decreased if $\sigma_i^+(x) < \sigma_j^+(x)$:

All $B_i - F_i$ and $B_j - F_j$ are independent sets, and therefore the number of K_p's either not containing x or containing x and only one vertex from $B_i \cup B_j$ remains the same, while the number of K_p's containing x, a $u \in B_i - F_i$ and a $v \in B_j - F_j$ is proportional to $\sigma_i^+(x) \cdot \sigma_j^+(x)$, that is, decreases. This contradiction proves the saturation principle.

Now we describe the structure of S in the case when $Q-V$ (or $S-V$) has no missing edges and at least two sets B_i contain horizontal lines. Assume that the indices are chosen so that $h_1 \geq h_2 \geq \ldots \geq h_f > 0$, $h_{f+1} = \ldots = h_d = 0$. So we deal with the case when $f \geq 2$.

Case (S1). Suppose that $|F_1| \geq 2$ and $\sigma_d^+(u_1) < b_1 - t_1$. It follows by (Q) that $|F_2| = \ldots = |F_f| = 1$. Also note that $p=3$ by (57). Shift as many horizontal edges of Q incident with $u_1, \ldots, u_{f-1}$ to u_f as possible. Since u_f is adjacent to $F_1 - u_1$ whose points have degree $b_1 - t_1 = \frac{n}{d} + O(\sqrt{k})$, by (37) its horizontal degree cannot grow too big during this shift, i.e.

$$\sigma_d^+(u_1) + \ldots + \sigma_d^+(u_f) = O(\sqrt{k}).$$

We shall prove that each $x \in V$ is joined to each $v \in B_2 \cup \ldots \cup B_d$. Here we need the

Strong saturation principle. If $f \geq 2$, then each $x \in V$ is joined to all the vertices of all but one sets B_1 $B_1 - F_1 + u_1$, $B_2, \ldots, B_d$. Indeed, fix a $u_i \in B_i - st(x)$ when $h_i = 0$. If e.g. $x \in V$ is not joined to a $v \in B_i$ and a $w \in B_j$, then it is neither joined to $u_i \in B_i$ and $u_j \in B_j$. Assume that $\sigma_i^+(x) \geq \sigma_j^+(x)$. First shift all the h_i edges incident with u_i to some u_l (where $l = j$ is also allowed if $h_j > 0$). This does not change the number of K_3's. Then replace an edge $(x, z) \in E(Q)$ by (x, u_i): now the number of K_3's decreases. This is a contradiction *proving that x is joined completely to each but one of $B_1, B_2, \ldots, B_d$.* A similar argument works if B_2 is replaced by $B_1 - F + u_1$.

We show that $x \in V$ cannot be joined to all points of $B_1 - F_1 + u_1$. As we have seen at the end of (Q8), any $x \in V$ is joined either to all the vertices of B_i or the neighbors of x are independent in S. If x is joined to all the vertices of $B_1 - F_1 + u_1$, then $st(x) \cap B_1$ contains edges in Q, thus in S as well. Therefore $st(x)$ contains a $u \in F_1 - u_1$, too. But $\sigma_Q^+(u) + \sigma_Q^+(x) \geqq \frac{3n}{2d} + o(n)$, contradicting (35) and (37*). Hence x must be adjacent to all points of $B_2 \cup \ldots \cup B_d$ in Q and in S as well. So considering the partition $D_1 = B_1 \cup V$, $D_2 = B_2$, $\ldots$, $D_d = B_d$, every edge between different classes will be in S. Thus $S \in U_0(n, E)$.

Case (S2). $\sigma_Q^+(u_1) \geqq \frac{3n}{4d}$. (We may have $|F_1| = 1$ or > 1.) By the saturation principle each $x \in V$ is connected to all points of all but one sets $B_1 - F_1, B_2, \ldots, B_d$. Suppose that x misses a point in B_2. Consider S. Since x is joined to all the vertices of $B_1 - F_1$ and by (37*) to none of F_1, it is joined to exactly $b_1 - t_1$ vertices of B_1 and these vertices are independent by the results of (Q8). Hence every horizontal edge of S in B_1 contains one of the t_1 points of B_1 non-adjacent to x in S. Thus for at least one of these vertices, say, for u, $\sigma_S^+(u) \geqq \frac{3n}{4d} + O(\sqrt{k})$, contradicting the definition of V. This proves that every $x \in V$ is connected to every y in $B_2 \cup \ldots \cup B_d$. We conclude as above.

Case (S3). $|F_1| = 1$ and $h_1 + \ldots + h_d \leqq \frac{3n}{4d}$. We know by the saturation principle that every $x \in V$ is adjacent to all points of all but at most one sets B_i. If x is not adjacent to all points of B_i, then it is not adjacent to u_i either. Also we have $p = 3$ again.

First let us assume that there exists an $x \in V$ not joined to u_1. Shift all but one horizontal edges of B_1 to B_2. (We know that B_1, B_2 contain horizontal edges!) This shifting results in another extremal graph Q'. Replace this last horizontal edge in B_1 (incident with u_1) by (u_1, x). This does not increase the number of triangles, moreover, it decreases, whenever at least one $y \in V$ is joined to u_1. This proves that no vertex of V is joined to u_1. Thus each $x \in V$ is joined to each $w \in B_2 \cup \ldots \cup B_d$. We conclude as in (S1).

So we may suppose that every $x \in V$ is adjacent to all points of B_1, and similarly to all points of $B_2, \ldots, B_f$. If every $x \in V$ is adjacent to all points in $B_1 \cup \ldots \cup B_d$ then we can again conclude as before. So suppose some $x \in V$ is non-adjacent to some point in B_{f+1} (say). Then $V = \{x\}$; in fact, if there exists another vertex $y \in V$, then we can shift edges connecting x to B_{f+1} to B_1 until the degree of u_1 becomes greater than $n - \frac{n}{3d}$. But then, being connected to y, it contradicts (37*). The case $V = \{x\}$ can be handled again in the same way as case (S1).

Case (S4). $|F_1| = 1$ and $h_1 + \ldots + h_d > \frac{3n}{4d}$, but $\sigma_Q^+(u_i) < b_i - 1$ $(i = 1, \ldots, f)$. Again, we have $p = 3$. It follows then (as before) that every $x \in V$ is joined to all points of all but at most one of $B_1-u_1, \ldots, B_f-u_f, B_{f+1}, \ldots, B_d$. Furthermore, since by shift we can increase the degree of any u_i $(i=1, \ldots, f)$ to more than $\frac{3n}{4d}$, no one of $u_1, \ldots, u_f$ is adjacent to any point in V.

Since the horizontal edges incident with $u_1, \ldots, u_f$ must be contained in the same number of triangles (by (58) in the definition of shift), it follows that

$$|B_1| = \ldots = |B_f|.$$

Shift now as many horizontal edges to B_i's with the smaller indices as possible. Even after this shifting each B_i $(2 \leq i \leq f)$ must contain horizontal edges, otherwise replacing an edge (x, u), $u \in B_1$ by (x, u_i) we could diminish $\sigma_1^+(x) \cdot \sigma_i^+(x)$, hence decreasing the number of K_3's of the resulting graph Q'. This is a contradiction, since Q' is extremal. Thus

$$h_1 + \ldots + h_f \geq (f-1)(|B_1|-1)+1 \geq |B_1|.$$

Hence

$$\sigma_Q^+(u_1) = |B_1| - 1$$

and

$$\sigma_Q^+(u_2) \geq 1.$$

Since u_1 and u_2 are adjacent, by (37) $\sigma_Q^+(u_2) = o(n)$. Thus we can replace (u_1, u_2) by an edge connecting u_2 to $B_2 - u_2$. This decreases the number of K_3's, a contradiction.

(T) Suppose still that $Q-V$ has no missing edges but let now $h_2 = \ldots = h_d = 0$. The argument in (S2) works unless $|F_1| = 2$ and $h_1 \leq \frac{n}{d} + O(\sqrt{k})$ or $|F_1| = 1$ and $h_1 \leq \frac{n}{2d} + O(\sqrt{k})$. Thus $\sigma_Q^+(u_1) < b_1 - t_1$. Assume first $h_1 > 0$.

As above, each $x \in V$ is joined to all points of all but at most one sets $B_1 - F_1$, $B_2, \ldots, B_d$. *Our aim is to show that no $x \in V$ is joined to u_1*. This implies (by (Q8)) that no $x \in V$ is joined to F_1 at all. Thus each $x \in V$ is joined to each $v \in B_2 \cup \ldots \cup B_d$. Hence S satisfies Definition 2 with $W = V \cup F_1$: $S \in U_2(n, E)$.

Let us assume (indirectly) that an $x \in V$ is joined to u_1. By (Q8) x must be connected to all points of B_1. Hence, by (37*), $F_1 - u_1 = \emptyset$, that is, $F_1 = \{u_1\}$.

We show that each $y \in V - x$ is joined to each $v \in Q - V - u_1$. Suppose $y \in V - x$ is not connected to some $v \in B_i$ $(i \geq 2)$, or some $v \in B_1 - u_1$. Then we can shift edges from y to u_1 and achieve $\sigma^+(u_1) \geq \frac{n}{2d} + \sqrt[4]{kn^2}$. This contradicts (37*) since u_1 is adjacent to x.

There are two cases: x is joined to all vertices of $B_2 \cup \ldots \cup B_d$ or not. In the first case the vertices of Q can be partitioned into the classes $V \cup B_1, B_2, \ldots, B_d$ so that vertices belonging to different classes are always adjacent. Unless $h_1 = 0$, the first class contains a K_3, therefore we can rearrange the edges in $V \cup B_1$ ruining all the K_3's and (consequently) diminishing the number of K_3's. This is a contradiction.

In the other case there is a $v \in B_i$ $(i \geqq 2)$ not joined to x. We can shift edges from x to u_1 until $\sigma^+(u_1) \geqq \frac{n}{2d} + \sqrt[4]{kn^2}$ is achieved. This proves (by (37*)) that no $y \in V-x$ is adjacent to u_1.

Also we can shift edges from u_1 to x. If this fills up x, i.e., in the resulting extremal graph Q' x is adjacent to all points of $Q-V$ and there is still a horizontal edge in Q' incident with u_1, then replacing (u_1, x) by a horizontal edge (u_1, w) we decrease the number of K_p's. This is a contradiction. Thus $Q'-V$ contains no horizontal edges. Carry out the same shifting of edges from u_1 to x but stop, when only one horizontal edge (u_1, w) is left. If $V-x \neq \emptyset$, we can replace (u_1, w) by (u_1, y), decreasing the number of K_p's, a contradiction. If $V=\{x\}$, then put x into the class B_i containing the v. Since x is joined to all the vertices of all the other classes, we are home: $Q \in U_0(n, E)$, which implies for $p \geqq 4$ that $Q \in U_1(n, E)$. The same holds for S, too.

The case $h_1=0$ is fairly simple, and left to the reader.

(U) We are left with the case when $Q-V$ has a missing edge (u_1, u_2) with, say, $u_i \in B_i$. Then we know that u_1, u_2 are not adjacent to any point in V. Then replacing V by $V+u_1$ and B_1 by B_1-u_1, the arguments in (S4) and (T) can be applied.

The proof of Theorem 3 is complete.

References

[1] Bollobás, B., Relations between sets of complete subgraphs, *Proc. of the Fifth British Combinatorial Conference (Aberdeen)*, 1975, 79–84.

[2] Erdős, P., On a theorem of Rademacher-Turán, *Illinois Journal of Math.*, **6** (1962), 122–127.

[3] Erdős, P., On the number of complete subgraphs contained in certain graphs, *Magy. Tud. Akad. Mat. Kut. Int. Közl.*, **7** (1962), 459–474.

[4] Goodman, A. W., On sets of acquaintances and strangers at any party, *Amer. Math. Monthly*, **66** (1959) 778–783.

[5] Lovász, L. and Simonovits, M., On the number of complete subgraphs of a graph, *Proc. of Fifth British Combinatorial Conference (Aberdeen)*, 1975, 431–442.

[6] Moon, J. W. and Moser, L., On a problem of Turán, *MTA Mat. Kut. Int. Közl.*, **7** (1962), 283–286.

[7] Turán, P., An extremal problem in graph theory (In Hungarian), *Mat. és Fiz. Lapok*, **48** (1941), 436–452.

EÖTVÖS LORÁND UNIVERSITY
H-1088 BUDAPEST, MÚZEUM KRT. 6—8.
HUNGARY