Cube-Supersaturated Graphs and Related Problems

P. Erdős

M. Simonovits

ABSTRACT

In this paper we shall consider ordinary graphs, that is, graphs without loops and multiple edges. Given a graph L, $ex(n,L)$ will denote the maximum number of edges a graph G^n of order n can have without containing any L. Determining $ex(n,L)$, or at least finding good bounds on it will be called

TURÁN TYPE EXTREMAL PROBLEM. Assume that a graph G^n has $E > ex(n,L)$ edges. Then it must contain some copies of L. Such a graph will be called supersaturated, or *L-supersaturated*. L-supersaturated graphs mostly contain not only one, but very many copies of L. The problem discussed here (and called "the problem of L-supersaturated graphs") is

Determine the minimum number of copies of L a graph G^n with $E > ex(n,L)$ edges must contain.

The main results of this paper are two "recursion theorems" motivated by the case when L is the graph determined by the vertices and edges of a cube.

Notation

Below we shall consider ordinary graphs, that is, graphs without loops and multiple edges. $G, H, \ldots, S$ and $G^n, H^n, \ldots, S^n$ will denote graphs and the upper indices will always denote the number of vertices. Also, we shall use $v(G)$, $e(G)$ and $\chi(G)$ to indicate the number of vertices, edges, and the chromatic number, respectively. K_p is the complete graph on p vertices, C_p is the cycle of length p, $K_{p,q}$ denotes the complete bipartite graph with p and q vertices in its color-classes. $G(A,B)$ denotes the bipartite subgraph of G induced by A and B, $(A \cap B = \emptyset)$.

Below $c_1, \ldots, c_j, \ldots$ will always denote positive constants, *independent of n*, but they may depend on the graphs in consideration and they may be different in different parts of the paper.

1. Introduction

In this paper we shall consider problems on supersaturated graphs. These type of problems are strongly related to extremal graph problems of Turán type.

Turán type extremal graph problems. Let $\mathbb{L}$ be a given family of (so called forbidden) graphs. Determine the maximum number of edges a graph G^n can have without containing any $L \in \mathbb{L}$ as a subgraph.

The maximum will be denoted by $ex(n, \mathbb{L})$, the family of graphs attaining this maximum by $EX(n, \mathbb{L})$. These graphs will be called *extremal*. If $p+1$ denotes the minimum chromatic number in G^n, then, by a theorem of Erdős and Simonovits [3],

$$ex(n, \mathbb{L}) = \left(1-\frac{1}{p}\right)\binom{n}{2} + o(n^2) \tag{1}$$

An extremal graph problem will be called *degenerate* if $p=1$, that is, *if there is a bipartite graph among the forbidden ones*. By (1), the problem is degenerate iff $ex(n, \mathbb{L}) = o(n^2)$. (By [8], if $L \in \mathbb{L}$ is bipartite and $c = 2/v(L)$, then $ex(n, \mathbb{L}) = O(n^{2-c})$.) In this paper we restrict our consideration to this degenerate case.

Let G^n have more than $ex(n, \mathbb{L})$ edges. Then it will be called *supersaturated* and obviously, it contains at least *one* forbidden subgraph. It is surprising that G^n contains in such cases not only one but very many forbidden subgraphs.

THE PROBLEM OF SUPERSATURATED GRAPHS. Given a graph G^n with $E > ex(n, \mathbb{L})$ edges, at least how many subgraphs of G^n are in $\mathbb{L}$? This minimum will be denoted by $f(n, \mathbb{L}, E)$.

First we mention a general theorem on supersaturated graphs.

THEOREM A. (Erdős-Simonovits, [5].) Given a family $\mathbb{L}$ of forbidden graphs, each of which has v vertices. For every $c > 0$ there exists a $c' > 0$ such that if

$$e(G^n) > ex(n, \mathbb{L}) + cn^2, \tag{2}$$

then G^n contains at least $c'n^v$ forbidden $L \in \mathbb{L}$.

Obviously, this result is sharp, since G^n has only $O(n^v)$ subgraphs of order v. The following problem seems much more difficult:

PROBLEM OF WEAKLY SUPERSATURATED GRAPHS. Let $\mathbb{L}$ be a family of forbidden graphs of order v. What is the minimum number of forbidden subgraphs in a graph G^n with

$$e(G^n)=ex(n,L)+k$$

edges if $k=o(n^2)$?

REMARK 1. The above definitions, Theorem A and problems carry over to the case of h-uniform hypergraphs without any difficulty. There, a problem is degenerate if $ex(n,il)=o(n^h)$.

As we mentioned, the problem of *weakly* supersaturated graphs is much more difficult than the case $k=cn^2$. The problem of *"very weakly supersaturated"* graphs seems even more intractable. Thus, for example, we can easily handle the problem of C_4-supersaturated graphs, but see no hope to prove the following

CONJECTURE 1. If $e(G^n)=ex(n,C_4)+1$, then G^n contains at least $\sqrt{n}+o(\sqrt{n})$ copies of C_4.

(If true, then this conjecture is sharp.)

Below we shall consider only the case, when $\mathbb{L}=\{L\}$, (and use the simpler notation $ex(n,L)$ and $f(n,L,E)$). In extremal graph problems and in many other similar situations we find that the extremal configurations tend to behave either in a very regular pattern, or in a very chaotic way: they have almost random structure. We feel that in case of bipartite L *the graphs* G^n with a given number E of edges and *having roughly the minimum number of copies of* L *tend to look like random graphs*. This is the background of the conjectures formulated below, and motivates the results of this paper and of some other ones [5], [6]. To formulate the main conjecture, let us count first the expected number of L's in a random graph G^n, where the edges are chosen independently, at random, and with probability $E/\binom{n}{2}$, so that the expected number of edges be E!).

Let $e=e(L)$, $v=v(L)$. The complete graph K_n contains $a_L\cdot\binom{n}{v}$ copies of L and each occurs in our random graph with probability $\left(E/\binom{n}{2}\right)^e$. Hence the expected number of occurrences of L in G^n is

$$a_L \cdot \binom{n}{v} \cdot \left(E/\binom{n}{2}\right)^e \approx c_L \cdot \frac{E^e}{n^{2e-v}}. \tag{3}$$

CONJECTURE 2. Given a bipartite graph L with $v=v(L)$ and $e=e(L)$ and a $c>0$, then there exists a $c'=c'(c)>0$ such that if

$$E=e(G^n)\geq(1+c)\cdot ex(n,L), \tag{4}$$

then G^n contains at least $c'\cdot\frac{E^e}{n^{2e-v}}$ copies of L.

One can easily show that if we use another random graph model, where the graphs G^n of E edges are chosen with the same probability, then (3) still holds. This shows that Conjecture 2 is sharp if true. The main goal of this paper is to prove Conjecture 2 in various cases. One of the main difficulties is that in most degenerate extremal graph problems, even if we have satisfactory upper bounds on $ex(n,L)$, *we are unable to prove good lower bounds*. Therefore we formulate a weaker form of the conjecture.

CONJECTURE 2^*. Assume that for a given forbidden L $ex(n,L)=O(n^{2-\alpha})$. Then there exist an $\tilde{\alpha}\leq\alpha$ and two constant c and $c'>0$ such that if $E=e(G^n)>cn^{2-\tilde{\alpha}}$, then G^n contains (for $e=e(L)$ and $v=v(L)$) at least $c'\cdot\frac{E^e}{n^{2e-v}}$ copies of L.

We cannot prove the assertion of Conjecture 2^* even under the stronger assumption that $E=e(G^n)>\frac{n^2}{\log n}$.

2. Main Results

PROPOSITION 1. If T is a tree, then Conjecture 2 holds with $L=T$.

In case, when L is an even cycle, Conjecture 2^* with $\tilde{\alpha}=1-\frac{1}{k}$ was proved by Simonovits, [12]. More precisely, a theorem of Erdős, (unpublished, see [1] or [2]) asserts that

$$ex(n,C_{2k})=0(n^{1+1/k}). \tag{5}$$

In [12] it is proved that there exist a $c_k>0$ and a $c_k'>0$ such that if

$$E=e(G^n)>c_k n^{1+1/k}, \tag{6}$$

then G^n contains at least $c_k'\cdot\left(\frac{E}{n}\right)^k$ copies of C_{2k}. This is sharp by

Remark 2, and proves Conjecture 2*. (For $k=2,3$, and 5 we know that (5) is sharp. Hence (6) proves Conjecture 2 for C_4, C_6 and C_{10}.)

If we wish to prove general degenerate extremal theorems, or corresponding results on supersaturated extremal graphs, one way to do this is to look for *recursion theorems*. A recursion theorem is a result asserting that

(a) if we have an upper bound on $ex(n,L)$ and construct a new graph L' from L in some given way, then we obtain an upper bound on $ex(n,L')$ in a simple way, in terms of $ex(n,L)$ and some other parameters of L';

(b) or, correspondingly, we can obtain lower bounds on $f(n,L',E)$ in terms of $f(n,L,E)$.

In [10] we have proved the following recursion theorem:

DEFINITION 1. Given a bipartite graph L and a fixed two-coloring of it by blue and red, L_t denotes the graph obtained by fixing a $K_{t,t}$ (which is vertex-disjoint from L and is colored by blue and red) and then joining all the red vertices of $K_{t,t}$ to all the blue vertices of L and all the blue ones of $K_{t,t}$ to all the red ones of L, (see Figure 1).

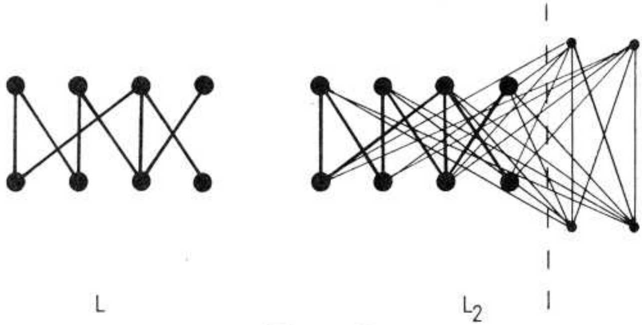

Figure 1.

THEOREM B. [4]. Let L be a bipartite graph with a fixed 2-coloring. Let $ex(n,L)=O(n^{2-\alpha})$ and define $\beta=\beta_t$ by

$$\frac{1}{\beta}-\frac{1}{\alpha}=t. \tag{7}$$

Then

$$ex(n,L_t)=0(n^{2-\beta}). \tag{8}$$

This theorem was needed primarily to prove the Cube theorem: Turán asked that if L denotes one of the five graphs corresponding to the five regular polyhedra, can one determine $ex(n,L)$? For tetrahedron (K_4) the answer is given by Turán's theorem, [13,14], for octahedron by us, for dodecahedron and icosahedron by Simonovits, see the survey [11], or [10]. The cube problem seems to be the most difficult. We [4] have proved:

THEOREM C. If Q denotes the graph determined by the vertices and edges of the cube and Q^* is the graph obtained by joining two opposite vertices of the cube by an edge, (see Figure 2), then

$$ex(n,Q)<ex(n,Q^*)=O(n^{8/5}). \tag{9}$$

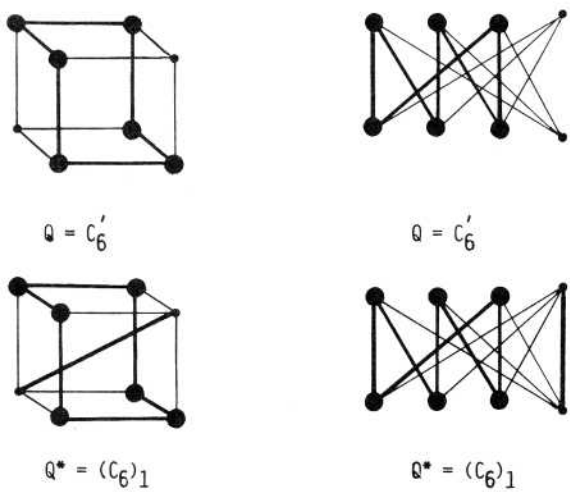

Figure 2.

(One can see in Figure 2 that if $L=C_6$, then $L_1=Q^*$. Hence the cube theorem immediately follows from $ex(n,C_6)=0(n^{4/3})$ and Theorem B.)

Here we shall prove

THEOREM 1. Let L be a bipartite graph with a fixed 2-coloring and $ex(n,L)=0(n^{2-\alpha})$ for some $\alpha\in(0,1)$. If β is defined by (7), and

Conjecture 2^* holds for L, then it also holds for L_t: there exists a constant $C>0$ such that if $e(G^n)=E>Cn^{2-\beta}$, then G^n contains at least $C_{L,t}\cdot\frac{E^{e'}}{n^{2e'-v'}}$ copies of L_t, where $e'=e(L_t)$, $v'=v(L_t)$.

THEOREM 2. Let L be a bipartite graph with a fixed 2-coloring in red and blue and L^* be the graph obtained from L by taking two new vertices, x and y and joining x to all the blue vertices of L and y to all the red ones (but not joining x to y). If $ex(n,L)=0(n^{2-\alpha})$ with some $\alpha\in(0,1)$ and β is defined by (7), and Conjecture 2^* holds for L, then it also holds for L^* in the sense that there exists a constant $C>0$ such that if $E=e(G^n)>Cn^{2-\beta}$ then G^n contains at least $C_L^*\cdot\frac{E^{e'}}{n^{2e'-v'}}$ copies of L^*, where $e'=e(L^*)=e(L)+v(L)$ and $v'=v(L^*)=v(L)+2$.

These theorems immediately imply the following

THEOREM 3. There exist three constant $C_Q,c,c^*>0$ such that if $E=e(G^n)>C_Q\cdot n^{8/5}$, then G^n contains at least cE^{12}/n^{16} copies of Q and at least $c^*\cdot E^{13}/n^{18}$ copies of Q^*.

REMARK. None of the theorems above imply the other, since e' is different in them. Hence Theorem 1 yields (generally) more copies of a smaller sample graph.

Using Proposition 1 and the above theorems we obtain

THEOREM 4. If L is a bipartite graph and x,y are two vertices of L such that $L-\{x,y\}$ is a tree, then $(ex(n,L)=O(n^{3/2})$, and there exist two constants C and c^* such that if $E=e(G^n)>Cn^{3/2}$, then G^n contains at least $c^*\cdot\frac{E^e}{n^{2e-v}}$ copies of L, where $e=e(L)$ and $v=v(L)$.

Theorem 4 has many applications. Thus e.g. if Q' is the graph obtained from the cube graph Q by deleting *one* edge, then Q' can be obtained from a path P^6 by the operation described in Theorem 2. Thus Theorem 2 can be applied with $\alpha=1$, $\beta=\frac{1}{2}$, yielding that if $e(G^n)>c_{Q'}\cdot n^{3/2}$, then G^n contains at least $c'\cdot\frac{E^7}{n^6}$ copies of Q'. There are many other cases, where Theorem 4 immediately yields the sharp result.

3. Proofs

The main tool of the proof of the Cube Theorem (Theorem C) is counting C_4's in G^n. Here we need a result sharper than that of [4]. We shall count C_4's in bipartite graphs where the sizes of the color-classes are m and $h=n-m$. Since our method works for $K_{p,q}$ as well, first we shall prove

LEMMA 1. Given p and q, there exist two constant c and $c'>0$ such that if $G^n \subseteq K_{m,h}$, and $E=e(G^n)>c\cdot m\cdot h^{1-1/p}$, then G^n contains at least

$$c'\cdot\frac{E^{pq}}{m^{(q-1)p}\cdot h^{(p-1)q}} \tag{10}$$

copies of $K_{p,q}$.

(Lemma 1 implies that Conjecture 2* holds for $K_{p,q}$ with $\tilde{\alpha}=\frac{1}{p}$, since for any L it is sufficient to prove Conjecture 2* for the case of bipartite G^n.)

PROOF. We shall use a convexity argument for which we extend $\binom{n}{k}$ to all the reals by

$$\binom{x}{k}=\begin{cases}\frac{x(x-1)\dots(x-k+1)}{k!} & \text{if } x\ge k-1\\ 0 & \text{if } x<k-1\end{cases}.$$

One can easily show that $\binom{x}{k}$ is convex in $(-\infty,+\infty)$, and in $(0,+\infty)$

$$\frac{x^k}{k!}\ge\binom{x}{k}=\frac{x^k}{k!}+O(x^{k-1})\ge\frac{1}{2}\cdot\frac{x^k}{k!}+O(1). \tag{11}$$

Let M and H be the color-classes of $K_{m,h}\supseteq G^n$. We shall call $(x,\{y_1,\dots y_q\})$ a *cap* if each y_j is joined to $x\in M$. Let us count them. If S is their number and $d_1,\dots,d_m$ are the degrees in M, then

$$S=\sum_{i\le m}\binom{d_i}{q}\ge m\cdot\binom{2E/m}{q}\ge c_1\cdot\frac{E^q}{m^{q-1}}-O(m)\ge c_2\cdot\frac{E^q}{m^{q-1}}, \tag{12}$$

Here we have used Jensen's Inequality and that $E/m\to\infty$. (The case $h=O(1)$ is trivial!)

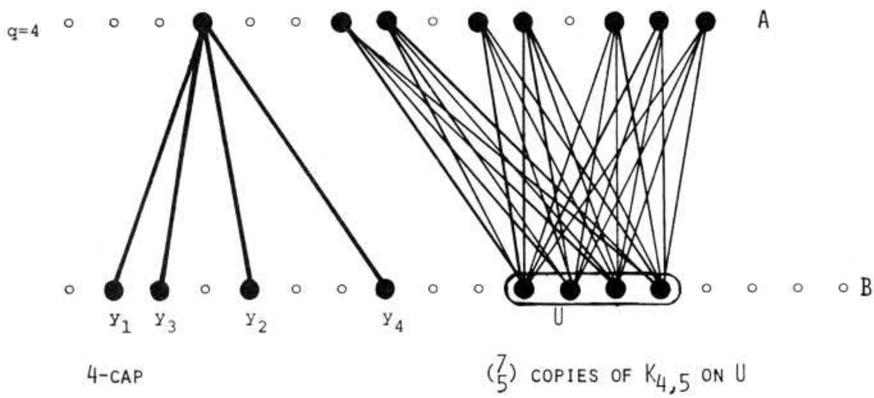

Figure 3

Let D_U denote the number of caps on a fixed $U=\{y_1,\ldots,y_q\}$. Then $\sum D_U=S$ and the number of $K_{p,q}\subseteq G^n$ is exactly

$$\sum\binom{D_U}{p}\geq\binom{h}{q}\binom{\frac{\sum D_U}{\binom{h}{q}}}{p}\geq c_3\cdot\frac{S^p}{h^{qp-p}}-O(h^q)$$

$$\geq c_4\cdot\frac{E^{pq}}{m^{(q-1)p}\cdot h^{(p-1)q}}.$$

(Here again we replaced $\binom{h}{q}$ by $c_5\cdot h^q$, which is permitted unless $h\leq q-1$. However, $h=O(1)$ have been excluded.)

LEMMA 2. Let $G(M,H)$ be a bipartite graph with the colour-classes M and H, where $|M|=m\geq|H|=h$. Assume that L is a bipartite graph which satisfies Conjecture 2^* with a given $\bar{\alpha}>0$. If $E=e(G(M,H))\geq c_6\cdot m\cdot h^{1-\bar{\alpha}}$ then $G(M,H)$ contains at least $\bar{c}\cdot\frac{E^e}{h^{e-v+1}\cdot m^{e-1}}$ copies of L, where $e=e(L)$, $v=v(L)$.

PROOF. We partition M into the subsets $U_1,\dots,U_{\lceil m/h\rceil}$, each of which is of size $\approx h$. Let F_i denote the number of edges joining U_i and H. On average, $G(U_i,H)$ has $\geq c_6 \cdot h^{2-\bar\alpha}$ edges, hence at least half of the edges belong to $G(U_i,H)$'s with $F_i \geq c_6 \cdot \frac{1}{2} \cdot h^{2-\bar\alpha}$: we may apply the assumption to them, they contain at least $c_7 \cdot (F_i)^e / h^{2e-v}$ copies of L. Thus we get at least

$$c_7 \cdot \sum \frac{F_i^e}{h^{2e-v}} = c_8 \cdot \left(\frac{Eh}{m}\right)^e \cdot \frac{m}{h} \cdot \frac{1}{h^{2e-v}} = c_8 \cdot \frac{E^e}{h^{e-v+1} \cdot m^{e-1}}$$

copies of L.

REMARK. This estimate is asymmetric. A symmetric weakening is that, under the assumptions of the lemma, $G(M,H)$ contains at least

$$c_8 \cdot \frac{E^e}{(hm)^{e-v+1} \cdot (h+m)^{v-2}} \tag{13}$$

copies of L.

PROOF OF THEOREM 1. (A) The operation $\mathbf{A}_t : L \to L_t$ is such that

$$\mathbf{A}_1(\mathbf{A}_{t-1}(L)) = \mathbf{A}_t(L).$$

A similar identity holds for the exponents: denote the β obtained from (7) by $\beta_t(\alpha)$. Then

$$\beta_1(\beta_{t-1}(\alpha)) = \beta_t(\alpha).$$

Using this, the reader can easily check that if Theorem 1 is proved for $t=1$, then a trivial induction on t proves Theorem 1 for every t. Hence *it is enough to prove Theorem 1 for t=1*

(B) Take a graph G^n with E edges and partition its vertex set into two parts M and H so that the bipartite graph $G(M,H)$ – defined by the edges of G^n joining M to H – has the maximum number of edges for all possible partitions. Let $d(x)$ denote the degree in G^n, $d'(x)$ the degree in $G(M,H)$. Then one can easily see that $d'(x) \geq \frac{1}{2} d(x)$ for every vertex x.

In the argument below, the vertices of *high valency are disturbing*, therefore we apply below a *two-step regularization procedure* to $G(M,H)$. We define the constants $r_i = 2^{i-2}/i^2$ for $i=4,5,\dots$ and $r_3=0$. Let $|M|=m$ and $|H|=h$ and

$$A_i=\{x\in M\colon r_{i-1}\cdot\frac{E}{m}\le d'(x)<r_i\cdot\frac{E}{m}\},\quad i=4,5,\ldots \tag{14}$$

Clearly, for at least one $i\ge 5$ $|A_i|\ge\frac{m}{2^i}$. We fix an $A^*\subseteq A_i$ with $|A^*|=\lceil\frac{m}{2^i}\rceil$. The bipartite graph $G(A^*,H)$ has at least $\frac{E}{4i^2}$ edges. (This means that the degrees in the first class of $G(A^*,H)$ are already roughly the same and still, $G(A^*,H)$ has almost (?) the original number of edges. Next, we partition H into the subsets

$$B_j=\{y\in H\colon r_{j-1}\cdot\frac{E}{4i^2h}\ge d''(y)<r_j\cdot\frac{E}{4i^2h}\}\ j=4,5,\ldots \tag{15}$$

where $d''(y)$ denotes the degree of y in $G(A^*,H)$. As in the previous case, we can fix a $j\ge 5$ such that $|B_j|\ge\frac{h}{2^j}$, then a $B^*\subseteq B_j$ of $\lceil\frac{h}{2^j}\rceil$ vertices. Now the graph $G(A^*,B^*)$ has $E^*\ge\frac{E}{16i^2j^2}$ edges. (This means that the degrees in B^* are now roughly the same and *at most* $r_j\cdot\frac{E}{4i^2h}$. At the same time, we have lost the "almost regularity" in A^*. However, we still have that the degrees in A^* are *at most* $r_i\cdot\frac{E}{m}$.)

(C) Now we shall count the number of C_4's in $G^*=G(A^*,B^*)$. Let $N(x,y)$ denote the number of paths P_4 joining an $x\in A^*$ to a $y\in B^*$. If (x,y) is an edge, then $N(x,y)$ is just the number of C_4's containing (x,y). Hence, by Lemma 1, applied with $p=q=2$, we get

$$\sum N(x,y)\ge c_1\cdot\frac{(E^*)^4}{(a^*\cdot b^*)^2} \tag{16}$$

C_4's, if $|A^*|=a^*$ and $|B^*|=b^*$. Of course, to get (16) we need that if e.g. $a^*\ge b^*$, then $e(G^*)\ge c_2\cdot a^*\cdot\sqrt{b^*}$. Observe that $\beta\le ½$ in all our applications, hence $E\ge c_3 n^{3/2}$ can be assumed with a "very large" c_3, which (since $a^*=\lceil\frac{m}{2^i}\rceil$, $b^*=\lceil\frac{h}{2^j}\rceil$ and $E^*\ge\frac{E}{16i^2j^2}$) implies that $e(G^*)\ge c_2\cdot a^*\cdot\sqrt{b^*}$. (The reader is reminded at this point that the constants c_i are not necessarily the same in different proofs.)

Let now $d^*(x)$ denote the degree of x in G^*. The edge (x,y) where $x\in A^*, y\in B^*$, will be called of *type A* if $d^*(x)\ge d^*(y)$. Otherwise, it will be called of *type B*. By symmetry, we may assume that (16) holds also for the edges of type A. Assume first (which is not always true) that we can apply Lemma 2 to *each* of these (at most E^*) edges. Each (x,y) defines a subgraph $G_{x,y}\subseteq G^*$ spanned by all the neighbours of x in G^* but y, and by all the neighbours of y but x. (See

Figure 4.) If $G_{x,y}$ contains an L, then G^* contains a corresponding L_1. Thus we obtain at least

$$c_4 \cdot \frac{N(x,y)^e}{d^*(x)^{e-1} \cdot d^*(y)^{e-v+1}} \geq c_5 \cdot \frac{N(x,y)^e}{\left(2^i \cdot \frac{E}{m}\right)^{e-1} \cdot \left(2^j \cdot \frac{E}{h}\right)^{e-v+1}} \quad (17)$$

copies of L_1 containing (x,y). (Here, we used only the weaker bounds $d^*(x) \leq 2^i \cdot \frac{E}{m}$ and $d^*(y) \leq 2^j \cdot \frac{E}{h}$.)

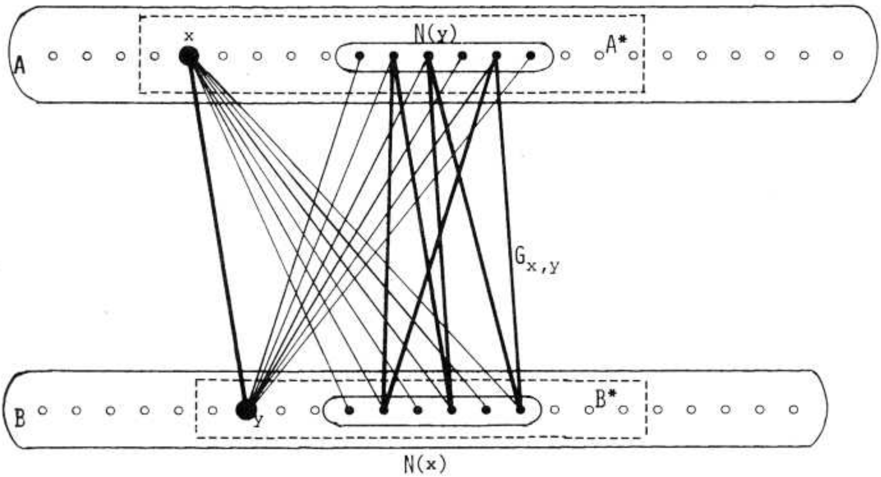

Figure 4

Since, by Jensen's Inequality,

$$\sum N(x,y)^e \geq c_6 \cdot \left(\frac{(E^*)^3}{(a^*b^*)^2}\right)^e \cdot E^* = c_6 \frac{(E^*)^{3e+1}}{(a^*b^*)^{2e}}, \quad (18)$$

and since $a^* = \lceil \frac{m}{2^i} \rceil$, $b^* = \lceil \frac{h}{2^j} \rceil$, we obtain – by adding up the contribution (17) for each edge of type A – at least

$$\sum c_4 \cdot \frac{N(x,y)^e}{d^*(x)^{e-1} \cdot d^*(y)^{e-v+1}} \tag{19}$$
$$\geq c_7 \cdot \frac{(E^*)^{3e+1}}{m^{2e} \cdot h^{2e}} \cdot \frac{1}{\left(\frac{E}{m}\right)^{e-1} \cdot \left(\frac{E}{h}\right)^{e-v+1}} \cdot \frac{2^{ie} \cdot 2^{je}}{2^{i(e-1)} \cdot 2^{j(e-v+1)}}$$

copies of $L_1 \subseteq G^*$. Since $E^* \geq \frac{E}{16i^2j^2}$, (19) yields at least

$$c_8 \cdot \frac{E^{e+v+1}}{m^{e+1} \cdot h^{e+v-1}} \cdot \frac{2^{i+j}}{P(i,j)} \tag{20}$$

copies of L_1, where $P(i,j)$ is some polynomial in i and j. Clearly, $\frac{2^{i+j}}{P(i,j)} \to \infty$ if either $i \to \infty$ or $j \to \infty$. Hence we may assume that i and j are bounded, which means that the last factor may be ignored. Since $e' = e+v+1$ and $v' = v+2$, and since $m^{e+1} \cdot h^{e-v+1}$ attains its maximum (for $m+h=n$) if $m = c_8 n$, (20) yields at least

$$c_9 \cdot \frac{E^{e'}}{n^{2e'-v'}}$$

copies of L_1, each obtained $O(1)$ times. This *would* complete our proof. Unfortunately, we do not know if Lemma 2 is applicable to each $G_{x,y}$.

(D) One can easily check that (16) guarantees that *"on average" Lemma 2 is applicable to* $G_{x,y}$. More precisely, the average of $N(x,y)$ is at least $c_1 \cdot \frac{(E^*)^3}{(a^* \cdot b^*)^2}$. We may fix the constant C of Theorem 1 so large that if $E = e(G^*) > Cn^{2-\beta}$, then either $d^*(x) \geq d^*(y)$ and

$$\frac{(E^*)^3}{(a^* \cdot b^*)^2} \geq C^* \cdot d^*(x) \cdot d^*(y)^{1-\alpha}, \tag{21}$$

or $d^*(x) < d^*(y)$ and

$$\frac{(E^*)}{(a^* \cdot b^*)^2} \geq C^* \cdot d^*(x)^{1-\alpha} \cdot d^*(y). \tag{21*}$$

Indeed, $E^* \geq \frac{E}{16i^2j^2}$ and $d^*(x) \leq 2^i \cdot \frac{E}{m}$, $d^*(y) \leq 2^i \cdot \frac{E}{h}$, from which a short calculation (using also (7)!) yields (21). Now we take only those edges of G^* which satisfy

$$N(x,y)\geq\frac{1}{2E^*}\sum N(x,y).$$

Let us call them "good" edges. Clearly (16) holds for these edges too, if c_1 is replaced by $\frac{1}{2}c_1$. Hence we may assume that if we take only the *"good" edges of type A*, (16) still holds, with $\frac{1}{4}c_1$ instead of c_1. Let E^{**} denote the number of these edges. Trivially, (18) and (19) remain valid if we use (as above) E^* – and not E^{**} – in our formulas, except that the summation is taken only for the E^{**} good edges of type A. Indeed,

$$\sum N(x,y)^e\geq c_6^*\left(\frac{(E^*)^4}{(a^*\cdot b^*)^2}\cdot\frac{1}{E^{**}}\right)^e\cdot E^{**}\geq c_6\cdot\frac{(E^*)^{3e+1}}{(a^*b^*)^{2e}}. \tag{18*}$$

In (19) we used only (18), thus it remains valid and the last part of the proof carries over without any change. Thus Theorem 1 is proved.

PROOF OF THEOREM 2. This proof is very similar to that of Theorem 1. We start exactly as above: choose a bipartite $G(M,H)$ in G^n and then apply the "two-step regularization", that is, find $G^*=G(A^*,B^*)$, as above. $N(x,y)$ again denotes the number of paths P_4 joining an $x\in A^*$ to a $y\in B^*$. However, here we shall count the P_4's in G^* instead of counting C_4's and will use the summation for all the pairs, not only for the joined ones, to establish a formula corresponding to (16). First we assert that

$$\sum N(x,y)\geq c_1\cdot\frac{(E^*)^3}{a^*\cdot b^*}. \tag{16'}$$

Indeed, $\sum N(x,y)$ counts the number of P_4's in G^*. If $E^*\geq 4a^*$, then the number of P_3's ($=K_{1,2}$'s) with the middle vertex in A^* is at least $c_2\cdot\frac{(E^*)^2}{a^*}$. (Here we use Lemma 1 in a sharper form: we use also that we can guarantee that the *first class of $K_{p,q}$ is in the second class of $K_{m,h}$*: as a matter of fact, we have proved this! Of course, counting the $K_{1,q}$'s in G^* is just "counting the caps" in the proof of lemma 1.)

Anyway, we have at least $c_2\cdot\frac{(E^*)^2}{a^*}$ P_3's in G^* with the middle vertex in A^*. The degrees in B^* are between $r_{j-1}\cdot\frac{E}{i^2h}$ and $r_j\cdot\frac{E}{i^2h}$, and therefore they are bounded from below by $c_3\cdot\frac{E^*}{b^*}$ (= the average degree in G^* in B^*). Hence we may extend each P_3 in at least $2c_3\cdot\frac{E^*}{b^*}$

ways. Thus we have at least $c_4\cdot\frac{(E^*)^3}{a^*b^*}$ 4-walks in G^*, and least $c_5\cdot\frac{(E^*)^3}{a^*b^*}$ of them are paths P_4 if E^*/b^* is sufficiently large (which can be assumed here).

From here on, the proof is roughly the same as in the previous case. Instead of (18) we get (summing for all the a^*b^* pairs (x,y))

$$\sum N(x,y)^e \geq c_6\cdot\left(\frac{(E^*)^3}{(a^*\cdot b^*)^2}\right)^e\cdot a^*\cdot b^* = c_6\cdot\frac{(E^*)^{3e}}{(a^*b^*)^{2e-1}}. \qquad (18')$$

Thus the number of L^*'s in G^* is at least

$$\sum c_7\cdot\frac{N(x,y)^e}{d^*(x)^{e-1}\cdot d^*(y)^{e-v+1}} \qquad (19')$$

$$\geq c_8\cdot\frac{(E^*)^{3e}}{(mh)^{2e-1}}\cdot\frac{1}{\left(\frac{E}{m}\right)^{e-1}\cdot\left(\frac{E}{h}\right)^{e-v+1}}\cdot\frac{2^{ie}\cdot 2^{je}}{2^{i(e-1)}\cdot 2^{j(e-v+1)}}.$$

Since $E^*\geq\frac{E}{16i^2j^2}$, (19') gives at least

$$c_9\cdot\frac{E^{e+v}}{m^e\cdot h^{e+v-2}}\cdot\frac{2^{i+j}}{P(i,j)} \qquad (20')$$

copies of L^* in G^*, where $P(i,j)$ is again a polynomial. Again, the last term in (20') tends to ∞, hence it can be deleted: G^* contains at least

$$c_{10}\cdot\frac{E^{e+v}}{m^e h^{e+v-2}} = c_{10}\cdot\frac{E^{e'}}{m^{e'-v'+2}\cdot h^{e'-2}}$$

copies of L^* (since $v'=v+2$, $e'=e+v$). Here the minimum is achieved (under the condition that $m+h=n$) if $m=c_{11}\cdot n$, $h=c_{12}\cdot n$, with $c_{11},c_{12}>0$. This would complete our proof.

As in the proof of Theorem 1, we still have a technical problem: "*Can we apply Lemma 2 to the graphs $G_{x,y}$?* Do the graphs $G_{x,y}$ have sufficiently many edges?" However, from (16') we get that $G_{x,y}$ has, on average, $c_1\cdot\frac{(E^*)^3}{(a^*\cdot b^*)^2}$ edges. This is exactly the same as in the previous proof. Therefore the last part of that proof, namely (D), can be repeated without any significant changes. (The insignificant changes are that the summation is taken for a^*b^* terms instead of E^* terms.) This completes the proof.

REMARK. The fact that the average number of edges in $G_{x,y}$ is the same in both proofs is surprising at first sight but becomes evident (?) if we remember that most parts of our proofs aim at proving that our graph behaves (from certain point of view) as a random graph. In a random graph $e(G_{x,y})$ is "independent" of whether (x,y) is an edge of G^* or not.

We can prove Conjecture 2^* in several other special cases, but shall not discuss them here.

The proof of Theorem 2 (or of Theorem 1) yields a proof of the Cube theorem as well. However, this is not radically different from the original one. Perhaps one interesting difference is that here we used a slightly weaker regularization method.

On the other hand, there are some instances where, in proving Conjecture 2^*, we obtain new (and often simpler) proofs of existing extremal graph theorems.

References

[1] J. A. Bondy and M. Simonovits: Cycles of even length in graphs, Journal of Combin. Theory, *16* (1974) 97-105.

[2] B. Bollobás: Extremal Graph Theory, London Math. Soc. Monographs, Nll, Academic Pres,, 1978.

[3] P. Erdös and M. Simonovits: A limit theorem in graph theory, Studia Sci. Math. Hungar. *1* (1966) 51-57.

[4] P. Erdös and M. Simonovits: Some extremal problems in graph theory, Coll. Math. Soc. J. Bolyai *4* (1969) 377-390.

[5] P. Erdös and M. Simonovits: Supersaturated graphs and hypergraphs, Combinatorica, 2(3) (1982) 275-288.

[6] P. Erdös and M. Simonovits: Compactness results in extremal graph theory, Combinatorica, 3 (2) (1983).

[7] R. J. Faudree and M. Simonovits: On a class of degenerate extremal problems, Combinatorica, 3 (1) (1983), 97-107.

[8] T. Kövári, V. T. Sós and P. Turán: On a problem of Zarankiewicz, Coll. Math., 3 (1954), 50-57.

[9] M. Simonovits: Extremal graph problems with symmetrical extremal graphs, additional chromatic conditions, Discrete Math. 7 (1974), 279-319.

[10] M. Simonovits: The extremal problem of the icosahedron, J. Combin. Theory, *B* 17 (1) (1974), 69-79.

[11] M. Simonovits: Extremal graph theory, a survey, to appear in " Selected Topics in Graph Theory, II", Ed. Beineke and Wilson, Academic Press (1983), 161-200.

[12] M. Simonovits: Cycle-supersaturated graphs, to be published

[13] P. Turán: On the theory of graphs, Coll. Math. 3 (1954), 19-30.

[14] P. Turán: On an extremal problem in graph theory, (in Hungarian) Mat. Fiz. Lapok 48 (1941) 436-452.