Annals of Discrete Mathematics 3 (1978) 29–41.

EXTREMAL GRAPHS WITHOUT LARGE FORBIDDEN SUBGRAPHS

B. BOLLOBÁS

Department of Pure Mathematics and Mathematical Statistics, Cambridge CB2 1SB, England

P. ERDÖS

Hungarian Academy of Sciences, Budapest, Hungary

M. SIMONOVITS

Department of Mathematics, Eötvös University, Budapest VIII, Hungary

E. SZEMERÉDI

Mathematical Institute, Hungarian Academy of Sciences, Budapest V, Hungary

The theory of extremal graphs without a fixed set of forbidden subgraphs is well developed. However, rather little is known about extremal graphs without forbidden subgraphs whose orders tend to ∞ with the order of the graph. In this note we deal with three problems of this latter type. Let L be a fixed bipartite graph and let $L+E^m$ be the join of L with the empty graph of order m. As our first problem we investigate the maximum of the size $e(G^n)$ of a graph G^n (i.e. a graph of order n) provided $G^n \not\supset L+E^{[cn]}$, where $c>0$ is a constant. In our second problem we study the maximum of $e(G^n)$ if $G^n \not\supset K_2(r, cn)$ and $G^n \not\supset K^3$. The third problem is of a slightly different nature. Let $C^k(t)$ be obtained from a cycle C^k by multiplying each vertex by t. We shall prove that if $c>0$ then there exists a constant $l(c)$ such that if $G^n \not\supset C^k(t)$ for $k=3, 5, \ldots, 2l(c)+1$, then one can omit $[cn^2]$ edges from G^n so that the obtained graph is bipartite, provided $n>n_0(c, t)$.

Our notation is that of [1]. Thus G^n is an arbitrary graph of order n, K^p is a complete graph of order p, E^p is a null graph of order p (that is one with no edges), C^m is a cycle of length m, $G_r(n_1, \ldots, n_r)$ is an r-partite graph with n_i vertices in the ith class, $K_r(n_1, \ldots, n_r)$ is a complete r-partite graph. $C^m(t)$ is a graph obtained from C^m by multiplying it by t, that is by replacing each vertex by t independent vertices. We use H^m, S^m, T^m, U^m to denote graphs of order m with properties specified in the text. We write $|A|$ for the cardinality of a set A, $|G|$ for the order of a graph G and $e(G)$ for the number of edges (the *size*) of G. The set of neighbours of a vertex x is denoted by $\Gamma(x)$ and $d(x)=|\Gamma(x)|$ is the degree of x. The minimum degree in G is $\delta(G)$.

Let $\mathcal{F}$ be a family of graphs, called the family of *forbidden* graphs. Denote by $\mathrm{EX}(n, \mathcal{F})$ the set of graphs of order n with the *maximal* number of edges that does not contain any member of $\mathcal{F}$. The graphs in $\mathrm{EX}(n, \mathcal{F})$ are the *extremal graphs of order n for* $\mathcal{F}$. Write $\mathrm{ex}(n, \mathcal{F})$ for the size of the extremal graphs: $\mathrm{ex}(n, \mathcal{F})=e(H)$, where $H \in \mathrm{EX}(n, \mathcal{F})$. The problem of determining $\mathrm{ex}(n, \mathcal{F})$ or $\mathrm{EX}(n, \mathcal{F})$ may be called a Turán type extremal problem. We shall prove some Turán type extremal results in which the forbidden graphs depend on n. The first deep theorem of this kind was proved by Erdös and Stone [8] in 1946. This theorem is the basis of the theory of extremal graphs without forbidden subgraphs (see [1, Ch. VI]). Considerable extensions of it were proved by Bollobás and Erdös [2] and by Bollobás, Erdös and Simonovits [3].

For fixed r and t the extremal graphs $\mathrm{EX}(n, K_3(2, r, t))$ were studied by Erdös and Simonovits [7]. Our first aim in this note is to describe $\mathrm{EX}(n, K_3(2, r, cn))$, where $r \geq 2$ and $c > 0$. In fact, we prove the following somewhat more general result.

Theorem 1. *Let L be a bipartite graph. Put*

$$q(n, L) = \max\{n_1 n_2 + \mathrm{ex}(n_1, L) + \mathrm{ex}(n_2, L) : n_1 + n_2 = n\}. \tag{1}$$

There exist $c > 0$ and n_0 such that if $n > n_0$ and

$$e(G^n) > q(n, L), \tag{2}$$

then G^n contains an $L + E^{[cn]}$. If in addition for every m there exists an extremal graph $S^m \in \mathrm{EX}(m, L)$ with maximum degree $< \frac{1}{2}cm$, then

$$\mathrm{ex}(n, L + E^{[cn]}) = q(n, L) \tag{3}$$

and every extremal graph $U^n \in \mathrm{EX}(n, L + E^{[cn]})$ can be obtained from an $S^m \in \mathrm{EX}(m, L)$ and an $S^{n-m} \in \mathrm{EX}(n-m, L)$ as $S^m + S^{n-m}$.

Remarks. (i) If $L = K_2(2, r)$, then the maximum degree of any $S^m \in \mathrm{EX}(m, L)$ is $o(m)$ and the same holds if L is not a tree, but there exists a vertex $v \in L$ for which $L - v$ is a tree. Thus Theorem 1 gives

$$\mathrm{ex}(n, K_3(2, r, [cn])) = q(n, K_2(2, r)).$$

It also gives information on the structure of the extremal graphs.

(ii) Theorem 1 states that $q(n, L)$ is an upper bound for $\mathrm{ex}(n, L + E^{[cn]})$. A lower bound for $\mathrm{ex}(n, L + E^{[cn]})$ can be obtained by observing that if $S^m \in \mathrm{EX}(n, L)$, then $S^m + E^{n-m} \not\supset L + E^{[cn]}$, so

$$\mathrm{ex}(n, L + E^{[cn]}) \geq \max\{n_1 n_2 + \mathrm{ex}(n_1, L) : n_1 + n_2 = n\}. \tag{4}$$

In some cases, for instance if L consists of independent edges, (4) is sharp.

(iii) The essential part of Theorem 1 states that a graph G^n not containing an $L + E^{[cn]}$ can not have more edges than $S^p + S^{n-p}$, where $S^p \in \mathrm{EX}(p, L)$, $S^{n-p} \in \mathrm{EX}(n-p, L)$ and p is suitably chosen. It is unfortunate that $S^p + S^{n-p}$ may contain an $L + E^{[cn]}$ and we need an additional condition to exclude this possibility.

The proof of Theorem 1 is based on five lemmas.

Lemma 2. $q(n+1, L) - q(n, L) \geq n/2$.

Proof. Let $q(n, L) = n_1 n_2 + \text{ex}(n_1, L) + \text{ex}(n_2, L)$, where $n_1 \leq n_2$. Then

$$q(n+1, L) \geq (n_1+1)n_2 + \text{ex}(n_1+1, L) + \text{ex}(n_2, L) \geq q(n, L) + n/2.$$

The next lemma is an immediate consequence of Lemma 2 and the straightforward Lemma V.3.2 of [1].

Lemma 3. *Given $c_1 > 0$ there exists $c_2 > 0$ such that if $e(G^n) > q(n, L)$ then G^n contains a subgraph G^p satisfying $p \geq c_2 n$, $e(G^p) > q(p, L)$ and $\delta(G^p) > (\frac{1}{2} - c_1)p$.*

Lemma 4. *There exists a constant $c_L > 0$ such that if $\delta(G^n) \geq (\frac{1}{2} - \frac{1}{100})n$ and $K = K_3(9r, 9r, 9r) \subset G^n$, where $r = |L|$, then G^n contains an $L + E^t$ with $t \geq c_L n$.*

Proof. Put $H = G^n - K$. Since at least $27r \cdot \frac{49}{100}n - (27\,r)^2$ edges join K to H, at least $\frac{1}{20}n$ vertices of H are joined to at least $11r$ vertices of K. Let $c_L = \frac{1}{20} \cdot 2^{-27r}$. Then H contains $t \geq c_L n$ vertices that are joined to the same set of at least $11r$ vertices of K. The subgraph of $K = K_3(9r, 9r, 9r)$ spanned by this set of vertices contains a $K_2(r, r)$ so $K_2(r, r) + E^t \subset G^n$. Since $L \subset K_2(r, r)$ we have $L + E^t \subset G^n$. □

The first part of the next lemma is a weak form of Theorem V.2.2 in [1], the second part is an immediate consequence of the first part.

Lemma 5. (i) *If $G = G_2(m, n)$ does not contain a $K_2(s, t)$ whose first class is in the first class of G then*

$$e(G) < t^{1/s} m n^{1-1/s} + sn.$$

(ii) *Given d and R, there exist $\varepsilon > 0$ and n_0 such that if $n \geq n_0$ and if in $G = G_d(n, n, \ldots, n)$ at least $(1-\varepsilon)n^2$ edges join any two classes then G contains a $K_d(R, R, \ldots, R)$.*

The last lemma needed in the proof of Theorem 1 is a slight extension of some results proved by Erdös and Simonovits [5, 6, 10].

Lemma 6. *Given c, $0 < c < 1$, and natural numbers d and R, there exist $M = M(c, d, R)$, $\delta = \delta(c, d, R) > 0$, and $n_0 = n(c, d, R)$ such that if $n > n_0$, $e(G^n) > (1 - 1/d - \delta)\frac{1}{2}n^2$ and $K_{d+1}(R, \ldots, R) \not\subset G^n$, then the vertices of G^n can be divided into d classes, say $A_1, A_2, \ldots, A_d$, such that the following conditions are satisfied.*

(i) $|n_i - n/d| < cn$, *where* $n_i = |A_i|$.

(ii) *The subgraph $G_i = G^n[A_i]$ of G^n, spanned by A_i, satisfies*

$$e(G_i) < cn^2$$

(iii) *Call a pair $\{x, y\}$ of vertices a missing edge if x and y do not belong to the same class A_i and xy is not an edge of G^n. The number of missing edges is less than cn^2.*

(iv) *Let B_i be the set of vertices in A_i joined to at least cn vertices of the same class A_i. Then $|B_i|<M$.*

Proof. Let $M_0=R$ and choose natural numbers $M_1<M_2<\ldots<M_d$ such that $M_i/M_{i+1}<\frac{1}{2}$. Put $M=M_d$. Pick η such that $0<\eta<(\frac{1}{2}c)^M$.

By Lemma 5 (ii) we can choose $\varepsilon, 0<\varepsilon<c$, and n_1 such that if $N=[\eta n]$, $n\geq n_1$ and in $H=G_d(N, N, \ldots, N)$ at most εn^2 edges are missing between any two classes then H contains a $K_d(R, R, \ldots, R)$.

The above mentioned theorem of Erdös and Simonovits (see Theorem V.4.2 in [1]) implies that there exist $n_0\geq n_1$ and $\delta>0$ with the following properties. If G^n is as in our lemma and $A_1, A_2, \ldots, A_d$ is a partition with the minimal number of missing edges (cf. condition (iii) of the lemma) then (i), (ii) and (iii) hold.

Suppose (iv) fails, say $|B_1|\geq M$. Then by the minimality of the partition each vertex of B_1 is joined to at least cn vertices in each A_i. Since

$$M_{i+1}cn>(\eta n)^{1/M_i}M_{i+1}n^{1-1/M_i}+M_i n,$$

repeated applications of Lemma 5 (i) imply that there are sets $\tilde{B}\subset B_1$, $\tilde{A}_i\subset A_i$, $i=1, 2, \ldots, d$, such that $|\tilde{B}|=R$, $|\tilde{A}_i|=N$ and each vertex of B is joined to each vertex of $A=\bigcup_1^d \tilde{A}_i$. Now it follows from (iii) and the choice of ε that $G[A]$ contains a $K_d(R, R, \ldots, R)$. Hence $G[A\cup B]$ contains a $K_{d+1}(R, R, \ldots, R)$. □

Proof of Theorem 1. It is easy to see that if G, H are graphs containing no L and no $K_2(1, cm/2)$, then $G+H$ contains no $L+E^{[cm]}$. Hence the second assertion of Theorem 1 is trivial. To prove the first assertion assume indirectly that G^n contains no $L+E^{[cn]}$ and $e(G^n)>q(n, L)$. We shall show that this is impossible if $c>0$ is sufficiently small. By Lemma 3 and Lemma 4 we may and will assume that $\delta(G^n)\geq(\frac{1}{2}-\frac{1}{10}r^{-1})n$, $R=gr$, and $G^n\not\supset K_3(9r, 9r, 9r)$. Applying Lemma 6 with $d=2$, we obtain a partition (A_1, A_2), satisfying (i)–(iv) of Lemma 6. For the sake of convenience in the sequel a subset H of the vertices of G^n and the corresponding spanned subgraph may be denoted by the same letter. Clearly, if m is the number of missing edges, then

$$e(G^n)=e(A_1)+e(A_2)+n_1n_2-m, \tag{5}$$

where $n_i=|A_i|$. Trivially, if neither A_1 nor A_2 contain L, then

$$e(G^n)\leq \mathrm{ex}(n_1, L)+\mathrm{ex}(n_2, L)+n_1n_2\leq q(n, L).$$

Thus $L\subset A_1$ may be assumed.

Let us assume that A_1-B_1 contains a subgraph L_0 isomorphic to L. To each $x\in L_0$ we find $\frac{1}{2}n(1-\frac{1}{2}r^{-1})$ or more vertices in A_2-B_2 joined to this x: since x is joined to cn or less vertices of A_1, it is joined to at least $(\frac{1}{2}-\frac{1}{10}r^{-1})n-\frac{1}{10}nr^{-1}$ vertices of A_2. Thus at least $n_2-r\cdot\frac{1}{4}nr^{-1}>\frac{1}{5}n$ vertices of A_2-B_2 are completely joined to L_0, yielding an $L+E^t$ for $t=[\frac{1}{5}n]$. This proves that $L\not\subset A_i-B_i$. Hence

$$e(A_i)\leq e(A_i-B_i)+|B_i|\,n_i\leq \mathrm{ex}(n_i, L)+|B_1|\,n_i.$$

Now we fix some constants and give the basic ideas of the proof. The details are given afterwards.

We fix a constant T such that $|B_i|\leq T\ (i=1,2)$. Lemma 6 guarantees the existence of such a T. A constant $c_L>0$ is fixed so that for $a_i=r^{2T-i}c_L$ we have $a_1<(10T)^{-2}$. Given a set $W\subset A_i$, denote by $F(W)$ the set of vertices of A_{3-i} not joined to at least one vertex of W. Observe that if W has at least $c_L n$ vertices, then $F(W)$ represents each $L\subset A_{3-i}$, for otherwise there would be an L completely joined to W and therefore $L+E^{c_L n}\subset G^n$ and we are home. Thus we may assume that $F(W)$ represents all the L's in A_{3-i}. Let

$$k_i=\frac{1}{n_i}\,(e(A_i)-\operatorname{ex}(n_i, L)).$$

Clearly, to represent all the L's in A_i we need vertices, the omission of which diminishes $e(A_i)$ by at least $e(A_i)-\operatorname{ex}(n_i, L)$, hence we have to omit at least k_i vertices: $F(W)$ has at least k_i vertices. This is the basic idea of the proof, but this in itself will not be enough. We shall prove the existence of a set Q_i of O(1) vertices in A_i such that the number of missing edges incident with this Q_i is at least $k_i n_i+\frac{1}{25}nT^{-1}$ if $L\subset A_i$. We have already checked the case, when no L occurs in A_1 and A_2. Let us consider the case, when $A_1\supset L$ but $A_2\not\supset L$. By (5) we have

$$e(G^n)\leq \operatorname{ex}(n_1, L)+k_1n_1+\operatorname{ex}(n_2, L)+n_1n_2-\left(k_1n_1+\frac{n}{25T}\right)<q(n, L).$$

If $A_1\supset L$, $A_2\supset L$, then the number of missing edges is estimated by the sum of the missing edges incident with Q_1 and Q_2 minus the number of missing edges between Q_1 and Q_2, which is only O(1). Hence

$$e(G^n)\leq \sum_i\left(\operatorname{ex}(n_i, L)+k_in_i-\left(k_in_i+\frac{n}{25T}\right)\right)+\mathrm{O}(1)<q(n, L).$$

This completes the sketch of the proof.

Let us see now how the argument above can be made precise. Recall that $L\subset A_1$. Let $L_1,\ldots,L_p,\ldots$ be subgraphs of A_1 isomorphic to L. For any $W=W_1\subset A_2$ and $\tilde{W}\subset W_1$, $|W_1|\geq a_1n$, $|\tilde{W}|\geq c_Ln$, $F(\tilde{W})$ represents all the L_p's, among them L_1, hence for at least a_1n-c_Ln vertices of W_1 there exists a vertex in L_1 not joined to it. Hence there exists an $x_1\in L_1$ and a $W_2\subset W_1$, $|W_2|\geq a_2n$, such that x_1 is not joined to W_2 at all. If x_1 does not represent all the L_p's, we may assume that $x_1\notin L_2$. Iterating this argument we find an $x_2\in L_2$ not joined to a $W_3\subset W_2$ at all, where $|W_3|\geq a_3n$, and if x_1, x_2 do not represent all the L_p's, we define x_3 and W_4 in the same way.

Generally, if x_p and W_{p+1} are already defined, we check whether the set $X_p=\{x_1,\ldots,x_p\}$ represents all the $L\subset A_1$. If it does or if $p=2T$, the procedure stops, otherwise we find an L_{p+1} and an x_{p+1} in it and a

$$W_{p+2}\subset W_{p+1},\quad |W_{p+2}|\geq a_{p+2}n$$

so that x_{p+1} is not joined to W_{p+2} at all. At the end of the procedure we have an $X = X_p$ and a $W' = W_{p+1}$ not joined at all to each other.

Let $B' \subset B_1$ be the class of vertices of degree at least $\frac{1}{2}n - \frac{1}{10}nT^{-1}$ in A_1. Let D be the set of vertices of A_2 joined to B' completely. D is relatively large. Indeed, by the minimum property of the partition (A_1, A_2) any $x \in B'$ is joined to at least $n(\frac{1}{2} - \frac{1}{10}T^{-1})$ vertices of A_2, hence at least $n(\frac{1}{2} + \varepsilon) - Tn(\frac{1}{10}T^{-1} + \varepsilon)$ vertices of A_i are joined to B' completely. Thus $|D| \geq \frac{1}{3}n$ if $\varepsilon \leq \frac{1}{20}T^{-1}$.

Now we define another procedure, in each step of which the above procedure is applied to a set $W_j \subset A_2$ yielding a pair of sets W'_j and X_j not joined to each other at all: $|W_j| \geq a_1 n$, $|W'_j| \geq c_L n$, $|X_j| \leq 2T$. Let

$$W_1 = D \quad (|D| \geq a_1 n),$$

$$W_j = D - \bigcup_{i<j} W'_i \quad \text{until} \quad |W_j| < a_1 n,$$

then

$$W_j = A_2 - \bigcup_{i<j} W'_i.$$

The corresponding sets in A_1 are $X_1, \ldots, X_j$. The procedure stops if for $W_j = A_2 - \bigcup_{i<j} W'_i$ we have $|W_j| < a_1 n$. By $|W'_1| \geq c_L n$ this will happen for some $j \leq c_L^{-1}$. Let $X = \bigcup X_j$. Clearly, $|X| \leq 2T/c_L = O(1)$. We shall show that there exist at least $k_1 n_1 + \frac{1}{25}nT^{-1}$ missing edges joining X to A_2. This will complete the proof.

We need a lower bound for the number of missing edges joining a W'_i to X: this lower bound is $|X_i|$. By the definition of D, if $W_i \subset D$, then each vertex of $X_i \subset A_1 - B'$ has degree $\leq \frac{1}{2}n - \frac{1}{10}nT^{-1}$ in A_1. These vertices represent all the L's in A_1, hence they represent at least $k_1 n_1$ edges:

$$|X_i| \geq \frac{k_1 n_1}{\frac{1}{2}n - \dfrac{n}{10T}} \geq k_1 + \frac{1}{6T},$$

if n is sufficiently large, ε sufficiently small and $k_1 \geq \frac{5}{6}$. If $k_1 \leq \frac{5}{6}$, we use $|X_i| \geq 1$, $|T| \geq 1$ (which can be assumed). Thus

$$|X_i| \geq k_1 + \frac{1}{6T}$$

again. In the other case, when $W_i \not\subset D$, we use a weaker lower bound. Since $|W'_i| \geq c_L n$ and no $x \in X_i$ is joined to W'_i, such an x is joined to at most $n_2 - c_L n$ vertices of A_2, and consequently, to at most $n_2 - c_L n = \frac{1}{2}n - c_L n + \varepsilon n$ vertices of A_1, we obtain now that

$$|X_i| \geq k_1 + k_1 c_L \geq k_1.$$

Thus the number of missing edges incident to X can be estimated from below by

$$\frac{n}{3}\left(k_1+\frac{1}{6T}\right)+\frac{n}{6}\cdot k_1-(a_1+\varepsilon)n\left(k_1+\frac{1}{6T}\right).$$

Here the third term stands for the vertices not belonging to any W_i' and for the difference between n_2 and $\frac{1}{2}n$. If ε is sufficiently small, T large, then by $k_1 \leq T$ and $a_1 \leq (10T)^{-2}$ we obtain that at least $k_1n_1+\frac{1}{25}nT^{-1}$ missing edges are between X and A_2. Thus the proof is complete. □

Remark. Theorem 1 can be generalized to higher chromatic numbers, that is, an analogous theorem holds for $L+K_{d-1}(r,\ldots,r,cn)$. The proof of this generalization is essentially the same as for the particular case considered above.

Our second theorem concerns $\mathrm{ex}(n, K^3, K_2(r,[cn]))$ and, more generally, $\mathrm{ex}(n, C^{2j+1}(t), K_2(r,[cn]))$. An interesting feature of the result is that the value does not really depend on j and t.

Theorem 7. *Let j, r, t be natural numbers, let $k=2j+1$ and let $c>0$. If $e(G^n)\geq cn^2$ and G^n does not contain a $C^k(t)$, then G^n contains a $K_2(r,m)$, where*

$$m=2^{2r-1}c^rn+o(n).$$

Proof. We shall show first that if instead of a $C^k(t)$ (and so a fortiori a C^k) we prohibit *all* odd cycles, then G^n contains a $K_2(r,m)$ with

$$m=2^{2r-1}c^rn+o(n),$$

but if $\varepsilon>0$ then G^n need not contain a $K_2(r,m')$ with

$$m'=(2^{2r-1}c^r+\varepsilon)n+o(n).$$

(This will show that the value of m given in the theorem is as large as possible and that the main thrust of the theorem is that the condition "G^n is bipartite" can be replaced by the much weaker condition "G^n does not contain a $C^k(t)$" without decreasing the value of m we can guarantee.)

The first assertion is an immediate consequence of Lemma 5. Instead of the second we prove the following stronger assertion.

Let n be even and let G^n be a random subgraph of $K_2(\frac{1}{2}n,\frac{1}{2}n)$ obtained by taking an edge of $K_2(\frac{1}{2}n,\frac{1}{2}n)$ with probability $4c$. Then, with probability tending to 1, G^n has $cn^2+o(n^2)$ edges and if $t=t(G^n)$ is the maximal number for which G^n contains a $K_2(r,t)$ then, again with probability tending to 1, we have

$$t=2^{2r-1}c^rn+o(n).$$

In order to prove this assertion, we denote by A and B the two classes of $K_2(\frac{1}{2}n,\frac{1}{2}n)$. We say that a vertex $x\in B$ forms a *cap* with a set U if $U\subset A$, $|U|=r$

and x is joined to every vertex in U. The expected number of vertices forming a cap with a given r-set in A is $2^{2r-1}c^r n$ and the variance of the event that an $x \in B$ forms a cap with U is $d_0^2 = 4c(1-4c)$. By the well known Bernstein inequality for binomial distributions (see p. 387 in [9]) the probability that U is joined to more than $2^{2r-1}c^r n + n^{2/3}$ or to less than $2^{2r-1}c^r n - n^{2/3}$ vertices $x \in B$ completely is $O(\exp(-c_1 n^{1/3}))$. Hence with probability tending to 1 on each U there is a $K_2(r, t)$ for $t = 2^{2r-1}c^r n - n^{2/3}$ but on no U for $t = 2^{2r-1}c^r n + n^{2/3}$, since

$$\binom{n}{r} \cdot O(\exp(-c_1 n^{1/3})) = o(1).$$

A similar application of Bernstein's inequality yields that $|e(G^n) - cn^2| \leq n^{5/3}$ with probability tending to 1.

Exactly the same argument gives that if G^n is a random subgraph of K_n of size $[cn]^2$ (or is obtained from K_n by choosing each edge with probability $2c$) then G^n will contain a $K_2(r, t)$ for $t = 2^r c^r n - n^{2/3}$ with probability tending to one, but for $t = 2^r c^r n + n^{2/3}$ only with probability tending to 0. This shows that prohibiting the odd cycles results in an increase of the constant from $2^r c^r$ to $2^{2r-1} c^r$ and that the main point of our theorem is that the same result can be obtained by prohibiting just one odd cycle.

The proof of our theorem is based on the following result of Szemerédi [11].

Lemma 8 (Uniform Density Lemma). *Given two subsets U, V of the vertex set of a graph G^n, denote by $e(U, V)$ the number of edges joining U to V and put*

$$d(U, V) = \frac{e(U, V)}{|U||V|}.$$

There exists for a given constant $\beta > 0$ an integer $M(\beta)$ such that for any G^n the vertices of G^n can be divided into disjoint classes $V_0, \ldots, V_k$ for some $k < M(\beta)$ so that $|V_i| = |V_j|$ if $i \neq 0$, $j \neq 0$, $|V_i| \leq \beta n$ if $i = 0, 1, \ldots, k$ and for all but βk^2 pairs (i, j) the following condition holds.

(*) *Whenever $U_i \subset V_i$, $U_j \subset V_j$ and $|U_i| > \beta |V_i|, |U_j| > \beta |V_j|$, then*

$$|d(U_i, U_j) - d(V_i, V_j)| < \beta^2.$$

Let us turn now to the main body of the proof of Theorem 7.

(A) Let $e(G^n) = cn^2$ and let $\beta > 0$ be an arbitrarily small constant, much smaller than c. Applying the Uniform Density Lemma to G^n we obtain the classes $V_0, V_1, \ldots, V_k$. Let $m = |V_i|$ $(i = 1, \ldots, k)$. Instead of G^n we consider a graph G' of $n - |V_0|$ vertices, obtained from $G^n - V_0$ by omitting all the edges

(i) joining vertices from the same V_i $(i = 1, \ldots, k)$;
(ii) joining a V_i to a V_j for an "exceptional pair", that is, (*) does not hold;
(iii) joining a V_i to a V_j, when $d(V_i, V_j) < \beta^{1/2}$.

Clearly,

$$n-\beta n \leq |G'| \leq n,$$

and

$$0 \leq e(G^n) - e(G') \leq 2\beta^{1/2} n^2$$

If β is sufficiently small. Therefore instead of proving Theorem 7 for G^n it is sufficient to prove it for G'. Hence we may and shall assume that $G' = G^n$.

(B) Let R^k be the graph whose vertices are the classes V_i ($i = 1, \ldots, k$) and V_i is joined to V_j in R^k if there exists an edge (u, v) in G^n joining V_i to V_j. We prove that R^k does not contain a triangle (K^3). Let us assume that V_1, V_2 and V_3 from a triangle in R^k. Put

$$U^+ = \{x \in V_3\colon\ d(x, V_1) \leq \beta\},$$
$$U^{++} = \{x \in V_3\colon\ d(x, V_2) \leq \beta\}.$$

For any $x \in U = V_3 - U^+ - U^{++}$ there exist a $U_{1,x}$ and a $U_{2,x}$ in V_1 and V_2 respectively, joined to x completely, where $|U_{i,x}| \geq \beta m$. Hence the number of edges joining $U_{1,x}$ to $U_{2,x}$ is at least

$$(\beta m)^2(\beta^{1/2} - \beta) > \beta^3 m^2.$$

This is a lower bound on the number of triangles on x, with the other two vertices in V_1 and V_2. Hence the total number of triangles (K^3's) of form (x, y, z), $x \in V_3$, $y \in V_1$, $z \in V_2$ is at least $(1-2\beta)\beta^3 m^3$: by (*) $|U^+| \leq \beta m$, $|U^{++}| \leq \beta m$. A theorem of Erdös [4] asserts, that if in an r-uniform hypergraph H of n vertices there are at least $cn^{r-(r-1)/t}$ hyperedges, then H contains a subgraph of the following form: $C_1, \ldots, C_r$ are vertex-disjoint t-tuples and we take all the r-tuples (= hyperedges) of form $(x_1, \ldots, x_r)$, $x_i \in C_i$ for $i = 1, \ldots, r$. Applying this theorem to the system of K^3's obtained above we get a $C_i \subset V_i$ ($i = 1, 2, 3$) with $|C_i| = t$ and such that each K^3 of the form (x, y, z), $x \in C_3$, $y \in C_1$, $z \in C_2$ belongs to G^n. Thus $K_3(t, t, t) = C^3(t) \subset G^n$. This contradiction proves the assertion of (B) for $k = 3$. In the general case we apply the theorem with kt instead of t and observe that $K_3(kt, kt, kt) \supset C^k(t)$, again completing the proof of (B).

(C) Now we fix a $c_1 \in (0, c)$ and assume indirectly that

$$e(G^n) = cn^2, \qquad G^n \not\supset C^k(t) \quad \text{and} \quad G^n \not\supset K_2(2^{2r-1} c_1^r n, r).$$

Let $d_i = d(V_i, V - V_i)$, where V is the vertex set of G^n. We may assume that $d_1 = \max d_i = d$. Let us permute the indices of V_i so that $V_2, \ldots, V_{s+1}$ are the classes joined to V_1, the others are independent of it. Clearly, $V_2, \ldots, V_{s+1}$ form a set of ms independent vertices. Hence

$$e(G^n) \leq \sum_{i=s+2}^{k} (d_i n)m + (d_1 n)m \leq (dn)(n-a). \tag{6}$$

where

$$a=\left|\bigcup_{1\le i\le s+1} V_i\right|=(s+1)m.$$

To obtain an upper bound of d in terms of a, we apply Lemma 5 to the bipartite graph determined by the classes $\bigcup_{2\le i\le s+1} V_i$ (= first class) and V_1 (= second class). We find that

$$G^n \supset K_2(r,t) \quad \text{with } t=(1-o(1))d^r n^r a^{-(r-1)}. \tag{7}$$

By the assumption $G^n \not\supset K_2(r, 2^{2r-1}c_1^r n)$ and by (7)

$$d^r n^{r-1} a^{-(r-1)} \le (1+o(1))2^{2r-1}c_1^r. \tag{8}$$

Let us assume that $d>2c_1$ (this will be shown later). From (8) and $c_1^r<\frac{1}{2}dc_1^{r-1}$ we obtain $d<(1+o(1))\,4c_1(a/n)$. This and (6) yield

$$cn^2\le e(G^n)\le dn(n-a)\le(1+o(1))n^2\cdot 4c_1\frac{a}{n}\left(1-\frac{a}{n}\right)\le c_1n^2,$$

which is a contradiction.

To prove $d>2c_1$ observe that "essentially, dn is the maximum degree":

$$cn^2\le e(G^n)=\frac{1}{2}\sum_i d(V_i, V-V_i)m(n-m)\le km\cdot d(n-m)=dn(n-m). \tag{9}$$

Until now β and c_1 were independent, now we may agree that β is chosen depending on c_1 and it is so small that $1-\beta>(c_1/c)$. This, $(m/n)<\beta$ and (9) yield the desired inequality $d>2c_1$. □

Remark. The method used to prove Lemma 6 and the method used to prove that K^3 does not occur in the graph R^k are equivalent: both can be used in both cases. The proof becomes slightly shorter if we consider only the case $t=1$.

Theorem 9. *Let t be a natural number and let $c>0$. Then there exists an n_0 such that if $n>n_0$ and $G^n \not\subset C^m(t)$ for $m=3,5,\ldots,2l(c)+1$, where $2l(c)+1>c^{-1}$, then G^n can be made bipartite by the omission of not more than cn^2 edges.*

Remark. Theorem 9 is sharp, apart from the value of $l(c)$ which is probably $O(c^{-1/2})$. This $l(c)=O(c^{-1/2})$ would be sharp if true. To see this put $n=(2l+3)$, and $G^n=C^{2l+3}(m)$. If $c=(2l+4)^{-2}$, then more than cn^2 edges must be omitted to turn G^n into a bipartite graph and C^n does not contain C^k if k is odd and smaller than $2l+3$.

Proof. Our proof consists of two parts. We shall give two versions of the second part.

Part I. Let $c'<c$ be fixed. We shall say that the edges are regularly distributed, if for every partition $V(G^n)=A\cup B$ we have $d(A,B)\geq 2c'$. If we have an arbitrary G^n, we shall find a G^m in it, in which the edges are regularly distributed and $h(G^m)\geq cm^2$, $m>n_0'$ also hold, where $h(G)$ denotes the minimum number of edges one has to omit to change G into a bipartite graph. Therefore it will be enough to prove the theorem for the case, when the edges are regularly distributed and this will be just Part II. Let us assume that the edges are not regularly distributed in G^n; $V(G^n)=A\cup B$ and $d(A,B)<2c'$. Clearly,

$$h(G^n)<h(G[A])+h(G[B])+d(A,B)\,|A|\,|B|,$$

therefore we may assume that

$$h(G[A])>c\,|A|^2+(c-c')\,|A|\,|B|. \tag{10}$$

Hence

$$\tfrac{1}{4}\,|A|^2>(c-c')\,|A|\,|B|,$$

that is, $|A|>c''\,|B|$ for $c''=4(c-c')$. This also shows that $|A|>(1+c'')^{-1}n$. Furthermore, by (10),

$$h(G[A])>c_1\,|A|^2 \quad \text{for } c_1=c+(c-c')\frac{|B|}{|A|}.$$

Put $G_0=G^n$, $c_0=c$, $G_1=G[A]$ and repeat the step above until either we arrive at a G_i in which the edges are regularly distributed or to a G_i with $\sqrt{n}$ or less vertices (and use always $c_i'=c_i-(c-c')$). It is easy to show that if n is sufficiently large, then G_i cannot go below $\sqrt{n}$, otherwise $c_i>1$ would occur. Hence the procedure will always stop with a graph G_i in which the edges are regularly distributed. This was to be proved.

Part II (First version). (A) We start with a graph G^n for which $h(G^n)\geq cn^2$, fix a $c''<c$ and then a $c'\in(c'',c)$ and a $\beta>0$, which is much smaller than c''. Using the first part we may assume that the edges are regularly distributed. We may repeat part (A) of the proof of Theorem 7 replacing $e(\ )$ by $h(\ )$ and c by c'. Then we may assume that $G'=G^n$, but have to decrease c': replace the original condition by condition $h(G^n)\geq c''n^2$. How we define the graph R^k as in the beginning of (B) of the proof of Theorem 7.

(B) We prove that if n is sufficiently large and $R^k\supset C^j$, then $G^n\supset C^j(t)$, where t is fixed, but arbitrarily large. Exactly as in the proof of Theorem 7, we can prove that G^n contains at least c_1n^j cycles C^j, where $c_1>0$ is a constant. Applying the theorem of Erdös on hypergraphs [4], we obtain j sets $X_1,\dots,X_j$ with $|X_j|=T\to\infty$, such that if $x_1\in X_1,\dots,x_j\in X_j$, then some permutation $(x_{i_1},\dots,x_{i_j})$ is a cycle of G^n (we consider here the hypergraph whose hyperedges are the j-sets of vertices of j-cycles in G^n). Unfortunately the cycles will not determine a $C^j(T)$, since the permutation $i_1,\dots,i_j$ may differ from j-tuple to j-tuple. However, let us apply the Erdös theorem again, now to the hypergraph whose vertices are in $X_1 \cup X_2 \cup \ldots \cup X_j$ and the hyperedges of which are some cycles of G^n of form $(x_{i_1}, \ldots, x_{i_j})$, $x_s \in X_s$, where we choose only one permutation $i_1, \ldots, i_j$, for which the number of cycles is at least $T^j/j!$. If T is large enough, we obtain j subsets $Y_i \subset X_i$ such that whenever $x_i \in Y_i$, then $(x_{i_1}, \ldots, x_{i_j})$ defines a cycle in G^n and $|Y_i| = t$. Thus we obtained a $C^j(t) \subset G^n$.

(C) Clearly, the only thing to prove is, that $R^k \supset C^{2s+1}$ for some $2s+1 \leqslant (c')^{-1}$. If e.g. $V_1, \ldots, V_j$ define a shortest odd cycle in R^k, by the assumption that the edges are regularly distributed in G^n, there must be a V_q, $q > j$, which is joined to at least $2c'j$ of the classes $V_1, \ldots, V_j$. If V_q is joined to a V_i and $V_{i'}$ for some i' farther from i than 2, then the arc $V_i V_q V_{i'}$ will create a shorter odd cycle. Hence either $C^3 \subset R^k$ or $2c''j \leqslant 2$, and, consequently, $j \leqslant (c')^{-1}$.

Part II (Second version). The difference between the two proofs is above all, that here we shall not use the Uniform Density Lemma.

(A) By the first part we may assume that the edges are regularly distributed. Let A_1 be an arbitrary set of $\sqrt{n}$ vertices. By $d(A_1, V - A_1) \geqslant c'$ (where V is the vertex set of G^n) and by Lemma 5 we can find a $B_1 \subset A_1$, for which $|B_1| = T = t^2$, and a set $B_2 \subset V - A_1$ for which $|B_2| \geqslant bn$ with $b = (c')^T$, so that B_1 and B_2 are completely joined. B_j is recursively defined:

$$\tilde{B}_j = \left\{ x \colon x \notin \bigcup_{i<j} B_i \quad \text{and} \quad d\left(x, \bigcup_{i<j} B_i\right) > c' \right\},$$

$$B_j = \tilde{B}_j - \bigcup_{i<j} B_i.$$

Clearly,

$$|B_j| \geqslant c' \left| V - \bigcup_{i<j} B_i \right|.$$

Hence for any fixed $\beta > 0$ we can find a $l_0 = l_0(c, \beta)$ such that

$$\left| V - \bigcup_{i<j} B_i \right| < \beta n \quad \text{if} \quad j \geqslant l_0.$$

Omitting all the edges between $\bigcup_{i \leqslant l} B_1$, and the rest of the graph we omit at most βn^2 edges. If now we omit all the edges (x, y) for which $x \in B_i$, $y \in B_{i+2p}$ for some i and p then we change the graph into a bipartite one. Hence there exists a pair (i, p) for which at least $(c'n^2 - \beta n^2)/l_0^2$ edges were omitted between B_i and B_{i+2p}. Hence there exists a $K_2(T, T)$ joining B_i to B_{i+2p} in the sense that the first (second) class of it is contained in B_i (B_{i+2p}). Let these classes be denoted by D_i and E_{i+2p}, respectively. If D_j is already defined, D_{j-1} can also be defined as follows: $|D_j| = T$, we find t vertices in D_j and $2T$ vertices in B_{j-1} joined to each other completely. By Lemma 5 this can be done if n is sufficiently large. The class

D_{i-1} contains T of these $2T$ vertices, E_{i-1} is obtained from E_i in the same way, but here we have to choose the T vertices outside of D_{i-1}. Finally we obtain a $C^{2i+2p-1}(t)$ in G^n, whose classes are E_0 in $B_1, E_2, \ldots, E_{i+2p}, D_i, D_{i-1}, \ldots, D_2$ in this cyclic order. This proves the theorem, except for the upper bound on the length of the cycle, which is very similar to that of the first version. We only sketch it here; if we already know the existence of a $C^{2s+1}(t)$ for any t and $s \le l_1$, then we take a $C^{2s+1}(t^2)$ for some very large t and find t vertices outside joined to the same $c'(2s+1)t^2$ vertices of this subgraph. If t is sufficiently large, at least $c'(2s+1)$ classes are joined to each of the considered t vertices by t or more edges. Thus we can find a shorter $C^{2s'+1}(t)$ if $2s+1>(c')^{-1}$. □

Remark. With essentially the same effort we could prove the existence of a $C^{2s+1}(t, cn, t, cn, t, cn, \ldots, t, cn, t)$ instead of the existence of a $C^{2s+1}(t)$, where $C^k(m_1, \ldots, m_k)$ is the graph obtained from the cycle C^k by replacing its ith vertex by m_i new independent vertices. In other words, we can guarantee that every second class of our graph contains cn vertices.

References

[1] B. Bollobás, Extremal Graph Theory (Academic Press, London, 1978).
[2] B. Bollobás and P. Erdõs, On the structure of edge graphs, Bull. London Math. Soc. 5 (1973) 317–321.
[3] B. Bollobás, P. Erdös and M. Simonovits, On the structure of edge graphs II, J. London Math. Soc. 12 (2) (1976) 219–224.
[4] P. Erdös, On extremal problems of graphs and generalized graphs, Israel J. Math. 2 (1965) 183–190.
[5] P. Erdös, Some recent results on extremal graph problems in graph theory, in: Theory of Graphs, Int. Symp. Rome (1966) 118–123.
[6] P. Erdös, On some new inequalities concerning extremal properties of graphs, in: Theory of Graphs, Proc. Coll. Tihany, Hungary (1966) 77–81.
[7] P. Erdös and M. Simonovits, An extremal graph Problem, Acta Math. Acad. Sci. Hung. 22 (3–4) (1971) 275–282.
[8] P. Erdös and A.H. Stone, On the structure of linear graphs, Bull. Amer. Math. Soc. 52 (1946) 1087–1091.
[9] A. Rényi, Probability Theory (North-Holland, Amsterdam, 1970).
[10] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, in: Theory of Graphs, Proc. Coll. Tihany, Hungary (1966) 279–319.
[11] E. Szemerédi, Regular partitions of graphs (to appear).