Studies in Pure Mathematics
To the Memory of Paul Turán

Extremal graph problems and graph products

by

M. SIMONOVITS (Budapest)

Abstract

In this paper we consider only graphs without loops and multiple edges. The product of two vertex disjoint graphs G_1 and G_2 is the graph obtained by joining each vertex of G_1 to each vertex of G_2. Given n and the sample graphs $L_1, \ldots, L_\lambda$, we shall consider those graphs on n vertices which contain no L_i as a subgraph and have maximum number of edges under this condition. These graphs will be called extremal graphs for the L_i's.

In many cases the extremal graphs are products of other extremal graphs (for some other families of sample graphs). The aim of this paper is to investigate, when are the extremal graphs products and when are not.

Notations

The graphs considered in this paper are undirected, have no loops and no multiple edges. They will be denoted by capitals, and the superscript will always denote the number of vertices. Thus G^n, H^n, S^n will all denote graphs of n vertices.

The number of vertices, edges, and the chromatic number of a graph G will be denoted by $v(G)$, $e(G)$ and $\chi(G)$, respectively. If x is a vertex of G, st(x) denotes the star of x, i.e. the set of vertices joined to it; $d(x)$ will denote the degree of x.

To simplify the definitions of graphs we shall use the following operations.

(a) $G=\sum G_i$, if the G_i's are spanned subgraphs of G the pairwise disjoint vertex sets of which cover G and no vertices belonging to different G_i's are joined. (SUM).

(b) $G=\times\, G_i$, if the G_i's are spanned subgraphs of G the pairwise disjoint vertex sets of which cover G and vertices belonging to different G_i's are *always* joined. (PRODUCT).

(c) If G_1 is a subgraph of G or a set of vertices and edges of it, then $G-G_1$ is the graph resulting by deleting all the vertices, edges, and also the vertices incident with some deleted edges of G_1 from G.

$K_d(r_1, \ldots, r_d)$ denotes the complete d-partite graph with r_i vertices in its ith class.

Introduction

A classical result of P. TURÁN [6, 7] asserts that if p and n are given integers and S^n is a graph not containing K_p as a subgraph and having maximum number of edges under this condition, then $S^n = K_{p-1}(n_1, \dots, n_{p-1})$ where $n_1, \dots, n_{p-1}$ is the most uniform partition of n into $p-1$ summands:

$$n_i = \left[\frac{n}{p-1}\right] \quad \text{or} \quad n_i = \left[\frac{n}{p-1}\right] + 1, \quad \text{and} \quad n_1 + \dots + n_{p-1} = n.$$

To generalize the above theorem one can ask the following general problems.

Problem 1. Let $\mathscr{L}$ be a given finite or infinite family of graphs and let $\mathscr{A}(n, \mathscr{L})$ denote the class of graphs on n vertices not containing any $L \in \mathscr{L}$ as a subgraph. What is the maximum number of edges a graph $G^n \in \mathscr{A}(n, \mathscr{L})$ can have.

(The graphs of $\mathscr{L}$ will be called sample graphs, the graphs attaining the maximum will be called extremal graphs, the maximum will be denoted by $\mathrm{ex}(n, \mathscr{L})$ and the class of extremal graphs will be denoted by $\mathrm{EX}(n, \mathscr{L})$.)

Problem 2. Describe the structure of the extremal graphs.

Some general results obtained by P. ERDŐS [1, 2] and the author [4] independently, give a fairly good description of the extremal graphs. Thus e.g. we have proved that

Theorem A. *For a given $\mathscr{L}$ let*

$$d = \min \{\chi(L) : L \in \mathscr{L}\} - 1. \tag{1}$$

There exists a $c > 0$ such that if S^n is an extremal graph for $\mathscr{L}$, then S^n can be obtained from a $K_d(n_1, \dots, n_d)$ by deleting from and adding to it $O(n^{2-c})$ edges. Further, $n_i = \frac{n}{d} + O(n^{1-c})$, $i = 1, 2, \dots, d$.

Corollary. *Under the conditions of Theorem A S^n can be obtained from some appropriate graphs $G_1, \dots, G_d$ by deleting $O(n^{2-c})$ edges from $\times_{i \le d} G_i$.*

Remark. 1. The basic content of Theorem A is that the extremal graphs depend only very loosely on $\mathscr{L}$, the minimum chromatic number determines their structure up to $O(n^{2-c})$ edges.

Problem 3. Under which condition is it true that $S^n = \times_{i \le d} G^{n_i}$ where $n_i = \frac{n}{d} + o(n)$?

Originally ERDŐS and I thought that whenever $\mathscr{L}$ is finite and n is sufficiently large, all the extremal graphs for $\mathscr{L}$ are products of graphs of almost equal size. Later we found some counter examples. However, we think that the following conjecture holds.

Conjecture 1. Let L be a $d+1$-chromatic graph which *cannot* be coloured by $d+1$ colours "1", "2", ..., "$d+1$" so that the subgraph $L_{1,2}$ spanned by the vertices of colours "1" and "2" is a tree or a forest. Then there exists an n_0 such that for any $n > n_0$, if S^n is an extremal graph for L, then $S^n = \underset{i \leq d}{\times} G^{n_i}$ where $n_i = \frac{n}{d} + o(n)$.

One can generalize Conjecture 1 to finite families of sample graphs as follows.

Definition 1. Let $\mathscr{L}$ be a given family of sample graphs. Let d be defined by (1). We say that M belongs to the *decomposition* family $\mathscr{M}$ of $\mathscr{L}$ if there exists an $L \in \mathscr{L}$ and an integer r for which

$$L \subseteq M \times K_{d-1}(r, \ldots, r). \tag{2}$$

Conjecture 2. Let $\mathscr{L}$ be a finite family of graphs and d be defined by (1). If the decomposition family $\mathscr{M}$ of $\mathscr{L}$ contains no trees or forests, then for any sufficiently large n each extremal graph S^n is a product: for some fixed integer t

$$S^n = S^{n_1} \times S^{n_2}, \quad \text{where} \quad n_1 = \frac{tn}{d} + o(n), \quad n_2 = n - n_1 .$$

Remark 2. One can ask, why to exclude the trees and forests in Conjectures 1 and 2. To motivate this we remark that

(a) as we shall see (Proposition 2 or Theorem 1 + Remark 3), Conjecture 2 does not hold if the decomposition is allowed to contain a path. This is, why we exclude the trees in Conjecture 2.

(b) It is known [2, 4], that the extremal graph S^n can be obtained from a $K_d(n_1, \ldots, n_d)$ $\left(\text{where } n_i = \frac{n}{d} + o(n)\right)$, by changing *only* $O(n)$ edges in it if and only if the decomposition contains a tree or a forest: in all the other cases we must alter at least cn^{1+a} edges in $K_d(n_1, \ldots, n_d)$, where $a > 0$ is a constant. There is a trivial, but very important difference between $f(n) = n$ and $f(n) = n^{1+a}$, namely, the latter one is *strictly* convex. Of course, this is only a heuristic motivation given in a very compact form.

It can be shown that if $\mathscr{L}$ contains more than one graph, the (stronger) assertion of Conjecture 1 does not necessarily hold.

Our assertions above are all trivial for $d = 1$. Hence we shall assume that $d \geq 2$. (The case $d = 1$ will be called *degenerate*.) The main idea of Conjectures 1 and 2 is to reduce the general case to the degenerate case as follows.

Proposition 1. *Let $\mathscr{L}$ be a finite family of sample graphs and $k = \max\{v(L) : L \in \mathscr{L}\}$. Let d be defined by* (1) *and $\mathscr{M}$ be the decomposition family of $\mathscr{L}$. If S^n is an extremal graph for $\mathscr{L}$ and*

$$S^n = \underset{i \leq d}{\times} G^{n_i}, \quad \textit{where} \quad n_i \geq k ,$$

then there exist d families of sample graphs, $\mathscr{M}_1, \ldots, \mathscr{M}_d$ for which

(a) $\max\{v(M): M \in \mathcal{M}\} \leqq k$.
(b) $\mathcal{M} \subseteq \mathcal{M}_i$ *and*

$$\min\{\chi(M): M \in \mathcal{M}_i\} = 2.$$

(c) *If* H_i *contains no* $M \in \mathcal{M}_i$ $(i = 1, \ldots, d)$, *then* $\underset{i \leqq d}{\times} H_i$ *contains no* $L \in \mathcal{L}$.
(d) G^{n_i} *is an extremal graph for* $\mathcal{M}_i$ $(i = 1, \ldots, d)$.

Proof. Let $\mathcal{M}_i$ be the family of graphs of at most k vertices not contained in G^{n_i}. Now (a) and (c)→(d) are trivial. If M is in the decomposition of $\mathcal{L}$ but $M \notin \mathcal{M}_i$, then there exists an $L \subseteq M \times K_{d-1}(k, \ldots, k)$ and by the definition of $\mathcal{M}_i$ $M \subseteq G^{n_i}$. Hence $M \times K_{d-1}(k, \ldots, k) \subseteq G^{n_i} \times \underset{j \neq i}{\times} G^{n_j} = S^n$, that is, $L \subseteq S^n$, which is a contradiction. This proves (b), since

$$\min\{\chi(M): M \in \mathcal{M}\} = 2$$

is obvious: we colour an appropriate $L \in \mathcal{L}$ by $d+1$ colours and denote by $L_{1,2}$ the subgraph spanned by the first two colours. Clearly, $L_{1,2}$ is bipartite and belongs to $\mathcal{M}$. To prove (c) observe that, since H_i contains no $M \in \mathcal{M}_i$, each subgraph of H_i of at most k vertices is also a subgraph of G^{n_i}. Thus each subgraph of $\underset{i \leqq d}{\times} H_i$ of at most k vertices is also a subgraph of $S^n = \underset{i \leqq d}{\times} G^{n_i}$. Thus it cannot belong to $\mathcal{L}$. This completes the proof.

Though the proof of Proposition 1 was fairly simple and straightforward, the proposition itself is worth some further explanation. Assume that Conjecture 1 holds. Then all the extremal problems satisfying the condition of Conjecture 1 can be reduced to degenerate extremal graph problems in the following sense:

Given a finite family $\mathcal{L}$ of sample graphs, the families $\mathcal{M}_1, \ldots, \mathcal{M}_d$ can be defined only in finitely many ways so that (a), (b) and (c) hold. Assume that we can solve the extremal problems corresponding to the degenerate families $\mathcal{M}_1, \ldots, \mathcal{M}_d$. If $H^{n_1}, \ldots, H^{n_d}$ are the corresponding extremal graphs, let $S^n = \times H^{n_i}$. Clearly,

$$e(S^n) = e(K_d(n_1, \ldots, n_d)) + \sum \operatorname{ex}(n_i, \mathcal{M}_i) = f(n_1, \ldots, n_d; \mathcal{M}_1, \ldots, \mathcal{M}_d). \tag{3}$$

At least in theory, we may find for each n and $\mathcal{M}_1, \ldots, \mathcal{M}_d$ the partition $n = n_1 + \ldots + n_d$ yielding the maximum in (3). Since there are only finitely many possible candidates for $\mathcal{M}_1, \ldots, \mathcal{M}_d$, we may find the one giving the highest maximum, and the corresponding S^n will be the extremal graph. In this sense we reduced the problem of $\mathcal{L}$ to the degenerate problems of $\mathcal{M}_1, \ldots, \mathcal{M}_d$.

This is, why Proposition 1 is important in theory. Another use of it is that in many cases we can guess the possible extremal graphs by assuming Conjecture 1, and finding the potentially possible sets $\mathcal{M}_1, \ldots, \mathcal{M}_d$, then the corresponding extremal graphs for $\mathcal{L}$. Knowing, which are the extremal graphs if Conjecture holds we can often prove that they are really extremal graphs, not using Conjecture 1 at all.

One can ask, whether Proposition 1 holds even if the decomposition family $\mathcal{M}$ contains a tree or a forest. The answer is that sometimes *yes* and sometimes *not*.

Proposition 2. *There exists a finite family $\mathcal{L}$ of sample graphs and an n_0 such that if $n > n_0$, then no extremal graph S^n (for $\mathcal{L}$) can be decomposed into the product of d nonempty graphs, where d is defined by* (1).

One way to prove Proposition 2 would be to show that for

$$\mathcal{L}^* = \{K_k(1, 3, 3), K_3 \times \bar{K}_3, (K_2 + K_2) \times \bar{K}_3, K_4)\}$$

Proposition 2 holds:

Let $\tilde{S}^n$ be obtained from $K_2(n_1, n_2)$, where $n_1 = \left[\frac{n}{2}\right]$ and $n_2 = n - n_1$, by adding two incident edges (x, y) and (y, z) and two further incident edges (x', y') and (y', z') to it and deleting (y', y), where x, y, z belong to the first class of $K_2(n_1, n_2)$ and x', y', z' to the other one. One can show that if n is sufficiently large, then this $\tilde{S}^n$ is the only extremal graph for $\mathcal{L}^*$ and $\tilde{S}^n$ cannot be decomposed into the product of two nonempty graphs. However, Proposition 2 will be derived as a consequence of a much deeper theorem, which could be called either an "inverse extremal graph theorem" or a compactness theorem.

An inverse extremal graph theorem

The aim of the next definition is to define a sequence of graphs which in some sense are very much alike and differ from each other only in size.

Definition 2. Let the graphs $A_1, \ldots, A_d$ and D be fixed and let also fix a subset B of the pairs $(x, y): x \in D, y \in \bigcup_{i \leq d} A_i$. Let us take an n for which $m_i = \dfrac{n - v(D)}{dv(A_i)}$ are all integers, take m_i vertex disjoint copies of A_i, denoted by $A_{i,j}$ and fix the isomorphisms $F_{i,j}: A_{i,j} \to A_i$. Let

$$Z^{n-v(D)} =: \mathop{\times}_{i=1}^{d} \sum_{j=1}^{m} A_{i,j}.$$

Let us join an $x \in D$ to a $y \in A_{i,j}$ iff $(x, F_{i,j}(y))$ belongs to B. Thus we obtain a graph S^n. A sequence

$$S^{n_1}, \ldots, S^{n_k}, \ldots$$

will be called a q-sequence if each S^{n_k} is obtained from the same $D, A_1, \ldots, A_d$ and B in the way described above. Sometimes we shall call D the head and $Z^{n-v(D)}$ the tail of the graph, respectively, though they are not uniquely defined by $\{S^{n_k}\}$.

Definition 3. Let $\{S^{n_k}\}$ be a q-sequence, obtained from $A_1, \ldots, A_d$, D and B. If for every q-sequence $\{T^{m_p}\}$ obtained from $A_1' \subseteq A_1, \ldots, A_d' \subseteq A_d$, $D' \subseteq S^{n_h}$ (for some fixed h) and $B' \subseteq B$,

$$e(S^{n_k}) \geqq e(T^{m_p}) \quad \text{if} \quad n_k = m_p, \quad k > k_0, \tag{4}$$

then $\{S^{n_k}\}$ will be called *dense*. If we have strict inequality in (4), then $\{S^{n_k}\}$ will be called *strictly dense*.

Theorem 1. *The following two assertions are equivalent:*

(i) $\{S^{n_0+km}\}$ *is a (strictly) dense q-sequence.*

(ii) *There exists a* **finite** $\mathcal{L}$ *such that* $\{S^{n_0+km}\}$ *is (the only) extremal graph for* $\mathcal{L}$ *and* $n = n_0 + km$, $(n > n_1)$.

Remark 3. One can easily find strictly dense q-sequences which are not products. Such a strictly dense sequence is e.g. $\tilde{S}^n$ defined after Proposition 2. By Theorem 1 these are sequences of extremal graphs. Thus Proposition 2 follows from Theorem 1. (To be quite precise, S^n is a q-sequence if n is even and another q-sequence if n is odd. Thus Theorem 1 yields two families $\mathcal{L}_{\text{even}}$ and $\mathcal{L}_{\text{odd}}$ and if $\mathcal{L} = \mathcal{L}_{\text{even}} \cup \mathcal{L}_{\text{odd}}$, then obviously S^n is the only extremal graph for $\mathcal{L}$ if n is sufficiently large.)

Remark 4. In [5] we proved that if $\mathcal{L}$ is a finite family of sample graphs and the decomposition family of $\mathcal{L}$ contains a path or a subgraph of a path, then there exists an integer t such that for every h there is a strictly dense q-sequence $\{S^h : n \equiv h \pmod t\}$ of extremal graphs for $\mathcal{L}$ if n is large enough. In other words, in every residue class mod t there is a sequence of extremal graphs of similar structure. Theorem 1 shows that this (main) theorem of [5] is sharp: each strictly dense q-sequence is an extremal sequence for some finite $\mathcal{L}$. In this sense Theorem 1 is an *inverse extremal graph theorem*. It is also an *inverse extremal graph theorem* in the following sense: we first fix the extremal graphs and then find the corresponding $\mathcal{L}$.

Remark 5. One part of Theorem 1, namely (ii)→(i) is trivial: let us fix an $\mathcal{L}$ satisfying (ii). If $\{T^{m_p}\}$ is a q-sequence obtained from a family $A_1' \subseteq A_1, \ldots, A_d' \subseteq A_d$, $D' \subseteq S^{n_h}$ and $B' \subseteq B$, then each T^{m_p} is a subgraph of an S^{n_k}. Thus T^{m_p} contains no $L \in \mathcal{L}$. Further, if $n_k = m_p$, then (4) holds (with strict inequality) since S^{n_k} is (the only) extremal graph for $\mathcal{L}$. Q.E.D.

Before turning to the proof of Theorem 1 we give some examples illustrating the notion of dense q-sequences.

Example 1. ERDŐS and RÉNYI called a graph A balanced if for every subgraph A'

$$\frac{e(A)}{v(A)} > \frac{e(A')}{v(A')}.$$

Let $A_1, \ldots, A_d$ be balanced and D arbitrary. If B is the whole direct product $\{(x, y) : x \in D, y \in \bigcup_{i \leq d} A_i\}$, then the corresponding q-sequence is strictly dense. (A tree, a complete graph, a complete bipartite graph, a cycle are all balanced. A balanced graph is always connected.)

Example 2. Let $A_1, \ldots, A_d$ be given graphs and A_1 be strictly unbalanced in the sense that it has a subgraph A' for which

$$e(A_1):v(A_1) < e(A'):v(A').$$

Let D be arbitrary and B be again the whole direct product. Then the corresponding q-sequence is *not* dense.

Example 3. Let $d=2$, $A_1 = A_2 = C^4$ be a four-cycle and Z^{8k} be obtained by taking the corresponding graph $G^{4k} \times G^{4k}$, where G^{4k} is the union of k disjoint C^4. Let $D = \{x\}$. The corresponding q-sequence S^{8k+1} is not dense if x is joined to one vertex of each C^4: we obtain a better sequence T^{8k+1} by omitting x and 3 further points from S^{8k+5}. If x is joined to 3 vertices of each C^4, then the obtained S^{8k+1} is a strictly dense q-sequence.

Proof of Theorem 1

Let us fix the sequence S^{n_k}. A graph G will be called "small" if it is contained in an S^{n_k}. By Remark 5 it is enough to prove (i)→(ii). We shall prove that if r is sufficiently large and

$$\mathcal{L}_r = \{L : L \text{ is not "small"}, v(L) \leq r\},$$

then S^{n_k} is (the only) extremal graph for $\mathcal{L}_r$ (for $n = n_k$).

We need the following two lemmas.

Lemma 1. *Let $\{S^{n_k}\}$ be a dense q-sequence. There exists an N_0 such that if $n_k > N_0$ and G^{n_k} is "small", then*

$$e(G^{n_k}) \leq e(S^{n_k}). \tag{5}$$

Further, if $\{S^{n_k}\}$ is strictly dense and the equality holds in (5), then $G^{n_k} = S^{n_k}$.

Lemma 2. *Given a dense q-sequence $\{S^{n_k}\}$, there exist two integers R and N_1 and a positive constant c such that if $n > N_1$, each vertex of G^n has valence $> \left(1 - \frac{1}{d} - c\right)n$ and each subgraph G^R of G^n is "small", then G^n is also "small".*

The proofs of these lemmas will be given in the next paragraph. Here we show, how to complete the proof of Theorem 1 assuming the lemmas.

First we remark that if R is sufficiently large, then $\mathscr{L}_R$ contains an L with $\chi(L)=$ $=d+1$. Indeed, if e.g. $b=v(D)+v(A_1)+\ldots+v(A_d)$ then K_{d+1} $(2bd, \ldots, 2bd)\notin S^{n_k}$ is trivial, thus K_{d+1} $(2bd, \ldots, 2bd)\in\mathscr{L}_R$ for $R\geqq 4bd^2$. On the other hand, each $K_d(p,\ldots,p)\subseteq S^{n_k}$ if $k\geqq k_0(d,p)$. Hence $\mathscr{L}_R$ contains no $\leqq d$-chromatic graphs:

$$\min\{\chi(L): L\in\mathscr{L}_R\}=d+1\,.$$

Hence, according to the main results of [1, 2] and [4], for $R\geqq 4bd^2$ if H^n is extremal for $\mathscr{L}_R$, then

$$e(H^n)=\mathrm{ex}\,(n,\mathscr{L}_R)=\frac{1}{2}\left(1-\frac{1}{d}+o(1)\right)n^2\,,$$

and each vertex of H^n is of valence $\geqq\left(1-\frac{1}{d}\right)n-o(n)$. Now we may apply Lemma 2 to H^n: there exist an $\tilde{R}$ and an N_1, further a $c>0$, such that if $N_1<n$ and the minimum degree of G^n exceeds $\left(1-\frac{1}{d}-c\right)n$ and each subgraph $G^{\tilde{R}}\subseteq G^n$ is "small", then G^n is also "small". Now we fix an $R=\max\{\tilde{R}, 4bd^2\}$ and get that if $n>N_2$, then the extremal graph H^n (for $\mathscr{L}_R$)

(a) contains no prohibited subgraphs $L\in\mathscr{L}_R$, hence each $G^{\tilde{R}}\subseteq H^n$ is "small", and therefore

(b) H^n itself is also "small".

Now, by Lemma 1, $e(H^{n_k})\leqq e(S^{n_k})$. On the other hand, S^{n_k} contains no prohibited subgraphs $L\in\mathscr{L}_R$, by the definition of $\mathscr{L}_R$, and H^{n_k} is extremal, thus $e(H^{n_k})\geqq e(S^{n_k})$, that is, S^{n_k} and H^{n_k} have the same number of edges and both are extremal for $\mathscr{L}_R$. If, in addition, $\{S^{n_k}\}$ is strictly dense, then, by Lemma 1, $S^{n_k}=H^{n_k}$. This completes the proof.

Proofs of the Lemmas

Proof of Lemma 1. Let $G^{n_k}\subseteq S^{n_l}$. To prove that $e(G^{n_k})\leqq e(S^{n_k})$ let us define $\tilde{G}^{n_k}$ as a "small" graph with the maximum number of edges on n_k vertices. It is enough to prove that

$$e(\tilde{G}^{n_k})\leqq e(S^{n_k}) \tag{6}$$

if $\tilde{G}^{n_k}$ is "small". Let S^{α} be defined as the infinite graph obtained from $Z^{\alpha}=:\bigtimes_{i=1}^{d}\sum_{j=1}^{\alpha}A_{i,j}$ and D in the way described in Definition 1. Since a graph is "small" iff it is a subgraph of S^{α}, $\tilde{G}^{n_k}$ is the spanned subgraph of S^{α} of n_k vertices with the maximum number of edges. Let us abreviate n_k by n, and denote by v_i the number of vertices of $\tilde{G}^n$ in $\sum A_{i,j}=:\tilde{A}_i$.

Fixing these numbers v_i we also fix the number of edges joining different $\tilde{A}_i$'s. Let $\tilde{A}_{i,j}$ denote the subgraph of $\tilde{G}^n$ spanned by the vertices belonging to $A_{i,j}$, too. We may assume that $\tilde{G}^n$ is a spanned subgraph of S^α, therefore

$$e(\tilde{G}^n)=e(K_d(v_1,\ldots,v_d))+\sum_i\sum_U c_i(U)\cdot(e(U)+e_U),\tag{7}$$

where $c_i(U)$ denotes the number of $\tilde{A}_{i,j}$'s isomorphic to U and in the same position as U, if U is a spanned subgraph of A_i. (Here the same position means that the mapping $F_{i,j}$ of Definition 1 maps $\tilde{A}_{i,j}$ onto U.) Further, e_U denotes the number of edges joining $\tilde{A}_{i,j}$ to D (or, in other words, the number of pairs (x, y) in B for which $y\in U\subseteq A_i$). The sum is taken for all the spanned subgraphs U of A_i.

Let $b=v(A_1)+v(A_2)+\ldots+v(A_d)+v(D)$. We assert that $e(U)+e_U$ is the same for all the graphs U such that $c_i(U)\geqq b!$ if i is fixed. Indeed, if $e(U)+e_U<e(U')+e_{U'}$ and $c_i(U)$, $c_i(U')\geqq b!$, then we may replace $b!/v(U)$ copies of U by $b!/v(U')$ copies of U', thus, by (7), increasing $e(\tilde{G}^n)$. This contradicts the maximality of $e(\tilde{G}^n)$. This very "replacement" argument also yields that we may assume that $c_i(U)\leqq b!$ for every U but one for each i: it may happen that this does not hold for the original $\tilde{G}^n$, but then it can be replaced by another one, $\tilde{\tilde{G}}^n$ for which this holds. After this replacement $\{\tilde{\tilde{G}}^n\}$ is already a q-sequence, and therefore (since S^{n_k} is a dense sequence) for $\tilde{G}^n=\tilde{G}^{n_k}$

$$e(\tilde{G}^{n_k})\leqq e(\tilde{\tilde{G}}^{n_k})\leqq e(S^{n_k}),\tag{8}$$

what was to be proved. The second part of Theorem 1 concerning the strictly dense sequences can easily be proved: we have to show that if $e(G^{n_k})=e(S^{n_k})$, then $G^{n_k}=S^{n_k}$ for $k>k_0$. Indeed, in this case, by (8) $G^{n_k}=\tilde{G}^{n_k}$ can be assumed. If there exists a $U\neq A_i$ for which $c_i(U)\geqq b!$, then the above replacement technique yields a q-sequence $\{\tilde{\tilde{G}}^{n_k}\}$ different from $\{S^{n_k}\}$, since the sum of the block-sizes is smaller) and this contradicts $e(G^{n_k})=eS^{n_k})$ or that $\{S^{n_k}\}$ is strictly dense. This shows that $c_i(U)<b!$ if $U\neq A_i$, that is, $G^{n_k}=\tilde{G}^{n_k}$ itself is a q-sequence. Therefore, by the definition of strict density, $G^{n_k}=S^{n_k}$ if k is sufficiently large.

Proof of Lemma 2. The basic idea of the proof is to partition first the vertices of G^n into $d+1$ classes $\bar{\mathscr{A}}_1,\ldots,\bar{\mathscr{A}}_d$ and $\bar{\mathscr{A}}_0$, then show that the subgraph of G^n spanned by $\bar{\mathscr{A}}_i$ $(i=1,\ldots,d)$ is the sum of components of at most $b=v(A_1)+\ldots+v(A_i)+v(D)$ vertices. If we take all the occurring components as many times as they occur in case if they occur at most $2bd!$ times, otherwise we take only $2bd!$ copies and we take $\bar{\mathscr{A}}_0$, for which $|\bar{\mathscr{A}}_0|=O(1)$, the if the subgraph G^R spanned by these $O(1)$ vertices is "small", then the original graph G^n is "small" as well. In details:

First we fix the constants c, M and R as follows: $r=(3b)^{b^2}$,

$$c=\frac{1}{100bd^2},\quad M=1000\frac{b}{c},\quad R=\max\{4bd^2g_b\cdot 2^{2br}, 30b^3d\}\tag{9}$$

where g_b denotes the number of graphs on $(3b)^{b^2}$ vertices. By the Erdős–Stone theorem [3] G^n contains a $K_d(M, \dots, M)$, if n is large enough. Let the classes of this $K_d(M, \dots, M)$ be $C_1, \dots, C_d$. Now we partition the vertices of $G^n - K_d(M, \dots, M) = G^{n-dM}$ into the following $d+2$ classes: P_i $(i=1, \dots, d)$ contains those vertices which are joined to each C_j $(j \neq i)$ by at least $M\left(1-\frac{1}{6b}\right)$ edges and by at most $3b-1$ edges to C_i. E is the class of vertices joined to each C_j $(i=j$ included) by at least $3b$ edges. V contains the rest.

We assert that $|E| \leq b$. Clearly, if $K_{d+1}(1, 3b, \dots, 3b) \subseteq S^{n_k}$, then its single vertex of the first class belongs to the "head", of S^{n_k}. Therefore, if L can be covered with $b+1$ copies of $K_{d+1}(1, 3b, \dots, 3b)$ with *different peaks*, then L is a prohibited subgraph. On the other hand, for each $x \in E$ we can find a $K_{d+1}(1, 3b, \dots, 3b)$ with the first class $\{x\}$, thus $|E| \leq v(D) \leq b$. A similar argument shows that if $Q_i \subseteq P_i$ is the set of vertices joined to at least $3b$ vertices of $P_i \cup C_i$, then $|\bigcup Q_i| \leq b$.

Next we show that

$$|V| \leq 7bcd \cdot n. \tag{10}$$

Indeed, if T denotes the number of edges joining $K_d(M, \dots, M)$ to G^{n-dM}, then on the one hand

$$T \geq dM\left(1-\frac{1}{d}-c\right)n-(dM)^2 \tag{11}$$

since each $x \in K_d(M, \dots, M)$ has valence $\geq \left(1-\frac{1}{d}-c\right)n$. On the other hand,

$$T \leq (n-dM)\cdot((d-1)M+3b)-|V|\frac{M}{6b}+bM \tag{12}$$

since the vertices of G^{n-dM} are generally joined by at most $(d-1)M+3b$ edges to $K_d(M, \dots, M)$, however, in case, when $x \in E$, it may be joined to $K_d(M, \dots, M)$ by dM edges and if $x \in V$, then it is joined to $K_d(M, \dots, M)$ by less than $(d-1)M+3b-\frac{M}{6b}$ edges. (10) follows easily from (11), (12) and (9).

Now that (10) is established, one can easily show that the classes P_i are approximately of the same size:

(a) Let $x \in P_i - Q_i$. Since x is joined to $\leq 3b$ vertices of its own class P_i and $d(x) \geq \left(1-\frac{1}{d}-c\right)n$, thus

$$|P_i| \leq n-d(x)+3b \leq \left(\frac{1}{d}+c\right)n+3b. \tag{13}$$

This means that none of the classes P_i can be much larger than the average $\frac{n}{d}$. But this implies that none of them can be much smaller:

$$|P_i| \geqq n-(d-1)\left(\frac{1}{d}+c\right)n-3(d-1)b-|V| \geqq$$

(14)

$$\geqq \frac{n}{d}-((d-1)+8bd)cn \geqq \frac{n}{d}-10bcd\cdot n\,.$$

Let us subdivide the class V into $d+1$ subclasses now:

— V_0 contains the vertices of V joined to at least $3b$ vertices of each P_i-Q_i;
— V_i is the set of vertices of V joined to at most $3b-1$ vertices of P_i-Q_i, $i=1, 2, \ldots, d$.

By the valency condition each $x \in V_i$ is joined to P_j-Q_j $(j \neq i)$ by at least

(15) $$\frac{n}{d}-10bcd\cdot n-(10bcd\cdot n+cn+3b) \geqq \frac{n}{d}-22bcd\cdot n$$

edges, since it misses at least $\frac{n}{d}-10bcd\cdot n-3b$ vertices of P_i-Q_i and it misses altogether $n-d(x) \leqq \frac{n}{d}+cn$ vertices. Thus the classes V_i $(i=0, 1, \ldots, d)$ are well defined. Let $\bar{\mathscr{A}}_i =: P_i \cup V_i \cup C_i$ and W_i be the set of vertices of $\bar{\mathscr{A}}_i$ joined to at least $3b$ vertices of the same $\bar{\mathscr{A}}_i$. A slight modification of the above argument shows that $\sum_i |W_i| \leqq b$: we replace K_{d+1} 1, $3b, \ldots, 3b)$ by L defined as follows: for $j=1, 2, \ldots, 3b$, $p=1, 2, \ldots, d$; we fix the vertices $y_{p,j}$ joined to a vertex x (which will be called the "peak", and $y_{p,j}$ is joined to $3b$ vertices of the jth class of a fixed $K_d(10b, \ldots, 10b)$ for every $j \neq p$. If this L is a subgraph of an S^{n_k}, then we omit the head, D from S^{n_k} and obtain, that the "tail" $S^{n_k}-D$ contains $L-D$. Since $v(D) \leqq b$, one can easily see, that the "peak" x was also deleted: $x \in D$. Thus $\sum_i |W_i| \leqq v(D) \leqq b$, since each $x \in W_i$ is the "peak" of an $L \subseteq G^n$. Let now L^* be defined as follows: we take a $K_d(3b, \ldots, 3b)$ and the vertices $y_{p,j}$ $(p=1, 2, \ldots, d$, $j=1, 2, \ldots, 3b)$ are joined to all the vertices of $K_d(3b, \ldots, 3b)$ except to the vertices of the pth class. Further, we take b^2+1 vertices x_i forming a path $(x_1x_2 \ldots x_{b^2+1})$ and join each x_i to each $y_{p,j}$ but for a $p=p_0$. One can easily check that deleting b vertices of L^* we get an L^{**} not occurring in the "tail" $Z^{n-v(D)}=S^n-D$. If on the other hand $\bar{\mathscr{A}}_{p_0}$ contained a path of length b^2+2b, then G^n contained an L^*. By $v(L^*) \leqq R$ this L^* would be "small", that is a subgraph of an S^{n_k} and therefore L^*-(at most b vertices)$=$ $=L^{**} \subseteq Z^{n-v(D)}$ would yield a contradiction, proving that $\mathscr{A}_p$ contains no path of b^2+1 vertices. Hence each connected component of the graph spanned by $\bar{\mathscr{A}}_p-W_p$ has at most $r=(3b)^{b^2}$ vertices. For a fixed p let us call two components equivalent iff they are isomorphic and connected to the vertices of $E\cup W_1\cup\ldots\cup W_d$ in the same way. The number of nonequivalent components is bounded by $g_b\cdot 2^{2br}$.

Let us take the subgraph of G^n defined as follows: we select all the vertices of $E\cup W_1\cup\ldots\cup W_d$ and for each $p=1,\ldots,d$ from each equivalence class of components (in $\bar{\mathscr{A}}_p-W_p$ we take $2bd$ copies of components if there exist that many members in the equivalence class, otherwise we take all of them, i.e. less than $2bd$ copies. These vertices define a graph $G^{\tilde{R}}\in G^n$ for some $\tilde{R}\leqq R$, therefore $G^{\tilde{R}}$ is "small". We embed this $G^{\tilde{R}}$ in S^α. This embedding yields automatically an embedding of G^n into S^α: if U is a component of $\bar{\mathscr{A}}_p$ occurring in $G^n-G^{\tilde{R}}$, then it has multiplicity $\geqq 2bd$, therefore it occurs at least $2bd$ times in $G^{\tilde{R}}\subseteq S^\alpha$. Hence it occurs at least b times in some class of the "tail" Z^α of S^α. Thus we may replace this b copies by an arbitrary number of copies from this class of the "tail". (If U and U' are two such components, joined to each other by at least an edge, then in the embedding of $G^{\tilde{R}}$ in S^α U and U' were put into different classes of the "tail", otherwise a class of the tail contained a K_b. Thus the increasing of the multiplicities of different connected components do not disturb each other!) Thus G^n is "small" as well. This proves Lemma 2.

References

[1] ERDŐS, P., Some recent results on extremal problems in graph theory (Results), *Theory of Graphs*, International Symposium, Rome (1966), 118–123.

[2] ERDŐS, P., New inequalities concerning extremal properties of graphs, *Theory of Graphs, Proc. Coll.*, held at Tihany, (Hungary, 1966), 77–81.

[3] ERDŐS P. and STONE, A. H., On the structure of linear graphs, *Bull. American Math. Soc.*, **52** (1946), 50–57.

[4] SIMONOVITS, M., A method for solving extremal problems in graph theory, stability problems, *Theory of Graphs, Proc. Coll.*, held at Tihany (Hungary, 1966), 279–319.

[5] SIMONOVITS, M., Extremal graph problems with symmetrical extremal graphs, *Discrete Math.*, **7** (1974), 349–376.

[6] TURÁN, P., Egy gráfelméleti szélsőértékfeladatról (In Hungarian, with German resume), *Mat. Phys. Lapok*, **48** (1941) 436–452.

[7] TURÁN, P., On the theory of graphs, *Colloq. Math.*, **3** (1954), 19–30.

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