COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
25. ALGEBRAIC METHODS IN GRAPH THEORY, SZEGED (HUNGARY), 1978.

SOME APPLICATIONS OF GRAPH THEORY AND COMBINATORIAL METHODS TO NUMBER THEORY AND GEOMETRY

P. ERDŐS

I have written several papers and notes on these subjects. To avoid repetitions and to keep this paper short, I will not try to give a systematic account of this subject but will only discuss a few recent results obtained by my collaborators and myself. First of all, I give a partial list of some of my older papers on this subject.

I. P. Erdős, On some problems in elementary and combinatorial geometry, *Annali di Mat. Pura et Applicata,* 53 (1975), 99-108. This is a survey paper with any references.

II. Some applications of graph theory to number theory, The many facets of graph theory, edited by G. Chartrand and S.F. Kapoor, Lecture notes in math 110, Springer-Verlag, 77-82.

III. On some applications of graph theory to number theory, Publ. Ramanujan Institute I.

IV. On the applications of combinatorial analysis to number theory geometry and analysis, *Actes Congrés Int. des Math., Nice,* 3 (1970), 201-210.

1.

In August 1977 the following question occurred to me. Let $x_1, \ldots$ $\ldots, x_n$ be n points in the plane, no three on a line. Determine the smallest $n = n_k$ such that there should always exist k of the x's which determine a convex k-gon which has no x's in its interior. It is easy to see that $n_4 = 5$, but it is not at all obvious that such an n exists for $k > 4$.

Ehrenfeucht in fact gave a simple and intuitive proof that n_5 exists, Harborth and independently Morris proved that $n_5 = 10$. At present it is not known if n_6 exists.

I arrived at the problem about n_k by adding a new condition to the well-known problem of E. Klein (Mrs. Szekeres). Determine or estimate the smallest integer $f(k)$ for which if $x_1, \ldots, x_{f(k)}$ are any $f(k)$ points, no three on a line, then one can always select k of them which form the vertices of a convex k-gon.

Szekeres conjectured $f(k) = 2^{k-2} + 1$. Szekeres and I proved

$$(1) \qquad 2^{k-2} + 1 \leqslant f(k) \leqslant \binom{2k-4}{k-2}.$$

(Some inaccuracies in our proof were corrected by Kalbfleisch). Makai and Turán proved $f(5) = 9$, $f(6) = 17$ has not yet been decided. For the literature on this problem see I.

In IV I stated without proof that to every $\epsilon > 0$ there is an $f_\epsilon(k)$ so that if $x_1, \ldots, x_{f_\epsilon(k)}$ are $f_\epsilon(k)$ points in the plane, no three on a line, then one can always find k of them which form a convex polygon all but two angles of which are greater than $\pi - \epsilon$. I outline the proof which uses Ramsey's theorem. Let t_ϵ be the smallest integer so that among any t_ϵ points there is always a triangle with an angle $> \pi - \epsilon$. It is well known and easy to see that t_ϵ always exists and we will discuss the exact determination of t_ϵ later.

Denote by $r_k(u, v) = m$ the smallest integer so that if we divide the k-tuples of a set S, $|S| = m$, into two classes, then either there is a set S_1, $|S_1| = u$, all whose k-tuples are in class I, or a set S_2, $|S_2| = v$, all whose k-tuples are in class II. Ramsey's theorem implies that $r(u, v)$ always exists.

From Ramsey's theorem we easily deduce

$$f_\epsilon(k) \leqslant r_3(t_\epsilon, r_4(5, k)) = m(\epsilon, k). \tag{2}$$

To prove (2), we split the triangles into two classes. A triangle is in the first class if its greatest angle is not greater than $\pi - \epsilon$ and in the second class otherwise. By the definition of t_ϵ every set of t_ϵ points contains a triangle of the second class. Thus by (2) there is a set of $r_4(5, k)$ points each triangle of which is in the second class. By E. Klein's old theorem every 5-tuple of these $r_4(5, k)$ points contains a convex quadrilateral. Hence there is a set of k points each quadrilateral of which is convex and hence it is a convex k-gon each triangle of which has an angle $> \pi - \epsilon$. But then all but two angles of this convex k-gon are greater than $\pi - \epsilon$, which completes the proof of (2).

There seems to be no doubt that $f_\epsilon(k)$ is much smaller than the value given by (2). It might be worthwhile to try to decide if $f_\epsilon(k) < C_\epsilon^k$ holds.

Let me tell now a few words about the more exact determination of t_ϵ. Szekeres and I proved that every set of 2^n points in the plane determine an angle $> \pi\left(1 - \frac{1}{n}\right)$. This is one of the few best possible results in this field, since an earlier theorem of Szekeres asserts that for every $\epsilon > 0$ there is a set of 2^n points in the plane all whose the angles are $< \pi\left(1 - \frac{1}{n}\right) + \epsilon$. Thus our result seems definitive, but this is not quite the case. Our results certainly imply

$$t_{\frac{1}{n}} \leqslant 2^n \quad \text{and} \quad t_{\frac{1}{n}+\eta} > 2^n \quad \text{for every} \quad \eta > 0$$

but $t_{\frac{1}{n}}$ could be less than 2^n, we only obtain

$$t_{\frac{1}{n}} \geqslant 2^{n-1} + 1 \tag{3}$$

and indeed there could be equality in (3) for $n > n_0$ (but this is certainly highly doubtful). All we proved is that $2^n - 1$ points in the plane always determine an angle $\geqslant \pi\left(1 - \frac{1}{n}\right)$.

The following question is perhaps of some interest. Let $\{x_1, \ldots, x_n\}$ be n points in the plane, no three on a line. Determine or estimate $\min C(x_1, \ldots, x_n)$ where $C(x_1, \ldots, x_n)$ denotes the number of convex subsets of $x_1, \ldots, x_n$. The exact determination of $\min C(x_1, \ldots, x_n)$ is probably hopeless, and even an asymptotic formula seems difficult. We now show

$$(4) \qquad n^{c_1 \log n} < \min C(x_1, \ldots, x_n) < n^{\frac{(1+o(1))\log n}{\log 2}} .$$

The upper bound in (4) follows immediately from (1).

To prove the lower bound observe that (1) implies that every set of t points contains a convex subset of size $[c \log t] = l$. Thus any set of n points contains at least

$$(5) \qquad \frac{\binom{n}{t}}{\binom{n-l}{t-l}} = \frac{n(n-1)\ldots(n-l+1)}{t(t-1)\ldots(t-l+1)}$$

convex l-tuples. To see (5) observe that a convex l-tuple occurs in $\binom{n-l}{t-l}$ t-tuples. Choose $t = [\sqrt{n}]$. Then from (5) we have by a simple computation

$$C(x_1, \ldots, x_n) > \left(\frac{n}{t}\right)^l > n^{\frac{c \log n}{2}}$$

which completes the proof of (4).

Very probably

$$\log \min C(x_1, \ldots, x_n) \sim c \log^2 n$$

holds with some constant c.

On the other hand, I could be wrong. Here is an example where I misjudged the situation. Let $x_1, \ldots, x_n$ be n points in the plane no three on a line. Denote by $f_k(n)$ the minimum number of convex k-tuples such a set must contain. I easily showed in 1934 that

$$\lim_{n=\infty} \frac{f_k(n)}{\binom{n}{k}} = c_k, \qquad 0 < c_k < 1$$

exists and that $f_4(n)$ equals the *rectilinear crossing number* of the complete graph $K(n)$. I did not realize the difference between crossing number, and without further thought I assumed that they probably will be the same. I also assumed that it will be hard or impossible to compute the c_k, thus I abandoned the whole question. I was both pleased and dismayed (at having overlooked the possibilities of the problem) when Guy told me (in Singapore in 1960) the conjecture that the crossing number of $K(n)$ equals

$$\frac{1}{4}\left[\frac{n}{2}\right]\left[\frac{1}{2}(n-1)\right]\left[\frac{1}{2}(n-2)\right]\left[\frac{1}{2}(n-3)\right].$$

The crossing numbers have now a large literature though Guy's conjecture is still open.

Hadwiger and Nelson define the chromatic number α_k of the k-dimensional space as follows. Join two points in k-dimensional space if their distance is 1. The chromatic number α_k of this graph is the chromatic number of k-dimensional space. It has been conjectured that $\alpha_2 = 4$, but now it is generally believed that $\alpha_2 > 4$. It is known that $4 \leqslant \alpha_2 \leqslant 7$. Larman and Rogers proved that $\alpha_k < 3^k$ and P. Frankl proved $\alpha_k > k^c$ for every c if $k > k_0(c)$. It is almost certain that $\alpha_k > (1+\epsilon)^k$ for some $\epsilon > 0$ (independent of k).

Added in proof. P. Frankl proved this conjecture.

Let $|S| = n$. Denote by $f(j, n)$ the largest integer for which there is a family $A_i \subset S$, $1 \leqslant i \leqslant f(j, n)$ of subsets of S so that for every $1 \leqslant i_1 < i_2 \leqslant f(j, n)$ $|A_{i_1} \cap A_{i_2}| \neq j$. Trivially $f(n, 0) = 2^{n-1}$.

P. Frankl proved that the value of $f(n, 1)$ is given by the family F_n, consisting of the sets having at least $\frac{n+1}{2}$ elements for n odd, and of the sets having at least $\frac{n}{2}$ elements in $S \setminus \{s\}$ for n even, where s is a particular element of S.

For $j > 1$ the value of $f(n,j)$ is not known. I conjecture that for every $\eta > 0$ there is an $\epsilon > 0$ so that if

$$(6) \qquad \eta n < j < \left(\frac{1}{2} - \eta\right) n \quad \text{then} \quad f(n,j) < (2-\epsilon)^n .$$

(6), if true, easily implies $\alpha_k > (1+\alpha)^k$ for a certain fixed $\alpha > 0$. It would be, of course, of interest to determine $\lim_{k \to \infty} \alpha_k^{\frac{1}{k}}$.

A well-known theorem of de Bruijn and myself states that if G is an infinite graph of finite chromatic number n then G has a finite subgraph G' of chromatic number n. Thus the determination of α_k is a finite problem and in particular if $\alpha_2 > 4$ there is a finite set S in the plane so that if we join every two points of S whose distance is 1, then the resulting graph $G_1(S)$ has a chromatic number > 4. It would be certainly interesting to find such a set if it exists.

Let now S be a set in the plane which contains no equilateral triangle of side 1. I thought it quite likely that then $G_1(S)$ has chromatic number < 4. I hoped that if this is not true then the following weaker conjecture holds: There is a k so that if $G_1(S)$ has girth $\geqslant k$ (i.e., the least circuit of $G_1(S)$ has k sides) then $G_1(S)$ has chromatic number < 4. Wormald in a recent paper (which is not yet published) disproved my original conjecture – he found an S for which $G_1(S)$ has girth 5 and chromatic number 4. Wormald's construction uses elaborate computations and is fairly complicated.

Let $u_1, \ldots, u_r$ be any r positive numbers. Join two points of the plane if their distance is one of the numbers u_i, $i_1 = 1, \ldots, r$. $\alpha_2(r)$ is the largest integer such that there is such a graph of chromatic number $\alpha_2(r)$. Can $\alpha_2(r)$ increase exponentially in r? It seems possible that $\alpha_2(r)$ increases polynomially in r. I can not disprove $\alpha_2(r) < r^{1+\epsilon}$.

Simonovits and I in a forthcoming paper which will appear in the Ars Combinatorica introduce a modified chromatic number of k-dimensional space α'_k as follows. Let $x_1, \ldots, x_n$ be again n distinct points in k-dimensional space; join two points whose distance is 1. α'_k is the largest integer so that our graph has chromatic number α'_k after the omission of

$o(n^2)$ arbitrarily chosen edges. Probably $\alpha'_k > (1+c)^k$, but we cannot prove even that $\frac{\alpha'_k}{k} \to \infty$. We proved that $\alpha'_4 = 2$, $\alpha'_5 > 2$. In fact if $k = 4$ we show that one can always omit $o(n^{\frac{7}{4}})$ edges so that the remaining graph should have chromatic number 2.

Here I just state an old problem which is also discussed in detail in I. Denote by $f_k(n)$ the largest integer such that among any n distinct points in k-dimensional space there are at most $f_k(n)$ pairs of points whose distance is 1. I conjectured $f_2(n) < n^{1+\epsilon}$ and offer a hundred dollars for a proof of disproof. All that is known is that

$$f_2(n) = o(n^{\frac{3}{2}}) \quad \text{and} \quad f_2(n) > n^{1 + \frac{c}{\log\log n}}.$$

The lower bound is probably close to the truth. G. Purdy and I are planning to write a book about this and related questions. Not much progress has been made here in the last few years. Let me state a problem where significant progress has been made very recently. Denote by $D(n)$ the minimum number of distinct directions determined by n distinct points in the plane. Scott proved

$$c_1 n^{\frac{1}{2}} < D(n) \leq 2\left[\frac{n}{2}\right].$$

and conjectured that the upper bound is exact (or is close to being so). Burton and Purdy recently proved Scott's conjecture. They also proved that n non-collinear points determine at least cn triangles of distinct areas. Previously Purdy and I only proved this with $n^{\frac{3}{4}}$. The paper of Burton and Purdy is not yet published.

To end this chapter, I state two more problems. Let $x_1, \dots, x_n$ be n distinct points in the plane. Denote by $c_1, \dots, c_m$ the set of all circles passing through at least three of the x_i's. Let $h(n)$ be the largest integer such that for suitable choice of the x_i there are $h(n)$ distinct circles of radious 1 among the c_i. I could only prove

$$\frac{3n}{2} < h(n) \leq n(n-1).$$

I would expect that

$$\frac{h(n)}{n^2} \to 0, \quad \frac{h(n)}{n} \to \infty. \tag{7}$$

Perhaps I overlook a simple idea but I could make no progress with this simple and attractive conjecture. I think an exact formula, or even an asymptotic formula for $h(n)$ might be difficult to get. Harborth and I tried to prove $\frac{h(n)}{n} \to \infty$ as follows: Consider the set of lattice points (x, y), $0 \leq x$, $y < n^{\frac{1}{2}}$. Denote by $h_r(n)$ the number of distinct circles of radius r which pass through at least three of our lattice points. Is it true that

$$\lim_{n \to \infty} \max_r \frac{h_r(n)}{n} = \infty \tag{8}$$

(8), if true, clearly implies the second conjecture (7), but we could not prove (8).

Finally let me state an old and completely forgotten question of Corrádi, Hajnal and myself: Is it true that if there are given n points in the plane, not all on a line, then they determine at least $n-2$ different angles? (0 and π are counted as different but angles greater than π are not allowed.)

2.

On combinatorial methods in number theory I published even more than on applications to geometry thus I mention only a few recent results, one of which I learned during the meeting. Let $1 \leq a_1 < \ldots$ be an infinite sequence of integers and denote by $f(n)$ the number of solutions of $n = a_i + a_j$. An old conjecture of Turán and myself states: If $f(n) > 0$ for all $n > n_0$ then $\limsup f(n) = \infty$. This conjecture seems rather instructable and I offer 500 dollars for a proof of disproof.

I observed many years ago that the multiplicative analog to this problem can be handled without too much difficulty: Let $b_1 < b_2 < \ldots$ be an infinite sequence of integers. Denote by $g(n)$ the number of solutions of $n = b_i b_j$. Assume $g(n) > 0$ for all $n > n_j$. Then $\limsup g(n) = \infty$. The proof was not very difficult but used extremal properties of hypergraphs and was not too simple either. During our meeting Nešetřil and Rödl told me their proof which is completely combinatorial and with their kind permission I now give their very ingenious proof. In fact we show the following stronger result: let $p_1 < p_2 < \dots$ be an infinite sequence of primes and let $u_1 < u_2 < \dots$ be the sequence of squarefree integers which are composed of exactly m of the p's. Let $a_1 < a_2 < \dots$ be a sequence of integers such that every u can be written in the form $a_i a_j$, then there is an integer t of $2m$ prime factors for which the number of solutions of $t = a_i a_j$ is at least $\binom{m+1}{\left[\frac{m+1}{2}\right]}$.

We split the m-tuples of integers into 2^m classes as follows. If $p_{i_1} \dots p_{i_m} = u$ $(i_1 < \dots < i_m)$ is one of the u's and $u = a_{j_1} a_{j_2}$, $a_{j_1} = p_{i_{s_1}} \dots p_{i_{s_k}}$, then the class of $\{i_1, \dots, i_m\}$ is determined by $\{s_1, \dots, s_k\}$. This is a subset of $\{1, \dots, m\}$, so there are 2^m classes. By Ramsey's theorem there is an infinite sequence of integers $i_1 < i_2 < \dots$ all whose m-tuples lie in the same class, characterized by a set $\{s_1, \dots, s_k\}$, $1 \leqslant s_1 < \dots < s_k \leqslant m$. Put $q_j = p_{i_j}$ and $r_j = q_{mj}$. The product of every k r_j's will be an a, as we can always choose m q_j's so that the s_1-th, ..., s_k-th of them should be our k r_j's. Regarding the set $\{1, \dots, m\} \setminus \setminus \{s_1, \dots, s_k\}$ we get that the product of $(m-k)$ r_j's must be also an a. Assume $m - k \geqslant k$. If t is the product of $(2m - 2k)$ r_j's, then $t = = a_i a_j$ has at least $\binom{2m-2k}{m-k} \geqslant \binom{m+1}{\left[\frac{m+1}{2}\right]}$ solutions, which completes the proof.

It is clear from the finite form of Ramsey's theorem that we obtain the same result if we assume only that every integer having m distinct prime factors formed from a finite set of primes $p_1 < \dots < p_s$, $s = s(m)$ is of the form $a_i a_j$, them there is an integer t for which the equation $t = a_i a_j$ has at least $\binom{m+1}{\left[\frac{m+1}{2}\right]}$ solutions.

Let $a_1 < a_2 < \ldots < a_n$ be n integers. Let $f(n)$ denote the largest integer such that there are at least $f(n)$ distinct numbers of the form $a_i + a_j$ and $a_i a_j$. Several years ago I conjectured that $f(n) > n^{2-\epsilon}$ for every $\epsilon > 0$ and $n > n_0(\epsilon)$. Szemerédi and I proved that

$$(1) \qquad n^{1+c} < f(n) < n^2 \exp\left(\frac{-c \log n}{\log \log n}\right).$$

Perhaps the upper bound gives the right order of magnitude for $f(n)$.

Consider all the integers of the form $a_{i_1} + \ldots + a_{i_k}$ and $a_{i_1} \ldots a_{i_k}$. Denote by $f_k(n)$ the minimal number of distinct integers of this form. We conjecture $f_k(n) > n^{k-\epsilon}$. Finally if we consider all the 2^n sums and 2^n products formed from the a's and $F(n)$ is the largest integer such that there are at least $F(n)$ distinct integers of this form, we conjecture that $F(n) > n^k$ for every k and $n > n_0(k)$. It is surprising that we seem to be unable to attach this very plausible conjecture – perhaps we overlook a simple idea. We proved

$$(2) \qquad F(n) < \exp \frac{c(\log n)^2}{\log \log n}.$$

Perhaps (2) gives the right order of magnitude for $F(n)$.

Graham and Rothschild conjectured that if we split the integers into two classes, then always there is an infinite sequence $a_1 < a_2 < \ldots$ such that all the sums

$$(3) \qquad \sum_k \epsilon_k a_k, \quad \epsilon_k = 0 \quad \text{or} \quad 1,$$

are in the same class. This conjecture was proved by Hindman.

I asked if it is true that there is an infinite sequence $a_1 < a_2 < \ldots$ such that all the sums and products

$$(4) \qquad \sum_k \epsilon_k a_k, \quad \prod_k a_k^{\epsilon_k}, \quad \epsilon_k = 0 \quad \text{or} \quad 1,$$

are in the same class. I also asked: if (4) is false, does it remain true if we only require the existence of a sequence $1 < a_1 < \ldots < a_n$ such that all the 2^n sums and products

$$(5) \qquad \sum_{k=0}^{n} \epsilon_k a_k, \qquad \prod_{k=0}^{n} a_k^{\epsilon_k}, \qquad \epsilon_k = 0 \quad \text{or} \quad 1$$

are in the same class? Hindman disproved (4), his paper will soon appear in the J. of Combinatorial Theory. (5) remains open for $n \geqslant 3$.

Pomerance and I investigated the following problem. Let $f(n)$ be the smallest integer such that n different integers $a_1, \ldots, a_n$ can be found, $n < a_t < nf(n)$, $t \mid a_t$.

We proved

$$c_1 \left(\frac{\log n}{\log \log n}\right)^{\frac{1}{2}} < f(n) < (2 - c_2) \log^{\frac{1}{2}} n.$$

We cannot decide whether

$$f(n) = o(\log^{\frac{1}{2}} n).$$

$F(n)$ is the smallest integer such that for every m there are n distinct integers $a_1, \ldots, a_n$ satisfying

$$a_t \equiv 0 \pmod t, \quad m < a_t < m + F(n)$$

for every t, $1 \leqslant t \leqslant n$. We could prove only $F(n) < n^{\frac{3}{2} + \epsilon}$.

One of our principle tools in all these results is the well known König – Hall theorem.

Define $F^*(n)$ as the smallest integer such that, for every m and every $p \leqslant n$ distinct integers $a_p^{(m)}$, $m < a_p^{(m)} < m + F^*(n)$ can be found satisfying $p \mid a_p$. We could not disprove $F^*(n) = O(n)$.

A curious result of Selfridge and myself seems to point in the other direction (but certainly does not decide the issue). For every $\epsilon > 0$ and k there is a set of k^2 primes $p_1 < \ldots < p_{k^2}$ and an interval $(x, x + (3 - \epsilon)p_{k^2})$ such that the number of distinct integers m in this interval which are multiples of any of the p_i's is $2k$; i.e. it is surprisingly small. We do not know what happens if the upper bound $(3 - \epsilon)p_{k^2}$ is replaced by $(3 + \epsilon)p_{k^2}$.

An older and somewhat related problem of D. Newman is it true that there is a one-to-one mapping φ of the integers $1 \leqslant t \leqslant n$ onto the integers $m \leqslant t \leqslant n + n$ such that $(t, \varphi(t)) = 1$ for every t? If $m = n + 1$, this was proved by Baines and Daykin, but as far as I know the general case is still open. One would expect that Hall's theorem will apply here but there seem to be unexpected difficulites.

Added in proof. Pomerance and Selfridge recently proved the general case.

REFERENCES

[1] P. Erdős – G. Szekeres, On some extremum problems in elementary geometry, *Annales Univ. Sci.*, Budapest 3-4 (1960-61), 53-62, see also.

[2] P. Erdős,– R. Guy, Crossing number problems, *Amer. Math. Monthly*, 80 (1973), 52-57.

[3] D.G. Larman – C.A. Rogers, The realization of distances within sets in Euclidean space, *Mathematica* 19 (1972), 9-24, see also D.R. Woodall, Distances realized by sets covering the plane, *J. Combinatorial Theory, A*, 14 (1973), 187-200.

[4] H. Harborth, Konvexe Funfecke in ebenen Punktmengen, *Elemente der Math.*, 33 (1978), 116-118.

[5] S. Jozsa – E. Szemerédi, The number of unit distances on the plane, Infinite and finite sets, *Coll. Math. Soc. J., Bolyai*, 1975, 2, 939-955.

[6] P. Frankl, An intersection problem for finite sets, *Acta Math. Acad. Sci. Hung.*, 30 (1977), 371-373.

P. Erdős

Mathematical Institute of the Hungarian Academy of Sciences, 1053 Budapest, Reáltanoda u. 13-15, Hungary.