SOME APPLICATIONS OF GRAPH THEORY TO NUMBERS THEORY

P. Erdös
Hungarian Academy of Sciences

Let $a_1 < \ldots < a_k \le n$ be a sequence of integers no one of which divides any other. It is not difficult to see that $\max k = [\frac{n+1}{2}]$ [1]. Assume now that no a_i divides the product of two others, then I proved that [2] ($\pi(x)$ denotes the number of primes not exceeding x)

$$(1) \qquad \pi(x) + \frac{c_1 x^{2/3}}{(\log x)^2} < \max k < \pi(x) + \frac{c_2 x^{2/3}}{(\log x)^2} .$$

The proof of both the upper and the lower bound used combinatorial methods. Probably

$$(2) \qquad \max k = \pi(x) + \frac{c x^{2/3}}{(\log x)^2} + o\left(\frac{x^{2/3}}{(\log x)^2}\right)$$

for a certain c; but I could not prove (2).

Assume next that the products $a_i a_j$ are all different. Then I proved [3]

$$(3) \qquad \pi(x) + \frac{c_3 x^{3/4}}{(\log x)^{3/2}} < \max k < \pi(x) + \frac{c_4 x^{3/4}}{(\log x)^{3/2}} .$$

I expect that here too

$$(4) \qquad \max k = \pi(x) + \frac{c x^{3/4}}{(\log x)^{3/2}} + o\left(\frac{x^{3/4}}{(\log x)^{3/2}}\right)$$

but again I can not prove (4). The proof of both the upper and the lower bound of (3) is combinatorial and graph theoretic.

Assume that the products taken r at a time $a_{i_1} \ldots a_{i_r}$ are all different. We have no completely satisfactory estimation of $\max k$ if $r > 2$.

Assume that all the products

$$\prod_{i=1}^{k} a_i^{\epsilon_i}, \qquad \epsilon_i = 0 \text{ or } 1$$

are different. I proved that [4]

$$(5) \qquad \pi(x) + \pi(\sqrt{x}) < \max k < \pi(x) + \frac{c_5 x^{1/2}}{\log x} .$$

The lower bound is obvious, it suffices to take the primes and their squares - the proof of the upper bound is more complicated. Probably

$$(6) \qquad \max k = \pi(x) + \pi(\sqrt{x}) + o\left(\frac{x^{1/2}}{\log x}\right)$$

holds and one can make plausible conjectures for sharper results than (6) [4].

Let $a_1 < \dots < a_k$ be a sequence of real numbers. Assume that for every four indices i, j, r, s

$$(7) \qquad |a_i a_j - a_r a_s| \ge 1,$$

if the a's are integers then (7) means that the products $a_i a_j$ are all different. I can not prove that (7) implies $k = o(x)$.

Let now $a_1 < \dots < a_k \le x$ and assume that the sums $a_i + a_j$ are all distinct. It is known that [7]

$$(1+o(1))x^{1/2} < \max k < x^{1/2} + x^{1/4} + 1 .$$

Turán and I conjectured

$$(8) \qquad \max k = x^{1/2} + O(1).$$

(8) if true seems rather deep. Assume now that all the sums taken r at a time $a_{i_1} + \dots + a_{i_r}$ are distinct. Bose and Chowla conjectured

$$\max k = (1+o(1))x^{1/r} ,$$

but they could only prove $\max k \ge (1+o(1))x^{1/r}$ [8].

Let us finally assume that $a_1 < \dots < a_k \le x$ is such that the sums $\Sigma\, \varepsilon_i a_i$, $\varepsilon_i = 0$ or 1 are all different. Moser and I proved that [5]

$$\max k = \frac{\log x}{\log 2} + \frac{\log\log x}{2 \log 2} + o(\log x).$$

Conway and Guy showed that for $x = 2^r$, $r > r_0$, $\max k \geq r+2$. Perhaps

(9) $$\max k = \frac{\log x}{\log 2} + O(1).$$

(9) is probably rather deep.

Let $a_1 < \dots < a_k \leq x$, $k > \pi(x)$. Then it is easy to see that the products $\Pi_{i=1}^{k} a_i^{\alpha_i}$ can not all be different. Let $k > \pi(x)$, denote by $f(k,x)$ the smallest integer so that there always are $f(k,x) = r$ primes $p_1 < \dots < p_s$ for which more than r a's are of the form $\Pi_{i=1}^{r} p_i^{\alpha_i}$. Clearly $f(k,x) \leq \pi(x)$, also $f(k,x)$ is a non increasing function of k. Straus and I proved

(10) $$f(\pi(x)+1, x) = (4+o(1))\frac{x^{1/2}}{(\log x)},$$

and in fact we obtained several sharper results than (10) the proof of which we will outline.

Let $k = cx$. I proved

(11) $$f(k,x) = \log\log x + (c_1+o(1))(2 \log\log x)^{1/2}$$

where

(12) $$\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{c_1} e^{-x^2/2}\, dx = c.$$

Now we prove (10). Let $p_1 < \dots < p_s$, $s = \pi(x^{1/2})$ be the primes not exceeding x, $q_1 < \dots < q_v$ are those primes greater than $x^{1/2}$ which divide more than one $\underline{a}$. Since $k > \pi(x)$ a simple argument shows that more than $s+v$ a's are composed of the primes $p_1,\dots,p_s$, $q_1,\dots,q_v$ (since all the other q's divide at most one $\underline{a}$ and $k > \pi(x)$). Thus

(13) $$f(\pi(x)+1, x) \leq s + v.$$

Now we show

(14) $$f(\pi(x)+1,\ x) \le 2s + 1.$$

The proof of (13) is indeed easy. If $s \ge v$, (13) implies (14). Assume next $v > s$, let $q_1 < \dots < q_{s+1}$ be the first $s+1$ q's. Clearly at least $2s+2$ a's are composed of the $2s+1$ primes $p_1,\dots,p_s,\ q_1,\dots,q_{s+1}$ (to each q there corresponds at least two a's and the a's corresponding to distinct q's are distinct). This completes the proof of (14).

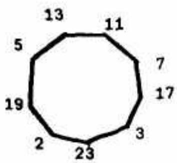
By the prime number theorem and (14) we obtain

(15) $$f(\pi(x)+1,\ x) \le (4+o(1))\frac{x^{1/2}}{\log x} .$$

Next we estimate $f(\pi(x)+1,\ x)$ from below. Let $p_1 < \dots < p_t$ be the set of primes not exceeding $(2-\varepsilon)x^{1/2}$. We define a set A_t of t integers as follows:

(16) $$A_t = \left\{p_tp_1,\ p_tp_2,\ p_{t-2r+1}p_{2r-1},\ p_{t-2r+1}p_{2r+1},\ p_{t-2r}p_{2r},\ p_{t-2r}p_{2r+2}\right\}\quad r = 1,\dots$$

and we close up the cycle so that eacy p_i, $1 \le i \le t$ should occur in exactly two integers of A_t. Let for example $t = 8$, then the set A_8 consists of the 8 integers $19\cdot2,\ 19\cdot3,\ 17\cdot2,\ 17\cdot5,\ 13\cdot3,\ 13\cdot7,\ 5\cdot11,\ 7\cdot11$. It is easy to give a geometric interpretation of A_t. Consider a polygon of t sides, the vertices are the primes $p_1,\dots,p_t$ and the edges which are interpreted as the products of the vertices are the elements of A_t, e.g. $t = 9$



It easily follows from the prime number theorem that for $x > x_0(\epsilon)$ all elements of A_t are less than x.

Now we define a set of $\pi(x) + 1$ integers as follows: The primes $p_j \le x$, $j \ge t+2$, the t elements of A_t and $2p_{t+1}$, $3p_{t+1}$. For $x \ge x_0(\epsilon)$ all these numbers are $\le x$ and $p_1, \ldots, p_{t+1}$ is clearly the smallest set of primes so that there are more a's composed of these primes than the number of these primes. Thus by the prime number theorem for $x > x_0(\epsilon)$

$$(17) \qquad f(\pi(x)+1,\ x) \ \ge\ t + 1 \ =\ (2-\epsilon+o(1))\frac{2x^{1/2}}{\log x}\,.$$

(15) and (17) imply (10). By using the prime number theorem with an error term the above proof gives

$$f(\pi(x)+1,\ x) \ =\ 2\pi(x^{1/2}) + o\left(\frac{x^{1/2}}{(\log x)^k}\right)$$

for every k.

We also observed that (13) is best possible for quite large values of x, e.g. $f(26,100) = 9$ $(\pi(100) = 25)$. To see this take the primes from 29 to 97 and the 10 numbers 34, 38, 39, 46, 55, 57, 69, 77, 85, 91. In fact there always is equality in (10) whenever the set of integers (16) formed with the primes $\le 2x^{1/2}$ are all not greater than x. This certainly happens for very much larger values of x than 100 but Straus and I conjectured that for $x > x_0$ this never happens and that in fact

$$(18) \qquad 2\pi(x^{1/2}) - f(\pi(x)+1,\ x) \ \to\ \infty.$$

We also made the following related conjecture: For every sufficiently large prime p_k there is an index i for which

$$(19) \qquad p_k^2 \ <\ p_{k+i}\, p_{k-i}.$$

(19) if true is certainly very deep. There certainly are fairly large values of k so that for all $i < k$, $p_k^2 > p_{k+i}p_{k-i}$ and we could perhaps try to find the largest such value by a computer, but even if one would succeed it will be very difficult to show that one really has found the largest such value.

Finally Straus and I proved that

$$(20) \qquad f(\pi(x)+1,\ x) = t$$

where t is the largest integer so that all the t integers of A_t are less than or equal to x. The proof of (20) follows easily from the remark that if $a_i = q_j z$ then all prime factors of z are $\leq x/q_j$.

Now we prove (11). A theorem of Kac and myself states [6] that the number of integers $n \leq x$ for which $V(n) > \ell og\ell og\ x + \alpha(2\ell og\ell og x)^{1/2}$ is ($V(n)$ denotes the number of distinct prime factors of n)

$$(21) \qquad x(1+o(1))\ \frac{1}{(2\pi)^{1/2}} \int_\alpha^\infty e^{-x^2/2} dx.$$

From (21) we immediately obtain that the number of integers $n \leq x$ for which

$$(22) \qquad V(n) > \ell og\ell og\ x + (c_1-\varepsilon_x)(2\ell og\ell og x)^{1/2}$$

is $> cx$ where c_1 is determined by (12) and ε_x tends to 0 as x tends to infinity. Let now $a_1 < \ldots < a_k \leq x$, $k > cx$ be the integers not exceeding x which satisfy (22). This set of integers clearly shows that for $k = cx$

$$(23) \qquad f(k,x) \geq \ell og\ell og\ x + (c_1+o(1))(2\ell og\ell og x)^{1/2}$$

(since no $\underline{a}$ is composed of fewer than $\ell og\ell og\ x + (c_1+o(1))(2\ell og\ell og x)^{1/2}$ prime factors).

Thus to complete the proof of (11) we have to estimate $f(k,x)$ from above. My first results were obtained by combinatorial methods. I proved that if

$a_1 < \dots < a_k \le x$, $k \ge cx$ then for every α there is a y and a sequence of primes

$$y < p_1 < \dots < p_r < 2y, \quad r > \alpha \log\log x$$

and integers

$$b_1 < \dots < b_s, \quad s > \frac{x}{(\log x)^k}, \quad k = k(\alpha)$$

so that all the numbers $p_i b_j$, $1 \le i \le r$; $1 \le j \le s$ are a's. From $s > (x/(\log x)^k)$ I then deduced that there are indices j_1, j_2 and primes $p_u p_v$ so that $b_{j_1} p_1 = b_{j_2} p_2$. But all these results only gave $f(k,x) < (2+o(1)) \cdot \log\log x$. Finally I found simpler number theoretic methods which gave the required upper bound for $f(k,x)$. I now outline my proof. Let $a_1 < \dots < a_k \le x$, $k \ge cx$ be any sequence of integers. It easily follows from (11) and (12) that for every $\epsilon > 0$ there is a $\delta = \delta(\epsilon)$ so that our sequence has a subsequence $a_{i_1} < \dots < a_{i_r}$ satisfying for every $1 \le j \le r$

$$V(a_{i_j}) < \log\log x + (c_1+\epsilon)(2\log\log x)^{1/2}, \quad r > \delta x. \tag{24}$$

Put $\exp\exp(\log\log x)^{1/3} = y$ ($\exp z = e^z$) and denote by $V_y(n)$ the number of distinct prime factors of n not exceeding y. It easily follows from the method of Turán[10] (or again from [6]) that for at least $\frac{r}{2}$ of the a_{i_j}'s we have

$$V_y(a_{i_j}) > \frac{2}{3}(\log\log x)^{1/3} = \frac{2}{3}\log\log y. \tag{25}$$

In (25) $\frac{2}{3}(\log\log x)^{1/3}$ could be replaced by $(\log\log x)^{1/3} - c(\log\log x)^{1/6}$ for sufficiently large c, but (25) suffices for our purpose.

Denote by $a_1 < \dots < a_t$, $t > \frac{\delta x}{2}$ the a's which satisfy (25). Denote further by $b_1 < \dots < b_z < y$ the integers for which

(26) $\frac{4}{3}(\log\log x)^{1/3} > V(b_i) > \frac{2}{3}(\log\log x)^{1/3}$.

From Turán's method [10] (or from [6]) $z = (1+o(1))y$. Consider now the integers

(27) $a_i b_j$, $1 \le i \le t$; $1 \le j \le z$.

Denote by $F_y(n)$ the number of prime factors of n not exceeding y where in $F_y(n)$ multiple factors are counted multiply. From (25) and (26) we have

(28) $F_y(a_i b_j) > \frac{4}{3}(\log\log x)^{1/3} = \frac{4}{3}\log\log y$.

From (28) it easily follows from the method of Hardy and Ramanujan [9] that the number of integers $m \le xy$ satisfying (28) in other words satisfying

(28a) $F_y(m) > \frac{4}{3}(\log\log x)^{1/3}$

is less than $xy\exp(-\eta(\log\log x)^{1/3})$ for a certain fixed $\eta > 0$. (Turán's method would give here only $(cxy/(\log\log x)^{1/3})$ which would not be enough for our purpose, but by using higher moments we would obtain $o(xy/\log\log x)$ which would suffice for our purpose.)

The number of the products of the form (27) is clearly

(29) $tz > \frac{\delta xy}{4}$.

From (29) and (28a) there is an $m < xy$ for which the number of solutions of $m = a_i b_j$ is greater than $(\log\log x)^2$, in other words m is divisible by at least $(\log\log x)^2$ distinct a's. (24) and (26) imply on the other hand that

(30) $V(m) < \log\log x + (c_1+\epsilon)(\log\log x)^{1/2} + \frac{4}{3}(\log\log x)^{1/3}$.

Thus clearly

$$f(k,x) < V(m)$$

or

$$(31) \qquad f(k,x) < \log\log x + (c_1+o(1))(\log\log x)^{1/2}.$$

(23) and (31) complete the proof of (11).

One could study $f(k,x)$ for $k = o(x)$ and $k > \pi(x)+1$, but I have not yet obtained as sharp results as (10) and (11).

References

1 Problem 3739 Amer. Math. Monthly 44(1937), 120.

2 P. Erdős, On sequences of integers no one of which divides the product of two others and on some related problems, MITT. FORSCH. INST. MATH. u. MECH. UNIV. TOMSK 2, 79-82 (1938).

3 P. Erdős, On some applications of graph theory to number theoretic problems, Publ. Ramanujan Institute No. 1(1969) 131-136.

4 P. Erdős, Extremal problems in number theory II. Mat. Lapok 17(1966), 135-155 (in Hungarian).

5 P. Erdős, Problems and results in additive number theory, Colloque sur la théorie des nombres, Bruxelles 1955, p. 136-137.

6 P. Erdős and M. Kac, The Gaussian law of errors in the theory of additive number theoretic functions, Amer. J. Math. 62(1940) 738-742.

7 H. Halberstam and K.F. Roth, Sequences, Oxford Univ. Press, New York-London 1966, p. 85-86 and B. Lindström, An inequality for B_2 sequences, J. Comb. Theory 6(1969), 211-212.

8 Ibid., 5, p. 81 and 85.

9 G.H. Hardy and S. Ramanujan, The normal number of prime factors of a number n, Quarterly J. Math. 48(1917), 76-92, see also Ramanujan, Collected papers 262-276.

10 P. Turán, On a theorem of Hardy and Ramanujan, J. London Math. Soc. 9 (1934), 274-276.