Problems and Results in Combinatorial Analysis and Combinatorial Number Theory

Paul Erdős

In this note, first of all, I state a few older problems whose solution or partial solution is long overdue and which in some cases were neglected. I also state a few more recent questions. I prove in detail a recent theorem of Selfridge and myself and of Ulam and myself. I try to give as exact references as possible but I am willing (and eager) to correct any mistakes which are pointed out to me.

1. Conjecture of Faber, Lovász and myself:

In September 1972 at Boulder at a party held for Lovász we conjectured: Let $|A_k| = n$, $1 \leq k \leq n$, $|A_{k_1} \cap A_{k_2}| \leq 1$, $1 \leq k_1 < k_2 \leq n$. Is it then true that we can color the elements of $\bigcup_{k=1}^{n} A_k$ by n colors so that each A contains an element of each color? We immediately realised that the conjecture fails for $n + 1$ sets, but did not immediately realise the difficulties of proving the conjecture. Greenwell and Lovász proved that the conjecture is true if the number of sets is $\leq \frac{n+1}{2}$. No further results are known, and I offer 300 dollars for a proof or disproof of our original conjecture.

The following modification of our conjecture is of interest. Let $A_1, \ldots, A_\ell$ be a family of sets. The graph $G(A_1, \ldots, A_\ell)$ is defined as follows: The vertices of our graph are the elements of $\bigcup_{i=1}^{\ell} A_i$. Two vertices are joined if they belong to the same A_i.

PROC. 9TH S-E CONF. COMBINATORICS, GRAPH THEORY, AND COMPUTING, pp. 29-40.

Put $\max_{1 \le i \le \ell} |A_i| = n$, $\max_{1 \le i < j \le \ell} |A_i \cap A_j| = u$. $f(n;\ell,u)$ denotes the maximal chromatic number of $G(A_1,\ldots,A_\ell)$. Our conjecture is equivalent to $f(n;n,1) = n$. It easily follows from a theorem of de Bruijn and myself that no graph with the parameters $n,n,1$ can contain a $K(n + 1)$. Perhaps it would be of interest to determine the smallest m_k for which $f(n;m_k,1) \ge n + k$. At the moment I do not even have a plausible conjecture, but perhaps this will not be hard to find. Is it true that the graphs for which the chromatic number of $G(A_1,\ldots,A_{m_k})$ is $\ge n + k$ always must contain a $K(n + k)^2$?

N. G. de Bruijn and P. Erdös, A colour problem for infinite graphs and a problem in the theory of relations, Nederl. Akad. Wetonnh. Proc. Ser. A. 54 (1951), 371-373.

2. Some old extremal problems.

Denote by $G(n;k)$ a graph of n vertices and k edges. V. T. Sós and I conjectured 15 years ago that every $G(n;[\frac{1}{2}(k - 1)n + 1])$ contains every tree with k edges. This is trivial for a star and Gallai and I proved it for a path. It is surprising that no progress has been made with this simple and useful conjecture. I offer 100 dollars for a proof or disproof.

Let G be a bipartite graph. $f(n;G)$ is the smallest integer for which every $G(n;f(n;G))$ contains G as a subgraph. Simonovits and I conjectured that for every rational α, $1 < \alpha < 2$ there is a G for which

$$(1) \quad \lim_{n=\infty} f(n;G)/n^{1+\alpha} = c_\alpha(G),\ 0 < c_\alpha(G) < \infty.$$

Conversely for every G there is a rational α which satisfies (1). I offer 300 dollars for a proof or disproof.

We have no guess for the possible values of $c_\alpha(G)$.

My reason for offering so much more for the second conjecture is that I am not entirely sure that a trivial counterexample can not be found to the first conjecture.

Finally I mention a problem of Sauer and myself. Denote by $f_3(n)$ the smallest integer for which every $G(n;f_3(n))$ contains a regular subgraph of valency three. Almost nothing is known about $f_3(n)$; almost certainly $f_3(n) < n^{1+\varepsilon}$. We do not even know if $f_3(n) < Cn$ is true or false. I offer 100 dollars for an answer.

For further problems and results see P. Erdös, Extremal problems on graphs and hypergraphs, Hypergraph Seminar, Lecture Notes in Math., Springer Verlag no. 411, p. 75-84 and Some recent progress on extremal problems in graph theory, Cong. Num. XIV, Proc. Sixth Southeastern Conf. on Combinatorics, Graph Theory and Computing, Florida Atlantic University, 1974, 3-14.

A comprehensive book on extremal graph problems by B. Bollobás will soon appear.

3. Some problems on probabilistic graph theory.

Let $G(n)$ be a graph of n vertices. It easily follows by probabilistic graph theory that for every $0 < \alpha < \frac{1}{2}$ there is a graph $G(n;[\alpha n^2])$ so that if $m/_{\log n} \to \infty$ then every spanned subgraph of m vertices has $(1 + o(1))\, \alpha\, m^2$ edges. It is easy to see that this result is best possible in the following strong sense: Every $G(n;[\alpha n^2])$ has two subgraphs G_1 and G_2 of $[C \log n]$ vertices each, so that ($e(G)$ is the number of edges of G)

$$\frac{e(G_1)}{(C \log n)^2} > \alpha + h_1(C) + o(1) \quad \text{and}$$

$$\frac{e(G_2)}{(C \log n)^2} < \alpha - h_2(C) + o(1).$$

The best values of $h_1(C)$ and $h_2(C)$ are, of course, not known ("of course", since their determination connects with Ramsey theory); $h_1(C) > 0$, $h_2(C) > 0$ holds for every C; also $h_1(C) \to 0$, $h_2(C) \to 0$ as $C \to \infty$.

By probabilistic methods it is not hard to show that for every $0 < c < \frac{1}{2}$ there is a graph $G(n;[n^{1+c}])$ which has no triangle and every spanned subgraph of αn vertices, $0 < \alpha < 1$, has $(1 + o(1))\alpha^{1+c}n^{1+c}$ edges. I do not know if this result remains true for $\frac{1}{2} \leq c < 1$.

The uniform distribution of edges becomes impossible if further conditions are imposed on the graph; e.g., it is easy to see that there is no $G(n;[cn^2])$ which contains no triangle, and for which every spanned subgraph of $[\frac{n}{2}]$ vertices has $(1 + o(1))cn^2/_4$ edges. In fact this uniform distribution probably implies that our graph must contain (for $n \to \infty$) arbitrarily large complete graphs.

It is true that almost all graphs $G(n;[Cn])$ contain a path of length $> cn^2$. ("Almost all" here means all but $o\left(\binom{\binom{n}{2}}{Cn}\right)$ of the graphs $G(n;[Cn])$. I conjectured this and in fact believed that c tends to 1 as C tends to infinity. Szemerédi disagrees; he believes that for every C the longest path is almost surely $o(n)$. At present we can not decide who is right.

P. Erdös and J. Spencer, Probabilistic methods in combinatorics, Academic Press and Hungarian Academy of Sciences, 1974.

P. Erdös, Some problems in graph theory, Hypergraph Seminar, Lecture Notes in Math 411, Springer Verlag 187-190.

P. Erdös, Some new applications of probability methods, to combinatorial analysis and graph theory, Cong. Num. X, Proc. Fifth Southeastern Conference on Combinatorics, Graph Theory and Computing, Florida Atlantic University, 1974, 39-51.

4. Further problems.

Now I state a few (of the many) problems which arose in our work with Faudree, Rousseau and Schelp. $\hat{r}(G_1,G_2)$ is the smallest integer for which there is a graph G of $\hat{r}(G_1,G_2)$ edges so that if we color the edges of G with two colors I and II, either, color I contains G_1, or color I, G_2. The most annoying problem is to determine or estimate $\hat{r}(P_n,P_n)$, where P_n is a path of length n. We could not even prove that

$$(2)\quad \lim_{n=\infty} \hat{r}(P_n,P_n)/n = \infty$$

and

$$(3)\quad \lim \hat{r}(P_n,P_n)/n^2 = 0.$$

I give 25 dollars for a proof or disproof of either (2) or (3) (i.e. 50 for both) and 100 dollars for an asymptotic formula for $\hat{r}(P_n,P_n)$.

Denote by $f(n)$ the largest integer for which there is a $G(n,f(n))$ so that

$$r(K_3;G(n;f(n)) \leq 2n - 1.$$

$f(n) > cn \log n/\log\log n$. $f(n) < n^{5/3+e}$ follows easily by the probability method. We have no idea of the true order of magnitude of $f(n)$. ($r(G_1,G_2)$ is the smallest integer so that if we color the edges of $K(r(G_1,G_2))$ (i.e. the edges of the complete graph of $r(G_1,G_2)$ vertices) by two colors, either color I contains G_1 or color I, G_2).

Let $F(n)$ be the largest integer so that for every $\ell \leq F(n)$ and every $G(n;\ell)$

$$r(K_3;G(n;\ell)) \leq 2n - 1.$$

Clearly $f(n) \geq F(n)$. It seems certain that $f(n)/F(n) \to \infty$, $F(n)/n \to \infty$. We have no idea of the true order of magnitude of $F(n)$ and $f(n)$.

Denote by C_k the circuit of k edges. Graver, Yackel and I proved that

$$(4) \quad c_1 n^2/(\log n)^2 < r(C_3, K_n) < c_2 n^2 \log\log n/\log n.$$

It would be interesting to obtain an asymptotic formula for $r(C_3, K_n)$, but this will probably be very difficult. It seems certain that for $n > n_0(\varepsilon)$

$$(5) \quad r(C_4, K_n) < n^{2-\varepsilon}$$

for some $\varepsilon > 0$ independent of n. I give 100 dollars for a proof or disproof of (5).

So far only one of our quadruple papers has appeared: P. Erdös, R. J. Faudree, C. C. Rousseau and R. H. Schelp, Generalized Ramsey theory for multiple colors, J. Comb. Theory, ser. B 20(1976), 250-264. Several more papers will soon appear, some jointly with S. Burr.

The reader can find lots of information in the excellent review article by S. A. Burr, Generalized Ramsey theory for graphs, A survey, in "Graphs and combinatorics 1973" (R. Bari and F. Harary, Editors) Springer Verlag, Berlin, 1974. A new survey by Burr, a survey of noncomplete Ramsey theory for graphs, will soon appear.

For the results of Graver, Yakel and myself, see e.g., my book with Spencer quoted in the previous chapter.

$r(C_m, K_n) \leq \{(m-2)(n^{1/k}+2) - 1\}n - 1; k = [\frac{m-1}{2}]$ is proved in our quadruple paper "On cycle-complete graph Ramsey theorems" which will soon appear in the Journal of Graph Theory.

5. Erdös-Rado conjecture on Δ-systems.

Let $f_3(n)$ be the largest integer with the property that $|A_i| = n$, $1 \leq i \leq f_3(n)$, and no three A's have pairwise the same intersection. Rado and I conjectured more than twenty years ago that there is an absolute constant C so that

$$(6) \qquad f_3(n) < C^n.$$

I offer 500 dollars for a proof or disproof of (6). The best upper bound is due to Joel Spencer; he proved that for $n > n_0(\varepsilon)$

$$f_3(n) < (1 + \varepsilon)^n n!$$

Abbott, Hanson and others obtained lower bounds, Abbott in particular proved that $f_3(3) = 20$.

P. Erdös and R. Rado, Intersection theorems for systems of sets I and II, J. London Math. Soc. 35 (1960), 85-90 and 44(1969), 13-17.

P. Erdös, E. Milner and R. Rado, Intersection theorems for systems of sets III, J. Australian Math. Soc. 18(1974), 22-41.

See also a forthcoming paper of Szemerédi and myself where several further related problems will be stated. Our paper will soon appear in the Journal of Combinatorial Theory.

6. Work with Ulam and Selfridge.

Now I give full details of some work which I did with Ulam on a combinatorial problem and with Selfridge on a problem in combinatorial number theory. Both results are very far from being complete--but if we live we hope at least partially to remedy this situation.

First I discuss my joint work with Selfridge which is closer to being complete: Let $p_0, p_1, \ldots, p_u$ be a set of $u + 1$ primes. An interval of length x can expect to contain roughly speaking $\sum_{i=0}^{u} \left[\frac{x}{p_i}\right]$

multiples of the p's; if the interval is short it may happen that all or most of the multiples of the p's may coincide. Thus to avoid trivial cases we henceforth assume $x > 2p_u$. We prove the following

Theorem 1. Let $u = k^2 - 1$. To every $\varepsilon > 0$ there is a sequence of primes $p_0 < \ldots < p_u$ and an interval I of length $(3 - \varepsilon)p_u$, which contains exactly $2k$ distinct multiples of the p's. In other words the number of distinct integers a_j in I for which $a_j \equiv 0 \pmod{p_r}$ for some r, $0 \le r \le u$ is $2k$.

We further show that this result is best possible in the following strong sense: Every interval of length $> 2p_u$ contains at least $2k$ distinct multiples of the p's.

This result is complete and best possible as it stands. Unfortunately we know next to nothing for intervals longer than $3p_u$. In particular, is it true that for every C and ε there are primes $p_0 < \ldots < p_u$ and an interval of length $> Cp_u$ which contains fewer than εu distinct multiples of the p's? At present we can not answer this question.

First we prove that our Theorem is best possible. In other words we prove that if I has length $> 2p_u$ then it contains at least $2k$ distinct multiples of the p's. Let (a,b) be the interval I, $b - a > 2p_u$. Denote by I_1 the interval $(a, a + \frac{|I|}{2})$ and by I_2 the interval $(a + \frac{|I|}{2}, b)$. Both of these intervals contain at least $\sum_{i=0}^{u} \left[\frac{|I|}{2p_i}\right] \geq k^2$ multiples of the p's (counted by multiplicity). If no m in I is a multiple of more than k p's then clearly there are at least $2k$ distinct multiples of the p's in I. Thus assume that, say in I_1, there is an integer m which is divisible by r p's, r is maximal and $r > k$.

Then in I_1 there are at least $\{\frac{k^2}{r}\}$ distinct multiples of the p's (where $\{x\}$ is the least integer not less than X).

Let $p_{i_1},\ldots,p_{i_r}$ be the p's which divide m. Consider the smallest $s_j \geq 0$ for which $m + 2^{s_j}p_{i_j}$ is in I_2 - such an s_j clearly exists. The numbers $m + 2^{s_j}p_{i_j}$ are clearly all distinct; thus I contains at least $r + \{\frac{k^2}{r}\} \geq 2k$ distinct multiples of the p's, as stated.

Now we prove the main part of our Theorem.

First we need a Lemma which is of some independent interest:

Lemma. Put $u = k^2 - 1$. For every k and arbitrarily large N there are k^2 primes

$$N < p_0 < \ldots < p_u < N + (\log N)^{k+3}$$

satisfying for every $1 \leq i \leq k - 1; 1 \leq t \leq k - 1$

$$p_i - p_0 = p_{i+tk} - p_{tk}.$$

In other words there are k sets of k primes whose internal structure is the same.

Probably very much more is true: there is an f(k) and infinitely many primes p so that all the numbers $p + tf(k)$, $0 \leq t < k^2$ are primes--in fact consecutive primes. Needless to say it is quite hopeless at present to prove this conjecture and fortunately we do not need it.

The proof of the Lemma is by a simple counting argument. It follows from the prime number theorem (or a more elementary theorem) that for every L there is an interval (x, x + L) which contains more than

$\frac{L}{2 \log x}$ primes. Denote these primes by

$$x < q_1 < \ldots < q_w < x + L, \; w > \frac{L}{2 \log x}$$

Consider the $\frac{w}{k}$ differences $q_{uk+1} - q_{(u-1)k+1}$. We only retain those differences which are less than $4k \log x$; clearly, there are at least $\frac{L}{4k \log x}$ such differences. The number of patterns for these k primes $\{q_{(u-1)\ell+1},\ldots,q_{uk}\}$ is clearly less than $(4k \log x)^{k+1}$. Thus if $L > (4k \log x)^{k+2}$ there are at least k k-tuples of primes giving the same pattern, which completes the proof of our Lemma.

Now using the Chinese remainder theorem we are ready to complete the proof of our Theorem. Put

$$\alpha_i = \prod_{j=0}^{k-1} p_{ik+j}, \quad \beta_i = \prod_{\ell=0}^{k-1} p_{\ell k+i}, \quad 0 \leq i \leq k - 1.$$

Clearly $\prod_{i=0}^{k-1} \alpha_i = \prod_{i=0}^{k-1} \beta_i = \prod_{j=0}^{u} p_j$. Now we determine $x \pmod{\prod p_j}$ as follows:

$$x + p_i \equiv 0 \pmod{\beta_i}, \; x + p_{ik} \equiv p_0 \pmod{\alpha_i},$$
$$0 \leq i \leq k - 1.$$

A simple argument shows that the interval $(x - p_0 + 1, x + 2p_0 - 1)$ of length $(3 - \varepsilon)p_u$ contains only $2k$ multiples of the p's; namely, the unique multiples of $\alpha_1,\ldots,\alpha_k;\beta_1,\ldots,\beta_k$.

Unfortunately if the interval has length $(3+\varepsilon)p_u$ so to speak "all hell breaks loose" and we lose control of the set of multiples of the p's. We hope to return to this subject at a later occasion. The following related problem is also interesting: Determine the smallest $h(u)$ so that if $p_1 < \ldots < p_u$ is a sequence of u primes, every interval of length $f(u)p_u$ contains an integer divisible by precisely one p. Many related questions can be asked.

Ulam and I considered the following combinatorial problem: Let $|S| = n$ be a set. Split the 2^n subsets of S into two classes. Determine the largest integer $f(n)$ so that there always is a family of $f(n)$ sets in the same class which is closed with respect to taking unions and intersections. Here we only could make the trivial observation that $f(n) \geq \frac{n+1}{2}$ (since there is a sequence of length $n + 1$ of nested subsets). We have no plausible conjecture for the true order of magnitude of $f(n)$. Denote by $F(n)$ the largest integer for which there is a family of $F(n)$ sets of the same class which is closed with respect to taking unions. Here we conjecture that $F(n) > n^c$ for every c if $n > n_0(c)$. From above we conjecture $F(n) < (1 + \varepsilon)^n$ for every $\varepsilon > 0$ if $n > n_0(\varepsilon)$. We have no good guess about the true order of magnitude of $F(n)$.

An older result substantially due to R. Rado and J. Sanders stated that for every k there is an n_k so that if $n \geq n_k$ and we divide the subsets of $|S| = n$ into two classes there are always k disjoint subsets so that all the $2^k - 1$ unions are in the same class. Unfortunately the proof gives for n_k an exorbitantly fast rate of growth which probably (?) does not describe the true state of affairs. We can show that n_k tends to infinity exponentially and in fact we prove the following more general

Theorem 2. Let $|S| = n$. There is a division of the subsets of S into two classes so that if $A_i \subseteq S$, $1 \leq i \leq k$ are such that all the $2^k - 1$ unions $A_{i_1} \dots A_{i_r}$ are distinct and belong to the same class then $k \leq (1 + o(1)) \log n / \log 2$.

To prove Theorem 2 observe that the subsets of S (not counting the empty set) can be divided into two classes in $2^{2^n - 1}$ ways. Now we estimate the number of those divisions into two classes for which there are k sets $A_1,\ldots,A_k$ in the same class so that all the $2^k - 1$ unions formed from them are distinct and are in the same class.

The k sets $A_1,\ldots,A_k$ can be chosen in at most

$$\binom{2^n}{k} < 2^{kn}$$

ways. Once the sets $A_1,\ldots,A_k$ have been chosen, since the $2^k - 1$ unions are assumed to be all distinct and in the same class there are

$$2 \cdot 2^{2^n - 1}\, 2^{-(2^k - 1)} = 2^{2^n - 2^k + 1}$$

ways of splitting the subsets into two classes so that all the unions of the sets $A_1,\ldots,A_k$ should be in the same class. Now

$$2^{kn} 2^{2^n - 2^k + 1} < 2^{2^n - 1}$$

if

$$(2) \quad 2^{2^k - 2} > 2^{kn}$$

A simple calculation gives that (2) holds if $k > (1 + o(1)) \frac{\log n}{\log 2}$. Thus there is a splitting of the subsets into two classes so that there should be no k subsets all whose $2^k - 1$ unions are distinct and in the same class if $k > (1 + \varepsilon) \frac{\log n}{\log 2}$.

We can not get at present a better upper bound even if we assume that the A's are disjoint, and in neither case has it been possible to obtain an acceptable lower bound.

Assume that we split the subsets into two classes in such a way that subsets of the same size belong to the same class. In this case Howorka proved that for every c and $n > n_0(c)$, $F(n) > n^c$.