Reprinted from JOURNAL OF COMBINATORIAL THEORY, Series A
All Rights Reserved by Academic Press, New York and London

Vol. 24, No. 3, May 1978
Printed in Belgium

Combinatorial Properties of Systems of Sets

P. ERDÖS AND E. SZEMERÉDI

Communicated by the Managing Editors

Received August 4, 1977

DEDICATED TO JOHN RIORDAN ON THE OCCASION OF HIS 75TH BIRTHDAY

A family of sets $\{A_k\}$ is called a strong Δ system if the intersection of any two of its members is the same, i.e., if $A_{k_1} \cap A_{k_2} = A_{k_3} \cap A_{k_4}$. It is called a weak Δ system if $|A_{k_i} \cap A_{k_j}|$ is the same for any two sets of our family. Δ systems have recently been studied in several papers. $f(n, r)$ is the smallest integer for which any family of $f(n, r)$ sets A_k, $1 \leqslant k \leqslant f(n, r)$ of size n $|A_k| = n$, $1 \leqslant k \leqslant f(n, r)$ contains a subfamily of r sets $\{A_{k_l}\}$ $1 \leqslant l \leqslant r$ which form a strong Δ system, $g(n, r)$ is the smallest integer for which every family of $g(n, r)$ sets A_k, $1 \leqslant k \leqslant g(n, r)$ of size n contains a subfamily of r sets $\{A_{k_l}\}$, $1 \leqslant l \leqslant r$ which form a weak Δ system.

Erdös and Rado [1] proved

$$2^n < f(n, r) < n!\, 2^n \tag{1}$$

and conjectured $f(n, r) < c_r^n$. This attractive and striking conjecture is open even for $r = 3$. Both the upper and the lower bound in (1) have been improved by Abbott, Hanson, and others but it is not yet known if

$$f(n, 3) < n!/A^n \tag{2}$$

for every A if $n > n_0(A)$. The sharpest upper bound is due to Spencer [4]; he shows $f(n, 3) < (1 + o(1))\, n!$. Thus it is not even known that for $n > n_0$ $f(n, 3) < n!$.

Trivially $f(n, r) \geqslant g(n, r)$. Erdös *et al.* [2] proved $g(n, r) \geqslant 5 \cdot 2^{n-2}$ and noted they cannot even prove $g(n, r) < (n!)^{1-\epsilon}$. Hanson determined $g(n, 3)$ for $n = 5$ and Abbott showed $f(3, 3) = 21$.

Denote by $H_n(3)$ the smallest integer with the property that if we color the edges of $K(H_n(3))$ by n colors ($K(H_n(3))$ is a complete graph on $H_n(3)$ vertices), there is a monochromatic triangle. In [2] $g(n, 3) < H_n(3) < en!$ is proved, but as far as we know no real progress has been made on these problems.

Let now S, $|S| = n$ be a set. $\mathbf{F}(n, r)$ is the largest integer so that there is a family $\{A_k\}$ of subsets of S, $1 \leqslant k \leqslant \mathbf{F}(n, r)$ which does not contain a strong Δ system of r elements. It is easy to see by the probabilistic method that

$$\mathbf{F}(n, r) > (1 + c_r)^n \qquad (c_r > 0 \text{ for } r > 3)$$

where $c_r \to 1$ as $r \to \infty$. It is easy to see that

$$\lim_{n\to\infty} \mathbf{f}(n, r)^{1/r} \to c_r + 1$$

exists but we cannot even prove $c_3 < 1$.

Abbott noticed that it is not easy to construct a family $\{A_k\}$ $A_k \subset S$, $1 \leqslant k \leqslant t_n$, $t_n/n \to \infty$ so that no three subsets $A_{k_1}, A_{k_2}, A_{k_3}$ form a weak Δ system. Define $G(n, r)$ as the largest family of subsets $\{A_k\}$, $1 \leqslant k \leqslant G(n, r)$ of S which do not contain a weak Δ system of r elements. We unfortunately cannot even prove

$$G(n, 3) < (2 - \epsilon)^n \tag{3}$$

for some $\epsilon > 0$ and all n. On the other hand, we are going to prove the following

THEOREM 1. *There exists a family* $\mathbf{F}$ *of subsets of a given set* S *so that* $\mathbf{F}$ *does not contain a weak* Δ *system, where*

$$|S| = n, \qquad |\mathbf{F}| \geqslant n^{\log n/4 \log\log n}.$$

Theorem 1 answers the question of Abbott but still leaves a tremendous gap in our knowledge.

Equation (3) follows from an older conjecture of Erdös which states the following: To every $\eta > 0$ there is an $\epsilon > 0$ so that if $A_k \subset S$, $1 \leqslant k \leqslant (2 - \epsilon)^n$, then for every j, $\eta n < j < (\frac{1}{2} - \eta)\, n$ there are two sets A_{k_1} and A_{k_2} of our family with $|A_{k_1} \cap A_{k_2}| = j$. This conjecture would have many applications. Here we outline the deduction of (3) from it.

Let $|S| = n$, $A_k \subset S$, $1 \leqslant k \leqslant (2 - \epsilon)^n$. Without loss of generality we can of course assume

$$(\tfrac{1}{2} - 2\epsilon^{1/2})\, n < |A_k| < (\tfrac{1}{2} + 2\epsilon^{1/2})\, n \tag{4}$$

since the number of the sets not satisfying (4) is easily seen to be $<\frac{1}{2}(2 - \epsilon)^n$. It is further easy to see by well-known asymptotic properties of the binomial coefficients that for one of the A's, say A_1, there are $(2 - 2\epsilon)^n$ A_k for which

$$n(\tfrac{1}{4} - 5\epsilon^{1/2}) < |A_1 \cap A_k| < (\tfrac{1}{4} + 5\epsilon^{1/2})\, n.$$

Hence there are at least $(2 - 2\epsilon)^n/n$ of them for which $|A_1 \cap A_k| = j$ where

$$(\tfrac{1}{4} - 5\epsilon^{1/2})\, n < j < (\tfrac{1}{4} + 5\epsilon^{1/2})\, n. \tag{5}$$

Now by the conjecture there are two sets A_{k_1} and A_{k_2} satisfying (4) and also $|A_{k_1} \cap A_{k_2}| = j$, but then A_1, A_{k_1}, A_{k_2} form a weak Δ system as stated.

We have no idea if Theorem 1 is best possible. There is a good chance that any set of $(1 + \epsilon)^n$ subsets of S, $|S| = n$ contains for $n > n_0(\epsilon)$ a weak Δ system of three (or more generally of r elements, for $n > n_0(\epsilon, r)$), and we do not even have a good guess for the true order of magnitude of $G(n, r)$.

The following general conjecture is probably relevant here. Let $A_k \subset S$, $|S| = n$, $1 \leqslant k \leqslant (1 + \epsilon)^n$. We conjecture that there is a subfamily $\{A_{k_i}\}$, $1 \leqslant i \leqslant (1 + \epsilon_1)^n$ and a γ so that

$$|A_{k_{i_1}} \cap A_{k_{i_2}}| = (\gamma + O(1))\, n. \tag{6}$$

Equation (5) is quite enough to deduce our conjecture, in fact we have to replace $\gamma + O(1)$ by

$$\gamma + \frac{1}{\log n} + O\left(\frac{1}{\log n}\right)$$

but the form (6) is of course more elegant. Without loss of generality we can assume that all the A_{k_i} are of size cn and if γ exists, it is, by well-known reasoning which goes back to at least Gillis and Khintchine, $\gamma \geqslant c^2$ [5].

Just one word about the difficulty of proving (6). If we have $f(n)$ sets $A_1, \ldots, A_{f(n)}$ of size cn where $f(n) \to \infty$, as n tends to infinity, it is easy to obtain by Ramsey's theorem that there are $f(n)^{c_\epsilon}$ sets $A_{i_1}, A_{i_2}, \ldots, A_{i_l}$, $l \geqslant f(n)^{c_\epsilon}$ so that

$$(a - \epsilon)\, n < |A_{i_{j_1}} \cap A_{i_{j_2}}| < (\alpha + \epsilon)\, n, \qquad 1 \leqslant j_1 < j_2 \leqslant l,\ \alpha > c^2. \tag{7}$$

c_ϵ tends to 0 as $\epsilon \to 0$. The proof only uses Ramsey's theorem; we could not utilize the fact that our sets are subsets of size cn of a set of n elements. To get sharper results we no doubt would have to use this fact. We omit the proof of (7) since it uses standard arguments.

For strong Δ systems we only can prove

THEOREM 2. *Let* $\{A_i\}$, $1 \leqslant i \leqslant t$, $t > 2^{(1-1/10(n^{1/2}))n}$, $A_i \subset S$, $|S| = n$. *Then there are three A's which form a strong Δ system.*

Obviously there is an l so that $|A_i| = l$ for at least t/n values of i. Let $\{A_i\}$, $1 \leqslant i \leqslant s$, $s > t/n$ be the subsets of size l of our system. For each A_i

consider all subsets of A_i of size $l - [\frac{1}{2}n^{1/2}]$. The total number of these subsets counted with multiplicity is clearly

$$s\binom{l}{[\frac{1}{2}m^{1/2}]}.$$

The total number of subsets of s of size $l - [\frac{1}{2}n^{1/2}]$ is clearly

$$\binom{n}{l - [\frac{1}{2}n^{1/2}]}.$$

Thus the same set occurs in at least u sets A_i where

$$u \geqslant \frac{s\binom{l}{[\frac{1}{2}n^{1/2}]}}{\binom{n}{l - [\frac{1}{2}n^{1/2}]}} \geqslant f([\tfrac{1}{2}n^{1/2}], 3). \tag{8}$$

Denote this set by B. Consider $A_i - B$ for all A_i which contain B. We have $|A_i - B| = [\frac{1}{2}n^{1/2}]$. By the Theorem of Erdös–Rado there are three A_i's, say A_1, A_2, A_3 for which the sets $A_1 - B$, $A_2 - B$, $A_3 - B$ form a strong Δ system and then clearly A_1, A_2, A_3 also form a strong Δ system, which completes the proof of Theorem 2.

The following questions are of some interest and use: Denote by $\mathbf{F}(n, k, r)$, respectively, $\mathbf{G}(n, k, r)$ the cardinality of the largest family of subsets of S, $|S| = n$ of sets of size k which do not contain a subfamily of size r forming a strong (respectively, weak) Δ system of size r. Let us restrict ourselves to $r = 3$. For $k < \log n/\log\log n$ we have $\mathbf{F}(n, k, 3) = f(n, 3)$ and $\mathbf{G}(n, k, 3) = g(n, 3)$ but as k increases we get interesting problems. It follows from the probability method that for $\epsilon n < k < (1 - \epsilon)\, n$

$$\mathbf{F}(n, k, 3) > \binom{n}{k}^{\eta} \qquad \eta = \eta(\epsilon)$$

but, say, if $k = (\log n)^c$, c large, or $k = \epsilon^{(\log n)^{1/2}}$ we have no useful upper or lower bounds for $\mathbf{F}(n, k, 3)$ or $\mathbf{G}(n, k, 3)$. Also, as will be seen later it would be very useful if we could prove $\mathbf{G}(n, \log n^2, 3) > n^{2+c}$ for some $c > 0$. Frankl observed that by the method of Erdös *et al.* [3] it is easy to prove that if we are given more than $s\binom{n-1}{k-1}$ k-element sets of an n set then there are at least $s + 1$ pairwise disjoint sets among them. Consequently

$$F(n, k, r) \leqslant r\binom{n-1}{k-1}. \tag{9}$$

Now we prove Theorem 1. First we need a lemma from [2].

LEMMA. *Let S be a set of size n. One can give a family of sets $A_k \subset S$, $|A_k| = c_1 \log n$, $1 \leqslant k \leqslant c_2 n$ so that the family $\{A_k\}$ does not contain a weak Δ system of three elements.*

The constants c_1 and c_2 could easily be determined but this is not worthwhile. On the other hand, it would be very useful if we could decide the following question. Let $|S| = n$, $A_k \subset S$, $1 \leqslant k \leqslant T$, $|A_k| = (\log n)^r$ be a family of sets which does not contain a weak Δ system of three elements. As will be seen our construction gives that T can be as large as $c_r n^r$. Can it be larger? We do not even know what happens for $r = 2$. Can one get more than $n^{2+\epsilon}$ sets of size $(\log n)^2$ which are all subsets of a set of size n no three of which form a weak Δ system? If we could do this we could immediately improve Theorem 1, but we feel that this problem is very interesting for its own sake.

The proof of the lemma is very simple. Let $2^{k+1} \leqslant n$ and consider a binary tree of length k; the vertices of the tree are the elements, the paths of length k are our sets. It is immediate that the sets corresponding to the paths do not contain a weak Δ system of three terms and this completes the proof of our lemma.

Now we are ready to prove Theorem 1.

Let $|S_r| = [(\log n)^r][n^{1/2}]$. In S_r we construct a binary tree as given by our Lemma—the individual vertices of our tree are sets of size $[(\log n)^r]$ and the length of tree is $\log n/2 \log 2$. Our set will be the $\cup S_r$, $1 \leqslant r \leqslant \log n/2 \log 2 \log \log n$, thus our set has fewer than n elements. Denote the sets defined by our binary tree in S_r by $B_j^{(r)}$, $1 \leqslant j \leqslant [n^{1/2}]$, $|B_j^{(r)}| = [(\log n)^r][\log n/2 \log 2]$. Now finally our sets which do not form a weak Δ-system are the sets

$$\bigcup_r B_j^{(r)}, \qquad 1 \leqslant r \leqslant \frac{\log n}{2 \log 2 \log \log n}, \qquad 1 \leqslant j \leqslant [n^{1/2}].$$

The number of these sets is

$$n^{1/2} \log n/2 \log 2 \log \log n$$

and it is easy to see that no three sets form a weak Δ system. To see this let A_1, A_2, A_3 be three sets of our system. We will refer to the $B_j^{(r)}$ as coordinates of our sets A_1, A_2, A_3. Assume that all three sets have the same coordinates for $r > r_0$ but no longer for r_0. Since $B_{j_1}^{(r_0)}$, $B_{j_2}^{(r_0)}$, $B_{j_3}^{(r_0)}$ do not form a weak Δ system we can assume without loss of generality

$$|B_{j_1}^{(r_0)} \cap B_{j_2}^{(r_0)}| - |B_{j_1}^{(r_0)} \cap B_{j_3}^{(r_0)}| \geqslant (\log n)^{r_0} \tag{10}$$

since the "elements" of $B_j^{(r_0)}$ are sets of size $(\log n)^{r_0}$. Observe that

$$\bigcup_{r<r_0} B_j^r < (\log n)^{r_0} .$$

Thus the "damage" done by (9) cannot be repaired.

REFERENCES

1. 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), 467–479; H. L. ABBOTT, Some remarks on a combinatorial theorem of Erdös and Rado, *Canad. Math. Bull.* **9** (1966), 155–160; H. L. ABBOTT AND D. HANSON, Intersection theorems for systems of sets, *J. Combinatorial Theory* **12** (1972), 381–389; P. ERDÖS, Problems and results on finite and infinite combinatorial analysis, *Coll. Math. Soc. Y. Bólyai* **10**; Infinite and finite sets, *Kegzthely* **1** (1973), 405–406.
2. P. ERDÖS, E. C. MILNER, AND R. RADO, Intersection theorems for systems of sets (III), *Austral. J. Math.* **18** (1974), 22–40.
3. P. ERDÖS, C. KO, AND R. RADO, Intersection theorem for systems of finite sets, *Quart. J. Math.* **12** (1961), 373–420.
4. J. GILLIS, Note on a property of measurable sets, *J. London Math. Soc.* **11** (1936), 139–141.