Europ. J. Combinatorics (1984) **5**, 127–131

Union-free Hypergraphs and Probability Theory

PETER FRANKL AND ZOLTAN FÜREDI

Let $F(n)$ denote the maximum number of distinct subsets of an n-element set such that there are no four distinct subsets: A, B, C, D with $A \cup B = C \cup D$. We prove that $2^{(n-\log 3)/3} - 2 \leq F(n) \leq 2^{(3n+2)/4}$. We use probability theory for the proof of both the lower and upper bounds. Some related problems are considered, too.

1. INTRODUCTION

In 1969 Erdös and Moser [4] raised the problem of estimating $f(n)$, the maximum number of distinct subsets of an n-element set such that all the $\binom{f(n)}{2}$ pairwise unions are different.

THEOREM 1. $$2^{(n-3)/4} \leq f(n) \leq 1 + 2^{(n+1)/2}. \tag{1}$$

Notice that the upper bound is an immediate consequence of $\binom{f(n)}{2} \leq 2^n$. To prove the lower bound we use an algebraic construction which is a modification of a construction of Babai and Sós [1]. How a family of sets can fail to have the union-free property? There are essentially two possibilities:

(a) there are four distinct sets A, B, C, D with $A \cup B = C \cup D$.
(b) there are three distinct sets A, B, C with $A \cup B = A \cup C$.

We call families for which (a) never holds *weakly union-free*, and those for which (b) never holds *cancellative* (the second name indicates that $A \cup B = A \cup C$ implies $B = C$). We denote by $F(n)(G(n))$ the maximum number of subsets of an n-set in a weakly union-free (cancellative) family, respectively.

Our main result is the following:

THEOREM 2. $$2^{(n-\log 3)/3} - 2 \leq F(n) \leq 2^{(3n+2)/4} \sim 2^{1/2} \cdot 1.68^n. \tag{2}$$

The lower bound is deduced by a non-constructive, probabilistic method. The proof of the upper bound uses information theory, it was inspired by the paper Kleitman, Shearer and Sturtevant [9]. For cancellative families we prove:

THEOREM 3. $$(8/9)^{\varepsilon(n)/3} 3^{n/3} \leq G(n) < n\,1.5^n \qquad (n \geq 14), \tag{3}$$

where $\varepsilon(n)$ is determined by $0 \leq \varepsilon(n) \leq 2$, $n + \varepsilon(n)$ is divisible by 3.

Erdös and Katona (cf. [8]) conjecture that the lower bound is exact. Their construction is simple: let $X_1, \ldots, X_q$ be pairwise disjoint sets with union of size n with $|X_i| = 2$ or 3 and with at most two sets of size 2 among the X_i. Let our family consist of all the transversals that is of those sets which intersect each X_i in one element. Clearly this family achieves the lower bound and it is cancellative.

2. RELATED AND OPEN PROBLEMS

Let k be an integer, $k \geq 2$. Let us denote by $f_k(n)$ the maximum number of k-subsets of an n-set forming a union-free family, $F_k(n)$, $G_k(n)$ are defined similarly. Then $f_2(n)$,

0195–6698/84/020127+05 \$02.00/0

$F_2(n)$, $G_2(n)$ denote the maximum number of edges in a graph without a cycle of length 3 or 4, of length 4, of length 3, respectively. The problem of determining $F_2(n)$ was raised by Erdös [3] already 45 years ago, but it is still unsolved. However it is known that

$$F_2(n)=\left[1(1+o(1))\frac{n^{3/2}}{2}+o(1)\right]. \tag{4}$$

Recently the second author determined the exact value of $F_2(n)$ for $n=4^s+2^s+1$. He proved: (cf. [7])

$$F_2(n)=2^{s-1}(2^s+1)^2. \tag{5}$$

For $f_2(n)$ it is only known that

$$\frac{1}{2\cdot 2^{1/2}}n^{3/2}<f_2(n)<\tfrac{1}{2}n^{3/2}. \tag{6}$$

The determination of $G_2(n)$ is a special case of Turan's theorem ([11]):

$$G_2(n)=[n^2/4]. \tag{7}$$

For $n=3$ the authors proved in [8]:

$$f_3(n)=[n(n-1)/6], \tag{8}$$

and

$$F_3(n)=n(n-1)/3 \quad \text{for } n>n_0 \quad \text{and} \quad n\equiv 1 \pmod 6. \tag{9}$$

Bollobás [2] proved:

$$G_3(n)=\left[\frac{n}{3}\right]\left[\frac{n+1}{3}\right]\left[\frac{n+2}{3}\right]. \tag{10}$$

For $k\geqslant 4$ no exact values are known. The authors have established several bounds for $f_k(n)$ and $F_k(n)$, e.g. (cf. [6]):

$$f_4(n)=[1+o(1)]n^3/24. \tag{11}$$

For $G_k(n)$ Bollobás [2] conjectures that

$$G_k(n)=\prod_{0\leqslant i<k}\left[\frac{n+i}{k}\right]. \tag{12}$$

It is easy to see that this is a lower bound for $G_k(n)$. We prove the conjecture for $n\leqslant 2k$.

PROPOSITION 2.1. *For $n\leqslant 2k$ we have*

$$G_k(n)=2^{n-k}. \tag{13}$$

COROLLARY 2.2. *For $n\geqslant 2k$ we have*

$$G_k(n)\leqslant\binom{n}{k}2^k\Big/\binom{2k}{k}. \tag{14}$$

For the problems considered in detail in this paper the most important would be to determine $\lim_{n\to\infty}\log h(n)/n$ where h is any of f, F and G. For f and F it is not even proved yet that this limit exists, for G it follows from $G(n_1+n_2)\geqslant G(n_1)G(n_2)$.

Let us note that equation (12) would imply $\lim_{n\to\infty}\log G(n)/n=3^{1/3}=1{\cdot}44\ldots$ The upper bound of Theorem 3 gives $1{\cdot}5$.

3. The Proof of the Upper Bound of Theorem 2

Let $\mathscr{F}=\{F_1,\ldots,F_m\}$ be any weakly union-free family of subsets of $\{1,\ldots,n\}$. Let $\mathbf{v}_i$ be the characteristic vector F_i: $\mathbf{v}_i$ is a (0, 1)-vector which has 1 in the jth position if and only if $j\in F_i$. The following proposition can be proved easily.

Proposition 3.1. *The* $\binom{m+1}{2}$ *sums* $\mathbf{v}_i+\mathbf{v}_{i'}$ $(1\leq i\leq i'\leq n)$ *are all distinct* $(0,1,2)$*-vectors of length* n.

Notice that this proposition already implies $\binom{m+1}{2}\leq 3^n$, in particular $m<3^{(m+1)/2}$. However, we want to show that the considerably stronger inequality (2) is valid. Let us give weights to the vectors $\mathbf{v}_i+\mathbf{v}_{i'}$. Let the weight, $w(\mathbf{v}_i+\mathbf{v}_{i'})$ be 1 if $i=i'$ and 2 if $i\neq i'$. Then the total sum of weights is m^2. Let us define a probability distribution $\mathbf{x}$ on these sums by setting $p(\mathbf{x}=\mathbf{v}_i+\mathbf{v}_{i'})=w(\mathbf{v}_i+\mathbf{v}_{i'})/m^2$. Then $\mathbf{x}$ can be considered as a random vector $\mathbf{x}=(x_1,\ldots,x_n)$ where $\mathbf{x}_j$ is the frequency distribution of 0s, 1s and 2s in the jth position. If d_j denotes the degree of j in $\mathscr{F}$, i.e., the number of sets containing j and $p_j=d_j/m$, then x_j is given by $p(x_j=2)=p_j^2$, $p(x_j=1)=2p_j(1-p_j)$, $p(x_j=0)=(1-p_j)^2$. Thus the information-theoretic entropy of x_j is:

$$H(x_j)=-p_j^2\log p_j^2-2p_j(1-p_j)\log 2p_j(1-p_j)-(1-p_j)^2\log(1-p_j)^2, \tag{15}$$

log means $\log_2$. The next proposition can be proved by elementary analysis:

Proposition 3.2. *The function in equation* (15) *takes its maximum value for* $p_j=\frac{1}{2}$ *where* $H(x_j)=\frac{3}{2}$.

The next proposition is from [10, p. 33].

Proposition 3.3 *If* $\mathbf{x}=(x_1,\ldots,x_n)$ *is a random vector then*

$$H(\mathbf{x})\leq\sum_{1\leq j\leq n}H(x_j). \tag{16}$$

Let us now count $H(\mathbf{x})$.

$$H(\mathbf{x})=-m\left(\frac{1}{m^2}\log(1/m^2)\right)-\binom{m}{2}\frac{2}{m^2}\log\left(\frac{2}{m^3}\right)=\log\left(\frac{m^2}{2}\right)+\frac{1}{m}\log 2>\log\left(\frac{m^2}{2}\right). \tag{17}$$

Now combining expressions (5), (6) and Proposition 2.2 we obtain $m^2/2<2^{3n/2}$, yielding the upper bound of expression (2).

4. The Lower Bound of Theorem 2

Let us consider a random (0, 1)-matrix of size $2m$ by n where each element is 1 with independent probability p (we shall fix m and p later). Each row of the matrix is the characteristic vector of a subset of $\{1,\ldots,n\}$. Let $\mathscr{F}$ denote the collection of the corresponding (not necessarily distinct) sets. The probability that some 4 sets in $\mathscr{F}$ satisfy (a) is $\{1-2(1-p)^2[1-(1-p)^2]\}^n$. This quantity becomes 2^{-n} for $p=(1-2^{1/2})/2$. If we choose m at most $2^{(n-\log 3)/3}$ then the expected number of four-tuples in $\mathscr{F}$, satisfying (a) is at most m. Omitting one set from each of these four-tuples we omit at most m sets, i.e. at least m sets remain and since (a) is impossible for these sets, at most one of them appears twice. Consequently, $F(n)\geq 2^{(n-\log 3)/3}-2$.

5. The Proof of Theorem 1

We only have to prove the lower bound. First let us note: arguing in the same way as for the lower bound of Theorem 2 but choosing $p=1/3$ we can get as many as $(1+o(1)($ $27/19)^{n/2}$ sets forming a union-free system, e.g. for $n>1000$ we obtain

$$f(n)>\tfrac{1}{2}(27/19)^{n/2}. \tag{18}$$

The inequality is actually stronger than that in Theorem 1, however it is non-constructive and valid only for large values of n.

To give the other bound it will be enough to show that for every positive integer n we have

$$f(4n)\geq 2^n. \tag{19}$$

To prove this inequality, let us consider 4 pairwise disjoint n-element sets: X, X', Y, Y' and let us fix 4 embeddings of $GF(2^n)$ into 2^X, $2^{X'}$, 2^Y, $2^{Y'}$, respectively: g, g', h, h'. Let 1 denote the element $(1, 1, \ldots, 1)$ in $GF(2^n)$. Now let us define:

$$\mathscr{A}=\{g(a)\cup g'(1-a)\cup h(a^3)\cup h(1-a^3)\colon a\in GF(2^n)\}.$$

We have to show that $\mathscr{A}$ is union-free. Suppose a, b, c, d are elements of $GF(2^n)$ for which the corresponding sets satisfy (a) or (b). Then $g(a)\cup g(b)=g(c)\cup g(d)$ and also $g'(1-a)\cup g'(1-b)=g'(1-c)\cup g'(1-d)$. The second equality yields $g'(a)\cap g'(b)=g'(c)\cap g'(d)$. We infer $a+b=c+d$. Similarly, from the equalities for h and h', it follows that $a^3+b^3=c^3+d^3$. However over a field of characteristic 2 we have: $a^3+b^3=(a+b)[(a+b)^2+ab]$. Since $a+b=c+d$, we infer $ab=cd$ from $a^3+b^3=c^3+d^3$. Thus $\{a, b\}$ and $\{c, d\}$ are both the set of roots of the equation $x^2-(a+b)x+ab=0$ i.e. $\{a, b\}=\{c, d\}$.

6. The Proof of the Bounds (13) and (14)

Let $\mathscr{A}$ be a cancellative family and let A be a member of $\mathscr{A}$ with maximal cardinality, say k. Then $A\cup B\neq A\cup C$ implies $B\cap(\{1,\ldots,n\}-A)\neq C\cap(\{1,\ldots,n\}-A)$ for $B, C\in(\mathscr{A}-\{A\})$. Thus

$$|\mathscr{A}|\leq 1+2^{n-k}. \tag{20}$$

Now assume that $\mathscr{A}$ is k-uniform that is all its members have the same size: k. Then $B\cap(\{1,\ldots,n\}-A)=\varnothing$ is impossible for $B\in(\mathscr{A}-\{A\})$, yielding equation (13), as an upper bound. To show that we have equality, let us partition $\{1,\ldots,n\}$ into k sets $X_1,\ldots,X_k$ such that $2k-n$ of them have size 1 and the remaining ones 2. Let $\mathscr{A}$ be the complete k-partite graph that is

$$\mathscr{A}=\{A\colon |A\cap X_i|=1 \text{ for every } 1\leq i\leq k\}.$$

We prove inequality (14) by a simple averaging argument. Suppose that $\mathscr{A}$ is a k-uniform, cancellative hypergraph on $X=\{1,\ldots,n\}$, $n\geq 2k$. Let Y be a random $2k$-element subsets of X. Set $\mathscr{A}_Y=\mathscr{A}\cap\binom{Y}{k}$. Then $\mathscr{A}_Y$ is cancellative. Thus equation (13) implies

$$|\mathscr{A}_Y|\leq 2^k. \tag{21}$$

Denoting by $E(|\mathscr{A}_Y|)$ the expected number of edges in $\mathscr{A}_Y$, we have

$$E(|\mathscr{A}_Y|)=|\mathscr{A}|\binom{2k}{k}\Big/\binom{n}{k}. \tag{22}$$

Since the expectation can not be greater than the maximum, expressions (21) and (22) imply inequality (14).

7. The Proof of Theorem 3

We need the following simple inequality:

$$\binom{2k}{k} > 2^{2k}/(2k)^{1/2}, \qquad \text{if } k \geq 7. \tag{23}$$

To prove expression (23), notice that it holds for $k=7$. Then apply induction. Passing from k to $k+1$ the LHS of expression (23) grows by a factor of $4(2n+1)/(2n+2)$, while the RHS by a factor of $4(2n/2n+2)^{1/2}$. Now, comparing these two, expression (23) follows from $2n+1 > (2n(2n+2))^{1/2}$.

Suppose now that $\mathscr{A}$ is a cancellative family on $\{1, \ldots, n\}$. Let A be a member of $\mathscr{A}$ having maximal size. If $|A| \geq n/2$ then inequality (20) yields expression (3). Thus we may suppose $|A| < n/2$. Let a_k denote the number of k-element subsets in $\mathscr{A}$. By definition we have:

$$a_k \leq G_k(n) \qquad \text{and} \qquad |\mathscr{A}| = \sum_{0 \leq k \leq n/2} a_k.$$

Thus inequality (14) implies

$$|\mathscr{A}| \leq \sum_{0 \leq k \leq n/2} \binom{n}{k} 2^k \Big/ \binom{2k}{k}.$$

Using expression (23), for $n \geq 14$ we infer

$$|\mathscr{A}| < n^{1/2} \sum_{0 \leq k \leq n} 2^{-k} \binom{n}{k} = n^{1/2} \left(\frac{3}{2}\right)^n.$$

References

1. L. Babai and V. T. Sós, Sidon sets in groups and induced subgraphs of Cayley graphs. *Eur. J. Combinatorics* (to appear).
2. B. Bollobás, Three-graphs without two triples whose symmetric difference is contained in a third. *Discr. Math.* **8** (1974), 21–24.
3. P. Erdös, On sequences of integers no one of which divides the product of two others and on some related problems. *Mitteilungen des Forschunginst. für Math. und Mechanik, Tomsk* **2** (1938), 74–82.
4. P. Erdös and L. Moser, Problem 35. *Proc. Conf. Combin. Structures and Appl. Calgary,* 1969, Gordon and Breach, New York, 1970, p. 506.
5. P. Frankl, Z. Füredi, A new extremal property of Steiner triple-systems (to appear).
6. P. Frankl and Z. Füredi, Union-free hypergraphs and equations over fields (to appear).
7. Z. Füredi, Graphs without quadrilaterals, *J. Combin. Theory, Ser. B* **34** (1983), 187–190.
8. G. O. H. Katona, Extremal problems for hypergraphs, Combinatorics, *Mathematical Centre Tracts* 56, *Part* 2, pp. 13–42.
9. D. J. Kleitman, J. Shearer, D. Sturtevant, Intersections of *k*-element sets. *Combinatorica* **1** (1981), 381–384.
10. R. McEliece, The theory of information and coding, *Encyclopedia of Mathematics, Vol.* 3, Addison-Wesley, Reading, MA, 1977.
11. P. Turan, On an extremal problem in graph theory (in Hungarian), *Mat. Lapok* **48** (1941), 436–452.

Received 14 January 1983 and in revised form 10 October 1983

Peter Frankl
C.N.R.S, E.R.175 'Combinatoire', C.M.S., 54, Boulevard Raspail, 75270 Paris, Cedex 06, France

Zoltán Füredi
Mathematical Institute of the Hungarian Academy of Sciences Budapest V., Realtanoda u. 13–15, Hungary