Graphs and Combinatorics 3, 341–347 (1987)

Graphs and Combinatorics

Bounding One-Way Differences

P. Frankl[1], Z. Füredi[2] and J. Pach[2]

[1] C.N.R.S. 15 Quai Anatole France, Paris, Cedex 75007, France

[2] Mathematical Institute of the Hungarian Academy of Science, 1364 Budapest, P.O. Box 127, Hungary

Abstract. Let $f(n,k)$ denote the maximum length of a sequence $(F_1, F_2, \dots)$ of distinct subsets of an n-element set with the property that $|F_i \backslash F_j| < k$ for all $i < j$. We determine the exact values of $f(n,2)$ and characterize all the extremal sequences. For $k \geq 3$ we prove that $f(n,k) = (1 + o(1))\binom{n}{k}$. Some related problems are also considered.

1. Introduction

Let $\mathscr{F}$ be a system of distinct subsets of an n-element set X, and let $k \geq 2$ be a fixed natural number. It is well-known (see [7], [12] or (3) below) that if $|F_i \backslash F_j| < k$ for all $F_i, F_j \in \mathscr{F}$ then $|\mathscr{F}| \leq \sum_{i=0}^{k-1} \binom{n}{i}$ and this bound cannot be improved.

In the present note we consider the following related question (raised in [1], [2] and [7]): What is the maximum length of a sequence $(F_1, F_2, \dots, F_m)$ of distinct subsets of an n-element set X with the property that

$$|F_i \backslash F_j| < k \quad \text{for all } i < j? \tag{1}$$

Let us denote this maximum by $f(n,k)$. We can clearly suppose without loss of generality that the F_i's are listed in increasing order of their cardinalities, i.e., $|F_i| \leq |F_j|$ for all $i \leq j$.

It is easy to show that

$$f(n,k) \geq \binom{n}{k} + 2\left(\binom{n}{k-1} + \binom{n}{k-2} + \cdots + \binom{n}{o}\right) - \binom{2k-1}{k}, \quad \text{if } n \geq 2k. \tag{2}$$

To this end fix a chain of subsets $E_1 \subset E_2 \subset \cdots \subset E_n = X$ with $|E_i| = i$ $(1 \leq i \leq n)$ and let $\mathscr{F}_j := \{F \subseteq X \,|\, |F| = j, F \supseteq E_{j-k+1}\}$ $(k \leq j \leq n-k)$. Then the number of elements of

$$\mathscr{F} := \{F \subseteq X \,|\, |F| < k\} \cup \left(\bigcup_{j=k}^{n-k} \mathscr{F}_j\right) \cup \{F \subseteq X \,|\, |F| > n-k\}$$

is equal to the right-hand side of (2), and enumerating them in increasing order of size they will obviously satisfy (1), too.

A set-sequence $\mathscr{F} = (F_1, F_2, \ldots, F_m)$ having property (1) is called *extremal* if $m = f(n,k)$. Set $f_i := |\{F \in \mathscr{F} \,|\, |F| = i\}|$. Two extremal sequences $\mathscr{F}$ and $\mathscr{F}'$ are said to be *essentially different* if $f_i \neq f_i'$ for some i $(0 \leq i \leq n)$.

Our next two theorems show that the lower bound (2) is sharp if $k = 2$ and is asymptotically sharp for all $k \geq 3$ $(n \to \infty)$.

Theorem 1. *If* $n \geq 4$ *then* $f(n,2) = \binom{n}{2} + 2n - 1$. *Furthermore, in this case there are exactly* 2^{n-3} *essentially different extremal sequences.*

Theorem 2. $f(n,k) < \binom{n}{k} + 5k^2\binom{n}{k-1}$ *for all* $n \geq 2k$.

The above problem can be reformulated in the following more general setting. Given a natural number n and a class $\mathscr{L}$ of (0, 1)-matrices (so called *'forbidden submatrices'*), determine the maximum integer m such that there exists an $m \times n$ (0, 1)-matrix M without repeated rows and containing no element of $\mathscr{L}$ as a submatrix. Let us denote this maximum by $\mathrm{ex}(n, \mathscr{L})$. In view of the condition that all rows of M should be distinct, we have $\mathrm{ex}(n, \mathscr{L}) \leq 2^n$.

Let A_k denote a $2 \times k$ matrix whose first and second rows contain only 1's and 0's, respectively. Using the above notation, we obviously get $\mathrm{ex}(n, \{A_k\}) = f(n,k)$.

Next, let $\mathscr{L}_k$ be the family of all $2^k \times k$ matrices which contain every (0, 1)-vector of length k (as a row) exactly once. The members of $\mathscr{L}_k$ differ in the order of rows only, hence $|\mathscr{L}_k| = 2^k$. A well-known theorem of Sauer [12] and Shelah [13] (see also [8], [9]) states that

$$\mathrm{ex}(n, \mathscr{L}_k) = \sum_{i=0}^{k-1} \binom{n}{i}. \tag{3}$$

From this we can easily deduce the following general result.

Theorem 3. *Let* $\mathscr{L}$ *be any family of forbidden* (0, 1)*-matrices, and suppose that there is a* $j \times k$ *matrix* $L \in \mathscr{L}$. *Then*

$$\mathrm{ex}(n, \mathscr{L}) \leq \left((j-1)\binom{n}{k} + 1\right)\left(\sum_{i=0}^{k-1}\binom{n}{i} + 1\right) - 1 \leq jn^{2k-1}$$

holds for every natural number n.

Proof. Let M be an $m \times n$ (0, 1)-matrix with distinct rows $M_1, M_2, \ldots, M_m$ and suppose that m exceeds the upper bound in the theorem. By repeated application of (3) we obtain that for every q $\left(1 \leq q \leq (j-1)\binom{n}{k} + 1\right)$ there exists a $2^k \times k$ submatrix L_q of M, which is equivalent to some member of $\mathscr{L}_k$ and whose rows are chosen from among

$$\left\{M_t \,\middle|\, (q-1)\left(\sum_{i=1}^{k-1}\binom{n}{i} + 1\right) < t \leq q\left(\sum_{i=1}^{k-1}\binom{n}{i} + 1\right)\right\}.$$

Now, by the pigeonhole principle, there are at least j submatrices (say, $L_{q_1}, L_{q_2}, \ldots, L_{q_j}$) sitting on the same set of k columns. Selecting a copy of the i-th row of L from $L_{q_i}(i = 1, 2, \ldots, j)$, we get a submatrix of M equivalent to L. □

A weaker upper bound for $\mathrm{ex}(n, \mathscr{L})$ was found by Anstee [2]. For more problems and results of this kind consult [3].

2. Proof of Theorem 1

Let $\mathscr{F} = \{F_1, F_2, \ldots, F_m\}$ be a system of subsets of an *n-element set* X satisfying condition (1), and put $\mathscr{F}_i := \{F \in \mathscr{F} \,|\, |F| = i\}$, $f_i := |\mathscr{F}_i|$, $i = 0, 1, \ldots. n$. For every pair $F_i \in \mathscr{F}_i$, $F_j \in \mathscr{F}_j$ $(0 \le i \le j \le n)$

$$|F_i \cap F_j| \ge i - k + 1. \tag{4}$$

In particular, any two members of $\mathscr{F}_i$ have at least $i - k + 1$ elements in common, i.e., $\mathscr{F}_i$ is $(i - k + 1)$ – *intersecting.*

From now on assume $k = 2$.

If $\mathscr{F}_i$ has at least two members F' and F'', say, then there are two possibilities. Either

(i) $\qquad F \supset F' \cap F''$ for every $F \in \mathscr{F}_i$, or

(ii) $\qquad F \subset F' \cup F''$ for every $F \in \mathscr{F}_i$.

In the first case we say that $\mathscr{F}_i$ is a *sunflower* with *centre* $F' \cap F''$ and the one element sets $F \backslash (F' \cap F'')$ are its *petals.* In the second case $\mathscr{F}_i$ is said to be an *inverse sunflower,* $F' \cap F''$ is its *centre* and the one element sets $(F' \cup F'') \backslash F$, $F \in \mathscr{F}_i$, are called *holes.*

Lemma 1. *Let $\mathscr{F}_i$ be a sunflower and $\mathscr{F}_j$ be an inverse sunflower for some $i < j$. Then* $\min\{f_i, f_j\} \le j - i + 2$.

Proof. Suppose, for contradiction, that both $\mathscr{F}_i$ and $\mathscr{F}_j$ have at least $j - i + 3$ members. Let C_i and C_j denote the centres of $\mathscr{F}_i$ and $\mathscr{F}_j$, respectively. Then $|C_i| = i - 1, |C_j| = j + 1$ and $|\bigcup\{F | F \in \mathscr{F}_i\}| \ge (j - i + 3) + (i - 1) = j + 2$, hence there is an $F \in \mathscr{F}_i$ such that $F \not\subset C_j$. If $|F \backslash C_j| > 1$, then taking any $F' \in \mathscr{F}_j$, the pair (F, F') will violate condition (1). So we can assume $|F \backslash C_j| = 1$. In this case $|C_j \backslash F| = (j + 1) - (i - 1) = j - i + 2$, thus there is a hole of $\mathscr{F}_j$ in $C_j \cap F$, i.e., there exists an $F' \in \mathscr{F}_j$ with $(F \backslash F') \cap C_j \ne \varnothing$, again a contradiction. □

Lemma 2. *Let q be a natural number,* $3 \le q \le n$. *Then*

$$|\{i | 2 \le i \le n - 2 \quad \textit{and} \quad f_i \ge q\}| \le n - q.$$

Proof. Suppose without loss of generality that $f_1 = f_{n-1} = n$. Let I_q (and I'_q) denote the set of all indices i $(1 \le i \le n - 1)$ for which $\mathscr{F}_i$ is a sunflower (an inverse sunflower, resp.) and $f_i \ge q$. Clearly $1 \in I_q$, $n - 1 \in I'_q$. Choose a pair $i \in I_q, j \in I'_q$, $i < j$, such that $j - i$ is minimal. Then there are no elements of $I_q \cup I'_q$ in the interval $J_q = \{i + 1, i + 2, \ldots, j - 1\}$, and by Lemma 1 we have $q \le j - i + 2$, i.e., $|J_q| \ge q - 3$. Hence $|I_q \cup I'_q| \le (n - 1) - |J_q| \le n - q + 2$. □

Lemma 2 shows that the i-th largest element of the sequence $(f_2, f_3, \ldots, f_{n-2})$ is at most $n - i$ $(i = 1, 2, \ldots, n-3)$, thus

$$|\mathscr{F}| \le f_0 + f_1 + f_{n-1} + f_n + \sum_{i=1}^{n-3} (n-i) \le 2n + 2 + \frac{(n-3)(n+2)}{2} = \binom{n}{2} + 2n - 1,$$

as desired.

If $|\mathscr{F}| = \binom{n}{2} + 2n - 1$, then $f_0 = f_n = 1$, $f_1 = f_{n-1} = n$ and the sequence $(f_2, f_3, \ldots f_{n-2})$ is some permutation of the numbers $3, 4, \ldots, n-1$. Furthermore,

$$J_q = \{i | 2 \le i \le n-2 \text{ and } f_i < q\}, \quad |J_q| = q - 3$$

and the only element $m_{q+1} \in J_{q+1} \backslash J_q$ is equal either to $\min J_{q+1}$ or to $\max J_{q+1}$ $(q = 3, 4, \ldots, n-1)$. In the first case $m_{q+1} \in I_q$, in the latter one $m_{q+1} \in I'_q$, i.e., $\mathscr{F}_{m_{q+1}}$ is a sunflower or an inverse sunflower, respectively. In view of the fact that $\min J_4 \ne \max J_4$ and for all $q = 3$ we have exactly 2 choices, we obtain that there are at most 2^{n-4} essentially different extremal sequences. It is easy to see that all of them can be realized in exactly 2 non-isomorphic ways ($\mathscr{F}_{m_4}$ can be a sunflower and an inverse sunflower as well). This completes the proof.

3. Proof of Theorem 2

Let $\mathscr{F} = \{F_1, F_2, \ldots\}$ be a system of distinct subsets of $X = \{1, 2, \ldots, n\}$ satisfying condition (1) for a fixed $k \ge 3$, let $\mathscr{F}_i := \{F \in \mathscr{F} \,\|F| = i\}$ and $f_i := |\mathscr{F}_i|$, $0 \le i \le n$. By (4), $\mathscr{F}_i$ is $(i - k + 1)$-intersecting, hence using a theorem of [6] (see also [10]) we obtain

$$f_i \le \binom{n}{k-1}, \quad 0 \le i \le n.$$

This immediately implies $f(n, k) \le n\binom{n}{k-1} \sim k\binom{n}{k}$.

To improve on this bound, we will apply first some simple operations (so-called left-shifts, cf. [4]) to our family $\mathscr{F}$. Given a pair i, j $(1 \le i < j \le n)$, let

$$C_{ij}(F) := \begin{cases} (F\backslash\{j\}) \cup \{i\} & \text{if } i \notin F, j \in F, (F\backslash\{j\}) \cup \{i\} \notin \mathscr{F} \\ F & \text{otherwise} \end{cases}$$

for any $F \in \mathscr{F}$. Further, set $C_{ij}(\mathscr{F}) := \{C_{ij}(F) | F \in \mathscr{F}\}$. The following statement can readily be checked.

Lemma 3. *$C_{ij}(\mathscr{F})$ also satisfies condition* (1). □

Repeating this operation for all pairs i, j (possibly several times), after a finite number of steps we obtain a *left-shifted* family $\mathscr{F}'$, i.e., one for which $C_{ij}(\mathscr{F}') = \mathscr{F}'$ $(1 \le i < j \le n)$.

Thus we can assume without loss of generality that $\mathscr{F}$ is left-shifted. Then $\mathscr{F}_i$ is left-shifted for all i and we can use the following.

Lemma 4. ([5]) *Suppose that $i \geq k$, $\mathcal{F}_i$ is left-shifted and $(i-k+1)$-intersecting. Then for any $F \in \mathcal{F}_i$ there exists a minimal integer $t = t(F)$, $0 \leq t \leq k-1$ such that*

$$|F \cap \{1,2,\ldots,i-k+1+2t\}| = i-k+1+t. \qquad \square$$

A set $F \in \mathcal{F}_i$ will be called *exceptional* if $i \geq k$ and at least one of the following four conditions is satisfied:

(i) $\{i-k-1+2t(F), i-k+2t(F), i-k+1+2t(F)\} \not\subseteq F$;
(ii) $\{i-k+2+2t(F), i-k+3+2t(F)\} \cap F \neq \varnothing$.
(iii) there exists an r $(1 \leq r \leq i-k+2t(F))$ such that $r, r+1 \notin F$;
(iv) there exists an r $(i-k+2t(F) < r < n)$ such that $r, r+1 \in F$.

Lemma 5. *The number of exceptional members of $\mathcal{F}$ is at most* $4k^2\binom{n}{k-1}$.

Proof. By a simple counting argument. The pair $(t(F), |F|)$ can take at most kn different values. In each case

$$|\{1,2,\ldots,|F|-k+1+2t(F)\} \backslash F| = t(F),$$
$$|F \cap \{|F|-k+2+2t(F),\ldots,n\}| = k-1-t(F).$$

Thus, e.g. the number of all members of $\mathcal{F}$ which are exceptional because of (i) does not exceed

$$\sum_{t,i} 3\binom{i-k+2t}{t-1}\binom{n-(i-k+1+2t)}{k-1-t} \leq 3k\binom{n}{k-1} \leq k^2\binom{n}{k-1}.$$

The other three cases can be treated similarly. $\square$

Each non-exceptional member $F \in \mathcal{F}$ will be assigned with a k-tuple

$$X_F := (\{1,2,\ldots,|F|-k+1+2t(F)\} \backslash F) \cup \{|F|-k+1+2t(F)\}$$
$$\cup \{x-1 \mid x \in (F \cap \{|F|-k+2+2t(F),\ldots,n\})\}.$$

Lemma 6. *Let F and G be two distinct non-exceptional members of $\mathcal{F}$. Then $X_F \neq X_G$.*

Proof. Suppose in order to obtain a contradiction that $|F| \leq |G|$ and $X_F = X_G$. $|F| = |G|$ implies $F = G$, thus we may assume that $|F| < |G|$ and $|F| - k + 1 + 2t(F) < |G| - k + 1 + 2t(G)$. Let

$$F' := F \cup (X_F \backslash \{|G|-k+1+2t(G)\}) \backslash \{x+1 \mid x \in X_F \backslash \{|G|-k+1+2t(G)\}\}.$$

Since $\mathcal{F}$ is left-shifted, obviously $F' \in \mathcal{F}$. On the other hand, all elements of the set

$$(X_F \backslash \{|G|-k+1+2t(G)\}) \cup \{|G|-k+2+2t(G)\}$$

belong to $F' \backslash G$. Hence $|F' \backslash G| \geq k$, contradicting (1). $\square$

By Lemma 6, $\mathcal{F}$ has at most $\binom{n}{k}$ non-exceptional members. Thus, in view of Lemma 5,

$$|\mathscr{F}| \le \binom{n}{k} + 4k^2 \cdot \binom{n}{k-1} + \sum_{i<k} f_i,$$

and the proof of Theorem 2 is complete.

4. Concluding Remarks and Problems

Conjecture 1. There exists a sufficiently large constant $n_o(k) \ge 2k$ such that if $n \ge n_o(k)$ then

$$f(n,k) = \binom{n}{k} + 2\left(\binom{n}{k-1} + \binom{n}{k-2} + \cdots + \binom{n}{0}\right) - \binom{2k-1}{k}. \quad (5)$$

As it was pointed out by N. Alon [1], (5) is not valid for $k = 3, n = 7$. Next we show that this is not an isolated example.

Proposition. *If k is large enough then $n_o(k) \ge 2k + \sqrt{k}/10$.*

Proof. Let $t \sim \sqrt{k}/10$ be an integer such that $k + t$ is odd, let $n = 2k + t$, $X = X_1 \cup X_2$ an n-element set, $|X_1| = k$, $|X_2| = k + t$. Then

$$\mathscr{F} := \{F \subset X \,|\, |F| < k \text{ or } |F| > n - k\}$$
$$\cup \left\{F \subset X \,|\, k \le |F| \le k + t \text{ and } |F \cap X_1| < \frac{k-t}{2}\right\}$$

(listed in increasing order of cardinality) obviously satisfies (1). Now by the formulas of Stirling and Moivre-Laplace we obtain that

$$|\mathscr{F}| > \frac{3}{2}\binom{n}{k} + 2\left(\binom{n}{k-1} + \binom{n}{k-2} + \cdots + \binom{n}{0}\right).$$

A set-system $\mathscr{F}$ is called a *Sperner family* if $F \not\subseteq G$ holds for every pair $F, G \in \mathscr{F}$.

Conjecture 2. Let $\mathscr{F} = \{F_1, F_2, \ldots\}$ be a Sperner family of subsets of an n-element set satisfying condition (1). Then $|\mathscr{F}| \le \binom{n}{k-1}$ holds for $n \ge 2k - 3$.

Remark. Let us mention that the above inequality follows from Sperner's theorem for $2k - 3 \le n \le 2k - 1$. We could prove it for $n = 2k$ as well. Their are four optimal families. Note that this is a stronger version of a conjecture of Frankl [7] which states that the same inequality holds under the condition that $|F_i \backslash F_j| < k$ for *all* i, j.

Conjecture 3. Let $\mathscr{L}, j, k$ denote the same as in Theorem 3. Then $\mathrm{ex}(n, \mathscr{L}) = 0(jn^k)$.

References

1. Alon, N.: Personal communication 1984
2. Anstee, R.P.: General forbidden configuration theorems. J. Comb. Theory (A) **40**, 108–124 (1985)

3. Anstee, R.P., Füredi, Z.: Forbidden submatrices. Discrete Math. **62**, 225–243 (1986)
4. Erdös, P., Ko, C., Rado, R.: Intersection theorems for systems of finite sets. Q. J. Math. Oxford (2) **12**, 313–320 (1961)
5. Frankl, P.: Families of finite sets satisfying a union condition. Discrete Math. **26**, 111–118 (1979)
6. Frankl, P.: The Erdõs-Ko-Rado theorem is true for $n = ckt$. Proc. Colloq. Math. Soc. J. Bolyai. **18**, 365–375 (1978)
7. Frankl, P.: Bounding the size of a family knowing the cardinality of differences. Studia. Sci. Math. Hung. (to appear)
8. Frankl, P.: On the trace of finite sets. J. Comb. Theory (A) **34**, 41–45 (1983)
9. Frankl, P., Pach, J.: On the number of sets in a null-t-design. Europ. J. Comb. **4**, 21–23 (1983)
10. Frankl, P., Wilson, R.M.: Intersection theorems with geometric consequences. Combinatorica **1**, 357–368 (1981)
11. Katona, G.O.H.: Open Problems (in Hungarian). Matematikus Kurir, Eötvös Univ. 1983
12. Sauer, N.: On the density of families of sets. J. Comb. Theory (A) **13**, 145–147 (1972)
13. Shelah, S.: A combinatorial problem, stability and order for models and theories in infinitary languages. Pacific J. Math. **41**, 247–261 (1972)

Received: September 25, 1985