COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA **25** (6) (2005) 633–643

UNAVOIDABLE TRACES OF SET SYSTEMS

JÓZSEF BALOGH, BÉLA BOLLOBÁS

Received December 15, 2002

Sauer, Shelah, Vapnik and Chervonenkis proved that if a set system on n vertices contains many sets, then the set system has full trace on a large set. Although the restriction on the size of the groundset cannot be lifted, Frankl and Pach found a trace structure that is guaranteed to occur in uniform set systems even if we do not bound the size of the groundset. In this note we shall give three sequences of structures such that every set system consisting of sufficiently many sets contains at least one of these structures with many sets.

1. Introduction

This note is concerned with finite systems of finite sets. Our notation and terminology are standard (see, e.g. [2]). In particular a *set system* is a (finite) collection of (finite) sets (with each set occuring with multiplicity at most 1). For brevity, we shall call a set system consisting of k sets a k-*system*. We shall use the standard notation $d(x)=|\{F:x\in F\in\mathcal{E}\}|$ for the degree of a vertex x and $[k]=\{1,\dots,k\}$.

The *trace* of a set system $\mathcal{F}$ on a set X is the set system

$$\mathcal{F}|X=\mathcal{F}_X=\{A\cap X : A\in\mathcal{F}\}.$$

Note that the trace is a simple set system i.e., each set in the trace is taken with multiplicity 1.

The classical result of set systems is the lemma of Sauer and Shelah [6] and [7] (see also Vapnik and Chervonenkis [8]) stating that if $\mathcal{F}\subset\mathcal{P}(n)$ and

Mathematics Subject Classification (2000): 05C35, 05C65, 05D05

$|\mathcal{F}| \geq \sum_{i=0}^{k-1} \binom{n}{i}$ then $\mathcal{F}|_X = \mathcal{P}(X)$ for some k-set X. Note that in this result the cardinality of $\mathcal{F}$ is assumed to be large compared to the cardinality of the ground set. This contrast with our main result, Theorem 1, where no assumption is made about the ground set. Results in similar vein to Theorem 1 were proved by Frankl and Pach [4] (see Theorem 8). We shall discuss these results in Section 6.

For $k = 0, 1, \ldots,$ let $\mathcal{F}_k$ be a k-system with $\cup_{F \in \mathcal{F}_k} = S_k$. We say that $(\mathcal{F}_k)_{k=0}^{\infty}$ is a *homogenous* sequence if for every $\ell \geq 0$ and set $X \subset S_k$, with $|X| = \ell$ we have

$$\mathcal{F}_\ell \subset \mathcal{F}_k|_X \subset \mathcal{F}_\ell \cup \{\emptyset, X\},$$

with the containments taken up to isomorphism. In fact, throughout the note, we shall not distinguish between isomorphic set systems. A typical example for a homogeneous sequence is for fixed integer ℓ

$$\mathcal{F}_k^{(\ell)} = \{A : A \subset [k], |A| \leq \ell\}.$$

The main aim of this note to show that there are three homogenous sequences such that, given $k \geq 2$, every sufficiently large set system contains a k-system from one of these sequences. Note that, *a priori*, it is not clear that there are sufficiently many homogenous sequences to ensure that, for given $k \geq 1$, every sufficiently large set system contains a k-system in one of these sequences. Let us introduce three homogenous sequences. For $k \geq 1$, let

$$\mathcal{C}_k = \{\{1, \ldots, i\} : 0 \leq i \leq k-1\},$$
$$\mathcal{S}_k = \{\{i\} : 1 \leq i \leq k\},$$
$$\mathcal{T}_k = \{[k] \setminus \{i\} : 1 \leq i \leq k\}.$$

We call $\mathcal{C}_k$ a *k-chain*, $\mathcal{S}_k$ a *k-star* and $\mathcal{T}_k$ a *k-costar*. Trivially, $(\mathcal{C}_k)$, $(\mathcal{S}_k)$ and $(\mathcal{T}_k)$ are homogenous sequences. Note that $\{[k] \setminus A : A \in \mathcal{S}_k\}$ is isomorphic to $\mathcal{T}_k$ and $\{[k-1] \setminus A : A \in \mathcal{C}_k\}$ is to $\mathcal{C}_k$.

Here is the main result of this note.

Theorem 1. *There is a function $f : \mathbb{N} * \mathbb{N} * \mathbb{N} \to \mathbb{N}, (k, \ell, m) \to f(k, \ell, m)$, such that every $f(k, \ell, m)$-system contains a k-star, an ℓ-costar or an m-chain.*

The rest of the paper is organized as follows. In Section 2 we prove Theorem 1. The third section is devoted to a variant of Theorem 1 for antichains, and in Section 4 we extend the main result. In the last two sections we embark on the past results and a connection between double chains and minimal antichains.

2. Proof of the main result

As usual, a collection $\mathcal{A}$ of sets is a *chain* if for all $A,B\in\mathcal{A}$, we have either $A\subset B$ or $B\subset A$, and it is an *antichain* if for all $A,B\in\mathcal{A}$ we have $A\not\subset B$ and $B\not\subset A$.

Let $f(k,\ell,m)$ denote the minimum number such that if $\mathcal{H}$ is a set system with at least $f(k,\ell,m)$ sets, then $\mathcal{H}$ contains either a k-star, or an ℓ-costar or an m-chain. Similarly, let $f'(k,\ell,m)$ denote the minimum number such that if $\mathcal{H}$ is an antichain set system with at least $f'(k,\ell,m)$ sets, then $\mathcal{H}$ contains either a k-star, or an ℓ-costar or an m-chain. *A priori* it is not clear that $f(k,\ell,m)$ or $f'(k,\ell,m)$ exists: our aim is to show that this is the case.

First we shall state and prove a lemma similar to the main result for antichains.

Lemma 2. *Let $k,\ell,m\in\mathbb{N}$. There is a number $f'(k,\ell,m)$ such that if $\mathcal{H}(V,\mathcal{E})$ is a set system which is an antichain with $|\mathcal{E}|\geq f'(k,\ell,m)$, then $\mathcal{H}$ contains either a k-star, an ℓ-costar, or an m-chain.*

Proof. For $k,\ell,m\geq 2$, it is easy to check that

$$f'(2,\ell,m)=f'(k,2,m)=f'(k,\ell,2)=2. \tag{1}$$

The fact that we know the function $f'(k,\ell,m)$ whenever k or ℓ or m is 2 enables us to use induction on $k+\ell+m$. Let us fix $k,\ell,m>2$. We claim

$$f'(k,\ell,m)\leq 2(m-2)(f'(k-1,\ell,m)-1)(f'(k,\ell-1,m)-1)+2. \tag{2}$$

Now note that $f'(k,\ell,m)=f'(\ell,k,m)$, as taking the complement of each element, a star produces a costar, and the complement of a chain is a chain. Suppose that, contrary to our assumption, there is a set system $\mathcal{H}$ with $2(m-2)(f'(k-1,\ell,m)-1)(f'(k,\ell-1,m)-1)+2$ sets that contains no k-star, ℓ-star or m-chain. Choose $\mathcal{H}$ to have a vertex set of minimal cardinality. Since $\mathcal{H}=(V,\mathcal{E})$ may be replaced by its completion $\bar{\mathcal{H}}=(V,\bar{\mathcal{E}})$, $\bar{\mathcal{E}}=\{V\setminus E: E\in\mathcal{E}\}$, we may assume that the average size of a set of $\mathcal{E}$ is at least $|V|/2$.

Since no vertex of $\mathcal{H}$ is in every set of $\mathcal{H}$, (by the minimality of $\mathcal{H}$) there is a vertex $x\in V$ such that $|\mathcal{E}|>d(x)\geq|\mathcal{E}|/2$ and $A\in\mathcal{E}$ be such that $x\notin A$. Set $\mathcal{E}_x=\{E\in\mathcal{E}: x\in E\}$. Then

$$|\mathcal{E}_x|\geq(m-2)(f'(k-1,\ell,m)-1)(f'(k,\ell-1,m)-1)+1. \tag{3}$$

Furthermore, set

$$\mathcal{F}_{x,A}=\{E\cap A: E\in\mathcal{E}_x\} \tag{4}$$

and, for all $B \in \mathcal{F}_{x,A}$, define

$$\mathcal{E}_B = \mathcal{E}_{x,A,B} = \{E \in \mathcal{E}_x : E \cap A = B\}. \tag{5}$$

Case 1. If $|\mathcal{E}_B| \geq f'(k-1,\ell,m)$ for some $B \in \mathcal{F}_{x,A}$, then $\mathcal{E}_B$ contains either a $(k-1)$-star, an ℓ-costar, or an m-chain. In the latter two cases we are done, as $\mathcal{E}_B \subset \mathcal{E}$. Otherwise, $\mathcal{E}_B$ contains a $(k-1)$-star $\mathcal{S}$ with sets $\{C_1,\dots,C_{k-1}\}$ and leaves $\{x_1,\dots,x_{k-1}\}$. As $k \geq 3$, we have $x_i \notin B$. Since $\mathcal{E}$ is an antichain, there is some $y \in A-B$, else $A \subset E$ for all $E \in \mathcal{E}_B$. This means, that the sets $\{C_1,\dots,C_{k-1},A\}$ form a k-star with leaves $\{x_1,\dots,x_{k-1},y\}$.

Case 2. $|\mathcal{E}_B| < f'(k-1,\ell,m)$ for all $B \in \mathcal{F}_{x,A}$. Then

$$|\mathcal{F}_{x,A}| \geq (m-2)(f'(k,\ell-1,m)-1)+1. \tag{6}$$

Suppose first that $\mathcal{F}_{x,A}$ contains an antichain $\mathcal{C}$ with $|\mathcal{C}| \geq f'(k,\ell-1,m)$.

By the induction hypothesis $\mathcal{C}$ contains either an $(\ell-1)$-costar, a k-star, or an m-chain. In the latter two cases we just take for each set E in the k-star (resp. in the m-star) a set F from $\mathcal{H}$ with $F \cap A = E$, and the set system what we obtain is a k-star (resp. an m-star) in $\mathcal{H}$, with the same vertices that belonged to the old system. If $\mathcal{C}$ contains an $(\ell-1)$-costar $\mathcal{S}$ with sets $\{C_1,\dots,C_{\ell-1}\}$ and vertices $\{x_1,\dots,x_{\ell-1}\}$, then for each set C_i, there is a $C'_i \in \mathcal{E}_x$ with $C_i \subset C'_i$ for all $1 \leq i \leq \ell-1$. Then clearly, $\{C'_1,\dots,C'_{\ell-1},A\}$ with $\{x_1,\dots,x_{\ell-1},x\}$ is an ℓ-costar in $\mathcal{H}$, and we are done.

If $F_{x,A}$ does not contain an antichain $\mathcal{C}$ with $|\mathcal{C}| \geq f(k,\ell-1,m)$, then by Dilworth's theorem, $\mathcal{F}_{x,A}$ contains a chain $\mathcal{D}$ with $|\mathcal{D}| \geq (m-1)$. In that case, adding A to the chain, clearly we can have easily an m-chain in $\mathcal{H}$. ∎

To deduce Theorem 1 from Lemma 2 we need to apply Dilworth's theorem. By Dilworth's theorem if

$$f(k,\ell,m) \geq (m-1)(f'(k,\ell,m)-1)+1$$

holds, then $\mathcal{H}$ contains either a chain with length m or an antichain with $f'(k,\ell,m)$ sets. In the first case the chain contains an m-chain, in the second case Lemma 2 can be applied. ∎

3. Double-chains

In this section we shall prove a slightly stronger result for antichains than Theorem 1, instead of m-chains we shall find m-double-chains. A *m-double-chain* is pair of collection of sets and vertices $(\mathcal{F}_m, F)$ with $\mathcal{F}_m \subset \mathcal{E}$, $F \subset V$ such that $\mathcal{F}_m = \{A_1,\dots,A_m\}$, $F = \{x_1,\dots,x_{m-1},y_1,\dots,y_{m-1}\}$ with

$$A_j = \{x_i : i < j\} \cup \{y_h : h \geq j\}.$$

Theorem 3. *Let $k,\ell,m\in\mathbb{N}$. Then there is a number $g(k,\ell,m)$ such that if $\mathcal{H}(V,\mathcal{E})$ is a set system which is an antichain with $|\mathcal{E}|\geq g(k,\ell,m)$, then $\mathcal{H}$ contains either a k-star, an ℓ-costar, or an m-double-chain.*

Proof. For the most part, we can imitate the proof of Lemma 2. In fact, only the treatment of our is different in the second part of the Case 2. First we should find the initial values of the function g:

$$g(2,\ell,m)=g(k,2,m)=g(k,\ell,2)=2.$$

Now we are going to prove the following recursion form:

$$g(k,\ell,m)\leq 2mg(k-1,\ell,m)^2 g(k,\ell-1,m).$$

In that case $\mathcal{F}_{x,A}$ contains a chain $\mathcal{C}=\{C_t\subset C_{t-1}\subset\ldots C_1\subset A\}$ with $t=(m+1)g(k-1,\ell,m)$. Let $y\in A-C_1$. For all i, $1\leq i\leq t$, let $B_i\in\mathcal{E}$ be a set with $B_i\cap A=C_i$. If the set system $\mathcal{H}_{V-A}$ with sets $\{B_1-A,\ldots,B_t-A\}$ contains a chain

$$\{B_{i_0}-A,\ldots,B_{i_m}-A\},$$

then $\{B_{i_0},\ldots,B_{i_m}\}$ with some properly chosen vertices is an m-double-chain. (The reason for this that $\mathcal{H}$ was an antichain and so if

$$\{B_{i_0}-A\subset B_{i_1}-A\subset\cdots\subset B_{i_m}-A\},\tag{7}$$

then

$$\{B_{i_0}\cap A\supset B_{i_1}\cap A\supset\cdots\supset B_{i_m}\cap A\}.)\tag{8}$$

Otherwise by Dilworth's theorem, the set system $\mathcal{H}_{V-A}$ contains an antichain with $g(k-1,\ell,m)$ sets. By the induction statement, this antichain contains either a $(k-1)$-star, an ℓ-costar or an m-double-chain. Clearly, in the latter two cases we are done. In the first case, if

$$\{B_{i_1}-A,B_{i_2}-A,\ldots,B_{i_{k-1}}-A\}$$

with $x_1,\ldots,x_{k-1}\in V-A$ is a $(k-1)$-star in $\mathcal{H}_{V-A}$, then the sets $\{B_{i_1},B_{i_2},\ldots,B_{i_{k-1}},A\}$ with vertices $\{x_1,\ldots,x_{k-1},y\}$ form a k-star in $\mathcal{H}$. ∎

4. The (k,t)-star, -costar case

A (k,t)-*star* is a k-system $\mathcal{F}$ whose restriction to a kt-set Y consists of k disjoint sets, all of order t. A (k,t)-*costar* is a k-system $\mathcal{F}$ whose restriction to a kt-set Y consists of the complement of k-disjoint sets, all with order t. Note that a k-star is precisely a $(k,1)$-star.

There is not much sense to define (k,t)-chain, because for example from a kt-chain a (k,t)-chain can be easily constructed.

An extension of Theorem 1 is the following.

Theorem 4. *Let t be a positive integer. Let $\mathcal{H}(V,\mathcal{E})$ be a set system such that for $A \neq B$, $A,B \in \mathcal{E}$, $|A\Delta B| \geq 4t-3$. Then for any positive integers k,ℓ,m there is a number $h(k,\ell,m)$ such that if $|\mathcal{E}| \geq h(k,\ell,m)$, then $\mathcal{H}$ contains either a (k,t)-star or an (ℓ,t)-costar or an m-chain.*

Proof. Let $\mathcal{H}(V,\mathcal{E})$ be a set system satisfying the conditions of the theorem. By Lemma 2 it contains a large star, or a large costar or an m-chain. In the third case we are done. Consider the case when we have a large K-star or a large costar $\{A_1,\dots,A_K\}$, $\{x_1,\dots,x_K\}$. Then construct (in both cases) the following set system with vertices $V\backslash\{x_1,\dots,x_K\}$ and sets $\{A_i\backslash\{x_1,\dots,x_K\}: 1,\dots,K\}$. Clearly for every pair of the sets of the new graph their symmetric difference is at least $4t-5$. Let us apply Lemma 2 to this set system (which is still an antichain). We can iterate this step $2t-1$ times. There are two possibilities. If during the process we find an m-chain then we are done. Otherwise, we find either at least t times a star, or at least t times a costar, from which the required configuration can be constructed. ∎

We can obtain a more general result from the previous result. Call the following configuration (r,t)-Δ-system: Let $K_1,\dots,K_r,L_1,\dots,L_r$ be disjoint sets, each with size at most t. Let $L=\cup_{i=1}^r L_i$, and A be a vertex set disjoint from all the other sets. Then the (r,t)-Δ-system is the following collection of sets:

$$A_i = A \cup K_i \cup L \setminus L_i.$$

Theorem 5. *Let k,ℓ,m,t be positive integers. Every set system with sufficiently many sets contains either a (k,t)-star, or an (ℓ,t)-costar, an m-chain, or an (r,t)-Δ-system.*

Proof. Let us imitate the proof of the previous lemma. At the point, when the new set system was constructed, here it is possible that there are multiple sets. But if there is a set with multiplicity at least r, then their ancestors will give a (r,t)-Δ-system. If this is not the case, then the set system contains many different sets, and the proof can be continued the same way as before. ∎

5. Minimal antichains

Recall that a set system $\mathcal{F}$ is an *antichain* if for all $A,B\in\mathcal{F}, A\neq B$, neither $A\subset B$ nor $B\subset A$. An antichain $(X,\mathcal{F})$ is a *minimal* antichain if for any vertex u the set system $\{F\setminus\{u\}:F\in\mathcal{F}\}$ is not an antichain.

Example. $(X_{k,\ell},\mathcal{F}_{k,\ell})$
Let k,ℓ be two positive integers. Let $\mathcal{F}_{k,\ell}$ be a set system on ground set $X_{k,\ell}=\{1,2,\dots,k\ell\}$ containing the sets

$$\mathcal{F}_{k,\ell}=\{(i-1)\ell+1,\dots,i\ell : i=1,\dots,k\} \\ \cup\{\{X\setminus\{i+j\ell : j=1,\dots,k\} : i=1,\dots,\ell\}.$$

Then $(X_{k,\ell},\mathcal{F}_{k,\ell})$ is a minimal antichain with $|X_{k,\ell}|=k\ell$ and $|\mathcal{F}_{k,\ell}|=k+\ell$.

It is rather suprising, but there is a common root of the existence of a double-chain in a hypergraph and the minimal antichain. Let us define the following function:

For positive integers $N>n$, let

$$f(N,n)=\max_{|\mathcal{A}|=N}\min_{S}\{|S| : \mathcal{A}_{|S} \text{ contains an antichain with } n \text{ sets}, \\ \mathcal{A} \text{ is an antichain}\}. \tag{9}$$

By the definition of the minimal antichain and Theorem 6, we have

$$f(n,n)=\lfloor n^2/4\rfloor. \tag{10}$$

From the other side, for any N we have

$$f(N,n)\geq 2n-2, \tag{11}$$

and this is best possible.

Proof. Let N be sufficiently large positive integer (at least $g(n,n,n)$ from Theorem 3). By Theorem 3, an antichain set system with N sets contains either an n-star, an n-costar, or an n-double-chain. Each of these structures spanned by at most $2n-2$ vertices, which gives (11). We cannot obtain a better bound, because for any N, consider an N-double-chain. One can check that this double-chain on any $2n-3$ vertices induce a set system with size at most $n-1$. ∎

To prove (10) we need to do some work.

In the next theorem we find for a minimal antichain of fixed number of elements (k), what the maximal possible size of the (spanning) ground set ($g(k)$) is. Let $f(n)$ be the following function:
for $n\leq 7$ let $f(n)=2n-2$, for $n\geq 7$ let $f(n)=\lceil n/2\rceil\cdot\lfloor n/2\rfloor$.

Theorem 6. *We have for all natural number n that $f(n)=g(n)$. Furthermore for $n\geq 8$, the unique extremal configuration is $\mathcal{F}_{k,k}$ in case $n=2k$ and $\mathcal{F}_{k+1,k}$ or $\mathcal{F}_{k,k+1}$ in case $n=2k+1$.*

Note that for $n\leq 7$ the extremal structures are the "tree-like" configurations (where the "tree edges" are "double edges". These concepts will become clearer from the proof below). For $n=7$ both type of constructions are extremal.

Proof. Let $\mathcal{F}$ be a minimal antichain. Construct the directed graph G as follows: Let $V(G)=\mathcal{F}$. For each $u\in X$, there is at least a pair (A,B), $A\neq B$, $A,B\in\mathcal{F}$ such that $A\setminus u\subset B$. If there are many such pairs, then choose one (A,B), and in the graph G put an edge from A to B. Do it for all $u\in X$. Note that the digraph G is not necessarily unique. Note that there are no parallel edges. A more general statement is true: Between two different vertices $A,B\in\mathcal{F}$, there cannot exist two edge disjoint oriented paths. Otherwise, if one of the paths is $u_1,\ldots,u_\ell$, the other is $v_1,\ldots,v_k$, then $\emptyset\neq A\setminus B\subset\{u_1,\ldots,u_\ell\}$ and $A\setminus B\subset\{v_1,\ldots,v_k\}$, but as the two paths were edge disjoint, this gives a contradiction. ∎

Let G_n be a directed graph on n vertices without parallel edges. We say that the graph G_n has property $(*)$, there are no two distinct vertices such that the first is connected to the second by two different edge disjoint paths. We shall show the following pure graph theoretical statement:

Lemma 7. *If G_n has property $(*)$, then $e(G_n)\leq f(n)$. The extremal graphs are listed below: For $n\geq 7$ let $A=\{1,\ldots,\lfloor n/2\rfloor\}$, $B=V(G)\setminus A$. For all $u\in A$ and $v\in B$ include the edge u,v. Or if we reverse the orientation of all of the edges, we get extremal graphs as well. (The difference what the reversing of the edges gives is when n is odd, and $|A|<|B|$.) For $n\leq 7$ take any tree on G_n and double each edge of the tree, giving different directions of the "parallel" edges.*

Proof of Lemma 7. Let the graph G_n have property $(*)$. Suppose, between two vertices, u and v there are two edges. Easy to check, if we contract u and v, and delete the two edges, then the new graph will have property $(*)$. Similarly, if there is an oriented triangle u,v,w, and the three vertices are contracted, and the edges of the triangle are cancelled, we obtain a graph with having property $(*)$. Now we proceed by induction. For $n=1$ or $n=2$ easy to establish the result. In general, suppose we know the statement for graphs with order less than n. Take an extremal graph G_n. If it contains a "double" edge, then by the contraction we have $f(n)\leq f(n-1)+2$. If G_n contains an oriented triangle then by the contraction and the induction hypothesis we have $f(n) \leq f(n-2)+3$. If neither of these cases holds then G_n can be considered as a triangle-free simple graph (it cannot contain a non-oriented triangle, because of property $(*)$), and by Turán Theorem the maximum number of possible edges and the structure of the extremal graphs are well-known. For $n=7$ both types of structures are possible. ∎

6. Remarks

The main result of Frankl and Pach [4] is the following: Let $T(n,r,k)$ denote the largest integer T such that there exists an r-uniform set system with T members on an n-element set X, which does not contain all r-tuples of a k-element subset of X.

Theorem 8. *Let $f_r(k)$ be the maximal number of sets in an r-uniform set system that does not contain a k-star. Then*

$$T(r+k-1,k-1,k) \leq f_r(k) \leq \binom{r+k-1}{k-1}.$$ ∎

Hajnal [5] suggested the following modification of the problem of finding $f_r(k)$. What is the maximal number of sets in an r-uniform set system if the system does not contain k sets that can be represented by t points each (i.e., what is the maximal number of sets in an r-system containing no $(k,t)-$star)? This question was answered by Frankl and Pach [4]. For general set systems, the corresponding result of ours is Theorem 5 answered this question for uniform set systems.

A collection $\mathcal{D}=\{D_1,\dots,\mathcal{D}_\ell\}$ of sets is a *Δ-system* if $D_i \cap D_j = K$ for all $i \neq j$ and some set K, the *kernel* of the Δ-system. The Frankl and Pach [4] result is the following.

Theorem 9. *Given four natural numbers, t,r,k,ℓ with $r \geq t, k \geq 3$ and $\ell \geq 2$, let $f_t^\ell(r,k) = max|\mathcal{E}(\mathcal{H})|$, where the maximum is taken over all r-uniform set systems $\mathcal{H}$ having the following two properties:*
(i) $\mathcal{H}$ does not contain an ℓ-member Δ-system whose kernel is of size greater than $r-t$;
(ii) $\mathcal{H}$ does not contain a (k,t)-star.
Then

$$c(k,\ell,t)r^{(k-1)t} \leq f_t^\ell(r,k) \leq c'(k,\ell,t)r^{(k-1)t},$$

where the positive constants c,c' depend only on k,ℓ and t (and not on r).

Having studied general set systems, let us see what we can say about the existence of our special structures in uniform set systems. Let us write $f_r(k,\ell,m)$ for the function above in the case of r-uniform graphs. Note first that no r-uniform set system contains an $(r+2)$-costar, or a $(r+2)$-chain or a $(r+2)$-double-chain. Hence, $f_r(k,\ell,m)=f_r(k,r+2,r+2)$ for all $\ell,m\geq r+2$. Let us write $f_r(k)$ for this common value. Thus $f_r(k)$ is the maximal number of sets in an r-uniform set system that contains neither a k-star, nor an $(r+2)$-costar, nor an $(r+2)$-chain. This function $f_2(k)$ is the maximal number of sets in a graph that contains no star-forest with k sets. The function $f_2(k+1)$ was determined by Vizing [9], although the result was formulated in a rather different way, in terms of domination numbers (see [3]). This result was later rediscovered by Frankl and Pach [4].

Theorem 10. *For all $k\geq 2$ we have*

$$f_2(k) = \binom{k+1}{2} - \left\lceil \frac{k+1}{2} \right\rceil .$$

For $r\geq 3$ it seems to be considerably harder to determine $f_r(k)$, see for example Theorem 8.

Set systems with possible multiple sets are easily identified with bipartite graphs. Indeed, given a set system $\mathcal{H}(V,\mathcal{E})$, we can construct a bipartite graph with classes $(V,\mathcal{E})$, $x\in V$ and $F\in\mathcal{E}$ are adjacent if and only if $x\in F$. This map is clearly invertible. Furthermore, if any two vertices have different neighborhoods, then the set system will not contain multiple sets. In view of this, our results about set system can be reformulated as results about bipartite graphs. We call the complement of a complete matching in a k by k bipartite graph a k-*comatching*. A k-*split* graph is a bipartite graph with bipartition $\{a_1,\dots,a_k\}$, $\{b_1,\dots,b_k\}$, and edges a_ib_j for $i>j$. If H is the bipartite graph corresponding to a set system $\mathcal{H}$ then $\mathcal{H}$ contains a k-star if and only if H contains an induced k-matching. Also, $\mathcal{H}$ contains a k-costar if and only if H contains an induced k-comatching, and $\mathcal{H}$ contains a k-chain if and only if H contains an induced k-split graph. Hence, Theorem 1 has the following immediate consequence.

Corollary 11. *Let k,ℓ,m be natural numbers and $f(k,\ell,m)$ as in Theorem 1. Let G be a bipartite graph with bipartition (A,B), where $|A|\geq f(k,l,m)$, with any two vertices of A having different neighborhoods. Then the graph G contains either an induced k-matching, or an induced ℓ-comatching or an induced m-split graph.*

This application was used as one of the tools in [1].

It would be interesting to determine the order of the function $f(k,k,k)$. The recursive bound given by Theorem 1 gives the upper bound $(2k)^{2^k}$. A naive random method or the example given by Frankl and Pach give the $k^{(1+o(1))k}$ lower bound. We suspect that the lower bound is closer to the truth; perhaps it is not too difficult to show that $f(k,k,k) \le k^{ck}$ for some constant c.

References

[1] J. Balogh, B. Bollobás and D. Weinreich: A jump to the Bell number for hereditary graph properties; to appear in *Journal of Comb. Theory Ser. B*.

[2] B. Bollobás: *Combinatorics*, Cambridge Univ. Press, Cambridge, 1986, xii+177 pp.

[3] T. W. Haynes, S. T. Hedetniemi and P. S. Slater: *Fundamentals of Domination in Graphs*, Marcel Dekker, New York, 1998; pp. 53–54.

[4] P. Frankl and J. Pach: On disjointly representable sets, *Combinatorica* **4** (1984), 39–45.

[5] A. Hajnal: personal communication, see [4].

[6] N. Sauer: On the densities of families of sets, *Journal of Comb. Theory Ser. A* **13** (1972), 145–147.

[7] S. Shelah: A combinatorial problem; stability and order for models and theories in infinitary languages, *Pacific J. Math.* **41** (1972), 247–261.

[8] V. Vapnik and A. Chervonenkis: On the uniform covergence of relative frequencies of events to their probabilities, *Theory of Probability and Its Applications* **16** (1971), 264–280.

[9] V. G. Vizing: A bound on the external stability number of a graph, *Dokl. Akad. Nauk SSSR* **164** (1965), 729–731.

József Balogh, Béla Bollobás

Department of Mathematical Sciences
University of Memphis
Memphis, TN 38152
USA

balogj@msci.memphis.edu, bollobas@msci.memphis.edu