Reprinted from JOURNAL OF COMBINATORIAL THEORY Vol. 12, No. 1, February 1972

A Generalization of Some Generalizations of Sperner's Theorem*

G. O. H. KATONA†

Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary

Communicated by W. T. Tutte

Received May 31, 1969

A theorem of Erdös says: If $\mathscr{A}$ is a family of subsets of a set S of n elements and no $h+1$ different members of the family form a chain $A_1 \subset \cdots \subset A_{h+1}$, then the maximum of the size of $\mathscr{A}$ is the sum of the h largest binomial coefficients of order n. The paper gives a weaker condition guaranteeing the same maximum. It is formulated in more abstract language.

INTRODUCTION

Sperner proved the following theorem [1]: Let $\mathscr{A} = \{A_1, ..., A_m\}$ be a family of subsets of a set S of n elements. If no two of them possess the property $A_i \subset A_j$ $(i \neq j)$, then

$$m \leqslant \binom{n}{\left[\frac{n}{2}\right]}.$$

Erdös answered the question: what is the maximum of m if no $h+1$ different elements of the family form a chain $A_{i_1} \subset \cdots \subset A_{i_{h+1}}$? The answer [2] is the sum of the h largest binomial coefficients of order n. Kleitman [3] and Katona [4] independently proved a sharpening of Sperner's theorem: Let $S_1 \cup S_2 = S$, $S_1 \cap S_2 = \varnothing$ be a partition of S. If $\mathscr{A} = \{A_1, ..., A_m\}$ is a family of subsets of S and no two different A_i, A_j satisfy the properties

$$A_i \cap S_1 = A_j \cap S_1 \quad \text{and} \quad A_i \cap S_2 \subset A_j \cap S_2$$

or

$$A_i \cap S_1 \subset A_j \cap S_1 \quad \text{and} \quad A_i \cap S_2 = A_j \cap S_2,$$

* This work was done while the author was at the Department of Statistics of the University of North Carolina at Chapel Hill.

† The author used earlier only the initial G.

then

$$m \leqslant \binom{n}{\left[\frac{n}{2}\right]}.$$

De Bruijn, Tengbergen, and Kruyswijk [5] generalized the original theorem of Sperner in the following manner: Let $f_1, \ldots, f_m$ be integer-valued functions defined on $S = \{x_1, \ldots, x_n\}$ such that $0 \leqslant f_i(x_k) \leqslant \alpha_k$, where α_k's are given positive integers. If no two different ones of them satisfy $f_i(x_k) \leqslant f_j(x_k)$ (for all k), then $m \leqslant M$, where M is the number of functions satisfying

$$\sum_{k=1}^{n} f(x_k) = \left[\frac{\sum_{k=1}^{n} \alpha_k}{2}\right].$$

Recently, Schönheim [6] gave a generalization of both Erdös's and Kleitman and Katona's results for integer-valued functions. The aim of this paper to give a common generalization of all these papers in a little more general language.

Definitions and the Theorem

Assume the directed graph G has the following property:

1. There exists a partition of its vertices into disjoint subsets $K_0, K_1, \ldots, K_n$ (they are called *levels*) of $k_0, k_1, \ldots, k_n$ elements and all the directed edges go from a vertex of K_i to a vertex of K_{i+1} $(0 \leqslant i \leqslant n)$. If $g \in K_i$ then we say that the *rank* of g is $r(g) = i$.[1]

A *symmetrical chain* in G is a set of vertices of a directed path, where for the starting point g and for the end point h the following equality holds:[2]

$$r(g) + r(h) = n.$$

We say that a directed graph G is a *symmetrical chain graph* if it satisfies property 1 and the following property 2:

2. There is a partition of its vertices into disjoint symmetrical chains.

It is easy to see that the following property is a consequence of property 2:

2a. $k_0 \leqslant \cdots \leqslant k_{\left[\frac{n}{2}\right]}$; $k_i = k_{n-i}$ $(0 \leqslant i \leqslant n)$.

Let us consider now a set S of n elements. Let its subsets be the vertices of the graph G and connect two vertices A and B (from A to B) if $B \supset A$

[1] This is equivalent to a partially ordered set with a rank function.

[2] The notion is introduced in [5].

and $| B - A | = 1$. This is the so-called *subset graph*. In this case K_i is the family of all subsets of i elements, $k_i = \binom{n}{i}$; thus property 1 (and (2a) easily holds. [5] shows that it has also property 2.

Similarly, if we consider the set of integer-valued function satisfying $0 \leqslant f(x_k) \leqslant \alpha_k$ as a vertex-set of a graph G, we connect two vertices f and g (from f to g) if $f = g$ except for one place x_k, where $f(x_k) = g(x_k) - 1$. K_i is in this case the set of functions for which $\sum_{k=1}^{n} f(x_k) = i$. It is easy to see that property 1 is satisfied. [5] proves that property 2 is also satisfied. This is called *function graph*.

Now we define the *direct sum* $G + H$ of two symmetrical chain graphs. Its vertices will be ordered pairs (g, h) $(g \in G,\ h \in H)$ and (g_1, h_1) is connected with (g_2, h_2) (in this direction) if and only if $g_1 = g_2$ and h_1, h_2 are connected in H (from h_1 to h_2), or $h_1 = h_2$ and g_1, g_2 are connected in G (from g_1 to g_2).[3]

If G is the subset graph of a set S_1 and H is a subset graph of a set S_2 (S_1 and S_2 are disjoint), then $G + H$ is the subset graph of $S_1 \cup S_2$. The situation is the same in the case of function graphs; the direct sum of two function graphs is again a function graph.

The generalization of Sperner's theorem (and also of the de Bruijn-Tengbergen-Kruyswijk theorem) in this language is the following: If we have a set $\{a_1, \ldots, a_m\}$ of vertices of a symmetrical chain graph and no two of them are connected with a direct path, then $m \leqslant k_{[n/2]}$.

The generalization of the Erdös theorem in this language is: if

$$\text{no } h + 1 \text{ different vertices from } \{a_1, \ldots, a_m\} \text{ lie in a directed path,} \tag{1}$$

then $m \leqslant$ the sum of the h largest k_i's. In a direct sum graph we will use a weaker condition rather than (1):

THEOREM. *Let G and H be symmetrical chain graphs with levels $K_0, \ldots, K_n$ (of $k_0, \ldots, k_n$ elements)* and *$L_0, \ldots, L_p$ (of $l_0, \ldots, l_p$) elements), respectively. If we have a set $(g_1, h_1), \ldots, (g_m, h_m)$ of vertices of $G + H$ such that*

$$\begin{aligned} &\textit{no } h + 1 \textit{ different ones of them satisfy the conditions:}\\ &g_{i_1} = \cdots = g_{i_w};\\ &h_{i_1}, \ldots, h_{i_w} \textit{ lie in a directed path in } H \textit{ in this order;}\\ &g_{i_w}, \ldots, g_{i_{h+1}} \textit{ lie in a directed path in } G \textit{ in this order;}\\ &h_{i_w} = \cdots = h_{i_{h+1}}\\ &\textit{for some } w\ (1 \leqslant w \leqslant h + 1), \end{aligned} \tag{2}$$

[3] This definition is equivalent to the usual definition of the direct sum of two partially ordered sets.

then $m \leqslant$ the number of vertices of the h largest levels of $G + H$, that is, the sum of the h largest numbers of type $\sum_{i=0}^{\alpha} k_i l_{\alpha-i}$.

Remark 1. If G and H are the subset graphs of the sets S_1 and S_2 of n and p elements, respectively, then the condition (2) becomes

$$\begin{aligned} &\text{there are no } h+1 \text{ different subsets } A_1, \ldots, A_{h+1} \text{ in} \\ &S = S_1 \cup S_2 \text{ such that} \\ &A_1 \cap S_1 = \cdots = A_w \cap S_1; \\ &A_1 \cap S_2 \subset A_2 \cap S_2 \subset \cdots \subset A_w \cap S_2; \\ &A_w \cap S_1 \subset A_{w+1} \cap S_1 \subset \cdots \subset A_{h+1} \cap S_1; \\ &A_w \cap S_2 = A_{w+1} \cap S_2 = \cdots = A_{h+1} \cap S_2 \\ &\text{hold for some } w \ (1 \leqslant w \leqslant h+1). \end{aligned} \tag{3}$$

It is clear that, if (3) would hold, then $A_1 \subset A_2 \subset \cdots \subset A_w \subset \cdots \subset A_{h+1}$ also should hold, that is, in this case we have a weaker condition than the Erdös theorem has, but we have the same result. The relation of this special case of our theorem to the Erdös theorem is the same as the relation of Kleitman and Katona's result to Sperner's theorem.

Remark 2. *If we put $h = 1$* in the preceding example we obtain Kleitman and Katona's result.

Remark 3. Theorems of Schönheim can be obtained if we use our theorem for function graphs and we put $h = 1$ or we change condition (2) by the stronger condition: no $h + 1$ functions satisfy $f_1 \leqslant \cdots \leqslant f_{h+1}$ for every x_k.

Remark 4. Let us consider now another important special case. Let S_1 be a one-element set, and let the vertices of G be the "functions" f on S_1, where $0 \leqslant f \leqslant n$ and f is an integer. There is a directed edge from f to g only if $g = f + 1$. Thus, G will be a directed path of length $n + 1$. Let H be the same graph with p instead of n. $G + H$ is in this case a rectangular $(n + 1) \times (p + 1)$ lattice (Fig. 1). It is a special case of the de Bruijn-Tengbergen-Kruyswijk theorem that, if we have a set of points of this rectangle no two of them connected with a directed path, then the

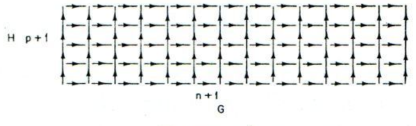

FIGURE 1

maximal number of these points is the length of the maximal diagonal (a *diagonal* is a set of vertices with the same coordinate sum), that is, $\min(n + 1, p + 1)$.

Schönheim's generalization of Erdös's theorem would state in this case that, if we have a set of vertices from this rectangle and

$$\text{no } h + 1 \text{ different one lie in one directed path,} \tag{4}$$

then the maximal number of these points is the sum of the lengths of the h largest different diagonals.

FIGURE 2

Our theorem says that, if we exclude the existence of $h + 1$ different points lying in a directed path which consists of two straight lines (Fig. 2) (instead of (4)), we obtain the same maximum. More exactly:

LEMMA. *Let R be a graph with vertices (i, j) $(0 \leqslant i \leqslant a;\ 0 \leqslant j \leqslant b;$ i and j are integers), where there are directed edges from (i, j) only to $(i, j + 1)$ and $(i + 1, j)$. If we have a set of vertices of m elements such that*

$$\begin{gathered} \textit{there are no } h + 1 \textit{ different vertices } (i_1, j_1), \ldots, (i_{h+1}, j_{h+1}) \\ \textit{with the properties for some } w \ (1 \leqslant w \leqslant h + 1) \\ \begin{aligned} & i_1 = \cdots = i_w ; \\ & j_1 < \cdots < j_w \qquad (\text{if} \quad w > 1); \\ & i_w < \cdots < i_{h+1} \qquad (\text{if} \quad w < h + 1); \\ & j_w = \cdots = j_{h+1} , \end{aligned} \end{gathered} \tag{5}$$

then $m \leqslant$ sum of the lengths of the h largest different diagonals

PROOFS

Proof of the Lemma. The set of the vertices (i_0, j), where i_0 is fixed and $0 \leqslant j \leqslant b$ is called a *column*. The *rows* are defined similarly. Let V be the set of vertices satisfying the conditions of the lemma and denote by c_t the number of columns having exactly t vertices from V. Obviously, by (5), $c_t = 0$ if $t > h$. Thus,

$$\sum_{t=0}^{h} c_t = a + 1. \tag{6}$$

Let us count in two different ways the number of vertices which are at least u-th elements of V in any column starting from below. In a column where the number of elements of V less than u we have to count 0, so counting column by column we obtain

$$c_u + 2c_{u+1} + 3c_{u+2} + \cdots + (h - u + 1)c_h .$$

On the other hand, counting row by row, we find that this number is at most $(h - u + 1)(b - u + 2)$ because we do not have to count the first $u - 1$ rows, and in the other rows we can have at most $h - u + 1$ such points by condition (5). Thus, we have the inequality

$$c_u + 2c_{u+1} + 3c_{u+2} + \cdots + (h - u + 1)\, c_h \leqslant (h - u + 1)(b - u + 2) \qquad (1 \leqslant u \leqslant h). \tag{7}$$

It therefore follows that the maximum of $\sum_{i=0}^{h} ic_i$ subject (only) to conditions (6) and (7) is an upper bound for m in the lemma.

If $a + 1 \leqslant b - h + 2$, then the choices $c_h = (a + 1)$, $c_i = 0$, $i = 1,2 ,..., (h - 1)$ maximize $\sum_{i=0}^{h} ic_i$ subject to (6) and (7) since for arbitrary $c_1 ,..., c_h$ satisfying (6) and (7) we have

$$\sum_{i=0}^{h} ic_i \leqslant h \sum_{i=0}^{h} c_i = h(a + 1).$$

Next, suppose that $a + 1 > b - h + 2$. We first show that, if $c_1',..., c_h'$ maximizes $\sum_{i=1}^{h} ic_i$ subject to (6) nad (7), it is no loss of generality to assume that $c_h' = b - h + 2$. For $c_h' \leqslant b - h + 2$ (in view of (7) with $u = h$) and, if $c_h' > b - h + 2$, then $c_h = b - h + 2$,

$$c_{h-1} = c_{h-1}' - 2(b - h + 2 - c_h'), \qquad c_{h-2} = c_{h-2}' + b - h + 2 - c_h' ,$$

$c_{h-3} = c_{h-3}' ,..., c_1 = c_1'$ satisfy $\sum_{i=0}^{h} ic_i = \sum_{i=0}^{h} ic_i'$, (6), and (7), as may be verified directly. For instance, verifying (7), we have for $u = h - 1$:

$$\begin{aligned} c_{h-1} + 2c_h &= c_{h-1}' - 2(b - h + 2 - c_h') + 2(b - h + 2) \\ &= c_{h-1}' + 2c_h' \leqslant 2(b - h + 3), \end{aligned}$$

and for $u \leqslant h - 2$:

$$\begin{aligned} & c_u + \cdots + (h-u-1)\, c_{h-2} + (h-u)\, c_{h-1} + (h-u+1)\, c_h \\ & \quad = c_u + \cdots + (h-u-1)(c'_{h-2} + b - h + 1 - c_h') \\ & \qquad + (h-u)(c'_{h-1} - 2(b-h+2-c_h')) \\ & \qquad + (h-u+1)(b-h+2) \\ & \quad = c_u' + \cdots + (h-u+1)\, c'_{h-2} + (h-u)\, c'_{h-1} + (h-u+1)\, c_h' \\ & \quad \leqslant (h-u+1)(b-u+2). \end{aligned}$$

For such a special choice $c_h = b - h + 2$, $c_{h-1}, \dots, c_1$, (7) is equivalent to

$$c_u + 2c_{u+1} + \cdots + (h-u)\, c_{h-1} \leqslant (h-u+1)(h-u) \tag{8}$$

for $1 \leqslant u \leqslant h-1$. If $a + 1 \leqslant b - h + 4$, then $c_1 = \cdots = c_{h-2} = 0$,

$$c_{h-1} = \begin{cases} 2, & \text{if } a+1 = b-h+4, \\ 1, & \text{if } a+1 = b-h+3, \end{cases}$$

maximizes $\sum_{i=1}^{h} ic_i$ subject to $c_h = b - h + 2$, (6), and (8) (=(7)). Let us assume now that $a + 1 > b - h + 4$. It is easy to see that there is a choice of $c_1, c_2, \dots, c_{h-1} = 2$ which maximizes $\sum_{i=1}^{h} ic_i$ subject to $c_h = b - h + 2$, (6), and (8) (=(7)). If we have another system $c_1, c_2, \dots, c_{h-1} < 2$, then we can change for a more appropriate one: $c_{h-1} = 2$, $c_{h-2} = c'_{h-2} - 2(2 - c'_{h-1})$, $c_{h-3} = c'_{h-3} + (2 - c'_{h-1})$, $c_{h-4} = c'_{h-4}, \dots, c_1 = c_1'$, which satisfies (6) (with $c_h = b - h - 2$) and (8) (=(7)) and gives the same value for $\sum_{i=1}^{h} ic_i$. Following this procedure we find a choice exists:

$$c_h{}^0 = b - h + 2, \qquad c^0_{h-1} = 2, \dots, c^0_{v+1} = 2,$$
$$c_v{}^0 = 1 \text{ or } 0, \qquad c^0_{v-1} = \cdots = c_1{}^0 = 0$$

which maximizes $\sum_{i=1}^{h} ic_i$ subject to (6) and (7), where v is determined by (6).

It is clear that $\sum_{i=1}^{h} ic_i{}^0$ gives an upper bound for m in the lemma. Now we shall show that we can achieve this bound under the stronger condition (5).

The set of points (i, j) satisfying $i + j = k$ is called the *k-th diagonal* of the rectangle and it is denoted by D_k. The *middle diagonals* are $D_y, \dots, D_{y+h-1}$, where

$$y = \left[\frac{a + b - h + 1}{2}\right].$$

It is easy to see, that the number of points of the h middle diagonals is thus $\sum_{i=1}^{h} ic_i^0$, that is, maximal. It means they are also the h largest diagonals, because arbitrary h diagonals satisfy the conditions of the lemma. The lemma is proved.

Proof of the Theorem. By property 2 of the symmetrical chain graphs, the vertices of G and H are divisible into symmetrical chains. Denote by G' and H' the graphs which have the same vertex-set as G and H, respectively, but they have edges only along these chains. Thus, G' and H' is a subgraph of G and H, respectively. It follows that $G' + H'$ is a subgraph of $G + H$. Hence, for a given set V of vertices, if h corresponds to V as a subset of $G + H$ while h' corresponds to V as a subset of $G' + H'$, then $h' \leqslant h$ and the number of vertices on the h' middle diagonals of $G' + H'$ is less than or equal to the number of vertices on the h middle diagonals of $G' + H'$ or, what is the same thing, the h middle diagonals of $G + H$. Thus it is sufficient to prove the theorem for $G' + H'$ instead of $G + H$. However, $G' + H'$ consists of disjoint rectangular lattices and condition (2) means simply condition (5) for every such rectangle.

We know that an optimal set of points in every rectangle is the union of the h middle diagonals. Define the levels of $G' + H'$ $(G + H)$ in the following manner. $(g, h) \in M_j$ iff $g \in K_i$ and $h \in L_{j-i}$ for some i $(0 \leqslant i \leqslant j)$. By definition of direct sum, it is easy to see that M_j's satisfy point 1 of the definition of a symmetrical chain graph. The *h middle levels* of $G' + H'$ are $M_z, \ldots, M_{z+h-1}$, where

$$z = \left[\frac{n + p - h + 1}{2}\right].$$

We will show that the union of the h middle diagonals for all the rectangles is just the h middle levels in $G' + H'$.

First we verify that an element of the h middle diagonals in a rectangle is an element of the h middle levels in $G' + H'$. Let us consider a fixed rectangle which is a direct sum of two symmetrical chains from G' and H' with vertices $g_0, \ldots, g_a$ and $h_0, \ldots, h_b$, respectively. If $r(g_0) = i$, then by the symmetricity $r(g_a) = n - i$; thus $i + a = n - i$, or

$$i = \frac{n - a}{2}. \tag{9}$$

(Obviously, n and a have the same parity.) Similarly, if $r(h_0) = j$ then

$$j = \frac{p - b}{2}. \tag{10}$$

If a point (g_k, h_1) is in D_r, in one of the h largest diagonal of the rectangle, then

$$k + 1 = r; \tag{11}$$

further, $r(g_k) = i + k$, $r(h_1) = j + 1$, thus $(g_k, h_1) \in M_{i+j+k+1}$, or, using (9), (10), and (11),

$$(g_k, h_1) \in M_{\frac{n+p-a-b}{2}+r}. \tag{12}$$

Since

$$y = \left[\frac{a + b - h + 1}{2}\right] \leqslant r \leqslant y + h - 1,$$

thus

$$z = \left[\frac{n + p - h + 1}{2}\right] \leqslant \frac{n + p - a - b}{2} + r \leqslant z + h - 1,$$

and (12) means that (g_k, h_1) is in one of the h middle levels of $G' + H'$.

Conversely, let (g, h) be an element of M_s, where $z \leqslant s \leqslant z + h - 1$. (g, h) is contained by one rectangular which is a direct sum of two symmetrical chains, say $g_0, \ldots, g_a$ and $h_0, \ldots, h_b$. Then by (9) and (10)

$$r(g_0) = \frac{n - a}{2}, \qquad r(h_0) = \frac{p - b}{2}.$$

If $(g, h) = (g_k, h_1)$ then

$$\frac{n - a}{2} + \frac{p - b}{2} + k + 1 = s,$$

that is, for

$$r = k + 1 = s - \frac{n - a}{2} - \frac{p - b}{2}$$

the following inequality holds:

$$\left[\frac{a + b - h + 1}{2}\right] = z - \frac{n + p - a - b}{2}$$
$$\leqslant r \leqslant \left[\frac{a + b - h + 1}{2}\right] + h - 1.$$

(g, h) is really an element of a diagonal from the h middle ones.

Thus, we proved that the points of the h middle levels form an optimal set. For the union of h arbitrary chosen levels of $G' + H'$ the conditions of the theorem are satisfied, so the middle levels must be the h largest ones (but there may exist h different levels with the same size-sum). The number of elements in M is obviously $\sum_{i=0}^{\alpha} k_i l_{\alpha-i}$; thus the optimal number is the sum of the h largest ones of these numbers. The proof is completed.

REFERENCES

1. E. SPERNER, Ein Satz über Untermengen einer endlichen Menge, *Math. Z.* **27** (1928), 544–548.
2. P. ERDÖS, On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* **51** (1945), 898–902.
3. D. KLEITMAN, On a lemma of Littlewood and Offord on the distribution of certain sums, *Math. Z.* **90** (1965), 251–259.
4. G. KATONA, On a conjecture of Erdös and a stronger form of Sperner's theorem, *Studia Sci. Math. Hungar.* **1** (1966), 59–63.
5. N. G. DE BRUIJN, C. A. VAN E. TENGBERGEN, AND D. KRUYSWIJK, On the set of divisors of a number, *Nieuw Arch. Wisk.* (2), **23** (1949–51), 191–193.
6. J. SCHONHEIM, A generalization of results of P. Erdös, G. Katona and D. Kleitman concerning Sperner's theorem, *J. Combinatorial Theory* (to appear).

Printed in Belgium by the St. Catherine Press Ltd., Tempelhof 37, Bruges