COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA **12** (1) (1992) 19–26

SMALL TRANSVERSALS IN HYPERGRAPHS

V. CHVÁTAL and C. McDIARMID

Received June 28, 1988

For each positive integer k, we consider the set A_k of all ordered pairs $[a,b]$ such that in every k-graph with n vertices and m edges some set of at most $am+bn$ vertices meets all the edges. We show that each A_k with $k \geq 2$ has infinitely many extreme points and conjecture that, for every positive ε, it has only finitely many extreme points $[a,b]$ with $a \geq \varepsilon$. With the extreme points ordered by the first coordinate, we identify the last two extreme points of every A_k, identify the last three extreme points of A_3, and describe A_2 completely. A by-product of our arguments is a new algorithmic proof of Turán's theorem.

1. The problem

A *k-graph* is an ordered pair (V,E) such that V is a finite set and E is a set of distinct k-point subsets of V. The elements of V are the *vertices* of the k-graph and the elements of E are the *edges* of the k-graph. We reserve the letters n and m for the number of vertices and for the number of edges, respectively, of a k-graph H, and similarly for n_i, m_i, H_i. A *transversal* (or a *cover* or a *blocking set*) in a k-graph is a set of vertices that meets all the edges; we let $\tau(H)$ denote the smallest size of a transversal in a k-graph H.

A problem of Turán [6] can be stated as the problem of determining the smallest $t(n,m,k)$ such that every k-graph H with n vertices and m edges has $\tau(H) \leq t(n,m,k)$. Trivially, $t(n,m,1)=m$. Turán [5] evaluated $t(n,m,2)$; the case of $k \geq 3$ remains unsolved. This is hardly surprising as Turán's problem subsumes other notoriously difficult combinatorial problems: for instance, $t(111,111,100) \geq 3$ if and only if a projective plane of order 10 exists. (To see this, consider any 100-graph H such that $H=(V,E)$ with $|V|=|E|=111$; define H^* to be the 11-graph (V,E^*) such that $A \in E^*$ if and only if $V-A \in E$. Now $\tau(H) \geq 3$ if and only if every two points of V lie in a common edge of H^*, which is the case if and only if H^* is a projective plane of order 10.)

We propose an easier variation on Turán's theme: for each fixed k, we consider the set A_k of all ordered pairs $[a,b]$ of real numbers such that

$$t(n,m,k) \leq am+bn.$$

To put it differently, we consider all theorems asserting that, for some fixed k, a, b, every k-graph H with n vertices and m edges has

$$\tau(H) \leq am+bn;$$

AMS subject classification (1991): 05 C 65, 05 B 40, 05 D 05

strongest theorems of this kind are in a one-to-one correspondence with extreme points of A_k. We show that each A_k with $k \geq 2$ has infinitely many extreme points (Theorem 4.1) and conjecture that, for every positive ε, each has only finitely many extreme points $[a, b]$ with $a \geq \varepsilon$ (Conjecture 4.2). When the extreme points of A_k are ordered by their first coordinate, $[1, 0]$ is trivially the last extreme point; we prove that

$$\left[\frac{\lfloor k/2 \rfloor}{\lfloor 3k/2 \rfloor}, \frac{1}{\lfloor 3k/2 \rfloor}\right]$$

is the next-to-last extreme point of every A_k with $k \geq 2$ (Theorem 6.2) and that $[1/6, 1/3]$ is the next-to-next-to-last extreme point of A_3 (Theorem 6.3). In addition, we describe A_2 completely (Theorem 5.1). A by-product of our arguments is a new algorithmic proof of Turán's theorem (Theorem 3.1).

2. A sequence of points in A_k

Lemma 2.1. *Let k, d be positive integers and let a, b real numbers; write*

$$\tilde{a} = \frac{k(1-b) - ad}{(k-1)d + k}, \; \tilde{b} = 1 - \tilde{a}(d+1).$$

If $[a, b] \in A_k$ and $a \geq \tilde{a} \geq 0$ then $[\tilde{a}, \tilde{b}] \in A_k$.

Proof. We wish to prove that $\tau(H) \leq \tilde{a}m + \tilde{b}n$ for all k-graphs H; for this purpose, we shall use induction on n. If $m \leq dn/k$ then, since $a \geq \tilde{a}$ and $d(a - \tilde{a}) = k(\tilde{b} - b)$, we have $am + bn \leq \tilde{a}m + \tilde{b}n$; since $\tau(H) \leq am + bn$ by assumption, we conclude that $\tau(H) \leq \tilde{a}m + \tilde{b}n$. If $m > dn/k$ then some vertex v in H is included in at least $d+1$ edges; by the induction hypothesis, and since $\tilde{a} \geq 0$, we have $\tau(H - v) \leq \tilde{a}(m - d - 1) + \tilde{b}(n-1) = \tilde{a}m + \tilde{b}n - 1$; and so

$$\tau(H) \leq 1 + \tau(H - v) \leq \tilde{a}m + \tilde{b}n. \quad \blacksquare$$

Theorem 2.2. *If k, d are positive integers and*

$$a = \prod_{i=1}^{d-1} \frac{(k-1)i}{(k-1)i + k}, \quad b = 1 - ad$$

then $[a, b] \in A_k$.

Proof. By induction on d, using Lemma 2.1. (Since the empty product equals 1 by definition, the basis of the induction amounts to observing that $[1, 0] \in A_k$.) ∎

3. Turán's theorem

A 2-graph is called simply a graph. Theorem 2.2 with $k = 2$ asserts that every graph H has

$$\tau(H) \le \frac{m}{\binom{d+1}{2}} + \frac{d-1}{d+1}n$$

for all positive integers d. A special case of this inequality is worth stating on its own.

Theorem 3.1. *If a graph H with n vertices has $\tau(H) \ge t$ for some nonnegative integer t then it has at last $\binom{d+1}{2}t - \binom{d}{2}n$ edges with $d = \lfloor n/(n-t) \rfloor$.* ∎

Note that $\binom{d+1}{2}t - \binom{d}{2}n$ with $d = \lfloor n/(n-t) \rfloor$ is the number of edges in the graph on n vertices that consists of $n-t$ vertex-disjoint cliques of sizes as nearly equal as possible; hence the bound of Theorem 3.1 is best possible. Note also that a graph H with n vertices has $\tau(H) \ge t$ if and only if it has no more than $n-t$ pairwise nonadjacent vertices. Hence Theorem 3.1 is nothing but the celebrated theorem of Turán [5]. Actually, our proof of Turán's theorem is algorithmic; we shall return to the algorithm in Section 7.

4. Extreme points of A_k

Since every 1-graph with n vertices and m edges has $m \le n$, we have $t(n,m,1) = m$ and

$$A_1 = \{[a,b] \in \mathbf{R}^2 : a+b \ge 1,\ b \ge 0\};$$

to avoid this trivial case, we shall assume that $k \ge 2$ from now on. Since

$$t(n,1,k) = 1 \quad \text{and} \quad t\left(n, \binom{n}{k}, k\right) = n-k+1$$

whenever $n \ge k$, we have $a \ge 0$, $b \ge 0$ whenever $[a,b] \in A_k$. To put it differently, A_k is a subset of the nonnegative quadrant $\mathbf{R}^2_+$. Note that $A_k + \mathbf{R}^2_+ = A_k$ and that $[0,1] \in A_k$, $[1,0] \in A_k$; note also that A_k, being the intersection of closed half-planes (one half-plane for each k-graph), is a closed convex set.

Recall that a point of a convex set is called *extreme* if it is not the midpoint of any line-segment that joins two distinct points in the set. Let $\text{conv}(S)$ denote the convex hull of a set S and let E_k denote the set of all extreme points of A_k. It is well known that every compact convex set is the convex hull of its extreme points; it follows easily that

$$A_k = \text{conv}(E_k) + \mathbf{R}^2_+.$$

In this sense, determining A_k reduces to determining E_k.

The purpose of this section is to point out that each E_k (with $k \ge 2$) is infinite.

Theorem 4.1. *For every integer k greater than one and for every positive ε there is a point $[a,b]$ in E_k such that $0 < a \le \varepsilon$.*

Proof. Assume the contrary: there is a positive ε such that A_k has no extreme point $[a,b]$ with $0 < a \le \varepsilon$. Letting C denote the intersection of A_k and the rectangle defined by $0 \le a \le \varepsilon$, $0 \le b \le 1$, observe that C has precisely three extreme points, namely, $[0,1]$, $[\varepsilon,1]$, and $[\varepsilon, 1-t\varepsilon]$ for some positive t. It follows that $b \ge 1 - ta$ whenever $0 \le a \le \varepsilon$ and $[a,b] \in A_k$. Now choose an integer d such that $d \ge 2$, $d \ge 1/\varepsilon$, $d > t$, and set

$$a = \prod_{i=1}^{d-1} \frac{(k-1)i}{(k-1)i+k}, \qquad b = 1 - ad.$$

Note that

$$a = \frac{k-1}{(k-1)(d-1)+k} \prod_{i=1}^{d-2} \frac{(k-1)(i+1)}{(k-1)i+k} \le \frac{(k-1)}{(k-1)(d-1)+k} < \frac{1}{d},$$

and so $0 < a < \varepsilon$. Since $b < 1 - ta$, we conclude that $[a,b] \notin A_k$, contradicting Theorem 2.2. ∎

Theorem 4.1 implies that A_k is not a polyhedron. However, it is conceivable that A_k is nearly a polyhedron: perhaps $[0,1]$ is the only accumulation point of E_k.

Conjecture 4.2. *For every positive integer k and for every positive ε, only finitely many points $[a,b]$ in E_k have $a \ge \varepsilon$.* ∎

In the following two sections, we provide meagre evidence in support of this conjecture: the assertion holds true for $k = 2$ (and all ε), it holds true for $\varepsilon = 1/3$ (and all k), and it holds true for $k = 3$, $\varepsilon = 1/6$.

5. All extreme points of A_2

The purpose of this section is to completely describe A_2 in two different ways.

Theorem 5.1. *Write*

$$E^* = \{[2/d(d+1), (d-1)/(d+1)] : d = 1,2,3,\ldots\} \cup \{[0,1]\}.$$

For every point $[a,b]$ in $\mathbf{R}^2$, the following three statements are equivalent:

(i) $[a,b] \in A_2$,

(ii) $\binom{n}{2}a + nb \ge n - 1$ *for all positive integers n,*

(iii) $[a,b] \in \text{conv}(E^*) + \mathbf{R}^2_+$.

Proof. To see that (i) implies (ii), substitute the clique on n vertices for H in the inequality $\tau(H) \le am + bn$.

To prove that (ii) implies (iii), consider an arbitrary point $[a,b]$ that satisfies (ii); letting n tend to infinity, note that $a \ge 0$; setting $n = 1$, note that $b \ge 0$. If $b \ge 1$ then (iii) holds as $[0,1] \in E^*$; if $0 \le b < 1$ then there are a positive integer d and a real number t such that $0 \le t \le 1$ and

$$b = t\frac{d-1}{d+1} + (1-t)\frac{d}{d+2}.$$

But then (ii) with $n = d+1$ gives

$$a \geq t\frac{2}{d(d+1)} + (1-t)\frac{2}{(d+1)(d+2)},$$

and so (iii) holds again.

To see that (iii) implies (i), note that $E^* \subseteq A_2$ by Theorem 2.2. ∎

Equivalence of (i) and (iii) guarantees that $E_2 \subseteq E^*$, and so Conjecture 4.2 holds true for $k = 2$ and all ε.

Actually, $E_2 = E^*$; to show this, we shall rely on a simple observation (which will be used again in the next section). Let us say that a point $[a, b]$ in A_k is *tight* at a k-graph H if $\tau(H) = am + bn$. The observation goes as follows:

if a point $[a, b]$ in A_k is tight at k-graphs H_1, H_2 with $m_1/n_1 \neq m_2/n_2$ then $[a, b] \in E_k$.

(To see this, note first that all $[a', b']$ in A_k have $a'm_i + b'n_i \geq \tau(H_i)$ for both $i = 1$, $i = 2$.) In particular, each point

$$[2/d(d+1), (d-1)/(d+1)]$$

belongs to E_2 since it is tight at the clique on d vertices and at the clique on $d+1$ vertices; point $[0, 1]$ belongs to E_2 simply because $A_2 \subseteq \mathbf{R}^2_+$ and because $[0, b] \in A_2$ implies $b \geq 1$ (by the equivalence of (i) and (ii) in Theorem 5.1).

6. The next-to-last extreme point of A_k

Our main result goes as follows:

Theorem 6.1. *If H is a k-graph with n vertices and m edges then*

$$\tau(H) \leq \frac{\lfloor k/2 \rfloor m + n}{\lfloor 3k/2 \rfloor}.$$

Proof. By induction on n. If a vertex v is included in at least three edges then, by the induction hypothesis,

$$\tau(H - v) \leq \frac{\lfloor k/2 \rfloor (m-3) + (n-1)}{\lfloor 3k/2 \rfloor} \leq \frac{\lfloor k/2 \rfloor m + n}{\lfloor 3k/2 \rfloor} - 1,$$

and we are done as $\tau(H) \leq 1 + \tau(H - v)$. Hence we may assume that each vertex of H is included in at most two edges. If p vertices are included in precisely one edge and q vertices are included in precisely two edges then $p + q \leq n$, $p + 2q = km$, and so

$$q \geq km - n.$$

Now consider the multigraph G whose vertices are the edges of H and whose edges correspond to the q vertices of H that are included in precisely two edges of H: if a vertex of H belongs to edges e and f of H then the corresponding edge of G joins vertices e and f of G. A theorem of Shannon [4] asserts that every multigraph of maximum degree k is edge-colorable by $\lfloor 3k/2 \rfloor$ colors; it follows that our multigraph G has a matching S such that

$$|S| \geq q/\lfloor 3k/2 \rfloor \geq (km - n)/\lfloor 3k/2 \rfloor.$$

Since S, seen as a set of vertices of H, meets $2|S|$ edges of H, we have

$$\tau(H) \leq |S| + (m - 2|S|) \leq m - \frac{km - n}{\lfloor 3k/2 \rfloor} = \frac{\lfloor k/2 \rfloor m + n}{\lfloor 3k/2 \rfloor}$$

as claimed. ∎

In deriving a corollary of Theorem 6.1, we shall rely on the observation made towards the end of the preceding section and on the following observation:

> if $[a_1, b_1]$ and $[a_2, b_2]$ are points in A_k with $a_1 < a_2$ and if there is a k-graph H such that both $[a_1, b_1]$, $[a_2, b_2]$ are tight at H then there is no $[a, b]$ in E_k with $a_1 < a < a_2$.
>
> (To see this, note first that all $[a, b]$ in A_k have $am + bn \geq \tau(H)$.)

Theorem 6.2. *Every A_k with $k \geq 2$ has precisely two extreme points $[a, b]$ with $a \geq \lfloor k/2 \rfloor / \lfloor 3k/2 \rfloor$, namely, $[\lfloor k/2 \rfloor / \lfloor 3k/2 \rfloor, 1/\lfloor 3k/2 \rfloor]$ and $[1, 0]$.*

Proof. Observe that there are k-graphs H_1, H_2 such that $n_1 = k$, $m_1 = 1$, $\tau(H_1) = 1$, and $n_2 = \lceil 3k/2 \rceil$, $m_2 = 3$, $\tau(H_2) = 2$. Trivially, $[1, 0] \in E_k$. The other point is in A_k by Theorem 6.1; to see that it is in E_k, consider H_1 and H_2. To see that there is no $[a, b]$ in E_k with $\lfloor k/2 \rfloor / \lfloor 3k/2 \rfloor < a < 1$, consider H_1. ∎

In case $k = 3$, we can do a little better.

Theorem 6.3. *A_3 has precisely three extreme points $[a, b]$ with $a \geq 1/6$, namely, $[1/6, 1/3]$, $[1/4, 1/4]$, and $[1, 0]$.*

Proof. Let H_1 denote the 3-graph with vertices 1,2,3,4 and edges

123, 124, 134, 234;

let H_2 denote the 3-graph with vertices 1,2,...,9 and edges

123, 456, 789, 147, 258, 369, 159, 267, 348, 168, 249, 357;

note that $\tau(H_1) = 2$, $\tau(H_2) = 5$. According to Theorem 6.2, $[1/4, 1/4]$ is the last but one extreme point of A_3. Then Lemma 2.1 with $a = b = 1/4$ and $d = 3$ shows that $[1/6, 1/3] \in A_3$; to see that $[1/6, 1/3] \in E_3$, consider H_1 and H_2. To see that there is no $[a, b]$ in E_3 with $1/6 < a < 1/4$, consider H_1. ∎

Incidentally, the 3-graph H_2 used in the proof of Theorem 6.3 is the affine plane $AG(2, 3)$. Jamison [3] and Brouwer and Schrijver [1] proved that $\tau(AG(d, q)) = d(q-1) + 1$.

7. Two greedy algorithms

Small transversals often (but not always) consist of vertices of large degrees. This observation motivates the following algorithm, which we shall call *MAX*.

MAX accepts any k-graph H as its input and returns a transversal T in H as its output. It is initialized by setting $T = \emptyset$ and $H^* = H$. As long as H^* has at least one edge, the following iterative step is repeated: select a vertex v whose degree in H^* is maximal, add v to T, and delete v from H^*.

The argument used to prove Theorem 2.2 yields readily the following result: if d is a positive integer then *MAX*, given any k-graph with n vertices and m edges, returns a transversal of size at most $am + bn$ with

$$a = \prod_{i=1}^{d-1} \frac{(k-1)i}{(k-1)i+k}, \quad b = 1 - ad.$$

In particular, if d and t are positive integers then *MAX*, given any graph with n vertices and fewer than $\binom{d+1}{2}t - \binom{d}{2}n$ edges, returns a transversal of size at most $t-1$. Thus *MAX* will find in any graph a transversal of the size that is guaranteed by Turán's theorem (to see this, set $d = \lfloor n/(n-t) \rfloor$).

It may be interesting to note that *MAX* shares this property with another algorithm, which we shall call *MIN*. Whereas *MAX* is driven by the idea of including vertices of large degrees, *MIN* is driven by the complementary idea of excluding vertices of small degrees.

MIN accepts any graph H as its input and returns a transversal T in H as its output. It is initialized by setting $T = \emptyset$ and $H^* = H$. As long as H^* has at least one edge, the following iterative step is repeated: select a vertex v whose degree in H^* is minimal, add all the neighbours of v to T, and delete all the neighbours of v as well as v itself from H^*.

Erdős [2] proved that *MIN* delivers a transversal of size t only if there is a graph G such that G and H have the same set of vertices, G consists of $n-t$ vertex-disjoint cliques, and the degree of each vertex in H is at least its degree in G. (It follows at once that *MIN* will find in any graph a transversal of the size that is guaranteed by Turán's theorem.) To make our exposition self-contained, we reproduce Erdős's argument here.

Let us proceed by induction in the number of iterations. If *MIN* terminates instantly after initialization then we set $G = H$. Now assume that *MIN* goes through at least one iteration, let H_0 denote the value of H^* after the first iteration, and let $d+1$ be the number of vertices in $H - H_0$. By the induction hypothesis, there is a graph G_0 such that G_0 and H_0 have the same set of vertices, G_0 consists of $n-t-1$ vertex-disjoint cliques, and the degree of each vertex in H_0 is at least its degree in G_0. Since each vertex in H has degree at least d, we may let G consist of G_0 along with a clique on the $d+1$ vertices of $H - H_0$.

Note added in proof: In December 1990, A. Sidorenko informed us that he had obtained results intimately related to our Theorems 6.1 and 6.3: see "On Turán's problem for 3-graphs" (In Russian), *Kombinatorii Analiz* **6** (1983), 51–57 and "On exact values of Turán's numbers" (in Russian), *Matematicheskie Zametki* **42** (1987), 751–759.

References

[1] A. E. BROUWER, A. SCHRIJVER: The blocking number of an affine space, *Journal of Combinatorial Theory* A **24** (1978), 251–253.

[2] P. ERDŐS: On the graph-theorem of Turán (in Hungarian), *Mat. Lapok* **21** (1970), 249–251.

[3] R. E. JAMISON: Covering finite fields with cosets of subspaces, *J. Combinatorial Theory* A **22** (1977), 253–256.

[4] C. E. SHANNON: A theorem on coloring the lines of a network, *Journal of Mathematics and Physics* **27** (1949), 148–151.

[5] P. TURÁN: Egy gráfelméleti szélsőértékfeladatról, *Mat. Fiz. Lapok* **48** (1941), 436–452 *(see also "On the theory of graphs", Colloquium Mathematicum 3 (1954), 19–30).*

[6] P. TURÁN: Research problems, *Magyar Tud. Akad. Kutató Int. Közl.* **6** (1961), 417–423.

V. Chvátal

Department of Computer Science
Rutgers University
New Brunswick, NJ; U.S.A.
`chvatal@cs.rutgers.edu`

C. McDiarmid

Corpus Christi College
Oxford, England.
`MCD@vax.oxford.ac.uk`