MATHEMATICAL CENTRE TRACTS 106 (1979) 99-105.

7

TURÁN THEORY AND THE LOTTO PROBLEM

A.E. BROUWER & M. VOORHOEVE

REPRINT FROM:
MATHEMATICAL CENTRE TRACTS 106 (1979)
"PACKING AND COVERING IN COMBINATORICS"
(A. SCHRIJVER, ed.)

MATHEMATICAL CENTRE TRACTS 106 (1979) 99-105.

7

TURÁN THEORY AND THE LOTTO PROBLEM

A.E. BROUWER & M. VOORHOEVE

1. TURÁN THEORY

Let $k,\ell,n \in \mathbb{N}$ such that $k \leq \ell \leq n$. We define the *Turán number* $T(n,k,\ell)$ as the smallest number of k-subsets of an n-set X such that any ℓ-subset of X contains at least one of these k-subsets. For example: $T(7,4,5) = 7$. (Take $X = \{0,1,\ldots,6\}$; the 4-subsets are all translates (mod 7) of $\{1,2,3,5\}$; this is easily seen to be optimal.) The relation between Turán numbers and covering numbers is discussed in Chapter 4 and 5. The above definition can be formulated in the language of hypergraphs (see Chapter 1) as follows: for a hypergraph $H = (X,E)$, let its *stability number* $\beta(H)$ be the maximal cardinality of a stable subset of H (i.e. a set containing no edge). Then $T(n,k,\ell)$ is the minimal number of edges of a k-uniform hypergraph H with n vertices such that $\beta(H) < \ell$. P. TURÁN [10] posed the problem of determining $T(n,k,\ell)$. In this section we give some estimates for this number. Notice that $T(n,k,\ell)$ is increasing in n and k and decreasing in ℓ. Trivially, $T(n,1,\ell) = n-\ell+1$. The numbers $T(n,2,\ell)$ and the corresponding graphs are determined by the following theorem of TURÁN [9].

THEOREM 1. *Let* $n \geq \ell \geq 2$ *and let* $G_{n,\ell}$ *be the graph on* n *points consisting of* $\ell-1$ *disjoint cliques of cardinality either* $\lfloor\frac{n}{\ell-1}\rfloor$ *or* $\lceil\frac{n}{\ell-1}\rceil$. *Then every graph* G *with* n *vertices and stability number less than* ℓ *that has the smallest possible number of edges is isomorphic to* $G_{n,\ell}$.

PROOF. If $\ell \leq n \leq 2\ell-2$, the theorem is immediate. We proceed by induction on n. Denote the number of edges of a graph G by m(G). Let G_X be a graph with vertex set X, $|X| = n+\ell-1$, and stability number $< \ell$ such that $m(G_X)$ is minimal. Then $\beta(G) = \ell-1$. Let S be a stable subset of X with $|S| = \ell-1$. Let $G_{X\setminus S}$ be the subgraph of G_X induced by $X\setminus S$ (see Chapter 1).

By the maximality of S, each point in $X\setminus S$ is adjacent to a point in S, so

$$m(G_X) - m(G_{X\setminus S}) \geq n.$$

By the induction hypothesis, $m(G_{X\setminus S}) \geq m(G_{n,\ell})$, so

$$m(G_X) \geq m(G_{n,\ell}) + n.$$

It is easily checked by counting edges that

$$m(G_{n+\ell-1,\ell}) = m(G_{n,\ell}) + n.$$

Hence $m(G_X) = m(G_{n+\ell-1,\ell})$ and all the inequalities must therefore have been equalities. So $m(G_{X\setminus S}) = m(G_{n,\ell})$ and, by the induction hypothesis, $G_{X\setminus S} = G_{n,\ell}$. Furthermore, each point of $X\setminus S$ is adjacent in G_X to one and only one point of S. If two points from different cliques were adjacent to the same point s in S, this would contradict the maximality of S, so G_X consists of $\ell-1$ disjoint cliques of size

$$\left\lfloor\frac{n}{\ell-1}\right\rfloor + 1 \text{ or } \left\lceil\frac{n}{\ell-1}\right\rceil + 1, \quad \text{so } G_X \cong G_{n+\ell-1,\ell}. \qquad \square$$

<u>REMARK</u>. The case $\ell = 3$ appeared in 1910 as problem 28, by W. Mantel, in "Wiskundige Opgaven" of the Dutch Mathematical Society.

<u>COROLLARY</u>. $T(n,2,\ell) = (q-1)(n-\frac{1}{2}(\ell-1)q)$, *where* $q = \left\lceil\frac{n}{\ell-1}\right\rceil$.

Generalizing the above idea of taking disjoint cliques, we find for general k the upper bound

$$(1) \qquad T(n,k,\ell) \leq \binom{n}{k} \left\lfloor\frac{\ell-1}{k-1}\right\rfloor^{1-k}.$$

(Partition X into $\left\lfloor\frac{\ell-1}{k-1}\right\rfloor$ subsets S_i of almost equal size and take for $\mathcal{E}$ the collection of all k-subsets of each S_i.)

KATONA, NEMETZ & SIMONOVITS [6] proved that

$$T(n,k,\ell) \geq \frac{n}{n-k}\, T(n-1,k,\ell).$$

(Proof: For each point $x \in X$ there are at least $T(n-1,k,\ell)$ k-sets not containing x. Now count pairs (x,E), where $x \notin E \in \mathcal{E}$.)

Since $T(\ell,k,\ell) = 1$, we find by induction

THEOREM 2. $T(n,k,\ell) \geq \lceil \frac{n}{n-k} \cdot \lceil \frac{n-1}{n-k-1} \cdot \ldots \cdot \lceil \frac{\ell}{\ell-k+1} \rceil \ldots \rceil \rceil \geq \binom{n}{k}/\binom{\ell}{k}$.

COROLLARY. *For any hypergraph* $H = (X,E)$ *such that each edge of* H *contains at least* k *points, we have* $\beta(H) \geq \lfloor |X|/\sqrt[k]{|E|} \rfloor$.

(Proof: Let $n = |X|$ and $m = |E|$. If $m \leq (n/\ell)^k$ then $m < \binom{n}{k}/\binom{\ell}{k} \leq T(n,k,\ell)$, so $\beta(H) \geq \ell$.)

ERDÖS & SPENCER [4] generalized Theorem 2 by proving

PROPOSITION. $T(n,k,\ell) \geq (a-(\ell-1))\binom{n}{k}/\binom{a}{k}$ *for* $\ell \leq a \leq n$.

PROOF. $T(n,k,\ell) \geq \frac{n}{n-k} T(n-1,k,\ell) \geq \ldots \geq (\binom{n}{k}/\binom{a}{k})T(a,k,\ell)$. Now notice that $T(a,k,\ell) \geq T(a,1,\ell) = a-\ell+1$. □

We can also use $T(a,k,\ell) \geq T(a,2,\ell)$ and Turán's theorem (Theorem 1) to obtain for $k \geq 2$

$$T(n,k,\ell) \geq (\lceil \frac{a}{\ell-1} \rceil -1)(a-\tfrac{1}{2}(\ell-1)\lceil \frac{a}{\ell-1} \rceil)\binom{n}{k}/\binom{a}{k}, \quad \text{for } \ell \leq a \leq n.$$

This is stronger than Theorem 2 and Erdös & Spencer's result, but only in extreme cases is it essentially stronger.

CHVÁTAL [3] showed how to use *lower* bounds on $T(n,k,\ell)$ in order to obtain *upper* bounds for the same function (with different parameters). He proved

THEOREM 3. $T(\binom{n}{k}, \binom{\ell}{k}, \binom{n}{k} - T(n,k,\ell)+1) \leq \binom{n}{\ell}$.

PROOF. Let $X = P_k(U)$, where U is an n-set and choose an $(\binom{n}{k}-T(n,k,\ell)+1)$-subset Z of X. Then $X \backslash Z$ has $T(n,k,\ell)-1$ elements, so there is a $Y_1 \in P_\ell(U)$ such that no k-subset of Y_1 is an element of $X \backslash Z$. Hence $P_k(Y_1) \subset Z$. This proves that each $(\binom{n}{k}-T(n,k,\ell)+1)$-subset of X contains a set of the collection $E = \{P_k(Y) | Y \in P_\ell(U)\}$. Since $|E| = \binom{n}{\ell}$, this proves the theorem. □

COROLLARY. $T(n,k,\ell) < 1 + \binom{n}{k}(1-\binom{n}{\ell}^{-1/t})$, *where* $t = \binom{\ell}{k}$.

PROOF. Set $M = \binom{n}{k}$, $N = \binom{n}{k}-T(n,k,\ell)+1$, $S = \binom{\ell}{k}$. By Theorems 2 and 3

$$\binom{n}{\ell} \geq T(M,S,N) \geq \binom{M}{S}/\binom{N}{S} > (M/N)^S.$$

Substituting the given expressions for M,N and S we obtain the corollary. □

For certain n,k,ℓ this is an improvement of Turán's bound (1).

LOREA [7] determines some Turán numbers with the help of the affine spaces AG(k,2). By a result of BROUWER & SCHRIJVER [2] the minimum cardinality of a vertex subset of AG(k,2) intersecting all hyperplanes is k+1. So each set of cardinality 2^k-k contains a hyperplane. Since there are $2.(2^k-1)$ hyperplanes, this proves

$$T(2^k,2^{k-1},2^k-k) \le 2.(2^k-1).$$

By a direct application of Theorem 2 we find

$$T(2^k,2^{k-1},2^k-k) \ge 2.(2^k-1).$$

Hence

$$T(2^k,2^{k-1},2^k-k) = 2(2^k-1),$$

and AG(k,2) with the hyperplanes form a so-called *Turán hypergraph*.

2. THE LOTTO PROBLEM

In this section we treat the problem of determining the minimal number of lotto forms one must fill in to be assured of winning a prize. Formalized, this becomes the question of finding the minimum number $L(n,k,\ell,t)$ of k-subsets of an n-set X, such that any ℓ-subset of X meets one of these k-subsets in at least t points. (Assume $0 \le t \le k,\ell \le n$.)

For lotto in Holland, $n = 41$, $k = 6$, $\ell = 7$, $t = 4$; in Germany $n = 49$, $k = \ell = 6$, $t = 3$. The number $L(n,k,\ell,t)$ is increasing in n and t and decreaseing in k and ℓ. Trivially, $L(n,k,\ell,0) = 1$ and $L(n,k,\ell,1) = \lceil\frac{n-\ell+1}{k}\rceil$. When $t = \ell$ we have the covering problem: $L(n,k,t,t) = C(t,k,n)$. When $t = k$ we have Turán's problem: $L(n,k,\ell,k) = T(n,k,\ell)$. Bounds for $C(t,k,v)$ and $T(n,k,\ell)$ usually can be generalized to bounds for $L(n,k,\ell,t)$. The analogue of Theorem 1 becomes

<u>THEOREM 5</u>. (HANANI, ORNSTEIN & SÓS [4])

$$L(n,k,\ell,2) \ge \frac{n(n-\ell+1)}{k(k-1)(\ell-1)}, \tag{2}$$

and

$$\lim_{n\to\infty} L(n,k,\ell,2) . \frac{k.(k-1)(\ell-1)}{n(n-\ell+1)} = 1.$$

Equality in (2) *holds iff* $n = m(\ell-1)$ $(m \in \mathbb{N})$ *and there exists an* $S(2,k,m)$ *Steiner system. (In particular when* $k \le 5$ *and* $m \equiv 1$ *or* k $(\mathrm{mod}\ k(k-1))$.)

PROOF. Suppose $H = (X,E)$ is a k-uniform hypergraph with n vertices and $L(n,k,\ell,2)$ edges such that each ℓ-subset of X meets some edge in at least 2 points. Construct the graph $G = (X,E^*)$ whose edges are all pairs of points contained in any edge of H. Then

$$|E^*| \ge T(n,2,\ell) \ge \tfrac{1}{2}n(n-\ell+1)/(\ell-1)$$

by Theorem 1, since each ℓ-set contains an edge of G. Since each edge E of H contains only $\binom{k}{2}$ pairs, we have

$$L(n,k,\ell,2) = |E| \ge \frac{n(n-\ell+1)}{k(k-1)(\ell-1)} .$$

If equality holds in (2), then necessarily $T(n,2,\ell) = \tfrac{1}{2}n(n-\ell+1)/(\ell-1)$, so $(\ell-1)|n$. The graph G then consists of $\ell-1$ cliques of cardinality $m = n/(\ell-1)$. For equality in (2) it is also necessary that the pairs in these m-cliques are covered by k-sets, each pair lying in precisely one k-set, so each m-clique carries an $S(2,k,m)$ Steiner system. These conditions are clearly also sufficient. For the asymptotic result, notice that

$$L(n,k,\ell,2) \le (\ell-1).C(2,k,\lceil\tfrac{n}{\ell-1}\rceil).$$

By Wilson's theorem (see Chapter 5)

$$\lim_{n\to\infty} \frac{C(2,k,m)}{\binom{m}{2}/\binom{k}{2}} = 1.$$

Combining these results we find

$$\lim_{n\to\infty} \frac{L(n,k,\ell,2).k(k-1)}{(\lceil\frac{n}{\ell-1}\rceil).(\lceil\frac{n}{\ell-1}\rceil-1)(\ell-1)} = 1$$

thus completing the proof. □

When $C(2,k,m)$ is close to the Schönheim bound for m near $\frac{n}{\ell-1}$ it is often possible to determine $L(n,k,\ell,2)$ exactly. For instance:

$$L(2m+1,3,3,2) = C(2,3,m) + C(2,3,m+1),$$

$$L(4m+2,3,3,2) = 2.C(2,3,2m+1),$$

$$L(4m,3,3,2) = C(2,3,2m-1) + C(2,3,2m+1)$$

(see BROUWER [1]). Generalizing the above idea, we find

THEOREM 5. $L(n,k,\ell,t) \geq T(n,t,\ell)/\binom{k}{t}$.

Hence, by Theorem 2, we have

COROLLARY. $L(n,k,\ell,t) \geq \dfrac{\binom{n}{t}}{\binom{\ell}{t}\binom{k}{t}}$.

F. STERBOUL [8] gives the following two estimates, which are sometimes stronger for small n, though weaker for $n \to \infty$, k,ℓ,t fixed.

THEOREM 6.

(i) $$L(n,k,\ell,t) \geq \max_{\ell \leq a \leq n} \left\lceil \left\lceil \frac{a-\ell+1}{k-t+1} \right\rceil \binom{n}{a} \Big/ \sum_{i=t}^{k} \binom{k}{i}\binom{n-k}{a-i} \right\rceil$$

(ii) $$L(n,k,\ell,t) \geq \max_{\ell \leq a \leq n} \left\lceil (a-\ell+1)\binom{n}{a} \Big/ \sum_{i=t}^{k} \binom{k}{i}\binom{n-k}{a-i}(i-t+1) \right\rceil.$$

Regarding upper bounds, no good general constructions are known. STERBOUL [8] gives a construction for the French (and German) lotto, proving that

$$L(49,6,6,3) \leq 175.$$

The reader is hereby invited to give a construction for the Dutch lotto.

REFERENCES

[1] A.E. BROUWER, *Some lotto numbers*, Math. Centre report (to appear).

[2] A.E. BROUWER & A. SCHRIJVER, *The Blocking Number of an Affine Space*, J. Combinatorial Theory (A) 24 (1978) 251-253.

[3] V. CHVÁTAL, *Hypergraphs and Ramseyan theorems*, Proc. Amer. Math. Soc. 27 (1971) 434-440.

[4] P. ERDÖS & J. SPENCER, *Probabilistic methods in combinatorics*, Academic Press, New York, 1974.

[5] H. HANANI, D. ORNSTEIN & V.T. SÓS, *On the lottery problem*, Magyar Tud. Akad. Mat. Kutató Int. Közl. 9 (1964) 155-158.

[6] G. KATONA, T. NEMETZ & M. SIMONOVITS, *On a graph-problem of Turán*, Mat. Lapok 15 (1964) 228-238 (in Hungarian; Russian and English summaries).

[7] M. LOREA, *On Turán hypergraphs*, Discrete Math. 22 (1978) 281-285.

[8] F. STERBOUL, *Le problème du loto*, Proc. Coll. Intern. "Mathematiques Discrètes: Codes et hypergraphes" (Brussels, 1978) (to appear).

[9] P. TURÁN, *Egy gráfelméleti szélsőértek feladatrol*, Mat. Fiz. Lapok 48 (1941) 436-452; see also *On the theory of graphs*, Colloq. Math. 3 (1954) 19-30.

[10] P. TURAN, *Research problems*, Magyar Tud. Akad. Mat. Kutató Int. Közl. 6 (1961) 417-423.