Discrete Mathematics 14 (1976) 157–163.

A NOTE ON THE EDGES OF THE n-CUBE

Sergiu HART
Department of Mathematics and Statistics, Tel-Aviv University, Tel-Aviv, Israel

Received 12 February 1975
Revised 28 April 1975

The following combinatorial problem, which arose in game theory, is solved here: To find a set of vertices of a given size (in the n-cube) which has a maximal number of interconnecting edges.

1. The combinatorial problem

Let C_n denote the n-dimensional cube, and let $V = V(C_n)$ be the set of its vertices.

For any $S, T \subset V$, let $E(S, T)$ denote the set of all edges connecting S and T, and let $E(S)$ stand for $E(S, S)$. For any finite set A, let $|A|$ be the number of elements in A.

Let k be a given integer, $0 < k \leqslant 2^n$.

Problem 1.1. Maximize $|E(S)|$ subject to the condition $|S| = k$.

Problem 1.2. Minimize $|E(S, V \setminus S)|$ subject to the condition $|S| = k$.

We will denote the maximum of Problem 1.1 by $f_n(k)$ and the minimum of Problem 1.2 by $g_n(k)$.

Problem 1.2 arose in game theory in connection with the Banzhaf value (see Section 3). The intuitive meaning of the two problems is as follows. In Problem 1.1, one is looking for a "closely knit" set of vertices of a given size, i.e., a set with a maximal number of inter-connecting edges. As for Problem 1.2, the object is to minimize the number of external connections, again for a set of vertices of a given size.

The two problems are equivalent, in the following sense.

Proposition 1.3. *Let n and k satisfy $0 < k \leqslant 2^n$. Then a set S with $|S| = k$ is a solution to Problem* 1.1 *if and only if it is a solution to Problem* 1.2. *Moreover,*

$$2f_n(k) + g_n(k) = n \cdot k .$$

Proof. Let S be any set of k vertices of the n-cube. Then the total number of edges "emanating" from the members of S is $n \cdot k$; hence

$$2|E(S)| + |E(S, V \setminus S)| = n \cdot k ,$$

from which the proposition follows immediately.

For any non-negative integer i, let $h(i)$ denote the sum *of its binary digits*, i.e., the number of ones there.

We will now define a class of sets of vertices which will solve these two problems. The definition will be by induction on the number of elements of the set.

The set S with $|S| = k$ is *good* if

(i) $k = 1$, or

(ii) $2^r < k \leqslant 2^{r+1}$ and there is an $(r + 1)$-dimensional cube C_{r+1} and a partition of it into two vertex-disjoint r-dimensional cubes C_r^1 and C_r^2, such that S is the union of a good set of $k - 2^r$ vertices from C_r^1 and of the set of all 2^r vertices of C_r^2.

E.g., if the vertices of C_n are denoted as $0, 1, 2, ..., 2^n - 1$ in the usual way (i.e., an edge connects two vertices if and only if their corresponding numbers differ in the binary representation by exactly one digit), then $\{0, 1, ..., k - 1\}$ is a good set of k vertices.

Theorem 1.4. *For any n and k satisfying $0 < k \leqslant 2^n$,*

$$f_n(k) = f(k) = \sum_{i=1}^{k-1} h(i) ,$$

and the maximum is achieved for all good sets S with $|S| = k$.

Theorem 1.5. *For any n and k satisfying $0 < k \leqslant 2^n$,*

$$g_n(k) = n \cdot k - 2f(k) ,$$

and the minimum is achieved for all good sets S with $|S| = k$.

Remark 1.6. (i) The two theorems are equivalent by Proposition 1.3; hence we will prove only Theorem 1.4 (in the next section).

(ii) The maximum in Problem 1.1 is independent of n.

2. Proof of the theorems

We define inductively the function F on the set of positive integers

$$F(1) = 0 ,$$

$$F(k) = \max_{0 < m \leqslant k/2} (F(m) + F(k - m) + m) .$$

Proposition 2.1. *Let k be a positive integer, and let $2^r < k \leqslant 2^{r+1}$. Then*

(i) *the maximum in the definition of $F(k)$ is achieved at $m = k - 2^r$;*

(ii) $F(k) = \sum_{i=0}^{k-1} h(i)$.

We first have to prove the following lemmas.

Lemma 2.2. *For all $j \geqslant 0, l \geqslant 1$,*

$$\sum_{i=0}^{l-1} h(i) \leqslant \sum_{i=j}^{j+l-1} h(i) .$$

Proof. On both sides we have the sum of the binary digits of l consecutive integers. The kth digit appears periodically as 2^k zeroes and 2^k ones. Hence the minimum number of ones is attained when beginning with 0, which is true for all k, and hence proves the lemma.

Lemma 2.3. *For all $0 < j \leqslant 2^n, l \geqslant 0$,*

$$\sum_{i=2^n-1-l}^{2^n-1} h(i) \geqslant \sum_{i=j-1-l}^{j-1} h(i) .$$

Proof. The proof is the same as for the previous lemma, maximizing here the number of ones.

Lemma 2.4. *For all $r \geqslant 0$,*

$$\sum_{i=0}^{2^r-1} h(i) = r \cdot 2^{r-1} .$$

Proof. Each of the r binary digits is equal to one for exactly 2^{r-1} numbers.

Proof of Proposition 2.1. For $k = 1$ the proposition holds.

Let $k = 2^r + s$, where $0 < s \leqslant 2^r$, and assume that we have proved (i) and (ii) for all integers less than k.

First, we will prove (i), i.e., for any $0 < m \leqslant \frac{1}{2}k$ we have to show that

$$F(m) + F(k-m) + m \leqslant F(s) + F(2^r) + s . \tag{1}$$

Case 1: $m < s$. Using the induction hypothesis we get that (1) is equivalent to

$$\sum_{i=0}^{m-1} h(i) + \sum_{i=0}^{2^r-1} h(i) + \sum_{i=2^r}^{k-m-1} h(i) + m \leqslant$$
$$\leqslant \sum_{i=0}^{m-1} h(i) + \sum_{i=m}^{s-1} h(i) + \sum_{i=0}^{2^r-1} h(i) + s ,$$

or

$$\sum_{i=2^r}^{k-m-1} h(i) - (s-m) \leqslant \sum_{i=m}^{s-1} h(i) .$$

But $h(i) - 1 = h(i-2^r)$ for all $2^r \leqslant i < 2^{r+1}$, hence we have to prove that

$$\sum_{i=0}^{s-m-1} h(i) \leqslant \sum_{i=m}^{s-1} h(i) ,$$

which is true by Lemma 2.2.

Case 2: $s < m \leqslant 2^{r-1}$. As in Case 1, we get from (1),

$$\sum_{i=s}^{m-1} h(i) + (m-s) \leqslant \sum_{i=2^r-m+s}^{2^r-1} h(i) .$$

Since $h(i) + 1 = h(i + 2^{r-1})$ for all $i < 2^{r-1}$, and $m - 1 < 2^{r-1}$, it follows from Lemma 2.3 that the inequality holds in this case too.

Case 3: $2^{r-1} < m \leqslant \frac{1}{2}k$. Let $m' = m - 2^{r-1}$; then the left-hand side of (1) is equal to

$$\sum_{i=0}^{2^{r-1}-1} h(i) + \sum_{i=0}^{m'-1} (h(i)+1) + \sum_{i=0}^{2^{r-1}-1} h(i)$$
$$+ \sum_{i=0}^{s-m'-1} (h(i)+1) + (m' + 2^{r-1}) .$$

By Lemma 2.4 and the induction hypothesis we get

$$(r-1)\cdot 2^{r-2}+F(m')+m'+(r-1)2^{r-2}+F(s-m')+s-m'+m'+2^{r-1}=$$

$$=(F(m')+F(s-m')+m')+r\cdot 2^{r-1}+s\ .$$

But $0<m'\leqslant \frac{1}{2}s$ and $r\cdot 2^{r-1}=F(2^r)$ (by Lemma 2.4); hence

$$F(m)+F(k-m)+m\leqslant F(s)+F(2^r)+s\ ,$$

which is (1).

Hence (i) is proved, and (ii) follows immediately using Lemma 2.4, which proves the proposition.

Remark 2.5. It can be easily computed that the maximum in the definition of $F(k)$ is also achieved at $m=[\frac{1}{2}k]$.

Proof of Theorem 1.4. Let C^1_{n-1} and C^2_{n-1} be a partition of C_n into two vertex-disjoint $(n-1)$-dimensional cubes. Let $S\subset V(C_n)$, $S^i=S\cap V(C^i_{n-1})$ for $i=1,2$, $k=|S|$, $k^i=|S^i|$. Since each vertex of C^1_{n-1} is connected in C_n to exactly one vertex of C^2_{n-1}, we get

$$|E(S)|\leqslant |E(S^1)|+|E(S^2)|+\min(k^1,k^2). \tag{2}$$

Assume by induction that the theorem is true for cubes of dimension less than n; then

$$|E(S)|\leqslant F(k^1)+F(k^2)+\min(k^1,k^2)\ ,$$

and from the definition of F we get

$$|E(S)|\leqslant F(k)\ ,$$

hence

$$f_n(k)\leqslant F(k)\ .$$

If S is a good set for k, where $2^r<k\leqslant 2^{r+1}$, then from the definition of a good set and applying the same arguments as above, we get

$$|E(S)|=|E(S\cap V(C^1_r))|+|E(S\cap V(C^2_r))|+(k-2^r)$$

$$=F(k-2^r)+F(2^r)+k-2^r=F(k)\ ,$$

which proves the theorem.

3. Application to game theory

Let (N, v) be an n-person simple game, i.e., $N = \{1, 2, ..., n\}$ is the set of players, and $v : 2^N \to \{0, 1\}$ is the characteristic function ($v(\emptyset) = 0$).

The player i is a *swing* for the coalition $S \subset N$, $i \notin S$, if $v(S) = 0$ and $v(S \cup \{i\}) = 1$. The number of coalitions for which i is a swing is denoted by η_i, and $\bar{\eta} = \sum_{i=1}^{n} \eta_i$. The *Banzhaf value* (or index) of the player i is then $\beta_i = \eta_i/\bar{\eta}$.

The work done on this concept (see [1]) has raised some problems about the total number of swings $\bar{\eta}$.

A coalition S is *winning* if $v(S) = 1$ and *losing* if $v(S) = 0$. Let k_w denote the number of winning coalitions and k_l the number of losing coalitions in the (simple) game (N, v); then

$$k_w + k_l = 2^n .$$

Corollary 3.1. *Let (N, v) be a simple game, and let $k = k_w$ or $k = k_l$. Then*

$$\bar{\eta} \geq k \cdot n - 2 \sum_{i=1}^{k-1} h(i) .$$

Proof. Represent each coalition as a vertex of the cube C_n; then obviously $\bar{\eta} \geq g_n(k)$ and the result follows from Theorem 1.5.

Next, let $p = \bar{\eta}/(n \cdot 2^{n-1})$. Then p is the *probability of a swing*, where each pair of coalitions differing by exactly one player has the same probability $1/(n \cdot 2^{n-1})$.

Let $n > 2$, and let $k = \min(k_w, k_l)$; then $2^r < k \leq 2^{r+1}$ for some $r \leq n - 2$. As before, $m = k - 2^r$, and we get from Proposition 2.1,

$$\begin{aligned} \bar{\eta} &\geq g_n(k) = k \cdot n - 2F(k) \\ &= m \cdot n + 2^r \cdot n - 2(F(m) + F(2^r) + m) \\ &= g_n(2^r) + g_{n-2}(m) \\ &\geq g_n(2^r) = (n - r) \cdot 2^r . \end{aligned}$$

But $r < \log_2 k$ and $2^r \geq \frac{1}{2} k$; hence

$$p = \frac{\bar{\eta}}{(n \cdot 2^{n-1})} > \frac{1}{n} \cdot \left(\frac{k}{2^n}\right) \cdot \left(-\log_2 \left(\frac{k}{2^n}\right)\right) .$$

Since $k/2^n$ is equal either to $k_w/2^n$ or to $k_l/2^n = 1 - k_w/2^n$, i.e., $k/2^n$ is the proportion of winning or losing coalitions in the game, we have the following corollary.

Corollary 3.2. *Let (N_n, v_n) be a sequence of simple games for $n = 1, 2, \ldots$, with $|N_n| = n$. If the proportion of winning coalitions has a limit other than 0 or 1 as $n \to \infty$, then*

$$\frac{p_n}{1/n} \geqslant c\,,$$

where p_n is the probability of a swing in the nth game, and c is a positive constant (depending on the sequence but not on n).

This corollary implies that p_n *cannot* decrease faster than $1/n$.

Added in proof. We thank Prof. M. Perles for pointing out the following relevant references to the combinatorial problem: L.H. Harper, Optimal assignments of numbers to vertices, J. SIAM 12 (1964) 131–135; A.J. Bernstein, Maximally connected arrays on the *n*-cube, SIAM J. Appl. Math. 15 (1967) 1485–1489.

Acknowledgment

The problem was suggested by Dr. L.S. Shapley, whom we thank for his help. The suggestions of an anonimous referee are also acknowledged, especially Proposition 1.3. The work has been done during the Colloquium on Mathematical Economics, University of California, Berkeley, August 1974.

Reference

[1] P. Dubey and L.S. Shapley, Some Properties of the Banzhaf Power Index (RAND, Santa Monica, 1975).