NOTE

A SHORT PROOF FOR A THEOREM OF HARPER ABOUT HAMMING-SPHERES

P. FRANKL

CMS, 54 bd. Raspail, Paris 75007, France

Z. FÜREDI

Mathematical Institute of the Hungarian Academy of Sciences, Realianoda u. 13–15, Budapest, Hungary

Received 27 June 1980

The Hamming-distance of two 0–1 sequences \( \alpha = (\alpha_i)_{i=1}^n \) and \( \beta = (\beta_i)_{i=1}^n \) is the number of different coordinates. In other terminology, the distance of two sets \( A \) and \( B \) is the cardinality of their symmetric difference, \( d(A, B) = |A \Delta B| \). (With this distance the set-system \( P(X) \) consisting of all subsets of the finite set \( X \) is a metric space).

A Hamming-sphere with center \( C \) is a set-system \( \mathcal{S} \subset P(X) \) such that for some \( k \):

\[
\{S \subseteq X : d(S, C) \leq k\} \subseteq \mathcal{S} \subseteq \{S \subseteq X : d(S, C) \leq k+1\}.
\]

The \( d \)-neighbourhood of a set-system \( \mathcal{A} \subset P(X) \) is

\[
\Gamma_d \mathcal{A} = \{Y \subseteq X : d(Y, \mathcal{A}) = \min_{A \in \mathcal{A}} d(Y, A) \leq d\}.
\]

It was Harper who first proved that the cardinality of \( \Gamma_d \mathcal{A} \) is at least as large as the \( d \)-neighbourhood of some appropriate Hamming-sphere with the same cardinality \( |\mathcal{A}| \). This theorem has important applications in information theory. Katona [3] gives a different proof. For a generalization see Margulis [5] (Blowing-up lemma). Here we give a new proof for Harper's theorem in an equivalent form.

**Theorem.** Let \( \mathcal{A} \) and \( \mathcal{B} \) be set-systems on \( X \) and

\[
d(\mathcal{A}, \mathcal{B}) = \min\{d(A, B) : A \in \mathcal{A}, B \in \mathcal{B}\} = d.
\]

Then there are two Hamming-spheres \( \mathcal{A}_0 \) with center \( X \) and \( \mathcal{B}_0 \) with center \( \emptyset \) such that \( |\mathcal{A}_0| = |\mathcal{A}|, |\mathcal{B}_0| = |\mathcal{B}| \) and \( d(\mathcal{A}_0, \mathcal{B}_0) \geq d(\mathcal{A}, \mathcal{B}) \).

**Proof.** Consider the set of pairs

\[
\{(A, A^*) : A \in \mathcal{A}, A^* \notin \mathcal{A}, |A| < |A^*|\}
\]
and
\[ (B, B^\ast): B \in \mathcal{B}, B^\ast \notin \mathcal{B}, |B| > |B^\ast|. \]
If there are no such pairs, then \( \mathcal{A} \) is an \( X \)-centered and \( \mathcal{B} \) is an \( \emptyset \)-centered Hamming-sphere, and then there is nothing to prove.

Otherwise let us choose a pair \((A, A^\ast)\) or \((B, B^\ast)\) with minimal symmetric difference \(|A \Delta A^\ast|\) or \(|B \Delta B^\ast|\) resp. Assume this minimal pair is \((A_0, A_0^\ast)\).

Set
\[ A_0 - A_0^\ast = U, \quad A_0^\ast - A_0 = V, \quad |U| < |V|. \]
For these sets \( U \) and \( V \) we define the following two operations (Up and Down).

\[ \mathcal{U}(A) = \begin{cases} A - U + V & \text{if } U \subseteq A, V \cap A = \emptyset, A - U + V \notin \mathcal{A}, \\ A & \text{otherwise.} \end{cases} \]
\[ \mathcal{D}(B) = \begin{cases} B - V + U & \text{if } V \subseteq B, U \cap B = \emptyset, B - V + U \notin \mathcal{B}, \\ B & \text{otherwise.} \end{cases} \]

It is clear that the mapping \( \mathcal{U} \) and \( \mathcal{D} \) are one-to-one and thus \(|\mathcal{U}(\mathcal{A})| = |\mathcal{A}|\), \(|\mathcal{D}(\mathcal{B})| = |\mathcal{B}|\), further \(|\mathcal{U}(A)| \geq |A|\), \(|\mathcal{D}(B)| \leq |B|\). Since \( \mathcal{U}(A_0) = A_0^\ast \), the joint application \( \mathcal{U} \) and \( \mathcal{D} \) strictly increases the quantity \( (\sum |A| - \sum |B|) \). We show \( d(\mathcal{U}(\mathcal{A}), \mathcal{D}(\mathcal{B})) \geq d(\mathcal{A}, \mathcal{B}) \), and thus the repeated applications of \( \mathcal{U} \) and \( \mathcal{D} \) finally lead to two Hamming-spheres.

If \( A \in \mathcal{U}(\mathcal{A}) \cap \mathcal{A} \) and \( B \in \mathcal{D}(\mathcal{B}) \cap \mathcal{B} \), then clearly \( d(A, B) \leq d \). Similarly, if \( A' \in \mathcal{U}(\mathcal{A}) \setminus \mathcal{A}, B' \in \mathcal{D}(\mathcal{B}) \setminus \mathcal{B} \), then \( A' = A - U + V, B' = B - V + U \) and thus \( A' \Delta B' = A \Delta B \) where \( A \in \mathcal{A}, B \in \mathcal{B} \). Therefore \( |A' \Delta B'| = |A \Delta B| \geq d \). This settles the cases of two old or two new sets.

If one set is new and the other is unchanged, e.g.
\[ A' \in \mathcal{U}(\mathcal{A}) \setminus \mathcal{A}, \quad B \in \mathcal{D}(\mathcal{B}) \cap \mathcal{B}, \]
then \( A' = A - U + V \) where \( A \in \mathcal{A} \).

If \( V \subseteq B \) and \( U \cap B = \emptyset \), then \( B \) has not been changed to a smaller set by the operation \( \mathcal{D} \) only because \( \bar{B} = (B - V + U) \in \mathcal{B} \). Thus \( A' \Delta B = A \Delta \bar{B} \) whence \( d(A', B) = d(A, \bar{B}) \geq d \).

If the condition \( (V \subseteq B, U \cap B = \emptyset) \) is not satisfied and \( U = \emptyset \), then \( V \notin \mathcal{B} \). Further \( A_0 \subseteq A_0^\ast \), thus the minimal choice of \((A_0, A_0^\ast)\) implies \(|V| = 1\). We infer
\[ A' \Delta B = (A + V) \Delta B = (A \Delta B) + V, \]
consequently \(|A' \Delta B| \geq d + 1\).

Finally, if \( 1 \leq |U| < |V| \) and the condition \((V \subseteq B, U \cap B = \emptyset)\) is not satisfied then there are two elements \( u \in U, v \in V \) such that at least one of the inclusions \( v \in V - B, u \in U \cap B \) holds. Since
\[ |\tilde{A}| = |A - (U - u) + (V - v)| = |A'| > |A| \quad \text{and} \quad |A \Delta \tilde{A}| < |A_0 \Delta A_0^\ast|, \]
the definition of \( A_0 \) implies that \( \tilde{A} \in \mathcal{A} \). Further \( A' = (\tilde{A} - u + v) \) and thus
$A' \Delta B = (\tilde{A} - u + v) \Delta B$. If we delete the element $u$ from $\tilde{A}$, then $|\tilde{A} \Delta B|$ increases or decreases by 1 according to whether $u \in B$ or not. Further if we adjoin the element $v$ to $(\tilde{A} - u)$ then $(\tilde{A} - u) \Delta B$ increases or decreases by 1 according to whether $v \notin B$ or not. Thus in any case

$$|A' \Delta B| = |(\tilde{A} - u + v) \Delta B| \geq |\tilde{A} \Delta B| \geq d. \quad \square$$

If

$$\sum_{j=k+1}^{n} \binom{n}{j} < a \leq \sum_{j=k}^{n} \binom{n}{j},$$

then the exact computation of $\min\{|I_d \mathcal{A}|: |\mathcal{A}| = a\}$ has been reduced to the following problem: Given a set-system $\mathcal{F}$ of $(a - \sum_{j=k}^{n} \binom{n}{j})$ $k$-element sets, at least how many $(k - d)$-element subsets are contained in the sets of $\mathcal{F}$? This well-known problem is answered by the theorem of Kruskal and Katona [2, 4] which states that if

$$|\mathcal{F}| = \binom{a_k}{k} + \binom{a_{k-1}}{k-1} + \cdots + \binom{a_t}{t},$$

where $a_k > a_{k-1} > \cdots > a_t \geq t$ (the representation of $|\mathcal{F}|$ in this form is unique), then

$$\{|Y| = k - d, \exists F \in \mathcal{F} \subset F\} \geq \binom{a_k}{k-d} + \binom{a_{k-1}}{k-d-1} + \cdots + \binom{a_t}{t-d}.$$

References


