.

ON THE UPPER BOUND OF THE SIZE OF THE r-COVER-FREE FAMILIES

MIKLÓS RUSZINKÓ †
RESEARCH GROUP for INFORMATICS and ELECTRONICS
HUNGARIAN ACADEMY of SCIENCES and
MATHEMATICAL INSTITUTE of the HUNGARIAN ACADEMY of SCIENCES
BUDAPEST, P.O.B. 127, 1364 HUNGARY; E-MAIL: h2942rus@ella.hu

ABSTRACT

Let $T(r,n)$ denote the maximum number of subsets of an n-set satisfying the condition in the title. It is proved in a purely combinatorial way, that for n sufficiently large

$$\frac{\log_2 T(r,n)}{n} \le 8 \cdot \frac{\log_2 r}{r^2}$$

holds.

1. Introduction

The notion of the r-cover-free families was introduced by Kautz and Singleton in 1964 [17]. They initiated investigating binary codes with the property that the disjunction of any $\le r$ ($r \ge 2$) codewords are distinct (UD_r codes). This led them to studying the binary codes with the property that none of the codewords is covered by the disjunction of $\le r$ others (*Superimposed codes, ZFD_r codes;* P. Erdős, P. Frankl and Z. Füredi called the correspondig set system r-cover-free in [7]).

Since that many results have been proved about the maximum size of these codes. Various authors studied these problems basically from three different points of view, and these three lines of investigations were almost independent of each other. This is why many results were found first in information theory ([1], [4], [5], [14], [15], [16], [17]), were later rediscovered in combinatorics ([2], [6], [7], [10], [18], [19]), or in group testing ([12], [13]), and vice versa.

We shall approach this area from the combinatorial side. Our main goal is to estimate the maximal size of the family of subsets of an n-element set with the property that no set is covered by the union of r others.

† This research was supported by the Hungarian National Foundation for Scientific Research, Grant No. T4271

2. Notation, definitions

Let S be an n-element set. 2^S is the set of all subsets of S. $\binom{S}{k}$ denotes the set of all k-subsets of S ($k \geq 0$). If $|S| = n$, then $|\binom{S}{k}| = \binom{n}{k}$. We denote by $[n]$ the set $\{1, 2, \ldots, n\}$, and $\log x$ is always of base 2. A set system $\mathcal{A} \subseteq 2^S$ is called k-uniform if its members are k- sets. It is usually supposed that the underlying set of the set systems is $[n]$.

We call $\mathcal{F}' \subset 2^S$ r-distinct, if

$$\bigcup_{i=1}^{k} A_i \neq \bigcup_{j=1}^{\ell} B_i$$

for any

$$\{A_1, A_2, \ldots, A_k\} \neq \{B_1, B_2, \ldots, B_\ell\},$$

$1 \leq k, \ell \leq r$; $A_1, A_2, \ldots, A_k, B_1, B_2, \ldots, B_\ell \in \mathcal{F}'$. $\mathcal{F} \subset 2^S$ is r-cover-free, if

$$A_0 \not\subseteq A_1 \cup A_2 \cup \ldots \cup A_r$$

holds for all distinct $A_0, A_1, \ldots, A_r \in \mathcal{F}$. $\mathcal{F}^\star \subset 2^S$ is $< r$ part intersecting, if

$$|A_i \cap A_j| < \frac{1}{r} \min\{|A_i|, |A_j|\}$$

for any distinct $A_i, A_j \in \mathcal{F}^\star$ holds. We denote by $T'(r,n)$, $T(r,n)$, $T^\star(r,n)$ and $T'(r,n,k)$, $T(r,n,k)$, $T^\star(r,n,k)$ the maximum cardinality of the corresponding set systems in general and in k-uniform case, resp. We will provide upper bounds on these functions for r fixed and n tending to infinity. We shall call

$$R(r) = \limsup_{n\to\infty} \frac{\log T(r,n)}{n}$$

the *rate* of the r-cover-free family. The following proposition is obvious from the definitions.

Proposition 2.1. *If $\mathcal{F}$ is $< r$ part intersecting, then $\mathcal{F}$ is r-cover-free; and if $\mathcal{F}$ is r-cover-free, then $\mathcal{F}$ is r-distinct. Hence*

$$T^\star(r,n) \leq T(r,n) \leq T'(r,n), \qquad \text{and} \qquad T^\star(r,n,k) \leq T(r,n,k) \leq T'(r,n,k).$$

The following upper and lower bounds were proved in [1], [4], [5], [7], [13]: there exist two (absolute) constants c_1, c_2 such that

$$\frac{c_1}{r^2} \leq \frac{\log T(r,n)}{n} \leq \frac{c_2}{r} \tag{1}$$

for any n. In most papers the lower bound is proved by probabilistic methods. In [13] V.T. Sós and F.K. Hwang used a greedy-type algorithm to generate $< r$ part intersecting families for proving the lower bound. The upper bound was proved using the observation that, by definition, $\sum_{i=1}^{r} \binom{T'}{i} \leq 2^n$. The gap between the upper and lower bounds is rather large. Dyachkov and Rykov obtained a better upper bound [4]:

$$\frac{\log T(r,n)}{n} \leq c_3 \frac{\log r}{r^2} \tag{2}$$

for some absolute constant c_3 and any n. Their proof is rather involved. Here we shall give a simple and purely combinatorial proof of this result.

3. The upper bound

First we have to prove some lemmata.

Lemma 3.1. *If r divides k then*

$$T(r,n,k) \leq r \frac{\binom{n}{\frac{k}{r}}}{\binom{k}{\frac{k}{r}}} \tag{4}$$

Proof.

Let $\mathcal{F}$ be a k-uniform r-cover-free family and let r divide k. Let A_0 be an arbitrary element of $\mathcal{F}$. By Baranyai's theorem [3] (asserting, that if r divides k then the r-uniform complete hypergraph on k vertices has one-factorisation) one can list all subsets of size $\frac{k}{r}$ of A_0 in the following way:

(a) in each row there is a partition of A_0;

(b) each subset is listed only once.

So each row will contain r subsets and the number of the rows will be $s = \frac{\binom{k}{\frac{k}{r}}}{r}$. This family of partitions can be represented in the following matrix form:

$$\begin{matrix} B_{1,1} & B_{1,2} & \dots & B_{1,r} \\ B_{2,1} & B_{2,2} & \dots & B_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ B_{s,1} & B_{s,2} & \dots & B_{s,r} \end{matrix}$$

where (a) means that $\bigcup_{j=1}^{k} B_{i,j} = A_0$ for any $1 \leq i \leq s$ and (b) means that $B_{i,j} \neq B_{k,\ell}$ for any $(i,j) \neq (k,\ell)$.

For any $1 \leq i \leq s$ the ith row contains at least one set $B_{i,j}$ which is not contained in any other $A_j \in \mathcal{F}$. Indeed, otherwise for each i we could fix a $B_{i,j} \subseteq A_j \neq A_0$, and therefore

$$A_0 = B_{i,1} \cup B_{i,2} \cup \ldots \cup B_{i,r} \subseteq A_1 \cup A_2 \cup \ldots \cup A_r,$$

which would violate the condition.

Hence A_0 has at least s subsets of size $\frac{k}{r}$, which are not contained in any other $A_j \in \mathcal{F}$ (*own subsets*). Since the number of all subsets of the underlying set is $\binom{n}{\frac{k}{r}}$, we get that

$$|\mathcal{F}|s = |\mathcal{F}|\frac{\binom{k}{\frac{k}{r}}}{r} \leq \binom{n}{\frac{k}{r}}$$

∎

From Lemma 3.1 the following stronger version of Lemma of [4] follows trivially.

Lemma 3.2.

$$T(r,n,k) \leq r\frac{\binom{n}{\lceil \frac{k}{r} \rceil}}{\binom{\lfloor \frac{k}{r} \rfloor r}{\lfloor \frac{k}{r} \rfloor}} \tag{5}$$

∎

Remark 1. Proposition 2.1 of Erdős, Frankl, Füredi [7] asserts that if we set $t = \lceil k/r \rceil$ then

$$T(r,n,k) \leq \frac{\binom{n}{t}}{\binom{k-1}{t-1}}. \tag{$*$}$$

This result was obtained by using the following

Lemma 1F. *(Frankl [8]). If $\mathcal{F}$ is a family of t-subsets of a k-set, and for any r sets*

$$F_1, \ldots, F_r \in \mathcal{F} \qquad \bigcap_{j=1}^{r} F_j \neq \emptyset,$$

if $rt/(r-1) \leq k$, then $|\mathcal{F}| \leq \binom{k-1}{t-1}$.

In the case when r does not divide k, $(*)$ is stronger than (5), but since it does not give a better exponent for $T(r,n,k)$, we shall use the inequality of Lemma 3.1.

It is also worth mentioning that in [9] (Theorem 3.4) Frankl and Füredi proved that for fixed k and r

$$\lim_{n\to\infty} T(r,n,k)/\binom{n}{t}$$

exists and equals to $\left(\binom{k}{t} - m(k,t,l)\right)^{-1}$, where $l = k - r(t-1) - 1$. (For the definition of $m(k,t,l)$ see [9].)

Obviously

$$T(r,n) \le n \cdot \max_{1\le k\le n} T(r,n,k) \tag{6}$$

and by (1) $T(r,n)$ is exponential in n. So the factor n is insignificant in (6). This leads to the following question: for which k attains $T(r,n,k)$ its maximum? If we knew this, we could estimate $T(r,n)$ from (5). But for $k = \frac{n}{2}$ (5) yields only

$$\frac{\log T(r,n,\frac{n}{2})}{n} \le \frac{c}{r},$$

(for some constant c).

Proposition 3.1. $T(r,n) \le T(r,2n,n)$

Proof.

Let $\mathcal{F} = \{A_1,\dots,A_T\}$, $A_i \subseteq [n]$. For $i = 1,\dots,T$ let $B_i = \{x+n : x \in \bar{A}_i\}$ ($B_i \subseteq \{n+1,\dots,2n\}$), $C_i = A_i \cup B_i$ ($C_i \subseteq [2n]$) and $\mathcal{F}^u = \{C_1,\dots,C_T\}$. $\mathcal{F}^u$ is an n-uniform r-cover-free family on a $2n$ element set, $|\mathcal{F}^u| = |\mathcal{F}|$. ∎

By Proposition 3.1,

$$R(r) \le 2 \cdot \limsup_{n\to\infty} \frac{\log T(r,2n,n)}{2n},$$

so the size of the $\frac{n}{2}$ uniform r-cover-free families can be very large. (The rate is at most twice less than in the non uniform case.) Thus, if we want to get a better upper bound for $\frac{\log T(r,n)}{n}$ than $\frac{c}{r}$, it is not enough to use the inequality (5). We have to compress somehow the elements of $\mathcal{F}$ without loosing the r-cover-free property, since for smaller k (5) gives a better bound.

Lemma 3.3. *Let A_i be an arbitrary element of $\mathcal{F} = \{A_1, A_2, \dots, A_T\}$ and $B_i \subseteq A_i$ an arbitrary subset of A_i. If $\mathcal{F}$ is r-cover-free, then*

(a) $\mathcal{F}^1 = \{A_j \setminus B_i\}_{j=1,\dots,T}^{j\ne i}$ *is $(r-1)$-cover-free,*

(b) $|\mathcal{F}^1| = T - 1$.

Proof.

(a) Suppose that

$$A_{j_0} \setminus B_i \subseteq (A_{j_1} \setminus B_i) \cup (A_{j_2} \setminus B_i) \cup \ldots \cup (A_{j_{r-1}} \setminus B_i)$$

for some $\{j_0, j_1 \dots, j_{r-1}\} \subseteq [T] \setminus \{i\}$. Then

$$A_{j_0} \subseteq A_{j_1} \cup A_{j_2} \cup \ldots \cup A_{j_{r-1}} \cup A_i,$$

which is a contradiction.

(b_1) For any $A_j \neq A_i$ we have $A_j \nsubseteq A_i$, so we threw out only A_i from $\mathcal{F}$.

(b_2) $A_k \setminus B_i \neq A_l \setminus B_i$ if $k \neq l$, so we didn't merge any two distinct members of $\mathcal{F}$. Indeed, suppose that $A_k \setminus B_i = A_l \setminus B_i$ and $k \neq l$. Then $A_k \subseteq A_l \cup A_i$, which is a contradiction $(r \geq 2)$. ∎

Theorem 3.1.

$$\frac{\log T(r,n)}{n} \leq 8 \cdot \frac{\log r}{r^2}. \tag{6}$$

Proof.

First we will assume that r^2 divides n and $\frac{n}{r}$ is even. Let $\mathcal{F}$ be an arbitrary r-cover-free family. We will use the following *set compression algorithm.*

1) $\mathcal{F}^0 = \mathcal{F}$.

2) If some element of $\mathcal{F}^i$ is of size $\leq \frac{2n}{r}$, then $\tilde{\mathcal{F}} = \mathcal{F}^i$. If $\mathcal{F}^i = \{A_1^{(i)}, A_2^{(i)}, \ldots, A_{T-i}^{(i)}\}$ contains a set $A_{j_0}^{(i)}$ of size $> \frac{2n}{r}$, then put $\mathcal{F}^{i+1} = \{A_j^{(i)} \setminus A_{j_0}^{(i)}\}_{j=1,\ldots,T-i}^{j_0 \neq j}$.

In each step of this algorithm we throw out more than $\frac{2n}{r}$ elements. Since the underlying set of $\mathcal{F}$ is of size n, our algorithm will stop in at most $\frac{r}{2}$ steps. Suppose that during this algorithm we threw out p elements from the underlying set in q steps. Let $T(r, n, \leq k)$ denote the maximum cardinality of an r-cover-free family of subsets of $[n]$ of size $\leq k$. Then, by Lemma 3.3 and set compression algorithm,

$$\begin{aligned}
T(r,n) &\leq T\left(r-q, n-p, \leq \frac{2n}{r}\right) + q \leq T\left(\frac{r}{2}, n, \leq \frac{2n}{r}\right) + \frac{r}{2} \\
&\leq \sum_{k=1}^{\frac{2n}{r}} T(\frac{r}{2}, n, k) + \frac{r}{2} \leq \sum_{k=1}^{\frac{2n}{r}} \frac{\frac{r}{2}\binom{n}{\frac{2k}{r}}}{\binom{k}{\frac{2k}{r}}} + \frac{r}{2} \\
&\leq \sum_{k=1}^{\frac{2n}{r}} \frac{r}{2} \cdot \binom{n}{\frac{2k}{r}} \leq n \cdot \binom{n}{\frac{4n}{r^2}}.
\end{aligned}$$

Let $h(x)$ $(0 < x < 1)$ be the binary entropy function, that is,

$$h(x) = x\log\frac{1}{x} + (1-x)\log\frac{1}{1-x}.$$

Then, by [11],

$$\binom{n}{cn} \leq 2^{n \cdot h(c)} \cdot n^{k_1},$$

where k_1 is an absolute constant. Therefore

$$\begin{aligned}
\log T(r,n) &\leq (k_1+1)\cdot\log n + n\cdot h\left(\frac{4}{r^2}\right) \\
&= o(n) + n\left(\frac{4}{r^2}\log\frac{r^2}{4} + \left(1-\frac{4}{r^2}\right)\log\left(\frac{r^2}{r^2-4}\right)\right) \\
&= o(n) + n\left(8\frac{\log r}{r^2} - \frac{8}{r^2} + \frac{1}{r^2}\log\left(1+\frac{4}{r^2-4}\right)^{r^2-4}\right) \\
&\leq o(n) + n\left(8\frac{\log r}{r^2} - \frac{8}{r^2} + \frac{4\log e}{r^2}\right) \\
&\leq o(n) + 8\frac{\log r}{r^2}\cdot n.
\end{aligned}$$

If r^2 does not divide n or r is odd, the same proof works, we only have to be more careful with the integer parts.

Remark 2. A more careful computation would give a better constant instead of 8.

4. Final remarks

The most important thing would be to narrow the gap between the upper and lower bounds on $T(r,n)$, (see e.g. V.T. Sós and F. K. Hwang, [13]). Of course, the same question applies also to $T^\star(r,n)$, $T(r,n)$ and $T'(r,n)$. In [17] one can find the proof of the following theorem.

If $\mathcal{F}'$ is r-distinct then it is $(r-1)$-cover- free.

So by this theorem and Proposition 2.1 $T(r,n)$ and $T'(r,n)$ are very closed. On the other hand, if $k \geq \frac{n}{r}$, then by Johnson's second bound [16] $T^\star(r,n,k)$ is only polynomial in n. The proof of the lower bound suggests that it attains the maximum in k about $\frac{n}{r+1}$. By the *set compression algorithm* we suppose that $T(r,n,k)$ attains the maximum in k about $\frac{n}{r}$, too. This suggests that $T(r,n)$ and $T^\star(r,n)$ are neither significantly different.

We consider the following problem. The estimations in the proof of Theorem 3.1 are very loose. Is it possible to prove $R(r) = o(\frac{\log r}{r^2})$ upper bound using the set compression algorithm?

The author is very thankful to P. Frankl for pointing out the arguments of Remark 1.

5. References

[1] Nguyen Quang A and T. Zeisel, Bounds on constant weight binary superimposed codes, *Probl. of Control and Information Theory* 17 (1988), 223-230.

[2] N. Alon, Explicit constractions of exponential sized families of k-independent sets, *Discrete Mathematics* 58 (1986), 191-193.

[3] Zs. Baranyai, On the factorization of the complete uniform hypergraph, *Proc. Colloq. Math. Soc. János Bólyai* (10. Infinite and finite sets, Keszthely, Hungary (1973).

[4] A. G. Dyachkov and V.V. Rykov, Bounds on the length of disjunctive codes, *Problemy Peredachi Informatsii,* Vol. 18, No 3 (1982), 7-13.

[5] A. G. Dyachkov and V.V. Rykov, A survey of superimposed codes theory, *Probl. of Control and Information Theory,* Vol. 12, No 4 (1983), 1-13.

[6] P. Erdős, P. Frankl and Z. Füredi, Families of finite sets in which no set is covered by the union of two others, *Journal of Combinatorial Theory, Series A* Vol. 33, No. 2 (1982), 158-166.

[7] P. Erdős, P. Frankl and Z, Füredi, Families of finite sets in which no set is covered by the union of r others, *Israel J. of Math.* Vol. 51. Nos. 1-2 (1985), 79-89.

[8] P. Frankl, On Sperner Families Satisfying an Additional Condition,*Journal of Combinatorial Theory, Series A* Vol. 20, No. 1 (1976), 1-11

[9] P. Frankl and Z. Füredi, Colored packing of sets, *Annals of Discrete Mathematics,* Vol. 34, (1987), 165-178

[10] P. Frankl and V. Rödl, Near perfect coverings in graphs and hypergraphs, *Europ. J. Combinatorics* 6 (1985), 317-326.

[11] R. G. Gallager, Information Theory and Reliable Communication, Wiley (1968), problem 5.8

[12] F. K. Hwang, A method for detecting all defective members in a population by group testing, *J. of the American Statistical Association,* Vol. 67, No 339 (1972), 605-608.

[13] F. K. Hwang and V.T. Sós, Non adaptive hypergeometric group testing, *Studia Sc. Math. Hungarica,* 22 (1987), 257-263.

[14] S. M. Johnson, On the upper bounds for unrestricted binary error-correcting codes, *IEEE Trans. on Inf. Th.,* Vol. it- 17, No. 4 (1971), 466-478.

[15] S. M. Johnson, Improved asymptotic bounds for error- correcting codes, *IEEE Trans. on Inf. Th.,* Vol. it-9, No 4 (1963) 198-205.

[16] S. M. Johnson, A new upper bound for error-correcting codes, *IRE Trans. on Inf. Th.,* Vol. it-8 (1962), 203-207.

[17] W. H. Kautz and R.C. Singleton, Nonrandom binary superimposed codes, *IEEE Trans. on Inf. Th.,* Vol. it-10 (October 1964), 363-377.

[18] V. Rödl, On a packing and covering problem, *Europ. J. Combinatorics,* 5 (1985), 69-78.

[19] V.T. Sós, An additive problem in different structures, *Proc. of the Second Int.*

Conf. in Graph Theory, Combinatorics, Algorithms, and Applications, San Fr. Univ., California (July 1989), 486-510.