Cent. Eur. J. Math. • 6(1) • 2008 • 1-11
DOI: 10.2478/s11533-008-0001-4

VERSITA

Central European Journal of **Mathematics**

Codes that attain minimum distance in every possible direction

Research Article

Gyula O.H. Katona[1*], Attila Sali[1†], Klaus-Dieter Schewe[2‡]

1 Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest, P.O.B.127, H-1364, Hungary

2 Massey University, Information Science Research Centre, & Department of Information Systems, Private Bag 11 222, Palmerston North, New Zealand

Received 30 November 2006; accepted 5 September 2007

Abstract: The following problem motivated by investigation of databases is studied. Let $\mathcal{C}$ be a q-ary code of length n with the properties that $\mathcal{C}$ has minimum distance at least $n-k+1$, and for any set of $k-1$ coordinates there exist two codewords that agree exactly there. Let $f(q,k)$ be the maximum n for which such a code exists. $f(q,k)$ is bounded by linear functions of k and q, and the exact values for special k and q are determined.

MSC: 05D99, 94B65, 94B25, 68P15

Keywords: Error correcting codes • database • functional dependency • key dependency • extremal problems of codes

1. Introduction

Arguably, the most important database constraint is the collection of functional dependencies that instances of a relational schema satisfy, in particular the key dependencies. Let X, Y be two sets of attributes, a database instance *satisfies* the functional dependency $X \to Y$, if whenever two records agree in the attributes of X, then they also agree in the attributes of Y. If R denotes the whole set of attributes, then $K \subseteq R$ is a *key*, if the functional dependency $K \to R$ holds. In what follows, we use the terminology of the book [1], with the restriction that only single relational schema is considered. Then a database instance can be viewed as a matrix, whose columns correspond to the attributes and rows to the individual records. Then a functional dependency $X \to Y$ is satisfied if and only if there exist no two rows of the database matrix that agree in columns of X but differ in some column of Y.

It is interesting from the point of view of schema design that given a collection Σ of functional dependencies, what other dependencies hold in a database instance that satisfies Σ. A way of solving this problem is the construction of

* *E-mail: ohkatona@renyi.hu*
† *E-mail: sali@renyi.hu*
‡ *E-mail: k.d.schewe@massey.ac.nz*

an *Armstrong instance* for Σ, that is, a database instance that satisfies a functional dependency $X \to Y$ if and only if $\Sigma \models X \to Y$. Silva and Melkanoff [17] developed a design aid that for a collection of functional and multivalued dependencies as input presents an Armstrong instance for that set. The existence of Armstrong instance for a set of functional dependencies was proved by Armstrong [2] and Demetrovics [3]. Later, Fagin [10] gave a necessary and sufficient condition for general dependencies.
Further investigations concentrated on the minimum size of an Armstrong instance, since it is a good measure of the complexity of the collection of dependencies or system of minimal keys in question [4–9, 11].
All papers cited above assumed that the *domain* of each attribute is unbounded, countably infinite. However, in the study of *Higher Order Datamodel* [12, 14–16] the question of bounded domains arises naturally. In fact, if a minimal key system contains only *counter attributes*, then the possible number of tuples in an Armstrong instance is bounded from above. Another reason to consider bounded domains comes from real life databases. In many cases the domain of an attribute is a well defined finite set, for example in car rental, the class of cars can take values from the set {subcompact, compact, mid-size, full-size, SUV, sportscar, van}. Same kind of finiteness may occur in case of job assignments, schedules, etc.
It is natural to ask what can be said about Armstrong instances if attribute A_i has a domain of size ℓ_i. The main question investigated in this paper was introduced in [16]. Let $\mathcal{K}_n^k$ denote the collection of all k-subsets of an n-element attribute set R.

Definition 1.1.
Let $q > 1$ and $k > 1$ be given natural numbers. Let $f(q,k)$ be the maximum such n that there exists an Armstrong instance using at most q symbols, for $\mathcal{K}_n^k$ being the system of minimal keys.

It is clear that for a meaningful Armstrong instance we need at least two distinct symbols, so $q > 1$ is necessary. On the other hand the minimal Armstrong instance for $\mathcal{K}_n^1$ uses only two symbols for arbitrary n [6], hence $f(q,k)$ is well defined only for $k > 1$.

Definition 1.2.
Let $\mathcal{K}$ be a Sperner system of minimal keys.

$$\mathcal{K}^{-1} = \{A \subset \mathbf{R}: \ \nexists K \in \mathcal{K} \text{ such that } K \subseteq A \text{ and } A \text{ is maximal with respect to this}\}$$

is the collection of *maximal antikeys* corresponding to $\mathcal{K}$.

The following basic fact is known [6].

Proposition 1.1.
A *is an Armstrong instance for* $\mathcal{K}$ *if and only if the following two properties hold:*

*(**K**) there exist no two rows of* **A** *that agree in all positions for any* $K \in \mathcal{K}$ *and*

*(**A**) for every* $A \in \mathcal{K}^{-1}$ *there exist two rows of* **A** *that agree in all positions of* A.

It is helpful to view an Armstrong instance for $\mathcal{K}_n^k$ as minimal key system using q symbols as a q-ary code $\mathcal{C}$ of length n, where codewords are the tuples, or rows of the instance. Using $(\mathcal{K}_n^k)^{-1} = \mathcal{K}_n^{k-1}$ we obtain

- $\mathcal{C}$ has minimum distance at least $n-k+1$ by (**K**).
- For any set of $k-1$ coordinates there exist two codewords that agree exactly there by (**A**).

A $k-1$-set of coordinates can be considered as a direction, so in $\mathcal{C}$ the minimum distance is *attained in all directions*. We give lower bounds for $f(q,k)$ in Section 2, while upper bounds are presented in Section 3. In particular, it is shown in Section 3 that Definition 1.1 is meaningful.

2. Lower bounds

For lower bounds we need constructions, and we work greedily. That is pick a pair of codewords for a given $k-1$ subset of positions such that they agree exactly at that $k-1$ positions. Then rule out the balls of radii $n-k$ around the two codewords. If enough codewords are left, then we can pick a pair for the next $k-1$ subset of positions, etc. In order to complete the plan above we need the following lemma.

Lemma 2.1.
Let $\mathcal{Q}$ be the set of q^n q-ary codewords of length n, furthermore let K be a $k-1$-subset of coordinate positions. Assume that $n \geq k$. Then $\mathcal{Q}$ can be partitioned into q^{n-1} classes of size q each, that any two codewords of the same class agree exactly on the positions of K.

Proof. By induction on n. The first interesting case is $n = k$ is trivial, since fixing the $k-1$ coordinate positions there is one position left, and a partition class contains the q codewords with the given fixed value on the positions in K. Now, assume that the partition exists for codewords of length n, and consider a class of q codewords. Each one of them has to be extended by one coordinate that takes values $1, 2, \ldots, q$. We want to do so in order to form q new classes are formed of the codewords of length $n+1$. It requires that in each new class the "extension coordinates" are all distinct. This could best represented by a *bipartite graph* $G(A, B, E)$, where A is the set of q codewords to be extended and $B = \{1, 2, \ldots, q\}$ and E consists of all possible edges between A and B. Now, one good extension is a *complete matching* in this bipartite graph. It is an easy exercise that $G(A, B, E)$, that is a complete bipartite graph, can be partitioned into q complete matchings. This partition into matchings gives the q new partition classes of codewords of length $n+1$. □

Lemma 2.1 tells us that as long as we have more codewords than the number of partition classes, i.e., q^{n-1} available, then for any given $k-1$ subset of the coordinates we can find two codewords that agree in exactly those positions. This implies that the greedy construction works if

$$\left[\binom{n}{k-1} - 1\right]\left[2\sum_{i=0}^{n-k}\binom{n}{i}(q-1)^i - B\right] < q^n - q^{n-1} \tag{1}$$

where B is the intersection of the two balls of radii $n-k$ around two codewords of distance $n-k+1$. The left-hand side of (1) is the total volume of the balls ruled out up to the last but one $k-1$-tuple of coordinates.

$$B = \sum_{\substack{a+b\geq k\\ a+b'\geq k\\ b+b'\leq n-k+1}} \binom{k-1}{a}(q-1)^{k-1-a}\binom{n-k+1}{b}\binom{n-k+1-b}{b'}(q-2)^{n-k+1-b-b'} \tag{2}$$

The expression for B in (2) is quite complicated, so to obtain a simple lower bound for n we may use that if

$$\left[\binom{n}{k-1} - 1\right]\left[2\sum_{i=0}^{n-k}\binom{n}{i}(q-1)^i\right] < q^n - q^{n-1} \tag{3}$$

holds, then (1) holds, as well.

Theorem 2.1.
Given $q > 4$, there is k_0 such that for every $k > k_0$ and for every $n < \frac{1}{2}k\log q$ there exists an Armstrong instance for $\mathcal{K}_n^k$ as minimal key system using at most q symbols.

Proof. The binomial coefficients in (3) can be bounded from above by the middle one and that in turn by $\frac{2^n}{\sqrt{n}}$, while the powers of $(q-1)$ can be replaced with the highest one. Write $n = \alpha k$ to obtain

$$\frac{2^{\alpha k}}{\sqrt{\alpha k}}\left(2[(\alpha-1)k]\frac{2^{\alpha k}}{\sqrt{\alpha k}}(q-1)^{(\alpha-1)k}\right) < q^{(\alpha k-1)}(q-1) \tag{4}$$

that implies (3). (4) is equivalent to

$$2^{2\alpha k}\cdot 2\left(\frac{\alpha-1}{\alpha}\right) < q^k\cdot\left(\frac{q}{q-1}\right)^{(\alpha-1)k-1} \tag{5}$$

(5) is a consequence of

$$2^{2\alpha k} < q^k \tag{6}$$

and

$$2\left(1-\frac{1}{\alpha}\right) < \left(1+\frac{1}{q-1}\right)^{(\alpha-1)k-1}. \tag{7}$$

Finally, (6) and (7) hold for $k > k_0$ if $1 < \alpha < \frac{1}{2}\log q$. □

Consider the case $q \leq 4$. It is shown by the $(k+1)\times(k+1)$ identity matrix that $f(2,k) \geq k+1$. However, we can get some better estimates. (3) can be written as

$$\left[\binom{n}{k-1}-1\right]\left[2\sum_{i=0}^{n-k}\binom{n}{i}\right] < 2^{n-1}. \tag{8}$$

Now, if $n < 2k$, then the left hand side of (8) can be bounded by $2\binom{n}{k-1}^2(n-k+1)$. That is if

$$2\binom{n}{k-1}^2(n-k+1) \leq 2^{n-1}, \tag{9}$$

then (8) holds. The binomial coefficient in (9) can be estimated [13] using the *entropy function* $H(x) = -x\log x - (1-x)\log(1-x)$ as $\log\binom{n}{k-1} = nH(\frac{k-1}{n}) + O(\log n)$. Taking the natural logarithm of both sides of (9) and substituting the previous estimate for $\log\binom{n}{k-1}$

$$2nH(\frac{k-1}{n}) + O(\log n) \leq (n-1)\log 2 \tag{10}$$

is obtained. Putting $n = c(k-1)$, (10) holds for $k > k_0$ if $H(\frac{1}{c}) < \frac{\log 2}{2}$. Thus, we have proved

Theorem 2.2.

There exists k_0 and $c > 1$ constants, that for $k > k_0$, and $n = \lfloor ck \rfloor$, there exists a binary code of length n of minimum distance $n-k+1$ that the minimum distance is attained in all possible directions.

The main point of Theorem 2.2 is that constant c is *strictly* larger than one. Theorem 2.2 gives lower bound for $f(3,k)$ and $f(4,k)$, as well, in an obvious way. To conclude this section we show a lower bound that turns out to be sharp.

Proposition 2.1.

$$f(q,2) \geq \binom{q+1}{2} \tag{11}$$

Proof. The lower bound is given by construction. Relation R has $q+1$ rows and $\binom{q+1}{2}$ columns (attributes). Since q symbols are allowed in each column we need to have exactly one pair of equal symbols and we do so that these pairs are all distinct. Since each column has a pair of rows that agree there, the minimum distance is at least $n-1$. On the other hand, no two rows agree in two positions, so the minimum distance is exactly $n-1$, and it is attained for every $n-1$ coordinate positions. Hence $f(q,2) \geq \binom{q+1}{2}$. □

3. Upper bounds

Here, we assume that a q-ary code of length n, minimum distance $n-k+1$ and of m codewords exists such that the minimum distance is attained in all possible directions. Also, we may assume without loss of generality that $q < m$.

Lemma 3.1.
If s is a sequence of length m containing elements from the set $\{1,2,\ldots,q\}$, then the number of equal pairs in s is at least

$$\frac{m}{2}\left(\frac{m}{q}-1\right) \tag{12}$$

Proof. Extend the concept of the binomial coefficient $\binom{m}{2}$ for real values x as $\binom{x}{2} = \frac{x(x-1)}{2}$. Since $f(x) = \frac{x(x-1)}{2}$ is a convex function (from below), the Jensen inequality

$$f\left(\frac{\sum_{i=1}^{q} x_i}{q}\right) \le \frac{\sum_{i=1}^{q} f(x_i)}{q} \tag{13}$$

implies

$$\binom{\frac{\sum_{i=1}^{q} x_i}{q}}{2} \le \frac{\sum_{i=1}^{q}\binom{x_i}{2}}{q}. \tag{14}$$

Let m_i denote the number of digits i in s. Then the number of equal pairs is

$$\sum_{i=1}^{q}\binom{m_i}{2}. \tag{15}$$

On the other hand, $\sum_{i=1}^{q} m_i = m$. Substituting these into (14), we obtain

$$\binom{\frac{m}{q}}{2} \le \frac{\sum_{i=1}^{q}\binom{m_i}{2}}{q} \tag{16}$$

and the desired (12). □

The following slight improvement of Lemma 3.1 will also be used. It is really an immediate corollary to Turán's Theorem [18], but we include a proof here for the sake of completeness.

Lemma 3.2.
If s is a sequence of length m containing elements from the set $\{1,2,\ldots,q\}$, where $q < m$, then the number of equal pairs in s is at least

$$q\binom{h}{2} + rh, \tag{17}$$

where $m = qh + r$ with $(0 \le r < q)$.

Proof. Let m_i be the number of digits i in s. Suppose that $m_1 < m_2 - 1$. Replace m_1 by $m_1' = m_1 + 1$ and m_2 by $m_2' = m_1 - 1$. This change does not change the sum of ms. The sum of the equal entries before the change is

$$\sum_{i=1}^{q}\binom{m_i}{2} \tag{18}$$

and it is

$$\binom{m_1+1}{2}+\binom{m_2-1}{2}+\sum_{i=3}^{q}\binom{m_i}{2} \tag{19}$$

after the changes. It is easy to see that the number of equal entries was decreased by $m_2-1-m_1>0$. The same can be said in every case when the difference of any two m_i's is at least 2. Otherwise, when the differences are 0 and 1 then r of them are $h+1$, on the other hand $q-r$ of them are h. The number of equal digits is

$$r\binom{h+1}{2}+(q-r)\binom{h}{2}=q\binom{h}{2}+rh. \tag{20}$$

□

Now, we are able to prove a general upper bound.

Theorem 3.1.
Let $q>1$ and $k>2$ be integers. Then

$$f(q,k)\leq q(k-1)\left(1+\frac{q-1}{\sqrt{\frac{2(qk-q-k+2)^{k-1}}{(k-1)!}}-q}\right) \tag{21}$$

holds.

Proof. Assume that a q-ary code of length n, minimum distance $n-k+1$ and of m codewords exists such that the minimum distance is attained in all possible directions and consider it as an $m\times n$ matrix. The number of pairs of equal entries in each column is at least (12). Altogether:

$$n\frac{m}{2}\left(\frac{m}{q}-1\right). \tag{22}$$

In one pair of rows at most $k-1$ of them can appear, otherwise, the two rows had k equal entries in contradiction with the assumption about the minimum distance. Hence, we have the inequality

$$n\frac{m}{2}\left(\frac{m}{q}-1\right)\leq(k-1)\binom{m}{2}. \tag{23}$$

For any choice of $k-1$ columns there must exist two (distinct) rows such that they have equal entries in these columns. It is easy to see that this pair of rows must be different for different choices of $k-1$ columns, otherwise the pair of rows would agree on the union of these two $k-1$-element sets, in contradiction with our assumptions on the minimum distance. Hence, the following inequality is obtained:

$$\binom{n}{k-1}\leq\binom{m}{2}. \tag{24}$$

(23) can be easily rewritten as an upper bound on n for fixed m:

$$n\leq\frac{(k-1)\binom{m}{2}}{\frac{m}{2}\left(\frac{m}{q}-1\right)}=q(k-1)\left(1+\frac{q-1}{m-q}\right)=a_{q,k}(m). \tag{25}$$

(24) gives another, but implicit upper bound, a somewhat weaker explicit bound will be formed. (24) implies

$$\frac{(n-k+2)^{k-1}}{(k-1)!} \leq \frac{m^2}{2}, \tag{26}$$

hence we obtain

$$n \leq \left(\frac{(k-1)!}{2}m^2\right)^{\frac{1}{k-1}} + k - 2 = b_{q,k}(m). \tag{27}$$

Notice that $a_{q,k}(m)$ is a decreasing, while $b_{q,k}(m)$ is an increasing function of m. Therefore, if α is the solution of the equation

$$a_{q,k}(m) = b_{q,k}(m) \tag{28}$$

in m then $a_{q,k}(\alpha) = b_{q,k}(\alpha)$ is a universal (independent of m) upper bound for n. Such a solution must exist if $a_{q,k}(q+1) \geq b_{q,k}(q+1)$ holds (at the smallest value where $a_{q,k}(m)$ is defined). That is, we have to show

$$q(k-1)\left(1 + \frac{q-1}{q+1-q}\right) \geq \left(\frac{(k-1)!}{2}(q+1)^2\right)^{\frac{1}{k-1}} + k - 2, \tag{29}$$

or equivalently

$$q^2(k-1) \geq \left(\frac{(k-1)!}{2}(q+1)^2\right)^{\frac{1}{k-1}} + k - 2. \tag{30}$$

Inequality (30) reduces to following in case of $q = 2$

$$4(k-1) \geq \left(\frac{(k-1)!}{2}9\right)^{\frac{1}{k-1}} + k - 2 \tag{31}$$

Using the inequality between geometric and arithmetic means (31) follows from

$$3k - 2 \geq \left(\frac{9}{2}\right)^{\frac{1}{k-1}} \frac{k}{2} \tag{32}$$

that holds trivially for $k > 2$. Considering the difference of the left and right hand sides of (30) for fixed k as a function of q one can realize that it is a monotone increasing function by observing that the derivative with respect of q is non-negative.
Since it is difficult to find the explicit solution of equation (28), we solve the easier equation

$$q(k-1) = \left(\frac{(k-1)!}{2}m^2\right)^{\frac{1}{k-1}} + k - 2, \tag{33}$$

replacing $a_{q,k}$ by a smaller function (what is actually a constant). Its solution β satisfies $\beta \leq \alpha$ by the monotonity of $b_{q,k}(m)$. We have

$$\beta = \sqrt{\frac{2(qk-q-k+2)^{k-1}}{(k-1)!}}. \tag{34}$$

Let us see that $q + 1 \leq \beta$ if $2 < k$. What we need is

$$(k-1)!(q+1)^2 \leq 2(qk-q-k+2)^{k-1} \tag{35}$$

or equivalently,

$$\left(\frac{(k-1)!}{2}\right)^{\frac{1}{k-1}}(q+1)^{\frac{2}{k-1}} \leq qk-q-k+2. \tag{36}$$

(36) holds with equality for $q=2$ and $k=3$. Keeping $q=2$, easy induction on k shows that (36) holds for $q=2$, $k\geq 3$. Now, fixing k, we find that the difference of the right hand side and the left hand side of (36) is a monotone increasing function of q, since its derivative with respect to q

$$k-1-\frac{2}{k-1}(q+1)^{\frac{2}{k-1}-1}\left(\frac{(k-1)!}{2}\right)^{\frac{1}{k-1}} \tag{37}$$

is nonnegative.
Then $a_{q,k}(\beta)$ is defined and is a universal upper bound on n and this is actually the bound formulated in the theorem. □

In the case $k=2$, our theorem is not valid. However, the method works. We only have to use somewhat better estimates rather than $a_{q,2}(m)$ and $b_{q,2}(m)$. These improved bounds lead to a better estimate for $k=3$, too. If Lemma 3.1 is replaced by Lemma 3.2 then

$$n \leq a^*_{q,k}(m) = \frac{(k-1)\binom{m}{2}}{q\binom{h}{2}+rh} \tag{38}$$

is obtained instead of (25). To be able to use this bound, we have to see that it is a decreasing function of m.

Lemma 3.3.

$$a^*_{q,k}(m) = \frac{(k-1)\binom{m}{2}}{q\binom{h}{2}+rh} \tag{39}$$

is decreasing in m for $q<m$.

Proof. Here $m=qh+r, 0\leq r<q$ implies $m+1=qh+r+1$, we have to verify

$$\frac{(k-1)\binom{m}{2}}{q\binom{h}{2}+rh} \geq \frac{(k-1)\binom{m+1}{2}}{q\binom{h}{2}+(r+1)h} \tag{40}$$

when $r+1<q$. This leads to the following inequality after carrying out the obvious cancelations.

$$(m-2r-1)h \geq 2q\binom{h}{2} = qh^2-qh \tag{41}$$

or equivalently: $qh-r-1\geq qh-q$ what is trivially true.
The above proof does not work perfectly when $r=q-1$ since then $m+1$ is obtained in the form $qh+q$. However, as it is easy to see, the formula for the minimum number of equal digits works in this case, too:

$$q\binom{h}{2}+qh = q\binom{h+1}{2}. \tag{42}$$

□

In the cases $k = 2, 3$, formula (24) has a nice form, there is no need to rewrite in the weaker form of (27). When $k = 2$, it is simply

$$n \le b^*_{q,2}(m) = \binom{m}{2}. \tag{43}$$

The solution of

$$a^*_{q,2}(m) = b^*_{q,2}(m) \tag{44}$$

that is of

$$\frac{\binom{m}{2}}{q\binom{h}{2} + rh} = \binom{m}{2} \tag{45}$$

is simply $q + 1$ as it can be seen by substitution since here $h = 1, r = 1$. The universal bound on n is $\binom{q+1}{2}$, which is sharp by Proposition 2.1
If $k = 3$, then (24) reduces to $n \le m$, that is $b^*_{q,3}(m) = m$. By (38), we have to solve the equation

$$a^*_{q,3}(m) = \frac{2\binom{m}{2}}{q\binom{h}{2} + rh} = m = b^*_{q,3}(m). \tag{46}$$

The solution is $3q - 1$. Indeed, $h = 2, r = q - 1$ implies $q\binom{h}{2} + rh = q\binom{2}{2} + (q-1)2 = 3q - 2$. The left hand side of (46) is really $3q - 1$.

Theorem 3.2.
If $k = 2$, then $f(q, k) \le \binom{q+1}{2}$, if $k = 3$ then $n \le 3q - 1$.

One can feel that if k and/or q are large, then the remainder term in Theorem 3.1 is less than 1, that is, the main term is the upper bound. Indeed, we can prove the following theorem.

Theorem 3.3.
If $5 \le k$ and $2 \le q$ then the upper bound in Theorem 3.1 can be improved to

$$f(q, k) \le q(k - 1) \tag{47}$$

with the following exceptions: $(k, q) = (5, 2), (5, 3), (5, 4), (5, 5), (6, 2)$.

Proof. We have to prove

$$\frac{q(q-1)(k-1)}{\sqrt{\frac{2(qk-q-k+2)^{k-1}}{(k-1)!}} - q} < 1 \tag{48}$$

for the desired cases. More precisely, it is sufficient to prove $\le$ since the denominator is a result of a non-sharp estimation. The inequality is equivalent to

$$q(q-1)(k-1) + q = q(qk - q - k + 2) \le \sqrt{\frac{2(qk - q - k + 2)^{k-1}}{(k-1)!}}, \tag{49}$$

that is

$$q^2(k-1)! \le 2(qk - q - k + 2)^{k-3}. \tag{50}$$

Replacing $qk - q - k + 2$ by $(q-1)(k-1)$ a somewhat stronger inequality is obtained:

$$q^2(k-1)! \le 2(q-1)^{k-3}(k-1)^{k-3}. \tag{51}$$

Suppose that $5 \leq k$. Then $(k-1)! \leq 2(k-1)^{k-3}$ (induction), (51) reduces to

$$q^2 \leq (q-1)^{k-3}. \tag{52}$$

Our strategy is to prove the statement of (52) or (51). If they are not true for some values then we go back to the original (50).
If $k = 6$, then (52) becomes $q^2 \leq (q-1)^3$. Analyzing this equation, one can see that it holds for $4 \leq q$. Since the right hand side of (52) is increasing in k, our statement is proved for the values $6 \leq k, 4 \leq q$.
Consider now the case $k = 5$. Then (51) has the form $q^2 24 \leq 2(q-1)4^2$. Solving the quadratic equation, we obtain that it holds for $8 \leq q$. The smaller values of q, namely $q = 4, 5, 6, 7$ can be checked for (50). It holds for $q = 6, 7$ but not for $q = 4, 5$.
The remaining cases are $q = 2, 3$ for all k. Fix first $q = 2$ and find the smallest k satisfying (51) for this case:

$$4(k-1)! \leq 2(k-1)^{k-3}. \tag{53}$$

It holds with $k = 8$, therefore (51) is true for $q = 2,\ 9 \leq k$ as it can be seen by easy induction. The smaller cases of k can be checked in the original (50). It holds for $k = 7$ and does not hold for $k = 5, 6$.
Let now $q = 3$. (51) reduces to $9(k-1)! \leq 2^{k-2}(k-1)^{k-3}$. It holds with $k = 6$, the larger values of k can be obtained by induction. The case $(q = 3,)k = 5$ is not true in (50). □

4. Conclusions

We have proved general lower and upper bounds on $f(q, k)$. However, there is a significant gap in them. It is pretty reasonable to assume that the lower bound can be improved, since the counting in the greedy construction seems to be wasteful. One expects the balls around different pairs of codewords not being disjoint. However, in the binary case one can prove that for small k, such as $k = 3, 4, 5$, the best construction is the $k+1 \times k+1$ identity. In particular, neither the Fano plane nor any of its extensions work. This makes it hard to find a "pattern" to generalize. On the other hand, the upper bound might also be improved, since having equality in (24) requires very strong structure.

References

[1] Abiteboul S., Hull R., Vianu V., Foundations of databases, Addison-Wesley, 1995
[2] Armstrong W.W., Dependency structures of data base relationships, Information Processing, 1974, 74, 580–583
[3] Demetrovics J., On the equivalence of candidate keys with Sperner systems, Acta Cybernet., 1978/79, 4, 247–252
[4] Demetrovics J., Füredi Z., Katona G.O.H., Minimum matrix representation of closure operations, Discrete Appl. Math., 1985, 11, 115–128
[5] Demetrovics J., Gyepesi G., A note on minimal matrix representation of closure operations, Combinatorica, 1983, 3, 177–179
[6] Demetrovics J., Katona G.O.H., Extremal combinatorial problems in relational data base, In: Fundamentals of computation theory, Springer-Verlag, Berlin-New York, 1981
[7] Demetrovics J., Katona G.O.H., Sali A., The characterization of branching dependencies, Discrete Appl. Math., 1992, 40, 139–153
[8] Demetrovics J., Katona G.O.H., Sali A., Design type problems motivated by database theory, J. Statist. Plann. Inference, 1998, 72, 149–164
[9] Demetrovics J., Katona G.O.H., A survey of some combinatorial results concerning functional dependencies in database relations, Ann. Math. Artificial Intelligence, 1993, 7, 63–82
[10] Fagin R., Horn clauses and database dependencies, J. Assoc. Comput. Mach., 1982, 29, 952–985
[11] Füredi Z., Perfect error-correcting databases, Discrete Appl. Math., 1990, 28, 171–176

[12] Hartmann S., Link S., Schewe K.-D., Weak functional dependencies in higher-order datamodels, In: Seipel D., Turull Torres J. M. (Eds.), Foundations of Information and Knowledge Systems, Springer LNCS, 2942, Springer Verlag, 2004
[13] Odlyzko A.M., Asymptotic enumeration methods, In: Graham R.L., Grötschel M., Lovász L. (Eds.), Handbook of combinatorics, Elsevier, Amsterdam, 1995
[14] Sali A., Minimal keys in higher-order datamodels, In: Seipel D., Turull Torres J. M. (Eds.), Foundations of Information and Knowledge Systems, Springer LNCS, 2942, Springer Verlag, 2004
[15] Sali A., Schewe K.-D., Counter-free keys and functional dependencies in higher-order datamodels, Fund. Inform., 2006, 70, 277–301
[16] Sali A., Schewe K.-D., Keys and Armstrong databases in trees with restructuring, Acta Cybernet., preprint
[17] Silva A.M., Melkanoff M.A., A method for helping discover the dependencies of a relation, In: Gallaire H., Minker J., Nicolas J.-M. (Eds.), Advances in Data Base Theory, Plenum Publishing, New York, 1981
[18] Turán P., On an extremal problem in graph theory, Mat. Fiz. Lapok, 1941, 48, 436–452 (in Hungarian)