On Codes for Partition Distance: Part I – Constructions and Bounds

A. D'yachkov [1], V. Rykov [2], D. Torney [3], S. Yekhanin [4]

Abstract

We discuss the distance concept between two q-ary n-sequences, $2 \leq q < n$, called a partition distance. This distance is a metric in the space of partitions of a finite n-set, where each partition contains $\leq q$ disjoint subsets of the n-set. For the metric, we study codes called q-partition codes and present a construction of these codes based on the first order Reed - Muller codes. A random coding bound is obtained.

1 Notation and Definitions

The symbol $\triangleq$ denotes definitional equalities. Let $n > q \geq 2$ be fixed integers, $A_q \triangleq \{0, \ldots, q-1\}$ be q-ary alphabet, $\mathcal{M}_q = \{\mu\}$, $\mu = \mu(x)$, be the set containing all $q!$ permutations of A_q. We denote elements of $\mathcal{M}_q$ by Greek letters: α, β, τ, ...

Let $[n] \triangleq \{1, 2, \ldots, n\}$ be the set of integers from 1 to n and $\mathbf{x} \triangleq (x_1, x_2, \ldots, x_n)$, $x_i \in A_q$, be an arbitrary fixed q-ary n-sequence. $\mathbf{x}$ identifies a *q-partition* $\{E_0, E_1, \ldots, E_{q-1}\}$ of the set

$$[n] = E_0 + E_1 + \cdots + E_{q-1}, \quad \text{where} \quad E_x = E_x(\mathbf{x}) \triangleq \{i : x_i = x\}, \quad x \in A_q.$$

Any q-partition contains q', $1 \leq q' \leq q$, nonempty parts. For any $\mu \in \mathcal{M}_q$, we define q-ary n-sequence $\mathbf{x}^\mu \triangleq (\mu(x_1), \mu(x_2), \ldots, \mu(x_n))$, called *$\mu$-complement of* $\mathbf{x}$. For the given n-sequence $\mathbf{x}$, all μ-complements $\mathbf{x}^\mu$, $\mu \in \mathcal{M}_q$, identify the same *q-partition.*

We denote this q-partition by symbol $\widetilde{\mathbf{x}} \triangleq \{E_0, E_1, \ldots, E_{q-1}\}$.

For arbitrary q-ary n-sequences $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $x_i \in A_q$, and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, $y_i \in A_q$, and arbitrary fixed elements $x, y \in A_q$ let $n(x, \mathbf{x}; y, \mathbf{y})$ denote the number of positions i where $\mathbf{x}_i = x$ and $\mathbf{y}_i = y$. Let

$$S(\mathbf{x}, \mathbf{y}) \triangleq \sum_{x \in A_q} n(x, \mathbf{x}; x, \mathbf{y}), \quad H(\mathbf{x}, \mathbf{y}) \triangleq n - S(\mathbf{x}, \mathbf{y}) = n - \sum_{x \in A_q} n(x, \mathbf{x}; x, \mathbf{y})$$

denote the *Hamming similarity* and *distance* between $\mathbf{x}$ and $\mathbf{y}$.

Properties of $H(\mathbf{x}, \mathbf{y})$ connected with the complement operation are given in

Proposition 1.

1. *For any $\alpha \in \mathcal{M}_q$ and $\beta \in \mathcal{M}_q$, the minimum*

$$\min_{\mu \in \mathcal{M}_q} H(\mathbf{x}^\alpha, \mathbf{y}^\mu) = \min_{\mu \in \mathcal{M}_q} H(\mathbf{x}^\mu, \mathbf{y}^\beta) = \min_{\mu \in \mathcal{M}_q} H(\mathbf{x}, \mathbf{y}^\mu) =$$

$$= n - \max_{\mu \in \mathcal{M}_q} \left\{ \sum_{x \in A_q} n(x, \mathbf{x}; \mu(x), \mathbf{y}) \right\} \triangleq \mathcal{P}_q(\mathbf{x}, \mathbf{y}) = \mathcal{P}_q(\mathbf{y}, \mathbf{x})$$

and the number $\mathcal{P}_q(\mathbf{x}, \mathbf{y}) \geq 0$. In addition, $\mathcal{P}_q(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x}$ and $\mathbf{y}$ identify the same unordered q-partition of the set $[n]$.

[1] Department of Probability Theory, Faculty of Mechanics and Mathematics, Moscow State University, Moscow 119992, Russia. E-mail: dyachkov@mech.math.msu.su. The work of A. D'yachkov was supported by the Russian Foundation of Basic Research, Grant 01-01-00495, and INTAS-00-738.

[2] Department of Mathematics, University of Nebraska at Omaha, 6001 Dodge St., Omaha, NE 68182-0243. E-mail: vrykov@mail.unomaha.edu

[3] MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545. E-mail: dct@lanl.gov

[4] Computer Science and Artificial Intelligence Laboratory, MIT 200 Technology Square, Cambridge, MA 02139. E-mail: yekhanin@mit.edu

2.

$$\mathcal{P}_q(\mathbf{x},\mathbf{y}) \;\le\; \frac{q-1}{q}\cdot n,$$

where, equality is achieved, for example, if $n = qk\cdot q$, $k = q, 2q, \ldots$ *and*

$$\mathbf{x} = (0,0,\ldots,0,\ldots,q-1,\ldots,q-1), \quad \mathbf{y} = (0,1\ldots,q-1,\ldots,0,\ldots,q-1).$$

Proof. 1) The proof is based on the identity $H(\mathbf{x}^{\beta}\,,\,\mathbf{y}^{\alpha}) = H(\mathbf{x},\,\mathbf{y}^{\beta^{-1}\alpha})$.

2) Consider q mappings $\mu_i = \mu_i(x)$, $x \in A_q$, $\mu_i \in \mathcal{M}_q$, $i = 0,1,2,\ldots,q-1$, having the form:

$$\mu_0(x) \triangleq x, \qquad \mu_i(x) \triangleq \mu_0\left(x + i \pmod q\right), \quad i = 1,2,\ldots,q-1.$$

It is easy to understand that for any pair of q-ary n-sequences $(\mathbf{x},\mathbf{y})$, the sum

$$\sum_{i=0}^{q-1}\sum_{x\in A_q} n\left(x,\mathbf{x};\,\mu_i(x),\mathbf{y}\right) \;=\; \sum_{x,y\in A_q} n\left(x,\mathbf{x};\,y,\mathbf{y}\right) \;=\; n.$$

Therefore, there exists an integer i such that the internal sum in the left-hand side is at least n/q. Taking into account the definition of $\mathcal{P}_q(\mathbf{x},\mathbf{y})$, we obtain the second statement.

Proposition 1 motivates

Definition 1.

$$\mathcal{P}_q(\widetilde{\mathbf{x}}\,,\,\widetilde{\mathbf{y}}) \;=\; \mathcal{P}_q(\mathbf{x}\,,\,\mathbf{y}) \triangleq \min_{\mu\in\mathcal{M}_q} H(\mathbf{x}\,,\,\mathbf{y}^{\mu}) = \min_{\mu\in\mathcal{M}_q} H(\mathbf{x}^{\mu}\,,\,\mathbf{y}) = \mathcal{P}_q(\mathbf{y},\mathbf{x}) \;=\; \mathcal{P}_q(\widetilde{\mathbf{y}}\,,\,\widetilde{\mathbf{x}})$$

is called a *partition distance* between q-ary n-sequences $\mathbf{x}$ and $\mathbf{y}$ or between unordered q-partitions $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ of the set $[n]$.

Remark 1. The distance $\mathcal{P}_q(\widetilde{\mathbf{x}}\,,\,\widetilde{\mathbf{y}})$ between q-partitions $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ of the set $[n]$ is equal to the minimum number of elements that must be deleted from $[n]$, so that two induced q-partitions ($\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ restricted to the remaining elements) are identical. The partition distance $\mathcal{P}_q(\widetilde{\mathbf{x}}\,,\,\widetilde{\mathbf{y}})$ is also equal to the minimum number of elements of $[n]$ that must be moved between subsets E_0, E_1, ..., E_{q-1} of q-partition $\widetilde{\mathbf{x}} = \{E_0, E_1, \ldots, E_{q-1}\}$, so that the resulting q-partition equals $\widetilde{\mathbf{y}}$. Such form of the partition distance definition was suggested in [1, 2].

Proposition 2. $\mathcal{P}_q(\mathbf{x}\,,\,\mathbf{y})$ *satisfies the triangle inequality*

$$\mathcal{P}_q(\mathbf{x}\,,\,\mathbf{y}) \;\le\; \mathcal{P}_q(\mathbf{x}\,,\,\mathbf{z}) + \mathcal{P}_q(\mathbf{z}\,,\,\mathbf{y}) \quad \textit{or} \quad \mathcal{P}_q(\widetilde{\mathbf{x}}\,,\,\widetilde{\mathbf{y}}) \;\le\; \mathcal{P}_q(\widetilde{\mathbf{x}}\,,\,\widetilde{\mathbf{z}}) + \mathcal{P}_q(\widetilde{\mathbf{z}}\,,\,\widetilde{\mathbf{y}}).$$

Hence, $\mathcal{P}_q(\widetilde{\mathbf{x}}\,,\,\widetilde{\mathbf{y}})$ *is a metric in the space of unordered* q*-partitions.*

Proof. By Definition 1, we have

$$\mathcal{P}_q(\mathbf{x}\,,\,\mathbf{z}) + \mathcal{P}_q(\mathbf{z}\,,\,\mathbf{y}) = \min_{\mu\in\mathcal{M}_q} H(\mathbf{x}\,,\,\mathbf{z}^{\mu}) + \min_{\mu\in\mathcal{M}_q} H(\mathbf{z}\,,\,\mathbf{y}^{\mu}).$$

Consider a mapping α for which $\min\limits_{\mu\in\mathcal{M}_q} H(\mathbf{x}\,,\,\mathbf{z}^{\mu}) \triangleq H(\mathbf{x}\,,\,\mathbf{z}^{\alpha})$. In virtue of Proposition 1,

$$\min_{\mu\in\mathcal{M}_q} H(\mathbf{z}\,,\,\mathbf{y}^{\mu}) = \min_{\mu\in\mathcal{M}_q} H(\mathbf{z}^{\alpha}\,,\,\mathbf{y}^{\mu}).$$

Let for the given α, $\min\limits_{\mu\in\mathcal{M}_q} H(\mathbf{z}^{\alpha}\,,\,\mathbf{y}^{\mu}) \triangleq H(\mathbf{z}^{\alpha}\,,\,\mathbf{y}^{\tau})$. We can write

$$\mathcal{P}_q(\mathbf{x}\,,\,\mathbf{z}) + \mathcal{P}_q(\mathbf{z}\,,\,\mathbf{y}) = H(\mathbf{x}\,,\,\mathbf{z}^{\alpha}) + H(\mathbf{z}^{\alpha}\,,\,\mathbf{y}^{\tau}) \ge H(\mathbf{x},\mathbf{y}^{\tau}) \ge \mathcal{P}_q(\mathbf{x}\,,\,\mathbf{y}),$$

where we apply the triangle inequality for Hamming distance and Definition 1.

Proposition 2 is proved.

Remark 2. A different definition of distance $\Delta(\widetilde{\mathbf{x}},\, \widetilde{\mathbf{y}})$ in the space of finite set partitions was suggested in [3], where $\Delta(\widetilde{\mathbf{x}},\, \widetilde{\mathbf{y}})$ is defined as one-half of the Hamming distance between *incidence matrices* of partitions $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$, namely:

$$\Delta(\widetilde{\mathbf{x}},\, \widetilde{\mathbf{y}}) \triangleq \sum_{1\le i<j\le n} |\delta_{ij}(\widetilde{\mathbf{x}}) - \delta_{ij}(\widetilde{\mathbf{y}})| \quad \text{and} \quad \delta_{ij}(\widetilde{\mathbf{x}}) \triangleq \begin{cases} 1, & \text{if } x_i = x_j, \\ 0, & \text{if } x_i \ne x_j. \end{cases}$$

Obviously, $\Delta(\widetilde{\mathbf{x}},\, \widetilde{\mathbf{y}})$ satisfies the triangle inequality.

2 Codes for Partition Distance

We will say that q-ary $(n \times N)$- matrix $X = \|x_i(j)\|$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, N$, $x_i(j) \in A_q$, is a q-ary *code of length n and size N*. Columns $\mathbf{x}(j) \triangleq (x_1(j), x_2(j), \ldots, x_n(j))$, $j = 1, 2, \ldots, N$, of X are called *codewords* and we will write $X \triangleq \{\mathbf{x}(j),\ j = 1, 2, \ldots, N\}$. In what follows, we will also interpret code X as a collection of the corresponding q-partitions, i.e.,

$$X \triangleq \{\widetilde{\mathbf{x}}(j),\ j = 1, 2, \ldots, N\}, \quad \text{where} \quad \widetilde{\mathbf{x}}(j) \triangleq \{E_0^j,\, E_1^j,\, \ldots,\, E_{q-1}^j\},\ E_x^j \triangleq \{i :\, x_i(j) = x\}.$$

Using an analogy with error-correcting [4] codes, we give the following

Definition 2. X is called an *q-partition code for the set* $[n]$ and $D(X) \triangleq \min\limits_{j\ne j'} \mathcal{P}_q(\mathbf{x}(j), \mathbf{x}(j'))$ is called a *distance* of X. From Proposition 1 it follows $D(X) \le \lfloor (q-1)n/q \rfloor$.

Below, we present a construction of q-partition codes for the set $[n]$ having the maximal possible distance $D = \left\lfloor \frac{q-1}{q} \cdot n \right\rfloor$. Let q be a prime or prime power and the q-ary alphabet A_q be interpreted as the field F_q with addition $(\oplus)$ and multiplication $(\cdot)$. The construction is based on the first order Reed-Muller code [4].

Proposition 3. *For $m = 1, 2, 3, \ldots$, there exists a family of q-partition codes X for the set $[n]$ with parameters: $n = q^m$, code size $N = \frac{q^m-1}{q-1} + 1$ and distance $D(X) = (q-1)q^{m-1} = n \cdot \frac{q-1}{q}$.*

Proof. Given a fixed integer $m = 1, 2, 3, \ldots$ we will use the following definition [4] of the first order Reed-Muller code C:

- messages for code C of size $|C| = q^{m+1}$ are q-ary sequences of length $m+1$ having the form $(\mathbf{a}; b) \triangleq (a_1, a_2, \ldots, a_m; b)$, where $a_i, b \in F_q$;
- each message represents a linear function of m variables $\mathbf{z} = (z_1, z_2, \ldots, z_m) \in F_q^m$

$$(\mathbf{a}, \mathbf{z}) \oplus b \triangleq \bigoplus_{i=1}^{m} (a_i \cdot z_i) \ \oplus\ b;$$

- codewords for messages $(\mathbf{a}; b)$ are q-ary sequences $\langle (\mathbf{a}, \mathbf{z}) \oplus b\,,\ \mathbf{z} \in F_q^m \rangle$ of length $n = q^m$, i.e., by definition, the first order Reed-Muller code C can be represented as follows [1]

$$C \triangleq \{\langle (\mathbf{a}, \mathbf{z}) \ \oplus b,\ \mathbf{z} \in F_q^m \rangle \mid (\mathbf{a}; b) \in F_q^{m+1}\}.$$

[1] It is easy to prove [4] that the code C is a q-ary linear (n, k)-code, $k = m+1$, of length $n = q^m$, size $|C| = q^{m+1}$ and the Hamming distance $d = (q-1) \cdot q^{m-1}$.

Let $\mathcal{A}$ denote the maximal set of vectors $\mathbf{a} \in F_q^m$ such that for any $b \in F_q$, $b \neq 0$, and any pair $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{A}$, where $\mathbf{a}_1 \neq \mathbf{a}_2$, the inequality $\mathbf{a}_1 \neq b \cdot \mathbf{a}_2$ holds. It is clear that the size of $\mathcal{A}$ is $|\mathcal{A}| = \frac{q^m-1}{q-1} + 1$. Consider the q-partition code

$$X \triangleq \{\langle (\mathbf{a}, \mathbf{z}),\ \mathbf{z} \in F_q^m \rangle \mid \mathbf{a} \in \mathcal{A}\} \tag{1}$$

of size $|X| = |\mathcal{A}| = \frac{q^m-1}{q-1} + 1$.

Lemma. *The distance of q-partition code X defined by* (1) *is* $D(X) = (q-1)q^{m-1}$.

Proof of Lemma. We actually prove that the partition distance between two arbitrary codewords of X is equal to $(q-1)q^{m-1}$. Thus, code X is an equidistant code in the partition metrics. Let $\mathbf{a}_1 \neq \mathbf{a}_2$ be two arbitrary fixed elements of $\mathcal{A}$ and $\mathbf{x}(j) = \langle (\mathbf{a}_1, \mathbf{z}),\ \mathbf{z} \in F_q^m \rangle$, $\mathbf{x}(j') = \langle (\mathbf{a}_2, \mathbf{z}),\ \mathbf{z} \in F_q^m \rangle$ be two distinct codewords of X. Obviously, for any $\mu \in \mathcal{M}_q$, the Hamming distance

$$H(\mathbf{x}(j), \mathbf{x}(j')^\mu) = q^m - S(\mathbf{x}(j), \mathbf{x}(j')^\mu)$$

where $S(\mathbf{x}, \mathbf{y})$ is the Hamming similarity, i.e., the number of coordinates where $\mathbf{x}$ and $\mathbf{y}$ coincide. One can easily check that

$$S(\mathbf{x}(j), \mathbf{x}(j')^\mu) = \sum_{x \in F_q} |\{\mathbf{z} \in F_q^m :\ (\mathbf{a}_1, \mathbf{z}) = x \ \text{ and } \ (\mathbf{a}_2, \mathbf{z}) = \mu^{-1}(x)\}|, \quad \mu \in \mathcal{M}_q. \tag{2}$$

Consider two cases.

1. If $\mathbf{a}_1 = \mathbf{0}$ and $\mathbf{a}_2 \neq \mathbf{0}$, then

   $$S(\mathbf{x}(j), \mathbf{x}(j')^\mu) = |\{\mathbf{z} \in F_q^m :\ (\mathbf{a}_2, \mathbf{z}) = \mu^{-1}(0)\}| \ = \ q^{m-1}, \quad \mu \in \mathcal{M}_q.$$

   The last equality follows because $\mathbf{a}_2 \neq \mathbf{0}$.

2. If $\mathbf{a}_1 \neq \mathbf{0}$ and $\mathbf{a}_2 \neq \mathbf{0}$, then we will show that for any $\mu \in \mathcal{M}_q$, all terms in (2) are equal to q^{m-2}. Fix an arbitrary $x \in F_q$. The term corresponding to x is the number of solutions to the following system of linear equations over F_q:

   $$\begin{cases} (\mathbf{a}_1, \mathbf{z}) &= x \\ (\mathbf{a}_2, \mathbf{z}) &= \mu^{-1}(x). \end{cases}$$

   For any $\mu \in \mathcal{M}_q$, the system is non-degenerate since $\mathbf{a}_1 \neq b \cdot \mathbf{a}_2$, $\mathbf{a}_1 \neq \mathbf{0}$, and $\mathbf{a}_2 \neq \mathbf{0}$. Thus, the number of solutions is equal to q^{m-2} and $S(\mathbf{x}(j), \mathbf{x}(j')^\mu) = q^{m-1}$.

Therefore, for any $\mu \in \mathcal{M}_q$, and any two codewords $\mathbf{x}(j) \neq \mathbf{x}(j')$ of code X, the Hamming distance

$$H(\mathbf{x}(j), \mathbf{x}(j')^\mu) = q^m - S(\mathbf{x}(j), \mathbf{x}(j')^\mu) = q^m - q^{m-1} = (q-1) \cdot q^{m-1}.$$

This completes the proof of Lemma. Proposition 3 is proved.

3 Bounds on the Rate

Let D, $1 \leq D \leq n(1-1/q)$, be an integer and $N_{\mathcal{P}}(q, n, D)$ denote the *maximal size* of q-partition codes X for the set $[n]$ having distance $D(X) \geq D$. If q is a prime or prime power, then Proposition 3 shows that

$$N_{\mathcal{P}}\left(q,\, q^m,\, (q-1)q^{m-1}\right) \ \geq \ \frac{q^m - 1}{q-1} + 1\,, \qquad m = 1, 2, 3, \ldots.$$

If d, $0 < d < 1 - 1/q$, is fixed, then

$$R_{\mathcal{P}}(q,d) \triangleq \varlimsup_{n\to\infty} \frac{\log_q N_{\mathcal{P}}(q,n,nd)}{n}$$

is called a *rate* of q-partition codes. A random coding lower bound on $R_{\mathcal{P}}(q,d)$ presented by Proposition 4 coincides with the classical Varshamov-Gilbert bound [4] for codes in Hamming metric.

Proposition 4. *For any d, $0 < d < 1 - 1/q$, the rate $R_{\mathcal{P}}(q,d) > 0$ and*

$$R_{\mathcal{P}}(q,d) \;\geq\; 1 - d\log_q(q-1) - h_q(d), \qquad h_q(d) \triangleq\; -d\log_q d - (1-d)\log_q(1-d).$$

Proof. Let $\mathbf{x}(j)$ and $\mathbf{x}(j')$, $j \neq j'$, be q-ary independent random codewords of code X with the same uniform distribution, i.e., for any q-ary n-sequence $\mathbf{x}$, the probability

$$\Pr\{\mathbf{x}(j) = \mathbf{x}\} = \Pr\{\mathbf{x}(j') = \mathbf{x}\} = q^{-n}.$$

For $\mu \in \mathcal{M}_q$, consider the random variable $\xi_\mu = \xi_\mu(\mathbf{x}(j), \mathbf{x}(j')) \triangleq \sum\limits_{x\in A_q} n\left(x, \mathbf{x}(j)\,;\, \mu(x), \mathbf{x}(j')\right)$.
One can easily check that ξ_μ has the following binomial distribution:

$$\Pr\left\{\xi_\mu(\mathbf{x}(j), \mathbf{x}(j')) = k\right\} \;=\; \binom{n}{k} \cdot \left(\frac{1}{q}\right)^k \cdot \left(1 - \frac{1}{q}\right)^{n-k}, \quad k = 0, 1, \ldots, n\,,$$

which does not depend on μ. In virtue of Definition 1 and Proposition 1, the partition distance $\mathcal{P}_q(\mathbf{x}(j), \mathbf{x}(j')) = n - \max\limits_{\mu\in\mathcal{M}_q} \xi_\mu(\mathbf{x}(j), \mathbf{x}(j'))$ and, therefore, for any $\mu \in \mathcal{M}_q$, the probability

$$\Pr\left\{\mathcal{P}_q(\mathbf{x}(j), \mathbf{x}(j')) \leq nd\right\} \;\leq\; q! \cdot \Pr\left\{\xi_\mu(\mathbf{x}(j), \mathbf{x}(j')) \geq n(1-d)\right\} =$$

$$= \; q! \sum_{k=n(1-d)}^{n} \binom{n}{k} \cdot \left(\frac{1}{q}\right)^k \cdot \left(1 - \frac{1}{q}\right)^{n-k}.$$

If d, $0 < d < 1 - 1/q$, is fixed, then the standard arguments used to obtain the random coding bound yield

$$R_{\mathcal{P}}(q,d) \geq \varlimsup_{n\to\infty} -\frac{\log \Pr\left\{\mathcal{P}_q(\mathbf{x}(j), \mathbf{x}(j')) \leq nd\right\}}{n} \geq$$

$$\geq \varlimsup_{n\to\infty} -\frac{q!}{n} + \varlimsup_{n\to\infty} -\frac{\log\left[\sum\limits_{k=n(1-d)}^{n} \binom{n}{k} \cdot \left(\frac{1}{q}\right)^k \cdot \left(1 - \frac{1}{q}\right)^{n-k}\right]}{n} =$$

$$= -d\log_q\left(1 - \frac{1}{q}\right) - (1-d)\log_q\frac{1}{q} - h_q(d) = 1 - d\log_q(q-1) - h_q(d).$$

Proposition 4 is proved.

References

[1] D. Gusfield, Partition-distance: A problem and class of perfect graphs arising in clustering, *Information Processing Letters*, **82** (2002), pp. 159-164.

[2] A. Almudevar, C. Field, Estimation of single generation sibling relationships based on DNA markers, *J. Agricultural, Biological Environment. Statist.*, **4** (1999), pp. 136-165.

[3] B.G. Mirkin, L.B. Tcherny, On measurement of proximity between various partitions of finite set, *Avtomatika i Telemekhanika*, 1970, No. 5, pp. 120-127 (in Russian).

[4] F.J. MacWilliams, N.J.A. Sloan, *The Theory of Error - Correcting Codes*, Amsterdam, The Netherlands: North Holland, 1977.