Andrey G. Bronevich and Natalia S. Spiridenkova

Some Characteristics of Credal Sets and Their Application to Analysis of Polls Results

National Research University
Higher School of Economics
(Moscow, RUSSIA)

ITQM 2017 in New Dehli (India)
8-10 December, 2017

The outline of the paper

The analysis of election results is usually based on classical statistics, but for some problems it would be reasonable to build statistical models based on imprecise probabilities that are more flexible for modeling uncertainty.
The main idea of this paper is to descibe election results by families of probability measures called credal sets. This helps us to describe adequately the choice of population located in large regions. We propose to use several characteristics to evaluate polls results like amount of contradiction between credal sets, inclusion indices and imprecision indices. This allows us to make conclusions about elections conducted in Russia.

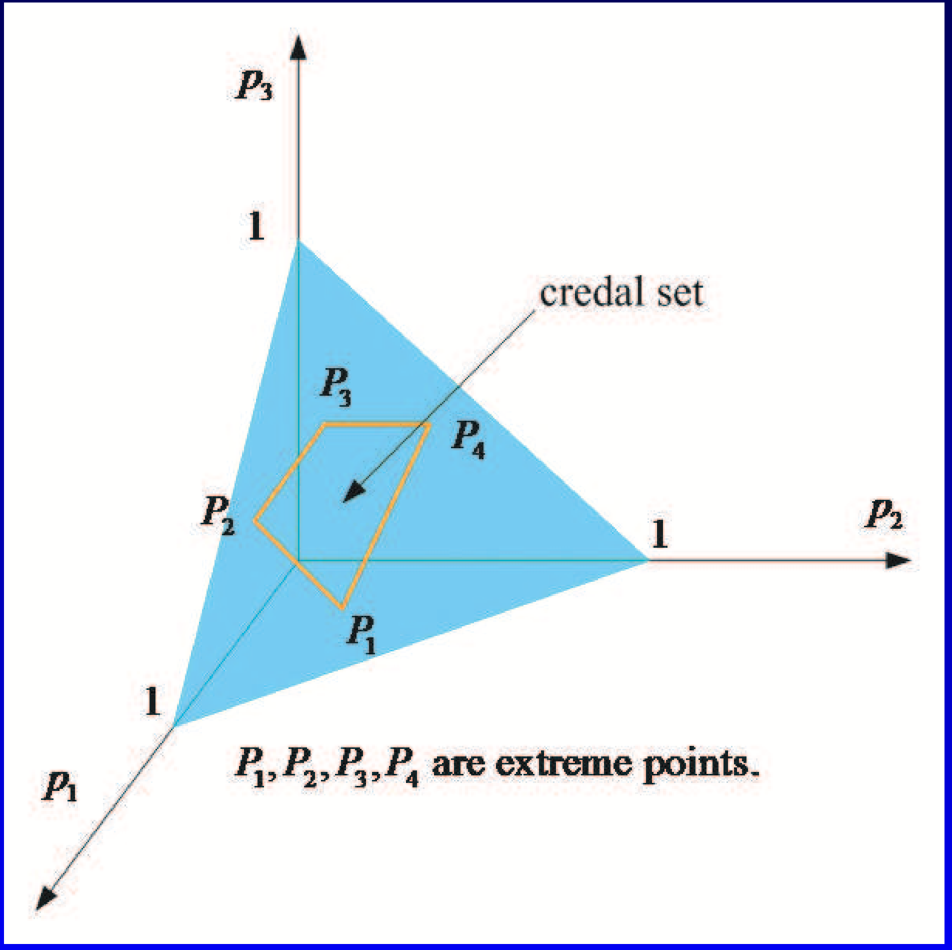
Credal Sets

- Let X be a finite set, 2^X be the powerset of X and M_{pr} be the set of all probability measures on 2^X.
- Any $P \in M_{pr}$ can be represented as a point $(P(\{x_1\}), ..., P(\{x_n\}))$ in $\mathbb{R}^n$.
- A *credal set* $\mathbf{P}$ is a non-empty subset of M_{pr}, which is convex and closed;
- if $\mathbf{P}$ is a credal set with a finite number of extreme points $P_k \in M_{pr}$, $k = 1, ..., m$, then

$$\mathbf{P} = \left\{ \sum_{k=1}^{m} a_i P_i | a_i \geqslant 0, \sum_{k=1}^{m} a_i = 1 \right\}.$$

Example

Let $X = \{x_1, x_2, x_3\}$ and denote $P(\{x_i\}) = p_i$, $i = 1, 2, 3$, then any credal set is convex subset of triangle (p_1, p_2, p_3): $p_i \geqslant 0$, $p_1 + p_2 + p_3 = 1$.



Contradiction in information

$P_1, P_2 \in M_{pr}$ are *fully contradictory* if there are $A, B \in 2^X$, such that $A \cap B = \emptyset$, $P_1(A) = P_2(B) = 1$, i.e. for the first source of information the event A is certain, but for the second it is impossible. For arbitrary $P_1, P_2 \in M_{pr}$ we can always divide P_1 and P_2 on two parts:

$$P_i = (1-a)P + aP_i^{(2)}, \;\; i = 1, 2, \qquad (1)$$

where $P, P_1^{(2)}, P_2^{(2)} \in M_{pr}$, $a \in [0,1]$, and $P_1^{(2)}, P_2^{(2)}$ are fully contradictory. The representation (1) is defined uniquely if $a \in (0,1)$. The value a is called the *amount of contradiction* between P_1 and P_2, and it is denoted by $Con(P_1, P_2)$.

Contradiction in information

We can compute $Con(P_1, P_2)$ using the formula:

$$Con(P_1, P_2) = 1 - \sum_{i=1}^{n} \min\{P_1(\{x_i\}), P_2(\{x_i\})\}.$$

The next proposition shows that $Con(P_1, P_2)$ is a metric on the set M_{pr}.

Proposition 1.

1. $Con(P_1, P_2) = 0.5 \sum_{i=1}^{n} |P_1(\{x_i\}) - P_2(\{x_i\})|;$
2. $Con(P_1, P_2) = \max_{A \in 2^X} |P_1(A) - P_2(A)|.$

Definition 1. The amount of contradiction between credal sets $\mathbf{P}_1$ and $\mathbf{P}_2$ is computed as

$$Con(\mathbf{P}_1, \mathbf{P}_2) = \inf \left\{ Con(P_1, P_2) | P_1 \in \mathbf{P}_1, P_2 \in \mathbf{P}_2 \right\}.$$

Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be credal sets with finite number of extreme points: $P_1^{(k)}, ..., P_{N_k}^{(k)}$ are extreme points of $\mathbf{P}_k$, $k = 1, 2$. Then any $P_1 \in \mathbf{P}_1$ can be represented as a linear combination of measures $P_1^{(1)}, ..., P_{N_1}^{(1)}$, i.e. $P_1 = \sum_{j=1}^{N_1} a_j P_j^{(1)}$, where $\sum_{j=1}^{N_1} a_j = 1$, $a_j \geqslant 0$, $j = 1, ..., N_1$.

In this case $Con(\mathbf{P}_1, \mathbf{P}_2)$ is a solution of the linear programming problem:

$$Con(\mathbf{P}_1, \mathbf{P}_2) = 1 - \sum_{i=1}^{n} c_i \to \min,$$

$$\begin{cases} \sum\limits_{j=1}^{N_1} a_j P_j^{(1)}(x_i) - c_i \geqslant 0, \ i = 1, \ldots, n, \\ \sum\limits_{j=1}^{N_2} b_j P_j^{(2)}(x_i) - c_i \geqslant 0, \ i = 1, \ldots, n, \\ a_j \geqslant 0, \, j = 1, \ldots, N_1, \ b_j \geqslant 0, \, j = 1, \ldots, N_2, \\ c_i \geqslant 0, \ i = 1, \ldots, n. \end{cases}$$

We can also compute the contradiction among several sources of information using the following definition.

Definition 2.

1. if $P_1, ..., P_m \in M_{pr}$, then

$$Con(P_1, ..., P_m) = 1 - \sum_{i=1}^{n} \min\{P_1(\{x_i\}), ..., P_m(\{x_i\})\};$$

2. if $\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_m$ are credal sets, then

$$Con(\mathbf{P}_1, ..., \mathbf{P}_m) = \inf\{Con(P_1, ..., P_m) | P_i \in \mathbf{P}_i, i = 1, ..., m\}.$$

Properties of $Con(\mathbf{P}_1, ..., \mathbf{P}_m)$

Information sources, described by credal sets $\mathbf{P}_1, ..., \mathbf{P}_m$, are called fully contradictory if $Con(\mathbf{P}_1, ..., \mathbf{P}_m) = 1$.

Proposition 2. Let $\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_m$ be credal sets, then

1. $Con(\mathbf{P}_1, ..., \mathbf{P}_m) = Con(\mathbf{P}_{i_1}, ..., \mathbf{P}_{i_m})$, where $(i_1, ..., i_m)$ is an arbitrary permutation of $\{1, ..., m\}$;
2. $Con(\mathbf{P}_1, ..., \mathbf{P}_{m-1}, \mathbf{P}_m) = Con(\mathbf{P}_1, ..., \mathbf{P}_{m-1})$ if $\mathbf{P}_{m-1} = \mathbf{P}_m$.

Proposition 3. $Con(P_1, ..., P_m) = 1$ for $P_1, ..., P_m \in M_{pr}$ iff there are $A_1, ..., A_m \in 2^X$ such that $P_j(A_j) = 1$, $j = 1, ..., m$, and $\bigcap_{j=1}^{m} A_j = \emptyset$.

Types of uncertainty

By credal sets we can described 2 types of uncertainty: non-specificity and conflict.

Non-specificity is connected with a with a choice of a probability measure among possible alternatives.

Conflict is related to probability measures. For example, among voters there is a conflict if one part of them gives their votes to one candidate, and the second part gives their votes to the opposite candidates.

Measuring Non-specificity

Definition 3. The amount of non-specificity based on the contradiction measure is

$$\nu_N^{(2)}(\mathbf{P}) = \sup\left\{Con(P_1, P_2) | P_1, P_2 \in \mathbf{P}\right\}.$$

Remark. Other possible choices of an non-specificity measure are to use the functionals:

$$\nu_N^{(k)}(\mathbf{P}) = \sup\left\{Con(P_1, ..., P_k) | P_1, ..., P_k \in \mathbf{P}\right\}$$

where $k = 3, 4, ...$

If in the expression of an non-specificity measure we take the amount of contradiction among all probability measures in $\mathbf{P}$, then we get the functional

$$\nu_N^{(\infty)}(\mathbf{P}) = 1 - \sum_{i=1}^{n} \inf \{P(\{x_i\})|P \in \mathbf{P}\}.$$

Proposition 3. Let $\mathbf{P}$ be a credal set with a finite set $ext(\mathbf{P})$ of its extreme points. Then

$$\nu_N^{(k)}(\mathbf{P}) = \sup \{Con(P_1, ..., P_k)|P_1, ..., P_k \in ext(\mathbf{P})\},$$

where $k = 2, 3, ...$

Corollary. Let $\mathbf{P}$ be a credal set with a finite set $ext(\mathbf{P}) = \{P_1, ..., P_m\}$ of its extreme points. Then $\nu_N^{(k)}(\mathbf{P}) = Con(P_1, ..., P_m)$ for $k \geqslant m$. In particular, also $\nu_N^{(\infty)}(\mathbf{P}) = Con(P_1, ..., P_m)$.

Measuring conflict

A probability measure describes information without conflict iff there is $x_k \in X$ such that $P(\{x_k\}) = 1$ and $P(\{x_i\}) = 0$ for every $x_i \neq x_k$. Such probability measures are called Dirac measures and we denote them by $\eta_{\langle\{x_k\}\rangle}$. Therefore, conflict described by a probability measure $P \in M_{pr}$ should reflect how P is close to a Dirac measure.

Definition 3. Let $P \in M_{pr}$, then the amount of conflict in P based on the contradiction measure is

$$\nu_C(P) = \min_{x_k \in X} Con(P, \eta_{\langle\{x_k\}\rangle}).$$

Because $Con(P, \eta_{\langle\{x_k\}\rangle}) = 1 - P(\{x_k\})$, the expression for $\nu_C(P)$ can be transformed to

$$\nu_C(P) = 1 - \max_{x_k \in X} P(\{x_k\}).$$

If we try to evaluate conflict in a credal set $\mathbf{P} \subseteq M_{pr}$, we should answer the question - whether conflict depends on all probability measures in $\mathbf{P} \subseteq M_{pr}$, or it is sufficient to find one probability measure with minimal conflict. This two possibilities lead to the following functionals:

$$\nu_C^{(1)}(\mathbf{P}) = \min_{x_k \in X} \sup_{P \in \mathbf{P}} Con(P, \eta_{\langle\{x_k\}\rangle}),$$

$$\nu_C^{(2)}(\mathbf{P}) = \min_{x_k \in X} \inf_{P \in \mathbf{P}} Con(P, \eta_{\langle\{x_k\}\rangle}).$$

Proposition 6. Let $\mathbf{P}$ be a credal set with a finite set $ext(\mathbf{P}) = \{P_1, ..., P_m\}$ of its extreme points. Then

1. $\nu_C^{(1)}(\mathbf{P}) = 1 - \max_{x_k \in X} \min_{i=1,...,m} P_i(\{x_k\});$
2. $\nu_C^{(2)}(\mathbf{P}) = 1 - \max_{x_k \in X} \max_{i=1,...,m} P_i(\{x_k\}).$

Distance between credal sets

We will use next the Hausdorff distance between credal sets based on the contradiction measure. Because the amount of contradiction between probability measures can be viewed as a metric on the set of probability measures, we can define the inclusion index of credal sets $\mathbf{P}_1$ and $\mathbf{P}_2$ as

$$\nu(\mathbf{P}_1 \subseteq \mathbf{P}_2) = \sup_{P_1 \in \mathbf{P}_1} \inf_{P_2 \in \mathbf{P}_2} Con(P_1, P_2),$$

and then the Hausdorff distance between $\mathbf{P}_1$ and $\mathbf{P}_2$ is defined as

$$d_H(\mathbf{P}_1, \mathbf{P}_2) = \max \left\{ \nu(\mathbf{P}_1 \subseteq \mathbf{P}_2), \nu(\mathbf{P}_2 \subseteq \mathbf{P}_1) \right\}.$$

Proposition 7. Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be an arbitrary credal sets. Then

1. $\nu(\mathbf{P}_1 \subseteq \mathbf{P}_2) = 0$ iff $\mathbf{P}_1 \subseteq \mathbf{P}_2$;
2. $d_H(\mathbf{P}_1, \mathbf{P}_2) = 0$ iff $\mathbf{P}_1 = \mathbf{P}_2$.

Proposition 8. Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be credal sets with finite number of extreme points, i.e. we assume that $P_1^{(k)}, \ldots, P_{N_k}^{(k)}$ are extreme points of $\mathbf{P}_k$, $k = 1, 2$. Then

$$\nu(\mathbf{P}_1 \subseteq \mathbf{P}_2) = \max_{j=1,\ldots,N_1} Con(P_j^{(1)}, \mathbf{P}_2),$$

$$\nu(\mathbf{P}_2 \subseteq \mathbf{P}_1) = \max_{j=1,\ldots,N_2} Con(\mathbf{P}_1, P_j^{(2)}).$$

Election Results in Russia

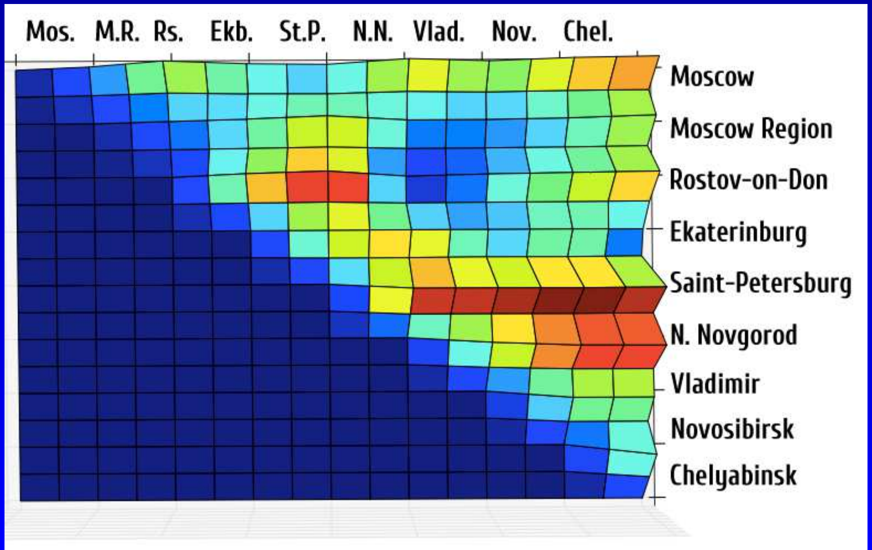
We can describe election results by credal sets considering the distribution of votes on districts. For example, consider the election results in Moscow city. Moscow consists of ten districts: Babushkinskiy, Kuntsevskiy, Leningradskiy, etc, and electors could vote for thirteen parties: "Motherland", "Communists of Russia", "Pensioners for Fairness", "United Russia", etc. Only four parties ("United Russia" (UR), "Fair Russia" (FR), Communist Party of Russian Federation (CPRF), and Liberal Democratic Party of Russia (LDPR)) went through the 5% threshold and came to the parliament by elections in 2016. The results of these parties and "Apple" in Moscow are shown in Table 1.

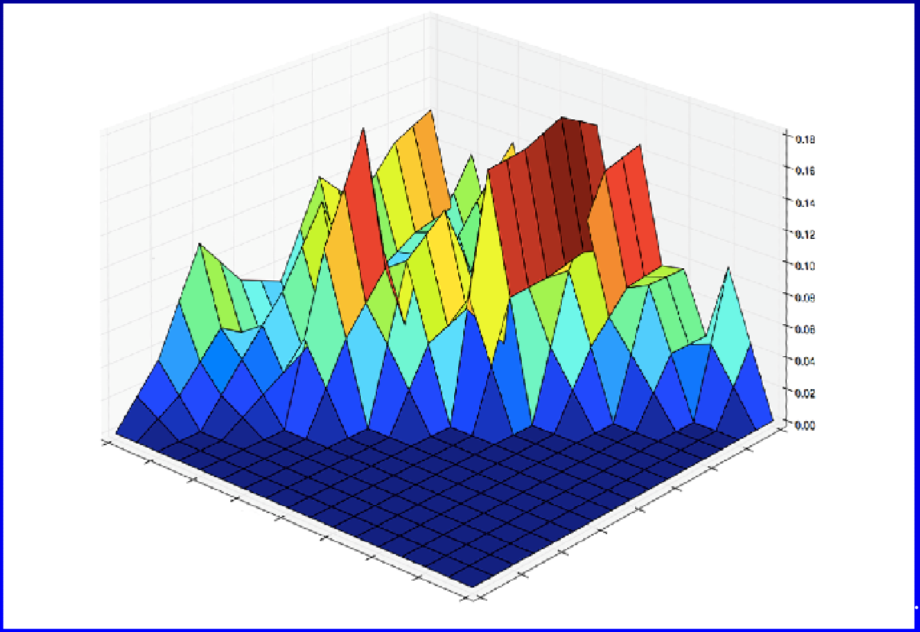
	UR	LDPR	Apple	CPRF	FR
Babushkinskiy	34.49	12.94	10.93	14.27	7.24
Kuntsevskiy	35.44	11.82	11.43	14.83	5.94
Leningradskiy	33.65	12.91	10.98	13.96	9.24

Thus, we can describe election results in Moscow by a credal set, in which extreme points correspond to probability measures that describe election results in districts. Then we analyze election results in nine regions: Moscow City, Moscow Region, Rostov-on-Don, Saint Petersburg, Yekaterinburg, Nizhny Novgorod, Chelyabinck, Novosibirsk, and Vladimir.

Results

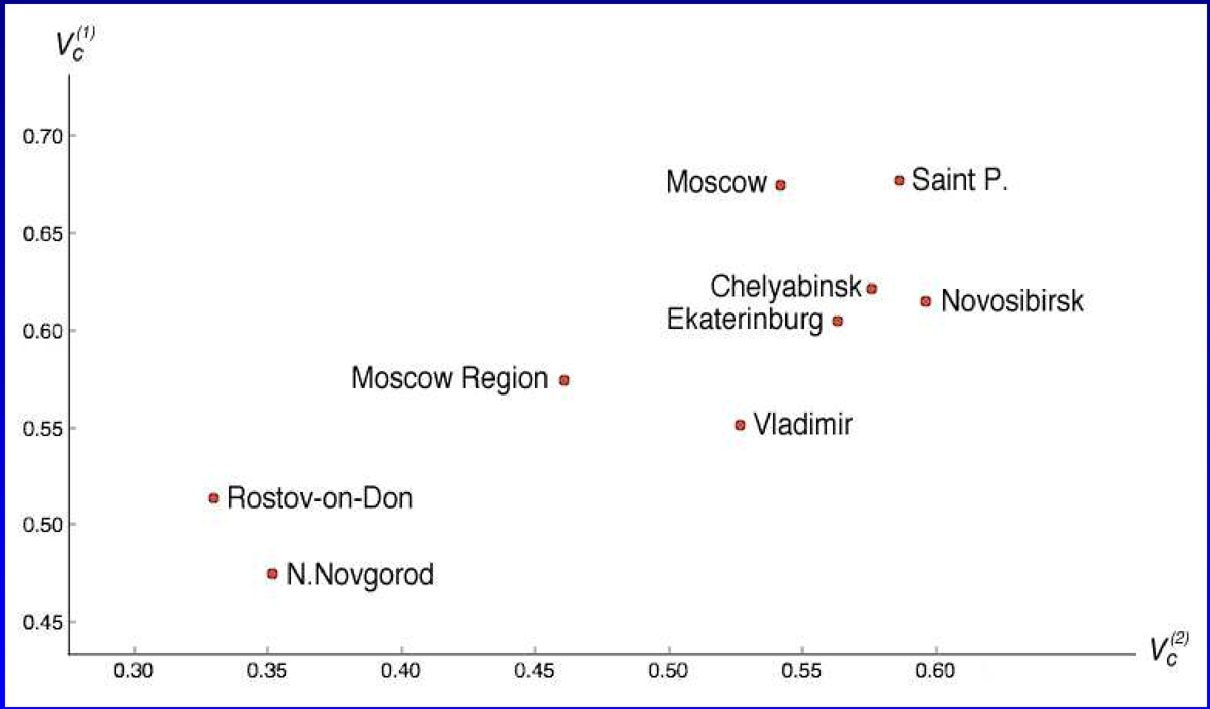
Next Figure depicts a map showing the amount of contradiction between voting results in the considered regions. We see a high contradiction zone (intensive red colour), where Saint-Petersburg intersects with other cities, while the two first rows (light green and blue), which correspond to Moscow and Moscow Region, show low contradiction between these two zones and other cities.





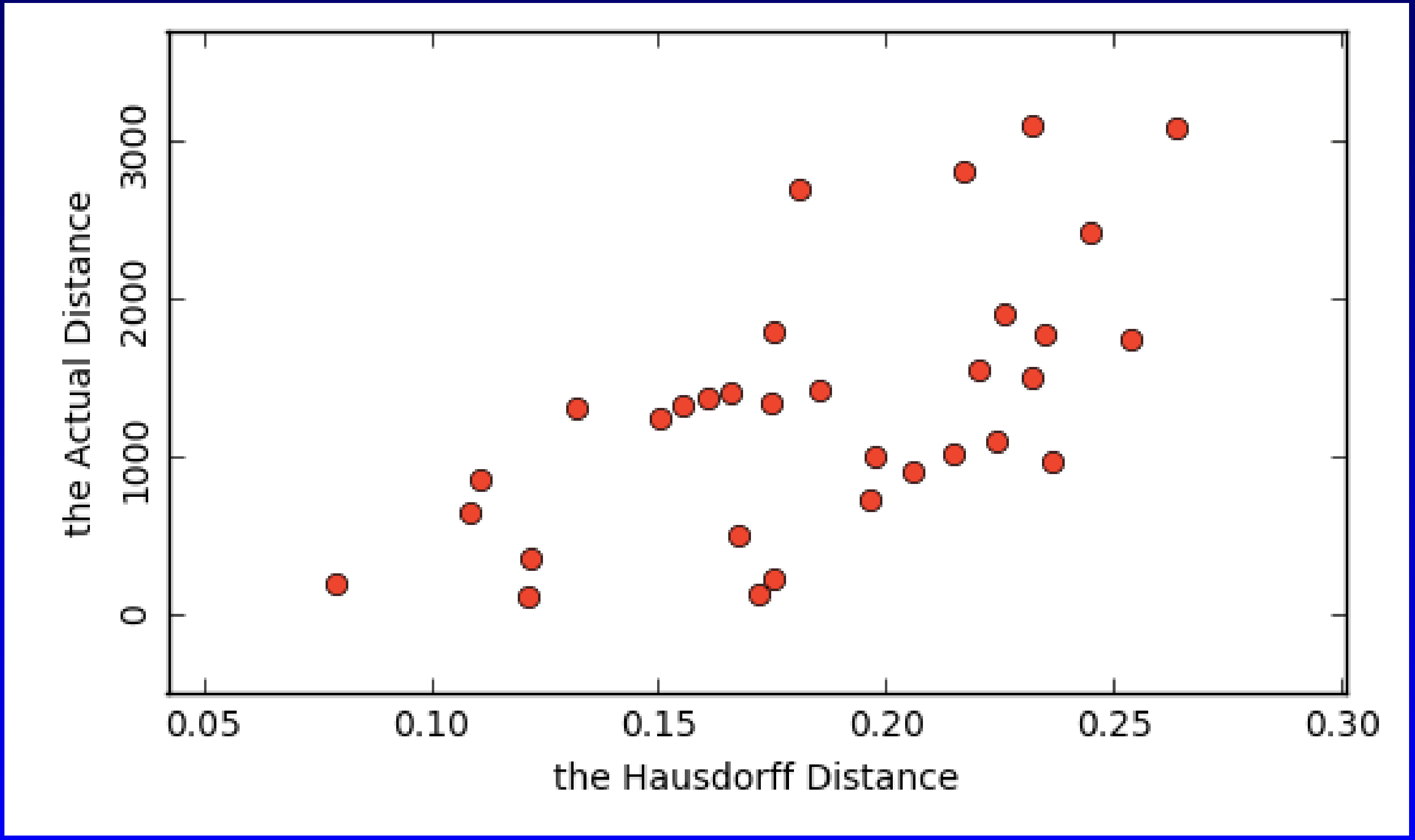
Results

Next figure shows how the cities are grouped according to the values of $\nu_C^{(1)}$ and $\nu_C^{(2)}$. We see that in Moscow, Saint-Petersburg and Novosibirsk the preferences of electorate are very ambiguous, while the citizenries of Rostov-on-Don and Nizhny Novgorod are quite solid.



Results

Finally, comparing the real distance between cities and the Hausdorff distance between polls results, depicted on Fig. 3, it is possible to note a certain correlation between values of these metrics.



Thank you for attention!!!