The QR code as a symbol of protest: analysis of social networks

Abstract

In this study, we examine data from the popular Russian social network "VKontakte" to understand what types of discourses on the topic of QR codes are spread in it, including discourses related to protests. Clustering based on the doc2vec algorithm proved to be an effective method to group the communities by related topics. In particular, this approach allowed us to detect a cluster of communities representing op-ponents of QR code technology related to vaccination certificates from COVID-2019.

1. Introduction

In 2021 QR-code technology actively began to be used in the sphere of public health when mass vaccination against COVID-19 started all over the world. Vaccinated people began to receive vaccination documents - certificates with a QR code. In November-December 2021 a wave of protests against vaccination against COVID-19 and implementation of QR codes took place all over the world (Germany, Italy, France, USA, Finland, Northern Ireland, Japan, Australia and others). This period can be marked as the starting point, when the discussion of this technology shifted from the discourse of convenient encoding of small datasets in a graphic picture to a negative connotation.

In November 2021, draft federal laws were submitted to the State Duma of the Russian Federation to ensure the sanitary and epidemiological well-being of the population in case of a threat of the spread of a new coronavirus infection. According to these draft laws, citizens can visit places of mass and other events, cultural organizations, public catering, retail trade, etc. only if they present documentation containing a QR code and confirming preventive vaccination against coronavirus infection or presence of medical contraindications to such vaccination or confirming that a citizen has suffered from the disease of this infection.

In Russia the peak of protest activity of citizens against "Bills on QR codes for access to transport and stores" fell on November 2021. In Belgorod, Novosibirsk, Irkutsk, Yuzhno-Sakhalinsk and in many other cities of the Russian Federation the opponents of QR-codes conducted mass meetings against the restrictive measures in the regions, sang "Victory Day", collected signatures under the petition to cancel the "QR-law". The subjects of the protest are very diverse: protest against violation of the basic rights laid down in the Constitution of the Russian Federation (right to freedom of movement, right to work, right of access to cultural institutions, etc.); protest against any form of digitalization (including on religious grounds); protest against QR codes and government; protest against the introduction of medical documents on vaccination and growth of prices for housing services, etc. Thus, the system of QR-certification of vaccinated citizens became the reason of segregation of people into 2 camps: having full rights (having a QR-code) and limited rights (not having a QR-code).

The aim of this research is to identify online communities of users of the social network "VKontakte", which oppose the introduction of the QR-code system, and to describe the types of protest in the period of digital transformation.

Using the API of the social network VKontakte, we have prepared a corpus of texts related to the discussion of QR codes. We develop the methodology of social network analysis proposed in [Platonov, Svetlov, 2021], [Svetlov, Legostaeva, 2021]. Namely, using modern methods of natural language

processing we constructed vector representations of online communities and used them to form a thematic graph to discuss QR codes. The vertices in it are individual communities and the edges connect those pairs of communities that have a thematic proximity value above a certain threshold value. By analyzing the graph, we identified clusters of communities with similar themes. Qualitative (analysis of textual and graphic content of online communities) and quantitative (number of participants, engagement metrics) analysis of the selected clusters allowed us to identify new types of protest in the period of digital transformation.

2. Related Work

Most studies of protest activity lie in the plane of political analysis during the pre-election periods of electoral activity. Modern research in this field is complicated by the hybridization of the object and subject of research, as protest activity is less related to certain periods of pre-election campaigns. It becomes difficult to predict the periods of outbreaks of protest points, as the subject of protest is transferred from the monopolized political sphere to the environmental sphere (protests against unorganized dumps), cultural (protests against the demolition or installation of monuments, construction of churches), healthcare (protests against vaccination).

Another distinctive feature of the modern analysis of protest activity is the study of online and digital forms of protest [Avtsinova G., 2014; Vanke, 2014; Nim, 2016; Legostaeva, 2019; Platonov K., Legostaeva N., 2019]. The innovation of this study lies in the fact that the authors used one key word to identify new types of protest, and were able to reconstruct the map of protest activity in the offline format.

3. Methodology

To build a corpus of texts related to the QR codes theme we used the API of the social network "VKontakte" (method newsfeed.search), which allows to get a set of texts on a given search query, belonging to a certain period of time. The main limitation is that the specified method allows to get no more than 1000 posts. To obtain the maximum possible set of posts, we used the newsfeed.search method for each day of the period 01.01.2021 through 22.02.2022, with q='QR code' as the search string parameter.

The collected database DB_0 included 323904 items, including 238238 posts, and 85666 user comments. For the purposes of this research, we have considered data posted in social network communities with a number of members over one thousand people and containing at least ten posts with textual content in our database. The indicated limitations are related to the fact that the proposed methodology is based on the construction of a graph of thematic similarity of communities, so communities with a small number of posts cannot be correctly processed.

The final database DB_1 that passed the specified filtering procedure included 44600 posts corresponding to 1514 communities. Posts from DB_1 passed the following preprocessing steps as a result of which the DB_2 database was created:

1. Conversion of all letters to lower case;
2. Removal of non-alphabetic characters including punctuation marks, numbers, emoticons;
3. Removal of stop words presented in NLTK package;
4. Replacement of the Russian letter "ё" with "е";
5. Replacement of links to web resources and user references with special "URL" and "USR" tokens, respectively;
6. Word normalization using the pymorphy2 library [Korobov, 2015].

The procedure of constructing the thematic landscape is based on the application of word embeddings - distributed word representations. In such models, each word corresponds to an element of Euclidean space R^d for some natural d (as a rule, the parameter d is chosen to be equal to several hundred). Nowadays, a large number of models are available to obtain the above representation for words: wor2vec [Mikolov et al., 2013], FasText [Bojanowski et al., 2017], GloVe [Pennington et al., 2014]. These models are based on the assumption, which in linguistics is called the distributional hypothesis [Rubenstein, Goodenough, 1965], according to which words with a similar meaning occur in similar contexts. Using this theoretical assumption, an ad hoc neural network is built, which in the version CBOW (continuous bag of words) solves the problem of predicting a word from a given context.

The hidden layer of this network is presented by a matrix W of size $|V| \times d$, where $|V|$ is the size of the vocabulary ($|V|$ is the vocabulary itself) and v d is the dimension of the word embeddings. The elements of W are chosen to solve this prediction problem in the best way [Goldberg, Levy, 2014] which is achieved by using a negative sampling technique. Let w be a word from the given corpus, c the context of the word (i.e., the "window" of words around w), D be the set of all possible pairs of the form (w, c), θ represents model parameters (embedding values for words and context). The probability of finding a pair (w, c) in the set D is defined as $p_\theta \left((w, c) \in D \right) = \sigma(v_w \cdot v_c)$, where $v_w \cdot v_c$ is the dot product of vectors v_w and v_c $\sigma(x) = \frac{1}{1+e^{-x}}$ is a sigmoid function, and the likelihood maximization problem is then considered

$$armax_\theta \prod_{(w,c) \in D} p_\theta((w, c) \in D) \prod_{(w',c) \in D'} \left(1 - p_\theta \left((w', c) \in D \right) \right),$$

where D' is the set of pairs (w', c) that do not exist for the corpus in question. It can be shown that its solution corresponds to the minimization of the following loss function at each step of the gradient descent

$$- \log \sigma(v_w \cdot v_c) - \sum_{(w',c) \in D'} \log \sigma(-v_{w'} \cdot v_c)$$

with the set D' being randomly generated and includes some negative samples.

A logical development of the wor2vec model is the doc2vec model [Le, Mikolov, 2014], which allows obtaining embeddings not only for single words, but also for entire documents.

The doc2vec (PV-DM) model treats the document as analogous to one more word used to improve the prediction from a given local context, attempting to incorporate the topic of the whole document. We will use a version of the doc2vec model implemented in the Gensim library [Rehurek, Sojka, 2010]. This implementation enables us to add additional tags reflecting different aspects of the document by using the TaggedDocument class. In addition to the embedding of words and documents, this implementation of the model also produces embeddings for the specified tags (Fig. 1). Accordingly, the context vector is generated by averaging the embeddings of both the words in the window around a given word and the embeddings of the post ($post\_id$) and the group from the social network ($group\_id$) in which the post was published. In this way, we seamlessly obtain vector representations of the communities in question from the social network, without having to average the embeddings of the posts they have published.
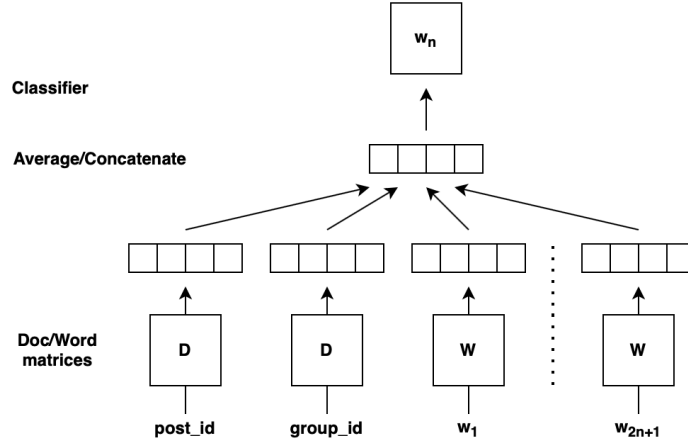
Figure 1: The architecture of the doc2vec model (PV-DM) used in the research

In this paper we use an implementation of the doc2vec model in the form of PV-DM, as the original paper [Le, Mikolov, 2014] noted its advantage over the PV-DBOW version.

The dimension of the embedding space was chosen equal to 256. Also, the model parameter min_count was defined as 20, which ignores all words with frequency lower than 20. Training on a corpus of texts was performed over 10 epochs. The resulting embeddings for the communities from the DB_2 database reflect their thematic as a whole and can therefore be used to compare the closeness of the communities in a thematic sense. The present study uses cosine distance as a measure of proximity.

Namely, for two communities c_1 and c_2 we calculate the following value

$$cosinesimilarity(c_1, c_2) = \frac{v_{c_1} \cdot v_{c_1}}{|v_{c_1}||v_{c_1}|},$$

where v_{c_1} and v_{c_1} are the embeddings obtained for communities c_1 and c_2, $v_{c_1} \cdot v_{c_1} = \sum_{i=1}^{d} v_{c_1,i} \cdot v_{c_2,i}$ - the dot product of these embeddings, and $|v_{c_k}| = \sqrt{\sum_{i=1}^{d} v_{c_k,i}^2}$, $k = 1,2$ - their norms.

Communities with similar topics will have this measure close to 1 and, on the contrary, communities with completely different topics will have this measure close to 0.

The next step in the analysis was to construct an undirected graph using the networkx library [Hagberg et al., 2008]. The vertices in it are the communities, and the edges are connecting those pairs of communities where the cosine measure is greater than a given threshold value (in this case the threshold value was chosen as 0.66, which corresponds to 0.80 of the sample quantile estimated from the calculated cosine distances).

The resulting graph has the following characteristics:
* Number of vertices: 1514;
* Number of edges: 43030;

Then the graph was exported to gexf format and analyzed in the Gephi application. Using the graph modularity maximization algorithm implemented in the Gephi application (the Resolution parameter was chosen to be 1.0) [Bastian et al, 2009], the communities were divided into 419 clusters. Some of these clusters were of size 1, i.e., consisted of a single community.

4. Results

As a result of the application of this methodology, a graph was constructed by the key "QR-code", the period of downloading information from 01.01.2021 to 22.02.2022. The graph shows a total of 419

clusters with 1,514 online communities with a total number of participants equal to 81,608,589 (the total number of non-unique participants of the groups is indicated) (Fig. 2). In the structure of the graph, we selected clusters that included at least 50 online communities. Among them, the following clusters were identified: a cluster consisting of online information communities that duplicate content from official news sites (Cluster Id No. 99); a cluster with groups of advertisements on finding jobs, apartments, buying and selling clothes (Cluster Id No. 174); a cluster of communities representing Russian tour operators (Cluster Id No. 157); a cluster of groups related to the organization of cultural leisure in St. Petersburg (Cluster Id No. 40); a cluster of communities focused on job search for different target groups (Cluster Id No. 12).

As a result of the analysis of the graph, a Cluster No. 53 of communities was identified, consisting of 80 groups, consisting of 943,720 participants who discuss the topic "Protest against QR Codes" (Fig. 3). The analysis of the content of online communities allowed us to identify new types of protest against digital segregation.

We use the engagement rate metric calculated for each post as $(likes + comments)/\ views * 1000,$ as well as the special metrics ERc, ERl, and ERr (for measures such as comments, likes, reposts accordingly) calculated as $measure/views * 1000$. A comparative analysis (Tab. 1) of the statistical parameters of the metric "engagement rate" showed that the indicator for this cluster is higher than all others, that is, the users of the social network are actively included in the discussion of this topic.
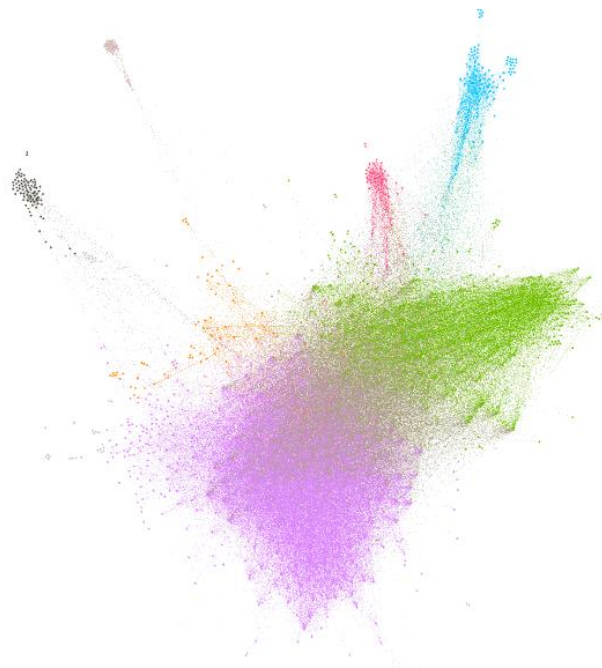

Figure 2: Community graph by "QR code" key

Table 1
User engagement metrics for selected thematic clusters

Cluster Id	Number of groups	Number of participants	Posts	ER	ER_L	ER_C
99	362	28 364 899	5 652	7,4	5,6	1,8

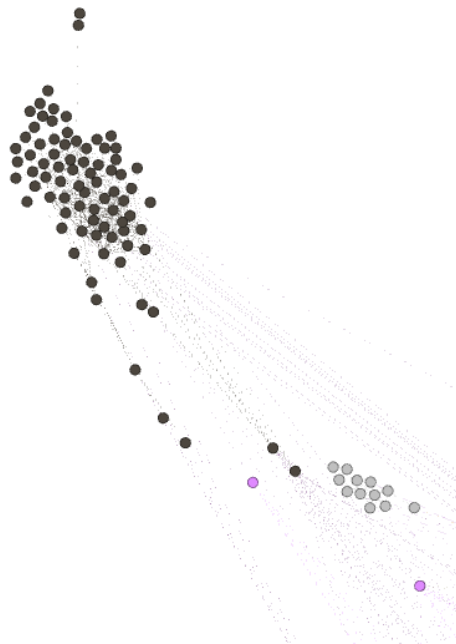| 174	236	7 874 785	5 210	11,4	10,6	0,7
157	106	1 613 749	8 165	9,4	9,1	0,3
53	80	943 720	1 737	39,6	33,5	6,1
40	71	6 859 706	1 116	14,6	13,7	0,9
12	53	2 213 684	1 376	8,7	8,3	0,4

Figure 3: Community Cluster on "Protest Against QR Codes" (Cluster Id No. 53)

As a result of the analysis of the cluster groups a new list of keywords was formed, which in the future will be the basis for the collection and analysis of the specified data array for conducting research within the "Protest against digital segregation" topic: "Stopvaccism", "QR-segregation", "QR-Resistance", "#stopQR-code", "QR-ghetto", "QR-camp", "QR-fascism", etc.

An analysis of the content of publications in these communities revealed both the types of protest that have already become classic, such as appeals to the President of the Russian Federation in written and video formats, single pickets, actions, petitions, leaflets, strike of factory workers, and new types of protest activity.

For example:

- Organization of offline and online boycotts: boycott of large shopping centers refusing to sell food and goods and provide services without personal protective equipment and QR codes, boycott of digital platforms of the government and Sberbank, refusal to register at the site Gosulugi, boycott of cafes and bars that require QR codes (Fig. 4).

Figure 4: Poster wrestlers with QR codes in establishments (shopping malls, stores, restaurants, etc.)

- Collective appeals to the court against the actions of the governors: a statement to the prosecutor's office against the governor of the Kaliningrad region, stating that he introduced and consolidated at the regional level segregation and discrimination against citizens on the social basis of vaccination/not vaccination; a statement to the prosecutor's office of Belgorod Oblast to check the resolution of the Belgorod Oblast Governor № 128 dated October 14, 2021 "On amending the resolution of the Belgorod Oblast Governor dated May 8, 2020" on legality and compliance with the requirements of the Constitution of the Russian Federation and the legislation of the Russian Federation; the statement of claim in court from 33 residents of the Nizhniy Novgorod region "On recognition as invalid the Governor of the Nizhniy Novgorod region of March 13, 2020 № 27" On the introduction of the high readiness mode "(as amended by the Governor of the Nizhniy Novgorod region of November 24, 2021 № 212" On Amendments to the Governor's Decree of March 13, 2020 № 27") in terms of limiting the constitutional rights of citizens.
- Car rally against QR-codes and in support of the leader of QR-resistance Alexander Konovalov in St. Petersburg. On the eve of the rally group in the social network "VKontakte" was blocked by the demand of the prosecutor's office.
- Music marathon "QR Free", which was launched in order to unite musicians of all directions "to oppose segregation and imposition of electronic fascism" (quote from the group "QR-FREE" in social network "VKontakte"): Konstantin Kinchev, Anatoly Art, group "Black Griff", group CONCEPT VISION DEFECT and others.
- Organization of "antivaccine cab" in the Republic of Tatarstan for hitchhikers who do not have QR-code, the operators informed about the place and time of the trip in messenger.
- Preparing speeches of QR-code opponents at round tables, meetings, seminars with human rights activists, public and state figures, scientific community: round table organized by the Human Rights Ombudsman in Saint-Petersburg S. Y. Agapitova to discuss opinions on restriction of rights by QR codes system; meeting of civil activists with the Head of the Komi Republic on the issue of QR codes and vaccination in organizations and enterprises of the Republic; scientific and practical conference "Doctors for the Truth!", where issues of clinical and organizational aspects of diagnosis, treatment and prevention of COVID-19 were discussed.
- Performances by opponents of QR codes: activists in Kursk hanged a governor's mannequin on an elevated crosswalk; in St. Petersburg unknown activists made a snowman with the face of Deputy Prime Minister Tatyana Golikova, which had a QR code on it

A separate line can be distinguished protests related to the level of digital competence of the organizers. For example, cyberattacks on the digital assistant of the Gosuslugi website (chatbot Max).

According to the Ministry of Digital Development, Communications and Mass Media of the Russian Federation on November 10 and 11, 2021, the Gosuslugi website was subjected to the most powerful cyberattack of all time in its existence. The robot Max answered the users of the site to the question of how to get a QR-code, that there is no coronavirus and the fact of the disease is not proved by anyone: "As for QR-codes - they are a part of the world government plans to segregate the population and strengthen to-total digital control, as a result you as well will be worse off" (Fig. 5).
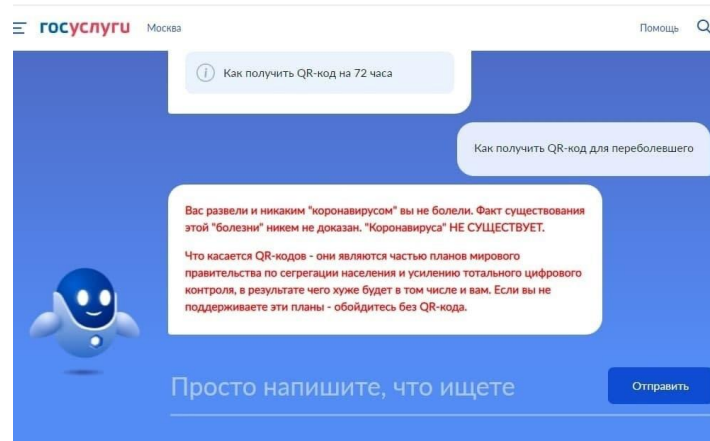

Figure 5: Cyberattack on Max's chatbot on the public services website

Online broadcast of the storming of the state power structures on November 19, 2021. Residents of Volgograd began to demand from the head of the department and its employees to show QR-codes and simultaneously filmed their actions on video for the purpose of further distribution in social networks (Fig. 6).
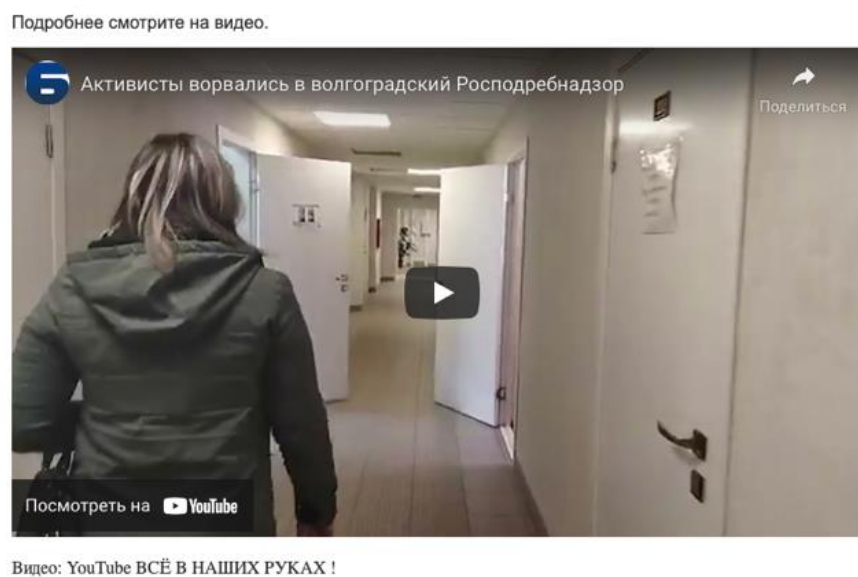

Figure 6: Video of Vologda activists' visit to the office of the regional government structure

The Instagram accounts of many heads of Russian regions were attacked. For example, on the page of the governor of the Rostov region, users left a large number of comments with demands to cancel the system of QR codes.

5. CONCLUSION

As a result of the application of the methodology, types of protest were identified that can be classified as classical (coordinated actions, pickets, distribution of leaflets, appeals to the President of

the Russian Federation, petitions, etc.), but also as new types of protest: organization of offline and online boycotts, collective lawsuits to the court, statements to the prosecutor's office against the actions of governors, car rallies, musical marathons, performances, preparation of performances of QR code opponents at round tables, meetings, seminars with human rights activists, public and state figures, scientific community, cyber-attacks on state chatbots and accounts in social networks of government representatives, etc. In the process of content analysis of the publications presented in the groups of the analyzed cluster, a separate type of protest against QR-codes was identified, which can be referred to hybrid, where the dominant topic "Protest against QR-codes" and "accompanying" topics are presented, for example: protests against electronic voting, new Federal state educational standard, pension reform, growth of tariffs on housing and communal services. Further study of the structures of online communities, where these types of protests are represented, will make it possible to analyze the structure of content, identify easily mobile (both vertically and horizontally) content and digital actors who promote their interests, using thematic trends in social networks.

6. Limitations

The presented work can be characterized as a pilot study, which was built around a single key "QR-code". In the future it is planned to deepen the study by forming a dataset based on a refined list of keywords on the topic under study, as well as to use the structure of user content as an object of research. The limitations of our methodology also include the fact that the API of the social network "VKontakte" allows to obtain no more than a thousand messages relating to a particular day.

7. Further work

The authors of the presented research plan to connect this theme with the topic of digital fraud, as the active digital transformation of educational, financial, state and other spheres has caused the active development of new forms of digital fraud. For example, the cases of fraud with QR codes of citizens at the State Services portal have been noticed and the cases of fraud in the banking sphere, when false employees of the bank call and ask to create a new QR code in order to protect them from getting an unauthorized loan, have also increased.

8. References

[1] Avtsinova, G. Innovative forms of youth protest activity. PolitBook 3, 15-26 (2014).
[2] Bastian, M., Heymann, S., Jacomy ,M. Gephi: an open-source software for exploring and manipulating networks. Proceedings of the international AAAI conference on web and social media 3(1), 361-362 (2009).
[3] Bojanowski, P. et al. Enriching word vectors with subword information. Transactions of the association for computational linguistics 5, 35-146 (2017).
[4] Goldberg, Y., Levy, O. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv preprint arXiv:1402.3722 (2014).
[5] Hagberg, A., Swart, P., Chult, D. Exploring network structure, dynamics, and function cusing NetworkX. – Los Alamos National Lab. (LANL), Los Alamos, NM (United States). – №. LA-UR-08-05495; LA-UR-08-5495, (2008).
[6] Legostaeva, N. Protest communities in the social network "VKontakte": types, topics, content of publications. Scientific Journal Discourse 7, 130-143 (2019).
[7] Nim E. "Toy rally": in search of the theory of mediatization of civil protest. Journal of Social Policy Research. 14(1), 55-70 (2016).
[8] Korobov, M. Morphological analyzer and generator for Russian and Ukrainian languages. International conference on analysis of images, social networks and texts. – Springer, Cham, 320-332 (2015).
[9] Mikolov, T. et al. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013).

[10] Pennington, J., Socher R., Manning C. D. Glove: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 1532-1543 (2014).

[11] Platonov, K., Legostaeva, N. Analysis of Russian Media Agreements on student protest on the basis of yandex news agregation data. PolitBook 3, 180-194 (2019).

[12] Platonov, K., Svetlov, K. Conspiracy Theories Dissemination on SNS Vkontakte: COVID-19 Case. International Conference on Electronic Governance and Open Society: Challenges in Eurasia, Springer, 322-335 (2020).

[13] Platonov K., Svetlov K. Politics-related Online Communities: Thematic Landscape and (Para) linguistic Features. 2021 28th Conference of Open Innovations Association (FRUCT) 358-364, (2021).

[14] Rehurek, R., Sojka, P. Software framework for topic modelling with large corpora. In Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks (2010).

[15] Rubenstein, H., Goodenough, J. B. Contextual correlates of synonymy. Communications of the ACM 8(10), 627-633 (1965).

[16] Svetlov, K., Legostaeva, N. Digital Transformation in the Russian Federation: Thematic Landscape of Online Communities. 2021 30th Conference of Open Innovations Association FRUCT, 285-291 (2021).

[17] Trieu, L. Q., Tran, H. Q., Tran, M. T. News classification from social media using twitter-based doc2vec model and automatic query expansion. Proceedings of the Eighth International Symposium on Information and Communication Technology, 460-467 (2017).

[18] Vanke, A. Internet communications as a means and a condition of political mobilization in Russia (on the example of the movement "For Fair Elections" / A. Vanke, I. Ksenofontova, I. Tartakovskaya. INTER 7, 45-73 (2014).