Digital Transformation in the Russian Federation: Thematic Landscape of Online Communities

Kirill Svetlov, Natalya Legostaeva
Saint Petersburg State University
Saint Petersburg, Russia
{k.svetlov, n.legostaeva}@spbu.ru

***Abstract*—The digital transformation of the economic, socio-political, and cultural spheres of Russian society is being actively discussed both at the state level by the country's top officials and by users of social networks. Proponents and opponents of digital technology discuss the positives and negatives of digitalization. What areas of digital transformation are of most interest to users of the popular Russian social network VKontakte? What new structures or processes related to the development of the digital society activate people to unite into online communities and develop online communication? In this study, the authors used the doc2vec model and constructed a graph of VKontakte social network communities related to digital transformation. This graph consists of 368 thematic clusters, which include online communities. The analysis of the content of the communities showed the most popular topics related to digital transformation in the Russian Federation. The proposed algorithm for searching and analyzing online communities on a wide range of topics allows for extensive scientific piloting with the prospect of intensifying scientific analysis on specific topics.**

I. Introduction

Since 2008, federal and national projects for the development of the information society have been actively implemented in the Russian Federation. One of the strategic documents is the Presidential Decree of 07.05.2018 "On the national goals and strategic objectives of the development of the Russian Federation until 2024" [1]. In this document, the national projects are grouped into three blocks: "Human Capital", "Comfortable Living Environment", and "Economic Growth". The "Economic Growth" block includes the "Digital Economy" national project, under which such federal projects as "Human Resources for the Digital Economy," "Information Security," "Digital Technology," "Digital Public Administration," and others have been developed. These federal projects include goals for the development of higher education programs in information technology, online programs to develop digital literacy, the introduction of elements of the model "Digital University" in universities, etc. Due to the active development of both television and Internet debates (on the platform "YouTube") on the topic of "Digital Transformation", the authors of this paper conducted a pilot study on the search for online communities in the popular Russian social network VKontakte, whose participants discuss the processes of digitalization in the online format. An analysis of the content of publications (posts) in the identified communities helped to describe popular topics related to digitalization (posts with high engagement metrics), as well as to identify new educational and business structures (for example, startup schools, technology parks, business incubators, etc.) created as part of the implementation of the national project "Digital Economy".

The paper consists of the following sections. In Related work we make an overview of foreign and domestic research in the field of digitalization of various areas of life. In Research question we formulate the purpose, object, subject, and main research questions. In Methodology we present the algorithm for collecting and processing the data set and describe the models that were used to construct the undirected graph. In Results we describe the thematic landscape and statistical description of online community clusters dedicated to digital transformation, and analyze engagement metrics of the VKontakte social network by highlighted community clusters. In Discussion we compare the results of the different engagement metrics. In Conclusion we summarize the results, discuss the limitations of the study, and present plans for future work. The study ends with acknowledgements and a list of references.

II. Related work

In the scientific literature, most studies are related to the study of digitalization of certain spheres of social life. For example, digitalization in the economic sphere is associated with the development of e-trade, e-banking, e-marketing [2], [3]. The digitalization of business explores such areas as new business models, online stores, online services, online marketing [4], [5]. Many researchers study the digital transformation of the government through the prism of the development of e-democracy, e-government, e-parliament, e-election, e-voting, e-petitioning, e-participation, secure E-Services [6]–[14]. Digitalization of the social sphere is related to the introduction of digital technology in education [15]–[17] and health [18]–[20]. Other studies focus both on the positive effects of digitalization, for example, at the level of society as a whole (improved quality of life, development of new forms of business, increased transparency of economic transactions, etc.), at the level of individual companies and industries (elimination of intermediaries, creation of new products and services, etc.), and on the risks and threats of digital transformation (cybercrime, reduction of jobs, cryptocurrency market risks, etc.) [21]–[23]. The uniqueness of this work lies in the fact that the authors did not limit the study to the analysis of online communities, whose participants discuss the process of

digitalization in a particular area of social interaction, from the perspective of a narrow understanding of digitalization (development of medicine, science, education, transportation, new industrialization, state regulation of the economy and planning, etc.) [24], but chose as the starting point of the study a broad understanding of digital transformation.

In terms of the machine learning methods used in our research, the following publications can be mentioned. In [25] the authors study the political landscape of the social network Vkontakte using topic modeling techniques, in particular ARTM. The use of the doc2vec model in the analysis of the social network Twitter is considered in [26]. The list of articles [27] on the construction of embeddings for various objects, in particular the tags assigned to documents [28], is also worth mentioning.

III. Research question

In this paper, the authors study the online communities of the social network VKontakte, whose participants discuss topics related to digitalization in the Russian Federation. The aim of the study is to describe the thematic landscape of online communities whose participants discuss digital transformation in the Russian Federation. The object of the research is the process of digital transformation, the subject of the analysis are online communities whose participants discuss digitalization. We submit the following research questions:

1) What thematic areas do online communities dedicated to the processes of digital transformation have?
2) How do the values of user engagement metrics vary in these areas?

IV. Methodology

To build a list $GroupList_1$ of communities on the specified topic we used the API of the VKontakte social network, which allows obtaining a relevant set of groups, as well as information about them (a short description and the number of members in a given group) for the given search queries formed in the Russian language. The search queries set $Query_1$ contained the following search phrases: digitalization, digital transformation, digital development, digital business, digital economy, risks of digitalization, challenges of digitalization, digital economy workforce (translated from Russian into English), etc. The resulting list of groups $GroupList_1$ included 8769 communities. Communities representing various digital equipment maintenance companies, digital printing service providers, etc. were excluded from this list. Additionally, communities whose information is private and available only to their members were excluded from the list. The filtered list $GroupList_2$ included 2194 groups. For each community from $GroupList_2$, we used VK API to retrieve 100 (or less, if their total number is less than 100) messages in the Russian language posted in it and available for viewing by all social network users. The resulting message database DB_1 consisted of 139 642 posts. 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 “ë” with “e”;
5) Replacement of links to web resources and user references with special ”URL” and ”USR” tokens, respectively;
6) Word normalization using the pymorphy2 library [29].

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 [30], FasText [31], GloVe [32]. These models are based on the assumption, which in linguistics is called the distributional hypothesis [33], 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 d is the dimension of the word embeddings. The elements of W are chosen to solve this prediction problem in the best way [34], 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 $p_\theta((w, c) \in D) = \sigma(v_w \cdot v_c)$, 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

$$\underset{\theta}{\operatorname{argmax}} \prod_{(w,c)\in D} p_\theta((w, c) \in D) \prod_{(w',c)\in D'} (1-p_\theta((w', c) \in D)),$$

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 [35], 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

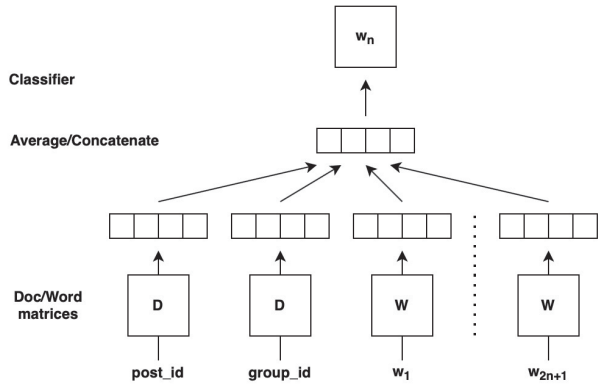

Fig. 1. The architecture of the doc2vec model (PV-DM) used in the research

implemented in the Gensim library [36]. 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. In this paper we use an implementation of the doc2vec model in the form of PV-DM, as the original paper [35] noted its advantage over the PV-DBOW version. The dimension of the embedding space was chosen equal to 128. 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

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

where v_{c_1} and v_{c_2} are the embeddings obtained for communities c_1 and c_2, $v_{c_1} \cdot v_{c_2} = \sum\limits_i v_{c_1,i} \cdot v_{c_2,i}$ – the dot product of these embeddings, and $\|v_{c_i}\| = \sqrt{\sum\limits_i v_{c_1,i}^2}$, $i = 1, 2$ – 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 [37]. 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 0.62, which corresponds to 0.75 of the sample quantile estimated from the calculated cosine distances). The resulting graph has the following characteristics:

1) Number of vertices: 2194;
2) Number of edges: 70 466;
3) Average degree: 53.44.

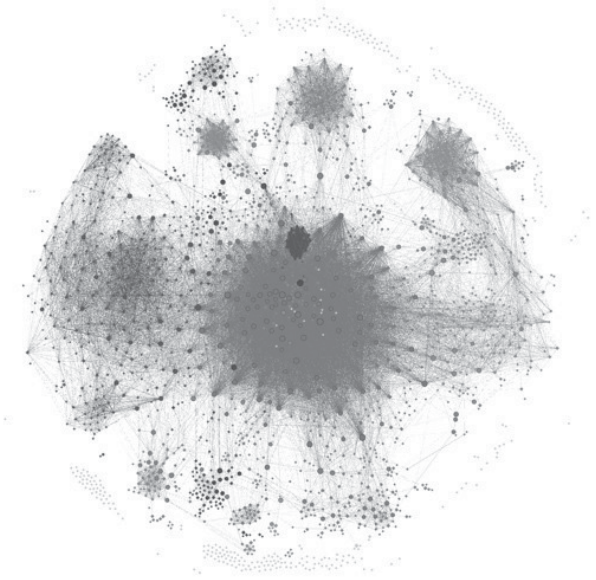

Fig. 2. Graph of the VKontakte social network communities, whose participants discuss digital transformation

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) [38], the communities were divided into 368 clusters (Fig. 2). Some of these clusters were of size 1, i.e. consisted of a single community. In the following analysis of the thematic landscape we will consider clusters containing at least 10 online communities.

V. Results

A. Thematic landscape and statistical description of online community clusters

Using this methodology allowed the authors to analyze 14 clusters (Tab. I) of online communities, with a total number of groups in each cluster ≥ 10, whose participants discuss the topic of digitalization.

A brief overview of clusters:

- Cluster 1. "Digital Economy": the number of groups in the cluster - 387, the number of participants – 348 824. Cluster participants discuss the digitalization of small and medium-sized businesses, startups for IT entrepreneurs, digital marketing, developments in digitalization and workplace automation.
- Cluster 2. "Digitalization of school education": the number of groups in the cluster - 290, the number of participants – 116 977. Cluster participants discuss teaching

TABLE I. THEMATIC CLUSTERS OF ONLINE COMMUNITIES

Cluster Id	Thematic focus of the cluster	Number of groups	Number of participants
1	Digital economy	387	348 824
2	Digitalization of school education	290	116 977
3	Cryptocurrency	266	1 829 363
4	IT companies	171	710 186
5	University business incubators	143	262 557
6	Government projects and structures related to digitalization	135	163 342
7	Small business incubators	119	83 942
8	Digitalization of higher education	108	141 202
9	Trainings on digitalization of business	80	341 216
10	Opponents of digitalization	42	27 537
11	Business digitalization trainings for young people (”Business Youth” online communities)	34	2 059
12	Innovative startups (”Legacy 21. Startup Show” online communities)	17	3 933
13	Investments	12	20 668
14	Co-working to create social and business projects (online communities ”New Space”)	10	4178

children and teenagers to create digital projects (websites, games, apps), and the development of additional education in the field of modern information technology.

- Cluster 3. ”Cryptocurrency”: the number of groups in the cluster - 266, the number of participants – 1 829 363. Participants in the cluster groups discuss crypto-assets, exchangers for buying and selling cryptocurrencies, crypto pawnshops, and the development of private investments.
- Cluster 4. ”IT companies”: the number of groups in the cluster - 171, the number of participants - 710 186. Cluster topics: modeling of complex technical systems, machine learning, big data analysis, security of information systems and technologies.
- Cluster 5. ”University business incubators”: the number of groups in the cluster - 143, the number of participants – 262 557. Cluster participants discuss the development of innovative businesses, training of innovation managers, online educational programs, web-quests.
- Cluster 6. ”Government projects and structures related to digitalization”: the number of groups in the cluster - 135, the number of participants – 163 342. Cluster participants discuss the development of international cooperation programs in information technology and programs for the development of digital entrepreneurship.
- Cluster 7. ”Small business incubators”: the number of groups in the cluster - 119, the number of participants – 83 942. Participants in cluster groups discuss support programs for current and future entrepreneurs (preferential rent, free consultations), trainings, seminars, webinars, forums, etc.
- Cluster 8. ”Digitalization of higher education”: the number of groups in the cluster - 108, the number of participants - 141 202. Cluster topics: the introduction of digital technologies in universities, the opening of graduate and postgraduate programs in innovation management and digital transformation.
- Cluster 9. ”Trainings on digitalization of business”: the number of groups in the cluster - 80, the number of participants - 341 216. Participants in the cluster groups discuss educational programs for business development and investor search.
- Cluster 10. ”Opponents of digitalization”: the number of groups in the cluster - 42, the number of participants – 27 537. Cluster members oppose the digital economy, electronic passports, and digital government.
- Cluster 11. ”Business digitalization trainings for young people (”Business Youth” online communities)”: the number of groups in the cluster - 34, the number of participants – 2 059. Cluster topics: small and medium business development, mentoring.
- Cluster 12. ” Innovative startups (”Legacy 21. Startup Show” online communities)”: the number of groups in the cluster - 17, the number of participants – 3 933. Cluster participants discuss trainings/workshops on business creation and acting, meetings with mentors.
- Cluster 13: ”Investments”: the number of groups in the cluster - 12, the number of participants – 20 668. Cluster topics: venture-engineering and social startups, cryptocurrency, blockchain, new types of transport and transport and logistics infrastructure.
- Cluster 14: ” Co-working to create social and business projects (online communities ”New Space”)”: the number of groups in the cluster - 10, the number of participants - 4 178. Participants of groups in the cluster discuss the development of social and business projects, forums, intensives, project creative sessions.

Analysis of the content of online communities allowed to identify the thematic focus of clusters.

B. Engagement metrics

The main characteristics of the popularity of posts in online communities are views, likes, comments, and reposts. Since large online communities with a large number of members can achieve publication popularity due to the size of their audience, in order to compare the popularity of posts in different groups researchers usually use relative measures that show how well views convert into likes, comments and reposts. The main

metric is the engagement rate (ER), calculated separately for each post

$$ER = \frac{Likes + Comments}{Views} \cdot 100 \ .$$

In this study, the authors use this metric when evaluating the popularity of posts. Engagement rates for likes, comments and reposts were also calculated.

$$ER_M = \frac{M}{Views} \cdot 100 \ .$$

where $M \in \{Likes, Comments, Reposts\}$. The resulting engagement rates for individual posts were averaged within online community clusters (Tab. II).

Thus, the first place according to the engagement rate by likes and comments is taken by the cluster "Investment" – 73.69, the second position was taken by the cluster "Innovative startups ("Legacy 21. Startup Show" online communities)" – 43.93. The third place was taken by the cluster "Co-working to create social and business projects (online communities "New Space")" – 41.90, in the fourth place the cluster "Cryptocurrency" – 41.30, on the fifth position is the cluster "Opponents of digitalization" – 40.60.

VI. Discussion

An analysis of the engagement coefficients for likes and comments shows that these parameters have a positive correlation (Fig. 3). This result is significant since these indicators often have an inverse relationship: messages that are approved by users of social networks have a high engagement rate by likes with a low engagement rate by comments. In contrast, posts on topics that provoke controversy among users have a low engagement rate by likes, but a high engagement rate by comments. According to the engagement rate by likes and comments, popular topics are related to investments, innovative startups, social and business projects, cryptocurrencies, and digital opponents. The result related to the engagement rate by reposts is interesting. According to the engagement rate by reposts the first place is taken by the cluster "Co-working to create social and business projects (online communities "New Space")" – 8.46, the second position was taken by the cluster "Opponents of digitalization" – 7.22. The third place was taken by the cluster "Digital economy" – 5.75, in the fourth place is the cluster "Cryptocurrency" – 4.21, on the fifth position is the cluster "Investments" – 5.16. Despite the fact that the popularity rating of topics differs by the engagement rate by likes and comments and the engagement rate by reposts, the authors of the study concluded that the topic of digital transformation is of high interest to users of the social network "VKontakte". This conclusion is confirmed by a positive correlation of engagement by likes and comments. Further development of the engagement rate metric will allow a more in-depth analysis of the digital communications of online community participants.

VII. Conclusion

Using the doc2vec model, which is based on the distributional hypothesis, the authors constructed a graph of VKontakte social network communities related to the topic of digital transformation. There are 368 clusters in this graph, 14 of which the authors analyzed for thematic focus, filtering them by the number of groups in the cluster (> 10 groups). An analysis of the content of posts in 14 clusters made it possible to see the most discussed thematic areas on the topic of "Digital Transformation": digitalization of the economy, digitalization of education, digitalization of business. Separately, we can identify clusters whose participants discuss digital currency, investment, and new IT companies. The authors of the study found a cluster whose participants are opponents of digitalization. As a result of applying this methodology to analyze the thematic landscape of online communities, the authors identified new structures, including digital ones, which were created as part of the national project "Digital Economy": a business incubator IT park, an online school of technological entrepreneurs, a platform "Digital pedagogical internship", a platform "Startup show for teens", virtual business incubator ideas, digital innovation studios, IT academies for school children. An interesting result was obtained based on the calculation of engagement rate. Thus, the "Cryptocurrency" cluster with a maximum number of participants (1 829 363) takes only the fourth position by the popularity of the content (engagement rate – 41.30), while the "Investments" and "Innovative startups ("Legacy 21. Startup Show" online communities)" clusters take first and second places with a small number of participants in the groups: 20 668 (engagement rate – 73.69) and 3 933 (engagement rate – 43.93), respectively.

VIII. Limitations and Further work

The limitations of the study include the fact that the analysis was based only on the social network "VKontakte". This social network is one of the most popular in Russia, but discussions about digitalization take place in other social networks (Facebook, Instagram, etc.). Also, for each community in $GroupList_2$, the last 100 messages were obtained, according to the time of publication. More data can expand the thematic landscape of online communities and form a larger number of clusters.

We analyzed clusters containing strictly more than 10 online communities. In future studies, we plan to reduce this limitation, as clusters with fewer groups may have other thematic areas related to the topic of digitalization, as well as more popular posts by the engagement rate. The authors also plan to detail the work on the formation of $GroupList_1$ and DB_1 in terms of the selection of search queries and the periods of data download.

Acknowledgment

The study was conducted within the framework of the project No. 121062300141-5 "Comprehensive study of factors

TABLE II. USER ENGAGEMENT METRICS FOR SELECTED THEMATIC CLUSTERS

Cluster Id	Thematic focus of the cluster	ER	ER_L	ER_C	ER_R
1	Digital economy	36.63	35.02	1.61	5.75
2	Digitalization of school education	31.74	30.64	1.10	2.51
3	Cryptocurrency	41.30	39.20	2.10	4.21
4	IT companies	25.17	23.68	1.49	1.96
5	University business incubators	16.50	15.43	1.07	1.97
6	Government projects and structures related to digitalization	28.96	28.22	0.74	2.74
7	Small business incubators	28.49	27.47	1.02	2.68
8	Digitalization of higher education	19.64	18.85	0.79	2.98
9	Trainings on digitalization of business	25.73	24.61	1.12	1.56
10	Opponents of digitalization	40.60	37.61	2.99	7.22
11	Business digitalization trainings for young people (”Business Youth” online communities)	1.83	1.70	0.13	0.04
12	Innovative startups (”Legacy 21. Startup Show” online communities)	43.93	43.73	0.21	0.55
13	Investments	73.69	71.80	1.89	5.16
14	Co-working to create social and business projects (online communities ”New Space”)	41.90	40.71	1.19	8.46

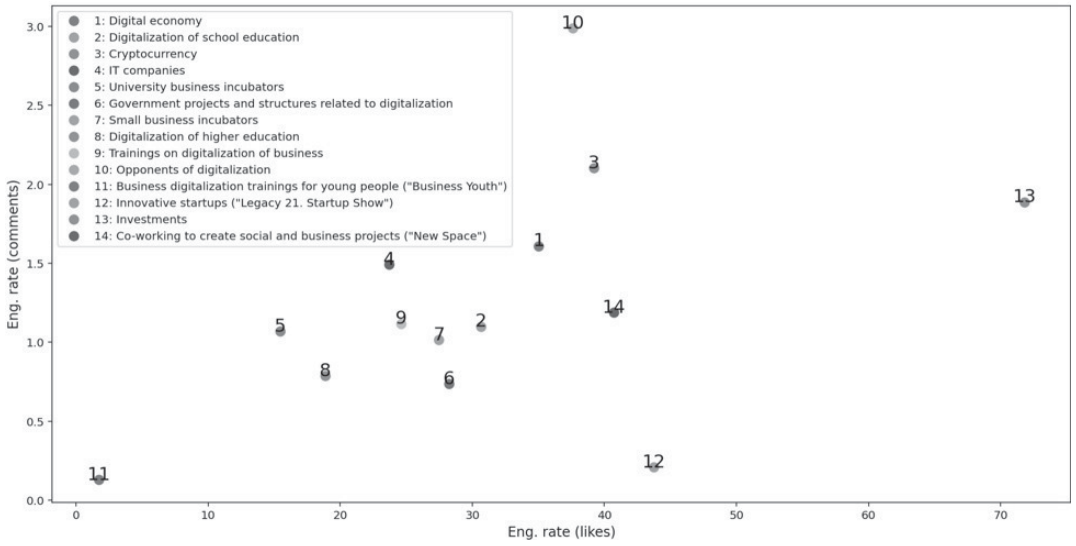

Fig. 3. The relationship between the engagement rate by likes and comments

and mechanisms of political and socio-economic sustainability in the transition to a digital society”.

REFERENCES

[1] Decree of the President of the Russian Federation of 07.05.2018 No 204 ”On the national goals and strategic objectives of the development of the Russian Federation until 2024”. [Online]. Available: http://www.kremlin.ru/acts/bank/43027 (In Russian).

[2] T. Ivanova, Y. Kochkova. Features of marketing in social media in the conditions of digitalization. *Ekonomicheskij zhurnal*, vol. 3, no. 51, pp. 67-76, 2018. (In Russian).

[3] A.N. Miroshnik, I.A. Khustochkin. Media-technologies in functioning of modern banking system of Russian Federation: digitalization of banking services. *Ekonomika: teoriya i praktika*, vol. 2, pp. 61-65, 2019. (In Russian).

[4] J. Naskali et al. Mapping Business Transformation in Digital Landscape: A Prescriptive Maturity Model for Small Enterprises. *International Conference on Well-Being in the Information Society*. Springer, Cham, pp. 101-116, 2018, July.

[5] H. Bouwman et al. The impact of digitalization on business models. *Digital Policy, Regulation and Governance*, Emerald Publishing Limited, vol. 20, no. 2, pp. 105-124, 2018, March.

[6] A.L. Shchur, A.M. Fedorov, I.O. Datiev. Digitalization in State and Municipal Governance: World Experience, Problems and Prospects of Using Online Social Networks. *Trudy Kol'skogo nauchnogo centra RAN*, vol. 10, no. 9, pp. 158-171, 2019. (In Russian).

[7] S. Khan, N.Z.A. Rahim, N. Maarop. A review on antecedents of citizen's trust in government social media services. *3C Tecnología. Glosas De Innovación Aplicadas a La Pyme*, pp. 109-120, 2019.

[8] A.E. Kon'kov. Digitalization of Political Relations: Frontiers of Perception and Mechanisms of Transformation. *Kontury global'nyh transformacij: politika, ekonomika, parvo*, vol. 12, no. 6, pp. 6-28, 2019. (In Russian).

[9] C.J. Cohen, J. Kahne. *Participatory politics. New media and youth political action.* Oakland: YPP Research Network 43 p., 2011.

[10] J. Kahne, E. Middaugh, D. Allen. 2 Youth, New Media, and the Rise of Participatory Politics. *From voice to influence*, University of Chicago Press, pp. 35-56, 2021.

[11] M. Solvak, K. Vassil. Could Internet Voting Halt Declining Electoral Turnout? New Evidence That E-Voting Is Habit Forming. *Policy & Internet*, vol. 10, no. 1, pp. 4-21, 2018.

[12] S.A.R. Barros, R.C. Sampaio. Do Citizens Trust Electronic Participatory Budgeting? Public Expression in Online Forums as an Evaluation Method in Belo Horizonte. *Policy & Internet*, vol. 8, no. 3, pp. 292–312, 2016.

[13] A. Makarov, A. Shabunin. On design of secure E-Services for public authority in the russian federation. *20th Conference of Open Innovations Association (FRUCT)*, pp. 250-257, 2017, April.

[14] A. Mitrakas, P. Hengeveld, D. Polemi, J. Gamper. Secure E-Government Web Services, *IGI Global*, 2007.

[15] T.V. Nikulina, E.B. Starichenko. Informatization and digitalization of education: concepts, technologies, management. *Pedagogicheskoe obrazovanie v Rossii*, vol. 8, pp. 107-113, 2018. (In Russian).
[16] N.P. Petrova, G.A. Bondareva. Digitalization and digital technologies in education. *Mir nauki, kul'tury, obrazovaniya*, vol. 5, no. 78, pp. 353–355, 2019. (In Russian).
[17] A.V. Morozov. Professional training of educational leaders using modern digital technologies. *Chelovek i obrazovanie*, vol. 4, no. 57, pp. 105–110, 2018. (In Russian).
[18] P. Näkki et al. *Social media for citizen participation: report on the Somus project*, VTT Technical Research Centre of Finland, 131 p., 2011.
[19] N.V. Zabolotnaya, I.N. Gatilova, A.T. Zabolotnyi Digitalization of health care: achievements and prospects for development. *Ekonomika. Informatika*, vol. 47, no. 2, pp. 380-389, 2020. (In Russian).
[20] A.V. Egorova. Digitalization of Regional Healthcare System in the Digital Economy. *Nauchnyj ezhegodnik Centra analiza i prognozirovaniya*, vol. 1, pp. 256-261, 2018.
[21] V.G. Khalin, G.V. Chernova. Digitalization and its Impact on Russian Economy and Society: Benefits, Challenges, Threats and Risks. *Upravlencheskoe konsultirovanie*, vol. 10, pp. 46-63, 2018. (In Russian).
[22] A.P. Kochetkov et al. The Concept of Digital Government as a Political Project for Russia: Prospects for Implementation in the Context of Challenges and Risks of Digitalization of Society. *Vlast*, vol. 1, pp. 316-331, 2021. (In Russian).
[23] A.A. Strokov. Digitalization of Education: Problems and Prospects. *Vestnik mininskogo universiteta*, vol. 2, no. 31, p. 15, 2020. (In Russian).
[24] T.N. Yudina. Digitalization as a trend of modern economic development of the Russian Federation: pro y contra. *Gosudarstvennoe i municipalnoe upravlenie uchenye zapiski*, vol. 3, pp. 139-143, 2017. (In Russian).
[25] K. Platonov, K. Svetlov. Politics-related Online Communities: Thematic Landscape and (Para) linguistic Features. *28th Conference of Open Innovations Association (FRUCT)*, IEEE, 2021.
[26] L. Q. Trieu , H. Q. Tran, M. T. Tran. 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*, pp. 460-467, 2017.
[27] Awesome-2vec, curated list of 2vec-type embedding models. [Online]. Available: https://github.com/MaxwellRebo/awesome-2vec .
[28] S. Chen, et al. DocTag2Vec: An Embedding Based Multi-label Learning Approach for Document Tagging. *Proceedings of the 2nd Workshop on Representation Learning for NLP*, 2017.
[29] M. Korobov. Morphological analyzer and generator for Russian and Ukrainian languages. In International Conference on Analysis of Images, *Social Networks and Texts*, Springer, Cham, pp. 320-332, 2015, April.
[30] T. Mikolov, K. Chen, G. Corrado, J.Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv*:1301.3781, 2013.
[31] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135-146, 2017.
[32] J. Pennington, R. Socher, C.D. Manning. Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532-1543, 2014, October.
[33] H. Rubenstein, J.B. Goodenough. Contextual correlates of synonymy. *Communications of the ACM*, vol. 8, no. 10, pp. 627-633, 1965.
[34] Y. Goldberg, O. Levy. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv*:1402.3722, 2014.
[35] Q.Le, T. Mikolov. Distributed representations of sentences and documents. *International conference on machine learning*, pp.1188-1196, 2014.
[36] R. Rehurek, P. Sojka. Software framework for topic modelling with large corpora. *In Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks*,2010.
[37] A. Hagberg, P. Swart, S Chult, D. *Exploring network structure, dynamics, and function using NetworkX*, Los Alamos, NM (United States), 2008.
[38] M. Bastian, S. Heymann, M. Jacomy. Gephi: an open source software for exploring and manipulating networks. *Third international AAAI conference on weblogs and social media*, pp. 361-362, 2009, March.