Politics-related Online Communities: Thematic Landscape and (Para)linguistic Features

Konstantin Platonov
Center for Sociological
and Internet Research
Saint Petersburg State University
Saint Petersburg, Russia
k.platonov@spbu.ru

Kirill Svetlov
Laboratory for research of social-economic
and political processes of modern society
Saint Petersburg State University
Saint Petersburg, Russia
k.svetlov@spbu.ru

***Abstract*—Politics-oriented online communities on social networking sites are becoming more influential in the conditions of lack of the independent media agenda. But how can the fragmented landscape of topics and ideologies represented by them be described using (para)linguistic features? In this research, we consider the sample of politics-related communities from the most popular Russian social networking site VKontakte. Additive regularization algorithm allowed us to classify the communities with similar topics and describe 9 salient clusters. It was shown how the thematic focus of the particular clusters is framed by linguistic and paralinguistic features, such as morphological specificities, tone, punctuation and so on. We propose a framework and a pipeline for the analysis of communicative strategies of communities discussing politics which could be used as a preliminary stage for the qualitative content analysis.**

I. Introduction

Online communities have a significant impact not only on agenda, but also on the political life of society as a whole. They are widely used for political mobilization [1], petition promotion [2], movement information support [1], as platforms for protest [3] and for many other political purposes. Politically-oriented online communities have been studied with use of the data from different social networks such as Twitter [4], Facebook [5], Reddit [6], LiveJournal [7].

Explanatory models of online communities have been considered in the context of social identity theory and social bond theory [8], also in theories of social trust and influence [9]. Today, online communities are the sources of alternative agenda that becomes of a great demand especially in the case of propagation of politics-related content. In the research related to agenda-setting theory [10] and echo chamber effect [11], the opportunities for creating particular agendas were analyzed. Agenda-setting is the function of media which is related to building a certain agenda by using covering particular topics (issues) and their characteristics (attributes) [12].

The research based on big data analysis shows not only that the public agenda is indeed reflected in the social media, but also that online discussion is more focused on the most topical and sensitive social problems, for instance abortion, as well as drugs and guns [13]. Obtained discrepancies between the representations of current events in various media inevitably lead to fragmentation of agenda [14] which can provoke polarization of opinions and an emergence of echo chambers [15].

In general, echo chambers are the structures of communication arising from «people's exposure to pro-attitudinal communication» [16]. Popular communities and their clusters connect people with similar views, opinions and political preferences, providing certain agenda to them. This process causes echo chamber effect which, on the one hand, can lead to decrease in quality and diversity of online discourse [17], and, on the other hand, intensifies opinion polarization [18], proliferation of fake news [19] and populist statements [16].

II. Related work

A. Political communicative strategy online

Social media are still efficient as a political marketing channel [20]. Online platforms, in a way, levelled the odds of wide-ranging political actors of making themselves heard. Even some marginal groups gained the opportunities to make statements with a huge reach [21]. This required politicians and movements to develop new communicative strategies that could be competitive in current conditions [22].

Marketing, mobilization and dialogue with users are among the most important motives for presence and publication activity online [23]. Different political actors choose various strategies according to the balance of these motives. These strategies are efficient in the context of the needs of particular audiences. For instance, Lilleker & Koc-Michalska analyzed online activity of the members of European Parliament and distinguished three such strategies: homestyle, impression management and participatory style [24]. Another classification includes informing, engaging, mobilizing and interacting strategies [25]. There are numerous classifications but the majority of them are based on two dimensions, namely, the actors differ in *what they write about* and *how they write about it* [26].

Thus, in the context of political communication, the ingredients of an efficient SMM (Social Media Marketing) strategy include not only the particular topics and their qualities (issues and attributes in terms of agenda-setting theory), but also lexical and stylistic choices on the level of texts. Communicative behavior of political actors forms their "lingual biography" [27]. In turn, perception of politics-related texts depends on linguistic and stylistic parameters [28], [29].

B. Politics-related online communities

According to Rheingold, virtual communities are social aggregations assembled by online communication of large number of people during a public discussion [30]. The term of political online communities received varying definitions in different research. On the one hand, the concept of community is represented at various scales from metacommunity, the complex of discussions in online space, to local cliques. On the other hand, the concept of «political» is also unclear applying to communities because various ideas could be behind it such as the link between group and ideology or movement, presence of politicized (communicative) goals or discourse.

The present paper focuses on communities which will be further referred to as "politics-oriented", rather than "political" or "politicized". Thus, we emphasized the primacy of thematic affiliation of these communities with political life rather than highlight their political aims. This approach affected the procedures of our sampling.

Although strategies can include implementation of various content types (e.g. photo and video), we assume that the key source of information about communicative strategies (at least on SNS VKontakte) is contained in texts published on behalf of online communities. These texts, their topics (primarily), word choice and other texts' features (secondly) characterize communities' discourse.

The online community on VKontakte is a special case of community with strict boundaries. Each group on this site has a screen name, a number of subscribers and a public wall page. The community agenda is given by a set of posts on its wall. For each post there are a number of likes, reposts, views and comments. VKontakte also has private communities, but we do not consider them in this study because access is limited and content reach is relatively low.

The methodology of this study is based on the assumption that allows to consider a community as an impersonal unit of observation, a quasi-subject involved in strategic communication, in other words, "not random or unintentional communication" [31], but "conscious" broadcasting of a certain agenda.

III. Research question

In this research, we investigate politics-oriented online communities presented in SNS (Social Networking Service) VKontakte (vk.com). The purpose of the study is to describe the main communicative strategies of these communities based on computational thematic and linguistic analysis. Since our sample is a selection of a number of politics-oriented online communities, we consider this analysis as a case study. For this reason, instead of applying the existing classifications, here we attempt to generalize the empirical data to specify strategies and describe them.

RQ. How the key communication strategies of popular politics-oriented online communities on SNS VKontakte can be described by using the texts of posts?

IV. Methodology

For the pre-processing of texts, we used MyStem lemmatizer [32], with the help of which tokens in posts were normalized to basic forms.

On the first stage of our analysis we built a topic model for the text corpus using BigARTM library [33], which implements the additive regularization algorithm [34]. From the mathematical point of view, the task of thematic modeling is the task of factorization of the term-document P matrix into the product of the Φ term-matrix and the Θ document-matrix.

Let D be a collection of documents and W be a set of tokens (words, n-grams) found in the documents. Matrices Φ and Θ must be restored from matrix P [35]. This objective is achieved as a solution to the optimization task of maximizing the logarithm of the likelihood function and regularizes $R_i(\Phi, \Theta)$

$$\sum_{d \in D} \sum_{w \in d} n_{dw} \ln \sum_{t \in T} \varphi_{wt} \theta_{td} + \sum_{1 \le i \le r} \tau_i R_i(\Phi, \Theta) \underset{\Phi, \Theta}{\longrightarrow} max$$

with restrictions on negativity and normalization of matrices Φ and Θ. Variable n_{dw} denotes how many times the word w meets in the document $d \in D$.

The above mentioned approach with $R_i(\Phi, \Theta) \equiv 0$ corresponds to probabilistic latent semantic analysis (PLSA) [36]. By adding different regulators to the likelihood function, it is possible to set different properties of the resulting model, including behavior identical to the common LDA approach [37].

In our analysis, we focused on PLSA, as the most appropriate approach, by setting the number of T (number of topics) to 200. Then, having obtained the matrix Θ, we averaged its columns corresponding to different documents d, placed in the given community c. The resulting matrix is $\mathrm{H} = \{\eta_{tc}\}_{t,c}$, its size is equal to $|T| \times |\mathrm{C}|$, where C is the number of unique communities in the given post corpus.

$$\eta_{tc} = \frac{\sum_{d \in c} \theta_{td}}{\sum_{d \in c} 1}$$

The resulting matrix H also has the property that the sum of elements in any of its columns equals 1. Any column c itself can be interpreted as a vector of thematic orientation of the corresponding community. Obtaining such vector embedding for our communities allows us to compare how similar their topics are, using, for example, a cosine similarity.

$$cosine\_similarity(c_1, c_2) = \frac{\sum_{t \in T} \eta_{tc_1} \cdot \eta_{tc_2}}{\sqrt{\sum_{t \in T} \eta_{tc_1}^2} \cdot \sqrt{\sum_{t \in T} \eta_{tc_2}^2}}$$

This measure is the value of the cosine of angle between vectors $\eta_{\cdot c_1}$ and $\eta_{\cdot c_2}$ in the Euclidean space $R^{|T|} = R^{200}$. Communities with similar topic distributions will have this indicator close to 1 and, conversely, communities with completely different topics will have this indicator close to 0. Using this metric, we have constructed a graph of politics-related communities. The nodes in it are the particular communities, and the edges connect the pairs of communities, in which the value of this metric is higher than a certain threshold value. We have chosen a threshold value equal to $\sqrt{2}/2$. As a result, all the communities with angles between topic vectors does not exceed $\pi/4$ radians became connected.

The resulting graph was further analyzed in Gephi [38]. Using the algorithm of community clustering, communities from Vkontakte were divided into 9 clusters via modularity maximizing algorithm (the Resolution parameter was selected as 0.9).

Having determined the clusters of politics-related communities, the analysis of text features was conducted. This procedure consists of calculating a number of metrics.

1. Formal morphological metrics of texts.
2. Tone of the texts.

3. Metrics of user engagement.
4. Other metrics such as the presence of hashtags, emoji, etc.

An interesting result concerning the study of publications was the basic graphematic analysis of texts from our dataset. For each post we have calculated the number of words and sentences in it, as well as the following indicators borrowed from mathematical linguistics [39].

$$objectiveness = \frac{NOUN + PRO}{ADJ + VERB},$$

$$qualitativeness = \frac{ADJ + ADVB}{NOUN + VERB},$$

$$activeness = \frac{VERB}{n_w},$$

$$dynamism = \frac{VERB}{NOUN + ADJ + PRO},$$

$$coherence = \frac{CONJ}{n_s},$$

where $NOUN$ is the number of nouns in a post, ADJ is the number of adjectives, $VERB$ is the number of verbs and participles, PRO is the number of pronouns, $ADVB$ is the number of adverbs, $CONJ$ is the number of prepositions and conjunctions, n_s is the number of sentences in a text and n_w is the number of words in a text.

Indicators were calculated using the morphological analyzer pymorphy2 [40]. The values of these indicators can give a basic understanding of the author's style and also what syntactic constructions were used. As a rule, high value of Objectiveness indicator is a characteristic of more formal writing style, where actors are clearly indicated. Qualitativeness metric characterizing the degree of concretization by adjectives and adverbs of nouns and verbs is also a characteristic of more formal style, for example, scientific articles. The values of Activeness and Dynamism are formal expressions of how much "action" occurs in the text, whether its narrative is dynamic or, conversely, statically describes phenomena. Finally, the Coherence value is responsible for how the text is written: whether it consists of individual short sentences or, conversely, is an articulation of a long, connected thought. Tone analysis was performed using FastTextSocialNetworkModel built in Dostoevsky library. Each post was assigned to one of the following tone classes: "negative", "neutral", "positive", "skip", "speech". The class "skip" corresponds to the unrecognized tone of the text, while "speech" corresponds to direct speech or addressing to someone. It is important to note that this model from Dostoevsky library was trained on the corpus of texts RuSentiment [41] from the social network Vkontakte, its value of the metric F1 reaches 0.71.

Basic characteristics of user engagement in posts are views as well as the indicators of audience involvement. These indicators reflect how well views are converted into likes, comments and reposts. The main metric is the engagement index, calculated separately for each post.

$$Eng.Rate = \frac{Likes + Comments}{Views} \cdot 1000$$

For a more detailed analysis of the responses on publications we also calculate the following indicators.

$$Eng.Rate_M = \frac{M}{Views} \cdot 1000$$

where $M \in \{Likes, Comments, Reposts\}$. They are of particular interest because different topics may have more potential for discussion and dissemination in a social network.

A. Data

The dataset was retrieved from SNS Vkontakte via publicly available data through VK API. This social networking site is the most popular SNS in Russia with monthly reach of more than 66% and daily reach more than 38% of population of Russia [42]. For all the posts collected from VK, data on publication time, number of likes, comments, reposts, and views were also acquired.

The sample was collected as follows.

1. Using the online service Kartaslov [43] we selected the top 100 terms associated with the keyword “Politics” (“Politika” in Russian).
2. The list was filtered manually taking into account the specifics of the task involving searching of relevant communities. Ambiguous, abstract, old-fashioned and evaluative terms were excluded. The new list included 60 terms.
3. Using this word list and requests to VK API, up to 100 URLs of communities for each term were collected. The set of communities included 5734 URLs (some of the keywords allowed to collect less than 100 URLs).
4. Doubles, irrelevant, regional, non-Russian, small (number of subscribers less than 10000) communities and communities which tend to overuse reposted content were filtered out.
5. For each of community 10000 last posts were collected (or all the accessible posts if there were less than 10000 of them published since its’ creation).
6. We excluded communities, which were not close thematically to the rest of the communities. After that we filtered out doubles, posts without text as well as unpopular posts with reach less than 1000 views. The dataset included 83 groups and 330754 posts.

B. Measures

Average metrics were calculated for all posts in all communities for each cluster separately.

- **Modularity_class**: a cluster identifier.
- **Eng_rate**: average engagement rate (ER = (comments+likes/views)*1000) for posts in a cluster.
- **Eng_L**: average ER_L (ER_L = (likes/views)*1000) for posts in a cluster.
- **Eng_C**: average ER_C (ER_C = (comments/views)*1000) for posts in a cluster.
- **Eng_R**: average ER_R (ER_R = (reposts/views)*1000) for posts in a cluster.
- **N_word**: average number of words for posts in a cluster.

- **N_sent**: average number of sentences for posts in a cluster.
- **Objectiveness**: average objectiveness for posts in a cluster.
- **Qualitativeness**: average qualitativeness for posts in a cluster.
- **Activeness:** average activeness for posts in a cluster.
- **Dynamism**: average dynamism for posts in a cluster.
- **Coherence**: average coherence for posts in a cluster.
- **Hashtag_%**: share of posts with hashtags.
- **Url_%**: share of posts with URLs.
- **Emoji_%**: share of posts with emojis.
- **Exclamation_%**: share of posts with exclamation mark in it.
- **Question_%**: share of posts with question mark in it.
- **Negative_%**: share of posts with negative tone.
- **Neutral_%**: share of posts with neutral tone.
- **Positive_%**: share of posts with positive tone.

V. RESULTS

A. Thematic landscape

The algorithm that uses the calculation of the angle between thematic vectors allowed us to select 9 community clusters with similar texts. Interpretation of the clusters was based on a qualitative generalization of the information about groups, including their names, their typical content and top 100 words describing the main topic of the cluster.

Fig. 1. Graph of thematic landscape

- Class 1: "Foreign affairs" (24%). This cluster of communities used to publish popular content related to foreign politics. It represents current media-like agenda, in some cases with value judgements.
- Class 2: "Infotainment" (20%). Influential populous communities with content close to infotainment such as "armchair analytics", humor, and memes.
- Class 3: "Opposition" (17%). Oppositional cluster, focused on criticism.
- Class 4: "Institutions, parties" (14%). Cluster of groups mostly connected with some political parties (Yabloko, Parnas, United Russia, Party of Growth, Civil platform) and power institutions (State Duma, Prime Minister).
- Class 5: "President" (13%). Cluster with posts devoted to the activity and the statements of Russian President Vladimir Putin.
- Class 6: "Police" (4%). Cluster related to the Ministry of Internal Affairs and Russian Police as well as current events and salient criminal cases.
- Class 7: "Geopolitics, SSC" (4%). This cluster is devoted to geopolitics and conspiracy theories associated with the Social Security Concept, SSC («Kontseptsiya obshchestvennoy bezopasnosti» in Russian), religious and political movement which was popular in 2000s.
- Class 8: "Ukraine, Belarus" (2%). Cluster corresponding to the topic of political relations with Ukraine and Belarus.
- Class 9: "Orthodox" (2%). Communities in this cluster are concerned with orthodox Christianity and Russian Orthodox Church.

As we can see, top 5 cluster together account for more than 90% of all the communities under consideration. Accordingly, these clusters are more complex comparing to the others.

B. Text characteristics

The analysis of the posts showed that the content of several clusters differs significantly from others in terms of text features. The particularities of the texts in clusters "Geopolitics" and "Belarus, Ukraine" is characterized by a high level of objectiveness, which corresponds to the "scientific-sounding" nature of the texts published in them, on the contrary, the texts in clusters "Infotainment" and "Orthodox" are characterized by low objectiveness, typical for "light" texts [39].

The clusters "Infotainment" and "Geopolitics, SSC" were leading in terms of qualitativeness. Texts in these clusters contained many adjectives and adverbs. The cluster "Ukraine, Belarus" was the leader in the level of dynamism, since it contained many verbs in the texts.

The differences in activeness by cluster are insufficient, besides, this index correlates with dynamism. Hereinafter we consider the levels as high if their values were in the first quartile.

TABLE I. TEXT CHARACTERISTICS

Cluster	Objectiveness	Qualitativeness	Activeness	Dynamism	Coherence
Foreign affairs	1,48	0,31	0,10	0,21	2,42
Infotainment	1,19	0,41	0,13	0,26	1,69
Opposition	1,69	0,32	0,13	0,27	3,05
President	1,80	0,32	0,12	0,23	2,82
Institutions, parties	1,91	0,32	0,12	0,21	4,38
Police	1,81	0,26	0,13	0,24	3,19
Geopolitics SSC	2,20	0,36	0,08	0,19	2,18
Ukraine, Belarus	2,09	0,21	0,10	0,38	1,59
Orthodox	1,07	0,29	0,10	0,21	1,76

C. Sentiment

Is was predictable that the tone analysis showed that "Infotainment" cluster had higher share of positive posts. On the contrary, "Orthodox" communities tended to be more negative.

TABLE II. SENTIMENT

Cluster	Negative_%	Neutral_%	Positive_%
Foreign affairs	0,04	0,85	0,07
Infotainment	0,09	0,72	0,12
Opposition	0,04	0,92	0,01
President	0,05	0,87	0,03
Institutions, parties	0,01	0,97	0,00
Police	0,03	0,95	0,01
Geopolitics, SSC	0,02	0,97	0,01
Ukraine, Belarus	0,01	0,98	0,00
Orthodox	0,14	0,69	0,03

D. Other features

The “Opposition” and “Geopolitics” clusters tended to “questioning” in published posts more comparing to other clusters. The style of “Orthodox” cluster was more emotional, its texts often included exclamations and also a lot of hashtags. The difference in the use of emoji found between the clusters reflected “visual” approach to their composition.

The high Emoji_% score for the “Institutions” cluster can be mainly explained by the presence of “Duma TV” and “United Russia” groups using emoji in almost every post. Even though cluster «Ukraine, Belarus» scored the most in URL_% metric, the vast majority of links were in fact posted in the community of a popular blogger Anatoly Shariy and led to his personal website Sharij.net.

TABLE III. OTHER FEATURES

Cluster	Hashtag_%	URL_%	Emoji_%	Exclamation_%	Question_%
Foreign affairs	0,34	0,30	0,18	0,18	0,19
Infotainment	0,23	0,06	0,15	0,22	0,16
Opposition	0,26	0,45	0,11	0,22	0,29
President	0,13	0,24	0,11	0,18	0,14
Institutions, parties	0,28	0,18	0,38	0,16	0,12
Police	0,38	0,16	0,06	0,12	0,04
Geopolitics, SSC	0,02	0,63	0,02	0,11	0,26
Ukraine, Belarus	0,00	0,92	0,00	0,01	0,01
Orthodox	0,58	0,06	0,06	0,38	0,22

It should also be noted that the clusters "Opposition", "Foreign affairs", "Orthodox" generally used longer texts.

TABLE IV. TEXT LENGTH

Cluster	N_word	N_sent
Foreign affairs	168	10
Infotainment	67	5
Opposition	168	10
President	110	7
Institutions, parties	101	5
Police	141	8
Geopolitics, SSC	42	3
Ukraine, Belarus	15	1
Orthodox	203	15

E. Engagement

Some variation in the engagement metrics was also found. Clusters “Foreign Affairs” and “Orthodox” had higher engagement rate than others. Meanwhile “Opposition” cluster had high levels of comments and reposts per post. The posts in clusters “Opposition”, “President”, and “Ukraine” are tended to spark discussions more active compared to the others.

TABLE V. ENGAGEMENT RATES

Cluster	Eng_rate	Eng_L	Eng_C	Eng_R
Foreign affairs	20,1	16,8	3,3	1,2
Infotainment	22,2	19,6	2,6	1,6
Opposition	30,4	25,2	5,1	4,2
President	37,6	31,8	5,8	2,8
Institutions, parties	22,1	18,8	3,3	3,0
Police	7,8	6,5	1,3	0,4
Geopolitics, SSC	13,3	12,4	0,8	2,8
Ukraine, Belarus	13,4	7,9	5,4	0,3
Orthodox	39,8	39,1	0,6	4,9

F. Generalization

Summarizing statistics for communities and posts in them, we can provide the key characteristics for each thematic cluster, reflecting the nature of strategic communication in communities that highlight their particular agenda and the specificity of content representation.

TABLE VI. KEY CHARACTERISTICS

Cluster	Key characteristics
Foreign affairs	Long texts
Infotainment	High qualitativeness More positive
Opposition	More questions Long texts
President	High ER
Institutions, parties	High coherence
Police	Low ER
Geopolitics, SSC	High objectiveness High qualitativeness More questions
Ukraine, Belarus	High objectiveness High dynamism More URLs Short texts
Orthodox	More negative More hashtags More exclamations Long texts High ER

VI. DISCUSSION

The results of the analysis showed patterns reflecting the differences in posting between particular clusters of online communities related to political issues.

The implementation of topic modeling (including ARTM) in combination with averaging of topics across posts in communities proved to be quite effective for clustering communities on topics with high interpretability.

Some features that we found were quite expected. For example, “Opposition” cluster that used long substantive posts and often called for discussion by questions addressed to the audience, or Orthodox Christian cluster that effectively engaged users through the negative tone and abundance of exclamations. Some of revealed features remain unclear and require further research into the details of individual communities.

Each communities’ cluster develops its communication strategy, adapting the style of texts to the audience demand. Factors such as increased involvement in the clusters related to the President and the Orthodox Church and the ostentatious formalism of the geopolitical conspiracy cluster seems to resonate with the specific nature of (social) media content in Russia.

Accordingly, the differences in the linguistic features of texts in communities' clusters were often linked with the topics forming these clusters which have to be framed in a certain form. Numerous echo chambers formed by the communities' clusters intrinsically differ not only in terms of the topic choice but in terms of published texts styles, and also the response of audiences (in fact, amount of feedback and, consequently, the ratios between lurkers and posters).

It should be emphasized that the presence and special style of texts in the cluster related to the Russian Orthodox Church in the landscape of considered communities play a salient role for politicized online communication in Russian social media as well as for the public political discourse in general [44].

The paper [45] states that communication strategies in social networks were more differentiated and had more positive, personalized and emotional content than similar strategies in traditional media. This result correlates with the findings of our study: the overall tone of messages is largely neutral and positive. Only one cluster of 9 had a significant share of negative messages.

VII. Conclusion

Popular politics-oriented online communities on SNS Vkontakte form a fragmented agenda, which is segmented in a complex pattern. Some clusters can be interpreted in relation to certain institutions of power, others by political orientation, and others by their content features.

On the other side, many communities create the illusion of diverse agenda that is actually mostly structured around a finite number of specified clusters that attract similar audiences. We determined that thematic clusters differed significantly in terms of (para)linguistic metrics as well as some other features and presented a methodology for analyzing the thematic landscape of politics-related communities.

The most interesting results that we found in the description of characteristics of clusters were related to the last three of them. Thus, the "Geopolitics, SSC" cluster met expectations of the increased Objectiveness and Qualitativeness characteristics common to academic texts were revealed. The hot topic of Russia's relations with Ukraine and Belarus were presented in clusters in the form of short substantial and robust texts with many links. Orthodox communities were shown to form their agenda using long texts with exclamations and hashtags, ensuring large ratio of likes and reposts on the post.

The fact that the linguistic features under consideration do not in themselves correlate with the engagement metrics shows that the success of a community (in the sense of providing a high level of feedback) depends not so much on *how they write* but rather on *what they write about* and *where*.

A. Limitations

Our research is a case study and is mostly methodological, since the sample of communities and posts was quite small. We used an associative field approach to collect a keyword list, this can also impose certain effects related to sample shifts.

This work focused on the most popular communities related to politics and on formal criteria describing their post publishing practices. The list of variables that we used describes only part of the aspects that characterize community publishing strategies. Qualitative analysis is required for the better understanding of described tendencies.

The proposed methodology relies heavily on the algorithm of topic modeling. It should also be taken into account that topic modeling tends to work well for sufficiently long texts. The same is true for the linguistic metrics we calculate – they will also be meaningful in texts of sufficient length, so applying the methodology we use to the data obtained from Twitter may not be quite correct and will require refinement of the methodology itself.

B. Further work

For a more detailed analysis of publishing practices and strategies it is necessary to add more characteristics not only for texts, but also for other types of content published on social networks. We plan further develop the analysis framework, as well as expand the sample by developing a more complex list of keywords.

Acknowledgement

The reported study was funded by RFBR and EISR according to the research project № 20-011-31318.

References

[1] J. L. Beyer, Expect us: Online communities and political mobilization. Oxford University Press, 2014.

[2] C. L. Dumas, D. LaManna, T. M. Harrison, S. S. Ravi, C. Kotfila, N. Gervais, ... and F. Chen, Examining political mobilization of online communities through e-petitioning behavior in We the People. Big Data & Society, 2(2), 2053951715598170, 2015.

[3] S. Eccles and C. Russell, Craftivism: online communities of practice as platforms for creative protest. In Media Education Summit, 2016, November.

[4] K. McKelvey, J. DiGrazia, and F. Rojas. Twitter publics: How online political communities signaled electoral outcomes in the 2010 US house election. Information, Communication & Society, 17(4), pp. 436–450, 2014.

[5] M. Conroy, J. T. Feezell, and M. Guerrero, Facebook and political engagement: A study of online political group membership and offline political engagement. Computers in Human behavior, 28(5), pp. 1535–1546, 2012.

[6] A. Soliman, J. Hafer, and F. Lemmerich, A characterization of political communities on Reddit. In Proceedings of the 30th ACM Conference on Hypertext and Social Media, pp. 259-263, 2019, September.

[7] A. Semenov, J. Veijalainen, M. Hajeer, and D. Dasgupta, Political Communities in Russian Portion of Livejournal. In 2014 International Conference on Computational Science and Computational Intelligence, IEEE, Vol. 1, pp. 314-319, 2014, March.

[8] Y. Ren, R. Kraut, and S. Kiesler, Applying common identity and bond theory to design of online communities. Organization studies, 28(3), pp.377-408, 2007.

[9] W. H. Chih, L. C. Hsu, and D. K. Liou, Understanding virtual community members' relationships from individual, group, and social influence perspectives. Industrial Management & Data Systems, 117(6), pp. 990-1010, 2017.

[10] Y. Su and P. Borah, Who is the agenda setter? Examining the intermedia agenda-setting effect between Twitter and newspapers. Journal of Information Technology & Politics, 16(3), pp. 236-249, 2019.

[11] K. Sasahara, W. Chen, H. Peng, G. L. Ciampaglia, A. Flammini, and F. Menczer, On the inevitability of online echo chambers, 2019, arXiv preprint arXiv:1905.03919.

[12] M. McCombs, A look at agenda-setting: Past, present and future. Journalism studies, 6(4), pp. 543-557, 2005.

[13] R. W. Neuman, L. Guggenheim, S. Mo Jang, and S. Y. Bae, The dynamics of public attention: Agenda-setting theory meets big data. Journal of Communication, 64(2), pp. 193-214, 2014.

[14] K. Searles and G. Smith, Who's the boss? Setting the agenda in a fragmented media environment. International Journal of Communication, 10, pp. 2074–2095, 2016.

[15] R. Chalif, Political media fragmentation: Echo chambers in cable news. Electronic Media & Politics, 1(3), pp. 46-65, 2011.

[16] S. Boulianne, K. Koc-Michalska, and B. Bimber, Right-wing populism, social media and echo chambers in Western democracies. New Media & Society, 22(4), pp. 683-699, 2020.

[17] K. Garimella, G.D.F. Morales, A. Gionis, and M. Mathioudakis, Reducing Controversy by Connecting Opposing Views. Proceedings of the 10th ACM International Conference on Web Search and Data Mining, pp. 81-90, 2017.

[18] V. Garimella, I. and Weber, A Long-Term Analysis of Polarization on Twitter. Proceedings of the 11th AAAI International Conference on Web and Social Media, 2017, arXiv preprint arXiv:1703.02769.

[19] P. Törnberg, Echo chambers and viral misinformation: Modeling fake news as complex contagion. PLoS one, 13(9), e0203958, 2018.

[20] M. Safiullah, P. Pathak, S. Singh, and A. Anshul, Social media as an upcoming tool for political marketing effectiveness. Asia Pacific Management Review, 22(1), pp. 10-15, 2017.

[21] B. McNair, An introduction to political communication. Taylor & Francis, 2017.

[22] H. Savigny, Public opinion, political communication and the internet. Politics, 22(1), pp. 1-8, 2002.

[23] G. S. Enli and E. Skogerbø, Personalized campaigns in party-centred politics: Twitter and Facebook as arenas for political communication. Information, communication & society, 16(5), pp. 757-774, 2013.

[24] D. G. Lilleker and K. Koc-Michalska, Online political communication strategies: MEPs, e-representation, and self-representation. Journal of Information Technology & Politics, 10(2), pp. 190-207, 2013.

[25] D. G. Lilleker, K. Koc-Michalska, E. J. Schweitzer, M. Jacunski, N. Jackson, and T. Vedel, Informing, engaging, mobilizing or interacting: Searching for a European model of web campaigning. European Journal of Communication, 26(3), pp. 195-213, 2011.

[26] K. Platonov and N. Legostaeva, Communicative Strategies of Russian Politicians on Social Networking Site Vkontakte. In International Conference on Electronic Governance and Open Society: Challenges in Eurasia. Springer, Cham, pp. 425-440, 2019, November.

[27] B. Johnstone, Lingual biography and linguistic variation. Language sciences, 21(3), pp. 313-321, 1999.

[28] A. Reyes, Building intimacy through linguistic choices, text structure and voices in political discourse. Language & Communication, 43, pp. 58-71, 2015.

[29] B. Park, M. Y. Kang, and J. Lee, Sustainable Political Social Media Marketing: Effects of Structural Features in Plain Text Messages. Sustainability, 12(15), 5997, 2020.

[30] H. Rheingold, The virtual community: Finding connection in a computerized world. Addison-Wesley Longman Publishing Co., Inc., 1993.

[31] K. Hallahan, D. Holtzhausen, B. Van Ruler, D. Verčič, and K. Sriramesh, Defining strategic communication. International journal of strategic communication, 1(1), pp. 3-35, 2007.

[32] I. Segalovich, A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine. In MLMTA, pp. 273-280, 2003, June.

[33] BigARTM. Web: http://bigartm.org

[34] K. Vorontsov and A. Potapenko, Tutorial on probabilistic topic modeling: Additive regularization for stochastic matrix factorization. In International Conference on Analysis of Images, Social Networks and Texts, Springer, Cham, pp. 29-46, 2014, April.

[35] A. Ianina, L. Golitsyn, and K. Vorontsov. Multi-objective topic modeling for exploratory search in tech news. In Conference on Artificial Intelligence and Natural Language. Springer, Cham, pp. 181-193, 2017, September.

[36] T. Hofmann, Probabilistic latent semantic indexing. 22nd Annual ACM SIGIR Conference (International) on Research and Development in Information Retrieval Proceedings. New York, NY: ACM. pp. 50–57, 1999.

[37] D. M. Blei, A. Y. Ng, and M. I. Jordan, Latent Dirichlet allocation. Journal of machine Learning research, 3(Jan), p. 993-1022, 2003.

[38] Gephi. Web: https://gephi.org/

[39] B. Golovin, Language and statistics. Prosveshchenie, 1971. (In Russian)

[40] M. Korobov, Morphological Analyzer and Generator for Russian and Ukrainian Languages. Analysis of Images, Social Networks and Texts, pp 320-332, 2015.

[41] A. Rogers, A. Romanov, A. Rumshisky, S. Volkova, M. Gronas, and A. Gribov, RuSentiment: An enriched sentiment analysis dataset for social media in Russian. In Proceedings of the 27th international conference on computational linguistics, pp. 755-763, 2018, August.

[42] Mediascope. Web: https://webindex.mediascope.net/report/general-statistics?byDevice=3&byDevice=1&byDevice=2&byGeo=1&byMonth=202004&id=16571

[43] Kartaslov. Web: https://kartaslov.ru/

[44] D. Zhuravlev. Orthodox Identity as Traditionalism: Construction of Political Meaning in the Current Public Discourse of the Russian Orthodox Church. Russian Politics & Law, 55(4-5), pp. 354-375, 2017.

[45] U. Klinger, J. and Svensson. The emergence of network media logic in political communication: A theoretical approach. New media & society, 17(8), pp. 1241-1257, 2015.