Hashtags as structural elements of digital socio-political agenda: folksonomy analysis

Olga Malysheva[1]*, and *Natalia* Ryabchenko[2]

[1]Kuban State University, Faculty of Modern Languages, 350040 Krasnodar, Russian Federation
[2]Kuban State University, Faculty of Management and Psychology, 350040 Krasnodar, Russian Federation

Abstract. The situation of the recent months, associated with constitutional transformational changes and struggle with coronavirus spread, showed that civic communications outperform the government as regards the speed, directions and the scale of discursive fields generation. The digital socio-political agenda forms new socio-cognitive patterns of civic behavior, which leads to the development of constructive or destructive socio-political practices. The government needs innovational approaches and complex methods to conduct timely predictive analysis of socio-political sentiment. It is necessary to develop computational linguistics that allows for Data Science and relational sociology methods, as well as linguo-discursive analysis to be used in order to assess the current state of the digital socio-political agenda and identify the transformational vector of social action in the offline space. The article reports the results of the study of 'Коронавирус' digital socio-political agenda that was conducted through the use of graph visualization method, folksonomy and linguo-discursive analysis. The empirical base for the study comprises a bulk of network data retrieved from Twitter via API. The findings of the study prove that 'Коронавирус' digital socio-political agenda has significantly transformed and expanded within a 5 months' period, with COVID-19 topic receiving twice as much attention from online users.

1 Introduction

Progressive transformations in the socio-political life, caused by the reinforced influence of information and communication technologies on all the spheres of civic life—especially relevant during reforms, crises and emergency situations—cause the necessity of studying the digital socio-political agenda that is formed by online technologies both in the online and in the offline space. According to Internet Live Stats [1], the amount of content generated in the online space per a second accounts for approximately 9000 tweets, 1024 photos on Instagram, more than 86000 search queries on Google and 86389 views of YouTube videos. The overall amount of Internet traffic totals to 102707 GB per second. This data is Big Data and is inherently creolized content. Creolized content, spreading in

---

* Corresponding author: malisheva_83@mail.ru

the online space, forms discursive fields that in turn trigger discussions. Thus, new ways of forming the digital socio-political agenda that predetermines the development of social action in the offline space are created.

The digital socio-political agenda, being formed by discursive fields which are inherently Big Data, with hashtags in particular, is a powerful counter propaganda tool. It is capable of neutralizing the development of destructive social practices in the offline space, as long as a timely identification of threats is conducted. The analysis of the current digital socio-political agenda and prediction of the its transformational vector in the online space is a priority for any government as regards civic support to political decisions, and maintaining power. Thus, it is essential to develop an innovative theoretical approach and complex research methods that would allow studying structural, procedural and contextual components involved into the digital socio-political agenda setting.

The digital socio-political agenda being of discursive fields is a complex construct, which at the moment is mainly formed by social media [2]. On the one hand, this is caused by the communicative paradox, and on the other hand, by advanced technological opportunities provided by social platforms and messengers. Social media, regardless of some limitations, allow instant sharing with creolized content, enhanced with the opportunity to use multiple channels to reach the broadest possible audience. Thus, socio-political communication in the online space is presently multimodal, fragmentary, digressive and multichannel, which reflects innovative trends in its development [3].

Social media are a means of communication and network communication services, which facilitates content spread through the networked structure of communicative interactions [4]. Social media are capable of building communities of users who have common interests, share attitudes and opinions, thus, enhancing civic communications, to accumulate pragmatic potential and trigger social action both in the online and offline space.

2 Problem statement

Modern trends of online socio-political communication, as well as technological opportunities of generating and multichannel content spread, have altered the ways the digital socio-political agenda is created. Currently, hashtags have become a revolutionary way to express active civic position, articulate your opinion and attract public attention to acute socio-political problems in a most laconic way. The relevance of hashtags, as well as their significance regarding the present state of social development and the format of communication, is indisputable. According to the American Dialect Society (ADS) [5], in 2012 a word '#hashtag' was called the word of the year. In 2014 ADS introduced a 'Most notable hashtag' category, which was subsequently renamed 'Hashtag of the Year'. According to ADS, the most notable hashtags are #blacklivesmatter (2014); #SayHerName (2015); #NoDAPL (2016); #MeToo, #NeverthelessShePersisted, #ReclaimingMyTime, #Resist (2017); #neveragain (2018); #thankunext and #timesup (2019). #BlackLivesMatter and #MeToo were contenders in the category Word of the Decade (2010-2019). This trends reflects a specific feature of modern communication which is text compression along with the technological opportunities which online space can provide—an increase in the range of meanings conveyed, the amount of content shared, and the speed of content spread.

Due to their structural and technological features, hashtags is a means of creating an emphasis and serve to attract attention. They classify information and facilitate its search in the online space. Besides, they categorize information and attach a particular set of meanings to a particular unit. Hashtags acquire additional meanings being accompanied by other hashtags. For instance, #trump when used in isolation has a neutral connotation.

However, when used with such hashtags as #maga, #donaldtrump, #keepamericagreat or #trumpsupporters, it is interpreted in highly positive way. Conversely, #trump followed by #biden, #funny, #memes, #fakenews, #thegreatawakening, #berniesanders, #obama, #notmypresident, #walkaway, #foxnews or #resist strongly suggests that the attitude expressed by the user is rather negative.

The recurrence of a hashtag in a discursive field in a particular socio-political context creates a stable connotative field for the given hashtag. The connotative field in turn influences the overall meaning of the content marked by a particular hashtag. Thus, hashtags categorize reality, attaching a specific set of meanings to a particular entity [6].

The use of socially significant hashtags in the online space is currently one of the most popular ways of attracting public attention to a wide range of socio-political issues [7]. Thus, hashtagization is an effective means of the digital socio-political agenda setting. Hashtags join the two parts of social space—the online and offline. In that respect they have dual and dynamic nature. The digital socio-political agenda, 'saturated' with hashtags that increase people's awareness about political, discrimination, women's rights, environmental issues, allows for social communities to shape and social action to be triggered in the offline space. The development and increase in popularity of such social movements as MeToo, Women's March, BlacklivesMatter, StayHome, МыВместе, whose activity stirred massive public outcry, is caused by the digital socio-political agenda 'saturated' with socially significant hashtags [8]. Thus, we conclude that the digital socio-political agenda through social media via hashtags is capable of producing and forwarding constructive (as well as destructive) social practices in the online and offline space.

3 Methods

The study of data, generated in the online space requires specific research approaches and methodology as the processes that take place in the online space are complex, non-linear and multidimensional. They occur across several dimensions, therefore, the data that are generated in the online space are defined by several parameters—belonging to a particular semiotic system, a community of users, the type of information dissemination technology (hypertext, author citation, retweet/ repost, the use of hashtags).

Big Data cannot be analyzed by means of traditional methods that imply continuous sampling of all relevant text corpora, as online data comprises bulks of unstructured information which are impossible to process due to limited working capacities. Continuous sampling of network data inevitably causes difficulty with Big Data processing. In our research, we apply an approach which means identification of the entry point, a criterion which allows us to identify a vector for retrieving data [9]. It enables targeted mining of network data, whose scope can be broadened once the processing of the first datasets is completed. Data mining criteria or entry points can range from a particular social networking site or a type of data to a key word, a topic, a hashtag, a named entity, or a date. Thus, identification of an entry point—a key word, a topic, or a platform for data mining— is absolutely essential. It is the most efficient approach to Big Data analysis that is particularly relevant when analyzing Fast Data. Research methodology that allows us to analyze topical data in real time should comprise Data Science and relational sociology methods, folksonomy analysis and graph visualization method, which are used to process empirical data for complex linguo-discursive analysis.

In our research, to study the digital socio-political agenda, identify the subject matter and its transformational vector, we apply folksonomy analysis and graph visualization method, enhanced by linguistic analysis.

Folksonomy is a practice of collaborative application of hashtags to label some content, which allows structuring, classification and categorization of content which constitutes the

subject matter of the digital socio-political agenda on a crowdsourcing principle. Being the result of collective work of independent users, hashtags accumulate sets of implicit denotative and connotative meanings, suggested by a number of users at the moment of labeling some content with a particular hashtag. Folksonomy analysis is a semantic analysis of hashtags and their interrelations. Together with linguo-discursive analysis, which allows interpretation of a combination of accumulated meanings and concepts online content contains, folksonomy analysis enables identification of the conceptual framework of the digital socio-political agenda through the analysis of hashtags and the specifics of their connections via graph visualization method.

Graph visualization method focuses on the visual representation of the elements as well as characteristic features of their relationships [10]. It allows for the digital socio-political agenda to be analyses as a social graph, where nodes (or vertices) represent hashtags used by online users for labeling and classifying content, and links (or edges) feature thematic proximity or remoteness regarding the key notion represented by the central node. Such approach allows us to see the current state of the digital socio-political agenda, predict its transformational vectors and foresee the development of constructive and destructive social practices in the offline space.

To study the transformation and change of 'Коронавирус' digital socio-political agenda, we allied folksonomy analysis, graph visualization method, enhanced by linguo-discursive method. The empirical base for the research into 'Коронавирус' digital socio-political agenda—which developed in the online space due to the crisis caused by coronavirus pandemic—comprises network data retrieved from Twitter. The choice of the platform is warranted by the fact that Twitter is one of the most cited information agencies, and political information is shared via this platform more frequently [11].

In March 2020 (the first week of the quarantine in Russia) and the first week of September 2020, our research team collected a set of network data via continuous sampling of messages containing 'коронавирус' key word. The empirical sample which we obtained comprises the following bulk of network data: messages posted by Twitter users; the dynamics of answers (retweets) to the posted messages; information about the users who posted these messages, with a registration of their interactions in order to analyze their activity as a social graph; hashtags used; frequently used words and collocations, with a registration of their interactions in order to analyze the dynamics of their use as a social graph and identify the themes that dominate the studied digital socio-political agenda; sets of emoji used in posts.

The article deals with the analysis of a part of the empirical sample which includes hashtags used by Twitter users in their posts together with 'коронавирус' keyword. We formed a dataset for March that includes 2038 hashtags (and 9768 links) and a dataset September, comprising 3850 hashtags (and 25712 links). The datasets contain information about: hashtag interactions, in order to analyze their activity as a social graph; hashtags used; frequently used words and collocations, with a registration of their interactions in order to analyze the dynamics of their use as a social graph and identify the themes dominating 'Коронавирус' digital socio-political agenda.

Via Gephi software program, using graph visualization method, we visualized the use of hashtags from the datasets. Thus, we obtained two social graphs '#Коронавирус-March' (Fig.1) and '#Коронавирус-September' (Fig.2). These graphs included all the hashtags used together with the keyword 'коронавирус' in the messages posted on Twitter within the corresponding periods.
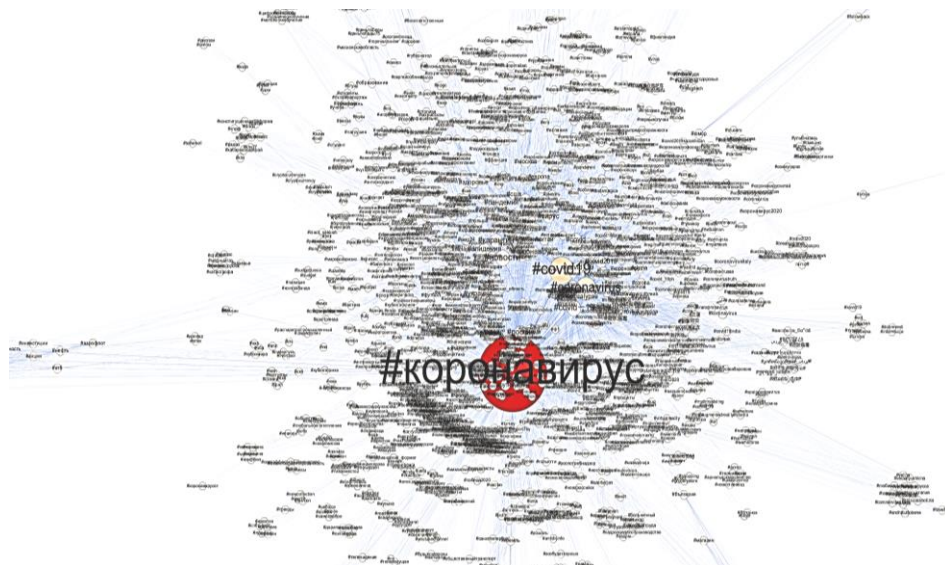
Fig.1. #Коронавирус (Coronavirus) – March social graph.

The given visualizations feature hashtags, which are represented by nodes, or vertices, and are linked to the key hashtag #коронавирус, which is the central hub of the given social graph, through lines that are the links, or edges, i.e. relationships between the hashtags. Using the ForceAtlas2 layout procedure via Gephi software program, we have carried out the distribution of nodes in space to visualize the nature of their relationships. By featuring close or distant location of nodes in the visualization, this layout reflects the thematic proximity or remoteness of hashtags and topics. The proximity of nodes indicates the thematic proximity of the messages in which the hashtags represented by these nodes were used. The diameter of a node reflects the importance of a given node in comparison with other nodes in the graph. A larger diameter shows a greater frequency of use, and an increased popularity of this hashtag compared to others. With an increase in the size of a node diameter, its importance also increases regarding its influence on other nodes in the graph, as well as on the semantic content and interpretation of all content published on the social networking site.
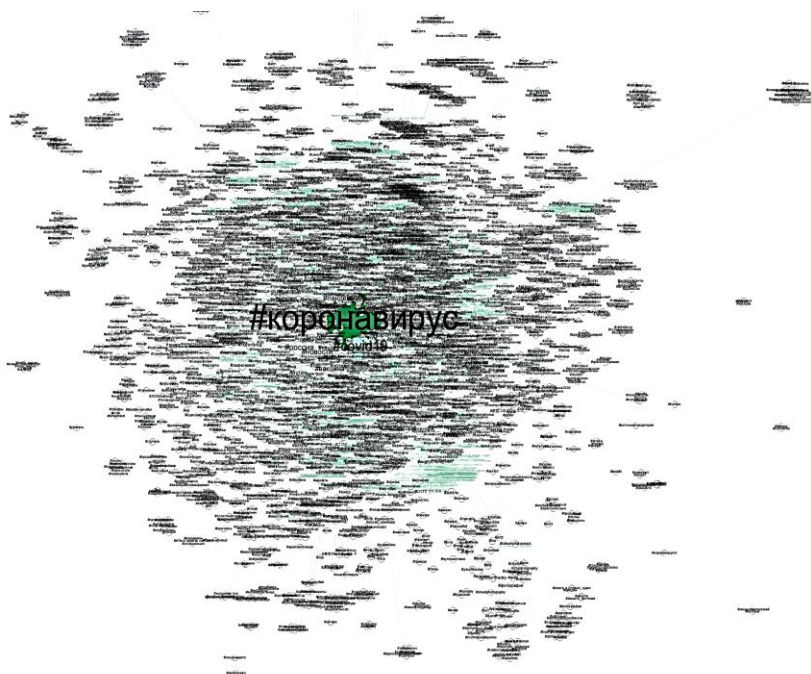
Fig.2. #Коронавирус (Coronavirus) – September social graph.

In the visualization, hashtags with a larger diameter are defined as hubs—the largest nodes and, accordingly, the dominant hashtags (and topics) in the analyzed digital socio-political agenda. To identify the aggregate of hubs that form the thematic basis of #Коронавирус-March and #Коронавирус-September social graphs, we applied a filter and found out that the percentage of hubs whose diameter significantly differs from the diameter of all other nodes in the graphs was 4%. Thus, we obtained #Коронавирус-March and #Коронавирус-September thematic bases, visualization of which is shown in Fig. 3 and 4.

4 Results

The analysis of the thematic basis of the #Коронавирус-March social graph (Fig. 3) showed that the distribution of topics is fairly even with respect to the presence of alternative or conflicting opinions, since in the visualization the hubs are located on one side of the central hub #коронавирус, and clear clustering is observed. The hubs that are closest to #коронавирус hashtag are #putin, #russia, and #formula 1. The clusters of large hubs #covid-#coronavirus-#covid-19 and #новости-#карантин-#эпидемия, #коронавирус-#тсн-#новины have a large number of links with each other and with the central hub, but are significantly remote. This means that the topic of coronavirus is directly associated with

these issues, but is not so relevant from users' perspective. The cluster comprising #pandemic-#italy-#us hashtags is indirectly linked to #коронавирус hashtag through #новости hub, which suggests that users were interested in the news reports in the mentioned countries, and defined the situation in those countries as #пандемии, unlike the situation in Russia (#эпидемия). The remoteness of these clusters from the central hub indicates that users did not feel the dependence of the problems they mentioned from coronavirus spread in Russia, with the exception of a topic regarding the president's actions in the current situation and Formula 1 race in Sochi. Notably, there is a cluster containing hashtags in Thai, Korean and Chinese, as well as a large cluster of hashtags containing #coronapocalypse hashtag, which indicates a lot of tension and pessimistic sentiment, which constructively transform in September.

Fig.3. Thematic basis of #Коронавирус (Coronavirus) **–** March social graph

Fig.4. Thematic basis of #Коронавирус (Coronavirus) **–** September social graph

The analysis of the thematic basis #Коронавирус-September social graph (Fig. 4) showed that the distribution and clustering of hubs and, respectively, main topics is uneven. This means that the digital socio-political agenda contains either diametrically opposing or mutually exclusive topics and opinions. This conclusion comes after the analysis of the

visualization of the thematic basis of #Коронавирус-September social graph, which shows the aggregate of hubs diametrically opposed to each other regarding #коронавирус hub. Notably, there occurs a significant reduction in the distance between large and smaller hubs, representing a variety of topics, and the central hub #коронавирус, which indicates the transformation in perception and attitudes of users. They became more awareness of the relevance of the coronavirus problem for Russia and its direct connection with a variety socio-political problems. Such topics as #минздрав, #реформа, #ученый, #пермь, #испытания, #днр, #донецкая область, #киев, #дагестан, #воз, #pandemic, #who came close to the central hub. The hubs and main themes of September are #covid19, #пандемия, #россия, #новости, #вакцина; #пандемия; #карантин. At the same time, #covid19 hub is defined by users through such hashtags as #тасстатистика_коронавирус, #тасситуация_коронавирус, #минздравтурции, which means that users became interested in the official verified information provided by a reputable news agency. This is a significant change compared to the situation in March, when users were interested various news resources which included Russian (#мир24), Ukrainian (#тсн#новини) and particularly Serbian (#trsvesti) TV and radio companies. The digital agenda for September includes, as one might expect, such hashtags as #дети, #образование, #школа, #1сентября, #статистика, #маска, #маски, #вакцина, #вакцинация, #здоровье. However, they are far from the central hub and are indirectly related to #коронавирус problem. Therefore, the relevance of these issues, according to the users, is not related to #коронавирус problem.

By comparing the findings received through the analysis of graph visualization of 'Коронавирус' digital socio-political agenda for March and September 2020, we can estimate how the agenda has changed over 5 months. The analysis showed that the digital socio-political agenda, along with a significant increase in interest to the situation '#коронавирус-Covid-19', has significantly transformed. The use of the #коронавирус -> #covid19 bundle almost doubled in September and amounted to 1129 times, which indicates a significant increase in user interest to this topic.

Notable changes to the 'March -> September' digital socio-political agenda are:

- attaching a wider range of socio-political problems to the topic of coronavirus spread;

- change of focus regarding a source of news about the coronavirus (various TV and radio companies -> news without specifying the source, or TASS);

- change in the status of the problem and its global nature recognition for Russia (#epidemic -> #pandemic);

- transformation of sentiment in a constructive direction which reflects determination to solve the problem and admits the absence of apocalyptical threat (#coronapocalypse—the end of the world caused by the coronavirus -> #ученые, #пермь, #испытания, #вакцина, #вакцинация);

- the emergence of new socio-political actors and topics (#минск, #беларусь, #мир24 -> #навальный, #беларусь, #лукашенко; #ковид19-#дезинформация; #школа; #маски);

- unification of spelling (covid19, covid-19, ncov2019, covid2019, covid -> predominance of 'covid19' spelling over other forms).

5 Discussion

During the period of reforms, crisis and emergency situations, to which the coronavirus pandemic refers, the research into digital socio-political agenda becomes highly warranted, since it affects the formation and the development of various processes in socio-political systems. The digital socio-political agenda consists of vast bulks of network data, the processing time of which affects the decision-making process in the political and social sphere. Therefore, the development of new methods of network linguistics, which allow us

to quickly and efficiently assess the transformation and development vectors of the digital socio-political agenda. This will allow neutralising destructive social practices and triggering constructive social action in both online and offline space.

Acknowledgments

References

1. Internet Live Stats (2020)
2. IFEX (2020)
3. N.A. Ryabchenko, O.P. Malysheva, VKL 101 (2020)
4. D. Stuart. Online Information Review **37**, 486 (2013)
5. American Dialect Society (2012)
6. A.A. Belovodskaia, Media Linguistics **6**, (2019)
7. Y. Al Hariri, W. Magdy, M. Wolters, 18 (2019)
8. Pew Research Center: Internet, Science & Tech (2018)
9. N.A. Ryabchenko, O.P. Malysheva, **86**, 1199 (2020)
10. R.M. Tarawaneh, P. Keller, A. Ebert, **27**, 151 (2011)
11. Brand analytics (2019)