



Differentiable characteristics of Telegram mediums during protests in Belarus 2020

Ivan Slobozhan¹ · Tymofii Brik² · Rajesh Sharma¹

Received: 14 September 2022 / Revised: 16 November 2022 / Accepted: 19 November 2022 / Published online: 4 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2023

Abstract

Online activism has been pivotal for anti-government protests all around the world. A lot of existing research studies how people use social media to mobilize and coordinate protests. However, some platforms offer more than one kind of medium for communication, in particular, Telegram. We found that a lot of research literature focuses on either one or another medium or sometimes merges them together. In this research, we performed a comparative analysis of three communication mediums in this platform: channels, groups, and local chats during protests in Belarus in 2020. We first collected a dataset which follows six months of protest, which allows us to study the temporal variation of online activism. We find that in Telegram, these three modes of communication are not equally important. Furthermore, they are responsible for different types of communication. For example, in groups, people discuss significant protests that occurred nationwide. In comparison, people discuss local and small demonstrations and protests in chats. Finally, Telegram channels are very distinct from local chats and channels in terms of content. Our finding suggests that future research should consider these modes of the communication independently rather than blending them.

Keywords Online protests · Social network · Telegram · Belarus · LDA

1 Introduction

More and more antigovernment protests have occurred in the world in recent years. Rallies against the arrest of Aleksei Navalny in Russia, Euromaidan in Ukraine, protests against police brutality in the USA, Germany, and Colombia are only a few examples of more than two hundred significant antigovernment demonstrations in more than one hundred countries.¹ A common denominator for such protests is that they all rely on social media. While Twitter, Facebook, and Reddit were extensively studied in the research community, the studies of Telegram and its role in the protest activity are

scarce. At the same time, a big part of computational social science is focused on English-speaking countries such as the USA, the United Kingdom, or countries in which English is common (for example, in India). However, computational social science does not pay equal attention to the research on protests in countries where the English or Latin alphabet is not common. This is surprising, given that such countries are often exposed to various forms of protests.

This paper fills these gaps by analyzing protests in Belarus in 2020 using Telegram's data from May to November 2020. Telegram has a unique feature that is not present in many other messengers. Telegram has three different communication tools (mediums) where users can share their messages with the audience. They are: (1) channels (broadcasting tool for admins where they share posts with the subscribers), (2) groups (communication tool for users to share messages with the rest of users), (3) and local chats (similar to groups, but the users share some specific geographical location). We collect more than four million messages from 654 local chats, more than 30 thousand posts from five large channels, and six million messages from two big groups

✉ Ivan Slobozhan
ivan.slobozhan@ut.ee

Tymofii Brik
tbrik@kse.org.ua

Rajesh Sharma
rajesh.sharma@ut.ee

¹ Institute of Computer Science, University of Tartu, Tartu, Estonia

² Policy Research Department, Kyiv School of Economics, Kyiv, Ukraine

¹ <https://carnegieendowment.org/publications/interactive/protest-tracker>.

from the period of 1st May 2020 to the 29th of November 2020. Using this rich dataset, we investigate the following Research Questions (RQs) to understand the role of each medium during the protests in Belarus in 2020:

RQ 1: Does the activity of users differ in mediums?

Firstly, we analyze the users' activity as the number of messages they post per day in each medium (see Sect. 4). Then we identify the top five spikes in each medium and plot them against the external offline events using open source information from Wikipedia and online news sources such as BBC,² DW³ and other.

Findings We discover notable differences in the activity of users and admins (for channels) in the number of daily messages in the mediums. We also find that the top spikes of users' activities are different in the three mediums. The dates of the highest spikes in channels match with the important political announcements. At the same time, spikes in local chats match with major protests and marches. Finally, the spikes in groups are related to both.

RQ 2: What topics do users discuss in each medium?

To investigate this question, we perform a three-stage analysis (see Sect. 5). Firstly, we analyze the most frequent words in each medium using WordClouds. Secondly, we extract topics using latent Dirichlet allocation (LDA) and compare them among the mediums. We compare topics in two different scenarios. First, we analyze topics that we extract for the whole period of the data. Second, we also select the most important topics which overlap between any two mediums during their most active days (i.e., spikes). Finally, we analyze the context surrounding the names of politicians, protests and specific locations utilizing Word2Vec embeddings. We train three models for each medium and then for each noun. We find the top ten most important nouns using cosine distances among words' embeddings. Then we infer the context surrounding using these ten words to find the main topics which users discussed.

Finding Using the data for the whole period, we find that users were concerned with three main subjects, namely COVID, elections and protests. However, we also find that each medium featured some unique topics during their respective spikes.

RQ 3: Do users communicate distinctly in different mediums?

We investigate whether users communicate distinguishable in the different mediums (see Sect. 6). In other words, whether it is possible to predict for a given message from which medium this message is. We train a classifier that predicts the medium by a given text input to check this question. We use TF-IDF with a bag on n-grams to create

features from text and then feed them to the logistic regression. Finally, we perform an error analysis and analyze the feature importance.

Finding We find that the messages on channels can be predicted much better than messages on other mediums. At the same time, messages from groups and local chants are not so easy to classify correctly. This finding has two possible interpretations. First, those people who post via channels use consistent words and language patterns which makes them recognizable. This could mean that people use channels to forward similar messages and news. At the same time, groups and chats are more disorganized and less homogeneous. Therefore, there is no consistent language pattern that unites them.

2 Related work

2.1 Online activism and protests across the globe

Online activism or cyber activism has been widely studied in the scholarship, especially with regards to Twitter and Facebook (Sandoval-Almazan and Gil-Garcia 2014). This activism extends across various topics including but not limited to #metoo movement (Goel and Sharma 2020), education movements, Scherman et al. (2015), environmental movements (Bastos et al. 2015), and global policy-based activism (Poell 2014). A lot of anti-government protests worldwide used online platforms as well (Jost et al. 2018; Theocharis et al. 2015). The most telling examples are Arab Spring in the Middle East and North Africa in 2010s Acemoglu et al. (2018); Steinert-Threlkeld et al. (2015), Russian protests against electoral frauds in 2011 Enikolopov et al. (2020), the Euromaidan Revolution in Ukraine (Metzger et al. 2016; Onuch 2015a). Some researchers used surveys to analyze how protesters used social media. For example, in Ukraine, a survey of students in Kyiv and Lviv showed that YouTube, VKontakte, Twitter, and Facebook assisted the students with the protests (Piechota and Rajczyk 2015). Other surveys conducted in Ukraine showed that protesters were invited by their friends and social ties (including online) rather than by parties, NGOs, or other formal organizations (Onuch 2015a, b). Similarly, a survey of participants in Egypt's Tahrir Square showed that social media played an impactful role in the protests by engaging users in information diffusion (Tufekci and Wilson 2012). Other researchers analyzed the content of hashtags or posts and also pointed out that online social media were crucial during the protests (Jost et al. 2018). For example, in Sinpeng (2021), researchers have analyzed the Twitter users network in Thailand. They observed an increase in the ties between them associated with the goal of taking the protests forward. Similar findings were observed in Turkey during the Taksim square

² <https://www.bbc.com/>.

³ <https://www.dw.com/>.

protests (Smith et al. 2015). Researchers in Enikolopov et al. (2020) used instrumental variables techniques to show a causal link between penetration of V Kontakte, online coordination, and the likelihood of protests in Russia.

2.2 The role of Telegram in protests

The platform Telegram recently gained attention in the research community. Researchers (Urman et al. 2020) analyzed the role of Telegram during protests in Hong Kong in 2019. They found that Telegram became popular among social activists and that it was mostly used by protesters to distribute information. At the same time, it was used to discuss future actions and coordination. However, they blend both channels and group messages in their analysis. In Su et al. (2022), the authors also examined protests in Hong Kong by analyzing the messages from a public channel through different forms of participatory activity. Akbari and Gabdulhakov (2019) analyzed the role of the platform during the protests in Iran and how the government demanded information and private messages from Pavel Durov. Last but not least, Schulze et al. (2022) performed a quantitative study about radicalization dynamics in Telegram during COVID-19 protests in Germany, where authors analyzed the contest of nine Telegram channels.

2.3 Protests in Belarus

There have been several works covering protests in Belarus, which happened in 2011 (Karaliouva 2013), 2017 (Hansen 2017), 2020 (Buzgalin and Kolganov 2020; Moshes and Nizhnikau 2021). Most of these works have not included social media analysis. For example, researchers in Karaliouva (2013) performed a comparative analysis of the protests covered by pro and anti-governmental news articles. In Hansen (2017), using the interviews with protesters, observers and opposition leaders, the author proposes that the very nature of how the city area is organized has an influence on the protest. He argues that the city centre does not have a preferable symbolic value to the opposition while also being avoided by the public. Recently, a study of protests in Belarus showed that pre-existing social networks significantly increased the likelihood of protests during the elections on August 9–15 Mateo (2022). The recent protests, which began in August 2020, have also attracted various studies. However, most of them have a rather sociological and historical vision of the protest. For example, in Buzgalin and Kolganov (2020), authors discussed historical reasons and social aspects of the protests. Similarly, authors in Moshes and Nizhnikau (2021) discussed reasons for the protests and possible outcomes. More recent studies analyzed the reasons for the actual outcome of the protests in Belarus 2022 (Mudrov 2021; Robertson 2022). Several research articles provided

quantitative analysis, such as Nikolayenko (2022), where the author studied the role of emotions in shaping mass mobilization. The closest research to our paper (Herasimenka et al. 2020) analyzed the protests in Telegram during the Belarus protests in 2020, but they investigated only the role of local chats and did not differentiate the role of different mediums on the protests in Belarus. Moreover, we found only one policy paper that addressed online activism during the protests in Minsk (Shelest 2020). In contrast to the previous studies, which investigated one specific medium, our paper is aimed to understand the different characteristics of channels, groups and local chats during the protests in Belarus and compare them.

3 Dataset

For our analysis we collect a set of messages from Telegram messenger⁴ using official Telegram API⁵ with a help of Python package Telethon.⁶

Telegram has different message communication tools between users, namely *channels*, *groups* and *location based chats (local chats)*. Channels and groups share many features, but the main difference is that in channels, there is one to many broadcasting (for example, from a channel creator or admins and moderators to the subscribers, but not vice versa). In groups, the subscribers are allowed to post messages in the news feed. Finally, local chats are designed for small-sized communities that share a specific location. Anyone close to the chat location can find these local chats using a nearby search, without knowing the exact group name or group id to search for it.

Our analysis is based on the data from a set of local chats located all across Belarus, large Belarusian groups without a specific location and active Belarusian news channels. We use a partial set of local chats listed here,⁷ which was created by activists. We use the word partial because some of the chats are private, so we could not scrape the information from them. Other chats changed their identifiers, so we would not find them using a direct search by chat id. The primal motivation of this map is to share the information about the local chats and encourage people to join them, discuss and share different information between each other.

All the data we collect for this work are from public mediums, which means it is either channel, group or local chat. If the medium is public, it means that anyone can join and read the content of that medium. We did not scrape any messages

⁴ <https://telegram.org>.

⁵ <https://core.telegram.org/>.

⁶ <https://docs.telethon.dev/en/latest/>.

⁷ <https://dze.chat>.

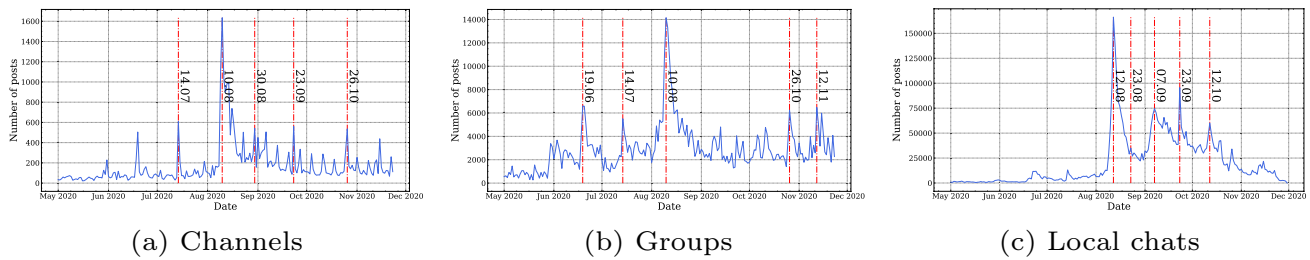


Fig. 1 Number of posts per day in each medium

or other information if the medium was private. Before analyzing the data, we anonymized all users IDs by assigning random but unique IDs instead of the original one. We also assigned random but unique IDs to each medium. Thus, our analysis and the results we report, do not harm the privacy of the users.

In total, we collect 4,482,070 messages from 654 local chats, 36,206 posts from five large channels and 6,061,56 messages from two big groups from the period of 1 May 2020 to the 29 of November 2020. We show channels', groups' and local chats' basic descriptive statistics in Table 1. We define users as the users with at least one comment during the period we analyze. As channels are one-sided medium (only admins or moderators of the channels can post) we do not have users' information for the channels. We observe that the number of posts is larger for the groups than for other mediums because the number of users is much larger. For the same reasons, the average delay per post (the difference in minutes between consecutive messages) is also the smallest for the groups. The average post length (number of characters), is the highest for the channels.

Table 1 Mediums basic descriptive statistics

	Mean	Std	Min	Median	Max
<i>Channels</i>					
Posts	7241	2227	5687	6304	11,118
Post length	233	361	1	119	5709
Delay	43	10	26	47	52
<i>Groups</i>					
Posts	303,078	271,438	111,142	303,078	495,014
Users	17,695	13,073	8451	17,695	26,939
Post length	102	201	1	52	6171
Delay	1	1	0	1	2
<i>Local chats</i>					
Posts	6853	15,665	1	1062	113,273
Users	431	1112	1	105	10,847
Post length	105	248	1	46	8273
Delay	71	51	0	55	212

4 Users activity in three mediums

This section investigates *RQ 1* (Does the activity of users differ in mediums?) by analyzing user activity in all three mediums. Specifically, we measure user activity in terms of messages appearing in each of the mediums daily. Figure 1 shows the number of messages daily in each of the mediums. There are multiple spikes in each plot, however, we discuss only the five most significant ones. To select top spikes, we first sort the dates by the number of messages. Then we select the top five dates, but with a constraint that the time window between the two highest days should be not less than ten days. We use this constraint because the most active days cluster together and usually correspond to the same or similar events. After that, we match the spikes' dates with the real events using open source information from Wikipedia and online news sources such as BBC,⁸ DW⁹ and other. More precisely, after selecting a date on which we consider spike appearance, we look at the Belarusian news media as a reference to find the event highlighted on the same date. After that, we report this event as the event corresponding to the spike on a given date of the spike.

4.1 Activity patterns in channels

Channels represent top-down communication from an author (admin or moderator) to the audience. Thus, channels are often used as online news feeds. Figure 1a shows the number of messages daily in channels and Table 2 matches the dates of the top five spikes with important events that happened in Belarus on the same date. One can observe that the admins/moderators made a significant number of posts on the dates of some important announcements (1, 2, 4, 5), and some of these announcements were followed by marches or protests (3, 4, 5)

⁸ <https://www.bbc.com/>.

⁹ <https://www.dw.com/>.

Table 2 Channels' significant events

Number	Date	Event
1	14th July 2020	Removal of two opposition candidates from the elections
2	10th August 2020	Lukashenko was announced winner of presidential elections
3	30th August 2020	Rally against Lukashenko
4	23th September 2020	State media announced that Lukashenko had been inaugurated. Series of protests
5	26th October 2020	Lukashenko refused step down from the presidency. Series of protests

Table 3 Groups' significant events

Number	Date	Event
1	19th June 2020	Lukashenko's announced he had "foiled a coup attempt". Arrest of main opposition rival
2	14th July 2020	Removal of two opposition candidates from the elections
3	10th August 2020	Lukashenko was announced winner of presidential elections
4	26th October 2020	Lukashenko refused step down from the presidency. Series of protests
5	12th November 2020	Gather to mourn death of protester Raman Bandarenka

Table 4 Local chat' significant events

Number	Date	Event
1	12th August 2020	Protests against government and policy brutality
2	23th August 2020	Rally against Lukashenko
3	7th September 2020	Protest against Lukashenko
4	23th September 2020	State media announced that Lukashenko had been inaugurated. Series of protests
5	12th October 2020	March of the Seniors

4.2 Activity patterns in groups

Similarly to the channels, Fig. 1b shows the number of messages per day in groups and Table 3 highlights the top five spikes in the user activity. Out of five spikes in groups, three of them match with the spikes in channels. We observe two significant pieces of news about severe human rights violations (1, 5) almost ignored by channels and picked up by groups. We can assume that Minsk activists were triggered and invested in human rights issues. They were able to raise this issue in the comments (in groups). However, such discussions were probably missing in channels or less highlighted.

4.3 Activity patterns in local chats

Local chats are designed for individuals connected by some geographical location to exchange messages with each other. Figure 1c shows the number of messages per day in local chats and Table 4 maps the peaks with real events. Interestingly, the activity in local chats is very different from the activity in channels or groups. First of all, compared to other mediums, the top five spikes in local chats might have

been triggered by protests or marches (1, 2, 3, 4, 5). In addition, the spike dates do not match the spike dates of other communication tools in almost all cases, except (4), which appears only in channels.

4.4 Discussion

As we observe from the analysis of top spikes, the users' activity in each medium is different. On the one hand, the quantity of messages is different. On the other hand, the spikes appear on different days across different mediums. More precisely, we find out that most of the top spikes in channels and groups can be matched with important political announcements. On the contrary, the top spikes in local chats can be matched with major protests or marches. In other words, the issue of protests and human activism could be considered central only for local chats from the very beginning of the political crisis. While previous studies showed that pre-existing social networks are important for future protest activities (Mateo 2022), we also find that active online communication keeps going after protests in August 2020.



Fig. 2 Word Clouds for each medium

To confirm our observations, we extend our analysis to the whole period of data instead of focusing on five data points (with the highest activity). We use a dynamic time wrapping algorithm to find the order of similarity between each pair of communication mediums represented as a time series of the number of comments per day. We use the following steps to calculate similarity between time series. Firstly, to calculate a measure of similarity between each pair of mediums, we align them by the dates on which we have a reported activity for each of the mediums. Next, the dynamic time wrapping algorithm was applied using Manhattan's distance. The results are the following: (1) distance between channels and groups is 5267, (2) distance between groups and local chats is 9304, (3) distance channels and local chats is 20,967. This could signify that channels and groups are more aligned in terms of the events that trigger activity, followed by channels and local chats and groups and local chats.

5 Topics discussed in three mediums

This section explores *RQ 2*: (What topics do users discuss in each medium?) by analyzing the most frequent words, topics and context of the specific words in each communication tool.

Before diving into each part of the analysis, we describe the preprocessing pipeline we used to clean raw text messages. The pipeline consists of lowercasing the text and removing punctuation and stopwords (English, Russian and Belarusian). After this part, we filter out messages with less than eight words (the median) to leave more meaningful messages. The reason is that most of the messages with less than this threshold, we find as simple and quick replies, for example, “so that am I talking about” or “yes, I agree with you”, which can bring more noise than a useful signal to our analysis.

5.1 Word clouds

Our analysis starts with comparing the most frequent words in channels, groups and local chats. We plot 100 most frequent words for each medium using WordCloud (see Fig. 2). We observe that the most frequent words have lots of overlaps. For example, in all mediums, words such as *Lukashenko*, *Belarus*, *people* are the top ones. A more careful investigation also shows that in channels words such as *news* have similar frequency as *people* and *anonymity*. However, in the local chats other most frequent words are *need*, *today*, *urgent*. This gives us additional evidence that channels were used to inform people of some nationwide events. However, groups and local chats in terms of most frequent words are more similar and both relate to coordination and protest discussion.

5.2 Topic modeling

This part analyzes the topics the users discussed in each communication tool using the topic modeling approach. We use the latent Dirichlet allocation (LDA) to find the topics. To select the number of topics, we use the grid search approach by the number of topics (from 1 to 20, with step 2) while finding the number that maximized the Coherence score. We analyze the topics for each medium separately for the whole period of data we collect and call this part of analysis *topic modeling in a global context*. We also analyze topics for each of the top five significant events we find in the previous section, focusing only on the spikes that overlap between any two mediums and call this part of analysis *topic modeling in a local context*.

5.2.1 Topic modeling in a global context

Topic modeling of the whole period of the data period is presented in Table 5 for channels. We present only channels because we do not observe too much variation for the other mediums. Most of the topics we observe can be related to

Table 5 Global topics (translated from the Russian and Belarusian languages to the English language)

Number	Topic	Tokens
1	Protests	Man, security officer, protest, today, belarusian, get out, time, detain, people, belarus
2	Clashes and losses	March, factory, new, grenade, disability, freeze, bot, bntu (university), timelapse, cascade
3	Information	Bchb (flag), package, special forces, viasna ^a , video, win, march, what, on, memes
4	Covid restrictions	Fashion, fashionable, forever, to be, gas, gomel (city), lockdown, today's march, belaes ^b , mask mode (covid restrictions)

^ahttps://en.wikipedia.org/wiki/Viasna_Human_Rights_Centre, ^bhttps://en.wikipedia.org/wiki/Astravets_Nuclear_Power_Plant

Table 6 Channels' events topics (translated from the Russian and Belarusian languages to the English language)

Spike	Number	Topic	Tokens
1	1	Protest information	person, news, anonymous, offer, people, detain, detention, queue, protest, riot police
	2	Protest location	Subway, station, coupling, internal, bandit, number, year, beat, hand, close
	3	Call for a political choice	Choice, necessary, registration, today, headquarters, right, street, want, car, refuse
2	1	Protest location	Uruchye (neighbourhood), house, stone hill, meadow, open, entrance, hospital, room, shooting
	2	Arrests	People, this, hand, bahdanovich, from, our cornfields, cornfields, people detained
4	3	Protest activity	Man, people, riot, protest, security officer, report, car, detain, protest, stand
	1	Restrictions	Girl, media, status, portal, detain, car, person, internet, police, sit
	2	Escape	Move, urgently, gray, instagram, attack, jay, potato, cruel, run, it
5	3	Political announcements	Inauguration, president, country, student, news, people, news, power, choice
	1	Student activism	Student, dispersal, bguir, video, photo, bgu, theater, close, grenade, walk
	2	Marches	Student, strike, man, today, solidarity, march, video, square, protest, column
	3	Casual	About, pvt, application, become, this, year, on, tie, time, shrink

Table 7 Groups' events topics (translated from the Russian and Belarusian languages to the English language)

Spike	Number	Topic	Tokens
2	1	Work	Live, plan, power, bot, business, work, very, have, politics, city
	2	Protest	Know, people, go, protest, think, until, point, become, belarusian, what
	3	Political choice	Person, choice, necessary, people, say, year, country, can, speak, want
3	1	Protests	Man, people, need, day, very, know, go, protest, now, power
	2	Riots	Riot, people, country, street, guys, belarus, year, beat, party, truth
	3	Revolution attitude	President, military, therefore, violence, channel, thank you, peace, belarusians, everyone, long live belarus
4	1	Student protests	Man, country, know, student, say, speak, power, do, child, due
	2	Support	People, want, now, well done, think, can, thank you, live, children, question
	3	Protest activity	Need, people, strike, year, work, go, time, work, factory, watch

Table 8 Local chats' events topics (translated from the Russian and Belarusian languages to the English language)

Spike	Number	Topic	Tokens
4	1	Coordination	Now, chat, need, be able, know, work, do, news, year, write
	2	Neighbourhood	Want, think, photo, city, flag, entrance, group, word, good, neighbor
	3	Coordination	Man, today, people, very, see, video, go, go out, time, do

protests, and one can be related to the covid restrictions. We hypothesise that before the active protests, the central topics in Belarus was Covid and Covid restrictions. Afterward, after the clashes began, the protest activity became dominant and displaced Covid.

5.2.2 Topic modeling in a local context

To perform topic modeling for the specific spike in a given communication tool, we select a day in which a particular spike has occurred and then take all the messages three days before this day and three days after the spike. Table 6 shows the topic modeling per spike for the most active events in channels. Later, Tables 7 and 8 show the topics and the tokens for the local events in groups and in local chats. We can observe that despite some similarities in the topic, we can clearly observe significant differences that align with our initial hypothesis about the role of different mediums during protests in Belarus 2020.

5.3 Contextual difference

Finally, we analyze the context of specific proper nouns such as the names of politicians, famous protesters and particular places, for example, where the significant protests took place. To understand the context of the nouns, we train Word2Vec with a skip-gram model for each medium. The words' embeddings (words' dense representations) generated by Word2Vec with similar contexts tend to lie closer in the embedding space. The cosine distance is frequently used to calculate the distance between the vectors. We use this observation to understand the context of the proper nouns of interest by finding the closest top ten words to a given proper noun. Then we train the Word2Vec model using gensim¹⁰ package and specify the embedding size equals 100 and a window size equals to five. After the model is trained, we feed specific words (see Table 9, column *Words*). These words we translate to Russian, find top ten words in Russian, and then translate closest ten words back into English for the sake of non-Russian readers. The context of the words is presented in Table 9.

5.4 Discussion

We observe that the overlaps in the spikes dates become distinguishable using the topic modeling approach. At the same time, topic modeling confirms our initial hypothesis that separates each of the mediums into announcement related (channels), global coordination (groups) and local coordination (local chats). At the same time, the results obtained

after analyzing the global context cannot differentiate these communication tools into specific categories, as most of them intersect in terms of the tokens related to some protest activity and Covid restrictions.

6 Predicting mediums

Finally, in *RQ 3* we examine whether it is possible to differentiate a message among different communication mediums. For this, we build a classifier that predicts the type of the medium (channel, group or local chat) from a textual message. Firstly, we filter out the dataset to the messages with at least eight words (median) after the preprocessing (including stopwords removal) to remove noisy signals from the users. That results in 1,588,963 messages in total, with 1,192,794 messages belonging to local chats, 374,344 messages belonging to groups and 21,825 to channels. Then, we split randomly our dataset into train and test subsets with a ratio of 80% and 20%, respectively. Because of the significant data imbalance, a stratified split was used to preserve the same ratio of different classes in train and test. For the classification model, we choose logistic regression trained on unigrams and bigrams of TF-IDF features obtained from the preprocessed messages. While building TF-IDF features, we consider only those n-grams that have at least five occurrences in each medium. Given data imbalance problem in our data, we re-weight the same based on the ratio of positives and negatives. We use the "one vs all" approach, where we iteratively train three models for each medium and consider messages from the other two mediums as negative examples.

To understand the models' performance, we use the ROC AUC score and report the metrics in Fig. 3. We observe that the highest metrics (ROC AUC = 0.92) we obtain for the channels. At the same time, the ROC AUCs for local chats and groups is somewhat similar, approximately 0.69.

As metrics show, the messages from channels are very easily differentiable. This could be because admins in channels use different wording compared to a general Telegram user. At the same time, using simple wording features is not enough to differentiate local chats and groups. However, it again shows a similarity between these two mediums.

6.1 Error analysis

We also perform the error analysis to understand better what types of messages generate false positives of the model. We analyze the errors based on where the classifier failed the most. To rank the classification errors, we used the probabilities generated by the classifier for each example of each class. Then we sorted the examples by the error rate for each class and analyzed the top example. The local chats

¹⁰ <https://radimrehurek.com/gensim/models/word2vec.html>.

Table 9 Context of specific words

Words	Channels	Groups	Local chats
Lukashenka	Quarantine, first, hardly, which, treat, human, full, summer, less, return	Win, coming, lose, get out, present, join, shift, shift, weakening, facto	Quarantine, first, hardly, which, treat, human, full, summer, less, return
Tikhanovskaya	Recognize, result, count, recalculation, deputy, calculation, resignation, early, legitimacy, resolution	Tikhanovskiy, Babariko, Tsepka, Tikhanovskaya, Kanopatskaya, Tikhanovskiy, re-elections, Tikhanovskiy, challenger, remaining	Recognize, result, consider, recalculation, deputy, calculation, resignation, early, legality, resolution
Covid	Tradition, village, spokoino, maximum, tyrant, neighbor, gift, honor, Volodarka, dawn	Covid, postpone, fixed, contributor, vaccination, covid, coronavirus, corona, get sick, pcr	Tradition, village, spokoino, maximum, tyrant, neighbor, gift, honor, Volodarka, dawn
Minsk	Bring, tikhare, air, uniforms, policemen, paddy wagons, rear, guys, aggressively, drivers	Baranovich, column, gallery, Oryol, bangalore, miller, gomet, Kamenka, Rusets, Sukharevo	Bring, tikhare, air, uniform, police, paddy wagons, rear, guys, aggressively, drivers
Tikhanovskiy	Mass, European, use, social, general, origin, used, organized, transport, issue	Tikhanovskaya, Tikhanovskaya, Babariko, Tikhanovskaya, withdraw, tenacious, Tikhanovskiy, Ermoshin, register, Tikhanovskiy	Mass, European, use, social, general, origin, used, organized, transport, issue
Constitution	Crisis, change, non-recognition, value, strengthen, term, fair, hockey, introduce, violate	Constitution, amendment, referendum, consitutional, change, legal, parliament, reform, impeachment, empowerment	Crisis, change, non-recognition, value, strengthen, term, fair, hockey, introduce, violate
Protest	Mobs, payments, peaceful, participant, elemental, protest, flash mobs, join, support, manage	Protest, demonstration, marches, resistance, festivities, procession, confrontation, obedient, meeting, actions	Mobs, payments, peaceful, participant, elemental, protest, flash mobs, join, support, manage
March	Grand, thousands, commemorate, Sunday, procession, women's, line up, districts, christ, memory	Procession, marches, picket, rally, walk, action, resurrection, demonstration, gathering, Saturday	Grand, thousands, commemorate, Sunday, procession, women's, line up, districts, christ, memory
babariko	Race, check, drawing up, ground, judge, conclusion, pre-election, chancellor, cease, investigation	Tenacious, tikhanovskaya, tikhanovskiy, tikhanovskaya, tikhanovskaya, tikhanovskiy, candidate	Race, check, drawing up, ground, judge, conclusion, pre-election, chancellor, cease, investigation

Fig. 3 Metrics

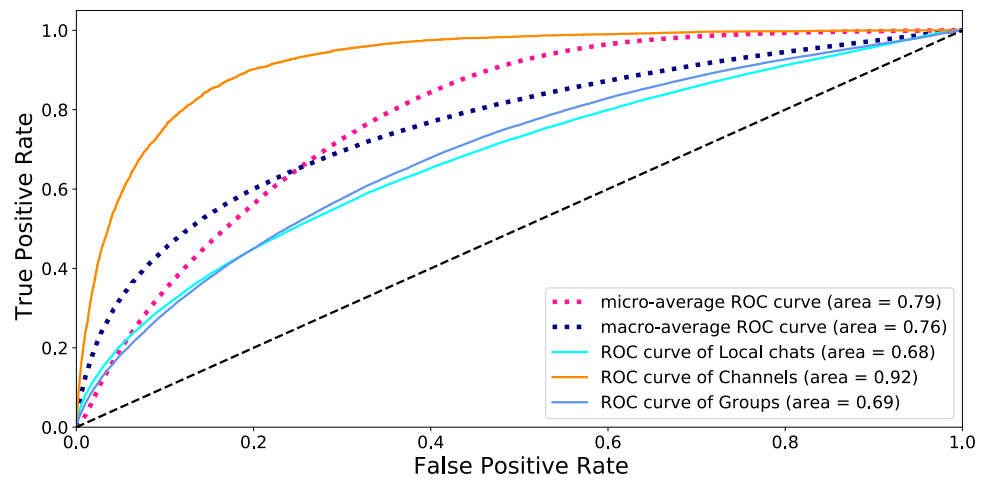


Fig. 4 Top features

y=0 top features		y=1 top features		y=2 top features	
Weight [?]	Feature	Weight [?]	Feature	Weight [?]	Feature
+6.460	горячие	+13.564	мкб	+5.508	мкб
+6.387	прислать	+10.458	котаны	+5.118	пп
+5.568	тнб	+8.345	бабарыкі	+5.001	думаю
+5.413	думаю	+7.923	кесулькендня	+4.571	що
+5.130	крыніца	+7.605	лукашэнка	+4.284	понимаю
+4.849	online	+7.342	фоты	+4.064	луки
+4.817	последствие	+7.097	ціханоўскай	+3.992	типа
+4.340	понимаю	+7.089	ціханоўская	+3.941	наогул
+4.310	чате	+6.901	каранавіруса	+3.905	инфа
+4.281	инфа	+6.578	urđ	+3.827	лука
+4.067	новость	+6.486	каранавірус	+3.622	смотрю
+3.892	предложить	+6.469	піша	+3.448	ня
+3.866	доброе	+6.464	чытач	+3.443	dissidentby
+3.863	зь	+6.395	цвк	+3.435	подскажите
+3.833	live	+6.380	белта	+3.427	майдана
+3.829	news	+6.300	мінску	+3.383	белоруссии
+3.742	flagshtok	+6.283	ціханоўскую	+3.376	крыніца
+3.608	предлагаю	+6.282	бабарыку	+3.333	elena
+3.533	чаты	+6.256	трёхпроцентного	+3.244	новость
+3.453	чата	+6.214	белаэс	+3.174	ды
+3.350	зубре	+6.000	мэс	+3.135	лулу
+3.273	этаже	+5.960	дэманстранты	+3.123	тады
+3.273	смотрю	+5.934	березино	+3.087	согласен
+3.237	типичная	+5.921	мінска	+3.083	дык
+3.237	согласна	+5.901	лукашэнка	+3.066	бойкот
+3.174	блин	+5.882	ант	+2.990	тыс
+3.162	ня	+5.875	акрэсціна	+2.967	це
+3.153	походу	+5.854	накопившиеся	+2.967	ви
+3.153	могу	+5.850	лігі	+2.963	блин
+3.135	сельмаш	+5.815	бабарыка	+2.929	жаль
+3.131	админ	+5.808	чытача	+2.924	луку
+3.127	инсайдер	+5.802	сказатб	+2.920	жабка
+3.122	id	+5.800	парада	+2.915	верно
+3.119	личку	+5.739	обнаружен	+2.899	лу
+3.114	подойду	+5.683	расказаў	+2.891	дарэчы
+3.104	согласен	+5.657	сутак	+2.880	мовы
+3.073	познакомиться	+5.624	памёр	+2.865	реально
+3.052	типа	+5.622	напоминаем	+2.845	украинцев
+3.034	подскажите	+5.613	советуем	+2.821	писал
+3.032	вроде	+5.562	скандуюць	+2.810	расеі
+3.020	гуляем	+5.520	паведамляе	+2.808	чате
+2.958	моему	+5.514	мінскі	+2.761	кремлебот

and groups are not distinguishable good enough based on the word TF-IDF because users there use casual language, which can be seen in Fig. 4. In addition, the top features from Groups and Local chats are heavily intersected. Finally, specific names of the locations are not occurred in the top words, probably due to their relatively low occurrences in day-to-day conversations that could potentially be useful in differentiating local chats and groups. At the same time, the channels' messages are yet very easily differentiable because they were had mostly the aim to inform people about some things and as it can be seen from the top words, they are mostly quite different from the rest of mediums. To this end, this is clear, that the words only are not yet enough to differentiate the mediums.

7 Discussion and conclusions

Computational social science is skewed towards either the English or Latin alphabet while paying little attention to other languages. This paper addresses this gap by studying online activities during the 6 months of the protest in Belarus. We collected data from three different communication tools (mediums) on Telegram: channels, groups, and chats. Our descriptive statistics and topic modeling show that people keep using social media after the active phases of the protest to discuss important political matters. A similar pattern was observed in Ukraine, where people used social media even after the end of the protest activities (Slobozhan et al. 2022). In what follows, we briefly discuss the results in terms of our three research questions.

Does the activity of users differ in mediums? The answer to this question is positive. Although the protest in Belarus relied on Telegram, protesters used it for different purposes depending on the medium (channels, groups, and chats). For example, users in groups mainly discussed announcements about national-level events. In contrast, local chats discussed local protests or demonstrations in particular neighbourhoods. What topics do users discuss in each medium? We observe that the topics vary by medium. Topics related to coordination were primarily raised in local chats (e.g., location and time of demonstrations), while channels and groups raised rather generic topics (e.g., news about the pandemic or Lukashenko's behaviour). While these findings are not surprising, they show that the online communication during the Belarus protests was well structured. Therefore, one should be careful when studying online communication on Telegram and consider analyzing mediums independently instead of blending them in one dataset. We also asked a question of whether users communicate distinctly in different mediums. In simple words, we wanted to understand if there are specific language patterns in each medium that can be easily recognized and predicted. It turns out that our models were able

to predict messages only from channels. At the same time, we were not able to differentiate messages from local chats and groups. Our interpretation of this finding is that the administrators of channels used templates for communication, they referred to similar sources and copied similar news, and perhaps were engaged in some coordination. Thus, messages in channels were more homogeneous in their topics and style, and our models were able to recognize them as belonging to the same category. In contrast, people who shared messages in local chats or groups were not homogeneous, they lived in different areas and cared about different (local) events. Respectively, they used some idiosyncratic language styles and references. Therefore, our models were not able to fit these messages into the same category. We believe that this finding is interesting because it shows that the communication from the administrators to broad masses (top-down communication) was well structured and perhaps coordinated, while the horizontal communication between local activists in local chats was more spontaneous and less structured. These findings provide new empirical evidence for the theory of "connective action", which is based on personalized content and is different from classic top-down communication. According to this theory, digital media facilitate "connective action" and influence the core dynamics of the protests (Bennett and Segerberg 2012).

Author contributions IS: Data Curation, Formal analysis, Investigation, Validation, Visualization, and Writing. TB: Problem formulation, and Writing—Review and Editing. RS: Supervision, and Writing—Review and Editing.

Funding This research is supported by H2020 SoBigData++ project (Grant Agreement No. 871042). and CHIST-ERA SAI project.

Data availability The dataset is collected from Telegram, and is sensitive in nature. Thus, dataset cannot be shared publicly.

Code availability Not applicable. As this work utilizes a non-shareable dataset.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Acemoglu D, Hassan TA, Tahoun A (2018) The power of the street: evidence from Egypt's Arab spring. *Rev Financ Stud* 31(1):1–42
- Akbari A, Gabdulhakov R (2019) Platform surveillance and resistance in Iran and Russia: the case of telegram. *Surveill Soc* 17(1/2):223–231
- Bastos MT, Mercea D, Charpentier A (2015) Tents, tweets, and events: the interplay between ongoing protests and social media. *J Commun* 65(2):320–350

- Bennett WL, Segerberg A (2012) The logic of connective action: digital media and the personalization of contentious politics. *Inf Commun Soc* 15(5):739–768
- Buzgalin A, Kolganov A (2020) The protests in Belarus: context, causes and lessons. *Crit Sociol* 47:441–453
- Enikolopov R, Makarin A, Petrova M (2020) Social media and protest participation: evidence from Russia. *Econometrica* 88(4):1479–1514
- Goel R, Sharma R (2020) Understanding the metoo movement through the lens of the twitter. In: *International conference on social informatics*. Springer, Berlin, pp 67–80
- Hansen A (2017) Public space in the Soviet city: a spatial perspective on mass protests in Minsk. *Nordlit* 39:33–57
- Herasimenka A, Lokot T, Onuch O, Wijermars M (2020) There's more to Belarus's 'telegram revolution' than a cellphone app. *Washington Post* 11
- Jost JT, Barberá P, Bonneau R, Langer M, Metzger M, Nagler J, Sterling J, Tucker JA (2018) How social media facilitates political protest: information, motivation, and social networks. *Polit Psychol* 39:85–118
- Karaliova T (2013) Two realities of one revolution: coverage of mass protests of 2011 in state-run and independent Belarusian media. *Media Transform* 10:118–140
- Mateo E (2022) "All of Belarus has come out onto the streets": exploring nationwide protest and the role of pre-existing social networks. *Post-Sov Aff* 38:1–17
- Metzger MM, Bonneau R, Nagler J, Tucker JA (2016) Tweeting identity? Ukrainian, Russian, and# Euromaidan. *J Comp Econ* 44(1):16–40
- Moshes A, Nizhnikau R (2021) The Belarusian revolution: sources, interim outcomes, and lessons to be learned. *Demokr J Post-Sov Democr* 29(2):159–181
- Mudrov SA (2021) Doomed to fail? Why success was almost not an option in the 2020 protests in Belarus. *J Contemp Central Eastern Eur* 29(1):109–120
- Nikolayenko O (2022) 'I am tired of being afraid': emotions and protest participation in Belarus. *Int Sociol* 37(1):78–96
- Onuch O (2015a) Euromaidan protests in Ukraine: social media versus social networks. *Probl Post-Communism* 62(4):217–235
- Onuch O (2015b) Facebook helped me do it': understanding the Euromaidan protester 'tool-kit. *Stud Ethn Natl* 15(1):170–184
- Piechota G, Rajczyk R (2015) The role of social media during protests on maidan
- Poell T (2014) Social media and the transformation of activist communication: exploring the social media ecology of the 2010 Toronto g20 protests. *Inf Commun Soc* 17(6):716–731
- Robertson G (2022) Protest, platforms, and the state in the Belarus crisis. *Post-Sov Aff* 38(1–2):146–149
- Sandoval-Almazan R, Gil-Garcia JR (2014) Towards cyberactivism 2.0? Understanding the use of social media and other information technologies for political activism and social movements. *Gov Inf Q* 31(3):365–378
- Scherman A, Arriagada A, Valenzuela S (2015) Student and environmental protests in Chile: the role of social media. *Politics* 35(2):151–171
- Schulze H, Hohner J, Greipl S, Girgnhuber M, Desta I, Rieger D (2022) Far-right conspiracy groups on fringe platforms: a longitudinal analysis of radicalization dynamics on telegram. *Converg Int J Res New Media Technol* 28(4):1103–1126
- Shelest O (2020) Local telegram groups (chats) in the summer-autumn of 2020: the dynamics of activity and the content of communication. *Center for European Transformation*
- Sinpeng A (2021) Hashtag activism: social media and the# freeyouth protests in Thailand. *Crit Asian Stud* 53:1–14
- Slobozhan I, Brik T, Sharma R (2022) Longitudinal change in language behaviour during protests: a case study of Euromaidan in Ukraine. *Soc Netw Anal Min* 12(1):1–12
- Smith BG, Men RL, Al-Sinan R (2015) Tweeting Taksim communication power and social media advocacy in the Taksim square protests. *Comput Hum Behav* 50:499–507
- Steinert-Threlkeld ZC, Mocanu D, Vespignani A, Fowler J (2015) Online social networks and offline protest. *EPJ Data Sci* 4(1):1–9
- Su CC, Chan M, Paik S (2022) Telegram and the anti-elab movement in Hong Kong: reshaping networked social movements through symbolic participation and spontaneous interaction. *Chin J Commun* 15(3):431–448
- Theocharis Y, Lowe W, Van Deth JW, García-Albacete G (2015) Using twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Inf Commun Soc* 18(2):202–220
- Tufekci Z, Wilson C (2012) Social media and the decision to participate in political protest: observations from tahrir square. *J Commun* 62(2):363–379
- Urman A, Ho J.C.-t, Katz S (2020) "No central stage": telegram-based activity during the 2019 protests in Hong Kong

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.