Data Collection and Sensemaking from Telegram: A Case Study of Ukrainian Political Leaders Channels and Chat Groups

Tuja Khaund, Mainuddin Shaik, and Nitin Agarwal

University of Arkansas at Little Rock, Little Rock AR 72204, USA
{txkhaund, mxshaik, nxagarwal}@ualr.edu

Abstract. Social media platforms such as Facebook, Twitter, etc., have become a valuable resource for marketing, public relations etc., over the years. One emerging mobile instant messaging medium, Telegram, has gained momentum over the years in countries such as Brazil, Indonesia, Iran, Russia, Ukraine, and Uzbekistan. Little work has been conducted on Telegram data which makes it a gold mine for researchers to explore and get insights, especially with an API to collect historical data. This paper takes a deeper look into the features of Telegram and gain insights using social network analysis along with text analysis. This paper analyses Telegram from a political perspective. Telegram is being actively utilized by Ukrainian parliamentary members to promote their work as well as ridicule their peers. As a result, channels are actively disseminating information of current political affairs and chat groups that discuss views on Ukrainian government are attracting more and more members to join. From our study, we conclude that Telegram has proved to be a rich data source to study social behaviors.

Keywords: Social Network Analysis, Sentiment Analysis, Telegram, Ukraine.

1 Introduction

Online social networks (OSNs) are dynamic social interaction platforms with billions of users worldwide. They attract people regardless of their age, gender, socioeconomic status, etc., and produce a tremendous amount of digital data for analysis [1]. The number of OSN users is increasing every year. According to the Pew Research Center's survey report, 65% of adult Americans use at least one social networking site [2]. Information is rapidly disseminated among these users through online social interactions. The interactions among OSN users generate a huge volume of data that provides the opportunity to study behaviors of our society [3]. An in-depth investigation of OSNs is important to enhance the understanding of the social and behavioral dynamics, as well as addressing pressing societal issues.

In recent years, Instant Messaging (IM) has become one of the fastest growing services among all other services provided by the mobile-based social media networks [4]. Owners of IM apps try hard to keep their current users and find new ones. Ease of use, high speed, minimized amount of spam, and reliability are the main reasons email is replaced by instant messengers. Telegram is the latest IM application emerging and

expanding very rapidly across the globe. Media Company "We are social" reported in the Digital 2019 Global issue that countries such as Iran, Ethiopia and Uzbekistan had more users from Telegram than any other IM applications[1].

1.1 What is Telegram and why should we study it?

Telegram is a cloud-based instant messenger founded by Russian entrepreneur Pavel Durov. It has more than 200 million monthly active users[2]. The new features introduced in Telegram such as channels, bots, super-groups, and advanced sharing mechanisms have risen up the instant messenger to a higher level [5]. A *group*, also known as *chat group*, on Telegram can have a maximum of 100,000 members. These groups can be used by companies and teams to coordinate work, families and friends to communicate, and ICOs to communicate effectively with their investors. Basic groups consist of 200 members while super-groups consist of 100,000 members and are optimized by the app developers to host large communities enabling super-groups to load quickly. *Channels* in Telegram can be used to broadcast messages to large audiences. A channel can have unlimited subscribers. Each post on a channel, has its own dedicated view counter, and the channels themselves have permanent URLs. Even a new subscriber can see all the messages ever shared on the channel, from day one. Also, links for any message of the channel, can be shared with other users, and once they open the link, they'll be taken to the channel, where they can view the message. On the plus side, if a channel is public, then the person who opens the link, doesn't require a Telegram account to view the message[3].

Telegram is now a new paradigm between social networks and instant messengers. It allows individuals to share news and information, coordinate political activity, and discuss politics [6]. Telegram is very popular and growing rapidly in countries such as Iran, Uzbekistan, Indonesia, Brazil, and Russia. In Russia, Telegram became the news hub for insider information and internal political discussions [7]. More recently, Telegram's popularity spiked in Ukraine. Amazon's Alexa Website Ranking service puts Telegram in the top 50 websites visited in Ukraine[4]. Telegram is not as strictly scrutinized as other social media outlets such as Facebook or Twitter, so it is hard to differentiate facts from misinformation.

Telegram offers two kinds of APIs for developers[5]. The Bot API allows users to easily create programs that use Telegram messages for an interface. The Telegram API and Telegram Database Library (TDLib) allow users to build their own customized Telegram clients and collect data. Both APIs are available for free.

In this paper, we take a deep dive into the Telegram messaging platform and explore their features for analysis. A medium that offers end-to-end encryption, an ability to

---

1 https://wearesocial.com/global-digital-report-2019

2 https://www.androidauthority.com/what-is-telegram-messenger-979357/

3 https://www.telegram-group.com/en/blog/the-difference-between-telegram-groups-and-telegram-channels/

4 https://www.alexa.com/siteinfo/telegram.org

5 https://core.telegram.org/api

self-destruct, no ads, high speed, and allows users to voice their opinions freely, should be studied. This paper analyses Telegram from a political perspective. Telegram is being actively utilized by Ukrainian parliamentary members to promote their work as well as ridicule their peers. As a result, channels are actively disseminating information of current political affairs and chat groups that discuss their views on Ukrainian government are attracting more and more members to join. Telegram has proved to be a rich data source for researchers to analyze and study social behaviors.

The rest of the paper is organized as follows. In the next section, we explore related work and examine how Telegram has helped other researchers to study group dynamics and misinformation, among other topics. Then we discuss our methodology in Section 3, where we collected Telegram data and performed social network analysis to study relations among actors of a chatroom. Section 4 presents our results and findings. Finally, we present our discussion and future work in section 5.

2 Literature Review

Existing research and studied cases have revealed how Telegram has been extensively used by individuals, educational organizations, companies and political parties for different purposes. Telegram features are used to study group dynamics to attract advertising and marketing companies to the platform and also help differentiate between high and low quality groups [8] based on users' social behavior. Bradshaw et al. [6] reported evidence of Telegram operating disinformation campaigns alongside other chat applications such as WhatsApp and WeChat. They identified evidence of political communication strategies such as targeting advertisements to specific segments of the population using demographic information or data on user attitudes, or gaming algorithms through search engine optimization techniques to make a content appear higher in search results. In 12 of 48 countries examined, the authors found evidence of cyber troop activity, with Iran being the most dominant. In a recent case study, the Atlantic Council's Digital Forensic Research Lab (DFRL) [7] analyzed the role of Telegram posts in the spread of misinformation about an upcoming prisoner exchange between Ukraine and Russia. These posts were subsequently confirmed, erroneously by a government official on Facebook. The DFRL analyzed the dissemination of the unsubstantiated reports and found that they first emerged in several anonymous Telegram channels, were later picked up by news outlets, and benefited from further amplification from influential public figures online. However, Russian media dedicated far less attention to the news of the alleged exchange and published brief reports, citing the Prosecutor General's Facebook page and Ukrainian media outlets.

Agur and Frisch [9] explored the catalytic effect of social media on digital and physical activism by interviewing participants in the Hong Kong's 2014 Umbrella Movement. They studied ways in which protesters used digital platforms during the debates about elections for Hong Kong's Chief Executive. While analyzing the extent of activists' social media usage to organize, mobilize, and persuade beyond the movement, Telegram played a significant role because of its robust security features. Protest leaders adopted Telegram for sensitive discussions, and developed guidelines for sensitive

deliberations, best conducted in person with phones stored in another location to protect against surveillance. Protest leaders were rarely responsive to Twitter DMs (direct messages) or email, however replied to Telegram's encrypted channels.

Ghaemi et al. [10] studied the impact of Telegram on Iranian Intermediate English as a Foreign Language (EFL) Learners. A PET proficiency test, two time-series vocabulary progress tests, and a post-test were applied to explore the effects of short messages services (SMS) on students' vocabulary learning process and the difference in mean scores was compared. The study revealed that Telegram as a teaching tool, had a positive effect on students' vocabulary learning. The experimental group obtained somewhat higher scores in the post-test than in two time-series progress tests, making the difference between the progress tests and post-test significantly different. Manna and Ghosh [5] conducted a comparative analysis of Telegram and WhatsApp to understand popular apps that provide library services with better facilities for modern library systems. Their study concluded that Telegram's ability to send bigger files like e-books, e-articles, or audio video lectures to its end user makes it the best choice for sending information and also a medium for communication with its users without any limitations.

3 Methodology

This section elaborates the steps followed to collect and process data successfully using the Telegram API. There are a few prerequisites to use the API:

1. Set up a Telegram developer account.
2. Extract *api_id*, *api_hash*, and *phone number* information to add to the code.
3. Once run, an authorization code will be sent to the Telegram app of the user (*Desktop version will not work*).

Once everything is set up correctly, the API will crawl and save all messages from all groups from a linked Telegram account. The initial data collection started in November 2019. The latest re-crawl date is February 09, 2020. Every time the script runs to collect data, it starts with the last message ID and appends the new data to the existing CSV file.

Data Collection. The first set of Telegram channels were identified manually through blogposts, tweets or Facebook posts. It is to be noted that some of the channel names had to be searched in Ukrainian as they were not found through English texts. We ended up with 8 active Telegram channels (see Table 1).

Table 1. Telegram channel information

Channel (ID)	Members	Message
Volodymyr Groysman (*volodymyrgroysman*)	102	37
Андрій Садовий (*andriysadovyi*)	287	158
Верховна Рада України (*verkhovnaradaukrainy*)	546	625
Гриценко 2019 (*grytsenko_2019*)	71	55

ЛЕЩЕНКО ТУТ (*LeshchenkoS*)	16741	633
Олексій Рябчин (*alexrbchnMP*)	302	186
Петро Порошенко (*PresidentPoroshenko*)	43060	644
Тимошенко Юлія підтримка (*Tymoshenko_Yulia*)	80	28

Using Telegram API, we extracted the following information from the channels:

1. message_id
2. message content
3. from_id (groups only)
4. datetime
5. url
6. reply_to_msg_id (groups only)

In order to extract URLs, we used an external Python library which immediately extracts and expands them. We extracted blogs and other social media links from these channels and found new Telegram channels (snowball sampling) which we added to our data collection pipeline. A few of these channels are listed on Table 2:

Table 2. New Telegram channels

Channel (ID)	Members	Message
Мустафа Найем (*mustafanayyem*)	19747	382
Sonya Koshkina (*sonyakoshkina*)	44001	1014
Шабунін депутатам (*Shabunin_RADI*)	1602	43
Володимир Зеленський (Президент України) (*PresidentZelenskij*)	6868	699
Легитимный (*legitimniy*)	66450	3110

We identified one chat group, Володимир Зеленський (*Volodimir Zelensky*) Chat, to conduct our social network analysis on a communication network. For the rest of the channels, we constructed timestamp frequency charts of posted messages to identify events of interest.

Translation. We used Google's API to translate all the Ukrainian text to English.

4 Analysis and Findings

In this section, we discuss the different analyses we performed with our dataset. There are two main features of Telegram, namely: Group and Channel. Since groups show interaction, we conducted social network analysis on their communication network. For channels, we extracted their text as well as their timestamp to identify events of interest.

4.1 Group Analysis

Social Network Analysis. We constructed a communication [user to user] network of the members in the chat group based on the '*reply_to'* property. We removed all entries where the '*reply_to_msg_id*' field was empty. We also discarded 80 entries, where the '*message_id*' was missing. This is due to the fact that if a user deletes any of their

messages, Telegram also deletes them from their servers. The final network consisted of 141 source nodes and 219 target nodes (see Figure 1).

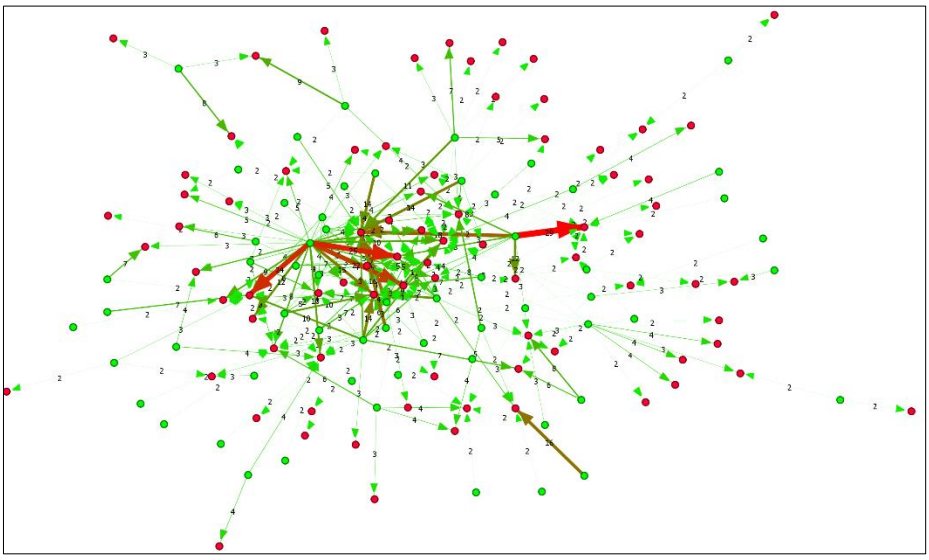

Fig. 1. User-User 'reply-to' network. Green nodes replied to red nodes.

The reply-to network is a directed network where the direction of the arrow travels from the source node to the 'replied-to' node. The source nodes are colored green and the target node is in red. The edges are weighted based on the number of replies between them. The edges are colored from green to red which corresponds to a weight range from low to high. This helps us identify actors who interact frequently with other members of the chat group. We found four active users including the group admin who made it to the top of the list:

1. Василина Дудла
2. D S
3. Makemake
4. Володимир Зеленський Chat (Admin)

For each of these users, we extracted their messages and conducted sentiment analysis using Linguistic Inquiry and Word Count (LIWC)[6]. The overall sentiment was positive. Василина Дудла's messages were highly positive while the rest of the users had neutral sentiments. The positive messages were in favor of president Zelensky showing praises and respect. Neutral sentiments could suggest either a lack of emotionality or different levels of ambivalence towards the members of the chat group or the content. While this analysis shows the overall sentiment of these users' interactions, it has a few limitations. For example, it is hard to detect whether these sentiments are addressed towards

[6] http://liwc.wpengine.com/

a member of the chat group or to the subject of discussion, president Zelenskiy. In that case, a targeted sentiment analysis will help us get better insights on these sentiments which will be conducted in our future work.

To understand what these prominent users are saying, we analyzed word clouds for their chat history (see Figure 2). Even though words such as war and death appear in our results, prominent words include Zelensky, Support, God, etc. This shows that the top actors were speaking in support of their president and comparing him to God. The bottom-left word cloud discusses various bills that was introduced by the Ukrainian government.

Fig. 2. Word cloud of the messages of the top four users, Василина Дудла (top-left), D S (top-right), Makemake (bottom-left) and Володимир Зеленський Chat (bottom-right) of the Zelensky chat group.

4.2 Channel Analysis

Telegram channels have limited forms of interaction between members. The admin is the content producer and its members are the communicators. Communicators can like

or dislike a post, but that feature is not enabled in all channels. So, we have ignored the analysis of likes/dislikes for this study. The message content of these channels was of interest. It is to be noted that we translated the messages from Ukrainian to English.

Posting Frequency. This analysis enabled us to identify events of interest based on user activity. For example, the activity of channel Олексій Рябчин (*alexrbchnMP*) peaked during the months of November 2018, January 2019, and May 2019. To understand the cause of these spikes, we created word clouds of all the messages that were shared on the channel during those three months. Ukraine introduced a new Martial Law in November 2018 and majority of the discussion revolved around it. Discussion in February involved President Poroshenko and talks about his impeachment bill. A few messages covered his speech during the NATO Summit at Brussels, Belgium. Many conversations were predominantly about Zelenskiy's inauguration and his upcoming roles as the new president of Ukraine. We conducted similar analyses on the other channels to understand their purpose and observed similar trends in posting during these three months. One channel 'grytsenko_2019' only posting messages during the month of February and May. This form of analysis helps understand the overall channel's user activities and a deeper analysis helps us understand the online discourse about a physical real-world event.

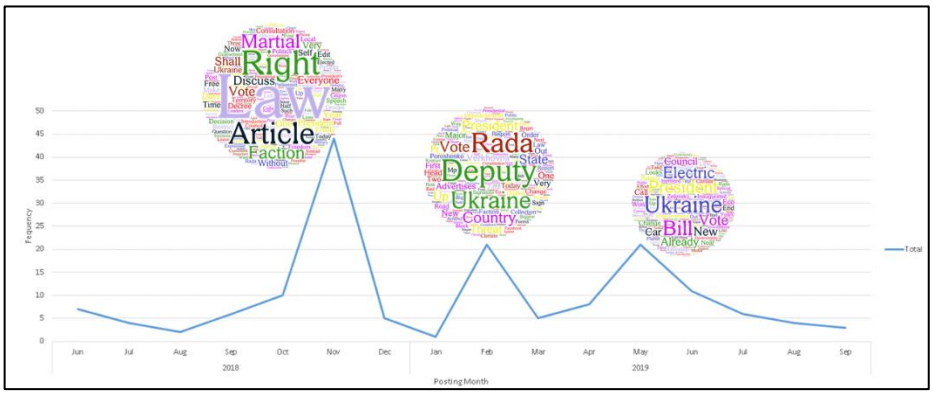

Fig. 3. Frequency of messages posted on the **Олексій Рябчин (*alexrbchnMP*)** channel. Word clouds of the messages for the peak months are also added.

Cross-media Link Analysis. Upon analyzing the links on our channel list, we found several URLs directing to various social media platforms. We extracted these links and ran a quick check to identify common URLs across different channels. This shows how social media platforms play different roles in information dissemination - platforms that host the information and the ones that drive users or views to the platform that is hosting the content. This is a common tactic used by most successful information campaigns. The common links shared by these Telegram users are shown in Table 3. For example, the YouTube link shared belongs to the official channel of ***Bihus.Info*** project, an anti-corruption journalistic investigation team. This video discusses an investigation led by

Lesia Ivanova about the secret correspondence of Poroshenko's entourage. The roll-backs and the scheme to steal millions on the defense, Russian smuggling and money laundering with the help of the son of one of the most influential officials and a friend of the President was covered in large scale.

Table 3. Cross-media Information sharing

Channel (ID)	Blogs	YouTube, Instagram
mustafanayyem	thebabel.net	youtube.com/watch?v=lGTf2nUyxfw
sonyakoshkina	lb.ua	
Shabunin_RADI	nv.ua	
alexrbchnMP	liga.net	
LeshchenkoS	epravda.com.ua	instagram.com/sergii_leshchenko/
PresidentPoroshenko	pravda.com.ua	
verkhovnarada	telegra.ph	

5 Discussion and Conclusions

Telegram is very popular and growing rapidly in countries such as Brazil, Indonesia, Iran, Russia, Ukraine, and Uzbekistan. Telegram is a rising platform for misinformation and social manipulation. However, it also plays a positive role in teaching students foreign languages and also provides better facilities of modern day library systems. There are several ways researchers can utilize Telegram to conduct studies that can benefit the society. It is a rich source of information and its ability to access historical data gives more flexibility to researchers to conduct in-depth analyses. In this study, we demonstrated a methodology to collect and analyze data from Telegram groups and channels. Analysis of these groups and channels helped us understand Ukrainian political affairs and people's reactions. For our chat group, we received positive reactions from its members praising their president.

For immediate future work, we would like to extend our work to study targeted sentiment analysis and examine community dynamics within the chat groups. Given Telegram's popularity in certain regions, we plan to investigate cyber-cultural influences on communications, as our long-term research.

Acknowledgements

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-17-S-0002, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock,

and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

1. Raghavan, S.: Digital forensic research: current state of the art. CSI Trans. ICT. 1, 91–114 (2013). https://doi.org/10.1007/s40012-012-0008-7.
2. Perrin, A.: Social Media Usage: 2005-2015. , Washington, DC (2015).
3. Ratkiewicz, J., Conover, M., Meiss, M.R., Gonc¸alves, B., Flammini, A., Menczer, F.: Detecting and Tracking Political Abuse in Social Media. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. pp. 297–304 (2011).
4. Sivabalan, K., Ali, Z.: Mobile Instant Messaging as Collaborative Tool for Language Learning. Int. J. Lang. Educ. Appl. Linguist. 99–109 (2019). https://doi.org/10.15282/ijleal.v9.297.
5. Manna, R.A., Ghosh, S.: A COMPARATIVE STUDY BETWEEN TELEGRAM AND WHATSAPP IN RESPECT OF LIBRARY SERVICES. Int. J. Libr. Inf. Sci. 7, (2018). https://doi.org/10.34218/IJLIS.7.2.2018.001.
6. Bradshaw, S., Howard, P.N.: Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation. 26.
7. @DFRLab: Ukrainian media jump the gun on Russia-Ukraine prisoner swap, https://medium.com/dfrlab/ukrainian-media-jump-the-gun-on-russia-ukraine-prisoner-swap-3216f860bc04, last accessed 2020/02/11.
8. Hashemi, A., Zare Chahooki, M.A.: Telegram group quality measurement by user behavior analysis. Soc. Netw. Anal. Min. 9, (2019). https://doi.org/10.1007/s13278-019-0575-9.
9. Agur, C., Frisch, N.: Digital Disobedience and the Limits of Persuasion: Social Media Activism in Hong Kong's 2014 Umbrella Movement. Soc. Media Soc. 5, 2056305119827002 (2019). https://doi.org/10.1177/2056305119827002.
10. Ghaemi, F.: The Impact of Telegram as a Social Network on Teaching English Vocabulary among Iranian Intermediate EFL Learners. Int. J. Inf. Commun. Sci. 2, 86 (2017). https://doi.org/10.11648/j.ijics.20170205.15.