

## Method for identifying Twitter accounts that have changed their opinion about politicians

© Taras Rudnyk<sup>[0000-0001-9492-0374]</sup> and © Oleg Chertov<sup>[0000-0003-0087-1028]</sup>

National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine

tarasrudnyk@gmail.com, chertov@i.ua

**Abstract.** Nowadays, social networks provide an opportunity to communicate with the world and share political views. Therefore, each politician has an account and many followers. Revealing political opinions on Twitter accounts can be a good alternative to regular social surveys. Identifying those who changed their opinions will help to understand the reasons for disappointment in a particular political figure. In this paper we introduce our approach to collect data from Twitter, detect accounts political views, opinion changes and the reasons people get frustrated. The experimental results showed that such research was successful in the analysis of Ukrainian President supporters.

**Keywords:** Twitter, social network, Ukraine, politics, machine learning, natural language processing, sentiment analysis.

### 1 Introduction

Social networks have caught the attention of people around the world. It is hard to imagine modern times without Twitter, Facebook or Instagram. While most users connect with friends and use social networks for entertainment others see them as a tool to gain authority or profit. The most obvious examples are politics and business. Every politician has a page on at least one social network. This is a kind of close communication with people, which sometimes even replaces social media. The approach has several advantages such as speed, convenience, and a reputation as a trusted source. Among the benefits for subscribers are analytical benefits. For example, using a business Twitter account, you will know exactly how many times people have seen, retweeted, liked, and replied to each tweet. Even more, you can export the data as a CSV file in order to perform advanced analytics. Instagram Insights tool can show you when your audience is online, which of your posts are most popular, impressions and reach of your account. Also, you can compare your original post with a promoted version of the post to see if your paid efforts are working. People increasingly tend to express their political preferences in social networks.

The automated approach has several advantages over traditional methods. Benefits include speed, price and visibility.

*Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).*

When we talk about social media as a political tool, the first thing that comes to mind is Donald Trump and Twitter. Trump is successfully using social media to rally his supporters and attract new ones [1]. The number of monthly Twitter users worldwide is growing continuously and as of the first quarter of 2020 has an average of 340 million monthly active users [2]. There is no doubt that politicians can influence large numbers of people who use Twitter on a daily basis.

Ukrainian president Volodymyr Zelenskyy won the presidency on April 21, 2019. The Twitter page was created in April 2019 and the very first tweet was written on April 25. Taking to account that Volodymyr Zelenskyy was chosen by 73.22 percent of voters, support in those days was overwhelming. Twitter was not a part of the elections campaign, but as we can see now, it is an important part of communicating with people. Ukrainian social polls and the media regularly talk about the decline in the president's ratings. The question this paper is addressing is the following: can we collect data from Twitter and use it as an alternative to conventional approaches to predicting election results and understand the most important reasons for people's choice? Is it possible to find people who have changed their opinions about Volodymyr Zelenskyy and his political party "Servant of the People"? Can we figure out which events affect people the most?

The plan of the paper is as follows. Section 2 presents the related paper and existing solutions for similar problems. Section 3 explains our approach to finding accounts that have changed their opinions about politicians. Starting with the identification of supporters, and their classification according to opinions that change over time. The section ends with an approach to identifying frustrated people through follower dynamics. Section 4 explains in detail the research carried out and the experimental results. The last section concludes the discussion of plans for future work.

## 2 Related Paper

The specificity of social media is manifested in the fact that, as a rule, only the most active members of society declare their political preferences. Nevertheless, some studies show that it is in social networks and largely due to the availability of its appropriate tools (setting preferences, filters, search history, etc.) the so-called "echo chambers" and "epistemic bubbles" are formed [3, 4]. And people who find the support of their like-minded people in such communities tend to be less active in discussing their political preferences offline, for example, in their workplace [5]. In fact, data for only the part of society that is most active on social media is available for research. But this data can give us a correct picture of the mood in the whole society, provided that fake news and accounts are weeded out [6].

Twitter is perhaps the most popular social networking site for research. This can be explained by three reasons: 1) the small size of the tweets, which makes them easier to collect and process; 2) active use of this social network by US President Donald Trump and a number of other politicians; 3) availability of accessible API. Unfortunately, in recent times, many of Twitter's API capabilities have been blocked or cut off.

From the point of view of studying the political preferences of Twitter users, there are two main approaches in the academic literature: based on sentiment analysis and based on the study of account links [7].

Various sentiment analysis methods that have been proposed for eliciting political opinions of Twitter users can be divided into three groups [8]: untargeted/overall sentiment, targeted sentiment, and stance detection methods.

Untargeted methods are based on either hand-curated [9, 10] or supervised learning-selected [11, 12] lists of positive and negative words. The final rating of a tweet takes into account the (possibly weighted) ratio of these two types of words.

Unlike untargeted sentiment analysis methods, target methods have a predefined person, entity or topic about which the sentiment of the author of the text should be determined [13]. In turn, stance detection methods differ from target sentiment analysis methods in that the target is indicated in them, but implicitly [14].

Hence, the sentimental analysis approach has two significant drawbacks:

- 1) the vast majority of Twitter users write very little themselves; for example, according to a 2018 Pew Research Center survey, 90% of American users wrote two or fewer tweets per month [15];
- 2) back in 2013, it was discovered that discriminating values of features for politically active and modest Twitter users should be different [16].

Article [17] was a pioneering study showing the importance of exploring the structure and internal links of social networks using the example of blogs. Different online social networks have their own means of expressing structural links between their members. However, in all of them, you can build social embedding either by specifying friends, or by specifying users followed by a given account [18].

Twitter has a number of additional features that can be viewed as associative links between accounts, namely: each Twitter account can answer, cite, comment or retweet a tweet from another one. A somewhat unexpected result was obtained in article [19]: it turned out that for political debates, the study of retweets is more important than the other “links” mentioned.

We use a combination of two approaches, downloading not only tweets about a specific topic but also account (profile) data. We download those who showed activity on the page of Volodymyr Zelenskyy or subscribed to him. After that, we download as much information as possible from the page of such users, which contains all the profile data and all the tweets. Similar approach was used in [20] to detect phony accounts of Ukrainian politicians on Facebook pages. This approach was based on finding groups of accounts that leave mostly positive comments, but at the same time start writing negative ones.

One of the reasons for constraining research in this domain is the lack of publicly available complete datasets containing the necessary information. This situation can be explained, on the one hand, by the complexity of collecting relevant data, since social networking APIs have a number of significant limitations. On the other hand, this is due to the strict requirements of the legislation to protect the privacy of personal data.

### 3 Method for finding accounts that have changed their opinion about a politician

#### 3.1 Problem formulation and solution

In this work, we develop an automated approach to find Twitter accounts that first supported politician but changed the opinion about him/her after some time. We assume that, during or immediately after elections, people express support or hope for politicians. After a while, frustration appears associated with unfulfilled promises or unexpected actions. We are trying to find the accounts of such people on Twitter. The search algorithm can be divided into several stages:

- *Data collection.* We collect as much data as possible about the politician and his/her followers from Twitter pages.
- *Pre-processing.* All data is converted into a uniform format. We use various data transformation techniques, such as converting date and time into one format, discarding stop words, stemming and lemmatization. After that we create datasets for further analysis.
- *Analysis.* In our case it is an iterative process of detection politician supporters, analyzing their personal information, then tweets and classification account to one of four groups: not active, do not show political views, supporters and haters. Iterations can be repeated an infinite number of times as we expand our datasets and people can change their opinions multiple times.
- *Reporting.* A summary of analysis results with visual representations.

#### 3.2 Advantages of automated approach

Compared to traditional methods for determining the ratings of politicians, the automated method has the following advantages:

- *Speed.* The approach allows you to track rating changes in real time and find the reasons. Social polls are carried out with some frequency, but it takes time to process the data, so for some time the public does not know the results and does not know what could cause an increase or decrease in ratings.
- *Price.* The developed algorithms allow to reduce the number of people working for the result. There is no need to interact directly with people and most of the work will be done by machines.
- *Visibility.* A visual representation of the data allows an understanding of how many people have changed their opinions and reasons.

#### 3.3 Twitter API and Selenium web driver

Twitter has official API which allows to download personal information of accounts and all tweets. To collect research data we used Python library tweepy which is based on official Twitter API. Not all data can be fetched as API returns only fresh tweet replies. For our research we need to collect replies left even more than one year before. As such information is available on user interface then we can imitate browser behavior

and find necessary data. For such purpose we use Python Selenium web driver and performed html parsing with downloaded data. Firstly, we download all Volodymyr Zelenskyy tweet identifiers using tweepy. Then Selenium opens each tweet scrolls down and extracts replies until the end of scroll line. All information stored to CSV datasets in order to perform analysis. System diagram is described on Fig. 1. As input user provides Tweeter account link and type of interested information. Based on inputs Python controller decides download account personal information such as user name, followers count, following count, location, description etc. or extract list of tweet identifiers and download all replies including their authors. All datasets are stored in CSV storage for advanced analysis.

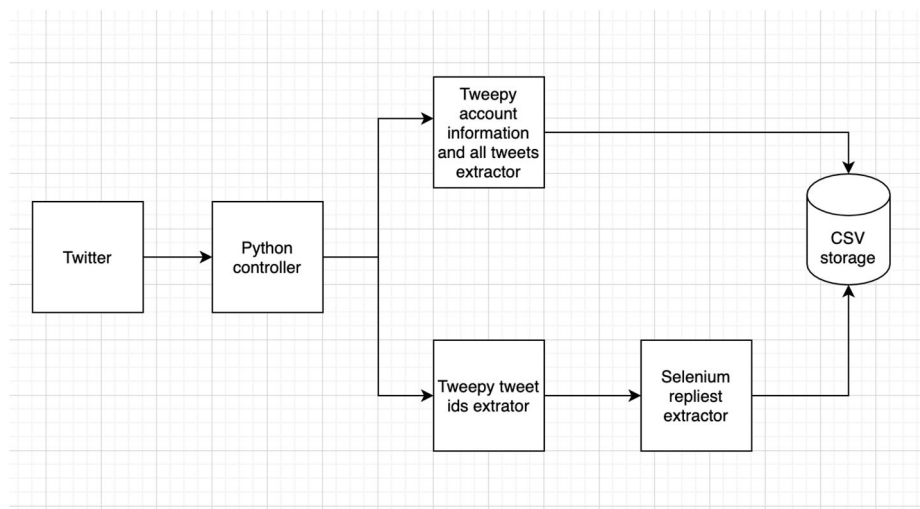


Fig. 1. Architecture of Twitter extractor

### 3.4 Identification of Volodymyr Zelenskyy supporters

To find accounts that have changed political positions, first of all we should find those who supported the political leader or party. Volodymyr Zelenskyy won the presidential elections in Ukraine on April 21, 2019. The Twitter page was created in April 2019 and very first tweet was written on April 25.

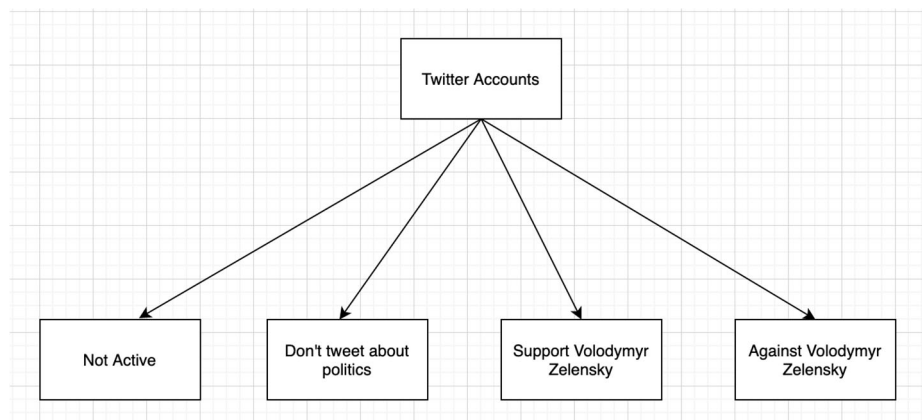
The approach of finding Zelenskyy supporters can be described as follows:

- Collect replies to the very first tweet of the President of Ukrainian.
- Detect language of replies and filter out replies not in Ukrainian, Russian or English as unpopular languages in Ukraine.
- Convert all messages to lower case, get rid of duplicated, stop words and punctuation marks.
- Transform all words to the root form – perform lemmatization.
- Create a set of words that can be confidently attributed to the supporters of Volodymyr Zelenskyy.

- Create word vectors for all tweets and the set of words.
- Measure how similar are tweet replies and our set of words using cosine similarity.
- Create a visual representation of classified messages.

### 3.5 Identification of accounts that changed political views

The next step is to find the accounts that have changed their opinion. In order to do this, we should take a closer look into the tweets of Volodymyr Zelenskyy supporters and find those who tweeted or retweeted some messages against the current president or his political party “Servant of the People”. Different people can be affected by different news, so the reasons can vary. Some news is very emotional and can affect people more than others. In most cases, a person does not change his opinion immediately; this may be the result of a series of unpleasant news. In this case, the last blow will be the latest news before the change of opinion. The initial supporters of the president can be divided into four groups, as shown in Fig. 2.



**Fig. 2.** Classification of the initial supporters of Volodymyr Zelenskyy

The search algorithm can be described as follows:

- Download all the tweets of detected supporters of Volodymyr Zelenskyy.
- Refer to a separate category those accounts that stopped tweeting on April or May 2019.
- By analogy with the previous algorithm, convert all messages to lower case, get rid of duplicates, stop words, punctuation marks and perform lemmatization.
- Create set of words which can be confidently classified as belonging to politics.
- Among the political tweets, those can be detected that are against Volodymyr Zelenskyy or “Servant of the People”, using separate set of words.
- Create a visual representation of results.

### 3.6 The unsubscribed accounts

Another way to find accounts that have changed their minds is to take a closer look at the presidential followers. On Twitter, Volodymyr Zelenskyy always has the dynamics of subscriptions. Some people do not show changes in their opinion but have stopped watching the actions of the president. We assume that the person who is disappointed by the incumbent president can unsubscribe from his page.

## 4 Experimental results

### 4.1 Volodymyr Zelenskyy supporters

First tweet contains 351 replies. Using Python library – fasttext, tweets were divided into four parts: Ukrainian, Russian, English and other. The last part was filtered out as it is very rare case. Tweets contain unnecessary objects like punctuation, stop words, and mentions that can affect the performance of an algorithms. For each tweet we filter common words that add noise or provide no value for analysis. Examples: the, a, an, in. Python library – nltk contains stop words for all interested languages. Also, punctuation and mentions were removed. Each word was converted to its base form with help of nltk.

Supporting tweets contain specific words. We build the set of words that can be confidently classified as belonging to Volodymyr Zelenskyy supporters. Such sets of words built for three interested languages. Examples of such words: congratulations, hope, trust, expect, adore, bow, courage, at last, reliance, stick it.

Classification performed by converting all words into vectors and measuring cosine similarity between tweet and our set of words.

As a result, 84 out of 351 accounts support Volodymyr Zelenskyy (see Fig. 3).

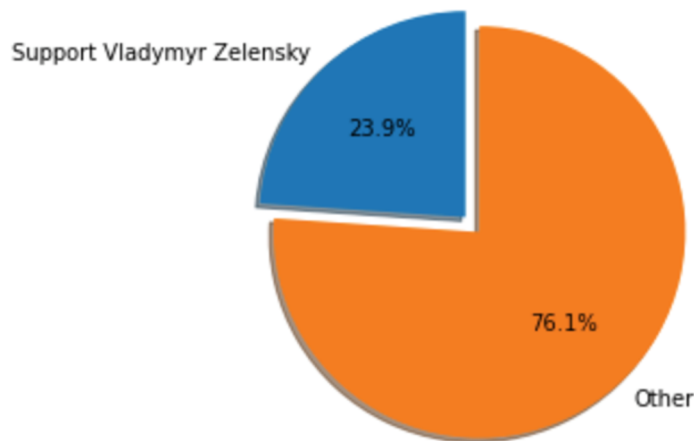


Fig. 3. Percent of Volodymyr Zelenskyy supporters under first tweet

## 4.2 Accounts that changed their political opinion

130 416 tweets were collected from 84 accounts that initially supported Volodymyr Zelenskyy. There is a group of accounts that became inactive after replies on the Presidents page. They could be filtered out simply by date of last tweet. 29 such accounts were removed from dataset as those who do not need additional analysis.

Second category of accounts are those who do not tweet about politics. In order to separate this group from those who require further investigation, we created a set of words related to Ukrainian politics. Example of words: Zelenskyy, Tishchenko, Poroshenko, Klitschko, Servant of the People, Fatherland, Opposition Platform – For Life, European Solidarity, For the Future, Our Land. Set contains long and shortened names of famous politicians and political parties. Same sets were created for Ukrainian, Russian and English languages. As a result, 26 out of 84 accounts were not politically active.

To separate the last two categories, we have reduced the set of words to those related to Volodymyr Zelenskyy and his party. In order to detect user opinion, we build a mapping of most annoying news about the president and "Servant of the People" decisions. Such as corruption, nepotism, lobbying for the interests of oligarchs and anti-Ukrainian laws. If a person tweets or retweets such news, then most likely he is against the current president. Using this approach, we found retweets of the following news:

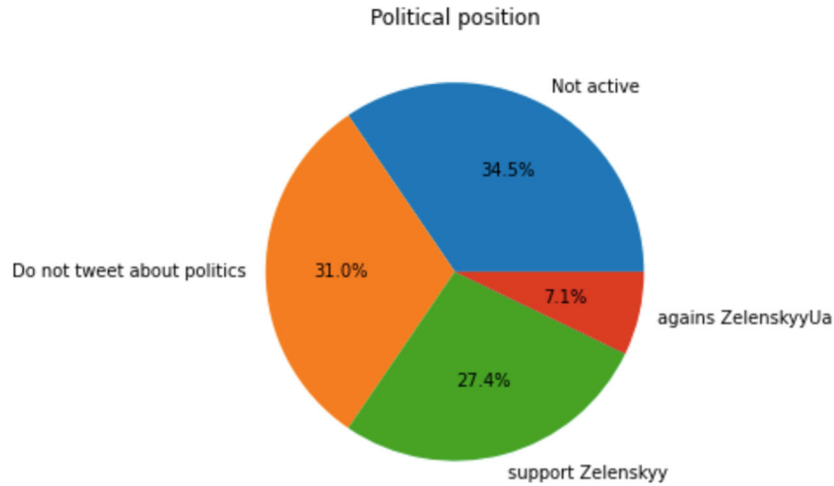
- Despite promises, nepotism spread in the Servant of the People.
- Corruption scandal associated with the deputy Alexander Yurchenko, who is a member of the Servant of the People party.
- Vladymyr Zelenskyy's interaction with Ukrainian oligarchs.

This finding confirms the importance of the retweets described in section 2.

The technique made it possible to identify those who were against the current President of Ukraine. As a result, 6 accounts were discovered, and the remaining 23 still support Volodymyr Zelenskyy.

Percentage of political positions of accounts which initially supported president of Ukraine is shown on Fig. 4.





**Fig. 4.** Percentage of political positions of accounts which initially supported president of Ukraine

### 4.3 Subscribers analysis

To detect those accounts that are disappointed in Vladimir Zelenskyy, we download the identifiers of all the president's followers on Twitter for 28 days. The overall dynamics show that the total number of subscribers is growing. This may be a result of the growing popularity of Twitter. The rate of new subscriptions has slowed down a lot over time. The more important fact is that people continuously unsubscribe with almost the same velocity (see Fig. 5). Loss of interest from subscribers can be the result of unjustified expectations. The following Table 1 gives a summary of subscriptions dynamics.

**Table 1.** Subscribers dynamics on Vladymyr Zelenskyy page

Date	Total Subscribers	New subscribers	Unsubscribed
08.11.2020	240408		
15.11.2020	244984	5472	896
22.11.2020	247146	2866	704
29.11.2020	248932	2514	728
06.12.2020	249869	1646	709

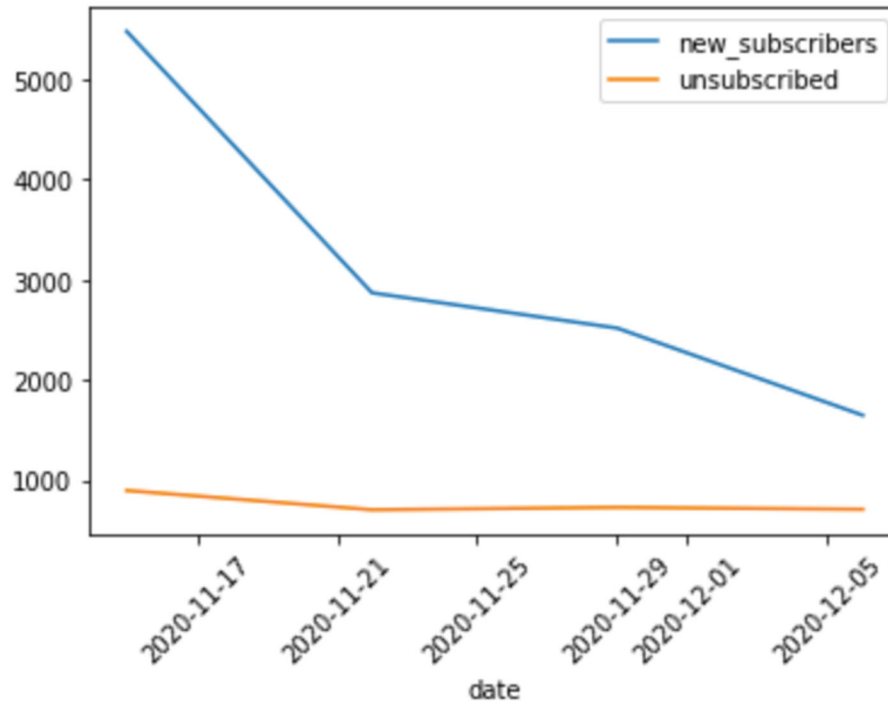


Fig. 5. Dynamics of subscriptions over time

It should be noted that for now period of time is pretty small and such accounts require further investigations. In future work we are planning to download personal information and all tweets of unsubscribed accounts.

## 5 Conclusions

In this research work, we have demonstrated an approach to collect Twitter dataset and classify accounts accordingly to their opinion about Volodymyr Zelenskyy and his political party "Servant of the People" during the period of his Twitter account creation to the 29<sup>th</sup> of November. Proposed approach allowed to detect four groups of people who initially were supporting Ukrainian president. Majority of accounts – 34.5 percent stopped any activity right after they replied to Volodymyr Zelenskyy first posts. It seems that they were created only with purpose of president popularization. Second group – 31 percent became politically not active and we could not classify their opinion. There is still big group of accounts – 27.4 percent which continue support of Volodymyr Zelenskyy. The last group – 7.1 percent are those who completely changed opinion about Ukrainian president which reflects results of local elections in Ukraine and social polls.

Such approach could be used not only for politics but in all fields where popularity matters such as companies and brands.

Another part of this research work is the analysis of the dynamics of subscriptions on the page of Volodymyr Zelenskyy. At the moment we have identifiers of subscribers for the last 4 weeks. In future work, we will extend the observation period. In addition, we plan to download personal information and all tweets of unsubscribed accounts in order to more accurately find out the reasons.

**Acknowledgment.** This study is funded by the NATO SPS Project CyRADARS (Cyber Rapid Analysis for Defense Awareness of Real-time Situation), Project SPS G5286.

## References

1. Ott, B. L., Dickinson, G.: *The Twitter Presidency: Donald J. Trump and the Politics of White Rage*. Routledge, Taylor & Francis, New York (2019).
2. Hootsuite. DIGITAL 2020. Global Digital Overview, <https://hootsuite.com/pages/digital-2020>, <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users>, last accessed 2020/12/01.
3. Dahlgren, P. Media, Knowledge and Trust: The Deepening Epistemic Crisis of Democracy. *Javnost - The Public* 25(1-2), 20–27 (2018).
4. Nguyen, C. T. Echo Chambers and Epistemic Bubbles. *Episteme*. 17(2), 141–161 (2020).
5. Hampton, K. N., Shin, I., Lu, W. Social media and political discussion: when online presence silences offline conversation. *Information, Communication & Society*, 20(7), 1090–1107 (2017).
6. Gurajala, S., White, J. S., Hudson, B., Matthews, J. N. Fake Twitter accounts: Profile characteristics obtained using an activity-based pattern detection approach. In: *SMSociety '15: International Conference on Social Media & Society*, ACM, Toronto (2015).
7. Gaumont, N., Panahi, M., Chavalarias, D. Reconstruction of the socio-semantic dynamics of political activist Twitter networks—Method and application to the 2017 French presidential election. *PLoS ONE* 13(9): e0201879 (2018).
8. Sen, I., Flöck, F., Wagner, C. On the Reliability and Validity of Detecting Approval of Political Actors in Tweets. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1413–1426. Association for Computational Linguistics (2020).
9. O'Connor, B., Balasubramanyan, R., Routledge, B. R., Smith, N. A. From tweets to polls: Linking text sentiment to public opinion time series. In: *Proceedings of the Fourth International Conference on Weblogs and Social Media (ICWSM 2010)*, DBLP, Washington (2010).
10. Pasek, J., Yan, Y. Y., Conrad, F. G., Newport, F., Marken, S. The stability of economic correlations over time: Identifying conditions under which survey tracking polls and twitter sentiment yield similar conclusions. *Public Opinion Quarterly*, 82(3), pp. 470–492 (2018).
11. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., Potts, C. Recursive deep models for semantic compositionality over a sentiment tree-bank. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642, Seattle, Washington, Association for Computational Linguistics (2013).

12. Thelwall, M. The heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In: Holyst J. (eds) *Cyberemotions. Understanding Complex Systems*, pp. 119–134. Springer, Cham (2017).
13. Tang, D., Qin, B., Liu, T. Aspect level sentiment classification with deep memory network. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 214–224, Austin, Texas. Association for Computational Linguistics (2016).
14. Mohammad, S. M., Sobhani, P., Kiritchenko, S. Stance and sentiment in tweets. *ACM Transactions on Internet Technology*, 17(3):26, pp. 1-23 (2017).
15. PEW RESEARCH CENTER. Survey of U.S. adult Twitter users conducted Nov. 21 - Dec. 17, 2018. Data about respondents' Twitter activity collected via Twitter API. "Sizing Up Twitter Users", [https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/pdl\\_04-24-19\\_twitter\\_users-00-06/](https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/pdl_04-24-19_twitter_users-00-06/), last accessed 2020/12/01.
16. Cohen, R., Ruths, D. Classifying Political Orientation on Twitter: It's Not Easy! In: *International AAAI Conference on Web and Social Media*, pp. 91–99 (2013).
17. Adamic, L. A., Glance, N. The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In: *Proceedings of the 3rd international workshop on Link discovery (LinkKDD'05)*, ACM, pp. 36–43 (2005).
18. Volkova, S., Bachrach, Y., Van Durme, B, Mining User Interests to Predict Perceived Psycho-Demographic Traits on Twitter. In: *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*, Oxford, pp. 36-43 (2016).
19. Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M. Quantifying Controversy in Social Media. In: *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM'16)*, pp. 33–42. New York, ACM Press (2016).
20. Chertov, O., Rudnyk, T., Palchenko, O.: Search of phony accounts on Facebook. Ukrainian case. In: *International Conference on Military Communications and Information Systems (ICMCIS)*, pp. 22–23, Warsaw (2018).