Current Trends in Computing
Vol.1 No.1 pp 75-82 (2023)
ISSN: 2980-3152
Received: 13.07.2023
Accepted: 19.07.2023

AN ATTEMPT TO VISUALISING TWITTER DATA FOR ANALYSIS

İLHAMİ M. ORAK [1*], MUHAMMED SELMAN DÖNMEZ[2]

[1,2]*Computer Engineering Department, Karabük Üniversitesi, Karabük, 78050, Türkiye*

Abstract. Social media is attracting so many people from different demographics. Facebook, YouTube, TikTok, WhatsApp, Instagram, and LinkedIn are some of those. Each one has some similarities as well as some differences. Twitter is one of the most popular social media platforms. It was publicly launched to be used for micro-blogging in 2006. Originally, it was limited to a maximum of 140 characters. It extended its usage capacity and became one of the most highly used social media platforms. Twitter users reached around 401 million by 2022. According to statistics, the number of daily tweets shared on Twitter is 500 million. The contents of shared tweets are so valuable for data analytic since they provide information about such as users' social lives, feelings, political opinions, and social environment. In this study, it is intended to introduce a visualization platform for Twitter data to provide better analysis. By using this tool, it will be possible to have a better understanding of the big data produced by Twitter posts. Analysis of all of the actions of Twitter such as hashtags and retweets is visualized. The visualization tool will also enable the reflection of the social network for targeted users and their followings and followers.

1. Introduction

Twitter has an important place among social media platforms in terms of the number of users and the data it produces [1]. Many different findings can be obtained using Twitter. As an example of these findings, people's opinions, environments, reactions to events, rising or falling trends, agendas, communities, and intuition about possible crises can be given. For these reasons, the use of Twitter data is considered important. When similar studies on Twitter data are examined, studies related to Twitter performance and impact analysis stand out. Examination of mention, hashtag, and retweet statistics, which are the basic actions of Twitter users, is among the most studied studies [2].

E-mail address: imorak@karabuk.edu.tr (*).

Key words and phrases. Twitter, Big data, Data visualization, Social media analysis.

It has been seen that there have been many studies on Twitter data, especially on ”Emotion” and ”Political” analysis since its establishment [3–6]. In his study, Kanbur aimed to reveal the social media appearances of political parties and to discover the sub-groups of political parties [7]. Terpstra tried to develop a crisis control mechanism using real-time data [8]. Pak and Patrick used Twitter for mind mining [9].

The number of studies that have combined data analysis and social network analysis needs to be improved in the literature. Therefore, this study aims to contribute to the existing literature significantly.

The main objective of this study is to analyze Twitter data and offer a comprehensive understanding of the social networks of users. A sophisticated software tool will be introduced to achieve this, capable of processing data using accurate methods and generating visual results. The outcomes of this research will enable predicting the characteristics and behaviours of the target audience or individuals.

In the study, from complex Twitter data meaningful information is drawn with powerful data visualization. One of the important abilities of the software developed is to query between layers. This helps users to find information about the followers and followers of the followers, as well as the followings and followers of the followings. It can search for a keyword among this data and create a relational path related to the target user and his environment. This results in to draw a meaningful conclusion for the end user with simple visualization techniques of these seemingly complex stages. Within the scope of the study, the data were collected using Beautiful Soup and JSON modules via the Application Programming Interface (API) of the Python language, stored in the MySQL database, and then visualized in 3D.

2. Material and Method

To produce meaningful results from the data, the steps of data manipulation, cleaning the data, producing results by applying mathematical operations to the data, and visualizing the data were followed (Figure 1). These processes were carried out through the open-source libraries of the Python language.

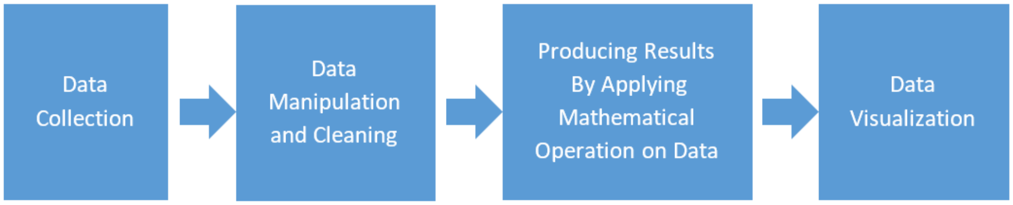

Data Collection
Data Manipulation and Cleaning
Producing Results By Applying Mathematical Operation on Data
Data Visualization

Figure 1. Steps followed in the study

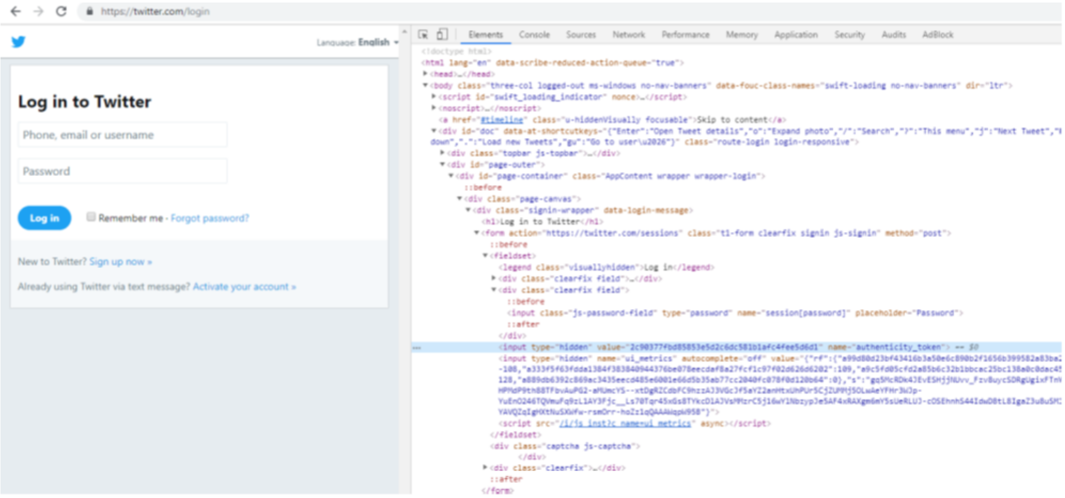
2.1. **Data Collection.** When the literature is reviewed, it is seen that Twitter's application programming interface (API) is frequently used during the data collection phase [10]. There are already tools on the internet that analyze Twitter. However, these tools use the API (Application Programming Interface) of the developer account offered by Twitter. The application programming interface (API) is based on the Representational State Transfer (REST) architecture. REST architecture refers to a collection of network design principles that define resources and ways to access data. The data targeted to be collected is user-related data such as all tweets, historical data, followings, and followers. In the study, a two-layer data query was carried out. Tier 1 includes only the user's followings and followers, while Tier 2 includes the followings of the user's followings and followers of his followers.

Data can be collected quickly by making queries over the API. However, the daily query limit and the fact that the obtained data does not arrive full-time (for example, the last 7 hours of data can be collected) prevent reaching an efficient result. For this reason, in the study, the process of parsing the Twitter page was determined as the method, and for this purpose, the Request, Beautiful Soup (BS4), and Selenium modules developed with the Python programming language were used.

Since it is necessary to login to Twitter with an account to view the following and followers of the person to be analyzed, the Request module was used. Then the cookie and session information obtained were use.

There are usernames, passwords, CSRF tokens, and some cookies in the JSON format variable created as a payload. In the headers' variable, there are header components created by browsers when requesting a basic web page. These were then forwarded to https://twitter.com/login as a post reques.

2.2. **Data Manipulation and Cleaning.** As a result of the processes, the data were collected through the selenium and request modules and then parsed with the BS4 library. The obtained data was written to the database via Laravel, a PHP framework. MySql, which works very well with PHP, is preferred as the database. However, on the PHP side, the ORM structure is preferred due to both security and flexibility of use. The Beautiful Soup library is used to extract HTML pages by selecting specific fields (Figure 2).



```
<!doctype html>
<html lang="en" data-scribe-reduced-action-queue="true">
▶<head>…</head>
▼<body class="three-col logged-out ms-windows no-nav-banners" data-fouc-class-names="swift-loading no-nav-banners" dir="ltr">
  ▶<script id="swift_loading_indicator" nonce>…</script>
  ▶<noscript>…</noscript>
   <a href="#timeline" class="u-hiddenVisually focusable">Skip to content</a>
  ▼<div id="doc" data-at-shortcutkeys="{"Enter":"Open Tweet details","o":"Expand photo","/":"Search","?":"This menu","j":"Next Tweet","
  down",".":"Load new Tweets","gu":"Go to user\u2026"}" class="route-login login-responsive">
    ▶<div class="topbar js-topbar">…</div>
    ▼<div id="page-outer">
      ▼<div id="page-container" class="AppContent wrapper wrapper-login">
          ::before
        ▼<div class="page-canvas">
          ▼<div class="signin-wrapper" data-login-message>
              <h1>Log in to Twitter</h1>
            ▼<form action="https://twitter.com/sessions" class="t1-form clearfix signin js-signin" method="post">
                ::before
              ▼<fieldset>
                  <legend class="visuallyhidden">Log in</legend>
                ▶<div class="clearfix field">…</div>
                ▼<div class="clearfix field">
                    ::before
                    <input class="js-password-field" type="password" name="session[password]" placeholder="Password">
                    ::after
                  </div>
...               <input type="hidden" value="2c90377fbd85853e5d2c6dc581b1afc4fee5d6d1" name="authenticity_token"> == $0
                  <input type="hidden" name="ui_metrics" autocomplete="off" value="{"rf":{"a99d80d23bf43416b3a50e6c890b2f1656b399582a83ba
                  -108,"a333f5f63fdda1384f383840944376be078eecdaf8a27fcf1c97f02d626d6202":109,"a9c5fd05cfd2a85b6c32b1bbcac25bc138a0c0dac4
                  128,"a889db6392c869ac3435eecd485e6001e66d5b35ab77cc2040fc078f0d120b64":0},"s":"gq5McRDk4JEvESHjjNUvv_FzvBuycSDRgUgixFTn
                  HPMdP9th88TFbvAuPG2-aMUmcYS--xtDgRZCdbFC9hzzAJ3VGcJf5aYZ2anHtxUhPUr5CjZUMMj5OLwAeYFHr3WJp-
                  YuEnO246TQVmuFq9zL1AY3Fjc__Ls70Tqr45xGs8TYkcD1AJVsMMzrC5j16wY1NbzypJe5AF4xRAXgm6mY5sUeRLUJ-cOSEhnhS44IdwD8tL8IgaZ3u8uSM
                  YAVQZqIgHXtNuSXWfw-rsmOrr-hoZz1qQAAAWqpW958"}">
                  <script src="/i/js_inst?c_name=ui_metrics" async></script>
                </fieldset>
                <div class="captcha js-captcha">
                    </div>
              ▶<div class="clearfix">…</div>
                ::after
              </form>
```

FIGURE 2. html codes of https://twitter.com/login page

2.3. **Result Generation and Data Visualization.** Bootstrap, which stands out with its responsive structure and was developed for mobile platforms, was used on the general template to benefit from customized CSS and JavaScript codes. The query page for the end user is intended to be as shown in Figure 2. According to different search types, the user will be able to create special keys and see the query on a single screen. A library named D3.js was used to visualize the data obtained from the query result. With the help of D3, drawings in SVG format are made and used on the page in HTML format. SVG format means drawings obtained with vectors. The biggest advantage of the SVG format is that there is no degradation in image quality. An example of SVG output is given in Figure 3.

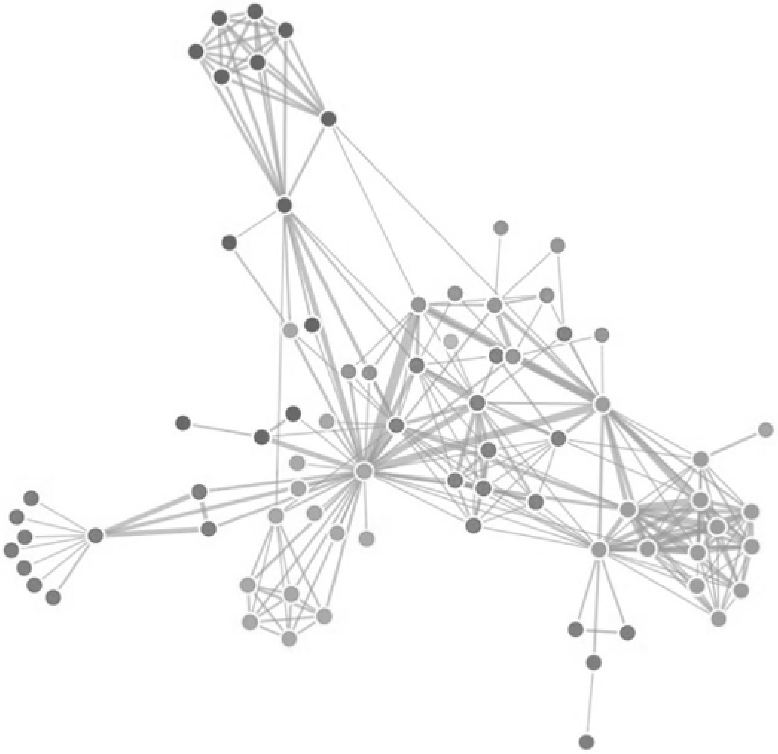


FIGURE 3. An output in SVG format

3. RESULTS

In the study, Twitter data were collected and analyzed. Thus, the social network analysis of the users was obtained. The resulting social networks were visualized using D3. It is worked on a Twitter account with around 14 million followers. First of all, as described in the methodology section, the data related to the person's Twitter data (followers and followings) were obtained via Python. Then, the data was processed in the database, and the interface in Figure 3 was created. The most frequently used tags, the most retweeted accounts, and the most mentioned accounts are displayed in the interface (Figure 4). Each part of the interface is displayed more clearly in Figure 5 to Figure 6.

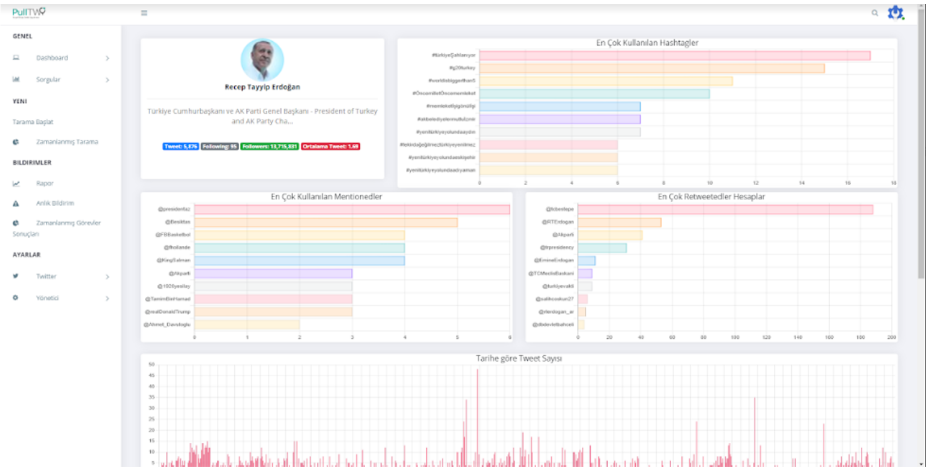


FIGURE 4. User Interface

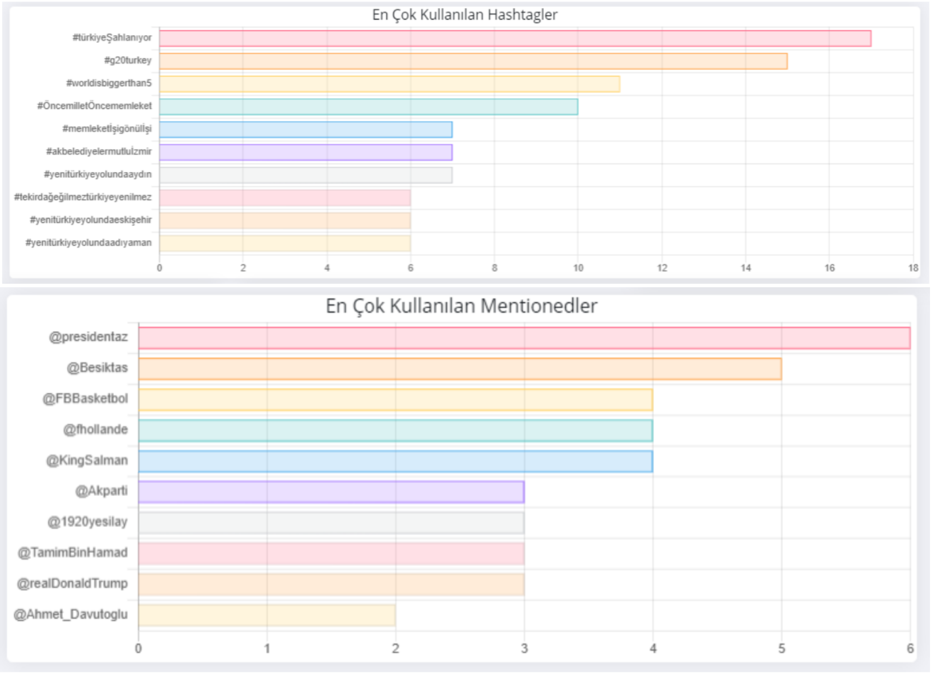


FIGURE 5. a. Most used hashtags in tweets, b. Most used mentions in tweets, respectively

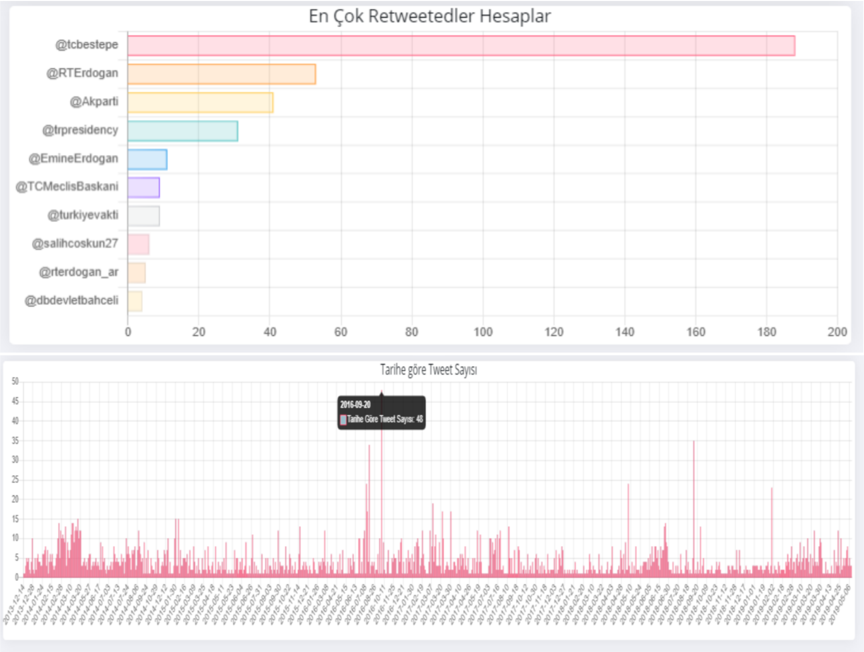


FIGURE 6. a. Most retweeted accounts in tweets, b. Number of Tweets by days, respectively

At the bottom of the interface, the historical change graph of the tweets is seen. In this way, the dates when the user uses Twitter the most can be determined. Then the followings and followers of the users (Layer 1), the followings of their followings, and the followers of their followers (Layer 2) are presented with the visualizations of social network analysis in Figure 7.

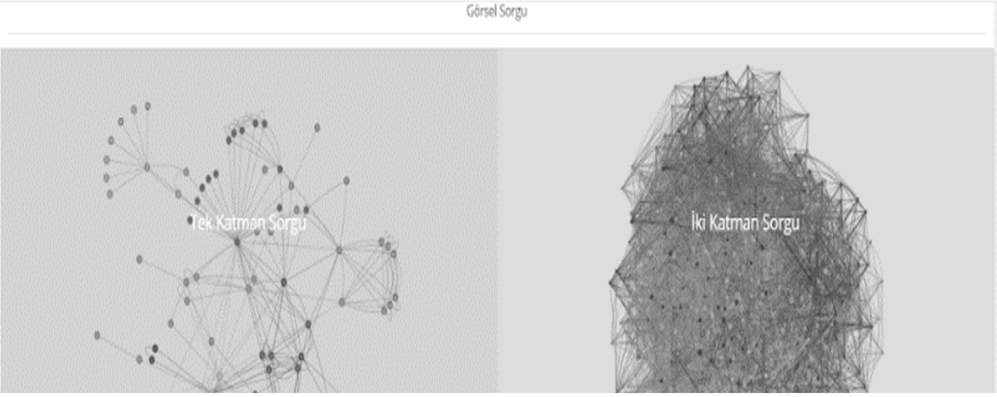


FIGURE 7. Social network analysis outputs

Through this software interpersonal connections can be seen. Each node represents a Twitter user, and interpersonal links are represented by links. In addition, links between different social networks and the sources of these links can be found. Figure 8 enables us to see which networks are associated with each other via which Tweet.

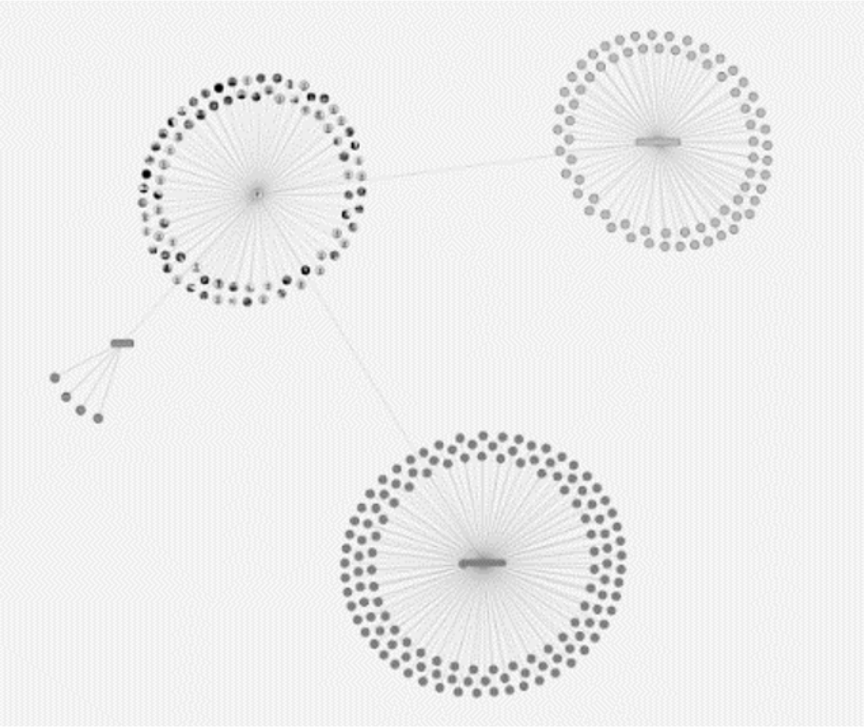


FIGURE 8. Network connections

As a result of the study, the behavior of the selected Twitter user can be discovered. In addition, it has been ensured that the users that have an indirect relationship with all other users to which the user under investigation is linked can be seen. This information obtained as the output of the study is considered useful in the context of social network analysis.

4. DISCUSSION

This study, carried out on Twitter data, is considered innovative because it has an interdisciplinary structure. The data collected from Twitter through Python and its various libraries was visualized with the D3 tool. Images related to the software developed during the study are included in the appendices. The visualization aims to examine the social network analysis of the users. Thus, it is aimed at making sense of Twitter data and being a decision support mechanism for the user. In the study, a two-layered query was made. In future studies, it is aimed at increasing the number of query layers. Thus, it is thought that a more detailed analysis will be achieved.

REFERENCES

[1] BusinessofApp. [Online]. Available: https://www.businessofapps.com/data/twitter-statistics/. [Accessed: 13 June 2023].

[2] V. Wisdom, An Introduction to Twitter Data Analysis in Python, 21 September 2016. [Online]. Available: [Accessed: 22 06 2023] DOI: 10.13140/RG.2.2.12803.30243.

[3] G. Göktürk, Analysis of Twitter to Identify Trends and Influentials with a Case Study on Turkish Twitter Users, 2014.

[4] M. H. T. G. E. Karadağ, Spor Taraftarlarının Twitter Kullanım Alışkanlıklarının İncelenmesi: Fırat Üniversitesi Örneği, Spor Eğitim Dergisi, Vol. 3, no. 1, pp. 44-53, 2019.

[5] P. Brassier, From Korea to the World: Women's Role as Peer-leaders in K-pop Transnational Online Brand Communities, Asia Pacific Business Review, 2023.

[6] D. Ediger, K. Jiang, J. Riedy, D. A. Bader, C. Corley, R. Farber ve W. N. Reynolds, Massive Social Network Analysis:Mining Twitter for Social Good, 39th International Conference on Parallel Processing, 2010.

[7] Y. Kanbur, The Use of Twitter as a Research Medium: An Example of an Application on Political Outlook, 2019.

[8] T. Terpstra, Towards a realtime Twitter analysis during crises for operational crisis management, Proceedings of the 9th International ISCRAM Conference, Vancouver, Canada, April 2012.

[9] P. P. Alexander Pak, Twitter as a Corpus for Sentiment Analysis and Opinion Mining, Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 2010.

[10] A. Kosorukoff, Social Network Analysis Theory and Applications, Passmore D. L., 2011.