Expert Systems With Applications 171 (2021) 114599

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Ranking influencers of social networks by semantic kernels and sentiment information

Berna Altınel Girgin

Computer Engineering Department, Marmara University, İstanbul, Turkey

ARTICLE INFO

Keywords:
Social network analysis
Opinion leader detection
Flow of influence
Sentiment polarity score
PageRank algorithm
Semantic kernels

ABSTRACT

Inspired by the importance of social media, a Social Network Opinion Leaders (SNOL) system has been proposed in this paper. The purpose of this system is to identify topic-based opinion leaders of social media. In order to accomplish this goal, several steps have been taken, such as data collection, data processing, data analysis, data classification, ranking of topic-based opinion leaders, and evaluation. The SNOL system has two main parts. In the first part, collected tweets are classified by semantic kernels for topic-based analysis. In the second part, leadership scores are given to each user in the network according to topic modeling and user modeling results. Leadership scores are then calculated with the formula generated and opinion leaders are determined for each category. Experiments are performed on data gathered from Twitter including 17,234,924 tweets from 38,727 users. The evaluation of opinion leader detection is a difficult job since there is no standard method for identifying opinion leaders. Therefore, the evaluation of the results of this study has been done using two different methods, retweet count and spread score, to prove that the suggested methodology outperforms the PageRank algorithm. The results have also been evaluated considering the user-topic sentiment correlation of the retrieved lists. Furthermore, SNOL has been compared against some opinion leader detection methods previously presented in the literature. The experimental results show that SNOL generates remarkably higher performance than the PageRank algorithm and other existing algorithms in the literature for nearly all topics and all selected top N opinion leaders.

1. Introduction

The rapid development of technology in today's world has started to diversify communication tools. The concept of social media, which people can use for communication as well as following local and world news, has settled into our lives. One of the widely used social networks, Twitter, is defined as a microblogging medium that allows individuals to instantly share content such as what is happening at the moment and their feelings, experiences, and thoughts with their followers by a short text called a "tweet" using a maximum of 280 characters. There are also many journalists, artists, politicians, and other famous people who can be reached directly on Twitter.

Social media is like a revolution since it has changed many aspects of people's lifestyles by bringing about new trends in communication, shopping, working, and other areas of life. On social media microblogging sites, increasing numbers of users meet daily and a wide range of ideas are quickly emerging, spreading, and creating interactive environments. In this context, social media can be seen as one of the most important sources of information affecting public opinion.

As the usage of social networks has become so widespread, the way people communicate has also changed. Interactions such as online followers/friends and the writing of comments have created wide user networks on social media. Within this network structure, some people prefer to produce their own content, sharing their experiences and thoughts with other users, and, thus, have more connections than others. These people, also known as opinion leaders, can address thousands of people with the content they share, and, more importantly, they can direct society's emotions and thoughts. This influence has attracted the attention of academics, and the identification of opinion leaders, or influencers, who are important in influencing and directing societies, has been the subject of many research projects.

The scientific and technological advancements described above will also have a variety of socioeconomic effects. For example, users who want to get information through social networks generally do not know who to follow in order to get the right information. Misinformation among the public can lead to various serious problems in the economic and social realms. In various commercial and marketing practices, the ideas and comments of opinion leaders on certain products seem to have

E-mail address: berna.altinel@marmara.edu.tr.

https://doi.org/10.1016/j.eswa.2021.114599
Received 30 September 2020; Received in revised form 22 December 2020; Accepted 9 January 2021
Available online 14 January 2021

an effect on the rest of society. It is necessary to carefully analyze the news coming from misleading/spam sources that should not be followed (including deceptive sources allied with foreign countries, etc.) before making critical decisions about political and military scenarios. These examples are just some of the situations that may be impacted by finding the correct opinion leaders in social networks.

The task of identifying influencers in social networks also has some societal/ethical considerations. These can be categorized as economic, ethical, and health and safety considerations:

1.1. Economic considerations

Consumers prefer to get advice from their close friends or experts about the products or services they will purchase. Today, these recommendations are provided by a group of people called influencers. Brands collaborate with influencers to sell and advertise their products or services. In this respect, it is important to find influencers who are relevant to the brand's sector, i.e. to classify influencers by topic.

1.2. Ethical considerations

Today, the widespread use of the Internet also causes information pollution. Opinion leaders are individuals who are followed by people who need information or advice. The attitudes and behaviors of opinion leaders are easily adopted by their followers. Successful identification of these people who are critical for their communities is extremely important.

1.3. Health and safety considerations

Extensive use of the Internet, as stated above, causes information pollution. Users can gain awareness of events affecting the world if they have access to the correct information. In some political and military scenarios, such as wars, news from misleading sources that should not be followed must be carefully analyzed before making critical decisions. Analyzing influencers and tweets is also crucial for many subjects that affect human health, such as increasing rates of non-immunized children or the so-called blue-whale game.

Various methods have been suggested in the literature for detection of opinion leaders and the flow of influence. These methods can be divided into 5 categories: 1) In diffusion-based approaches (e.g. Van Eck et al., 2011), attempts are made to detect opinion leaders by simulating how information spreads in the user network. 2) In graph-based approaches, networks are created based on user relationships and opinion leaders are detected using these graphs. In studies (Cui & Pi, 2017; Gökçe et al., 2014; Meltzer et al., 2010) in which the network was studied as a graph, opinion leaders were identified based on centrality measurements. 3) In statistical and stochastic approaches, various statistical measurements and features are used to discover dependencies within networks. In one study, for example (Alp & Öğüdücü, 2019), the matrix factorization method was used by calculating user features. 4) In PageRank-based approaches, efforts are made to improve the PageRank algorithm or use it as the baseline for other proposed algorithms. PageRank was used as a baseline in one study (Luo et al., 2018) in which the weighted InfluenceRank was proposed and in another (Song et al., 2007) where InfluenceRank was proposed. In a study in which the Personalized PageRank algorithm was evaluated (Alp & Öğüdücü, 2018), user properties were calculated and given to PageRank as parameters. 5) Finally, machine learning approaches are used in tasks such as classifying texts on microblogging sites. In these studies (e.g. Aleahmad et al., 2016; Hu et al., 2012), as opinion leaders were detected, their posts were used in classification algorithms like support vector machine (SVM).

In this paper, a Social Network Opinion Leaders (SNOL) system is proposed. The purpose of this system is to identify topic-based opinion leaders of social media. In order to accomplish this goal, research is undertaken with several subparts, including data collection, data processing, data analysis, data classification, ranking of topic-based opinion leaders, and evaluation. The SNOL system consists of two main parts. In the first part, a network was created by collecting 17,234,924 tweets from 38,727 Twitter users between December 1, 2019, and January 31, 2020. After the preprocessing step, latent Dirichlet allocation (LDA) including a proper pooling strategy was applied to decide the categories of the dataset using the Machine Learning Language Toolkit (MALLET) (McCallum, 2002). Word clusters generated in LDA were reviewed by human experts and some tweets were extracted from their categories. Finally, five different topic categories were obtained: economy, culture and arts, politics, sports, and technology. In order to classify tweets based on these categories, several semantic kernel classifiers were applied. In the second part, some user feature calculations including focus rate, activeness, authenticity, and follower/following ratio were also computed and added into the feature set in order to enrich it. Furthermore, some centrality measures were calculated and incorporated into the feature set. Evaluation of opinion leader detection is actually a difficult job since there is no standard method for identifying opinion leaders. In the literature, different approaches are implemented in many different studies since this evaluation can be subjective and specific to the problem domain. Therefore, the evaluation of the results in this study is done using two different methods, retweet count and spread score, to prove that the suggested methodology outperforms Google's PageRank algorithm (Page et al., 1999). In addition to these methods, sentiment analysis is incorporated into the evaluation process. In order to calculate the sentiment polarity scores of tweets, some dictionary-based approaches are used.

The main contributions of this work are as follows:

- Several supervised and semi-supervised semantic classifiers are used in order to classify tweets under predefined labels. With the help of these semantic classifiers, it is possible to see the effect of the usage of class-based semantic values of words in discriminating classes in the text classification field. The details of these classifiers are given in Section 3. To the best of our knowledge, this is the first effort to use these supervised and semi-supervised semantic classifiers in order to classify tweets in such a social network analysis environment.
- This SNOL system is a very rich package. For example, LDA is used with different pooling strategies. Furthermore, degree centrality, betweenness centrality, closeness centrality, and several user features such as focus rate, activeness, authenticity, and follower/following ratio are also included in the feature set. In addition, sentiment information has been used in the evaluation process. To the best of our knowledge, this is the first study to include all of these facilities into one single system that aims to identify topic-based opinion leaders.
- The evaluation of the results in this study is done using two different methods, retweet count and spread score, to prove that the suggested methodology outperforms Google's PageRank algorithm (Page et al., 1999). In addition to these methods, sentiment analysis is also incorporated into the evaluation process. In order to calculate the sentiment polarity scores of tweets, some dictionary-based approaches are used. Again, to the best of our knowledge, this is the first effort to include all of these evaluation techniques in the same system that aims to identify topic-based opinion leaders.
- A scalable and flexible big data infrastructure including distributed cloud deployment is constructed for SNOL, which lets the system adopt to environments with different sizes of data and run on these platforms easily.
- The performance improvement is one of the important contributions of the SNOL system. According to the experimental results reported in this paper, the SNOL system generates notable performance compared to the PageRank algorithm for nearly all topics and all selected top N opinion leaders.

The remainder of the paper is organized as follows: Section 2 includes the related work and background information. The proposed methodology, including the data collection and preprocessing steps, topic modeling step, user modeling step, calculation of sentiment polarity scores of tweets, and evaluation, is detailed in Section 3. The experimental setup and the corresponding experimental results are presented in Section 4 together with some discussion points. Finally, Section 5 offers concluding remarks and future directions.

2. Identifying opinion leaders in social networks

Various methods have been suggested in the literature for detection of opinion leaders and the flow of influence. These methods can be grouped into five categories: 1) diffusion-based approaches, 2) graph-based approaches, 3) statistical and stochastic approaches, 4) PageRank-based approaches, and 5) machine learning approaches.

2.1. Diffusion process-based approaches

In diffusion-based approaches, an attempt is made to understand the structure of the network and analyze how information is spread by simulating social networks. The influence maximization problem was first mentioned by Kempe et al. (2003). Their work rapidly attracted the attention of researchers studying opinion leader detection. For instance, Zhao et al. (2016) proposed a method called IM-LPA by combining the influence maximization algorithm with label propagation to rank opinion leaders. In another study (Van Eck et al., 2011), the role of opinion leaders in the diffusion of a product was researched. An empirical survey was conducted and, as a result, 3 characteristics of opinion leaders were revealed. A 3-step agent-based simulation model was then constructed, in which hypotheses would be tested. These steps were mass media, word of mouth, and adoption. As a consequence of the study, it was observed that the adoption speed of the product increased according to the presence of opinion leaders in the network and the appropriate comments of opinion leaders about the product. Numerous centrality measures such as in-degree links, out-degree links, and betweenness centrality were calculated and applied in the methodology for detecting opinion leaders in a recent study (Rehman et al., 2020). The community evolution Louvain method was employed to analyze the experimental results. Higgs boson data from Twitter were used as a dataset. This dataset contains 256,491 nodes and 328,132 edges in the retweet network, 116,408 nodes and 150,818 edges in the mention network, and 38,918 nodes and 32,523 edges in the reply network. According to the analysis of Rehman et al. (2020), for a more detailed and adequate examination of a network, a dataset encompassing a longer period is suggested. In addition, incorporating the contents of tweets might increase the classification performance of the system.

2.2. Graph-based approaches

In graph-based approaches, network graphs are created based on the relationships between users, and efforts are made to identify opinion leaders utilizing user features and centrality measurements inferred from the graphs. In one such study (Gökçe et al., 2014), political opinion leaders were attempted to be detected by using the degree centrality, eigenvector centrality, and betweenness centrality of user nodes in the social network structure. The sample dataset created with 6000 users was increased to 15 million using 2 degrees of separation. The users in the dataset were filtered considering attributes such as language and location and 10 million users then remained. The top 100 users were listed using 3 centrality measurements and these measurements were positively correlated. In another study (Cui & Pi, 2017), a method was developed that offers a probabilistic generate-graph model using user features and outbreak nodes instead of static features such as number of good friends. Users were categorized using user attributes and it was found that efficient nodes had higher values by calculating outbreak index values (Cui & Pi, 2017). In the last layer of the framework, which consisted of 3 layers (i.e. data layer, topic discovery layer, and opinion leader discovery layer), relation matrices were created with the help of the UCINET tool. The method applied in that study performed better when compared to SVM and Bayesian algorithm results. In another work (Meltzer et al., 2010), the authors explored how to build effective clinical quality improvement teams using social network analysis (SNA). People were often placed in the same cluster because they interacted with those with whom they shared the same sociodemographic, professional, or other features. Net degree, betweenness measure, and network density were used to build an empirical methodology. One of the results of this paper was that the teams with the highest net degree included people with different backgrounds. Due to the insufficient data, however, the ways in which the effectiveness of the team could be impacted in terms of its structure and quality improvement context could not be analyzed.

In a recent study (Jain & Katarya, 2019), the modified firefly algorithm was suggested to find local and global opinion leaders in a social network. The Louvain method was used to determine communities within the social network. Two different datasets were applied in the experimental environment: a synthesized dataset including 20 nodes and 70 edges and “small slashdot” including 13,182 nodes and 34,621 edges. According to the experimental results reported by Jain and Katarya (2019), the proposed methodology was superior to standard SNA methods. In another very recent study (Li et al., 2019), an opinion community identification approach and opinion leader identification approach were presented for ranking opinion leaders in social networks. These approaches were improved by the user influence, sentiment analysis, content similarity, time similarity, and topology structure of users. Using 11,713 posts and 8976 topics from a world forum, the authors constructed their dataset. The single-pass (SP) algorithm, online time-based opinion leader discovery (OTOLD) algorithm, experience-based opinion leader discovery (EOLD) algorithm, K-means algorithms, and PageRank algorithm were used as benchmark algorithms in order to see the effect of the presented approach. According to the experimental results given by Li et al. (2019), their novel approach can effectively rank opinion leaders.

2.3. Statistical and stochastic approaches

In statistical and stochastic approaches, various calculations and features are used to discover dependencies within networks and use them in opinion leader detection. In one such study (Li & Du, 2011), a framework called BARR was proposed, which determines opinion leaders and gives marketers a chance to determine their strategies according to the posts of bloggers. Blogs were searched using user-defined keywords and web pages were analyzed. Domain ontology was extracted by calculating entropy values from the collected ontologies. Relations between bloggers were found using domain name ontology, centrality, and prestige. The so-called Technique for Order Preference by Similarity to Ideal Solution, which summarizes the Euclidean distance between measurements and ideal solutions, was used for selection of hot blogs by topic. Whether or not a user is an opinion leader was decided based on the calculated quantity and quality values. Another study (Cho et al., 2012) investigated which opinion leader is best for a selected market in terms of diffusion speed and the maximum cumulative number of adopters. Based on social network theory, this work examined how opinion leaders affect networks and product diffusion. A simulation with 3 different scenarios was repeated 100 times with a network of 10,000 entities. Alp and Öğüdücü (2019) suggested the influence factorization method. This method was developed using topic-based user feature matrices and it utilizes the matrix factorization method. Matrix factorization provided the advantage of detecting potential influencers as well as current ones. For topic modeling, LDA was used by applying the pooling method, in which the tweets shared each day by each user were combined. While the alternating least squares approach was used for factorization, PageRank, TwitterRank, and Personalized PageRank were chosen for the baseline. Amor et al. (2016) suggested an approach for community identification and associated role classification for retweet networks and followers separately. The logic behind it is constructed on the role difference of users in the diffusion of information that they outline by in-flow and out-flow communication designs. Based on the work of Zhao et al. (2015), an opinion leader might be identified in a social network with the expected number of retweets. The theoretical SEISMIC framework (Zhao et al., 2015) uses a dataset collected from Twitter. Amor et al. (2016) reported that, using retweets, it is possible to find the influencers in social networks.

Saia et al. (2014) proposed a novel dynamic coherence-based technique that analyzes the information in user profiles based on their coherence in a recommendation system. They used the Webscope movie dataset[1]. They extracted similarities among words from WordNet. With their methodology, which is called the dynamic coherence-based modeling approach, they aimed to remove the corresponding incoherent items in a user profile. According to their experimental results, the methodology improved the F1 measure of the state-of-the-art recommendation system.

2.4. PageRank-based approaches

In PageRank-based approaches, researchers try to improve their results by making changes to the PageRank algorithm or using it as the baseline for algorithms that they have developed themselves. In a recent work (Luo et al., 2018), for example, an improved weighted LeaderRank algorithm was presented. User weights were calculated using not only replies but also posts, reading, likes, etc. The link between users is based on whether one user is replying to another. After calculating the influence score of each user, results were compared with the influence values of the PageRank and LeaderRank algorithms, and more accurate results were obtained with the weighted LeaderRank. In another study (Song et al., 2007), a novel algorithm named InfluenceRank, which ranks the blogs in a blog network according to their importance and currentness, was studied. The topic space was created by accepting each entry in the blogs as a document and using LDA. The feature vector of each entry was then created in the topic space and dissimilarities were calculated using cosine similarity. The proposed algorithm performed better than PageRank, random sampling, time-based ranking, and information novelty-based ranking algorithms used as the baseline. In a more recent work (Alp & Öğüdücü, 2018), the Personalized PageRank (PPR) algorithm based on Google's PageRank algorithm was proposed to analyze opinion leaders in a topic-based manner. The algorithm combines network topology and user features such as focus speed and originality and uses these features as a parameter of the PageRank algorithm. In order to label users' tweets, word clusters were created with LDA and topic titles were determined. The PPR algorithm was run and the top 25 users were retrieved for each topic. After the users were obtained in the same way from the algorithms used as the baseline, the results were compared according to the spread score formula used for evaluation. The results showed that at least one PPR result was better than the baseline algorithms in every case.

The rank after clustering (RaCRank) algorithm was presented to discover opinion leaders in social networks in a recent study (Zhang et al., 2020). This method has two parts: a modified version of K-means is utilized with in-degree, betweenness, and center features, and then a 2-hop clustering coefficient is suggested. Users' leadership scores are calculated depending on user influence, center features, and user activeness. A dataset was built by collecting 49,613 users, with 59,957 edges among these users. The authors compared their results with AllUserRank, ClusterRank, and UI-LR. Although the RaCRank algorithm performs slightly worse than UI-LR, it outperforms AllClusterRank (Zhang et al., 2020).

2.5. Machine learning approaches

Machine learning approaches are used in tasks like classifying texts in social networks to try to detect opinion leaders. In one such work (Page et al., 1999), the authors focused on how news spreads on Twitter. Two different SVM classifiers were trained with the bag-of-words method for determining whether news shared on Twitter before reaching mass media was a rumor or not. While the first classifier determines the tweet's relevance to the topic, the second classifier decides whether the tweet is certain or uncertain. It was discovered that people prefer a small group of people called opinion leaders who share information with others on Twitter instead of learning from news sources. In another work (Aleahmad et al., 2016), OLFinder was proposed to find influential users by analyzing important topics in the domain. Popularity scores according to users' links on the network and competency scores based on topics were calculated. According to the results obtained, the proposed method outperformed the basic algorithms in the literature. LDA was used for topic extraction and was found to be better than TF-IDF.

In a recent work (Jain et al., 2020), a novel approach was proposed for community detection and the social network-based nature-inspired whale optimization algorithm. Global and local top N opinion leaders are detected using different standard benchmark optimization functions. The community-partitioning algorithm is used to discover communities in a social network. The experiments were performed using two different datasets; the first one was a synthesized dataset consisting of 100 nodes and 467 edges, and the second was the "wiki-vote" dataset consisting of 7115 nodes and 103,689 edges. As the number of users in the network increased, the performance of the algorithm increased.

A cluster-based opinion leader identification approach was offered in another recent work (Chen, 2019). This approach first initializes a social network by considering the post/reply relationships of Mobile01 forum posts. The authors discovered the important communities in the network with the parameter-free method that they implemented. K-means was used on the important communities to create clusters and each cluster was given a score. They selected final opinion leaders from each high-performing cluster. Experiments were performed by using forum discussions from Mobile01 related to four different car brands. According to the experimental results, the presented method was superior to the baseline algorithm for each dataset (Chen, 2019).

Another very recent work (Carta et al., 2020) proposed a novel approach based on machine learning techniques in order to forecast future popularity and especially the engagement factor of Instagram posts. This engagement factor is defined as the ratio of expected likes to number of followers of the account (Carta et al., 2020). The authors designed the prediction of the future popularity problem as a classification task, which is intended to be solved by regression techniques. They collected a comprehensive dataset from Instagram including 106,404 rows for a total of 2545 different users. They started with a preliminary list of Instagram account identifiers, and then they iteratively extended the list of accounts until reaching a sufficiently large number of accounts. They then applied appropriate preprocessing and feature engineering steps to the data by using JSON and Python. They used XGBoost as a supervised classification algorithm. The authors stated that the XGBoost algorithm offers efficient implementation and improved classification performance compared to gradient boosting. They built a scalable and flexible big data infrastructure. According to their experimental results, the implementation based on gradient boosting has good effectiveness with a balanced accuracy of 64.72% in predicting future posts on Instagram.

In another recent work (Phan et al., 2020), a new system was presented in order to detect the sentiment polarity of tweets. This system includes fuzzy sentiment using the feature ensemble model. In comparison to most of the existing studies in the literature, this study adds syntactic information and positioning of words into a sentiment analysis system. These authors used a dataset from their previous study (Phan et al., 2019) as their first dataset. This dataset was constructed by

[1] http://webscope.sandbox.yahoo.com

collecting all English tweets from Twitter for all hashtags related to the fuzzy semantic words and negation words such as #quite, #too, #not, #no, etc. between May 1 and November 30, 2018, for all topics. This dataset has 7368 tweets. They also built a second dataset by adding 14,865 English tweets of airline companies obtained from the Kaggle website.[2] The authors created 5 different types of vector models and added them into their tweet embedding methodology: lexical vector, word-type vector, polarity sentiment vector, semantic vector, and position vector. Then, by using this novel tweet embedding methodology, they analyzed the sentiment of tweets with the help of a convolutional neural network (CNN). They designed all layers of the CNN, including the convolutional layer, max-pooling layer, and their softmax layer, according to their novel tweet embedding model. They reported their experimental results in their study (Phan et al., 2020), showing that method significantly improved the performance in the sentiment analysis of tweets.

3. Methodology

In this paper, a method is introduced to identify topic-based opinion leaders based on user features and the network structure on Twitter, i.e. the follower and friend relations among users. Fig. 1 shows the basic flow of the suggested approach. SNOL starts with data collection from Twitter. By using the Twitter API, user information and tweets are collected for a specific period. This information is used to construct a global network. A very detailed preprocessing approach is then applied, including cleaning, removing, and clustering. After that, the topic modeling task is accomplished with the classification of tweets by several supervised and semi-supervised semantic classifiers. In the user modeling part, many characteristic features of the users are extracted. These features are used in order to get some important information like degree centrality, betweenness centrality, closeness centrality, focus rate, activeness, authenticity, and follower/following ratio. In the candidate opinion leaders step, the leadership scores of the users in the social network are calculated. Finally, these results are compared with the results of the baseline algorithms in the literature, such as Google's PageRank algorithm. Additionally, sentiment information is added in order to improve the validation step.

The main steps will be explained in the following respective subsections: data collection and preprocessing, topic modeling, user modeling, and getting the sentiment polarity information of tweets.

3.1. Data collection and preprocessing

In order to detect influencers on Twitter, two different Turkish datasets were collected. The first dataset consists of only tweets while the second dataset consists of both tweets and the users' follower/following relationships. Fig. 2 represents the main steps of data collection and preprocessing. At the beginning, 350 specific Twitter users were manually selected according to their various tweets on different topics and the common sense of the social network. The first dataset consists of 1,842,499 tweets. The Twitter API was used for the tweet collection operations. For the second dataset, 38,727 Twitter users and 17,234,924 tweets were collected between December 1, 2019, and January 31, 2020. These datasets were stored in a MongoDB dataset with indexing.

A cleaning process was applied to the tweets, removing mentions, hashtags, URLs, emojis, and punctuation from them. Furthermore, stop word filtering and stemming were applied to the tweets with Zemberek (Akin & Dundar, 2007). After cleaning the data, LDA was applied to decide the categories for the first dataset using MALLET (McCallum, 2002). As tweets are short text documents, one of the pooling methods suggested by Alp and Ögüdücü (2015) was applied to get more coherent clusters. Word clusters generated in LDA were reviewed by human experts and some tweets were extracted from their categories. Finally, topic labels were added to the pooled tweets to build the final dataset. Five different topic categories were thus obtained: economy, culture and arts, politics, sports, and technology. The number of tweets for each category are as follows: 300,335 tweets for the economy category, 349,859 tweets for culture and arts, 414,404 tweets for politics, 459,083 tweets for sports, and 318,818 tweets for technology.

3.2. Topic modeling

In order to classify tweets based on the predefined categories of economy, culture and arts, politics, sports, and technology, several semantic kernel classifiers are applied. Three of these semantic kernels are the semantic meaning classifier (SMC) (Altınel, Diri, & Ganiz, 2015), abstract feature classifier (AFC) (Altınel, Ganiz, & Diri, 2015), and relevance value classifier (RVC) (Altınel et al., 2019). These semantic kernel classifiers attempt to add semantic values of terms into the classification process. Additionally, sprinkled and adaptive sprinkled versions of these semantic classifiers were also developed and are applied to the experimental environment. Finally, two composite semantic kernels formed from these semantic kernels are also implemented and run in the experimental environment.

3.2.1. Sprinkled (S) process

In this method, the class label of the tweet is added as a separate feature to the word-document matrix (Chakraborti et al., 2006), as shown in Fig. 3.

For example, according to Fig. 3(b), the first three documents belong to the c_1 class; they take a value of 1 for the c_1 feature and 0 for the c_2 feature. The remaining three documents take a value of 0 for the c_1 feature and 1 for the c_2 feature because they belong to the c_2 class.

3.2.2. Adaptive sprinkled (AS) process

In this method, the amount sprinkled (i.e. the number of new columns to be added) is added as a parameter, which is directly proportional to the separability of two classes (Chakraborti et al., 2007). The number of sprinkled terms specific to the classes is determined by probabilistic calculations based on the confusion matrix in Fig. 3.

3.3. User modeling (Constructing the feature set of each User)

In the topic modeling step, appropriate category labels are assigned to unlabeled tweets. After completing this step, several subnetworks are constructed based on the topic distribution of the tweets of each user. If the tweet ratio of a user for any topic exceeds the predefined limit, that user is added to the subnetwork of that topic. The distribution of users obtained after removing the users who do not obey the threshold is shown in Table 1.

Using these subnetworks, the degree centrality (DC), betweenness centrality (BC), and closeness centrality (CC) are calculated. DC shows the number of nodes with which a node in a network has a direct relationship. BC represents the importance of a node that connects different nodes thanks to their locations in a network. People with high BC values are critical people because they control the information passing among others. Meanwhile, the direct and indirect connections of users with high CC values in a network allow them to reach other users more quickly. The more central a node is, the closer it will be to all other nodes.

Some user feature calculations including focus rate, activeness, authenticity, and follower/following ratio are also computed and added into the feature set in order to enrich it, as follows:

3.3.1. Focus rate

This represents the topic distribution of the user's tweets (Alp &

[2] https://www.kaggle.com/crowd_ower/twitter-airline-sentiment

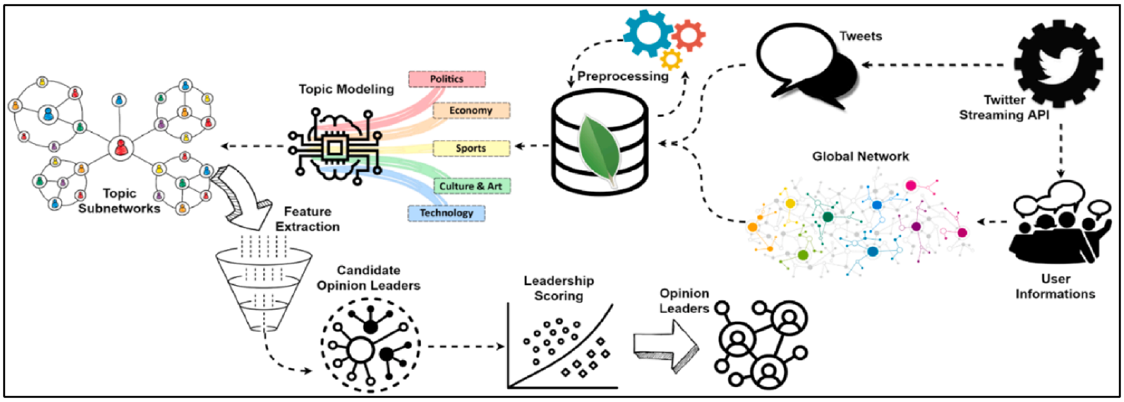

Fig. 1. Flowchart of SNOL.

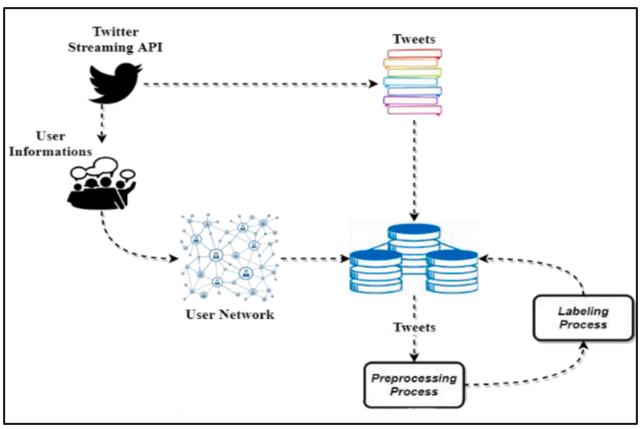

Fig. 2. Dataset construction steps.

Öğüdücü, 2019). For each topic, the user's focus rate (fr^t_u) is computed by dividing the number of tweets posted for the topic (p^t_u) by the total number of tweets (p_u). If it exceeds the predefined threshold for any topic, it is stated that the user is focused on that topic.

$$fr^t_u = \frac{|p^t_u|}{|p_u|} \tag{1}$$

3.3.2. Activeness

This shows how often a user tweets on a topic (Alp & Öğüdücü, 2019). For each topic, the user's activeness (ac^t_u) is computed by dividing the number of days on which that topic is tweeted about (d^t_u) by the total number of days (d).

$$ac^t_u = \frac{|d^t_u|}{|d|} \tag{2}$$

3.3.3. Authenticity

This represents the originality of tweets about a topic (Alp & Öğüdücü, 2019). The user's retweets (rt^t_u) on a topic are subtracted from all tweets related to that topic (p^t_u), and then the result is divided by all tweets about that topic (p^t_u).

$$au^t_u = \frac{|p^t_u| - |rt^t_u|}{|p^t_u|} \tag{3}$$

3.3.4. Follower/following ratio

This provides a comparison of the number of users following a user with the number of users followed by that user (Anger & Kittl, 2011). For each topic, the user's follower/following ratio (ff^t_u) is computed by dividing the user's in-degree for the topic (id^t_u) by the out-degree (od^t_u).

$$ff^t_u = \frac{|id^t_u|}{|od^t_u|} \tag{4}$$

Table 1
Number of users in subnetworks.

Category	# Users
Economy	692
Culture and Arts	3529
Politics	25,193
Sports	3067
Technology	452

$$\begin{array}{c|cccccc} & t_1 & t_2 & t_3 & t_4 & t_5 & t_6 \\ \hline d_1 & 1 & 1 & 1 & 0 & 0 & 0 \\ d_2 & 1 & 0 & 1 & 0 & 0 & 0 \\ d_3 & 1 & 1 & 1 & 0 & 0 & 0 \\ d_4 & 0 & 0 & 0 & 1 & 1 & 1 \\ d_5 & 0 & 0 & 0 & 1 & 0 & 1 \\ d_6 & 0 & 0 & 0 & 1 & 1 & 1 \end{array}$$

(a)

$$\begin{array}{c|cccccccc} & t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & c_1 & c_2 \\ \hline d_1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ d_2 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ d_3 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ d_4 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ d_5 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ d_6 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{array}$$

(b)

Fig. 3. (a) Standard word-document matrix, (b) word-document matrix with class labels (Altınel et al., 2019).

3.4. Obtaining the sentiment polarity information of tweets and users

3.4.1. Obtaining sentiment polarity information of tweets

In order to calculate the sentiment polarity scores of tweets, a dictionary-based method is used. The SentiTurkNet (Dehkharghani et al., 2016) and SenticNet (Cambria et al., 2010) polarity lexicons are used in SNOL. The sentiment lexicon for emojis is created with "1″, "0", and "−1" scores for selected emojis. The score "1" shows positive sentiment polarity, "−1" shows negative sentiment polarity, and "0" shows neutral sentiment polarity. The average polarity score of a tweet is computed with the following equation (Dehkharghani et al., 2016):

$$P(t) = \sum_{wordsintweet} \frac{positivity(w) - negativity(w)}{numberofwordsinthesentimentlexicon} \tag{5}$$

Here, *positivity(w)* represents the positive sentiment polarity of the word (w) in the tweet (t) and *negativity(w)* shows the negative sentiment polarity of the word (w) in the tweet (t); *number of words in the sentiment lexicon* signifies the number of words whose polarity scores are found in lexicons; and *P(tweet)* represents the average sentiment polarity score of the tweet.

For calculating the sentiment polarity of a tweet (Dehkharghani et al., 2016), positive and negative sentiment scores of words are calculated by searching for both the word itself and its stem in our lexicons. The average polarity score of the tweet is then calculated with Eq. (5). In order to handle linguistic issues in Turkish (Dehkharghani et al., 2015) like negation, negativity, and absence, suffixes are detected, and if a word contains a relative suffix, the polarity of that word is reversed. Lastly, the coverage rate (i.e. the number of words whose polarity scores are found in the lexicons/number of total words) is calculated. For the dataset containing nearly 470,000 tweets, this rate was 81%, and for the dataset with 17 million tweets, this rate was 66%.

When combining scores, if a word is found in both the SenticNet and SentiTurkNet dictionaries, the value in SentiTurkNet is taken into consideration since SenticNet is a dictionary translated from English and the values in SentiTurkNet are more reliable. Afterwards, the emoji numbers and scores of the tweet are calculated using the emoji dictionary. This technique is then applied and run for small portions of the dataset, and the results are examined manually. The average scores returned by each dictionary and the combined version of these three dictionaries are observed. It is seen that using the three dictionaries together yields the most logical results. In addition, the problem of negativity was also analyzed to increase accuracy, and it was decided that if words containing negative or absence suffixes are found in the dictionary, the polarity score would be used by multiplying it by minus.

3.4.2. Obtaining sentiment polarity information of users

The average sentiment polarity score for each topic is calculated by selecting 100 users for each category among the opinion leaders determined using Eq. (6):

$$SPS_{topic\ t} = \sum_{user} (\alpha \times \text{Centrality Measure} + \beta \times \text{User Score} + \Omega \times \textit{Sentiment Polarity Score}) \tag{6}$$

Here, Centrality Measure denotes the centrality measure of the user, User Score denotes the total user score including all features (i.e. focus rate, activeness, authenticity, and follower/following ratio), *Sentiment Polarity Score* denotes the sentiment polarity score of the user, and SPS $_{topic}$ denotes the total sentiment polarity score of topic t. Furthermore, α, β, and Ω are coefficients whose total is equal to 1.

3.5. Detection of topic-based opinion leaders

First, a feature vector is created for each user by considering the subnetwork that a user is in. The K-means clustering algorithm is then used to identify user groups that are more likely to be opinion leaders since opinion leaders show similar behaviors. For instance, influencers generally have a high focus rate and authenticity value since they prefer to focus on a particular topic and express their own opinions. In order to decide which clusters to choose as candidates, fuzzy-based techniques are applied (Duan et al., 2014). A candidate opinion leader is expected to have high values in the feature vector. Therefore, normal cumulative distribution is used for each feature as a fuzzy membership function:

$$f(x) = \int_0^{x_{max}} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{7}$$

Here, μ indicates the mean of the feature and σ represents the variance, x_{max} shows the maximum value of the feature in the dataset, and x represents the consistent attribute. Later, a score is computed for each cluster by multiplying the function result of each attribute in the feature vector. The clusters are sorted according to their scores and potential clusters are selected as candidate influencers. Consequently, Eq. (8) is used to detect actual topic-based influencers:

$$OLscore^t_u = (w_1 \times dc) + (w_2 \times bc) + (w_3 \times cc) + (w_4 \times fr^t_u) + (w_5 \times ac^t_u) + (w_6 \times au^t_u) + (w_7 \times ff^t_u) \tag{8}$$

Users in possible clusters are labeled as influencers, others are labeled as normal users, and feature weights obtained by training the SVM are considered as coefficients in the formula. User lists are detected after each user's opinion leader score is computed. Finally, the top N users are chosen as the influencers. N is tuned to an optimal value after a series of experiments.

4. Experiments

4.1. Experimental setting and evaluation

Because of the imbalanced structure of the first dataset, the F1 measure is applied as the evaluation metric. In addition, the Student t-test is performed on each fold with a significance level of 0.05 to observe the statistical difference between the baseline algorithms and classifiers. The user modeling module is implemented utilizing the Apache Spark ecosystem. In order to store data, Neo4J is used, and in order to calculate centrality metrics, Neo4J's Cypher query language is used. Two different methods are used to evaluate the lists of opinion leaders. Both methods utilize the idea of information diffusion with retweets. In the first method, the total retweet count of a user is the number of times that the user retweets the tweets of other users on a given topic:

$$RTcount^t(u) = \sum_{\substack{tweet \in \\ (|p^t_u| - |rt^t_u|)}} \sum_{\substack{u' \in \\ subnet.}} rt^{tweet}_{u'} \tag{9}$$

Here, *tweet* shows a user's own tweets, while u' indicates other users in the topic subnetwork. $rt^{tweet}_{u'}$ represents whether the tweet was retweeted by user u' and it takes a value of 0 or 1.

Then, using the *RTcount* values, users are sorted in descending order and the top N of them are chosen. In the same manner, the top N users are selected according to their scores, which are calculated by the proposed method. Finally, the agreement between the two lists of users is calculated as the correctness accuracy of the suggested approach.

The second approach utilizes the spread score method. The spread score of a user for a given topic is the total retweet count normalized by retweet rate (Alp & Öğüdücü, 2019):

$$spread(t) = \sum_{u \in inf_t} \frac{|p^t_{u,rt}|}{|p^t_u|} \sum_{p \in p^t_{u,rt}} |retweets_p| \tag{10}$$

In this approach, the top N users are selected via the proposed method and their spread scores are summed to calculate the information diffusion. In order to observe the behavior of both the baseline and proposed methods with different numbers of N, the evaluation procedure is repeated with N as 20, 30, 40, 50, 60, 70, 80, 90, and 100.

In addition to these methods, sentiment analysis is also incorporated into the evaluation. The polarity scores of users and topics are calculated according to the related tweets' polarity scores. The polarity scores of the determined opinion leaders are expected to be close to the polarity score of the society on that topic. To assure this condition, polarity scores of users and society are translated into time series. The Pearson correlation coefficient between users' polarity scores and topic polarity scores is then calculated, and by taking the absolute value of this coefficient, the users with this score being less than 0.5 are eliminated from further analysis.

The server on which we run our experiments has 32.0 GB of RAM and Intel Core i7-8750H CPU 2.20 GHz. The Python 3.6 library named TWINT is used to collect the data. Incoming data are stored on a local server, being indexed on MongoDB. The Java 8.0 library called Zemberek is used for preprocessing of the collected data, and the Python 3.6 library named Py4J is used to connect with Java. Our classification algorithms are coded on Python 3.6. Furthermore, during matrix calculations in SVM, we decided to use Apache Spark because all the words in the dataset occurred frequently in the topics and the mathematical operations required for the algorithms are heavy in terms of workload. Thanks to Spark, these calculations are performed in a distributed way and the execution times of the algorithms are shortened. Since we are working with big data, we work with the LinearSVC classifier 'C' = 1, 'dual' = 'false', 'penalty' = 'l2' parameters of the scikit-learn library. For the skip-gram model of the Word2vec model, the Python 3.6 library named Gensim is used. The NetworkX Python library is used to extract the features of the social network.

The SentiTurkNet (Dehkharghani et al., 2016) and SenticNet (Cambria et al., 2010) polarity lexicons are used to detect the sentiment polarities of words in SNOL. Parallel computing libraries are used to run experiments on a 17-million-tweet dataset in order to increase the performance. The Swifter and Dask libraries for Python were considered for the experimental environment. Since the Swifter library works faster, this library is used. In addition, to prevent memory errors, small chunks are formed from several portions of the tweet dataset and computations are done. The user modeling module is implemented by utilizing the Apache Spark ecosystem. In order to store data, Neo4J is used, and in order to calculate centrality metrics, Neo4J's Cypher query language is used.

4.2. Baseline algorithms

SVM with the linear kernel function is chosen as a baseline algorithm to compare the results of the previously mentioned semantic kernel classifiers. The main objective of the SVM approach is to maximize the margin between two classes by finding the best separating hyperplane (Boser, 1992; Vapnik, 1995). Experiments for the baseline algorithm are performed with the same training portions as all supervised semantic kernel and semi-supervised semantic kernel classifiers. Different values including 10^{-2}, 10^{-1}, 1, 10^{1}, are 10^{2} are tried for the SVM's misclassification penalty parameter C and it is tuned to 1 for all experiments.

In order to compare the performance of the suggested algorithm, PageRank is selected as a baseline algorithm. The main logic behind the PageRank algorithm is to rank websites by computing their significance. It is actually a mathematical formula that calculates the value of a web page according to the amount and quality of links with other pages (Page et al., 1999). In the literature, the PageRank algorithm is also used as a baseline in studies (Alp & Öğüdücü, 2018, 2019), where the objective is to identify opinion leaders of specific social network structures. Candidate opinion leaders for each category are sorted based on their PageRank scores and then the top N users are chosen as opinion leaders.

4.3. Experimental results and discussion

4.3.1. Experimental results of the topic modeling step

By using the supervised and semi-supervised semantic classifiers described in Section 3.2, tweet classification experiments are performed using the first dataset. After that, according to the experimental results from the first dataset, the best performing classifier is determined and it is used in order to classify unlabeled instances in the second dataset, as shown in Table 2.

Experimental results of the supervised classifiers show that the default versions of the semantic approaches exceed their sprinkled and adaptive sprinkled versions. At very small training percentages (i.e. 1%, 5%), all supervised classifiers except Sprinkled-RVC outperform the baseline method, as seen in Table 2. We may conclude that, given a very small amount of data, semantic kernels are better at capturing the patterns represented in the dataset. Semantic classifiers achieve higher scores than the baseline method until a 70% training percentage. Across all the experiments, the RVC classifier outperforms the baseline method. Interestingly, the SMC classifier yields worse results compared to the baseline and other approaches, especially at training percentages greater than 10%. In all the experiments, the adaptive sprinkled versions of the semantic approaches give better results; consequently, their sprinkled versions fit the expectations (Chakraborti et al., 2007). The experimental results for the composite approaches show that the kernel matrix created by using more than one kernel classifier produces higher scores than using a single kernel classifier.

Experimental results for the semi-supervised experimental settings show that all of the approaches are better than the baseline method, as can be seen in Table 3. A remarkable point here is that the RVC semi-supervised classifier achieves an F1 score of 80.18% at 1% labeled set percentage, whereas the linear SVM classifier produces an F1 score of 73.94%. The RVC classifier is selected as the best model since it performs better than any other approach including the baseline model, and the difference between the baseline method and RVC classifier is statistically significant at the 0.05 level with a p-value of 0.0000907. Next, the RVC classifier is trained on the whole of the first dataset to assign labels to the

Table 2
F1 score results of supervised classifiers on the first dataset.

Classifier name	Training percentage						
	1%	5%	10%	30%	50%	70%	90%
Linear SVM (Baseline)	73.23	77.53	79.02	81.09	82.07	82.71	83.20
AFC	76.06	79.62	80.68	81.97	82.48	82.76	82.97
Sprinkled-AFC	75.76	79.49	80.59	81.91	82.43	82.72	82.94
Adaptive Sprinkled-AFC	76.08	79.62	80.67	81.96	82.47	82.75	82.96
SMC	76.08	79.84	80.85	81.62	81.83	81.94	82.03
Sprinkled-SMC	75.57	78.81	78.86	78.58	78.58	78.27	78.00
Adaptive Sprinkled-SMC	75.80	78.50	79.02	79.15	78.85	78.59	78.40
RVC	77.08	80.40	81.39	82.60	83.08	83.31	83.53
Sprinkled-RVC	71.22	75.80	77.02	78.47	78.90	79.29	79.47
Adaptive Sprinkled-RVC	76.66	80.10	81.07	82.20	82.54	82.77	82.73
Composite (AFC + RVC + SMC)	76.51	80.06	81.10	82.28	82.77	83.06	83.28
Sprinkled Composite (AFC + RVC + SMC)	75.44	79.21	80.33	81.63	82.14	82.47	82.68
Adaptive Sprinkled Composite (AFC + RVC + SMC)	76.49	79.90	80.95	82.23	82.71	83.03	83.23

Table 3
F1 score results of semi-supervised classifiers on the first dataset.

Labeled Data %	Unlabeled Data %	Test Data %	Baseline Linear SVM	SMC	RVC	AFC
1%	79%	20%	73.94	78.02	80.18	78.88
5%	75%	20%	78.03	79.96	81.29	79.91
10%	70%	20%	79.45	80.37	81.66	80.29
15%	65%	20%	80.14	80.59	81.90	80.54

second dataset.

4.3.2. Experimental results of the sentiment analysis step

4.3.2.1. Sentiment analysis of users and topic categories. After calculating the sentiment polarity score for each tweet, the tweets are classified as positive-sentiment or negative-sentiment. Next, the positive-sentiment tweet rate of each user is calculated. In addition, the polarity score for each user is calculated by taking the average of the polarity scores of the tweets of each user.

The positive-tweet rate and the polarity score are calculated for each category. According to the positive-sentiment polarity tweet rate in Table 4, it is seen that the highest positive-sentiment polarity tweet rates among the categories of science, education, economy, culture/arts, and sports are those for the education and culture/arts categories.

Sentiment polarity analysis of opinion leaders

Average positive-sentiment and negative-sentiment tweets and rates of opinion leaders calculated according to Eq. (6) are reported in Table 5. Two approaches are discussed and analyzed for using sentiment polarity to detect opinion leaders. The first possible approach involves adding the sentiment polarity score directly to the opinion leader detection formula by multiplying it by a coefficient as a feature. In this way, the intent would be to use the sentiment polarity score as a threshold to remove any opinion leader from the list of candidate opinion leaders if his/her score were very irrelevant according to the sentiment polarity score for the topic. It was observed in this work that analyzing the opinion leader and topic sentiment polarity scores according to time would give results that are more accurate. Using the Pandas library, the sentiment polarity scores of each topic and each opinion leader were therefore translated into a time series. These results were then examined and plotted.

The second possible approach would involve calculating the Pearson correlation coefficient between the sentiment polarity score of the opinion leader and the sentiment polarity score for the topic. By taking the absolute value of this coefficient, the opinion leaders whose Pearson correlation coefficients are less than 0.5 would be eliminated and it could be seen whether the accuracy increased or not. According to the experimental results, a very slight performance improvement was observed. Moreover, instead of Pearson correlation coefficients, the same comparisons were done by calculating the dynamic time warping score between the leaders and the topic. However, the Pearson correlation coefficient was preferred in this case because dynamic time warping calculations are very complex and costly and give almost the same results as Pearson correlation coefficients.

4.3.3. Experimental results of the opinion leader detection step and discussion

Evaluation of opinion leader detection is actually a difficult job since there is no standard method to estimate opinion leaders. In the literature, different approaches are tried in many different studies since this evaluation can be subjective and specific to the problem domain. Therefore, the evaluation of the results in this study is done using two different methods to prove that the suggested methodology outperforms PageRank. The results are also evaluated by considering the user-topic sentiment correlation of the retrieved lists.

The similarity scores (%) of the PageRank and SNOL algorithms in terms of *RTcount* results are shown in Table 6 and Table 7. Table 6 reports the scores without and Table 7 reports the scores with user-topic sentiment correlation. According to the experimental results in Table 6, the SNOL and PageRank algorithms seem to give the same similarity scores in the Sports category for the top 20. However, SNOL exceeds PageRank in the Culture and Arts category and the Politics category, while PageRank exceeds SNOL in both the Economy and Technology categories for the top 20. For the top 30, equality between PageRank and SNOL is seen only in the Culture and Arts category. In all other categories, SNOL exceeds the PageRank algorithm for the top 30. The same trend is seen for the top 40, 50, 60, 70, 80, 90, and 100. The reason for the results increasing as the N value rises is that the number of people considered opinion leaders in the lists approaches the number of users present in the network. Another reason is that, since users are ranked based on the values they have in the algorithm, their orders in the lists can be different, and as N increases, these different ranks are better captured.

SNOL shows better results than PageRank without user-topic sentiment correlation results, whereas, as seen in Table 7, the results decrease when sentiment correlation is introduced to the methodology. The reason for this could be that the categories of topics were created as general content instead of focusing on specific events while the datasets were constructed. While an event in one category has a positive sentiment, another in the same category may have a negative sentiment. This could lead to poor results in the opinion leader lists.

Comparisons between SNOL and PageRank using the spread score, which is the second evaluation method, are shown in Figs. 4–11. Spread scores of the PageRank and SNOL lists without user-topic sentiment correlation for the technology, culture and arts, economy, and politics datasets are shown in Figs. 4–7, while spread scores of the PageRank and SNOL lists with user-topic sentiment correlation for the technology, culture and arts, economy, and politics datasets are shown in Figs. 8–11.

In almost all categories, the top N users selected by SNOL spread information more than the users selected by PageRank. Results for the economy and technology categories show that PageRank and SNOL give almost the same spread scores as N increases. This can be explained by the fact that these networks have a small number of users, and so, eventually, both algorithms pick the same people.

In the categories of culture and arts and politics, SNOL notably outperforms PageRank as N increases according to Fig. 5 and Fig. 7.

The results of spread score evaluation with user-topic sentiment correlation show that SNOL achieves a higher spread score than PageRank for the categories of economy, culture and arts, and technology according to Figs. 8–10.

Spread scores of the PageRank and SNOL lists with user-topic sentiment correlation for the economy dataset are shown in Fig. 8. According to the experimental results illustrated in Fig. 8, SNOL is superior to PageRank with the economy dataset. For example, with 80 opinion leaders, the spread score results of SNOL and PageRank are 1800 and 1600, respectively.

Spread scores of PageRank and SNOL lists with user-topic sentiment correlation for the culture and arts dataset are shown in Fig. 9. According to the experimental results illustrated in Fig. 9, the superiority of SNOL over PageRank is clear. For instance, the spread score difference between SNOL and PageRank is about 1500 with 100 opinion leaders.

The superiority of SNOL over PageRank for the technology dataset is especially visible between 20 and 50 opinion leaders, as shown in Fig. 10. The spread score of SNOL is higher than the spread score of PageRank between 22 and 62 opinion leaders for the politics dataset according to Fig. 11. With between 62 and 100 opinion leaders, PageRank yields a higher spread score compared to SNOL for the politics dataset. This may be explained by the insufficient coverage of the semantic lexicons used in SNOL. In future work, we will use different semantic lexicons with wider coverage.

We also compared our experimental results against those of three

Table 4
Average sentiment polarity score of the users for the categories of the first dataset.

	Number of positive-sentiment tweets	Number of negative-sentiment tweets	Number of tweets	Positive-tweet ratio	Negative-tweet ratio	Average value of polarity score
Science	2209	1012	3221	0.685812	0.314188	0.046532
Education	21,278	6372	27,650	0.769548	0.230452	0.072963
Economy	59,179	34,931	94,110	0.628828	0.371172	0.025099
Culture and Arts	24,609	9243	33,852	0.726959	0.273041	0.049095
Politics	143,650	96,160	239,810	0.599016	0.400984	0.01839
Sports	60,522	29,342	89,864	0.673484	0.326516	0.041598

Table 5
Average positive-sentiment and negative-sentiment tweets and rates of opinion leaders calculated according to Eq. (6).

	Average Number of positive-sentiment tweets	Average number of negative-sentiment tweets	Average number of tweets	Average ratio of positive-sentiment tweets
Science	141.74	83.04	224.78	0.677053
Education	178.43	75.28	253.71	0.733759
Economy	233.55	121.81	355.36	0.699804
Culture and Arts	175.25	79.85	255.1	0.722094
Politics	349.81	205.45	555.26	0.665777
Sports	270.91	129.18	400.09	0.705134

other opinion leader detection methods previously presented in the literature. The first one is the OLMiner algorithm presented by Chen (2019). The datasets used in that work were collected from Mobile01,[3] a platform hosting discussions about different car brands. We selected the Audi car brand in order to construct a dataset. We collected discussions about Audi from Mobile01 using Web Crawler and Google Translate. We collected about 6000 articles, 85,000 reply relationships, and 18,000 users. We prepared this dataset and applied it in our experimental environment. We used all textual materials instead of tweets in the SNOL system. We applied the same preprocessing to these textual materials as we apply to tweets in the SNOL system. We also set the users in the Audi dataset up similarly to the users of the SNOL system. We compared our information spread scores to the information spread scores of OLMiner as reported by Chen (2019) for numbers between 10 opinion leaders and 100 opinion leaders. According to the experimental results depicted by

Table 6
Similarity score (%) of PageRank and SNOL algorithms in terms of *RTcount* results without user-topic sentiment correlation.

Method	Top-N									Category
	20	30	40	50	60	70	80	90	100	
PageRank	**25**	30	35	40	40	44	44	48	48	Economy
SNOL	20	40	43	42	43	50	53	58	54	
PageRank	0	**7**	8	6	3	7	8	9	11	Culture and Arts
SNOL	5	7	10	14	12	17	23	24	27	
PageRank	35	30	30	30	26	24	24	23	21	Politics
SNOL	40	33	38	32	33	31	30	31	28	
PageRank	**15**	13	15	16	15	19	19	23	26	Sports
SNOL	15	20	20	22	27	30	29	29	29	
PageRank	**30**	**30**	30	34	40	43	48	48	48	Technology
SNOL	25	30	35	36	43	49	49	49	49	

Table 7
RTcount results of the PageRank and SNOL lists with user-topic sentiment correlation.

Method	Top-N									Category
	20	30	40	50	60	70	80	90	100	
PageRank	**23**	28	33	35	36	39	40	42	46	Economy
SNOL	17	36	37	38	39	41	46	50	52	
PageRank	0	6	7	6	3	7	8	9	10	Culture and Arts
SNOL	4	6	8	13	11	16	21	23	25	
PageRank	32	27	28	29	24	22	23	23	21	Politics
SNOL	38	31	33	31	32	30	29	30	27	
PageRank	12	11	13	14	15	17	18	21	25	Sports
SNOL	12	18	19	21	25	27	28	29	28	
PageRank	**26**	27	29	31	32	34	35	37	40	Technology
SNOL	24	27	33	34	35	36	37	39	42	

[3] http://www.mobile01.com

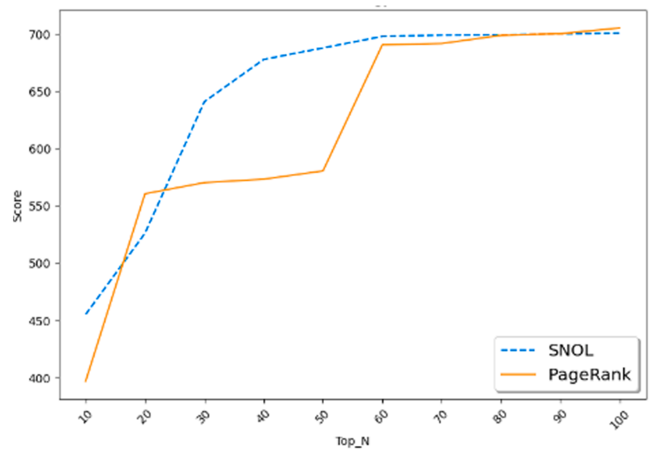

Fig. 4. Spread scores of the PageRank and SNOL lists without user-topic sentiment correlation for technology dataset.

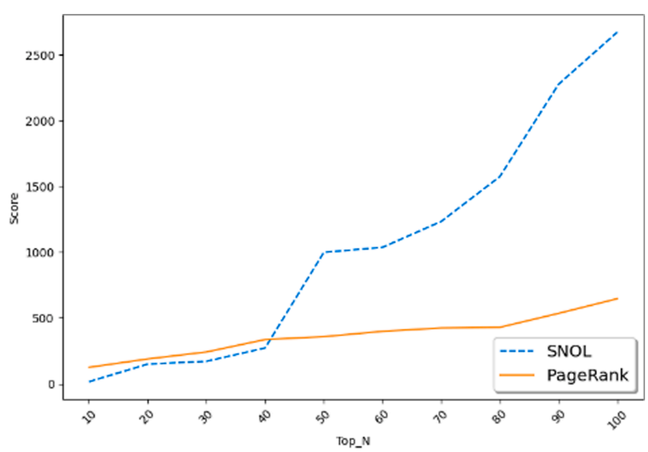

Fig. 5. Spread scores of the PageRank and SNOL lists without user-topic sentiment correlation for culture and arts dataset.

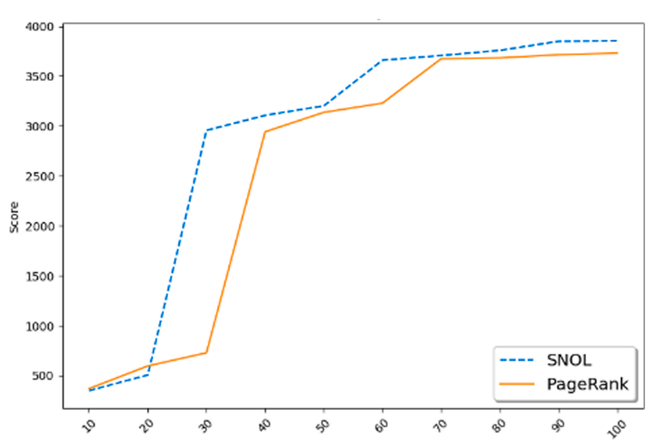

Fig. 6. Spread scores of the PageRank and SNOL lists without user-topic sentiment correlation for economy dataset.

Chen (2019), the information spread scores of OLMiner for 20 opinion leaders, 50 opinion leaders, and 100 opinion leaders are 6000, 8000, and 10,000, respectively. On the other hand, the information spread scores of SNOL for 20 opinion leaders, 50 opinion leaders, and 100 opinion leaders are 7500, 9700, and 11,150, respectively.

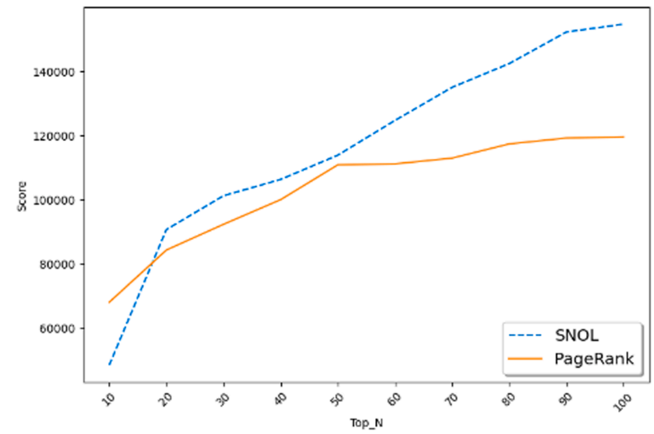

Fig. 7. Spread scores of the PageRank and SNOL lists without user-topic sentiment correlation for politics dataset.

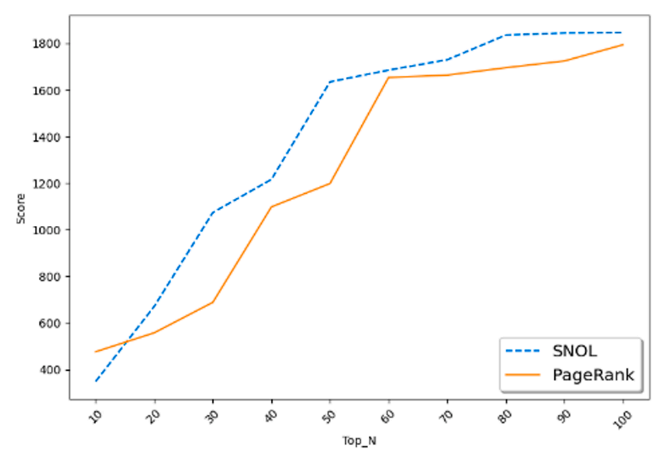

Fig. 8. Spread scores of the PageRank and SNOL lists with user-topic sentiment correlation for economy dataset.

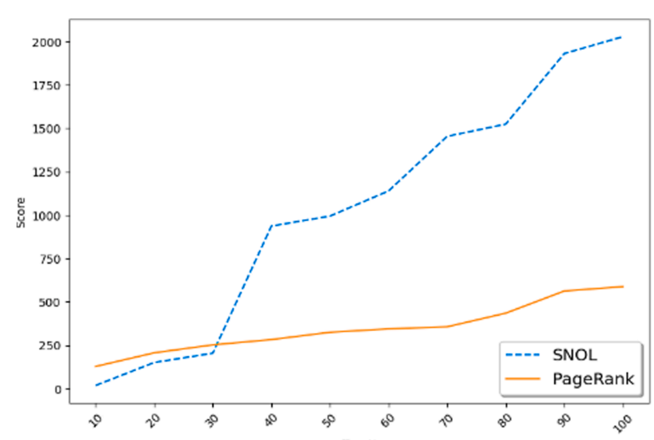

Fig. 9. Spread scores of the PageRank and SNOL lists with user-topic sentiment correlation for culture and arts dataset.

The second previous study in the literature that we compared our results to is that of Jesus et al. (2014). In that study, a learning system was developed for the RepLab 2014 author profiling task at UNED. This system is based on a voting model. The features used for this system are tweet texts' POS tags, number of hashtags, number of links, number of mentions, number of followers, number of emoticons, and retweet

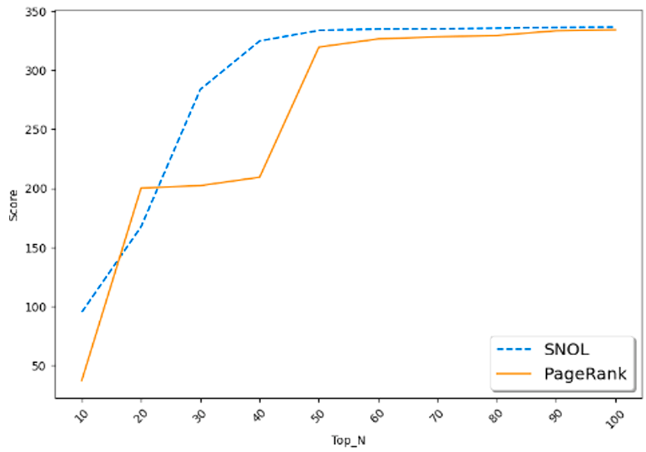

Fig. 10. Spread scores of the PageRank and SNOL lists with user-topic sentiment correlation for technology dataset.

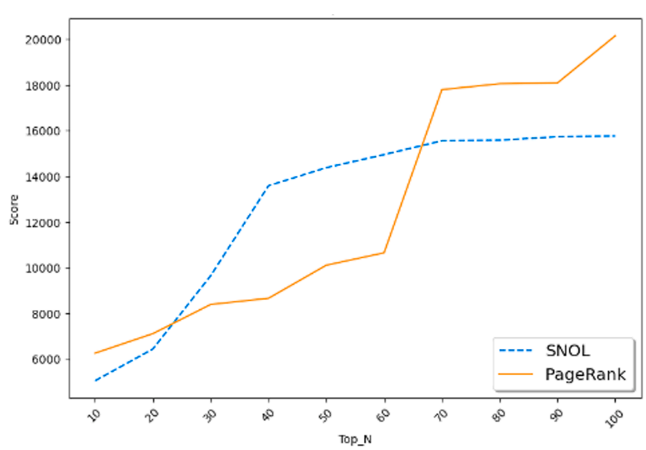

Fig. 11. Spread scores of the PageRank and SNOL lists with user-topic sentiment correlation for politics dataset.

Table 8
The precision values of the random forest algorithm, naive Bayes algorithm, and SNOL on the economy dataset for different numbers of opinion leaders.

Algorithm	Number of opinion leaders			
	10	20	30	100
Random forest algorithm	0.3	0.45	0.52	0.74
Naive Bayes algorithm	0.23	0.39	0.47	0.65
SNOL	0.38	0.49	0.55	0.82

speed. These features are used in different weighting schemes. Since we already have similar features in our dataset in SNOL, we tried to implement the methodology presented by Jesus et al. (2014). Those authors applied the random forest algorithm and naive Bayes algorithm from WEKA. We applied these algorithms to our economy dataset, implementing the algorithms in Python with the scikit-learning library. The precision values of these algorithms are shown in Table 8. According to the experimental results listed in Table 8, the precision values of the random forest algorithm, naive Bayes algorithm, and SNOL are 0.74, 0.65, and 0.82 with 100 opinion leaders. The superiority of SNOL can be observed at each number of opinion leaders.

The third study from the literature against which we compared our results is that of Jain et al. (2020). Two real datasets were used in their study, one of which is called "Wiki-vote dataset." We utilized the Wiki-vote dataset in our experimental environment. This dataset takes the form of an undirected graph and it can be accessed via a GitHub link.[4]

We applied this undirected graph structure in the experimental environment of the SNOL system. The firefly optimization algorithm presented in this study obtains a precision score of 0.84 with the Wiki-vote dataset as reported by Jain et al. (2020). According to our experimental results, SNOL obtains a precision score of 0.87 with this dataset.

The superiority of SNOL compared to different studies from the literature could be explained by the capability of SNOL using sentiment information in both the methodology and evaluation parts.

5. Concluding remarks

The SNOL system has been proposed in this paper. The purpose of this system is to identify topic-based opinion leaders of social media. In order to accomplish this goal, several steps are undertaken, including data collection, data processing, data analysis, data classification, the ranking of topic-based opinion leaders, and evaluation. SNOL has two main parts, namely topic modeling and user modeling. In topic modeling, tweets are classified with several semantic kernels and their sprinkling and adaptive sprinkling versions. In user modeling, the feature set of each user is constructed in the social network, built from the collected tweets. After the calculation of degree centrality, betweenness centrality, and closeness centrality, several user features such as focus rate, activeness, authenticity, and follower/following ratio are also included in the feature set. The experiments are performed on data collected from Twitter including 17,234,924 tweets from 38,727 users. According to the topic modeling and user modeling results, leadership scores are given to each user in the network. Users with the highest scores are said to be opinion leaders. In order to evaluate SNOL's performance, the PageRank algorithm is also run on the same dataset.

Comparison of the results obtained from PageRank and SNOL is a very difficult task because there is no standard method to evaluate opinion leaders. In the literature, different methods are applied in many different studies since this evaluation can be subjective and specific to the problem domain. Therefore, the evaluation of the results in this paper is done using two different methods (i.e. retweet count and spread score) to prove that SNOL outperforms PageRank. The results obtained for *RTcount* are shown in Table 6, where it is clear that SNOL exceeds or gives equal results to PageRank for almost all N values in each category. The results increase as the N value rises because the number of people considered as opinion leaders in the lists accordingly approaches the number of users present in the network. Another reason for this result is that, since the users are ranked according to the values they have in the algorithm, their order in the lists can be different, and as N increases, these different rankings are better captured.

The comparisons between SNOL and PageRank using the spread score, which is the second evaluation method of the present work, are shown in Figs. 4–13. In almost all categories, the top N users selected by SNOL spread information more than the users selected by PageRank. Results for the categories of economy and technology show that PageRank and SNOL give almost the same spread scores as N increases. This is the case because these networks have a small number of users and so, eventually, both algorithms pick the same people. In the politics and the culture and arts categories, according to the experimental results, our SNOL framework generates remarkable performance compared to the PageRank algorithm for nearly all topics and all selected top N opinion leaders. These preliminary results motivate us to improve our model with the contribution of some other user features, especially in the user modeling part.

SNOL particularly outperforms PageRank as N increases. Spread score results show that SNOL achieves a higher spread score than PageRank in the categories of economy, culture and arts, and technology. In the politics category, SNOL is better than PageRank, but after a certain

[4] https://github.com/nsitlokeshjain/opinionleader

point PageRank gives a higher spread score.

SNOL has also been compared to some other opinion leader detection methods previously presented in the literature. The experimental results show that SNOL generates remarkably better performance than the PageRank algorithm and the other compared algorithms from the literature for nearly all topics and all selected top N opinion leaders.

As future work, these operations performed on the static network can be applied to dynamic networks to identify real-time opinion leaders. In this way, an environment suitable for Twitter's natural dynamics can be created and adapted to changing follower/friend relations.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by the Scientific and Technological Research Council of Turkey (TÜBİTAK) with grant number 118E315. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of TÜBİTAK.

References

Akin, A.A. & Dundar, M. (2007). Zemberek, an open source nlp framework for turkic languages. Structure, 10.

Aleahmad, A., Karisani, P., Rahgozar, M., & Oroumchian, F. (2016). OLFinder: Finding opinion leaders in online social networks. *Journal of Information Science, 42*(5), 659–674.

Alp, Z. Z., & Öğüdücü, Ş. G. (2019). Influence factorization for identifying authorities in twitter. *Knowledge-Based Systems, 163*, 944–954.

Alp, Z. Z., & Öğüdücü, Ş. G. (2018). Identifying topical influencers on twitter based on user behavior and network topology, Knowledge-Based Systems, 141, 211–221.

Alp, Z.Z., Ögüdücü, S.G. (2015). Extracting topical information of tweets using hashtags, in: 14th IEEE International Conference on Machine Learning and Applications, ICMLA 2015, Miami, FL, USA, December 9–11, 2015, 2015, pp. 644–648.

Altınel, B., Ganiz, M. C., Şipal, B., Erkaya, E., Yücedağ, O. C., & Doğan, M. A. (2019). Word sense disambiguation using semantic kernels with class-based term values. *Turkish Journal of Electrical Engineering & Computer Sciences, 27*(4), 3180–3194.

Altınel, B., Diri, B., & Ganiz, M. C. (2015). A novel semantic smoothing kernel for text classification with class-based weighting. *Knowledge-Based Systems, 89*, 265–1177.

Altınel, B., Ganiz, M.C., Diri, B. (2015). A Corpus-Based Semantic Kernel for Text Classification by using Meaning Values of Terms, Elsevier, Engineering Applications of Artificial Intelligence Volume 43, August 2015, Pages 54–66. doi:10.1016/j.engappai.2015.03.015.

Amor, B. R., Vuik, S. I. , Callahan, R. , Darzi, A. , Yaliraki, S. N. , & Barahona, M. (2016). Community detection and role identification in directed networks: Understand- ing the twitter network of the care. Data debate. In Dynamic networks and cy- ber.

I. Anger C. Kittl Measuring influence on Twitter In Proceedings of the 11th international conference on knowledge management and knowledge technologies 2011 1 4.

Boser, B.E., Guyon, I.M., Vapnik, V.N. (1992). A training algorithm for Optimal margin classifier. In Proceedings of the 5th ACM Workshop. Comput. Learn. Theory, 144–152.

Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010). Senticnet: A publicly available semantic resource for opinion mining. In 2010 AAAI Fall Symposium Series.

Carta, S., Podda, A. S., Recupero, D. R., Saia, R., & Usai, G. (2020). Popularity Prediction of Instagram Posts. *Information, 11*(9), 453.

Chakraborti, S., Lothian, R., Wiratunga, N., & Watt, S. (2006). Sprinkling: supervised latent semantic indexing, In European Conference on Information Retrieval (pp. 510-514). Springer, Berlin, Heidelberg.

Chakraborti, S., Mukras, R., Lothian, R., Wiratunga, N., Watt, S. N., & Harper, D. J. (2007). Supervised Latent Semantic Indexing Using Adaptive Sprinkling, In IJCAI (Vol. 7, pp. 1582-1587).

Chen, Y. C. (2019). A novel algorithm for mining opinion leaders in social networks. *World Wide Web, 22*(3), 1279–1295.

Cho, Y., Hwang, J., & Lee, D. (2012). Identification of effective opinion leaders in the diffusion of technological innovation: A social network approach. *Technological Forecasting and Social Change, 79*(1), 97–106.

Cui, L., & Pi, D. (2017). Identification of micro-blog opinion leaders based on user features and outbreak nodes. International Journal of Emerging Technologies in Learning, 12(1).

Dehkharghani, R., Saygin, Y., Yanikoglu, B., & Oflazer, K. (2016). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation, 50*(3), 667–685.

Dehkharghani, R., Yanikoglu, B., & Saygin, S. (2015). Sentiment Analysis in Turkish : Towards a Complete Framework.

Duan, J., Zeng, J., & Luo, B. (2014). Identification of opinion leaders based on user clustering and sentiment analysis. In 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 1.

Gökçe, O. Z., Hatipoğlu, E., Göktürk, G., Luetgert, B., & Saygin, Y. (2014). Twitter and politics: Identifying Turkish opinion leaders in new social media. *Turkish Studies, 15*(4), 671–688.

Hu, M., Liu, S., Wei, F., Wu, Y., Stasko, J., & Ma, K.L. (2012). Breaking news on twitter, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2751–2754). ACM.

Jain, L., & Katarya, R. (2019). Discover opinion leader in online social network using firefly algorithm. *Expert Systems with Applications, 122*, 1–15.

Jain, L., Katarya, R., & Sachdeva, S. (2020). Opinion leader detection using whale optimization algorithm in online social network. Expert Systems with Applications, 142, 113016.

Jesus J, Lomena M, and Lopez Ostenero F. UNED at CLEF RepLab 2014: Author Profiling, in Proceedings of the Fifth International Conference of the CLEF initiative. Springer, 2014, pp. 1537–1549.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In Proceedings of the ninth acm sigkdd international conference on knowledge discovery and data mining (pp. 137–146). ACM.

Li, F., & Du, T. C. (2011). Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs. *Decision Support Systems, 51*(1), 190–197.

Li, C., Bai, J., Zhang, L., Tang, H., & Luo, Y. (2019). Opinion community detection and opinion leader detection based on text information and network topology in cloud environment. *Information Sciences, 504*, 61–83.

Luo, L., Yang, Y., Chen, Z., & Wei, Y. (2018). Identifying opinion leaders with improved weighted LeaderRank in online learning communities. International Journal of Performability Engineering, 14(2), 193–201.

McCallum, A. K. (2002). "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. McCallum, A. K. (2002). "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu.

Meltzer, D., Chung, J., Khalili, P., Marlow, E., Arora, V., Schumock, G., & Burt, R. (2010). Exploring the use of social network methods in designing healthcare quality improvement teams. *Social Science & Medicine, 71*(6), 1119–1130.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web, Stanford InfoLab.

Phan, H. T., Nguyen, N. T., Van Cuong, T., & Hwang, D. (2019). *A method for detecting and analyzing the sentiment of tweets containing fuzzy sentiment phrases* (pp. 1–6). IEEE.

Phan, H. T., Tran, V. C., Nguyen, N. T., & Hwang, D. (2020). Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model. *IEEE Access, 8*, 14630–14641.

Rehman, A. U., Jiang, A., Rehman, A., Paul, A., & Sadiq, M. T. (2020). Identification and role of opinion leaders in information diffusion for online discussion network. *Journal of Ambient Intelligence and Humanized Computing*, 1–13.

Saia, R., Boratto, L., & Carta, S. (2014). Semantic Coherence-based User Profile Modeling in the Recommender Systems Context. In KDIR (pp. 154–161).

Song, X., Chi, Y., Hino, K., & Tseng, B. (2007). Identifying opinion leaders in the blogosphere. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (pp. 971–974). ACM.

Van Eck, P. S., Jager, W., & Leeflang, P. S. (2011). Opinion leaders' role in innovation diffusion: A simulation study. *Journal of Product Innovation Management, 28*(2), 187–203.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.

Zhang, B., Bai, Y., Zhang, Q., Lian, J., & Li, M. (2020). An opinion-leader mining method in social networks with a phased-clustering perspective. *IEEE Access, 8*, 31539–31550.

Zhao, Q. , Erdogdu, M. A. , He, H. Y. , Rajaraman, A. , & Leskovec, J. (2015). Seismic: A self-exciting point process model for predicting tweet popularity. In Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data min.

Zhao, Y., Li, S., & Jin, F. (2016). Identification of influential nodes in social net- works with community structure based on label propagation. *Neurocomputing, 210*, 34–44.