Sigma J Eng Nat Sci, Vol. 40, No. 2, pp. 441–463, June, 2022

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2022.00043

Review Article

From past to present: Spam detection and identifying opinion leaders in social networks

Ayşe Berna ALTINEL GİRGİN*[1], Gizem GÜMÜŞÇEKİÇCİ[2]

[1]Department of Computer Engineering, Faculty of Technology, Marmara University, Istanbul, Türkiye
[2]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Işık University, Istanbul, Türkiye

ARTICLE INFO

Article history
Received: 17 November 2020
Accepted: 08 October 2021

Key words:
Social Network Analysis; Opinion Leader Detection; Flow of Influence; Spam Filtering

ABSTRACT

On microblogging sites, which are gaining more and more users every day, a wide range of ideas are quickly emerging, spreading, and creating interactive environments. In some cases, in Turkey as well as in the rest of the world, it was noticed that events were published on microblogging sites before appearing in visual, audio and printed news sources. Thanks to the rapid flow of information in social networks, it can reach millions of people in seconds. In this context, social media can be seen as one of the most important sources of information affecting public opinion. Since the information in social networks became accessible, research started to be conducted using the information on the social networks. While the studies about spam detection and identification of opinion leaders gained popularity, surveys about these topics began to be published. This study also shows the importance of spam detection and identification of opinion leaders in social networks. It is seen that the data collected from social platforms, especially in recent years, has sourced many state-of-art applications. There are independent surveys that focus on filtering the spam content and detecting influencers on social networks. This survey analyzes both spam detection studies and opinion leader identification and categorizes these studies by their methodologies. As far as we know there is no survey that contains approaches for both spam detection and opinion leader identification in social networks. This survey contains an overview of the past and recent advances in both spam detection and opinion leader identification studies in social networks. Furthermore, readers of this survey have the opportunity of understanding general aspects of different studies about spam detection and opinion leader identification while observing key points and comparisons of these studies.

Cite this article as: Girgin Altınel AB, Gümüşçiçekçi G. From past to present: Spam detection and identifying opinion leaders in social networks. Sigma J Eng Nat Sci 2022;40(2):441–463.

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

This paper was recommended for publication in revised form by Regional Editor Ahmet Selim Dalkilic

INTRODUCTION

In recent years social platforms have gained popularity and with this popularity people tend to share their ideas on important events and topics in these platforms, creating social networks. Due to these reasons, many studies have begun to use the data and networks from social platforms [1,2]. Therefore, different types of study began to appear in the literature about spam detection and identification of opinion leaders. With the increase of studies about spam detection and identification of opinion leaders, surveys about these topics also have started to take place in the literature. Most of these surveys cover the application of different methodologies to a certain degree. However, to the best of our knowledge, there is no survey that contains approaches for both detection of influencers and filtering of the spam content in social networks in a single item of content. In order to fill this gap, we performed a comprehensive study. This survey explored the past and recent advancements in the domains of identifying opinion leaders and finding spam content in social networks.

There are individual surveys that focus on a single distinct topic [1,2,3]. In the survey created by Wu et al. (2018) studies were collected about spam detection based on only Twitter data and categorized studies based on approaches and accuracy [3]. The authors state that Twitter is one of the most popular microblogging platforms [3]. In the study they discussed the features of Twitter spam detection and categorized studies according to their proposed approaches [3]. In the reviewing process the advantages and disadvantages of the studies were considered [3]. This study is helpful due to including comparisons of studies on Twitter spam detection while presenting studies aspects such as their methodologies, accuracies etc. [3]. The advantages, disadvantages and comparison of studies are gathered into this study, overall it presents a good review about Twitter spam detection.

Another study by Chakraborty et al (2016), social spam studies were reviewed [1]. In the survey studies were classified according to the features, properties and platforms on which it is posted and studies were analyzed according to approaches, and accuracies of the studies and general methodology of the studies were reviewed [1]. Authors also addressed the challenges faced in social spam detection, and suggested a road map on how to use current social spam detection approaches to handle the problems that might occur in the detection process [1]. For analysis of influence in social networks a study was performed by Peng et al (2018) [2]. They reviewed studies about influence in social networks at many different levels such as definition, properties, architecture, applications, and diffusion models [2]. When our study is compared with the surveys mentioned above, some similarities and differences can be detected. Our study differs from [1,2,3] in terms of the number of topics covered. This survey contains studies from both spam detection and opinion leader identification topics. These studies are specifically analyzed and reviewed in this survey. Since this survey is more up to date and covers both spam detection and identification of opinion leaders, it stands out more than other existing surveys with adding innovation to the literature by covering two different topics in a single study. The reviewing style of studies behind our study and studies [1,2,3] are similar which consists of methodology, dataset and evaluation.

Another similarity between this survey and other surveys is that it analyzes and reviews related studies on different levels and mostly focuses on summarizing the reviewed studies. To explain in more detail, this survey collects important studies from past to present about spam detection and identification of opinion leaders in social networks. It then reviews the related studies and categorizes them according to its approaches and features. The existing approaches for the detection of opinion leaders are divided into five fundamental categories: diffusion-based approaches, graph-based approaches, statistical and stochastic approaches, PageRank-based approaches, and machine learning approaches. Like the surveys mentioned above this survey also addresses challenges that occur in the processes. Furthermore, our survey attempts to organize existing approaches to the detection of spam content under four fundamental categories: methods using user-based and content-based features, methods using honeypot features, and sentiment and graph-based approaches. Multi-process applications are usually more difficult to implement than single process applications. Regardless of their approach, there are multiple steps used in both spam detection and identifying opinion leaders in social networks. Thus, there are some challenges that might occur in these processes. These include the complexity of computations, the availability of datasets, the cost of training, etc. Usually the approaches used in opinion leader identification are complex processes. Different approaches require different resources. For example, Graph-based and PageRank-based approaches require more resources while diffusion-based approaches are expensive to compute due to having complex mathematical computations. Some machine learning approaches are expensive to train. For spam detection approaches there is a general problem in the realization process. The problem is the versatility of spam content. Spammers constantly change its content so for spam detection systems detecting the spam content is a big challenge. Besides the computation problems, there are some issues about the availability of datasets that are used in the studies. Usually finding a proper and sufficient social network dataset is hard. There are limited resources for different languages. Mostly English datasets are more common and easier to find. The evaluation analysis of opinion leader identification is also explained in Section 2. Briefly, there are ways for evaluation of opinion leader identification systems. Mostly the number of retweets a user gets can be an

important indicator for evaluation. Another indicator is the number of in-degree and out-degree values a user gets in a social network. The present study detailly explains and analyzes the most common challenges faced in spam detection and identification of opinion leaders in social networks.

Social Network Analysis (SNA)

Online social networking sites have increased in popularity in recent years. Worldwide users use these sites to make new friends, keep up to date with their friends, and follow current ideas and activities. The prevalence of such sites among Internet users in Turkey and their use are increasing with each passing day. Comments on companies, institutions, and individuals on social networks have a significant impact on social network users. Due to the prevalence and intensive use of social networks, the content shared on these networks today can be considered one of the most important sources of information affecting the opinion of the public.

Identifying the Influencers on Social Media

Are the messages that affect the behavior and thoughts of individuals in society transferred to society directly, or are these messages first received by certain people in the community and transferred to others? [4] Those who are interested in the process of entry and diffusion of messages to society have determined that messages do not directly reach large masses but are transmitted to others through a few people in the receiving mass and proposed a two-step flow of communication model. Is it possible to verify this with a simple observation? When we actually examine a community in a more detailed and careful way, we observe that the behavior and preferences of people in their daily lives (going to the cinema, fashion, shopping, choosing a holiday place, political ideas, etc.) are influenced by their close friends, families, relatives, and certain people in their business circles. We can define these people as opinion leaders who are followed by the majority of the community when making choices. They have certain characteristics that distinguish them from other people in the community: First of all, they are the people who need information or need advice. The attitudes and behaviors of opinion leaders are easily adopted by groups. For example, members of an "X" group living in an area where girls do not go to school for social reasons can send their daughter to school when the group leader sends her daughter to school. Opinion leaders have the ability to influence other people; they are known to be trusted in their environment. Successful identification of these people, who are critical in their communities, is an important and worthy task.

In the literature on social networks, the concept of "electronic influencers", which shows the people followed by the majority of the society, was first introduced by Hon and his team in 1999 [5], and then Johnson and his research team in 2007 analyzed electronic influencers in various dimensions [6]. According to Johnson and his team, electronic influencers are considered users that affect the diffusion rate and volume of information in a society.

Twitter was founded in 2006 and can be defined as a microblogging environment that allows individuals to learn what is happening and express their feelings, experiences, thoughts, and so on. There are many journalists, artists, athletes, politicians, and other famous people who can be reached directly on Twitter. Twitter gives users the freedom to share (retweet) and re-share their favorite tweets with their own accounts. This viral effect directly results in "micro-celebrity", which means that ordinary people become celebrities in the virtual environment [7]. These people have the power of addressing thousands of people at the same time with a single tweet. Due to these features, these people can also be regarded as opinion leaders of the social network in which they are involved. Therefore, the presence of opinion leaders is a very important issue for research with many social benefits. While academic studies on this subject are at an early stage, there has been increasing activity in recent years [8,9].

Filtering Spam Content of Social Media

The prevalence of social networks is directly proportional to the quality of content created by users. As in the email environment, unfortunately, social media environments are a tempting medium for spammers. Malicious and misleading content, or spam in general, can be defined as misleading and sometimes damaging content related to an organization, product, or person transmitted to a large audience by automated computer programs called bots. The misleading content of social bots has the potential to disrupt the content on social networking sites, leading to users leaving the social network or using it less. In addition to this kind of damage, it is also possible to mislead public opinion about institutions, products, or persons in general. Spam user accounts and content created by social bot programs, which are becoming increasingly complex by imitating the behavior of real social network users, are difficult to identify and the efforts of social networking sites are insufficient [10]. For example, recent research has shown that at least 23 million Twitter users are spam users [11]. Another popular social network, Facebook, has recently decided to use feedback from users to support automated methods of fighting spam content [12]. The academic studies on this subject are in the very early stages but there has been intense activity especially in the last few years [10].

It has been observed that even though spam accounts' tweet, retweet, follower, friend, and favorite tweet numbers are above the normal account average, they do not provide enough information for spam detection [13]. Because of the frequent use of email and comment fraud, online social network (OSN) data use traditional machine learning methods, but because short sentences and abbreviations are frequently used in social networking content, the dataset

is transformed into a rather rare matrix [14], and alternative methods for methods of interpreting basket-like data properties are studied [15]. Social network analysis is done in addition to the word and sentence analysis on the information created by the users [16,17], as well as using URL blacklist and URL routing data detection [18], creating profiles they call honey cubes, and expecting spam users to interact with these profiles [19,20].

In addition, it has been investigated whether emotional data can be used to solve the problem of spam detection. As a result, significant differences have been identified [15]. Unsupervised methods were also explored, and fake messages were extracted by extracting messages from a trusted source using messages in the shared URL content and hashtag set [21]. Although most of the research has been done for English language content, research has been done for different languages and countries [22], but, as far as we know, there is no research on Turkish content. The existing studies are fairly new and there are many ways to get more information about spam users and content detection. One of the most interesting features of this area is that spammers always discover new ways to emulate normal users and avoid spam detection. All these reasons and the importance of the subject motivated us to work on this issue.

Here, the studies that address identifying opinion leaders and filtering spam content are reviewed in depth in the light of the latest developments and research. In other words, this survey explores the past and recent advancements regarding the problem of ranking the opinion leaders and filtering spam content of social media.

The rest of the survey is organized as follows. There is a detailed description of the studies that address the finding of opinion leaders in a social network in Section 2. After that, a detailed discussion about the studies that aim to filter spam content in social networks is given in Section 3. Next, considerations, current challenges, future directions, and the concluding remarks of the researchers are presented in Section 4.

IDENTIFYING OPINION LEADERS ON SOCIAL NETWORKS

In order to detect opinion leaders in social media, a number of methods have been proposed. These methods can be grouped into five categories, namely 1.) Diffusion based approaches, 2.) Graph-based approaches, 3.) Statistical and stochastic approaches, 4.) Page-rank based approaches, 5.) Machine learning approaches.

Diffusion Process Based Approaches:

This section includes the studies that use diffusion process based approaches to identify opinion leaders in social networks. In diffusion process based approaches the social network structure is analyzed using some measures and algorithms. After the network structure is analyzed typically key users are selected according to different measures. User interaction patterns are recognized to determine the opinion leaders.

All the following studies in this section tries to identify opinion leaders on various different social networks such as Twitter, blogs etc. using different diffusion based approaches. The influence maximization problem was firstly introduced by [23]. The study was done in English and received attention immediately from researchers studying opinion leader detection. The influence maximization problem is settled with the initial users. These initial users from the network are the seeds for spreading information to the rest of the network. For example, [24] suggests a methodology that consists of the combination of an influence maximization algorithm and label propagation, named IM-LPA, in order to rank social networks' opinion leaders.

In [25], the effects of opinion leaders on the diffusion processes of a product are investigated. They first conducted an empirical survey in English among children who play a certain free online game. The survey topics can be categorized as follows: 1.) Status of each user in the adoption process, 2.) Opinion leaders use the product more, 3.) They involve others in the use of the product, 4.) Opinion leaders do not know more about the product than their followers, 5.) Sources that children used to obtain information about the product, 6.) Interpersonal influence type of user. Based on the results of this survey they inferred three characteristics of opinion leaders: 1. They are better at figuring out if the product is good. 2. Normative influence has less impact on opinion leaders than it has on their followers. 3. They are more innovative than their followers. Opinion leaders take more central positions in the network [24,25] and they build an agent-based model to test their hypotheses. The authors of the paper state that these hypotheses can be grouped into three categories, namely the importance of the influence type of the opinion leader, the importance of mass media, and the importance of a number of opinion leaders in the network. The simulation had the following three steps: mass media, word of mouth (WoM), and adoption. At first, none of the agents know about the product and at the mass media stage a predefined percentage of agents are informed. Non-leaders can only give 0 or 1 to product quality; on the other hand, leaders might know the correct product quality is 0.5. At the WoM stage agents start to hear about a product from their neighbors (followers) but they will only accept it if their neighbor is an opinion leader or has experienced real product quality. In the last stage, the agents decide whether or not to adopt the product [25]. According to [25], opinion leaders' effect on adoption percentage depends on their innovative behavior and their lower sensitivity to normative influence. If opinion leaders can judge product quality properly adoption speed is increased, which shows how strong an effect informational influence has on product diffusion. To identify the opinion

leaders the authors use an opinion leadership scale developed in another study. The opinion leaders are detected by different metrics. Opinion leaders are more involved with a product than their followers and this means they talk more about the product, and usually opinion leaders have more followers than a regular user. All these differences between opinion leaders and regular users are key in opinion leader detection. Results also show that adoption speed is much higher in a network with opinion leaders in comparison to a network without any leaders [25].

Cho and his research team (2012) studied diffusion processes with the assumption of opinion leaders' initiation [26]. They also suggest that opinion leaders have distinct features, especially regarding their degree of sociality. Moreover, according to the experimental results reported in their study, sociality centrality is the most important feature to achieve higher diffusion.

The following study uses a Twitter dataset to perform opinion leader identification. In the study conducted by Rehman et al. (2020), several centrality measures were applied in order to detect opinion leaders [27]. To detect opinion leaders a Twitter dataset was used. The dataset used is a directed Twitter network named Higgs-Twitter, which is publicly available. For instance, their in-degree and out-degree links are extracted and ranked based on their betweenness centrality values in the network. In order to analyze community evolution, the Louvain method is employed. With these processes they detected key users via user interaction patterns. Five different kinds of users were detected; they named these key users as the "conversation starter", "influencer", "active engager", "network builder", and "information bridge" [27]. To detect these key users, "in-degree" and "out-degree" connections were used. The conversation starter is a user that has many "in-degree" connections and no "out-degree" connections; this means it receives many retweets but retweets none; so this type of key user is determined as a conversation starter [27]. Another key user type is the influencer, this key user has many "in-degree" and "out-degree" connections; so the influencer is very active by generating a large number of tweets and receiving many tweets from many users, and it has the potential of being an opinion leader [27]. Another key type of user is the active engager; this type of user has plenty of "out-degree" connections but very few or no "in-degree" connections. These types of users spread information to the network by sending many tweets; so even though this type of user sends many tweets, it receives very few retweets due to this reason. The active engager type of key user cannot be the opinion leader in a network [27]. The network builder type of key user has a significant impact on a network by connecting influencer users together to create a bigger network. The last key user is the information bridge. The role of this type of user is to connect the active engager user to the influencer users, which is an important aspect for a network [27]. The experiments were conducted to analyze the network on the Twitter dataset with 256,491 nodes and 328,132 edges in the retweet network, 116,408 nodes and 150,818 edges in the mentioned network, and 38,918 nodes and 32,523 edges in the reply network. According to the results, the influencer users are the opinion leaders in a network since many types of users mention and retweet influencer users in their tweets [27]. Based on the discussion in [27], it is not satisfactory to perform analysis on a 7-day dataset with regard to getting sufficient connection patterns among the users of the network. Hence, for more detailed and satisfactory analysis of a network, a longer duration is recommended. In the present study, the contents of the tweets from the dataset were not included in the detection process; only the user connections were included but according to the authors of [27], including the contents of the tweets might be a good idea to get better relationship patterns among the users of the network.

Graph-Based Approaches:

This section includes the studies that use graph-based approaches to identify opinion leaders in social networks. Graph-based approaches are used to extract features of social networks.

The following two studies identifies opinion leaders mainly on a Twitter dataset using graph-based approaches. Cui and Pi (2017) developed a new method based on the user features and outbreak nodes, which are thought to be more effective than static features (registration time, number of good friends, etc.) that are used to determine opinion leaders [28]. This method offers a probabilistic generate-graph model. The user features contain the input values and behavior characteristics of the user, while the outbreak nodes contain values that are observed. The outbreak nodes are variants that are used to identify whether a user is an opinion leader or not. The outbreak nodes are defined in a network. Besides defining the outbreak nodes, the importance of the outbreak nodes are also calculated using outbreak index. This calculation depends on the shortest distance between the source node and the outbreak node. If the distance of an outbreak node to the source node is shorter than other outbreak nodes, the outbreak node which has the shortest distance is classified as more important than other outbreak nodes. The importance of the outbreak node directly affects the users' importance in a network and the opinion leaders are selected according to this variant. In the study two datasets are used which are the Sina micro-blog and Twitter datasets [28]. In the study, user features were also separated into user attributes and user behaviors. The user attributes were handled under four main headings and they categorized the users according to their importance and gave weight to each category. The degree to which the user is active was another title. The third was the value of the user's fans; a high value of this is a sign of good quality fans. The last one is the value of good friends. They also identified 5 user characteristics:

micro-blog original ratio, non-empty forwarding ratio, the original micro-blog interaction, non-marketing activity participation, and URL usage rate. The framework they provide consists of 3 layers: data layer, topic discovery layer, and opinion leader's discovery layer. In the first layer, they prepare the data, followed by steps like text segmentation and removing stop word, and the data are ready for layer 2. In the second layer, words that are not important in application are deleted and the processed data are obtained as source text and label set. They then categorized it by subject. In the last layer, the opinion is first calculated in the network relation matrices. Index values in relation matrices are calculated with the UCINET analysis tool. From the values of matrices the maximum values are selected then the weight of each index is determined according to the maximum values and finally the comprehensive indexes are concluded. Finally for the evaluation of the system, 60 percent of the data were used as the training set. In the dataset for Sina and Twitter, prediction accuracy is calculated as 90%. They then compared their own methods with the Bayesian algorithm and the support vector machine (SVM) algorithm and concluded that their algorithms are much better in terms of precision and recall values [28].

There are some concepts used while identifying opinion leaders on social networks with graph-based approaches. Some of these concepts are Degree Centrality, Betweenness Centrality, Closeness Centrality and Eigenvector centrality. These concepts are explained below including the concepts formulation and explanations of symbols in the formula of each concept.

Degree Centrality: This indicates the number of nodes which a node has directly relations with on a network graph. Degree Centrality shows the relationship structure of nodes in a social network. In other words the Degree Centrality finds the number of neighbors a node has on a graph. In Figure 1, Diane acts as a "connector" or "hub", as the person with the highest degree of centrality (orgnet). Equation (1) is the formula for the degree of centrality:

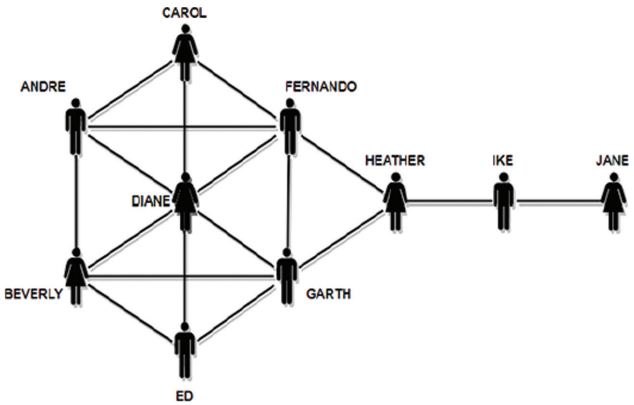

Figure 1. Social network between Diane and her friends (orgnet) [29].

$$Degree\ Centrality = CD(i) = \sum_{j=1}^{n} \left(Aij\right) \qquad (1)$$

Symbol A represents the node, to find the neighbors of the node A, matrix form of the graph is used. Matrix form of a graph directly shows the relation of nodes in the graph. In this example; each row and column specifies a node in the matrix, "j" represents the columns and "i" represents rows. Using the Equation (1) number of neighbors of a node is calculated. From the equation both the in-degree (the number of incoming edges of a node) and out-degree (the number of outgoing edges of a node) neighbors of a node can be calculated.

Betweenness Centrality: The ability to be connected to different nodes as they are located in the network graph represents the relationship structure in a social network. From this structure the importance of nodes in a social graph can be determined. Betweenness Centrality detects the influence of a node using the relationship structure of a graph. It is stated that the nodes which act as a bridge in a network have more importance and influence over other nodes in the network. A node can be classified as a bridge node if it is located in a place that connects one part of the graph to another or it connects blocks of nodes together. So Betweenness Centrality tries to detect these bridge nodes to determine which nodes have influence on a social network. For example, in Figure 1, Heather connects two different user blocks to each other, despite the relatively small number of direct connections to other users. Without Heather, there will be no communication between these blocks. Therefore, people with a high betweenness centrality value are critical people and are the nodes (orgnet) that need to be examined regarding what information is distributed in the network.

$$Betweenness\ Centrality = g(v) = \sum_{0}^{n} \frac{\sigma_{st}(v)}{\sigma_{st}} \qquad (2)$$

σ_{st} is the total number of shortest paths from node s to node t so $\sigma_{st}(v)$ is the number of the paths that passes through the node v

Closeness Centrality: In Figure 1, Fernando and Garth have fewer connections with other users than Diane. However, their direct and indirect connections allow them to reach other users in the social network more quickly (orgnet).

The degree centrality, betweenness centrality, and closeness centrality values are listed in Table 1.

$$\text{Closeness Centrality} = C(x) = \frac{1}{\Sigma_y \ d(y,x)} \qquad (3)$$

d(y,x) is the distance between vertices y and x.

Eigenvector centrality: The logic behind eigenvector centrality states that a node gets a higher centrality value as

it is connected to larger important (central) nodes [27,30]. The centrality of a specific node is proportional to the centrality of nodes to which it is connected in a graph.

$$\begin{aligned} Eigenvector\ Centrality = x_v &= \frac{1}{\lambda}\sum_{t\in M(v)}^{n} \\ x_t &= \frac{1}{\lambda}\sum_{t\in G}^{n} \quad a_{v,t}x_t \end{aligned} \tag{4}$$

For a given class G: = (V,E) with vertices |V|, where $a_{v,t}$ is the adjacency matrix, x is the relative centrality. M(v) is the set of neighbors of v and λ is a constant.

Gökçe et al. (2014) attempted to identify Turkish opinion leaders in the domain of politics [31]. In the study, degree centrality, eigenvector centrality, and betweenness centrality measurements are used to identify political opinion leaders. They used degree centrality to identify the size of the audience for a user and eigenvector centrality to analyze each user in relation to the entire network. Betweenness centrality has been applied for indicating the importance of a user for the flow of communication [31]. First of all, they created an initial sample of 6,000 users to start creating the dataset. By adopting a holistic approach that can be used to create online populations, they collected a theoretically complete population of Turkish Twitter users instead of using a predetermined Twitter hashtags approach. Due to Twitter's API limits, the team had to develop an application that granted the utilization of many users' rate limit [31]. Finally, the remaining 10 million users were mapped as a network graph that shows the relation between them and it gave over 451 million connections within Turkish Twitter users. They obtained the results by analyzing the nodes on the graph. One of the most important disadvantages of this research is that it is quite difficult to determine whether the ghost opinion shapers, which are not found in traditional media channels like the printed media, are real people or made-up accounts [31]. They listed the top 100 Twitter users who are political opinion leaders using the three main measurements of centrality. Centrality formulas tend to correlate positively and these users in the list are high in terms of all three central measures [31].

Table 1. Centrality Values of Social Networking between Diane and Friends (orgnet) [29]

	Degree Centrality	Betweenness Centrality	Closeness Centrality
Diane	0.667	0.102	0.600
Fernando	0.556	0.231	0.643
Garth	0.556	0.231	0.643
Andre	0.444	0.023	0.529
Beverly	0.444	0.023	0.529
Carol	0.333	0.000	0.500
Ed	0.333	0.000	0.500
Heather	0.333	0.389	0.600
Ike	0.222	0.222	0.429
Jane	0.111	0.000	0.310

The following two studies feature social network analysis (SNA) methods. Meltzer et al. (2010) investigated how SNA can be used to design effective clinical quality improvement teams [32]. The relationships in a team are "social capital" and should be optimized along with quantity and types of humans. It is stated in the paper that the SNA term cluster corresponds to pockets of homogeneous individuals within the organization and within these pockets SNA metrics like degree, betweenness, and density can be applied [32]. The authors design an empirical methodology based on the following definitions: net degree, betweenness measure, and network density. Net degree is applied as the total number of unique people that the team can reach by using one or more team members; therefore, it depends on the nature of the team's task. Betweenness is calculated as the number of times an actor lies on the shortest path between other actors. High betweenness can indicate one's ability to coordinate projects in a team; hence it can be a good metric to identify the potential leaders in emergency situations [32]. Furthermore, network density is described in the paper as a fraction of potential connections in a network that are actual connections. They state that if movement of information is important for success then high density would be ideal and, like betweenness, it depends on the team's purpose. According to their paper, the teams with high density might have a lower net degree. If information must be carried over a network, betweenness might be useful. For instance, some physicians have a below-average degree while having above-average betweenness. This makes them good candidates for carrying information. They also mentioned that diversity and betweenness are both dependent on the purpose of the team [32]. Due to the lack of sufficient outcome data, they were unable to analyze the team's effectiveness which might be influenced by team structure and quality improvement context. In addition to these limitations, they also lack data on the quality of interactions [32].

In the study by Jain and Katarya (2019), a modified firefly algorithm was suggested to detect global opinion leaders in a social network and local opinion leaders in communities [33]. The Louvain method is applied to discover communities in the social network. The Louvain method uses a greedy algorithm to detect communities in large networks. In the Louvain algorithm, the greedy approach uses hierarchical clustering that consistently removes edges with higher betweenness centrality [33]. The experiments are performed using two different datasets. The first one is a synthesized dataset consisting of 20 nodes and 70 edges, and the second is called 'small slashdot', consisting of 13,182 nodes and 34,621 edges. According to the experimental

results reported in [33], the proposed algorithm outperforms standard SNA methods. Although better results are obtained with a static network compared to dynamic networks, both global and local opinion leaders can still be found accurately even with increasing network sizes.

The following study identifies opinion leaders using a relatively large forum dataset. In the work by Li et al. (2019), opinion community detection technique and opinion leader detection technique suggested to rank influencers in social networks [34]. Opinion community is a community formed by various people that support the same opinion or idea. Opinion leaders are individuals who have the ability to set or shape people's opinions in a social network. Finding influencers in social networks makes it possible to detect these opinion communities and opinion leaders. These opinion community detection and opinion leader detection techniques used in this study include emotional analysis model, user influence model, time similarity, content similarity, and topology structure of users to efficiently detect opinion leaders in a social network. The emotional analysis model is applied to detect meaning of users posts, the user influence model used is built based on content similarity and network topology structure. Time similarity is calculated based on users posting times. Using the methods mentioned above are used to detect opinion communities. For the evaluation of the system a dataset was constructed by collecting 11713 posts and 8976 topics from a world forum. In order to see the effects of the suggested techniques a single-pass (SP) algorithm, the K-means (KM) algorithm, online-time-based opinion leader discovery (OTOLD) algorithm, experience-based opinion leader discovery (EOLD) algorithm, and PageRank algorithms are used as benchmark algorithms in their experimental environment. Based on the experiment results reported in the experiments, it is determined that the study can efficiently detect opinion communities in social networks [34].

Statistical and Stochastic Approaches:

This section includes studies that use statistical and stochastic approaches to identify opinion leaders in social networks. Statistical and stochastic approaches consist of mathematical calculations to determine features of the social networks. After the feature extraction opinion leaders are determined.

The following study attempted to identify opinion leaders using a blog dataset. [35] attempted to provide a framework named BARR for the identification of opinion leaders in blogs. By using the framework, identification of opinion leaders gives marketers the opportunity to market their products or services and allows companies to determine their strategies according to positive or negative shares of bloggers. They used topic detection and tracking analysis to understand messages and identify hot topics [35]. The workflow of their framework consists of 5 steps. The first step involves a blog search using user-defined keywords. In the second step, the framework analyzes web pages and extracts ontologies. To determine which word is to be saved to build the domain ontology, it calculates the entropy value and records the ontology that exceeds the predefined threshold. For the entropy formula which is referenced as formula (5), the frequency of the words in the documents (frequency (word)), the total number of words (NumOfWords) and the total number of documents (NumOfPages) are used [35].

$$Entropy(word) = \frac{\frac{frequency(word)}{NumOf\ Words}.In\left(\frac{frequency(Word)}{NumOf\ Words}\right)}{In\left(NumOf\ Pages\right)} \quad (5)$$

In the third step, the ontology of a word is found, blog and blogger instances are constructed, and it identifies relationships between blogger instances. In the fourth step, hot blogs are detected by using centrality and prestige. For these two criteria, 4 information sources are used. The popularity of a blog is determined according to the number of visits, number of reviews, and rank parameters that are related to blog content. The parameters expertise, number of blogs, and number of comments are similar for readers and author resources. For expertise, vector spaces are created with frequency values calculated in the second step. The cosine between vectors gives the author's degree of expertise. The value can be between 0 and 1 [35]. For the relationship between the author and readers source, homophily and tie strength are used. When calculating homophily the formula used for expertise is used. Vector space is created by word entropies instead of word frequency. In the final step, since a blogger with few blog entries cannot be considered influential, influence is formulated considering quantity and quality values and it is determined whether the blogger is an opinion leader or not [35].

The following study uses a hybrid model approach to identify opinion leaders in a network. This hybrid model is the combination of diffusion process approaches and statistical and stochastic approaches. Cho et al. (2012) investigated which opinion leader is best for marketing purposes of a product considering the diffusion speed and the maximum cumulative number of adopters in social networks [26]. Opinion leaders carry an important role for marketing of products. Selecting the best opinion leader for a specific product requires analysis of the opinion leaders which has effects on the marketing of the product. To be able to achieve that first they base their research on social network theory and redefine opinion leaders accordingly. Then it is examined how the opinion leader's effect changes based on a network with different types and characteristics. Lastly, they examine how the percentage of initial opinion leaders affects product diffusion [26]. The network model used is a stochastic cellular automata (SCA) model. A network comprises N entities. Entities with similar attributes are

grouped together. An entity can have at most 6 neighbors at a distance of 1.

The strength of a tie between two entities is defined as "intimacy". Intimacy can take values from 1 to 5 with a larger value meaning deeper intimacy or a stronger tie. They defined the sociality of each entity as the total sum of the intimacy that an entity has. If an entity has a degree of connections, it has a high probability to be at the center of the network. Measuring how far each entity is located from the center is the problem and the authors tackle this problem by considering five different concepts of centrality: send-nomination, receive-nomination, sociality, distance, and rank-nomination centralities [26]. Time is chosen to be the continuity of "period". Initial adopters learn about the product in period 0. In period 1 these entities either propagate the product or not. The probability of propagation is calculated with formula (6).

$$P_{propogate} = \frac{intimacy}{sociality} \tag{6}$$

Entities that are aware of products at a period of 1 either decide to adopt the product or not according to the predetermined threshold.

Formula (7) states the probability of satisfying the threshold condition that rises as more and more neighbors adopt the product.

$$P_{adaptusers}(i) = \begin{cases} 0, when\ threshold > \dfrac{number\ of\ adoptusers\ among\ neighbors}{number\ of\ neighbors} \\ 1, when\ threshold \leq \dfrac{number\ of\ adoptusers\ among\ neighbors}{number\ of\ neighbors} \end{cases} \tag{7}$$

Their network used in the simulation contains 10000 entities. They repeated this simulation 100 times. In the first scenario, they run the simulation with initial adopters with different centralities. According to the results, distance centrality is the best, followed by rank-nomination centrality [26]. In the second scenario, they changed the network attributes and ran the simulation as a scenario one. It is observed that when the threshold is small (0.3) the opinion leader with the longest nomination will reach out to more entities. When the threshold is increased rank-nomination and nomination-send centralities outperform others. Next, they run the simulation with different average sociality and sociality levels to observe changes in the final cumulative number of adopters (FCNA). When they ran the simulation with a higher average sociality [26] FCNA increased; yet when they increased the sociality FCNA decreased. Distance and send-nomination perform better when average sociality falls below the longest-nomination. Though when they increased sociality the final cumulative number of adopters decreased. Average nomination length affects innovation diffusion in the following ways: distance centrality outperforms the rest when the average length is 2, 3, or 5. When the average length is 7 send-nomination centrality performs better [26]. In the third scenario, they only change the percentage of initial adopters. Next, they increase the initial adopters to 150 and distance centrality starts to outperform the others [26]. In conclusion, the effectiveness of opinion leaders varies with some metrics such as distance centrality; so, according to the authors, an opinion leader with a high centrality rate is considered more effective when compared with one with a low centrality rate.

The following two studies detect opinion leaders in a Twitter network using statistical and stochastic approaches. The authors of [36] proposed a methodology for community detection and associated role categorization in retweet networks and followers independently. Markov stability is used to detect the communities in the network [36]. The goal in the network is to partition the network graph into meaningful subgroups [36]. These subgroups can be classified as communities [36]. The authors applied a multiscale flow-based community detection approach with Markov stability to the network collected from Twitter users. According to [37], an influencer might be detected in a social network by the estimated number of retweets. The theoretical framework SEISMIC [37] performs experiments for this assumption on Twitter. The authors report in [36] that by the number of retweets it is possible to find the influencers in social networks.

In the study conducted by Alp et al. (2019) a new method for identifying influencers called influencer factorization was proposed [38]. For this purpose, 20 Twitter users were determined manually based on topics, and their followers and friends were gathered with a breadth-first search until a sufficient number of users was obtained. Then, those users with protected accounts were deleted and 20 tweets were collected from the remaining users. For topic modeling, the Latent Dirichlet Allocation (LDA) from the Machine Learning for Language Toolkit (Mallet) tool was used. Since tweets are short texts, pooling, a method applied by combining the tweets that each user sends each day, was applied. After creating word clusters with LDA, the topic labels were determined by selecting the clusters in which the words are related to each other and the tweets were labeled. If users have a topic that exceeds the threshold in their tweets, it is added to this user as the user's topic label. In this way, a global user network and user networks based on the topics were created. For user modeling, user-specific features were used. With focus rate, how focused the user has been is measured in any topic. It is assumed that the people who are considered influencers focus on one topic and do not post tweets much about other topics.

For each topic, the user's focus rate fr_u^t is calculated by dividing the number of tweets posted for the topic p_u^t by the total number of tweets (pu).

$$fr_u^t = \frac{|p_u^t|}{|p_u|} \tag{8}$$

Activeness measures how often a user tweeted about a topic. Attempts were made to determine whether a user who tweeted about a topic continuously was influential in that topic. Activeness ac_u^t is calculated by dividing the number of days the user is active on that topic d_u^t by the total number of active days (d) [38]:

$$ac_u^t = \frac{|d_u^t|}{|d|} \tag{9}$$

Authenticity is used to measure how much a user speaks about a topic. They thought that influencers prefer to share their own thoughts rather than conveying others' thoughts. User's retweets rt_u^t about a topic are subtracted from all tweets about that topic p_u^t, and then the result is divided by all tweets about that topic p_u^t [38]:

$$au_u^t = \frac{|p_u^t| - |rt_u^t|}{|p_u^t|} \tag{10}$$

In addition, hybrid features were calculated by combining these features with coefficients. Using the existing sparse matrices in their calculations, they generated their own influencer factorization method from the matrix factorization method, which also makes predictions for unseen data. This enabled them to identify potential influencers as well as identify existing influencers. Using users' retweeting rates of each other, the matrix was created to be used in user–user influence factorization and, using user-specific features, matrices were created for each topic to be used in user–topic influence factorization. The alternating least squares (ALS) algorithm was used for factorization. PageRank, TwitterRank, and Personalized-PageRank were used as baseline methods to compare the results. As a result, the authenticity feature was seen to have a positive effect when used with other features [38].

The following study conducted an opinion leader identification study on a survey dataset using various types of analysis methods. According to the study conducted by Tam (2020), states that social media which is widely used today, significantly changes society, institutions, and individuals [39]. The study stated that the question of how the behaviors of people who are described as social media influencers are perceived by the target audiences of these influencers and how they are seen as opinion leaders and affected by them should be studied [39]. The author obtained the datasets by presenting a Turkish online survey to the users. While 634 people participated in the survey, 63 were not included in the data because it was determined that the survey was filled out randomly [39]. A total of 571 questionnaires were analyzed, with data not included [39]. Regarding the scales in the research, after the factor loads and numbers were tested with exploratory factor analysis (EFA) and the accuracy was tested with the confirmatory factor analysis (CFA), the analysis was carried out through the Analysis of Moment Structures (AMOS) program [39]. The author analyzed the results of the survey data collected from 571 participants; "Bilgi" ("Information"), "Taklit" ("Imitation"), and "İletişim" ("Communication") variables had an effect on the participants, while "Yakınlık" ("Intimacy"), "Güven" ("Trust"), and "Eğlence" ("Entrainment") variables did not have any effect [39]. Referring to this situation, it has been observed that social media users obtain information from their communication with influencers and imitate them in the light of this information, and see them as opinion leaders. It has been determined that the influencers mentioned throughout the research interact with users mostly on Instagram, followed by YouTube and Twitter [39]. On the other hand, it has been concluded that Facebook, besides these platforms, although it was quite popular at one time, had little effect [39]. According to the author, more detailed research on virtual opinion leadership can be done by utilizing people who were chosen as opinion leaders in the future [39]. There are some advantages and disadvantages of this study. An advantage is that the author investigated the factors in the perception of influencers on social media as opinion leaders and divided them into 6 hypotheses and by using the structural equation model (SEM) the effect of the independent variables and the dependent variables was revealed. A disadvantage was that it was mentioned that influencers who are trusted and loved are not seen as opinion leaders, but the reason for this was not investigated.

PageRank Based Approaches:

This section includes the studies that use PageRank based approaches to identify opinion leaders in social networks. PageRank algorithm is a ranking algorithm which weights each element and ranks it accordingly to the maximum to minimum. This section includes reviews of studies that use similar approaches to PageRank or use PageRank as a baseline algorithm. To identify opinion leaders, first the users in the network are weighted and ranked according to some measures and the determined weight then leaders are selected.

The following study uses a distinct initial comprehensive influence model to determine opinion leaders. Luo et al. (2018), presented an improved weighted LeaderRank algorithm to identify opinion leaders [40]. When the LeaderRank algorithm is compared with other algorithms such as degree centrality, k-shell decomposition, and PageRank, etc., it gives better accuracy. Every user is considered a node in the weighted LeaderRank algorithm. The algorithm has limitations. For example, a link weight is calculated by the in-degree of each node [40]. In the improved weighted LeaderRank algorithm, weight is not

calculated just using replies; it uses posting, reading, being praised, etc. According to [40], every user has a different initial influence. Furthermore, positive influence represents how active a user is, while indirect influence represents how much a user is taken into consideration. Their initial comprehensive influence model is shown in Figure 2.

They used the following formula to calculate the users' initial comprehensive influence:

$$f = \sum_{i=1}^{2} w_i * a_i + \sum_{i=1}^{4} v_i * b_i \qquad (11)$$

Here f is the initial comprehensive influence, w_i and v_i are weight, a_1,a_2 are the active influence, and b_1,b_2,b_3,b_4 are indirect influences [40]. They used the entropy weight method, which was proposed by Shannon in 1948, to figure out the optimal weight wi and vi. Every user has a different relationship with another user. For example, if user A replies to user B more than to user C then user A has a stronger relationship with B than with C. Here the replies are represented by weight [40]. They used Massive Open Online Courses' forums, on which people can interact with each other to discuss the courses, in order to collect their dataset. They collected 1215 records and 302 users, and they found that only 22.5% of the users are really active, which means they are both posting and replying [40]. They extracted the mutual replying relationships as a metric for the users in which they are in the dataset. If a user replies to another user there is a link between them. They obtained the number of replies. The second metric is the number of times for each learner posting, reading, being read, being replied to, being praised, and being concerned. Then they calculated the appropriate weight for all elements (posting, replying, being read, being replied to, etc.) according to their metrics and the entropy weight method. After that they calculated the influence value for all users and compared them with PageRank and LeaderRank influence values. As a result, it is seen that their algorithm is better than the others [40].

The following two studies use the LDA algorithm to detect opinion leaders on a social network dataset. In a study by Song et al. (2007), a novel algorithm called InfluenceRank whose aim is to identify opinion leaders in the blogosphere is presented [41]. They rank blogs based on how important they are in the blog network and how new the information they carry is [41]. According to the researchers, when a blog generates a post its source is either other blog posts, mass media, or original ideas. In that paper, they model this source of information as a hidden node. With this node, the blog can create new information by itself. Next, they calculate the information novelty of one blog. They define a document as an entry in the blog. Then, they apply the LDA to reduce data dimensionality and to create a topic space for representing entries. Next, they generate a feature vector by projecting each entry onto the topic space. They use cosine similarity in order to compute dissimilarity since it outperforms Kullback Leibler [41]. They collected the dataset between July 2005 and October 2006 by using an NEC focused blog crawler to evaluate their algorithm. Next, they preprocessed the crawled data by removing stop words and removing entries with less than ten terms. As a result, they obtained 407 English language blogs with 67,549 entries and with 11,187 key terms. They used four algorithms as the baseline: PageRank (PR), random sampling (RS), time-based ranking (Time), and information novelty-based ranking (IN) [41]. According to their experimental results reported in the paper, the InfluenceRank algorithm outperforms all other baseline algorithms in terms of performance [41]. In the research by Alp et al. (2018), a methodology called Personalized PageRank was proposed to identify topical influencers based on Google's PageRank [9]. They combined user-specific features with network topology. These user features are focus rate, activeness, authenticity, and

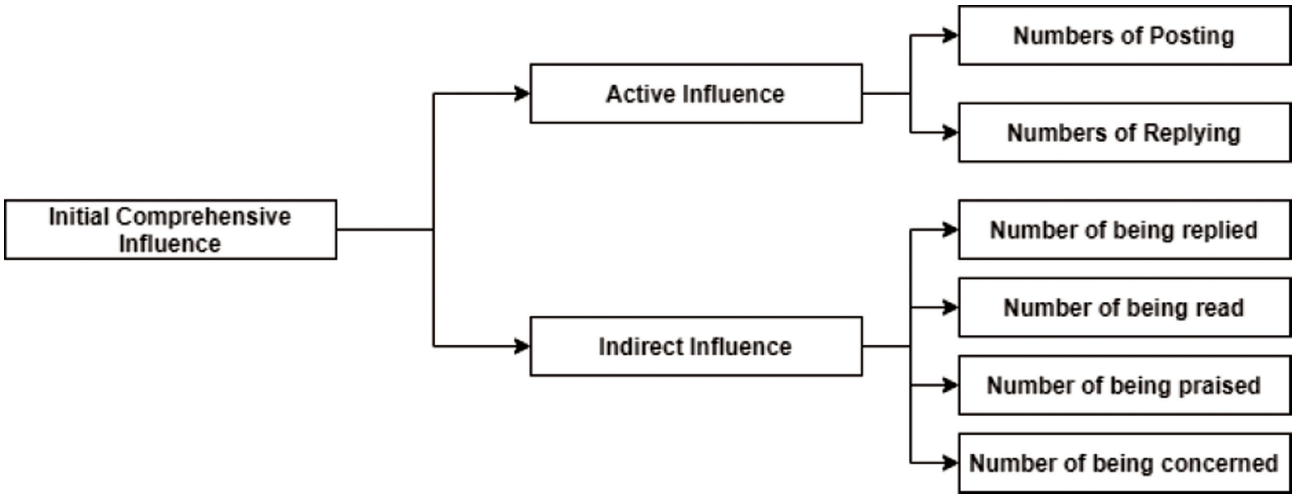

Figure 2. Users' initial comprehensive influence model [40].

speed of getting reaction. They calculate 3 different activeness measurements for each user on each topic:

- The number of days that the user posted on a specific topic.
- The average number of tweets for each day on a specific topic
- The average number of tweets on a specific topic multiplied by the active days.

Speed of getting a reaction: This feature measures the average time between the users posts a tweet and gets its first retweet.

These features are integrated into Google's PageRank formula by replacing the dumping factor "d". w_u^t indicates either one of the user-specific features mentioned previously [9].

Google's PageRank:

$PR^{(0)}(v) = 1$,

$$PR^{(i+1)}(v) = (1-d) + d\sum_{u \in Fw} \frac{PR^{(i)}(u)}{Nu} \quad (12)$$

Personalized PageRank:

$PPR^{(0)}(v) = 1$,

$$PPR^{(i+1)}(u) = \left(1 - w_u^t\right) + w_u^t \sum_{v \in F_u} \frac{PPR^{(i)}(v)}{N_v} \quad (13)$$

The algorithms are implemented in a distributed manner using Apache Spark. Space complexity is linear to user amount.

For data collection, they picked 20 Twitter users who focus on different topics such as politics, sports, TV, and religion. They collected 20 tweets from each user and eliminated the users who do not post in Turkish 80% of the time. By using Twitter Streaming API, they collected tweets from these users between November 4th, 2015 and January 12th, 2016 [9].

For preprocessing they performed stemming and stopwords, punctuation, and mentions removal. Next, the LDA enhanced with pooling was performed on the dataset. Afterward, 3 human experts examined the output of the LDA, which is word clusters, and named each cluster with appropriate topic titles. For each user, they calculated their topical tweet rate and labeled the user with the topic title highest rate. Since experimenting on the global network is computationally expensive, they divided their network into sub-networks. Moreover, a user is not added to any of the networks if their tweet rate is not higher than a certain threshold for any of the sub-networks.

To evaluate their algorithm, they use measurement based on the retweet rate of a user on the test data. For each sub-network, the output of their algorithm is the top 25 users. To estimate information diffusion of the 25 users they calculate normalized tweet rates for each user and sum them. Spread score is calculated as follows [9]:

$$spread(t) = \sum_{u \in \{in f_t\}} \frac{\left|p_{u,tr}^t\right|}{\left|p_u^t\right|} \sum_{p \in p_{u,tr}^t} \left|retweets_p\right| \quad (14)$$

In the second evaluation method, they conducted a human survey. It performed as showing each volunteer their algorithms' top 25 pick user's tweets and asking them if the user is a 'high influencer', 'influencer' or 'not an influencer'. In addition to Personalized PageRank (PRR) with 6 different user features they also run experiments with Google's PageRank, Twitter Rank, Random Influencer Selection, and Most Followed User. All algorithms are run for each topic. Each algorithm outputs the highest performing 25 users. For each of these users, the potential spread score is calculated. Experiment results show that at least one PRR algorithm outperforms the baseline algorithms in each case. PRR performs worse than the baseline algorithms when it runs with the speed of getting reaction user features [9].

The following study proposes a similar algorithm baselining PageRank algorithm by utilizing the K-means algorithm. In a study conducted by Zhang et al. (2020), a rank after clustering (RaCRank) algorithm is proposed to detect opinion leaders in social networks [42]. The algorithm consists of two phases. In the first phase, a modified version of K-means is utilized with the following features: in degree, betweenness, and center. Then they proposed a two-hop clustering coefficient. In the second phase of the algorithm, users' leadership scores are calculated based on user activeness, user influence, and center. Experiments are conducted by using a social network with 49,613 users, and 59,957 edges among these users. The suggested method is compared with AllUserRank, ClusterRank, and UI-LR. Although the RaCRank algorithm performs slightly worse than UI_LR, it outperforms AllClusterRank. According to [41,42], future studies can focus on detection of topic-based opinion leaders.

Machine Learning Approaches:

This section includes review of studies that use machine learning approaches to identify opinion leaders in social networks. Supervised or unsupervised machine learning methods can be used in the identification process. Smart learning systems can be realized using machine learning methods which receive high accuracy.

The following four studies identify opinion leaders using machine learning techniques on a Twitter dataset. In [43] a deep analysis was performed of how Twitter plays a role in breaking news. In order to examine this, they chose the time period during which Osama bin Laden was killed. Because the news was spread on Twitter before the mass media, people were divided into two opposing poles about whether Twitter was reshaping journalism. Throughout

the study, it was discovered that people prefer Twitter and a small group of people called opinion leaders who share information with people instead of learning from the news source. The study was expanded when it was discovered that news was spread thanks to a group of people who play key roles as well as opinion leaders [43]. Due to the size of the data to be collected, a sampling method was applied. In this method, within a 2-hour period from the moment the news broke, 30% of shared tweets are collected for 40 seconds every two minutes. While tweets were collected, they accepted non-English tweets that did not include the determined keyword but were related to the subject, and also assumed that the tweets containing the keyword but were irrelevant to the subject made up a small percentage. Once the data were collected, they added "certain", "uncertain", and "irrelevant" labels to the tweets and decided to use 235 tweets: 54.9% of these tweets are certain, 42.1% are uncertain, and 3% are irrelevant. They trained 2 classifiers for determining whether the news shared on Twitter before the mass media was a rumor or not. While the first classifier determines the tweet's relevance, the second classifier decides whether the tweet is certain or uncertain. They used the SVM classifier with the bag-of-words representation technique. They obtained the result of 75.8% overall confidence [42,43]. In addition, to find out who produced the biggest reaction as soon as the news began to spread, they collected the most mentioned 100 users' information within a 2-hour period. As a result of this review, they saw that most of the information consumed on Twitter was produced by a small group called opinion leaders. When opinion leaders were grouped manually, it was seen that they affected Twitter users in different ways. Furthermore, when the links shared in tweets are examined, people still trust the content produced by the mass media more than other sources [43].

In the study by Safalı (2020), the opinions of the users about the People's Alliance were classified [44]. The dataset used in this study consists of 4000 users and a total of 800,000 data shared by these users [44]. So, 20 data were selected from each of the 4000 users and this way 800,000 sized dataset was formed. Natural language processing methods were used to clean up the distortions [44]. Feature extraction was done using the term frequency method [44]. The extracted terms list was converted to numeric values [44]. The algorithms applied to the training set were K nearest neighbor (KNN), decision tree, ordinal minimum optimization, and Bayes [44], the KNN algorithm was determined as the most successful [44]. Different data collection, analysis, and classification studies were examined and compared. A comprehensive literature analysis was performed. In the data preprocessing stage of the study, the Zemberek Library was used to correct spelling mistakes and stem the words. Then the feature extraction process was performed. After the extraction process 80% of the data were selected as the training data and 20% as the test data and the data were labelled according to the extracted features [44]. Kappa statistics were used while calculating the model accuracy rates of the 4 different algorithms used. In the continuation of the article, the classification process is explained. The Weka Library was used while calculating the kappa values in the study. As a result the algorithm achieved the highest accuracy of 97%. Looking at the results, the most successful model according to the kappa statistics is the KNN algorithm. In the study the authors also explained how to choose the best model [44]. Thus, in conclusion, this article is concluded by mentioning neutrality and neutral sharing. The biggest advantage of the study is that the dataset is very large, but while using this advantage, bots that share on social media should be considered and not included in the study.

Another study performed on a formed Twitter dataset like the studies [43,44]. In the study by Gümüşsu and Murat (2019), the authors investigated the relationship between the People's and Nation Alliances based on the tweets that had "tamam" or "devam" tags [45]. The study used only Turkish tweets that were posted on the June 24th election day on the words of President Recep Tayyip Erdoğan, "Milletimiz tamam derse o zaman kenara çekiliriz..." ("If our nation says done then we will step aside...")[45]. There were 4886 "tamam" ("done") tagged and 7430 "devam" ("continue") tagged tweets making up the dataset [45]. The aforementioned data were obtained with the KNIME program via the Twitter API and then the collected data were saved in Excel [45]. In the text preprocessing phase of the data, the open source "zemberek" library written in the Java programming language was used [45]. The "Text2ArffV5" software "Kemik Doğal Dil işleme" was used in root finding and word weighting processes [45]. To determine which Alliance was the opinion leader, word clusters were formed [45]. Two types of word clusters were created, in the first cluster the "devam" tagged tweets were collected and from this word cluster supporters of the People's Alliance could be found [45]. In the second cluster the "tamam" tagged tweets were collected and from this word cluster supporters of the National Alliance could be found [45]. This study also examines relationships between the People's and National Alliances, so to measure whether there was a significant relationship between the Nation and the People's Alliances correlation analysis was performed [45]. It is shown that users who tweeted "devam" tagged tweets support the People's Alliance and those who tweeted "tamam" support the National Alliance, and, as a result, a weak relationship was found between these two groups [45]. In addition, from the results obtained from this study some predictions can be made; for example, with the data obtained from the tweets sent, the election with the highest participation can be predicted, the effect of the arguments selected for election propaganda on the voters can be researched, and arguments can be developed in this way, or the arguments used in the election propaganda can be presented to the public

by maliciously distorting them and manipulating the elections through this process. Therefore, it can be ensured that the countries/peoples/elections are unpredictable.

In the study by Aleahmad et al. (2016), a method was presented for finding users with a high impact on users in a particular domain of social media [46]. To do this, first of all, important topics in a domain are extracted. The competency score is then calculated according to these topics and the popularity score is calculated based on the number of users' in-links [46]. First, they use the LDA to extract the main topics discussed in some domains (e.g., automotive and banking). Secondly, the competency of each user for this subject is calculated. For this purpose, firstly, the most relevant posts are found by using an information retrieval system. Then, it is analyzed how much each user shares these relevant posts for each topic, and how important a user is to a topic. Each topic has different degrees of importance for a domain and they attempt to give suitable weights to the topics. After that, the in-degree centrality metric was calculated to measure the impact of users on the corresponding social media [46]. A popularity score is calculated:

$$Popularity\ (a) = 1 - e^{-\lambda F(a)}, \tag{15}$$

where F(a) is the in-degree centrality of the user a and λ is a constant factor that is used for tuning the popularity score of the user [46]. Finally, the opinion leadership score of a user is computed as below:

$$Leadership(a,d) = \beta \times Compe(a,d) + (1-\beta) \times Popularity(a) \tag{16}$$

They used RepLab 2015 as their dataset. This dataset contains 4.5 million tweets and 7491 users. While 1185 of them are automotive, 1315 of them are banking related profiles [46]. According to the experimental results reported in the paper; OLFinder algorithms work better than the baseline algorithm in the literature. In addition, topic extraction was performed with TF-IDF instead of LDA and the results were compared and LDA was found to be better [46].

The following study used a forum dataset to propose an opinion leader detection method. In the work by Chen (2019), a cluster-based opinion leader detection method was suggested [47]. The suggested method first initializes a social network by considering the post-reply relationship of the Mobile01 forum posts that have Chinese content. Then the authors in [47] detect the significant communities in the network with the parameter-free method they developed. "Kmeans" is used on the significant communities to create clusters and each cluster is given a score. They choose final opinion leaders from each high performing cluster. Experiments are performed by using forum discussions from Mobile01 related to 4 different car brands. The results are evaluated against a leadership quality clustering algorithm (att_clustering) [48] by using information spread. The results show that the proposed algorithm outperforms att_clustering on each dataset [47].

The following study proposes a different way to find opinion leaders when compared with the studies in this category. This study adapts a nature inspired algorithm to detect opinion leaders. In a recent work conducted by Jain et al. (2020), a novel approach was proposed for community detection and a social network-based nature-inspired whale optimization algorithm [49]. The whale optimization algorithm states that each user in a network behaves like a whale. The whale aka user having more reputation is considered as prey for other users. All the other whales want to catch the prey due to the prey's reputation. This simulation is inspired by nature and it means the preys are the popular users on a network and other users try to connect with these users on the network which forms the communities on the network. This way the communities can be detected. Global and local top-N opinion leaders are detected. The community partitioning algorithm is used to discover communities on the social network. The experiments are performed using two different datasets. The first one is a synthesized dataset consisting of 100 nodes and 467 edges and the second is called a 'wiki-vote dataset', consisting of 7115 nodes and 103,689 edges. As the number of users on the network increases, the performance of the algorithm increases.

From the reviewed studies, studies about opinion leader identification were collected. All the collected studies listed in Table 2 based on authors of the publication, year of publication, the used methodology, the used dataset, the approach and accuracy.

DETECTING SPAMMERS ON SOCIAL NETWORK

In order to detect spam accounts in social media, a number of methods have been proposed. These methods can be grouped into three categories, namely 1.) Methodologies using user-based and content-based features, 2.) Methodologies using honeypot features, 3.) Methodologies using sentiment information and graph-based methodologies.

Methodologies Using User-Based and Content-Based Features:

This section includes reviews of spam detection studies which use user-based and content-based features. Studies in this section detect spammers in social networks by analyzing users' properties and content that is shared by these users in a social network using various algorithms.

The following three studies used mainly a Twitter dataset to detect spammers using user-based and content-based features. In [50] the authors created their own dataset by collecting from selected trending topics on Twitter.

Table 2. Comparison of different opinion leader identification studies.

Authors	Year	Methodology	Dataset	Performance Metric	Results
M. Hu et al.[43]	2012	Support Vector Machine	Twitter	Accuracy	75.8%
Aleahmad et al.[46]	2016	Latent Dirichlet allocation	Twitter	Precision -Recall,	95.2%
Safalı [44]	2020	KNN, Decision Tree, Ordinal Minimum Optimization, Naive Bayes	Twitter	Accuracy	97%
Cui and Pi [28]	2017	Support Vector Machine, Naive Bayes	Twitter	Precision -Recall,	90%
Jain and Katarya [33]	2019	Louvain Method	Microblogging sites	Precision, F1-score	94%
Zhao et al.[37]	2015	Statistical methods	Twitter	Accuracy	60%
Jain et al.[49]	2020	Whale Optimization Algorithm	Survey	Accuracy	83%
Duan et al. [48]	2014	Fuzzy-based Clustering-means, EM-based clustering	Forum	Average accuracy	70%

Then they manually labeled data as spammers or non-spammers. They used only tweets in English in the dataset. According to the dataset, spammers are categorized according to two main attributes, which are behavioral and content attributes. Content attributes are properties of the posted tweets such as the number of hashtags per number of words on each tweet, the number of URLs per word, etc. They noticed three of them are more efficient to distinguish between spammers and non-spammers. An SVM (LibSVM) supervised learning algorithm was used with the Radial Basis Function (RBF) kernel in order to detect spammers. According to their experimental results, LibSVM and SVMlight (Support Vector Machine Light) gave the same results. The dataset used underwent 5-fold cross-validation. As a result, they correctly classified 70% as spammers and 96% as non-spammer. The 30% misclassified as non-spammer is because of the dual behavior of spammers. This kind of spammers post non-spam tweets to act like non-spammers, so they cannot be easily detected [50]. The importance of the attributes was measured on Weka using two feature selection methods, which are information gain and chi-squared. Content attributes and user behavior attributes are homogeneously distributed through the top positions. It provides good results even if content attributes become ineffective against the spammers [50]. Spammers can develop new techniques that can be seen as non-spammer, so with these new techniques spam detection systems can miss some spam content which means some of the attributes spammers use can become useless for spam detection systems since it cannot be detected [50]. Their results indicate that even with a different mix of their attributes the classification approach can give high accuracy. Another study that use a Twitter dataset like the previous study is conducted by McCord and Chuah (2011), the following features are used to detect spam [17]: user-based features, i.e., followers, following, and user behaviors such as time periods and tweet frequency of a user; and content-based features, i.e., number of URLs, replies/mentions, retweets/tweet length, and hashtags. The dataset was built from follower/following information of Twitter users and it includes their most recent 100 tweets in English. They [17] also used a reputation formula to represent one of Twitter's spam and abuse policies:

$$R(j) = \frac{n_i(j)}{n_i(j) + n_o(j)}, \tag{17}$$

where $n_i(j)$ and $n_o(j)$ represent the number of followers and the number of followings user j has, respectively. According to their experiments, reputation is not a good metric to detect spammers but number of following/followers can be useful to detect them. Spammers make more mentions than normal users. They also observed that spammers are more active in the early hours of the day [17]. They use several traditional machine learning classification algorithms with these user-based and content-based features. Their experimental results show that the random forest (RF) classifier is the best one, with 95.7% precision and 95.7% F-measure [17].

In the study by Ferrara et al. (2016) developed a framework to detect bots on Twitter [51]. The main goal was to separate human-like behavior from bot-like behavior while describing some features. They did it using 6 classes, namely user, network, friends, timing, content sentiment, and their 1500 features [51]. According to their methodology, to classify an account as a bot the model must be trained with all feature examples. It is very hard to find bot instances, so they used bots as described by Caverlee's team. Then they collected more than 200 recently posted tweets and more than 100 mentions; also they worked with data from Mobile01, which is a Chinese blog. As a result, some features such as user meta-data were very useful to separate the bot-like behaviors. According to the findings, bots retweet more than humans and have longer usernames but

they post fewer tweets than humans; also their accounts are usually newly created accounts [51]. They implemented a web-based application and it collects tweets and other data from an account to detect whether it is a bot or not in real time by calculating the probability of being a bot separately [51]. According to the discussion in [51], it is hard to recognize cyborgs, which are a mix of humans and bot. These kinds of accounts could be stolen or a human can give them their account [51].

Following four studies perform spam detection on emails using similar types of datasets. In the study by Şahin and Demirci (2020), the authors proposed a study to detect and filter spam emails using the KNN algorithm [52]. For this purpose, three different datasets were used: Enron-Spam, Ling-Spam and SMS-Spam [52]. The Enron-Spam dataset has 17,171 spam and 16,545 regular mails. The Ling-Spam dataset consists of 481 spam and 2412 regular mails [52]. The SMS-Spam dataset includes 4825 regular and 747 spam mails [52]. The datasets are splitted as 70% for training, 30% for testing of the system [52]. But first the datasets are preprocessed by being separated from stopwords and punctuations and stemmed to be processed for the ready for the KNN algorithm [52]. Then the words were weighted according to their values and more meaningful data were tried to be obtained [52]. The F-measure was used for the evaluation of the system [52]. As a result, the system was run with different k values in 3 datasets, the most successful results were always obtained at k=1 value and the value decreased as the k value increased [52]. The future goal of this study is to achieve better results by properly weighting all the data and improving the intermediate stages such as preprocessing [52].

Şeydanur and Soğukpınar (2020) conducted a study to detect malicious (phishing) emails using deep learning approaches [53]. They used the Jose Nazario phishing email dataset and the Enron email dataset which is the same dataset that is used by the study [52]. The final dataset had 4512 emails of which 2256 are phishing emails and the remaining 2256 emails are safe emails [53]. The emails in the dataset are in English. The authors divided the emails in the dataset into two parts, the header and the body [53]. Some features are extracted from the header data and converted into digital format with StandardScaler. After extracting the text content of the body part, it is converted into a vector with Word2vec and LSTM (long short-term memory) is given as input to the neural network. Finally, the outputs from the MLP and LSTM neural networks are transformed into a matrix and given as input to the MLP neural network, which will make the decision [53]. While 3609 emails from the dataset were used for training, the remaining 903 were used for testing [53]. The authors achieved an accuracy of 96.84% [53]. Some studies in the literature only deal with text processing, extending the time of the study. Some studies, on the other hand, only deal with the feature extraction method, reducing the accuracy of the study. This study, on the other hand, combined the two, resulting in a fast and high-accuracy study. As the authors plan for the future, more efficient parallel algorithms can be used to add more training data and reduce computation times to improve the performance of the project. An advantage is that it is a study with a high accuracy rate, and a disadvantage is that the training data are scarce.

In study [54], a spam detection model was proposed for Turkish emails. For this study a dataset named "TurkishEmail", which has 800 emails, was used. As the first process in the methodology, the texts in the dataset were edited by reducing capital letters, deleting numbers, deleting stop words, etc. [54]. Then using the Turkish natural language processing (NLP) library Zemberek, all the words in the emails were stemmed [54]. In the study different machine learning algorithms were used to measure the success of the proposed system. These are RF, C45, sequential minimal optimization (SMO), KNN, logistic regression (LR), naive Bayes (NB) and multilayer perceptron (MLP) algorithms; the "WEKA" library is used for these algorithms [54]. For feature selection two types of tests were used, which were chi square (CHI) and information gain (IG) [54]. When the evaluation of feature selection was made using the CHI test, the SMO algorithm received the best result for spam classification but when the evaluation of feature selection was done with the information gain test, the MLP algorithm received the best result as a spam classification algorithm [54]. So it can be seen that algorithms performance change according to the tests used in the evaluation process. The advantage of this research is that it is one of the few reliable studies in the literature; in contrast, the disadvantage is that it uses a small dataset.

The following study performs spam detection on short messages. Örnek (2019) described a study for the detection of spam messages on the short message service (SMS) [55]. TurkishSMS message and UCI SMS spam collections were used as the database for spam detection [55]. Spam detection was performed for Turkish and English SMSs. For the methodology, the Orange 3 application was used for spam SMS identification [55]. In practice, different algorithms were tried on two different datasets [55]. In this way, the most appropriate and correct working algorithm was selected for the dataset [55]. Text mining was used to classify SMS messages as spam and non-spam. Before classification, the texts went through some preliminary stages [55]. These stages are tokenization, lemmatization, term weighting, and feature selection. A spam collection is a dataset containing spam text messages. It contains 5574 samples in total. There are examples of messages with 4827 non-spam and 747 spam messages. The TurkishSMS collection is the first Turkish message collection. It contains 850 samples in total; 430 messages are non-spam and the remaining 420 are spam [55]. The classification phase consists of two phases, training and testing [55]. With the model obtained in the training phase, the classification

process is performed in the testing phase. In this study, a 10-fold cross validation method was used for all algorithms [55]. With this method, the dataset was divided into 10 parts; 9 of them were used for training and 1 was used for testing. Many algorithms have been tried with the Orange 3 application for the two data collections [55]. The highest accuracy rate for the SMS dataset was obtained in neural networks and the highest accuracy for the UCI SMS spam dataset was obtained with the naive Bayes algorithm [55]. It has been observed that the accuracy and error rates were different for different algorithms [55]. The advantage of this study is that it is helpful to determine which algorithms work more accurately in detecting spam words since the study uses many different algorithms to determine which algorithm works better. The disadvantage of the study is that the dataset used was small when compared with other datasets and also there are limited resources for Turkish datasets.

Methodologies Using Honeypot Features:

This section includes reviews of spam detection studies which use honeypot features. Honeypot features are used to detect spam content and information about spam users. Studies in this section baselines the methodology Honeypots uses to detect spammers. The spam users and content is analyzed to detect future spam content and spammers.

Stringhini et al. (2010) investigated how spammers act in social networks (Twitter, Facebook, and Myspace) and what their characteristics and behaviors are [20]. To do that they created accounts that are called honey profiles, and they logged all activities of the honey profiles such as friend requests and invitations. Thus, they were able to examine spammers' interactions between social media users for 3 types of social network. After that, spammers are categorized according to their strategies, which are Displayer, Bragger, Poster, and Whisperer.

Displayer: They show spam content in their profiles, which is not effective.

Bragger: They share status updates or tweets according to the type of social media, and these feeds are shown just to the victims. Friends of the victims cannot see spam messages.

Poster: In addition to status updates and tweets, they post the spam content directly to the victims' wall, and users' friends are able to see the spam content. This one is the most effective behavior.

Whisperer: This kind of spammer sends direct messages that the only victim can see.

To detect spammers they used some metrics that they developed, which are FF ratio (R), URL ratio (U), message similarity (S), friend choice (F), message sent (M), and friend number (FN). According to [20], the definitions are as follows:

FF Ratio (R): number of friend requests that users send to their honey profiles.

URL Ratio (U): messages that contain URLs.

Message Similarity (S): Number of messages sent by a user that include similar content. There is a formula to measure this.

Friend Choice (F): How do spammers detect real users in a network? Maybe there is a list of users which spammers will send friend requests to? Maybe spammers are selecting their victims according to names. This metric is obsolete; which means it is not used in current spam detection models.

Message Sent (M): Number of messages sent by a user. Spammers send less than 20 messages.

Friend Number (FN): Number of friends that a user has.

The authors in [20] built two systems using these features and worked with English data. The RF classification algorithm is used in the Weka framework. They manually picked 500 spam profiles and 500 real users in order to train the classifiers. When choosing spam profiles, they paid attention to at least one of the R, S, U features [20]. In the end, as a concrete result, 15,857 spambots were detected and deleted from Twitter. According to the discussion in the study [20], they were able to capture more extensive spammers because the honeypot is more diverse in terms of both age and nationality compared to other studies. Furthermore, they state that their dataset is larger than the datasets in the literature and includes 3 social networks: Twitter, Facebook, and MySpace [20].

Lee et al. (2010) created social honeypot profiles to collect spammers and log their information from the network [19]. Then they analyzed the properties and the behaviors of the collected spammers to create spam classifiers. Like the study [20] this study also used English MySpace and English Twitter profiles [19]. Honeypots are triggered according to suspicious activity such as suspicious friend requests; then its bot logs information about the spam candidate. In their work, there is a human inspector to validate the accuracy of the spammers [19]. According to their observations from MySpace and Twitter, they attempt to categorize the spammer profiles based on the strategies while spreading the spams [19]. They try to explore if there is a signal other than spam behaviors. Their honeypots are triggered by spam behaviors. After that, they select some features such as age, gender, tweet frequency, and tweet content to train the classifiers to detect and distinguish the spammers. They evaluated more than 60 classifiers in the Weka machine learning toolkit without changing the default values [19]. The experimental results from the MySpace dataset show that marital status and sex are the least discriminative features and "about me" content is a more discriminative feature. Moreover, according to their work, the best classifier is Decorate among the ones they

use in their study. In addition, classification results from the Twitter spam dataset show URLs per tweet, account age, and text-based features extracted from tweets are more discriminative features in comparison to other features. According to their analysis, with the content similarity for spammers, promoters, and legitimate users, spammers can be detected with the legitimate users who have the least similar content. For the average URL per tweet comparison, spammers and promoters have the same behavior. They state their future directions [19] are as follows: 1.) Traditional email and web approaches can be useful for spam detection on social media; 2.) Honeypots can be widened; 3.) New approaches can be added to the honeypots to detect spammers.

Methodologies Using Sentiment Information and Graph Methodologies:

This section includes reviews of spam detection studies which use sentiment information and graph methodologies. The studies in this section analyzes the network graph and its users using graph methodologies then performs sentiment analysis on the content of the users of the network to determine if the user content includes spams. After these analyses, spam users and content are detected.

Hu et al. (2014) built their methodology using psychological findings [15]. They state that the sentiment difference between randomly selected users must be greater than the sentiment difference between the two users with the same identity, i.e., both are spammers or normal users. For example, if both users are spammers or normal users then their sentiments might be consistent. Then they used the X(u)w formula to find the sentiment of the users. To calculate sentiment differences between two users the following formula is used [15]:

$$d(i,j) = \left\| s(i) - s(j) \right\| \tag{18}$$

They first calculated the sentiment difference with the same identity and then they also calculated the sentiment difference of two different randomly selected users. They did this calculation for each user in their dataset. Then they group the users based on the calculated sentiment differences. After that, they realize that this sentiment difference can be a determination feature for spammers [15]. They used 3 English datasets, namely TAMU Social Honeypots (TUSH), Twitter Suspended Spammers (TSS), and Stanford Twitter Sentiment (SENT). While TAMU and TUSH dataset contain labels for social spammer detection, SENT has sentiment labels. The supervised sentiment analysis model was trained in the labeled SENT dataset and this learned model was applied to calculate the sentiment score of the TUSH and TSS datasets [15].

They proposed a model with a graph Laplacian. They constructed an undirected graph with edges and nodes [15]. For their adjacency matrix, Eq (3) is used:

$$A(i,j) = \begin{cases} 1 \ if \ \ u_i \in N(u_j) \ or \ u_j \in N(u_i) \\ 0 \ otherwise. \end{cases} \tag{19}$$

Here *ui* and *uj* are nodes and *N*(*ui*) and *N*(*uj*) are their k-nearest neighbors, respectively. They modeled content, sentiment, and social network information. Then they used these formulas to construct an algorithm to detect social spammers. The basic idea is to optimize the target by targeting the variable while correcting the other [15]. According to the experimental results, the sentiment information approach compared with the other baseline methods helped to improve the accuracy of spammer detection [15].

In [14] the directed graph Laplacian was used to model social network information, the same as the study [15]. There were four types of relationship in this graph: [spammer, spammer], [normal, normal], [normal, spammer], and [spammer, normal]. However, they did not include the fourth one because it can be easily manipulated [14]. Considering a model based on *k* users, they proposed in their methodology to update the **U** and **H** factor matrices by adding Online Social Spammer Detection (OSSD) to the *(k + 1)* user without over-calculating [14]. With this formula, they only updated the columns of the encoding matrix and it decreased the computational cost. With their proposed approach, OSSD, time complexity was reduced compared to Non-Negative Matrix Factorization (NMF). Their time complexities are O (nm*r*2) and O (n*r*2) [13]. They used two real-world datasets, i.e., the TAMU Social Honeypots Dataset (TwitterT) and Twitter Suspended Spammers Dataset (TwitterS). While the first one is a balanced dataset, the second one is an imbalanced dataset. This means that spammers and legitimate users' proportions are nearly the same in the TwitterT dataset. [14] According to their experimental results, their proposed approach with online learning did not bring any negative impact compared with batch-mode learning. Moreover, Batched-Mode Social Spammer Detection (BSSD), which is a variant of the proposed method, and OSSD gave better results compared to other methods [14].

From the reviewed spam detection studies, studies that use Twitter data were collected. All the collected studies listed in Table 3 based on authors of the publication, year of publication, methodology, the used dataset, performance metrics and accuracy of the publication.

CONSIDERATIONS, CURRENT CHALLENGES, CONCLUDING REMARKS

Considerations:

The task of finding influencers in social networks also involves some considerations. As 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

Table 3. Comparison of different spam detection studies on the Twitter dataset

Authors	Year	Methodology	Dataset	Performance Metric	Results
McCord and Chuah[17]	2011	Support Vector Machine, Naive Bayes, K-nearest Neighbor, Random Forest	Twitter	Precision and F-measure	95.7%
Hu et al.[14]	2014	Laplacian matrix factorization	Twitter	F1-measure	91.8%
Hu et al.[15]	2014	Laplacian matrix factorization	Twitter	F1-measure	97.7
Ferrara et al. [51]	2016	Supervised, unsupervised, hybrid methods	Twitter	Accuracy	95%
Benevenuto et al. [50]	2010	SVM	Twitter	Micro-F1	87.6%
Stringhini et al. [20]	2010	Random Forest	Twitter	Accuracy	90.93%
Lee et al. [19]	2010	Decorate, LogitBoost, FT, SimpleLogistic, LibSVM, Classification, Regression	Twitter	Accuracy	99.21%

influencers to sell and advertise their products or services. In this respect, it is important to find influencers that are relevant to the brand's sector, i.e., to classify influencers by topics.

As health and safety considerations, users gain awareness of events affecting the world if they have access to the correct information. In some political and military scenarios, such as wars, various news stories from misleading sources that should not be followed must be carefully analyzed before making critical decisions. Furthermore, analyzing influencers and tweets is crucial in many subjects that affect human health, such as the non-immunization of children and the Blue Whale Challenge.

Current Challenges and New Horizons:

Commonly, identifying influencers is a task including some steps, such as data collection and selecting important features, data preprocessing, network structuring, modeling, finally ranking opinion leaders, and evaluating the experimental results. Implementing such a system causes several challenges for researchers:

The Complexity of Computations: Graph-based and PageRank-based approaches usually require more resources. On the other hand, diffusion-bases usually use expensive mathematical calculations in order to apply diffusion processes on the network while training the corpus. Consequently, all these computations will increase the overall complexity. In order to handle this issue, researchers need to find ways to reduce the complexity of their algorithms.

Availability of Datasets: Only a small number of databases are available for a limited number of languages (i.e., English, German, Chinese, etc.). There are some public datasets based on some different purposes; the SNAP-LIM dataset can be regarded as one of the largest information diffusion datasets, consisting of 500 million tweets [56]. The Enron dataset is among the popularly used datasets [57,58]. The Enron dataset contains 255,000 emails and 1 million authors and infers social relationships. Furthermore, the Higgs Twitter dataset [59] is an influential user detection dataset containing 456,626 nodes. This dataset includes the following relation among users and user reactions to posted messages. There is another influential node identification dataset, namely the ISIS Twitter dataset, containing 17,000 tweets [60]. The ISIS Twitter dataset is built by collecting tweets from 112 users. There is also the OpinionRank dataset [61], which contains the reviews on Edmund.com and Tripadvisor.com.

Processing Complexity of a Large Dataset: There exists a considerable processing cost for a large collected dataset. This processing cost includes data preprocessing, identifying and selecting important features, network structuring, and modeling (i.e., running machine learning/deep learning algorithms) of the massive corpus. Researchers who implement influencer systems on social networks need to optimize the processing time/complexity of their algorithms.

Dealing with Spam/Bot Content: Identifying influencers on social networks involves a big challenge: spam/bot content on social networks. It is classified as a challenge because there are many spam users in social networks. To be able to identify influencers, spam users must be identified and removed from the network. So to overcome this challenge, researchers need to develop methods to filter spam users from their influencer dataset. There is a wide range of algorithms suggested in the literature to filter spam users on social networks. These algorithms are mentioned in detail in Section 3.

Computation of Hardware Systems: Making an influencer system work is computationally expensive. In particular, modeling and visualizing the network require expensive hardware resources.

Analysis/Evaluation: There are several ways to evaluate an opinion leader system:

1. The reputation of social network users will be calculated. The most important indicator of this is the number of retweets (the ability to share the tweets of the people they like in their own accounts). Since

the number of retweets is directly proportional to the amount of information spread, it is used for the confirmation of opinion leaders [9,62,63]. It will be assessed whether the retweet amount and the opinion leaders determined by the Social Ranking Social Counting (SRSC) system are the same persons.

2. Calculation of the in-degree values of social network users is another validation technique used in studies in the literature to find opinion leaders [22]. The in-degree values of the nodes in the social network will be detected and calculated in the SRSC system. The in-degree value of a node is the number of edges coming into the node. Then the people who are opinion leaders will be identified in a social network. With the results of this method, it will be evaluated whether the opinion leaders to be determined by the SRSC system are the same persons.

Concluding Remarks

More and more people are interacting through microblogging sites. Ideas emerge and spread quickly with the easy usage of microblogging sites. Due to the efficiency of these sites, it is observed that important events and news were published on these sites before even being published by the necessary sources. Due to these reasons, social media can be considered among the primary sources of information that affects the communities' opinions. Furthermore, detecting opinion leaders and filtering spam content are two important and attractive tasks for both commercial and academic platforms. Consequently, we prepare a detailed survey that includes not only recent advancements but also past studies regarding the problems of opinion leader identification and spam filtering. In order to detect opinion leaders in social media, a number of methods have been proposed. These methods are grouped into five different categories: 1.) Diffusion-based approaches, 2.) Graph-based approaches, 3.) Statistical and stochastic approaches, 4.) Page-Rank-based approaches, 5.) Machine learning approaches. The advantages and disadvantages of each method are analyzed, compared, and reported in Section 2. On the other hand, we also review techniques for spam filtering. These methods can be grouped into four categories: 1.) Methodologies using user-based and content-based features, 2.) Methodologies using honeypot features, 3.) Methodologies using sentiment information and graph-based methodologies. The advantages and disadvantages of each method are analyzed, compared, and reported in Section 3. We also reviewed two surveys that have properties similar to those of this study, we analyzed the differences and similarities of this study with the other surveys and made a comparison in addition to briefly explaining the other surveys' contents. We also report general considerations. Identifying influencers is a task including some steps such as data collection and selecting important features, data preprocessing, network structuring, modeling, finally ranking opinion leaders, and evaluating the experimental results. Implementing such a system entails several challenges for researchers. We analyze and report these challenges in Section 4.

This survey investigates the advancements in the identification of opinion leaders and the detection of spam content fields and highlights their strengths in comparison to each other. Nevertheless, determining whether to use a diffusion process-based approach, a graph-based approach, a statistical approach, or a PageRank-based approach for identifying opinion leaders is still a challenging task that depends on the availability and the size of the dataset and the nature of the problem being investigated. The same is valid for detecting spammers on social networks since there are many methods such as user-based and content-based techniques, methodologies using honeypot features, methodologies using sentiment information, and graph-based methodologies. As we have learned in this paper, there are many challenges that can occur in the process of spam detection and opinion leader identification in social networks. We hope that this paper will be helpful for interested readers for their further exploration on spam detection and opinion leader identification on social networks.

ACKNOWLEDGMENT

This work is supported in part by the Scientific and Technological Research Council of Turkey (TÜBİTAK) through grant number 118E315 and grant number 120E187. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of TÜBİTAK.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Chakraborty M, Pal S, Pramanik R, Ravindranath Chowdary C. Recent developments in social spam detection and combating techniques: A survey. Inf Process Manag 2016;52:1053–1073. [CrossRef]

[2] Peng S, Zhou Y, Cao L, Yu S, Niu J, Jia W. Influence analysis in social networks: A survey. J Netw Comput Appl 2018;106:17–32. [CrossRef]

[3] Wu T, Wen S, Xiang Y, Zhou W. Twitter spam detection: Survey of new approaches and comparative study. Comput Secur. 2018;76:265–284. [CrossRef]

[4] Pepitone A, Katz E, Lazarsfeld PF. Personal influence: The part played by people in the flow of mass communications. Am J Psychol 1957;70:157. [CrossRef]

[5] Hon LC, Grunig JE. Guidelines for measuring relationships in public relations. Florida: Institute for Public Relations; 1999.

[6] Johnson TJ, Kaye BK, Bichard SL, Wong WJ. Every blog has its day: Politically-interested internet users' perceptions of blog credibility. J Comput Mediat Commun 2007;13:100–122. [CrossRef]

[7] Marwick A, Boyd D. To see and be seen: Celebrity practice on Twitter. Converg Int J Res New Media Technol 2011;17:139–158. [CrossRef]

[8] Kaymaz G. Detection of topic-based opinion leaders in microblogging environments. [Master Thesis] Istanbul: Bogaziçi University Graduate Program in Computer Engineering; 2013.

[9] Alp ZZ, Öğüdücü ŞG. Identifying topical influencers on twitter based on user behavior and network topology. Knowl Based Syst 2018;141:211–221. [CrossRef]

[10] Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. 2014.

[11] Seward ZM. Twitter admits that as many as 23 million of its active users are automated Quartz. 2014. Available at: https://qz.com/248063/twitter-admits-that-as-many-as-23-million-of-its-active-users-are-actually-bots/ Accessed on Nov 24, 2021.

[12] Ha A. Facebook says it's cracking down on clickbait. TechCrunch. 2014 Aug 25 Available at: http://techcrunch.com/2014/08/25/facebook-vs-clickbait/ Accessed on Nov 24, 2021.

[13] Yardi S, Romero D, Schoenebeck G. Detecting spam in a twitter network. First Monday 2010;15:2793. [CrossRef]

[14] Hu X, Tang J, Liu H. Online social spammer detection. Proceedings of the AAAI Conference on Artificial Intelligence. 2014:28.

[15] Hu X, Tang J, Gao H, Liu H. Social spammer detection with sentiment information. In: 2014 IEEE International Conference on Data Mining. IEEE; 2014. [CrossRef]

[16] Ghosh S, Viswanath B, Kooti F, Sharma NK, Korlam G, Benevenuto F, et al. Understanding and combating link farming in the twitter social network. In: Proceedings of the 21st international conference on World Wide Web - WWW '12. New York, New York, USA: ACM Press; 2012. [CrossRef]

[17] McCord M, Chuah M. Spam detection on twitter using traditional classifiers. In: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011:175–186. [CrossRef]

[18] Lee S, Kim J. WarningBird: A near real-time detection system for suspicious URLs in twitter stream. IEEE Trans Dependable Secure Comput 2013;10:183–195. [CrossRef]

[19] Lee K, Caverlee J, Webb S. Uncovering social spammers: social honeypots+ machine learning. SIGIR '10: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. 2010:435–442. [CrossRef]

[20] Stringhini G, Kruegel C, Vigna G. Detecting spammers on social networks. In: Proceedings of the 26th Annual Computer Security Applications Conference on - ACSAC '10. New York, USA: ACM Press; 2010. [CrossRef]

[21] Anantharam P, Thirunarayan K, Sheth A. Topical anomaly detection from Twitter stream. In: Proceedings of the 3rd Annual ACM Web Science Conference on - WebSci '12. New York, USA: ACM Press; 2012. [CrossRef]

[22] Weng J, Lim EP, Jiang J, He Q. Twitterrank: finding topic-sensitive influential twitterers. WSDM '10: Proceedings of the third ACM international conference on Web search and data mining. 2010:261–270.

[23] Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03. New York, New York, USA: ACM Press; 2003. [CrossRef]

[24] Zhao Y, Li S, Jin F. Identification of influential nodes in social networks with community structure based on label propagation. Neurocomputing. 2016;210:34–44. [CrossRef]

[25] Van Eck PS, Jager W, Leeflang PSH. Opinion leaders' role in innovation diffusion: A simulation study: Opinion leaders' role in innovation diffusion. J Prod Innov Manage 2011;28:187–203. [CrossRef]

[26] Cho Y, Hwang J, Lee D. Identification of effective opinion leaders in the diffusion of technological innovation: A social network approach. Technol Forecast Soc Change 2012;79:97–106. [CrossRef]

[27] Rehman AU, Jiang A, Rehman A, Paul A, Din S, Sadiq MT. Identification and role of opinion leaders in information diffusion for online discussion network. J Ambient Intell Humaniz Comput 2020.

[Online ahead of print]. doi: 10.1007/s12652-019-01623-5. [CrossRef]

[28] Cui L, Pi D. Identification of micro-blog opinion leaders based on user features and outbreak nodes. Int J Emerg Technol Learn 2017;12:141. [CrossRef]

[29] Social & Organizational Network Analysis software & services for organizations, communities, and their consultants. Available at: http://www.orgnet.com. Accessed on Nov 24, 2021.

[30] Bonacich P. Some unique properties of eigenvector centrality. Soc Networks 2007;29:555–564. [CrossRef]

[31] Gökçe OZ, Hatipoğlu E, Göktürk G, Luetgert B, Saygin Y. Twitter and politics: Identifying Turkish opinion leaders in new social media. Turk Stud 2014;15:671–688. [CrossRef]

[32] Meltzer D, Chung J, Khalili P, Marlow E, Arora V, Schumock G, et al. Exploring the use of social network methods in designing healthcare quality improvement teams. Soc Sci Med 2010;71:1119–1130. [CrossRef]

[33] Jain L, Katarya R. Discover opinion leader in online social network using firefly algorithm. Expert Syst Appl 2019;122:1–15. [CrossRef]

[34] Li C, Bai J, Zhang L, Tang H, Luo Y. Opinion community detection and opinion leader detection based on text information and network topology in cloud environment. Inf Sci 2019;504:61–83. [CrossRef]

[35] Li F, Du TC. Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs. Decis Support Syst 2011;51:190–197. [CrossRef]

[36] Amor BR, Vuik SI, Callahan R, Darzi A, Yaliraki SN, Barahona M. Community detection and role identification in directed networks: understanding the Twitter network of the care. data debate. In: Adams N, Heard N, editors. Dynamic Networks and Cyber-Security. London: World Scientific; 2016: 111–136. [CrossRef]

[37] Zhao Q, Erdogdu MA, He HY, Rajaraman A, Leskovec J. SEISMIC: A self-exciting point process model for predicting tweet popularity. arXiv [cs. SI]. 2015. Preprint. doi:10.1145/2783258.2783401. [CrossRef]

[38] Alp ZZ, Öğüdücü ŞG. Influence Factorization for identifying authorities in Twitter. Knowl Based Syst 2019;163:944–954. [CrossRef]

[39] Tam MS. Opinion leadership role of social media influencers. Gumushane Univ e-journal Faculty Commun 2020;8:1325–1351.

[40] Luo L. Identifying opinion leaders with improved weighted LeaderRank in online learning communities. Int J Perform Eng 2018;14:193–201. [CrossRef]

[41] Song X, Chi Y, Hino K, Tseng B. Identifying opinion leaders in the blogosphere. In: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management - CIKM '07. New York, New York, USA: ACM Press; 2007. [CrossRef]

[42] Zhang B, Bai Y, Zhang Q, Lian J, Li M. An opinion-leader mining method in social networks with a phased-clustering perspective. IEEE Access. 2020;8:31539–550. [CrossRef]

[43] Hu M, Liu S, Wei F, Wu Y, Stasko J, Ma K-L. Breaking news on twitter. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2012. [CrossRef]

[44] Safali Y. Classification of social media users' opinions on the cumhur alliance with machine learning techniques. Bilgisayar Bilimleri ve Teknolojileri Dergisi 2020;1:51–57.

[45] Gümüşsu E, Murat N. To examine of the relationship between shared tweets of the 'tamam' and 'devam' tags and the people's alliance and national Alliance. Bilişim Teknolojileri Dergisi. 2019;12:287–298. [CrossRef]

[46] Aleahmad A, Karisani P, Rahgozar M, Oroumchian F. OLFinder: Finding opinion leaders in online social networks. J Inf Sci 2016;42:659–674. [CrossRef]

[47] Chen Y-C. A novel algorithm for mining opinion leaders in social networks. World Wide Web 2019;22:1279–1295. [CrossRef]

[48] Duan J, Zeng J, Luo B. 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). IEEE; 2014:377–383. [CrossRef]

[49] Jain L, Katarya R, Sachdeva S. Opinion leader detection using whale optimization algorithm in online social network. Expert Syst Appl 2020;142:113016. [CrossRef]

[50] Benevenuto F, Magno G, Rodrigues T, Almeida V. Detecting spammers on twitter. Collaboration, electronic messaging, anti-abuse and spam conference (CEAS). 2010;6:12.

[51] Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Commun ACM 2016;59:96–104. [CrossRef]

[52] Sahin DO, Demirci S. Spam filtering with KNN: Investigation of the effect of k value on classification performance. In: 2020 28th Signal Processing and Communications Applications Conference (SIU). IEEE; 2020. [CrossRef]

[53] Şeydanur AHİ, Soğukpinar İ. Phishing e-mail detection with deep learning models. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi 2020;13:7–31. [Turkish]

[54] Eryilmaz EE, Şahin DÖ, Kiliç E. Detection of Turkish spam emails with machine learning algorithms using different feature selection methods. Türkiye

Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi 2020;13:57–77. [Turkish]

[55] Örnek Ö. Spam detection in Turkish and English SMS messages with orange 3. Journal of ESTUDAM Information 2020;1:1–4.

[56] Yang J, Leskovec J. Modeling information diffusion in implicit networks. In: 2010 IEEE International Conference on Data Mining. IEEE; 2010. [CrossRef]

[57] Jaber M, Wood PT, Papapetrou P, Helmer S. Inferring offline hierarchical ties from online social networks. In: Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion. New York, New York, USA: ACM Press; 2014. [CrossRef]

[58] Tang W, Zhuang H, Tang J. Learning to infer social ties in large networks. In: Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011:381–397. [CrossRef]

[59] De Domenico M, Lima A, Mougel P, Musolesi M. The anatomy of a scientific rumor. Sci Rep 2013;3:2980. [CrossRef]

[60] Li Q, Kailkhura B, Thiagarajan JJ, Zhang Z, Varshney PK. Influential node detection in implicit social networks using multi-task Gaussian copula models [Internet]. arXiv [cs.SI]. 2016. Preprint. https://doi.org/10.48550/arXiv.1611.10305.

[61] Ganesan K, Zhai C. Opinion-based entity ranking. Inf Retr Boston 2012;15:116–150. [CrossRef]

[62] Cha M, Haddadi H, Benevenuto F, Gummadi K. Measuring user influence in twitter: The million follower fallacy. Proceedings of the international AAAI conference on web and social media. 2010;4:10–17.

[63] Sankar CP, Asharaf S, Kumar KS. Learning from bees: An approach for influence maximization on viral campaigns. PLoS One 2016;11:e0168125. [CrossRef]