IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 16, 2020, accepted May 26, 2020, date of publication June 1, 2020, date of current version June 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2999033

Measuring Reputation and Influence in Online Social Networks: A Systematic Literature Review

SAMI AL-YAZIDI, JAWAD BERRI, MUHAMMAD AL-QURISHI, AND MAJED AL-ALRUBAIAN
Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: Jawad Berri (jberri@ksu.edu.sa)

This work was supported by the DataPlume Lab, Riyadh, Saudi Arabia, info@dataplume.com, http://www.dataplume.com.

ABSTRACT Online Social Networks (OSNs) are an integral part of the modern lifestyle, and the use of OSN is highly prevalent today. OSNs have enabled the generation of unprecedented quantities of data on users, their interests, relationships, actions, opinions, locations visited, items purchased, and their experiences in various aspects of life. Data and information related to reputation and influence exist on OSNs and are hidden. The objective of this study is to provide a better and deeper understanding and a detailed review of the current state of research pertaining to the measurement of reputation and influence in OSNs, specifically Twitter. We conducted a Systematic Literature Review to collect, analyze, and synthesize data on the accuracy and value of previous literature that has focused on this field, has addressed a variety of topics, and has been published in digital databases between 2010 and September 2019. We followed a predefined review procedure that integrated both automatic and manual search strategies to cover all relevant research papers published in this period.

INDEX TERMS Influence, online social networks, reputation, systematic literature review, trust, twitter.

I. INTRODUCTION

The reputation and influence of a user on social media is an emerging area of research that is progressively gaining attention [1]. Although several studies have been carried out in this field, there is no systematic and reliable model to measure reputation and user influence on social media. The use of social media differs from one individual to another based on the extent of exposure. As a result, it can affect a user positively or negatively. This paper presents a systematic review of the literature on the reputation and influence of a user on social media, produced from 2010 to September 2019. It synthesizes the findings of the existing literature and identifies areas that require further study and opportunities for future research in this field. The following research questions (RQs) are considered in this study:

- RQ1: What are the most important features for measuring reputation and influence?
- RQ2: Why is measuring reputation and influence on Twitter relevant?
- RQ3: What measurement metrics and methodologies have been used to measure reputation and influence? Why are they insufficient?

This study provides a more comprehensive picture of various emergent topics, methodologies, and theories to provide guidelines for research and to enhance future studies in this area. This study also contributes knowledge to the field of social media, with a focus on reputation and user influence measurement through a careful analysis of the existing studies in the field. The findings obtained will be used to present an approach for the measurement of user reputation and influence.

II. BACKGROUND

As Internet use has risen dramatically over the recent few years, online social networks (OSNs) have grown to become challenging areas of study, where massive content is rapidly generated and complex relations among actors converge [2]. Knowing the reputation and influence of users and being able to predict both can be useful for many areas of business such as viral marketing, information broadcasting, recommendation systems, searching, and

The associate editor coordinating the review of this manuscript and approving it for publication was Chunsheng Zhu.

social customer relationship management, among several others [3].

A. OVERVIEW OF OSNs

A social network can be defined as a website that provides its users with the opportunity to be linked and connected with their relatives, friends, and family [28]. An account or profile on an OSN is the online representation of a user (individual or organization) on the OSN. The terms profile, user, and account are used interchangeably across OSNs. In most user profiles, we find descriptive information such as names, gender identities, education, interests, and contact information. Connections among users may be one- or two-sided. A one-sided connection is where a user requests a connection with another user, and then the connection is established immediately without any action from the other side. A typical example of this is the follower relationship on Twitter [13]. Two-sided connection comprises the acceptance of the relationship between two users by both of them. An example of this is the relationship friendship on Facebook. Social networks enable users to share photos, videos, music, and other information with selected friends or with the public at large [14], [15]. Some of the well-known social networks such as Facebook, Twitter, LinkedIn, Instagram, YouTube, and Myspace provide convenient options to maintain connections with other users and friends that have shared interests [16]–[18].

B. STRUCTURE OF AN OSN

An OSN is structured in the form of a graph in which nodes representing users are connected to each other, whereas interactions between are represented as connections between the nodes. Formally, an OSN is represented as a graph G= (V, E) where V is a set of nodes representing users, and E is a set of edges representing relationships between users [13]. A graph has three important related concepts:

Centrality: A measure indicating the importance of a node in the network.
Graph diameter: The longest path among the shortest paths between any two nodes.
Graph vertex degree: The total number of edges associated with a node.

A social graph depicts the personal relationships among users in an OSN, and is the cornerstone for the creation of a strong model of trustworthiness that represents all kinds of social activities [19]. Social graphs are heterogeneous because they capture different kinds of social activities such as posting, sharing, and messaging. Through the use of social graphs, it is possible to develop different measures of trustworthiness and to adopt a simple approach that can help understand the nodes that mostly affect others and the nodes that see other nodes in terms of trust. Users, their social relationships, and activities are the most important entities in OSNs. Malicious activities can be differentiated from legitimate ones by measuring the trustworthiness of those activities and by classifying each activity into a category of conformity. Every entity can be assigned a trustworthiness score using social graphs. If the trustworthiness of an entity is low, there is a high possibility that the social activities associated with it will turn out to be less reliable. This is why social graphs are used along with any suggested model of trust to validate the results acquired. [20]–[22].

TABLE 1. Twitter entities relationships.

	User	Tweet
User	Follow, Mention	Tweet, Retweet, Like and Reply
Tweet	Retweeted by, liked by and replied by	Replies / is replied from Retweets / is retweeted from Retweet with comments / being retweeted with comments

C. TWITTER

Twitter is considered as one of the main OSNs today [4]. It was established in 2006 and since 2010 has been studied extensively in the contexts of social network analysis, computer science, and sociology [51], [77], [80], [85], [94]. It is a microblogging site on which people broadcast short messages (up to 280 characters), called tweets. Tweets may contain plain text, images, videos, URLs, mentions of other users (preceded by the "@" symbol), and hashtags formed by using the "#" sign in front of a word with no punctuation or spaces between them. "Trending topic" is a term coined by Twitter to refer to the most used keywords that appear in a large number of tweets within a given period of time.

There are two main entities in Twitter: users and tweets. Public relationships between these entities are in the form of user-to-user, user-to-tweet, tweet-to-tweet, or tweet-to-user. The actions permitted for each type of relationship are presented in Table 1 [23]. Both user-to-tweet and tweet-to-user relationships are symmetric. We may say that user X is a follower of user Y when X carries out the action of following Y. Then, Y becomes the "followee." In the relationship between a user and a tweet, four actions are available for users: retweeting (sharing the tweet of another user), retweeting with comment (where the user can add a comment to the retweeted tweet), reply to a tweet and liking (denoting that the user liked the tweet). These actions can be considered as implicit interactions between the user's action and the author of the original tweet [23].

1) TWITTER API

Twitter offers public data though its Application Programming Interface (API). An API is a set of functions, protocols, and tools that are used to build an application or to facilitate communication with provided services. Twitter provides APIs to developers who aim to provide access to Twitter data and information. Twitter data are obtained through different kinds of requests that allow researchers and developers to search for historical and/or real-time information.

2) REPUTATION AND INFLUENCE ON TWITTER

Reputation and influence have been used for different purposes on Twitter, such as political activities, rumormongering, human mobility, transportation, and epidemiology, among others. Catching a small set of active users who can propagate trends, ideas, rules, and ambitions is very critical to study their effect and influence on Twitter. The identification and measurement of the reputation and influence of users on Twitter represent areas that have been studied considerably [23], [160], [133]. This is especially important, considering the high percentage of users who are often inactive or do not provide additional information [24]. The criteria for identifying influential users and their reputation are as many as the growing number of techniques to rank them.

This problem is manifold. Social media platforms, especially Twitter, generate large amounts of content within very short periods of time. Human analysis of the incoming data is not always feasible. Therefore, reputation management systems have come up to help businesses understand and manage issues pertaining to reputation on social media [25]. These systems identify important feedback through various types of quantitative analyses, and help keep businesses up-to-date with respect to reputation and related issues. One approach to this is filtering out the so-called "high-impact" responses based on factors such as the reputation of the user [26]. While this approach is excellent for singling out tweets that are likely to have significant impact and decreasing the computational analysis required, it also leaves the system open to an inadequate understanding of overall sentiments, which does not change as a result of high-impact tweets alone [27].

III. THE REVIEW METHOD

This study uses Systematic Literature Review (SLR), which according to Kitchenham (2007) is a repetitive process that combines all existing literature on a specific topic or research question [5]. SLR aims to address problems by identifying, critically evaluating, and integrating the findings of all relevant high-quality individual studies addressing one or more research questions. This method determines how much the current studies have progressed toward clarifying the use and impact of social media on their users by means of reputation/influence assessment. It identifies relations, contradictions, gaps, and inconsistencies in the literature, and explores reasons to keep studying the area. It also helps formulate general statements, develop theories, and describe directions for future research [9]. SLR is very important for this study as it is able to address a much broader question than single empirical studies can, because it has the potential to provide the most important practical implications [6]. The authors of this paper followed the steps that Kitchenham and Charters [5] suggested for conducting an SLR. These steps are:

- Identify the review questions;
- Formulate a review protocol;
- Set up inclusion and exclusion criteria;
- Review selection procedures and strategies;
- Conduct quality assessment of studies reviewed; and
- Extract data and synthesize evidence.

The authors identified the following review questions:

RQ1: What are the most important features for measuring reputation and influence?

RQ2: Why is measuring reputation and influence on Twitter relevant?

RQ3: What measurement metrics and methodologies have been used to measure reputation and influence? Why are they insufficient?

A. REVIEW PROTOCOL

The steps deployed in this SLR have been defined to help realize comprehensive results. The review protocol is vital for the realization of an SLR as it identifies the criteria that can be employed in pursuing the goals of the review while also limiting the chances of bias [7]. The review plan entails the following steps: (1) research setting, (2) search strategy, (3) review questions, (4) review selection process, (5) quality assessment, (6) data extraction, and (7) the synthesis of the data extracted [5].

B. INCLUSION AND EXCLUSION CRITERIA

The main objective behind using inclusion and exclusion criteria is to ensure that all primary studies selected for the SLR are relevant and are related to this study. Data were collected from various documents such as journals, conference papers, book chapters, and workshops, all written in English and published in digital databases between 2010 and 2019. The authors did not include articles with content that did not apply to this study. This was achieved through a peer-review process. The following are the criteria for this review:

Inclusion criteria:

The text was available in full.
It is written in English.
It was published between 2010 and 2019.
It is related to the research questions.
It focused on measuring user reputation and influence on Twitter.
It was published in selected digital databases.

Exclusion Criteria:

It is not available in full text.[1]
It is written in a language other than English.
It is outdated from specified range.
It has no relationship with the research questions.
It is a duplicated study written by the same authors.

[1] Some journal papers may be announced for upcoming journal issues but not released yet as full text.

C. SEARCH STRATEGY

Search strategy refers to an organized structure of key terms used to search a database. This technique involves the use of manual and automatic search efforts to retrieve accurate results. This technique was used to explore the sources for review in this study. Manual search was carried out for primary study references while automatic search was conducted to identify keywords defined by authors of the reference studies and as an electronic search using online scientific databases to address the research questions [5]. Next, five online databases including ISI Web of Knowledge (includes all databases such as Derwent Innovations Index, KCI-Korean Journal Database, MEDLINE, Russian Science Citation Index, and SciELO Citation Index), IEEE explorer, ACM Digital Library, and Science Direct, were examined and selected as the main sources of data. These databases were selected because they were seen as being the most relevant and up-to-date. They are also the most prominent among journal publications and attract the most reputable conferences and articles in the field.

The keywords of interest were searched for among both titles and research questions to define the limits to place under consideration. As a result, the keywords were matched with the published research and relevant literature by the use of a mix of common measuring reputation and its influence on users in Twitter. These keywords were used to search the selected databases to identify important studies published in the area:

"Measuring reputation on Twitter,"

"Measuring user influence on Twitter," and

"Sentiment analysis for reputation on Twitter."

We started with an automated search and followed it up with a manual search to ascertain the initial findings of the study. To collect references to primary studies, we used a forward and backward approach [8] to guarantee additional reference scanning. Through this, the SLR accomplished its goals and identified answers for the research questions. The second manual search aimed to ascertain whether the research carried out was complete in relation to the objectives of the study [8], [10]. EndNote, Mendeley, and Microsoft Excel were used to organize and arrange all primary studies and helped identify duplicate studies so that they could be removed.

D. STUDY SELECTION PROCESS

The results of the first two rounds presented 605 papers. Among these, 112 were eliminated using End Note application since they were duplicates. The inclusion and exclusion criteria were used on the remaining 493 papers to exclude those that were not available in full and those that were not written in English. Thus, 125 papers were excluded.

The remaining 368 papers were examined based on the title and abstract to make sure they are relevant to our study. Filtering out irrelevant papers during this step was made using the following procedure: i) All authors read the title and abstract of papers and provide a decision whether to consider the paper or reject it from consideration in the literature review; ii) Papers for which all authors have unanimously voted "Not Consider" are excluded; iii) Papers for which at least two authors have voted "Consider" are considered in the literature review; iv) Papers for which only one author has voted "Consider" goes through a second round of selection where another author (not the author who recommended the paper) reads the full paper and provides a final decision on considering or not considering the paper. The result of this step ended with 116 papers that are relevant to the subject under study and that have been examined thoroughly in this SLR.

TABLE 2. Study selection results.

Online DB used in SLR	Initial Results	Final Results
ISI Web of Knowledge (including all databases)	247	46
IEEE Explore	104	34
ACM Digital Library	72	14
Science Direct	182	22
Summary	**605**	**116**

Table 2 presents the number of primary studies drawn from various online databases through the selection process. Most articles gathered before the selection process were found on ISI Web of Knowledge (247), Science Direct (182), IEEE Explore (104), and ACM Digital Library (72).

In the second round, most primary studies were drawn from ISI Web of Knowledge (46), IEEE Explore (34), Science Direct (22), and ACM Digital Library (14). The figure below compares the results from each database.

E. QUALITY ASSESSMENT

We examined the selected review papers by applying a set of techniques that provide a decision on the interpretation and findings of the primary studies [5]. It is an evaluation to extent to which trials design and management are likely to have prevented systematic errors and biasness. To evaluate the quality of the results in these papers, we applied a set of six quality criteria to sharpen our results:

QA1. How does this paper address the measurement of reputation and influence of users on Twitter?

QA2. Does the topic pertain to the fields of computer science, information systems, or marketing/management?

QA3. Does it explain the context in which the research was carried out?

QA4. Is the research methodology described extensively?

QA5. Does the paper explain the methodology used for data collection clearly?

QA6. Does the paper employ data analysis approach correctly?

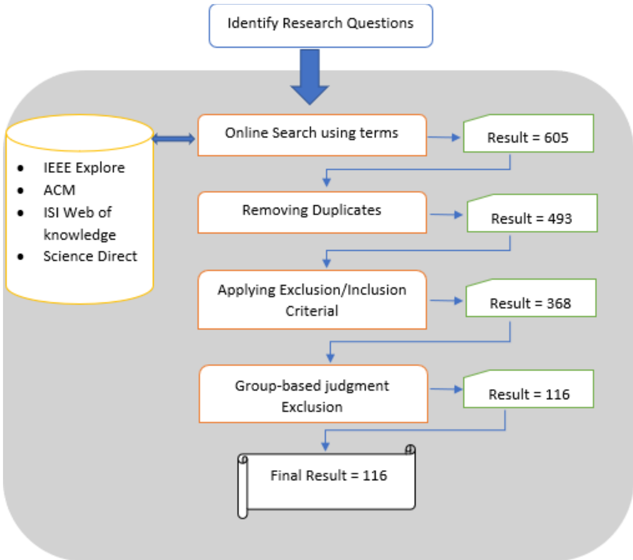

FIGURE 1. **Study selection process.**

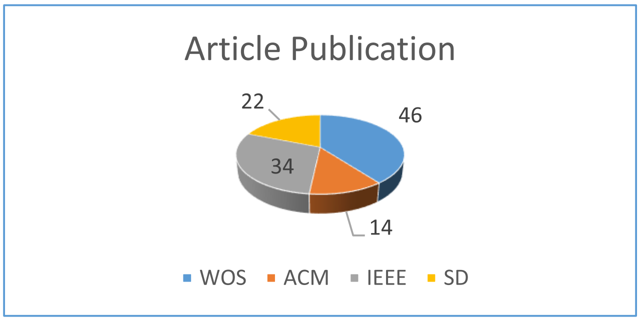

FIGURE 2. **Databases that published the articles.**

IV. DATA EXTRACTION AND SYNTHESIS FOR SLR

In any systematic review, data extraction occurs before synthesis. One has to read the studies included and extract the relevant results to respond to the review questions. At this stage, we developed a data extraction form in which the findings from the 104 papers were recorded accurately. We used these forms as a means to record the data collected through the review method accurately [5]. We collected information by scanning each study using Microsoft Excel spreadsheets and End Note. This resulted in the elements that were proposed by the first process of data extraction studies which entailed research domain, theories, research methods, and social media activities [11]. Paper Title, Authors, Publication Year, Type of Paper, Digital Database, National/Regional, Contexts/Area, Research Methodology, and Research Domain were the columns included in the collected data for all sources in this study.

A. OVERVIEW OF PUBLICATION SOURCES

The 116 primary studies that were considered relevant to this study following the SLR were published with the intent of treating them as reference points for our subject. The primary studies were evaluated against inclusion and exclusion criteria that required the qualitative examination of the findings. A total of 69 journal articles and 47 conference papers were drawn out. Journal papers dominated among the publications. Figure 3 represents this finding.

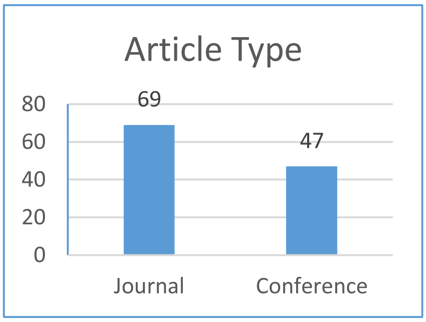

FIGURE 3. **Article type.**

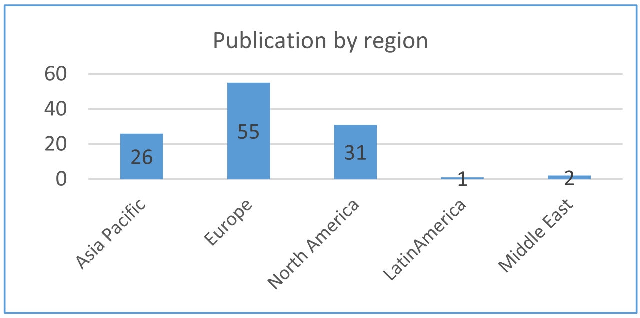

FIGURE 4. **Articles per region.**

B. COVERAGE OF RESEARCH REGIONS

The primary studies emanated from 28 different countries spread out in 5 regions, indicating a relatively wide geographical coverage. Figure 4 shows that Europe leads with 55 articles, followed by North America with 31 articles, and then the Asia Pacific region with 26 articles. The Middle East had published only two articles, whereas Latin America had published only one. Together, Europe, North America, and the Asia Pacific region constituted nearly 97% of the total studies.

The USA took the lead in the number of publications, with 28 articles, accounting for nearly 24% of the total number. UK follows with 21 articles, accounting for 18% of the total. Netherlands ranked third with 17 articles. These three countries together represent 56.9% of the total number of studies, as seen in Figure 5. Note here that this geographical classification is based on the first author's affiliation country.

V. RESULTS BY RESEARCH QUESTION

As seen in Section III we defined some research questions that needed to be addressed in our SLR. In this section, we respond to all the research questions in detail. Each question is presented under a separate subsection.

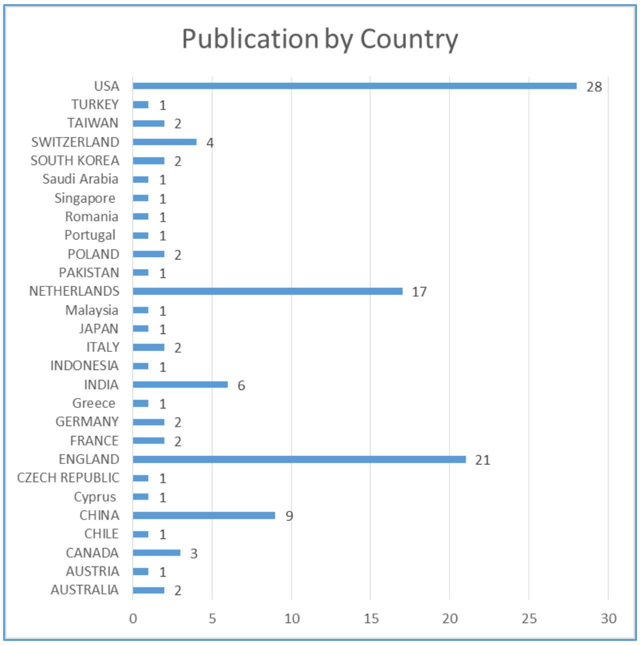

FIGURE 5. Quantity of publications per country.

A. RQ1: WHAT ARE THE MOST IMPORTANT FEATURES FOR MEASURING REPUTATION AND INFLUENCE?

Measuring reputation in OSNs can include a myriad of features and dimensions. Some of these are presented here.

1) INFLUENTIAL USERS

Several researchers have focused primarily on the number of responses elicited from particular social media posts. However, Sun and Ng [40] sought to measure the reputation or influence of the positive or negative sentiment of a post. They contrasted public and private posts and influencers and influence. Oro *et al.* [41] proposed a method to exploit the contents of the messages posted by influencers to express their opinions on various things. They modeled these contents with a three-layer network in which the layers represented users, items, and keywords, along with intra-layer interactions among the actors in the same layer.

Daniel *et al.* [42] attempted to leverage the power of OSNs to improve the accuracy of the prediction of stock value. The authors developed an algorithm to measure the popularity of company-related events on Twitter and turned this information into relevant insights with predictive value. This task was complicated by the fact that Twitter contains a lot of noise, so it was necessary to filter out irrelevant messages and focus only on those that actually are related to the financial markets. They used sentiment analysis to detect the most influential events and captured public reactions to them.

2) USER'S SENTIMENTALITY AS A FEATURE

Very few works have considered the user's feelings while evaluating their credibility as a source. This element is very significant particularly when the user is inclined to a certain demographic. Sentiment describes the degree to which a user's tweet is antagonistic, or affects social relations, or other users' psychological states, and their orientations toward a particular theme. Alrubaian *et al.* [43] applied a reputation-based technique to each user's profile and introduced a sentiment score based on the user's history. The assessment of the connection structure and the user's feelings on the OSN were mixed to define and assess the topically appropriate and reliable sources of data on the OSN. They measured the reputation or credibility of a user on Twitter based on how popular he/she was, and how sentimental he/she was toward a particular topic. They introduced a sentiment score for a user that was calculated using the following formula, where T+ represents the user's positive tweets and T- represents the user's negative tweets.

$$\Delta_{u_i} = \frac{\sum T_{ui}^{+}}{\sum Tu_i^{+} + \sum |T_{ui}^{-}|}$$

Then, they calculated the reputation of the user with respect to a given topic using the following formula:

$$R^p_{(u_i)} = \Delta_{u_i} \times I^p_{(u_i)} + (1 - \Delta_{ui}) \times \omega^p(u_i)$$

where $I^p_{(u_i)}$ is the activity and $\omega^p(u_i)$ is the influence of the user u_i on a given topic P.

3) USER'S INTERESTS AS A FEATURE

Shi *et al.* [44] examined the gradual changes in users' interests throughout the trajectory of a significant event. Their model tackled the data sparsity issue and utilized short text retrieved from the microblogging network to model the data. The researchers introduced a cosine measure similarity detection method to discover user interest communities. An automatic hot event filtering algorithm was designed to completely remove the influence of general events and increase the efficiency of the mining event. Analyzing and interpreting the data acquired from microblogging networks can be difficult since most posts include a limited amount of text and are usually not posted in a coherent time sequence. Thus, they created a long document and included all short texts from the posts analyzed with the aim of discovering the essential keywords for each hot topic. This way, the authors solved the short text sparsity issue, increased the model's learning ability, and improved its topic analysis quality. The cosine measurement and user interest discovering methods they employed helped increase the efficiency and accuracy of their evaluation of the event.

Focusing on the drawbacks associated with topic detection on Twitter, Cigarrán *et al.* [45] proposed a novel methodology known as Formal Concept Analysis (FCA). This technique organized content by drawing upon shared terms and creating thematic concepts to operate as topics. FCA is a mathematical technique used to form concepts using lattices and ordered sets of theories to organize information and identify relationships. At the heart of this model is a formal concept derived from a formal context, which comprises objects, attributes, and the relationships between the objects and the attributes. A formal concept is thus a pair (A, B) with A being the set of objects and B being the set of relationships. A is otherwise referred to as the extent, while B is the intent. The author applies this concept by taking tweets as objects and using their terms as attributes. Every tweet in the same formal concept shares terms and are thus related. This will create concept lattices whose upper parts will be made of generic topics for a given set of tweets while the lower parts will be more specific.

Momeni *et al.* [46] employed a review-based approach that surveyed the existing approaches to assessment and ranking user-generated content (UGC). They used two methods for these human-centered and machine-centered techniques. The former relied on the end user, a group of end users, or a platform designer to use a system and create the default rankings. The latter used machine learning algorithms to rank and assess UGC. On the basis of these, the authors classified frameworks to assess and rank the UGC into four groups: community-based, end-user-based, designer-based, and hybrid frameworks.

The community-based framework uses either machine-based or human-based approaches to rank and assess UGC based on group preferences, with a well-defined value; for example, separating useful and non-useful comments on platforms such as YouTube. End-user-based frameworks are designed to accommodate individual differences in the evaluation and ranking of the UGC via human-centered or mechanical techniques so that individual consumers can explore the content, define their own concept, or communicate in accordance with preferences expressed by the scheme to amend their rankings, ratings, implicit and detailed behavior. The designer-based framework uses a software designer's encoded values to assess and rank UGC. The hybrid framework does not fall explicitly into the groups mentioned above as it is usually a blend of different approaches in the way it works. A good example is a system that learns from community traits to develop computational systems to study an end user [46].

4) EXTENT OF ACTIVITY, TWEETING BEHAVIOR, AUTHENTICITY, AND SPEED OF GETTING RESPONSES AS FEATURES

Alp and Ötğüdücü [47] proposed a method that included several preparatory steps to conduct a topical analysis of influence. First, they collected relevant data on 20 high-profile Twitter users on different topics, focusing exclusively on those who tweeted in Turkish. Next, they identified the topic in each tweet. Since the same root word can assume many different forms in Turkish, it became necessary to use an NLP tool built specifically for this language and optimized to extract topics from small volumes of text. Human experts reviewed the output of this tool and shaped the final division of tweets into topics. User modeling came next. It required the calculation of several features for each user based on publicly available data. The most important features were i) focus rates that defined the narrow specialization for a small number of topics; ii) extent of activity, expressed through the number of active days and the average number of tweets per active day on a given topic; iii) authenticity, which was calculated as the number of original (non-retweeted) posts divided by the total number of posts; and iv) the speed of getting reactions, understood as the average time before a follower reacted to the user's message. While the first three had a positive correlation with perceived influence, the last was assumed to have a negative correlation (i.e., higher value indicated lower influence).

The algorithm for calculating user influence, named Personalized Page Rank (PPR) accounted for both nodal and user-based features. Map and reduce functions similar to Hadoop Map-Reduce framework were applied to execute distributed processing. In the resultant matrix, the top layer contained initial user weights obtained from the user modeling stage. The second layer included mappers that determined the contributions of individual users, with contributions originating from the user's followers counted along with one's own. Finally, the bottom layer employed reducers to find new ranks for each user based on additional information acquired through processing. After a large number of iterations, the algorithm could separate nodes that seemingly had the same amount of influence with followers, balancing weights with both full and partial contributions. The cycle could be executed quickly because of the distributed model, which practically meant that PPR algorithms can be used as a near real-time tool to determine high authority users on social networks. Zhou *et al.* [48] used interactional information behaviors among communities on Twitter together with UGCs as the features for formulating the model.

5) RELATIONSHIPS

Relationships are a critical element of OSNs. Li *et al.* [49] developed a measurement tool called CASINO to measure influence and conformity in social interactions across different topics. CASINO extracts a set of topic-based social interactions on an OSN, labels the edges (relationships) among individuals with positive or negative signs, and computes the influence and conformity indices of the sample [49]. Cataldi *et al.* [50] proposed a system to explore the changes in user relationships with respect to a specific area of discussion.

6) CONNECTIONS AND CENTRALITY

Lu and Wan [71] proposed a method to determine the centrality of individual users on a microblogging network named WeiboRank, and several key conditions had to be met to account for the nature of the platforms analyzed. First, the relationships among users are one-directional, which means that it is always clear which user is a follower and which is located "more centrally." This relationship is typically represented with a pointed arrow, while users are represented as individual nodes, and the entire network is formulated as an interconnected spreading matrix. When the initiating user creates a new message (i.e., tweet), the followers have a choice to read and republish it, or to ignore it if it does not attract their attention.

Two important model parameters must be defined to successfully determine the value of WeiboRank. These are the direct capacity to influence other users, which is denoted as F1(v), and the region of potential influence, which is denoted as R. Based on the distance between the two nodes, a layer marked N1 is created for every user. It contains nodes that are away from the center with a distance of one or less. These nodes obviously represent the direct contacts and delineate the direct capacity for influencing. The next N2 layer contains nodes, each with a distance of two. These nodes are connected to the center through direct followers. A repetition of this process leads to the creation of Nm, a layer containing the most distant followers within the user's network. The user's region of influence can be defined as the sum of the number of nodes in each layer, representing the probability that all of his followers, no matter how distant, will see his tweets on their feeds if they choose to do so. Finally, those two measures are combined to acquire the Weibo Rank (WR) value, which is a single output that can be used to compare users based on their influence. This variable can be calculated using the following formula:

$$\mathrm{WR(v)} = \mathrm{F1(v)} \cdot \sum\nolimits_{jw1}^{R} /\mathrm{Rja}$$

for any given user v based only on public information about his direct and indirect connections. It can be said that WR captures the extent of the average information load and expresses it in a practical format that refers to the network in a global sense. Higher values of WR are indicative of a more influential user; thus ranking can be carried out quickly and intuitively while fully appreciating the complex spread patterns of information through the microblog network. Loucif *et al.* [51] blended both the explicit and implicit features of Twitter to build the proposed model. The explicit features were replies, retweets, and mentions. The implicit features were similarity in user interest and user connections.

7) NODE ATTRIBUTES AND INTERACTIONS AS FEATURES

Yu *et al.* [52] created a directed weighted information interactive network to calculate node attributes and interaction influence. They incorporated the time dimension to construct a Weibo user influence computing model. The users of Weibo were referred to as nodes, while the inter-user relationship was defined as an edge. The researchers defined the user relationship within the Weibo system as UR = (V, E, P, W) where V indicated the user node set, E stood for inter-user directional edge set, P denoted the node weight set, and W designated the directional edge weight set. Weibo's unique network topology structure only shows the potential user information dissemination ability and not the actual user influence or their ability to spread information. Yu *et al.* [52] and Alp and Ötğüdücü [47] explored posting and forwarding behavior and used it to create an interactive information network. Dugué *et al.* [53] identified the role of nodes in relation to communities and their work was based on extending Guimera and Amaral [54], who coined two complementary measures that placed every node in the social network in a 2D role space. The original work proposed a number of thresholds to discretize the subspace, with each such space corresponding to a given role. Two role measures were used to define the role of a node within its community: the first was known as within-module degree based on the concept of z-score, and the second was a participation coefficient that analyzed the connections that the node had with the community. Dugué *et al.* [53] identified limitations in the original approach and proposed new measures to reduce the effects of these limitations. They introduced three new measures in place of the participation coefficient used in the original approach, namely diversity, intensity, and heterogeneity. They also added another measure to evaluate internal connectivity. All these measures were expressed as z-scores that can adapt to different community sizes. Diversity evaluates the communities to which a given node is connected with respect to others in its community. External Intensity measures the number of links a node has with communities aside from its own, and with respect to the nodes in its own community. Heterogeneity of a given node is a measure of the number of links that the node has from one community to another. Internal Intensity represents the internal connectivity of a given node. The author used an automatic method by leveraging unsupervised classification to define roles in the community. The measures defined were processed and clustered, where each cluster identified in an 8-dimensional role space was considered as a community role.

del Campo-Ávila *et al.* [55] analyzed three of the most popular tools used to measure influence on Twitter: Klout, PeerIndex, and TwitterGrader. The authors used data mining techniques on a Twitter dataset containing a sample of top Tweeters, well-known companies, and categories on Twitter. The authors were more vested in dual approximation, from both qualitative and quantitative perspectives. For quantitative purposes, regression was the most palatable option, bearing in mind that the information from the tools was numerical. For the purposes of making things simpler for the Tweeter, it was also vital to reduce the numeric values to discrete levels of influence (low, medium, and high). Classification algorithms were used for this purpose. To represent the data, the author used a model that induced decision trees as they were easy to consume and could also be translated easily into action. The author used regression (REPTREE) and M5P and classification (C4.5) algorithms implemented in Weka. These algorithms used the variance reduction and information gain ratio as the criteria to pick the most relevant traits to expand the tree.

8) RESPONSE AND POSTS (REPLIES, RETWEETS, AND MENTIONS) AS A FEATURE

Einwiller and Steilen [56] analyzed OSN pages of some of the largest companies in the US with the aim of identifying how well they employed the possibilities provided by social media to handle complaints submitted by dissatisfied customers. Social media platforms such as Twitter and Facebook enable companies to restore a complainant's trust and persuade observers of the online correspondence that they have their best interests in mind. The study was based on a number of premises that were drawn from the works of some of the leading experts in the field of corporate complaint handling. The authors began with the assumption that defensive strategies deflected responsibility for the crisis while accommodative strategies were more compliant and comprised corrective action. The authors selected 34 entries from the Forbes Global 2000 list with the aim of analyzing their Facebook and Twitter accounts thoroughly. They analyzed various complaints expressed in the form of tweets and Facebook posts, responses, and all other correspondence between the complainant and the company. The sample tested over a one-month period included 5023 complaints, 4154 corporate responses, 3335 follow-up posts shared by complainants, and 2534 posts by regular Facebook and Twitter users. H-index has been used for over a decade as a reliable measure of authority for scientists and other influential people. Cerchiello *et al.* [57] decided that the definition of the H-index has to be suitable for the measurements it would serve to enable; therefore, the formulation was developed specifically for financial markets. They used parameters that were harvested directly from Twitter data, such as the volume of tweets and the related retweet vector. Given that Twitter data are of a discrete nature, it also became necessary to introduce statistical backgrounds and to allow it to be used for inferential rather than just descriptive purposes.

Random variables that represent expressions of retweet activity can be summarized as $X1, \ldots Xn$, and they collectively form function F. Assuming that F is more or less continuous, it becomes possible to come up with a full statistical definition of h-index in the form $h: 1 - F(h) = h/n$. After performing order statistics with the goal of adjusting the model to the nature of data, the formula can be expressed in the following way: $h = \max \{t: X(t) \geq t\}$. This operation accounts for the fact that stochastic variability is considerably greater for retweets than for scientific papers. Thus, it is feasible to derive confidence intervals that are large enough to enable the evaluation of individual tweet authors.

Retweet vector X, which is based on a raw number of retweets, plays a significant role in calculating the index. This can be seen as a setting in which each incident of retweeting has a low probability of occurring, making it logical to use loss data modeling for the statistical processing of the vector. Within the Loss Data Approach (LDA) technique, any loss is described by two factors: frequency and severity. Frequency can be defined as the number of events within a certain period, and severity represents their cumulative impact. After completing the operations described above, the authors were able to generate the distribution of retweets for each user, described as $k(x1)$ and understood as a part of function Ci. The distribution for the volume of tweets per author was also calculated and expressed as $p(n1)$ and was combined with the other parameter through binomial convolution. By combining the likelihoods for each factor, the authors were able to calculate the quantity $\lambda i1$, which represents the reliability estimation for each individual Twitter author. This variable is the summarized expression of confidence in a particular author based on his previous activity on the social network platform.

Servi and Elson [58] presented an approach that entailed two steps: using a program to analyze social media text and then using a mathematical algorithm to detect emotional shifts in the resulting dataset. For the first step, a program known as "Linguistic Inquiry and Word Count" (LIWC) was used. This program has been used extensively for content analysis tasks (college writing samples, science articles, blogs, novels, speeches, and newspapers), a feat that made it easy to use. From the results produced by LIWC, the mathematical algorithm can detect changes in emotion through breakpoint analysis. The points where emotions shift substantially are known as breakpoints. Thus, the mathematical algorithm is tailored to estimate these points, coupled with the trends in emotions between the points. The algorithm can be used to gauge influence by detecting situations in which a user on Twitter posted a tweet that had so much influence that it caused a breakpoint in the levels of an emotion expressed, whether positive or negative. Using the breakpoint, it is possible to understand the timeframe to analyze for the possibility that a single user created a breakpoint especially if the tweet was retweeted. The algorithm can also be deployed on retweets.

The model proposed by Loucif *et al.* [51] used commonly known explicit features on Twitter, such as tweets, replies, retweets, and mentions. The model also used an implicit feature that can be inferred from the tweets posted by a given user on Twitter. The author narrowed down the approaches to those that can be applied to small texts, or rather, when the sample of text is limited, as in the case of a tweet. Some approaches of interest here include those employed by Rocha *et al*. and Anderson *et al.* [60], [161], who used character-level statistics to find information in small text content. The work evaluated the structural traits of messages, such as finding items specific to greetings, farewells, and signatures in the content. Koppel *et al.* [59] focused on analyzing error patterns in emails, such as fragmented sentences, wrong tenses, and inversion of letters [60]. Sanderson and Guenter [161] used a method known as Author Unmasking, which postulates that the difference between two texts from the same authors will be reflected in a limited number of features. They reviewed other closely related approaches such as sequence-based approaches of character-level Markov Chains and character-level sequence kernels, source code authorship profiling, and functional n-gram among others.

Momeni *et al.* [46] noted the importance of UGC in the context of improving systems such as search, retrieval, and recommendation systems. It is particularly useful in dissecting and finding useful information on user opinions online, where it has grown quite popular especially with the growth of OSNs.

9) CONTEXT

Context is often challenging to gauge within social media posts. Reyes *et al.* [69] developed a model that operated with a system of features, wherein each feature can be analyzed separately in terms of information gain to determine its individual contributions to the overall discrimination power of the system. Their model is organized according to four types of conceptual features, namely signatures, unexpectedness, style, and emotional scenarios to detect the irony within a post on an OSN [69]. Khrabrov and Cybenko [70] produced a metric system to find people of growing influence on OSNs based purely on the structure and dynamics of their conversations. They understood context to be a high-intensity ecosystem with its own "mind economy," adapting to maximize the participants' rankings and to promote their shared message [70]. External out-of-network sources such as the mainstream media can impact the ecosystem [67]. Thus, Meyers *et al.* [67] presented a model in which information can reach a node through the links of the OSN or through the influence of external sources.

10) COMMON INTEREST AMONG USERS AS A FEATURE

Ji *et al.* [68] used the dataset collected from the Epinions social network to study the effects of common interests on trust formation among online users. They pinpointed four patterns of online activities such as writing, commenting on, rating reviews, and expressing their trust to other users on the platform. The reviews that the two users u and v commented on in this study were considered as a unit through which their common interests were measured. They were used to calculate the overlap rate with the following formula:

$$\rho_{uv}(t) = \sum\nolimits_{i=1}^{t} n(u, v, i)/k(u, t)[20]$$

The overlap rates were calculated before and after trust formation so that the impact of common interests could be measured separately for each case. Two null models were also presented and compared with the empirical results obtained through the overlap rate calculation to prove that random temporal activities of users have no effect on trust formation. The researchers divided the dataset into smaller groups of users that varied in size from 100 to 1000, in order to get more precise user interest overlap rates. They discovered differences in overlap rates before and after the trust relation was formed. These results are only valid when user comments are considered. However, things changed significantly when user tastes are considered. This is why the authors utilized the Pearson Correlation Coefficient (PCC) to determine the similarity in tastes for a pair of users before and after the creation of a trust relationship. The ratings given by users in the Epinions dataset, that go from 1) Not Helpful to 5) Most Helpful were used to calculate the correlation between the rating vectors for each pair of users. The authors came up with the following equation:

$$\theta(t_c) = \frac{1}{E}\sum\nolimits_{u=1}^{N}\sum\nolimits_{v=a}^{N}\theta uv(tc)\ [20]$$

Comment patterns can play a large role in trust formation, especially when users of different degrees are considered. The authors investigated the correlation between the user degree K_u and the taste similarity θ with the aim of determining the significance of user degree on formation of trust relationships. Users were divided into 16 groups, so that the authors could capture the average taste similarities for different user degrees. The exact moment when a trust relationship was formed was also considered, since taste similarity or overlap rates of users can change in the long run. The method used in this paper measured the overlap rate and taste similarity and the user degree to determine the role of common interests in online user trust formation.

Zhou *et al.* [48] presented the dynamically socialized user networking (DSUN) model that focuses on analyzing potential and dynamic user correlations based on the similarities in topic awareness and behavioral influence. Similarity-based relationships were calculated by fetching time-varying features and computing similarities in topic awareness. To extract features of topic awareness, the author introduced two types of interests: transient and durative. The former describes user interests that change over time while the latter describes user interests that remain constant throughout. The authors used the associative ripple approach to calculate similarities in topic awareness. Influence-based relationships were calculated by evaluating user interactional behavioral traits. The authors used social tags posted along with content on OSNs. For example, the "@" tag on Twitter can help show the influencing behavior or can show that user A has a connection with user B. The social tag "RT @name" is used to show influenced behavior or that user A has received information from user B.

Ma *et al.* [61] presented an algorithm known as MapMe. The first step in this algorithm is to divide a big graph, which is a characteristic feature on most OSNs, into smaller subgraphs. This is done using the spectral method. The subgraphs of the OSNs are then matched based on the similarity, and the node pairs of the subgraphs are mapped. These node factors were mapped in both profile-based and network structure-based matching models. Profile-based matching was done using the Doc2vec method, which represents a document as a vector, and fetches features from the user profile. From the network similarity perspective, MapMe uses a number of techniques such as k depth degree, user's clustering coefficient, and User's Eigenvector Centrality. Profile-based and network structure-based similarities are then balanced bearing in mind that neither approach is perfect in estimating the similarity between the nodes in different OSNs. Balancing entails tweaking a parameter α, known as a smooth factor, which belongs to the [0, 1] set. The author found the optimal value for this by tuning the value in an experimental setup. The author used k depth degree, the user's clustering coefficient and User's Eigenvector Centrality to measure the similarities in the network structure.

11) NEIGHBOR ACTIVITY AS A FEATURE

By discovering the role of the node's neighbors in spreading influence and by tracking their activity history, Sheikhahmadi *et al.* [62] developed a more accurate method of identifying influential users on OSNs. Their two-level method ranks the nodes based on the neighbors' roles and the strength of the relationship between the node and its neighbor. A network graph is partitioned into communities that comprise sets of nodes that have both edges in the set, by using an algorithm designed to extract community structures from large networks quickly. The process starts with an N-node network and an initial number of communities that is equal to the number of nodes in the network. Afterward, neighbors (w) of node (v) are considered, and after calculations they are removed from their original communities and are added to the new community (w). The node (v) is placed into the community that has a maximum calculated modularity value. Once the network graph is partitioned into communities, the newly generated clusters are weighted in accordance with the action log table. Low influence nodes are removed, and the activities of the remaining nodes and their neighbors are extracted based on the numbers of likes, comments, and shares. However, as the impacts of these likes, comments, and shares on the process of information spreading are not the same, each action was ranked differently. The time of the activity during which the node's followers interacted with a post was also measured, since influential posts tend to attract attention over prolonged periods of time. By classifying the communities, the researchers determined the role of the node's neighbors in spreading messages. The study shows that the delayed reaction to a user's action can reveal the level of influence a tested user has and this leads to a higher level of accuracy in detecting influential neighbors. A semi-local method was employed to trace the spreading influence of a node [55].

B. RQ2: WHY IS MEASURING THE REPUTATION AND INFLUENCE ON TWITTER RELEVANT?

Social science researchers believe that Twitter is an important lens that can be used to view communication and perception. Thus, Twitter has been increasingly evaluated in terms of sentiment analysis, data application, and influence. However, these components of social media touch upon the over-saturation of media messaging rather mildly. As Chang *et al.* [63] noted, Twitter users can become overwhelmed by the abundance of noise and simply disengage, and may miss relevant information as a result.

1) SENTIMENT ANALYSIS

Sentiment analysis is applicable to marketing, public relations, and consumerism. It is thus important to be able to trace an exchange of information within a social network and understand the influence that social media users have on the broader community [50]. This infiltration of social media on perception has marketers seeking to utilize not only analytics to inform their campaigns and initiatives but also sentiment analysis. Bae and Lee [64] proposed a measure of positive-negative influence on Twitter to be used as an indicator for identifying a real-world audience and their sentiments, thus not only providing new insights on influence but also facilitating a better understanding of popular users.

Sentiment analysis can also benefit the public relations sector. The need to monitor reputation online is essential because real-time online opinions and comments are key to understanding how an organization should manage its external communications [65]. Yet, as Spina *et al.* [65] explained, existing processes are time consuming and expensive.

Sentiment can not only impact perception and credibility but also influence how people react to such stimuli. Sun and Ng [40] asserted that the sentiment of social media posts can serve as an impetus to consumer behavior and purchasing habits. Influential people posting on social media can influence others to either comply with the intent behind a given social media post or oppose its purpose, which can be detrimental to sales [40]. Ruan *et al.* [66] contended that the public mood or sentiment expressed on social media is directly related to the ups and downs in the financial markets. However, as in the case of most innovations, this inherent power of social media sentiment can also be abused [40]. Meyers *et al.* [67] said that it is important to delineate between tweets that expose people to information versus those that infect people with information in order to pinpoint and thus model influence within social networks. There is little commonality and consistency across social media posts, which makes it difficult to not only navigate them but also predict the directions they will follow. One such characteristic that is challenging to detect is irony, as explored in [69].

2) DATA APPLICATION

Another tool for measuring social media reputation and influence is the application of newly acquired data. Ho *et al.* [72] believed that ideas and inference can be propagated even more powerfully online and specifically through microblogs rather than through traditional channels. The researchers contended that being able to measure the level of information propagation in a social network can enhance research on social science significantly [72]. Li *et al.* [49] agreed that OSNs provide a unique opportunity to analyze a wealth of data. By identifying the ''influential'' users on an OSN, users may be able to maximize the influence of a piece of information for gains such as advancing marketing, enhancing public relations, or encouraging consumerism [49]. Li *et al.* [73] asserted that the most notable application of predicting user influence is viral marketing [73].

3) ONLINE INFLUENCE

If online influence can be tracked and measured, it can also be applied to the enhancement of online reputation. Khrabrov and Cybenko [70] devised a metric system to enable individuals to track changes and dynamics of their own and their peers' influence online. They suggested that competitive intelligence analysts can analyze the dynamics of consumer opinions about their products and services with a deeper understanding of who plays a key role in influencing attitudes and opinions to inform corporate decisions [70]. They said that an organization can also improve its analysis of its own evolution and changing roles by using the metric system they proposed [70]. Subbian and Melville [74] noted that engaging with influential social media authors is critical for corporations since such influencers have a reach that is both far and quick [74].

Despite the inherent advantages of capturing social media influencers, the debate continues on whether these authors are actually capturing the full scope of the audience. As Zhaoyun *et al.* [75], [76] asserted, having many followers on Twitter is perceived as a means of introducing diversity into various conversations; however not all users are equally influential. Stewart [77] took this concept a step further by evaluating the ways in which influence is developed, circulated, and understood among scholars. Existing research has failed to consider the means of academic influence outside peer-reviewed publications [77].

Oro *et al.* [41] took a different approach, by tackling ways to better understand the popularity of the topics and attitudes through social media and by identifying influencers online. Agah and Dasari [78] explained that understanding the impact of OSNs on online influence can help organizations narrow down their digital marketing strategies. The use of social media influencers and local proximity, targeted advertising, trend analysis, news, and events can be leveraged [79].

4) TO DISCOVER THE CENTRAL NODES

Several researchers have tried to identify how central nodes can be discovered in large-scale OSNs, particularly in microblogging platforms. One of the earliest attempts was made by Kwak *et al.* [80], who used the number of followers as the primary criterion. Zhang *et al.* [81] took the effort to the next level by providing a formal definition of localized social influence while also explaining typical behavior of Twitter users with respect to influence. Kitsak *et al.* [82] proposed a different approach and searched for the deepest nodes located in the inner network by using the K-shell decomposition method. Kang *et al.* [83] postulated a new model based on diffusion centrality that relied on both semantic information and analytic estimates. However, this solution is very demanding in terms of computational power. Gao *et al.* [84] provided an alternative centrality measure that is only suitable for undirected networks and cannot be applied on microblogging websites such as Twitter or Weibo. Lu and Wan [71] were inspired by all the aforementioned studies and developed a ranking method to overcome the inherent limitations of the existing methods. They used the PageRank method to evaluate webpages based on the number of inbound links leading to it.

Previous works in the area of determining online influencers can be divided into two major groups. Non-graph solutions ignore the structure of followers and focus on other types of information. For example, Cha *et al.* [85] tried to extract influence based on variables such as the total number of followers and tweets. Another solution from this group was proposed by Pal and Counts [86] who focused on the Gaussian clustering method and leveraging parameters such as the originality of tweets and topics of interest. On the contrary, graph-based approaches try to use the structure of the network as a source of information on the user's authority in the network. Some of the classic methods used for this purpose include Independent Cascades (IC) and Linear Threshold (LT), both of which are based on the classification of network nodes according to activity level. However, determining accurate parameters for such methods can present a significant challenge, and this limits their practical value. Long and Freese [87] attempted to overcome this limitation through the introduction of an algorithm that keeps the number of seeds at a minimum. Romero *et al.* [88] argued that influence is best defined as a combination of activity and popularity, developing hybrid models that take both network topology and content-related factors into account.

Influence has become a vital component that has grown alongside the rising importance of OSNs, particularly from business and political perspectives. Every marketer and politician want their posts to reach a wide audience. One of the ways to spread messages via Twitter is to rely on influencers. Identifying these influencers is still a challenge for many researchers, and Loucif *et al.* [51] presented a method of measuring influence based on structural, interactional, and semantic dimensions of the social network.

Most tools used to calculate influence on Twitter have reduced this task to a single number. This is not an easy task. These efforts have not resulted in positive effects in response to user needs. Users do not have access to the processes involved in calculating their scores. Thus, they cannot tweak aspects of their social interactions to improve their influence. del Campo-Ávila *et al.* [55] presented an interesting study to bridge the gap between the tools and the users, thus helping users on Twitter understand how to enhance their influence by using these tools.

5) TO DETERMINE THE INTERPERSONAL TRUST AMONG USERS AND CORRELATIONS

OSNs like Facebook and Twitter help their users establish relationships easily based on mutual interests. A significant amount of research has already focused on this topic. Dodds *et al.* [89] studied the correlations between users' interests and formation of contacts. Massa and Avesani [90] used local trust matrix techniques to study the effect of the initial interpersonal trust relations among users on predictions of trust formation. Lewis *et al.* [91] attempted to examine the relations between peer influence, social homophily, and the formation of relationships. Most papers in this field focus on specific groups such as students or drug users that are more likely to share the same sets of values, cultures, or behavioral norms. However, social media also enables users to make new friends because users can discover things they have in common with other users. Such systems are known as Coupled Social Networks or (CSN) and they are essential for the Recommender System analysis. Nie *et al.* [92] introduced a hybrid algorithm that relies on a two-layer network, where the first layer analyzes how users relate to each other and the second reveals how users relate to different items. Zhou *et al.* [48] presented a model to analyze potential and dynamic user correlations to construct dynamic user networking. They looked at inspiring not just collaborations among users on OSNs but also facilitating the flow of information among communities.

6) TOPIC DETECTION

Topic detection on both social media and online is important as it is gaining traction among companies that want to mine opinions on their brands, products, and/or services offered online. This work is thus important both in terms of making improvements in how topics are detected on Twitter and in enhancing the accuracy in tracking and managing user opinions.

7) TO DETECT SOCIAL CAPITALISTS

Social capitalists are not healthy for Twitter because they are more or less not interested in the content generated by the users they follow. This behavior is commonly exhibited by spammers on Twitter, and makes it hard for genuine users to find genuine content. Finding social capitalists on Twitter is a means to improve the integrity of the platform and forms a basis for designing better rules and search engines [53].

8) TO IDENTIFY SHIFTS IN EMOTIONS AMONG SOCIAL MEDIA USERS

Servi and Elson [58] presented a new way of defining and measuring influence on OSNs like Twitter, which comprised a quantitative technique that uses both text processing and a mathematical algorithm to identify trends in both emotion and influence. They sought to extend the work done previously in terms of quantifying influence on OSNs. The main target is to conceptualize influence and measure it quantitatively. This is a shift from previous techniques which uses metrics such as directed links, retweets, and mentions to measure influence on platforms like Twitter.

9) TO IDENTIFY AUTHORS OF POSTS

Social media forensics support the investigation of crime by enabling author attribution and identification. This is gaining importance especially because malicious users are growing in number as a result of the opportunity to anonymize activity online. There is thus a necessity for means to help narrow down the search for authors of malicious content and those propagating misinformation or trolling campaigns on Twitter. Automatic techniques to find these types of users through stylometry becomes necessary as a solution to this problem [60].

Given the rising popularity of OSNs, it is common for users to create profiles on more than one platform simultaneously. Owing to this, identifying a user across platforms offers several opportunities to identify users, study migration patterns, estimate influence, and find experts on OSNs. Existing solutions in this domain have mostly focused on network topology. Ma *et al.* [61] aimed at extending this through a novel solution known as MapMe, a tool that factors in the user profile and social network structure to improve the accuracy in mapping users across different OSNs. Suppose a user has accounts on three different OSNs, say Aminer, Sina Weibo, and ScienceNet. The connection of a single user on these networks is known as the anchor link, and refers to a pair of accounts that belong to the same user. Identifying anchor links accurately across multiple OSNs helps identify users, detect patterns in migration, estimate levels of influence, and find experts.

C. RQ3: WHAT ARE THE MEASUREMENTS, METRICS, AND METHODOLOGIES THAT HAVE BEEN USED TO MEASURE REPUTATION AND INFLUENCE ON OSNs? WHY ARE THEY INSUFFICIENT?

The literature shows that a variety of strategies have been used to measure reputation and influence on OSNs. These strategies are discussed in this section.

1) MODEL DEVELOPMENT

Social media researchers have generated several models to measure reputation and influence. Reyes *et al.* [69] proposed a model to capture both low-level and high-level properties of textual irony. Sun and Ng [40] built a dictionary of emoticons used on popular OSNs in both English and Chinese, and categorized them into five groups: happy, tense, sad, angry, and fatigue, to determine the emotion of a post. Li *et al.* [49] proposed a model to evaluate conformity within a social group. Khrabrov and Cybenko [70] presented techniques to track all those who have influence on an OSN. Cataldi *et al.* [50] adopted a method for categorizing the content shared by the community they studied. Spina *et al.* [65] suggested a model combining two tasks, namely learning tweet similarity and applying a clustering algorithm using the confidence of the classifier as a measure of similarity. Agah and Dasari [78] used a customized divide-and-conquer algorithm to measure the influence of each of these nodes using a linear function. Zhaoyun *et al.* [75], [76] offered a new method to determine the transition probabilities of uncertain relational networks. Li *et al.* [73] identified three dimensions of social influence (Monomorphism vs. Polymorphism, High Latency vs. Low Latency, and Information Inventor vs. Information Spreader), introduced existing influence models, proposed an information diffusion model, demonstrated experimental results, and evaluated the prediction power of their proposed information diffusion model [73].

2) EXPERIMENTAL INVESTIGATION

Experimental investigation is one of the longest lasting traditions in social science research and has been used to study and measure reputation and influence online. Meyers *et al.* [67]

experimented by testing their algorithm against common sense baselines to examine the retweeting of URLs. They used different combinations of network topologies, exposure curves, event profiles, internal hazard functions, and synthetic and real data to measure retweets. By exploiting multilinear algebra, Oro *et al.* [41] presented a method to extract the most active users stating their points of view on dominant themes tagged with dominant keywords. They conducted a series of experiments on different real-world datasets collected from Twitter and Yelp Social Networks on different issues [41]. Wei *et al.* [79] evaluated influential Twitter users in Boston, Bristol, and Seattle, to calculate unique pairs of latitude/longitude coordinates for users with multiple pairs of geographical coordinates [79]. Ho *et al.* [72] collected Plurk data to create propagation trees that they constructed for each relevant user and then displayed the information propagated by the top-ranked users showing the top five micro-bloggers for each measure. Subbian and Melville [74] compared a Supervised Kemeny Ranking to various techniques to predict the virility potential of tweets surrounding the controversial Pepsi app. The authors combined several algorithms to make a robust measurement tool and then applied the tool to the Twitter dataset [74]. Bae *et al.* [64] distinguished between the positive and negative audiences of popular users with the help of three correlation methods, namely Pearson correlation analysis, Spearman rank correlation analysis, and Granger causality analysis. Chang *et al.* [63] evaluated celebrity-initiated tweets that had the potential to lead to large context trees. The authors manually selected 10 Twitter context trees to compare 7 text-based summarization methods, namely Centroid, SimToRoot, Linear, Mead, LexRank, SVD, and GBDT, using four different types of user interaction features, namely ContentOnly, ContentAttribute, AllNoGranger, and All [63]. Ruan *et al.* [66] used their trust network-based reputation mechanism, which amplifies the correlation between a specific firm's Twitter sentiment valence and the firm's stock abnormal returns to test the possible auto-correlation property of abnormal stock returns and the relationship between Twitter sentiment valence and abnormal stock returns.

3) OBSERVATION

Observation is also a valid method of studying social media influence. Stewart [77] conducted a qualitative ethnographic study to explore the situated knowledge of open scholars regarding academic influence. The study used participant observation, semi-structured interviews, and document analysis as ethnographic methods to evaluate the professional use of Twitter by scholars.

4) RESPONSE TIME AND LIFETIME OF POST

Sheikhahmadi *et al.* [62] attempted to determine how influence spreads over time. Spreading velocity has been one of the primary parameters used as a measure of influence. By measuring time units necessary for a user's neighbors to interact with the user's post and the total lifetime of every post, the authors managed to capture the temporal dimension of the problem. Another important parameter is the ratio between various types of actions by neighbors such as replies, mentions, and retweets, as it described the structure of the interaction between a user and his followers. All quantities were presented as distributions, typically in conjunction with a time stamp. As social networks continue to grow and evolve, further studies that include even more parameters may be necessary. The study treated all posts equally and failed to establish a relationship between the content of the post and the extent of influence of the network user. More research is necessary to fully understand the role of the message in the process of detecting influential users since most relevant studies only consider parameters such as network topology or the response time. The proposed model has been tested only on Twitter. There is no indication that it can function with the same degree of efficiency when applied to an OSN with a smaller number of users.

PPR score is the composite measurement that served as a key ranking criterion in Alp and Ötğüdücü [47]. This measurement is calculated independently for each topic; thus, a single user can have a range of scores. It is based on user features and activity, which can be seen as functions of raw statistics such as the total number of active days, tweets, connections, and retweets. The time elapsed between a post and a response to it is also a parameter in one of the features. The study used score spread and active node count measurements for evaluation. The algorithm is heavily dependent on the choice of user features that affect how PPR scores are calculated. As experimental results have shown, the chosen blend of features appears relevant for the analysis of certain topics, but inadequate for others. In other words, the definition of influence seems arbitrary, which is illustrated by the fact that human observers assign different values to the same aspects that the algorithm is based on. The study was limited to content on Twitter exclusively in Turkish, and it is questionable whether the same model would be effective in the study of content in another language or a different social network.

5) USER CENTRALITY

The main value calculated for all users in the sample was WeiboRank, which is a composite measure of a user's centrality within a network [71]. This variable is calculated based on the volume and structure of connections for each user. To compare WeiboRank and alternative methods for determining centrality, the authors of [71] used Kendall's coefficient, which implies a high level of correlation when its absolute value approaches 1. This measure allows various methods to be evaluated based on their propagation range density, and provides a neutral framework for assessing the accuracy of key node identification in microblogging networks. Though WeiboRank was found to be more accurate than other methods, its performance declined when the number of top-ranked users required from the algorithm began to grow. This feature limited its applicability, even if the solution was suitable for most ranking tasks. Further, the method relied exclusively on the network topology and ignored other factors such as history of activity on the network, semantic content, etc. Thus, it is possible to imagine more complex methods that achieve higher accuracy by accounting for additional factors. Finally, the performance of the algorithm was evaluated strictly by a comparison with competing methods without any external confirmation of influence [51], [71].

6) OVERLAP RATE, TASTE SIMILARITY, AND USER DEGREE/RANKING

Ji *et al.* [68] aimed at measuring three main parameters: the overlap rate ρ, the taste similarity θ, and the influence of user degree/ranking, measuring the third on the previous two parameters with the aim of determining the influence of common interests on the formation of trust relationships. They considered the time taken to create trust relationships since it helped the researchers distinguish between the common interests before and after the creation of such relationships. The PCC was used to quantify the taste similarity through rating vectors of common interests for each pair of users. User degree K_u was measured to investigate its influence on developing taste similarity.

The empirical analysis of the Epinions dataset (Epinions.com is an OSN on which users contribute reviews and share their opinions on any theme or topic) has proven to be easily affected by external factors that cannot be controlled. Fashion trends, for example, may determine the items that the users on these platforms like, and they are often the reason why users leave a positive or a negative comment in response to a particular item they are shown. Besides the external factors, the study did not consider the maximum number of relations a human being can maintain, which is why the results for some of the tested users may be unrealistic. Furthermore, the common interest overlap rate can easily change in the long run, making it an unreliable indicator.

Loucif *et al.* [51] used the ranking of egos based on the three compared simplistic approaches studied (Twinf, the proposed model; In-degree Centrality; and TrustRank) to measure performance. The authors recognized the shortcomings of the model presented, stating that even though it outperformed TrustRank and in-degree centrality, it was still too simplistic unless other functions were added into the model. The authors proposed adding social influence factors such as interpersonal trust in the future.

7) RELIABILITY AND SENSITIVITY

Cigarrán *et al.* [45] used measurements proposed by Replab Organizers, namely reliability and sensitivity. The two measures factor in the relationship between two pairs of documents, or rather the relationships between two items categorized in the same cluster. This means that reliability is the precision of the binary relationship predicted by the algorithm in relation to that derived from the gold standard. Sensitivity, on the other hand, is the recall of this relationship. However, this work relied on the textual representation of tweets. This has drawbacks particularly in terms of the lack of structure. Thus, this work has left it to future research to explore the use of high-level features to describe tweets based on sematic and linguistic features.

Servi and Elsong [58] used the sum of squared errors to measure the performance of the forecasting algorithms in their experiments. This involved comparing the values predicted by each forecasting algorithm to the expected value. While monitoring breakpoints to establish how a user emotion has shifted, one can examine how the trend lines rise or fall, and how sharp such rises and falls are (i.e., the slopes). Further research is necessary to mathematically determine how sharp a rise or fall should be for it to be deemed significant.

8) CLUSTER AND COMMUNITY ANALYSIS

Dugué *et al.* [53] used cluster analysis as a means of identifying the position of social capitalists in the network. Through ANOVA and post hoc tests, they depicted the differences for all clusters in the different measures used. The author noted the difficulties in applying the proposed approach to the network as a result of the large size of data. The author proposed the extension of this work in two ways: adding weights to links in the network and adapting the approach to address overlapping communities. For performance measurement, Zhou *et al.* [48] compared the proposed method with two other models to discover communities, that is, the LDA-based community discovery and influencer-based community discovery presented in previous work. For this comparison, the author used top-10 accuracy to evaluate the accuracy of the models. For the application scenario-based simulation, the author compared the number of negotiations from a sample of 100 active participants in the realm of collective decision making in the simulation. This was compared to that of the Delphi method. The authors proposed that future work should aim to extend the work presented, particularly in the form of more experiments to tweak and improve metrics in the proposed model. This entailed adjusting weights in equations to also adjust them to fit complex scenarios. The author also proposed to get a better understanding of the collective intelligence and communities within OSNs.

9) CLASSIFICATION ACCURACY

Rocha *et al.* [60] used accuracy as a metric to compare the feature sets in order to find the most relevant one as well as the best combination. They used training time as a metric to measure different characterization techniques in line with three major tasks, namely feature extraction, vector creation, and classification learning. They used classification accuracy as a metric to compare the classification algorithms reviewed. The authors recommended that future work should use techniques such as random projection to reduce computational power needs. There is also a need to rely on other information on Twitter that goes beyond the text, including metadata, images, and videos in order to improve the accuracy of author attribution on OSNs. Three metrics were used in the experiments to measure the performance of MapMe, namely

Precision, Recall, and F1. Precision is the number of anchor links between OSNs that are mapped out correctly by the algorithm under evaluation. Recall is the portion of the actual anchor links that are included in the result. F1 is a weighted mean of both precision and recall. The solution presented by Ma *et al.* [61] has been used on synthetic and small real-world datasets. They suggested scaling and applying MapMe to larger OSNs to prove its scalability and advantages in research.

10) TREE SIZE, CORRELATION COEFFICIENT, AND RELATIVE ABSOLUTE ERROR

del Campo-Ávila *et al.* [55] focused on a range of features that may be important while calculating influence on Twitter through three popular tools: Tree Size of the regression trees generated, correlation coefficient, and the relative absolute error of the algorithms in relation to the values in the dataset. In future works, the author plans to design interactive tools where the end user could facilitate the learning algorithm to undertake some specific actions, thus tailoring the recommendations according to one's purpose. The author also noted problems in collecting information from the corresponding APIs of Klout, TwitterGrader, and PeerIndex. Some of the problems noted here include fail whale error, cancelation because of time delay, and users without information.

VI. CHALLENGES, GAPS, AND LIMITATIONS

A. WHAT ARE THE KEY CHALLENGES FACED WHILE USING OSNs TO MEASURE REPUTATION AND INFLUENCE?

Researchers have faced challenges in their attempts to measure reputation and influence on OSNs. Some of these challenges are explained here.

1) SUBJECTIVITY OF MESSAGE AND SENSEMAKING

One obstacle impacting the measurement of online reputation and influence is the subjectivity of the message and the ambiguity involved in sensemaking. Reyes *et al.* [69] highlighted the challenge of irony in posts on OSNs as it cuts through every aspect of language, from pronunciation to lexical choice, syntactic structure, semantics, and conceptualization. The authors stated that a general solution will not be found in any single technique or algorithm [69]. Spina *et al.* [65] added that topics of interest for reputation monitoring are usually fine-grained and suffer from data sparsity. Ho *et al.* [72] explained that the quantification of a person's capability in disseminating ideas, the measurement of the extent of propagation of a concept, and the demonstration, visualization, and propagation of information, all through a microblog, are among the most difficult tasks [72]. Stewart [77] demonstrated these obstacles by monitoring how academics make sense of each other's profiles, influence, and credibility on social media. Bae and Lee [64] highlighted the significant correlation between sentiment and phenomena while Ruan *et al.* [66] sought to find a correlation with financial markets, which can be especially difficult while attempting to narrow down the influencer.

2) NON-TEXT POSTS

Implied intent and all that is not expressed through the written word can pose a challenge to the measurement of reputation and influence, despite their impact on both. Sun and Ng [40] confirmed that sentiment analysis is more complex than traditional analysis because it utilizes symbols and emoticons. Traditional summarization techniques only consider textual information, which is insufficient for Twitter context summarization tasks, since information on Twitter is more than text [63].

3) ENGAGEMENT

The degree to which one is engaged with online content is not straightforward [46], [152], [158]. One challenge is measuring influence adequately by considering both negative relationships and the conformity of people [49]. Determining the specific characteristics of influential users on OSNs and identifying aspects that make them influential are challenging [70]. The relative influence of each Twitter user is not measured easily especially because of fake followers and because the reintroduction of tweets instead of retweeting makes the numbers unreliable [79]. When someone retweets a tweet with a URL, it can be unclear whether it was tweeted because of the person's influence or because of the influence of an exogenous source such as the news media [67]. The challenge lies in tracing the spread of information within the network along with the cause of the spread [50].

Previous studies have utilized variables such as follower/follow ratio, unique re-tweeters, unique messages retweeted, and username mention count, which are not necessarily accurate measures of a user's online influence [78]. Li *et al.* [73] were at an impasse when it came to creating a model that considers the dynamic probabilities of influence. Zhaoyun *et al.* [75] [76] attempted to meet the challenge of creating a collaborative algorithm by taking both social media influence and location into account. Previous works have not provided direct measures of influence and authority on OSNs [74].

B. WHAT ARE THE LIMITATIONS AND GAPS IN THE PREVAILING LITERATURE ON OSNs CONCERNING THE MEASUREMENT OF REPUTATION AND INFLUENCE?

Existing literature addressing reputation and influence on OSNs has several limitations particularly because of the limitations in the measurement tools used, the approaches followed, and the new perspectives engaged with.

1) MEASUREMENT TOOLS

Most studies have relied on measures based on centrality indices and graph theory rather than the perceived authoritarian stature of OSN users [41]. Meyers *et al.* [67] explicitly modeled the activity and influence of an external source, unlike Subbian and Melville [74], who found that online influence tends to be measured through followers or location and not with reference to the context of particular tasks, and this has made the resulting data unreliable.

2) STUDY APPROACH

The greatest limitation in current research is the stagnation in the approaches used. Existing approaches rely on measures based on centrality indices and graph theory rather than the perceived authoritarian stature of social media users [70]. Previous work has also focused on information diffusion through the study of information propagation, the structure and evolution of information, the dynamics of massive viral communication, and the link prediction in social network environments [50]. Research has focused on topic and event detection on Twitter, topic models, and state-of-the-art online reputation monitoring [65]. However, according to Spina *et al.* [65], none of these works include clustering relationships between tweets. Instead, they align topics with online news and manually evaluate the aggregated output to create ground truths to decide whether or not a topic is emerging. Most previous studies have been based on a single relationship [75] [76]. Further, the transition probability of the random surfer has been referred to in the previous literature as the number of tweets and the topical similarity according to relationships with followers, but other relational networks such as retweets, replies, and so on have been neglected [75], [76]. Current research efforts lack depth in their analysis of the ''characteristics'' of the output influence and most tools compute user influence based on their static network properties [73]. According to Agah and Dasari [78], Klout is a website and mobile app that uses social media analytics to measure social influence of its users. They pointed out that Klout scores, which use a fixed range of nodes, are not fully representative of the influence of a person.

3) NEW PERSPECTIVES

New perspectives and characteristics of online reputation and influence have been highlighted in the most recent research. For example, irony has received little serious computational treatment in the past [69]. Reyes *et al.* [69] differed in that they dealt with non-factual information that is linguistically expressed, such as sentiment, attitude, humor, and mood to analyze irony in terms of a multidimensional model of textual elements. Scholars have studied public mood resulting from social media posts by using the profile of mood states (POMS), an additional rating system that assesses prevailing mood states. However, no scholar who has used POMS has compared the results and findings across different social media platforms [40]. While previous research on Twitter data summarization has centered on topic level summarization, it has failed to leverage user interaction information on Twitter [63]. Farhani *et al.* [63] proposed a Twitter-specific summarization technique to generate a Twitter context tree. Another key limitation of existing social influence analysis techniques according to Li *et al.* [49] is that they only utilize positive interactions (e.g., agreement, trust) between individuals, ignoring two equally important factors, namely negative relationships (e.g., distrust, disagreement) between individuals and conformity of people, which refers to a person's inclination to be influenced.

Whereas the ranking of peer-reviewed citations through academic publishing has traditionally been the primary form of influence among scholars, there is no formal system to judge the quality or credibility of an academic's work [77]. Whereas there have been many sentiment-based analyses of Twitter to predict real-world phenomena, there has been no research on the use of positive-negative sentiments between popular users on Twitter and their audiences as a new measure of influence [64]. Bae *et al.* [64] believed that their sentiment analysis provided new insights into the influence of a user and was related to the real world. Many existing works use historical stock market data without considering the media or public sentiment [66]. As explained by Ruan *et al.* [66], most sentiment analysis produces only binary results and does not consider influential users. Thus, they proposed using trust as a method of filtering and amplifying social media to increase its correlation with financial data in the stock market [66]. Few existing studies determine social media influencers based on their location and geographic proximity with each other [79]. According to Wei *et al.* [79], other metrics being utilized did not reflect the flow of information on Twitter accurately, and did not represent a credible portrayal of influence among users who are not directly connected. Ho *et al.* [72] went a step further by designing and testing an online demo system to determine how to measure or quantify the propagation of information on microblogging platforms.

VII. DISCUSSION AND DIRECTIONS FOR FUTURE STUDY

Daniel *et al.* [42] framed a very original approach that attempted to overcome the limitation of social media inputs in evaluating the financial perspectives of a company. Previously, this line of research was blocked by the high level of noise that is typical of social networks, which prevented researchers from differentiating between impactful users and those whose opinions carry little weight. The authors introduced a complex algorithm with several modules to reduce the noise level and single out the most relevant messages on Twitter using sentiment analysis to study the content of those messages. Several different tools for sentiment analysis were used for this purpose, providing some measure of variation in the approach, but all of them returned very similar results, thus confirming the original hypothesis. Another key aspect of this study is the demonstrated ability to handle large collections of data with the test sample that consisted of 12 million messages and the period covered by research extending for around two years. This is very important for the analysis of macroeconomic trends, which can be slow to develop and thus difficult to notice in studies focusing only on short periods of time and on a limited number of social media posts.

Using the Sina Weibo platform as their test subject, Yu *et al.* [52] developed a model that accurately computed the degree of user influence based on user attention rates, activity levels, message content influence, and the number of shared posts. By introducing the time component, the authors equipped the model with the capacity to differentiate between microblogs shared in the past and those posted recently, and by doing so, they enhanced the model's ability to calculate the user influence accurately. Higher user influence also suggests higher attention rates and a greater impact on information dissemination on the network. Fake and zombie fans can inflate attention rates, which can lead to inaccurate results. Thus, the authors focused on user ability to spread information more than on the number of fans. The model did not produce the expected results while calculating the indirect influence of users, mostly because it was applied to large-scale datasets.

Most studies do not address the issue of event evolution, which is why tracking important events as they evolve is still a slow and complicated process. Shi *et al.* [44] developed the HEE technique that was based on the event evolution model. Their technique is capable of discovering user interest communities, organizing posts in event categories, and automatically filtering out general events. An automatic topic clustering algorithm is used to combine all text and discover essential keywords for each post. By producing a long text document that solves the problem of sparse data and increases the quality of topic definition, they managed to solve a problem that has plagued scholars in the past. A cosine measure-based event similarity method was employed to predict the evolution of an event into a hot topic. The HEE model succeeded in delivering higher accuracy rates of event evolution, but it still cannot predict user behavior during the evolution of an event. Further, the researchers did not take all aspects of the tweets into consideration, which leaves room for the improvement of the HEE model and provides a foundation for additional research.

With the advent of the Internet and social media platforms, voicing an opinion on a company or its products and services has become a simple process. Recommendations or positive reviews can be viewed by millions of potential clients, but unfortunately the same is true for complaints and negative reviews. Einwiller and Steilen [56] focused on the strategies of major US companies in handling complaints through Twitter and Facebook. The results showed that companies did not fully employ the opportunities provided by social media because complainant dissatisfaction rates are high, and the average amount of time it takes a company to respond to a complaint is still quite long. However, the study only focused on 32 companies, which is why the results cannot be generalized to other cultures and organizations. The data used in the study were collected over a period of one month, which means that the results only show how well companies handled complaints at a certain point in time. However, the study also showed that the most frequently used complaint handling strategies do not appease complainants. On the other hand, response strategies that improve customer satisfaction were only used in a small number of cases.

Cerchiello andSteilen [57] utilized h-index, a measure that was originally used for describing the reliability of scientific authors and for differentiating among influential Twitter users belonging to the financial community. To this end, they modified the original metric and made it suitable for ranking purposes through a series of statistical operations, so that new metrics could be described accurately as a ''statistical h-index.'' Distributions of tweet volume and the retweet vector were used to calculate the h-index and confidence factor for each user with a discrete convolutional model used to solve the problem of discontinuous data input. Data loss function was also used as a source of inspiration for the model, accounting for the temporal dimension and rewarding authors who had become active only more recently. The resulting model has a distinct advantage of providing a quick and convenient way of ranking influential users, while the accuracy of predictions can be improved by introducing confidence intervals. It is a valuable tool that can help a large number of users who are interested in receiving financial advice to identify the most trustworthy sources of information.

Marketing and a number of other industries have been transformed completely by social media. Today, spreading information does not mean spending substantial amounts of money on promotion, because influential social media users can help a company reach its target audience much faster. This is why detecting influential users can be profitable for online advertising professionals. Sheikhahmadi *et al.* [62] succeeded in creating a model to monitor the behavior of the user's neighbors to determine their level of influence. Parameters such as the number of likes, comments, and shares have been used to measure the role of neighbors in spreading a user's influence. Furthermore, they also measured the amount of time spent by the neighbors in interacting with a post with the aim of establishing a connection between the longevity of a post and the corresponding user influence. The study did not take the content of a post into consideration. During the experiments, all posts were treated equally, which is why it failed to show how important the message was in identifying influential users on social media. However, the study comprised significant improvements over earlier ones since it drastically improved the accuracy of influential user identification.

Lu and Wan [71] aimed to provide a new, reliable, and simple method to help identify the most influential individuals within a large group of OSN users. Similar efforts were undertaken in the past, but the authors wanted to overcome some of the limitations of past studies by introducing a novel model to study how information spreads through the network. The notion of centrality is essential for the study, and it was formalized through several parameters that can be determined for each user based on his public contact network. From this data, the algorithm is able to calculate WeiboRank value for every user, indicating how influential his microblog posts are. To prove that WeiboRank actually measured the intended properties, empirical testing was conducted using live data from the Sina Weibo network. The model was tested on this sample, and findings were analyzed both quantitatively and qualitatively, thus revealing several interesting trends that mostly supported the initial hypothesis. WeiboRank was found to be particularly effective in identifying small groups of most influential nodes out of a much larger sample, a context that closely resembles the typical expectations of a ranking tool. Hence, it can be said that the authors succeeded in introducing a new practical tool to analyze user relationships on microblogging networks.

Improving our understanding of how user online trust relations are formed helps us create better Recommender Systems that can easily recognize potential users who can form a trust relationship based on their common interests. Comprehending how trust relations are created can help identify users who are more likely to form such relationships. Ji *et al.* [68] used the Epinions dataset to measure and calculate the overlap rate and the similarity in tastes. The impact of user degree on taste similarity was also measured to determine its influence on both parameters in the process of user trust formation. The empirical results of the study show that similar tastes with respect to common interests increase the chances of forming trust relationships. Despite providing insights into the role of common interests in the creation of trust among online users, they failed to examine how external factors impact the formation of trust relationships. The study was conducted using data collected from a single social network, and there was no indication that the results could be applied to other OSNs. However, the paper did offer valuable information on how trust relations are formed on OSNs and the factors that must be considered so that we can fully understand how the trust between two social media users is established.

Alp and Ötğüdücü [47] provided a hybrid solution to the question of accurate identification of influences or opinion makers on microblogging platforms. The solution originated from the idea that different users may be considered experts for certain topics, but not for others. Thus, influence must be determined based on topics. Two major sources of information were analyzed, namely the structure of the follower network and nodal features such as focus rate, activeness, authenticity, and speed of getting reactions. This complex algorithm was executed in a distributed manner, reducing the time needed to process raw data and return topical PPR scores. Experimental verification of this approach was successful, indicating that the PPR algorithm was more effective in identifying top influencers in the most topical categories than baseline solutions operating on different principles. The accuracy of the selection was confirmed by human observers who rated the tweets extracted from topical influencers selected by the algorithm. It was found that the spread score is the most optimal evaluation measure that can quickly identify influencers in a data set. However, the PPR algorithm can have other applications with minor modifications, and Alp and Ötğüdücü [47] would like to explore the possibility of using it to identify users who intentionally troll others by promoting inaccurate information.

Momeni *et al.* [46] aimed to create algorithms to help solve the problem of bias in the crowd using blockchain technology. Expanding the study of dimensions of quality to cover a wide area of application, most dimensions according to the author have been analyzed on a limited domain. An example is ''truthfulness'', which is a dimension of quality that has mostly been looked at through the lens of microblogging platforms. Improving techniques to encourage users to generate high quality content exploring the potential of hybrid approaches to ranking and assessing UGC.

VIII. CONCLUSION

In conclusion, this study uses Systematic Literature Review (SLR), which is a repetitive process that combines all existing literature on a specific topic or research question [5]. SLR aims to address problems by identifying, critically evaluating, and integrating the findings of all relevant high-quality individual studies addressing one or more research questions. This method determines how much the current studies have progressed toward clarifying the use and impact of social media on their users by means of reputation/influence assessment. It identifies relations, contradictions, gaps, and inconsistencies in the literature, and explores reasons to keep studying the area. It also helps formulate general statements, develop theories, and describe directions for future research.

We conducted a Systematic Literature Review to collect, analyze, and synthesize data on measuring organizations' reputation and influence in social networks. As a result, we studied 116 articles deeply and extensively and we have answered the research questions, analyze them, discussed the gaps, challenges, limitations and future directions.

This review has some limitations that future research can address. First, it has a time boundary for the articles that it reviewed; that is, articles were sourced from 2010 to September 2019. However, this topic is still in its early stages, and new publications will continue to arise. Thus, future research should consider newer research and published papers on this topic. This will help increase awareness of the field among professionals and researchers, especially with the continuous proliferation of new social networks and their use for measuring reputation and influence.

APPENDIX: PRIMARY STUDIES REFERENCES

SID	References
S1	R. P. Mohammadiani, S. Mohammadi, and Z. Malik. "Understanding the relationship strengths in users' activities, review helpfulness and influence," *Computers in Human Behavior*, Elsevier, vol. 75, pp. 117–129, 2017.
S2	D. Garcia, P. Mavrodiev, D. Casati, and F. Schweitzer. "Understanding popularity, reputation, and social influence in the twitter society," *Policy Internet*, Wiley, vol. 9, no. 3, pp. 343–364, 2017.

S3 M. Alrubaian, M. Al-Qurishi, M. Al-Rakhami, M. M. Hassan, and A. Alamri. "Reputation-based credibility analysis of Twitter social network users," *Conc. Comput.: Pract. Exp.*, Wiley, vol. 29, no. 7, 2017.

S4 G. Gadek, A. Pauchet, N. Malandain, K. Khelif, L. Vercouter, and S. Brunessaux. "Topical cohesion of communities on Twitter," *Proc. Comput. Sci.*, Elsevier, vol. 112, pp. 584–593, 2017.

S5 J. Benthaus, M. Risius, and R. Beck. "Social media management strategies for organizational impression management and their effect on public perception," *J. Strat. Inf. Syst.*, Elsevier, vol. 25, no. 2, pp. 127–139, 2016.

S6 P. Cerchiello and P. Giudici, "How to measure the quality of financial tweets," *Qual. Quant.,* Springer, vol. 50, no. 4, pp. 1695–1713, 2016.

S7 M. H. Peetz, M. de Rijke, and R. Kaptein. "Estimating reputation polarity on microblog posts," *Inf. Proc. Manag.*, Elsevier, vol. 52, no. 2, pp. 193–216, 2016.

S8 B. Stewart. "Open to influence: What counts as academic influence in scholarly networked Twitter participation," *Learn, Media Technol.*, vol. 40, no. 3, pp. 287–309, 2015, doi:10.1080/17439884.2015.1015547.

S9 H. Kim, J. Park, M. Cha, and J. Jeong. "The effect of bad news and CEO apology of corporate on user responses in social media," *PloS one*, vol. 10, no. 5, 2015, doi: 10.1371/journal.pone.0126358.

S10 N. A. Vidya, M. I. Fanany, and I. Budi, "Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers," Procedia Computer Science, Elsevier, vol. 72, pp. 519–526, 2015.

S11 M. Gandhi and A. Muruganantham. "Potential influencers identification using multi-criteria decision making (MCDM) methods," *Proc. Comput. Sci.*, Elsevier, vol. 57, pp. 1179–1188, 2015.

S12 M. Rouprêt, T. M. Morgan, P. J. Bostrom, M. R. Cooperberg, A. Kutikov, K. D. Linton,... and A. Winterbottom. "European Association of Urology (@ Uroweb) recommendations on the appropriate use of social media," *Eur. Urol.*, Elsevier, vol. 66, no. 4, pp. 628–632, 2014.

S13 L. Weitzel, J. P. M. de Oliveira, and P. Quaresma. "Measuring the reputation in user-generated-content systems based on health information,", *Proc. Int. Conf. Comput. Sci.*, vol. 29, pp. 364–378, 2014.

S14 S. F. Schultz, S. Utz, and A. Göritz. "Is the medium the message? Perceptions of and reactions to crisis communication via twitter, blogs and traditional media," *Public Relat. Rev.,* Elsevier, vol. 37, no. 1, pp. 20–27, 2011.

S15 X. Zhang, D. D. Han, R. Yang, and Z. Zhang. "Users' participation and social influence during information spreading on Twitter," *PloS one*, vol. 12, no. 9, e0183290, 2017, doi: 10.1371/journal.pone.0183290.

S16 J. Riddell, A. Brown, I. Kovic, and J. Jauregui. "Who are the most influential emergency physicians on Twitter?" *West. J. Emerg. Med.*, vol. 18, no. 2, p. 281, 2017, doi: 10.5811/westjem.2016.11.31299.

S17 L. L. Shi, L. Liu, Y. Wu, L. Jiang, and J. Hardy. "Event detection and user interest discovering in social media data streams," in *IEEE Access*, vol. 5, pp. 20953–20964, 2017.

S18 A. M. Abirami and A. Askarunisa. "Sentiment analysis model to emphasize the impact of online reviews in healthcare industry," *Online Inf. Rev.*, vol. 41, no. 4, pp. 471–486, 2017.

S19 A. Rocha, W. J. Scheirer, C. W. Forstall, T. Cavalcante, A. Theophilo, B. Shen,... & E. Stamatatos. "Authorship attribution for social media forensics," *IEEE Trans. Inf. Forens. Sec.*, vol. 12, no. 1, pp. 5–33, 2016.

S20 G. Ghosh, S. Banerjee, and N. Y. Yen. "State transition in communication under social network: An analysis using fuzzy logic and Density Based Clustering towards big data paradigm," *Future Gen. Comput. Syst.*, Elsevier, vol. 65, pp. 207–220, 2016.

S21 H. Loucif, A. Boubetra, and S. Akrouf. "A simplistic model for identifying prominent web users in directed multiplex social networks: a case study using Twitter networks," *New Rev. Hypermedia Multi.*, vol. 22, no. 4, pp. 287–302, 2016, doi: 10.1080/13614568.2016.1152311.

S22 X. Teng, S. Pei, F. Morone, and H. A. Makse. "Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks," *Sci. Rep.*, vol. 6, 36043, 2016.

S23 J. Zhang, R. Zhang, J. Sun, Y. Zhang, C. Zhang, J. Zhang,... and C. Zhang. "TrueTop: A Sybil-Resilient System for User Influence Measurement on Twitter," *IEEE/ACM Trans. Network. (TON),* vol. 24, no. 5, pp. 2834–2846, 2016.

S24 F. Riquelme and P. González-Cantergiani. "Measuring user influence on Twitter: A survey," *Inf. Proc. Manag.*, Elsevier, vol. 52, no. 5, pp. 949–975, 2016.

S25 J. Y. Nip and K. W. Fu. "Networked framing between source posts and their reposts: an analysis of public opinion on China's microblogs," *Inf., Commun. Soc.*, vol. 19, no. 8, pp. 1127–1149, 2016.

S26 D. Surian, D. Q. Nguyen, G. Kennedy, M. Johnson, E. Coiera, and A. G. Dunn. "Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection," *J. Med. Internet Res.,* vol. 18, no. 8, 2016, doi: 10.2196/jmir.6045.

S27 D. Xia, S. Mankad, and G. Michailidis. "Measuring influence of users in twitter ecosystems using a counting process modeling framework," *Technom,* vol. 58, no. 3, pp. 360–370, 2016, doi: 10.1080/00401706.2016.1142906.

S28 J. Lu and W. Wan. "Identification of key nodes in microblog networks," *Etri J.*, vol. 38, no. 1, pp. 52–61, 2016, doi:10.4218/etrij.16.0115.0732.

S29 D. Yu, N. Chen, and X. Ran. "Computational modeling of Weibo user influence based on information interactive network," *Online Inf. Rev.*, vol. 40, no. 7, pp. 867–881, Nov.2016, doi: 10.1108/OIR-12-2015-039.

S30 F. Celli, A. Ghosh, F. Alam, and G. Riccardi. "In the mood for sharing contents: Emotions, personality and interaction styles in the diffusion of news," *Inf. Proc. Manag.*, Elsevier, vol. 52, no. 1, pp. 93–98, 2016, doi:10.1016/j.ipm.2015.08.002.

S31 N. Dugué, V. Labatut, and A. Perez. "A community role approach to assess social capitalists' visibility in the Twitter network," *Soc. Net. Anal. Mining*, vol. 5, no. 1, p. 26, 2015, doi: 10.1007/s13278-015-0266-0.

S32 D. H. Alahmadi and X. J. Zeng. "ISTS: Implicit social trust and sentiment based approach to recommender systems," *Expert Syst. App.*, Elsevier, vol. 42, no. 22, pp. 8840–8849, 2015, doi:10.1016/j.eswa.2015.07.036.

S33 L. Ji, J. G. Liu, L. Hou, and Q. Guo. "Identifying the role of common interests in online user trust formation," *PloS one*, vol. 10, no. 7, e0121105, 2015, doi: 10.1371/journal.pone.0121105.

S34 L. Weng and F. Menczer. "Topicality and impact in social media: diverse messages, focused messengers," *PloS one*, vol. 10, no. 2, e0118410, 2015, doi: 10.1371/journal.pone.0118410.

S35 J. Li, W. Peng, T. Li, T. Sun, Q. Li, and J. Xu. "Social network user influence sense-making and dynamics prediction," *Expert Syst. App.*, Elsevier, vol. 41, no. 11, pp. 5115–5124, 2014, doi: 10.1016/j.eswa.2014.02.038.

S36 Z. Y. Ding, Y. Jia, B. Zhou, Y. Han, L. He, and J. F. Zhang. "Measuring the spreadability of users in microblogs," *J. Zhejiang Univ. Sci. C*, Elsevier, vol. 14, no. 9, pp. 701–710, 2013, doi: 10.1016/j.eswa.2014.02.038.

S37 D. Zhaoyun, J. Yan, Z. Bin and H. Yi, "Mining topical influencers based on the multi-relational network in micro-blogging sites," in *China Communications*, vol. 10, no. 1, pp. 93-104, Jan. 2013.

S38 Y. Bae and H. Lee. "Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers," *J. Am. Soc. Inf. Sci. Technol.*, vol. 63, no. 12, pp. 2521–2535, 2012, doi: 10.1002/asi.22768.

S39 P. Saleiro, E. M. Rodrigues, C. Soares, and E. Oliveira. "Texrep: a text mining framework for online reputation monitoring," *New Gen. Comput.*, vol. 35, no. 4, pp. 365–389, 2017, doi: 10.1007/s00354-017-0021-3.

S40 A. Bukhari, U. Qamar, and U. Ghazia. "URWF: user reputation based weightage framework for Twitter micropost classification," *Inf. Syst. e-Bus. Manag.*, vol. 15, no. 3, pp. 623–659, 2017, doi: 10.1007/s10257-016-0320-0.

S41 V. P. Lijo and H. Seetha. "Text-based sentiment analysis," *Intern. J. Knowl. Learn*, vol. 12, no. 1, pp. 1–26, 2017.

S42 F. Pérez-Tellez, J. Cardiff, P. Rosso, and D. Pinto. "Disambiguating company names in microblog text using clustering for online reputation management," *Revista Signos*, vol. 48, no. 87, pp. 54–76, 2015, doi: 10.4067/S0718-09342015000100003.

S43 A. Reyes, P. Rosso, and T. Veale. "A multidimensional approach for detecting irony in Twitter," *Lang. Resour. Eval.*, vol. 47, no. 1, pp. 239–268, 2013, doi: 10.1007/s10579-012-9196-x.

S44 S. K. Bansal and A. Bansal, "Reputation-Based Web Service Selection for Composition," 2011 IEEE World Congress on Services, Washington, DC, 2011, pp. 95-96.

S45 S. Garg, A. Saini and N. Khanna, "Is sentiment analysis an art or a science? Impact of lexical richness in training corpus on machine learning," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, 2016, pp. 2729-2735.

S46 F. Neri, C. Aliprandi, F. Capeci, M. Cuadros and T. By, "Sentiment Analysis on Social Media," 2012 *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Istanbul, Turkey, 2012, pp. 919-926.

S47 L. Servi and S. B. Elson, "A Mathematical Approach to Gauging Influence by Identifying Shifts in the Emotions of Social Media Users," in *IEEE Transactions on Computational Social Systems*, vol. 1, no. 4, pp. 180-190, Dec. 2014.

S48 Z. Yin and Y. Zhang, "Measuring Pair-Wise Social Influence in Microblog," *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, Amsterdam, Netherlands, 2012, pp. 502-507.

S49 M. Danisch, N. Dugué, and A. Perez. "On the importance of considering social capitalism when measuring influence on Twitter," in *Proc.* BESC-ICBES-CC, Shanghai, China, 2014.

S50 T. Rao, and S. Nagpal, "Real-time geo influence in social networks," *2011 3rd International Conference on Electronics Computer Technology*, Kanyakumari, India, 2011, pp. 246–250.

S51 A. Agah and P. K. Dasari. "Measuring influence in online social networks: With a focus on Gowalla and Brightkite," in IEEE ICC, Paris, France, 2017, pp. 1–5.

S52 E. Oro, C. Pizzuti, N. Procopio, and M. Ruffolo. "Detecting topic authoritative social media users: a multilayer network approach," *IEEE Transact. Multi*, vol. 20, no. 5, pp. 1195–1208, 2017, doi: 10.1109/TMM.2017.2763324.

S53 K. Lerman, P. Jain, R. Ghosh, J. H. Kang, and P. Kumaraguru, "Limited attention and centrality in social networks," in *Proc.* ICSIT, State College, PA, USA, 2013, pp. 80–89.

S54 K. Subbian and P. Melville. "Supervised rank aggregation for predicting influencers in Twitter," in IEEE ICPSRT – ICSC, Boston, MA, USA, 2011, pp. 661–665.

S55 B. Sun and V. T. Ng. "Analyzing sentimental influence of posts on social networks," in *Proc.* IEEE ICSCWD, Hsinchu, Taiwan, 2014, pp. 546–551.

S56 J. Chen, H. Li, Z. Wu, and M. S. Hossain. "Sentiment analysis of the correlation between regular tweets and retweets," in IEEE ISNCA, Cambridge, MA, USA, 2017, pp. 1–5.

S57 Y. Blanco-Fernández, M. Lopez-Nores, J. J. Pazos-Arias, and M. I. Martín-Vicente, "Spreading influence values over weighted relationships among users of several social networks," in *Proc.* IEEE ICPCCW, Lugano, Switzerland, 2012, pp. 149–154.

S58 Z. Jianqiang, G. Xiaolin, and T. Feng. "A new method of identifying influential users in the micro-blog networks," *IEEE Access*, vol. 5, pp. 3008–3015, 2017, doi: 10.1109/ACCESS.2017.2672680.

S59 R. Mehmood, H. Maurer, and M. T. Afzal, "Knowledge discovery in hashtags," in *Proc.* IEEE ICET, Islamabad, Pakistan, 2013, pp. 1–6.

S60 J. Ma, Y. Qiao, G. Hu, Y. Huang, M. Wang, A. K. Sangaiah,... and Y. Wang. "Balancing user profile and social network structure for anchor link inferring across multiple online social networks," *IEEE Access*, vol. 5, pp. 12031–12040, 2017, doi: 10.1109/ACCESS.2017.2717921.

S61 C. T. Ho, C. T. Li, and S. D. Lin. "Modeling and visualizing information propagation in a micro-blogging platform," in ICASNAM, Kaohsiung, Taiwan, 2011, pp. 328–335.

S62 Y. Zhang, J. Mo, and T. He, "User influence analysis on micro blog," in *Proc*. IEEE ICCC-IS, Hangzhou, China, 2012, pp. 1474–1478.

S63 H. S. Farahani, A. Bagheri, and E. H. K. M. Saraf. "Characterizing behavior of topical authorities in Twitter," in ICIMIA, Bangalore, India, 2017, pp. 581–586.

S64 A. Khrabrov and G. Cybenko. "Discovering influence in communication networks using dynamic graph analysis," in 2010 IEEE ICSC, Minneapolis, MN, USA, 2010, pp. 288–294.

S65 X. Zhou, B. Wu, and Q. Jin. "Analysis of User Network and Correlation for Community Discovery Based on Topic-Aware Similarity and Behavioral Influence," *IEEE Transact. Human-Machine Syst.*, vol. 48, no. 6, 559-571, Dec. 2018, doi: 10.1109/THMS.2017.2725341.

S66 S. Yoshida, J. Kitazono, S. Ozawa, T. Sugawara, T. Haga, and S. Nakamura, "Sentiment analysis for various SNS media using Naïve Bayes classifier and its application to flaming detection," in *Proc.* IEEE CIBD, Orlando, FL, USA, 2014, pp. 1–6, doi: 10.1109/CIBD.2014.7011523.

S67 T. Ohbe, T. Ozono, and T. Shintani, "Developing a sentiment polarity visualization system for local event information analysis," in *Proc.* IIAI-ICAAI (IIAI-AAI), Kumamoto, Japan, 2016, pp. 19–24.

S68 A. Tedeschi and F. Benedetto, "A cloud-based big data sentiment analysis application for enterprises' brand monitoring in social media streams," *2015 IEEE 1st International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, Turin, Italy, 2015, pp. 186-191, doi: 10.1109/RTSI.2015.7325096.

S69 F. Peleja, J. Santos, and J. Magalhães, "Ranking linked-entities in a sentiment graph," in *Proc.* IEEE/WIC/ACM IJC-WI-IAT, Warsaw, Poland, 2014, pp. 118–125.

S70 S. Hirokawa and J. Zeng, "Comparison of tourism data using double ranking," in *Proc*. ACIS-ISSNE, Seoul, South Korea, pp. 67–72.

S71 E. Koutrouli, G. Kanellopoulos, and A. Tsalgatidou, "Reputation mechanisms in on-line social networks: the case of an influence estimation system in Twitter," in *Proc.* SEED ACE-CNSMC, Kastoria, Greece, 2016, pp. 98–105.

S72 D. Spina, J. Gonzalo, and E. Amigó. "Learning similarity functions for topic detection in online reputation monitoring," in *Proc.* ACM SIGIR RDIS, Queensland, Australia, 2014, pp. 527–536.

S73 M. Cataldi, N. Mittal, and M. A. Aufaure. "Estimating domain-based user influence in social networks," in *Proc.* ACM SAC, Coimbra, Portugal, 2013, pp. 1957–1962.

S74 Y. Chang, X. Wang, Q. Mei, and Y. Liu. "Towards twitter context summarization with user influence models," in *Proc.* ACM ICWSDM, Rome, Italy, 2013, pp. 527–536.

S75 R. Belkaroui and R. Faiz. "Conversational based method for tweet contextualization," *Vietnam J. Comput. Sci.*, vol. 4, no. 4, pp. 223–232, 2017.

S76 M. de Maertelaere, T. Li, and G. Berens, "Social influence: the effect of Twitter information on corporate image," in *Proc.* AICEC, Singapore, Singapore, 2012, pp. 292–293.

S77 H. Li, S. S. Bhowmick, and A. Sun. "Casino: towards conformity-aware social influence analysis in online social networks," in *Proc.* ACM ICIKM, Glasgow Scotland, UK, 2011, pp. 1007–1012, doi: 10.1145/2063576.2063721.

S78 T. Afroze, "Exploring the similarities of influencers in online brand communities," in *Proc.* ICSMS, London, United Kingdom, 2016, p. 9.

S79 P. Y. Huang, H. Y. Liu, C. H. Chen, and P. J. Cheng, "The impact of social diversity and dynamic influence propagation for identifying influencers in social networks," in *Proc.* IEEE/WIC/ACM IJCW-IAT, Atlanta, GA, USA, 2013, pp. 410–416.

S80 B. Markscheffel, and J. Schmidt, "The Scientific Influence Passivity Score (SIP): Bibliometrics learning from Twitter," in *Proc. ICMEDES*, Buraidah, Saudi Arabia, 2014, pp. 20–24.

S81 S. A. Myers, C. Zhu, and J. Leskovec. "Information diffusion and external influence in networks," in *Proc.* ACM SIGKDD ICKDDM, Beijing, China, 2012, pp. 33–41.

S82 J. Hopcroft, T. Lou, and J. Tang. "Who will follow you back?: reciprocal relationship prediction," in *Proc. ACMIC-IKM*, Glasgow, Scotland, UK, 2011, pp. 1137–1146.

S83 R. Belkaroui and R. Faiz "Towards events tweet contextualization using social influence model and users conversations," in *Proc. ICWIMS,* Larnaca, Cyprus, 2015, p. 3.

S84 H. Wei, J. Sankaranarayanan, and H. Samet. "Measuring spatial influence of Twitter users by interactions," in *Proc.* ACM SIGSPATIAL WALEN, Redondo Beach, CA, USA, 2017, no. 2, p. 1–10, doi: 10.1145/3148044.3148046.

S85 M. Jeragh, E. Al-Quraishi, and E. Al-Dwaisan, "A twitter-based weighted reputation system," *Proc. Comput. Sci.*, Elsevier, vol. 10, 902–908, 2012, doi: 10.1016/j.procs.2012.06.119.

S86 R. Page, "Saying 'sorry': Corporate apologies posted on Twitter," *J. Pragmat.*, Elsevier, vol. 62, pp. 30–45, 2014, doi: 10.1016/j.pragma.2013.12.003.

S87 H. S. Manaman, S. Jamali, and A. AleAhmad, "Online reputation measurement of companies based on user-generated content in online social networks," *Comput. Human Behav.*, Elsevier, vol. 54, pp. 94–100, 2016, doi: 10.1016/j.chb.2015.07.061.

S88 N. D. Van Norel, P. A. Kommers, J. J. Van Hoof, and J. W. Verhoeven, "Damaged corporate reputation: Can celebrity Tweets repair it?" *Comput. Human Behav.,* Elsevier, vol. 36, pp. 308–315, 2014, doi: 10.1016/j.chb.2014.03.056.

S89 R. Rutter, S. Roper, and F. Lettice, "Social media interaction, the university brand and recruitment performance," *J. Bus. Res.*, Elsevier, vol. 69, no. 8, pp. 3096–3104, 2016, doi: 10.1016/j.jbusres.2016.01.025.

S90 J. Cigarrán, A. Castellanos, and A. García-Serrano. "A step forward for Topic Detection in Twitter: An FCA-based approach," *Exp. Syst. App.*, Elsevier, vol. 57, pp. 21–36, 2016, doi: 10.1016/j.eswa.2016.03.011.

S91 S. A. Einwiller and S. Steilen. "Handling complaints on social network sites–An analysis of complaints and complaint responses on Facebook and Twitter pages of large US companies," *Public Relat. Rev.*, Elsevier, vol. 41, no. 2, pp. 195–204, 2015, doi: 10.1016/j.pubrev.2014.11.012.

S92 K. Ntalianis, N. Papadakis, and P. Tomaras, "Reputation monitoring over rule-stringent social media based on advanced wrapper technologies," *Proc. Soc. Behav. Sci.*, Elsevier, vol. 148, pp. 559–566, 2014, doi: 10.1016/j.sbspro.2014.07.079.

S93 S. Wan, R. Koh, A. Ong, and A. Pang, "Parody social media accounts: Influence and impact on organizations during crisis," *Public Relat. Rev.*, Elsevier, vol. 41, no. 3, pp. 381–385, 2015, doi: 10.1016/j.pubrev.2015.01.002.

S94 A. Zhang, M. Zheng, and B. Pang. "Structural diversity effect on hashtag adoption in Twitter," *Phy. A : Stat. Mech. App.*, Elsevier, vol. 493, pp. 267–275, 2018, doi: 10.1016/j.physa.2017.09.075.

S95 Z. Z. Alp and S. G. Öğüdücü. "Identifying topical influencers on twitter based on user behavior and network topology," *Knowl. Based Syst.*, Elsevier, vol. 141, pp. 211–221, 2018, doi: 10.1016/j.knosys.2017.11.021.

S96 S. Jendoubi, A. Martin, L. Liétard, H. B. Hadji, and B. Yaghlane. "Two evidential data based models for influence maximization in Twitter," *Knowl. Based Syst.*, Elsevier, vol. 121, pp. 58–70, 2017, doi: 10.1016/j.knosys.2017.01.014.

S97 M. Daniel, R. F. Neves, and N. Horta. "Company event popularity for financial markets using Twitter and sentiment analysis," *Expert Syst. App.*, Elsevier, vol. 71, pp. 111–124, 2017.

S98 J. del Campo-Ávila, N. Moreno-Vergara, and M. Trella-Lopez. "Bridging the gap between the least and the most influential Twitter users," *Proc. Comput. Sci.*, *4th Int. Conf. Ambient Sys, Net & Tech (ANT 2013)*, Halifax, Nova Scotia, Canada, vol. 19, pp. 437–444, 2013.

S99 A. Sheikhahmadi, M. A. Nematbakhsh, and A. Zareie. "Identification of influential users by neighbors in online social networks," *Physica A : Stat. Mech. App.*, Elsevier, vol. 486, pp. 517–534, 2017, doi: 10.1016/j.physa.2017.05.098.

S100 Y. Bai, Z. Yao, and Y. F. Dou, "Effect of social commerce factors on user purchase behavior: An empirical investigation from renren.com," *Intern. J. Inf. Manag.*, Elsevier, vol. 35, no. 5, pp. 538–550, 2015, doi: 10.1016/j.ijinfomgt.2015.04.011.

S101 B. Choi and I. Lee. "Trust in open versus closed social media: The relative influence of user-and marketer-generated content in social network services on customer trust," *Tele. Informat.*, Elsevier, vol. 34, no. 5, pp. 550–559, 2017, doi: 10.1016/j.tele.2016.11.005.

S102 M. A. Al-garadi, K. D. Varathan, and S. D. Ravana. "Identification of influential spreaders in online social networks using interaction weighted K-core decomposition method," *Physica A : Stat. Mech. App.*, Elsevier, vol. 468, pp. 278–288, 2017, doi: doi.org/10.1016/j.physa.2016.11.002.

S103 A. Z. Bahtar and M. Muda, "The impact of User–Generated Content (UGC) on product reviews towards online purchasing–A conceptual framework," *Proc. Econ. Finance*, Penang, Malaysia, vol. 37, pp. 337–342, 2016.

S104 Y. Han, B. Fang, and Y. Jia, "Predicting the topic influence trends in social media with multiple models," *Neurocomputing*, Elsevier, vol. 144, pp. 463–470, 2014, doi: 10.1016/j.neucom.2014.03.054.

S105 P. Dey and S. Roy. "Centrality based information blocking and influence minimization in online social network," in IEEE IC ANTS, Bhubaneswar, India, 2017, pp. 1–6.

S106 A. K. Bhowmick, G. S. B. Chandra, Y. Singh, and B. Mitra. "Constructing influence trees from temporal sequence of retweets: An analytical approach," in IEEE ICBD, Seattle, USA, 2018, pp. 624–633.

S107 W. Zhao, Z. Guan, Y. Huang, T. Xi, H. Sun, Z. Wang, and X. He. "Discerning influence patterns with beta-poisson factorization in microblogging environments," IEEE TKDE, 2019, doi: 10.1109/TKDE.2019.2897932.

S108 F. Al-Raisi, S. Alam, B. Vavala, and M. S. Liu. "Do actions speak louder than words? Predicting influence in Twitter using language and action features," in FedCSIS, Poznan, Poland, 2018, pp. 457–464.

S109 Q. Liao and S. Sinthupinyo. "Exclusiveness centrality: A novel centrality measure to find behavioral influencers on Twitter," in ECAI, Iasi, Romania, 2018, pp. 1–6.

S110 M. A. Zia, Z. Zhang, L. Chen, M. Hashim, and S. Su. "Exploration of influential people for viral marketing," *China Commun.*, IEEE, vol. 15, no. 5, pp. 138–148, 2018, doi: 10.1109/CC.2018.8387993.

S111 P. A. Joshi, G. Simon, and Y. P. Murumkar. "Generation of brand/product reputation using Twitter data," in *ICICET*, Pune, India, 2018, pp. 1–4, doi: 10.1109/ICICET.2018.8533743.

S112 B. Tidke, R. Mehta, and J. Dhanani. "SIRIF: Supervised influence ranking based on influential network," *J. Intel. Fuzzy Syst.*, vol. 35, no. 2, pp. 1225-1237, 2018, doi: 10.3233/JIFS-169667.

S113 S. Jendoubi and A. Martin. "A reliability-based approach for influence maximization using the evidence theory," in ICBDAKD, Lyon, France, 2017, pp. 313–326.

S114 S. Chung, M. Chong, J. S. Chua, and J. C. Na. "Evolution of corporate reputation during an evolving controversy," *J. Commun. Manag.*, vol. 23, no. 1, pp. 52–71, 2019, doi: 10.1108/JCOM-08-2018-0072.

S115 R. Syed. "Enterprise reputation threats on social media: A case of data breach framing," *J. Strat. Inform. Syst.*, Elsevier, vol. 28, no. 3, pp. 257-274, 2019, doi: 10.1016/j.jsis.2018.12.001.

S116 Z. Z. Alp and S. G. Öğüdücü. "Influence factorization for identifying authorities in Twitter," *Knowl. Based Syst.*, Elsevier, vol. 163, pp. 944–954, 2019, doi: 10.1016/j.knosys.2018.10.020.

REFERENCES

[1] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi, and S. Nerur, "Advances in social media research: Past, present and future," *Inf. Syst. Frontiers*, vol. 20, no. 3, pp. 531–558, Jun. 2018, doi: 10.1007/s10796-017-9810-y.

[2] A. Delfanti and A. Arvidsson, *Introduction to Digital Media*. Hoboken, NJ, USA: Wiley, 2019.

[3] U. Sivarajah, Z. Irani, S. Gupta, and K. Mahroof, "Role of big data and social media analytics for business to business sustainability: A participatory Web context," *Ind. Marketing Manage.*, vol. 86, pp. 163–179, Apr. 2020, doi: 10.1016/j.indmarman.2019.04.005.

[4] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of social media," *Bus. Horizons*, vol. 53, no. 1, pp. 59–68, Jan. 2010, doi: 10.1016/j.bushor.2009.09.003.

[5] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, Apr. 2007, doi: 10.1016/j.jss.2006.07.009.

[6] R. F. Baumeister and M. R. Leary, "Writing narrative literature reviews," *Rev. Gen. Psychol.*, vol. 1, no. 3, p. 311, 1997, doi: 10.1037/1089-2680.1.3.311.

[7] B. Kitchenham, *Procedures for Performing Systematic Reviews*, vol. 33. Keele, U.K.: Keele Univ., 2004, pp. 1–26.

[8] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quart.*, vol. 26, no. 2, pp. 13–23, 2002.

[9] M. Unterkalmsteiner, T. Gorschek, A. K. M. Moinul Islam, C. Kian Cheng, R. Bayu Permadi, and R. Feldt, "Evaluation and measurement of software process Improvement—A systematic literature review," *IEEE Trans. Softw. Eng.*, vol. 38, no. 2, pp. 398–424, Apr. 2012.

[10] H. Zhang, M. A. Babar, and P. Tell, "Identifying relevant studies in software engineering," *Inf. Softw. Technol.*, vol. 53, no. 6, pp. 625–637, Jun. 2011, doi: 10.1016/j.infsof.2010.12.010.

[11] B. Zeng and R. Gerritsen, "What do we know about social media in tourism? A review," *Tourism Manage. Perspect.*, vol. 10, pp. 27–36, Apr. 2014, doi: 10.1016/j.tmp.2014.01.001.

[12] R. Gunturu, "Survey of sybil attacks in social networks," 2015, *arXiv:1504.05522*. [Online]. Available: http://arxiv.org/abs/1504.05522

[13] M. Wani, M. A. Alrubaian, and M. Abulaish, "A user-centric feature identification and modeling approach to infer social ties in OSNs," in *Proc. Int. Conf. Inf. Integr. Web-Dased Appl. Services IIWAS*, 2013, p. 107, doi: 10.1145/2539150.2539194.

[14] S. Y. Bhat and M. Abulaish, "Analysis and mining of online social networks: Emerging trends and challenges," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 3, no. 6, pp. 408–444, Nov. 2013, doi: 10.1002/widm.1105.

[15] C. T. Butts, "Social network analysis: A methodological introduction," *Asian J. Social Psychol.*, vol. 11, no. 1, pp. 13–41, Mar. 2008, doi: 10.1111/j.1467-839X.2007.00241.x.

[16] A. Chinchore, F. Jiang, and G. Xu, "Intelligent Sybil attack detection on abnormal connectivity behavior in mobile social networks," in *Proc. Int. Conf. Knowl. Manage. Org.*, Maribor, Slovenia, Springer, 2015, pp. 602–617.

[17] Y. Hu and L. Shi, "Visualizing large graphs," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 7, no. 2, pp. 115–136, Mar. 2015, doi: 10.1002/wics.1343.

[18] M. Oliveira and J. Gama, "An overview of social network analysis," *Wiley Interdiscip. Rev., Data Mining Knowl. Disc., Wiley*, vol. 2, no. 2, pp. 99–115, 2012.

[19] J. Piersa and T. Schreiber, "Spectra of the spike-flow graphs in geometrically embedded neural networks," in *Artificial Intelligence and Soft Computing*, Zakopane, Poland, 2012, pp. 143–151.

[20] P. A. Dow, L. A. Adamic, and A. Friggeri, "The anatomy of large Facebook cascades," in *Proc. ICWSM*, Cambridge, MA, USA, 2013, pp. 145–154.

[21] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the facebook social graph," 2011, *arXiv:1111.4503*. [Online]. Available: http://arxiv.org/abs/1111.4503

[22] A. M. Greenberg, W. G. Kennedy, and N. D. Bos, "Social computing, behavioral-cultural modeling and prediction," in *Proc. ICP*, Washington, DC, USA, 2013, p. 238.

[23] F. Riquelme and P. González-Cantergiani, "Measuring user influence on Twitter: A survey," *Inf. Process. Manage.*, vol. 52, no. 5, pp. 949–975, Sep. 2016.

[24] I. Anger and C. Kittl, "Measuring influence on Twitter," in *Proc. 11th Int. Conf. Knowl. Manage. Knowl. Technol. i-KNOW*, 2011, p. 31.

[25] M. El Marrakchi, H. Bensaid, and M. Bellafkih, "Scoring reputation in online social networks," in *Proc. 10th Int. Conf. Intell. Systems: Theories Appl. (SITA)*, Oct. 2015, pp. 1–6.

[26] H. Dikow, O. Hasan, H. Kosch, L. Brunie, and R. Sornin, "Improving the accuracy of business-to-business (B2B) reputation systems through rater expertise prediction," *Computing*, vol. 97, no. 1, pp. 29–49, Jan. 2015, doi: 10.1007/s00607-013-0345-x.

[27] C. Dijkmans, P. Kerkhof, A. Buyukcan-Tetik, and C. J. Beukeboom, "Online conversation and corporate reputation: A two-wave longitudinal study on the effects of exposure to the social media activities of a highly interactive company," *J. Comput.-Mediated Commun.*, vol. 20, no. 6, pp. 632–648, Nov. 2015.

[28] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, "SybilLimit: A near-optimal social network defense against sybil attacks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 3–17, May 2010.

[29] W. Zhao, Z. Guan, Y. Huang, T. Xi, H. Sun, Z. Wang, and X. He, "Discerning influence patterns with beta-Poisson factorization in microblogging environments," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1092–1103, Jun. 2020, doi: 10.1109/TKDE.2019.2897932.

[30] S. Chung, M. Chong, J. S. Chua, and J. C. Na, "Evolution of corporate reputation during an evolving controversy," *J. Commun. Manage.*, vol. 23, no. 1, pp. 52–71, Feb. 2019, doi: 10.1108/JCOM-08-2018-0072.

[31] Z. Z. Alp and S. G. Öğüdücü, "Influence factorization for identifying authorities in Twitter," *Knowl.-Based Syst.*, vol. 163, pp. 944–954, Jan. 2019, doi: 10.1016/j.knosys.2018.10.020.

[32] A. Zhang, M. Zheng, and B. Pang, "Structural diversity effect on hashtag adoption in Twitter," *Phys. A, Stat. Mech. Appl.*, vol. 493, pp. 267–275, Mar. 2018, doi: 10.1016/j.physa.2017.09.075.

[33] A. K. Bhowmick, G. Sai Bharath Chandra, Y. Singh, and B. Mitra, "Constructing influence trees from temporal sequence of retweets: An analytical approach," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 624–633.

[34] F. Al-Raisi, S. Alam, B. Vavala, and M. S. Liu, "Do actions speak louder than words? Predicting influence in Twitter using language and action features," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, Sep. 2018, pp. 457–464.

[35] Q. Liao and S. Sinthupinyo, "Exclusiveness centrality: A novel centrality measure to find behavioral influencers on Twitter," in *Proc. 10th Int. Conf. Electron., Comput. Artif. Intell. (ECAI)*, Jun. 2018, pp. 1–6.

[36] M. A. Zia, Z. Zhang, L. Chen, M. Hashim, and S. Su, "Exploration of influential people for viral marketing," *China Commun.*, vol. 15, no. 5, pp. 138–148, May 2018, doi: 10.1109/CC.2018.8387993.

[37] P. A. Joshi, G. Simon, and Y. P. Murumkar, "Generation of brand/product reputation using Twitter data," in *Proc. Int. Conf. Inf., Commun., Eng. Technol. (ICICET)*, Aug. 2018, pp. 1–4, doi: 10.1109/ICICET.2018.8533743.

[38] B. Tidke, R. Mehta, and J. Dhanani, "SIRIF: Supervised influence ranking based on influential network," *J. Intell. Fuzzy Syst.*, vol. 35, no. 2, pp. 1225–1237, Aug. 2018, doi: 10.3233/JIFS-169667.

[39] R. Syed, "Enterprise reputation threats on social media: A case of data breach framing," *J. Strategic Inf. Syst.*, vol. 28, no. 3, pp. 257–274, Sep. 2019, doi: 10.1016/j.jsis.2018.12.001.

[40] B. Sun and V. T. Ng, "Analyzing sentimental influence of posts on social networks," in *Proc. IEEE 18th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2014, pp. 546–551.

[41] E. Oro, C. Pizzuti, N. Procopio, and M. Ruffolo, "Detecting topic authoritative social media users: A multilayer network approach," *IEEE Trans. Multimedia*, vol. 20, no. 5, pp. 1195–1208, May 2018.

[42] M. Daniel, R. F. Neves, and N. Horta, "Company event popularity for financial markets using Twitter and sentiment analysis," *Expert Syst. Appl.*, vol. 71, pp. 111–124, Apr. 2017.

[43] M. Alrubaian, M. Al-Qurishi, M. Al-Rakhami, M. M. Hassan, and A. Alamri, "Reputation-based credibility analysis of Twitter social network users," *Concurrency Comput., Pract. Exper.*, vol. 29, no. 7, Apr. 2017, Art. no. e3873.

[44] L.-L. Shi, L. Liu, Y. Wu, L. Jiang, and J. Hardy, "Event detection and user interest discovering in social media data streams," *IEEE Access*, vol. 5, pp. 20953–20964, 2017.

[45] J. Cigarrán, Á. Castellanos, and A. García-Serrano, "A step forward for topic detection in Twitter: An FCA-based approach," *Expert Syst. Appl.*, vol. 57, pp. 21–36, Sep. 2016, doi: 10.1016/j.eswa.2016.03.011.

[46] E. Momeni, C. Cardie, and N. Diakopoulos, "How to assess and rank user-generated content on Web," in *Proc. Companion Web Conf. Web Conf. WWW*, 2018, pp. 489–493.

[47] Z. Z. Alp and S. G. Öğüdücü, "Identifying topical influencers on Twitter based on user behavior and network topology," *Knowl.-Based Syst.*, vol. 141, pp. 211–221, Feb. 2018, doi: 10.1016/j.knosys.2017.11.021.

[48] X. Zhou, B. Wu, and Q. Jin, "Analysis of user network and correlation for community discovery based on topic-aware similarity and behavioral influence," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 6, pp. 559–571, Dec. 2018, doi: 10.1109/THMS.2017.2725341.

[49] H. Li, S. S. Bhowmick, and A. Sun, "CASINO: Towards conformity-aware social influence analysis in online social networks," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage. CIKM*, 2011, pp. 1007–1012, doi: 10.1145/2063576.2063721.

[50] M. Cataldi, N. Mittal, and M.-A. Aufaure, "Estimating domain-based user influence in social networks," in *Proc. 28th Annu. ACM Symp. Appl. Comput. SAC*, 2013, pp. 1957–1962.

[51] H. Loucif, A. Boubetra, and S. Akrouf, "A simplistic model for identifying prominent Web users in directed multiplex social networks: A case study using Twitter networks," *New Rev. Hypermedia Multimedia*, vol. 22, no. 4, pp. 287–302, Oct. 2016, doi: 10.1080/13614568.2016.1152311.

[52] D. Yu, N. Chen, and X. Ran, "Computational modeling of Weibo user influence based on information interactive network," *Online Inf. Rev.*, vol. 40, no. 7, pp. 867–881, Nov. 2016, doi: 10.1108/OIR-12-2015-0391.

[53] N. Dugué, V. Labatut, and A. Perez, "A community role approach to assess social capitalists visibility in the Twitter network," *Social Netw. Anal. Mining*, vol. 5, no. 1, p. 26, Dec. 2015, doi: 10.1007/s13278-015-0266-0.

[54] R. A. Guimer and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, no. 7028, p. 895, 2005, doi: 10.1038/nature03288.

[55] J. del Campo-Ávila, N. Moreno-Vergara, and M. Trella-Lopez, "Bridging the gap between the least and the most influential Twitter users," *Proc. Comput. Sci., 4th Int. Conf. Ambient Sys, Net & Tech (ANT)*, Halifax, NS, Canada, vol. 19, 2013, pp. 437–444.

[56] S. A. Einwiller and S. Steilen, "Handling complaints on social network sites—An analysis of complaints and complaint responses on facebook and Twitter pages of large US companies," *Public Relations Rev.*, vol. 41, no. 2, pp. 195–204, Jun. 2015, doi: 10.1016/j.pubrev.2014.11.012.

[57] P. Cerchiello and P. Giudici, "How to measure the quality of financial tweets," *Qual. Quantity*, vol. 50, no. 4, pp. 1695–1713, Jul. 2016.

[58] L. Servi and S. B. Elson, "A mathematical approach to gauging influence by identifying shifts in the emotions of social media users," *IEEE Trans. Comput. Social Syst.*, vol. 1, no. 4, pp. 180–190, Dec. 2014.

[59] M. Koppel, J. Schler, and S. Argamon, "Computational methods in authorship attribution," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 1, pp. 9–26, Jan. 2009, doi: 10.1002/asi.20961.

[60] A. Rocha, W. J. Scheirer, C. W. Forstall, T. Cavalcante, A. Theophilo, B. Shen, A. R. B. Carvalho, and E. Stamatatos, "Authorship attribution for social media forensics," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 1, pp. 5–33, Jan. 2017.

[61] J. Ma, Y. Qiao, G. Hu, Y. Huang, M. Wang, A. K. Sangaiah, C. Zhang, and Y. Wang, "Balancing user profile and social network structure for anchor link inferring across multiple online social networks," *IEEE Access*, vol. 5, pp. 12031–12040, 2017, doi: 10.1109/ACCESS.2017.2717921.

[62] A. Sheikhahmadi, M. A. Nematbakhsh, and A. Zareie, "Identification of influential users by neighbors in online social networks," *Phys. A, Stat. Mech. Appl.*, vol. 486, pp. 517–534, Nov. 2017, doi: 10.1016/j.physa.2017.05.098.

[63] Y. Chang, X. Wang, Q. Mei, and Y. Liu, "Towards twitter context summarization with user influence models," in *Proc. ACM ICWSDM*, Rome, Italy, 2013, pp. 527–536.

[64] Y. Bae and H. Lee, "Sentiment analysis of Twitter audiences: Measuring the positive or negative influence of popular twitterers," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 12, pp. 2521–2535, Dec. 2012.

[65] D. Spina, J. Gonzalo, and E. Amigó, "Learning similarity functions for topic detection in online reputation monitoring," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. SIGIR*, 2014, pp. 527–536.

[66] Y. Ruan, A. Durresi, and L. Alfantoukh, "Using Twitter trust network for stock market analysis," *Knowl.-Based Syst.*, vol. 145, pp. 207–218, Apr. 2018, doi: 10.1016/j.knosys.2018.01.016.

[67] S. A. Myers, C. Zhu, and J. Leskovec, "Information diffusion and external influence in networks," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining KDD*, 2012, pp. 33–41.

[68] L. Ji, J.-G. Liu, L. Hou, and Q. Guo, "Identifying the role of common interests in online user trust formation," *PLoS ONE*, vol. 10, no. 7, Jul. 2015, Art. no. e0121105, doi: 10.1371/journal.pone.0121105.

[69] A. Reyes, P. Rosso, and T. Veale, "A multidimensional approach for detecting irony in Twitter," *Lang. Resour. Eval.*, vol. 47, no. 1, pp. 239–268, Mar. 2013.

[70] A. Khrabrov and G. Cybenko, "Discovering influence in communication networks using dynamic graph analysis," in *Proc. IEEE 2nd Int. Conf. Social Comput.*, Aug. 2010, pp. 288–294.

[71] J. Lu and W. Wan, "Identification of key nodes in microblog networks," *ETRI J.*, vol. 38, no. 1, pp. 52–61, Feb. 2016, doi: 10.4218/etrij.16.0115.0732.

[72] C.-T. Ho, C.-T. Li, and S.-D. Lin, "Modeling and visualizing information propagation in a micro-blogging platform," in *Proc. Int. Conf. Adv. Social Netw. Anal. Mining*, Jul. 2011, pp. 328–335.

[73] J. Li, W. Peng, T. Li, T. Sun, Q. Li, and J. Xu, "Social network user influence sense-making and dynamics prediction," *Expert Syst. Appl.*, vol. 41, no. 11, pp. 5115–5124, Sep. 2014, doi: 10.1016/j.eswa.2014.02.038.

[74] K. Subbian and P. Melville, "Supervised rank aggregation for predicting influencers in Twitter," in *Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust IEEE 3rd Int. Conf. Social Comput.*, Oct. 2011, pp. 661–665.

[75] Z.-Y. Ding, Y. Jia, B. Zhou, Y. Han, L. He, and J.-F. Zhang, "Measuring the spreadability of users in microblogs," *J. Zhejiang Univ. Sci. C*, vol. 14, no. 9, pp. 701–710, Sep. 2013.

[76] D. Zhaoyun, J. Yan, Z. Bin, and H. Yi, "Mining topical influencers based on the multi-relational network in micro-blogging sites," *China Commun.*, vol. 10, no. 1, pp. 93–104, Jan. 2013.

[77] B. Stewart, "Open to influence: What counts as academic influence in scholarly networkedTwitterparticipation," *Learn., Media Technol.*, vol. 40, no. 3, pp. 287–309, Jul. 2015, doi: 10.1080/17439884.2015.1015547.

[78] A. Agah and P. K. Dasari, "Measuring influence in online social networks: With a focus on gowalla and brightkite," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–5.

[79] H. Wei, J. Sankaranarayanan, and H. Samet, "Measuring spatial influence of Twitter users by interactions," in *Proc. 1st ACM SIGSPATIAL Workshop Analytics Local Events News LENS*, 2017, pp. 1–10, doi: 10.1145/3148044.3148046.

[80] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, A social network or a news media?" in *Proc. 19th Int. Conf. World Wide Web WWW*, 2010, pp. 591–600.

[81] J. Zhang, B. Liu, J. Tang, T. Chen, and J. Li, "Social influence locality for modeling retweeting behaviors," in *Proc. IJCAI*, Beijing, China, 2013, pp. 1–7.

[82] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Phys.*, vol. 6, no. 11, pp. 888–893, Nov. 2010, doi: 10.1038/nphys1746.

[83] C. Kang, C. Molinaro, S. Kraus, Y. Shavitt, and V. S. Subrahmanian, "Diffusion centrality in social networks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2012, pp. 558–564.

[84] S. Gao, J. Ma, Z. Chen, G. Wang, and C. Xing, "Ranking the spreading ability of nodes in complex networks based on local structure," *Phys. A, Stat. Mech. Appl.*, vol. 403, pp. 130–147, Jun. 2014, doi: 10.1016/j.physa.2014.02.032.

[85] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *Proc. IAAAI WSM*, Washington, DC, USA, 2010, pp. 1–8.

[86] A. Pal and S. Counts, "Identifying topical authorities in microblogs," in *Proc. 4th ACM Int. Conf. Web Search Data Mining WSDM*, 2011, pp. 45–54.

[87] J. L. Scott and J. Freese, *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX, USA: Stata Press, 2006.

[88] D. M. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter," in *Proc. 20th Int. Conf. World Wide Web WWW*, 2011, pp. 695–704.

[89] P. S. Dodds, "An experimental study of search in global social networks," *Science*, vol. 301, no. 5634, pp. 827–829, Aug. 2003.

[90] P. Massa and P. Avesani, "Controversial users demand local trust metrics: An experimental study on epinions.com community," in *Proc. AAAI*, 2005, pp. 121–126.

[91] K. Lewis, M. Gonzalez, and J. Kaufman, "Social selection and peer influence in an online social network," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 1, pp. 68–72, Jan. 2012, doi: 10.1073/pnas.1109739109.

[92] D.-C. Nie, Z.-K. Zhang, Q. Dong, C. Sun, and Y. Fu, "Information filtering via biased random walk on coupled social network," *Sci. World J.*, vol. 2014, pp. 1–10, Jul. 2014.

[93] R. P. Mohammadiani, S. Mohammadi, and Z. Malik, "Understanding the relationship strengths in users' activities, review helpfulness and influence," *Comput. Hum. Behav.*, vol. 75, pp. 117–129, Oct. 2017.

[94] D. Garcia, P. Mavrodiev, D. Casati, and F. Schweitzer, "Understanding popularity, reputation, and social influence in the Twitter society," *Policy Internet*, vol. 9, no. 3, pp. 343–364, Sep. 2017.

[95] G. Gadek, A. Pauchet, N. Malandain, K. Khelif, L. Vercouter, and S. Brunessaux, "Topical cohesion of communities on Twitter," *Procedia Comput. Sci.*, vol. 112, pp. 584–593, Jan. 2017.

[96] X. Zhang, D.-D. Han, R. Yang, and Z. Zhang, "Users' participation and social influence during information spreading on Twitter," *PLoS ONE*, vol. 12, no. 9, Sep. 2017, Art. no. e0183290, doi: 10.1371/journal.pone.0183290.

[97] J. Riddell, A. Brown, I. Kovic, and J. Jauregui, "Who are the most influential emergency physicians on Twitter," *West. J. Emerg. Med.*, vol. 18, no. 2, p. 281, 2017, doi: 10.5811/westjem.2016.11.31299.

[98] A. M. Abirami and A. Askarunisa, "Sentiment analysis model to emphasize the impact of online reviews in healthcare industry," *Online Inf. Rev.*, vol. 41, no. 4, pp. 471–486, Aug. 2017.

[99] P. Saleiro, E. M. Rodrigues, C. Soares, and E. Oliveira, "TexRep: A text mining framework for online reputation monitoring," *New Gener. Comput.*, vol. 35, no. 4, pp. 365–389, Oct. 2017, doi: 10.1007/s00354-017-0021-3.

[100] A. Bukhari, U. Qamar, and U. Ghazia, "URWF: User reputation based weightage framework for Twitter micropost classification," *Inf. Syst. e-Bus. Manage.*, vol. 15, no. 3, pp. 623–659, Aug. 2017, doi: 10.1007/s10257-016-0320-0.

[101] V. P. Lijo and H. Seetha, "Text-based sentiment analysis," *Intern. J. Knowl. Learn., Inder Sin.*, vol. 12, no. 1, pp. 1–26, 2017, doi: 10.1504/IJKL.2017.088163.

[102] J. Chen, H. Li, Z. Wu, and M. S. Hossain, "Sentiment analysis of the correlation between regular tweets and retweets," in *Proc. IEEE 16th Int. Symp. Netw. Comput. Appl. (NCA)*, Oct. 2017, pp. 1–5.

[103] Z. Jianqiang, G. Xiaolin, and T. Feng, "A new method of identifying influential users in the micro-blog networks," *IEEE Access*, vol. 5, pp. 3008–3015, 2017, doi: 10.1109/ACCESS.2017.2672680.

[104] H. S. Farahani, A. Bagheri, and E. H. K. M. Saraf, "Characterizing behavior of topical authorities in Twitter," in *Proc. Int. Conf. Innov. Mech. Ind. Appl. (ICIMIA)*, Feb. 2017, pp. 581–586.

[105] R. Belkaroui and R. Faiz, "Conversational based method for tweet contextualization," *Vietnam J. Comput. Sci.*, vol. 4, no. 4, pp. 223–232, Nov. 2017, doi: 10.1007/s40595-016-0092-y.

[106] S. Jendoubi, A. Martin, L. Liétard, H. B. Hadji, and B. B. Yaghlane, "Two evidential data based models for influence maximization in Twitter," *Knowl.-Based Syst.*, vol. 121, pp. 58–70, Apr. 2017, doi: 10.1016/j.knosys.2017.01.014.

[107] B. Choi and I. Lee, "Trust in open versus closed social media: The relative influence of user- and marketer-generated content in social network services on customer trust," *Telematics Informat.*, vol. 34, no. 5, pp. 550–559, Aug. 2017, doi: 10.1016/j.tele.2016.11.005.

[108] M. A. Al-garadi, K. D. Varathan, and S. D. Ravana, "Identification of influential spreaders in online social networks using interaction weighted K-core decomposition method," *Phys. A, Stat. Mech. Appl.*, vol. 468, pp. 278–288, Feb. 2017, doi: 10.1016/j.physa.2016.11.002.

[109] P. Dey and S. Roy, "Centrality based information blocking and influence minimization in online social network," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS)*, Dec. 2017, pp. 1–6.

[110] S. Jendoubi and A. Martin, "A reliability-based approach for influence maximization using the evidence theory," in *Proc. ICBDAKD*, Lyon, France, Aug. 2017, pp. 313–326.

[111] J. Benthaus, M. Risius, and R. Beck, "Social media management strategies for organizational impression management and their effect on public perception," *J. Strategic Inf. Syst.*, vol. 25, no. 2, pp. 127–139, Jul. 2016.

[112] M.-H. Peetz, M. de Rijke, and R. Kaptein, "Estimating reputation polarity on microblog posts," *Inf. Process. Manage.*, vol. 52, no. 2, pp. 193–216, Mar. 2016.

[113] H. Kim, J. Park, M. Cha, and J. Jeong, "The effect of bad news and CEO apology of corporate on user responses in social media," *PLoS ONE*, vol. 10, no. 5, May 2015, Art. no. e0126358, doi: 10.1371/journal.pone.0126358.

[114] N. A. Vidya, M. I. Fanany, and I. Budi, "Twitter sentiment to analyze net brand reputation of mobile phone providers," *Procedia Comput. Sci.*, vol. 72, pp. 519–526, Jan. 2015.

[115] M. Gandhi and A. Muruganantham, "Potential influencers identification using multi-criteria decision making (MCDM) methods," *Procedia Comput. Sci.*, vol. 57, pp. 1179–1188, Jan. 2015.

[116] M. Rouprêt, T. M. Morgan, P. J. Bostrom, M. R. Cooperberg, A. Kutikov, K. D. Linton, J. Palou, L. Martínez-Piñeiro, H. van der Poel, C. Wijburg, A. Winterbottom, H. H. Woo, M. P. Wirth, and J. W. F. Catto, "European association of urology (@Uroweb) recommendations on the appropriate use of social media," *Eur. Urol.*, vol. 66, no. 4, pp. 628–632, Oct. 2014.

[117] L. Weitzel, J. P. M. de Oliveira, and P. Quaresma, "Measuring the reputation in user-generated-content systems based on health information," *Procedia Comput. Sci.*, vol. 29, pp. 364–378, Jan. 2014.

[118] F. Schultz, S. Utz, and A. Göritz, "Is the medium the message? Perceptions of and reactions to crisis communication via Twitter, blogs and traditional media," *Public Relations Rev.*, vol. 37, no. 1, pp. 20–27, Mar. 2011.

[119] G. Ghosh, S. Banerjee, and N. Y. Yen, "State transition in communication under social network: An analysis using fuzzy logic and density based clustering towards big data paradigm," *Future Gener. Comput. Syst.*, vol. 65, pp. 207–220, Dec. 2016.

[120] X. Teng, S. Pei, F. Morone, and H. A. Makse, "Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks," *Sci. Rep.*, vol. 6, no. 1, Dec. 2016, Art. no. 36043, doi: 10.1038/srep36043.

[121] J. Zhang, R. Zhang, J. Sun, Y. Zhang, and C. Zhang, "TrueTop: A sybil-resilient system for user influence measurement on Twitter," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2834–2846, Oct. 2016.

[122] J. Y. M. Nip and K.-W. Fu, "Networked framing between source posts and their reposts: An analysis of public opinion on China's microblogs," *Inf. Commun. Soc.*, vol. 19, no. 8, pp. 1127–1149, Aug. 2016, doi: 10.1080/1369118X.2015.1104372.

[123] D. Surian, D. Q. Nguyen, G. Kennedy, M. Johnson, E. Coiera, and A. G. Dunn, "Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection," *J. Med. Internet Res.*, vol. 18, no. 8, p. e232, Aug. 2016, doi: 10.2196/jmir.6045.

[124] D. Xia, S. Mankad, and G. Michailidis, "Measuring influence of users in Twitter ecosystems using a counting process modeling framework," *Technometrics*, vol. 58, no. 3, pp. 360–370, Jul. 2016, doi: 10.1080/00401706.2016.1142906.

[125] F. Celli, A. Ghosh, F. Alam, and G. Riccardi, "In the mood for sharing contents: Emotions, personality and interaction styles in the diffusion of news," *Inf. Process. Manage.*, vol. 52, no. 1, pp. 93–98, Jan. 2016, doi: 10.1016/j.ipm.2015.08.002.

[126] D. H. Alahmadi and X.-J. Zeng, "ISTS: Implicit social trust and sentiment based approach to recommender systems," *Expert Syst. Appl.*, vol. 42, no. 22, pp. 8840–8849, Dec. 2015, doi: 10.1016/j.eswa.2015.07.036.

[127] L. Weng and F. Menczer, "Topicality and impact in social media: Diverse messages, focused messengers," *PLoS ONE*, vol. 10, no. 2, Feb. 2015, Art. no. e0118410, doi: 10.1371/journal.pone.0118410.

[128] F. Pérez-Tellez, J. Cardiff, P. Rosso, and D. Pinto, "Disambiguating company names in microblog text using clustering for online reputation management," *Revista Signos*, vol. 48, no. 87, pp. 54–76, 2015.

[129] S. K. Bansal and A. Bansal, "Reputation-based Web service selection for composition," in *Proc. IEEE World Congr. Services*, Jul. 2011, pp. 95–96.

[130] S. Garg, A. Saini, and N. Khanna, "Is sentiment analysis an art or a science? Impact of lexical richness in training corpus on machine learning," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2016, pp. 2729–2735.

[131] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By, "Sentiment analysis on social media," in *Proc. IEEE/ACM ICASNAM*, Istanbul, Turkey, Aug. 2012, pp. 919–926.

[132] Z. Yin and Y. Zhang, "Measuring pair-wise social influence in microblog," in *Proc. Int. Conf. Privacy, Secur., Risk Trust Int. Conf. Social Comput.*, Sep. 2012, pp. 502–507.

[133] M. Danisch, N. Dugué, and A. Perez, "On the importance of considering social capitalism when measuring influence on Twitter," in *Proc. BESC-ICBES-CC*, Shanghai, China, 2014, pp. 1–9.

[134] T. Rao and S. Nagpal, "Real-time geo influence in social networks," in *Proc. 3rd Int. Conf. Electron. Comput. Technol.*, Apr. 2011, pp. 246–250.

[135] K. Lerman, P. Jain, R. Ghosh, J.-H. Kang, and P. Kumaraguru, "Limited attention and centrality in social networks," in *Proc. Int. Conf. Social Intell. Technol.*, May 2013, pp. 80–89.

[136] Y. Blanco-Fernandez, M. Lopez-Nores, J. J. Pazos-Arias, and M. I. Martin-Vicente, "Spreading influence values over weighted relationships among users of several social networks," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. Workshops*, Mar. 2012, pp. 149–154.

[137] R. Mehmood, H. Maurer, and M. T. Afzal, "Knowledge discovery in hashtags#," in *Proc. IEEE 9th Int. Conf. Emerg. Technol. (ICET)*, Dec. 2013, pp. 1–6.

[138] Y. Zhang, J. Mo, and T. He, "User influence analysis on micro blog," in *Proc. IEEE 2nd Int. Conf. Cloud Comput. Intell. Syst.*, Oct. 2012, pp. 1474–1478.

[139] S. Yoshida, J. Kitazono, S. Ozawa, T. Sugawara, T. Haga, and S. Nakamura, "Sentiment analysis for various SNS media using Naïve Bayes classifier and its application to flaming detection," in *Proc. IEEE Symp. Comput. Intell. Big Data (CIBD)*, Dec. 2014, pp. 1–6, doi: 10.1109/CIBD.2014.7011523.

[140] T. Ohbe, T. Ozono, and T. Shintani, "Developing a sentiment polarity visualization system for local event information analysis," in *Proc. 5th IIAI Int. Congr. Adv. Appl. Informat. (IIAI-AAI)*, Jul. 2016, pp. 19–24.
[141] F. Peleja, J. Santos, and J. Magalhaes, "Ranking linked-entities in a sentiment graph," in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. (WI) Intell. Agent Technol. (IAT)*, Aug. 2014, pp. 118–125.
[142] S. Hirokawa and J. Zeng, "Comparison of tourism data using double ranking," in *Proc. 1st ACIS Int. Symp. Softw. Netw. Eng.*, Dec. 2011, pp. 67–72.
[143] E. Koutrouli, G. Kanellopoulos, and A. Tsalgatidou, "Reputation mechanisms in on-line social networks: The case of an influence estimation system in Twitter," in *Proc. SouthEast Eur. Design Autom., Comput. Eng., Comput. Netw. Social Media Conf. SEEDA-CECNSM*, 2016, pp. 98–105.
[144] M. de Maertelaere, T. Li, and G. Berens, "Social influence: The effect of Twitter information on corporate image," in *Proc. 14th Annu. Int. Conf. Electron. Commerce ICEC*, 2012, pp. 292–293, doi: 10.1145/2346536.2346595.
[145] T. Afroze, "Exploring the similarities of influencers in online brand communities," in *Proc. 7th Int. Conf. Social Media Soc. SMSociety*, 2016, pp. 1–9, doi: 10.1145/2930971.2930981.
[146] P.-Y. Huang, H.-Y. Liu, C.-H. Chen, and P.-J. Cheng, "The impact of social diversity and dynamic influence propagation for identifying influencers in social networks," in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. (WI) Intell. Agent Technol. (IAT)*, Nov. 2013, pp. 410–416.
[147] B. Markscheffel and J. Schmidt, "The scientific influence passivity score (SIP): Bibliometrics learning from Twitter," in *Proc. 6th Int. Conf. Manage. Emergent Digit. EcoSyst. MEDES*, 2014, pp. 20–24.
[148] J. Hopcroft, T. Lou, and J. Tang, "Who will follow you back?: Reciprocal relationship prediction," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage. CIKM*, 2011, pp. 1137–1146.
[149] R. Belkaroui and R. Faiz, "Towards events tweet contextualization using social influence model and users conversations," in *Proc. 5th Int. Conf. Web Intell., Mining Semantics WIMS*, 2015, pp. 1–9.
[150] M. Jeragh, E. AlQuraishi, and E. AlDwaisan, "A Twitter-based weighted reputation system," *Procedia Comput. Sci.*, vol. 10, pp. 902–908, Jan. 2012, doi: 10.1016/j.procs.2012.06.119.
[151] R. Page, "Saying 'sorry': Corporate apologies posted on Twitter," *J. Pragmatics*, vol. 62, pp. 30–45, Feb. 2014, doi: 10.1016/j.pragma.2013.12.003.
[152] H. Shad Manaman, S. Jamali, and A. Aleahmad, "Online reputation measurement of companies based on user-generated content in online social networks," *Comput. Hum. Behav.*, vol. 54, pp. 94–100, Jan. 2016, doi: 10.1016/j.chb.2015.07.061.
[153] N. D. Van Norel, P. A. M. Kommers, J. J. Van Hoof, and J. W. M. Verhoeven, "Damaged corporate reputation: Can celebrity tweets repair it?" *Comput. Hum. Behav.*, vol. 36, pp. 308–315, Jul. 2014, doi: 10.1016/j.chb.2014.03.056.
[154] R. Rutter, S. Roper, and F. Lettice, "Social media interaction, the university brand and recruitment performance," *J. Bus. Res.*, vol. 69, no. 8, pp. 3096–3104, Aug. 2016, doi: 10.1016/j.jbusres.2016.01.025.
[155] K. Ntalianis, N. Papadakis, and P. Tomaras, "Reputation monitoring over rule-stringent social media based on advanced wrapper technologies," *Procedia Social Behav. Sci.*, vol. 148, pp. 559–566, Aug. 2014, doi: 10.1016/j.sbspro.2014.07.079.
[156] S. Wan, R. Koh, A. Ong, and A. Pang, "Parody social media accounts: Influence and impact on organizations during crisis," *Public Relations Rev.*, vol. 41, no. 3, pp. 381–385, Sep. 2015, doi: 10.1016/j.pubrev.2015.01.002.
[157] Y. Bai, Z. Yao, and Y.-F. Dou, "Effect of social commerce factors on user purchase behavior: An empirical investigation from renren.Com," *Int. J. Inf. Manage.*, vol. 35, no. 5, pp. 538–550, Oct. 2015, doi: 10.1016/j.ijinfomgt.2015.04.011.
[158] A. Z. Bahtar and M. Muda, "The impact of user—Generated content (UGC) on product reviews towards online purchasing—A conceptual framework," *Procedia Econ. Finance*, vol. 37, pp. 337–342, Jan. 2016.
[159] Y. Han, B. Fang, and Y. Jia, "Predicting the topic influence trends in social media with multiple models," *Neurocomputing*, vol. 144, pp. 463–470, Nov. 2014, doi: 10.1016/j.neucom.2014.03.054.
[160] S. A. Williams, M. M. Terras, and C. Warwick, "What do people study when they study Twitter? Classifying Twitter related academic papers," *J. Document.*, vol. 69, no. 3, pp. 384–410, May 2013.
[161] C. Sanderson and S. Guenter, "Short text authorship attribution via sequence kernels, Markov chains and author unmasking: An investigation," in *Proc. Conf. Empirical Methods Natural Lang. Process. EMNLP*, 2006, pp. 482–491.
[162] A. Tedeschi and F. Benedetto, "A cloud-based big data sentiment analysis application for enterprises' brand monitoring in social media streams," in *Proc. IEEE 1st Int. Forum Res. Technol. Soc. Ind. Leveraging Better Tomorrow (RTSI)*, Sep. 2015, pp. 186–191, doi: 10.1109/RTSI.2015.7325096.

SAMI AL-YAZIDI received the B.E. degree in software engineering from the King Fahad University of Petroleum and Minerals (KFUPM) in Dhahran, Saudi Arabia, in 2008, the master's degree with the Information Systems Department of Computer and Information Sciences College (CIS), King Saud University (KSU), in 2011, where he is currently pursuing the Ph.D. degree in information systems working on social networks and machine learning fields. His research interests include social media analysis, data analytics and mining, social computing, information credibility, data science, and service-oriented architecture (SOA). He published an articles in International *Journal of Software and Web Sciences* (IJSWS) titled as *Monitoring SOA Based Applications According to Business Level Agreement*. He is Sun Certified Java Programmer, Project Manager Professional, and ITIL certified.

JAWAD BERRI received the Ph.D. degree in computer science from Paris-Sorbonne University, France, in 1996. His research interests focus on learning technologies, context-aware mobile systems, and natural language processing. He has been involved in many projects related to e-learning, mobile learning, semantic web, automatic summarization, web information filtering, and mobile agents for web information discovery. He worked as a Researcher at the CNRS–French National Research Centre and at the Computer Science Institute, University of Zurich, Switzerland. He worked in the industry sector at Sonatrach–the Algerian Petroleum and Gas Corporation as a Senior Researcher to initiate the Department Of Research And Development. In academia, he has been a Faculty Member with the United Arab Emirates University, Khalifa University of Science Technology and Research. He is currently at King Saud University, Saudi Arabia. His contributions in research projects in the industry and academia led to the publication of articles in numerous journals and conferences.

MUHAMMAD AL-QURISHI received the Ph.D. degree from the College of Computer and Information Sciences (CCIS), King Saud University (KSU), Riyadh, Saudi Arabia, in 2017. He is currently a Postdoctoral with the Chair of Pervasive and Mobile Computing (CPMC), CCIS, KSU. He is one of the founding members of CPMC. He has published several articles in refereed journals (IEEE, ACM, Springer, and Wiley). His research interests include data science, big data analysis and mining, pervasive computing, and machine-learning. He received the Innovation Award for a mobile cloud serious game from KSU 2013.

MAJED AL-ALRUBAIAN received the M.S. degree in information systems and the Ph.D. degree in computer and information sciences from the College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia, in 2004 and 2016, respectively. He is currently involved in the academic field of one of the major colleges with security responsibilities in Saudi Arabia. His research interests include social media analysis and mining, social computing, information credibility, human–computer interaction, the social-IoT, and cyber security.