IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON APPLICATIONS OF BIG DATA IN SOCIAL SCIENCES

Received October 21, 2018, accepted November 5, 2018, date of publication December 11, 2018,
date of current version January 7, 2019.

*Digital Object Identifier 10.1109/ACCESS.2018.2886314*

# Credibility in Online Social Networks: A Survey

**MAJED ALRUBAIAN[1,2], (Member, IEEE), MUHAMMAD AL-QURISHI[1,2], (Member, IEEE), ATIF ALAMRI[1,3], (Member, IEEE), MABROOK AL-RAKHAMI[1,2], (Student Member, IEEE), MOHAMMAD MEHEDI HASSAN[1,2], (Senior Member, IEEE), AND GIANCARLO FORTINO[4], (Senior Member, IEEE)**

[1]Research Chair of Pervasive and Mobile Computing, King Saud University, Riyadh 11543, Saudi Arabia
[2]Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[3]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[4]Department of Informatics, Modeling, Electronics, and Systems, University of Calabria, 87036 Rende, Italy

Corresponding author: Mohammad Mehedi Hassan (mmhassan@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research and Research Chair of Pervasive and Mobile Computing at King Saud University.

**ABSTRACT** The importance of information credibility in society cannot be underestimated given that it is at the heart of all decision-making. Generally, more information is better; however, knowing the value of this information is essential for the decision-making processes. Information credibility defines a measure of the fitness of the information for consumption. It can also be defined in terms of reliability, which denotes the probability that a data source will appear credible to the users. A challenge in this topic is that there is a great deal of literature that has developed different credibility dimensions. In addition, information science dealing with online social networks has grown in complexity, attracting interest from researchers in information science, psychology, human–computer interaction, communication studies, and management studies, all of whom have studied the topic from different perspectives. This work will attempt to provide an overall review of the credibility assessment literature over the period 2006–2017 as applied to the context of the microblogging platform, Twitter. The known interpretations of credibility will be examined, particularly as they relate to the Twitter environment. In addition, we investigate levels of credibility assessment features. We then discuss recent works, addressing a new taxonomy of credibility analysis and assessment techniques. At last, a cross-referencing of literature is performed while suggesting new topics for future studies of credibility assessment in a social media context.

**INDEX TERMS** Online social networks, credibility assessment, Twitter.

## I. INTRODUCTION

Every day, millions of people join online social networks (OSNs) regardless of gender, age, social, economic, or religious classifications [1]. Among existing social platforms, clear market leaders are Facebook followed by Twitter, hosting 2 billion and 640 million respective users, including 1.33 1.32 billion and 328 271 million active monthly users [2]–[4]. Accordingly, through user activities, i.e., interactions of the users with the OSN, terabytes of data are generated every second [5]–[7]. This vast and rich collection of user-generated content includes people's opinions about events and products; personal ideas, feelings, and interests; opinions about current societal debates and governmental policies; and much more [8].

Access to such data may be very interesting to a wide range of organizations [3], since patterns of action can be deducted based on the input originating on social networks [5], [6]. Furthermore, social media can be used by political parties for collecting funds and appealing to the voters. Per [9], one of the reasons that Barack Obama was so successful in mobilizing the youth vote was his campaign team's social media proficiency. The remarkable amount of money raised all over the US, along with the tapping into the public sentiment of wanting ''change,'' was undoubtedly helped by OSNs. Similarly, in the 2014 Indian parliamentary election, there was a tremendous surge in the use of social media by political parties for campaigning, sentiment polarization, mass interaction, manifesto propagation, and fundraising. Thus, it is evident that online social media assumes a rightful place in the chain of distribution of news, advertisements, fundraising, facilitating political campaigns, and even in revolutions, such as the Arab Spring [10].

Despite the immense potential of OSNs, this technology is also misused to execute a number of undesirable acts,

for example generating spam, rumors, fake messages, and fake accounts, to gain stronger influence, create chaos, or destabilize homeland security [9], [11]–[13]. Spammers employ a myriad of techniques [12] to bombard social media members with irrelevant and unsolicited content. Messages of this type either pretend to be advertisements or attempt to perform some other fraud and help execute phishing attacks or spread viruses through included links. For example, in August 2009, approximately 11% of all Twitter posts were considered spam, and in May 2009, several users' Twitter accounts were hacked and advertisements were propagated from them [14]. In addition, Twitter is becoming a hotbed for rumor dissemination and sharing [15].

Data have been misused, intentionally or unintentionally, to generate fake and dubious content. In the case of unintentionally created fake data, many users promote and share important news without verifying it. Hence, a rumor can be quickly transformed into new, official-looking content that claims to be a true news item. For example, Twitter posts (called tweets) about swine flu caused widespread public panic in 2009 [16], [17]. In another case, appearance of false claims about health insurance reform prompted negative public opinion and forced U.S. administration to issue a clarification [17].

Research on credibility analysis in OSNs has increased enormously over the last eight years, as shown in Figure 1. However, there are many challenges in determining the credibility of a user in a social network. First and foremost is the tremendous magnitude of OSN users and their highly clustered structure [8]. By their nature, OSNs evolve dynamically to grow to a tremendous size, and may include certain features that obscure the information used to discern users' credibility. Another challenge is that the reliability of any social platform member is affected by his relationships with other members, as well as temporary social alignment [18]. A third challenge is that ill-intentioned members are able to circumvent currently used defenses. For instance, within Twitter environment it's relatively easy to buy instant popularity or employ software to create a large number of bot accounts and large amounts of spammy content [11].

With this in mind, it's very challenging to ascertain trustworthiness of social platform members and the content they generate. As OSNs are increasingly essential for distribution of information to the general population, solving the aforementioned challenges in user credibility determination in OSNs requires developing strong techniques for measuring user and content credibility. This work presents a comprehensive survey of the literature related to user and content credibility assessment in the well-known microblog, Twitter. One similar survey has been done, though it focused mostly on measuring the credibility of high-impact events, such as the London riots (2011) and Assam riots (2012), and the adverse impact of online reports to happenings in the real world [19]. Hence, a comprehensive survey concentrating on various methodologies of and issues regarding credibility assessment and analysis is still needed. This study aims to contribute such a survey. Measuring and analyzing content/user credibility have been addressed over different web content services [20]–[22] and blogs [23]; however, the focus of this survey is restricted to assessing the credibility of content and users on Twitter. Accordingly, this work comments on the state-of-the-art literature that employ various credibility assessment methods, for instance, machine learning and human-based approaches. It also provides several classifications of the current literature based on the approaches used and the level of credibility assessment. Several issues that influence the credibility assessment process, such as user reputation, context/event-based assessment, and trust-based assessment, are also discussed critically.

The rest of this study is structured in the following manner. Section 2 gives a short summary of credibility definitions from different disciplines (followed by their properties and measurements) and the relationship between credibility and trust. Section 3 presents a summary of existing related surveys. Section 4 discusses the status of credibility assessment features extracted from the literature. Section 5 presents approaches to credibility assessment and their performance. We also outline the importance of measuring user and content credibility in OSNs in this section. Section 6 examines the literature based on the new taxonomy. Section 7 summarizes the most important projects and systems. In Section 8, we identify the benchmark datasets and matrices used. Section 9 rounds up the study, finalizing the entire work and indicating future directions for the field.

## II. BACKGROUND

Credibility as a concept has been attracting attention since the Internet revolution of the late 1990s allowed users to interact, communicate, and generate content with little reference to sources. Thus, studies and analyses of credibility have been performed by researchers from diverse disciplines and different perspectives [24], such as information science, marketing [25], management information, communications, web engineering, information retrieval (IR) [26], human-computer interaction (HCI) [27], and psychology [28].

One must note that there are different types of credibility, such as source, media, and message credibility, and assessments of these objects differ [24]. Concepts discussed along with credibility include trust, reliability, and reputation [24], [29], [30].

### A. ONLINE SOCIAL NETWORKS (OSNs)

Social networks are internet-based platforms that enable users to create broadly or partially visible profiles with a set of clearly defined rules. Social networks can also be understood as open or partially open internet services that simplify communication and relationships between individual with similar mindsets, or connecting large number of such individuals into thematic groups [31].

OSNs allow people to connect to users with whom they share interests and to move through their first-level and

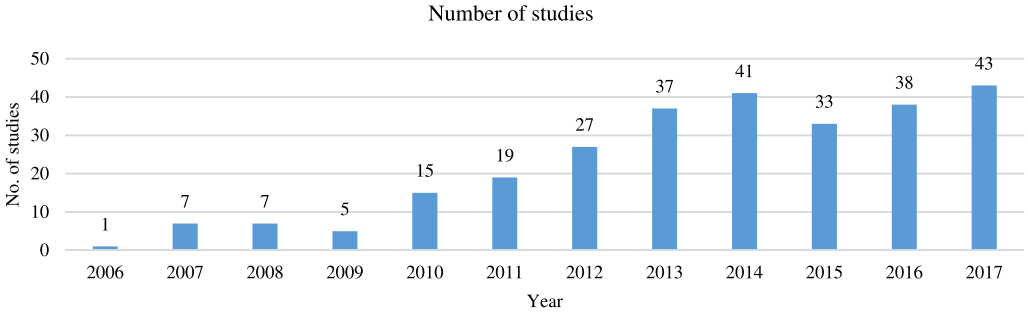


**FIGURE 1.** Research on credibility assessment in social networks, 2006–2017.

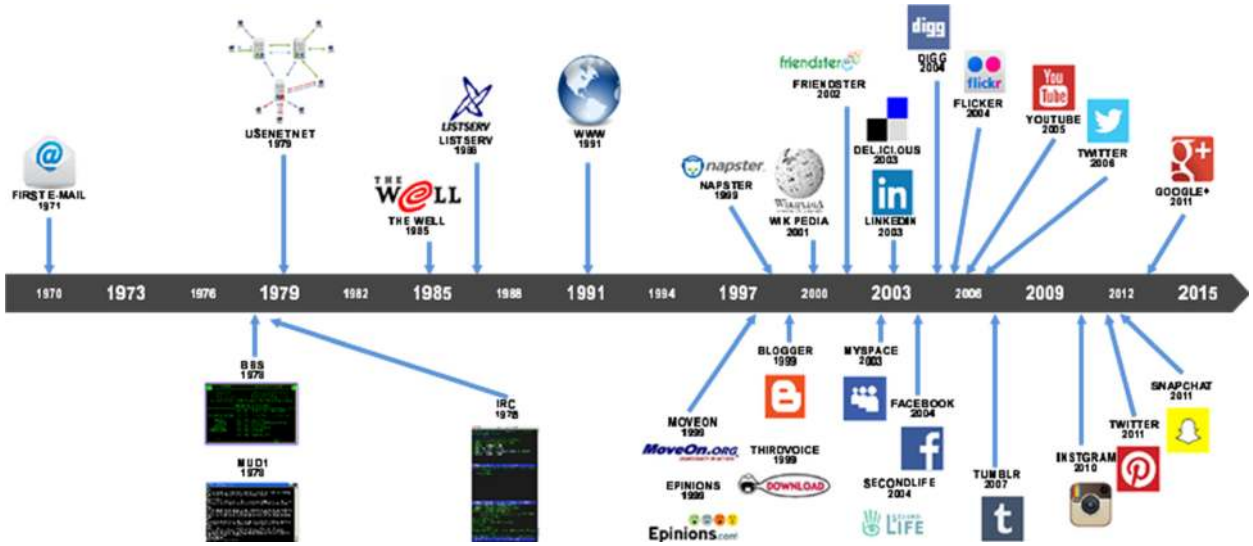


**FIGURE 2.** Evolution of online social media and networks.

second-level contacts within the network. In other words, social platforms are constructs consisted of numerous points bound together by links that symbolize a certain type of mutual relationship [3], [6], [11]. Points stand for people, informal societies and companies, while the tying links symbolize interactions, such as agreements, personal closeness, or business ties [32]. The idea of a social platform gradually changed, so right now analytic description of such networks represents its own science with distinct methodologies, tools and specialists [33]. Many scientific works addressed various characteristics of social media platforms and their responses to external factors. Figure 2 shows the proliferation of OSNs as technology has evolved.

Most relevant work in assessing the credibility of content and users on social networks focuses on the role of OSNs during unfolding news and real-world events, such as earthquakes, volcanic eruptions, and other high-profile occurrences that instigate interactions between users. Therefore, in our survey, we focus on credibility analysis studies of microblog networks such as Twitter.

The importance of Twitter in society today cannot be overstated. Twitter is by far the most popular microblogging site in the world, with hundreds of millions of users. Twitter is an excellent platform for real-time information dissemination owing to its characteristic 140-character-long messages, making it an invaluable source of news. Members of this network can publish or see short messages, most of which are public and can be viewed without registering for a Twitter account. To receive tweets from a user automatically, others can follow that user. Twitter's follow relationship is asymmetrical: one user following another does not require the converse to be true. Twitter deserves to be described as a news media channel as much as a social platform, with certain members relaying information to others [13]. This characteristic is specific for Twitter and can't be found on typical social platforms, which allow two-way communication between the connected members. Still, two-way communication can be found on Twitter; it can occur if two members become each other's followers. Starting from the aforementioned characteristic, we posit that Twitter is general and mirrors real-life information propagation [13].

### B. DEFINITIONS

The word credibility dates to the middle of the 16th century, from the medieval Latin credibilitas, which is in turn from the Latin credibilis. It also has an origin in American English in reference to official statements about the Vietnam War. As stated in the Oxford English Dictionary, this term has

the meaning of ''the concept of eliciting confidence,'' or ''inherent persuasiveness and truthfulness'' [34]. The most correlated synonyms of credibility are trustworthiness and believability. In our review of information credibility on Twitter, we formulate three types of use for this term:

1. *Single post credibility* occurs when a single post (tweet) is believable and reliable, which means that the message includes relevant and accurate information about a certain topic.
2. *Member credibility* is the reliability of a user account, as measured with a score calculated for every user in an OSN. The less reliable a member is, the less likely it is that messages coming from this account are trustworthy,
3. *Topic-level credibility* is the believability, reliability, and acceptance of a topic or event, calculated as a numerical score for each tweet regarding that topic/event.

Researchers have expended great effort in studying aspects of information credibility on Twitter. Kang [35] defines a fourth type of Twitter credibility called social credibility, which is the expected believability of a user based on his/her status in the social network on a topic domain, given all available metadata.

Castillo *et al.* [36] produced a concept of credibility similar to our single post trustworthiness, though it lacked the topic-level constraint that we developed. Those three types of credibility are synchronized with each other; as trustworthiness of a single message affects overall credibility of its author.

### C. CREDIBILITY AND TRUST

Credibility is synonymous with believability [37]. Credibility beliefs stem from evaluating the attributes of an attitude object, resulting in perceptive knowledge that guides feelings and actions [38]. From this perspective, credibility beliefs derive from a cognitive process. Assessing content/user credibility helps people to determine whether the given content/user can be trusted for information. Because social network visitors are interested in finding reliable information, the issue of trusting users and their content is rooted in credibility [23]. Despite their different meanings, the terms credibility and trust are often erroneously interchanged in academic and professional literature because their meanings are related. Nonetheless, the differences between credibility and trust relate, respectively, to believability and dependability. This notion of credibility as an antecedent to trust is supported in the literature [39]–[41]. Social marketing research has found that credibility has a significant association with trust that leads to an intention to engage in campaign efforts [39]. Using interviews, Arnott *et al.* [42] found that credibility had a high explanatory power in an analysis of the predecessors of trust in an organizational brand. In addition, Wakefield and Whitten [41] observed a positive relationship between high credibility scores and buyer confidence in an online shopping context.

In a related assessment methodology, Kang [35] cited several recent studies on adopting reputation-based and policy-based trust methods for measuring the trust in and credibility of information. Moreover, Kang [35] executed a trust-finding procedure to judge the reliability of microblogging content. The policy, articulated in the form of a scientific model, tracked historical metadata, such as the number of contacts, how old the Twitter account is, and how well it's connected with influential members. To develop trusts within a policy-based trust approach, some specific criteria should be evaluated. In this respect, the verification of credentials is frequently involved [43]. Using a reputation-based trust approach, this assessment methodology depends on the historical interactions of a user, sometimes first-level communication with the person making the assessment or based on recommendations of independent, unrelated members. In general, both policy- and reputation-based trust methods share the involvement of a third-party verification process; nevertheless, they have different assessment conditions [35].

## III. PREVIOUS SURVEYS

In this survey, a large body of research is discussed and studied in terms of credibility in online social networks. We investigate 192 research papers in the credibility assessment domain and more than 92 social media analysis papers from several aspects and different approaches. As far as we are informed, there has never been a similar study that aimed to classify recent work related to trustworthiness of social media content. In this section, we review the most important studies close to the subject of our survey; these are not directly related, but converge in their content or aim to measure credibility. In this sense, the nearest scientific topic is trust.

The field of information credibility has increasingly received attention from researchers interested specifically in online credibility after social media platforms became globally popular [20], [22], [28]. Currently there is no wide-ranging agreement regarding possible courses of action to improve overall credibility. This topic has been studied as part of several other topics, for example, information diffusion, trust, recommendation, and reputation; nevertheless, it requires further, specialized study. Only one survey so far has focused on assessing credibility on Twitter when high-impact events occur [19]. It discussed some aspects of spam and phishing detection on Twitter and proposed techniques to remove spam from Twitter. In addition, it described trust/credibility assessment in terms of systems developed to determine how trustworthy information is on Twitter.

Ali Shah *et al.* [20] center their attention on web credibility pertaining to digital content. There is widespread awareness of digital content dissemination owing to the emergence of blogs, wikis, and social networking platforms. This has in turn spurred the growth of valuable communication through online channels. However, on the downside, content bias, and demagogic or false information is also becoming a major challenge from these channels. This paper thus surveys the

dynamics of web credibility assessment through an analysis of three major subtopics. This is achieved through:

1. Analyzing users' perceptions of web credibility.
2. Analyzing the factors adopted in web credibility assessment.
3. Crafting a hybrid model to cater to an array of credibility judgment techniques

Trust is the basis of decision-making and forms the basis of creating and maintaining collaborations [44]. It thus becomes important to find algorithms that model and measure trust using sufficient detail and framework-based adequacy. The clear challenge in this area is that trust quantification has grown very complex, as illustrated by the need to derive trust from complex or composite networks. Deriving trust from modern networks is challenging, in the sense that it usually calls for sifting through up to four distinct layers: communication protocols, information exchange, social interactions, and cognitive motivations.

Cho *et al.* [44] use a stepwise approach in their survey, starting by providing the reader with different definitions of trust from a multidisciplinary perspective. The authors then develop an algorithm that can be used to analyze trust assessment. Here, they describe terms like trustor, trustee, and risk assessment based on Romano's classification of the concept of trust. This section also defines the factors that affect trust between entities, e.g., social relationships or psychological states. In analyzing trust assessment, the concept of trust is classified into phenomenon-, sentiment-, and judgment-based trust. Phenomenon-based trust affects how a trustor forms trust with the trustee. Sentiment-based trust involves an analysis of why a trustor should or should not trust a trustee. Judgment-based trust defines how trust is measured and updated. In their methodology, they illustrate methods of trust measurement, starting by differentiating trust from mistrust, distrust, undistrust, and misdistrust. These distinctions are made through mathematical notations accrued from various literature. For example, trust is defined through the notation T(i,j, $\alpha$) (i.e., the entity i trusts j in situation $\alpha$). The authors then scale trust using binary, discrete, and nominal scales, with discrete and continuous scales favored for the ease they afford users in spotting outliers.

The next part of the methodology from Cho *et al.* classifies factors affecting trust considering that trust can be derived from individual and relational features. In this classification, the authors establish two broad trust constructs: relational and individual trust attributes. Individual trust attributes are further categorized into logical and emotional trust. The methodology then delves more deeply into these factors while providing models that estimate them. In logical trust, the authors model factors affecting an entity's trust that are derivable from observations and evidence, including belief, confidence, experience, frequency, certainty, competence, honesty, integrity, recency, stability, relevance, credibility, completeness, cooperation, rationality, reliability, dependability, and availability. In emotional trust, expectation, hope, fear, frustration, disappointment, relief, disposition, and regret are modeled. Relational trust attributes dwell on modeling trust as attributes of collective units using similarity, centrality, and importance.

A proposal of composite trust is then made. Composite trust depicts features like communication trust, social trust, information trust, and cognitive trust, which can be attained from a complex, composite social network. Communication trust is simply the trust that is a consequence of communication networks consisting of links and nodes connected via cable or wireless connectivity protocols. Information trust, on the other hand, indicates trust in information networks that offer information services; it includes aspects like quality of information and credibility. Cognitive trust is the trust derived from a thought process. Lastly, social trust is indicative of trust between humans. There have been efforts to illustrate composite trust in human-machine research, a good example being automation as defined by Muir in 1994. In this model, trust can be defined by the following mathematical formula:

$$Trust = Regularity + Consistency + Belief + Knowledge + Accountability + Availability$$

Their discussion ends with an analytic overview of the properties of trust, including subjectivity, asymmetry, dynamicity, incomplete transivity, and context dependency. Cho *et al.* [44] analyze the concepts of trust in different domains. They do not, however, close all the loopholes, given that there are still design challenges in developing trust models. The first challenge is that it is difficult to verify and validate trust models. The dynamics of the network environment may also affect estimation of trust. These are among the issues that should be covered in next surveys related to trust models.

More closely related to credibility, trust has been examined in areas such as informatics, social science, psychology, and economics [44], [45]. The effectiveness of these endeavors depends on the reliance individuals place on one another and on businesses or governing bodies. Hence, reliance is a primary requirement in a social setup. Sherchan *et al.* [45] consider the status quo regarding the meaning of trust and describe trust in social networks. They classify trust into six categories. Calculative trust is the solution to a computation the trustor makes to optimize his investment in the relationship. Relational trust is reliance that increases as time goes on. Emotional trust is the safety and peace of mind felt by relying on someone trusted. Cognitive trust is founded on rationality and a sane disposition. Institutional trust is nurtured by an institution giving rewards and punishments for good and bad behavior, respectively. Dispositional trust is based on an individual's previous experiences of trusting others. Despite this literature, an all-inclusive survey concentrating on approaches and issues related to assessing credibility on online social networks (specifically Twitter) is still missing. Sherchan *et al.* [45] tried to bridge the gap resulting from the failure to integrate sociological, psychological, and computer-based aspects of trust in social networks. They

analyzed social trust literature from these perspectives, but do not provide a comprehensive answer for the question of confidence on social platforms.

A critically important challenge in mass emergencies is that they usually present uncertainties coupled with minimal time in which to react. As stated earlier, social media channels have grown in members, meaning that if they are to be used during a crisis, then information sifting must extract information to suit the specific situation. Imran et al [46] gave a literature review of systems used in social media monitoring. The systems provided in this case are based on a dashboard that delivers visual data during disasters. They highlight key techniques used to collect, represent, and process social media data, first by showing the characteristics of social media messages posted during a crisis (Twitter is used in this scenario). The authors show how data from social media platforms can be acquired using an application programming interface (API). Two APIs are currently available to users: streaming and search APIs. However, the Twitter information that can be queried is limited compared with other social media platforms.

Imran *et al.* [46] illuminate the preprocessing of data using natural language processing (NLP) algorithms such as use of tokens, speech tags, label assignment to interdependent semantic structures, identification of defined elements and their connections, filtering, deduplication, and feature extraction. The importance of attaching geographical locations to the data in a process usually known as geotagging and geocoding is also reiterated. Geotagging is important in the sense that it facilitates the retrieval of information pertaining to a given location. Data can be provided for analysis either in archived form or in live feeds, and data analysis of these forms are known, respectively, as retrospective and real-time analysis. Imran *et al.* [46] took a step forward by discussing the challenges involved in processing social media data, noting that scalability, content, and privacy issues pose major challenges. The major problem identified in this survey is that social media adoption for disaster management is still very new to large organizations. Considerations like privacy issues and limited data make the approach difficult to adopt. The future of this approach relies on the development of more efficient and effective data processing algorithms.

Mansour *et al.* [18], [47] gave a short review of current credibility assessments in microblogs. Mansour [48] discussed credibility in general and provided a simple comparison of various automatic information credibility assessment systems grounded in techniques and features. These authors also proposed a new credibility assessment model that considers contextual and cultural differences. One important shortcoming of this work is that its methodology and classification method were not explained and no results were shown. This work doesn't introduce any new technique for ascertaining credibility, but instead creates an overall framework and experimentally examines in which ways and to which degree the established methods can be useful to determine trustworthiness of online data from different sources.

## IV. LEVELS OF CREDIBILITY ASSESSMENT FEATURES

The credibility of Twitter content is generally assessed at three levels, the post-, topic-, and user-levels [36], [49], [50]. With each level, studies vary in their approaches, techniques, and methodologies, and are based on different models, features, levels of human involvement, and datasets [48]. Some researchers have taken hybrid approaches to credibility assessment at the topic level likewise at the post level, and some have conducted their experiments at all levels.

Figure 3 shows the number of studies done at each of the three levels of credibility assessment. We now describe those levels of credibility assessment and elucidate their advantages and disadvantages. Several works that have utilized different credibility assessment models at different levels are discussed in Section 6.

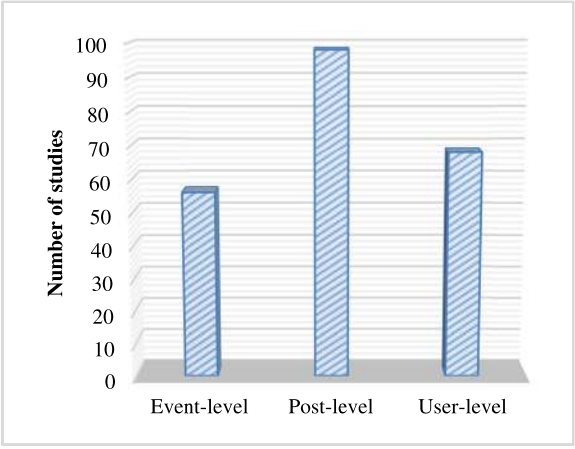


**FIGURE 3.** Levels of credibility assessment features used in the literature.

### A. POST LEVEL

At the post level, the task is to analyze the content attributes of a tweet to assess its credibility score and determine whether it is trustworthy [51], [52]. Research on this topic is divided into offline systematization of already present input [29], [53]–[57] and real-time systems that use only the data accessible in each post (not considering complete historical, user, or topic data) [58]–[61]. Starting from tweet attributes, characteristics like total # if characters in the tweet, total # of words it contains, total # of questions, and total # of uppercase characters are computed. At this level of credibility assessment, the number of features extracted from various post attributes may be diverse, but can be classified as follows:

1. *Message characteristics* consist of the semantic body of the tweet, with parameters like Tweet length; the volume of responses and/or republishing (which may indicate the relevance of the message); whether a tweet contains hashtags (#), @ mentions, or links; and

presence of standard of dynamic emojis. Moreover, the number of duplications, verbs, and nouns used to describe an event may affect the credibility of a tweet.

2. *Multimedia features* based on images, videos, and audio may be considered. Researchers extract only the metadata [62]–[64] of the media, such as description, title, size, video duration, average number of tags per photo, and average upload time between any two consecutive uploads. Some studies have shed light into the activities [58] related to propagating manipulated pictures in posts by using automated systems to distinguish genuine images from fake images posted on Twitter. Ginsca *et al.* [65] checked the credibility of image tagging on the photo sharing site Flickr. They assumed that credible users produce credible content and vice versa; nevertheless this premise could be misleading since there are instances of famous and credible users who spread fake information by mistake [183].
3. *Sentiment features* can derive from counting the volume of affirmative and critical phrases in a tweet starting from a prepared table of phrases. Park *et al.* [66] found that most high-impact events are negative and include extreme negative sentiment words and opinions.

Working at the tweet level is beneficial in that it can help in automatically measuring credibility [36], [55], as it is necessary only to adopt a method based on artificial intelligence and importance assessment to rank posts based on trustworthiness rating. Moreover, such assessment processes can be done in real time, such as with TweetCred [58]. However, at this level it is difficult to pass judgment on the credibility of an event or topic since the need to contain the whole message in 140 spaces creates some problems discussing universally relevant conversations. In other words, single messages fail to provide enough data to discern the themes that might be referred to in the messages.

### B. TOPIC LEVEL

Events are usually trending topics that attract many users who in turn start tweeting, commenting further, and retweeting about them. When a high-impact event occurs, thousands of posts are created each minute [6]. Most researchers [36], [52], [60], [66]–[71] start by collecting tweets about topics or events to analyze them and try to mitigate the spread of misinformation on Twitter during crisis events like earthquakes or explosions. Topic-based features focus on aggregating tweet-based features [36], [54]. Examples of this include the URL and hashtag fractions of the messages, the volume of affectionate phrases in a sample, and the median affection rating in the messages. Presence of duplicate tweets indicates that some members could sometimes repeat publishing of their messages.

In addition, the number of verbs and nouns used to describe an event are considered. Some authors have measured credibility at this level using topic and opinion classification [68], [69]. This method assumes that, to assess the information credibility of an event, one must account for the different opinions [72] of Twitter's many users worldwide. Nevertheless, using this approach to credibility assessment at the topic level tends to be improper, especially when involving crowds of unknowledgeable users or those who have no experience with a topic. Furthermore, assessing credibility at the topic level is more efficient than at the post level because the former works with enough content to develop the correct judgment. However, this level may suffer from fake accounts that can post misleading content on the same topic or for the same event.

### C. USER LEVEL

This level of credibility assessment depends on features extracted from user accounts and user-generated content. Certain characteristics from this group are invisible, while others are obvious from member accounts. Latent attributes have become a topic of significant interest among social media researchers and the industries built around utilizing and monetizing online social content [73], focusing, for instance, on age group, sex, school degrees, ideological affiliation, and even beverage choices [74]–[78]. The number of followers, friends, tweets, and retweets are explicit properties that determine how broadly tweets spread and how much they affect a user's reputation [29]. Despite their ability to measure user reputation and influence, user base features are critical in determining credibility for any news event or topic. However, users can easily obtain thousands of followers within minutes from Twitter follower markets [79], [80].

### D. HYBRID LEVEL

To utilize the advantages of post-, topic-, and user-level credibility assessments, many researchers have adapted hybrid credibility measures that combine the three levels. Hybrid-level credibility assessment is illustrated in [4], [36], and [81], in which credibility assessment models maintain complete entity (topic, post, and user) and relation (network formation) awareness to precisely judge information credibility.

A hybrid level can utilize the advantages of all three levels and their features. Therefore, many researchers [4], [55], [82]–[84] have used the hybrid level to eliminate different obstacles in trustworthiness analysis.

## V. CREDIBILITY ANALYSIS METHODS

This chapter is focused on multiple credibility analysis methods previously adopted by online social media analysis researchers. Figure 4 shows the share of each method type in the studies we surveyed. The strengths and weaknesses are also discussed. The credibility assessment methods are divided into three major categories: automation-based methods, human-based methods, and hybrid methods, which in turn are divided into further subcategories (shown in Figure 5). This categorization is based on the perspectives of the researchers and their understanding of the problem. A few methods treat this issue as a classification that needs

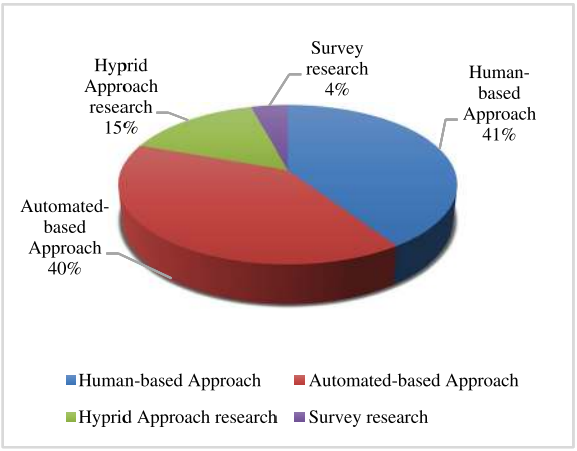


**FIGURE 4.** Numbers of researches regarding credibility analysis methods.

to be standardized through artificial intelligence and smart programming. In different methods this is understood as a cognitive task that demands human involvement. The hybrid methods combine methods from the main categories or from different subcategories. The next three sections are devoted to the classes of credibility analysis methods mentioned above.

### A. AUTOMATION-BASED APPROACHES

Automatically assessing the content credibility of microblogging information, which is created and disseminated at an unprecedented rate, is a crucial task [85]. There are insufficient resources for human operators to search for misleading messages related to universally relevant, significant and fast-changing events.

Many recent studies analyzing information credibility on Twitter use automated and semi-automated techniques, including supervised and unsupervised machine learning algorithms, weighted algorithms, and graph-based methods. Figure 6 depicts the subdivision of automation-based credibility analysis methods. In the following section, we describe some works that have adopted automation-based approaches.

### B. MACHINE-LEARNING-BASED TECHNIQUES

This part is dedicated to a summary of the various supervised techniques previously employed to perform credibility analysis tasks. Machine learning can autonomously acquire and integrate knowledge acquired from various sources. [86]. Such techniques can generally be classified in two groups:

1. *Supervised techniques*, including Support Vector Machine (SVM), Logistic/Linear regression models, Bayesian theory, Decision Tree methods.
2. *un-supervised techniques* for example cluster formation (e.g., k-means, fuzzy c-means, or hidden Markov models).
3. We add a third type, utilized in several studies [28], [50], [59], [87], [88]: semi-monitored techniques.

Machine-learning techniques were invented to complete tasks that include staggering volumes of information and demand tracking numerous parameters. Techniques from this group are often employed for voice and picture identification or determining credit ratings in banking [89].

#### 1) SUPERVISED APPROACHES

In the literature, various supervised techniques are used to analyze the features extracted from user profiles and interactions and adopt classification methods to identify non-credible and spam information [18], [28], [36], [49], [54], [89]–[93]. Most of the supervised learning approaches [47] are comprised of decision trees, decision rules, SVMs, and Bayesian algorithms.

Castillo *et al.* [36] were pioneers of trust assessment on Twitter. They showed that, based on various features, the J48 multiple path, SVM, and basic Bayesian techniques can be solid indicators of trustworthiness of topical content. Their method was able to successfully recognize 89.121% of topic appearances and their credibility classification accuracy reached almost 86%. Gupta *et al.* [51] relied on a couple of standardized methods for classification, the J48 decision tree and naïve Bayes approaches, to distinguish between fake and real images. J48 decision trees gave the best results with a prediction accuracy of about 97% in distinguishing fake images from real ones. Moreover, they demonstrated that programmed methods can recognize inaccurate images. Based on this foundation, a different study by the same team [54] used SVM and IR techniques to assess post-level credibility by sorting messages based on included words and member characteristics. Their results showed that information about a topic garnered an average of 30% of tweets, whereas 14% corresponded to spam. Only 17% of tweets were credible.

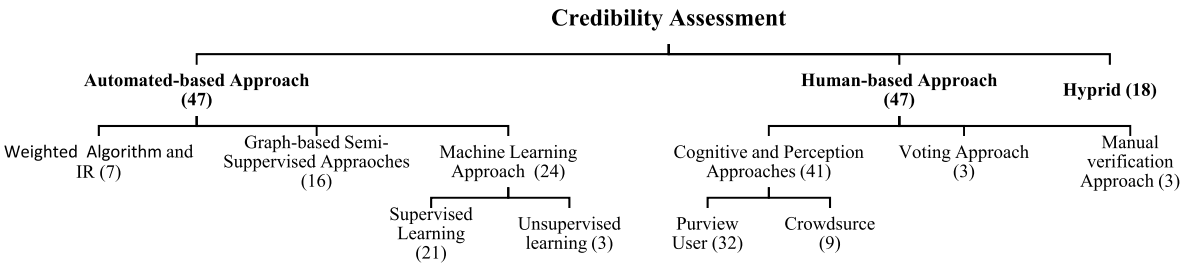


**FIGURE 5.** New classification of credibility assessment on microblogs.

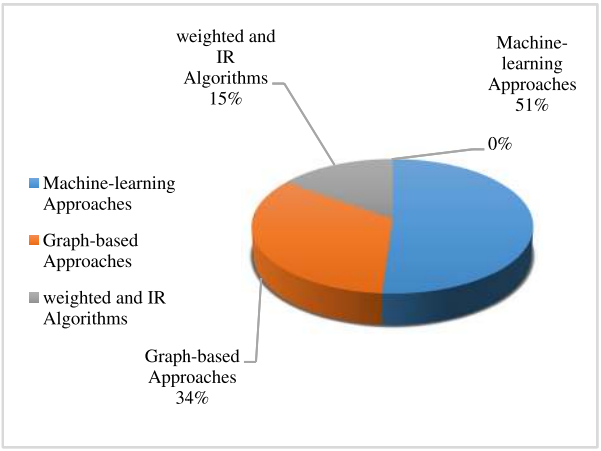


**FIGURE 6.** Studies of automated credibility analysis methods.

In other efforts [93], a monitored method with importance indication was chosen to classify messages, showing that these approaches give excellent results regarding crises.

In 2013, Castillo *et al.* published further work [94], [94] on credibility assessment of user tweets during crises. The method they used involves a sequence of monitored classifiers composed of two stages. First, they determined whether a message was relevant to the situation. Second, they labeled a message as credible or not. In the detection stage, they used Bayesian model, logistic regression, and J48 approaches to attain the best performance.

A number of systematization schemes were tried in order to obtain characteristics relevant for determining credibility. Castillo et al. [36] stated four main feature classes, including: (1) content-focused characteristics, pertaining to message itself; (2) member-focused characteristics, pertaining to the posting member; (3) promotion-focused characteristics, pertaining to the social media platform; and (4) topic-based features, which are aggregated and attributed to a specific topic. Castillo *et al.* utilize these features and a branching path method to come up with their systematization. However, using only these features without any awareness of latent attributes, such as the typical length of membership for members posting about a certain theme, may lead to low-accuracy results [95]. Some researchers [92], [96], [97] postulated that connections count can be a predictor of that member's social status. Such characteristics have a lot of impact on the assessment of member's trustworthiness. Therefore, precise tracking of members and their conections gives more accurate credibility classification results.

Similarly, using topic/event features involves aggregating message-based and user-based features [36], and leads to more accurate results [98]. This includes intuitive relationships, such as the volume of member's messages, the average volume of credible posts for a credible event in contrast to that for a noncredible event, and the number of credible followers; thus, the way a user creates a tweet is more important for identifying its credibility than the user's characteristics [52]. However, it is notable that content features are generally more effective in detecting fake content than user-based features, as shown in [52], in which credibility accuracy using content features is 81%, while that based on user features is 70%.

Regarding topic and opinion classification, [99] examined tweets and categorized them by topic after determining their credibility score. However, this work did not explain its methodology for finding credible content, mentioning utilizing past knowledge. Ikegami *et al.* [68] introduced an interesting method for classifying topics and opinions based on information credibility. They measured the credibility of posts regarding the major quake that hit Japanese islands in 2011 by counting opinions of each tweet. The credibility of a tweet was based how many positive/negative opinions that user received on his/her post: the more positive opinions, the more credible the post, and vice versa. The $\kappa$ statistic between their method and human scoring was greater than 0.6. Table 1 (see Appendix) summarizes the credibility analysis papers that utilize supervised learning techniques.

#### 2) UNSUPERVISED APPROACHES

Table 2 in Appendix lists three studies utilizing unsupervised approaches to determining social media content credibility. The table also shows the type and number of features used. Abbasi and Liu [100] proposed a new algorithm called CredRank to evaluate user behavior in OSNs and rank their credibility. CredRank [100] was designed to group the connected members together and estimate the groups based on the number of members. The clustering step measures the similarity of users' behaviors with respect to social network type. For example, tweets are clustered by calculating users' tweet similarity. They used a CredRank Jaccard coefficient to measure the similarity of users' behavior. In their experiment, Abbasi and Lui used a dataset crawled from the US Senate's website and analyzed the correlation between voters; they found six highly correlated senators' votes. They then sorted them in descending order upon counting the senatorial seats. They claimed that the solution they proposed had a broad range of applications, including in stopping the propagation of false news, preventing large-scale actions, and foiling inaccurate product descriptions; however, they did not show the method's effectiveness with real cases. Conversely, Al-Sharawneh at al. [101], [102] argued that information in Twitter disseminates through influential people known as leaders or pioneers. In order to identify leaders, they propose an approach to utilizing social networks for assessing their credibility through their impressions and roles in events such as crises.

### C. GRAPH-BASED AND SEMI-SUPERVISED ALGORITHMS

One group of methodologies presented in credibility analysis research based on automated classifiers or ranking systems is the group of graph-based methods [95], [103], [104].

Gupta et al. [95] presented a credibility analysis method using event-graph-based optimization. To model the relationships among users, microblogs, and topics, they used a PageRank-like algorithm called EventOptCA to iteratively calculate these relations and compute credibility scores. Other graph-based algorithms use semi-supervised learning to utilize multiple structures in marked as well as unmarked sample. Typically, this group of techniques stretches the marks through a prepared matrix, with each group of marked or unmarked events can be defined as a node, while the connections indicate the relationship between each couple of events. The goals of these approaches are twofold: (i) refined labels should be close to the annotated labels and (ii) the refined labels should be smooth over the completely defined graph.

The described techniques can be improved if these concerns were addressed:

1. Advanced neuro-linguistic techniques should be utilized for better message assessment.
2. Proof that the message is based on false information can usually be found within the message. Still, semantic analysis of the content could fail to detect it.
3. Show business reports leave an impression of low credibility since they are frequently written in a casual way. Credibility for such events may need to be studied separately.

Several other researchers have implemented machine learning techniques as systems to assess credibility for both real-time and offline social network content. Al-Eidan *et al.* [56] developed an automatic credibility measurement system for Arabic web content. The system classifies Arabic weblogs based on selected features into three classes: believable, not believable, and doubtful. This work illustrates that credibility does not depend on a single object, user, or report; it is a feature that becomes clear only after checking several factors. These authors used a weighted feature approach, checking the existence of two features: connections to confirmed and well established media pages and the existence of verification for that member profile. Moreover, they considered URLs, user mentions, retweets, and hashtags at the tweet level, and location, biography, web site information, the size of the following, and the verification of the profile in question.

Ravikumar *et al.* [105] proposed a three-layer graph model for the microblog ecosystem. The model utilizes the implicit relationships among tweets to achieve good computational performance and high precision in assessing the trustworthiness of tweets. Truthy [61] is a unified framework designed to enable authors to analyze user behavior and idea diffusion in a broad variety of data feeds. Truthy tracked approximately 305 million tweets and detected general interest memes to reduce this number to 600 thousand messages that were kept for a later review. Using Truthy, scores for collections of tweets can be calculated, reflecting the probability that those tweets are misleading.

TweetCred [58] is a Google Chrome plugin used to assign a credibility score between 1 (not credible) and 7 (fully credible) to tweets in a Twitter account's timeline in real time. This plugin is based on two well known algorithms. More than 1000 members tried the plugin during the 90 day test period. One of its features is that it can be effective without access to member's track record or complete data about the instance. TweetCred's response time, effectiveness, and usability were evaluated on about 5.4 million tweets. Although the reaction was slowed down by the need to communicate with Twitter's API to obtain tweet details, reaction interval was shorter than six seconds in more than 80% of cases and less than 10 seconds for 99% of users. Table 3 in Appendix lists graph-based credibility analysis studies.

### D. WEIGHTED AND IR ALGORITHMS

In this group of techniques, researchers calculate the ratings for every individual characteristic. Al-Khalifa *et al.* [29] presented an evidence-based technique for assessing the credibility of Twitter content. Technique is limited to media reports and starts from the premise that all media reports can be trusted, which doesn't correspond to reality. Similarly, Xia *et al.* [106] try to measure credibility using credible independent entities. Unlike Al-Khalifa and Al-Eidan [29], the authors don't automatically accept that independent entities, such as Wikipedia, must be trustworthy in every case. Consequently, they take advantage of checking only credible external sources. However, both methods depend on measuring the similarity between the examined tweets and external news sources, which means that they cannot calculate credibility values if news does not exist.

There has been little research using IR [29] and NLP [63] to check the credibility of Arabic language web content. Most investigations have automatically measured credibility for English, German, and Japanese web content. Al-Eidan *et al.* [56], published in 2009, was the first work for Arabic in this domain, followed by another paper by Al-Khalifa and Al-Eidan [29] in 2011 that improved upon their previous work. They used an evidence-based method that allowed them to compare content with trusted news sources using NLP. However, this approach requires profound semantic analysis to understand context, and is applicable only to text content, whereas OSNs also contain media content. Furthermore, their approach defines only a binary score for trustworthiness (good or bad) and it fails to cover some of the elements identified by earlier studies, for example hash tags, republishing, and emojis. Additionally, the work does not explain the complexity of the trustworthiness scoring system nor its effectiveness with introduction of additional characteristics [47].

To determine media credibility using IR to extract evidence from tweets, Middleton [63] established a project called REVEAL (http://revealproject.eu/), which detects false and legitimate messages coming from well regarded or unknown members with a set of regex patterns matching both terms

and POS tags. He ranked Twitter evidence based on the most trusted and credible sources, as would human journalists. This project analyzes individual tweets (post level) which, as discussed in Section 4.1, is insufficient for determining a tweet's credibility. In addition, it is based on well-known journalistic verification principles that reduce its adaptability. Finally, this method is still not fully automated, works on small datasets, and requires some human intervention. Table 4 (in Appendix) summarizes the research on Twitter credibility that has adopted weighted and IR techniques.

### E. HUMAN-BASED CREDIBILITY ASSESSMENTS

In this section, we review all works that manually assess content, user, and topic credibility. We categorize these works by their evaluation approaches, which rely on the final judgment of a human subject. These categories are divided into voting, cognitive, and manual verification approaches. Figure 7 shows the representation of these methods in the human-based credibility analysis literature. In the next chapter, we will discuss every technique in depth.

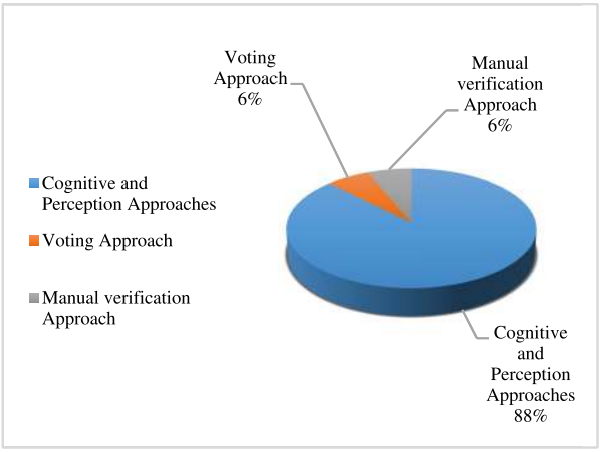


**FIGURE 7.** Studies regarding human-based credibility analysis methods.

#### 1) THE VOTING METHOD

Voting is an efficient method for establishing credibility at the user level, where we can rank users/accounts by their credibility scores for a given piece of information [81], [90], [100], [107]. Neither Twitter nor other OSNs can verify all users. Just a tiny fraction of all members might be validated by these services; for example, only 6% of registered Twitter users are verified. In this sense, and considering that many people prefer to participate anonymously to facilitate their freedom of expression, it is predictable that most users on Twitter or any other OSN are unverified [100]. The real problem is that even verified users can disseminate rumors and misinformation. Gupta and Kumaraguru [82] found that many verified accounts have propagated fake content.

Information credibility models proposed by Canini *et al.* [81], [90] deal with how follower relationships as votes of confidence by the network can be refined further by distinguishing social ties from those based on reputation. Canini *et al.* developed an interesting algorithm to determine the trustworthiness of user-generated messages on Twitter for any given topic. Their algorithm works as follows:

1. Use the Twitter API to search for tweets about a topic.
2. Identify users related to the queried topic and consider those users as voters.
3. Rank the user list by finding the qualifying members. Those members can often pop up in the query results.
4. Rerank the user list per the analysis of each user's textual content and then use topic modeling to find the highest-scoring users.
5. Evaluate the algorithm using the crowdsourced Amazon Mechanical Turk (MTurk) to assign scores to most credible members for five search topics.
6. Finally, compare results provided by both the proposed algorithm and the WeFollow website.

Canini *et al.* perceived that their algorithm has the potential to support users in identifying interesting users to follow on Twitter. The idea of combining topic-based textual analyses with network-based information might produce powerful tools for valuing Twitter content. However, one important issue in this work is that algorithms require knowledge of the follower graph. In addition, people who follow/friend users might vote for them regardless of the credibility of their content because they know them. In other words, we cannot guarantee the credibility of the voters. Therefore, members have to be able to estimate trustworthiness of messages arriving from unknown members.

#### 2) COGNITIVE AND PERCEPTION APPROACHES

Cognitive ability can be defined ''psychological function of knowledge that consists of elements like vigilance, sensory system, learning, and thought processes'' [109], correlating exactly which concerns are being regarded, including beliefs, reasons, intuitions, and expertise [110]. Beyond social network analysis, cognitive psychology helps to identify the mental ability tasked with data dissemination [111]. This method addresses four main questions: message consistency, message coherence, source credibility, and general message acceptability. Many researchers [84], [109]–[118] using this approach employ human perceptions and knowledge to evaluate fake content and misinformation in tweets.

In this section, studies from two approaches are investigated. The first applies statistical analysis based on questionnaires, interviews, and opinions of members (e.g., general population, acquaintances, authority figures, and news reporters). The second applies statistical analysis based on a crowdsourced group such as MTurk. The following sections discuss these aspects in detail.

*User Purview:* In crisis events, studies use this approach coupled with experimental and statistical methodologies to

analyze Twitter. They try to justify the importance of Twitter as a tool for communication during crises. Usually, any study in this direction involves different types of news events, such as crises or political upheaval. Researchers have tried to analyze how news and rumors are propagated during an event of interest and to establish the dynamics of the Twitter community in such a scenario [53]. In addition, they have investigated whether a social network can discriminate between rumors and true information and compared the effect of authorities' accounts on the flow of information.

Most of the research in this approach centers interest on surveying the behavior of Twitter users [53], [119], [120]. Mendoza et al. [53] chose the case study of an earthquake in Chile in their survey moving to establish the reliability of Twitter for disseminating information in emergencies. This was done by comparing how confirmed news and false rumors flow through the network. Another related study [119] aimed to glean insight on Twitter use during a forest fire in Marseille, France. It established that Twitter users vary in behavior: there are those whose aim is to aggregate information, such as normal citizens, and users with links to media outlets. Kireyev *et al.* [120] conducted a study of Twitter during the two quakes that happened in 2009. This study emphasized modeling tweets to find relevant information. Qiu *et al.* [121] established that tweets prefixed with "@" (indicating direct replies to other users) were less frequent during such events. Also established here is that the number of tweets including a URL tends to decrease in crises.

O'Donovan *et al.* performed a statistical evaluation of characteristics and their distribution [122] for four classes described in Table A.6 (in Appendix). They revealed that URLs, mentions, retweets, and tweet length are more relevant and give good indications of credibility. Throughout our survey, we explore more than 130 studies of credibility assessment in crisis and non-crisis events and found that those features are most important relative to credibility over OSNs (see Section 4 and Fig. 4).

Morris *et al.* worked on a model [123] that includes characteristics that are highly relevant for measuring confidence, positing that reviewers could be affected by the framing of a certain conversation. They adopted an experimental scheme to find out which characteristics shape member's beliefs about credibility. They found that there are several features that enhance credibility, including the influence of the user, the user's topical expertise judged from his/her biography, and the user's reputation (e.g., whether a user has a Twitter verification seal). In the second experiment, they found that the message topic influences the perception of tweet credibility, with science topics receiving a higher rating, followed by politics and entertainment. User images had no significant impact on tweet credibility, but user names had a significant effect on tweet credibility. A follow-up experiment established that use of the default Twitter icon lowered the perception of credibility. Other images did not show any differences in perceived credibility. Kawabe *et al.* [123] present an advanced way of assessing user credibility perceptions in Twitter. This serves as a step up from previous studies done by other researchers. The authors did, however, neglect demographic aspects in their research; this is subject to future work that will aim to fine tune the research.

Schmierbach and Oeldorf-Hirsch [124] conducted a similar survey by displaying digital articles from the major media outlet alongside Twitter messages from the same source. Their study found that many participants were more doubtful about reports originating on social networks versus the stories coming from the site. In other words, the users evaluated news items as significantly less credible when reading them as tweets compared with reading them on the newspaper's website. Thus, they eliminated the topic of posts as a variable that might lower overall credibility ratings.

Yang *et al.* [125] organized an internet survey that compared member impressions about US-based Twitter and its Chinese counterpart, the microblogging site Weibo. They compared the relevance of multiple characteristics for both SN's, for example sex, username, personal photo, country of origin, and areas of interest. The researchers concluded that the relevance for each factor depends on the network. On top of that conclusion, Chinese members as a group were found to be more trustful towards Weibo than US-based members are prone to trust Twitter. Other research used cognitive models, such as the ACT-R model with statistical analysis, to measure the credibility of microblogging [84].

The outcome of this approach varies between users because the cognitive approach utilizes a user's cognitive abilities for assessing Twitter content. This provides a different instrument for determining the trustworthiness of Twitter messages, although the findings must be taken relatively if we consider who is evaluating the content.

Most of these studies have worked to illustrate the dynamics of Twitter communities in cases of mass convergence and emergencies, but they have done little to assess Twitter as a tool for accruing truthful information in such events. In the future, microblogging sites are expected to provide a way for users to assess trust in information. This can be achieved through state-of-the-art classifiers that show whether a tweet is asking for more information. For example, users could be alerted if others are questioning the information posted.

*Crowdsourcing:* With respect to crowdsourcing, Kumar [126] analyzed user perceptions on a decentralized online network called CrowdFlower to estimate the trustworthiness of messages on media topics. The study showed that users considered eight significant features in judging the credibility of information. Their results showed prominent features; for example, URL, identity and user confidence were relevant for estimating a message to be truthful. To guarantee evaluator reliability, the authors examined evaluators by presenting two golden questions; if the evaluator answered them correctly, then his/her judgment

was accepted, otherwise he/she was eliminated from consideration. Saez-Trumper [92] proposed an application with three techniques; one is a crowdsourcing approach used to detect malicious users on Twitter. He developed a panel with all the information about a photo and its Twitter account. Then users were asked to tag the account as fake or credible. However, he did not determine how to guarantee users' evaluations. In contrast, some studies used crowdsourcing platforms in the annotation phase; for example, in several studies [36], [58], [122], authors used MTurk and/or CrowdFlower to annotate users that live in the US.

Several studies have been performed regarding the use of crowdsourcing to examine user perceptions of information credibility on well-known social media services [126]–[128]. Kumar [126] used crowdsourcing platforms to examine how users behave when they judge the credibility of political news over Twitter. They found that there are eight features driving users' judgments; further, they showed that the topic type (such as political topics) affects the consistency of users' judgments. Similar studies [127], [128] have depended on crowdsourcing to study human perceptions. However, they tried to compare user credibility judgments of political issues over OSNs (e.g., Facebook, blogs, and Twitter) to traditional media (e.g., Fox News and CNN). In addition, they studied the impact of credibility in motivating people to use social networks to collect data about politics. They determined that established outlets such as CNN or FOX are seen as more reliable than recently devised social platforms.

Several researchers [36], [46], [49], [58], [91], [94], [122], [129] used crowdsourcing to annotate and label training data. Sikdar at al. [49], [129] emphasize that building the ground truth is the most important stage in analyzing the credibility of any content that needs stable ground truth measures. They used MTurk to build this stable ground truth; although we agree with establishing a stable ground truth, using crowdsourcing to do this must be reviewed for several reasons. Labeling data and building a ground truth, as stated in [49], must be justified and robust. Depending on the wisdom of the crowd isn't ideal, since a majority of participants may be devoid of related knowledge, particularly on certain topics (e.g., crisis management or political events). Secondly, crowdsourcing is inherently inconsistent [81], [90]. Canini *et al.* [81], [90] built their ground truth using MTurk. They tried to confirm that each recruited person had an active Twitter account by demanding to know their handle, age, and availability frequency; however, even with experts, the consistency of judgments should be controlled [4].

#### 3) MANUAL VERIFICATION APPROACH

A common approach for social network corporations to encounter malicious user activities is through the manual evaluation and verification [130]–[132]. For example, on Twitter, some users have been manually verified by Twitter following administrative processes cross-checking their profile with their identity. Twitter states that ''Validation is the procedure we use to confirm real identities for people and companies present on the network'' [130]. However, only a very small fraction (0.006%) of users is verified by Twitter. Another approach taken by OSNs is to publish detailed lists of malicious members [132] in order to warn the legitimate members about possible danger. All profiles found to take part in forbidden actions can be suspended or deleted at any time. This approach has several drawbacks, such as inefficiency and the inability to distinguish malicious activities in a timely fashion.

In general, a majority of techniques that involve the human factors share two major weaknesses: variability and inaccuracy. Even with golden questions, the problem becomes worse if a large part of participants gives the wrong answer. We consulted more than 1,000 studies on this topic, and learned that automated techniques were able to deliver more precise results than those that relied on live evaluators. Like Table 1, Table 2 in Appendix lists human-based credibility assessment studies classified by their methodology, type, and the number of features used.

### F. HYBRID APPROACHES

To utilize the advantages of both the automation- and human-based approaches, several researchers [4], [18], [47], [48], [67], [133]–[139] have analyzed the credibility of social media content using hybrid approaches. The specific combination of techniques varies from one study to another, as does the level of combination. For example, some authors have used supervised approaches, such as the naïve Bayes approach (which is a machine learning approach at the topic level in our classification), with a human perception survey method [67]. Another used a clustering approach (also a topic-level machine learning approach in the proposed taxonomy) with a method from the same class that included weighted and IR algorithms [133]. Table 9 in Appendix shows the recent studies using hybrid approaches.

## VI. DISCUSSION

In this survey, we focus on research that measures credibility in OSNs in terms of published content and users, especially on Twitter. Past research has shown how information can be accrued from Twitter, but most has failed to highlight how this information can be filtered to separate real information from false. This topic has attracted interest from researchers worldwide. We have discussed studies that are closely related to the topic in detail and found that most related work on Twitter content credibility assessment was performed at four levels of feature extraction: post, topic, user, and hybrid levels [36], [49], [50]. With each level, works in the literature vary in their approaches, techniques, and methodologies and are based on different models, features, datasets, and amounts of human involvement [48]. Some researchers have also followed other types of tweet features, such as dissemination

feature modes, coupled by fractions of retweets and the overall number of tweets. We consider these to be hybrid-level features that assess credibility relating to single messages as well as broader themes.

Our study examines three main approaches in-depth—automation-based [28], [50], [59], [87], [88], human-based [81], [90], [100], [107], and hybrid [4], [18], [47], [48], [67], [133]–[139] approaches—to assessing the credibility of Twitter content. Each methodology is divided further, as stated in Section 5. Some researchers [36], [46], [53], [58], [87], [94] emphasized that automatic methods of assessing credibility are better than human-based methods owing to the huge amount of data and the nature of Twitter streaming. Dissemination using 140-character-long messages is very useful in emergency situations (e.g., terrorist attacks, hurricanes, earthquakes, or riots) whereby information can be accrued from first-person sources, but may cause panic. Consequently, misinformation may spread widely and can cause a lot of damage before human operators detect it. However, others [84], [109]–[118] have stressed that manual methods of credibility assessment are more accurate than automation-based methods. They used statistical analyses to interpret results and measure user behavior regarding particular news. However, techniques using live evaluators are typically highly variable and change from case to case, especially those using crowdsourcing. Therefore, some researchers [53], [119], [120] accentuate that users selected for assessment tasks should be users with a specific purview (e.g., citizens, witnesses, experts, or media reporters) whereas others [81], [90] only stress that users should have technical expertise with Twitter.

By merging two methodologies, some researchers [4], [18], [47], [48], [67], [133]–[139] have tried to leverage the precision of live operators to determine basic premises and mark the datasets. However, they have not overcome the consistency and reliability problem. Some examined assessors to guarantee their reliability by posing two golden questions; if the assessor answered those questions correctly, then his/her judgment was accepted. Therefore, the problem is exacerbated if most evaluators' answers are wrong [92]. Canini *et al.* [81], [90] partially solved the reliability problem by using domain experts associated with a given topic. They concentrated on evaluating the source of information by analyzing the use of automated ranking strategies to measure them. The authors found that messages and connection matrixes are good indicators for calculating the trust factor for Twitter members. However, the consistency of domain experts needs to be measured and justified [1].

Some existing related studies reveal that the first common method used for credibility analysis of members of a microblogging online platform is manual evaluation/verification [130]–[132]. However, this approach has several drawbacks. Furthermore, there are many systems developed for credibility detection that are limited in their applications; for instance, TweetCred [56] is implemented as a Chrome plugin. However, Gupta *et al.* [56] noted that their tool is experimental and will improve over time. In addition, its feedback is not automatic, as it is received from users. Castillo *et al.* [36], [94] established feature analysis for credibility tasks and noted that:

1. New Twitter users and the most active users spread credible information;
2. Tweets with positive sentiments propagate credible information;
3. Tweets with question marks and emoticons tend to spread non-credible information.

From these findings, the authors showed that it is possible to separate newsworthy tweets from those that are conversational. Newsworthy topics are shown to have deep propagation trees coupled with included URLs. In credibility assessment at the feature level, the metrics obtained show that the top element subset and propagation set are critical, as the linear interdependence of feature pairs is weak. Consequently, feature importance varies between events and is not the same for measuring credibility [1]. For example, Canini et al. [81], [90] observed that network formation and tweet content are suitable indicators that help to measure credibility. The relevance of each characteristic will greatly contribute to the efficiency of the method [1]. Hence, they also proposed a method to measure the consistency of expert evaluations and control the evaluation process. In summary, the accuracy of automated and semi-automated techniques still has much room for improvement.

In a work by Al-Khalifa and associates [29], [56] a new technique is discussed grounded on overlapping of tweets with already confirmed sources. The technique faces several significant issues. To begin with, it is dependent on the success of the initial semantic analysis. Also, it can return inaccurate results if the analyzed topic absent from the mainstream media reports. Next, it is poorly suited for multimedia content, for example pictures of short clips. Fourth, it does not consider the sentiments of the compared pairs; using preprocess stemming and pruning statements may lead to the opposite results. Only one study [95] attempted to calculate their algorithm's computation time, which was O(IT2/E), where T denotes volume of unique tweets, I is the amount of iterations, and E is the total volume of events.

We can see that some existing works consider calculating and evaluating credibility based on a given topic [41], [46], [53], [54], [60], [70], [88], [93], [95], [121], [140]–[142]. They overlook calculating credibility for a user based on his/her reliability of information/content as well as inter-entity relationships [93]. Therefore, we still haven't seen a decisive work that could establish a systematic credibility metric in OSNs to detect and evaluate misinformation that integrates reliability and reputation, though most of the research provides a good foundation for future efforts. In principle, future work should include extending experiments to larger datasets and partial datasets;

some studies used small datasets [29], [57], [59], [72], [81], [133], [143]. Other open problems include assessing the credibility of target pages to which URLs redirect. The assessment of other factors that influence credibility on Twitter, such as avatar images, is also open for future research [36]. One study mentioned that microblogging sites should provide a way of allowing users to assess trust in information.

Overall, the shortcomings of both human- and computer-based evaluation were highlighted in this review. For instance, human evaluation is impractical in that users tend to avoid assessing content credibility. Human-based approaches are affected by the evaluators' experiences and motivations. Computer-based techniques, however, are handicapped by the dynamic nature of web technology, for example, the lack of extensive Web ontologies. The arguments given in later sections form a basis for the need for a hybrid Web credibility assessment model that leverages both computer- and human-based models to optimize its results.

## VII. CREDIBILITY PROJECTS AND SYSTEMS

Pheme is a European-Union-funded, 36-month research project that seeks to ascertain veracity in OSNs. To cite the Pheme website, during the 2011 England riots, there were false claims that the iconic London Eye was on fire. Yet, the fact that some entity on the internet used social media to make this claim and that it was believed shows the remarkable power of OSNs. One can use several examples to show that social media is a platform that allows for the proliferation of unverified information. Moreover, there is no system or body that analyses messages created by members from a more systemic perspective. Therefore, data is created, distributed, and accepted by end users.

The remarkable uptake of unsubstantiated information is the reason for Pheme's existence. The name is an homage to the Greek goddess of fame and rumors. In the context of the project, Phemes "are meme messages that contain bits of truth." Fundamentally, the project is an attempt to use data analytics with linguistic and visual methods to validate data so that they can be used for healthcare or media reporting platforms. Accordingly, the project has seven partners, including the European universities King's College London and MODUL University Vienna, and companies like Atos and Swissinfo.

The methods used to realize the project's objective include NLP and data retrieval, online research, social media tracking, and data presentation. For example, PubMed, the world's largest online database for original medical publications, is used as a reference to verify medical data, while SwiftRiver, online tool for checking the veracity of reports, is used for digital journalism verification.

Reveal is a European Union project that seeks to create, refine, and use services and tools to analyze and thus verify social media. Moreover, the project aims to analyze at a higher level. For example, the project wants to give end users the chance to engage with social media by looking

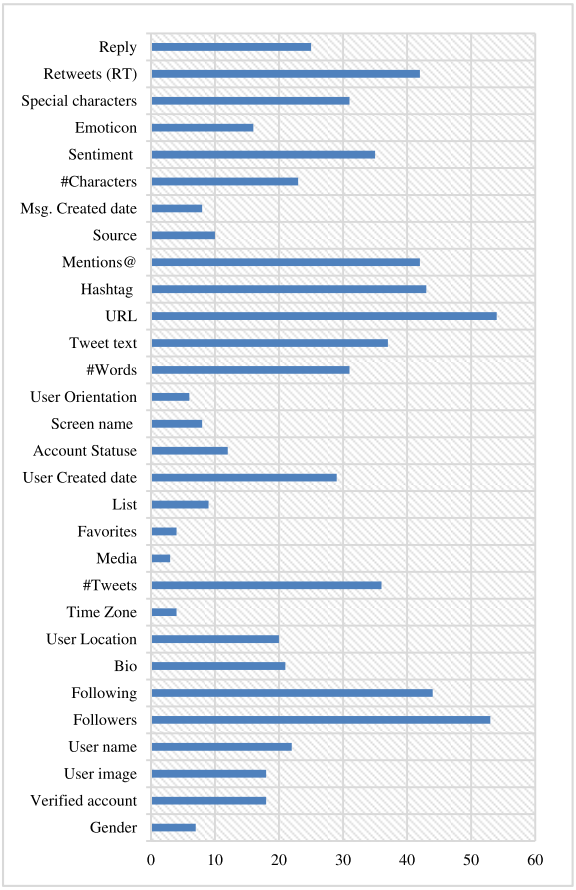


**FIGURE 8.** The most important features used in the literature.

at reputations, influence, and credibility. At the crux of the project is a desire to provide verification from journalistic and enterprise points of view. Therefore, Reveal is, in essence, an attempt to analyze the transition from print media to social and digital media. To quote the Reveal website, "No longer can a selected few (i.e., media organizations) act as gatekeepers. . . Individuals now have the opportunity to access information directly from primary sources [such as] Social Media."

The transition from traditional, verified, and thus unchallenged media to unconventional, unverified, peer social media has created the fundamental problem of excess, useless, and misleading information, which is termed noise. The proliferation of noise means that analysts must sift through vast amounts of data. Reveal aims to go beyond this data shifting by creating and refining methods that allow end users to quickly discover whether data are reliable. These are hidden dimensions or modalities with which Reveal plans to engage. Examples of data cases that have modalities include the Arab Spring in Egypt and the state of the Great Barrier Reef, which involve conflicting perspectives. Reveal partners include academic, research, commercial, and public enterprises. For example, Alcatel-Lucent, a French global telecommunications company, is interested in content generator profiles and modalities. Alternatively, German radio

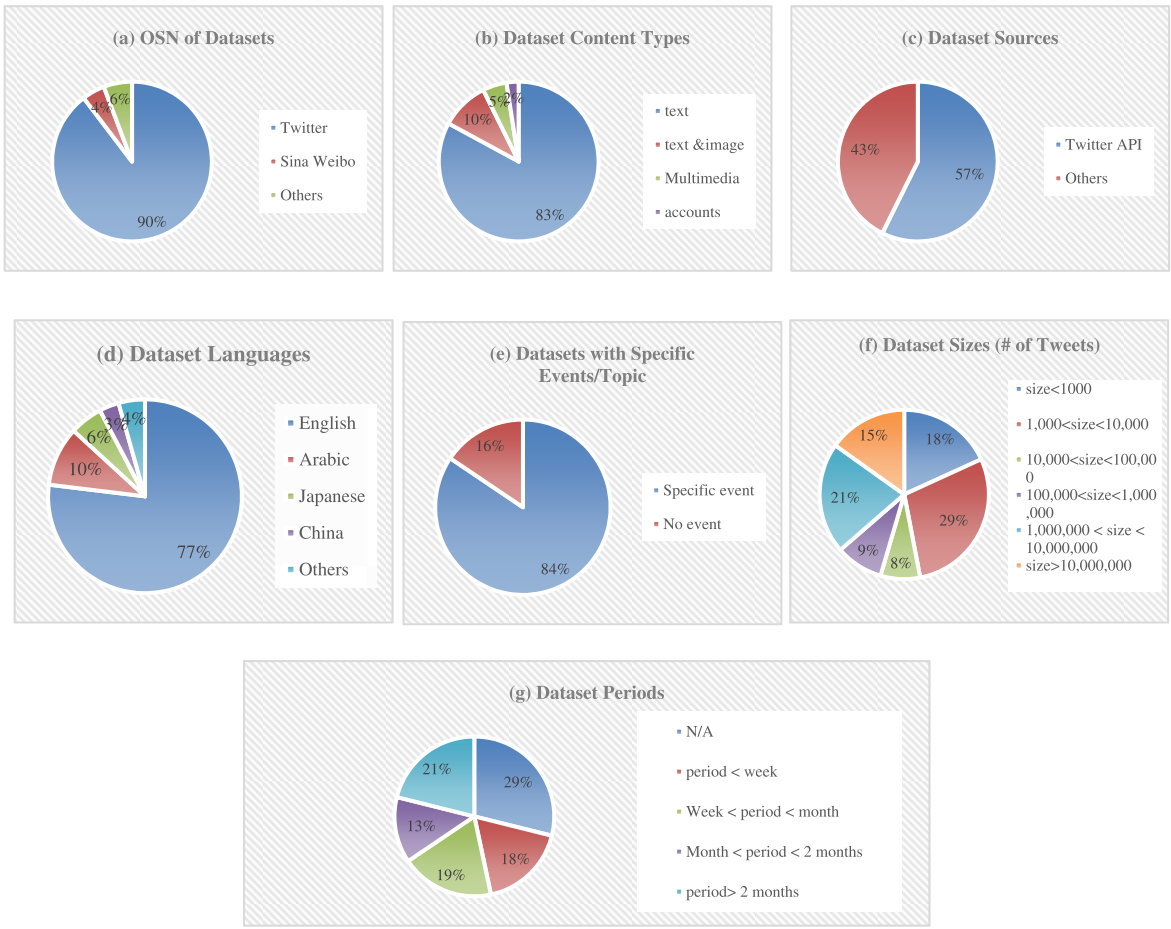


**FIGURE 9.** Statistical analysis for the datasets used in credibility analysis literature.

service for international news is interested in the entire data generation process from creation to dissemination. Thus, Reveal is a fundamentally multidisciplinary and international project.

Truthy is an Indiana University research project that seeks to understand how information spreads through social media networks. Indiana University uses public data from social media platforms, which is widespread and abundant, to analyze and create models of how information spreads from creation to acceptance. Per the Truthy website, the project analyzes all data types ''from political tendencies to financial statistics, from reports to emerging cultures, and from hot online themes to quantifiable research, with amazing level of depth.'' For example, the analysis entails exploring all factors that affect data propagation. For example, ''majority opinion, member's status, alertness, connection matrix, as well as other quantities,'' all impact how public knowledge is formed, spread, and accepted.

Currently, successful Truthy cases include the ''relationship between Instagram and the success of models during fashion week.'' Fundamentally, the sudden appearance of a model and related trend can affect tangible events and even consumer habits. Additionally, Truthy successfully showed that social media popularity in tweets or reposts could affect electoral results. Thus, political candidates must be cognizant of the power and reach of social media. Finally, the project identified how sybils, defined as ''automated software capable of producing messages and contacting live users on social platforms, imitating and in some cases directing user's activities,'' can affect elections, financial markets, and even personal relationships.

The Truthy project is financed by the US Government, the Lilly Foundation, one of the world's largest philanthropic foundations with interests in religion and pharmaceuticals; and the James S. McDonnell Foundation, the bequest of an aviator whose company is now part of Boeing. Thus, the project has attracted the interest of major governmental and international partners, as data propagation can have profound effects on everything from finance to personal relationships. Table 10 in Appendix summarizes the most important projects and systems used to estimate the trustworthiness of messages on social platforms.

## VIII. BENCHMARK DATASETS AND PERFORMANCE MEASUREMENTS

### A. DATASETS

In this review, we mainly focused on Twitter, moving to other social networks if needed in the domain of credibility

**TABLE 1.** Research papers that used supervised learning technique to analyze credibility.

| Authors | Event-level | Post-level | User-level | Models/Algorithm/Approach | # Features |
|---|---|---|---|---|---|
| Sarna, G. and Bhatia [184] | - | + | + | Naïve Bayes, Bayes, SVM, KNN | 6 |
| Gupta et al. [100] | + | + | - | Naïve Bayes decision tree | N/A |
| Gupta et al. [51] | + | + | - | Naïve Bayes decision tree | 25 |
| Saikaew and Noyunsan [101] | - | + | - | SVM | 8 |
| Castillo et al. [93] | + | + | - | Naïve Bayes, Bayes network, logistic regression, and random forest | 68 |
| Wang [102] | - | + | - | AND, OR, majority voting, Bayesian model | N/A |
| Sun et al. [60] | + | + | + | Naïve Bayes, Bayesian network, neural network, and decision tree | 7 |
| Metaxas and Mustafaraj [103] | + | - | - | Machine learning and data mining | N/A |
| Gupta et al. [104] | - | + | - | Supervised machine learning and IR | N/A |
| Sharf and Saeed [105] | - | + | - | J48 decision tree | 8 |
| Boididou et al. [52] | - | + | + | Naïve Bayes, J48 decision tree, Kstar, and random forest | 25 |
| A and C [109] | + | + | - | A modular approach and LDA | N/A |
| Castillo et al. [36] | + | + | + | J48 decision tree | 67 |
| Xia et al. [106] | + | + | + | Bayesian network and sequential k-means | 26 |
| Finn et al. [107] | - | + | - | Naïve Bayes and decision tree | 2 |
| Liu et al. [108] | - | + | - | LDA, C4.5 decision tree, SVM, and hierarchical clustering | 15 |
| Wu et al. [119] | - | + | - | Logistic regression | N/A |
| Ginsc [71] | - | - | + | SVM, random forest classifier | N/A |
| Thandar and Usanavasin [71] | + | - | + | SVM | N/A |
| Al-Eidan et al. [56] | - | + | - | SVM | 9 |
| Yang et al. [85] | + | + | + | SVM | 19 |
| Yang et al. [35] | + | + | + | J48 and C4.5 decision tree | 53 |
| Thandar, M. and Usanavasin, S [185] | - | + | + | SVM | 19 |

**TABLE 2.** Research papers that used unsupervised learning to analyze credibility.

| Authors | Event-level | Post-level | User-level | Models/Algorithm/ Approach | # Features |
|---|---|---|---|---|---|
| Abbasi and Liu [110] | - | - | + | Credrank and k-means | N/A |
| Al-Sharawneh et al. [111] | - | - | + | Credibility-weighted importance, degree of centrality, quality analysis measurements | 7 |
| Al-Sharawneh and Williams [112] | - | - | + | Collaborative filtering | NA |

**TABLE 3.** Research papers that used graph-based methods to analyze credibility.

| Authors | Event-level | Post-level | User-level | Models/Algorithm/Approach | # Features |
|---|---|---|---|---|---|
| Ratkiewicz et al. [61] | + | + | - | Klatsch data model and Google-based profile-of-mood-states sentiment analysis | N/A |
| Ulicny and Kokar [116] | - | + | - | RDF graph and Tunkrank | N/A |
| McKelvey and Menczer [117] | + | + | - | Trustworthiness score | N/A |
| Nguyen et al. [118] | - | + | - | Ranking- and optimization-based algorithms, Monte Carlo | N/A |
| Ravikuma et al. [115] | + | + | - | TF-IDF, top-k precision, and NDCG | N/A |
| Gupta et al. [98] | + | + | + | Decision trees (J48), SVM, naïve Bayes, k-nearest neighbors | 12 |
| Pasternack and Roth [119] | - | + | - | A PageRank-like credibility propagation approach and latent credibility analysis | N/A |
| Nagy and Stamberger [59] | - | + | - | PageRank and naïve Bayes classifier | 68+ |
| Gupta et al. [58] | + | + | + | SVM-Rank | 45 |
| Zou et al. [88] | + | + | + | Online prediction, decision tree, and SVM | 12 |
| Schaffe et al. [120] | + | + | - | Information filtering, expectation maximization, and context-sensitive models | N/A |
| Han et al. [28] | - | + | + | k-means and SVM | N/A |
| Qiu et al. [121] | + | + | - | Vector space model, linear regression, and voting | N/A |
| Namihira et al. [122] | + | + | - | LDA | N/A |
| Ikegami et al. [67] | + | + | - | LDA | N/A |
| Kawabe et al. [123] | + | + | - | LDA | N/A |

analysis. Figure 9a shows the online social network services used to analyze content credibility. Most studies have concentrated on text content, as shown in Figure 9b, whereas small fractions addressed multimedia and users. Many studies, projects, and systems of credibility analysis use Twitter APIs that give access to Twitter's REST API and stream API, and some researchers use commercial software to collect the required tweet data. Figure 9c gives statistics on the percentage of studies that use the Twitter API. These APIs enable programmers to harvest information from the networks and develop apps, overcoming the inherent obstacles. Two major classes of APIs exist, an ask-reply API that offers query

**TABLE 4.** Research papers that used weighted and IR algorithms to analyze credibility.

| Authors | Event-level | Post-level | User-level | Models/Algorithm/Approach | # Features |
|---|---|---|---|---|---|
| Al- Khalifa and Al- Eidan [29] | - | + | + | Similarity between Twitter posts and authentic and verified news sources | 6 |
| Al-Eidan et al. [57] | - | + | - | Similarity between Twitter posts and authentic and verified news sources | 5 |
| Middleton [63] | - | + | + | IR | N/A |
| O'Donovan et al. [125] | - | - | + | TF-IDF and LDA | N/A |
| Widyantoro and Wibisono [89] | - | + | - | TF-IDF | 6 |
| Gupta and Kumaraguru [126] | + | + | - | TF-IDF | N/A |
| Par et al. [66] | + | + | - | Linguistic inquiry and word count | 4 |

**TABLE 5.** Research papers that used voting approaches to analyze credibility.

| Authors | Event-level | Post-level | User-level | Models/Algorithm/Approach | # Features |
|---|---|---|---|---|---|
| Canini et al. [90] | - | + | - | A new ranking method, TF-IDF, LDA, and ANOVA | N/A |
| Canini et al. [81] | - | + | - | NumVotes, DivF, TF-IDF and LDA | N/A |
| Sirivianos et al. [127] | - | - | + | Social tagging and graph analysis | N/A |

**TABLE 6.** Statistical analysis of features distributions by O'Donovan et al. [142].

| Class | Description | Number or Context |
|---|---|---|
| Diverse topics | Diverse topics on Twitter, e.g., #Romney #Facebook | 8 topics (see Table II) |
| Credibility | Manually provided tweet assessments | Credible or noncredible |
| Chain length | Mined retweet chains classified by length | Long or short |
| Dyadic pairs | Mined interpersonal interaction and classified | Dyadic or not dyadic |

interfaces from the site and a streaming API that transfers processed data to the interested user as soon as it becomes available. [3]. The prepared data is divided into three groups: open, semi-open, or closed. Despite the preeminence of social media and its impact on people's lives, only a few languages have been investigated in terms of credibility analysis, as depicted in Figure 9d.

In data collection, some studies focused on time-sensitive data with a Twitter monitor being used for automatic event detection [36], [46], [53], [58]. Figure 9e shows the amount of research focused on global catastrophes (such as the Boston bombings or Hurricane Sandy) that prompted Twitter members to spontaneously share a huge volume of information of varying credibility. Among credibility studies, the only dataset we found is CREDBANK [83], which is available in different forms as of May 28, 2015. A quick overview of these datasets can be found at the CREDBANK website. Depending on the release version, data were collected from autumn of '14 to the second month of the next year, thus creating a compilation of incoming messages sent in this interval, with a separation of messages pertaining to real-world events, which were rated for credibility. These datasets contain four files. First, the database of messages consists of nearly 170 million items distributed in several temporal clusters. Second, the database of topics (event/non-event) holds over 62 thousand entries. Every entry is accompanied by 3 words, chosen as the most relevant three keywords indicated by the latent Dirichlet allocation (LDA) technique based on incoming messages. Next, the credibility trustworthiness score as well as its detailed explanation. Lastly, the searched tweet database is comprised of contains the astronomical 80 million messages related to the events from a previous database. Messages are harvested with the help of the official search API. Time span is very important when collecting such data. Researchers have tried to find a way for collecting time-sensitive data from OSN platforms so they can benefit emergency workers, local activists, or other relevant parties. The data collection period varies in the literature, as explained in Figure 9e.

**TABLE 7. Research papers that used a user-purview cognitive approach to analyze credibility.**

| Authors | Event-level | Post-level | User-level | Models/Algorithm/Approach | # Features |
|---|---|---|---|---|---|
| Suzuki [146] | - | + | + | Reputation-based credibility degree assessment | N/A |
| AlRubaian M. et al., 172] | + | + | + | User sentiment/ popularity measurement | 18 |
| Westerman et al. [132] | - | + | + | ANOVA and MANOVA | 2 |
| Liao et al. [84] | - | + | + | LDA and ANOVA | N/A |
| Jaho et al. [133] | - | + | - | Rating-based, statistical methods, and cumulative distribution function | N/A |
| Kwon et al. [147] | + | - | - | Intraclass correlation coefficient | N/A |
| Rieh et al. [148] | - | + | - | Empirical study | N/A |
| Kang et al. [149] | - | + | + | Heatmap visualization and statistical analysis | 12 |
| Armstrong and McAdams [150] | - | + | + | Empirical study | N/A |
| Llamero [150] | + | + | - | Qualitative approach | N/A |
| Metzger and Flanagin [130] | - | + | + | Cognitive heuristics | N/A |
| Gao et al. [152] | - | + | + | Cronbach's α and ANOVA | N/A |
| Edwards et al. [153] | - | + | - | MANOVA and ANOVA | 12 |
| Kang [135] | - | + | + | AMOS 18.0 with maximum likelihood and confirmatory factor analysis | N/A |
| Yang et al. [145] | + | + | + | ANOVA | 5 |
| Ulicny and Baclawsk [154] | - | + | - | Google News ranking | 9 |
| Finn and Zúñiga [155] | - | + | + | Pearson's correlation | N/A |
| Rieh [156] | - | + | + | Diary study and experience sampling method | N/A |
| Briscoe et al. [97] | - | + | - | Friedman's test and ANOVA | 6 |
| Bakker et al. [157] | - | + | - | Survey-embedded experiment | N/A |
| Johnson [158] | - | + | - | Kruskal-Wallis test and Cronbach's α | N/A |
| Chesney and Su [159] | - | + | + | Kolmogorov-Smirnov test and Cronbach's α | N/A |
| Xu et al. [160] | - | - | + | Statistical methods | N/A |
| Thomson et al. [161] | + | + | - | Statistical methods | 3 |
| Pattanaphanchai et al. [162] | - | + | - | Thematic analysis and expert panel | N/A |
| Zubiaga and Ji [163] | + | + | - | Statistical methods (Krippendorff's test) | N/A |
| Morris et al. [143] | + | + | + | ANOVA | 31 |
| Mendoza et al. [53] | + | - | - | Filter-based heuristic approach | N/A |
| DeGroot et al. [158] | - | + | + | Statistical methods | N/A |
| Nurse et al. [164] | - | - | - | Statistical methods | N/A |
| AlMansour and Iliopoulos [165] | + | + | + | Statistical methods and Krippendorff 's α | 34 |
| Pal and Counts [77] | + | + | + | Statistical methods | N/A |
| Go et al. [166] | - | + | - | Statistical methods and structural equation modeling | N/A |

**TABLE 8.** Research papers that used a crowdsourced cognitive approach to analyze credibility.

| Authors | Event-level | Post-level | User-level | Models/Algorithm/Approach | # Features |
|---|---|---|---|---|---|
| Johnson and Kaye [168] | + | + | - | Amazon MTurk and Cronbach's α | N/A |
| Shariff et al. [167] | + | + | + | CrowdFlower and association rules mining | 8 |
| O'Donovan et al. [142] | + | + | + | Amazon MTurk and distribution analysis | 34 |
| Johnson and Kaye [169] | + | - | - | Online survey, Cronbach's α, and hierarchical regression | N/A |
| Sikdar et al. [49] | + | + | + | Amazon MTurk, statistical methods, and various ground truth measures | 45 |
| Sikdar et al. [170] | + | + | - | Amazon MTurk, statistical methods, and various ground truth measures | 45 |
| Ghosh et al. [137] | + | + | + | Crowd-sourcing and language processing | N/A |
| Schaffer et al. [138] | - | + | - | Crowdsourcing, maximum-likelihood estimation, and ANOVA | N/A |
| Aladhadh et al. [171] | + | + | + | CrowdFlower, crowd sourcing, and statistical methods | N/A |
| Johnson, T.J. and Kaye, B.K. [170] | + | + | + | Amazon (MTurk) and statistical methods | N/A |

Of the more than 181 research papers read for this survey, 112 are in the domain of microblog credibility analysis. We tried to explore the most commonly used features regardless of the levels of credibility assessed. It was found that size of the contact network is the parameter with the highest value, and was used in 68 papers in the literature regardless of the social network type, followed by message URLs, which were used in 63 papers. In Twitter credibility analysis, we found that, from 112 papers, 54 papers used URLs and their aggregations as the main features, whereas the number of followers was used by 53 papers. In contrast, time zone, number of favorites, and media were the least popular, with only 4, 4, and 3 papers using them, respectively, like it was demonstrated on Figure 8. Hence, relevance of the chosen characteristics greatly determines how successful the technique can be. The problem here is that, although numerous studies used an enormous amount of data for various credibility analyses, there is no standardization effort towards a common dataset for credibility assessment research.

### B. PERFORMANCE MEASUREMENT

Performance metrics differ according to the methodology used in assessing and analyzing credibility. Research in human-based methodologies used statistical analysis to evaluate their performance. Some researchers evaluated hypotheses by observing the accuracy of a hypothesis over a limited sample of data to perceive how well this estimates its accuracy in additional tests. This is also motivated by finding the best approach for using these data to learn a hypothesis. For example, Sikdar *et al.* [49] used 10-fold cross validation to measure the ground truth, breaking data into 10 ground truth bins of size n/10; they trained on nine bins and tested on one.

Numerous evaluation metrics are usually used to evaluate performance in credibility analysis tasks. Some of the most common measures are precision and recall, which are used specifically for IR [144]. Similarly, accuracy, retrieval time, and F-measure metrics are used to determine the accuracy of post, topic, and user credibility classifiers [34], [36]. F-measure is relied upon to determine equivalence and balance for precision-recall tradeoffs. Other researchers compute prediction accuracy, the $\kappa$ statistic, and the receiver operating characteristic (ROC) area of the validation data used [145]. The $\kappa$ statistic shows how the classifier outperforms a random guess (-1 as the lowest value and 1 as the highest). ROC space denotes the possibility for the evaluator to differentiate the real success ratio (effectiveness) from the false recognition ratio (negative differentiation). The predictive power is estimated through regression-based methods, which is typically chosen because it can be accurate in predicting the outcomes.

Other researchers use error-based measures to examine the performance of their algorithms [94], [94]. Gupta and Kumaraguru [54] used the normalized discounted cumulative gain (NDCG). NDCG engages with machine learning approaches to measure the cumulative gain of SVM-ranking outputs. In [54], the authors measured

**TABLE 9.** Research papers that used hybrid approaches to analyze credibility.

| Authors | Event-level | Post-level | User-level | Models/Algorithm/Approach | # Features |
|---|---|---|---|---|---|
| Kang et al. [67] | + | + | - | Bayesian, J48 tree, perception, and social graph | 19 |
| Ito et al. [175] | + | + | + | LDA, clustering, and random forest classifier | 27 |
| AlMansour et al. [48] | + | + | + | User surveys and automatic classification techniques | N/A |
| AlMansour et al. [18] | - | + | - | Decision tree, naïve Bayes classifier using weighting-based features, and use perception | N/A |
| Bhattacharya et al. [176] | - | + | - | Use perception, J48 decision tree, naïve Bayes classifier, and SVM | 10 |
| Qazvinian et al. [177] | - | + | + | Crowdsourcing and Bayes classifiers | 9 |
| AlMansour [49] | + | - | - | Decision tree, naïve Bayes classifier, and user survey | 29 |
| Sikdar et al. [91] | - | + | + | Machine learning and expectation maximization | N/A |
| Gün and Karagöz [178] | + | + | + | Random forest tree, J48 tree, ADTree, random tree, BFTree, naïve Bayes classifier, KStar and AdaBoost, and graph-based | 43 |
| Lorek et al. [55] | + | + | + | Manual tagging and random forest | 12 |
| Mitra and Gilbert [83] | + | + | - | Crowdsourcing, Euclidean distance similarity metric, LDA, and clustering | N/A |
| AlRubaian et al. [4] | + | + | + | Human expert, naïve Bayes classifier, and pairwise comparison | 22 |
| Wang et al. [179] | + | - | + | Fact-finding, maximum likelihood estimation (expectation maximization) | N/A |
| Mukherjee et al. [181] | - | + | + | Probabilistic graphical model, Markov random field, and SVM | N/A |
| Kumar and Geethakumari [131] | + | - | + | Statistical method, PageRank algorithm | N/A |
| Gupta and Kumaraguru [54] | - | + | + | Ranking SVM and linear logistic regression analysis | 27 |
| Kang et al. [180] | + | + | + | Amazon M Turk and a J-48 tree (C4.5 rules) | 55 |
| Briscoe et al. [134] | - | + | + | Graph-based and principal components analysis | N/A |

Rank-SVM effectiveness for the messages receiving the highest scores. To initialize the rankings, they used the time spent as the highest ranked message as the parameter for sorting the messages. In general, we perceived that researchers have used several evaluation criteria for various analysis tasks; there is no a common evaluation guideline for evaluation the content that originated from OSN's.

## IX. CONCLUSION

We have carried out a comprehensive literature review of credibility assessment studies of a reputable OSN, Twitter. We discussed these works from different levels of feature extraction and methodology. In addition, we discussed a summary of the existing works in this field, which could be of great value for researchers who wish to gain an understanding

**TABLE 10.** Projects and systems developed to analyze credibility.

| Name | URL | Type |
|---|---|---|
| Stanford Web Credibility Research | http://credibility.stanford.edu/ | Project |
| Pheme | http://www.pheme.eu/ | Project |
| REVEAL | http://revealproject.eu/ | Project |
| Truthy | http://www.truthy.indiana.edu/ | Project and system |
| TweetCred | https://chrome.google.com/webstore/detail/tweetcred/fbokljinlogeihdnkikeeneiankdgikg?hl=en | System |
| TwitterBOT | Proposed Solution (Future work) | System |
| Fake Tweet Buster | http://grupoweb.upf.edu/fake-buster | Web application |
| TRAILS | http://twittertrails.com/ | System |
| TopicNets system, compared with "Fluo" | http://cs.ucsb.edu/~jod/topicnets.html | System |

of the relevant credibility analysis methodologies and the level of features in these assessments. The motivation for this study was the rise of online social media information, regardless of the difficulty in filtering credible information sources. The analysis aimed to highlight the importance of the topic, providing previous studies related to the issue, outlining different approaches to solving the problem, and giving suggestions for future work. Our recommendations include the following:

1. Twitter credibility researchers commonly use text analysis tasks. However, analysis of multimedia (images, audio, and video) must be explored further.
2. Text analysis has been employed effectively; nonetheless, semantic analysis of text content has not been explored.
3. The feature levels of credibility assessments require further investigation, especially in terms of relative importance. In some cases, certain features are more important than others, leading to misjudgment of the trustworthiness of the content and source.
4. We believe that a hybrid model can leverage the advantages of both the human- and automation-based models. We hope that future research will expand the hybrid models to formulate automation relevant to social media content credibility judgment.
5. The studies in this area lack experiments with larger datasets and high-performance algorithms. In addition, there is a lack of publicly available standard) datasets with which to benchmark the different methodologies used.

Credibility analysis of OSN content is an encouraging research field. This survey has covered the current top-notch contributions related to this field and identified many relevant issues that deserve further study.

## APPENDIX

See Tables 1–10.

## ACKNOWLEDGMENTS

The authors would like to thank Prof. Huan Liu from Arizona State University and Prof. Reda Alhajj from Calgary University for their valuable comments, which improved the content of this paper.

## REFERENCES

[1] W. Jiang, J. Wu, F. Li, G. Wang, and H. Zheng, "Trust evaluation in online social networks using generalized network flow," *IEEE Trans. Comput.*, vol. 65, no. 3, pp. 952–963, Mar. 2016.

[2] S. Ranganath, S. Wang, X. Hu, J. Tang, and H. Liu, "Facilitating time critical information seeking in social media," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, pp. 2197–2209, Oct. 2017.

[3] C.-T. Li, Y.-J. Lin, and M.-Y. Yeh, "Forecasting participants of information diffusion on social networks with its applications," *Inf. Sci.*, vol. 422, pp. 432–446, Jan. 2018.

[4] J. Kim and M. Hastak, "Social network analysis: Characteristics of online social networks after a disaster," *Int. J. Inf. Manage.*, vol. 38, no. 1, pp. 86–96, 2018.

[5] Z. Zhang, R. Sun, X. Wang, and C. Zhao, "A situational analytic method for user behavior pattern in multimedia social networks," *IEEE Trans. Big Data*, vol. 24, no. 1, pp. 1, Jan. 2017.

[6] M. Al-Qurishi, M. S. Hossain, M. Alrubaian, S. M. M. Rahman, and A. Alamri, "Leveraging analysis of user behavior to identify malicious activities in large-scale social networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 799–813, Feb. 2017.

[7] Z. Du, Y. Yang, Q. Cai, C. Zhang, and Y. Bai, "Modeling and inferring mobile phone users' negative emotion spreading in social networks," *Future Gener. Comput. Syst.*, vol. 78, pp. 933–942, Jan. 2018.

[8] 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. Inform. Integr. Web-Based Appl. Services*, 2013, p. 107.

[9] F. Sáez-Mateu, "Democracy, screens, identity, and social networks: The case of donald trump's election," *Amer. Behav. Sci.*, vol. 62, no. 3, pp. 320–334, 2017.
[10] A. Paradise *et al.*, "Creation and management of social network honeypots for detecting targeted cyber attacks," *IEEE Trans. Comput. Social Syst.*, vol. 4, no. 3, pp. 65–79, Sep. 2017.
[11] M. Al-Qurishi, M. Al-Rakhami, A. Alamri, M. Alrubaian, S. M. M. Rahman, and M. S. Hossain, "Sybil defense techniques in online social networks: A survey," *IEEE Access*, vol. 5, pp. 1200–1219, 2017.
[12] C. Dong and B. Zhou, "Spam detection, E-mail/social network," in *Encyclopedia of Social Network Analysis and Mining*. New York, NY, USA: Springer, 2014, pp. 1954–1960.
[13] E. Morozov and M. Sen, "Analysing the Twitter social graph: Whom can we trust?" M.S. thesis, Dept. Comput. Sci., Univ. Nice Sophia Antipolis, Nice, France, 2014.
[14] C. Yang, R. C. Harkreader, and G. Gu, "Die free or live hard? Empirical evaluation and new design for fighting evolving Twitter spammers," in *Recent Advances in Intrusion Detection*. Berlin, Germany: Springer, 2011, pp. 318–337.
[15] J. Maddock, K. Starbird, H. J. Al-Hassani, D. E. Sandoval, M. Orand, and R. M. Mason, "Characterizing online rumoring behavior using multidimensional signatures," in *Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Social Comput.*, 2015, pp. 228–241.
[16] N. P. Nguyen, G. Yan, M. T. Thai, and S. Eidenbenz, "Containment of misinformation spread in online social networks," in *Proc. 4th Annu. ACM Web Sci. Conf.*, 2012, pp. 213–222.
[17] D. T. Nguyen, N. P. Nguyen, and M. T. Thai, "Sources of misinformation in online social networks: Who to suspect?" in *Proc. Mil. Commun. Conf. (MILCOM)*, 2012, pp. 1–6.
[18] A. A. AlMansour, L. Brankovic, and C. S. Iliopoulos, "A model for recalibrating credibility in different contexts and languages—A Twitter case study," *Int. J. Digit. Inf. Wireless Commun.*, vol. 4, pp. 53–62, Jan. 2014.
[19] A. Joshi, U. D. S. Bedathur, and V. Goyal, "A survey on analyzing and measuring trustworthiness of user-generated content on Twitter during high-impact events," Ph.D. dissertation, Indraprastha Inst. Inf. Technol., New Delhi, India, 2013.
[20] A. A. Shah, S. D. Ravana, S. Hamid, and M. A. Ismail, "Web credibility assessment: Affecting factors and assessment techniques," *Inf. Res.*, vol. 20, no. 1, pp. 1–28, Mar. 2015.
[21] X. Si, J. G. Deng, H. Ke, D. Zhang, Z. I. Gyongyi, and E. Y. Chang, "Ranking user generated Web content," U.S. Patent 8 965 883, Feb. 24, 2015.
[22] W. Choi and B. Stvilia, "Web credibility assessment: Conceptualization, operationalization, variability, and models," *J. Assoc. Inf. Sci. Technol.*, vol. 66, pp. 2399–2414, Dec. 2015.
[23] W. Weerkamp and M. de Rijke, "Credibility-inspired ranking for blog post retrieval," *Inf. Retr.*, vol. 15, nos. 3–4, pp. 243–277, 2012.
[24] S. Y. Rieh and D. R. Danielson, "Credibility: A multidisciplinary framework," *Annu. Rev. Inf. Sci. Technol.*, vol. 41, no. 1, pp. 307–364, 2007.
[25] K. Kwon, J. Cho, and Y. Park, "Multidimensional credibility model for neighbor selection in collaborative recommendation," *Expert Syst. Appl.*, vol. 36, pp. 7114–7122, Apr. 2009.
[26] P. Kamthan, "A framework for the active credibility engineering of Web applications," *Int. J. Inf. Technol. Web Eng.*, vol. 3, no. 3, pp. 17–27, 2008.
[27] J. Lazar, G. Meiselwitz, and J. Feng, "Understanding Web credibility: A synthesis of the research literature," *Found. Trends Hum.-Comput. Interact.*, vol. 1, no. 2, pp. 139–202, 2007.
[28] H. Han, H. Nakawatase, and K. Oyama, "Evaluating credibility of interest reflection on Twitter," *Int. J. Web Inf. Syst.*, vol. 10, no. 4, pp. 343–362, 2014.
[29] H. S. Al-Khalifa and R. M. Al-Eidan, "An experimental system for measuring the credibility of news content in Twitter," *Int. J. Web Inf. Syst.*, vol. 7, no. 2, pp. 130–151, 2011.
[30] B. Hilligoss and S. Y. Rieh, "Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context," *Inf. Process. Manage.*, vol. 44, pp. 1467–1484, Jul. 2008.
[31] K. S. K. Chung, M. Piraveenan, and L. Hossain, "Topology of online social networks," in *Encyclopedia of Social Network Analysis and Mining*, R. Alhajj and J. Rokne, Eds. New York, NY, USA: Springer, 2014, pp. 2191–2202.
[32] M. O. Jackson, *Social and Economic Networks*, vol. 3. Princeton, NJ, USA: Princeton Univ. Press, 2008.
[33] M. AlRubaian, M. Al-Qurishi, M. Al-Rakhami, S. M. M. Rahman, and A. Alamri, "A multistage credibility analysis model for microblogs," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2015, pp. 1434–1440.
[34] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, "Finding high-quality content in social media," in *Proc. Int. Conf. Web Search Data Mining*, 2008, pp. 183–194.
[35] B. Kang, T. H. Höllerer, M. Turk, X. Yan, and J. O'Donovan, "An analysis of credibility in microblogs," M.S. thesis, Dept. Comput. Sci., Univ. California, Santa Barbara, Santa Barbara, CA, USA, 2012.
[36] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," presented at the 20th Int. Conf. World Wide Web, Hyderabad, India, 2011.
[37] N. C. Burbules, "Paradoxes of the Web: The ethical dimensions of credibility," *Library Trends*, vol. 49, no. 3, pp. 441–453, 2001.
[38] M. R. Solomon, D. W. Dahl, K. White, J. L. Zaichkowsky, and R. Polegato, *Consumer Behavior: Buying, Having, and Being*. Upper Saddle River, NJ, USA: Prentice-Hall, 2014.
[39] B. Cugelman, M. Thelwall, and P. Dawes, "Website credibility, active trust and behavioural intent," in *Proc. 3rd Int. Conf. Persuasive Technol.*, 2008, pp. 47–57.
[40] T. R. Cosenza, M. R. Solomon, and W.-S. Kwon, "Credibility in the blogosphere: A study of measurement and influence of wine blogs as an information source," *J. Consum. Behav.*, vol. 14, no. 2, pp. 71–91, 2015.
[41] R. L. Wakefield and D. Whitten, "Examining user perceptions of third-party organizations credibility and trust in an e-retailer," *J. Org. User Comput.*, vol. 18, no. 2, pp. 1–19, 2007.
[42] C. Sichtmann, "An analysis of antecedents and consequences of trust in a corporate brand," *Eur. J. Marketing*, vol. 41, pp. 999–1015, Sep. 2007.
[43] D. Artz and Y. Gil, "A survey of trust in computer science and the semantic Web," *J. Web Semantics*, vol. 5, pp. 58–71, Jun. 2007.
[44] J.-H. Cho, K. Chan, and S. Adali, "A survey on trust modeling," *ACM Comput. Surv.*, vol. 48, no. 2, 2015, Art. no. 28.
[45] W. Sherchan, S. Nepal, and C. Paris, "A survey of trust in social networks," *ACM Comput. Surv.*, vol. 45, no. 4, 2013, Art. no. 47.
[46] M. Imran, C. Castillo, F. Diaz, and S. Vieweg. (Jul. 2014). "Processing social media messages in mass emergency: A survey." [Online]. Available: https://arxiv.org/abs/1407.7071
[47] A. A. AlMansour, L. Brankovic, and C. S. Iliopoulos, "Evaluation of credibility assessment for microblogging: Models and future directions," in *Proc. 14th Int. Conf. Knowl. Technol. Data-Driven Bus.*, 2014, Art. no. 32.
[48] A. A. AlMansour, "Towards customizing credibility in different contexts: Languages, topics and locations—A Twitter case study," in *Proc. Int. Conf. Digit. Inf. Process., e-Bus. Cloud Comput.*, 2013, pp. 218–224.
[49] S. Sikdar, B. Kang, J. O'Donovan, T. Höllerer, and S. Adal, "Cutting through the noise: Defining ground truth in information credibility on Twitter," in *Proc. HUMAN*, vol. 2, 2013, pp. 151–167.
[50] S. Sikdar *et al.*, "Finding true and credible information on Twitter," in *Proc. 17th Int. Conf. Inf. Fusion*, 2014, pp. 1–8.
[51] H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: Characterizing and identifying fake images on Twitter during hurricane sandy," presented at the 22nd Int. Conf. World Wide Web Companion, Rio de Janeiro, Brazil, 2013.
[52] C. Boididou, S. Papadopoulos, Y. Kompatsiaris, S. Schifferes, and N. Newman, "Challenges of computational verification in social multimedia," in *Proc. 23rd Int. Conf. World Wide Web Companion*, 2014, pp. 743–748.
[53] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: Can we trust what we RT?" in *Proc. 1st Workshop Social Media Anal.*, 2010, pp. 71–79.
[54] A. Gupta and P. Kumaraguru, "Credibility ranking of tweets during high impact events," in *Proc. 1st Workshop Privacy Secur. Online Social Media*, 2012, p. 2.
[55] K. Lorek, J. Suehiro-Wici ski, M. Jankowski-Lorek, and A. Gupta, "Automated credibility assessment on Twitter," *Comput. Sci.*, vol. 16, no. 2, pp. 157–168, 2015.
[56] R. M. B. Al-Eidan, H. S. Al-Khalifa, and A. S. Al-Salman, "Towards the measurement of Arabic Weblogs credibility automatically," in *Proc. 11th Int. Conf. Inf. Integr. Web-Based Appl. Services*, 2009, pp. 618–622.
[57] R. M. B. Al-Eid, R. S. Al-Khalif, and A. S. Al-Salman, "Measuring the credibility of Arabic text content in Twitter," in *Proc. 5th Int. Conf. Digital Inf. Manage.*, 2010, pp. 285–291.

[58] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, "TweetCred: Real-time credibility assessment of content on Twitter," in *Social Informatics*. New York, NY, USA: Springer, 2014, pp. 228–243.
[59] A. Nagy and J. Stamberger, "Credo: A framework for semi-supervised credibility assessment for social networks," in *Proc. Int. Conf. Data Mining*, 2012, p. 1.
[60] S. Sun, H. Liu, J. He, and X. Du, "Detecting event rumors on Sina Weibo automatically," in *Web Technologies and Applications*. New York, NY, USA: Springer, 2013, pp. 120–131.
[61] J. Ratkiewicz *et al.*, "Truthy: Mapping the spread of astroturf in microblog streams," in *Proc. 20th Int. Conf. Companion World Wide Web*, 2011, pp. 249–252.
[62] C. Boididou *et al.*, "Verifying multimedia use at MediaEval 2015," in *Proc. MediaEval Workshop*, Wurzen, Germany, 2015, p. 3.
[63] S. E. Middleton, "Extracting attributed verification and debunking reports from social media: MediaEval-2015 trust and credibility analysis of image and video," in *Proc. MediaEval Workshop*, Wurzen, Germany, 2015, p. 3.
[64] C. Boididou, S. Papadopoulos, D.-T. Dang-Nguyen, G. Boato, and Y. Kompatsiaris, "The CERTH-UNITN participation @ verifying multimedia use 2015," in *Proc. MediaEval Workshop*, Wurzen, Germany, 2015, pp. 1–3.
[65] A. L. Ginsca, A. Popescu, M. Lupu, A. Iftene, and I. Kanellos, "Evaluating user image tagging credibility," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. New York, NY, USA: Springer, 2015, pp. 41–52.
[66] J. Park, M. Cha, H. Kim, and J. Jeong, "Sentiment analysis on bad news spreading," in *Proc. ACM CSCW Workshop Design, Influence, Social Technol.*, 2012, pp. 1–7.
[67] B. Kang, J. O'Donovan, and T. Höllerer, "Modeling topic specific credibility on Twitter," in *Proc. ACM Int. Conf. Intell. User Interfaces*, 2012, pp. 179–188.
[68] Y. Ikegami, K. Kawai, Y. Namihira, and S. Tsuruta, "Topic and opinion classification based information credibility analysis on Twitter," in *Proc. IEEE Int. Conf. Syst, Man, Cybern.*, Oct. 2013, pp. 4676–4681.
[69] D. H. Widyantoro and Y. Wibisono, "Modeling credibility assessment and explanation for tweets based on sentiment analysis," *J. Theor. Appl. Inf. Technol.*, vol. 70, no. 3, pp. 540–548, 2014.
[70] C. Buntain, "Discovering credible events in near real time from social media streams," in *Proc. 24th Int. Conf. World Wide Web Companion*, 2015, pp. 481–485.
[71] A. L. Ginsca, "Estimating user credibility in multimedia information flows," in *Proc. 5th BCS-IRSG Symp. Future Directions Inf. Access (FDIA)*, 2013, pp. 51–52.
[72] M. Thandar and S. Usanavasin, "Measuring opinion credibility in Twitter," in *Recent Advances in Information and Communication Technology*. New York, NY, USA: Springer, 2015, pp. 205–214.
[73] R. Cohen and D. Ruths, "Classifying political orientation on Twitter: It's not easy!" in *Proc. 7th Int. AAAI Conf. Weblogs Social Media*, 2013, pp. 91–98.
[74] J. D. Burger, J. Henderson, G. Kim, and G. Zarrella, "Discriminating gender on Twitter," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2011, pp. 1301–1309.
[75] F. Al Zamal, W. Liu, and D. Ruths, "Homophily and latent attribute inference: Inferring latent attributes of Twitter users from neighbors," in *Proc. 7th Int. AAAI Conf. Weblogs Social Media*, 2013, pp. 387–390.
[76] W. Liu and D. Ruths, "What's in a name? Using first names as features for gender inference in Twitter," in *Proc. AAAI Spring Symp., Anal. Microtext*, 2013, pp. 10–16.
[77] A. Pal and S. Counts, "What's in a @name? How name value biases judgment of microblog authors," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011.
[78] D. Rao and D. Yarowsky, "Detecting latent user properties in social media," in *Proc. NIPS MLSN Workshop*, 2010, pp. 1–7.
[79] G. Stringhini *et al.*, "Follow the green: Growth and dynamics in Twitter follower markets," in *Proc. Conf. Internet Meas.*, 2013, pp. 163–176.
[80] D. Micheli and A. Stroppa, "Twitter and the underground market," in *Proc. 11th Nexa Lunch Seminar*, 2013, pp. 5–9.
[81] K. R. Canini, B. Suh, and P. Pirolli, "Finding relevant sources in Twitter based on content and social structure," in *Proc. NIPS MLSN Workshop*, 2010, pp. 1–7.
[82] A. Gupta and P. Kumaraguru, "Designing and evaluating techniques to mitigate misinformation spread on microblogging Web services," Tech. Rep., 2015.
[83] T. Mitra and E. Gilbert, "CREDBANK: A large-scale social media corpus with associated credibility annotations," in *Proc. 9th Int. AAAI Conf. Web Social Media*, 2015, pp. 258–267.
[84] Q. V. Liao, P. Pirolli, and W.-T. Fu, "An ACT-R model of credibility judgment of micro-blogging Web pages," in *Proc. Int. Conf. Cogn. Modeling*, vol. 103, 2012, pp. 103–108.
[85] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on Sina Weibo," in *Proc. ACM SIGKDD Workshop Mining Data Semantics*, 2012, Art. no. 13.
[86] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
[87] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, "TweetCred: A real-time Web-based system for assessing credibility of content on Twitter," in *Proc. 6th Int. Conf. Social Inf. (SocInfo)*, 2014, p. 1405.5490
[88] J. Zou, F. Fekri, and S. W. McLaughlin, "Mining streaming tweets for real-time event credibility prediction in Twitter," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2015, pp. 1586–1589.
[89] B. Batrinca and P. C. Treleaven, "Social media analytics: A survey of techniques, tools and platforms," *AI Soc.*, vol. 30, no. 1, pp. 89–116, 2015.
[90] K. R. Canini, B. Suh, and P. L. Pirolli, "Finding credible information sources in social networks based on content and social structure," in *Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust IEEE Int. Conf. Social Comput.*, Oct. 2011, pp. 1–8.
[91] S. Sikdar *et al.*, "Finding true and credible information on Twitter," in *Proc. 17th Int. Conf. Inf. Fusion (FUSION)*, 2014, pp. 1–8.
[92] D. Saez-Trumper, "Fake tweet buster: A Webtool to identify users promoting fake news on Twitter," in *Proc. 25th ACM Conf. Hypertext Social Media*, 2014, pp. 316–317.
[93] A. Gupta and P. Kumaraguru, "@Twitter credibility ranking of tweets on events # breakingnews," Indraprastha Inst. Inf. Technol., Delhi, India Tech. Rep., 2012.
[94] Castillo, M. Mendoza, and B. Poblete, "Predicting information credibility in time-sensitive social media," *Internet Res.*, vol. 23, pp. 560–588, Oct. 2013.
[95] M. Gupta, P. Zhao, and J. Han, "Evaluating event credibility on Twitter," in *Proc. SIAM Int. Conf. Data Mining*, 2012, pp. 153–164.
[96] E. J. Briscoe, D. S. Appling, and H. Hayes, "Social network derived credibility," in *Recommendation and Search in Social Networks*, E. J. Briscoe, D. S. Appling, and H. Hayes, Eds. New York, NY, USA: Springer, 2015, pp. 59–75.
[97] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and passivity in social media," in *Machine Learning and Knowledge Discovery in Databases*. New York, NY, USA: Springer, 2011, pp. 18–33.
[98] A. Olteanu, S. Peshterliev, X. Liu, and K. Aberer, "Web credibility: Features exploration and credibility prediction," in *Advances in Information Retrieval*. New York, NY, USA: Springer, 2013, pp. 557–568.
[99] R. V. K. A. and I. S. C., "Topical categorization of credible microblog content," *Int. J. Sci. Res. Dev.*, vol. 2, pp. 655–658, 2014.
[100] Joshi, U. D. S. Bedathur, V. Goyal, and P. K. PK, "Analyzing and measuring trustworthiness of user-generated content on Twitter to study high-impact events," Tech. Rep., 2013.
[101] K. R. Saikaew and C. Noyunsan, "Features for measuring credibility on Facebook information," *Int. J. Comput., Autom., Control, Inf. Eng.*, vol. 9, no. 1, pp. 174–177, 2015.
[102] D. Wang, "Analysis and detection of low quality information in social networks," in *Proc. IEEE 30th Int. Conf. Data Eng. Workshops*, Mar./Apr. 2014, pp. 350–354.
[103] P. T. Metaxas and E. Mustafaraj, "Trails of trustworthiness in real-time streams (extended summary)," presented at the Design, Influence Social Technol. DIST Workshop ACM Comput. Supported Cooperat. Work (CSCW), Seattle, WA, USA, 2012.
[104] A. Gupta, M. Gupta, and P. Kumaraguru, "Twit-digest: An online solution for analyzing and visualizing Twitter in real-time," Tech. Rep.
[105] Z. Sharf and A. U. Saeed, "Twitter news credibility meter," *Int. J. Comput. Appl.*, vol. 83, pp. 49–51, Jan. 2013.
[106] X. Xia, X. Yang, C. Wu, S. Li, and L. Bao, "Information credibility on Twitter in emergency situation," in *Intelligence and Security Informatics*. New York, NY, USA: Springer, 2012, pp. 45–59.
[107] S. Finn, P. T. Metaxas, and E. Mustafaraj, "Spread and skepticism: Metrics of propagation on Twitter," in *Proc. ACM Web Sci. Conf.*, 2015, Art. no. 39.

[108] X. Liu, R. Nielek, A. Wierzbicki, and K. Aberer, "Defending imitating attacks in Web credibility evaluation systems," in *Proc. 22nd Int. Conf. World Wide Web Companion*, 2013, pp. 1115–1122.

[109] X. Wu, Z.-M. Feng, W. Fan, J. Gao, and Y. Yu, "Detecting marionette microblog users for improved information credibility," in *Machine Learning and Knowledge Discovery in Databases*. New York, NY, USA: Springer, 2013, pp. 483–498.

[110] M.-A. Abbasi and H. Liu, "Measuring user credibility in social media," in *Social Computing, Behavioral-Cultural Modeling and Prediction*. New York, NY, USA: Springer, 2013, pp. 441–448.

[111] J. Al-Sharawneh, S. Sinnappan, and M.-A. Williams, "Credibility-based Twitter social network analysis," in *Web Technologies and Applications*. New York, NY, USA: Springer, 2013, pp. 323–331.

[112] J. Al-Sharawneh and M.-A. Williams, "Credibility-based social network recommendation: Follow the leader," in *Proc. AIS*, 2010, pp. 1–9.

[113] D. Wang *et al.*, "On Bayesian interpretation of fact-finding in information networks," in *Proc. 14th Int. Conf. Inf. Fusion (FUSION)*, 2011, pp. 1–8.

[114] H. Huang *et al.*, "Tweet ranking based on heterogeneous networks," in *Proc. 24th Int. Conf. Comput. Ling.*, 2012, pp. 1239–1256.

[115] S. Ravikumar, R. Balakrishnan, and S. Kambhampati, "Ranking tweets considering trust and relevance," in *Proc. 9th Int. Workshop Inf. Integr. Web*, 2012, Art. no. 4.

[116] B. Ulicny and M. M. Kokar, "Automating military intelligence confidence assessments for Twitter messages," in *Proc. 6th Annu. Network Sci. Workshop*, 2012, pp. 1–9.

[117] K. R. McKelvey and F. Menczer, "Truthy: Enabling the study of online social networks," in *Proc. Conf. Comput. Supported Cooperat. Work Companion*, 2013, pp. 23–26.

[118] D. T. Nguyen, N. P. Nguyen, and M. T. Thai, "Sources of misinformation in online social networks: Who to suspect," in *Proc. Mil. Commun. Conf. (MILCOM)*, 2012, pp. 1–6.

[119] J. Pasternack and D. Roth, "Latent credibility analysis," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 1009–1020.

[120] J. Schaffer *et al.*, "Interactive interfaces for complex network analysis: An information credibility perspective," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2013, pp. 464–469.

[121] Q. Qiu, R. Xu, B. Liu, L. Gui, and Y. Zhou, "Credibility estimation of stock comments based on publisher and information uncertainty evaluation," in *Machine Learning and Cybernetics*. Berlin, Germany: Springer, 2014, pp. 400–408.

[122] Y. Namihira, N. Segawa, Y. Ikegami, K. Kawai, T. Kawabe, and S. Tsuruta, "High precision credibility analysis of information on Twitter," in *Proc. Int. Conf. Signal-Image Technol. Internet-Based Syst.*, 2013, pp. 909–915.

[123] T. Kawabe *et al.*, "Tweet credibility analysis evaluation by improving sentiment dictionary," in *Proc. IEEE Congr. Evol. Comput.*, May 2015, pp. 2354–2361.

[124] Y. Suzuki and A. Nadamoto, "Credibility assessment using Wikipedia for messages on social network services," in *Proc. IEEE 9th Int. Conf. Dependable, Auto. Secure Comput.*, Dec. 2011, pp. 887–894.

[125] B. Kang, J. O'Donovan, and T. Höllerer, "A framework for modeling trust in collaborative ontologies," in *Proc. 6th Graduate Student Workshop Comput.*, 2011, p. 39.

[126] S. Kumar, "Ranking assessment of event tweets for credibility," *Int. J. Sci. Eng. Res.*, vol. 1, pp. 45–50, 2013.

[127] M. Sirivianos, K. Kim, and X. Yang, "FaceTrust: Assessing the credibility of online personas via social networks," in *Proc. 4th USENIX Conf. Hot Topics Secur.*, 2009, pp. 1–6.

[128] A. Gupta, H. Lamba, and P. Kumaraguru, "1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on Twitter," in *Proc. eCrime Researchers Summit (eCRS)*, 2013, pp. 1–12.

[129] E. Conte, "What path monitor: A brief note on quantum cognition and quantum interference, the role of the knowledge factor," *Psychology*, vol. 6, no. 3, pp. 291–296, 2015.

[130] M. J. Metzger and A. J. Flanagin, "Credibility and trust of information in online environments: The use of cognitive heuristics," *J. Pragmatics*, vol. 59, pp. 210–220, Dec. 2013.

[131] K. P. K. Kumar and G. Geethakumari, "Detecting misinformation in online social networks using cognitive psychology," *Hum.-Centric Comput. Inf. Sci.*, vol. 4, p. 14, Dec. 2014.

[132] D. Westerman, P. R. Spence, and B. Van Der Heide, "A social network as information: The effect of system generated reports of connectedness on credibility on Twitter," *Comput. Hum. Behav.*, vol. 28, no. 1, pp. 199–206, 2012.

[133] E. Jaho, E. Tzoannos, A. Papadopoulos, and N. Sarris, "Alethiometer: A framework for assessing trustworthiness and content validity in social media," in *Proc. 23rd Int. Conf. World Wide Web Companion*, 2014, pp. 749–752.

[134] E. J. Briscoe, D. S. Appling, R. L. Mappus, IV, and H. Hayes, "Determining credibility from social network structure," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2013, pp. 1418–1424.

[135] M. Kang, "Measuring social media credibility: A study on a measure of blog credibility," Inst. Public Relations, Univ. Florida, Gainesville, FL, USA, Tech. Rep., 2010.

[136] K. Štepánová, "Computational cognitive modeling," Czech Tech. Univ. Prague, Jugoslávských partyzán , Czech Republic, Tech. Rep., 2015.

[137] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi, "Cognos: Crowdsourcing search for topic experts in microblogs," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2012, pp. 575–590.

[138] J. Schaffer *et al.*, "Truth, lies, and data: Credibility representation in data analysis," in *Proc. IEEE Int. Inter-Disciplinary Conf. Cogn. Methods Situation Awareness Decis. Support*, Mar. 2014, pp. 28–34.

[139] B. De Longueville, R. S. Smith, and G. Luraschi, "'OMG, from here, I can see the flames!': A use case of mining location based social networks to acquire spatio-temporal data on forest fires," presented at the Int. Workshop Location-Based Social Netw., Seattle, WA, USA, 2009.

[140] K. Kireyev, L. Palen, and K. M. Anderson, "Applications of topics models to analysis of disaster-related Twitter data," in *Proc. NIPS Workshop Appl. Topic Models, Text and Beyond*, 2009, pp. 1–4.

[141] A. L. Hughes and L. Palen, "Twitter adoption and use in mass convergence and emergency events," *Int. J. Emergency Manage.*, vol. 6, nos. 3–4, pp. 248–260, 2009.

[142] J. O'Donovan, B. Kang, G. Meyer, T. Höllerer, and S. Adalii, "Credibility in context: An analysis of feature distributions in Twitter," in *Proc. Int. Conf. Privacy, Secur., Risk Trust Int. Conf. Social Comput.*, 2012, pp. 293–301.

[143] M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz, "Tweeting is believing? Understanding microblog credibility perceptions," in *Proc. ACM Conf. Comput. Supported Cooperat. Work*, 2012, pp. 441–450.

[144] M. Schmierbach and A. Oeldorf-Hirsch, "A little bird told me, so I didn't believe it: Twitter, credibility, and issue perceptions," *Commun. Quart.*, vol. 60, pp. 317–337, Jul. 2012.

[145] J. Yang, S. Counts, M. R. Morris, and A. Hoff, "Microblog credibility perceptions: Comparing the United States and China," in *Proc. Conf. Comput. Supported Cooperat. Work*, 2013, pp. 575–586.

[146] Y. Suzuki, "A credibility assessment for message streams on microblogs," in *Proc. Int. Conf. P2P, Parallel, Grid, Cloud Internet Comput.*, 2010, pp. 527–530.

[147] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Aspects of rumor spreading on a microblog network," in *Social Informatics*. NewYork, NY, USA: Springer, 2013, pp. 299–308.

[148] S. Y. Rieh, G. Y. Jeon, J.-Y. Yang, and C. Lampe, "Audience-aware credibility: From understanding audience to establishing credible blogs," in *Proc. 8th Int. AAAI Conf. Weblogs Social Media*, 2014, pp. 1–12.

[149] B. Kang, T. Höllerer, and J. O'Donovan, "Believe it or not? Analyzing information credibility in microblogs," presented at the IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, Paris, France, Aug. 2015.

[150] C. L. Armstrong and M. J. McAdams, "Blogs of information: How gender cues and individual motivations influence perceptions of credibility," *J. Comput. Mediated Commun.*, vol. 14, no. 3, pp. 435–456, 2009.

[151] L. Llamero, "Conceptual mindsets and heuristics in credibility evaluation of e-Word of Mouth in tourism," *Online Inf. Rev.*, vol. 38, no. 7, pp. 954–968, 2014.

[152] Q. Gao, Y. Tian, and M. Tu, "Exploring factors influencing Chinese user's perceived credibility of health and safety information on Weibo," *Comput. Hum. Behav.*, vol. 45, pp. 21–31, Apr. 2015.

[153] C. Edwards, A. Edwards, P. R. Spence, and A. K. Shelton, "Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter," *Comput. Hum. Behav.*, vol. 33, pp. 372–376, Apr. 2014.

[154] B. Ulicny and K. Baclawski, "New metrics for newsblog credibility," in *Proc. Int. Conf. Weblogs Social Media*, 2007, pp. 1–2.

[155] J. Finn and H. G. de Zúñiga, "Online credibility and community among blog users," in *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 48, no. 1, pp. 1–9, 2011.

[156] S. Y. Rieh, "Participatory Web users' information activities and credibility assessment," Tech. Rep., 2010.

[157] T. Bakker, D. Trilling, C. de Vreese, L. Helfer, and K. Schönbach, "The context of content: The impact of source and setting on the credibility of news," *Recherches Commun.*, vol. 40, no. 40, pp. 151–168, 2013.
[158] K. A. Johnson, "The effect of *Twitter* posts on students' perceptions of instructor credibility," *Learn., Media Technol.*, vol. 36, no. 1, pp. 21–38, 2011.
[159] T. Chesney and D. K. S. Su, "The impact of anonymity on weblog credibility," *Int. J. Hum.-Comput. Stud.*, vol. 68, pp. 710–718, Oct. 2010.
[160] K. Xu, Y. Liu, X. Zhao, and X. Dong, "Trust them or not? A study on media credibility of newspapers accounts on Sina Weibo," *SSRN Electron. J.*, pp. 1–29, Apr. 2013.
[161] R. Thomson *et al.*, "Trusting tweets: The Fukushima disaster and information source credibility on Twitter," in *Proc. 9th Int. ISCRAM Conf.*, 2012, pp. 1–10.
[162] J. Pattanaphanchai, K. O'Hara, and W. Hall, "Trustworthiness criteria for supporting users to assess the credibility of Web information," in *Proc. 22nd Int. Conf. World Wide Web Companion*, 2013, pp. 1123–1130.
[163] A. Zubiaga and H. Ji, "Tweet, but verify: Epistemic study of information verification on Twitter," *Social Netw. Anal. Mining*, vol. 4, p. 163, Dec. 2014.
[164] J. R. C. Nurse, I. Agrafiotis, M. Goldsmith, S. Creese, and K. Lamberts, "Two sides of the coin: Measuring and communicating the trustworthiness of online information," *J. Trust Manage.*, vol. 1, p. 5, May 2014.
[165] A. A. AlMansour and C. S. Iliopoulos, "Using Arabic microblogs features in determining credibility," in *Proc. Int. Conf. Adv. Social Netw. Anal. Mining*, 2015, pp. 1212–1219.
[166] E. Go, K. H. You, E. Jung, and H. Shim, "Why do we use different types of websites and assign them different levels of credibility? Structural relations among users' motives, types of websites, information credibility, and trust in the press," *Comput. Hum. Behav.*, vol. 54, pp. 231–239, Jan. 2016.
[167] S. M. Shariff, X. Zhang, and M. Sanderson, "User perception of information credibility of news on Twitter," in *Advances in Information Retrieval*. New York, NY, USA: Springer, 2014, pp. 513–518.
[168] T. J. Johnson and B. K. Kaye, "Reasons to believe: Influence of credibility on motivations for using social networks," *Comput. Hum. Behav.*, vol. 50, pp. 544–555, Sep. 2015.
[169] T. J. Johnson and B. K. Kaye, "Credibility of social network sites for political information among politically interested Internet users," *J. Comput. Mediated Commun.*, vol. 19, no. 4, pp. 957–974, 2014.
[170] S. Sikdar, B. Kang, J. O'Donovan, T. Höllerer, and S. Adah, "Understanding information credibility on Twitter," in *Proc. Int. Conf. Social Comput.*, 2013, pp. 19–24.
[171] S. Aladhadh, X. Zhang, and M. Sanderson, "Tweet author location impacts on Tweet credibility," in *Proc. Australas. Document Comput. Symp.*, 2014, p. 73.
[172] 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, pp. 1–3, Jan. 2017.
[173] M. Alrubaian, M. Al-Qurishi, M. Al-Rakhami, and A. Alamri, "A credibility assessment model for online social network content," in *From Social Data Mining and Analysis to Prediction and Community Detection*. Cham, Switzerland: Springer, 2017, pp. 61–77.
[174] Twitter. (2014). *How to Report Spam in Twitter?*. [Online]. Available: http://support.twitter.com/articles/64986-how-to-report-spam-on-twitter
[175] J. Ito, J. Song, H. Toda, Y. Koike, and S. Oyama, "Assessment of Tweet credibility with LDA features," in *Proc. 24th Int. Conf. World Wide Web Companion*, 2015, pp. 953–958.
[176] S. Bhattacharya, H. Tran, P. Srinivasan, and J. Suls, "Belief surveillance with Twitter," in *Proc. 4th Annu. ACM Web Sci. Conf.*, 2012, pp. 43–46.
[177] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, "Rumor has it: Identifying misinformation in microblogs," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2011, pp. 1589–1599.
[178] A. Gün and P. Karagöz, "A hybrid approach for credibility detection in Twitter," in *Hybrid Artificial Intelligence Systems*. Cham, Switzerland: Springer, 2014, pp. 515–526.
[179] D. Wang, L. Kaplan, T. Abdelzaher, and C. C. Aggarwal, "On credibility estimation tradeoffs in assured social sensing," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 6, pp. 1026–1037, Jun. 2013.
[180] B. Kang, S. Sikdar, T. Höllerer, J. O'Donovan, and S. Adali, "Deconstructing information credibility on Twitter," in *Proc. 22nd Int. World Wide Web Conf.*, 2013.
[181] M. Viviani and G. Pasi, "Credibility in social media: Opinions, news, and health information—A survey," in *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. Hoboken, NJ, USA: Wiley, 2017.
[182] D. Hawking, A. Moffat, and A. Trotman, "Efficiency in information retrieval: Introduction to special issue," *Inf. Retr. J.*, vol. 20, no. 3, pp. 169–171, 2017.
[183] E. Djafarova and C. Rushworth, "Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users," *Comput. Hum. Behav.*, vol. 68, pp. 1–7, Mar. 2017.
[184] G. Sarna and M. P. S. Bhatia, "Content based approach to find the credibility of user in social networks: An application of cyberbullying," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 2, pp. 677–689, 2017.
[185] M. Thandar and S. Usanavasin, "Measuring opinion credibility in Twitter," in *Recent Advances in Information and Communication Technology*. Cham, Switzerland: Springer, 2015, pp. 205–214.

**MAJED ALRUBAIAN** received the Ph.D. degree from the Department of Information Systems, College of Computer and Information Sciences (CCIS), King Saud University (KSU), Riyadh, Saudi Arabia, in 2017. He has published several papers in refereed journals, including the IEEE, ACM, Springer, and Wiley. His research interests include social media analysis and mining data, information credibility, security informatics, and machine learning. He received the Best Ph.D. Thesis Award from CCIS, KSU, in 2017.

**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 Post-Doctoral with the Chair of Pervasive and Mobile Computing (CPMC), CCIS, KSU, and is one of the founding members of CPMC. He has published several papers 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 an Innovation Award for a mobile cloud serious game from KSU 2013. He also received the Best Ph.D. Thesis Award from CCIS, KSU, in 2018. He got the IBM Data Science Professional Certificate and deep learning certification from deeplearning.ai.

**ATIF ALAMRI** received the Ph.D. degree in computer science from the School of Information Technology and Engineering, University of Ottawa, Canada, in 2010. He is currently an Associate Professor with the Software Engineering Department, College of Computer and Information Sciences (CCIS), King Saud University (KSU), Riyadh, Saudi Arabia. He is one of the founding members of the Chair of Pervasive and Mobile Computing (CPMC), CCIS, KSU, successfully managing its research program, which transformed the chair as one of the best chairs of research excellence in the college. His research areas of interest are multimedia assisted health systems, ambient intelligence, service-oriented architecture. multimedia cloud, sensor-cloud, the Internet of Things, Big data, mobile cloud, social network, and recommender system.

**MABROOK AL-RAKHAMI** received the master's degree in information systems from King Saud University, Saudi Arabia. He is currently pursuing the Ph.D. degree with the Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He has authored several papers in refereed IEEE/ACM/Springer conferences and journals. His research interests include edge Intelligence, social networks, cloud computing, and the Internet of Things.

**MOHAMMAD MEHEDI HASSAN** received the Ph.D. degree in computer engineering from Kyung Hee University, South Korea, in 2011. He is currently an Associate Professor with the Information Systems Department, College of Computer and Information Sciences (CCIS), King Saud University (KSU), Riyadh, Saudi Arabia. His research areas of interests include cloud computing, multimedia, mobile cloud, sensor-cloud, the Internet of Things, and Big data. He received the Excellence in Research Award from CCIS, KSU, in 2015 and 2016.

**GIANCARLO FORTINO** has been a Professor of computer engineering with the Department of Informatics, Modeling, Electronics and Systems, University of Calabria, Rende, Italy, since 2006. He holds the Italian National Habilitation for Full Professorship. He has authored over 230 publications in journals, conferences, and books. His research interests include distributed computing, wireless sensor networks, software agents, cloud computing, and the Internet of Things systems. He is the Founding Editor of the Springer Book Series *Internet of Things: Technology, Communications and Computing* and serves in the editorial board of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, *Journal of Networks and Computer Applications*, *Engineering Applications of Artificial Intelligence*, *Information Fusion*, and *Multi Agent and GRID Systems*.