IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 11, 2021, accepted September 18, 2021, date of publication September 22, 2021, date of current version October 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3114417

Social Media and Microblogs Credibility: Identification, Theory Driven Framework, and Recommendation

KHUBAIB AHMED QURESHI[1], **RAUF AHMED SHAMS MALICK**[2], **AND MUHAMMAD SABIH**[3]
[1]Department of Computer Science, DHA Suffa University, Karachi 75500, Pakistan
[2]Department of Computer Science, National University of Computer and Emerging Sciences, Karachi 75230, Pakistan
[3]Department of Electrical Engineering, DHA Suffa University, Karachi 75500, Pakistan

Corresponding author: Khubaib Ahmed Qureshi (k.ahmed@dsu.edu.pk)

This work was supported by DHA Suffa University, Karachi, Pakistan.

ABSTRACT Social media microblogs are extensively used to get news and other information. It brings the real challenge to distinguish that what particular information is credible. Especially when user authenticity is hidden, due to the microblog's anonymity feature. Low credibility content creates an imbalance in society. Therefore many research studies are conducted to assess automatic microblog's credibility but the majority of them offer different concepts of credibility and the problem seems unresolved. Credibility is multidisciplinary, hence there is no generalized or accepted credibility concept with all its necessary and detailed constructs/components. Therefore, it is necessary to understand the complete anatomy of information credibility from different disciplines. It is accomplished here through an in-depth and organized study of all the problem dimensions for the identification of comprehensive and necessary credibility constructs. The framework is also proposed based on the identified constructs. It adheres to these constructs and presents their inter-relationships. It is believed that the framework would provide the necessary building blocks for implementing an effective automatic credibility assessment system. The framework is generic to social media and specifically implemented for microblogs. It is completely transformed up to features level, in the context of microblogs. Regarding automatic credibility assessment, it is proposed after detailed analysis that the attempt should be made for hybrid models combining feature-based and graph-based approaches. It is observed that quite a few surveys in the literature focus on some limited aspects of microblogs credibility but no literature survey and fundamental study exists that consolidates the work done. To understand the broader domain of credibility and consolidate the work in this area that can lead us to a suitable framework, we explored the existing literature from different disciplines for the said objectives. We categorized them along various dimensions, developed taxonomy, identified gaps and challenges, proposed a solution, developed a theory-driven framework with its transformation to microblogs, and suggested key areas of research.

INDEX TERMS Social media credibility, twitter information credibility, credibility features, automatic credibility assessment models, proposed solution, credibility framework, credibility taxonomy, credibility levels dimensions constructs, credibility studies, credibility dataset.

I. INTRODUCTION

Microblogs are intensively used to share news [1], opinions, observations, health issues, entertainment, experiences, and many more [2]. It is therefore becoming an imperative source of information but on the other hand, not-credible [3], [4] and cumbersome [5]. Taking an example of microblogs such as Twitter is steadily achieving gigantic consideration [6] as an important form of information media [7]. A large number of users throughout the world spread a wide range of information in real time [8]. Millions of Tweets are posted per hour on Twitter. Currently, it is the growing social medium and prevalent news media source as well [9]. Users massively share news headlines and also report real-time events of varying nature, well before official sources [8].

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

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

Twitter users are of many kinds, such as citizens, companies, governments, famous personalities, politicians, and many more, and such a wide range of users heavily depend on it for their business, political, social, and educational communications. Therefore on the dark side of this beautiful picture spammer also exploits the anonymity feature of microblogs to propagate their spam messages and scam URLs. It is quite vulnerable and turns into a medium of wrongdoers to spread rumors, fake news and other forms of misinformation [10]–[13]. Spread of hate speech [14], [15], political astroturf memes [16], extreme biases [17] are also found. Low credibility content creates an imbalance in society by damaging the reputation, public trust, freedom of expression, journalism, justice, truth, and democracy. Consequently, microblogs' users often need to judge the information's credibility. It becomes more challenging when source/user authenticity is hidden from the viewer, though user anonymity is one of the prose of microblogs. Unfortunately, it also welcome some other issues like: user's coordinated behavior [18], follower's fallacy [19], etc. It not only affects the quality of microblogs content but also introduces another challenge for gauging the source credibility.

There are many studies conducted at different aspects of credibility in many fields, such as; psychological factors affecting credibility, credibility types, dimensions, constructs, theoretical credibility frameworks, user's perceptions of credibility, suggested credibility features, automatic credibility assessment studies, and experimental studies of ranking information based on credibility, etc. Even then there is neither comprehensive nor accepted credibility attempt exist [20], [21], nor there is a standard definition of credibility found, though there are some related terms used to define credibility [22]. Considering the broader domain of credibility, having related terms or even having definitions only, never provides us that these are the necessary aspects that must be considered when credibility is assessed. Though it is required and extremely important in doing such assessments. In continuation with these challenges. It is also discovered that no literature survey and fundamental study exists that consolidates the work done from different fields. Therefore to fulfill the objectives. The literature is explored to identify such necessary credibility components. These identified components also lead us to propose a suitable framework of automatic credibility assessment.

Another very obvious fact to be highlighted to understand the importance and need of such broad and in-depth study; is about different types of malicious profiles or simply called malicious accounts. Which are completely ignored in all credibility studies. Though there are separate bot-detection studies found but not under the umbrella of credibility or not considered as a necessary aspect of credibility. Examples of such malicious profiles are; Bots, Trolls, Cyborgs, etc. All such forms of malicious profiles are usually believed to aggravate the wrong sense of credibility indicators and play a key role in the spread of low credibility contents [23]. It became very evident in investigations into Russian attempts to influence the 2016 US election [24]. It has also been observed that a massive amount of low credibility contents have already been shared over social media and microblogs before and after the US Election 2016 despite many efforts of credibility assessments [25]–[28]. It shows that some important and necessary aspects were ignored in available credibility assessment methods, as discussed earlier. Although credibility has been studied since ancient times, and in different research fields to date, such as psychology, media science, information science, communication, journalism, social sciences, and information retrieval, etc. [29]. It is noticed in literature that, due to being multi-perspective nature the diversity in the definition and perception of credibility reflects different viewpoints in different work studies. These studies only stick to just a single or only a few aspects of credibility. Some studies consider only Relevance as a criterion of being credible, some assume just Reputation as the major driver of Credibility, whereas the majority only stick that Fake and Rumor identification is credibility identification. It is also perceived by researchers, that Rankings concerning author Influence and Topic Expertise are strongly treated as credibility ranking. The majority of studies exploit just Informativeness as a credibility indicator. Few found examining Trust level as true credibility judgment. It is observed and quite evident in many research studies as well, that the credibility notion needs to be standardized because many studies only cover either one or some aspects of credibility (see figure 1) and a majority are left undiscovered. Some potential aspects are not even explored though much affect the credibility (see figure 2). Effective and comprehensive credibility concept may conforms some combined aspects presented in both figure 1 and figure 2. It means that low credibility contents may have a variety of forms presented in both figures. There is another strong observation developed through a majority of research studies, that credibility is assessed for news contents only (fake/real), though it equally exists in non-news contents as well, with a different set of aspects. Therefore those necessary set of credibility related aspects need to be identified which must be evaluated for any

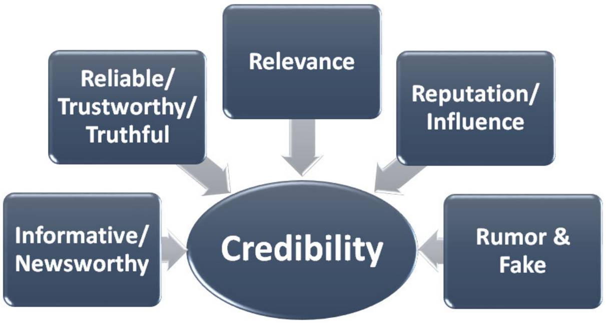

FIGURE 1. Majority of the studies only cover either one or only some of the above aspects of credibility and a majority of the aspects are left undiscovered.

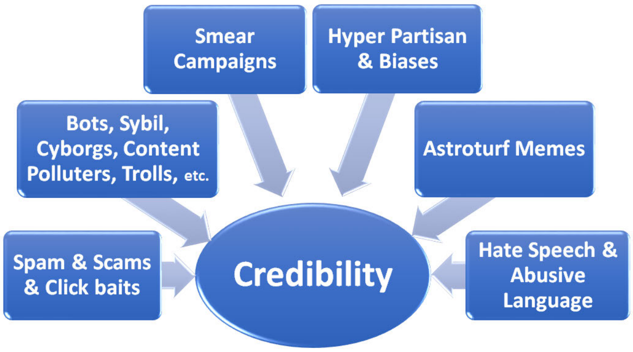
Smear Campaigns
Hyper Partisan & Biases
Bots, Sybil, Cyborgs, Content Polluters, Trolls, etc.
Astroturf Memes
Spam & Scams & Click baits
Credibility
Hate Speech & Abusive Language

FIGURE 2. Above are some general aspects of credibility which are completely missed in literature within the context of credibility. Low credibility contents may have the above forms, which should also be considered when credibility assessment is made.

piece of information in terms of its credibility assessment. It is already discussed that credibility is multi-disciplinary, hence there is no generalized or accepted credibility concept with all its necessary and detailed constructs/ components. It is extremely necessary and quite challenging, to understand the broad domain of information credibility to extract its complete anatomy from different disciplines. It could be accomplished through an in-depth and organized study of all the problem dimensions and identification of comprehensive and necessary credibility constructs under credibility's definition first. Further, the development of a concrete framework that adheres to those basic constructs/components could be possible. The framework will be theory-driven and provide a complete relationship/connection between different identified credibility components. In this study, we are concerned with the said identification followed by the development of a generic and comprehensive framework of information credibility. The framework will be generic to social media and specifically implemented for microblogs. It will be completely transformed up to features level, in the context of microblogs.

Nowadays numerous applications use a vast amount of microblogs data, such as; recommendation systems, event detection systems, social bookmarking systems, disaster response applications, campaign management systems, business monitoring applications, different types of prediction systems, and microblog search engines, etc. Each one of them only requires credible data to make these systems more effective [30]–[32]. Therefore dealing with information credibility problems in microblogs and social platforms, is necessary [33]. Once we would be able to develop an efficient and comprehensive credibility framework, which is missing and required, then there could be many applications in which the credibility framework would successfully contribute. For example; one of the most obvious applications could be the determination of the credibility of various posts during major global or local events. This can help for example in disaster response situations where the important information such as the extent of damage and need for action, can be figured out based on a large amount of microblogs posts and the trust ratings of their posters.

It is observed that quite a few surveys in the literature focus on some limited and individual aspects of microblogs credibility like health info. credibility [29], user influence/source credibility [34], trust in social networks [35], relevance-trust and influence [36]. There is a surface level or extremely short survey conducted over twitter information credibility in [37] and another general survey over information credibility of social media is done in [38]. As far as we discovered that there is no literature survey and fundamental study exists that consolidates the work on credibility similar to this study.

The remaining of the paper is organized as follows: problem formulation is done in section II. Credibility Taxonomy is developed as table:1 and figure: 3. The same is discussed from section 3-7, such as: in section III different definitions of credibility with its necessary and related components (levels, dimensions, and constructs, etc.) are presented. It helps us to understand credibility in the broader sense. Section IV highlights theoretical credibility frameworks. The most important section V presents many research areas which must be considered in credibility study and found extremely supportive, therefore named as supported research. Taxonomy's main section VI purely focuses only on all social media and microblogs specific information credibility studies. Last section VII of taxonomy is about standard credibility datasets. Section VIII literature-based important features are presented. Section IX summarizes the study through important findings and discussions. In section X we presented first, all theories in support of credibility framework identification and then our proposed theory-driven credibility framework is presented in section XI followed by section XII as Recommendations. Section XIII is about future research directions and section XIV concludes our study. Challenges and limitations are presented within different sections. Important terms used in the study are defined in appendix.

II. PROBLEM FORMULATION

To better understand the problem, in this section, we have formulated the credibility assessment as a classification problem and scoring/ranking problem. The mathematical problem formulation is done as following:

Let $P = \{p_1, p_2, \ldots, p_n\}$ be the set of n Posts, and $U = \{u_1, u_2, \ldots, u_m\}$ be the set of m Users on microblog. Each p_i consists of series of features including text domain, text sentiment score, text length, post spread score, no. of comments and replies, etc. Similarly each u_i consists of series of features like: influence score, name, domain, date creation, etc.

Classification Problem: Given Post P, and User U goal is to learn prediction function, such as $f(p_i, u_j) \rightarrow \{0,1\}$ satisfying:

$$f(p_i, u_j) = \begin{cases} 1 & \textit{if p is credible} \\ 0 & \textit{otherwise} \end{cases} \tag{1}$$

TABLE 1. Simplified credibility taxonomy: only top level and lowest levels are presented in this tabular form, intermediate levels are explicitly omitted for simplicity and better understanding. Detailed taxonomy with complete levels are shown in credibility taxonomy figure 3.

S. No.	Category	Sub-Category	Reference	Description
1	Credibility Definitions	Believability, Trust, Reliability, Accuracy, Fairness, Objectivity	[39]	How credibility is defined and its related components, e.g.: Levels, Dimensions, & Constructs,etc. and what is the relationship between credibility and trust.
		Quality of being Trusted and Believed	[40]	
		Quality of being Believed	[41]	
		Credibility has components: Message, Source and Media	[42]	
		Expertise and Trustworthiness	[43]–[46]	
		Believable Person and information	[47]	
	Credibility Components	Levels, Dimensions, and Constructs	see table 2, 3	
	Credibility VS Trust	Credibility is antecedent to Trust	[48]–[51]	
2	Theoretical Credibility Assessment Framework	Media-based Framework	[52]–[55]	These conceptual or theoretical frameworks provides: 1. Categorization similar to evolutionary generations. 2. Understanding of credibility assessment process & related concepts & how it is affected. 3. Underlying process involved behind people to perform assessment of credibility.
		Website-based Framework: Fogg's Prominence Interpretation Theory	[43]	
		Content-based Framework	[56], [57]	
		Interaction-based Framework: Rieh's Predictive and Evaluative Judgment	[58]	
		Interaction-based Framework: Wathen, Burkell- First Medium is rated, then source and message, third interaction of presentation and content	[59]	
		Interaction-based Framework: Sundar's MAIN model (Modality, Agency, Interactivity, and Navigability) four "affordances" in digital media	[60]	
		Interaction-based Framework: Elaboration Likelihood Model (ELM) of Persuasion	[61]	
		Interaction-based Framework: Heuristic Systematic Model (HSM) of information processing	[62]	
		Interaction-based Framework: Controlled and Automatic Processing Models (CAPM)	[63]	
		Interaction-based Framework: Social Information Processing Theory (SIPT)	[64], [65]	
		Interaction-based Framework: Dual processing model for Web	[66]	
		Unifying Framework: Provides basic levels: Interaction, Heuristics, and Construct	[67]	
		Unifying Framework: Rieh et al- Extension	[68]	
3	Supported Work	Misinformation/Disinformation: Rumor and Fake News Detection	[3], [69]–[76] and [25], [71], [77]–[84]	These are all studied as separate research areas in the literature, though each one of them are different construct/ aspect of credibility, therefore we consider them as important building blocks of credibility or necessary components of credibility framework & picture of credibility will be considered incomplete if not incorporated in the study.
		Political Astroturf Meme Detection	[16], [85], [86]	
		Spam and Phishing Detection	[87]–[90]	
		Topic specific Expert Identification	[91], [92]	
		Personality Specific Behavior Identification	[93]	
		Suspicious Behavior: Bot/Troll/Cyborg/Sybil/Content Polluter, Social Spambots, etc.	[80], [94]–[96] and [23], [26], [97]–[102]	
		Influence and Diffusion	[103]–[106]	
		Trust and Distrust Propagation	[107], [108]	
		Post Ranking	[109]–[112]	
		Hate Speech, Offensive and Abusive Language Detection	[113]–[115]	
		Hyper-partisan/Bias/Polarization Detection	[17], [101], [116], [117]	
4	Information Credibility of Social Media and Microblogs	User Perception	[37], [118]–[121]	Many organic surveys are conducted in which user perceptions or other elements have been studied, to explore all possible and important features of information credibility specifically with respect to the perception, judgement and heuristic of user.
		Explanatory Studies	[30], [119], [122]	Wide range of features are studied, and many explanatory studies are conducted regarding broad feature analysis. To conclude what serves best for credibility assessment data is collected from microbloging sites and tagged either by means of crowed sourcing environments or experts.

TABLE 1. *(Continued.)* Simplified credibility taxonomy: only top level and lowest levels are presented in this tabular form, intermediate levels are explicitly omitted for simplicity and better understanding. Detailed taxonomy with complete levels are shown in credibility taxonomy figure 3.

S. No.	Category	Sub-Category	Reference	Description
		Source Credibility	[19], [34], [123]–[129]	Researches where information credibility assessment is done through greater focus towards source /user of information
		Feature Based Models	[21], [113], [130]–[137]	ML/IR based models are used which use features commonly related to Topic, Posts, Authors, and Network, etc. Either atomic level of information is used, means contents contained within the tweet or Varying level of information with aggregated and historic features, to assess the Information Credibility
		Graph Based Models	[127]	Uses SNA/ Graph based models by utilizing friends-followers network, user-tweet-retweet and retweet networks, etc.
		Hybrid Models	[31], [138], [139]	Some combination of Feature based and Graph/SNA based methods used
5	Standard Credibility Dataset	Credibility benchmarks are not predefined therefore its related gold standard dataset is missing. The difficulty of collecting large amount of such data has not yet received the attention it deserves [29].		

Scoring/Ranking Problem: This could also be ranking/ scoring function, such as: f(p_i, u_j)→{0,1,2,3,4,5} satisfying:

$$f(p_{\rm i}, u_{\rm j}) = \begin{cases} 0 & if\ p\ is\ not-credible \\ 1 & if\ p\ is\ low-credible \\ . & . \\ . & . \\ 5 & if\ p\ is\ highly-credible \end{cases} \tag{2}$$

A. INFORMATION CREDIBILITY TAXONOMY

In the following sections, from section III to section VII, complete information credibility is presented. The taxonomy is also drawn in figure 3. In this hierarchy the first branch named 'Credibility and its Components' presents different types of credibility, credibility dimensions,, credibility constructs, credibility definitions, etc. Second branch named 'Theoretical Frameworks of Credibility Assessment', which actually presents evolution of Credibility, till date. In the field of communication and psychology such concepts are best presented as frameworks. In third branch named 'Supported Research' where different aspects of credibility i.e.: Deception, Hate Speech, and Influence Identification, etc are presented. Fourth branch named 'Information Credibility of Social Media and Microblogs', presents types of information credibility experiments, related to social media and microblogs only. The last branch named 'Standard Credibility Dataset' presents details about available datasets.

III. CREDIBILITY AND ITS COMPONENTS

As an important objective with many challenges, this section not only presents credibility definitions (as related terms) but also extends them systematically and forms the basis of the credibility framework's building blocks (e.g.: levels, dimensions, constructs) through related research studies from different fields. Different credibility components are comprehensively explored and presented.

A. CREDIBILITY DEFINITIONS

Many efforts have been made to define Credibility. It is a complex and multi-dimensional concept. There is no clear definition, it has been defined through several related concepts [22]. Therefore such definitions are taken from both, strong research studies and standard dictionaries:

It is defined as: "believability, trust, reliability, accuracy, fairness, objectivity, and other concepts and combination" [39], Oxford dictionary defines credibility as "the quality of being trusted and believed in" [40], as Merriam Webster dictionaries it is defined as "the quality of being believed" [41]. Many researcher's core references of studies in communication examining credibility as message credibility, source credibility, and media credibility [42]. The majority of researchers are agreed that there are two attributes of credibility: expertise and trustworthiness [22], [43]–[46]. Similarly, across multiple definitions credibility is believability. Credible information means believable information similarly credible persons are believable persons [47].

After going through the above formal definitions we can divide credibility into two main components: message and source. Where the source is further examined through trustworthiness and expertise. This forms the basis of credibility framework.

1) CREDIBILITY COMPONENTS

After an in-depth exploration of research studies conducted in psychology, communication and information science, and to understand the broad domain of credibility, the following major credibility-related components (e.g.: levels, dimensions, and constructs) are found. They all are comprehensively discussed in following sub-sections and summarized in table 2 and 3 as well. These components are in varying sizes/levels of hierarchy. The top most (levels of credibility) is defined first and the lowest most (constructs) is defined last. The order is also maintained in table columns. The outcome of the credibility components section would be resulted in section X and to some extent, section XI. The following components are explored from various studies to propose a generic credibility framework for social media. The framework simply exposes the relationships found in these components. In the last portion of section XI where generic social media framework is further transformed for microblogs, using microblog specific features is not concerned as an outcome of this section.

B. LEVELS OF CREDIBILITY

There are different levels of credibility assessed in literature, which should be known for a better understanding of the

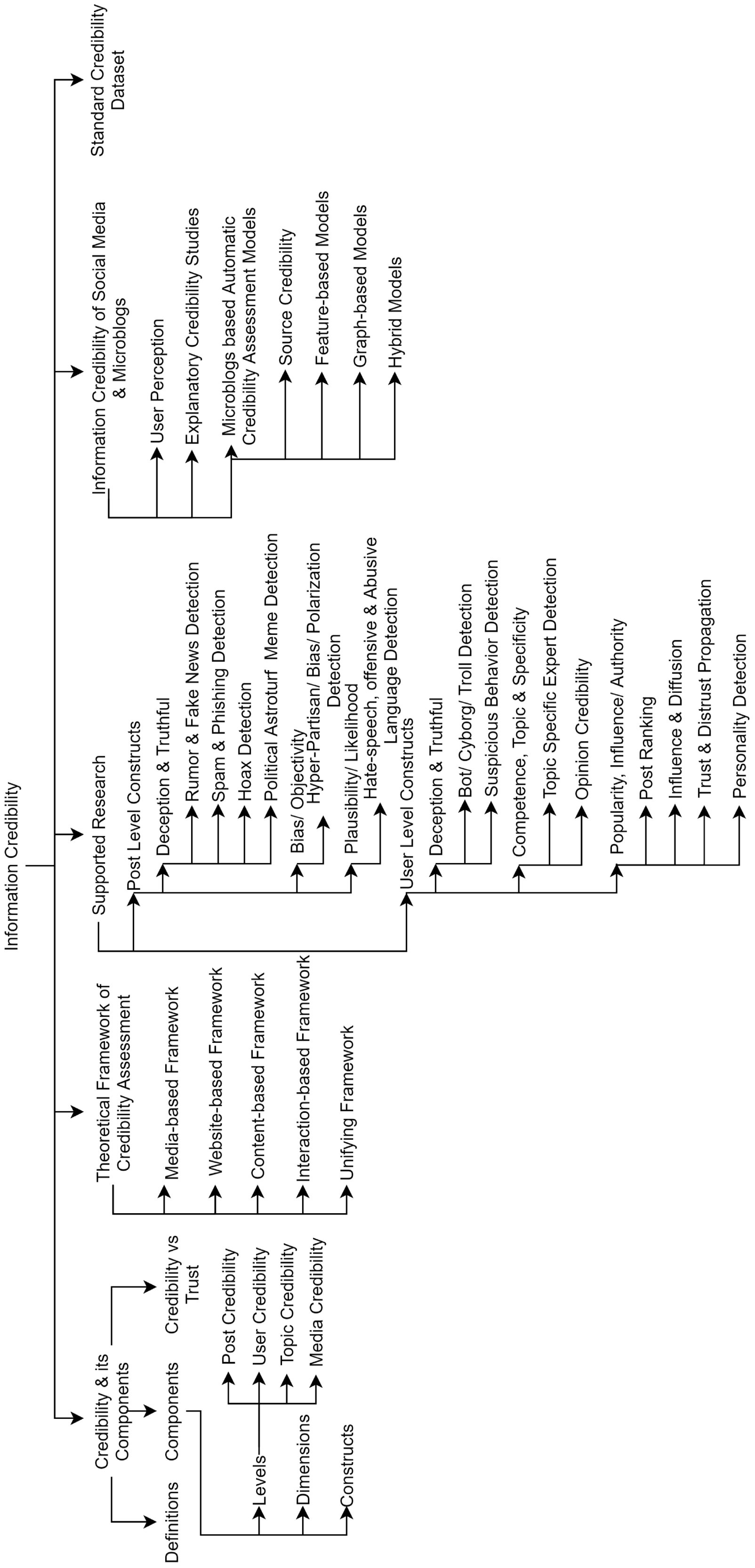
Information Credibility
Credibility & its Components
Definitions
Components
Credibility vs Trust
Levels
Dimensions
Constructs
Post Credibility
User Credibility
Topic Credibility
Media Credibility
Theoretical Framework of Credibility Assessment
Media-based Framework
Website-based Framework
Content-based Framework
Interaction-based Framework
Unifying Framework
Supported Research
Post Level Constructs
Deception & Truthful
Rumor & Fake News Detection
Spam & Phishing Detection
Hoax Detection
Political Astroturf Meme Detection
Bias/ Objectivity
Hyper-Partisan/ Bias/ Polarization Detection
Plausibility/ Likelihood
Hate-speech, offensive & Abusive Language Detection
User Level Constructs
Deception & Truthful
Bot/ Cyborg/ Troll Detection
Suspicious Behavior Detection
Competence, Topic & Specificity
Topic Specific Expert Detection
Opinion Credibility
Popularity, Influence/ Authority
Post Ranking
Influence & Diffusion
Trust & Distrust Propagation
Personality Detection
Information Credibility of Social Media & Microblogs
User Perception
Explanatory Credibility Studies
Microblogs based Automatic Credibility Assessment Models
Source Credibility
Feature-based Models
Graph-based Models
Hybrid Models
Standard Credibility Dataset

FIGURE 3. Detailed credibility taxonomy: the organized and complete taxonomy with all its levels is presented in this figure.

TABLE 2. Credibility components identified from research studies: different research studies related to credibility levels, dimensions, and constructs (table 1 of 2).

Ref	Levels	Dimensions	Constructs	Description
[140]	Source, Message	Quality, Trustworthiness	Source: Competence/ Expertise, Proximity/ Location, Popularity. Message: Recency, Corroboration/Agreement	Trustworthiness metrics proposed through survey research.
[141]	Topic, Source, Message	NA	Source: Authority/ Influence, Expertise, Popularity. Contents: Info. Quality, Popularity	Exploratory credibility feature analysis conducted on Twitter data, tagged by crowd-sourcing and experts
[142]	Source, Message	NA	Source: Expertise, Community. Message: Clarity, Emotions/Valance, Consensus (Consistency, User Judgment)	Social media based credible marketing related electronic word of mouth (eWOM) framework is proposed based on research theories.
[143]	Topic, Source, Message	Information Quality, Expertise, Trustworthiness	Survey covering many constructs used in studies.	Complete literature survey presenting different Levels, Dimensions, and Constructs of credibility.
[144]	Media Credibility	NA	1. Trustworthiness, 2. Un-Biased, 3. Accuracy, 4. Completeness, 5. Fairness	Defining and measuring media credibility.
[145]			6. Balanced (added)	Effects of balanced and imbalanced conflict story structure on perceived story bias and news media credibility explored through experimental study.
[146]			7. Factual (added)	Many constructs are measured through experimental study.
[53]			8. Expertise (added) 9. Social Concerns (added)	Literature review of credibility in the contemporary media environment.
[147]			Only 1-4	Survey on media credibility of newspapers accounts on Sina Weibo.
[123]	Source Credibility	Expertise, Trustworthiness	NA	Seminal work on source credibility: Survey & Controlled Group Study.
[148]		Goodwill/Caring (added)		First suggested perceived caring/goodwill as source credibility aspect.
[149]				Aspect of 'caring' fully studied in survey.
[150]				Reexamination of the construct and its measurement done and Goodwill added through survey study.
[151]				Endorsing through theories
[152]	General Credibility	Expertise, Trustworthiness	NA	Seminal work in Attitudes & Comm., reporting series of experiments on credibility.
[153]	Source, Contents	Quality, Expertise, Trustworthiness, Reliability/ Relevance/ Consistent	NA	Literature based, proposed contents/IR Credibility Framework

subject area. Levels of credibility are treated at the highest level of the component's hierarchy or they are a macro-level component. They are classified as following and also summarized in table 2 and 3:

1) POST CREDIBILITY

It is the most important and primitive form. It means the message or post itself is credible [135], [156]. It may effects the credibility of the user or event, etc. It is the most suitable for online/ real-time credibility identification systems because no historic data is needed. On the dark side, it poses a weak credibility assessment based on a limited scope.

2) USER/SOURCE CREDIBILITY

It corresponds to the poster (e.g.:speaker, organization, govt., news organization, etc.) or user of the post [125], [127]. In most studies, it is presumed that if the source is credible then the message associated with the user is also credible [34], [123], [129]. Somehow it is treated as the higher level, which means user credibility may be based on the user's post collections [34]. Which makes it a historic/ offline assessment system, because we need all historic data for evaluation. Online/ real-time or immediate assessment is not possible. Hence combined post and user information presents better credibility identification.

Social/Domain Expert Credibility: In [157] a variant or subset of source/user credibility is identified. It is based on the social status of a user in a social network on a certain domain. A similar concept is also used for Opinion Credibility [158]. Source credibility is known to be a super-set of such subsets. Source credibility could be measured in terms of a broad set of credibility aspects like influence, popularity, truthfulness, expertise, biasness, etc. whereas such subsets are measured on just a single aspect e.g.: expertise.

3) TOPIC/EVENT CREDIBILITY

Event comprises all related posts to a specific event/topic. Whereas topic/event could be identified by a set of

TABLE 3. Credibility components identified from research studies: different research studies related to credibility levels, dimensions, and constructs (table 2 of 2).

Ref	Levels	Dimensions	Constructs	Description
[67]	Credibility Constructs (Media, Source, Content)	NA	1. Believable/ Plausibility, 2. Truthful, 3. Trustworthy 4. Objectivity/Un-Biased 5. Reliability/ Accuracy/ Relevance/ Consistent	Unifying framework defined constructs
[68]			Found Best:(2-5 above) & 6. Recency/Timeliness, Found Good (for other Information Objects): 7. Completeness 8. Official, 9. Un-Biased, 10. Authority/Influence, 11. Expertise, 12. Scholarly/ Reference/ Educational Endorsement	Extension to Unifying framework to make it global
[154]	Content (Content Trustworthiness)	NA	1. Topic 2. Context and criticality 3. Popularity 4. Authority/Influence 5. Experience/ Reputation 6. Recommendation 7. Related Resources 8. Provenance/ Source 9. User expertise 10. Bias 11. Incentive 12. Limited resources 13. Agreement/ Corroboration 14. Specificity 15. Likelihood/ Believable/ Plausibility 16. Age/ Timeliness/ Validity 17. Appearance 18. Deception 19. Recency/Recent Image	Comprehensive study describing content trustworthiness: means how end-users make decisions regarding trusting information. Exhaustive literature review and simulation study supported.
[155]	Source, Message	Expertise: (Source, Content), Trustworthiness	Expertise: Quality, Accuracy, Authority, Competence Trustworthiness: Reputation, Reliability, Trust	Study from communication domain enlightening emergent and Modern concepts related to credibility.

keywords [31], [133], [159], [160]. The specific event comprises a collection of posts and associated posters as well. An example of such topic/event credibility is the Credibility of posts during COVID-19.

4) MEDIA CREDIBILITY

It is also multidimensional (high level) construct. Comprised of source credibility and medium credibility. Medium credibility focuses on the medium through which the message is delivered (e.g.: newspaper, radio, television, etc.- In the context of our study it is just an underlying social network used for information propagation) [161].

In our case of microblog, the microblog's credibility is Media Credibility which is based on the poster and underlying social network used for information propagation (as the medium). A very important and distinct notion presented in [59] that in modern scenario medium is also replaced with source only. Therefore only source (including all chain of message propagators) credibility could easily be used in place of media credibility.

The above types are somehow synchronized with each other. Therefore media credibility assessment system will require examination of the post, source, and underlying information propagation social network, to claim its microblog credibility system. Therefore for our proposed credibility framework only post-level and source-level credibility would be enough.

C. CREDIBILITY DIMENSIONS AND CONSTRUCTS

It is quite challenging to define credibility in terms of its necessary components/elements, because there is no standardization due to its multidisciplinary [162] and emerging [155] nature. In the field of psychology and communication, the orientation of credibility is source-based and therefore called source credibility whereas in information science it is message oriented and called information credibility [162].

Dimensions are considered at middle and constructs are at the lowest level of credibility components hierarchy.

1) CREDIBILITY DIMENSIONS

Despite all above challenges it is observed through literature exploration that the majority of researchers accepts that there are at least two major dimensions (dimensions are also called topics, factors, etc. in literature) of credibility: Expertise and Trustworthiness [123], [152], [155], other many studies endorse with minor addition [148]–[151]. Another important Dimension named: Information/ Data/ Content Quality is also found in [143], [153], [154], [163].

It could be concluded that the most agreed upon dimensions are Expertise, Trustworthiness, and Quality of Information. These could be the necessary dimensions of the proposed framework.

2) CREDIBILITY CONSTRUCTS

Under the above dimensions, there are some constructs (constructs are also called sub-topics, sub-factors, etc. in literature) proposed in different credibility studies. The list of constructs could be different concerning information object or media, etc. A very detailed survey discussing factors/sub-factors (topics/sub-topics) studied in variety of research studies [143]. Some basic credibility constructs are proposed in the most popular and highly concerned 'unifying framework' [67] (will be discussed in next section) which were extended concerning the varying type of information object (e.g.: Social Networks/ Media, Microblogs, Web Blogs, Search Engines, General Websites, Electronic Commerce Sites, News Sites, Educational Portals, etc.) or media contents (TV, radio, podcast, music, photo, video, etc.) in [68]. Detailed constructs specific to Data/ Content Quality are presented in [154], [163]. Constructs to assess Media Credibility are proposed in [53], [144]–[146].

Regarding our proposed credibility framework which will be generic to social media but specific to microblogs. The levels and dimensions would be generic to social media only. Constructs must be compatible with both social media and microblogs and then further lower-level components (e.g.: features) must be microblogs specific or information object-specific only. Keeping the specific attributes of social media and microblogs both, the following few constructs could be shortlisted from table 2 and 3 in addition to the following two criteria. 1. These constructs are common to both post and source levels, and 2. They are also common to trustworthiness, expertise, and information quality dimensions. These constructs are; 1. Recency, 2. Truthful, 3. Deception, 4. Topic, 5. Specificity, 6. Unbiased/Objectivity, 7. Popularity, 8. Plausibility, 9. Authority/Influence, 10. Competence/Reputation, 11. Uniqueness/Completeness, etc.

Complementing the above recommended key Levels, Dimensions, and Constructs, some frameworks (comprised of levels, dimensions, and constructs) are developed and experimental studies are conducted to adhere to the findings discussed. For example, the electronic word of mouth (eWOM) framework for marketing related to social networks credibility is presented in [142]. The credibility framework for Information Retrieval systems is presented in [153].

In addition to the above frameworks and basic component related studies there are few exploratory studies conducted which also support and confirm the identified components. An exploratory study for credibility feature analysis conducted on Twitter data, tagged by crowd-sourcing and experts [141] (see table 2 and 3, for these frameworks).

Summarized Levels, Dimensions and Constructs are presented in table 2 and 3. There are numerous studies found in psychology, communication and Information science on credibility-related components e.g.: levels, dimensions, and constructs; but only some representative studies are presented in the table for understanding and support.

D. RELATIONSHIP OF CREDIBILITY AND TRUST

The concept of credibility and trust must be clarified and their relationship should be presented. Credibility and trust are mistakenly used interchangeably. Credibility is believability while Trust is dependability. Credibility is an antecedent to trust [48]–[51].

IV. THEORETICAL FRAMEWORK OF CREDIBILITY ASSESSMENT

For the past many years, there have been so many research studies on credibility. All mostly in the field of information science, psychology, and communication. However, to better understand people's credibility assessment within various information contexts, modern credibility research has started to take a multidisciplinary approach [162] and becoming emergent [155]. In various research communities, different conceptual and theoretical frameworks have emerged regarding the conceptions of credibility, due to increasing concerns about the credibility of online information. There are the following distinct conceptual or theoretical frameworks categorized and described in order (similar to evolutionary generations), for examining the credibility of online information. They provide an understanding of the credibility assessment process and related concepts and how it is affected in general or discuss the underlying process involved behind people to assess credibility. One can easily understand that how these frameworks are evolved concerning the modern requirements and challenges:

A. MEDIA-BASED FRAMEWORK

It is the earliest framework, developed within the field of communication. Researchers within this framework have long been interested, since the 1950s, to know the relative credibility [52] of different media channels (e.g.: Radio, TV, Magazine, Newspapers, and now Web is also included). Communication scholars investigated various factors affecting media credibility [53] including people's perception of Web-based information, and Web vs traditional media [54], [55].

The major limitation of this framework was that it considers people's general perception regarding medium instead of focusing on what use of information, which is obtained from it. For example, if someone considers the Web as the bad medium in terms of credibility doesn't mean that every website will be considered poor in credibility.

B. WEBSITE-BASED FRAMEWORK

In this framework complete website is examined for credibility. In Stanford Web Credibility project [164] various elements of the website are examined which affects user's credibility assessments. After many studies Fogg's: Prominence Interpretation Theory is developed; which talks about the following, that needs to occur for people to assess web credibility: Prominence (likelihood of an element noticed) and Interpretation (value assigned to that element based on user's judgment). Factors affecting prominence as well as interpretation are also discussed [43]. There are few other studies [165], [166] found on website credibility under the website-based framework, all have the common strength that it covers both contents with peripheral cues (e.g.:appearance, design, presentation, etc.) as components of credibility. But on the other hand side, there is a weakness that every piece of information contained in the website is not separately considered.

C. CONTENT-BASED FRAMEWORK

Website contains many information objects therefore each information object is individually assessed in this framework. This framework assumes that information credibility may vary even within the same website. The main focus of the framework is: When we access any piece of information we emphasize assessing its quality. Therefore the chief aspect of information quality is defined as credibility [56]. It is reported in [57] that social-Q&A type of sites, users evaluate credibility primarily on contents because of having limited cues to source credibility.

The weakness of the framework includes missing the emotional effects of interaction with information and aesthetic aspects of the information object.

D. INTERACTION-BASED FRAMEWORK

This framework assumes that instead of discrete evaluative event credibility assessment is best expressed through an interactive and iterative process. It also guides that assessment of credibility could easily be chalked out through observation during user's information seeking process with their selections made for searching that information.

The interaction framework also emphasizes the fact that credibility assessment is subjective means highly depends on the user's current knowledge and experience. Limitation to this framework seems that most of the studies only focus on the human information searching and navigating process.

Rieh's model explains that when a user starts the information-seeking process, it begins earlier from predictive judgment, which leads the user to access information resources and then go towards evaluative judgment [58]. Hilligoss and Rieh added the third type of judgment as Verification [167], later through their empirical study.

Wathen and Burkell define an interactive and stage process where the first Website's surface-level characteristics (content organization, interactivity, interface design, speed, appearance, etc.)/medium credibility is rated, then the user rates the source and message (trustworthiness, competence, expertise, etc.) and the third aspect is the interaction of presentation and content [59] which is finally assessed as per user's cognitive states.

Sundar's credibility assessment also adheres interaction framework and presents the MAIN model (Modality, Agency, Interactivity, and Navigability) having four technical ''affordances'' in digital media [60]. Affordances can increase or decrease content effects on credibility, like moderators; in several psychological ways. It is therefore recommended by Sundar, that role of heuristics in credibility assessment should be explored. To understand the role of the heuristic in understanding credibility assessment is presented in Elaboration Likelihood Model (ELM) of persuasion and Heuristic Systematic Model (HSM) of information processing. Both models share many of the same concepts. Therefore Dual Processing model of information processing and credibility evaluation [66] has taken motivation into account like dual-process theories [168] and also based on both.

ELM of persuasion [61] is dual-process theory and the general theory of attitude change (e.g.: What attitudinal changes in user will occur when user come across messages and sources). It provides a general framework for understanding the basic processes underlying the effectiveness of persuasive communications.

Similarly, HSM of information processing [62] is a popular communication model which explains how people receive and process persuasive messages. Similar to all dual-process theories: ELM, Controlled and Automatic Processing Models (CAPM) [63], it is also defined in this model that individual can process messages in either ways, systematically or heuristically.

Another widely used interpersonal communication and media studies theory named Social Information Processing Theory (SIPT) [64], [65] which explains online interpersonal communication and how people develop and manage relationships in a computer-mediated environment. It says that the community exploits any piece of information that the channel provides them to make assessments about others.

Among dual-process theories (ELM, HSM, CAPM) and SIPT, there are few other fairly general theories and frameworks that are often adopted by credibility researchers to characterize the credibility assessment process and its constructs and components.

E. UNIFYING FRAMEWORK

Finally most important unifying framework of credibility assessment is proposed for a different type of media, information objects, and contents for a variety of information activities. It provides very basic levels of credibility judgments: Interaction (credibility judgments in which sources or information examined), Heuristics (general rule of thumb, could be applied to a wide range of situations), Construct (how credibility conceptualized) as basic levels and an additionally defined Context (surrounding the user) of credibility assessment [67]. Later the framework was fully extended by Rieh *et al.* [68] to cater to the need of current and modern participatory web environment (include Web 2.0 means all kinds of modern social media services and others). It could be concluded that Unifying Framework is the most relevant and therefore should be followed to fulfill the modern requirements. The proposed framework is also enriched with the constructs presented in Unifying Framework.

V. SUPPORTED RESEARCH

Many of the supported or closely related and somehow different dimensions of microblogs-based information credibility, have already been studied separately. Unfortunately, they are not considered as directly related to credibility in the literature, but all of them are comprising different constructs/aspects of credibility and therefore need to be augmented, holistically. The mapping of all supported research studies with appropriate constructs is done in this section. All these constructs/aspects are also shown in the proposed high-level credibility framework's table:14 and then these aspects are mapped to individual features in table:15, where all these studies are highly contributing. We consider these supported research studies as important building blocks of credibility or necessary components of the credibility framework. Picture of credibility will be considered incomplete if they are not incorporated in the study. Each one of them is considered a completely separate research area therefore details are omitted but only the research area name together with important references are mentioned. Important terms are defined in the Appendix for basic understanding and clarity. What we have done for simplicity and increased productivity that we go through all supported research studies and list down all important features. These features are then proposed for implementing microblogs specific credibility framework. They are presented in the table: 15 which provides the implementation of our generic framework to microblog specific framework. All these features are added with their supported references and reason in table 15 of our proposed credibility framework section XI.

In this section, to support the understanding of credibility components, each area of research (which is named as supported research in the study) is categorized with respect to its respective level and appropriate construct. For example, 'fake news detection' is an area of research which is classified under construct, named 'Deception and Truthful', presented within the level, called 'post level'. The area of research will not be discussed, only the name of the area with respective references will be included. There are few constructs shortlisted in section III-C2 related to social media and microblogs-based information credibility. Examples of those few constructs are; 1. Recency, 2. Truthful, 3. Deception, 4. Topic, 5. Specificity, 6. Unbiased/Objectivity, 7. Popularity, 8. Plausibility, 9. Authority/Influence,
10. Competence/Reputation, 11. Uniqueness/Completeness, etc. Different areas of research considered related to this section are categorized under these relevant constructs. Those areas of research under each construct's heading are as follows. The constructs are also grouped under respective levels like post level and user level.

Post Level Constructs: It is discussed earlier that post is the most basic and lowest level in all other levels of credibility. Though we have considered only two levels, post, and user. Aggregation of many post-level constructs will automatically result in user-level constructs, e.g.: if the majority of posts are biased then the user will automatically be biased. The same will be the case of fake posts. It means that few constructs will be common in both levels. Those common constructs are Deception, Truthful, Unbiased, and Popularity, etc. Despite that few constructs are common, only those constructs are repeated where detection mechanism found different at both post and user levels, e.g.: Deception and Truthful. The techniques detecting deception at post level are discussed as fake news detection, rumor detection, etc. but techniques detecting deception at the user level are called bot-detection, suspicious behavior detection, etc.

A. DECEPTION AND TRUTHFUL

Detection of all deceptive and untruthful contents must be done at each post level. This section includes all such studies which provide the understanding and also suggest ways and means of their detection.

It is discussed in [169]–[171] that false Information [82], [172] or deceptive information [33] has variety of flavors: Fake/ False News, Misinformation, Disinformation, Hoaxes, Propaganda, Satire, Rumors, Click-Bait, and Junk News, etc. Though an agreed and standardized definition is completely missing but is generally considered that misinformation is information that is inaccurate and misleading which could spread unintentionally in contrast to disinformation which is false information and spread deliberately to deceive people.

False Information Detection: Following are studies related to deception and false information detection including their different forms. Only name of the field/area and related references will be provided.

Misinformation/ Disinformation and its detection: [20], [170], [173]–[176], Rumor and its detection: [3], [69]–[76], Fake News and its detection: [25], [71], [77]–[84], Stance Detection is basically identification of the relevance of news article's contents with title. Its now assumed as sub-category of fake news detection, such that for fake news identification first stance is evaluated: [177], Hoax Detection: [178], Spam and Phishing Detection: [87]–[90] spam and phishing detection techniques can also automatically filter click-bait, fake reviews, and some political astroturfs. Because they are similar in structural or strategical patterns and may called modern form of spams.

Damage of Reputation Detection: There are some types of deceptive and false information that damage one's reputation and naturally affect one's credibility, they are called smear campaigns which may include: satire, conspiracy, propaganda [179], political astroturf memes, etc. There are different Political Astroturf Meme Detection studies also found: [16], [85], [86].

B. BIAS/OBJECTIVITY

It is found that some post may have a piece of such information which come from a particular point of view and may rely on propaganda, decontextualized information, and opinions distorted as facts. These posts are categorized as extremely biased. They must be identified or detected in the early stages of their spread otherwise have associated grave repercussions. They create highly polarized groups, in terms of religion, politics, race, etc. Therefore following are few example studies which can identify 'bias/objectivity' construct of credibility, they are: Hyper-partisan/ Bias/ Polarization Detection: [17], [101], [116], [117].

C. PLAUSIBILITY/LIKELIHOOD

Freedom of expression is a human right but hate speech towards a person or group based on race, caste, religion, ethnic or national origin, sex, disability, gender identity, etc. is an abuse of this sovereignty. Hate speech is essentially a discourse that might be extremely harmful to the feelings of a person or group and may contribute towards brutality or insensitivity which shows irrational and inhuman behavior. It seriously promotes violence or hate crimes and creates an imbalance in society by damaging peace, emotions, reputation, trust, credibility, human rights, justice, and democracy, etc. In addition, to hate speech some other related concepts must also be considered like Hate, Cyberbullying, Discrimination, Flaming, Harassment, Abusive Language, Profanity, Toxic Language or comment, Extremism, Radicalization, etc [14]. These all are some general information quality-related constructs that must be considered for detection. Following are few example studies which can fulfill the requirements, such as; Hate Speech, Offensive and Abusive Language Detection: [113]–[115].

User Level Constructs: User level is higher than post level. Many user-level constructs could be accumulated through their respective post-level constructs. Therefore they are omitted from this section. Considering the case of fake posts, if the majority of posts posted by a user are fake then that user will not be trustworthy. In the following user-level constructs, only those constructs are presented where detection mechanism is found different concerning the user. Following are all supported research studies categorized under user-level constructs. Only the name of the field/area and related references will be provided, details of the field are not included:

D. DECEPTION AND TRUTHFUL

It is worth mentioning that majority of the incredible contents are spread through different types of Bots, Trolls, Cyborgs, Sybils, Content Polluters, or Social Spambots, etc. There are almost 15-17% accounts which are bots [23], presenting human impersonation and perform many malicious and suspicious activities, e.g.: Spread of misinformation and fake news, fake support, fake product reviews, advertise for doubtful legality, hashtag, and other promotions, spread unsolicited spam, scam URLs, terrorist propaganda, manipulate the stock market, rumor dissemination and support, conspiracy, astroturf political campaigns, and religious activism, bias public opinion, sponsor public character and many similar activities [20], [26], [94]. Therefore once they are identified and blocked then all such contents will automatically be filtered and the remaining large portion of contents will be treated as legitimate and credible.

There are studies found for such malicious profiles identification and detection, for example Bot/ Trolls/ Cyborg/ Sybils/ Content Polluters/ Social Spambots and its detection: [23], [26], [97]–[102] and Suspicious Behavior Detection: [80], [94]–[96].

E. COMPETENCE, TOPIC AND SPECIFICITY

Following are a few examples of supported research studies that could help in the determination of the above group of constructs. Dealing with the social status of a user in microblog's social network on the certain domain such as politics, education, sports, science and technology, social issues, etc. It is simply called Topic Specific Expert identification, here the competence within a specific domain or topic is concerned, which could be done with the help of these studies: [91], [92], [157]. A very similar concept is used as Opinion Credibility in [158]

F. POPULARITY, INFLUENCE/AUTHORITY

Every user in a microblog's social network has certain influence/ authority/ popularity. Highly influential/ authoritative/ popular users can affect an individual's attitudes, beliefs, and subsequent actions or behaviors. We need to identify an appropriate way to measure user influence/ popularity/ authority score. Most authoritative/ influential/ popular users are assumed more credible. There could be different ways of measuring such scores. Making use of the follower-following network or user-tweet/retweet network, and then apply modified page rank like model or some form of authority transfer or some centrality measure for calculating highly influential/ popular/ authoritative/ reputed user. It could be measured by applying some ratios of followers count, followings count, with some form of popularity measures e.g.: no of times a user is mentioned, retweeted, replied, listed, favorited, etc. by other users of microblog's social network.

The above methods are commonly considered in computing source credibility. Some good variants can enrich these methods with a quite different perspective. Using the following concepts will provide required value addition, such as Post Ranking, which is done concerning relevance of user and content, as well as source popularity: [109]–[112]. Influence and Diffusion Methods: [103]–[106]. Trust and Distrust Propagation: [35], [107], [108]. Personality specific behavior [93] identification, which greatly helps in detecting different behaviors. Personality Detection, which provides big-five personality traits (i.e.: 1. Open/Closed, 2. Spontaneous/ Conscientious, 3. Introvert/ Extrovert, 4. Hostile/ Agreeable, 5. Stable/ Neurotic) that help predicting behavior and influencing ability. [180], [181].

VI. INFORMATION CREDIBILITY OF SOCIAL MEDIA & MICROBLOGS

The outcome of our study is two-fold. Understanding the broader domain of credibility with basic components identification and then the development of compatible social media generic framework. This will further be transformed to microblog-specific implementation. Considering the first objective: credibility related various generic studies from different fields have already been explored in former sections (III-B, III-C and IV). Presenting frameworks, models/theories (see section IV), and its macro components (see sections III-B, III-C, e.g.: levels, dimensions and general constructs) for broad range of information objects (e.g.: General Websites, News Media Sites, Search Engines, etc.). Moving forward towards the second objective: it is needed to exhaust only information object-specific studies. Characteristics of our information object (social media and microblogs) are quite distinct from other information objects, such as the authenticity of the source is hidden from the user. Contents are massively shared. User engagements and responses are shown. Content has a long propagation path that is hidden from the user. User-generated content, which is noisy. Having spelling mistakes, free from grammar,

small in size, have little context, contain language variations, furnished with special meaning in form of emoticons, hashtags, user mentions, re-tweets, and capitalization, etc. Therefore we have only considered social media and microblogs specific studies in this section. Considering any other type of information object (e.g: General Websites, News Media Sites, Search Engines, etc.) related credibility studies will not be productive for this section. Credibility constructs are somehow information object-related and need transformation [68] which is done in many studies, like [182]. It is also done in our proposed credibility framework presented in section XI, where constructs are only social media-specific and then corresponding features are microblogs
specific.

There are some domain-specific studies of information credibility found, like: Health [29], Disaster [183], Fake Review/ Opinion [184], Image/ Media [185], Geographic Information [186], Language Specific [187], [188], Country-Specific Perceptions [121], etc. All such studies are not considered much relevant because the challenge here is to understand correctly what is information credibility concerning social media in general first and then how will it be achieved for microblogs. Once this general understanding of information credibility will be developed then these very specific studies will be effective. The outcome of this section has resulted in the last segment of section XI as well, where the generic credibility framework of social media is further transformed to microblogs. Studies of this section guide that what are the set of microblog's features which are recommended for specific aspect (e.g.: Hate, Bot, Fake, Influence, etc.) evaluation.

Studies conducted specifically on information credibility related to social media and microblogs can easily be classified as studies that present User Perceptions of Information Credibility, Explanatory Studies, Source Credibility, Feature-Based Models, Graph-Based Models, and Hybrid Models of Information Credibility.

A. USER PERCEPTION

This section presents extremely important, Social Media and Microblogs Credibility specific variety of hypothesis. These hypothesis are related to human cognitive heuristics, judgments, perceptions, and assessments. They are identified and examined through different methods like surveys, interviews, empirical & experimental studies, observations, and statistical methods. All such studies are very comprehensively presented under different columns in table 4 and 5. In each study of the table, a very organic survey is conducted. In these surveys user perceptions, judgments, assessments, and heuristics have been studied, to explore possible and important features of information credibility for social media and microblogs. Researchers use these recommended features as a starting point and conduct explanatory studies to conclude what serves best for credibility assessment.

B. EXPLANATORY CREDIBILITY STUDIES

There is no accepted credibility standard [20], [21] and it is very difficult to judge different researches and generalize the findings. In this section, such studies are included in which different efforts have been made for identifying important microblogs specific credibility indicators through the wide range of factors studied (see studies conducted in section VI-A to explore important credibility indicators), and then these explanatory studies are conducted. They conclude that what serves best for credibility assessment. In such studies, to provide detailed features exploration and analysis, mostly data is collected from microblogs sites and tagged either through crowed sourcing environments or experts. A complete list of explanatory studies is presented and summarized under different columns in table 6. Following are only a few studies discussed for a basic understanding of such studies.

In [30] manually tagged dataset having three classes of features: social, context, and behavioral are analyzed, within 8 different topics and concluded the best credibility indicators.

In [122] an effort has been made and a wide range of factors are studied and an explanatory study is conducted.

Another very important explanatory study together with user perceptions has been conducted in [119], which examines the relationship between reader's demographics and related credibility features with user perceptions. Over 1317 attributed news tweets were collected and annotated using both TweetCred and manually; for examination of the relationship between eight tweet level features (including source) having reader's perception of credibility, news attributes, and reader demographics features. Further correlation among the attributes was also explored using Cohen's Kappa, chi-square, and association rule mining.

1) MICROBLOGS BASED AUTOMATIC CREDIBILITY ASSESSMENT MODELS

Following are all such categories in which only microblogs-based automatic credibility assessment systems are considered. They are classified as; Source Credibility, Feature-Based Models, Graph-Based Models, and Hybrid Models of Credibility. The outcome of this section is resulted in section IX and section XII. In section IX all these automatic assessment studies are summarized in four groups and their important findings are discussed. Findings include common features, strengths, and shortcomings. In section XII recommendations are presented, based on important findings of section IX.

C. SOURCE CREDIBILITY

There are many research studies where information credibility assessment is done through greater focus towards source /user of information [34], [129]. Ranking microblog users regarding their credibility could also be a candidate approach [123], therefore ways for source determination is also studied [124] and what affects the source

TABLE 4. Following are many surveys conducted. In these surveys user perceptions, judgment, heuristic, assessment or other elements have been studied. These elements are identified and examined through different methods like: surveys, interviews, empirical & experimental studies, observations and statistical methods (table 1 of 2).

Paper	Features Level Covered	Approach	Technique	Variables/ Features	Remarks
[189]	Topic, Post	Survey: Amazon MTurk	Variance Inflation Factor(VIF), Correlation, Hierarchical Regressions, Cronbach's α	NA	Social media sites vs traditional news media
[190]	Topic	Survey: Online	Cronbach's α, 9 Hierarchical Regressions	NA	Politically interested online users view for social networks as credible
[191]	Post	Survey: Amazon MTurk, Fluo, Apollo	Maximum Likelihood Estimation, ANOVA	NA	Human cognitive limit vs effect of automated system recommendation
[192]	Topic, Post, User	Survey: Online	Variance Inflation Factor (VIF), Correlation, Hierarchical Regressions, Cronbach's α	NA	Political blog credibility and selective exposure, avoidance
[120]	User, Post	Survey: Online	Statistical Methods, ANOVA, MANOVA	NA	Effect of follower-following over source credibility
[193]	Topic, Post, User	ACT-R Model Memories	Correlation, LDA, ANOVA	NA	Human credibility judgments
[8]	Post	Interviews, categorization using content analysis	Empirical Study	NA	Audience aware credibility constructs
[194]	User, Post	Survey: Amazon MTurk	Correlation Analysis, Statistical Methods	22	Factors influencing credibility perceptions for micro-blogs.
[195]	User, Post	Credibility Judgments	Cognitive Heuristics	NA	Cognitive heuristics for credibility judgment in online environments
[196]	Topic, Post, User	Survey: Mock Site, Interviews, Three Experiments	3 Way ANOVA, Cronbach's α	5	Factor influencing credibility of health and safety information on Weibo
[197]	User, Post	Controlled Experiment of 2 Treatment Group	1 Way K-Group MANOVA, ANOVA	7	Twitter's human agent vs bots
[121]	Topic, Post, User	Survey: Online	ANOVA	5	Country specific credibility perceptions
[198]	User, Post	Empirical Study, Web-based information activity diary survey, Experience Sampling	Statistical Methods	11	Various credibility constructs
[199]	User, Post	3 Surveys using Mock Site	Post Hoc Wilcoxon Rank, Omnibus F, Tukey, Friedman's Test, 1 Way ANOVA, PCA	6	Social network derived credibility

credibility [128]. For example, in [126] US Senate voting history data is used and the user is ranked to measure information credibility based on their online behavior. CredRank Algo (based on IR tech.) is developed by the authors to detect Coordinated Behavior. If it is found then those users were marked as not-credible. In [19] researchers proposed that user influence can be measured through characteristics like In-degree, Retweets, and Mentions.

Focusing on source credibility, tweet timelines of 10 general and 10 highly influential Twitter users of five areas each like: car, investment; are fetched and then making use of topic-related user's social structure, they try to find most influential/centric users within each topic as credible [125]. It is the combination of topic models over message contents and link structure analysis of the underlying social network.

User ranking based on authoritative user scores considering friend network and user-tweet/retweet network is implemented using ObjectRank in [127].

The research study performed in [111] focused on exploring indicators of credibility during eight diverse events. They concluded that URLs, Tweet length, Mentions, and Retweets are the best credibility indicators. The system proposed a ranking strategy based on content relevance and account authority considering: followers, mentions, list membership, and user-retweet graph. The system was trained using a learning to rank algorithm named RankSVM.

In short: the majority of researches in this category make use of the follower-following network or user-tweet/retweet network, etc.; with some form of popularity measures e.g.: the number of times a user is mentioned,

TABLE 5. Following are many surveys conducted. In these surveys user perceptions, judgment, heuristic, assessment or other elements have been studied. These elements are identified and examined through different methods like: surveys, interviews, empirical & experimental studies, observations and statistical methods (table 2 of 2).

Paper	Features Level Covered	Approach	Technique	Variables/ Features	Remarks
[200]	Post	Survey Embedded Experiment	Statistical Methods, Regression, Cronbach's α	11	Credibility of news: source, context
[201]	User, Post	Survey: Online	Kruskal–Wallis, Mann–Whitney U, the Wilcoxon Matched Pairs Test, Spearman Rank Correlation, Cronbach's α, Statistical Methods	NA	Credibility perception of social, teacher, scholarly tweets
[147]	User	Survey: Online	Correlation, P-Value, T-Test	4	Media cedibility of newspapers accounts on Sina Weibo
[202]	User	Data Collection: Twitter, Coded: Research Team	Statistical Methods	4	Tweet's source credibility : fukushima nuclear disaster
[203]	Topic, Post, User	Tagging: Professionals, Features Rated (Perception Based): Amazon MTurk Survey	Krippendorff's α, Pearson correlation, Box Plots, Scatter Plots, P-Value, Precision, Recall, F1 Scores	6	Epistemic study of information verification: features for Hurricane Sandy pictures real/fake
[118]	Topic, Post, User	Think aloud, Elaborative Questions (Verbal)	Statistical Methods, ANOVA	31	Microblog credibility perceptions
[204]	Post, User	Survey: Online	Tukey's HSD Test, Hierarchical Regression Model, Constant-Comparative Method, Statistical Methods	3	Student perceptions of instructor credibility and beliefs about Twitter as a communication tool
[205]	Post, User	Two Software based Surveys	Linear Regression, Statistical Methods, Pearson Product Moment Correlation	5+3	Visualization perception of five + three factors of trustworthiness
[206]	Topic, Post, User	Survey Questions/Ratings: Office Users	Pearson Correlation, KMeans, Linear Regression, Feature Distributions, T-Test, Density Estimation: Gaussian Kernel, Outlier: K-Divergence, Statistical Methods	10	Study bias amongst microblog users due to the value of an author's name.
[207]	Post	Survey	Maximum Likelihood Estimation, Structural Equation Modeling, Statistical Methods, Error Methods: Chi-Square, RMSs, GFI, CFI, AGFI, CI.	8	Credibility and trust in online media use
[208]	Post, User	Survey: Online	Tool: G-Power, Paired Sample T-Test, ANCOVA, Wilks' Lambda, Levene's Test	6	Journalistic credibility on twitter

retweeted, replied, listed, favorited, etc. by other users of the social network, and then apply modified page rank like model or some form of authority transfer for calculating highly influential/popular/authoritative/reputed user as credible. In source credibility identification, Social Network Analysis (SNA)/Graph-based methods are exploited most of the time, except few studies, which found some weighted ratios of a different combination of popularity measures, effective. There is no consideration towards post quality therefore labeling the post as credible or not credible is completely ignored. Primarily the efforts are being made to rank the user therefore there is strong overlap with both ranking and graph base credibility assessment methods.

D. FEATURE BASED MODELS FOR CREDIBILITY

Studies in this category usually build models which are either Machine Learning (ML) based or Information Retrieval (IR) based. They use features related to 'topics', 'posts', 'authors', 'network', etc., and of different types as well, such as Aggregated and Historic. Examples of topic-level aggregated features are the number of positive sentiment tweets, Avg. length of a tweet in a topic, etc. Historic features are difficult to extract and next level to aggregated features. For example, A user will be known as Topic Expert if his number of tweets under that topic is greater than the average number of tweets of that topic tweeted by all users. Calculating such features requires exhausting the complete dataset for that feature level (e.g.: user in this example). Such types of features are used to

TABLE 6. Explanatory studies: Many efforts have been made for identifying important microblogs specific credibility indicators through wide range of factors studied in previous survey studies.

Paper	Features Level	Approach	Technique	Variables/ Features	Remarks
[209]	Topic, Post, User	Tagging: CrowdFlower	Predictive Association Rule Analysis	8	News related tweet's credibility perception
[30]	Topic, Post, User	Tagging: Amazon MTurk	Distribution Analysis	34	Credibility related features distribution of twitter
[122]	Topic, Post, User	Tagging: Amazon MTurk	Statistical Methods, Kappa-statistic, Correlation, Forward Subset Selection Regression (FSS), Logistic Regression	45	Twitter credibility feature exploration and various ground truth analysis
[210]	Topic, Post, User	Tagging: Amazon MTurk	Statistical Methods, Kappa-statistic, Correlation, Forward Subset Selection Regression (FSS), Logistic Regression	45	Twitter feature exploration with network context and ground truth selection for credibility
[211]	Topic, Post, User	Tagging: CrowdFlower, Author and Post	Mean, Pearson Correlation	5	Impact of author's location on credibility
[212]	Post, User	10M Tweets Rated using proposed equations	Correlations, CDF, Statistical Methods	18	Scored features are statistically explored for trustworthiness assessment
[3]	Topic, Post, User	Manually (keyword search) Tagged for Ground Truth	Descriptive statistics, Filter Based Heuristic Approach	6	Understanding rumor/fake patterns/behavior/features in crisis
[141]	Topic, Post, User	Credibility Rating: Crowdsourcing and Experts	Krippendorff's α, Feature Distributions, Statistical Methods	44	Determining features of credibility in Arabic microblogs determining credibility
[119]	Topic, Post, User	TweetCred: Rating, CrowdFlower: Perception Survey	Chi-square Correlation Analysis, Cohen's Kappa, Association Rule Mining	Twitter:11, Demographic:4	Perception of reader vs news related microblog credibility features

explicitly exploit inter-entity relationships, which are inherent in graph/ network.

These feature-based assessment studies are also summarized in tables: 7 and 8. Each study is comprehensively presented across many important attributes. They are salient qualitative dimensions of these research studies which should be known for efficient exploration of the research area. These attributes are as following:

1. Paper: provides the reference of the concerned study. 2. Algo.: provides the name of the best performing algorithm of the study. 3. Learning Type: presents what type of learning or method is used like supervised classification, semi-supervised, unsupervised, ranking, etc. 4. Approach: presents what type of approach is used like feature-based, graph-based, information retrieval (IR) similarity measure based, weighted equations for scoring, user-defined ratios, etc. 5. Features Level: specifies that what different types and levels of features are used. There could be different levels of features e.g.: topic, user, post/tweet, or if its graph-based method then what type (directed, undirected) of the graph is developed over what entities/nodes (topic, user, post). Similarly, there are different types of features like historic, aggregated, or temporal. User+Historic means that user-level historic features are used. 6. Dataset: shows summary/statistics of data collected in the study. All of the studies extract their own dataset, because of the unavailability of the standard dataset. 7. Outcome: what was predicted in the study is expressed in the outcome. e.g.: credible (credible, not-credible), credibility levels (high, medium, low, not-credible), rank/score (0-10), etc. 8. Label Method: provides that who labeled the data, like domain experts, crowed source workers, automatically tagged through computations, by authors, evaluators (means team working for data extraction and labeling), manual (means labeling source is not defined). The labeling method defines the quality of the system. The best labeling is done by experts while labeling done by crowdsource workers is weak. 9. Focus: exposes major and special focus of the study, or system tile, e.g.: real-time assessment system, if the system is developed for 'emergency situation', if the system is produced for 'high impact events', 'fact-checking and scoring system', 'topic credibility', etc. 10. Product: either study is providing product as 'browser plug-in' or 'Twitter plug-in', or its just a research. 11. Distinct Attribute: it provides highlights of the study or some distinct features of the study, or if the system uses some distinct components, or some methodology, like 'online emergency monitoring component' is provided, 'Experimental study' is also provided, post 're-tweet network' is exploited for assessment, 'topic-based' method is provided, the system explicitly works on 'user expertise and reputation' for assessment, the system provides idea of how 'topic-based expert user with biasness' is assessed, etc. 12. Category: this attribute provides fine-grained classification of the study, either system uses ML, or IR, Learn to Rank (ranking), Mathematical, or Hybrid methods for assessment.

There are generally two classified groups where first includes such studies in which scientist worked at the atomic level of information means only or mostly on tweet [135], [156]; to assess the Information Credibility, such as: In [156] it is assumed that credibility can be judged from tweet text, a credible tweet always has many retweets with original text remain. However in a low credible message several terms are added with user opinions, deleted or edited, and has

TABLE 7. Feature based & hybrid microblog credibility assessment models classification (table 1 of 2). Each study is comprehensively presented across following important attributes. They are salient qualitative dimensions of these research studies which should be known for efficient exploration of the research area.

Paper	Algo	Learning Type	Approach	Features Level	Dataset	Outcome	Label Method	Focus	Product	Distinct Attributes	Category
[137]	Feature Rank Naïve Bayes	Classification	Feature Based	Tweet, Aggregated (Topic,User), User+ Historic	Yemen Civil War: Keywords (Taiz & Aden) Tagged 11000 sample tweets	Credible	Experts	Four Component System, CDF Based Feature Analysis	No	User Expertise & Reputation	ML
[134]	Proposed CIT Bayesian Network	Classification	Feature Based	Tweet, User + Aggregated, Topic+ Aggregated, Temporal	UK-Riots related topics, 350 Tweets Tagged	Credible	Experts	Emergency Situation	No	Online Emergency Monitoring Component	ML
[135]	SVM Rank	Semi-Supervised Ranking	Feature Based	Mostly Tweet Level	6 High Impact Crisis Events (tagged 500 tweets for each event)	Score/Rank	Crowd Worker	Real-time Credibility Score	Browser Plugin	Mostly features extracted from tweets	Learn to Rank
[133]	J-48 Decision Tree (3-Fold)	Classification	Feature Based	Tweet, User, Topic + Aggregated, Retweet Tree Propagation	2524 Trending Topics, Classes: News +Chats, Topic are tagged using 10 tweet samples	Credible	Crowd Worker	Topic Credibility	No	Retweet Net. tree used	ML
[131]	SVM Rank with PRF Re-ranking	Ranking	Feature Based	Tweet, User	14 News events of finance, politics like domains, 3586 Trending topics, 35M tweets + 6M Users, Tagged 500 Tweets for each topic	Rank	Crowd Worker	High Impact Event	No	Pseudo Relevance Feedback for Re-ranking	Learn to Rank
[132]	Similarity Score TF-IDF	Un-Supervised	IR Similarity & Feature Based	Tweet, User	2 News Topics (Iran/Yemen & Houthi), 600 Arabic Tweets, 179 Authorized News Articles for Verification, 29 Tagged Tweets for Evaluation	Credibility Levels	29 Tweets Tagged for Evaluation & Thresholds are set by Experts to Classify	Similarity Based News Fact checking Score+ Feature Score: Link, User Authority, Inappropriate Words	No	Experimental Study	IR Based
[113]	Random Forest (n-estimator:100)	Classification	Feature Based	Tweet, Topics, User + Historic	10 Trending Topics of Japan, 200 Tweets of each Topic Tagged	Credible	Crowd Worker	LDA used for Topic Extraction	No	Topic Based User Expertise & Bias Extracted	ML
[130]	J48 Decision Tree	Classification & Weighted Equations	Feature Based, Social Context Based Authoritative User using Ratios	Tweet, Topic+ Aggregated, User+ Historic	7 Topics of Libya only, 37K Users, 126K Tweets, Tagged 5000 only	Credibility: +Ve, -Ve, True, Null	Evaluators	Topic Based Assessments, LDA is used for Topic	No	1. Social Model: Social NW's important indicator's ratios are weighted. 2. Supervised Model based on Contents. 3. Hybrid 1& 2.	ML & Mathematical

TABLE 8. Feature based & hybrid microblog credibility assessment models classification (table 2 of 2). Each study is comprehensively presented across following important attributes. They are salient qualitative dimensions of these research studies which should be known for efficient exploration of the research area.

Paper	Algo	Learning Type	Approach	Features Level	Dataset	Outcome	Label Method	Focus	Product	Distinct Attributes	Category
[136]	Random Forest	Classification	Feature Based	User, Tweets	7000 Tweets on Nature Environment Preservation, 100 Terms used, Tagged 1206 Tweets	Credible	Manual	Trying to make time efficient Twitter Plugin	Twitter Plugin	Reconciliation System for Tagging Evaluation & Retagging	ML
[21]	Naïve Bayes + Relative Feature Importance	Classification	Feature Based & Weighted Score	User+ History, Tweets	1.2M Tweets of Topic Iraq & Levant (ISIS) DAISH, Tagged 1000 of 700 Users	Credible	Experts	1.Complete User Level Sentimental & Credible Tweets Ratio is computed. 2. Tweet's Credibility Probability value predicted. 3. Weighted 1 & 2	No	Multi- Stage Model having: Relative Importance, Classification, Opinion Mining Components	ML
[31]	Random Forest (10 fold CV)	Supervised & Unsupervised	Feature Based & Graph Based	Topic+ Aggregated, User, Tweet, Directed Weighted Graph of: User, Tweet, Topic	Turkey's 25 Trendy Topics & 100 Tweets each & also Tagged	Labels: Important, Newswor-thy, Correct	Crowd Worker	1.Different Models trained separate for user, topic, tweet and combined. 2.Labels are converted in scores. 3. Authority transfer 4. Threshold is used for getting labels	No	1. Authority Transfer Model by means of equations. 2. Weighted Directed Network of user-tweet-topic.	Hybrid
[138]	SVM	Classification & Unsupervised	Feature Based & Graph Based	User, Tweets, Events+ Aggregated, Weighted Directed Hierarchical Net. of: Event, Sub-event, Msg.	Cina Weibo's Two Datasets: 1.SW2013 (Topic Independent) 18 Fake, 171Real News (79K Tweets, 63K Users) 2.SW-MH370 (Topic Dependent) 32 Rumor, 103 Real (31K Tweets, 24K Users)	Credible	Fake, Rumor, & Real News are Collected from Authentic Sources	1.Focus on News Credibility. 2. Sub-event & Message Layers has Inter & Intra Layer Links. 3. Event & Sub-event Credibility Initial Scores are Calculated By Avg. of Message	No	Graph Optimization Method Used for Proposed Model Named NewsCP	Hybrid
[139]	Decision Tree(J48)-D2010 & KNN-D2011	Classification & Unsupervised	Feature Based & Graph Based	User, Tweet, Event, Directed Weighted Network of: Event, Tweets, Users	1. D2010-47K Users, 76K Tweets, 2K Topics, 207 Events. 2. D2011-9K Users, 76K Tweets, 2K Topics, 250 Events. (Events are tagged by 10 sample tweets)	Credible	Authors	1. Focus on Event Credibility. 2. Event & Tweet Implications computed as +ve/-ve weights within layers. 3. Event & Tweet layers have inter & Intra links	No	1. Initially used PageRank like Algo. as BasicCA. 2. Graph Optimization used to enhance results as: EventOptCA	Hybrid

low retweets. Based on the said concept user credibility is also calculated with tweets. The reputation-based credibility degree assessment method developed for wikis is applied for tweets. The study has no experiments and Evaluation. It just uses ratios/ mathematical scores.

A browser plug-in named TweetCred, is a real-time system, build over semi-supervised learning using SVM-Rank and trained through 45 tweet level features (only data provided in tweet object is used). These features are generally classified in Meta-data, Content-based linguistic features, Author (only #follower-following and age), Content-based lexical features, URL reputation score, and Tweet Network features. The system is developed through six high-impact crisis events of the year 2013. Only US-based annotators were used to annotate 500 tweets. The system was widely downloaded and used [135].

Other group includes studies which exploit other level features too in addition to tweet level with all features types e.g.: Aggregated and Historic, such studies include A Hybrid model combining two models through averaging and filtering: the first model, named social model measure social credibility, deals with credibility at the user level, combining many dynamics of topic-specific content flow within its social network; and second model named content model measure content credibility, calculates fine-grained tweet level content based credibility [130]. In short: a total of 19 features are used to generate a score first and then making use of user friendship network user transfer that score to their followers. Dataset was generated through 7 topic-specific ‘‘Libya’’ and a total of 5000 manually annotated tweets of 37K users.

14 high impact news events of 2011 are considered and investigate the tweets based on supervised learning, with RankSVM + Pseudo Relevance Feedback over content-based and user-based static features, and then credibility is ranked [131].

An experimental system was developed with two approaches: One was based on the similarity of tweet news text and verified/authentic news text, and the other was combined with similarity-based features and other proposed (tweet and user level) features. Only IR-based methods were used and the system was developed on two hot news topics having 600 tweets which were verified through 179 authentic news articles [132].

It’s a seminal study [133] where tweets belong to trending topics are collected and a wide variety of features related to Topic, User, Propagation, and Message; are extracted for supervised learning using J48 Decision Tree as best ML Algo. [134] Aims to measure credibility in an emergency situation using Bayesian Network over features based on: Diffusion, Topic, Content, and User.

An effort was made to develop a time-efficient twitter plugin in [136]. A dataset of 7000 tweets fetched on Nature Environment Preservation, with the help of more than100 related terms, then 1206 tweets tagged and Random Forest classifier was trained over user and tweet level features. Results were improved through a reconciliation system for tagging evaluation and re-tagging.

A classification system consisting of four components: Reputation Component - based on user popularity and sentimentality; it initially helps to filter neglected information for further assessment. Classifier Component - classify credible/incredible, using four ML-based classifiers. User Expertise Component - rate user expertness for the topic. Finally, the Feature Rank algorithm best ranks the features for best credibility assessment. The system was trained and tested on two fetched datasets [137].

A Multi-Stage Model [21] having: Relative Importance, Classification, and Opinion Mining Components. The system’s Dataset was constructed using 1.2M Tweets of Topic: Iraq and Levant (ISIS) DAISH. Only 1000 tweets of 700 Users were tagged to train the Naïve Bayes classifier with Relative Feature Importance implemented over user and tweet level features. First of all complete User’s Sentimental and Credible Tweets Ratio was computed, then Tweet’s Credibility probability value predicted using a trained classifier and finally, both values are combined as weighted credibility score.

Total 2000 trendy tweets of 10 topics posted in japan were annotated through four questions and trained a Random Forest classifier. Four distinct features: tweet topic, user topic, user’s expertness, and bias are additionally assessed. Tweet topic and user topic features were extracted from LDA and concluded that topical features improve credibility assessment [113].

Following are some serious observations, first: it has been observed across all automatic credibility assessment systems of any type (e.g.: Source based, Feature based, Graph based, and Hybrid) and even in explanatory studies, that majority of these studies get their dataset labeled either considering that: post seems ’informative/newsworthy’ or ’trustworthy/truthful’ to the evaluators. Only couple of studies considered Real and Fake news from authentic sources and get their dataset labeled on authentic basis rather than on evaluator’s perception. Second: Many important aspects regarding evaluation criteria discussed in [213] are also fully ignored. Third: another important observation regarding every research study that they just consider news event for credibility, any other piece of information is not even considered for credibility assessment, though information credibility exist in every piece of information.

E. GRAPH BASED MODELS FOR CREDIBILITY

Such studies of Source Credibility type, are classified in this category which uses Social Network Analysis (SNA)/ Graph-based models [214] by utilizing friendship (follower/following) network, user’s tweet/retweet propagation network, etc. The majority of Source Credibility studies are graph-based (see section VI-C). An academic research (TURank) [127] is discussed as an example case which is classified as source credibility using the graph-based method. In this study, the original Twitter information network flow is used to find the authoritative user. The philosophy of TURank says that: user becomes more authoritative when followed by another authoritative user. Likewise, tweets become more important when retweeted and it also affects its user's authority. Therefore types of such authority transfer in TURank are: user-user, tweet-tweet, tweet-user, and user-tweet.

Other graph-based models are intentionally not discussed for these few reasons. 1. They are too many in quantity because a majority of source credibility studies are all graph-based. 2. There are surveys available for these graph-based models which are explicitly discussed under source credibility. 3. Important concepts and techniques are completely covered in the other two types of models like feature-based models and hybrid models.

F. HYBRID MODELS FOR CREDIBILITY

Hybrid models combine the strength of both feature-based and graph-based models, therefore a much better approach has resulted in very few shortcomings. It is commonly observed that studies in this area initially exploit feature-based models to get User, Tweet, etc. seed scores which become nodes of some user-defined network. Afterward, the network of such entities like Topic, Tweets, Users, or Events, having inter and intralayer-directed links with signed weights, are made. Event/Topic initial scores may be generated through aggregated values of their decedents. Finally, graph-based or graph optimization methods are used for score convergence, and some thresholds are used for credibility prediction. It is explored that simply linking entities as a network enable hybrid models to best exploit implicit entity relations.

Following are the studies categorized as hybrid models for credibility. In the study, [31] a total of 41 features for Topic, Tweet, and User are used for learning and score generation. As each tweet refers to the user as well as the topic, therefore initial score is used in authority transfer for calculating the credibility of each tweet. Dataset was generated through 25 trending topics of Turkey having 100 tweets in each.

Another hybrid approach is used in [138]. Two Datasets (topic dependent and independent) were used. Both extracted from Cina Weibo's messages having Rumors, Fake, and Real News which were selected from authentic sources. The SVM classifier was trained first on the user, tweet, and event (aggregated) level features, and then a weighted directed hierarchical network of entities as Event, Sub-event, and Messages was constructed with inter and intralayer links. Inter-layer links represent explicit relations between network entities. Messages' initial credibility scores were generated by a trained SVM classifier and then Event and Sub-event credibility initial scores are calculated by respective averages. Finally, the graph optimization method was used for the proposed model named NewsCP.

In [139] two datasets having 76K Tweets and 2K topics each, of 457 total Events, all were tagged with 10 sample tweets. First of all two separate classifiers: Decision Tree (J48) and KNN were trained for each dataset, on User, Tweets, and Event level features then a weighted directed network of entities having Event, Messages, and Users was constructed. Entities were linked with their explicit relations. Event and Tweet Implications are computed as positive/ negative weights within each respective layer for their intralayer links. Initial tweet scores were obtained from the respective classifier and then a PageRank-like algorithm named BasicCA was executed over the network. The final optimized results were obtained from Event Graph optimization-based algorithm named: EventOptCA.

All above hybrid credibility assessment studies are summarized in Table: 8 (see last three entries of the table), across different attributes.

VII. STANDARD CREDIBILITY DATASET

It is extremely important to discuss that one more challenging issue which is unsolved. It is the absence of predefined credibility benchmarks and its related gold standard dataset. The difficulty of collecting a large amount of such data has not yet received the attention it deserves [29].

Though there are many Deception related (e.g.: fake news, Rumor, Hoax, Spam, etc.) datasets (e.g.: LIAR [215], FakeNewsNet [17], BuzzFeedNews [216], DeClare [217], FakeNewsAMT [218], Hoaxy [219], Kaggle's- BSDetector [220], SemEval Task8 [221], Rumors [222], etc.) [169] are available. Web site's contents related credibility dataset [182], Event Credibility dataset [160], Bot and Malicious Profiles Detection dataset [100] and similarly few other credibility related components datasets are also available.

We have developed Credibility Taxonomy in table: 1 and figure: 3, summarizing all above sections (3-7) and the detailed classified tables: 7 and 8 to summarize and categorize automatic credibility assessment approaches across various dimensions, for all feature-based/ML/IR and Hybrid models. Graph-based models are intentionally not included for few reasons: one they are too many in quantity, second there are surveys available for only source credibility, and last; important concepts and techniques are completely covered in the other two as well.

VIII. LITERATURE BASED IMPORTANT FEATURES

It is very important to know that what features are being used in microblogs credibility assessment studies, throughout the literature. Therefore, in this section most common and important features are extracted without any specific consideration of type, and methodology used. In this research study, there were almost 50 papers which were focusing specifically on microblogs. These were all discussed under section VI: Information Credibility of Social Media and Microblogs. There are two components in every information shared at microblogs: Post and Poster. At poster level: it is found that user's followers and followings, number of posts, age of account were found dominating in many papers. Location, picture in profile, description in profile were moderately used. It can also be observed that in the same user object, that time zone and gender are not much used (see figure 4).

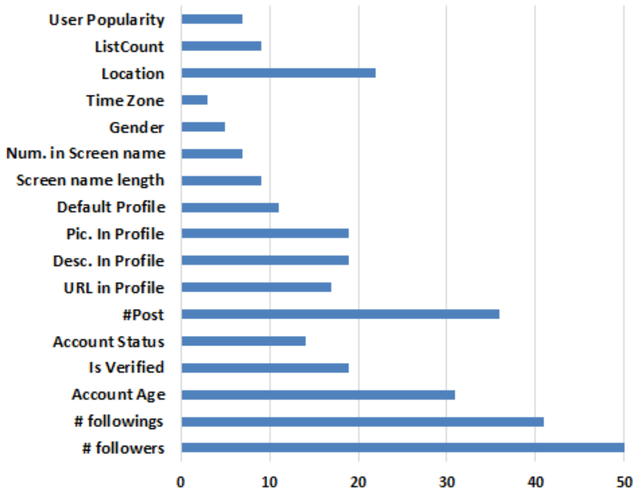

FIGURE 4. Mostly used user-related features in literature: in 50 papers.

In post object: URL, retweet, hashtags, mentions are found strongly dominant in the majority of papers. Sentiment score, and post content/text which was mostly used as bag of word (BOW) form, were also considered good features for assessing the microblogs' credibility. The number of words, number of characters, and number of replies are moderately used. Similarly, in post object few features like: number of media, isreply, isretweet, special characters, and day of the week are less utilized (see figure 5).

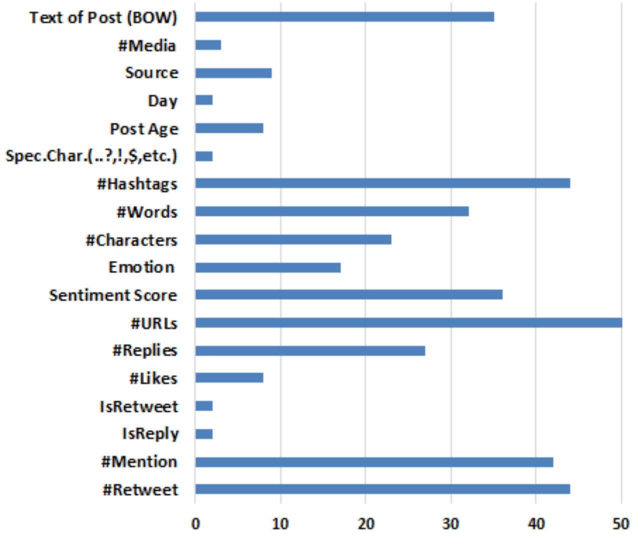

FIGURE 5. Mostly used post-related features in literature: in 50 papers.

It is found that in addition to raw features (e.g.: #retweet, is_reply, #mentions, #hash-tags, etc.) aggregated features and historic features performed better in assessing credibility [223].

The above most commonly used features are also adopted for the proposed framework's features presented in table 15.

IX. FINDINGS AND DISCUSSIONS

After a deep exploration of the literature and in-depth study of many automatic credibility assessment models (discussed from section VI-C to VI-F). These models or research studies are broadly categorized as following four types, to understand and briefly discuss their distinct features (see table 9), strengths, shortcomings (see table 10) and for recommendations (see table 11). Each category is briefly discussed under following respective headings and summarized in table: 9 as well:

A. FEATURE BASED - TWEET CREDIBILITY

A large number of researches only extract features based on authors, contents of tweet, topic, and underlying network-related static features, e.g.: number of followers and followings, etc. (available in a tweet only) and apply Machine Learning models to identify credibility score or label. Such models completely ignore the influence of user's friends-network and post propagation networks, etc. On the other hand, they are also unaware that some very important credibility features like the number of retweets, likes, followers, etc. are generally inflated by malicious profiles/ bots, hence produce a completely false sense of credibility. Similar to all other categories they are also affected by the absence of post quality-related many credibility aspects (e.g.: Fake, Bias, Spam, Rumor, Smear Campaigns, Conspiracy, etc.) proposed in our credibility framework's table 14.

B. GRAPH BASED - USER CREDIBILITY

It is assumed in this category that if the post is authored or propagated through a highly authoritative or influential/central user then it is likely to be more credible. Many attempts are made using graph-based (un-supervised) methods to identify the user influence within the social network and then credibility is judged for the author's influence, and therefore infected with things like fake followers/ follower's fallacy, coordinated behavior, etc. These manipulations are mostly done by malicious profiles/bots. Focus is fully shifted towards the source of the message and therefore post itself is completely ignored and similarly, post-quality-related important credibility aspects discussed earlier are also ignored. Regarding automatic credibility assessment, we need to carefully set a threshold for identification of our credibility label, as data will be unlabeled for such un-supervised problems.

C. FEATURED BASED - TWEET + USER CREDIBILITY

It is the extension of 1st Category (named as Feature Based-Tweet Credibility), discussed earlier with additional focus on user-related credibility aspects. In addition to message/post credibility, by using different ways and means, the credibility of the user is also measured to label or score the final credibility. For example, assessing the credibility of the user, different historic or aggregated or weighted features could be used (historic, aggregated features are discussed in subsection VI-D). It is generally observed that; each message

TABLE 9. Summary/ important aspects of microblogs based automatic credibility assessment models categories (table 1 of 4). There were four types of microblogs based automatic credibility assessment models (see sections: VI-C, VI-D, VI-E, VI-F).

1st Category of Research	2nd Category of Research	3rd Category of Research	4th Category of Research
Feature Based – Tweet Credibility	**Graph Based – User Credibility**	**Feature Based – Credibility: Tweet + User**	**Hybrid - Feature Based + Graph Based**
Using only or mostly Tweet level features, ML model is trained.	1. Using friends-following or User-tweet-RT Network, apply modified Page Rank like model or some form of Authority transfer, etc. for calculating highly influential/popular/reputed user as credible.	User level (Historic, Aggregated/ Weighted Features: Sentiments, Favorites, Mentions, Retweet, Listed, Friends NW influence, etc.) & Tweet level features are used to train ML model.	1. Initially feature based models are used to get user, tweet score, then network of entities (like: Topic, Tweets, Users, Events) having inter and intra layer links with weights, are made and finally some graph based methods used for score convergence.
	2. Ranking User.		2. Best exploits implicit entity relations.
	3.Un-supervised.		3. Much better approach with minor shortcomings.

TABLE 10. Shortcomings of microblogs based automatic credibility assessment models categories (table 2 of 4).

1st Category of Research	2nd Category of Research	3rd Category of Research	4th Category of Research
Feature Based – Tweet Credibility	**Graph Based – User Credibility**	**Feature Based – Credibility: Tweet + User**	**Hybrid - Feature Based + Graph Based**
Shortcomings			
Credibility not captured.	Assumption that Post quality is just based on user.	Effected with followers fallacy, and bot manipulations like problems.	Assessing Event/Topic credibility is complex. Somehow tweet credibility.
Fakeness not assessed.	Post Quality is completely ignored.	Post Fakeness not assessed.	Respective Feature Based shortcomings are inherent.
Bot manipulations in RTs, Likes, Mentions, #Tags etc.	Bots/Cyborgs seems highly influenced here.	Bot manipulations in RTs, Likes, Mentions, #Tags etc.	Thresholds may be different for different nature of topics/events in real-time.
	Credible/Not-Credible not labeled		

TABLE 11. Proposed or recommended features selected or considered under each research category of microblogs based automatic credibility assessment models (table 3 of 4).

1st Category of Research	2nd Category of Research	3rd Category of Research	4th Category of Research
Feature Based – Tweet Credibility	**Graph Based – User Credibility**	**Feature Based – Credibility: Tweet + User**	**Hybrid - Feature Based + Graph Based**
Proposed/Recommended Credibility System			
Hybrid (Feature Based + Graph Based) – Tweet Credibility Score: User + Tweet			
Comprehensive Tweet level Features (in addition to others) are used to assess post quality rank through Learn to Rank Model.	Un-usual to this category, Graph Based models are applied at both User + Tweet levels. Graph Based models are applied at each tweet's retweet network and user followers-following network.	Many User + Tweet level features, including Historic, Aggregated and simple features are considered.	Both User (Friends Network-Influence) and Tweet level (Retweet Network-Spread & Propagation) features having scores, which are used as features.
	User influence score (using only trustworthy or Non-Bot followers-following network)		Finally all features including Network Scores and remaining normal User Features + Tweet Features are used to rank tweet, using Learn to Rank models.
	Tweet Spread, and Propagation scores (using retweet network) are also calculated.		

affects the credibility of its author and vice versa. It's an extremely important phenomenon but very rarely identified. We observed that only one study tries to identify the credibility score through identification of the topic of the tweet and then identify the number of topic-specific influential users involved in re-tweeting and then determine the credibility score. One study identifies credible tweets only when if it remains original and then scores its source and then the tweet score is calculated but it's all about theoretical. One study proposed that if more authoritative/centric people are involved in retweeting then score of credibility is increased.

D. HYBRID - FEATURE BASED + GRAPH BASED

The modern method, which isn't sufficiently explored in studies till now, exploits the power of both Feature-Based and Graph-Based models, known as a hybrid. They attempt for Feature-Based models for initial credibility prediction of respective entities, for example, predict credibility of the tweet, user, topic, etc. and then further boost the results through incorporating their scores/predictions to an inter-connected network of participating entities like Post, Poster, Topic, and Event. There is an obvious observation, as we discussed in the above 3rd category, that each entity affects the credibility of others and gets affected, which means all

TABLE 12. Summary of completely distinct recommendations which are not considered in any of the four research categories presented in tables 9,10, 11 (table 4 of 4).

Completely Distinct Considerations Recommended (not included in any research category)
Identify if A/C behavior is like malicious Bot/Troll/Cyborg/Sybil, etc. (such A/C will be omitted from friends network for correct influence calculations)
Identify post fakeness (Post Level) and also update User's fake producer counter (User Level)
Score of post is computed (using all User & Post Level features + actual retweet network's propagation and spread measures + user rank over friend's network)
Users features includes: Domain (area of expertise), correct influence calculated only over trustworthy friends network, etc.
User includes: Fake Produced % age, Spread Score & Propagation Score (Avg. of Tweet Spread & Propagation Scores)

are interdependent, which is implicitly exploited through network models in hybrid settings. Despite the strength we discussed, they inherently suffer from some shortcomings of feature-based models as well. Few other shortcomings include difficulty in assessing real event credibility or topic credibility values, which somehow primarily, tweet credibility again, e.g.: the credibility of a topic is computed through all their tweet credibility values. Once they are calculated, then they are again used in their interconnected network of participating entities, where these values are mostly amplified with some scalar effect. Another limitation is threshold settings which differ for the different domains (e.g.: politics, education, entertainment, sports, etc.).

1) SHORTCOMINGS

(other than above categories): besides all above category-specific shortcomings there are some other extremely important shortcomings that are not discussed in any category because they don't fall in any category and they are also considered as our recommendations (see table 12). They are also discussed in section XII with other associated details and enlisted as following, like:

1. A very vital aspect that is completely ignored that the credibility of a message can't be determined without going into the underlying credible and trustworthy friend's network, to measure the correct influence of the user. If malicious profiles exist in a friend's network then they must be omitted before examining the user rank/influence. Malicious profiles/bots identification and their rectification must be done for credibility assessment initialization, to prevent their serious manipulations at various places.

2. Chain of narrators is extremely important in assessing the message's credibility. Once a post is identified as fake then its producer must be penalized by incrementing its fake producer counter. Similarly, each fake propagator involved in post propagation within the post's chain of narrators must also be updated.

3. Credibility of the post must be calculated using a comprehensive list of features provided in table 15. This proposed list of features covers the majority of aspects like, post quality (which is ignored in the majority of studies, see figure 6 for a post-quality-related group of aspects), veracity and different forms of deception, hate speech, post spread and propagation, user's veracity, expertise, rank, and malicious profile identification. All of these features are extremely important for automatic credibility assessment, e.g.: the spread and propagation pattern of a message is an important feature for credibility assessment. Computing user's influence or rank on a followers-following network comprising of non-malicious users/profiles only. After computation of all such features, an appropriate Machine learning model could be trained over these features for score/rank prediction.

4. Two extremely important features which are fully ignored in credibility studies are user domain/topic-specific expertise and true user influence score computation without bot manipulations.

5. As it has been discussed earlier that many post-level features could compute user-level features. Therefore many user-level scores could easily be computed like, User's Avg. Post Credibility Score, User's Fake Post Produced %age, User's Fake Post Propagated %age, User's Spread and Propagation Score Avg., etc. Computation of all such scores at the user level will implicitly reduce the dissemination of low-credibility contents, over microblogs.

Detail recommendations are presented in section XII.

We have also presented a summary of the above observations in table 9 with their shortcomings in table 10 and our those recommendations which are based on already defined research categories, presented in the table: 11, whereas recommendations which are fully distinct or completely missing in all the categories are proposed in table 12.

X. THEORY DRIVEN CREDIBILITY FRAMEWORK

The framework has theoretical foundation. How the framework is driven and what are the basis of our proposed framework is presented as following:

1. Basic components (Levels, Dimension, Constructs) of credibility are identified through detailed literature exploration from different disciplines of credibility like physiology, communication, information sciences, etc. (see section III-A under heading 'Credibility Components', and table 2 and 3).

2. All credibility supported research studies were identified first, after detailed literature exploration, then each concerned research study is categorized and discussed under its respective construct. Example: Fake News Detection studies are categorized and discussed under Deception, Truthful constructs (see complete section V).

3. Necessary credibility components identified in step 1 and 2, are presented in the form of a framework, presenting their inter-relationships (see table 14).

TABLE 13. Economics, social sciences basic theories, and credibility studies driven credibility framework components.

Basic	Framework	Components	Theory	Description	Research Based Ref.
Post Related Theories	Contents	Quality	Information Manipulation Theory [224]	Too many or too few refers to deception	It is primary component so too many ref. are found, just few are: [135], [153], [38], [163], [225], [154]
			Reality Monitoring [226]	Real events are identified by sensory perceptual information	
			Four Factor Theory [227]	Emotion, arousal, thinking, and behavioral control are expressed differently in lies and truth.	
			Undeutsch Hypothesis [228]	Factual contents differ in quality and style from fallacy	
Source (User) Related Theories	Expertise	Community/ Peer Influence	Rare Behavior [18]	Unusual behavior than majority	Expertise: [229], [230]
			Synchronized Behavior [18]	All such user show/ follow the similar behavior patterns.	
			Coordinated Behavior [18]	All such chain of users are developed to perform some pre-defined task of their master.	
			Collective Behavior [18]	Actions performed by presence oriented mass (crowds, mobs, riots, cults)/ distance oriented (rumors, mass hysteria, moral panics, fads, crazes)	
			Social Identity Theory [231]	portion of an individual's self-concept derived from perceived membership in a relevant social group	Combined Expertise & Trustworthiness [43]–[46], [232], [38], [128], [150], [233], [234], [22], [36], [225], [235]
			Emperor's Dilemma [236]	Alternative possibility, that members of a group may enforce to act in ways that few if any group members actually want or need.	
			Normative Influence Theory [237]	People change to form a good impression and fear of embarrassment or to be liked or accepted by others	
			Availability Cascade [238]	Self-reinforcing process in which a collective belief gains more and more plausibility through its increasing repetition in public discourse within their social circles	
		Individual Influence	Overconfidence Effect [239]	One's subjective confidence in his or her judgments is reliably greater than the objective ones.	
			Illusion of Asymmetric Insight [240]	We understand others better than they understand themselves	
			Naïve Realism [241]	A believe that we see the world objectively, and people who disagree, must be irrational, or biased.	
	Trustworthiness		Selective Exposure [242]	Prefer information based on pre-existing attitude.	
			Confirmation Bias [243]	Trust information based on pre-existing beliefs.	
			Desirability Bias [244]	Accept information that please them.	
		Community/ Peer Influence	Bandwagon Effect [245]	Do something because others are doing.	Trustworthiness: [36], [246]–[248]
			Conservative Bias 158	Revise one's belief insufficiently when presented with new evidence.	
			Validity Effect [249]	Believe that information is correct after repeated exposures.	
			Semmelweis Reflex [250]	When something contradicts with well established norms then reject such new evidences	
			Attentional Bias [251]	failure to consider alternative possibilities when occupied with an existing train of thought	
			Echo Chamber Effect [252]	Within a close system, belief are amplified by communication and repetition	
		Driven By Benefits	Contrast Effect [253]	When compression enhances differences.	
			Prospect Theory [254]	People decide between alternatives like gains or losses, and just think in terms of expected utility rather than absolute outcomes.	
			Optimism Bias [255]	Overestimate the probability of positives and underestimate the probability of negatives	

4. To strengthen our framework components we identify the basic theories of Economics and Social Sciences which are supporting or leading towards individual framework components (see table 13).

5. To strengthen our framework components we identify the basic credibility studies which are supporting or leading towards individual framework components (see table 13).

It has already been discussed that outcome of our study was two-fold. Understanding the broader domain of credibility with basic components identification (i.e.: levels, dimensions, and constructs) and then the development of compatible social media generic framework will be carried out. This will further be transformed to microblog-specific implementation. Considering the first objective, a theory-driven generic framework of social media is going to be identified in this section, consisting of levels and dimensions. These two components are completely generic to social media only. Constructs must be carefully identified for both social media and microblogs. Therefore they are identified in the next section XI in addition to our microblog specific implementation as our second objective fulfillment. The generic (levels and dimensions) and specific (constructs) framework components have already been identified in previous section III, under the heading of 'Credibility Components', through strong and detailed literature exploration.

In addition to the literature explored in previous sections, to form the strong basis of credibility framework. A comprehensive and dual study is also conducted as follows. Table 13 completely map our framework components (see first merged column for framework components) with the following Social Sciences & Economics Theories (see second column for these theories with short description) and then with Credibility Studies in the last column (see research-based references of these studies):

A. SOCIAL SCIENCES & ECONOMICS THEORIES DRIVEN

We have surveyed many related basic behavioral and human cognition theories defined across varied disciplines: like economics and social science. Each theory with its short description is presented in table 13. They provide important guidelines for the required level (post or poster) of credibility and deception. Such theories simply lead towards building efficient models of credibility identification or assessment. High-level analysis of these selected theories resulted that they are either related to the post itself or posters. Hence two pillars or levels of credibility could be identified first which are 'post' and 'poster'. Further considering the important dimensions of credibility. These theories are also classified under 'content quality' of post and two types of influence (e.g.:community and individual) which directly affect either 'poster's expertise' or 'trustworthiness'. Some specific theories are driven by benefits that affect the poster's trustworthiness as well. Therefore three major dimensions are also identified: content quality, expertise, and trustworthiness (see table 13).

B. CREDIBILITY STUDIES DRIVEN

In strong support of our framework components, we had already explored detailed credibility studies and credibility macro components (e.g. Levels, Dimensions and Constructs) had also been identified in section III (see table 2 and 3). It has also been discussed that considering social media credibility, only two main levels of credibility named message credibility and source credibility are feasible [42]. Regarding media credibility, it is also discussed earlier that in modern scenario medium is also replaced with source only [59]. In this case, the source has to be thoroughly examined including all chain of narrators involved in message propagation. Many leading credibility related research studies highlighted 'Trustworthiness' and 'Expertise' as major dimensions of source credibility (see last column named 'Research based Ref.' of table 13) and also table 2, 3. It could also be seen in table 13 that content quality is the most important and primary dimension of message/post (see table 2 references [140], [143], [153], etc. and table 13 references [38], [135], [154], [163], [225], etc.)

Identified credibility framework which is completely generic to social media, is theory-driven. The framework is fully supported through social sciences & economics theories and credibility-related research studies. Complete mapping of these theories and important research studies are all provided in a single comprehensive table 13. Considering the primary objective of the study, the extract of credibility framework is theory and research studies driven. High level credibility framework picture is further presented in figure:6, which will be completely understood after section XI.

XI. PROPOSED SOCIAL MEDIA CREDIBILITY FRAMEWORK

The framework was identified through supported theories in the previous section. That theory-driven framework is presented with further necessary details, in this section:

A. MOTIVATION AND OBJECTIVE

Determining the credibility of information in microblogs is becoming one of the most challenging issues day by day and still, unresolved [20]–[22]. Even though it has been studied much, since last many years. It is observed that too much work is done on theoretical or conceptual aspects of credibility in other related fields but they are not properly considered in microblogs related automatic credibility assessment studies conducted in computer science. These theoretical or conceptual aspects of credibility are mostly studied in psychology, communication and information sciences, where as microblogs related automatic credibility assessment studies are done in computer sciences fields. Unfortunately, no work had been done on mapping these general constructs of credibility for microblogs, which should be considered minimally when developing the respective system to assess the credibility. Due to being multi-perspective nature, the diversity in the definition and perception of credibility reflects different viewpoints in different work studies. Some studies consider 'Relevance' as the criterion of being credible. Some assume 'Reputation' as the major driver of credibility, whereas the majority only stick that 'Fake' identification is credibility identification. It is also perceived by researchers, that 'Ranking' concerning author influence and topic expertise are strongly treated as credibility ranking. The majority of studies exploit 'Informativeness' as a credibility indicator. Few found examining 'Trust' level as true credibility judgment. It is observed and quite evident in many research studies as well, that the credibility notion needs to be standardized because

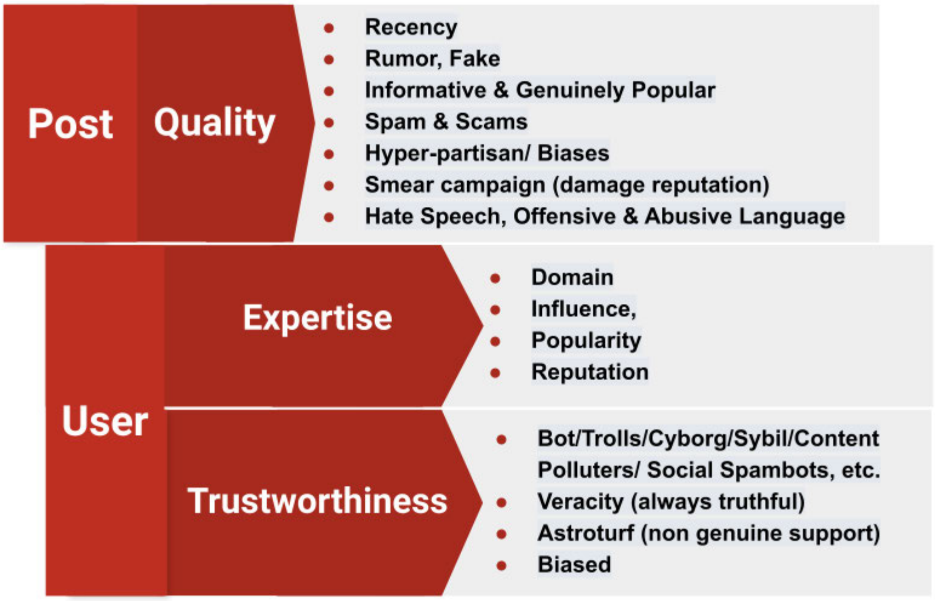
Post
Quality
Recency
Rumor, Fake
Informative & Genuinely Popular
Spam & Scams
Hyper-partisan/ Biases
Smear campaign (damage reputation)
Hate Speech, Offensive & Abusive Language
User
Expertise
Domain
Influence,
Popularity
Reputation
Trustworthiness
Bot/Trolls/Cyborg/Sybil/Content Polluters/ Social Spambots, etc.
Veracity (always truthful)
Astroturf (non genuine support)
Biased

FIGURE 6. Generic social media credibility framework's high-level component diagram: Constructs are intentionally omitted from the picture for simplicity and understanding.

each one of them only covers some aspect of credibility, and the majority are left undiscovered.

One important objective of the study was to fill the specified gap and propose a theoretical framework with a similar approach followed in many similar studies like [51], [59], [142], [153], [162].

B. FINDINGS

Investigating and exploring the credibility studies found in different fields, like psychology, communication, and information sciences, etc. identified extremely important credibility constructs under the dimensions and levels. There were some critical constructs also identified which were completely missing in many credibility assessment studies. Therefore challenge of credibility assessment was unresolved. Many research studies are now considered under these constructs. Following are some example studies considered under their respective constructs: Hyper-partisan, Hate speech & offensive language, and Smear campaign which are considered under post quality's constructs Bias/Objectivity, Plausibility, Deception/Truthful respectively. Similarly some other studies like malicious profiles (bots, cyborgs, Sybils, etc), and astroturf (non-genuine support) which are considered under user trustworthiness constructs Deception/Truthful, Truthful respectively (see table 14).

It is also discovered that credibility is composed of many constructs, which are identified in section III-C2. All these constructs must be considered in assessment instead of considering only one or two. Majority of earlier credibility assessment studies only consider one or two constructs, like relevance, deception, truthful, popularity. Only these some construct were mostly considered in majority of the studies in isolated manner and remaining all were ignored.

C. CONSTRUCTS FOR SOCIAL MEDIA AND MICROBLOGS

The proposed framework is comprised of specific credibility levels, dimensions, and constructs and simply presenting their relationships. Credibility levels and dimensions identified in section III are general to social media credibility and therefore could serve as standard social media credibility framework, regardless of a very specific information object, whereas constructs will be information object-specific means both social media and microblogs specific [68]. Therefore in this section, such important constructs will be identified, and then a proposed social media framework will be presented. Distinct social media and microblogs characteristics are discussed in section VI. Important list of constructs are selected from table 2 and 3 considering the specified characteristics and presented as following. The same list of constructs was already shortlisted in section III-C2 with the same preferences.

The list is presented again for easy reference. These constructs together with associated aspects are also shown in the table 14, presenting high-level credibility framework:

1. Recency, 2. Truthful, 3. Deception, 4. Topic, 5. Specificity, 6. Unbiased/Objectivity, 7. Popularity, 8. Plausibility, 9. Authority/Influence, 10. Competence/ Reputation, 11. Uniqueness/ Completeness, etc.

Finally to complete our proposed generic social media-based credibility framework. Specific aspects/ characteristics elaborating each construct are also presented.

Considering third/ second last column of table 14. All constructs are specified in bold and aspects are written adjacent

TABLE 14. Proposed high-level generic credibility framework for social media: presenting relationships between credibility levels, dimensions, constructs, and aspects.

High-level Credibility Framework	**Dimensions (Level)**	**Constructs:** Aspects	**Descriptions**
	Quality (Post)	**Deception, Truthful:** Rumor and Fake	Misinformation, Disinformation, Hoaxes, etc.
		Uniqueness/Completeness: Informative. **Popularity:** Genuinely Popular **Recency:** Recency	General quality related attributes (Informative, Recent, etc.). Popularity must be clean from Bot manipulations.
		Deception, Truthful: Spam & Scams	Phishing, click-bait, Political Astroturf Meme, Fake Reviews, etc.
		Unbiased: Hyper-partisan/ Biasness	Polarization, etc.
		Deception, Truthful: Smear campaign (damage reputation)	Satire, Meme, Propaganda, Conspiracy, etc.
		Plausibility: Hate Speech, Offensive & Abusive Language	All types of Hates: Ethnic based, Xenophobia, Islamophobia, racism, misogyny, etc.
	Expertise (User)	**Competence/Topic/Specificity:** Domain	Top three areas in which he message, e.g: Politics, Sports, Health, etc.
		Authority: Influence. **Popularity:** Popularity. **Competence:** Reputation	Different measures of Centrality, Authority Transfer, User Defined Ratios & Influence
	Trustworthiness (User)	**Deception, Truthful:** Bot/Trolls/Cyborg/ Sybil/Content Polluters/Social Spambots, etc.	Fake A/Cs, Non Human Behavior, etc.
		Truthful: Veracity (always truthful)	Not fake and rumor producer/propagator, don't like them as well.
		Deception: Astroturf (non genuine support)	Followers fallacy, Bot Nets, Troll Factories/ Troll Farm, Link Farming, etc.
		Unbiased: Biased	If greater no of Hyper-partisan posts found

to the constructs. For example considering the first line, Deception, Truthful (constructs): Rumor and Fake (aspects). It simply means that 'Deception, Truthful' constructs could be implemented through 'Rumor' detection and 'Fake' detection. These 'Rumor' and 'Fake' are aspects, which need to be implemented for fulfilling respective constructs (e.g.: Deception, Truthful). All other remaining aspects are specified under their constructs, in the same manner.

The framework presents all components like Levels, Dimensions, Constructs, and related Aspects (see the complete framework in the table: 14). This framework will be further transformed for microblogs using microblog specific features, in the last segment of this section.

D. OVERVIEW OF PROPOSED FRAMEWORK

After conducting detailed and organized literature exploration it is proposed, that true Credibility is measured through narrator (user level) and their narration (post level) both (see figure 6). Narrator assessment may be done on its 'Expertise' and 'Trustworthiness' (see dimensions of the user), which are further assessed on multiple bases (see aspects, e.g.: domain, influence, popularity, reputation under 'expertise' dimension of the user). The narrator's 'expertise' could be judged through its genuine 'influence' based only on trustworthy social network context, level of expertise with relevant 'topic/domain', together with his/her 'popularity' and good 'reputation'. The narrator's trustworthiness could be assessed through the following aspects: the narrator should always be 'truthful', must not be 'biased'. The narrator should not behave like malicious profiles (e.g.: Bot/ troll/ cyborg/content polluter, etc.), etc. Similarly, narration may also be assessed on its 'Quality' (see the dimension of post). Quality may have different bases for assessment (see aspects, e.g.: recency, fake & rumor, hate speech, offensive & abusive language, biasness, informative, popular, etc. under the quality dimension of post). The quality of the post could be judged through different aspects: post 'truthfulness', level of 'informativeness', and 'popularity'. Post must also be clear from hate speech, and biasness, etc (see figure 6).

An effort is being made to present a proposed generic credibility framework (see table: 14) for social media. Comprising the levels (see column II - e.g.: Post, User) at which the credibility should be assessed together with respective dimensions which completely adhere to the credibility related research studies and theories (see same column II – 1. Post: Quality. 2. User: Expertise, Trustworthiness) which need to be addressed. Finally what aspects/attributes under specified constructs (see column III – 1. Post Quality: Fake, Spam, Hyper-partisan, etc. 2. User Expertise: Domain, Influence. 3. User Trustworthiness: Bot, Veracity, Biased, etc.) are comprising each construct under the dimensions. Last column (see column IV) of our framework presents related or similar attributes which will be automatically covered if someone just considers the main aspects/attributes presented in column III. It could be noticed that the comprehensive set of aspects/attributes have mostly resulted from a thorough study of a large set of supported researches presented in section V. High-level credibility framework picture is also presented in figure:6. Important terms are also defined in the appendix section of the paper for clarity and understanding.

E. FRAMEWORK MAPPING TO MICROBLOG'S FEATURES

Considering the second objective of the study: the social media generic framework will be transformed to microblog-specific implementation through microblog's specific features. Therefore after presenting the most important baseline of our work as Proposed Social Media Credibility Framework. We are now presenting in the table: 15, that how each aspect/attribute of our social media credibility framework could be implemented over microblogs, through our proposed list of sample features. These features have mostly resulted from a detailed study of researches presented in section V and, section VI. Each feature is then justified by appropriate reference of research (covering a wide range of literature review, two complete sections of the study, section V and, section VI), together with its significance and judgment.

The proposed list of sample features are furnished with two different levels (e.g.: User-Level, Post-Level), network features (e.g.:Friends network's Influence or Rank, Retweet network's Spread and Propagation), aggregated features (e.g.: Reciprocity, Reputation, etc.), and historic features at user level (e.g.: Domain, Veracity, Biased, etc.).

Features presented in table 15 have varying levels of complexity. Few features are very simple and they are known as raw features, e.g.: number of followers, number of followings, age of account, is-verified, number of posts, URL in profile, description in profile, etc. Few features will be computed either through a separately trained machine learning system or by putting some extra effort, like the use of some lexicon, dictionary, etc. Examples of such features are, Bot/Cyborg Likelihood Score, Hate-speech (Y/N), Abusive Language (Y/N), Sentiment Score, Emotion Valance-Arousal- Dominance (VAD) Score, Bias (Y/N), Fake (Y/N), Topic of the Post (e.g.: Politics, Sports, Education, Social Issues, etc.), Psycho-linguistic features calculated through Linguistic Inquiry and Word Count (LIWC) lexicon with following categories: Informality, Cognitive Process, Perceptual Process, and Diversity. Some features could be computed by calling API, e.g.: Web Of Trust (WOT) Score, Informative: Alexa Rank, Likes or Dislikes of YouTube Videos, and Ground Truth Labels for the URL's found in the post. Some features could be computed through standard libraries or self-made programs. This list of features is: 'User Ranks' which could be computed using page rank or modified page rank-like algorithms, or different centrality measures of social network analysis, etc. Other such features are Spread and Propagation features of the post's re-tweet network which could be computed using tree libraries.

XII. RECOMMENDATIONS

For better understand-ability, this section will present all the recommendations as a blueprint, sketch, or glimpse of the real system as if the system should look like this. Our basic recommendations are presented under two tables. Table 11 presenting category-specific common properties which are also found or picked in our proposed solution, whereas table 12 presets completely distinct and new properties which are unique to our recommended solution.

The proposed solution is overcoming all identified shortcomings and further strengthening itself with extra proposed features.

A. GUIDED DATA TAGGING

Data tagging is most important for automatic credibility assessment systems. Following are few serious issues found in these studies. Those are addressed as following:

The required level of reliability needed in labeling the credibility dataset would require a completely different process. Data could not be tagged only based on the evaluator's/expert's perception about the post. It is very challenging to correctly label such multi-perspective data without discovering hidden facts about the post. Data will be tagged in a completely guided environment. Each post will be tagged after various flags indicated by the variety of available tools. All aspects of credibility must also be considered. Expert/evaluator will be indicated about poster's likelihood score of the malicious profile, top 3 domains of the poster, Avg. number of malicious profiles found in poster's friend network, post's WOT score, Alexa rank, Ground Truth labels, etc. if URL is found in the post, etc.

During tagging/scoring, all aspects of credibility must be examined instead of only a few aspects which are mostly examined in most of the studies. The majority of studies either consider only fake/real as credibility. Some consider that only popular, topic expert is representative of credibility, etc.

Evaluators must be given clear guidelines for tagging, like what will be the credibility label/score/rank if the post is posted by a topic expert and the topic of the post is completely matched with the expertise of the poster. What label/score will be assigned to a post that is fake and posted by a malicious profile, etc. What about the post that has extreme bias and suffering from hate speech with abusive language.

In the presence of such indicators with clear guidelines post will be ranked/labeled finally by the expert/evaluator.

B. HYBRID SYSTEM - GRAPH BASED + FEATURE BASED

It is supposed to be a Hybrid system of a different kind. In our proposed solution: The graph-Based method will be executed first on two network-based features. There are two distinct sets of network features based on retweet network, and friends network, presented in Table: 15, at no 1, 46,

TABLE 15. Implementing social media credibility framework to microblog's: following are mapping of social media credibility framework's aspects to proposed microblog's sample features. Transforming the generic framework to micriblogs specific implementation, just need the generic aspects to be transformed to microblog's features.

S.No.	Feature Name	Feature Level	Cred. Framework Aspects	Reference/Reason
1	User Ranks: Influence, other Centrality Scores, etc.	User Level (Friends Network)	Expertise, Quality	Measure of user influence and rank [34]
2	No. of followers	User Level	Bots, Expertise, Trustworthiness, Fake	Too few and too many: less expertise and trustworthiness. Less gap b/w 2 and 3: high competence, Ratio determines nature of A/C e.g.: broadcast, etc. Too many 2 and 3: Bot [120], High rate of friend/followers: Fake post producer [256], significant no of connections: active user [133]
3	No. of friends			
4	Age of a/c		Trustworthiness, Veracity, Expertise	Old A/C: produce less misinformation [256] and more trustworthy, Expertise, Competence [122] and New A/C: produce more misinformation and less trustworthy [256].
5	IsVarified and Protected		Bots, Fake	Verified a/c means real a/c notbot and Fake post producer [256].
6	No. of Tot.Posts		Trustworthiness, Expertise	High no of posts: credible post producer, active user [133]. User posting behavior:tweets/re-tweets [133]
7	URL in Profile (Y/N)		Trustworthiness	User perception based features visible at a glance, if yes then user perceive as credible [118]
8	Desc in Profile, Pic (Y/N)			
9	Bot/Cyborg Likelihood		Bot, Misinformation	Covers many aspects of credibility [23].
10	List Count			Bot: 0 or Very Less [257].
11	Reputation: Followers/Followers + Followings			Bot:0, Human:1; Celebrities and popular org: high, more followers than followings. Bots: More followings than followers [97]
12	Reciprocity: fraction of friends who are also followers (overlap)			Bot: Low, Human: High [97]
13	Default Profile			Mostly non active , new user uses default [257].
14	Domain (Top 3 domains extracted from post of user having topics)		Expertise, Quality	Once expert's domain and tweet topic is matched, fully reflects credibility [113].
15	Hate Speech, Abusive and offensive Language. (Y/N)	Tweet / Post Level	Hate, Quality, Smear	Potentially harmful to specific group/community, could promote violence and social disorder, to humiliate or insult [113]
16	Get Ground Truth Labels for each URL in the Post.		Fake, Satire, Bias, Hate, Rumor, Spam Conspiracy	Varying level of Reliability and Bias labeling, URL's could be used for post identification as: Fake, Satire, Extreme Bias, Conspiracy, Rumor, Click-bait, Hate Group, Junk Science, etc. [179], [258]–[261]
17	Network: #Retweet		Quality, Fake, Rumor	Popularity, symbol of quality, msg endorsement [133], [137]. Considered important [30]. Fake has high retweets. [3], [83]
18	#mentions		Quality, Spam	Considered very important feature [30]. Too many mentions low credibility [122], in emergency also [131]
19	IsReply		Quality	One of some user perception based features visible at a glance, if yes: seems credible [118], it shows that User listen,agree/disagree and validate [256]
20	IsRetweet			
21	No. of Likes		Quality, Fake	Treated as good reputation [137]. Real news has more likes where as Fake has less [83].
22	No. of Replies		Fake, Bot	Bot: Very Less [262], Fake Post: High [20], Less [83]
23	Links: No. of URL		Fake	URL presence: High Credible [30], [133], Fake posts: large no of URLs [256]
24	WOT Score for URLs		Fake, Spam	Site reputation Score: Low score bad reputation [135] and spam, etc.Internet Trust Tool [263]
25	Likes/Dislikes (if YouTube Video(s)), etc.	Tweet/ Post Level	Quality	High values good reputation and credibility
26	Psycho-linguistic (Informality): No. of Swear words/ Netspeak/ Assent/ Non-fluencies/ Fillers/ Typos		Fake, Quality, Spam	Psycho-linguistic LIWC [264](Informality) features. In news: Non Fake [133], [265], Identify type of tweet: Non News. Presence shows bad quality, Presence: Non Spam
27	No. of Self Words(i,my,mine)			Word like "I saw" more credible, Identify tweet type: Non News, Non Spam
28	Pronoun (1st, 2nd, 3rd Person) Present (y/n)			Identify tweet type: Non News, Non Spam
29	Sentiments: Sub/ Obj Score,etc.		Quality, Bias, Fake	Negative Sentiments are more credible in news. Generally either positive/ neutral is credible [133]. Real News: High ratio of neutral replies and Fake: High –ve replies. [83]. Bias Language Corpus [266] High Bias Language: High Fake.
30	Emotion VAD Scores			
31	Language Bias			
32	Text: Length		Quality, Fake	More length : more credible [30], [111], [133]
33	No. of words			
34	Fraction of upper case letters		Spam	High fraction leads to spam [160]

TABLE 15. *(Continued.)* **Implementing social media credibility framework to microblog's: following are mapping of social media credibility framework's aspects to proposed microblog's sample features. Transforming the generic framework to micriblogs specific implementation, just need the generic aspects to be transformed to microblog's features.**

35	No. of Hashtags		Spam	3 or more are considered as spam [160]
36	?, !, Stock Symbol ($) (Y/N). Contain multiple ?, ! (Y/N).		Fake, Quality, Spam	Identify tweet type: Non News, Non Spam (completely varying behavior in different aspects)
37	Smile icon, frown icon (:(), etc. (y/n)			
38	MetaData: Age(sec)		Quality	Capture all time dependent aspects
39	Day of the Week			
40	Source (API 3rd Party, Un-Reg., mob, web)		Bot, Trustworthiness	Human: Web/Mob. Bot: API 3rd Party [97]. Source as Mobile is more credible.
41	IsGeo-Coordinates (Y/N), etc.		Fake, Quality	Represent Location: More credible [135]
42	Fake: Yes/No		Fake, Misinformation	Used for assessing truthiness, controls misinformation. 200 Fact Checking Web Sites [267]
43	Topic: Politics, Health, Sports, Education, etc.		Expertise, Quality, Misinformation	If user's domain and tweet topic is matched, fully reflects credibility [113]. Some Topics are less credible [31], [130]. Misinformation is more diffused in some topics [175], [268].
44	Informative: Alexa Rank		Quality	High Rank means informative and credible.
45	Psycho-linguistic/LIWC: Cognitive Process, Perceptual Process and Diversity		Fake, Misinformation, Quality	Different classes of attributes are identified in LIWC [264] to identify Fake [265] .
46	Spread: Level No., No. of RTs at each level (apply spread model)	Tweet/ Post Level (Retweet Network)	Fake, Misinformation	High spread and propagation lies in fake news, misinformation, etc. [20], [81], [269]. Very specific patterns are found in majority of Misinformation type contents within the Retweet Network [101], [171].
47	Propagation: Root Degree, Max Subtree, Avg. Subtree, Tree Max Degree and Avg. Degree (excluding root), Tree Max Depth, Avg. Depth			

and 47. Using friends network, where malicious profiles/bots will be eliminated, and the influence scores for each user will be calculated and saved as User level feature (see feature no: 1 in table 15). It is an extremely important aspect that is completely ignored that the credibility of a message can't be determined without going into the underlying credible and trustworthy friend's network, to measure the correct influence of the user. If malicious profiles exist in the friend's network then they must be omitted before examining the user rank/influence. Malicious profiles/bots identification and their rectification must be done before the credibility assessment initialization, to prevent their serious manipulations at various places. Likewise, using tweet-retweet propagation network, in which all malicious profiles/bot will be eliminated and then spread and propagation scores will be calculated and saved as tweet feature (see feature no: 46 and 47 in table 15). The spread and propagation pattern of the message is an important indicator of credibility assessment. After calculating all user-level features and tweet-level features, different machine learning models could be executed over these features for the prediction of post label/ rank / score. Therefore our model is following a hybrid approach combining both graph-based methods and feature-based methods.

C. POST CREDIBILITY SCORE

It could easily be observed that our proposed list of features completely covers all quality-related aspects of a tweet. These quality-related aspects were mostly missed from the majority of studies in the literature. After calculating all user-level features and tweet level features (see the recommended list of features in table 15) either any conventional Machine Learning Regression model (e.g.: Gradient Boosting, Ada-Boost, CAT Boost, LightGBM, SVM, Random Forest, Linear Regression, etc.) or any modern Learn to Rank model (e.g.: Lambda Rank, SVM Rank, Lambda Mart, etc.) could be executed to predict tweet's credibility score.

D. USER LEVEL SCORES

Referring to our recommended solution, in addition to the basic tweet credibility score, different user level scores could easily be calculated based on the tweets of the user. It has been discussed earlier that many post-level features could compute user-level features. Examples of such User level scores are as follows. Computation of all such scores at the user level will implicitly reduce the dissemination of low-credibility contents, over microblogs.

1. User %age of fake produced and propagated: which will be a historic feature computed through no of fake tweets produced or propagated by that user.

2. User Avg. Spread and Propagation Scores: which will also be historic features, computed through avg. of all tweets spread and propagation scores of the user.

3. User Avg. Credibility Score: similarly User Avg. Credibility Score will be calculated by taking avg. of credibility score of all tweets of that user.

4. User Top 3 Domains: it could also be computed through all tweet topics tweeted by that user and the top 3 could be accumulated.

E. SCORES CONVERGENCE

Above all user-level, scores and post-level scores could easily be calculated in real-time and displayed at respective entity levels. The chain of narrators is extremely important in assessing the message's credibility. Once a post is identified as fake then its producer must be penalized by incrementing its fake producer counter. Similarly, each fake propagator involved in post propagation within the post's chain of narrators must also be updated. It is worth mentioning that every post's credibility score will affect the respective user-level score and user score will also be affected by its post credibility. For example tweet's final credibility score will only be accumulated through all its chain of narrators and vice versa.

XIII. FUTURE RESEARCH DIRECTIONS

There is a need for benchmark/gold-standard credibility dataset construction. The dataset will include different forms of deceptions [33], like rumor, fake news, spam & scam, hoax, click-bait, junk science, conspiracy, and different forms of smear campaigns, etc. The dataset must also be enriched with hate speech, with its related concepts like abusive language, offensive language, general hate, cyberbullying, discrimination, flaming, harassment, profanity, toxic language or comment, extremism, radicalization, etc. [14]. There should also be sufficient malicious profiles (e.g.: Bots/Cyborgs, etc.). It must contain a good mix of news and non-news pieces of information. The dataset tagging should be done exactly in the way which is presented in section XII's heading 'Guided Data Tagging'. Regarding the features of the dataset. The following necessary features must be included for credibility assessment. The dataset must have a three-degree friends network (followers/following directed graph), user profiles, complete tweets of all users involved in the datasets with the number of replies & number favorites & who has favorited, etc., in addition to actual tweets which will be considered for credibility assessment. Actual and complete tweet-retweet multi-level propagation network (generally Twitter API provides flat retweeter's list), information of the list/ groups, media files, etc. The dataset's post should have a balanced number of domains e.g.: Politics, Entertainment, Sports, Education, etc. The dataset should also be developed through multiple microblogs and in different languages.

There are many challenges involved in the development of such a dataset because of accessibility privileges, the huge amount of data collection and management, strict tagging requirements, etc. Fortunately, there are few components of such dataset that are already available (see section VII) that need to be compiled concerning credibility, and missing components will be added.

In addition to the real-world labeled dataset. We need to implement the recommended system presented in this study, for its efficacy and performance evaluation.

After the necessary understanding of information credibility for microblogs presented through this study. There is a need to explore the literature regarding information credibility using multi-modal data and, explainable credibility assessment methods. It is very important that whatever credibility assessment is done by the system needs to be explained, that how the contents are categorized as not-credible or credible. Similarly, credibility assessment should make use of voice, image, and video from the post, in addition to text.

Regarding the challenges and limitations, which are presented in different sections of the study therefore not discussed separately.

XIV. CONCLUSION

An effort of presenting the anatomy of information credibility for social media and microblogs was made, through a detailed and, organized study. Many research studies were conducted to assess automatic microblog's credibility but the majority of them had different concepts of credibility. Credibility is multi-disciplinary, hence there was no generalized or accepted credibility concept with all its necessary and detailed constructs/components. Therefore, it was necessary to understand the complete concept of information credibility from different disciplines. It could be accomplished through an organized study of all the problem dimensions and identification of comprehensive and necessary credibility constructs under credibility's definition. Such literature exploration and the fundamental study was missing regarding the work done. Therefore to consolidate, standardized, identify gaps, propose solutions and recommendations in this area. We deeply explore the existing literature first, categories them along various dimensions, identify gaps and shortcomings then suggest important recommendations. As a result of a successful explorational study, a complete information credibility framework for social media is proposed. It is the first framework considering all necessary constructs of credibility identified in this study. Afterward, the presented framework is also transformed for microblogs credibility assessment. The transformation is done to individual features level for understanding and clarity. Therefore the framework can simply be implemented as a successful system. Another important aspect which we noticed missing in previous researches and therefore proposed, that Credibility should be measured through the narrators and narrations both, considering their important aspects or bases of assessments. The narrator's assessment should be done on multiple bases such as its genuine social network influence, should always be truthful and unbiased, its area of expertise, popularity, and good reputation, etc. Similarly, narration could be assessed on its quality basis like it must be true, clean from spam & scams, rumors, and smear campaigns, etc. It should be informative, clear from the variety of hate speeches, and extreme biases, etc. Our credibility framework is based on both user and post. Which could provide two-fold benefits: information credibility ratings as well as user credibility ratings. Later credibility (user credibility ratings) will be extremely helpful in other applications for example to assess the reviews of credible authors, considerations of credible user's recommendations, etc.

APPENDIX A TERMS DEFINED

A. CLAIM

Un-verified piece of news/ article/ information/ opinion in question, which could be rumor, hoax, satire, and fake news, etc.

B. FACT-CHECKING

Process of claim evaluation through authentic publish media, journalists, and domain experts, etc., and resulted as Fake, Real, etc.

C. SATIRE

is characterized by humor, irony, absurdity, exaggeration, and ridicule. They can mimic genuine news, primarily written to criticize.

D. HOAX

Deliberately fabricated falsehood made to masquerade as truth, intentionally conceived to deceive readers.

E. PROPAGANDA

Information that tries to influence the emotion, the opinions, and the actions of target audiences through deceptive, selectively omitting, and one-sided messages. The purpose could be political, ideological, or religious, etc.

F. RUMOR

Claim that has not been verified (may be true or false), apparently credible but hard to verify and spread from one person to another.

G. CLICK-BAIT

Low-quality journalism intended to attract traffic and monetize via advertising revenue.

H. MEME

A piece of information that replicates among people (Dawkins 1989). It bears similarities to infectious diseases, as both travel through social ties from one person to another. Piece of information mostly spread widely on the internet, often altered for humorous effect. Meme types are hashtags, URLs, Mentions, and Phrases.

I. ASTROTURFING

A particular type of abuse disguised as spontaneous ''grassroots'' behavior, but that is in reality carried out by a single person or organization. Non-genuine public support of an issue. Quiet related to spam.

J. SYBIL'S

Suspicious accounts, no malicious contents are posted, creating many fake identities to unfairly increase the power or influence of someone, therefore, produce a false sense of credibility. This concept is called Link Farming. Some similar terms to Sybil's are also popular e.g.: Sockpuppet, Zombie Followers, and Fake followers, etc.

K. BOTS, TROLLS, CYBORG

''Bots'' are fully automated accounts and completely distinct from professional ''trolls'', which are human-run accounts, and the ''Cyborg'' accounts which combine human-generated content with automated posting.

L. BOTNETS

connected bots network.

M. SOCIAL SPAMBOTS

More sophisticated bots, mimic human-like behavior.

N. SPAMBOTS/CONTENT POLLUTERS

Traditional and simple type of bots, e.g.: Duplicate Spammers, Malicious Promoters, Self-Promoters, Friend Infiltrator, etc.

O. COORDINATED BEHAVIOR

Chain of users which are developed to perform some pre-defined task of their master (example of pre-defined task could be: always like the post, add specific hashtag and mention, then forward post to others).

P. FOLLOWERS FALLACY

Users with manipulated followers count. These untrustworthy users use bot activities to increase followers count for having high influence, popularity, or reputation. There are different ways, like online black-market services, they help the users to increase their followers/likes. Users can purchase bulk followers and likes from these markets. Users exploit such services to inflate followers, likes, and shares of the post to become more influential and popular.

Q. EXTREME BIAS

Piece of information come from a particular point of view and may rely on propaganda, decontextualized information, and opinions distorted as facts.

R. LINGUISTIC INQUIRY AND WORD COUNT (LIWC)

Psycholinguistic features are very important in credibility analysis through text, which could be computed by LIWC. It is a text analysis lexicon and a program that calculates the percentage of words in a given text that fall into one or more of over 80 linguistic, psychological, and topical categories indicating various social, cognitive, and affective processes. i.e.: the word 'cried' is part of four-word categories: sadness, negative emotion, overall affect, and a past tense verb.

REFERENCES

[1] E. Shearer and B. J. Gottfried, "News use across social media platforms 2017," Pew Res. Center, Washington, DC, USA, Tech. Rep. 202.419.4372, 2017, p. 17.

[2] A. Perrin, "Social media usage," Pew Res. Center, Washington, DC, USA, Tech. Rep. 202.419.4372, 2015, pp. 52–68.

[3] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: Can we trust what we RT?" in *Proc. 1st Workshop Social Media Anal. (SOMA)*, 2010, pp. 71–79.

[4] E. J. Briscoe, D. S. Appling, and H. Hayes, "Social network derived credibility," in *Recommendation and Search in Social Networks*. Cham, Switzerland: Springer, 2015, pp. 59–75.

[5] A. Java, X. Song, T. Finin, and B. Tseng, "Why we Twitter: Understanding microblogging usage and communities," in *Proc. 9th WebKDD 1st SNA-KDD Workshop Web Mining Social Netw. Anal. (WebKDD/SNA-KDD)*, 2007, pp. 56–65.

[6] A. Lenhart. *Teens, Privacy and Online Social Networks. How Teens Manage Their Online Identities in the Age of Myspace, Pew Internet & American Life Project Report 2007*. Accessed: Aug. 12, 2021. [Online]. Available: http://www.pewinternet.org/PPF/r/211/report_display.asp

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

[8] A. Hermida, "Twittering the news: The emergence of ambient journalism," *J. Pract.*, vol. 4, no. 3, pp. 297–308, 2010.

[9] S. Laird, "How social media is taking over the news industry," Apr. 2012. [Online]. Available: http://mashable.com/2012/04/18/social-media-and-the-news/

[10] N. Newman, R. Fletcher, A. Kalogeropoulos, D. Levy, and R.-K. Nielsen, *Reuters Institute Digital News Report 2017*. Oxford, U.K.: Univ. Oxford, Reuters Institute, 2017.

[11] D. Acemoğlu, A. Ozdaglar, and A. ParandehGheibi, "Spread of (mis) information in social networks," *Games Econ. Behav.*, vol. 70, no. 2, pp. 194–227, 2010.

[12] P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities," *Expert Syst. Appl.*, vol. 153, Sep. 2020, Art. no. 112986.

[13] B. Collins, D. T. Hoang, N. T. Nguyen, and D. Hwang, "Trends in combating fake news on social media—A survey," *J. Inf. Telecommun.*, vol. 5, no. 2, pp. 247–266, 2021.

[14] P. Fortuna and S. Nunes, "A survey on automatic detection of hate speech in text," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–30, 2018.

[15] K. A. Qureshi and M. Sabih, "Un-compromised credibility: Social media based multi-class hate speech classification for text," *IEEE Access*, vol. 9, pp. 109465–109477, 2021.

[16] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, and F. Menczer, "Detecting and tracking the spread of astroturf memes in microblog streams," 2010, *arXiv:1011.3768*. [Online]. Available: http://arxiv.org/abs/1011.3768

[17] K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, Jan. 2019, pp. 312–320.

[18] M. Latah, "The art of social bots: A review and a refined taxonomy," 2019, *arXiv:1905.03240*. [Online]. Available: http://arxiv.org/abs/1905.03240

[19] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in Twitter: The million follower fallacy," in *Proc. 4th Int. AAAI Conf. Weblogs Social Media*, 2010, pp. 1–8.

[20] C. Shao, P.-M. Hui, L. Wang, X. Jiang, A. Flammini, F. Menczer, and G. L. Ciampaglia, "Anatomy of an online misinformation network," *PLoS ONE*, vol. 13, no. 4, Apr. 2018, Art. no. e0196087.

[21] 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.

[22] S. Y. Rieh, M. R. Morris, M. J. Metzger, H. Francke, and G. Y. Jeon, "Credibility perceptions of content contributors and consumers in social media," *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 51, no. 1, pp. 1–4, 2014.

[23] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," in *Proc. 11th Int. AAAI Conf. Web Social Media*, 2017, pp. 1–10.

[24] C. Grimme, M. Preuss, L. Adam, and H. Trautmann, "Social bots: Human-like by means of human control?" *Big Data*, vol. 5, no. 4, pp. 279–293, Dec. 2017.

[25] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 36–211, 2017.

[26] M. Hindman and V. Barash, "Disinformation, and influence campaigns on Twitter," Knight Found., Miami, FL, USA, Tech. Rep. 238, 2018.

[27] R. Faris, H. Roberts, B. Etling, N. Bourassa, E. Zuckerman, and Y. Benkler, "Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election," Berkman Klein Center Internet Soc. Harvard Univ. Res. Paper, Aug. 2017, vol. 6. [Online]. Available: https://dash.harvard.edu/handle/1/33759251

[28] A. Bovet and H. A. Makse, "Influence of fake news in Twitter during the 2016 U.S. Presidential election," *Nature Commun.*, vol. 10, no. 1, pp. 1–14, Dec. 2019.

[29] M. Viviani and G. Pasi, "Credibility in social media: Opinions, news, and health information—A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 7, no. 5, p. e1209, Sep. 2017.

[30] J. Odonovan, 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.*, Sep. 2012, pp. 293–301.

[31] A. Gün and P. Karagöz, "A hybrid approach for credibility detection in Twitter," in *Proc. Int. Conf. Hybrid Artif. Intell. Syst.* Cham, Switzerland: Springer, 2014, pp. 515–526.

[32] S. Kumar, F. Morstatter, and H. Liu, *Twitter Data Analytics*. New York, NY, USA: Springer, 2014.

[33] E. Aïmeur, H. Hage, and S. Amri, "The scourge of online deception in social networks," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2018, pp. 1266–1271.

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

[35] W. Sherchan, S. Nepal, and C. Paris, "A survey of trust in social networks," *ACM Comput. Surv.*, vol. 45, no. 4, pp. 1–33, 2013.

[36] E. Ciceri, R. Fedorov, E. Umuhoza, M. Brambilla, and P. Fraternali, "Assessing online media content trustworthiness, relevance and influence: An introductory survey," in *Proc. KDWeb*, 2015, pp. 29–40.

[37] J. An, W. J. Li, L. N. Ji, and F. Wang, "A survey on information credibility on Twitter," *Appl. Mech. Mater.*, vols. 401–403, pp. 1788–1791, Sep. 2013.

[38] M. Alrubaian, M. Al-Qurishi, A. Alamri, M. Al-Rakhami, M. M. Hassan, and G. Fortino, "Credibility in online social networks: A survey," *IEEE Access*, vol. 7, pp. 2828–2855, 2019.

[39] C. C. Self, "Credibility," in *An Integrated Approach to Communication Theory and Research*, M. B. Salwen and D. W. Stacks, Eds. Mahwah, NJ, USA: Lawrence Erlbaum," 1996.

[40] *Credibility Definition Oxford*. Accessed: Aug. 12, 2021. [Online]. Available: https://en.oxforddictionaries.com/definition/ credibility

[41] *Credibility Definition Merriam Webster*. Accessed: Aug. 12, 2021. [Online]. Available: https://www.merriam-webster.com/dictionary/credibility

[42] P. J. Kalbfleisch, "Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment," in *Communication Yearbook*. Evanston, IL, USA: Routledge, 2003, pp. 307–350.

[43] B. J. Fogg, "Prominence-interpretation theory: Explaining how people assess credibility online," in *Proc. CHI Extended Abstr. Hum. Factors Comput. Syst. (CHI)*, 2003, pp. 722–723.

[44] W. Choi and B. Stvilia, "Web credibility assessment: Conceptualization, operationalization, variability, and models," *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 12, pp. 2399–2414, Dec. 2015.

[45] S. Y. Rieh, "Credibility and cognitive authority of information," in *Encyclopedia of Library and Information Sciences*, 3rd ed., M. Bates and M. N. Maack, Eds. New York, NY, USA: Taylor & Francis, 2010, pp. 1337–1344.

[46] B. J. Fogg, "Persuasive technology: Using computers to change what we think and do," *Ubiquity*, vol. 2002, p. 2, Dec. 2002.

[47] S. Tseng and B. J. Fogg, "Credibility and computing technology," *Commun. ACM*, vol. 42, no. 5, pp. 39–44, May 1999.

[48] B. Cugelman, M. Thelwall, and P. Dawes, "Website credibility, active trust and behavioural intent," in *Proc. Int. Conf. Persuasive Technol.* Berlin, Germany: Springer, 2008, pp. 47–57.

[49] R. L. Wakefield and D. Whitten, "Examining user perceptions of third-party organizations credibility and trust in an E-retailer," *J. Org. End User Comput.*, vol. 18, no. 2, pp. 1–19, Apr. 2006.

[50] C. Sichtmann, "An analysis of antecedents and consequences of trust in a corporate brand," *Eur. J. Marketing*, vol. 41, nos. 9–10, pp. 999–1015, Sep. 2007.

[51] G. Sahut and A. Tricot, "Wikipedia: An opportunity to rethink the links between sources-credibility, trust and authority," *1st Monday*, vol. 22, no. 11, pp. 1–32, 2017.

[52] B. W. Roper, *Public Attitudes Toward Television and Other Media in a Time of Change: The 14. Report in a Series by the Roper Organization Inc*. New York, NY, USA: Television Information Office, 1985.

[53] M. J. Metzger, A. J. Flanagin, K. Eyal, D. R. Lemus, and R. M. Mccann, "Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment," *Ann. Int. Commun. Assoc.*, vol. 27, no. 1, pp. 293–335, Jan. 2003.

[54] T. J. Johnson and B. K. Kaye, "Cruising is believing?: Comparing internet and traditional sources on media credibility measures," *J. Mass Commun. Quart.*, vol. 75, no. 2, pp. 325–340, Jun. 1998.

[55] J. Mashek, L. McGill, and A. C. Powell, *Lethargy'96: How the Media Covered a Listless Campaign*. Washington, DC, USA: Freedom Forum First Amendment Center at Vanderbilt Univ., 1997.

[56] N. Roberts, "Second-hand knowledge: An inquiry into cognitive authority," *Social Sci. Inf. Stud.*, vol. 5, no. 3, p. 147, Jul. 1985.

[57] S. Kim, "Questioners' credibility judgments of answers in a social question and answer site," *Inf. Res.*, vol. 15, no. 2, pp. 2–15, 2010.

[58] S. Y. Rieh, "Judgment of information quality and cognitive authority in the web," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 53, no. 2, pp. 145–161, 2002.

[59] C. N. Wathen and J. Burkell, "Believe it or not: Factors influencing credibility on the web," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 53, no. 2, pp. 134–144, 2002.

[60] S. S. Sundar, "Technology and credibility: Cognitive heuristics cued by modality, agency, interactivity and navigability," in *Digital Media, Youth, and Credibility* (MacArthur Foundation Series on Digital Media and Learning), M. Metzger and A. Flanagin, Eds. Cambridge, MA, USA: MIT Press, 2007, pp. 73–100.

[61] R. E. Petty and J. T. Cacioppo, "The elaboration likelihood model of persuasion," in *Communication and Persuasion*. New York, NY, USA: Springer, 1986, pp. 1–24.

[62] S. Chaiken, "The heuristic model of persuasion," in *Proc. Social Influence, Ontario Symp.*, vol. 5, 1987, pp. 3–39.

[63] R. M. Shiffrin and W. Schneider, "Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory," *Psychol. Rev.*, vol. 84, no. 2, p. 127, 1977.

[64] J. B. Walther, "Interpersonal effects in computer-mediated interaction: A relational perspective," *Commun. Res.*, vol. 19, no. 1, pp. 52–90, 1992.

[65] J. B. Walther, "Social information processing theory," in *Engaging Theories in Interpersonal Communication: Multiple Perspectives*, 2nd ed., D. O. Braithwaite and P. Schrodt, Eds. Newbury Park, CA, USA: Sage, Oct. 2014, ch. 29, pp. 391–404, doi: 10.4135/9781483329529.n29.

[66] M. J. Metzger, "Making sense of credibility on the web: Models for evaluating online information and recommendations for future research," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 13, pp. 2078–2091, 2007.

[67] B. Hilligoss and S. Y. Rieh, "Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context," *Inf. Process. Manage.*, vol. 44, no. 4, pp. 1467–1484, Jul. 2008.

[68] S. Y. Rieh, Y.-M. Kim, J. Y. Yang, and B. St. Jean, "A diary study of credibility assessment in everyday life information activities on the web: Preliminary findings," *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 47, no. 1, pp. 1–10, Nov. 2010.

[69] A. Friggeri, L. Adamic, D. Eckles, and J. Cheng, "Rumor cascades," in *Proc. 8th Int. AAAI Conf. Weblogs Social Media*, 2014, pp. 1–10.

[70] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 1103–1108.

[71] S. Vosoughi, M. N. Mohsenvand, and D. Roy, "Rumor gauge: Predicting the veracity of rumors on Twitter," *ACM Trans. Knowl. Discovery Data*, vol. 11, no. 4, pp. 1–36, Jul. 2017.

[72] E. Seo, P. Mohapatra, and T. Abdelzaher, "Identifying rumors and their sources in social networks," *Proc. SPIE*, vol. 8389, May 2012, Art. no. 83891I.

[73] 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.* Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 1589–1599.

[74] S. Vosoughi, "Automatic detection and verification of rumors on Twitter," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2015.

[75] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," in *Proc. 25th Int. Joint Conf. Artif. Intell. (IJCAI)*. New York, NY, USA: AAAI Press, 2016, pp. 3818–3824. [Online]. Available: https://ink.library.smu.edu.sg/sis_research/4630

[76] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, and P. Tolmie, "Analysing how people orient to and spread rumours in social media by looking at conversational threads," *PLoS ONE*, vol. 11, no. 3, Mar. 2016, Art. no. e0150989.

[77] L. Sydell, "We tracked down a fake-news creator in the suburbs. Here's what we learned," Nat. Public Radio, CA, USA, Tech. Rep. 503146770, 2016, vol. 23. [Online]. Available: http://n.pr/2nuHN1T

[78] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," *Nature Commun.*, vol. 9, no. 1, pp. 1–9, Dec. 2018.

[79] A. Kucharski, "Study epidemiology of fake news," *Nature*, vol. 540, no. 7634, p. 525, 2016.

[80] E. Beğenilmiş and S. Uskudarli, "Organized behavior classification of tweet sets using supervised learning methods," in *Proc. 8th Int. Conf. Web Intell., Mining Semantics*, Jun. 2018, pp. 1–9.

[81] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, pp. 1146–1151, May 2018.

[82] S. Kumar and N. Shah, "False information on web and social media: A survey," 2018, *arXiv:1804.08559*. [Online]. Available: http://arxiv.org/abs/1804.08559

[83] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context and spatial-temporal information for studying fake news on social media," 2018, *arXiv:1809.01286*. [Online]. Available: http://arxiv.org/abs/1809.01286

[84] B. Ghanem, P. Rosso, and F. Rangel, "Stance detection in fake news a combined feature representation," in *Proc. 1st Workshop Fact Extraction Verification (FEVER)*, 2018, pp. 66–71.

[85] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, and F. Menczer, "Truthy: Mapping the spread of astroturf in microblog streams," in *Proc. 20th Int. Conf. Companion World Wide Web (WWW)*, 2011, pp. 249–252.

[86] J. Ratkiewicz, M. D. Conover, M. Meiss, B. Gonçalves, A. Flammini, and M. F. Menczer, "Detecting and tracking political abuse in social media," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 1–8.

[87] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on Twitter," in *Proc. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf. (CEAS)*, vol. 6, 2010, p. 12.

[88] S. Chhabra, A. Aggarwal, F. Benevenuto, and P. Kumaraguru, "Phi.sh/$oCial: The phishing landscape through short URLs," in *Proc. 8th Annu. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf.*, 2011, pp. 92–101.

[89] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: The underground on 140 characters or less," in *Proc. 17th ACM Conf. Comput. Commun. Secur.*, 2010, pp. 27–37.

[90] S. Yardi, D. Romero, G. Schoenebeck, and D. Boyd, "Detecting spam in a Twitter network," *1st Monday*, vol. 15, no. 1, Dec. 2010, doi: 10.5210/fm.v15i1.2793.

[91] 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. (SIGIR)*, 2012, pp. 575–590.

[92] R. Yeniterzi and J. Callan, "Constructing effective and efficient topic-specific authority networks for expert finding in social media," in *Proc. 1st Int. Workshop Social Media Retr. Anal. (SoMeRA)*, 2014, pp. 45–50.

[93] S. Adali, F. Sisenda, and M. Magdon-Ismail, "Actions speak as loud as words: Predicting relationships from social behavior data," in *Proc. 21st Int. Conf. World Wide Web (WWW)*, 2012, pp. 689–698.

[94] M. Jiang, P. Cui, and C. Faloutsos, "Suspicious behavior detection: Current trends and future directions," *IEEE Intell. Syst.*, vol. 31, no. 1, pp. 31–39, Jan. 2016.

[95] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Social fingerprinting: Detection of spambot groups through DNA-inspired behavioral modeling," *IEEE Trans. Dependable Secure Comput.*, vol. 15, no. 4, pp. 561–576, Jul. 2018.

[96] M. Jiang, P. Cui, A. Beutel, C. Faloutsos, and S. Yang, "Catching synchronized behaviors in large networks: A graph mining approach," *ACM Trans. Knowl. Discovery From Data*, vol. 10, no. 4, pp. 1–27, Jul. 2016.

[97] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?" *IEEE Trans. Dependable Secure Comput.*, vol. 9, no. 6, pp. 811–824, Nov. 2012.

[98] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," *Nature Commun.*, vol. 9, no. 1, pp. 1–9, Dec. 2018.

[99] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "BotOrNot: A system to evaluate social bots," in *Proc. 25th Int. Conf. Companion World Wide Web (WWW Companion)*, 2016, pp. 273–274.

[100] K. Lee, B. D. Eoff, and J. Caverlee, "Seven months with the devils: A long-term study of content polluters on Twitter," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 185–192.

[101] L. G. Stewart, A. Arif, and K. Starbird, "Examining trolls and polarization with a retweet network," in *Proc. ACM WSDM, Workshop Misinformation Misbehavior Mining Web*, 2018, pp. 90–96.

[102] P. Galán-García, J. G. D. L. Puerta, C. L. Gómez, I. Santos, and G. P. Bringas, "Supervised machine learning for the detection of troll profiles in Twitter social network: Application to a real case of cyberbullying," *Log. J. IGPL*, vol. 24, no. 1, pp. 42–53, 2016.

[103] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: Quantifying influence on Twitter," in *Proc. 4th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2011, pp. 65–74.

[104] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM SIGMOD Rec.*, vol. 42, no. 2, pp. 17–28, 2013.

[105] M. Farajtabar, Y. Wang, M. Gomez-Rodriguez, S. Li, H. Zha, and L. Song, "Coevolve: A joint point process model for information diffusion and network evolution," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 1305–1353, 2017.

[106] P.-M. Hui, C. Shao, A. Flammini, F. Menczer, and G. L. Ciampaglia, "The Hoaxy misinformation and fact-checking diffusion network," in *Proc. 12th Int. AAAI Conf. Web Social Media*, 2018, pp. 1–10.

[107] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust," in *Proc. 13th Conf. World Wide Web (WWW)*, 2004, pp. 403–412.

[108] M. Jamali and M. Ester, "TrustWalker: A random walk model for combining trust-based and item-based recommendation," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2009, pp. 397–406.

[109] W. Feng and J. Wang, "Retweet or not? Personalized tweet re-ranking," in *Proc. 6th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2013, pp. 577–586.

[110] S. Ravikumar, R. Balakrishnan, and S. Kambhampati, "Ranking tweets considering trust and relevance," in *Proc. 9th Int. Workshop Inf. Integr. Web (IIWeb)*, 2012, pp. 1–4.

[111] Y. Duan, L. Jiang, T. Qin, M. Zhou, and H.-Y. Shum, "An empirical study on learning to rank of tweets," in *Proc. 23rd Int. Conf. Comput. Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 295–303.

[112] H. Huang, A. Zubiaga, H. Ji, H. Deng, D. Wang, H. Le, T. Abdelzaher, J. Han, A. Leung, J. Hancock, and C. Voss, "Tweet ranking based on heterogeneous networks," in *Proc. COLING*, 2012, pp. 1239–1256.

[113] 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*, May 2015, pp. 953–958.

[114] S. Malmasi and M. Zampieri, "Detecting hate speech in social media," 2017, *arXiv:1712.06427*. [Online]. Available: http://arxiv.org/abs/1712.06427

[115] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, and R. Kumar, "Predicting the type and target of offensive posts in social media," 2019, *arXiv:1902.09666*. [Online]. Available: http://arxiv.org/abs/1902.09666

[116] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," 2017, *arXiv:1702.05638*. [Online]. Available: http://arxiv.org/abs/1702.05638

[117] A. Dey, R. Z. Rafi, S. Hasan Parash, S. K. Arko, and A. Chakrabarty, "Fake news pattern recognition using linguistic analysis," in *Proc. Joint 7th Int. Conf. Informat., Electron. Vis. (ICIEV), 2nd Int. Conf. Imag., Vis. Pattern Recognit. (icIVPR)*, Jun. 2018, pp. 305–309.

[118] 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 (CSCW)*, 2012, pp. 441–450.

[119] S. M. Shariff, X. Zhang, and M. Sanderson, "On the credibility perception of news on Twitter: Readers, topics and features," *Comput. Hum. Behav.*, vol. 75, pp. 785–796, Oct. 2017.

[120] 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.

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

[122] S. Sikdar, B. Kang, J. ODonovan, T. Höllerer, and S. Adah, "Understanding information credibility on Twitter," in *Proc. Int. Conf. Social Comput.*, Sep. 2013, pp. 19–24.

[123] C. I. Hovland and W. Weiss, "The influence of source credibility on communication effectiveness," *Public Opinion Quart.*, vol. 15, no. 4, pp. 635–650, 1951.

[124] G. Barbier and H. Liu, "Information provenance in social media," in *Proc. Int. Conf. Social Comput., Behav.-Cultural Modeling, Predict.* Berlin, Germany: Springer, 2011, pp. 276–283.

[125] 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 3rd Int. Conf. Social Comput.*, Oct. 2011, pp. 1–8.

[126] M.-A. Abbasi and H. Liu, "Measuring user credibility in social media," in *Proc. Int. Conf. Social Comput., Behav.-Cultural Modeling, Predict.* Berlin, Germany: Springer, 2013, pp. 441–448.

[127] Y. Yamaguchi, T. Takahashi, T. Amagasa, and H. Kitagawa, "Turank: Twitter user ranking based on user-tweet graph analysis," in *Proc. Int. Conf. Web Inf. Syst. Eng.* Berlin, Germany: Springer, 2010, pp. 240–253.

[128] D. Westerman, P. R. Spence, and B. Van Der Heide, "Social media as information source: Recency of updates and credibility of information," *J. Comput.-Mediated Commun.*, vol. 19, no. 2, pp. 171–183, Jan. 2014.

[129] L. Huang and Y. Xiong, "Evaluation of microblog users' influence based on PageRank and users behavior analysis," *Adv. Internet Things*, vol. 3, no. 2, pp. 34–40, 2013.

[130] B. Kang, J. O'Donovan, and T. Höllerer, "Modeling topic specific credibility on Twitter," in *Proc. ACM Int. Conf. Intell. User Interfaces (IUI)*, 2012, pp. 179–188.

[131] A. Gupta and P. Kumaraguru, "Credibility ranking of tweets during high impact events," in *Proc. 1st Workshop Privacy Secur. Online Social Media (PSOSM)*, 2012, pp. 2–8.

[132] 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, Jun. 2011.

[133] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, 2011, pp. 675–684.

[134] X. Xia, X. Yang, C. Wu, S. Li, and L. Bao, "Information credibility on Twitter in emergency situation," in *Proc. Pacific-Asia Workshop Intell. Secur. Inform.* Berlin, Germany: Springer, 2012, pp. 45–59.

[135] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, "Tweetcred: Real-time credibility assessment of content on Twitter," in *Proc. Int. Conf. Social Informat.* Cham, Switzerland: Springer, 2014, pp. 228–243.

[136] K. Lorek, J. Suehiro-Wiciński, M. Jankowski-Lorek, and A. Gupta, "Automated credibility assessment on Twitter," *Comput. Sci.*, vol. 16, no. 2, pp. 157–168, 2015.

[137] M. Alrubaian, M. Al-Qurishi, M. M. Hassan, and A. Alamri, "A credibility analysis system for assessing information on Twitter," *IEEE Trans. Dependable Secure Comput.*, vol. 15, no. 4, pp. 661–674, Jul./Aug. 2016.

[138] Z. Jin, J. Cao, Y.-G. Jiang, and Y. Zhang, "News credibility evaluation on microblog with a hierarchical propagation model," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 230–239.

[139] M. Gupta, P. Zhao, and J. Han, "Evaluating event credibility on Twitter," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2012, pp. 153–164.

[140] J. R. C. Nurse, I. Agrafiotis, S. Creese, M. Goldsmith, and K. Lamberts, "Building confidence in information-trustworthiness metrics for decision support," in *Proc. 12th IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Jul. 2013, pp. 535–543.

[141] A. A. AlMansour and C. S. Iliopoulos, "Using Arabic microblogs features in determining credibility," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2015, pp. 1212–1219.

[142] G. Moran and L. Muzellec, "EWOM credibility on social networking sites: A framework," *J. Marketing Commun.*, vol. 23, no. 2, pp. 149–161, Mar. 2017.

[143] J. R. C. Nurse, S. S. Rahman, S. Creese, M. Goldsmith, and K. Lamberts, "Information quality and trustworthiness: A topical state-of-the-art review," in *Proc. Int. Conf. Comput. Appl. Netw. Secur. (ICCANS)*, IEEE, 2011.

[144] P. Meyer, "Defining and measuring credibility of newspapers: Developing an index," *J. Quart.*, vol. 65, no. 3, pp. 567–574, Sep. 1988.

[145] F. Fico, J. D. Richardson, and S. M. Edwards, "Influence of story structure on perceived story bias and news organization credibility," *Mass Commun. Soc.*, vol. 7, no. 3, pp. 301–318, Jul. 2004.

[146] C. Gaziano and K. McGrath, "Measuring the concept of credibility," *J. Quart.*, vol. 63, no. 3, pp. 451–462, Sep. 1986.

[147] K. Xu, Y. Liu, X. Zhao, and X. Dong, "Trust them or not? A study on media credibility of newspapers accounts on Sina Weibo," Apr. 2013, doi: 10.2139/ssrn.2258551.

[148] J. McCroskey, *An Introduction to Communication in the Classroom*. Edina, MN, USA: Burgess International Group," 1992.

[149] J. J. Teven and J. C. McCroskey, "The relationship of perceived teacher caring with student learning and teacher evaluation," *Commun. Educ.*, vol. 46, no. 1, pp. 1–9, Jan. 1997.

[150] J. C. McCroskey and J. J. Teven, "Goodwill: A reexamination of the construct and its measurement," *Commun. Monogr.*, vol. 66, no. 1, pp. 90–103, 1999.

[151] D. O'Keefe, "Theories of behavioral intention," in *Persuasion Theory & Research*, 2nd ed. Thousand Oaks, CA, USA: Sage, 2002, pp. 101–135.

[152] C. I. Hovland, I. L. Janis, and H. H. Kelley, "Communication and persuasion: Psychological studies of opinion change," New Haven, Yale Univ. Press, New Haven, CT, USA, 1976.

[153] A. L. Ginsca, A. Popescu, and M. Lupu, "Credibility in information retrieval," *Found. Trends Inf. Retr.*, vol. 9, no. 5, pp. 355–475, 2015.

[154] Y. Gil and D. Artz, "Towards content trust of web resources," *J. Web Semantics*, vol. 5, no. 4, pp. 227–239, 2007.

[155] A. J. Flanagin and M. Metzger, "Digital media and youth: Unparalleled opportunity and unprecedented responsibility," in *Digital Media, Youth, and Credibility* (Foundation Series on Digital Media and Learning), M. J. Metzger and A. J. Flanagin, Eds. Cambridge, MA, USA: MIT Press, 2008, pp. 5–28, doi: 10.1162/dmal.9780262562324.005.

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

[157] B. Kang, T. Höllerer, M. Turk, X. Yan, and J. O'Donovan, "An analysis of credibility in microblogs," Ph.D. dissertation, Dept. Comput. Sci., Univ. California, Santa Barbara, Santa Barbara, CA, USA, 2012.

[158] 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.

[159] M. Gupta, P. Zhao, and J. Han, "Evaluating event credibility on Twitter," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2012, pp. 153–164.

[160] 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.

[161] G. J. Golan, "New perspectives on media credibility research," *Amer. Behav. Sci.*, vol. 54, no. 1, pp. 3–7, Sep. 2010.

[162] S. Y. Rieh and D. R. Danielson, "Credibility: A multidisciplinary framework," *Annu. Rev. Inf. Sci. Technol.*, vol. 41, no. 1, pp. 307–364, 2007.

[163] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *J. Manage. Inf. Syst.*, vol. 12, no. 4, pp. 5–33, 1996.

[164] B. J. Fogg, P. Swani, M. Treinen, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, and J. Shon, "What makes web sites credible: A report on a large quantitative study," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI)*, 2001, pp. 61–68.

[165] T. Hong, "The influence of structural and message features on web site credibility," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 57, no. 1, pp. 114–127, Jan. 2006.

[166] D. Robins and J. Holmes, "Aesthetics and credibility in web site design," *Inf. Process. Manage.*, vol. 44, no. 1, pp. 386–399, 2008.

[167] S. Y. Rieh and B. Hilligoss, "College students' credibility judgments in the information-seeking process," in *Digital Media, Youth, and Credibility* (Foundation Series on Digital Media and Learning), M. J. Metzger and A. J. Flanagin, Eds. Cambridge, MA, USA: MIT Press, Jan. 2008, pp. 49–72, doi: 10.1162/dmal.9780262562324.049.

[168] S. Chaiken and Y. Trope, *Dual-Process Theories in Social Psychology*. New York, NY, USA: Guilford Press, 1999.

[169] F. Pierri and S. Ceri, "False news on social media: A data-driven survey," *SIGMOD Rec.*, vol. 48, no. 2, pp. 18–27, Dec. 2019.

[170] J. Roozenbeek and S. van der Linden, "Fake news game confers psychological resistance against online misinformation," *Palgrave Commun.*, vol. 5, no. 1, pp. 1–10, 2019.

[171] S. Volkova, K. Shaffer, J. Y. Jang, and N. Hodas, "Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2017, pp. 647–653.

[172] B. Guo, Y. Ding, L. Yao, Y. Liang, and Z. Yu, "The future of false information detection on social media: New perspectives and trends," *ACM Comput. Surv.*, vol. 53, no. 4, pp. 1–36, 2020.

[173] F. Jin, W. Wang, L. Zhao, E. Dougherty, Y. Cao, C.-T. Lu, and N. Ramakrishnan, "Misinformation propagation in the age of Twitter," *Computer*, vol. 47, no. 12, pp. 90–94, Dec. 2014.

[174] X. Wang, Y. Lin, Y. Zhao, L. Zhang, J. Liang, and Z. Cai, "A novel approach for inhibiting misinformation propagation in human mobile opportunistic networks," *Peer-Peer Netw. Appl.*, vol. 10, no. 2, pp. 377–394, Mar. 2017.

[175] N. A. Karlova and K. E. Fisher, "A social diffusion model of misinformation and disinformation for understanding human information behaviour," *Inf. Res.*, vol. 18, no. 1, p. 573, 2013. [Online]. Available: http://InformationR.net/ir/18-1/paper573.html

[176] X. Chen, L. Y.-H. Lo, and H. Qu, "SirenLess: Reveal the intention behind news," 2020, *arXiv:2001.02731*. [Online]. Available: http://arxiv.org/abs/2001.02731

[177] D. Küçük and F. Can, "Stance detection: A survey," *ACM Comput. Surv.*, vol. 53, no. 1, pp. 1–37, 2020.

[178] E. Tacchini, G. Ballarin, M. L. D. Vedova, S. Moret, and L. de Alfaro, "Some like it hoax: Automated fake news detection in social networks," 2017, *arXiv:1704.07506*. [Online]. Available: http://arxiv.org/abs/1704.07506

[179] (2020). *Credibility Ground Truths*. [Online]. Available: https://didyoucheckfirst.wordpress.com/opensources-co-news-sites/

[180] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from Twitter," in *Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust IEEE 3rd Int. Conf. Social Comput.*, Oct. 2011, pp. 149–156.

[181] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our Twitter profiles, our selves: Predicting personality with Twitter," in *Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust IEEE 3rd Int. Conf. Social Comput.*, Oct. 2011, pp. 180–185.

[182] M. Kakol, R. Nielek, and A. Wierzbicki, "Understanding and predicting web content credibility using the content credibility corpus," *Inf. Process. Manage.*, vol. 53, no. 5, pp. 1043–1061, 2017.

[183] R. Thomson, N. Ito, H. Suda, F. Lin, Y. Liu, R. Hayasaka, R. Isochi, and Z. Wang, "Trusting tweets: The Fukushima disaster and information source credibility on Twitter," in *Proc. 9th Int. ISCRAM Conf.* Vancouver, BC, Canada: Simon Fraser Univ., 2012, pp. 1–10.

[184] M. Viviani and G. Pasi, "Quantifier guided aggregation for the veracity assessment of online reviews," *Int. J. Intell. Syst.*, vol. 32, no. 5, pp. 481–501, May 2017.

[185] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: Characterizing and identifying fake images on Twitter during hurricane sandy," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 729–736.

[186] N. H. Idris, M. Jackson, and M. Ishak, "A conceptual model of the automated credibility assessment of the volunteered geographic information," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 18, no. 1, 2014, Art. no. 012070.

[187] R. M. B. Al-Eidan, H. S. Al-Khalifa, and A. S. Al-Salman, "Measuring the credibility of Arabic text content in Twitter," in *Proc. 5th Int. Conf. Digit. Inf. Manage. (ICDIM)*, Jul. 2010, pp. 285–291.

[188] A. A. Al Mansour, 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, no. 1, pp. 53–62, 2014.

[189] 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.

[190] 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.

[191] J. Schaffer, T. Abdelzaher, D. Jones, T. Höllerer, C. Gonzalez, J. Harman, and J. O'Donovan, "Truth, lies, and data: Credibility representation in data analysis," in *Proc. IEEE Int. Inter-Disciplinary Conf. Cognit. Methods Situation Awareness Decis. Support (CogSIMA)*, Mar. 2014, pp. 28–34.

[192] T. J. Johnson and B. K. Kaye, "The dark side of the boon? Credibility, selective exposure and the proliferation of online sources of political information," *Comput. Hum. Behav.*, vol. 29, no. 4, pp. 1862–1871, Jul. 2013.

[193] Q. V. Liao, P. Pirolli, and W.-T. Fu, "An ACT-R model of credibility judgment of micro-blogging web pages," in *Proc. ICCM*, vol. 103, 2012, pp. 1–6.

[194] B. Kang, T. Höllerer, and J. O'Donovan, "Believe it or not? Analyzing information credibility in microblogs," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2015, pp. 611–616.

[195] 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.

[196] 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.

[197] 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.

[198] S.-Y. Rieh, "Participatory web users' information activities and credibility assessment," *J. Korean Soc. Library Inf. Sci.*, vol. 44, no. 4, pp. 155–178, 2010.

[199] E. J. Briscoe, D. S. Appling, and H. Hayes, "Social network derived credibility," in *Recommendation and Search in Social Networks*. Cham, Switzerland: Springer, 2015, pp. 59–75.

[200] 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, pp. 151–168, Dec. 2013.

[201] K. A. Johnson, "The effect of Twitter posts on students' perceptions of instructor credibility," *Learn., Media Technol.*, vol. 36, no. 1, pp. 21–38, Mar. 2011.

[202] R. Thomson, N. Ito, H. Suda, F. Lin, Y. Liu, R. Hayasaka, R. Isochi, and Z. Wang, "Trusting tweets: The Fukushima disaster and information source credibility on Twitter," in *Proc. ISCRAM*, 2012, pp. 1–10.

[203] A. Zubiaga and H. Ji, "Tweet, but verify: Epistemic study of information verification on Twitter," *Social Netw. Anal. Mining*, vol. 4, no. 1, p. 163, Dec. 2014.

[204] J. M. DeGroot, V. J. Young, and S. H. VanSlette, "Twitter use and its effects on student perception of instructor credibility," *Commun. Educ.*, vol. 64, no. 4, pp. 419–437, Oct. 2015.

[205] J. R. 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, no. 1, p. 5, 2014.

[206] A. Pal and S. Counts, "What's in a @ name? How name value biases judgment of microblog authors," in *Proc. ICWSM*, 2011, pp. 1–8.

[207] 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.

[208] M. R. Jahng and J. Littau, "Interacting is believing: Interactivity, social cue, and perceptions of journalistic credibility on Twitter," *J. Mass Commun. Quart.*, vol. 93, no. 1, pp. 38–58, Mar. 2016.

[209] S. M. Shariff, X. Zhang, and M. Sanderson, "User perception of information credibility of news on Twitter," in *Proc. Eur. Conf. Inf. Retr.* Cham, Switzerland: Springer, 2014, pp. 513–518.

[210] S. K. Sikdar, B. Kang, J. O'Donovan, T. Hollerer, and S. Adal, "Cutting through the noise: Defining ground truth in information credibility on Twitter," *Human*, vol. 2, no. 3, pp. 151–167, 2013.

[211] S. Aladhadh, X. Zhang, and M. Sanderson, "Tweet author location impacts on tweet credibility," in *Proc. Australas. Document Comput. Symp. (ADCS)*, 2014, pp. 73–76.

[212] 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*, 2014, pp. 749–752.

[213] 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, pp. 1–4.

[214] R. Mihalcea and D. Radev, *Graph-Based Natural Language Processing and Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[215] W. Yang Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," 2017, *arXiv:1705.00648*. [Online]. Available: http://arxiv.org/abs/1705.00648

[216] C. Silverman. *This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook, Buzzfeed News 2016*. Accessed: Aug. 12, 2021. [Online]. Available: https://zenodo.org/record/1239675

[217] K. Popat, S. Mukherjee, A. Yates, and G. Weikum, "DeClarE: Debunking fake news and false claims using evidence-aware deep learning," 2018, *arXiv:1809.06416*. [Online]. Available: http://arxiv.org/abs/1809.06416

[218] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," 2017, *arXiv:1708.07104*. [Online]. Available: http://arxiv.org/abs/1708.07104

[219] C. Shao, G. L. Ciampaglia, A. Flammini, and F. Menczer, "Hoaxy: A platform for tracking online misinformation," in *Proc. 25th Int. Conf. Companion World Wide Web (WWW Companion)*, 2016, pp. 745–750.

[220] M. Risdal. *Fake News Dataset 2017*. Accessed: Aug. 12, 2021. [Online]. Available: https://www.kaggle.com/mrisdal/fake-news

[221] L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. W. S. Hoi, and A. Zubiaga, "SemEval-2017 task 8: RumourEval: Determining rumour veracity and support for rumours," 2017, *arXiv:1704.05972*. [Online]. Available: http://arxiv.org/abs/1704.05972

[222] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," in *Proc. 25th Int. Joint Conf. Artif. Intell. (IJCAI)*. New York, NY, USA: AAAI Press, 2016, pp. 3818–3824. [Online]. Available: https://ink.library.smu.edu.sg/sis_research/4630

[223] A. Olteanu, S. Peshterliev, X. Liu, and K. Aberer, "Web credibility: Features exploration and credibility prediction," in *Proc. Eur. Conf. Inf. Retr.* Berlin, Germany: Springer, 2013, pp. 557–568.

[224] S. A. McCornack, K. Morrison, J. E. Paik, A. M. Wisner, and X. Zhu, "Information manipulation theory 2: A propositional theory of deceptive discourse production," *J. Lang. Social Psychol.*, vol. 33, no. 4, pp. 348–377, Sep. 2014.

[225] O. W. Makinde, "Assessing the credibility of online social network messages," Ph.D. dissertation, Univ. Derby, Derby, U.K., 2018.

[226] M. K. Johnson and C. L. Raye, "Reality monitoring," *Psychol. Rev.*, vol. 88, no. 1, p. 67, 1981.

[227] M. Zuckerman, B. M. DePaulo, and R. Rosenthal, "Verbal and nonverbal communication of deception," in *Advances in Experimental Social Psychology*, vol. 14. Amsterdam, The Netherlands: Elsevier, 1981, pp. 1–59.

[228] U. Undeutsch, "Beurteilung der glaubhaftigkeit von aussagen," *Handbuch Psychologie*, vol. 11, pp. 26–181, Nov. 1967.

[229] P. M. Homer and L. R. Kahle, "Source expertise, time of source identification, and involvement in persuasion: An elaborative processing perspective," *J. Advertising*, vol. 19, no. 1, pp. 30–39, Mar. 1990.

[230] R. Li and A. Suh, "Factors influencing information credibility on social media platforms: Evidence from Facebook pages," *Proc. Comput. Sci.*, vol. 72, pp. 314–328, Jan. 2015.

[231] B. E. Ashforth and F. Mael, "Social identity theory and the organization," *Acad. Manage. Rev.*, vol. 14, no. 1, pp. 20–39, Jan. 1989.

[232] G. Cronkhite and J. Liska, "A critique of factor analytic approaches to the study of credibility," *Commun. Monogr.*, vol. 43, no. 2, pp. 91–107, Jun. 1976.

[233] X. Hu, "Assessing source credibility on social media—An electronic word-of-mouth communication perspective," Ph.D. dissertation, Bowling Green State Univ., Bowling Green, OH, USA, 2015.

[234] 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 3rd Int. Conf. Social Comput.*, Oct. 2011, pp. 1–8.

[235] J. Jessen and A. H. Jørgensen, "Aggregated trustworthiness: Redefining online credibility through social validation," *1st Monday*, vol. 17, no. 1, Jan. 2012. [Online]. Available: https://firstmonday.org/ojs/index.php/fm/article/view/3731/3132, doi: 10.5210/fm.v17i1.3731.

[236] D. Centola and R. Wilier, "The emperor's dilemma," in *Theories of Social Order: A Reader*. USA: Stanford Univ. Press, 2009, p. 276.

[237] M. Deutsch and H. B. Gerard, "A study of normative and informational social influences upon individual judgment," *J. Abnormal Social Psychol.*, vol. 51, no. 3, p. 629, 1955.

[238] T. Kuran and C. R. Sunstein, "Availability cascades and risk regulation," *Stanford Law Rev.*, vol. 51, no. 4, p. 683, Apr. 1999.

[239] D. Dunning, D. W. Griffin, J. D. Milojkovic, and L. Ross, "The overconfidence effect in social prediction," *J. Personality Social Psychol.*, vol. 58, no. 4, p. 568, 1990.

[240] E. Pronin, J. Kruger, K. Savtisky, and L. Ross, "You don't know me, but I know you: The illusion of asymmetric insight," *J. Personality Social Psychol.*, vol. 81, no. 4, p. 639, 2001.

[241] L. Ross and A. Ward, "Naive realism in everyday life: Implications for social conflict and misunderstanding," in *Values and Knowledge*, 1st ed., E. S. Reed, E. Turiel, and T. Brown, Eds. New York, NY, USA: Psychology Press, Mar. 1996, ch. 6, pp. 103–136, doi: 10.4324/9780203773772.

[242] J. L. Freedman and D. O. Sears, "Selective exposure," in *Advances in Experimental Social Psychology*, vol. 2. Amsterdam, The Netherlands: Elsevier, 1965, pp. 57–97.

[243] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises," *Rev. Gen. Psychol.*, vol. 2, no. 2, pp. 175–220, 1998.

[244] R. J. Fisher, "Social desirability bias and the validity of indirect questioning," *J. Consum. Res.*, vol. 20, no. 2, pp. 303–315, 1993.

[245] H. Leibenstein, "Bandwagon, snob, and Veblen effects in the theory of consumers' demand," *Quart. J. Econ.*, vol. 64, no. 2, pp. 183–207, 1950.

[246] S. T. Moturu and H. Liu, "Quantifying the trustworthiness of social media content," *Distrib. Parallel Databases*, vol. 29, no. 3, pp. 239–260, Jun. 2011.

[247] S. T. Kate, "Trustworthiness within social networking sites: A study on the intersection of HCI and sociology," B.S. thesis, Univ. Amsterdam, Amsterdam, The Netherlands, 2009.

[248] A. Algarni, H. Al Makrami, and A. Alarifi, "Toward evaluating trustworthiness of social networking site users: Reputation-based method," *Arch. Bus. Res.*, vol. 7, no. 3, pp. 27–41, Mar. 2019.

[249] L. E. Boehm, "The validity effect: A search for mediating variables," *Personality Social Psychol. Bull.*, vol. 20, no. 3, pp. 285–293, Jun. 1994.

[250] P. Bálint and G. Bálint, "The semmelweis-reflex," *Orvosi Hetilap*, vol. 150, no. 30, p. 1430, Jul. 2009.

[251] C. MacLeod, A. Mathews, and P. Tata, "Attentional bias in emotional disorders," *J. Abnormal Psychol.*, vol. 95, no. 1, p. 15, 1986.

[252] K. H. Jamieson and J. N. Cappella, *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. London, U.K.: Oxford Univ. Press, 2008.

[253] C. I. Hovland, O. Harvey, and M. Sherif, "Assimilation and contrast effects in reactions to communication and attitude change," *J. Abnormal Social Psychol.*, vol. 55, no. 2, p. 244, 1957.

[254] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," in *Handbook of the Fundamentals of Financial Decision Making: Part I*. Singapore: World Scientific, 2013, pp. 99–127.

[255] N. H. Frijda, *The Emotions*. Cambridge, U.K.: Cambridge Univ. Press, 1986.

[256] J. Amador, A. Oehmichen, and M. Molina-Solana, "Characterizing political fake news in Twitter by its meta-data," 2017, *arXiv:1712.05999*. [Online]. Available: http://arxiv.org/abs/1712.05999

[257] M. Alsaleh, A. Alarifi, A. M. Al-Salman, M. Alfayez, and A. Almuhaysin, "TSD: Detecting sybil accounts in Twitter," in *Proc. 13th Int. Conf. Mach. Learn. Appl.*, Dec. 2014, pp. 463–469.

[258] (2020). *Wiki Fake Sites*. [Online]. Available: https://en.wikipedia.org/wiki/List_of_fake_news_websites

[259] (2020). *News Guard Tech*. [Online]. Available: https://www.newsguardtech.com/

[260] (2020). *News Media Bias*. [Online]. Available: https://www.allsides.com/

[261] (2020). *Media Bias Fact Check*. [Online]. Available: https://mediabiasfactcheck.com/

[262] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proc. 26th Int. Conf. World Wide Web Companion (WWW Companion)*, 2017, pp. 963–972.

[263] *Internettrusttoolnewsguardwebsite*. Accessed: Aug. 12, 2021. [Online]. Available: https://www.newsguardtech.com/

[264] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *J. Lang. Social Psychol.*, vol. 29, no. 1, pp. 24–54, Mar. 2010.

[265] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, "Fake news early detection: A theory-driven model," *Digit. Threats, Res. Pract.*, vol. 1, no. 2, pp. 1–25, Jul. 2020.

[266] M. Recasens, C. Danescu-Niculescu-Mizil, and D. Jurafsky, "Linguistic models for analyzing and detecting biased language," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2013, pp. 1650–1659.

[267] *Factcheckingwebsiteslist*. Accessed: Aug. 12, 2021. [Online]. Available: https://reporterslab.org/fact-checking/

[268] W. H. Dutton, G. Blank, and D. Groselj, *Cultures of the Internet: The Internet in Britain: Oxford Internet Survey 2013 Report*. Oxford, U.K.: Oxford Internet Institute, 2013.

[269] D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, and J. L. Zittrain, "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

KHUBAIB AHMED QURESHI was the Head of the Computer Science Department, HIMS, Hamdard University, Karachi, Pakistan. He is currently affiliated with the Department of Computer Science, DHA Suffa University, Karachi, as an Assistant Professor, and the Head of the Data Science Program. He is having 20 years of comprehensive research and teaching experience, and continuing his research in the area of computer science named data science, complex networks, and social computing. He has authored several research articles along with chapters in different books.

RAUF AHMED SHAMS MALICK received the Ph.D. degree from the University of Karachi. He has been a Visiting Scholar at NİG, Japan, and UCLA, USA. He has founded several companies with state of the art products related to social media, location-based analytics, and organizational networks. He is involved in complex system research and pursuing problems in the area of biological networks, networked economics, and personality traits. He has distinguished background in designing novel solutions for complex systems. He is currently affiliated with the Department of Computer Science, National University of Computer and Emerging Sciences, as an Assistant Professor, and continuing his research in the specialized scientific area of computer science, complex networks, social computing, bioinformatics, and integrated systems. He has authored several articles along with chapters in different books.

MUHAMMAD SABIH received the B.E. degree in industrial electronics from IIEE-NED University, Pakistan, in 2000, and the M.S. and Ph.D. degrees in systems engineering from KFUPM, Dhahran, Saudi Arabia, in 2009 and 2014, respectively. During his research period at KFUPM, he worked for several applied research collaborations between KFUPM and MIT. His data driven research work expanded to a funded industry-academia collaboration project between KFUPM and Yokogawa, Saudi Arabia, and turned into U.S. patent. He worked as an Algorithm Specialist and developed anomaly detection algorithms using Python on the pipeline inspection data at leading the Research and Technology Center (RTRC), German ROSEN Group, Dhahran Techno-Valley (DTV), Saudi Arabia. Since 2009, he has been working in the field of computer and electrical engineering and a Professional Member of the International Society of Automation (ISA). He is currently an Assistant Professor with DHA Suffa University and actively engaged in developing solutions from industrial data utilizing machine learning methods for estimation, modeling, and compensation. He has one U.S. patent and around ten peer-reviewed papers. His current research interests include industry 4.0, data science, modeling, and estimation for real world problems.