IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 2, 2019, accepted January 13, 2019, date of publication January 17, 2019, date of current version February 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2893657

Info-Trust: A Multi-Criteria and Adaptive Trustworthiness Calculation Mechanism for Information Sources

YALI GAO[1], XIAOYONG LI[1], JIRUI LI[1], YUNQUAN GAO[1], AND PHILIP S. YU[2,3], (Fellow, IEEE)

[1]Key Laboratory of Trustworthy Distributed Computing and Service, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]University of Illinois at Chicago, Chicago, IL 60607, USA
[3]Institute for Data Science, Tsinghua University, Beijing 100084, China

Corresponding author: Xiaoyong Li (lxyxjtu@163.com)

This work was supported in part by the NSFC-General Technology Fundamental Research Joint Fund U1836215, in part by the National Nature Science Foundation of China under Grant 61672111, and in part by the BUPT Excellent Ph.D. Students Foundation under Grant CX2018216.

ABSTRACT Social media have become increasingly popular for the sharing and spreading of user-generated content due to their easy access, fast dissemination, and low cost. Meanwhile, social media also enable the wide propagation of cyber frauds, which leverage fake information sources to reach an ulterior goal. The prevalence of untrustworthy information sources on social media can have significant negative societal effects. In a trustworthy social media system, trust calculation technology has become a key demand for the identification of information sources. Trust, as one of the most complex concepts in network communities, has multi-criteria properties. However, the existing work only focuses on single trust factor, and does not consider the complexity of trust relationships in social computing completely. In this paper, a multi-criteria trustworthiness calculation mechanism called Info-Trust is proposed for information sources, in which identity-based trust, behavior-based trust, relation-based trust, and feedback-based trust factors are incorporated to present an accuracy-enhanced full view of trustworthiness evaluation of information sources. More importantly, the weights of these factors are dynamically assigned by the ordered weighted averaging and weighted moving average (OWA-WMA) combination algorithm. This mechanism surpasses the limitations of existing approaches in which the weights are assigned subjectively. The experimental results based on the real-world datasets from Sina Weibo demonstrate that the proposed mechanism achieves greater accuracy and adaptability in trustworthiness identification of the network information.

INDEX TERMS Multi-criteria, adaptive weight, trust calculation mechanism, information sources, social media.

I. INTRODUCTION

With the boom of social media and mobile devices, people can now easily create, publish, and access user-generated content, leading to great information exposure for almost everyone in the world. Participants in social media become producers and consumers of user-generated information, hence the shift of the role of information sources from a few dedicated entities to a diverse and distributed group of individuals. However, given the side effects of freedom of speech and the existence of online users with anonymous or fake identities, the trustworthiness of information sources has become a serious problem for social media. Untrustworthy or even malicious information sources, which send spam; spread malware, online rumors, unverified claims, fraudulent or fake reviews, and deceptive marketing; and launch other underground illicit activities, exert harmful effects on individuals and bring about inflamed sentiment, economic losses, and other negative impacts on society. The trustworthiness calculation mechanism is a tool for facilitating decision making in diverse applications. As a result of the complexity of social media and the concept of trust itself, quantifying trust in social media is a difficult and significant problem, which has spurred a significant amount of research from the academia and industry [1]–[6].

A. MOTIVATIONS

In the past few years, academic research communities including social computing, cybersecurity, data mining, etc. have been attracted to the problem of evaluating the trustworthiness of information sources, and many state-of-the-art studies have been carried out, such as [2]–[4], [6]–[9]. Some of them are very creative and elaborate, but most of them still face two key limitations that need to be solved.

1) FEW STUDIES HAVE FOCUSED ON A FULL VIEW OF TRUSTWORTHINESS ASSESSMENT OF INFORMATION SOURCES

From the perspective of information consumers, trust is a comprehensive index for information quality guarantee, and a trust management system comprises multi-criteria trust factors. The same is validated by the following observation in real life: people will check multiple views to obtain a clear idea about certain information sources. For example, people may investigate basic profiles (who is he/she?), posting history (what did he/she post?), social structure (how about his/her social relation?), and user feedback (how about user feedback?) to conduct a comprehensive evaluation of the trustworthiness of information sources. This setting highlights the fact that knowledge from a single view could contain noisy information, but by combining information from different perspectives, we can obtain a full view to improve the accuracy of trustworthiness evaluation.

To the best of our knowledge, most current studies either ignore feedback-based factor or identity-based factor in trust evaluation, which may lead to inaccurate trustworthiness perceptions. For example, Golbeck [10] considered only the network structure of information sources and did not take note of other trust factors. In [3], user profile, history behavior, and network structure were considered, but the feedback-based trust factor was ignored. In reality, real-time feedback from information consumers is significant and effective to evaluate the trustworthiness of information sources. A major limitation of current studies is that they ignored one or several criteria and thus failed to cope with the concealment of adversaries (e.g., purchasing followers, leveraging URL shortening services), which will greatly hinder the acceptance of trustworthiness evaluation results [11].

2) MANY SCHEMES LACK ADAPTABILITY TO A TRUST FUSION CALCULATION, IN WHICH SUBJECTIVE METHODS OR WEIGHTED AVERAGE METHODS ARE USED TO ASSIGN WEIGHTS TO TRUST FACTORS

Avoiding the effect of individual favoritism on weight allocation and confirming the weight allocation of multi-criteria adaptively are important in trust fusion calculation [12], [13]. Some previous studies are based on expert opinion to weigh trust factors; however, this approach lacks adaptability and may lead to inaccurate results in trust evaluation. In a recent work [4], trustworthiness was defined as $R(D) = \lambda_1 \cdot R_1(D) + \lambda_2 \cdot R_2(D) + \lambda_3 \cdot R_3(D) + \lambda_4 \cdot R_4(D)$, where $\lambda_1, \lambda_2, \lambda_3$, and λ_4 with $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$ are weights to allow tradeoff among the four factors; however, the authors only provided the default setting (i.e., 1/4, 1/4, 1/4, 1/4). Although the model in [2] is a weighted sum-based multi-criteria based evaluation scheme, it does not describe how to define the weights clearly. Thus, these schemes lack the adaptability to weigh these trust factors.

B. OUR CONTRIBUTIONS

Focusing on above issues of trust evaluation in social media and on the basis of previous work on trustworthiness evaluation [4], [6], [14]–[16], this study presents an innovative trustworthiness evaluation mechanism based on human cognitive behavior. Multiple criteria are incorporated to reflect the characteristics of complexity and uncertainty of trust. The weights of these trust factors are dynamically allocated by ordered weighted averaging - weighted moving average (OWA-WMA) combination algorithm [17]–[20]. This mechanism overcomes limitations of previous approaches in which weights are allocated subjectively. Results of simulation experiments demonstrate that the proposed mechanism achieves greater accuracy and adaptability in trust evaluation. The main innovations and key features of the proposed mechanism can be summarized as follows:

- **Multi-criteria trust factors are leveraged to build an accuracy-enhanced trust calculation mechanism for information sources.** In open social media systems, trust is one of the most complex concepts in network communities, and trustworthiness is difficult to quantify and predict. To the best of our knowledge, this work is the first to comprehensively and properly combine data from multiple perspectives, including identity, posting history, network structure, and user feedback, to evaluate the trustworthiness of information sources. This combination reduces network risk while significantly enhancing the accuracy of trust evaluation.
- **An adaptive and robust overall trust degree aggregation algorithm is proposed.** Many previous studies used artificial or weighted average means to assign weights to different trust factors. However, the adaptability of these models faces limitations. According to multi-source information fusion theory, the OWA-WMA combination algorithm is leveraged to integrate multiple trust factors into an overall trust evaluation, which can overcome the limitation of human subjectivity in weight allocation.

A series of simulation experiments based on real-world data sets from Sina Weibo were conducted to evaluate the effectiveness and adaptability of the proposed multi-criteria and adaptive trustworthiness calculation mechanism for information sources. Experimental results demonstrate that the proposed Info-Trust model can significantly outperform the state-of-the-art approaches for the task of trustworthiness evaluation of information sources.

The remainder of this paper is organized as follows. Section II provides an overview of related work.

TABLE 1. Comparison of existing trust models in social media.

Model	IF	BF	RF	FF	Dimension	Computation Model	Weight Allocation	Adaptive Model	Trust Value
TidalTrust [10]	✗	✗	✓	✗	1	linear model	weighted average	✗	discrete, [1,10]
Adali's model [21]	✗	✓	✗	✗	1	linear model	weighted average	✗	continuous, [0,1]
Jia's model [22]	✗	✗	✓	✗	1	linear model	weighted average	✗	continuous, [0,1]
SybilSCAR [23]	✗	✗	✓	✗	1	based on RW and LBP	equally weighted	✗	continuous, [0,1]
SWTrust [24]	✗	✓	✓	✗	2	linear model	weighted average	✗	continuous, [0,1]
Pichon's model [2]	✓	✓	✗	✗	2	choquet integral	artificially weighted	✗	continuous, [0,1]
Canini's model [3]	✓	✓	✓	✗	3	linear model	weighted average	✗	continuous, [0,1]
Zhao's model [4]	✗	✓	✓	✗	2	linear model	artificially weighted	✗	continuous, [0,1]
Info-Trust	✓	✓	✓	✓	4	hybrid	adaptively weighted	✓	continuous, [0,1]

[1] In the second, third, fourth, and fifth columns, IF, BF, RF, and FF represent the identity factor, behavior factor, relation factor, and feedback factor, respectively. The symbol ✓indicates that the related factor is taken into account, and the symbol ✗means that the related factor is not considered.
[2] In the SybilSCAR model, RW denotes random walk, and LBP denotes loop belief propagation.

Section III presents the problem formation and an overview of the architecture model. Section IV details our mechanism for assessing the trustworthiness of information sources in social media. Section V describes the experiments and performance results of the proposed mechanism. Finally, Section VI summarizes the paper and outlines future work.

II. RELATED WORK

Trust management and computing systems are successfully used in numerous application scenarios, helping consumers identify trustworthy and reliable service providers [13], [16], [25]. Similar approaches are needed to support social media systems in identifying trustworthy information sources. In general, trustworthiness is a measure of confidence that a node will behave in an expected manner. Recently, a number of innovative trust evaluations (some may interpret it as trust computing or trust inference) on information sources for social media have been proposed by researchers in the academia and industry [2]–[4], [7]–[9], thereby helping people decide whether to trust an unknown information source. In this subsection, we take an in-depth look at the proposed trust models in the related literature. Table 1 outlines the comparison of these representative models in detail.

Depending on the evaluation methods used, we roughly categorized the existing trust evaluation approaches into *behavior-based methods* [2], [21], *network structure-based methods* [6], [22], [23], [26], and hybrid schemes [2]–[4]. *Behavior-based methods* extract valuable information, such as volume and frequency, from historical posting behavior to evaluate trustworthiness [21], [27], [28]. However, a key limitation of behavior-based methods is that they are not adversarially robust, i.e., fraudulent information sources have evolved to mimic normal sources and evade existing detection features via manipulation of their futures as desired. *Network structure-based methods* extract structural information, such as in-degree, out-degree, tie-strength, etc.; leverage the global structure and/or the local structure of a social graph; and uncover how trust propagates in the social graph [29]. These methods are often based on the intuition that although an malicious source can arbitrarily control the connections between Sybils, it is difficult for the malicious source to manipulate the connections between benign source and malicious source, which requires actions from benign sources. Wang *et al.* [26] designed a pairwise Markov random field to model the joint probability distribution of the states of all nodes on the basis of a set of labeled fraudulent nodes and normal nodes.

Egele *et al.* [7] presented the COMPA system to detect compromised accounts in social networks. The authors used statistical models to analyze user behavior and leveraged anomaly detection techniques to identify sudden changes in behavior. Ruan *et al.* [30] studied and proposed a set of social behavioral features (e.g., first activity, activity sequence, browsing preference, and visit duration) that characterize a user behavioral profile and accurately reflect a user's online social network activity pattern. Their proposed profiles can accurately differentiate individual users and detect compromised accounts. However, in [7] and [30], the authors only focused on user behavioral features, and they did not consider the user feedback information and social structure-based trust factor.

Canini *et al.* [3] proposed a hybrid approach to automatically identify and rank reputable and credible sources in social networks, on the basis of their social network structure and expertise in a given topic. Their research results indicate that the topical content of information sources and social network structures affect judgments of credibility. The trust mechanism, as one of the most complex and dynamic concepts in social relationships, should take multiple decision-making factors into account [11], [31]. However, in indicating the underlying trustworthiness of information sources, considering only three trust factors is not enough. The disregard for other informative trust factors, such as user feedback-based trust factors, may bring about inaccurate or unfair results when making trust decisions.

Feedback, also known as recommendation or reputation, provides an efficient and effective way to build reputation-based trust in social media. Most social media systems can provide additional feedback information. However, feedback analysis in social media for source trustworthiness evaluation has been barely studied. In our proposed model, comments submitted by users and feedback from the feedback platform are collected and aggregated to yield feedback-based trust.

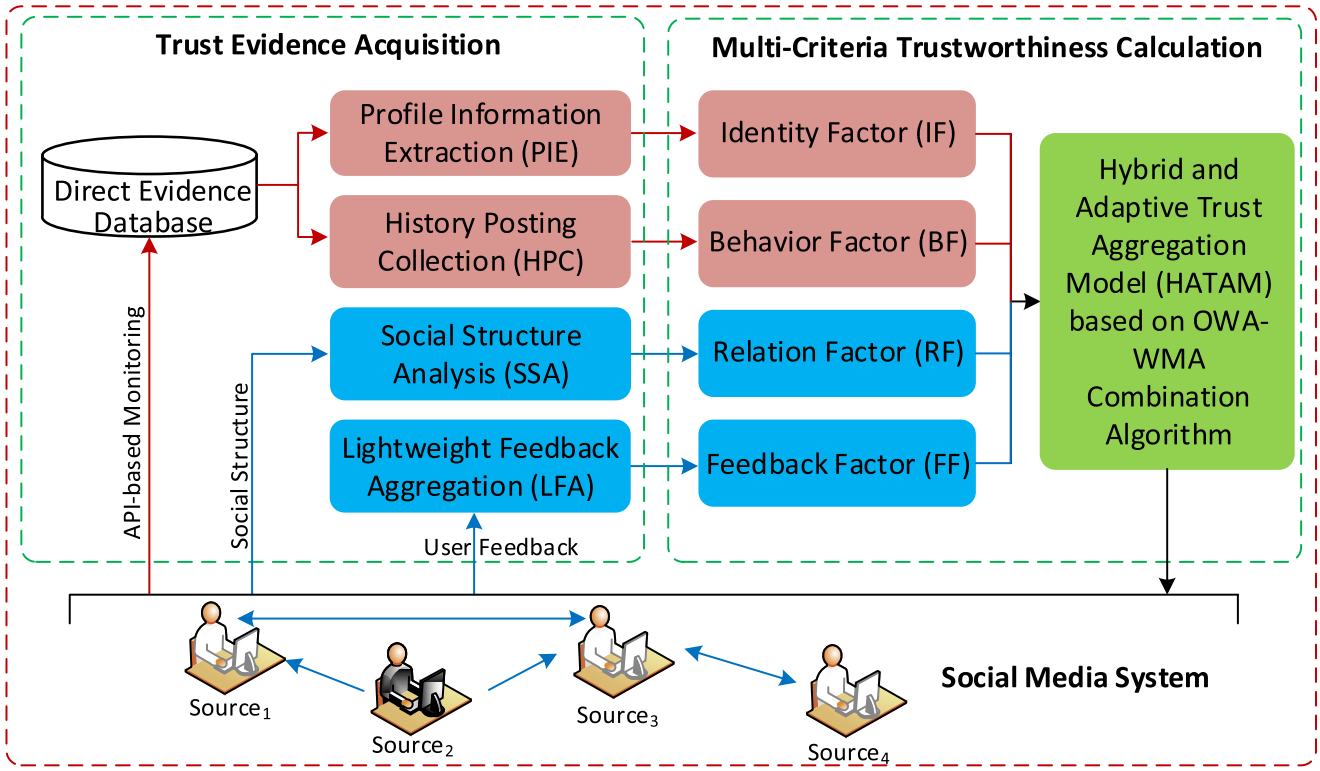

FIGURE 1. Info-Trust's architecture and main function modules.

Pichon *et al.* [2] presented a general approach to evaluate the reliability of information sources; in this approach, the richness of expression, user engagement, and legitimacy of theme are combined to evaluate the reliability of typical sources, i.e., Twitter accounts. Although this model is a multi-attribute approach, in which the Choquet integral is used to combine different criteria, it uses a manual method to assign weights to criteria. The lack of adaptability to weight distribution for trust factors hinders the generation of accurate trust decisions in dynamic environments. Zhao *et al.* [4] proposed a topic-focused, similarity-based trust model to assess the trustworthiness of users on Twitter. Different from traditional graph-based trust ranking approaches, their method is not only scalable but also capable of considering the heterogeneous contextual properties of textual, temporal, and spatial features to rank the trustworthiness of tweets and users. However, the weights for each trust factor are assigned by human users, which is quite subjective and inflexible in a dynamic network environment.

III. PRELIMINARIES AND OVERVIEW

In this section, we first present the problem formulation and then give an overview of the proposed architecture.

A. PROBLEM FORMULATION

We consider the problem of calculating the trustworthiness of information sources. We define a directed graph $G \in \langle S, E \rangle$, where S is the set of information sources and $E \subseteq S \times S$ is the set of edges between information sources. Node $s_i \in S$ represents an information source, and edge $e_{ij} \in E$ means that source s_i follows source s_j. Let $M(s_i)$ be a set of information published by source s_i. Trustworthiness evaluation approaches and models are the core technologies of trust management. Before introducing the details of trustworthiness calculation mechanism, we first present some basic definitions on trustworthiness.

Definition 1 (Trustworthiness of Information Source): The trustworthiness of an information source s_i (denoted as T_i) is a measure of the degree to which an information source s_i is believed to provide information that conforms to fact. The most correlated synonyms of trustworthiness are credibility and believability.

Definition 2 (Value Domain of Trustworthiness): The trustworthiness of an information source is represented by continuous numerical values in the range of [0,1], with 1 representing full trust (upper bound) and with 0 denoting no trust (lower bound).

B. ARCHITECTURE OVERVIEW

Before introducing the details of the proposed trust calculation mechanism for information sources, we first present the basic architecture of the proposed mechanism. As depicted in Fig. 1, the proposed model, called Info-Trust, consists of two major modules labeled as follows: (1) trust evidence

acquisition module, and (2) multi-criteria trustworthiness calculation module. First, the profile information of information sources, such as registration age, is extracted to evaluate the degree of identity-based trust. Then, as a key part of direct evidence database, information sources' history posting together with its number of likes, shares, and comments, is collected in the form of real-time monitoring by streaming application program interface (API), search API, and other available tools [32]. In addition, information sources' social structure is acquired to evaluate the relation-based trust. Finally, user feedback information, such as the reports collected by the reporting platform of social media and the comment list of the information, is leveraged to calculate feedback-based trust, which is often ignored in other research.

As shown in Fig. 1, the trustworthiness value of information source i, marked as T_i, is calculated by the OWA-WMA combination algorithm. The fusion function includes four subfunctions, namely, identity factor (IF), behavior factor (BF), relation factor (RF), and feedback factor (FF). Thus, the trustworthiness of information source i can be determined by the following vector:

$$D = (T_i^I, T_i^B, T_i^R, T_i^F), \tag{1}$$

where T_i^I is used to evaluate source i's identity trust on the basis of the user profile. T_i^B is used to evaluate source i's trustworthiness on the basis of posting history, especially the social influence of fake information posted by source i. T_i^R is used to evaluate source i's trustworthiness on the basis of its social relation. T_i^F is used to evaluate source i's trustworthiness on the basis of the feedback information from users. In short, the trust evaluation vector includes four complementary factors, namely, identity trust factor T_i^I, behavior trust factor T_i^B, relation trust factor T_i^R, and feedback trust factor T_i^F. Through the OWA-WMA combination algorithm, these four trust factors can be adaptively combined into one overall trustworthiness metric.

Definition 3: In general, the overall trust degree (OTD) of information source s_i is calculated by the following equation:

$$T_i = W \times \widehat{D} = \sum_{j=1}^{4} w_j \times T_i^{X_j}, \tag{2}$$

where $W = (w_1, w_2, w_3, w_4)$ is the weight vector of trust factors, with $1 \geq w_1 \geq w_2 \geq w_3 \geq w_4 \geq 0$, $\sum_{j=1}^{4} w_j = 1$; we set $\widehat{D} = (T_i^{X_1}, T_i^{X_2}, T_i^{X_3}, T_i^{X_4})$ where $T_i^{X_j}$ is the j^{th} largest value in $D = \{T_i^I, T_i^B, T_i^R, T_i^F\}$.

In existing research, three kinds of subjective methods were used to assign values to these weights [4], [33], [34], i.e., random allocation, average weight, and expert opinion. However, these methods share a common drawback, that is, the lack of dynamic adaptability. Once the value of a weight is given, the value cannot be dynamically and adaptively adjusted. Therefore, allocating the values to $(w_1, \ldots, w_4)$ adaptively is one of the key tasks in current work. The OWA-WMA combination algorithm, which integrates an ordered weighted averaging (OWA) operator and a weighted moving average (WMA) model [17]–[20], not only considers the degree of varying influence among various data but also focuses on dynamic weighting problems. In the proposed mechanism, the OWA-WMA combination algorithm is used to assign weights to these trust factors. This capability allows the model to provide a detailed and accurate evaluation of the trust calculation process.

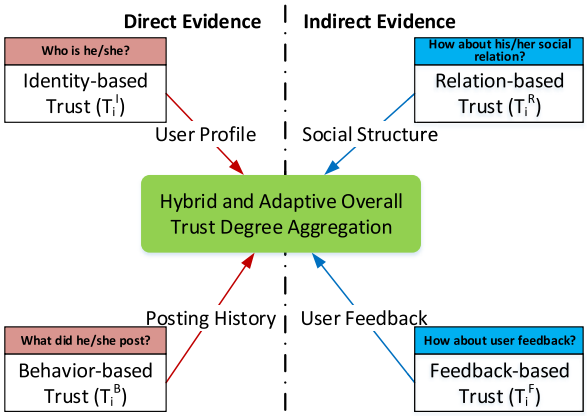

FIGURE 2. Proposed trust model aggregates comprehensive trust evidence to provide a full view of trustworthiness assessment of information sources.

IV. CALCULATION MECHANISM OF MULTI-CRITERIA TRUST FACTORS

A quantified trustworthiness value of an information source is an integrated opinion rating of multi-criteria. As mentioned in Section II, most existing trustworthiness evaluation approaches for information sources are not comprehensive enough because of various limitations, such as the disregard for feedback-based trust factors [3], which greatly hinder the acceptance of trustworthiness evaluation results in social media. In the present work, we present a generic model to aggregate comprehensive trust factors to give a full view trustworthiness assessment of information sources. The proposed model is called Info-Trust. As shown in Fig. 2, the four trust factors proposed in this paper are considered to be complementary to one another in the characterization of an information source. The advantage of this model is its capability of fusing evaluation results that combine identity-based and behavior-based trust factors, with relation-based and feedback-based trust factors simultaneously. Before the detailed presentation of the calculation mechanism, we first list the key notations and their descriptions in Table 2.

A. IDENTITY-BASED TRUST

Profile information on social media have been shown to be correlated with the trustworthiness of information sources [35]. Research has also shown that disinformation are likely to be created and spread by social bots [36], [37]. Generally, authenticated information sources are more trusted than anonymous information sources. Thus, we define authenticated score $AS(s_i)$ to quantify trustworthiness from

TABLE 2. Notations and their descriptions.

Notation	Description
S	set of information sources
E	set of edges between information sources
N	number of information sources
T_i^X	trustworthiness of source s_i in criterion X, where $X = \{I, B, R, F\}$
T_i	overall trust degree (*OTD*) of source s_i
W	weight vector of four trust factors, $W = \{w_1, w_2, w_3, w_4\}$
$AS(s_i)$	authenticated score of source s_i
$RS(s_i)$	registration age score of source s_i
$SP(s_i)$	social popularity of source s_i
$TS(s_i)$	authority score of source s_i
$Follower(i)$	follower set of source s_i
Q_i	fake information set of source s_i
$I_{i,f}$	total influence of source s_i's fake information f
$NoFlw(s_i)$	number of source s_i's followers
$NoLik(f)$	number of fake information f's likes
$NoShr(f)$	number of fake information f's shares
$NoMet(f)$	number of fake information f's mentions
$LC(s_i)$	local cluster coefficient of source s_i
e_{s_i}	sum of the in-degree and out-degree built by source s_i's neighbors
K_{s_i}	number of source s_i's neighbors
$BC(s_i)$	betweenness centrality of source s_i
δ_{st}	total number of shortest paths from vertex s to vertex t
$\delta_{st}(s_i)$	number of shortest paths from s to t that pass through vertex s_i
ϱ_i	number of positive feedback toward source s_i
ϑ_i	number of negative feedback toward source s_i
λ	situation parameter in the calculation of OWA-based weight vector

the perspective of whether source s_i is authenticated (set to 1, i.e., given a full mark) or anonymous (set to 0.2, i.e., given very low marks). Given the fact that a number of newly created accounts are created intentionally to spread disinformation such as social bots [36], sources that disseminate truthful information tend to have longer register time than those disseminate disinformation [38]. The registration age of source s_i, which is determined by the time range of source s_i's account register time with the current date, cannot be changed artificially and is relatively difficult for malicious sources to evade. Generally, the older the registration age, the more trustable it is. We then take the *registration age score* of source s_i as an evaluation factor, which is denoted as $RS(s_i)$ and can be calculated in the following formula:

$$RS(s_i) = \frac{R(s_i) - \mu_R}{\sigma_R}, \tag{3}$$

where $R(s_i)$ is the registration age of source s_i, μ_R represents the average registration age of all sources, and σ_R represents the standard deviation of the registration age of all sources.

In general, the *number of followers* of a source reflects its popularity and trustworthiness. A larger number of followers of a source (a larger in-degree) commonly implies that more users trust this source and would like to receive information from it. Thus, we calculate the *social popularity* of source s_i, denoted as $SP(s_i)$, using the following equation:

$$SP(s_i) = \frac{log(NoFlw(s_i) + 1)}{log(\max\limits_{s_j \in S}(NoFlw(s_j)) + 1)}. \tag{4}$$

Inspired by the PageRank algorithm [39], which calculates node authority on the basis of the topology of the entire webpage, we define the *authority score* of source s_i as follows:

$$TS(s_i) = d \times \sum_{s_j \in Follower(i)} \frac{TS(s_j)}{NoFlw(s_j)} + \frac{1-d}{N}, \tag{5}$$

where $Follower(i)$ is the follower set of information source s_i, $NoFlw(s_j)$ is the number of source s_j's followers, N is the total number of sources, and $d \in (0, 1)$ is the damping factor. Thus, identity-based trust T_i^I can be calculated with **Algorithm 1**.

Algorithm 1 Calculation of Identity-Based Trust

Require: S, E, and the set of sources' profile information.
Ensure: Sources' profile-based trust.

```
for each s_i ∈ S do
    if s_i is authenticated then
        AS(s_i) = 1;
    else
        AS(s_i) = 0.2;
    end if
    Calculate registration age score RS(s_i) using Eq. (3);
    Calculate social popularity SP(s_i) using Eq. (4);
    Calculate authority score TS(s_i) using Eq. (5);
    T_i^I = (AS(s_i) + RS(s_i) + SP(s_i) + TS(s_i))/4.
end for
```

B. BEHAVIOR-BASED TRUST

There is no doubt that the posting-behavior is a kind of significant direct evidence for the trustworthiness evaluation of sources in social media. Meanwhile, no significant difference exists between trustable and trustless sources in features such as "like count per post" and "share count per post" [40], but the same is not true for the features of fake information.

Definition 4: In the model presented here, the behavior-based trust of source s_i, which is denoted as T_i^B, considers the quantity and influence of source s_i's fake information history and is calculated as follows:

$$T_i^B = 1 - \frac{\sum_{f \in Q_i} I_{i,f}}{\sum_{i=1}^{N} \sum_{f \in Q_i} I_{i,f}}, \tag{6}$$

where Q_i is the fake information set of source s_i, f is a piece of fake information in Q_i, and I_f is the influence of fake information f.

On the basis of our observations, we consider the number of likes, shares, and mentions of fake information f as the best

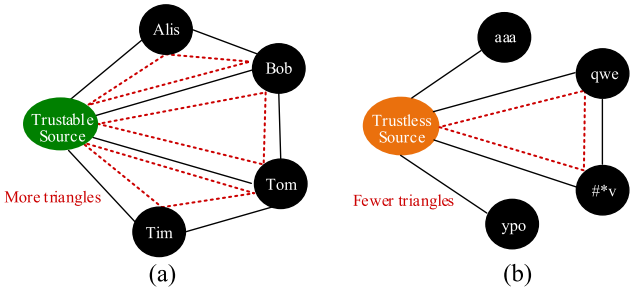

FIGURE 3. Illustration of the differences in local clustering coefficient between trustable sources and trustless sources. (a) Trustable sources, $LC(s_i) = \frac{2\times3}{4\times3} = \frac{1}{2}$. (b) Trustless sources, $LC(s_i) = \frac{2\times1}{4\times3} = \frac{1}{6}$.

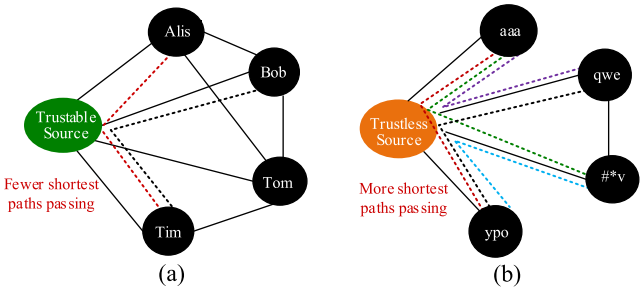

FIGURE 4. Illustration of the differences in betweenness centrality between trustable sources and trustless sources. (a) Trustable sources, $BC(s_i) = 2/C_4^2 = \frac{2}{6} = \frac{1}{3}$. (b) Trustless sources, $BC(s_i) = 5/C_4^2 = \frac{5}{6}$.

indicators from a quantitative perspective for the evaluation of I_f. First, we calculate the post influence of source s_i using the number of likes $NoLik(f)$, denoted as $LK_{i,f}$, as shown in Eq. (7). Similarly, we use the number of shares $NoShr(f)$ and the number of mentions $NoMet(f)$ to calculate the post influence of source s_i, denoted as $SH_{i,f}$ and $MT_{i,f}$ and as shown in Eq. (8) and Eq. (9), respectively. Finally, we obtain the total influence of source s_i's fake information f, as shown in Eq. (10).

$$LK_{i,f} = \frac{log(NoLik(f)+1)}{log(\max_{f\in Q_i}(NoLik(f))+1)}, \tag{7}$$

$$SH_{i,f} = \frac{log(NoShr(f)+1)}{log(\max_{f\in Q_i}(NoShr(f))+1)}, \tag{8}$$

$$MT_{i,f} = \frac{log(NoMet(f)+1)}{log(\max_{f\in Q_i}(NoMet(f))+1)}, \tag{9}$$

$$I_{i,f} = (LK_{i,f} + SH_{i,f} + MT_{i,f})/3. \tag{10}$$

C. RELATION-BASED TRUST

The common understanding is that trustable information sources are usually followed by family members, colleagues, and friends. Thus, these nodes are likely to have a strong relationship with one another. However, trustless sources usually blindly follow other nodes, which usually do not know one another and share a loose relationship [41]. To quantify how close a source's neighbors are to being a clique, we utilize a measure in graph theory, i.e., *local clustering coefficient* [42], which is determined by the proportion of links between nodes within the neighborhood divided by the number of links that could possibly exist between them [43]. Thus, for each information source s_i in the social media graph (i.e., source s_i is a vertex in the graph), its *local clustering coefficient* can be computed with the following equation:

$$LC(s_i) = \frac{|e_{s_i}|}{K_{s_i}\cdot(K_{s_i}-1)}, \tag{11}$$

where $|e_{s_i}|$ is the sum of the in-degree and out-degree for directed graphs (or twice the number of edges for undirected graphs) built by source s_i's neighbors and K_{s_i} is the number of source s_i's neighbors. As depicted in Fig. 3(a), the three triangles in different colors represent three relations between the green node's neighbors, and $LC(s_i) = \frac{2\times3}{4\times3} = \frac{1}{2}$. As depicted in Fig. 3(b), the red dotted triangle represents the only relation between the orange node's neighbors, and $LC(s_i) = \frac{2\times1}{4\times3} = \frac{1}{6}$. From Fig. 3 and Eq. (11), compared with trustable sources, trustless sources have lower values of local clustering coefficients.

Compared with trustable information sources, malicious sources will typically use a shotgun approach to find victims (i.e., randomly follow many unrelated sources to gain social relations), thereby creating more shortest paths between their following sources passing through them [41]. To quantify this feature, we leverage *betweenness centrality*, a measure of centrality in a graph based on shortest paths [44]. In a directed graph, the *betweenness centrality* of each vertex s_i can be calculated with Eq. (12):

$$BC(s_i) = \sum_{s\neq s_i\neq t} \frac{\delta_{st}(s_i)}{\delta_{st}}, \tag{12}$$

where δ_{st} is the total number of shortest paths from vertex s to vertex t and $\delta_{st}(s_i)$ is the number of shortest paths from s to t that pass through vertex s_i. This metric reflects the position of a vertex in the graph, and the vertices that occur on many shortest paths between vertices have higher values of betweenness centrality than those that do not. As depicted in Fig. 4(a), the red dashed line and the black dashed line represent two different shortest paths that pass through the green node, and $BC(s_i) = 2/C_4^2 = \frac{2}{6} = \frac{1}{3}$. As depicted in Fig. 4(b), dashed lines in five different colors represent five shortest paths that pass through the orange node, and $BC(s_i) = 5/C_4^2 = \frac{5}{6}$. From Fig. 4 and Eq. (12), compared with trustable sources, trustless sources have higher values of betweenness centrality.

Definition 5: In the model presented here, the relation-based trust factor of source s_i is defined as:

$$T_i^R = \frac{LC(s_i) + (1 - BC(s_i))}{2}. \tag{13}$$

Informed malicious information sources may be able to carefully choose the sources to follow and make their values of local clustering coefficient and betweenness centrality close to those of trustable sources. However, this process not only requires more time, money, and skills to implement but

also limits the number of their potential victims. Furthermore, considering that precisely calculating the values of such two graph metrics on large graphs (e.g., the whole Twitter graph) is time consuming, we leverage a neighbor-sampling technique, which allows us compute these metrics piece by piece, to calculate the metrics in an approximate and lightweight way.

D. LIGHTWEIGHT FEEDBACK-BASED TRUST

Given the large-scale social media network environment, which hosts millions of information sources and handles thousands of posts per second, the delay induced by trust systems could be a challenging problem. Hence, the feedback aggregation mechanism with high computational efficiency is the most fundamental requirement. In this work, we design a lightweight feedback aggregating mechanism.

Most social media platforms provide users with the function of reporting, and once users find something malicious with an information source, they could report it to the platform. Feedback from users, which is usually not given full attention or is even ignored, is vital to the evaluation of the trustworthiness of sources. In view of malicious feedback, we only take advantage of feedback from honest and trustable sources. In other words, only the feedback from reporters whose overall trust degree is no less than the predefined threshold (we set the value to 0.6 empirically) can be taken into account.

Definition 6: In the model presented here, the feedback-based trust T_i^F can be calculated with the following formula:

$$T_i^F = \frac{\varrho_i + 1}{\varrho_i + \vartheta_i + 2}, \text{where } T_i \geq 0.6, \tag{14}$$

where ϱ_i is the number of positive feedback toward information source s_i and ϑ_i is the number of negative feedback toward information source s_i.

In Eq. (14), when ($\varrho_i + \vartheta_i = 0$), the information source s_i has not received any feedback or has just joined the social media system. We set $T_i^F = 0.5$ when $\varrho_i + \vartheta_i = 0$. This idea is based on the research result in [45], in which the authors pointed out the inefficiency of suspecting new accounts because only a few users are malicious in the social media system. This approach can give new accounts a chance to join social media until they are proven to be malicious.

E. HYBRID AND ADAPTIVE TRUST AGGREGATION

After the identity-based trust factor T_i^I, behavior-based trust factor T_i^B, social relation-based trust factor T_i^R, and indirect feedback-based trust factor T_i^F are calculated separately for information sources, we need to integrate them together to obtain the aggregated *OTD*. An intuitive way is to average all the trust factors; however, the issue is that doing does not distinguish their levels of importance because not all aspects contribute to trustworthiness evaluation equally.

The OWA-WMA algorithm, which is a combination of an OWA operator and a WMA model [17], [18], considers the influence degrees among different factors and the dynamical weighting problems. Decision makers simply need to dynamically change the weights of input data according to the aggregation situation. The system then provides them the results of fusion computing. Therefore, the OWA-WMA algorithm is leveraged to weigh these trust factors in our presented trust mechanism. The OWA operator can assign different weights to different trust factors. Mathematically, the WMA model is the accumulation of the latest history of trust degrees with a moving average function.

Algorithm 2 Calculation of OWA-Based Weight Vector

Require: n, λ; /* For different n and λ, we can get different OWA weights; λ is the situation parameter. */
Ensure: Weight vector $W = (w_1, w_2, w_3, w_4)$.
1: **if** $\lambda < 0.5$ **then**
2: $\quad \lambda = 1 - \lambda$;
3: **end if**
4: Calculate w_1 according to Eq. (16);
5: Calculate w_n according to Eq. (17);
6: **for** $t = 2$ to $(n-1)$ **do**
7: $\quad$ Calculate w_t according to Eq. (18).
8: **end for**

Definition 7: Formally, an OWA operator with n dimensions is a mapping $F : R_n \rightarrow R$, which has an associated weight vector $W = \{w_1, w_2, \ldots, w_n\}$ in the unit interval and with a sum to one with the following equation:

$$F(p_1, p_2, \ldots, p_n) = \sum_{j=1}^{n} w_j p_{\sigma(j)}, \tag{15}$$

where $p_{\sigma(j)}$ is the j^{th} largest in the set $\{p_1, p_2, \ldots, p_n\}$.

To determine the value of weight vector W, we can leverage different aggregation operators. The OWA operator provides a parameterized class of mean-type aggregation operators [17]–[19]. The OWA operator is a non-linear operator that results from the process of determining w_j. Fuller and Majlender [46] proposed a Lagrange multipliers-based method to determine a special class of OWA operators having maximal entropy of the OWA weights, and derived a polinomial equation to determine the optimal weighting vector. The weights of $p_{\sigma(j)}$ can be calculated using the following equations:

$$w_1[(n-1)\lambda + 1 - nw_1]^n = [(n-1)\lambda]^{n-1}[((n-1)\lambda - n)w_1 + 1], \tag{16}$$

$$w_n = \frac{((n-1)\lambda - n)w_1 + 1}{(n-1)\lambda + 1 - nw_1}, \tag{17}$$

$$w_j = \sqrt[n-1]{w_1^{(n-j)} w_n^{(j-1)}}, \quad 1 \leq j \leq n. \tag{18}$$

In the above equations, parameter λ, whose range is between 0 and 1, is treated as a tool for the trust mechanism to determine the most important factor on the basis of set $\{p_1, p_2, \ldots, p_n\}$. According to [46], the optimal value of w_1 should satisfy Eq. (16). When w_1 is obtained, w_n can be calculated using Eq. (17), and the values of the other weights

TABLE 3. Mapping relation between overall trust degree and trust assessment level.

No.	Overall Trust Degree	Trust Assessment Level
1	[0.90,1.00]	Very High
2	[0.75,0.90)	High
3	[0.50,0.75)	Medium
4	[0.25,0.50)	Low
5	[0.00,0.25)	Very Low

can be calculated with Eq. (18). Subsequently, the OWA operator is used to calculate the weight vector, as depicted in **Algorithm** 2. If we set $n = 4$ in **Algorithm** 2, we will obtain the weights of the four trust factors $W = (w_1, w_2, w_3, w_4)$.

Definition 8: The WMA model has a specific meaning of weights that changes arithmetically; and it is defined as follows:

$$F(U) = \sum_{i=1}^{n} \omega_i U_i, \tag{19}$$

where $F(U)$ is the fusion function of series U, i is the number of data items used to calculate the weighted average, U_i is the actual data item, and ω_i is the weight allocated to U_i (with $\sum \omega_i = 1$). The four trust factors are obtained: $D = (T_i^I, T_i^B, T_i^R, T_i^F)$. Conceptually, let $U = D$, then Eq. (2) becomes a WMA model.

In brief, the calculation process of sources' *OTD* is shown in **Algorithm** 3. In addition, to ensure an intuitive understanding of source s_i's trustworthiness, as shown in Table 3, we define a simple mapping relation between overall trust degree and trust assessment level. The information source is deemed as trustless if its *OTD* value is lower than 0.25 and as trustable if its *OTD* value is no less than 0.5.

Algorithm 3 Calculation of Sources' *OTD*

1: **for** $s_i \in S$ **do**
2: Calculate identity-based trust T_i^I using **Algorithm 1**;
3: Calculate behavior-based trust T_i^B using Eq. (6);
4: Calculate relation-based trust T_i^R using Eq. (13);
5: Calculate feedback-based trust T_i^F using Eq. (14);
6: Calculate the weight vector W using **Algorithm 2**;
7: Calculate the *OTD* of source s_i, T_i using Eq. (2).
8: **end for**

Ensure: T_i // *OTD* of source s_i.

V. EVALUATIONS

To demonstrate the overall performance of our proposed multi-criteria and adaptive trust calculation mechanism in terms of accuracy and adaptability in handling dynamic malicious behavior, we conducted a set of extensive simulation experiments to evaluate the trustworthiness of Sina Weibo accounts on the basis of our crawled real-world dataset. The simulation is based on NetLogo [47], which is particularly suitable for modeling complex systems and exploring the connection between the micro-level behavior of peers and macro-level patterns. Performance was evaluated from two perspectives: (1) accuracy, which is used to assess the accuracy of the trustworthiness evaluation; (2) adaptability, which is used to evaluate our proposed mechanism's response capacity in dealing with sources' dynamic and complicated posting behavior. For comparison purpose, three typical trust models, direct trust model [2], Canini's trust model [3] and averaged weight model are also implemented in the simulator.

A. EXPERIMENTAL SETUP

To demonstrate the effectiveness of our proposed trust mechanism for social media, we conducted experiments on China's leading microblog service provider, Sina Weibo. Hundreds of millions of microblogs are posted in Sina Weibo by registered users. Therefore, Sina Weibo is an excellent case to evaluate the trustworthiness of information sources.

1) DATASET DESCRIPTION

In order to achieve our research goal, we need to create a rich dataset by crawling real social media data. In this study, to crawl Sina Weibo data, We have developed a Sina Weibo crawler that taps into Sina Weibo open API [48]. Given the sheer volume of Sina Weibo data and the restricted API, crawling all individual microblogs and evaluating the trustworthiness of all users are impractical. Therefore, we specially concentrated on and selected a number of users who had posted microblogs about information that was later confirmed to be rumor. The official Sina Weibo account, "Weibo Refutes Rumors," provides users with recent disinformation, making the labeling of our dataset of high-quality. We also asked two annotators to give overall trustworthiness scores to accounts on the given topic as actual values. We collected microblogs posted from July 2017 to September 2017. These microblogs involved topics about health science, deceptive advertising, and social events. To present a fine-grained evaluation [4], we divided the dataset into three parts on the basis of topics, and conducted experiments separately. The three datasets, namely, health science, advertising, social events, contain 15761, 16451, 44715 nodes respectively, and the average out-degrees are 61.57, 110.41, 113.49 respectively.

2) COMPARED METHODS

We compare the performance of Info-Trust with the following approaches:

- Direct trust model [2], which considers only direct trust factors, including identity-based and behavior-based trust factors.
- Canini's trust model [3], which leverages identity information, posting behavior, and social structure to find credible information sources in social networks. However, the feedback-based trust factor is ignored in this mechanism.
- Averaged weight model, which considers the four trust factors simultaneously but allocates their weights equally.

TABLE 4. Simulation parameters and their possible values.

Parameter	Description	Possible Values
I_d	Interval of dynamics	5, 10, 20
P_b	Proportion of benign sources	20%, 50%
P_m	Proportion of malicious sources	20%, 50%
W	Weights in models	0.25, 0.3, 0.5
λ	Parameter in **Algorithm** 2	[0.5,1]

3) SIMULATOR SETTINGS

In the simulator, there are benign sources and malicious sources. Benign sources always publish trustable information, but malicious sources change their posting quality in a specific interval (I_d). Decreasing the value of I_d can make a malicious source change its posting quality faster. P_b is the proportion of benign sources, while P_m is the proportion of malicious sources in social media. Time-step is the running steps of the simulator. The parameters discussed above and their possible values used in the simulator are summarized in Table 4.

B. ACCURACY EVALUATION

All trust calculation mechanisms should have good accuracy in trustworthiness evaluation. In this work, we use the mean absolute deviation (MAD) to evaluate the accuracy of the proposed model [49].

$$\Gamma(t) = \frac{\sum |A_t - F_t|}{\sum t}, \tag{20}$$

where A_t is the actual value calculated at time stamp t, F_t is the predicted value of Info-Trust calculated at timestamp t and $|A_t - F_t|$ is the evaluation error at time-stamp t, $\sum t$ is the total running timestamps. $\Gamma(t)$ is an indicator of accuracy in the trust evaluation and is used to check whether an error is within the acceptable control limit. The closer the value is to zero, the better the trust model.

To analyze the impact of different values of situation parameter λ in **Algorithm** 2, we conducted experiments on the basis of the three datasets. Fig. 5 depicts the experimental results. When situation parameter $\lambda = 0.6$, we get the optimum value of MAD. Therefore, in the later experiments, we set the value of λ to 0.6 as the basic value of situation parameter.

An experiment we are interested in is an examination of how trust is changed when some factors, especially social structure-based and feedback-based trust factors, are disregarded. Table 5 presents trust factors used by different trust models and examples of weights. As depicted in Table 5, the direct trust model, as a representative two-dimensional trust model, considers identity-based and behavior-based trust factors in trust evaluation, and the weight of each factor is artificially configured. In Canini's trust model [3], three trust factors, namely, identity-based trust factor, behavior-based trust factor, and relation-based trust factor, are considered except feedback-based trust factor. What's more, to measure the effectiveness of Info-Trust's adaptive weights,

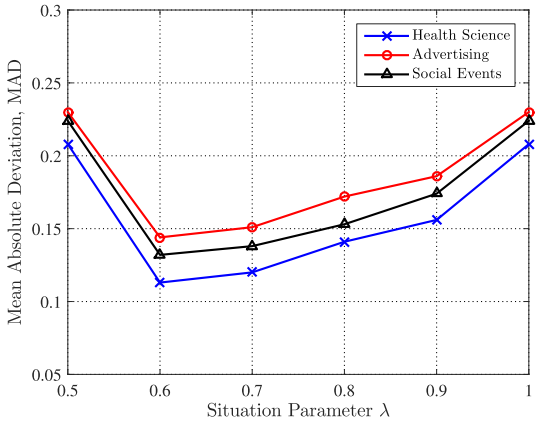

FIGURE 5. MAD under varying situation parameter λ.

TABLE 5. Trust factors used by different trust models and examples of weights.

Weights / Factors / Models	T_i^I	T_i^B	T_i^R	T_i^F
Direct trust model [2]	0.5000	0.5000	-	-
Canini's trust model [3]	0.4000	0.3000	0.3000	-
Averaged weight model	0.2500	0.2500	0.2500	0.2500
proposed Info-Trust model	0.3474	0.2722	0.2133	0.1671

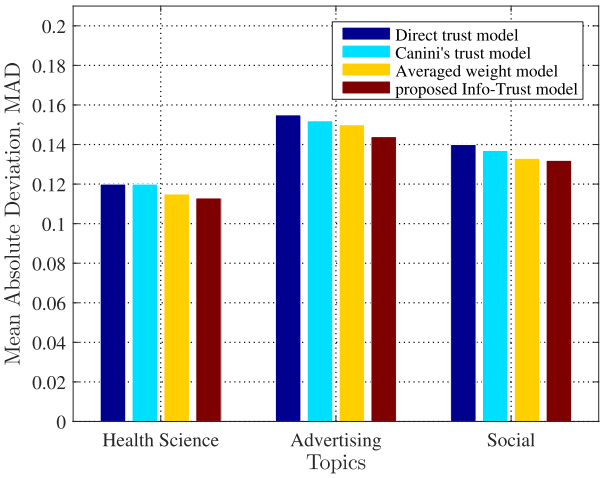

FIGURE 6. Comparison of mean absolute deviation with different numbers of trust factors under different topics in a relatively stable community.

we consider a four-dimensional trust model, in which the weights are equally set to 0.25, i.e., not a dynamic weight allocation. From the previous introduction, proposed Info-Trust model contains four trust factors, whose weights are dynamically calculated by **Algorithm 2**. The last line of Table 5 depicts one of the weight scenarios.

The mean absolute deviation Γ in this work can reflect the unbiasedness of the trust calculation mechanism. A small value of $\Gamma(t)$ indicates that the trust calculation mechanism achieves satisfactory and unbiased accuracy. Figs. 6 and 7 describe the experimental results of $\Gamma(t)$ with different

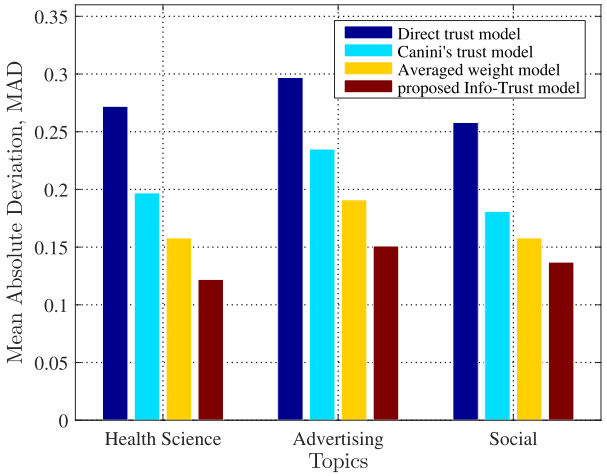

FIGURE 7. Comparison of mean absolute deviation with different numbers of trust factors under different topics in a malicious community.

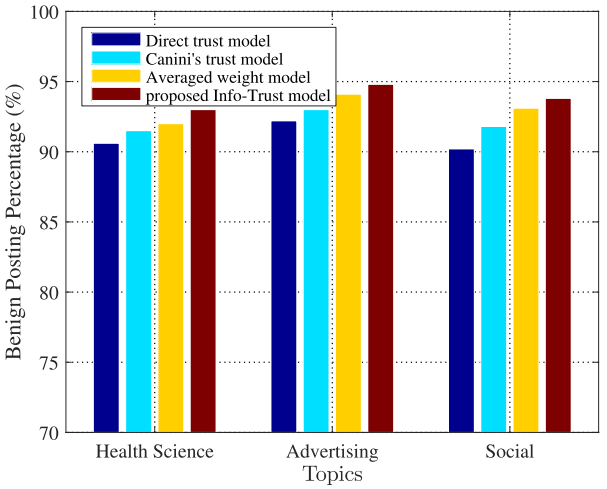

FIGURE 8. Simulation results of benign posting percentage under an idle and stable network environment. (PF = 0.2, PDF = 0.2).

numbers of trust factors. Obviously, the number of trust factors has a direct impact on the accuracy of the trust model. In Figs. 6 and 7, if the simulator leverages all the four trust factors, the value of $\Gamma(t)$ is smaller than the other two models. With such a rich set of trust factors, a highly accurate trust evaluation can be achieved.

Fig. 6 depicts the experimental results with different trust factors under a relatively stable environment. In the simulation, the total percentage of malicious information sources is 20%, which indicates that the community is a relatively good community (i.e., with few malicious sources). As shown in Fig. 6, all these models have a relatively close performance (the difference is less than 0.1), which reflects that these mechanisms perform well when faced with few malicious sources.

Fig. 7 depicts the comparison of the results of $\Gamma(t)$ in a malicious community. In the simulation, the total percentage of malicious sources is 50%, which indicates that the community is an uncomfortable community. Fig. 7 clearly shows that in a malicious environment, the number of trust factors has a direct impact on the accuracy of the trust calculation mechanism. In Fig. 7, if the simulator uses the trust mechanism with all four trust factors, then the value of $\Gamma(t)$ is the lowest one among the four models. This outcome implies that our multi-criteria trust mechanism retains its robust service capability in a malicious environment. As an example, under the condition of ''advertising'' topic, the MAD of our method is 0.151, which is smaller than that of the direct trust model whose $\Gamma(t)$ value is 0.297. As shown in Figs. 6 and 7, the proposed multi-criteria trust mechanism performs better than the other mechanisms do, as we expected. From the perspective of application, the computation of multiple trust factors results in increased computational overhead and thus affects the overall performance of the trust system. However, in view of the significant improvement in the credibility and security of a social media system, an additional overhead can be negligible.

C. ADAPTABILITY EVALUATION

Generally, the dynamics of social media networks are caused by the following two parts: the dynamism of sources' posting behavior (normal information sources are compromised by malicious ones and start to provide malicious information service), and the dynamism of sources' posting quality. In the simulator, two parameters are used to reflect the dynamics of the simulated social media system. (1) Posting frequency (PF, in the range of [0,1]). Each information source publishes a post with a PF. A greater PF means that more postings will be published in a certain amount of time. (2) Posting dynamic factor (PDF). After a random time, i.e., several time stamps in the simulation, the information sources oscillate to publish either benign or malicious posts.

Actually, a high benign posting percentage (BPP) reflects the system's good adaptability. Thus, we define BPP, denoted as $\varphi(\Delta t)$, and leverage the following function to evaluate the adaptability of these trust mechanisms,

$$\varphi(\Delta t) = \frac{\sum_{t=1}^{\Delta t} B(\Delta t)}{\sum_{t=1}^{\Delta t} S(\Delta t)} \times 100\%, \tag{21}$$

where $B(\Delta t)$ is the total number of benign postings counted by the simulator in period Δt and $S(\Delta t)$ is the total number of postings in Δt.

In the experiments, the proportion of honest feedback raters (HFRs) is set to 80%, and the proportion of malicious feedback raters (MFRs) is set to 20%. These settings are consistent with the actual social media system. In a real social media system, most participants are honest (HFR = 80%), and only a small number of participants are malicious. According to the following four network environments, the relevant issues will be discussed: idle and stable environment, busy and stable environment, idle and dynamic environment, and busy and dynamic environment.

We first observe the performance in the case of an idle and stable environment, where the dynamic factors are PF = 0.2 and PDF = 0.2. Fig. 8 shows that the four

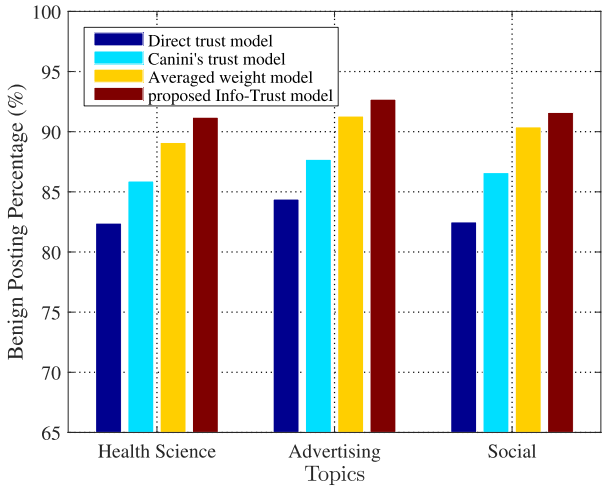

FIGURE 9. Simulation results of benign posting percentage under a busy and stable network environment. (PF = 0.8, PDF = 0.2).

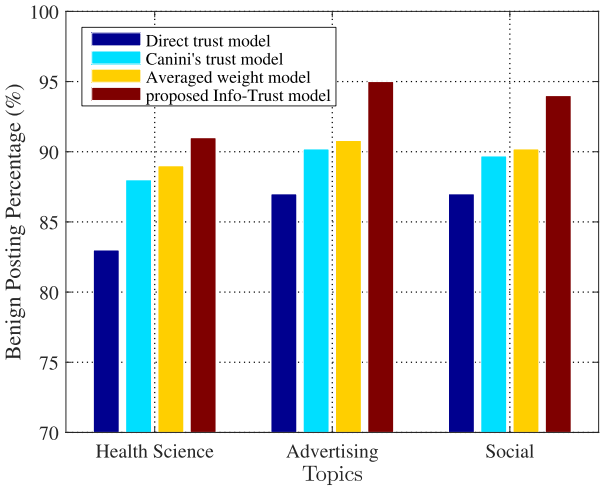

FIGURE 10. Simulation results of benign posting percentage under an idle and dynamic network environment. (PF = 0.2, PDF = 0.8).

models exhibit good robustness, with the values of $\varphi(\Delta t)$ exceeding 90%. However, $\varphi(\Delta t)$ of our proposed model is slightly higher than that of the other three models. Fig. 9 depicts the simulation results under a busy and stable environment, in which the dynamic factors are PF = 0.8 and PDF = 0.2. In such a network environment, $\varphi(\Delta t)$ of the direct trust model decreases by 8%, that of Canini's trust model decreases by 5%, and that of the averaged weight model decreases by 3%. The decrease in the $\varphi(\Delta t)$ value in the proposed model is less than 2%, which reflects that it is more robust than the other three models are under a busy but stable environment.

To study adaptability under a highly dynamic environment, the PDF is set to 0.8. Fig. 10 shows that in an idle and highly dynamic environment, where PF = 0.2 and PDF = 0.8, our proposed mechanism has the highest adaptability, and its $\varphi(\Delta t)$ reaches 93%. When we set the value of PF to 0.8, the social media system is not only a highly dynamic system but also a busy system. As depicted in Fig. 11, where PF = 0.8 and PDF = 0.8, compared with the direct trust model and Canini's model [3], our proposed multi-criteria trust mechanism has higher BPP under such a highly dynamic environment. However, compared with the results in the idle and highly dynamic environment, the performance of each model in the busy and highly dynamic environment obviously decreases.

An application-level trust calculation mechanism should be able to quickly respond to malicious behavior. To further evaluate the adaptability of the proposed mechanism, we consider the case in which a good information source suddenly becomes malicious. In this group of experiments, the total number of observations is set as 100, and the values of situation parameter λ are set to 0.5, 0.6, and 0.7. Fig. 12 shows how our multi-criteria trust model describes and predicts the dynamic change of *OTD* values under different situation parameters λ. The *OTD* of the information source drops quickly when it is detected posting fake information.

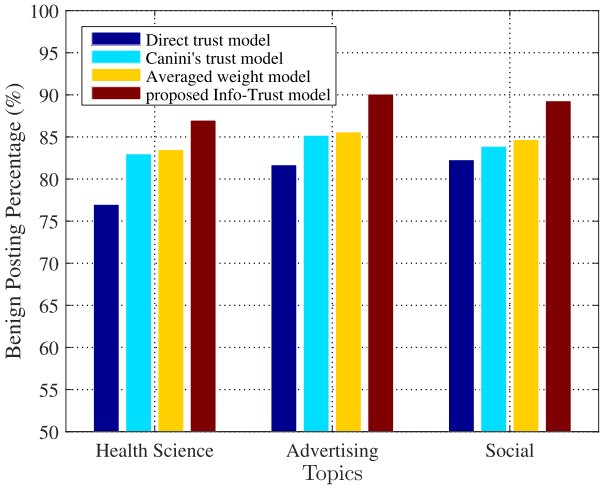

FIGURE 11. Simulation results of benign posting percentage under a busy and dynamic network environment. (PF = 0.8, PDF = 0.8).

As shown in Fig. 12, after $t = 50$, the smaller the value of λ, the lower the value of the *OTD* due to the fake information's influence on behavior-based trust T_i^B and feedback-based trust T_i^F. The results show that our method can robustly identify information sources' dynamic posting behavior.

Through the comparison of these results, the proposed Info-Trust model is found to have a robust adaptability in a stable network community and in a dynamic network community. The main reason for this difference is that in the direct trust model and Canini's trust model, subjective methods are used to weigh the trust factors. However, these methods cannot capture the adaptability and complexity of the trust evaluation process. Consequently, they may lead to misinformation and hinder an accurate trustworthiness evaluation. In the model presented in this work, the OWA-WMA combination algorithm is innovatively used to weigh the multiple trust factors. This algorithm can overcome the limitation of weight allocation in the other two models. Therefore, with the

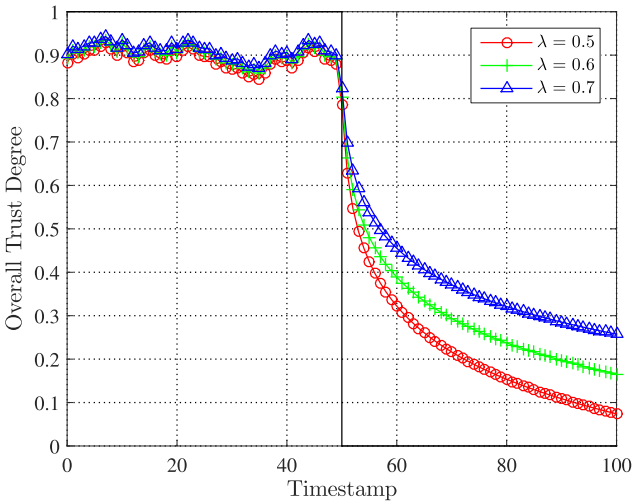

FIGURE 12. Overall trust degree of a good information source that turns malicious at $t = 50$ with different situation parameters λ.

adaptive weighting method, the accuracy of trust evaluation substantially improves. In sum, the results of this model based on multi-criteria trust factors are considerably better than of those models with simple trust factors.

VI. CONCLUSION AND FUTURE WORK

In this paper, a multi-criteria trustworthiness calculation mechanism called Info-Trust is proposed for information sources, in which identity-based trust, behavior-based trust, relation-based trust, and feedback-based trust factors are incorporated to present a accuracy-enhanced full view of trustworthiness evaluation of information sources. This method surpasses the limitations of existing approaches in which weights are assigned subjectively. The results of extensible simulation experiments on the basis of real-world dataset demonstrate that the proposed Info-Trust yields good results in terms of evaluation accuracy and adaptability in trustworthiness identification of network information.

As for future work, we are interested in motivating social users to submit their feedback to the user feedback platform and in implementing our proposed mechanism in other social media systems such as Twitter to verify the universality of the proposed mechanism. Another research direction we wish to explore is the consideration of other sophisticated attacks, such as random, insidious, and opportunistic attack behavior, to test the robustness of our trust calculation framework.

REFERENCES

[1] L. Ge, J. Gao, X. Li, and A. Zhang, "Multi-source deep learning for information trustworthiness estimation," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 766–774.

[2] F. Pichon, C. Labreuche, B. Duqueroie, and T. Delavallade, "Multidimensional approach to reliability evaluation of information sources," in *Wiley Online Library, Information Evaluation*. 2014, pp. 129–159.

[3] 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 (PASSAT) IEEE 3rd Int. Conf. Social Comput. (SocialCom)*, Oct. 2011, pp. 1–8.

[4] L. Zhao, T. Hua, C.-T. Lu, and R. Chen, "A topic-focused trust model for Twitter," *Comput. Commun.*, vol. 76, pp. 1–11, Feb. 2016.

[5] T. Lucassen and J. M. Schraagen, "Trust in Wikipedia: How users trust information from an unknown source," in *Proc. 4th Workshop Inf. Credibility*, 2010, pp. 19–26.

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

[7] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "Towards detecting compromised accounts on social networks," *IEEE Trans. Dependable Secure Comput.*, vol. 14, no. 4, pp. 447–460, Jul. 2017.

[8] K. S. Adewole, N. B. Anuar, A. Kamsin, K. D. Varathan, and S. A. Razak, "Malicious accounts: Dark of the social networks," *J. Netw. Comput. Appl.*, vol. 79, pp. 41–67, Feb. 2017.

[9] Z. Wang, W. Dong, W. Zhang, and C.-W. Tan, "Rooting our rumor sources in online social networks: The value of diversity from multiple observations," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 663–677, Jun. 2015.

[10] J. A. Golbeck, "Computing and applying trust in Web-based social networks," Ph.D. dissertation, Univ. Maryland, Dept. Comput. Sci., College Park, MD, USA, 2005.

[11] X. Li, F. Zhou, and X. Yang, "A multi-dimensional trust evaluation model for large-scale P2P computing," *J. Parallel Distrib. Comput.*, vol. 71, no. 6, pp. 837–847, 2011.

[12] X. Li, H. Ma, F. Zhou, and X. Gui, "Service operator-aware trust scheme for resource matchmaking across multiple clouds," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 4, pp. 1419–1429, May 2014.

[13] X. Li, F. Zhou, and J. Du, "LDTS: A lightweight and dependable trust system for clustered wireless sensor networks," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 924–935, Jun. 2013.

[14] W. Sherchan, S. Nepal, and C. Paris, "A survey of trust in social networks," *ACM Comput. Surv.*, vol. 45, no. 4, p. 47, 2013.

[15] W. Jiang, G. Wang, M. Z. A. Bhuiyan, and J. Wu, "Understanding graph-based trust evaluation in online social networks: Methodologies and challenges," *ACM Comput. Surv.*, vol. 49, no. 1, p. 10, 2016.

[16] X. Li, H. Ma, W. Yao, and X. Gui, "Data-driven and feedback-enhanced trust computing pattern for large-scale multi-cloud collaborative services," *IEEE Trans. Services Comput.*, vol. 11, no. 4, pp. 671–684, Jul. 2018.

[17] A. Emrouznejad and M. Marra, "Ordered weighted averaging operators 1988–2014: A citation-based literature survey," *Int. J. Intell. Syst.*, vol. 29, no. 11, pp. 994–1014, 2014.

[18] R. R. Yager and J. Kacprzyk, Eds., *The Ordered Weighted Averaging Operators: Theory And Applications*. Springer, 2012.

[19] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-18, no. 1, pp. 183–190, Jan. 1988.

[20] B. S. Ahn, "On the properties of OWA operator weights functions with constant level of orness," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 4, pp. 511–515, Aug. 2006.

[21] S. Adali *et al.*, "Measuring behavioral trust in social networks," in *Proc. IEEE Int. Conf. Intell. Secur. Inform. (ISI)*, May 2010, pp. 150–152.

[22] J. Jia, B. Wang, and N. Z. Gong, "Random walk based fake account detection in online social networks," in *Proc. 47th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw. (DSN)*, Jun. 2017, pp. 273–284.

[23] B. Wang, L. Zhang, and N. Z. Gong, "SybilSCAR: Sybil detection in online social networks via local rule based propagation," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.

[24] W. Jiang, G. Wang, and J. Wu, "Generating trusted graphs for trust evaluation in online social networks," *Future Gener. Comput. Syst.*, vol. 31, pp. 48–58, Feb. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X1200146X

[25] B. Fang, Y. Jia, X. Li, A. Li, and X. Wu, "Big search in cyberspace," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1793–1805, Sep. 2017.

[26] B. Wang, N. Z. Gong, and H. Fu, "GANG: Detecting fraudulent users in online social networks via guilt-by-association on directed graphs," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 465–474.

[27] Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai, "Uncovering social network sybils in the wild," *ACM Trans. Knowl. Discovery Data*, vol. 8, no. 1, p. 2, 2014.

[28] Q. Cao, X. Yang, J. Yu, and C. Palow, "Uncovering large groups of active malicious accounts in online social networks," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2014, pp. 477–488.

[29] Y. Zhang, H. Chen, and Z. Wu, "A social network-based trust model for the semantic Web," in *Proc. Int. Conf. Auton. Trusted Comput.*, 2006, pp. 183–192.

[30] X. Ruan, Z. Wu, H. Wang, and S. Jajodia, "Profiling online social behaviors for compromised account detection," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 176–187, Jan. 2016.
[31] X. Li, H. Ma, F. Zhou, and W. Yao, "T-broker: A trust-aware service brokering scheme for multiple cloud collaborative services," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 7, pp. 1402–1415, Jul. 2015.
[32] M. Al-Qurishi, M. Al-Rakhami, M. Alrubaian, A. Alarifi, S. M. M. Rahman, and A. Alamri, "Selecting the best open source tools for collecting and visualzing social media content," in *Proc. 2nd World Symp. Web Appl. Netw. (WSWAN)*, Mar. 2015, pp. 1–6.
[33] 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.
[34] 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.
[35] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Apr. 2018, pp. 430–435.
[36] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newslett.*, vol. 19, no. 1, pp. 22–36, 2017.
[37] C. Shao, G. L. Ciampaglia, O. Varol, A. Flammini, and F. Menczer. (2017). "The spread of fake news by social bots." [Online]. Available: https://arxiv.org/abs/1707.07592
[38] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu. (2018). "FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media." [Online]. Available: https://arxiv.org/abs/1809.01286
[39] *Pagerank*. Accessed: Sep. 3, 2018. [Online]. Available: https://en.wikipedia.org/wiki/PageRank
[40] M. Alrubaian, M. Al-Qurishi, M. Al-Rakhami, M. M. Hassan, and A. Alamri, "Reputation-based credibility analysis of Twitter social network users," *Concurrency Comput., Pract. Exper.*, vol. 29, no. 7, p. e3873, 2017.
[41] C. Yang, R. Harkreader, and G. Gu, "Empirical evaluation and new design for fighting evolving Twitter spammers," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 8, pp. 1280–1293, Aug. 2013.
[42] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.
[43] (2018). *Clustering Coefficient*. Accessed: Sep. 3, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Clustering_coefficient
[44] (2018). *Betweenness Centrality*. Accessed: Sep. 3, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Betweenness_centrality
[45] P. Resnick and E. J. Friedman, "The social cost of cheap pseudonyms," *J. Econ. Manage. Strategy*, vol. 10, no. 2, pp. 173–199, 2001.
[46] R. Fullér and P. Majlender, "An analytic approach for obtaining maximal entropy OWA operator weights," *Math. Preprint Arch.*, vol. 124, pp. 53–57, Nov. 2001.
[47] U. Wilensky. *Netlogo*. Accessed: Sep. 2018. [Online]. Available: https://ccl.northwestern.edu/netlogo/
[48] *Sina Weibo Api Documation*. Accessed: Sep. 2017. [Online]. Available: https://open.weibo.com/wiki/API%E6%96%87%E6%A1%A3/en#Weibo_Access_API
[49] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series—Part I," *Fuzzy Sets Syst.*, vol. 54, no. 1, pp. 1–9, 1993.

YALI GAO is currently pursuing the Ph.D. degree with the School of Software, Beijing University of Posts and Telecommunications, China. She has published some papers in journals and conference proceedings. Her current research interests include social networks, network security, and trusted systems.

XIAOYONG LI received the Ph.D. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 2009. He is currently a Professor of computer science with the Beijing University of Posts and Telecommunications. His current research interests include cloud computing, network security, and trusted systems. As the first author, he has published over 60 papers in journals and conference proceedings. He has obtained five patents and five software copyrights in network security, cloud computing, and other fields. In 2009, he was awarded an Outstanding Doctoral Graduate in Shaanxi Province, China. In 2012, he was awarded a New Century Excellent Talent in University, China. In 2015, he was a recipient of the IET Information Security Premium Award.

JIRUI LI received the M.S. degree in computer application technology from the Wuhan University of Technology, Wuhan, China, in 2006. She is currently pursuing the Ph.D. degree with the School of Cyberspace Security, Beijing University of Posts and Telecommunications, China. She has published over 20 papers in journals and conference proceedings, and participated in several projects above the provincial level. Her current research interests include mobile cloud computing, distributed computing and trusted services, and the Internet of Things.

YUNQUAN GAO received the M.S. degree in computer software and theory from South China Normal University, Guangzhou, China, in 2006. He is currently pursuing the Ph.D. degree with the School of Cyberspace Security, Beijing University of Posts and Telecommunications, China. His current research interests include the Internet of Things, and trusted services.

PHILIP S. YU received the B.S. degree in electrical engineering from National Taiwan University, the M.S. and Ph.D. degrees in EE from Stanford University, and the MBA degree from New York University. He is currently a Distinguished Professor of computer science with the University of Illinois at Chicago (UIC), and also holds the Wexler Chair in information technology. Before joining UIC, he was with IBM, where he was the Manager of the Software Tools and Techniques Group, Watson Research Center. His research interest is on big data, including data mining, data stream, database, and privacy. He has published more than 970 papers in refereed journals and conferences. He holds or has applied for over 300 US patents. He was a member of the Steering Committee of the IEEE Data Engineering and the IEEE Conference on Data Mining. He is a Fellow of the ACM and the IEEE. He is on the Steering Committee of the ACM Conference on Information and Knowledge Management. He received the ACM SIGKDD 2016 Innovation Award for his influential research and scientific contributions on mining, fusion, and anonymization of big data, the IEEE Computer Society's 2013 Technical Achievement Award for "pioneering and fundamentally innovative contributions to the scalable indexing, querying, searching, mining, and anonymization of big data", and the Research Contributions Award from the IEEE International Conference on Data Mining (ICDM), in 2003, for his pioneering contributions to the field of data mining. He also received the ICDM 2013 10-year Highest-Impact Paper Award, and the EDBT Test of Time Award (2014). He has received several IBM honors, including two IBM Outstanding Innovation Awards, an Outstanding Technical Achievement Award, two Research Division Awards, and the 94th plateau of Invention Achievement Awards. He was an IBM Master Inventor. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING (2001–2004), and is the Editor-in-Chief of the *ACM Transactions on Knowledge Discovery from Data*.