Trustworthiness Assessment of Users in Social Reviewing Systems

Christian Esposito, Vincenzo Moscato, and Giancarlo Sperlí

***Abstract*—Social Networks represent a cornerstone of our daily life, where the so-called social reviewing systems (SRSs) play a key role in our daily lives and are used to access data typically in the form of reviews. Due to their importance, social networks must be trustworthy and secure, so that their shared information can be used by the people without any concerns, and must be protected against possible attacks and misuses. One of the most critical attacks against the reputation system is represented by mendacious reviews. As this kind of attacks can be conducted by legitimate users of the network, a particularly powerful solution is to exploit trust management, by assigning a trust degree to users, so that people can weigh the gathered data based on such trust degrees. Trust management within the context of SRSs is particularly challenging, as determining incorrect behaviors is subjective and hard to be fully automatized. Several attempts in the current literature have been proposed; however, such an issue is still far from been completely resolved. In this study, we propose a solution against mendacious reviews that combines fuzzy logic and the theory of evidence by modeling trust management as a multicriteria multiexpert decision making and exploiting the novel concept of time-dependent and content-dependent crown consensus. We empirically proved that our approach outperforms the main related works approaches, also in dealing with sockpuppet attacks.**

***Index Terms*—Credibility, Dempster–Shafer (D-S) theory, feature ranking, fuzzy logic, multicriteria decision making, reputation, social reviewing systems (SRSs), user trustworthiness.**

I. Introduction

AS WELL known, the online social networks [1] are Internet-enabled applications used by people to establish social relations with the other individuals sharing similar personal interests and/or activities. Apart from exchanging personal data, such as photographs or videos, mainly all these applications allow their users to share comments and opinions on specific topics, so as to suggest objects or places of interest (e.g., TripAdvisor, Foursquare, etc.) or to provide social environments able to facilitate particular tasks (e.g., the search of a job as in LinkedIn, the answer to research questions as in ResearchGate, purchases on Amazon, etc.). Due to this comment/opinion sharing, these social applications, which we will refer to as social reviewing systems (SRSs) have been extensively used when people need to make daily decisions, increasing their popularity. As a concrete example, most of us access to a preferable SRS before choosing a restaurant or buying something so as to get reviews and feedback. People are progressively and symbiotically dependent on them as proved by the advanced opinion modeling and analysis, exploiting the impact of neighbors on user preferences or approaching the existing information overload in SRS, such as [2], [3]. For this reason, the trustworthiness of SRS is particularly important, and a key concern for effective opinion dynamics and trust propagation within a community of users [4]. In fact, SRSs suffer from forged messages and camouflaged/fake users that are able to avoid individuals take the right decision. This may raise several issues about privacy and security [5], mainly due to the fact that several personal and sensitive information are shared, and leaked, throughout SRS [6], [7], and that a person may choose to hide its true self and intentions behind a totally false virtual identity [8] or a Bot (short for software robots) may mimic human behavior in SRS [9]. In addition, threats in SRS, such as data leaks, phishing bait, information tampering, and so on, are never limited to a given social actor, but spread across the network like an infection by obtaining victims among the friends of the infested actors. So, an SRS provider needs to provide proper protection means to guarantee its trustworthiness.

Some works in the current literature, such as [10], mostly deal only with forging messages as this can be easily resolved by using cryptography. However, the second kind of malicious behavior caused by camouflaged/fake users is still an open issue. During the last decade, several solutions have been proposed in order to deal with the problem of camouflaged/fake users [11]–[13]. The issue of providing privacy has led to the adoption of access control means, while counteracting forging nodes/identities and social links/connections demanded authentication of users and exchanged messages [14], [15]. Mostly, such mechanisms aim at approaching external attackers or intruders, while thwarting legitimate participants in the SRS acting in a malicious way is extremely challenging. A naive way to protect against malicious individuals is to have users being careful when choosing with whom to have a relationship. Two users in social networks may have various kinds of relationships: 1) in Facebook-like systems users can indicate others as "friends," or 2) in Instagram-like systems a user can "follow" others.

Manuscript received March 16, 2020; revised August 22, 2020 and October 9, 2020; accepted December 28, 2020. This article was recommended by Associate Editor E. Chen. *(Corresponding author: Giancarlo Sperlí.)*

Christian Esposito is with the Department of Computer Science, University of Salerno, 84084 Fisciano, Italy (e-mail: esposito@unisa.it).

Vincenzo Moscato and Giancarlo Sperlí are with the Department of Electrical and Information Technology, University of Naples "Federico II," 80125 Naples, Italy (e-mail: vincenzo.moscato@unina.it; giancarlo.sperli@unina.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSMC.2020.3049082.

Digital Object Identifier 10.1109/TSMC.2020.3049082

However, users are typically not so careful when accepting received joining requests, and selecting other users to be connected with is typically extremely difficult (as malicious users are also experts in camouflaging themselves). Despite the relationships among the social actors within an SRS should be based on the direct knowledge in the real life of the people behind such actors (such as former classmates, colleagues, or member of the same family or group of friends), the majority of the relationships are typically made without such a face-to-face knowledge but among users that have never been met in person. *Trust management* is among the most popular solution to fight against such inside attackers [16]. It consists to assign a "trust" value to users based on the direct analysis of their behaviors or indirect trust relationship among social actors. To this aim, it is a soft secure measure implying the revocation of a social link toward those actors with a low trust value, or to strengthen the protection measures for those actors exhibiting a low trust degree, by limiting the data/functionalities that they can have access to. Despite being a powerful protection means [17], trust management is not explicitly provided by the main SRS platforms, due to the issues related to its automatic computation.

There is the problem to select the data of interest upon which computing the trust degree among the vast amount of shared information, which shows the main features (volume, variety, and velocity) of *big data*. To simplify the problem, a well-investigated aspect is the study of trust network [18]. Specifically, an SRS is seen as a graph, where each vertex is a social actor, and each link models a social relationship between two actors where a trust value is assigned by one to another by means of the previous computation approach [19], [20]. It is not rare that actors may interact with nonadjacent other actors, so it is important to find a trust path among nonadjacent actors to compute trust transitivity (so that they can interact). However, there is still the problem on how measuring the mutual trust of two actors connected by a social relationship, which is further used for trust transitivity for unrelated users having some related users in common. This can be roughly measured as the ratio of the good iterations over the total number of elapsed interactions, even if more complex models have been proposed. Possible violations against the ethical norms cannot be objectively determined (meaning that such judgments are absolutely true or false and it is possible to assign them a 0 or 1 value) but are strongly based on or influenced by the person making the judgment (i.e., are subjective) and expressed in a partial truth way (i.e., judgement can range from completely true to completely false and it is possible to assign a real number going from 0 to 1), due to uncertain and vague natures of the behavioral data collected on human interactions. This makes the overall trust management an example of the so-called fuzzy decision-making problem [21] and make its fully automation extremely complex. Protecting the overall aggregation from the impact of malicious or fake reputations is an issue to consider [22], and falls within the literature of security and privacy of *Recommender Systems* [23].

Our work aims to contribute to the on-going efforts on the trust computation in SRSs, where the behaviors of social actors are described by means of their submitted "reviews" on specific objects of interest. Specifically, a genuine review may reflect a correct behavior of the actor, while a deceitful review is a sign of a malicious behavior. To deal with the mentioned big data problem, we have considered those social application to recommend objects of interest, since they restrict the kind of behavioral data to reviews as text content in natural language. Despite the current SRS providers are reluctant to disclose their data (since mainly sensitive for their users and/or business), reviews are publicly accessible and user review datasets are largely available. To deal with the subjective review judgement, we leverage on the fuzzy theory as widely used in [24]. To deal with the problem of computing trust computation, we propose a proper robust aggregation means of the outcomes of review analyzers, each evaluating incoming reviews based on a specific criterion. To deal with the problem of mendacious reviews, we estimate which reviews contains opinions deviating from the evaluation of an objective of interest from the majority, as only a small portion of the reviewers is malicious.

The contribution of our work consists in the definition of a proper process to estimate the *trustworthiness* of social actors based on their published reviews and to achieve robustness against possible false opinions as follows.

1) Identifying possible mendacious reviews by exploiting a multicriteria decision making and introducing the novel concept of time-dependent and content-dependent crown consensus, where various criteria are used to evaluate the quality of a given review.
2) Performing reputation aggregation based on the Dempster–Shafer (D-S) combination rule [25] so as to infer the user trustworthiness.
3) Implementing the proposed approach in a cloud-based platform by crawling reviews from heterogeneous datasets, preprocessing (by performing data cleaning) and storing the acquired reviews in a NoSQL database, and realizing the envisioned trust computation by using an analytics engine for big data processing.
4) Experimenting the proposed approach and implemented solution on two different datasets, one from the Yelp Dataset Challenge and the other from Amazon Customer Review Datasets. We have also adopted the *YelpNYC* dataset so as to run the effectiveness evaluation of the approach against some of the main works within the literature. Such experiments proved the higher degree of precision and the user ranking challenges than the other approaches.

Several similar approaches have been proposed in the literature for evaluating user trustworthiness by using the textual review, especially in the context of spam detection. They can be generally classified in three groups: 1) linguistic based [26], focusing on the identification of linguistic features of malicious reviews; 2) behavioral based [27], leveraging metadata information of submitted review and user profile for identifying fake reviews; and 3) graph-based approaches [28], analyzing users and objects ties. The proposed framework exploits a behavioral analysis by combining in a novel way reviews' metadata, user compliments and rate's variation over time by leveraging fuzzy logic and the theory of evidence. A

novel set of criteria has been formulated in order to determine the trustworthiness of reviews, and they are aggregated so as to determine the overall user trust degree. The evidence theory has been used to compute trust by aggregating binary evidences, such as in [29] and [30]. Our novelty is represented by its application to users' trustworthiness assessment based on reviews' quality scores.

In the proposed approach, users' trustworthiness is not computed by considering their relationships but only the reviews' features. This is because users' relationships have a limited consideration for SRSs. However, it is possible to integrate our approach with one of those in the literature computing user trustworthiness based on the established relationships, as it may be seen as an additional criterion to be aggregated with the D-S combination rule within our proposed multicriteria decision making process.

Our experiments show that we are able to achieve an higher identification degree of malicious users in the considered datasets than the existing solutions. The proposed approach can be useful to cope with account hijacked thanks to the Evil Twin or Phishing attacks to compromise recommendation systems exploiting SRS [31], and in protecting the system against the sockpuppet attacks [32].

This article is organized as follows. In Section II, the relevant literature is presented and analyzed so as to highlight pros and cons of the related works, and compare them with our proposed solution so as to highlight the novelty of our work. Section III presents in details our proposed approach, while Section IV describes the implemented prototype, used to assess the effectiveness of the proposed approach, and discusses the obtained experimental results. We conclude this article in Section V by summarizing the findings of the proposed approach and planning the future work.

II. State-of-the-Art

Nowadays, the exponential growth of mobile devices enables a large number of people to pervasively use on-line services every time and everywhere in order to support their decisions. Indeed, the pervasive availability of the Internet allows the continuous use of these services, such as SRSs (Facebook, Twitter, and so on) and e-commerce (Amazon, Zalando, and so on) in any moment of our daily life. Throughout, SRS user behaviors are strongly influenced by the actions made by other ones and their opinions. The recent phenomenon of *influencers* [33] represents a concrete example of such a social behavior. They are people being highly active in SRS and capable of moving and directing the followers' opinion to their liking by means of their tweets, or posts. In fact, a user may choose to eat in a given restaurant or to buy a particular item based on the review and the assigned vote made by other users.

Despite being extremely useful to avoid nasty surprises, such a mechanism is vulnerable to malicious adversaries making fake reviews so as to harm honest entities (so as to blackmail them to stop the negative reviews) or to promote fraudulent or terrible entities (at the expense of unsuspecting customers). Therefore, it is important to define a mechanism for estimating the user trustworthiness based on the comparison of its behavior with respect to those exhibited by other users, so as to detect fake comments and malicious users. This falls within the broader research topic of trust management, which has been approached in many computing contexts and disciplines, from information systems, such as peer-to-peer systems or mobile ad hoc networks [34], [35], to those human processes targeted by managerial and social sciences [36]. Trust management in the case of SRSs lies at the intersection of such contexts, where the vagueness and subjectivity of the human processes are coupled with objective and precise measures coming from the automatic computations done by computers. Specifically, reviews are written by humans (or software mimicking humans) by using natural language, whose validity must cope with certain ethical guidelines (also written in natural language) and a set of decision criteria more related to affective states and personal judgments. Therefore, it is extremely subjective if a given review is valid or not, with a possibility that there are no black-and-white answers. However, due to the massive amount of data to be processed, reviews must be mined by the computers so as to have measures against a set of criteria, to be further aggregated to come up with a single score to express the trustworthiness of users (visualized as a number or stars or a given traffic light color).

Several approaches have been proposed in the current literature to estimate the user trustworthiness in order to properly support a wide range of applications. An interesting survey on SRS security and trustworthiness issues has been recently proposed in [37], where an architecture for trust management is defined based on signaling theory and crowd computing. In a big data context, Yu *et al.* [38] described an approach for computing user trustworthiness by leveraging on the "familiarity" and "similarity" concepts and considering the influence of user actions on the trustworthiness computation. The aim of this methodology is to detect malicious users-based also on a security queue to record users' historical trust information. Afterward, Yu *et al.* [39] proposed an approach based on deep learning techniques in conjunction with user trustworthiness characterization for configuring privacy settings for social image sharing. In addition, a two-phase trust-based approach based on deep learning techniques has also been proposed by Deng *et al.* [40] for social network recommendation, so as to determine the users' interests and their trusted friends' interests together with the impact of community effect for recommendations. Rayana and Akoglu [27] presented a system, namely, *SpEagle*, that uses metadata (i.e., text, timestamp, and rating) in conjunction with relational data to spot suspicious users and reviews.

Other related approaches exploit reviews' evaluation for detecting and/or characterizing spam in social media. Shehnepoor *et al.* [28] proposed a framework named *NetSpam* that models reviews in online social media, as a case of heterogeneous networks, by using spam features for detection purposes. Ye *et al.* [41] described an approach based on the temporal analysis by monitoring selected indicative signals of opinion spams over the time, for detecting and characterizing abnormal events in real time.

TABLE I
APPROACHES FOR EVALUATING THE TRUSTWORTHINESS OF USERS AND/OR REVIEWS

Ref.	Entities	Techniques	SRS
[26]	Reviews	Analysis of the temporal patterns and their relationships with the rate of posting fake	Yelp
[37]	Users	Crowd evaluation and measurement based on signaling theory in economics and information management and crowd computing	No Dataset
[38]	Users	Probabilistic method for analyzing the influence of users' actions on trustworthiness	Artificial Dataset
[39]	Users	Deep learning for extracting features from images and identifying privacy objects and events.	PicAlert Flickr
[40]	Users	Deep Learning and matrix factorization for trust-aware social recommendation	Epinions Flixster
[27]	Users/ Reviews	Classification approach based on behavioral and textual features for users, products, and reviews.	Yelp
[28]	Reviews	Unsupervised classification approach based on weighted schema of behavioral and linguistic features for users and products	Amazon Yelp
[41]	Reviews	Time series analysis of eight indicative signals for each product for detecting and characterizing spam campaigns.	SWM FLIPKART
[42]	Tweets	Classification approach based on credibility of content, user reputation and expertise for assessing information.	Twitter
[43]	Reviews	Multidimensional model by mining feedback comments for computing reputation score	Ebay Amazon
[44]	Users	Classification approach based on the fairness of a user the reliability of a rating and the goodness of a product incorporating also users' behavioral properties.	Flipkart Epinions Amazon Bitcoin
[45]	Users	Combined use of consensus and trustworthiness techniques	SQUARE benchmark
[46]	Users	Classification approach based on six axioms to define the interdependency among three intrinsic quality metrics concerning a user, reliability and goodness of a product by combining network and behavior properties.	Flipkart Epinions Amazon Alpha OTC
[48]	Users/ Reviews	Bayesian inference on Bayesian model for computing the likelihood-based suspiciouness metric to identify fake reviews and users.	Flipkart SWM
[47]	Users	An approach aims to be resistant to the camouflage attacks for identifying fraud activities, providing also an upper bounds on the effectiveness of fraudsters.	Amazon Trip Advisor Epinion Wiki-vote
[32]	Users/ Reviews	An approach based on a new class of users (*trusted users*), considering also reviews left by verified users.	OTC Alpha Epinions Amazon OTC

A system based on four integrated components, specifically: 1) a reputation-based component; 2) a credibility classifier engine; 3) a user experience component; and 4) a feature-ranking algorithm, has been designed and implemented by Alrubian *et al.* [42] for assessing information credibility on Twitter. In [43], the *CommTrust* framework has been introduced for trust evaluations by mining feedback comments. More in detail, it is based on a multidimensional trust model for computing reputation scores from user feedback comments, which are analyzing combining natural language processing techniques, opinion mining, and topic modeling. Furthermore, another framework, namely, *LiquidCrowd*, has been proposed by Castano *et al.* [44] exploiting consensus and trustworthiness techniques for managing the execution of collective tasks. Kumar *et al.* [45] proposed a system, namely, *FairJudge*, to identify fraudulent users based on the mutually recursive definition of the following three metrics: 1) the user trustworthiness in rating products; 2) the rating reliability; and 3) the goodness of a product. Moreover, Kumar *et al.* [46] described a system for identifying fraudulent users based on six axioms to define the interdependency among three intrinsic quality metrics concerning a user, reliability and goodness of a product by combining network and behavior properties.

Hooi *et al.* [47] developed an algorithm, called *FraudAR*, aiming at being resistant to the camouflage attacks, for identifying fake reviews and users. Furthermore, *Birdnest*, an approach combining Bayesian model of user rating behavior and a likelihood-based suspiciousness metric [normalized expected surprise total (NEST)], has been proposed in [48]. Liu *et al.* [32] investigated the sockpuppet attacks on reviewing systems by proposing a fraud detection algorithm, called RTV, that introduces trusted users and also considers reviews left by verified users.

Despite the current literature on the addressed topic is extremely vast and heterogeneous, the above mentioned approaches (summarized in Table I) present some drawbacks. First, such as in [37]–[40], [45], and [46], most of them are based on the concept of crowd consensus to detect malicious reviews. If a given review diverges from the option of the majority within a group of users, it must be malicious and contains misleading comments. Most of them do not consider when a review has been posted and the possible evolution over the time of the quality of a given object of interest. So, a diverging review may erroneously detected as malicious if an evolution over the time of the perceived quality is not considered. Second, other approaches rely only on the textual content of the user reviews [43] or simply combine these textual features with some contextual metadata about the user [27], [28], [42]. Finally, KC and Mukherjee [26] studied the temporal pattern of postings for a given user in order to detect the time when it triggers its malicious behavior. However, an adversary may have intermittent malicious behaviors, making such an analysis meaningfulness.

The proposed framework moves beyond the current state-of-the-art by introducing the novel principal of temporal- and content-dependent crown consensus: how the examined user rates a given business object matches the opinion of the majority based on its rate variation over time and textual content variations (also known as *consensus* sociological principle [49]). Moreover, the review analysis combines human

evaluations (expressed by using fuzzy logic techniques) and automatically computed review features and the innovative consensus divergence, through a multicriteria decision making supported by the D-S theory. The novelty is indeed represented by a larger set of criteria and the application of D-S aggregation, allowing a higher degree of reliability in trust assessment. In fact, on the one hand, most of the approaches mainly consider review utility and user activity, fewer ones also include review quality, but we have also added two different formulations of crown consensus. On the other hand, other related works preferred to use the simpler minimum, maximum or average operations applied over the received reputation scores

III. Methodology

The modern SRSs offer several features allowing users to interact among each other by using different channels (such as posts, tweets, or streaming) to exchange various multimedia content (such as text, images, video, and audio). Such SRSs have been designed to provide a communication channel over the Internet for people; however, as their use has considerably increased in the last decade, their business value emerged. Targeting and connecting with potential customers by exploiting an SRS as a simple method of advertising is the first example of such a business-related use, but actually it is their weakest business use. One of the most powerful and popular uses is related to reviewing a given business object, where users provide their opinions and past experiences so that other users are able to evaluate a given business object and take business-related decisions based on the reviews made by others. Generally speaking, such reviews are made of two distinct parts. On the one hand, there is a visual aspect that quickly provides an eye-catching summarizing general opinion of the business object, mainly in terms of stars (usually from 0 to 5) or a color associated to a number within an interval (which can be computed as the normalized number of positive reviews over the total number of reviews submitted by the users). On the other hand, there is a bunch of texts written by the users that have reviewed the given business object. When a user makes a review of a business object, there is the establishment of a social relationship among them, so that a social business network (SBN) can be built. An SBN depicts the social network of a company with their customers, and can be formally defined as follows.

Definition 1 (Social Business Network): Let U and BO be, respectively, a set of users and of business objects, the *business social network* can be defined as a graph $G = (V, E)$ in which $V = U \cup BO$ and E is composed by the set of reviews, with the related metadata $(\lambda_0, \lambda_1, \ldots, \lambda n\_o)$, made by users on different business objects, whose number is n_o. $\rho_o(t)$ indicates a specific review made at time instance t for the business object o.

The idea behind our approach is to estimate the trustworthiness of a user leveraging the information involved in a SBN. Ideally, we can find trustworthy a user that, in all of his review, perfectly expresses the value of a business object without any malicious falsification.

Definition 2 (Real User Trust): Let $\rho_o(t)$ and $q_o(t)$ be, respectively, the numeric representation for the review of a given business object done by the user at time t and its real value within a 2-D Cartesian space (using metadata λ_i), defined along the dimensions of quality and price. We can assume such a user as trustworthy if the distance between these elements is 0 for all the reviewed N objects and $\tau(t)$ needs to be formulated so that for trustworthy users, it returns 1. Therefore, we can formulate it as 1 minus the sum of the distance between the reviewed and real value for all objects normalized by the total number of objects, obtaining the following expression:

$$\tau(t) = 1 - \frac{\sum_{i=1}^{N}(\rho_i(t) - q_i(t))}{N}. \tag{1}$$

Such a value in (1) should be computing for the overall observation time: $\tau = \int_{t_0}^{t_f} \tau(t)dt$. Despite correct, such a definition is not viable to be used in practice for two main reasons, there may be a divergence between a review and the real value of an object related to subjective criteria adopted during the reviewing process (what is valuable and perceived as high quality to a human individual, may be worthless to other people). The ground truth of the real value for a business object is not available. For these reasons, the previous definition is not adequate, and we should adopt a different approach. This leads to an aligned group-based definition where a user is trustworthy if its reviews are aligned with the ones of the majority of a group of people, based on the assumption that only a minority of people may have malicious intentions. Such an assumption comes from the well-known Byzantine generals problem of the academic literature in distributed systems [50], in a consensus problem, the agreement among N actors with F of them being malicious can be reached only if $N \leq 3F$. Therefore, this turns out that malicious behaviors are detectable if only few members of N entities are Byzantine (i.e., $N >> F$, or more precisely $N = 3F + 1$). Based on this, we can say that a user is trustworthy if his/her divergence to the opinion of the majority of the other users in his group is acceptable (so as to consider the subjectivity of the judgment), or more formally as follows.

Definition 3 (Majority-Based User Trust): Let $\rho_o(t)$ and $\hat{\rho}_o(t)$ be, respectively, the representation for the review of a given business object done by the user at time t and the review made by the majority of the other users, we can assume such a user as trustworthy [i.e., $\hat{\tau}(t) = 1$] only if the sum of the distances between $\rho_o(t)$ and $\hat{\rho}_o(t)$ for any object o, normalized by the number of objects, is lower than a certain small threshold σ. This can be expressed as follows:

$$\hat{\tau}(t) = \begin{cases} 1, & \text{if } \frac{\sum_{i=1}^{N}(\rho_i(t) - \hat{\rho}_i(t))}{N} \leq \sigma \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

The information exposed by an SBN can be exploited in order to estimate the opinion of the majority, and proficiently determine a user trustworthiness. However, this is easier said than done. First, even for computerized experts, it is hard to measure the quality of a review with a single numeric value (or an interval), but it is simpler to express such a quality against multiple distinct criteria, such as its utility, correctness, and

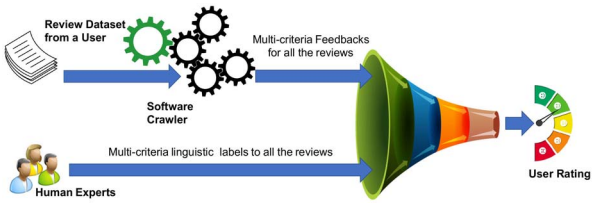

Fig. 1. Schematic representation of the proposed approach.

so on. How properly aggregating these multivalue feedbacks coming from human and computerized experts represents the first problem. Furthermore, reviews cannot be represented in a Cartesian space by means of numeric values as they are expressed in natural language if we consider the feedback coming from human experts, characterized by a certain degree of fuzziness in terms of imprecision as humans have trouble to give objective numeric judgments. This poses the problem of how properly representing them so as to make it easy to be processed by a computer, and aggregating them so as to have the opinion of the majority, and calculating the distance of a given review from such a global opinion. The problem of detecting malicious reviews is not simple, as their fuzziness obstacles this. However, it is possible to determine a set of criteria against which the reviews can be assessed, supporting the thesis of being malicious or not. Such criteria can be automatically computed by using the finding of the research on text mining, sentiment analysis and natural language processing. However, such objective consideration must be jointly considered with more subjective assessment given by human experts, which may be the users of SRS stating the utility and plausibility of such reviews after having verified the reviewed business object by their own.

Our work has the goal of dealing with these problems as follows. The first highlighted problem implies the modeling of the envisioned trust estimation as a multicriteria multiexpert decision-making (MCME-DM) method, which must be based on the fuzzy logic theory and a proper aggregation rule so as to deal with the fuzziness introduced by human experts. Fig. 1 depicts the scheme of the proposed approach. The top part of the figure represents the reviews of all the users being processed by proper crawling software that analyzes the set of past user reviews (and feedbacks that other users have given on them as done by any SRS) to analytically assess quality measures. The left part of the figure represents the reviews being judged by both human experts, such as the other users. The linguistic labels assigned to all the reviewers by human experts and the crisp values computed by the software crawler are aggregated so as to determine the trustworthiness of a given user (available at the top of the figure).

The representation of subjective review judgments gave by the humans is described in Section III-A by leveraging on the fuzzy set theory [51], while how a software crawler assesses the reviews is presented in Section III-B. These subjective and objective assessments are combined according to the method illustrated in Section III-C, exploiting the combination rule taken from the D-S theory [25]. A summarizing algorithm of the proposed approach is presented in Section III-D, with a precise description of what represented in Fig. 1.

A. Representation of Subjective Judgment

Computers are able to easily perform complex computations on numeric values, while meeting some trouble to deal with linguistic expressions due to their fuzziness, on the contrary to humans. However, if the objective assessments coming from computer-based text analysis are plausible to be numeric, the human experts (especially if we consider that are the generic users of SRS) are likely to feel uncomfortable to deal with numbers. In order to let these two distinct worlds to meet and work along side, we can use linguistic labels [52], [53], such as given adjectives, such as LOW or HIGH, to let humans easily express the assessment of a given review against a certain criterion, e.g., a review is highly or lowly fair, poorly or highly useful, and so on. Such labels are typically adjectives spanning from an extremely negative one to an extremely positive one, as follows:

$$S = \{s_0 : N(\text{NONE}), s_1 : L(\text{LOW}), s_2 : M(\text{MEDIUM}) \\ s_3 : H(\text{HIGH}), s_4 : P(\text{PERFECT})\} \quad (3)$$

where $g = 5$ is the number of terms in the set, and is called as its granularity. The approach is similar to those user satisfaction applications, where each user is asked to assign starts or points to a given quality measure concerning a received service. The lowest linguistic label represent the single star, while the highest one is the maximum number of stars.

Typically, the adopted granularity is odd [54], so that the central term indicates a neutral situation, where no preference is expressed, while the other terms are placed symmetrically around the central one. Furthermore, it is evident from the previous example that the terms are ordered, based on their positiveness, where $s_i \leq s_j \iff i \leq j$. To let computers be able to process such linguistic labels, we can associate them with a representation in terms of fuzzy sets, i.e., each label has associated a proper membership function, drawn from the literature of the fuzzy logic, modeled as a trapezoid or Gaussian function or other ones. Such membership functions represent the main building blocks of the fuzzy set theory, as they determine the fuzziness in a fuzzy set. Accordingly, the selection of the best shapes of membership functions strongly depends on the particular problem of interest, but there is no criteria to consider in such a selection. Within the current literature, there are many academic publications and books giving directions of how to choose membership functions, such as [55]. Triangular functions typically represent the starting point, and bell-shaped functions provide the best results, generally. Trapezoidal functions result from a crisp interval-based rule applied to inputs with uniform uncertainty, and represent a tradeoff between the simplicity of the triangular one, and the complexity of the bell-shaped one, and for these reasons they have been applied in this work. The admissible numeric interval to be associated to a given criterion is uniformly covered by these functions, as in Fig. 2 for the case of the trapezoid membership function.

To this aim, vectors filled with linguistic fuzzy sets (one per each assessment criterion) can be associated to reviews

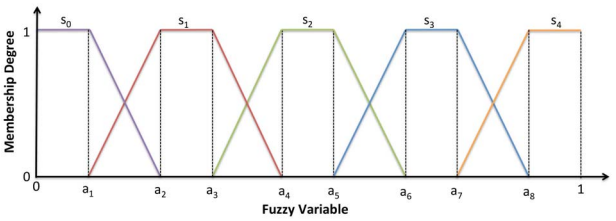

Fig. 2. Set of five terms in (3) and its semantics as fuzzy sets.

from the human experts, while numeric values are given by computerized experts. In order to bring them to the same representation, each linguistic label, seen as a fuzzy variable, can be transformed to a crisp number within the interval [0, 1] by applying a defuzzification operation. An example can be using the centroid method that determines the geometric center over the x-axis of the membership function associated to the linguistic term, according to the formula available in [56]. Considering the numeric values, they can be reduced within the interval $[0, 1]$ by means of a trivial normalization (by dividing those numbers against their maximum).

B. Assessment Criteria

Each review of a given user must be evaluated against a set of well defined criteria both by human experts and a crawler. In our work we have considered four criteria, but the proposed approach is generic enough to be easily adapted to a greater number of them. By properly aggregating the scores obtained from these criteria, the trustworthiness degree can be computed. When a human expert is called to assess some review, he/she needs to assign a linguistic label to each of such a criterion, based on its personal and subjective judgment; while the crawler assigns a numeric crisp value, which is converted into a linguistic label by means of fuzzification. The list of used criteria is as follows. Each of them can be automatically computed based on the existing feedback that users provide after having read a review of user i for the business object o published at time t, namely, $r_{i,o}(t) \in R$:

1) a binary function belong$(u_i, r_{i,o}(t))$ which returns 1 if the review as the second input has been published by the user passed as its first input;
2) votes indicating if a review has been useful, represented by a binary variable useRcvd$(u_j, u_i, r_{j,o}(t))$ assuming 1 if the review $r_{j,o}(t)$ published by user u_j received a vote from the user u_i with respect to its utility;
3) compliments related to how the review has been written and structured, represented by a binary variable compRcvd$(u_j, u_i, r_{i,o}(t))$ assuming 1 if the review $r_{j,o}(t)$ published by user u_j received a compliment from the user u_i.

Definition 4 (Usefulness): A useful review is the one that allowed a user to take the best decision with respect to business object. For a concrete example, if such an object is a restaurant, a review has been useful if it is allowed to avoid a bad restaurant or to pick up a good one. However, when reading a review, it is not possible to claim beforehand if a review is useful. On the contrary, a useful review is the one that lets the reader understand if the reviewer object is good or not, so as to support a possible decision. An unuseful review is typically too vague, imprecise, or short so that after having read it, a decision cannot be taken. Therefore, a human expert has to rate if after reading the review, he/she was able to gather an overall glimpse of the reviewed object so as to make a decision. For the crawler, we have the following formulation: let U be a set of users in the considered SBN and T is the observation time window, the *usefulness* criteria is the ratio of the total number of "useful" votes received by a user with respect to the maximum of "useful" votes assigned to a single user in U set, expressed as follows:

$$c_1(u_j) = \frac{\sum_{i \neq j, o \in O, t \in T} \text{useRcvd}(u_j, u_i, r_{j,o}(t))}{\max_{u \in U} \left\{ \sum_{i \neq j, o \in O, t \in T} \text{useRcvd}(u_j, u, r_{j,o}(t)) \right\}}. \tag{4}$$

This criterion states that a user is considered trustworthy if it has received many "useful" votes.

Definition 5 (Quality): A good review is the one that is well written, has some photographs attached to it and/or provides evidences and details to support its claim. The human expert, therefore, has to judge how well a review is written and structured, but for the crawler we have the following formulation. Let U be a set of users in SBN and T is the observation time window, the *quality* criteria are computed as the total number of compliments received normalized on the maximum of the sum of compliments assigned to a single user in U set, expressed as follows:

$$c_2(u_j) = \frac{\sum_{i \neq j, o \in O, t \in T} \text{compRcvd}(u_j, u_i, r_{j,o}(t))}{\max_{u \in U} \left\{ \sum_{i \neq j, o \in O, t \in T} \text{compRcvd}(u_j, u, r_{j,o}(t)) \right\}}. \tag{5}$$

This criterion states that a user is trustworthy if it has received a high number of compliments.

Definition 6 (User Activity): A user of a given SRS is characterized by a certain frequency of published reviews, spanning from the sporadic ones to the very active users. If we assume that a very active reviewer is generally more expert, we can formulate a rating criterion dependent on how active a user is within the SRS. A human expert can judge such a measure based on how often he/she reads posts, tweets, or review from such a user, while the crawler makes its calculations from the statistics within the considered dataset. Specifically, let U be a set of users in SBN and *voteSent*(u) and *reviewsLeft*(u) be, respectively, the number of sent votes and left reviews by a given user $u \in U$, expressed as follows:

$$\begin{aligned} \text{voteSent}(u) &= \sum_{v \neq u \in U, o \in O, t \in T} \text{compRcvd}(u, v, r_{j,o}(t)) \\ &\quad + \sum_{v \neq u \in U, o \in O, t \in T} \text{useRcvd}(u, v, r_{j,o}(t)) \\ \text{reviewsLeft}(u) &= \sum_{o \in O, t \in T} \text{belong}(u, r_{j,o}(t)). \end{aligned} \tag{6}$$

The *user activity* criteria (c_{ua}) are defined as follows:

$$c_3(u_j) = \frac{\text{voteSent}(u_j) + \text{reviewsLeft}(u_j)}{\max_{u \in U} \{\text{voteSent}(u) + \text{reviewsLeft}(u)\}}. \tag{7}$$

This criterion analyzes the user activity that describes how a given user interacts with other ones through the evaluation of its sent votes and published reviews weighted by useful received votes.

Definition 7 (Time-Dependent Crown Consensus): According to our definition of trustworthiness given in (2), a given review needs to be aligned with the opinion of the majority, with a slight divergence. However, it is important to also consider the possibility that a review can influence the other users, so we can formulate an assessment of a given review with respect to the opinion of the rest of the users before the review has been published and afterward, and consider how the review diverges with respect to the prereview and post-review opinion of the others. The human expert should subjectively measure such a divergence, but the crawler works as follows. Let t_1 and t_2 be two different time instances ($t_1 < t_2$), $r_{i,o}(u_i, t_r)$ be a review made by a given user (u_i) at the time instance t_r within the time interval $[t_1, t_2]$, with $\alpha_o(t_1, t_r)$ and $\beta_o(t_r, t_2)$ being the rate received by the given business object o with the review publisher before and after the instance t_r. The *time-dependent crown consensus* criterion is a set of fuzzy rules that assign a proper linguistic label based on the distance between $r_{i,o}(t_r)$, $\alpha_o(t_1, t_r)$, and $\beta_o(t_r, t_2)$

$$c_4(u_i) = \begin{cases} S_2, & \alpha_o(t_1,t_r) = \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) = x \\ S_1, & \alpha_o(t_1,t_r) = \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) < x \\ S_1, & \alpha_o(t_1,t_r) = \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) > x \\ S_0, & \alpha_o(t_1,t_r) < \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) \geq \\ & \geq \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) - \beta_o \geq \delta_i \\ S_1, & \alpha_o(t_1,t_r) < \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) \geq \\ & \geq \beta_o(t_r,t_2) \\ S_3, & \alpha_o(t_1,t_r) < \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) < \\ & < \beta_o(t_r,t_2) \\ S_4, & \alpha_o(t_1,t_r) < \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) < \\ & < \beta_o(t_r,t_2) \wedge \beta_o(t_r,t_2) - r_{i,o}(u_i,t_r) < \\ & < \delta_d \\ S_4, & \alpha_o(t_1,t_r) > \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) > \\ & > \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) - \beta_o(t_r,t_2) < \\ & < \delta_d \\ S_3, & \alpha_o(t_1,t_r) > \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) > \\ & > \beta_o(t_r,t_2) \\ S_1, & \alpha_o(t_1,t_r) > \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) \leq \\ & \leq \beta_o(t_r,t_2) \\ S_0, & \alpha_o(t_1,t_r) > \beta_o(t_r,t_2) \wedge r_{i,o}(u_i,t_r) \leq \\ & \leq \beta_o(t_r,t_2) \wedge \beta_o(t_r,t_2) - r_{i,o}(t_r) > \delta_i \end{cases} \tag{8}$$

where δ_i and δ_d are, respectively, the maximum and minimum difference of votes, while x being the rate of the majority of users in the considered time interval. Equation (8) is interpreted according to the indications in [48] to detect the review fraud by measured how mendacious reviews skewed rating distributions. The linguistic term S_0 (associated with the lowest level of trust as in Fig. 2) is given if after the reviewing time, the object gets a different value and the user review is far from the new object value. The linguistic term S_1 is given if the object value is not changed but the user review differs from the object value and the value assigned by the majority. The linguistic term S_2 is assigned if the object value is not changed and the user review matched with the one of the majority. Finally, the linguistic term S_4 (associated with the highest level of trust as in Fig. 2) is assigned if the object value is changed and the user review matched with the one of the majority causing the change.

The *time-dependent crown consensus* criterion analyzes how the rating of a given user about a business object varies based on the actions made by the other ones in a specific time interval.

Definition 8 (Content-Dependent Crown Consensus): According to the previous definition of crown consensus, we consider also the review's content divergence with respect to other ones made by other users on the same business object. Let t_1 and t_2 be two different time instances ($t_1 < t_2$), $\vec{r}_{i,o}(t_r) = \{f_1, f_2, \ldots, f_N\}$ be the vector representation of the review's content, obtained by mapping the review into a N-dimensional space by using the cosine similarity, made by a given user (u_i) at the time instance t_r within the time interval $[t_1, t_2]$, with $\alpha_o(t_1, t_r)$ and $\beta_o(t_r, t_2)$ being the comment received by the given business object o with the review publisher before and after the instance t_r. We define as the set R the group of users forming the majority expressing the degree x and the centroid of the set of vectors $\vec{r}_{i,o}(t_r) = \{f_1, f_2, \ldots, f_N\}$ with $i \in R$ is indicated with c. The *content-dependent crown consensus* criterion is a set of fuzzy rules that assign a proper linguistic label based on the respect distance between $\vec{r}_{i,o}(t_r)$ of the ith user and the $\vec{r}_{c,o}(t_r)$ of the centroid c meant as the Cartesian distance among vectors in an N-dimensional space, $\alpha_o(t_1, t_r)$ and $\beta_o(t_r, t_2)$

$$c_5 = \begin{cases} S_0, & \alpha_o(t_1,t_r) \neq \beta_o(t_r,t_2) \wedge \\ & \wedge\ d\big(\vec{r}_{i,o}(t_r), \vec{r}_{c,o}(t_r)\big) > \epsilon \\ S_1, & \alpha_o(t_1,t_r) = \beta_o(t_r,t_2) \\ & \wedge\ d\big(\vec{r}_{i,o}(t_r), \vec{r}_{c,o}(t_r)\big) > \epsilon \\ S_2, & \alpha_o(t_1,t_r) = \beta_o(t_r,t_2) \\ & \wedge\ \epsilon \geq d\big(\vec{r}_{i,o}(t_r), \vec{r}_{c,o}(t_r)\big) \leq \epsilon + c \\ S_3, & \alpha_o(t_1,t_r) = \beta_o(t_r,t_2) \\ & \wedge\ d\big(\vec{r}_{i,o}(t_r), \vec{r}_{c,o}(t_r)\big) < \epsilon \\ S_4, & \alpha_o(t_1,t_r) \neq \beta_o(t_r,t_2) \\ & \wedge\ d\big(\vec{r}_{i,o}(t_r), \vec{r}_{c,o}(t_r)\big) < \epsilon \end{cases} \tag{9}$$

where c is a given constant and ϵ is equal to the maximum distance of the vector representation of the review's content from a member of the majority with the centroid

$$\epsilon = \max_{r \in R} d\big(\vec{r}_{r,o}(t_r), \vec{r}_{c,o}(t_r)\big). \tag{10}$$

While the assessments of such last two criteria return linguistic values, the outputs of the first three criteria are crisp values that can be easily transformed into a fuzzy one by using different fuzzification function, such as the above-mentioned centroid method.

C. D-S Aggregation

At this level, we have a set of linguistic label (specifically, four of them, one per each of the previous criteria), assigned to each user by each of the adopted expert. We have the issue of properly aggregating them so that each user has assigned a

single linguistic label representing its trustworthiness degree. To this aim, we can leverage on the D-S theory, dealing with the problem of having a certain decision to be taken and a set of possible hypotheses, supported by multidimensional evidences, i.e., a vector where each element is the degree of satisfaction of a specific criterion assigned by one of the contacted experts. The problem is to find the hypotheses with the strongest support from the obtained evidences. The first stage concerns with the definition of the discernment frame Θ that is set of hypotheses within the decision process. In our case, it represents the elements of the selected linguistic term sets to be assigned to the review to be assessed. It can be described by the following equation, in which N is the number of hypotheses:

$$\Theta = \{H_1, H_2, \ldots, H_N\}. \tag{11}$$

It is important to note that the discernment frame, whose hypotheses are not mutually exclusive, is exhaustive, and in our case, it is composed of all the linguistic labels assumed in (3). Once defined the discernment frame, we can determine the corresponding power set as its superset composed by 2^N combinations of N hypotheses in Θ

$$P(\Theta) = \{\emptyset, H_1, H_2, \ldots, H_N, H_1 \cup H_2, H_1 \cup H_3, \ldots, \Theta\}. \tag{12}$$

Thus, the power set can be composed by singleton elements, that is a single hypothesis in Θ, and a combination of these elements. Furthermore, it is possible to define a mass function $[m(H_i) : H_i \in P(\theta)]$, whose support assumes values in $[0, 1]$, by using the evidences received from the different experts and the Lagrangian definition of probability.

Definition 9 (Mass Function): Let E and θ being, respectively, the set of evidences to support the hypothesis within the power set and one of such hypotheses, the mass function of the hypotheses θ is the ratio of the evidences supporting such a hypothesis over the total number of available evidences.

Thus, we can define a counting function of power set's elements based on a given criterion C_j, the linguistic terms set assigned to it (e_{C_j} $e_{C_j} = \{DT(C_j), \cup_{i=0}^{R} i(C_j)\}$) and the ith provider as follows:

$$c\big(H_i, e_{C_j}, i\big) = \begin{cases} 1, & \text{if } e_{c_j}(i) = H_i \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

The mass function for a given hypothesis for a certain criterion C_j can be defined as

$$m_{C_j}(H_i) = \frac{c\big(H_i, e_{C_j}\big)}{\sum_{\forall H_i \in P(\Theta)} \big(c\big(H_i, e_{C_j}\big)} \tag{14}$$

where $c(H_l, e_{C_j}) = \sum_{i=0}^{R-1} C(H_l, e_{C_j}, i)$.

Since we encompass multiple criteria, there is a mass function for each hypothesis related to a given criteria. These functions can be combined by the D-S rule of combination (also called the orthogonal sum or D-S combination rule) to obtain a single function. Thus, let C_1 and C_2, m_{C_1} and m_{C_2} be, respectively, two criteria and the respective mass functions, the overall mass function can be defined as follows:

$$m(H_i) = m_{c_1} \oplus m_{c_2} = \frac{\sum_{X \wedge Y = A} m_{c_1}(X) m_{c_2}(Y)}{(1 - k)} \tag{15}$$

where $k = \sum_{X \wedge Y = 0} m_{c_1}(X) m_{c_2}(Y)$ is the conflict degree between the two analyzed criteria, whose value varies between 0 (no conflict) to 1 (no agreement). Thus, the obtained overall mass value is used to quantify the believe of a given hypotheses

$$\text{bel}(H_i) = \sum_{\emptyset = B \subset A} m(B) \quad \forall A \subseteq \Omega. \tag{16}$$

The believe value is bounded between belief and plausibility, as follows:

$$\text{bel}(H_i) \geq P(H_i) \geq pl(H_i) = \text{bel}(\omega) - \text{bel}(H_i). \tag{17}$$

It is possible to derive a crisp number from the belief interval by means of the so-called pignistic probability transformation, which is built based on the expected utility theory to represent beliefs when taking the optimal decision that maximizes the expected utility within the context of a decision-making process. Such a transformation determines the pignistic probability as follows:

$$\text{Bet}P(A) = \sum_{W \subseteq \Omega A \in W} \frac{m(W)}{|W|} \tag{18}$$

where $|W|$ is the number of entities contained in the W set. Therefore, the outcome of our aggregation is given by the hypothesis that has obtained the greatest value of the pignistic probability.

To better describe the aggregation process, let us consider an example, where a review dataset is given as input, as well as a set of linguistic labels assigned by the human experts based on the four criteria to a given user (by making a fuzzy average of the assessments to each review published by such a user). Based on data extracted from the dataset, a crawler obtains the numeric assessment of the criteria for the user, which are fuzzified later on. The evidences, as linguistic labels, are used to compute the mass functions per each criterion, and an aggregated mass function is computed from them. Finally, the pignistic probability for each label is determined and the label with the highest probability is assigned to the user as its trust degree. Our work assumed that a single linguistic label must be assigned to a given criterion, but even this is challenging for a human expert (which may be uncertain which label to select between two adjacent ones or not) and may cause errors during the fuzzification. To this aim, it is possible to apply the so-called hesitant fuzzy linguistic term sets (HFSs), described in [57], which consists of an interval on the adopted linguistic term set encompassing two labels, where the first one has a lower mean of the second one, and the relative membership function is the union of the membership functions on the constituting terms. Our approach does not need to be changed, since in this case, not only the singleton elements will have supporting evidences but also their permutations. To simplify our implementation and consequent experiments, we limit our work to avoid the use of HFS.

The following example [22] has been provided to better explain the proposed aggregation methodology. Let u_i, e_1, and e_2 be, respectively, a given user, that published different reviews on business objects, and two experts, used by the proposed methodology for evaluating a given user. In Table II,

TABLE II
OBTAINED TRUST DEGREES OVER FIVE CRITERIA FROM THREE EXPERTS ABOUT u_i USER

Criteria	TRU	e_1	e_2
c_1	S_3	S_2,S_3	S_1,S_2,S_3
c_2	S_3	S_2	S_2
c_3	S_2,S_3	S_3	S_2,S_3
c_4	S_2,S_3,S_4	S_3	S_2
c_5	S_2, S_3	S_2,S_3	S_3

TABLE III
MASS FUNCTION VALUES COMPUTED BY USING THE EVIDENCES IN TABLE II AND (14)

Criteria	$m(S_2)$	$m(S_3)$	$m(S_2, S_3)$	$m(S_1, S_2, S_3)$	$m(S_2, S_3, S_4)$
c_1	0	0.33	0.33	0.33	0
c_2	0.66	0.33	0	0	0
c_3	0	0.33	0.66	0	0
c_4	0.33	0.33	0	0	0.33
c_5	0	0.33	0.66	0	0

TABLE IV
AGGREGATED MASS FUNCTION VALUES OBTAINED BY USING THE MULTIPLE MASS FUNCTIONS IN TABLE III AND THE D-S COMBINATION RULE IN (15)

Criteria	$m(S_2)$	$m(S_3)$	$m(S_2, S_3)$	$m(S_1, S_2, S_3)$	$m(S_2, S_3, S_4)$
$c_1 \oplus c_2$	0.56	0.42	0	0	0
$c_3 \oplus c_4$	0,25	0,25	0,25	0	0
$(c_1 \oplus c_2) \oplus (c_3 \oplus c_4)$	0.37	0.28	0	0	0
$c_5 \oplus (c_1 \oplus c_2) \oplus (c_3 \oplus c_4)$	0.28	0.32	0	0	0

we report the trust values provided by our methodology (*TRU*) and experts over the five described criteria about the user u_i. In the second step, we compute the mass functions, whose values are shown in Table III, by using the evidences in Table II and resolving (14). Thus, it is possible to compute the aggregated mass function, shown in Table IV, based on value shown in Table III and the D-S combination rule in (15). By looking (18), the pignistic probabilities in the example can be computed as follows:

$$\text{Bet}P(S_2) = 0.28,\ \ \text{Bel}P(S_3) = 0.32,\ \ \text{Bel}P(S_2, S_3) = 0. \quad (19)$$

Therefore, the computed trust degree is equal to the linguistic term S_3.

D. Summary

The overall proposed approach is described in Algorithm 1. At the beginning, among the collected reviews, those made by the user of interest are extracted. Then, the criteria illustrated in Section III-B are computed automatically by the crawler, while a set of human experts are contacted to have their subjective judgement. This can happen synchronously by having the software implementing the approach to ask a request and wait for a reply from each human expert, or even asynchronously, by having the human experts to associate to each of the review that they have read a personal judgement. After the assessment of the reviews are available, they are used to compute the mass function and to aggregate all these functions in a unique mass function, as described in Section III-C. Finally, a pigmistic probability can be computed and associated to each the linguistic fuzzy terms presented in Section III-A. The term with the highest of these probability is selected as the final result of the approach, and its defuzzification is returned.

Algorithm 1: Pseudocode Describing the Proposed Approach for Trust Computation

```
Result: τ for given user u
R ← getReviews();
R4U ← ∅;
while ρ in R do
    if ρ.author == u then
        R4U ← ρ
    end
end
cA ← computeCriteria(R4U, R);
cB ← askCriteria(R4U);
mA ← computeMassFunction(cA);
mB ← computeMassFunction(cB);
m ← aggregate(mA, mB);
p ← computePigmisticProb(m);
τ ← defuzzify(max(p));
```

IV. EXPERIMENTAL ASSESSMENT

A. Protocol and Dataset

Our evaluation is mainly focused on assessing if our approach achieves the following three criteria.

1) *Efficiency:* To evaluate the running times to compute user trustworthiness with respect to the number of users and reviews, and the average ratio between reviews and users, with respect to the state-of-art approaches.
2) *Cost*: To measure the deployment cost of our framework on the Microsoft Azure[1] cloud platform.
3) *Effectiveness:* To examine the accuracy of the proposed approach by varying expert and criteria weights and to compare our technique with the other ones proposed in the literature.
4) *Robustness With Respect to Sockpuppet Attacks:* For analyzing how the proposed approach deals with this particular attack by varying the percentage of most suspicious accounts considered fraudsters and to compare the obtained results with respect to the state-of-the-art ones.

The *Yelp Dataset Challenge*,[2] a subset of data provided by *Yelp* mainly for research purposes, has been used to carry out the described analysis. In particular, it is composed by 1.3 million of users with different metadata (such as votes, stars, registration year, and so on), 174.000 business objects having over 1.2 thousand business attributes (such as hours, parking, ambient and so on), 5.2 million of reviews, including users wrote the review and the related business object. Moreover, we have evaluated the performance of the proposed approach also on *Amazon Customer Reviews Dataset*,[3] a subset of reviews

[1] https://azure.microsoft.com
[2] https://www.yelp.com/dataset
[3] https://s3.amazonaws.com/amazon-reviews-pds/readme.html

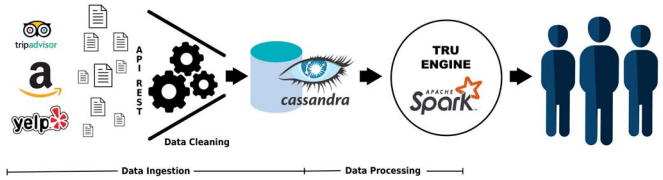

Fig. 3. Overall architecture.

about Amazon products over a period of two decades. In particular, the *quality* criteria have been computed on the basis of *Useful*, *funny* and *cool* votes of *Yelp* dataset and *total votes* attribute of the *Amazon* dataset whilst the *Usefulness* criteria has been defined according to the *Useful* attribute in both datasets. Furthermore, the *User Activity* criteria has been computed by combining the same attributes as the first two criteria and the number of reviews performed by the user. In addition, the Crown consensus criteria have been computed according to the review's content and time for both datasets. In addition, concerning the effectiveness evaluation we decided to use the *YelpNYC* dataset [27], containing information on 360.000 reviews about restaurants in New York City. We note that this dataset has been chosen because it is the most used one for measuring accuracy performances of user trustworthiness assessment techniques. Finally, the robustness evaluation has been performed on the *Amazon* dataset[4]—composed by 256 059 users, 74 258 products and 560 804 ratings—because it is widely used for analyzing the sockpuppet attack, as shown in [32].

B. Approach Implementation

Our envisioned approach has been implemented according to the system architecture depicted in Fig. 3, composed by two main components performing *data ingestion* and *data processing* tasks. The ingestion module realizes first the data crawling from heterogeneous sources (such as the above-mentioned Yelp, but also TripAdvisor, Foursquare) by using the related native application program interface (API). Thus, such information is opportunely cleaned by removing reviews composed by a small number of words [58], as they usually correspond to fake reviews, and successively they are stored into the NoSQL columnar database named *Cassandra*,[5] for easily supporting data aggregation operations for user reviews manipulation. More in detail, information is stored into two tables named, respectively, *User*, including five columns concerning the main statistic of a given user (*user_id*, *average_star*, *useful_votes*, *funny_votes*, *cool$_v$otes*), and *Review*, containing the main features about a given review (*review_id*, *user_id*, *business_id*, *average_star*, *useful_votes*, *funny_votes*, *cool$_v$otes*).

The data processing engine is mainly based on the Apache Spark framework[6] for properly supporting the proposed approach. The processing task leverages the *Spark SQL* module that extends the well-known database operations in a distributed environment for handling the required actions on the two described tables.

The proposed framework has been deployed on the Microsoft Azure *HDInsight*,[7] a cloud-based platform using a cluster composed by two D12v2 head nodes for managing the entire cluster, and by four D13v2 workers for executing the distributed jobs; finally, the technological stack is based on Spark 2.1.0 and Hadoop 2.7.

TABLE V
ACCURACY EVALUATION ON YELP DATASET CHALLENGE. H IS AN HIGH VALUE WHILE L IS A LOW VALUE. MORE IN DETAIL, OUR FRAMEWORK IS NAMED e_1 WHILE e_2 AND e_3 ARE HUMAN EXPERTS, RESPECTIVELY

w_{e_1}	w_{e_2}	w_{e_3}	w_{c_1}	w_{c_2}	w_{c_3}	w_{c_4}	w_{c_5}	ϵ
H	L	L	H	L	L	L	L	0.71
H	L	L	L	H	L	L	L	0.76
H	L	L	L	L	H	L	L	0.84
H	L	L	L	L	L	H	L	0.94
H	L	L	L	L	L	L	H	0.87
L	H	L	H	L	L	L	L	0.69
L	H	L	L	H	L	L	L	0.76
L	H	L	L	L	H	L	L	0.81
L	H	L	L	L	L	H	L	0.89
L	H	L	L	L	L	L	H	0.84
L	L	H	H	L	L	L	L	0.70
L	L	H	L	H	L	L	L	0.77
L	L	H	L	L	H	L	L	0.83
L	L	H	L	L	L	H	L	0.90
L	L	H	L	L	L	L	H	0.86

C. Results

As it can be observed from Fig. 4(a)–(c), the efficiency of our solution is compared with respect to *SpEagle+* [27] and *NetSpam* [28] by varying the number of users, reviews, and ratio between users and reviews. We can notice as performances strongly depend on the average number of reviews for user because our approach has mainly been focused on a set of aggregation operations over the reviews themselves. Furthermore, our approach shows better performances in terms of running times with respect to the other ones because they are essentially based on identifying metapaths on heterogeneous information networks (operations that are computationally onerous).

Fig. 5 reports cost analysis made varying the cluster configuration in terms of number of workers (i.e., configuration with two, four, or six workers), whose deployment costs ranges from 6.46 to 14.78 €/h. More in detail, we analyze the running times on three different datasets, namely, *Low*, *Medium*, and *High*, corresponding, respectively, to 10.000, 1.000.000, and 10.000.000 of reviews. It is important to note that an increase of resource number does not always disclose benefits in terms of ratio between running times and overall costs.

Then, we performed the parameter tuning of the proposed approach, whose results are shown in Tables V and VI, for examining its effectiveness varying both experts (w_{e_1}, w_{e_2}, and w_{e_3}) and criteria (w_{c_1}, w_{c_2}, w_{c_3}, w_{c_4}, and w_{c_5}) weights. Indeed, we computed the accuracy of our approach on the basis of the following formula:

$$\epsilon = 1 - \frac{\sum_{i=1}^{N} (\hat{\tau}_i - \tau_i)}{N} \tag{20}$$

[4]http://snap.stanford.edu/data/web-FineFoods.html
[5]http://cassandra.apache.org/
[6]https://spark.apache.org/
[7]https://azure.microsoft.com/it-it/services/hdinsight/

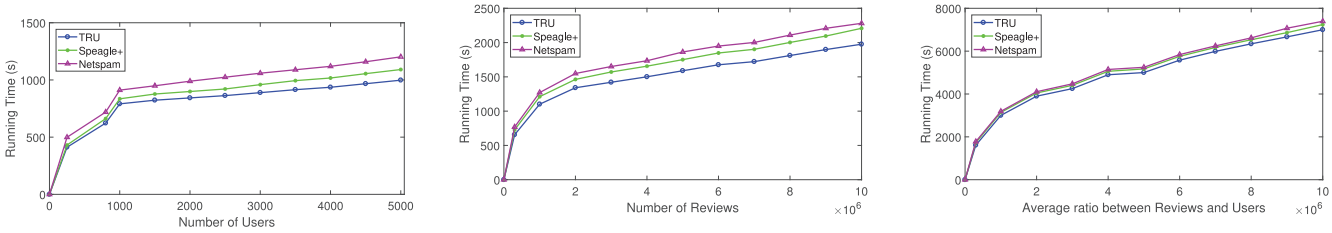

Fig. 4. Efficiency evaluation. (a) Running time varying number of users. (b) Running time varying number of reviews. (c) Running time varying average ratio between users and reviews.

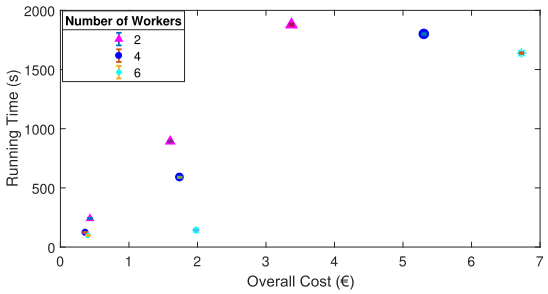

Fig. 5. Cost analysis varying cluster configuration. The markers' width denotes the size of the used dataset.

TABLE VI
ACCURACY EVALUATION ON AMAZON DATASET REVIEW. H IS AN HIGH VALUE WHILE L IS A LOW VALUE. MORE IN DETAIL, OUR FRAMEWORK IS NAMED e_1 WHILE e_2 AND e_3 ARE HUMAN EXPERTS, RESPECTIVELY

w_{e_1}	w_{e_2}	w_{e_3}	w_{c_1}	w_{c_2}	w_{c_3}	w_{c_4}	w_{c_5}	ϵ
H	L	L	H	L	L	L	L	0.68
H	L	L	L	H	L	L	L	0.72
H	L	L	L	L	H	L	L	0.79
H	L	L	L	L	L	H	L	0.83
H	L	L	L	L	L	L	H	0.86
L	H	L	H	L	L	L	L	0.66
L	H	L	L	H	L	L	L	0.69
L	H	L	L	L	H	L	L	0.74
L	H	L	L	L	L	H	L	0.78
L	H	L	L	L	L	L	H	0.82
L	L	H	H	L	L	L	L	0.70
L	L	H	L	H	L	L	L	0.73
L	L	H	L	L	H	L	L	0.77
L	L	H	L	L	L	H	L	0.81
L	L	H	L	L	L	L	H	0.84

where $\tau_i = 1 - [(n^i_{\text{fake}})/(n^i_{\text{review}})]$ is the trustworthiness of each user u_i measured using the number of the fake reviews of u_i (namely, n^i_{fake}) with respect to the totality of user reviews (namely, n^i_{review}), and $\hat{\tau}_i$ is the trustworthiness of the same user in according to our approach. The number of fake reviews for each user has been evaluated by using the approach shown in [27]. More in detail, in Tables V and VI, we show the accuracy of the proposed approach for the YELP and Amazon datasets according to the (20), by considering different combinations of experts and criteria weights, in which only one for each category has been considered high. We run several tests by varying the considered criteria in the trust computation and concluded that the configuration where all our criteria are included allows us to achieve the higher degree in terms of accuracy. In particular, our results achieves better performance considering crown consensus criteria as high.

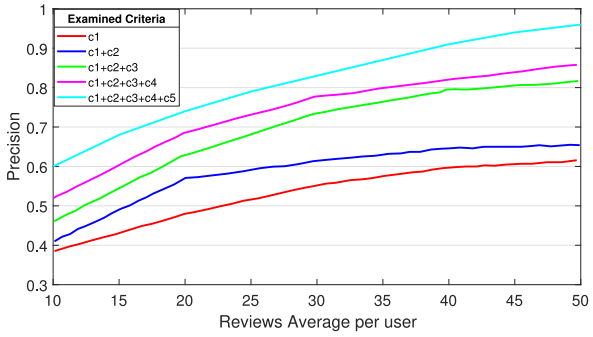

Fig. 6. Precision on the YELP dataset varying reviews average number per user where $c1$, $c2$, $c3$, $c4$, and $c5$ corresponds, respectively, to the above-described criteria.

Table V shows how combining different features in *YELP* allows us to increase the quality of our approach. More in detail, we can observe how the ways used by human for choosing a given business object (i.e., an hotel or a pub) are mainly focused on the analysis of how opinions and votes that are assigned by other users vary over the time. In fact, we obtain the best results by considering the fourth criteria (*time-dependent crown consensus*) as high. In Table VI, it is possible to note, in turn, how the lack of interest in the analysis of user reviews for the choice of a given product affects the quality of our approach when using Amazon data. More in detail, the number of assigned votes to each review is less then *YELP* dataset emphasizing the relevance of the criteria based on the reviews' semantic analysis (*content-dependent crown consensus*).

In addition, we evaluated the effectiveness of the proposed approach in terms of precision varying the average number of reviews per user. In particular, we compared five different versions of the proposed methodology by including a subset of the defined criteria.

As easy to note in Fig. 6, the runs-based mainly on the number of useful or received votes have an initial increase in quality by increasing of review average number, but successively they suffer of low rating, while the analysis of the activity of a user allows a better characterization and better results. Finally, the use of crown consensus criteria allows to evaluate how the user behaves with respect to the majority, which obviously requires a large amount of data to operate.

Eventually, we provide the precision analysis for the Amazon Dataset Review in the same way as made for Yelp in

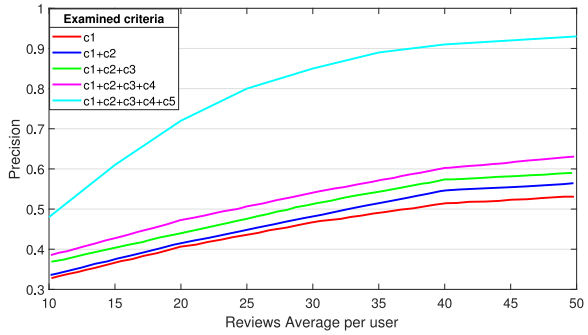

Fig. 7. Precision on Amazon dataset varying reviews average number per user where $c1$, $c2$, $c3$, $c4$, and $c5$ corresponds, respectively, to the above-described criteria.

TABLE VII
EFFICACY ANALYSIS OF THE PROPOSED APPROACH ON THE *YelpNYC* DATASET WITH THE RESPECT TO OTHER ONES PROPOSED IN LITERATURE IN TERMS OF AP AND AUC VARYING THE NUMBER OF EXAMINED REVIEWS (60%, 80% AND ENTIRE DATASET)

	60%		80%		100%	
	AP	**AUC**	**AP**	**AUC**	**AP**	**AUC**
Random	0.1482	0.4793	0.1577	0.4891	0.1782	0.5000
FraudEagle	0.1898	0.5324	0.2122	0.5781	0.2233	0.6062
Wang et al.	0.1845	0.5436	0.2278	0.5812	0.2381	0.6207
SpEagle	0.2399	0.5812	0.2588	0.6212	0.2680	0.6575
Alrubaian et al.	0.2723	0.5789	0.3412	0.6345	0.3725	0.6701
SpEagle+	**0.2912**	**0.6013**	0.3245	0.6412	0.3480	0.6828
NetSpam	0.2844	0.5875	0.3341	0.6123	0.3634	0.6947
Proposed	0.2878	0.5901	**0.3511**	**0.6576**	**0.3882**	**0.7144**

Fig. 7. It is possible to note that the user activity and divergence criteria do not have great impact on the quality of the approach, given the average number of reviews of users on such networks is very low. In turn, the analysis of reviews' content allows to improve the precision results analyzing users reviews divergence from other user reviews.

Finally, we evaluated the effectiveness of the proposed approach on the *YelpNYC* dataset with respect to *SpEagle* [27], *SpEagle*$^+$ [27], NetSpam [28], FraudEagle[59], Wang *et al.* [60], and Alrubaian *et al.* [42] according to average precision (AP) and area under the curve (AUC) metrics, concerning the user ranking challenge. In a similar manner to [27], we separated the users into two classes on the base of S-value: 1) *spammers* (authors of unreliable reviews belonging to S_0, S_1 and S_2) and 2) *benign* (authors with reliable reviews corresponding to S_3 and S_4). As shown in Table VII SEagle+ provides better performance considering 60% of the dataset because it better fits using small reviews number. In turn, the proposed approach shows better efficacy values in terms of AP and AUC. In conclusion, the use of our proposed time-dependent crown consensus allows to assess more precisely the trustworthiness of a given user than the reported state-of-the-art.

D. Robustness With Respect to Sockpuppet Attack

In this section, we describe the evaluation made for analyzing how the proposed approach deals with the *sockpuppet* attack [32], which concerns the use of several accounts to artificially post high star ratings of a product whilst they publish real reviews about a set of other products. To this aim, we have evaluated the proposed method against several state-of-the-art approaches (i.e., REV2 [46], BIRDNEST [48], FraudEagle [59], FRAUDAR [47], RTV [32], and RTV-SUP [32]) on the *Amazon* dataset in terms of $F1$ score, Precision, Recall varying the percentage (q%) of the most suspicious accounts considered fraudsters. This evaluation has been performed according to the experimental protocol shown in [32], labeling 10 000 users as verified that could be either malicious or benign. In particular, after each experiment using a fake review detection algorithm, we measured how precision, recall, and $F1$ score varied after the attack by computing their relative change rc_A (in the worst case), defined as $rc_A = [(v_a - v_b)/v_b]$, v_b being the value of the measure obtained before applying the attack and v_a the value after applying the attack.

TABLE VIII
RELATIVE CHANGES OF PRECISION, RECALL AND F1-MEASURE VARYING THE PERCENTAGE OF MOST SUSPICIOUS ACCOUNTS (q)

	q	REV2	BIRDNEST	FraudEagle	FRAUDAR	RTV	RTV-SUP	Proposed
F1-score	0.5%	-0.98	-0.93	-1.0	-0.84	-0.02	-0.08	-0.15
	1%	-0.99	-0.92	-1.0	-0.81	-0.96	-0.03	-0.18
	3%	-0.96	-0.84	-1.0	-0.53	-0.05	-0.24	-0.35
	5%	-0.97	-0.78	-1.0	-0.28	-0.07	-0.21	-0.25
Recall	0.5%	-0.99	-0.95	-1.0	-0.91	-0.01	-0.05	-0.13
	1%	-0.99	-0.93	-1.0	-0.90	-0.97	-0.04	-0.18
	3%	-0.99	-0.91	-1.0	-0.74	-0.06	-0.25	-0.36
	5%	-0.99	-0.87	-1.0	-0.61	-0.07	-0.21	-0.26
Precision	0.5%	-0.46	0.83	-1.0	5.82	-0.03	-0.07	-0.12
	1%	-0.48	0.87	-1.0	5.87	-0.94	-0.08	-0.15
	3%	-0.78	0.82	-1.0	5.28	-0.03	-0.12	-0.19
	5%	-0.81	0.79	-1.0	5.25	-0.02	-0.14	-0.21

Table VIII shows the obtained results, where it is possible to note that RTV and *RTV-SUP* outperforms other approaches because these two techniques are specifically designed for sockpuppet attack. Nevertheless, the proposed approach achieves comparable results with respect the majority of the state-of-the-art techniques although it was specially designed to be generic. Moreover, the proposed approach does not require the presence of trusted users who are difficult to identify within a social network.

V. FINAL REMARKS

This study proposed a solution to the problem of trust management within the context of the social networks, where it is important to deal with the subjectivity of the detection of malicious behaviors and the need of objectivity in order to design an automatic process to assign trust degrees to users based on their activity in the social network. To this aim, we have approached the vagueness and subjectivity in the review analysis from the social network by means of the fuzzy theory. We have leveraged on the theory of evidence so as to device a MCME-DM process to aggregate the judgments from multiple perspectives and optimize the trust estimation. We have performed a realistic experimental campaign considering the YELP and Amazon dataset and showed that aggregating the output of multiple criteria allows achieving higher accuracy in detecting malicious reviews. We have also compared our approach against the main related works in the existing literature and showed that our approach obtained better efficacy by using 80% and 100% of the considered dataset.

As future work, we plan to investigate more in detail the influence of common attacks toward a recommendation system so as to enhance the security of such a solution, in addition to the study of the privacy concerns of such systems, by considering the key legal frameworks, such as the The EU General Data Protection Regulation (GDPR). Moreover, the main critics to D-S aggregation are to return counterintuitive results when combining unreliable evidences [61] and/or conflicting evidences from independent sources [62]. In order to improve the detection of a potential problem in the aggregation process, special formulations of the mass functions and other concepts of the D-S theory emerged over the last decade, such as the evolutionary combination rule (ECR) in [63]. We have left as future work the investigation of this approach within the context of our work.

REFERENCES

[1] M. Faloutsos, T. Karagiannis, and S. Moon, "Online social networks," *IEEE Netw.*, vol. 24, no. 5, pp. 4–5, Sep/Oct. 2010.

[2] J. Castro, J. Lu, G. Zhang, Y. Dong, and L. Martinez, "Opinion dynamics-based group recommender systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 12, pp. 2394–2406, Dec. 2018.

[3] F. Xiong, X. Wang, S. Pan, H. Yang, H. Wang, and C. Zhang, "Social recommendation with evolutionary opinion dynamics," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 10, pp. 3804–3816, Oct. 2020.

[4] R. Ureña, G. Kou, Y. Dong, F. Chiclana, and E. Herrera-Viedma, "A review on trust propagation and opinion dynamics in social networks and group decision making frameworks," *Inf. Sci.*, vol. 478, pp. 461–475, Apr. 2019.

[5] Y. Xiang, E. Bertino, and M. Kutylowski, "Security and privacy in social networks," *Concurrency Comput. Practice Exp.*, vol. 29, no. 7, 2017, Art. no. e4093.

[6] D. Irani, S. Webb, K. Li, and C. Pu, "Modeling unintended personal-information leakage from multiple online social networks," *IEEE Internet Comput.*, vol. 15, no. 3, pp. 13–19, May/Jun. 2011.

[7] A. Nosko, E. Wood, and S. Molema, "All about me: Disclosure in online social networking profiles: The case of Facebook," *Comput. Human Behav.*, vol. 26, no. 3, pp. 406–418, 2010.

[8] K. Krombholz, D. Merkl, and E. Weippl, "Fake identities in social media: A case study on the sustainability of the Facebook business model," *J. Service Sci. Res.*, vol. 4, no. 2, pp. 175–212, 2012.

[9] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Commun. ACM*, vol. 59, no. 7, pp. 96–104, 2016.

[10] X. Wang *et al.*, "Game theoretic suppression of forged messages in online social networks," *IEEE Trans. Syst., Man, Cybern., Syst.*, early access, Mar. 5, 2019, doi: 10.1109/TSMC.2019.2899626.

[11] M. A. Ferrag, L. Maglaras, and A. Ahmim, "Privacy-preserving schemes for ad hoc social networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 3015–3045, 4th Quart., 2017.

[12] I. Kayes and A. Iamnitchi, "Privacy and security in online social networks: A survey," *Online Soc. Netw. Media*, vols. 3–4, pp. 1–21, Oct. 2017.

[13] S. R. Sahoo and B. B. Gupta, "Classification of various attacks and their defence mechanism in online social networks: A survey," *Enterprise Inf. Syst.*, vol. 13, no. 6, pp. 832–864, 2019.

[14] C. Zhang, J. Sun, X. Zhu, and Y. Fang, "Privacy and security for online social networks: Challenges and opportunities," *IEEE Netw.*, vol. 24, no. 4, pp. 13–18, Jul. 2010.

[15] H. Gao, J. Hu, T. Huang, J. Wang, and Y. Chen, "Security issues in online social networks," *IEEE Internet Comput.*, vol. 15, no. 4, pp. 56–63, Jul./Aug. 2011.

[16] F. Buccafurri, G. Lax, D. Migdal, S. Nicolazzo, A. Nocera, and C. Rosenberger, "Contrasting false identities in social networks by trust chains and biometric reinforcement," in *Proc. Int. Conf. Cyberworlds*, 2017, pp. 17–24.

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

[18] G. Liu *et al.*, "TOSI: A trust-oriented social influence evaluation method in contextual social networks," *Neurocomputing*, vol. 210, pp. 130–140, Oct. 2016.

[19] H. Xia, F. Xiao, S.-S. Zhang, X.-G. Cheng, and Z.-K. Pan, "A reputation-based model for trust evaluation in social cyber-physical systems," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 792–804, Apr.–Jun. 2020.

[20] X. Niu, G. Liu, and Q. Yang, "Trustworthy website detection based on social hyperlink network analysis," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 54–65, Jan.–Mar. 2020.

[21] R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," *Manag. Sci.*, vol. 17, no. 4, pp. B141–B164, 1970.

[22] C. Esposito, A. Castiglione, and F. Palmieri, "Information theoretic-based detection and removal of slander and/or false-praise attacks for robust trust management with Dempster–Shafer combination of linguistic fuzzy terms," *Concurrency Comput. Practice Exp.*, vol. 30, no. 3, 2018, Art. no. e4302.

[23] S. K. T. Lam, D. Frankowski, and J. Riedl, "Do you trust your recommendations? An exploration of security and privacy issues in recommender systems," in *Emerging Trends in Information and Communication Security*. Berlin, Germany: Springer, 2006, pp. 14–29.

[24] R. Katarya, "A systematic review of group recommender systems techniques," *Proc. Int. Conf. Intell. Sustain. Syst. (ICISS)*, Dec. 2017, pp. 425–428.

[25] M. Casanovas and J. Merigó, "Fuzzy aggregation operators in decision making with Dempster–Shafer belief structure," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7138–7149, 2012.

[26] S. Kc and A. Mukherjee, "On the temporal dynamics of opinion spamming: Case studies on yelp," in *Proc. 25th Int. Conf. World Wide Web*, 2016, pp. 369–379.

[27] S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2015, pp. 985–994.

[28] S. Shehnepoor, M. Salehi, R. Farahbakhsh, and N. Crespi, "NetSpam: A network-based spam detection framework for reviews in online social media," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1585–1595, Jul. 2017.

[29] Y. Liu, Y. Sun, S. Liu, and A. C. Kot, "Securing online reputation systems through trust modeling and temporal analysis," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 936–948, Jun. 2013.

[30] Y. Ruan and A. Durresi, "A survey of trust management systems for online social communities–trust modeling, trust inference and attacks," *Knowl. Based Syst.*, vol. 106, pp. 150–163, Aug. 2016.

[31] C. Timm and R. Perez, *Seven Deadliest Social Network Attacks*. Rockland, MA, USA: Syngress, 2010.

[32] R. Liu, R. Liu, A. Pugliese, and V. Subrahmanian, "STARS: Defending against sockpuppet-based targeted attacks on reviewing systems," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 5, p. 56, 2020.

[33] M. R. Subramani and B. Rajagopalan, "Knowledge-sharing and influence in online social networks via viral marketing," *Commun. ACM*, vol. 46, no. 12, pp. 300–307, 2003.

[34] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decis. Support Syst.*, vol. 43, no. 2, pp. 618–644, 2007.

[35] J. Cho, A. Swami, and I. Chen, "A survey on trust management for mobile ad hoc networks," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 562–583, 4th Quart., 2011.

[36] W. M. Grudzewski, A. Sankowska, and M. Wańtuchowicz, *Trust Management in Virtual Work Environments: A Human Factors Perspective*. Hoboken, NJ, USA: CRC Press, 2008.

[37] Z. Zhang and B. B. Gupta, "Social media security and trustworthiness: Overview and new direction," *Future Gener. Comput. Syst.*, vol. 86, pp. 914–925, Sep. 2018.

[38] J. Yu, K. Wang, P. Li, R. Xia, S. Guo, and M. Guo, "Efficient trustworthiness management for malicious user detection in big data collection," *IEEE Trans. Big Data*, early access, Oct. 10, 2017, doi: 10.1109/TBDATA.2017.2761386.

[39] J. Yu, Z. Kuang, B. Zhang, W. Zhang, D. Lin, and J. Fan, "Leveraging content sensitiveness and user trustworthiness to recommend fine-grained privacy settings for social image sharing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1317–1332, May 2018.

[40] S. Deng, L. Huang, G. Xu, X. Wu, and Z. Wu, "On deep learning for trust-aware recommendations in social networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1164–1177, May 2017.

[41] J. Ye, S. Kumar, and L. Akoglu, "Temporal opinion spam detection by multivariate indicative signals," in *Proc. 10th Int. AAAI Conf. Web Soc. Media*, 2016, pp. 743–746.

[42] M. Alrubaian, M. Al-Qurishi, M. M. Hassan, and A. Alamri, "A credibility analysis system for assessing information on Twitter," *IEEE Trans. Depend. Secure Comput.*, vol. 15, no. 4, pp. 661–674, Jul./Aug. 2018.

[43] X. Zhang, L. Cui, and Y. Wang, "CommTrust: Computing multi-dimensional trust by mining e-commerce feedback comments," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp. 1631–1643, Jul. 2014.
[44] S. Castano, A. Ferrara, L. Genta, and S. Montanelli, "Combining crowd consensus and user trustworthiness for managing collective tasks," *Future Gener. Comput. Syst.*, vol. 54, pp. 378–388, Jan. 2016.
[45] S. Kumar, B. Hooi, D. Makhija, M. Kumar, C. Faloutsos, and V. Subrahamanian, "FairJudge: Trustworthy user prediction in rating platforms," 2017. [Online]. Available: arXiv:1703.10545.
[46] S. Kumar, B. Hooi, D. Makhija, M. Kumar, C. Faloutsos, and V. Subrahmanian, "REV2: Fraudulent user prediction in rating platforms," in *Proc. 11th ACM Int. Conf. Web Search Data Min. (WSDM)*, 2018, pp. 333–341.
[47] B. Hooi, H. A. Song, A. Beutel, N. Shah, K. Shin, and C. Faloutsos, "FRAUDAR: Bounding graph fraud in the face of camouflage," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2016, pp. 895–904.
[48] B. Hooi *et al.*, "BirdNest: Bayesian inference for ratings-fraud detection," in *Proc. SIAM Int. Conf. Data Min.*, 2016, pp. 495–503.
[49] R. B. Cialdini, *Influence: Science and Practice*, vol. 4. Boston, MA, USA: Pearson Educ., 2009.
[50] L. Lamport, R. Shostak, and M. Pease, "The Byzantine generals problem," *ACM Trans. Program. Lang. Syst.*, vol. 4, no. 3, pp. 382–401, 1982.
[51] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.
[52] F. G. Marmol, J. Marin-Blazquez, and G. M. Perez, "LFTM, linguistic fuzzy trust mechanism for distributed networks," *Concurrency Comput. Practice Exp.*, vol. 24, no. 17, pp. 2007–2027, 2012.
[53] C. Martinez-Cruz, C. Porcel, J. Bernabe-Moreno, and E. Herrera-Viedma, "A model to represent users trust in recommender systems using ontologies and fuzzy linguistic modeling," *Inf. Sci.*, vol. 311, pp. 102–118, Aug. 2015.
[54] P. Bonissone and K. Decker, "Selecting uncertainty calculi and granularity: An experiment in trading-off precision and complexity," in *Proc. 1st Annu. Conf. Uncertainty Artif. Intell. (UAI)*, Jul. 1985, pp. 217–248.
[55] W. Duch, "Uncertainty of data, fuzzy membership functions, and multilayer perceptrons," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 10–23, Jan. 2005.
[56] C. Esposito, A. Castiglione, F. Palmieri, and M. Ficco, "Trust management for distributed heterogeneous systems by using linguistic term sets and hierarchies, aggregation operators and mechanism design," *Future Gener. Comput. Syst.*, vol. 74, pp. 325–336, Sep. 2017.
[57] R. Rodriguez, L. Martinez, and F. Herrera, "Hesitant fuzzy linguistic term sets for decision making," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 1, pp. 109–119, Feb. 2012.
[58] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?" in *Proc. 7th Int. Conf. Weblogs Soc. Media (ICWSM)*, Jan. 2013, pp. 409–418.
[59] L. Akoglu, R. Chandy, and C. Faloutsos, "Opinion fraud detection in online reviews by network effects," in *Proc. 7th Int. AAAI Conf. Weblogs Soc. Media*, 2013, pp. 1–10.
[60] G. Wang, S. Xie, B. Liu, and S. Y. Philip, "Review graph based online store review spammer detection," in *Proc. IEEE 11th Int. Conf. Data Min.*, 2011, pp. 1242–1247.
[61] D. Han, Y. Deng, and C. Han, "Sequential weighted combination for unreliable evidence based on evidence variance," *Decis. Support Syst.*, vol. 56, pp. 387–393, Dec. 2013.
[62] L. A. Zadeh, "Review of a mathematical theory of evidence," *AI Mag.*, vol. 5, no. 3, p. 81, 1984.
[63] X. Deng, D. Han, J. Dezert, Y. Deng, and Y. Shyr, "Evidence combination from an evolutionary game theory perspective," *IEEE Trans. Cybern.*, vol. 46, no. 9, pp. 2070–2082, Sep. 2016.

Christian Esposito received the Ph.D. degree in computer engineering and automation from the University of Naples "Federico II," Naples, Italy in 2009.

He is a Tenured Assistant Professor with the University of Salerno, Fisciano, Italy, and was an Assistant Professor with the University of Naples "Federico II." He has authored more than 50 journal publications and 30 conference papers, and has been involved in several international and national research and industrial projects. His research interests include reliable and secure communications, middleware, distributed systems, multiobjective optimization, and game theory.

Dr. Esposito has served as a reviewer or a guest editor for several international journals and conferences, and has been a PC member or a organizer of about 50 international conferences/workshops. He is also a member of three journal editorial boards.

Vincenzo Moscato received the Ph.D. degree in computer science and engineering from the University of Naples "Federico II," Naples, Italy in 2005.

He is currently an Associate Professor of Database and Information Systems with the DIETI Department of Electrical Engineering and Information Technologies, University of Naples "Federico II." He was involved in several international, national, and local research projects. He is currently an author of more than 100 publications on international journal and conference proceedings. His current research interests lie in the area of multimedia, knowledge management, and big data analytics.

Giancarlo Sperlí received the Ph.D. degree in information technology and electrical engineering from the University of Naples "Federico II," Naples, Italy in 2018.

He is a Research Fellow with the University of Naples "Federico II." He has authored more than 60 publications on international journal and conference proceedings. His main research interests are in the area of cybersecurity, semantic analysis of multimedia data, and social networks analysis.