IJNRD.ORG ISSN : 2456-4184

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG

An International Open Access, Peer-reviewed, Refereed Journal

International Research Journal IJNRD Research Through Innovation

Trustworthiness Assessment of Users in Social Reviewing Systems

C. Tamilselvi, Dr. P. Malathi, Ph.D.

Computer Science Engineering, Madha Engineering College, Tamilnadu, India.

r.tamilselviwaran@gmail.com

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 multi criteria multi expert 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 sock puppet attacks.

1. INTRODUCTION

As well known, the online social networks are Internet-enabled applications used by people to establish social relations with the other individuals sharing similar personal interests and/or activities. A part 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., Trip Advisor, Four square, etc.) or to provide social environments able to facilitate particular tasks (e.g., the search of a job as in Linked In, the answer to research questions as in Research Gate, 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. 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. 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, mainly due to the fact that several personal and sensitive informationare shared, and leaked, throughout SRS, and that a person may choose to hide its true self and intentions behind a totally false virtual identity or a Bot (short for software robots) may mimic human behavior in SRS. 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, 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. The issue of providing privacy has led to the adoption of access control means, while counter acting forging nodes/identities and social links/connections demanded authentication of users and exchanged messages. 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. 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. 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, 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. 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. It is not rare that actors may interact with non adjacent other actors, so it is important to find a trust path among non adjacent 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 user shaving 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 and make its fully automation extremely complex. Protecting the overall aggregation from the impact of malicious or fake reputations is an issue to consider, and falls within the literature of security and privacy of *Recommender Systems*.

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 judgment, we leverage on the fuzzy theory as widely used in 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 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 *Yelp NYC* 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, focusing on the identification of linguistic features of malicious reviews; 2) behavioral based, leveraging metadata information of submitted review and user profile for identifying fake reviews; and 3) graph-based approaches, 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 overtime 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 and 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, a sit may be seen as an additional criterion to be aggregated with the D-S combination rule within our proposed multi criteria 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, and in protecting the system against the sock puppet attacks.

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 proto type, 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.

2. LITERATURE SURVEY

Liu *et al.* investigated the sock puppet attacks on reviewing 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.* for assessing information credibility on.

Twitter. In, the *Comm Trust*framework has been introduced for trust evaluations by mining feedback comments. More in detail, it is based on a multi dimensional trust model for computing reputation scores from user feedback comments, which are analyzing combining natural language processing techniques, opinion mining, and topic modeling. Further more, another framework, namely, *Liquid Crowd*, has been proposed by Castano*et al.* exploiting consensus and trustworthiness techniques for managing the execution of collective tasks.

Kumar *et al.* 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.* described a system for identifying fraudulent users based on six axioms to define the inter dependency among three intrinsic quality metrics concerning a user, reliability and goodness of a product by combining network and behavior properties.

Hooi*et al.* developed an algorithm, called *Fraud AR*, aiming at being resistant to the camouflage attacks, for identifying fake reviews and users. Further more, *Bird nest*, 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.* 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.

3. 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* ∪ *BO* and *E* is composed by the set of reviews, with the related metadata (*λ*0, *λ*1, . . . , *λn_o*), made by users on different business objects, whose number is *n_o*. *ρo(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 *ρo(t)* and *qo(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 *λi*), 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 *τ(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:

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

Such a value in (1) should be computing for the overall observation time: $\tau = \int_{t_0}^{t_1} \tau(t)dt$. 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* ≤ 3*F*. 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* = 3*F* +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 $\rho\hat{}_o(t)$ 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 $\rho_o(t)$ = 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:

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

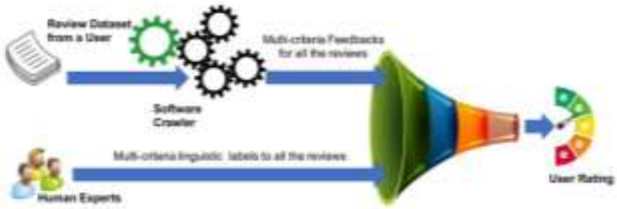

Review Dataset from a User
Software Crawler
Multi-criteria Feedbacks for all the reviews
Multi-criteria linguistic labels to all the reviews
Human Experts
User Rating

Fig. 1. Schematic representation of the proposed approach.

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 so on. How properly aggregating these multi value 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 multi criteria multi expert 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, 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. 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, 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:

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

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, 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 $si \leq sj \Longleftrightarrow i \leq j$. 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. 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 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.

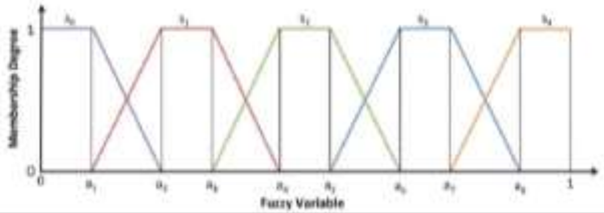


Fig. 2. Set of five terms in (3) and its semantics as fuzzy sets.

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 $(u_i, r_{i,o}(t))$ assuming 1 if the review $r_{i,o}(t)$ 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 $(u_i, r_{i,o}(t))$assuming 1 if the review $r_{i,o}(t)$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:

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

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:

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

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* ∈ *U*, expressed as follows:

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

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

$$c_3(u_j) = \frac{\mathrm{voteSent}(u_j) + \mathrm{reviewsLeft}(u_j)}{\max_{u \in U} [\mathrm{voteSent}(u) + \mathrm{reviewsLeft}(u)]}. \quad (7)$$

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 pre-review 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 $\beta_o(t_r, t_2)$ $c_4(u_i)$ =

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

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 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 *S4* (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 ($t_1 < t_2$), $\_r_{i,o}(t_r) = \{f_1, f_2, . . . , f_N\}$ 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 $r_{i,o}(t_r) = \{f_1, f_2, . . . , f_N\}$ 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 $r_{i,o}(t_r)$ of the ith user and the $r_{c,o}(t_r)$ 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 =

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

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

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

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.

4. 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 Azure1 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 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

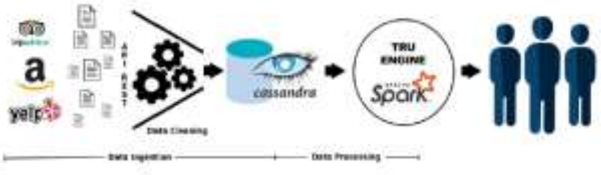


Fig. 3. Overall architecture.

Fig. 3. Overall architecture.

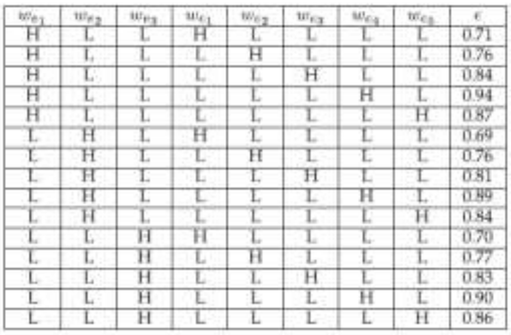
use the *YelpNYC* dataset, containing information on 360.000 reviews about restaurants in New York City. We note that this dataset has been chosen because it is the most used one for measuring accuracy performances of user trustworthiness assessment techniques. Finally, the robustness evaluation has been performed on the *Amazon* dataset4—composed by 256 059 users, 74 258 products and 560 804 ratings—because it is widely used for analyzing the sockpuppet attack.

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, 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*, *coolvotes*), and *Review*, containing the main features about a given review (*review_id*, *user_id*, *business_id*, *average_star*, *useful_votes*, *funny_votes*, *coolvotes*).

The data processing engine is mainly based on the Apache Spark framework6 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.

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

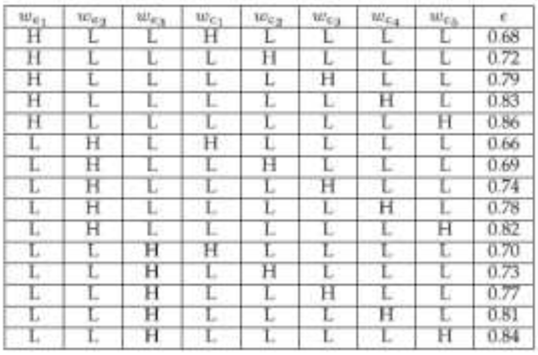
The proposed framework has been deployed on the Microsoft Azure *HDInsight*, 7 a cloud-based platform using a cluster composed by two D12v2 head nodes for managing the entire cluster, and by four D13v2 workers for executing the distributed jobs; finally, the technological stack is based on Spark 2.1.0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

C. Results

As it can be observed from Fig. 4(a)–(c), the efficiency of our solution is compared with respect to *SpEagle+* and *NetSpam* 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 e/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.

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

Then, we performed the parameter tuning of the proposed approach, whose results are shown in Table, for examining its effectiveness varying both experts (*we*1 , *we*2 , and *we*3) and criteria (*wc*1 , *wc*2 , *wc*3 , *wc*4 , and *wc*5) weights. Indeed, we computed the accuracy of our approach on the basis of the following formula:

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

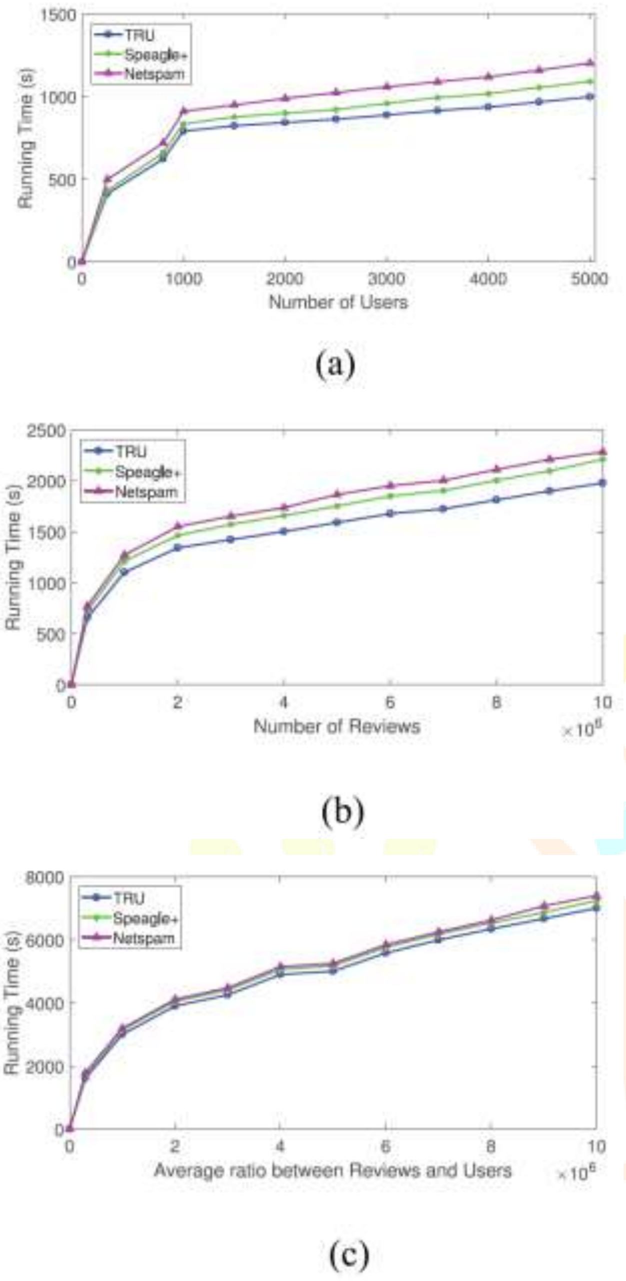
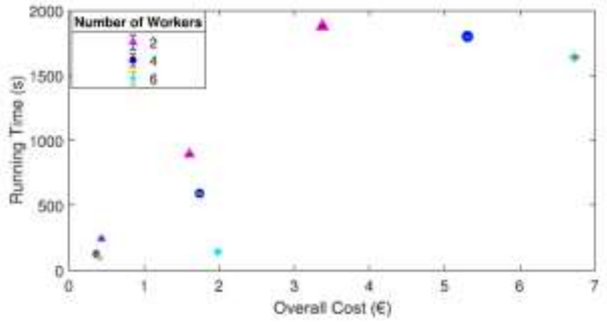


Fig. 4. Efficiency evaluation. (a) Running time varying number of users. (b) Running time varying number of reviews. (c) Running time varying average ratio between users and reviews.



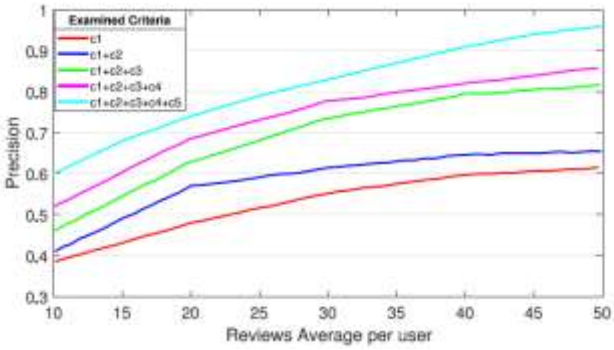


Fig. 5. Cost analysis varying cluster configuration. The markers' width denotes the size of the used dataset.

5. 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 and/or conflicting evidences from independent sources. 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).

REFERENCES

[1] 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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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