# A Graph-Based Comprehensive Reputation Model: Exploiting the Social Context of Opinions to Enhance Trust in Social Commerce

Su-Rong Yan[a,b], Xiao-Lin Zheng[a,d*], Yan Wang[c], William Wei Song[e],Wen-Yu Zhang[b]

[a]*College of Computer Science, Zhejiang University, 310027, Hangzhou, China*
[b]*College of Information, Zhejiang University of Finance and Economics, 310018, Hangzhou, China*
[c]*Department of Computing, Macquarie University, Sydney, NSW 2109, Australia*
[d]*Stanford University, Stanford, CA 94305, USA*
[e]*Dalarna University, Sweden*

**ABSTRACT:** Social commerce is a promising new paradigm of e-commerce. Given the open and dynamic nature of social media infrastructure, the governance structures of social commerce are usually realized through reputation mechanisms. However, the existing approaches to the prediction of trust in future interactions are based on personal observations and/or publicly shared information in social commerce application. As a result, the indications are unreliable and biased because of limited first-hand information and stakeholder manipulation for personal strategic interests. Methods that extract trust values from social links among users can improve the performance of reputation mechanisms. Nonetheless, these links may not always be available and are typically sparse in social commerce, especially for new users. Thus, this study proposes a new graph-based comprehensive reputation model to build trust by fully exploiting the social context of opinions based on the activities and relationship networks of opinion contributors. The proposed model incorporates the behavioral activities and social relationship reputations of users to combat the scarcity of first-hand information and identifies a set of critical trust factors to mitigate the subjectivity of opinions and the dynamics of behaviors. Furthermore, we enhance the model by developing a novel deception filtering approach to discard "bad-mouthing" opinions and by exploiting a personalized direct distrust (risk) metric to identify malicious providers. Experimental results show that the proposed reputation model can outperform other trust and reputation models in most cases.

***Keywords***: social commerce, trust and reputation, social context, theory of reasoned action, risk tolerance

## 1. Introduction

Buyers in e-commerce can neither physically examine a product nor verify the reliability of the seller given their temporal and spatial separation from sellers [34]. In such a context, buyers usually have limited information regarding the sellers and the goods; hence, they experience a high degree of uncertainty [35]. Meanwhile, social media tools such as social networking services (SNS) enable people to share their opinions regarding a product and transaction [39,53]. To this end, an increasing number of e-commerce industries have adopted SNS to encourage user interactions, including eBay.com, Amazon.com, and Taobao.com. Product review sites such as Epinions.com utilize the same

* Corresponding Author. Tel.: 1-650-319-5822, 0086-13588029069; Fax: 0086-571-87951453
E-mail address: xlzheng@zju.edu.cn ;zhengxl@stanford.edu

tools, which are part of the larger emerging phenomenon of social commerce [3,30,41,50] wherein the business activities of companies are supported by the voluntary effort of external partners [9]. Therefore, the use of such social media generates new revenue opportunities for marketers and businesses in online shopping while providing consumers with product information and advice. These parties obtain both economic and social rewards for sharing [17].

With the aid of social media tools, each user within a community should ideally have the same communicative potential. However, interested parties or stakeholders can easily manipulate online reviews for their strategic interests given the open and dynamic nature of social media infrastructure. Consumers also have external incentives to misreport and thus misrepresent the reviews available to other users [2,12]. Consequently, potential buyers discount actual reviews heavily as a result of the veracity of reviews questioned under deceptive environments.

Consumers are driven to value the decisions and opinions of social relationship members in product purchasing as per basic behavioral psychology [28,33]. However, new consumers usually have limited or virtually no direct interaction or relationship with other consumers in the context of social commerce because of the strong community structure in social networks [59]. This observation implies that new entrants often serve only the small communities (cliques) of direct acquaintances. A review is the subjective perspective of a consumer regarding his/her experiences in community activities and may merely represent his/her individual preference and opinion. Furthermore, trusted friends may not have similar tastes. Consequently, users who increasingly rely solely on their acquaintance communities for information readily encounter collective community bias [37]. In this case, the objectivity of reviews is not fully guaranteed under subjective environments.

The assessment of information credibility is a more challenging problem in social media than in conventional media because the username (virtual identity [25]) is the sole piece of information on its source. Social media is characterized by a lack of strong governance structures [37]. As a form of social governance, reputation mechanisms are among the most successful and widespread incentive mechanisms on the Internet. A reputation system uses a specific method (e.g., averaging, social network-based, game-theoretic, probabilistic-based, or belief-based) to compute the reputation values for a set of objects (e.g., users, goods, or services) within a community based on rating data collected from others [21,43].

Social voting (user voting) is a simple but widely used reputation mechanism that has been applied to indicate the usefulness of reviews and the popularity of reviewers. Thus, the cognitive load of users is reduced. However, cheating behaviors such as vote-buying, vote-exchanging, and fake news reduce the reliability of voting results [29]. In addition, social voting mechanisms suffer from various types of bias, including the imbalance vote bias (users tend to value others' opinions positively rather than negatively), winner circle bias (reviews with many votes attract much attention and therefore accumulate votes disproportionately), and early bird bias (the first reviews to be published tend to obtain more votes, but newly posted reviews are most likely to receive no votes or only a few votes) [31,46]. Without proper measures, the reputation mechanism obtains and produces unreliable information.

User credibility is generally addressed by solving the information credibility problem in social media. The veracity and objectivity of opinions are mainly influenced by the honesty and volume of the opinion contributors or so-called "raters" who share their experiences or opinions with others. In line with this information, different approaches have been proposed to determine and filter out deceptive information provided by raters [8,11,52]. The approaches that trust public information

typically assume that the majority of raters are honest, whereas those that depend only on personal observations may fail in settings where in the observations of raters are inadequate[40]. Social network-based methods [16] that extract trust values from social links among users can improve the performance of reputation mechanisms based solely on the information obtained from limited personal observations or from unknown users. However, these links may not always be available and are typically sparse in social commerce, especially for new users [57]. Furthermore, malicious users can still exploit the perceived social connections among users to easily publicize misinformation in social commerce if social links are available [1,7]. Some incentive-based proposals [13,27] are usually set up based on the rational behavior model of economics theory. These proposals have a sound logical foundation but do not seriously consider user behavior strategies (such as collusion or bad-mouthing). These problems are intrinsic in the processes of obtaining genuine feedback about actual interactions and estimating reputations accurately.

To circumvent the aforementioned problems, this paper proposes a new graph-based comprehensive reputation (CR) model for social commerce based on consumer behavior and psychological theory to improve the veracity and objectivity of opinions and to enhance consumer trust in (product and opinion) providers even in the presence of malicious and new users. This study is also motivated by the power of graph theory in data representation. The main salient features of the model proposed in this study are listed below.

- The (activity) history of opinion contributors and their social network interactions instinctively generate a social context for opinions (e.g., reviews and votes) given the limited direct interactions among numerous consumers in the social commerce context [32]. All of the data representation methods used in existing trust and reputation systems are application-specific and therefore limit the data inputs and representations that can be employed [19,43]. This study models the social context of opinions as users' activities and social relationship networks (UASRNs in the present paper) through a graph-based data representation model. The proposed reputation model fully exploits the social context of opinions and incorporates both the behavior and social relationship reputations of a user to combat the scarcity of first-hand information.
- Furthermore, raters can be malevolent and providers can behave dynamically or inconsistently toward different consumers, which results in a biased reputation estimating. In addition, a set of trust factors are identified by deep-mining the context of user opinions. These factors are used to refine reputation assessment and to mitigate the subjectivity of opinions and dynamics of behaviors.
- Trust and perceived risk are essential constructs of transaction intention in times of uncertainty [38]. Thus, the reputation model is enhanced and personalized by leveraging the limited risk tolerance of rational consumers toward negative outcomes of transaction behavior to improve the accuracy and robustness of trust decision-making in deceptive and collusive environments. On the one hand, the enhanced model develops a novel deception filtering approach to discard bad-mouthing opinions; on the other hand, the personalized reputation model applies a personalized risk metric to detect malicious providers.

The remainder of this paper is organized as follows: Section 2 provides a detailed literature survey of the existing reputation mechanisms. Section 3 describes the graph-based CR model. Section 4 details the enhancement and personalization of the model. Section 5 experimentally evaluates the performance of the CR model. Section 6 summarizes the key contributions of this study.

## 2. Background and related work

The importance of trust is elevated in e-commerce because of the high degree of uncertainty and risk present in most online transactions [40]. Therefore, reputation-based trust mechanisms have been widely used in various domains. Many reputation systems are constantly under attack and must address strategic problems upon detection[18]. Researchers have identified several classes of typical attacks [18]: (i) Self-promoting, in which attackers collude with a group of users to give them unfairly high ratings and boost their personal reputations. (ii) Slandering (bad mouthing), wherein attackers collude with a group of users to provide unfairly low ratings to their competitors to ruin the reputations of these competitors. (iii) Whitewashing, in which attackers abuse the system for short-term gain by allowing their reputations to degrade quickly and then re-entering the system with a new identity. (iv) Orchestrating, wherein attackers orchestrate their efforts and employ several of the aforementioned strategies. In this type of attack, colluders balance the maximization of selfish or malicious behavior and the avoidance of detection. In the face of malevolent opinion contributors and product providers who behave dynamically or inconsistently toward different consumers, the feedback dispersion caused by taste differences is difficult to distinguish from that induced by other factors (including unfair ratings by raters and the discriminatory and dynamic behaviors of providers) in terms of feedback reports. In the design of reputation mechanisms, an important and relevant challenge is ensuring that honest feedback is obtained on the actual interactions. To reduce the effect of opinions that are judged to be unreliable on estimated reputation, they are either adjusted or ignored the opinions [44].

Numerous studies have aimed to identify and detect deceptive information(unfair ratings). Some methods obtain unreliable reputation information by considering only the statistical properties of the reported opinions [8,11,51]. For example, Dellarocas [11] proposed mechanisms to detect and filter out the feedbacks in certain scenarios using cluster-filtering techniques. The technique can be applied to feedback-based reputation systems for the detection of suspiciously unfair ratings prior to actual aggregation. Chen and Singh [8] and Whitby et al.[51] assumed that inaccurate or unfair raters are generally a minority among reputation sources. Hence, they consider the opinion contributors whose opinions deviate in some way from the mainstream to be most likely inaccurate. Adversaries who submit dishonest feedback can still build good reputations as raters simply by submitting much feedback and by becoming the opinion leader.

Some methods detect unreliable feedback information based on other information, such as the reputation of the source or its relationship with the trustee [44,45,40]. For instance, Teacy et al.[44] proposed the Trust and Reputation model for Agent-based Virtual OrganisationS (TRAVOS),which computes trust using beta probability density functions. This model is similar to the (modified) Bayesian Reputation System (BRS) developed by Whitby et al. These models mainly differ in terms of how they filter out unfair opinions; TRAVOS identifies unreliable reputation information using only personal observations. Although it does not assume that the majority of ratings is fair, it may fail when the personal observations of consumers are inadequate. Moreover, it is vulnerable to the subjectivity of individual user opinions. An integrated approach for context-aware service selection in deceptive environments ( POYRAZ) [40] integrates the advantages of both BRS and TRAVOS in deceptive and subjective environments. It filters out deceptive experiences by combining two different sources of information, namely, the private and public credits of the reviewer, to generate a superior result. Teacy et al.[45] also applied a Hierarchical and Bayesian Inferred Trust model recently to assess the extent to which an agent should trust its peers based on direct and third-party information.

The emerging channel of social media not only permits users to express their opinions, but also enables them to build various explicit or implicit social relationships. Therefore, social network-based methods [16] can improve the performance of reputation mechanisms based solely on the information from limited personal observations or from unknown users, as mentioned previously. Such methods extract trust values from social links among users. Some proposals [4,19,20] capitalize on social knowledge to propagate the sourcing of reputation from trusted sources along the edges of a directed graph, thereby producing a "web of trust". These approaches limit the effect of malicious nodes by reducing the number of unknown and potentially malicious nodes to which honest nodes extend trust at the expense of requiring social interaction. However, malevolent users impinge within a chain of social connections in the context of social commerce [1,54]. Therefore, Caverlee et al.[7] presented a SocialTrust framework to aggregate trust in online social networks. SocialTrust distinguishes the relationship quality of the user based on trust and tracks his/her behavior; however, it does not consider the context of the transaction. As a result, it is unsuitable for social e-commerce, in which social relationships may not be adequate or available at any given time. Nonetheless, the web of trust derived from social links may not always be available as well and is typically sparse in social commerce applications, especially for new users [57], because the (intentional or involuntary) provision of personal information to social media risks the loss of privacy. If the business intrudes into the customer information space, then social network-based methods can fail. The preservation of privacy while using social networks is a significant issue. Thus, a fair balance must be struck between identity protection and accountability [25].

Some other proposals also seek to address the issue of incentive-based feedback quality or credibility. They assume that users are rational and build a reputation system according to game theory. Jurca and Faltings [24] established an incentive-compatible mechanism by introducing payment for reputation-building. Specifically, agents are only paid for reputation provision if the subsequent agent reports a similar result. This approach is interesting because it contains the heuristic to address dishonest opinions. However, it disregards collusion, the nodes of which can influence the next report. The Pinocchio approach [13] therefore encourages the exchange of reputation information and rewards participants that advertise their experience to others. It uses a probabilistic honesty metric to detect dishonest users and to deprive them of these rewards. However, the Pinocchio approach does not intend to protect against conspiracies or bad-mouthing. Kerr and Cohen [27]consider trust to be a tradable commodity that provides an incentive for honesty. Their model demonstrates that rational sellers choose to be honest given the commodity *trunits* because it is a profit maximization strategy. However, *trunits* neither regulates nor predicts buyer behavior. Thus, the model operates according to the honesty of buyer feedback and adopts a parallel mechanism to elicit honest responses.

Therefore, this study aims to build a reputation mechanism to improve the veracity and objectivity of opinions and to enhancing consumer trust in (product and opinion) providers even in the presence of malicious and new users. Trust has long been regarded as a catalyst in consumer—marketer relationships because it generates expectations of successful transactions [38]. McKnight et al.[36] explained that initial trust is formed among individuals in an unfamiliar entity on the basis of framework of the Theory of Reasoned Action (TRA) [14]. This theory postulates that belief influences attitude, which in turn shapes behavioral intention. When individuals form a positive attitude toward a given transaction behavior, their intention to engage in such behavior strengthens. Trust induces positive attitudes toward transactions with provider when it is viewed as a salient behavioral belief, thus reducing uncertainty and generating expectations for satisfactory transactions. As a result, the

behavioral intentions of a consumer toward transaction are positively influenced [38]. TRA has been successfully applied to the study of consumer behavior, technology acceptance and system use, and various instances of human behavior. Research that follows TRA consistently reports high correlations between intentions and actual use [22,26].

Researchers consistently focus on customer repurchase behavior because the proportion of second or n-time purchases is larger than that of initial acceptances. Marketing studies have determined that the primary reason for a consumer to repurchase products or to patronize services lies in his/her level of satisfaction. Much of the literature in satisfaction survey research indicate that repurchase intention is positively correlated with high overall customer satisfaction [28,58]. To clarify how customer satisfaction can affect customer retention and loyalty, the Expectation Confirmation Model (ECM) explains the customer consumption decision in the post-purchase process and constantly dominates academic research and managerial practice [26]. According to this model, customer satisfaction with past outcomes positively influences future behavior.

Our current work is inspired by the concepts of the aforementioned works, such as modified BRS, SocialTrust, TRA, and ECM. Thus, it develops a graph-based CR model by exploiting the social context of opinions on the basis of consumer behavior and psychological theory. This study differs from the previous studies in the following ways: First, the constructed data model (UASRNs) emphasizes the links and interactions among users and their values and context, unlike both the graph-based reputation[19,43] and behavior-specific, knowledge-based trust models [48]. Furthermore, we combine multiple reputation components (trust sources) based on the constructed data model to combat the scarcity of first-hand information. We then identify a set of trust confidence factors by mining the context of opinions based on UASRNs to mitigate the subjectivity of these view points and the dynamics of behaviors. Finally, we enhance and personalize the proposed model for accurate and robust trust decision-making under deceptive and collusive environments. The enhanced and personalized model leverages the limited risk tolerance of consumers for negative outcomes of transaction behavior to exclude “bad-mouthing” opinions prior to ratings aggregation and to detect malicious providers. Therefore, the reputation model proposed in this study is expected to generate a comprehensive perspective of the honesty, expertise, and influence levels of users. It is also predicted to contribute to accurate and tamper-resistant reputation establishment in social commerce environments.

## 3. Graph-based CR model

Numerous providers can provide the same or a similar type of product or service in social commerce scenarios but at different qualities and price levels. For example, when a potential buyer Alice has a new transaction demand and only a few or no direct previous interactions with providers $u_j$ ($j$=1,2,…), she must collect information regarding the providers from other consumers. This information is used to compute the expected behavior of provider $u_j$ in relation to her current transaction demand. First, Alice can learn about related experiences from her friends and/or other consumers $u_i$ ($i$=1,2,…). As previously described, Alice can face three challenges:

- *The veracity of opinions*. Some malicious (dishonest) consumers may wish to either defame or promote providers by reporting and propagating deceptive experiences [18,40]. As a result, they may mislead Alice if she cannot differentiate these experiences from truthful ones.
- *The subjectivity of opinions*. Consumers can evaluate their interactions with the same providers according to different satisfaction criteria even though most of them share their experiences

honestly, especially Alice's friends (if applicable). Therefore, recommendations regarding the expertise level, experience level, and preference of opinion contributors may be biased under subjective environments.

- *The dynamics of behaviors*. Furthermore, the dynamic or inconsistent behaviors of providers toward various consumers worsen the aforementioned situation. Thus, distinguishing among honest, deceptive, and biased experiences is difficult for Alice.

Therefore, the reputation model proposed in this study must reflect the honesty, expertise, and influence level of the user in the target domain. Moreover, a mechanism must be developed to filter out deceptive opinions.

In social commerce scenarios that report only a small fraction of interactions between many consumers, Alice may instinctively search for additional contextual information about $u_i$ ($i$=1,2,…) given limited first-hand data. The roles of a registered user $u_i$ vary based on the nature of activities in a social commerce context, such as buying a product or service, reviewing his/her transaction, and advising (voting) on the reviews of other consumers (Fig. 1). $u_i$ can also subscribe to a social network composed of members with similar interests through following or a followed list. User $u_i$ can add people to his/her friend list by mutual agreement. Furthermore, he/she can quit the social network and block people from his/her own social network. User $u_i$ presumably possesses a unique and stable ID. In real life, methods or policies (e.g., a social network [55,56] and a meta-identity system [42]) must be employed to limit the Sybil attacks and whitewashing by malicious users [18]. The (activity) history of $u_i$ and his/her social network interactions generate a social context for his/her opinions. Consequently, Alice may first gather the available transaction information of $u_i$ with provider $u_j$ (including transaction price, amount, and time) and the other transaction histories of $u_i$. Alice may then collect voting information from other consumers (e.g., $u_k$) regarding the review posted by $u_i$. She may also determine the transaction history, expertise level, and social position of $u_k$. Finally, she may also obtain data regarding the social relationships (friends and followers) of $u_i$. She may then make a trust decision based on the performance of $u_i$ in related activities and on the status of his/her social relationships.

Accordingly, the method proposed in this study first models the social context of an opinion in the form of the UASRN within the context of social commerce with given characteristics. All UASRNs constitute a complex and heterogeneous network. Thus, the proposed method combines the behavior and social reputations of users to produce a graph-based CR model based on the constructed UASRNs and is enabled by the power of graph theory in data representation. The notions used in this study are listed in the Appendix below.

### *3.1 Preliminary definition*

The UASRN of a user represents other related users and captures the activities and social relationships between both parties. UASRNs track all of the activities conducted by each user on others and all of the relationships among the users. We introduce the following preliminary definitions.

**Definition 1.** (*UASRNs*):UASRNs denote a quintuple composed of users $U$, activities B, relationships $\Re$, context $C$, and values $V$. UASRNs=<$U$, B, $\Re$, $C$, $V$>, where

- $U=\{u_1,u_2,\ldots,u_n\}$ is a set of users, and $u_i$=<uid, profile>, where the user identifier (uid) is unique and the user profile is associated with a set of tags;
- $C=\{c_1,c_2,\ldots,c_n\}$ is a set of contexts that are relative to specific activities and social relationships.

The context is defined as any piece of information that can characterize these activities and social relationships. In this study, context mostly represents the content and time of specific activities and social relationships, i.e., $c_k$= <content, time>, $k$=1, 2,...,$n$;

• $V=\{v_1,v_2,\ldots,v_n\}$ is a set of values associated with activities and relationships, such as transaction ratings, vote ratings, and the number of interactions;

• $B=\{beh_x(u_i,u_j,c_k,v_z)|u_i\in U,\ beh_x\in BEH,c_k\in C,v_z\in V\}$ is a set of all of the activities that have been performed; $beh_x(u_i,u_j,c_k,v_z)$ is an activity between $u_i$ and $u_j$ in activity set B and characterizes the contextual view of this activity for each participant; and $BEH=\{\ldots,beh_x,\ldots\}$is a set of potential activity types for a user to engage in ($x$=1, 2,...);

• $\Re=\{rel_y(u_i,u_j,c_k,v_z)|u_i,u_j\in U,rel_y\in REL,c_k\in C,v_z\in V\}$ is a set of all of the social relationships to which users have subscribed; $rel_y(u_i,u_j,c_k,v_z)$ is the social relationship between $u_i$ and $u_j$ in the social relationship set $\Re$, and characterizes the contextual view of the social relationship of each participant; $REL=\{\ldots,rel_y,\ldots\}$is a set of potential social relationship types to which a user can subscribe ($y$=1,2,…).

In this definition, users are represented by nodes. The corresponding activities and social relationships are denoted by directed edges, which are valued as the weights of the corresponding edges in the heterogeneous graph. A UASRNs ample is plotted in Fig. 1.

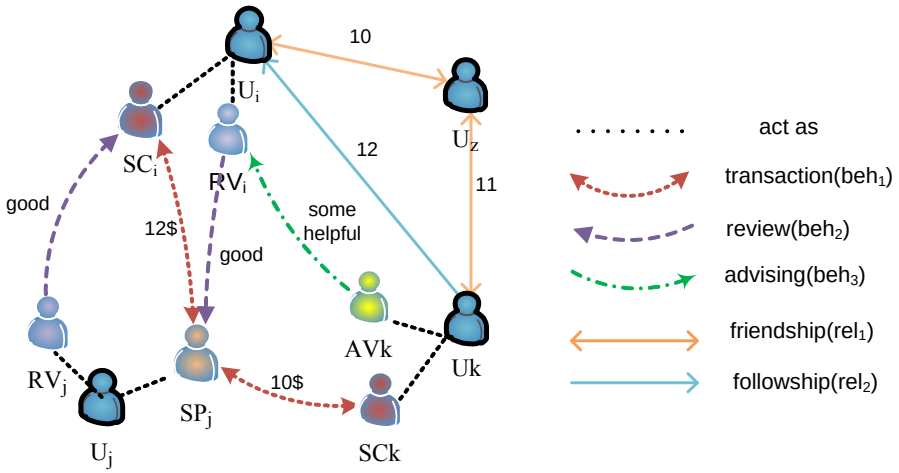


**Fig.1.** Schematic of a simple UASRNs ample

The reputation values of each behavior type $beh_x\in BEH$ ($x$=1, 2,...) displayed by user $u_i$ ($i$=1,2,…) in the UASRNs are derived from the performance of $u_i$ in corresponding activities. Similarly, the reputation values of each social relationship type $rel_y\in REL$ ($y$=1, 2,...) are obtained from the performance of the corresponding relationship circle of $u_i$ along the edge of the UASRNs. Both values contribute to a portion of the comprehensive reputation score. The comprehensive reputation score is ultimately aggregated by suitable operators and arguments.

**Definition 2.** (*Comprehensive reputation score R*): $R\in[-1,1]$ represents a public opinion on the ability and willingness of a user to perform a particular action. This score reflects the honesty, expertise, and influence levels of a user in the target domain and is measured by incorporating both the behavior and social relationship reputations of the user. That is, the score incorporates the performance of a user in different activities and that of related members.

Feedback data are an essential resource for the evaluation of the corresponding behavioral qualities of the users. Two kinds of explicit feedback are provided within the social commerce scenarios of the

given characteristics as shown in Fig.1.

- *Transaction ratings*. The review activity ($beh_2$) of users $u_i$ and $u_j$, who act as reviewers ($rv_i$ and $rv_j$) and generates feedback ratings (e.g., "good") on transaction activity ($beh_1$), which is referred to as transaction ratings.
- *Vote ratings*. The advising(voting) activity ($beh_3$) of user $u_k$, who acts as either an advisor or a voter ($av_k$) and gives feedback ratings (e.g., "some helpful") on the review activity ($beh_2$) of $u_i$, which is labelled as vote ratings.

Users usually do not feedback advising(voting) activity ($beh_3$) explicitly. In this study, we therefore consider only the activities with explicit feedback in the estimation of user performance with regard to these activities although implicit user feedback can be leveraged to generate ratings on advising activities. Textual feedback (e.g., comments) must be mapped to the corresponding numerical ratings by using opinion mining and fuzzy relationship to facilitate computer processing.

The *behavior reputation* $R_{beh_x}(i)$ of the CR of user $u_i$ consists of his/her transaction and review activity reputations. The transaction reputation of $u_i$ as a consumer represents his/her transaction experience/history and reveals the veracity of his/her opinions to some extent, whereas the transaction reputation of provider $u_j$ corresponds to his/her performance in transaction activities. Similarly, the review reputation of $u_i$ reflects the quality of his/her posted consumer reviews, as well as his/her active, expertise, and acceptable levels to some extent. *The social relationship reputation* $R_{rel_y}(i)$ of $u_i$ consists of his/her friend and follow relationship reputations. This friend relationship reputation is derived from the established reliability of his/her friends(e.g., $u_z$) and reflects the degree of confidence other users have in his/her reviews. The follow relationship reputation of $u_i$ is derived from the established reliability of his/her followers (e.g., $u_k$) and establishes the extent of user influence.

Accordingly, the corresponding CR score of user $u_i$ in time $n$ is defined as follows

$$R^n(i) = \partial\left(\sum_{x=1}^{2}(\varpi_x \times R^n_{beh_x}(i)) + \sum_{y=1}^{2}(\omega_y \times R^n_{rel_y}(i))\right) \quad , \sum(\varpi_x + \omega_y) = 1 \tag{1}$$

where $R^n_{beh_x}(i)$ ($x$=1,2) is the behavior reputation score of the $beh_x$ activity of user $u_i$ in time $n$; $R^n_{rel_y}(i)$ ($y$=1,2) is the social relationship reputation score of the $rel_y$ relationship of user $u_i$ in time $n$; $\partial$ is the scale argument used to adapt to different rating scales; and $\varpi_x$ and $\omega_y$ are the weights of the behavior and social relationship reputation values, respectively. These weights can be determined based on the preference of potential buyer Alice or on the policies of the reputation system.

*3.2 Assessment of user behavior reputation*

The assessment of the behavior reputation of a user should incorporate features that reward long-term good behavior and that penalize users who dynamically change their behavior for their strategic interests (e.g., whitewashing and its variants) [18]. Thus, we model the effect of the historical and recent behavior of user $u_i$ in relation to his/her behavior reputation using two factors.

***Maturity factor*** $A\mu$: Sharing the historical behavior reputation of a user is critical in encouraging other users in the network to behave well at all times and in limiting the ability of exploitative users to whitewash their reputation ratings by leaving and re-entering the network repeatedly. The maturity

factor is expected to provide an incentive for a community user to emphasize behavior history and to maintain a stable identity. A large number of effective activities performed by $u_i$ may enhance her/his maturity. With the factor, a new user will have to behave honestly for an initial period of time to build up a positive reputation before starting the self-serving attack.

$$A\mu_x^n(i) = \kappa \left| beh_x^n(i) \right| \Big/ \left( \left| beh_x^n(i) \right| + 1 \right) \tag{2}$$

where $beh_x^n(i)$ is a set of the $beh_x$ activities that have been performed by user $u_i$ until time $n$. The real number $\kappa$ in the interval (0, 1) is the learning rate.

***Behavior shift adaptation factor*** *Af*: Exploitative users may build up a good initial reputation rating in a reputation-based trust-enabled transaction community. However, such users suddenly display harmful behavior or a change in their qualities either intermittently or randomly. This occurrence can affect the system significantly because the malicious user continues to nurse a good reputation for a substantial period of time, during which the system is slow to identify this negative behavior and is incapable of sufficiently lowering the reputation of the malicious user [18]. The proposed model adopts the behavior shift adaptation factor, which scales the changes in recent user behavior to detect sudden shifts in behavior. This factor takes the ratings $\in$ [-1,1] of honest users as its argument and returns as its output [0, 1], when the rating is less than 0 indicates that it is a negative behaviors and needs to be enlarged to detect and vice versa. With the factor, the system can sufficiently lower the reputation of the malicious user, and the malicious behavior of an old user can be quickly identified.

$$Af_x(i) = \frac{1}{\left| beh_x(i) \right|} \begin{cases} \sum_{l \in beh_x(i)} r_{l \to i} \times r_{l \to i} & \text{if } r \in [0,1] \\ \sum_{l \in beh_x(i)} \sqrt{-r_{l \to i}} & \text{if } r \in [-1,0) \end{cases} \tag{3}$$

where $beh_x(i)$ represents a set of $beh_x$ (type) activities performed by user $u_i$ in a time slot; and $r_{l \to i}$ is the rating of $u_l$ on user $u_i$ in $beh_x$ (type) activities.

To reflect the dynamic changes in the behaviors and social relationships of a user, timeline $t$ is divided into timeslots. If a user provides more than one feedback on the same user in a timeslot, his/her final feedback is adopted to avoid the "flooding" of the system by users.

Let $R_{beh_x}^n(i)$ denote the behavior reputation score of the $beh_x$ activity of $u_i$ in time $n$. $R_{beh_x}^0(i)$ represents the initial reputation value of $u_i$ in $beh_x$ type activity in a cold-start state ($n$=0). A new user can build different initial behavior reputations based on the trust policies in the given application domains. When $n$=1,2,3,...$\infty$, the assessment of user behavior reputation maintains this reputation through the iterative revision of reputation score according to the historical behavior reputation in time $n$-1 and the recent behavior reputation in the $n^{\text{th}}$ timeslot.

Let $R_{beh_x}^{n-1}(i)$ denote the historical behavior reputation of $u_i$ in time $n$-1. $\hat{R}_{beh_x}^n(i)$ corresponds to the behavior reputation of $u_i$ in the $n^{\text{th}}$ timeslot ($n$=1,2,...) and $\hat{Af}_x^n(i)$ indicates the behavior shift adaptation factor of $u_i$ in the $n^{\text{th}}$ timeslot. The assessment of behavior reputation can optimize the historical reputation by balancing the weights of the historical and recent reputations through the

maturity $A\mu_x^n(i)$ of $u_i$.

$$R_{beh_x}^n(i) = (1 - A\mu_x^n(i)) \times R_{beh_x}^{n-1}(i) + A\mu_x^n(i) \times A\hat{f}_x^n(i) \times \hat{R}_{beh_x}^n(i) \qquad n = 1,2,... \Big| x = 1,2 \tag{4}$$

Once the $\hat{R}_{beh_x}^n(i)$ values in all timeslots are known, the behavior reputation score of the $beh_x$ activity of $u_i$ in time $n$ can be calculated. In the following section, we assess the behavior reputation $\hat{R}_{beh_x}^n(i)$ ($x$=1,2) of user $u_i$ in the $n^{\text{th}}$ timeslot.

*3.2.1 The transaction reputation component and its influential factors*

The computation of transaction reputation should be bound to the context of reviews (opinions) that affect the contribution of transaction ratings to the transaction reputation of user $u_i$ ($i$=1,2,…). The transaction reputation considers the critical factors related to the price and amount of transactions to measure either the honesty or influence level of a transaction rating.

*The effect of transaction amount Wa*: In some cases, a malicious provider may build a good reputation initially by selling good-quality and low-cost services or products. However, they will then deceive consumers by providing poor-quality and expensive services or products before disappearing [49]. This scenario may depict a typical attack on a reputation-based trust-enabled transaction community. Thus, reputation fraud may occur if transaction reputation assessment does not consider the amount of transactions performed. The effect of the transaction amount defines the degree of its contribution to transaction reputation. A high transaction amount between $u_i$ and provider $u_j$ may enhance its influence on their transaction reputations.

$$Wa(i, j) = \frac{Tr(a)}{D} \tag{5}$$

where $Tr(a)$ denotes the transaction amount of the transaction activity between $u_i$ and $u_j$ in time $t$; flag $a$ indicates that the return value is the transaction amount; and $D$ is related to the maximum transaction amount of a transaction or to the upper bound of the transaction amount.

*The expected utility of transaction price Wp*: Potential buyers do not always select the highest-quality product or service. Other things being equal, the purchasing behavior of a potential buyer is influenced by his/her justification of the cost and value (whether the offer is fair, appropriate, or deserved)[28]. An increase in transaction price and promised service quality may enhance perceived service quality expectations based on expected utility theory [10]. Moreover, a significant difference between the transaction price and the average market price can heighten the possibility that the provider is fraudulent. The expected utility of transaction price models the partial influence of transaction and average marketprices on the transaction reputations of $u_i$ and $u_j$.

$$Wp(i, j) = \exp\left(-\left|\frac{(Tr(p) - Tr(p)^*)}{Tr(p)^*}\right|\right) \tag{6}$$

where $Tr(p)$ is the transaction price and $Tr(p)^*$ is the average market price of a product or service (category) in the transaction activity between $u_i$ and $u_j$ in time $t$. Flag $p$ suggests that the return value is the transaction price.

The harmonic mean of the two crucial factors is used to evaluate the contribution of transaction ratings to the transaction reputations of consumer $u_i$ and provider $u_j$. Let $SP(i)$ represent a set of

providers who have transacted with $u_i$ in the given $n^{th}$ timeslot and let $r_{j\to i}$ be the transaction rating of provider $u_j$ on $u_i$ in this timeslot. The transaction reputation of $u_i$ as a consumer in the $n^{th}$ timeslot is defined as follows

$$\hat{R}^n_{beh_1}(i) = \frac{1}{|Sp(i)|}\sum_{j\in SP(i)} r_{j\to i}\frac{2Wp(i,j)Wa(i,j)}{Wp(i,j)+Wa(i,j)} \tag{7}$$

Let $SC(j)$ represent a set of consumers who have transacted with provider $u_j$ in the $n^{th}$ timeslot and let $r_{i\to j}$ be the transaction rating of consumer $u_i$ on $u_j$ in this timeslot. Let $R^n(i)>0$ be the current CR score of $u_i$ (in time $n$).The transaction reputation of provide*r* $u_j$ in the given $n^{th}$ timeslot is expressed as follows

$$\hat{R}^n_{beh_1}(j) = \frac{1}{\sum R^n(i)}\sum_{i\in SC(j)} r_{i\to j}\frac{2Wp(i,j)Wa(i,j)}{Wp(i,j)+Wa(i,j)}R^n(i) \tag{8}$$

*3.2.2 The review reputation component and its influential factors*

The calculation of review reputation should also consider the context of voting (opinions) that affects the contribution of vote ratings to the review reputation of user $u_i$ ($i$=1,2,…). The review reputation measures the level of either the expertise or influence of vote ratings from advisors $u_k$ ($k$=1,2,…) by highlighting two important factors related to the price difference between two transactions and to the recognition degree of an advisor (voter).

*The feedback utility of transaction price Wρ′*: The expected utility of the aforementioned transaction price can also result in a cascade of user reactions in advising (voting) activities. This process is known as the feedback utility of transaction price. That is, the difference in transaction price between reviewer $u_i$ and advisor $u_k$ with respect to the same provider partly influences the vote rating of $u_k$. The feedback utility of transaction price models the influence of this variation on voting. A considerable difference in transaction price may increase the bias of advisor $u_k$ during the provision of voting ratings on the review posted by $u_i$.

$$W\rho'(i,k) = \exp(-|\Delta|) \tag{9}$$

where $\Delta$ is the difference in transaction price between $u_i$ and $u_k$ given the same provider $u_j$ in time $t$. When $\Delta$ approaches 0, $W\rho'$ should be approximately equal to 1. When $\Delta$ is greater than or less than 0, $W\rho'$ should be less than 1.

*The recognition degree of a user Wr*: In social commerce scenarios, users with similar interests and tastes form a clique through lists of friends, followers, and followees. Users with large groups of either followers or friends are usually the opinion leaders in the clique. The recognition degree of user $u_k$ denotes his/her social position, popularity, and level of influence in the target domain.

$$Wr(k) = \frac{|rel_1(k)|+|rel_2(k)|}{2N} \tag{10}$$

where $|rel_1(k)|$ is the number of friends of user $u_k$; $|rel_2(k)|$ is the number of followers; and $N$ is the average number of active community users.

The harmonic mean of the two crucial factors is used to evaluate the contribution of the vote ratings of $u_k$ to the review reputation of $u_i$. The review reputation of $u_i$ as a reviewer in the given $n^{th}$ timeslot is defined as follows

$$\hat{R}^{n}_{beh_2}(i)=\frac{1}{\sum R^{n}(k)}\begin{cases}\sum\limits_{k\in AV(i)} r_{k\to i}\dfrac{2Wr(k)W\rho'(i,k)}{Wr(k)+W\rho'(i,k)}R^{n}(k) & \text{if } Wr(k)>0\\ \sum\limits_{k\in AV(i)} r_{k\to i}\dfrac{W\rho'(i,k)}{2}R^{n}(k) & \text{else}\end{cases} \tag{11}$$

where $AV(i)$ represents a set of advisors who have rated $u_i$ in the $n^{\text{th}}$ timeslot; $r_{k\to i}$ is the voting rating of $u_k$ regarding the review activity between $u_i$ and $u_j$ in this timeslot; and $R^n(k)$ (>0) is the current CR of $u_k$ (in time $n$).

*3.3 Assessment of social relationship reputation*

As the saying goes, "Birds of a feather flock together." Therefore, the reputation of the social relationship of user $u_i$(e.g., friends and followers) provides the basis for trust inference in the social commerce context, especially given limited first-hand interactive information. However, some followers may be either "zombie fans" or fake followers. Furthermore, some of these friends and (or) the followers may have established a mutual benefit collusion. Thus, the quality of the social relationship of $u_i$ must be considered in the computation of social relationship reputation calculation.

Accordingly, let $rel^{n}_{y}(i)$ represent a set of social relationship members who have $rel_y$ (type) relationships with $u_i$ in time $n$. The social relationship reputation $R^{n}_{rel_y}(i)$ of the $rel_y$ (type) relationship of $u_i$ in time $n$ should be determined by the following components.

- ***The number*** $\left|rel^{n}_{y}(i)\right|$ of members who have $rel_y$ relationships with user $u_i$ in time $n$.

- ***The comprehensive reputation*** $R^n(z)$ of the relationship members $u_z$ ($z\in rel^{n}_{y}(i)$) of $u_i$ in time $n$.

- ***The intimacy level*** $Rp^n$ of these members with $u_i$ in time $n$. Intimacy level is mostly determined by the number of interactions $q$ and hops $m$ between the two sides ($u_i$ and $u_z$).

$$Rp^{n}(i,z)=\exp(-q^{-1})\big/(m+1) \tag{12}$$

In this scenario, a relationship edge expressed by a highly reputable/highly intimate user counts more than that expressed by a highly reputable/low-intimacy level user.

- ***The relationship quality*** $Rq^{n}(z)^{m}$ of each of the members in time $n$. The world of social networks is small. Thus, the social relationship quality of $u_z$ should be determined primarily based on his/her direct social relationships and marginally based on his/her indirect social relationships (up to a small number ($m$>1) of hops away). The quality of the $rel_y$ (type) relationship is related to the overall feedback ratings on the $rel_y$ relationship members of $u_z$ and their number.

$$Rq^{n}(z)^{m}=\begin{cases}F(z) & m=0\\ Rq^{n}(z)^{m-1}\sum_{l\in rel^{[m]}_{y}(z)}F(l)\Big/\left|rel^{[m]}_{y}(z)\right| & m=1,2,...\propto\end{cases} \tag{13}$$

where $rel^{[m]}_{y}(z)$ ($m$=0, 1, 2,…) represents the $rel_y$ relationship circle of $u_z$ at a distance of $m$ hops.

The relationship quality of user $u_z$ merely refers to the overall feedback rating received in the base

case ($m$=0). $F(z)=\frac{1}{\sum R^n(k)}\sum_{k\in AV(z)} R^n(k)\left|r_{k\to z}^{+}\right| \Big/ \left(\left|r_{k\to z}^{+}\right|+\left|r_{k\to z}^{-}\right|\right)$, where $R^n(k)>0$ and $AV(z)$ represents a set of advisors who have rated $u_z$. In addition, $|r_{k\to z}^{+}|$ and $|r_{k\to z}^{-}|$represent the amounts of positive and negative ratings, respectively. The relationship quality of a user is beneficial in that it provides an incentive for users to monitor the quality of their social relationships and isolate misbehaving users. A rational user retains social relationships with well-behaving users.

The assessment of social relationship reputation by a user maintains social relationship reputation through the dynamic revision of reputation score according to the number, relationship quality, CR, and intimacy level of relationship members. Accordingly, the social relationship reputation score of user $u_i$ in time $n$ is defined as follows

$$R_{rel_y}^n(i)=\frac{1}{\left|rel_y^n(i)\right|}\sum_{z\in rel_y^n(i)}\frac{2Rq^n(z)^m Rp^n(i,z)}{Rq^n(z)^m+Rp^n(i,z)}R^n(z)\Bigg| \quad n=1,2,...,y=1,2 \tag{14}$$

$R_{rel_y}^0(i)$ denotes the initial reputation value of the $rel_y$ (type) social relationship of user $u_i$ in a cold-start state ($n$= 0). The initial social relationship reputation of new users without any social relationships is set to a default value.

## 4. Enhancement and personalization of the CR model

Trust and perceived risk are essential constructs of transaction intention when uncertainty is present [38]. The CR model is enhanced according to TRA and ECM, and it is personalized by leveraging the limited risk tolerance of the rational consumer towards negative outcomes of transaction behavior.

### *4.1 Deception filter for "bad-mouthing" ratings*

As mentioned above, $u_i$ can be a negative (dishonest) consumer who wishes to either defame or promote providers by releasing and propagating deceptive opinions [18,40] that may mislead potential buyer Alice. In addition, slandering (“bad-mouthing”) attacks are simplified although the behavior shift adaptation factor *Af* (Section 3.2) facilitates quick detection when user behavior worsens to abuse the system for personal short-term gains [18].Therefore, the reputation model should at least be capable of filtering out deceptive ratings from “bad-mouthing” consumers.

Distinguishing honest feedback from “bad-mouthing” opinions or from the highly biased feedback is difficult from the perspective of feedback reports because opinion contributors can be malevolent or critical and because product providers can behave either dynamically or inconsistently towards various consumers. Thus, we propose a “bad-mouthing” rating filter enabled by TRA and ECM to address the limited first-hand information. The major building blocks of the TRA model are salient beliefs, which are used to ascertain attitudes and consequently, to determine intentions and behavior. Belief is the subjective probability of an individual suffering the consequence of a particular behavior. Satisfaction reveals the actual evaluation of the outcomes of previous behavior. Hence, the high proportion of dissatisfaction with the outcomes of previous behavior heightens the subjective probability of the negative consequences of future transaction behavior, i.e., high perceived risk. This risk in turn generates negative attitudes toward future transaction behavior and discourages transaction intention further. The repurchase intention of a customer is primarily determined by customer satisfaction, according to ECM. Therefore, the overall dissatisfaction of customers with previous transaction

behavior in relation to the given provider negatively affects their repurchase intention and behavior in line with the synthesis of TRA and ECM. This scenario is unlikely in a rational consumer who continually engages in transaction behavior with the same provider in a short period of time even if a high proportion of their transactions are extremely unsatisfactory. Thus, the number of extremely negative ratings (high levels of dissatisfaction) are rarely high. Consequently, the increased proportion of extremely negative user ratings regarding the same provider in the past corresponds to an increased fraud risk in the current rating of the user. In other words, an honest and rational $u_i$ does not always repeat transactions with a provider who usually provides poor service or product quality.

Changes in provider behavior can also generate extremely negative ratings from honest users. To limit the misjudgment, a certain proportion of extremely negative ratings must be retained.

**Definition 3.** (*Maximum risk tolerance MRT*):*MRT* is precisely defined as the maximum percentage of the extremely negative ratings of a user regarding the same provider before the extremely negative ratings of the user are excluded from the estimation of provider reputation.

As a result, a novel deceptive rating filtering approach can be developed from the viewpoint of the system to filter out suspicious "bad-mouthing" ratings that exceed *MRT* value prior to reputation assessment. Specifically, the transaction timeline of $u_i$ with $u_j$ is first divided into time windows. Let $|tw|^n$ represent the length of $n^{\text{th}}$ elemental time window. This length may be fixed or adapted by the frequency of the transactions of $u_i$ with $u_j$ [11]. We then calculate the local percentage of the extremely negative ratings in each time window and the global percentage in all time windows. Let *Ntr r*epresent the threshold of extremely negative ratings. Each user can have different satisfaction thresholds [5, 22] and rating scales. However, a consensus is often reached in instances of extreme dissatisfaction. For instance, $r \in [-1, -0.9]$ suggests that a user is highly unhappy with a person or an object given the rating scale [−1, 1]. Finally, $r_{i\to j}^t$ is filtered out (set to 0) given the current extremely negative rating $r_{i\to j}^t$ of user $u_i$ regarding user $u_j$ in time $t$ ($t \in n^{\text{th}}$ elemental time window, $n$=1,2,…) if either the local percentage of $r_{i\to j}^t$ in the current time window or the global percentage in all time windows is greater than the threshold *MRT* (obtained by dynamically learning the behavior of the majority of users).

$$r_{i\to j}^t = \begin{cases} 0 & \text{if } \left|r_{i\to j}^{-*}\right|^n \Big/ |tw|^n > MRT \text{ and } n == 1 \\ 0 & \text{else if } \left|r_{i\to j}^{-*}\right|^n \Big/ |tw|^n > MRT \text{ or } \left|r_{i\to j}^{-*}\right| \Big/ \left|r_{i\to j}\right| > MRT \text{ and } n > 1 \\ r_{i\to j}^t & \text{else} \end{cases} \tag{15}$$

where $|r_{i\to j}^{-}*|^n$ and $|r_{i\to j}|^n$ denote the numbers of extremely negative and total ratings of $u_i$ regarding a certain provider $u_j$ in the given $n^{\text{th}}$ time window, respectively. $|r_{i\to j}^{-}*|$ and $|r_{i\to j}|$ represent the number of extremely negative ratings of $u_i$ regarding $u_j$ in all time windows.

*4.2 Personalized trust decision-making with personalized CR*

Potential buyer Alice can assess the global trust (reputation) of an unknown provider according to the aggregated, community-wide (i.e., community consumers $u_i$, $i$=1,2,...) reputation values in the current CR model. Alice may have also had some direct experience with a low percentage of the small providers of all community members. Thus, we aim to balance the personal experiences of each user

with the larger view of the entire community through a personalized extension of the reputation model, which is labelled as the personalized CR (PCR) model. It customizes the adjustment of the quality component of reputation.

To generate this personalized view over global trust values, a natural approach involves combining the direct experiences of the user with these global ratings linearly. In e-commerce, negative experiences usually influence the trust decision-making of consumers significantly with respect to provider selection [47]. We estimate personalized direct distrust (risk) based on the previous negative interaction experiences of the user rather than predict the trust in future interactions directly given that both trust and perceived risk are essential constructs of transaction intention when uncertainty is present [38].

The tolerance of different users can vary in relation to the negative outcomes of transaction behavior and the ability to perceive risk because satisfactory standards differ from person to person.

**Definition 4.** (*Personalized maximum risk tolerance PMRT*): The risk tolerance of each consumer varies with respect to negative outcomes of transaction behavior in trust decision-making.

In the study conducted by Buchegger and Le Boudec [5], users can specify a tolerance for negative behaviors in p2p and ad-hoc environments. Users in social commerce cannot specify their *MRT* levels explicitly, but their behavior can reflect their trust decisions.

Risk is difficult to capture objectively; thus, the literature predominantly addresses the notion of perceived risk, which is defined as the subjective belief of the consumer in relation to suffering loss in pursuit of a desired outcome [38]. Each consumer evaluates its transactions with the providers through an internal risk function. This function considers the previous experiences of the consumer with the provider to be its argument and the returns to be its output [0, 1], where 0 indicates a low risk in future transactions and 1 represents a high risk. Let *PMRT* denote the *MRT* of Alice for negative transaction experiences. The risk $Risk^{t}_{Alice}(j) \in [0,1]$ perceived by Alice with respect to provider $u_j$ in time $t$ ($t \in n^{\text{th}}$ elemental time window, $n$=1,2,…) is

$$Risk^{t}_{Alice}(j) = \begin{cases} \left| r_{Alice \to j}^{-*} \right| \Big/ \left( \left| r_{Alice \to j}^{-*} \right| + 1 \right) & \text{if } \left| r_{Alice \to j}^{-*} \right|^n \Big/ |tw|^n < PMRT \text{ and } n == 1 \text{ or} \\ & \left| r_{Alice \to j}^{-*} \right|^n \Big/ |tw|^n < PMRT \text{ and } \left| r_{Alice \to j}^{-*} \right| \Big/ \left| r_{Alice \to j}^{-*} \right| < PMRT \\ 1 & \text{else} \end{cases} \quad (16)$$

where $|r_{Alice \to j}^{-*}|^n$, $|r_{Alice \to j}|^n$, $|r_{Alice \to j}^{-*}|$, $|r_{Alice \to j}|$, and $|tw|^n$ are as defined previously.

Accordingly, the PCR score of provider $u_j$ as estimated by Alice at time $t$ is

$$R^{t}_{Alice}(j) = R^{t}(j) - Risk^{t}_{Alice}(j) \quad (17)$$

We assume that the ability and willingness of a user to provide services or products are independent of his/her ability to provide reviews or advice. A consumer depends heavily on the transaction reputation of the provider in provider selection decision-making. Thus, Eq.16 is revised as $R^{t}_{Alice}(j) = R^{t}_{beh_1}(j) - Risk^{t}_{Alice}(j)$ .

## 5. Experimental results and discussion

We simulate a social commerce scenario and evaluate the validity of the CR and PCR models in various malicious settings given the separation of transaction and social relationship data and the difficulty of distinguishing between the ratings data derived from honest users and those obtained from malicious users in real life. The simulation platform was developed by Netlogo, which is a popular multi-agent simulation tool in the artificial intelligence community (http://ccl.northwestern.edu/netlogo/). In the simulations, the following special emphases are considered: (i) an evaluation of CR&PCR in the face of strategies that attempt to subvert its effectiveness, including deceptive raters and malicious providers; and (ii) a comparison between CR&PCR and alternative trust and reputation models with given algorithm-independent factors.

*5.1 Simulation environment and experimental setup*

To simulate a social commerce scenario that supports a cold-start system, the runtime environment is constructed based on the following settings:

*Simulation setup*: We generated 50 providers, 500 consumers (raters), and 2,000 round traces for the experiments. The simulation begins with a cold start by assigning a default reputation score for each user in the community. At the beginning of the simulations, we assume that the consumers do not have any prior experiences with providers. Thereafter, users (consumers) are randomly selected to initiate a purchase session with a provider. Realistically, a consumer may not use the same service or product in every round. The probability of a consumer requesting for a transaction is known as the activity level and is denoted by $\alpha \in [0.1, 1]$. We presume that each consumer (as an active user) keeps at least 10% transaction needs. The consumer then rates the transaction with the provider based on the service quality experienced or on his/her personal behavioral strategies. Other users can browse and vote for consumer reviews according to their own transaction experiences with the provider and behavioral strategies. Providers with no collusive intentions typically provide good feedback (1.0) regarding each transacted consumer for reciprocity (in our experimental framework, we assume that this occurrence is observed 9 out of 10 times).

The reputation system calculates the reputation score for each user in the community at regular intervals (each tick). This score is applied in the subsequent steps. Users can then opt to follow some of the reviewers they appreciate with a probability of $b$ in each round because the similar rating patterns (taste) of two consumers may induce a social relation between them [43]. If two users reach a mutual agreement in each round (the probability of the existence of such an agreement is $c$), the bidirectional follow is recorded in their respective friend lists. Each of them may also quit his/her friend and follow cliques (the probability of quitting a clique is $d$). We assume that probabilities $b$, $c$, and $d$ are related to activity level $\alpha$ and to user type.

Each user can engage in four activities (product provision, provider selection, review, and advice) and can play four different roles (provider, consumer, reviewer, and advisor. The latter two terms correspond to raters) in social commerce. Each user presumably gains a unique and stable ID through policy mechanism. We parameterized our simulation environment after considering some of the important real-life factors, namely, risk tolerance, price, subjectivity, and deception. We briefly explain our parameters in the section below in relation to the factors.

- *PMRT*. Each consumer displays different risk tolerances for negative transaction experiences in provider selection, as observed frequently in the real world. A probability is selected at random (*PMRT* is set to [0.2, 0.8] in the experiments). This parameter is introduced to mimic variations in the real-life risk tolerances of consumers for negative transaction experiences.

• *Price*. Transaction price is an important factor for a potential consumer in selecting and rating providers [28]. We assume that the varied service quality levels *Sl* correspond to different price levels *Pl* in a given transaction. Various types of users cultivate different price preferences and corresponding quality expectations. Therefore, we consider three types of prices: low, median, and high prices.

• *Subjectivity*. Consumers with similar demands may possess varied satisfaction criteria. Thus, two consumers may display different degrees of satisfaction (e.g., ratings) for the same demand and supplied service quality depending on their satisfaction criteria [40]; that is, consumer subjectivity. We define subjectivity as a parameter *sub* in the experiments. A user cannot be required to provide a ground truth rating ($r$) in an open network without a full global view of the system. We suppose that a subjective rating $r^*$ follows a normal distribution $N(r, sub)$.

• *Deception*. Both providers and consumers can be deceptive. A provider with the probability of dishonesty $PD \in [0, 1]$ will be dishonest $PD \times 100\%$ of the time and honest $(1 - PD) \times 100\%$ of the time. The poor service quality level delivered by dishonest providers is categorized into "no response" and "low grade", as shown in Table 1. The parameter *liar* is defined as the ratio of liars in the consumer society. Liars modify their reviews before sharing them so that they mislead other consumers the most. We use rating-based approaches for benchmark comparisons of the experiences. Formally, if the true rating of a liar regarding a provider is $r$, the liar modifies the rating as $N(-r, \sigma)$ before sharing this rating with the other consumers. This deception formulation is frequently applied in trust and reputation literature [44,51].

*Benchmark algorithms*: The experiments compare the proposed methods (CR&PCR) using four state-of-the-art trust and reputation algorithms: three typical trust and reputation algorithms with a filtering mechanism and one social relationship-based trust and reputation algorithm (SocialTrust) [7]. Specifically, the three typical algorithms applied are (modified) BRS [51], TRAVOS [44], and POYRAZ [40].

*Evaluation metrics*: The type of metrics used depends on the object to which trust and reputation are applied. The reputation values may be used by the entities in the community for decision-making purposes. Therefore, the success of a reputation system is measured by its accuracy in predicting the quality of future interactions based on the calculated reputations. We measure the main performance metric as the percentage of the satisfactory transaction experience reported by honest consumers. Precision is the mean of all of the step corrections from tick 1 to tick $n$ as follows:

$$\text{Precision} = \frac{1}{n}\sum_{i=1}^{n} \frac{\text{Number of good provisions recived by honest users}}{\text{Number of transactions attempted by honest users}}. \tag{18}$$

Various settings are applied in our experiments. For each setting, simulations are repeated a minimum of 10 times to increase reliability. We averaged the precision of many different approaches throughout the simulations, and their mean values are shown in the following figures. Nonetheless, the reported mean values may vary across samples. To determine the confidence intervals of the mean values, our simulation results are analyzed through a*t*test with a 95% confidence interval as suggested in [40]. At this probability, the reported mean values for the evaluation metrics deviate by approximately 2% in most cases. These values occasionally vary by almost 4% for the evaluation metrics related to CR&PCR. The mean values of BRS deviate by roughly 7%, whereas those of TRAVOS and SocialTrust differ by approximately 9%. These findings imply that our results are statistically significant and that our conclusions generally may not change significantly under different

simulation runs.

*Baseline*: We develop a normal configuration as the baseline to illustrate the simulation results of the experiments clearly. The baseline parameter configuration is listed in Table 1. Furthermore, the parameter values are used in all cases unless otherwise specified. We generated 50 providers, 500 consumers (500 raters), and 2,000 rounds. The percentage of *liar* ranges between 10% and 90%. The "typical" provider population may also differ across various applications. The space of possibilities is vast; thus, it cannot be explored completely. Given this drawback, we consider a typical provider population that consists of approximately 50% honest and 50% malicious providers, including intermittent providers. Nonetheless, the number of interactions $q$ is difficult to simulate because of the interactions associated with complex and subjective factors such as mood and personality. We set $q^{-1}$as the product of the activity level of interactive parties. The average number of active community users denotes a portion of community raters multiplied by a function of activity $a$.

**Table 1**

Major experimental variables

| Variable | Symbol | Value |
|---|---|---|
| Rating scale | $R$ | [−1,1] |
| Total number of providers | | 50 |
| Total number of raters | | 500 |
| Number of simulation rounds | | 2,000 |
| Range of user activity level | $a$ | [0.1, 1] |
| Average number of active users | $N$ | 1,000 × (1 − 0.2) |
| Number of interactions | $q$ | 1/(activity of $u_i$ × activity of $u_j$) |
| Hop | $m$ | 1 |
| (Delivered)poor service quality level | | {0 ("no response"), $Sl$ − 1("low grade")} |
| (Promised) service quality level | $Sl$ | {1, 2, 3} |
| Average market price | $Tr(P)$* | 20×$Sl$ |
| Maximum transaction amount | $D$ | 700 |
| Personalized maximum risk tolerance | $PMRT$ | [0.2, 0.8] |
| Probability of dishonesty | $PD$ | [0, 1.0] |
| The ratio of liars | $liar$ | [0, 1.0] |
| The subjective rating | $r$* | N($r$, $sub$) |
| Threshold of extremely negative ratings | $Ntr$ | [−1, −0.9) |
| Length of elemental time window | $tw$ | 10 |

*Parameter setting of algorithms*: Certain parameters must be defined in accordance with the compared algorithms in all of the experiments. In the case of SocialTrust, α=0.2, β=0.8, λ=0.85, ψ=0.5 [7],$\gamma_1$=1, $\gamma_2$=4, $k$=1, and the initial reputation is set to 0.1. In the case of modified BRS, $q$= 0.02. In the case of TRAVOS [44], $bin_1$= [0, 0.2],…,$bin_5$= [0.8, 1] and ε= 0.2. In the case of POYRAZ, φ= 0.1, ε= 0.4,γ = 0.6, and a rater is regarded as a liar if his/her trustworthiness value is less than 0.5 [40]. The initial reputations of all other reputation models being compared are set to 0 with the exception of SocialTrust.

The experiments on the CR&PCR model tested the effects of the parameters $\omega_1$, $\omega_2$, $\varpi_1$, $\varpi_2$, $\kappa$,

and *MRT* on the estimated transaction reputation of providers with the probability of dishonesty $PD \in$ [0, 0.9]. These experiments are conducted in reliable feedback environments without liars and subjectivity. The estimated reputation values usually converge starting from the$10^{th}$ round to the $100^{th}$ round and gradually stabilize in value. In the following sections, we set the default settings as follows: *MRT*=0.6, $\kappa$=0.5 (as argument $\kappa$ increases from 0.2 to 0.8, the weight of the recent reputation and the accuracy curve of the estimated reputation increase smoothly), $\varpi_1$ =0.3, $\varpi_2$ =0.3, $\omega_1$=0.2, $\omega_2$=0.2 (we expect the friend-and-follow relationship reputation to supplement or witness to the actual performance of a rater; hence, the weights of the components of this relationship reputation are low), and *sub*=0.5. The following experiments revise a small fraction of the assumptions.

The success of a reputation system is measured by its accuracy in predicting the quality of future interactions based on calculated reputations. However, this accuracy is difficult to achieve in an environment wherein any party can attempt to exploit the system for its own benefit. Some attacks that are motivated by selfish intent are narrowly focused and affect only the reputation of the misbehaving users or of a few selected targets (*first-type attack*). The influence of other attacks motivated by malicious intent are broader in comparison and affect large percentages of users within the system (*second-type attack*) [18].

*5.2 Comparison and discussion of CR and the alternative models against the lack of first-hand information under deceptive and subjective environments*

In this section, we request consumers to make selection decisions based on information from others rather than on their own previous experiences to test the effectiveness of our approach against the lack of first-hand information. To select providers under deception, we must filter deceptive information. Thus, we experimentally evaluate CR and then compare it with alternative models of trust and reputation to validate the deception filtering approach that is employed in the CR.

A typical example of the *first-type attack* (orchestrated self-promoting and slandering) is when exploitative provider B boosts his/her own reputation through liars and defames the reputation of his/her competitor A. The liars divide themselves into two teams and each team selects either A or B at random for transaction and misreport, thereby manipulating the opinions available to other users. We verify the reputation models against the first-type attack in deceptive and subjective environments.

*Deceptive environments without subjectivity*: Consumers with similar demands also share tastes (*sub*= 0) in this setting. As such, such consumers are similarly satisfied with the consistent *Sl* service quality provided by the same providers. To determine the effect of *MRT* under deception in this environment, we repeat our experiments for different percentages of liars given 50% honest (*PD*=0) and 50% malicious providers (*PD*=1.0). The service quality level delivered by malicious providers follows a normal distribution N(0, $\sigma$)(“no response”) with 50% probability or N(*Sl* − 1, $\sigma$)(“low grade”) with 50% probability so as to disguise negative behavior. Provider B displays one of these two kinds of poor service qualities at random. The test result shows that the precision of our algorithm (CR) increases as the value of argument *MRT* increases from 0.2 to 0.8 when the percentage of liars is less than or equal to 50% (*liar*≤50%). This precision also increases as the value of argument *MRT* increases from 0.2 to 0.5 when the percentage of liars is greater than 50% (*liar*>50%). However, the precision level decreases when *MRT* value increases further from 0.5 to 0.8. Optimal precision is achieved at *MRT*=0.6.

To understand the rationale behind this observation, we ran traces and recomputed the precision

accumulated over every round from tick =0 to 2000 when *MRT*=0.6. The error rate of CR with respect to the identification of liars increases when the ratio of liars (*liar*) exceeds than 0.5 (left-hand graph in Fig. 2). Thus, the CR misclassifies liars as honest raters and honest consumers as liars under this condition. Fortunately, CR can identify the "bad-mouthing" ratings of liars in relation to provider A with increasing effectiveness with the increment in the number of interactions (tick >100). Thus, accumulated precision increases accordingly. In addition, the error rate of CR with regard to the identification of liars increases slightly with the increments in the numbers of interactions and of liars when the ratio of liars (*liar*) is less than 0.5 because the CR begins misclassifying honest consumers as liars when the interactive number of liars is sufficiently large. Meanwhile, the "low grade" service quality level delivered by provider B disguises its malicious behaviors and generates increased noise in deception filtering.

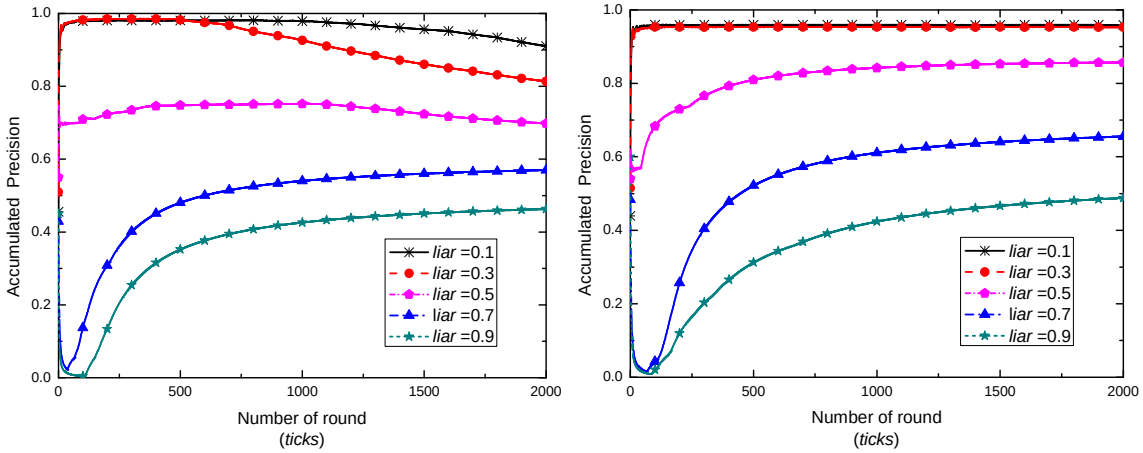


**Fig.2**.Trust convergences of CR with (right) and without (left) subjectivity in deceptive environments

*Comparison with other methods*: To determine the effect of deception filtering approaches based on public information in such a setting, we compared CR with alternative models (*MRT*=0.6). Fig. 3 presents the experimental comparison results of BRS, POYRAZ (which considers only the public credit of raters), and SocialTrust (which is not personalized). The precision value obtained by all algorithms decreases considerably with a significant increase in the percentage of liars, and our reputation model (CR) outperforms BRS, POYRAZ, and SocialTrust given that the percentage of liars in society (*liar*) is greater than 60%. This outcome can be attributed to the following two factors: First, CR incorporates the multi-reputation components of raters, which increases the reputation evidence that originates from raters. Second, the proposed deception filtering approach derives the "bad-mouthing" ratings above the *MRT* independently of both the behavior patterns of the provider and the ratio of liars in society. Therefore, it can gradually resist the malicious providers who defame the reputation of their competitors and aim to boost their own reputations when the percentage of liars is greater than 50%.

POYRAZ (which considers only the public credit of raters), SocialTrust (which is not personalized), and BRS are sensitive to the percentage of liars. In particular, the performance of BRS decreases more significantly than that of performance of and SocialTrust when *liar* is greater than 0.4. This result is expected because BRS is designed for use in environments wherein a considerable majority of consumers is honest. By contrast, POYRAZ is designed for settings in which both the personal observations (private information) of the consumers and public information are available. Thus, the reputation calculated with POYRAZ considers additional liar ratings and decreases in proportion with

the increase in the percentage of liars given the lack of personal observations. SocialTrust is developed for environments wherein trust groups (social relationships) are present. Thus, SocialTrust cannot detect misbehaving trust groups when the percentage of liars increases significantly.

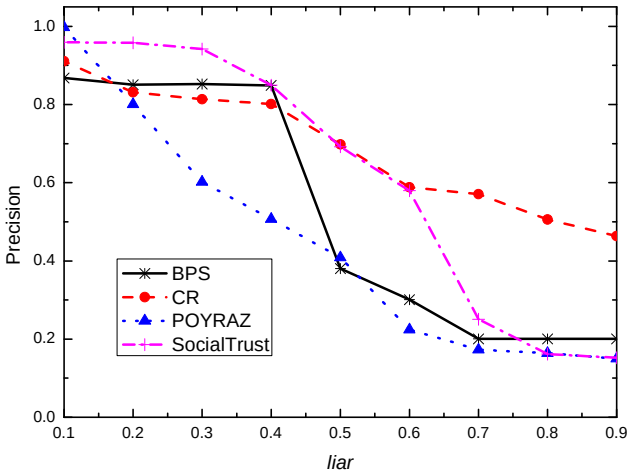


**Fig. 3**. Comparison of the filtering approaches in deceptive environments

*Deceptive and subjective environments*: In many real-life settings, deceptive and subjective information coexist. To determine the combined effect of subjectivity and deception as the deception ratings are filtered, we repeat our experiments for the different ratios of liars under subjectivity. The corresponding experimental results indicate that the precision of CR increases slightly with the increase in the subjectivity of raters from 0 to 0.5. This finding suggests that the subjectivity of raters partly aggravates the "low grade" service quality level delivered by provider B and that this subjectivity favors the deceptive filter. As mentioned previously, precision increases as the value of argument *MRT* increases from 0.2 to 0.6. However, precision decreases when *MRT* value increases further from 0.6 to 0.8. Precision is generally optimal at *MRT* =0.6. The precision accumulated over every round is plotted in the right-hand graph of Fig. 2 from tick =0 to 2000 when *MRT*=0.6. The speed of trust convergence depicted in this graph (with subjectivity) is higher than that in the left-hand graph when the percentage of liars is less than 50%.

*Comparison with other methods*: Fig. 4 displays the comparison of the experimental results with those obtained with BRS, POYRAZ (which considers only the public credit of raters), and SocialTrust (which is not personalized) in the same setting. The experimental results change with an increase in the subjectivity of raters from 0 to 0.5. Furthermore, the precision levels of CR, SocialTrust, and POYRAZ increase slightly, but that of BRS decreases somewhat. This result can mainly be attributed to the fact that the subjectivity of raters partly reduces the effect of the disguised "low-grade" service quality level delivered by malicious provider B. Overall, CR (*MRT*=0.6) is more robust than BRS, POYRAZ, and SocialTrust in relation to deception and subjectivity. In addition, all algorithms perform at the similar level in environments with both subjective and non-subjective populations.

The vulnerability of the BRS approach to subjectivity is expected because this rating-based approach assumes that consumers are not subjective [23]. That is, this approach infers that every consumer rates "good" providers as good and "bad" providers negatively. Under subjectivity (e.g., *sub*= 0.5), however, the definitions of "good" and "bad" depend on each consumer and may change significantly across consumers, as in real life. Therefore, consumers (raters) with different tastes are difficult to distinguish from the liars. That is, the ratings of an honest consumer and a liar of the same providers need not be in conflict at all times. Moreover, if two consumers are both honest but they

differ in terms of satisfaction criteria under the subjectivity condition, then their ratings conflict in terms of consumer satisfaction. If two honest consumers consistently rate other providers negatively, the providers cannot satisfy any of them. Therefore, consumers with different tastes are difficult to differentiate from liars and disguised providers who provide a "low grade" level of service.

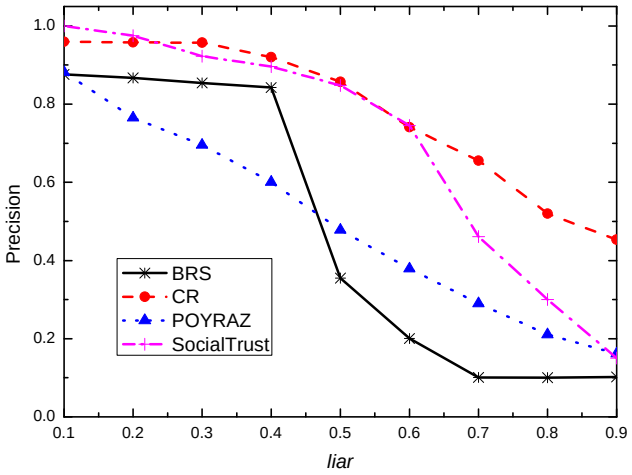


**Fig.4**. Comparison of the filtering approaches in deceptive and subjective environments

*5.3. Comparison and discussion of PCR and the alternative models under various environments*

In this section, we presume that consumers make decisions based on both information derived from their own previous experiences and that obtained from the experiences of others. We also compare PCR ($MRT$ = 0.6) with alternative models of trust and reputation. To begin with, we verify the trust and reputation models against the *first-type attack* in various environments.

*Deceptive environments without subjectivity*: We compare the PCR model with the alternative algorithms BRS, POYRAZ, TRAVOS, and the personalized SocialTrust in this setting to determine the effect of deception. Fig.5 indicates that the performance levels of the trust and reputation models PCR and POYRAZ do not decrease significantly with the increase in the percentage of liars in society. Although POYRAZ is slightly more sensitive to the ratio of liars than PCR, both algorithms identify satisfactory providers effectively. Unlike POYRAZ and PCR, however, SocialTrust, TRAVOS, and BRS are extremely sensitive to the percentage of liars and the deceptive information they disseminate to the society. Nonetheless, the personalized SocialTrust is more effective than BRS and TRAVOS. This difference in performance can be ascribed to the fact that the personalized SocialTrust balances the ratings by liars with personal observations. TRAVOS performs poorly in this environment because this model labels raters with conflicting tastes as liars and disregards their ratings accordingly. Thus, TRAVOS cannot distinguish the dispersion of ratings induced by taste differences from that caused by other factors (including the deceptive ratings of raters and the malicious behaviors of provider B) when malicious provider B delivers a disguised level of service that follows a normal distribution N(0, $\sigma$)("no response") with 50% probability or N($Sl-1$, $\sigma$)("low grade") with 50% probability in deceptive environments.

When TRAVOS applies the personal observations of consumers regarding providers to detect and filter out deceptive ratings, it is more effective than or is equally effective to the BRS algorithm that filters out deceptive ratings about the providers based on the majority of ratings, as exhibited in Fig. 5. Moreover, the performance of TRAVOS improves with the increase in the number of consumers to reputation source interactions, whereas the performance of BRS remains constant because it does not

learn from past experience.

*Deceptive and subjectivity environments*: In this setting, we compare the PCR model with the alternative algorithms BRS, POYRAZ, TRAVOS, and SocialTrust to determine the collective effect of subjectivity and deception. The precision levels of both POYRAZ and PCR improve slightly when the subjectivity of raters increases from 0 to 0.5. This result is expected because these algorithms factor in the subjectivity of raters in different ways. Fig.6 depicts the enhanced performance of TRAVOS. This observation is primarily attributed to the fact that the subjectivity of raters partly aggravates the "low grade" service quality level delivered by malicious provider B. Furthermore, this subjectivity favors TRAVOS because this algorithm labels raters with conflicting tastes as liars and disregards their ratings accordingly, as mentioned previously.

*Deceptive and subjective environments with intermittent providers*: Provider performance may change over time in many environments. In open networks, the detection and tracking of intermittent behavior is a major challenge. Malicious providers that display such behavior patterns strategically alternate their behavior to disrupt the application scenario in which the reputation system is deployed. In this setting, the exploitative provider B expresses intermittent behavior patterns ($PD$=0.5), and its delivered service quality level follows a normal distribution N(0, $\sigma$)("no response") with 50% probability or N($Sl$, $\sigma$)("promised") with 50% probability. Provider B displays one of these two kinds of service qualities (good and poor) in sequence. Thus, only 50% of its total provisions are good. As a result, the intermittent behaviors of provider B is liable to be mistaken for malevolent rater behaviors.

The experimental results depicted in Fig.7 indicate that PCR is insensitive to intermittent providers and that the precision value remains greater than 0.9. PCR can track the change in provider behaviors quickly because it amplifies the shift in recent user behavior using the behavior shift adaptation factor $Af$. In addition, both POYRAZ and SocialTrust are more effective than BRS and TRAVOS because they model trust dynamics with more sophisticated characteristics.

*Deceptive and subjectivity environments with collusion and intermittent providers*: In the experiments described above, the liars are isolated. In fact, their misbehavior is effectively only with additional capabilities. Liars typically form cliques by colluding and leveraging their tightly coupled relationship structure to overpower trust and reputation models. The size of the collusive group is equal to the total number of deceptive users in the system in our simulation. Collusive raters give an extremely positive rating of +1 to the collusive members and a highly negative rating of −1 to users outside the collusive group. The service quality level delivered by collusive providers follows a normal distribution N(0, $\sigma$) with 90% probability and N($Sl$, $\sigma$) with 10% probability. Their reputation is boosted, whereas the reputation of honest users is corrupted by the collusive members. Thus, the derived reputation ratings favor collusive users. In this scenario, the collusive users in the network can promote other collusive members. Fig.8 indicates that the precision levels obtained by all of these algorithms (except PCR) decrease significantly with the considerable increase in the percentage of collusive liars. The experimental results also suggest that PCR is robust to collusive liars and that precision values remain above 0.95. This outcome can be attributed to the following three factors: First, PCR incorporates the multi-reputation components of the raters, which enhances the source of reputation evidence. Second, the adopted deception filtering approach can partially filter out the "bad-mouthing" ratings on honest provider A and differentiate the honest raters from collusive liars in the computation of CR score. Consequently, the degradation of the reputation of provider A as a result of the malevolent behaviors of collusive liars can be mitigated to some extent. Third, the estimated risk in the computation of PCR score, can limit the boosting of the reputation of the exploitative

provider B (*PD*=0.9) as a result of the collusive behaviors of liars during provider selection. This risk in the computation of PCR score is based on the direct and extremely negative interaction experience of the consumer.

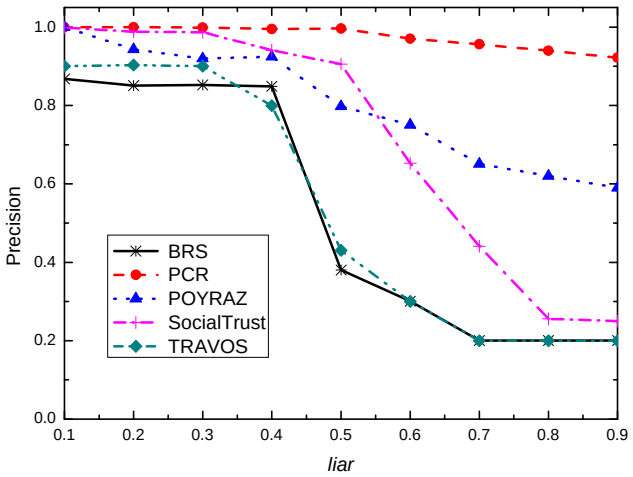


**Fig.5**. Comparison of trust and reputation approaches in deceptive environments

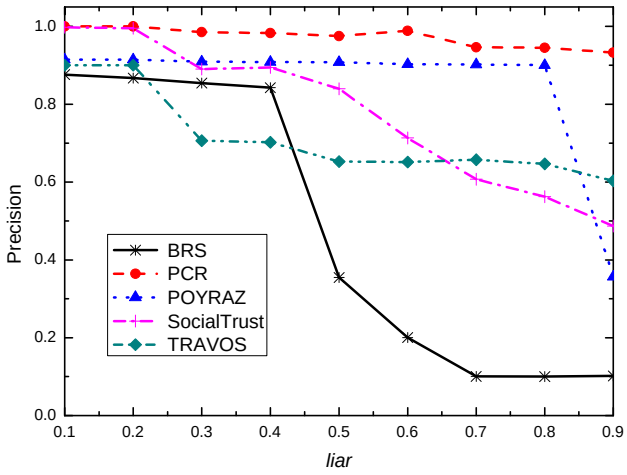


**Fig.6**. Comparison of trust and reputation approaches in deceptive and subjective environments

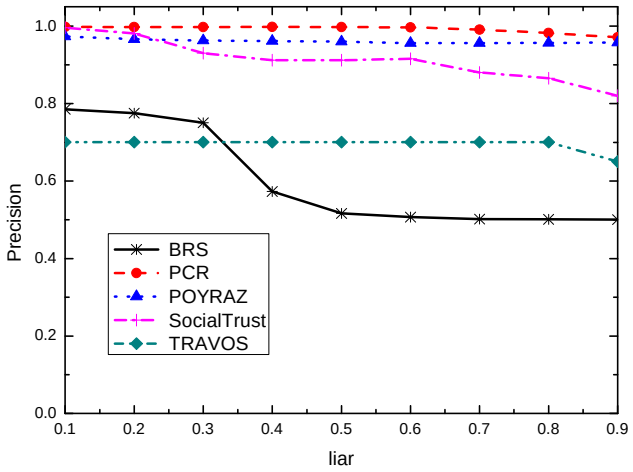


**Fig.7**. Comparison of trust and reputation approaches in deceptive and subjective environments given intermittent providers

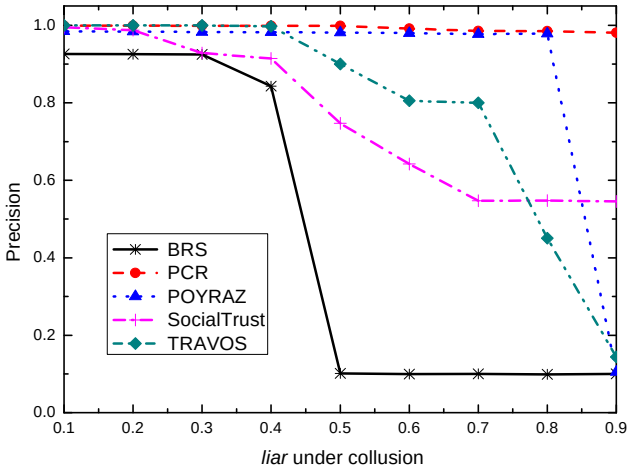


**Fig.8**. Comparison of trust and reputation approaches in deceptive and subjective environments under collusion

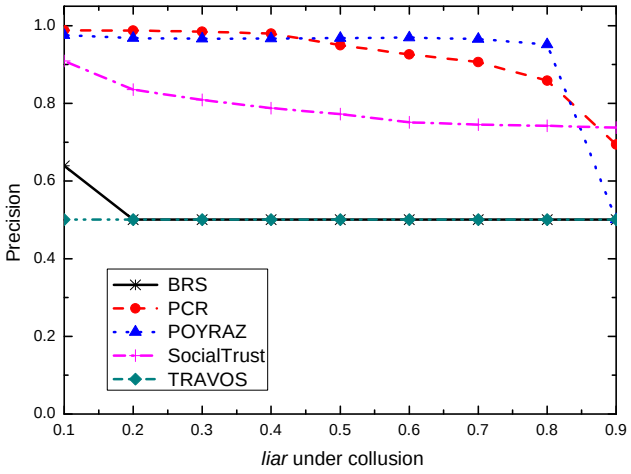


**Fig.9**. Comparison of trust and reputation approaches in deceptive and subjective environments given both intermittent providers and collusion

Similar results can be obtained for populations with malicious providers who display intermittent behavior patterns. Fig.9 exhibits the results for all of these algorithms when provider B expresses intermittent behavior patterns ($PD$=0.5) and suggests that intermittent providers who collude with liars induce significant noise and fluctuation.

In addition, we verify the trust and reputation models against the *second-type attack* with a broad influence over a large percentage of the identities within the system under various environments. A typical example of the *second-type attack* is the scenario in which honest consumers (raters) always recommend a good provider, whereas liars randomly select an exploitative provider to boost or a good provider to defame in each round, thus ruining the system. In this setting, the direct interaction of honest consumers with others in the system is limited.

The experimental results show that precision level decreases gradually and continuously as the proportion of liars increases. Fig.10 depicts the results for all of these algorithms given coexisting deceptive and subjective information. The service quality level delivered by malicious providers ($PD$=0.9) follows a normal distribution N(0, $\sigma$)("no response") with 90% probability and N($Sl$, $\sigma$)("promised") with 10% probability. The $PD$ of the honest provider is set to 0. The decrease in precision (except POYRAZ) is significant when the ratio of liars increases as per this figure. POYRAZ

is more effective than the other algorithms because it detects deception using personal and shared public information. By contrast, PCR, BRS, and SocialTrust identify liars according to shared public information alone, and TRAVOS depends solely on personal observations. POYRAZ is superior even at the start of the simulations because it combines inadequate personal observations with public information provided by others. Public information normally confuses this algorithm as the ratio of liars increases, but it follows the application of public data with private inspections. Thus, POYRAZ is not significantly affected by misleading public information. This benefit is enhanced further as the skepticism of consumers increases with experience. Nonetheless, POYRAZ remains superior because it combines the private and public credits of the raters. However, the private credits of raters may not always be available in real social commerce. Thus, the performance of TRAVOS worsens with the decrease in the number of interactions between consumer and reputation source (limited private observations). In addition, PCR performance improves significantly with the increased interaction between consumers and providers.

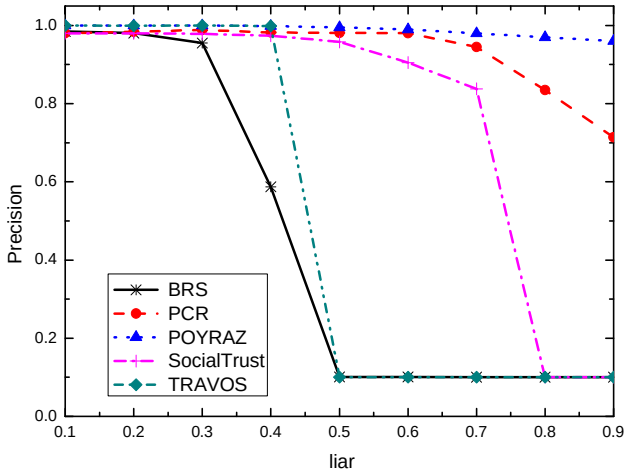


**Fig.10**. Comparison of trust and reputation approaches in deceptive and subjective environments under large-scale manipulation

In summary, we conclude that the model proposed in this study outperforms other methods in most cases based on the simulation results and discussions presented previously, especially with respect to the detection of dynamics and deception. In addition, the 50 providers, 500 consumers (raters), and 2,000 round traces analyzed above suggest the following order of algorithm speeds: POYRAZ >TRAVOS> SocialTrust >=CR&PCR> BRS (CPU Intel core i5 3.0 GHz, memory16 GB). The runtime of all of the methods increases gradually as the number of users increases. However, that of CR&PCR decreases gradually as the number of bad-mouthing users increases because of the deception filter for “bad-mouthing” ratings.

## 6. Conclusion

This study designs and evaluates a graph-based CR model to aggregate reputation in social commerce scenarios, to improve the veracity and objectivity of opinions, and to enhance the trust of consumers in providers. The proposed model supports the accurate and robust establishment of reputations even in the presence of new and malicious-manipulation users. On the basis of consumer behavior and psychological theory, the main contributions of this study are listed as follows:

(i) A (new) consumer usually has limited or virtually no direct interaction or relationship with other consumers in the context of social commerce. We believe that the scarcity of first-hand information

can be addressed using the information made available in various contexts. The proposed model can generate resilient reputation ratings to combat this scarcity on the basis of constructed UASRNs and by exploiting the evidence of opinions from relative social context through the integration of user behavior reputation with his/her social relationship reputation.

(ii) Raters can be malevolent, and providers can behave dynamically or inconsistently toward different consumers to balance the maximization of selfish or malicious behavior and the avoidance of detection. We believe that e-commerce users are sensitive to transaction price. They act rationally during transaction but rate subjectively. Effective trust factors can be identified by deep-mining the context of user opinions, which can in turn mitigate the subjectivity of opinions and the dynamics of behaviors. As a result, reputation can be estimated accurately.

(iii) Trust and perceived risk are essential constructs of transaction intention in the face of uncertainty. The behavioral intention of a customer toward repeat intention is clarified through the full synthesis of ECM and TAR. We believe that a rational consumer has a limited risk tolerance toward the negative outcomes of transaction behavior as per the theoretical rationale of TRA and ECM. That is, overall dissatisfaction with past outcomes negatively influences the repeat purchase intention and behavior of a rational consumer. Thus, a rational consumer is unlikely to resume transactions with the same provider if a high proportion of their transactions end negatively. The enhanced and personalized model leverages the limited risk tolerance of the consumer for negative transaction experiences and can enhance the accuracy and robustness of trust and reputation decision-making under deceptive and collusive environments. This result is attributed to the fact that the combination of the novel deception filtering approach with the personalized direct distrust (risk) metric can discredit “bad-mouthing” opinions and detect malicious providers to a certain extent.

The experimental results show that the model proposed in this study facilitates a personal and satisfying provider selection process in most cases, unlike alternative trust and reputation methods.

However, the performance of the proposed model can still be improved as follows: (i) As a result of the separation of transaction and social relationship data in real life, the experiments do not strictly follow real-world social network examples in which the degree of connection in scale-free networks obeys a power law. “Free text” reviews are crucial in reputation assessment in actual social e-commerce, which is dictated not only by numeric ratings but by these reviews as well. Thus, textual reviews should be converted into numeric ratings when CR&PCR is applied to real situations. (ii) Malicious users provide deceptive feedback through unfairly high and/or unfairly low ratings; thus, the deception filtering approach should also consider “inflated” ratings to enhance the quality of the effect. In addition, the behavior strategies of providers have not been examined in-depth for the PCR model in consideration of the fact that providers can vary their behaviors either intentionally or unintentionally. (iii) The CR does not fully develop the role-based reputation mechanism to boost the initial reputation of new users. The concept proposed by Burnett et al. [6] may be especially valuable for this model because it can complement the current work through methods that can boost user reputation.

**Acknowledgments**

This work is supported in part by the Natural Science Foundation of China (Grant Nos. 61379034, 60903038,61202197, and 51375429), the Zhejiang Provincial Natural Science Foundation of China (Grant No.LQ14F010006), and the Science and Technology Project of Zhejiang Province (Grant No. 2010R50041).

## References


[1] S.Al-Ou, H.N.Kim, A.E.Saddik, A group trust metric for identifying people of trust in online social networks, Expert Systems with Applications 39 (18) (2012)13173–13181.

[2] P.Bedi, R.Sharma, Trust based recommender system using ant colony for trust computation, Expert Systems with Applications 39 (1) ( 2012) 1183-1190.

[3] D.Beisel, The emerging field of social commerce and social shopping, In: Proceedings of Genuine VC, 2006.

[4] T.Beth, M.Borcherding, B.Klein, Valuation of trust in open networks, Springer Berlin Heidelberg,1994.

[5] S.Buchegger, J.Y.Le Boudec, A robust reputation system for P2P and mobile ad-hoc networks, In: the Second Workshop on the Economics of Peer-to-Peer Systems (P2PEcon), 2004, pp. 1-6.

[6] C.Burnett, T.J.Norman, K.Sycara, Bootstrapping trust evaluations through stereotypes, In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent System (AAMAS), 2010, pp. 241-248.

[7] J.Caverlee, L.Liu, S.Webb, The SocialTrust framework for trusted social information management: architecture and algorithms, Information sciences 180 (1) (2010) 95-112, special Issue on Collective Intelligence.

[8] M.Chen, J.P.Singh, Computing and using reputations for internet ratings, In: the 3rd ACM Proceeding on Electronic Commerce(ACM-EC), Tampa, USA, 2001, pp. 154-162.

[9] S.Cook, The contribution revolution: letting volunteers build your business, Harvard Business Review, (2008) 60-69.

[10] V.Deora, J.Shao, W.Gray, J.Fiddian, A quality of service management framework based on user expectations, In: the International Conference on Service-oriented Computing (ICSOC), Trento, Italie, 2003, pp. 104-114.

[11] C.Dellarocas, Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior, In: the ACM Proceeding on Electronic Commerce (ACM-EC)), ACM, Minneapolis, MN, USA, 2000, pp. 150-157.

[12] C.Dellarocas, Strategic manipulation of internet opinion forums: implications for consumers and firms, Management Science 52 (10) (2006) 1577-1593.

[13] A.Fernandes, E. Kotsovinos, S.Östring, B.Dragovic, Pinocchio: Incentives for honest participation in distributed trust management, In:Trust Management, Springer Berlin Heidelberg, 2004, pp. 63-77.

[14] M.Fishbein, I.Ajzen, Beliefs, attitude, intention and behavior: an introduction to theory and research, Addison-Wesley, Reading, MA, 1975.

[15] Z.B. Gan, Q.Ding, V.Varadharajan, Reputation-based trust network modeling and simplification in multiagent-based e-commerce systems, In: Fifth International Conference on Next Generation Web Services Practices (NWESP'09),2009, pp.60-67.

[16] J.Golbeck, Generating predictive movie recommendations from trust in social networks, In: the International conference on Trust Management, Springer-VerlagBerlin, Heidelberg, 2006, pp. 93-104.

[17] S.Guo, M.Q.Wang, J.Leskovec, The role of social networks in online shopping: Information passing, price of trust, and consumer choice, In: the ACM Conference on Electronic Commerce(ACM-EC), San Jose, California, USA, 2011, pp. 157-166.

[18] K.Hoffman, D.Zage, C.Nita-Rotaru, A survey of attack and defense techniques for reputation systems, ACM Computing Surveys 42(1) (2009) 1-19.

[19] Z.Huang, W.Y.Chung, H.C.Chen, A graph model for e-commerce recommender systems,Journal of the American Society for Information Science and Technology, 55(3) (2004) 259–274.

[20] M.Jamali, Probablisitic model for recommendation in social networks,Phd Thesis,Simon Fraser University,2013.

[21] X.Jin, S.H.G.Chan, Reputation estimation and query in peer-to-peer networks, Communications Magazine, IEEE, 48(4) (2010) 122-127.

[22] Y.Jin, M.Su, Recommendation and repurchase intention thresholds: a joint heterogeneity response estimation, International Journal of Research in Marketing 26 (3) (2009) 245-255.

[23] A.Jøsang, R.Ismail, C. Boyd, A survey of trust and reputation systems for online service provision, Decision Support Systems 43 (2) (2007) 618-644.

[24] R.Jurca, B.Faltings, An incentive compatible reputation mechanism, In: E-Commerce, 2003. CEC 2003. IEEE Conference on, Newport Beach, CA, USA, 2003, pp. 285-292.

[25] G.Kambourakis,Anonymity and closely related terms in the cyberspace: An analysis by example,Journal of Information Security and Applications (2014), http://dx.doi.org/10.1016/j.jisa.2014.04.001.

[26] C.Liao, J.L.Chen, D.C.Yen, Theory of planning behavior (TPB) and customer satisfaction in the continued use of e-service: An integrated model, Computers in Human Behavior, 23(6) (2007) 2804-2822.

[27] R.Kerr, R.Cohen, Trust as a tradable commodity: afoundation for safe electronic marketplaces, Computational Intelligence 26(2) (2010) 160-182.

[28] H.Kim, R.D.Galliers, N.Shin, J.H.Ryoo, J.Kim, Factors infiuencing Internet shopping value and customer repurchase intention, Electronic Commerce Research and Applications 11 (14) (2012) 374–387.

[29] Y.M.Li, C.P.Kao, TREPPS: a trust-based recommender system for peer production services, Expert Systems with Applications 36 (1) (2009) 3263-3277.

[30] S.L.Linda, Social commerce–E-commerce in social media context, Word Academy of Science, Engineering and Tchnology 50 (2011) 39-44.

[31] J.Liu, Y.Cao, C.-Y. Lin, Y.Huang, M.Zhou, Low-quality product review detection in opinion summarization, In: the Joint Conference on Empirical Methods in Natural Language Processingand Computational Natural Language Learning (EMNLPCoNLL), 2007, pp. 334–342.

[32] Y.Lu, P.Tsaparas, A.Ntoulas, L.Polanyi, Exploiting social context for review quality prediction, In: the 19th international conference on World Wide Web(WWW), ACM, 2010, pp. 691-700.

[33] Y.Lu, L.Zhao, B.Wang, From virtual community members to C2C e-commerce buyers: trust in virtual communities and its effect on consumers purchase intention, Electronic Commerce Research and Applications 9 (4) (2010) 346-360.

[34] D.Lucking-Reiley, Auctions on the Internet: what's being auctioned, and how?, Journal of Industrial Economics 48 (3) (2000) 227-252.

[35] J.F.Luo, S.L.Ba, H. Zhang, The effectiveness of online shopping characteristics and well-designed websites on satisfaction, MIS Quarterly 36 (4) (2012) 1131-1144.

[36] D.H.McKnight, V.Choudhury, C.Kacmar, Developing and validating trust measures for e-commerce: an integrative typology, Information Systems Research 13 (3) (2012) 334-359.

[37] M.Parameswaran, A.B.Whinston, Social computing: an overview, Communications of the Association for Information Systems 19 (2007) 762-780

[38] P.A.Pavlou, Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model, International journal of electronic commerce, 7(3) (2003) 101-134.

[39] C.Riegner, Word of mouth on the web: the impact of web 2.0 on consumer.purchase decisions, Journal of Advertising Research 47 (4) (2007) 436-248.

[40] M.Sensoy, J.Zhang, P.Yolum, R.Cohen, POYRAZ:Context-aware service selection under deception, Computational Intelligence 25 (4) (2009) 335-366.

[41] A.T.Stephen, O. Toubia, Deriving value from social commerce networks, Journal of Marketing Research 47 (2) (2010) 215-228.

[42] J.M.Such, A.Espinosa, A.Garcia-Fornes, V.Botti, Partial identities as a foundation for trust and reputation, Engineering Applications of Artificial Intelligence 24 (7) (2011) 1128-1136.

[43] M.Tavakolifard, K.C.Almeroth, Social computing: an intersection of recommender systems, trust/reputation systems, and social networks, Network,IEEE 26 (4) (2012) 53-58.

[44] W.T.L.Teacy, J.Patel, N.R. Jennings, M. Luck, TRAVOS: Trust and reputation in the context of inaccurate information sources, Autonomous Agents and Multi-Agent Systems 12(2) (2006)183–198.

[45] W.T.Teacy, M.Luck, A.Rogers, N.R.Jennings, An efficient and versatile approach to trust and reputation using hierarchical Bayesian modelling, Artificial Intelligence 193 (2012) 149-185.

[46] O.Tsur, A.Rappoport, REVRANK: A fully unsupervised algorithm for selecting the most helpful book reviews, In: the International Conferences on Weblogs and Social Media (ICWSM), 2009, pp. 154-161.

[47] S.Utz, P.Kerkhof, J.Van den Bos, Consumers rule: how consumer reviews influence perceived trustworthiness of online stores, Electronic Commerce Research and Applications 11(1) (2012) 49-58.

[48] C.Von Der Weth, K.Böhm, A unifying framework for behavior-based trust models,In: on the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE, Springer Berlin Heidelberg, 2006, pp. 444-461.

[49] Y.Wang, D.S.Wong, K.J. Lin, V.Varadharajan, Evaluating transaction trust and risk levels in peer-to-peer e-commerce environments, Information Systems and E-Business Management 6 (1) (2007) 25-48.

[50] C.Wang, P.Zhang, The evolution of social commerce: an examination from the people, business, technology, and information perspective, Communications of the AIS (CAIS) 31 (2012) 105-127.

[51] A.Whitby, A.Jøsang, J.Indulska, Filtering out unfair ratings in bayesian reputation systems, In: Proceedings of the Workshop on Trust in Agent Societies, at the Autonomous Agents and Multi Agent Systems Conference (AAMAS), New York, 2004.

[52] F.Wu, H.H. Li, Y.H.Kuo, Reputation evaluation for choosing a trustworthy counter party in C2C e-commerce, Electronic Commerce Research and Applications 10 (4) (2011) 428-436.

[53] K.Q.Xu, X.T. Guo, J.X. Li, R.Y.K.Lau, S.S.Y. Liao, Discovering target groups in social networking sites: an effective method for maximizing joint influential power, Electronic Commerce Research and Applications 11(4) (2012) 318–334.

[54] K.Q.Xu, J.X.Li, Y.X.Song, Identifying valuable customers on social networking sites for profit maximization, Expert Systems with Applications 39 (17) (2012) 13009-13018.

[55] H.Yu, P. Gibbons, M.Kaminsky, F.Xiao, SybilLimit: A near-optimal social network defense against sybil attacks, IEEE/ACM Transactions on Networking 18 (3) (2010) 885-898.

[56] H.Yu, M.Kaminsky, P. B. Gibbons, A. Flaxman, SybilGuard: defending against Sybil attacks via social networks, In: Proceedings of the ACM Special Interest Group on Data Communicatio (SIGCOMM), New York, NY, USA, 2006, pp. 267-278.

[57] W.Yuan, D.Guan, Y.K.Lee, S.Lee, S.J.Hur, Improved trust-aware recommender system using small-worldness of trust networks, Knowledge-Based Systems 23 (3) (2010) 232-238.

[58] Y.Zhang, Y.Fang, K.K.Wei, E.Ramsey, P.McCole, H.Chen, Repurchase intention in b2c e-commerce—a relationship quality perspective, Information & Management 48 (6) (2011)192-200.

[59] .H.T.Zou, Z.G.Gong, N.Zhang, W.Zhao, J.Z.Guo, TrustRank: a Cold-Start tolerant recommender system, Enterprise Information Systems (2013) (ahead-of-print) 1-22.

**Appendix**

| Notions | Illustration |
|---|---|
| UASRNs | All Users' activities and relationships networks |
| $U=\{u_1, u_2,\ldots,u_n\}$ | A set of users |
| $C=\{c_1, c_2,\ldots,c_n\}$ | A set of context |
| $V=\{v_1,v_2,\ldots,v_n\}$ | A set of values associated with activities and relationships, |
| $B=\{...,beh_x(u_i,u_j,c_k,v_z),...\}$ | A set of all the activities that have occurred |
| $\Re=\{...,rel_y(u_i, u_j,c_k, v_z),...\}$ | A set of all the social relationships between users |
| $BEH=\{\ldots, beh_x, \ldots\}$ | A set of activity types, $x$=1(transaction), 2 (review),3 (advising) |
| $REL=\{\ldots,rel_y, \ldots\}$ | A set of social relationship types, $y$=1(friendship),2 (followship),… |
| $beh_x^n(i)$ | A set of the $beh_x$ (type) activities performed by user $u_i$ from time 1 to $n$ |
| $rel_y^n(i)$ | A set of social relationship members who have $rel_y$ (type) relationship with $u_i$ in time $n$ |
| $R^n(i)$ | The comprehensive reputation (CR) score of user $u_i$ in time $n$ |
| $R_i^t(j)$ | The personalized comprehensive reputation (PCR) score of provider $u_j$ estimated by consumer $u_i$ at time $t$ |
| $\hat{R}_{beh_x}^n(i)$ | The behavior reputation of user $u_i$ in the $n^{th}$ timeslot. |
| $R_{beh_x}^n(i)$ | The behavior reputation score of $beh_x$ activity of user $u_i$ in time $n$, $x$=1(transaction), 2 (review) |
| $R_{rel_y}^n(i)$ | The social relationship reputation score of $rel_y$ relationship of user $u_i$ in time $n$ , $y$=1(friendship),2 (followship) |
| $\varpi_x$ | The weights of the behavior reputation value |
| $\omega_y$ | The weights of the social relationship reputation value |
| $r_{i\to j}^{+}$ , $r_{i\to j}^{-}$, $r_{i\to j}^{-}$* | The positive rating, the negative rating, and the extremely negative rating |
| $A\mu$ | The maturity factor |
| $Af$ | The behavior shift adaptation factor |
| $Wa$ | The impact of transaction amount |
| $Wp$ | The expected utility of transaction price |
| $W\rho$' | The feedback utility of transaction price |
| $Wr$ | The recognition degree of a user |
| $Tr(a)$ | The transaction amount of a transaction activity |
| $Tr(p)$ | The transaction price of a product or service (category) |
| $Tr(p)$* | The average market price of a product or service (category) |
| $SP(i)$ | A set of providers who have transactions with user $u_i$ |
| $SC(j)$ | A set of consumers who have transactions with provider $u_j$ |
| $AV(i)$ | A set of advisors who have rated the reviewer $u_i$ |
| $Rq$ | The social relationship quality |
| $Rp$ | The intimacy level |
| $MRT$ | The maximum risk tolerance from the viewpoint of system |
| $PMRT$ | The personalized maximum risk tolerance of consumers |