Systematic analysis of centralized online reputation systems

Ling Liu *, Malcolm Munro

School of Engineering and Computing Sciences, Durham University, South Road, Durham, DH1 3LE, UK

ARTICLE INFO

ABSTRACT

Centralized online reputation systems have been widely adopted by Internet companies to help users build trust, reduce information asymmetry and filter information. Research in this area to date has focused on analyzing the effectiveness of single-type systems, while less attention has been paid to the comparison of different systems. This paper proposes an analysis model that can classify and measure different reputation systems in the same context. The model divides reputation systems into five underlying components: input, processing, output, feedback loop and storage. A series of benchmark criteria is then defined based on the characteristics of each component. The model comprehensively analyzes most characteristics of centralized reputation systems and it takes both performance and costs of systems into consideration.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Since the end of the 20th century, centralized online reputation systems have been widely adopted by Internet companies. These systems use Internet technologies to collect people's opinions on products, transactions and events then aggregate and publish the information [5,33,28]. For example, to build trust between strangers, eBay.com, one of the largest marketplaces on the Internet, allows buyers and sellers to leave positive, neutral or negative feedback for each other. Amazon.com encourages their users to write product reviews on the website, so that potential consumers can gather more information about the products [3,5]. Furthermore, by taking advantages of 'the wisdom of the crowd' [39], reputation systems can be used to filter information. Digg.com is a website that allows people to share Internet contents by submitting links of stories. Voting stories up ('digging') or down ('burying') is the site's cornerstone function. Each story and comment have a number associated with it, which is calculated by the number of 'diggs' minus the number of 'buries'. Larger numbers indicate more interesting stories.

Based on the information storage location, reputation systems can be divided into two main types [28,19,13]. *Centralized reputation systems* employ central servers to gather, process and disseminate information, whereas *distributed reputation systems*, rely on decentralized solutions, where users store reputation information at their own locations [19]. This paper concentrates on centralized systems only, hence in the following sections, all 'reputation system' refers to a centralized system unless otherwise noted.

2. Related work

A great deal of research in the area of reputation systems focuses on analyzing single-type systems. Because eBay has become a great success, much research has concentrated on eBay-like 'trust-building' mechanisms. Results in Lucking-Reiley et al. [29] and Utz et al. [41] showed that although reputation systems may not be perfect, trust can be built among strangers as long as the sellers value their reputation (i.e., the feedback scores in eBay). Furthermore, sellers, who have a good reputation (i.e., higher scores) may sell their products at a higher price [34,15]. Similarly, researchers found that product reviews can have a substantial influence on online retailers' sales and the sites' usefulness [2,20]. Lampe and Resnick [21] and Lerman [23] discussed how reputation systems can be used as 'social information processing' and thus help online social news centers rate and recommend stories.

There are a small number of papers that provide general reviews of different types of reputation systems. Liang and Shi [26] concentrated on the aggregation algorithms of reputation systems and divided them into five categories according to how they weight the ratings and how the weights are decided. Then, the authors used a simulation tool to evaluate their performances (algorithm complexity, system running cost and system benefits). The results showed that most of the time a better performance often accompanies with a high system cost and complexity. Sabater and Sierra [37] focused on the computational models and used a series of criteria to compare systems. Their criteria included the conceptual model, information source, granularity, accessibility, information format, agent behavior assumptions and trust/reputation reliability measures. Hoffman et al. [14] surveyed a number of academic models and commercial systems and measured their weaknesses to known attack strategies. The authors developed a classification framework which divided reputation systems into three

main processes: formulation, calculation and dissemination. *Formulation* measures the source of information and information type. The *calculation* assesses the aggregation algorithms, finally, *Dissemination* considers the distribution and storage durability of information. Ruohomaa et al. [35] proposed an evaluation model which focusing on the nature of the information, the gathering of the information and the aggregation of the information. Most of these cross-type evaluations concentrated on the reputation systems designed for Peer-to-Peer (P2P) systems, which are distributed systems. Malaga [30] focused on comparing 11 commercial sites. With the analysis, the authors found some common problems of reputation systems: inaccurate algorithms, barrier to entry, no incentives to rate, inability to filter or search, no categorization and unlimited memory of information.

In addition to the lack of research focusing on cross-type evaluation of centralized systems, cost is another factor that has long been ignored. As Dellarocas [4] indicated even the simplest rating incurs a cost. When analyzing reputation systems both their performance and costs should be considered.

This paper proposes a systematic analysis model which aims to measure both performance and cost of different kinds of reputation systems. The rest of the paper is organized as follows. Section 3 analyzes the structural model of all reputation systems. It is followed by Section 4, which introduces the criteria of the model. Section 5 discusses the validation of the criteria. Finally Section 6, summarizes the results of the paper and discusses the limitation and possible future research.

3. The terminology and structure of reputation systems

3.1. Reputation system terminology

Before further discussion, some terms need to be clarified. These terms, which represent the entities that are involved in reputation systems, have been widely used by researchers.

Definition 1. Information source

An information source provides information to the reputation systems.

Most of the time, information is provided by a person, i.e., an evaluator.

Definition 2. Evaluator

An evaluator is an information source, i.e., the person who provides reputation information.

In other cases, information sources can be systems. For example, some reputation systems collect information from other systems rather than from evaluators.

Definition 3. Target

A target refers to the entity for which evaluators provide information. A target may be a product, one transaction or even a story.

Definition 4. End user

An end user uses reputation systems for seeking information about a target.

Most of the time, end users are the visitors to the website, whereas evaluators may or may not be visitors. There will be more discussion in Section 4 on the set of evaluators.

Definition 5. Reputation information

Reputation information refers to information related to a target's reputation, such as reviews or ratings of a product.

Reputation systems usually collect two main kinds of reputation information: explicit information and implicit information.

Definition 6. Explicit information

Explicit information indicates the information that evaluators actively provide, such as a rating on the product and a text comment.

Definition 7. Implicit information

Implicit information is usually generated from evaluators' activities. For example, the total number of views of a video or a book selling figures.

The problem with the implicit information is that the true opinions of all the evaluators may not be reflected in the information. For example, a person may buy a book which eventually they dislike. Furthermore, when buying the book they may not be aware of the consequences of their activities, while evaluators, who provide ratings and reviews, know that the information will have influences on the target's reputation. Usually implicit information has more 'evaluators' than explicit information, because not all the people who buy a product will provide a rating. This paper concentrates on explicit information only.

3.2. Structure of reputation systems

In the area of Information Systems (IS), researchers tend to divide systems structure into four components (Fig. 1a): input, processing, output and feedback [22]. Input is the process of gathering data and processing transforms raw data into information. Output then transforms information into meaningful results with certain formats and feedback is used to provide information to change the input or processing activities [38].

It is commonly accepted that reputation systems are a specific kind of information systems. A number of researchers have proposed a similar structure for reputation systems (see Fig. 1b): information collection, processing and dissemination. Information collection indicates the activities of collecting ratings and reviews from information sources, processing refers to the aggregation of the ratings, and dissemination refers to the distribution of reputation information to the end users [14,45,40,6,12].

Following the IS structure analysis, our previous work [28] has proposed a more comprehensive structure for online reputation systems. Regardless of their interfaces, functions or roles, all reputation systems should have the following five components:

Definition 8. Input

Input is the process of collecting reputation information from information sources.

Definition 9. Processing

Processing is the procedure of computing and aggregating the reputation information.

Definition 10. Output

Output indicates the dissemination of the reputation information.

Definition 11. Feedback loop

A feedback loop is the collection of feedback of the review, which can be seen as the 'review of the review'.

The content of the review is an important factor in the reputation system. However it is difficult to measure the quality of each review; therefore, some systems adopt the feedback loop component to filter
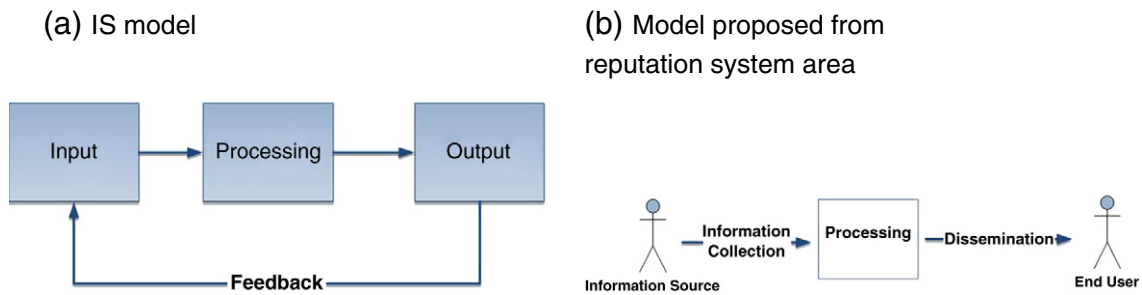
(a) IS model

(b) Model proposed from reputation system area

Fig. 1. Structure of reputation systems.

reviews. For example, Amazon allows their end users to vote on product reviews as 'helpful to you' or 'not helpful'. These reviews can then be ranked by the number of 'helpful' votes they have received.

Unlike the other components, the feedback loop is not included in all reputation systems. It should be noted that some websites use the word 'feedback' to refer to the reviews (reputation information). To avoid confusion in this paper, 'feedback' will be used to indicate the information collected in the feedback loop component only.

Definition 12. Storage

The storage refers to the process of storing all the collected and processed information.

Fig. 2, which is slightly revised from Liu et al. [28], presents the interrelationships between the five components. It shows that reputation information flows from sources to the Processing component. After being aggregated, it will be published. If end users are interested, they may be allowed to leave feedback (the dotted lines indicate that the feedback loop is an optional component). During the whole process, all information needs to be stored in the Storage component.

Based on the structure model, all reputation systems can be divided into these five components. Then, a series of benchmark criteria can be defined according to each component's characteristics. Thus, reputation systems can be assessed regardless of their different interfaces or functions.

4. Analysis model

In this section, a number of criteria are defined for each component. The paper will first discuss the characteristics of each component and then propose the relevant criteria. The paper will also discuss the possible quantification or the influential factors of each criterion.

Before further discussion, it should be noted that online reputation systems do not solely exist on the Internet. They are integrated within websites. The evaluation of reputation systems should focus on the system's own characteristics rather than evaluating the website that adopted the system. In other words, even though some factors may have influence on the quality of reputation systems, such as web page design or usability, they will not be assessed as long as they are not directly related to the system.

4.1. Input

Input refers to the collection of ratings, text reviews and other relevant reputation information. It is one of the most important components because the other four components rely on the information collected from the input.

Three essential elements are involved in the input: collection channels, information sources and reputation information. Reputation systems
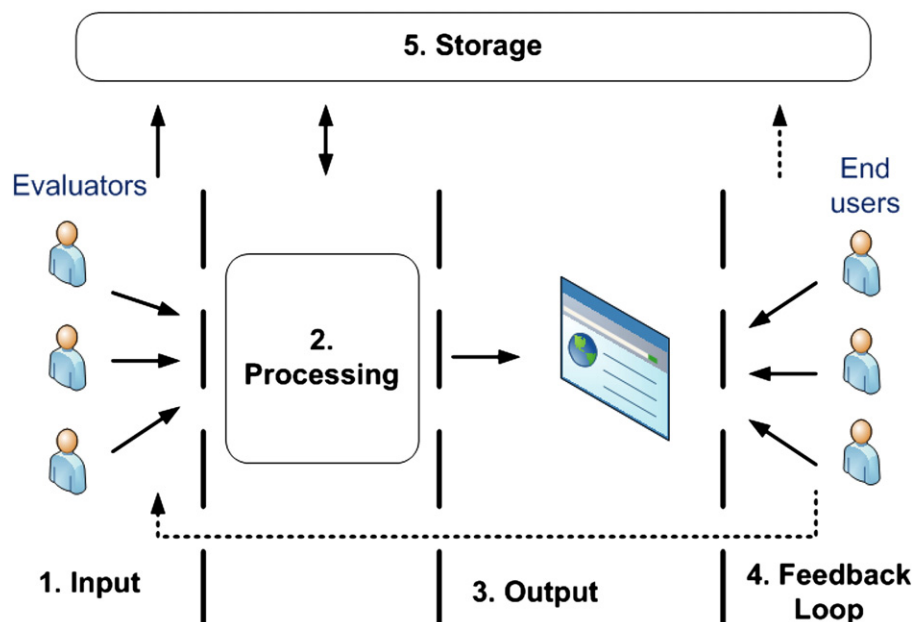

Fig. 2. Reputation system structure model.

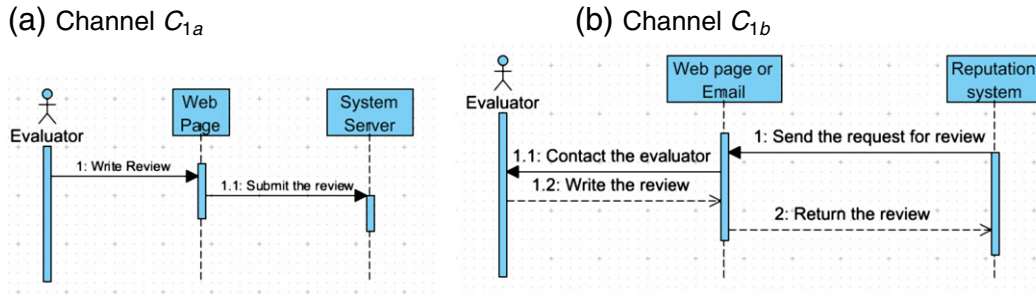## (a) Channel C_{1a}　　　　　## (b) Channel C_{1b}

Fig. 3. Sequence diagram for direct channels.

use collection channels to gather the reputation information provided by information sources.

4.1.1. Collection channel

4.1.1.1. Criterion I1. Collection channel. The collection channel is the method of gathering information from sources. There are two main kinds of channels: direct channels and indirect channels. *Direct channels* refer to those that collect information directly from the evaluators. Some of them passively wait for the evaluators to write reviews (Channel C_{1a}). Other systems choose to invite evaluators via email or web page links (Channel C_{1b}). Fig. 3 illustrates the UML sequence diagrams for the direct channels.

Unlike direct channels, *indirect channels* collect reviews from other reputation systems (Channel C_2). For instance, a number of reputation systems have agreed to allow Google to retrieve their reviews and publish partial or full reviews on the Google Shopping page.

4.1.2. Information sources
Information sources are vital to reputation systems because they provide the information.

4.1.2.1. Criterion I2. Set of evaluators. It is essential for a target to get a sufficient number of ratings/reviews before reputation can reflect its true quality [5,33]. Each target can attract a set of different evaluators. The size of the set (U_e) can be used to estimate the number of reviews that the target can receive (N_{tr}). U_e can be calculated by:

$$U_e = U_q * p_e \tag{4.1}$$

U_q is the number of people who are qualified to leave reputation information, i.e., those who are eligible to be evaluators. Considering that when sending out surveys, only a small number of which will be returned. Similarly, not all qualified evaluators will leave reviews. Thus, p_e denotes the proportion of people who actually provide reviews, which is similar to the response rate in surveys.

Who are qualified to be evaluators depend on the systems' regulations. For example, Amazon allows evaluators to leave reviews on any products, once they have registered with the site and bought one item. However eBay only allows the parties of the transaction (buyers and sellers) to rate each others. U_q can be classified into five sets (Fig. 4).

First, consider systems that limit evaluators to their own site visitors. Although it is very rare, a system may allow people to leave reviews without even registering. In this case, everybody on the Internet who has ever visited the site (U_v) can be an evaluator, which means, $U_q = U_v$ (Set1). A more common case is that the system requires evaluators to register first, i.e., the system accepts all registered users (U_r) to be evaluators, therefore, $U_q = U_r$ (Set2). Moreover, systems may limit their evaluators to an even smaller set with further restrictions (U_t). For example, Reevoo, a product review center, asks

evaluators to provide a proof of purchase before leaving product reviews. Thus, $U_q = U_t$ (Set3).

Second, in addition to collecting reputation information from site visitors, systems may collect reviews from the users of other sites. Some reputation systems work with a number of online shops, which allow the system to collect reputation information directly from their customers after purchases. Therefore, the qualified evaluators are the summation of all the shop's customers: $U_q = \sum_{i=1}^{N_s} U_{s,i}$ (Set4). $U_{s,i}$ is the number of customers of the ith shop and N_s is the number of shops that have cooperated with the reputation system.

As discussed in Section 4.1.1 some systems, such as Google Shopping, collect information from other reputation systems rather than from evaluators (Set5); therefore, the number of evaluators is the summation of the numbers of evaluators of all cooperating systems: $\sum_{i=1}^{N_r} U_{e,i}$ (N_r is the number of cooperating reputation systems and $U_{e,i}$ denotes the number of evaluators of the ith system).

In summary:

$$U_e = \begin{cases} U_v * p_e & \text{Set 1 : all system visitors can be evaluators} \\ U_r * p_e & \text{Set 2 : if only registered users can be evaluators} \\ U_t * p_e & \begin{array}{l}\text{Set 3 : if only people have registered and are} \\ \text{qualified for further restrictions can leave reviews}\end{array} \\ \sum_{i=1}^{N_s} U_{s,i} * p_{e,i} & \text{Set 4 : if systems cooperate with online shops} \\ \sum_{i=1}^{N_r} U_{e,i} & \begin{array}{l}\text{Set 5 : if systems collect information from} \\ \text{other reputation systems.}\end{array} \end{cases}$$

$$\tag{4.2}$$

The proportion of people who actually leave reviews (p_e) can be influenced by many factors, most of which are related to the web site rather than the system, for example, the nature of the targets and the website design. However, the collection channel is considered to have an influence to the p_e, which is because that the evaluators are more likely to leave reviews if they receive a reminder from the system (systems use C_{1b} channel). Moreover, if the system requires evaluators to have direct interactions with the target ($U_q = U_t$), e.g., buyers (evaluators) and sellers (targets), it might have a higher p_e. Research has shown that on eBay, whose $U_q = U_t$, after each transaction, 67–77% buyers or sellers leave ratings for each other [7], whereas on Amazon ($U_q = U_r$), the proportion is much smaller [2].

Most of the time, reputation systems only allow an evaluator leave reputation information to the same target once. Thus, $U_e = N_{tr}$. However, some systems, in particular online auction sites and marketplaces, allow buyers and sellers leave reviews about each other after every transaction. That is to say, evaluators can leave reputation information to the same target repeatedly. It can be imagined that one may take advantage of this policy to increase own ratings rapidly, by exchanging ratings after fake transactions with same person. To avoid this problem, reputation systems usually do not count every rating. For example, in eBay if a seller receives multiple ratings from same buyers within the same week, the seller's reputation score will
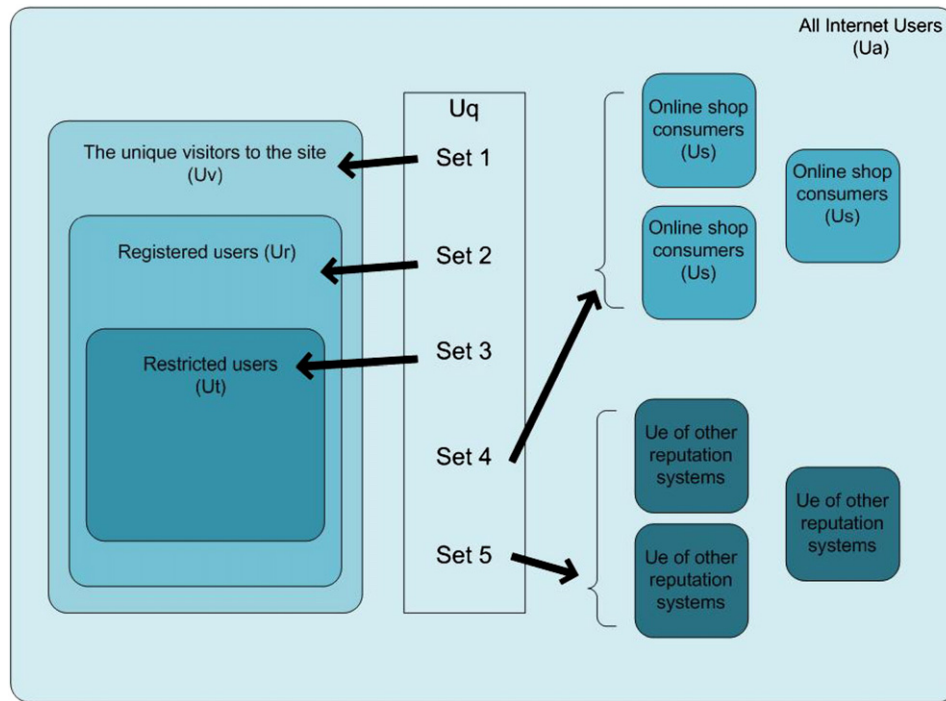
Fig. 4. The five sets of U_q.

only be affected by 1 rating.[1] Therefore, for these systems, $N_{tr} = C * U_e$. C is the factor to calculate multiple ratings.

4.1.2.2. Criterion I3. Granularity. Granularity identifies how evaluators associate with the targets. There are two kinds of granularities between an evaluator and a target.

- *The expertise granularity* refers to the evaluator's level of expertise in the target's area. An individual may enjoy a high reputation for their expertise in one domain while having a low reputation in another [44]. Reputation systems can take different approaches to identify the evaluator's expertise in a specific domain. Some systems choose to illustrate the expertise based on the evaluator's activities within the system. For example, if an evaluator has good reputation on writing reviews for digital cameras, then he/she has a high level expertise granularity with digital cameras. Verifying the evaluator' real world identity can also help increase the expertise granularity, because their off line world reputation then can be transferred to the system. For instance, if a famous chef writes a review on a cookery book, end users may trust his/her review than the other ones.
- *The interaction granularity* indicates whether an evaluator has any direct interactions with the target. The interaction granularity can be identified by the reputation systems. The eBay and Reevoo examples discussed previously showed how reputation systems can control the interaction granularity by regulating the requirements of evaluators.

Based on the analysis, when assessing the level of granularity of a reputation system, it should consider 1) whether the system can present the expertise of the evaluators in a specific domain and 2) whether the system requires evaluators have direct interaction with the targets.

4.1.2.3. Criterion I4. Evaluator credibility. It is important for reputation systems to have the ability of assessing the evaluators credibilities (EC), which can be seen as the reputation of the evaluator. Although some systems use the evaluators credibility to identify the evaluator's expertise granularity (Criterion I3), this criterion is different from it. That is because the expertise granularity only focuses on the credibility in the specific domain, while this criterion concerns the credibility in all domains.

This section concentrates on the EC providers and Criterion *P2* in Section 4.2.2 identifies how the credibilities are calculated. In reputation systems, an evaluator's credibility is associated with the quality and quantity of the reviews they have written. Within the entities that have been discussed in Section 3, three of them can be EC providers: feedback providers, targets and the end users.

1. Feedback providers. Some systems allow end users to give feedback on the reviews. The results of the feedback influence the credibility of evaluators. For example, Amazon lets end users rate the reviews as 'helpful' or 'not helpful'. The evaluator's credibility score will rise with the increase of the 'helpful' votes they received.
2. In C2C marketplaces, where buyers rate sellers, the sellers also have opportunities to rate the buyers, which means that the rating given by the seller (the *target*) influences the credibility of the buyer (the *evaluator*). Then each agent's score can be seen as a reputation score or a credibility score.
3. Reputation systems can also allow end users to rate the evaluators on their credibility directly. For example, end users may rate an evaluator as a trustable evaluator, then the higher score the evaluator gets, the better credibility they have.

It should be noted that some systems have a special ranking mechanism for their users, called the 'Karma' mechanism. It records every activity a user has done within the system, then gives points to it [11]. For example, Yahoo! Answers is a website which allows people to ask and answer questions within the community. Each time users answer a question they will get 2 points. With this Karma mechanism, users have scores. Usually, the higher the score,

the more active they are in the community. Because most sites use Karma mechanisms to identify the behavior of evaluators rather than reflect the credibilities of evaluators; therefore, the paper does not consider it as a credibility mechanism.

4.1.3. Reputation information

4.1.3.1. Criterion I5. Information format. When collecting reputation information, reputation systems usually supply a form for evaluators to fill in (like a survey). It contains different format information, including ratings, text comments or even rich media (photos and videos) formats.

Different information formats have different roles in reputation systems. For example, ratings, which are easily to be aggregated to an overall score, can provide a comparable meaning between targets. Text reviews, however, contain more detailed information.

4.1.3.2. Criterion I6. Information breadth. Information breadth specifies the number of properties that a system collects. The breadth is an important dimension for assessing the completeness of information. More information can illustrate a clearer image of a target. For example, Tripadvisor.com, a travel-related review center, encourages their evaluators to rate hotels for their 'value', 'rooms', 'location', 'cleanness' and 'service' separately.

Although end users may desire more information, too much information may reduce the evaluators' motivation on leaving reviews. Most reputation systems let evaluators choose how much information they want to provide by marking the properties as '*Required*' and '*Optional*'.

4.1.3.3. Criterion I7. Input collection cost. The *input collection cost* refers to how much time it takes to collect a single unit of reputation information (*a unit* of reputation information includes all the information a system needed for the target).

Our previous work has proposed an approach to calculate the collection costs (T_{ip}) [28]. It indicated that collection channels have major impact on the costs. In C_{1a} and C_{1b} systems, reputation information is provided by evaluators, which means, the collection cost is how much time it takes an evaluator to complete the 'review questions'. C_2 systems collect information from other systems, thus, the cost is therefore depended on the indexing speed of the system.

4.1.3.3.1. Collection cost of C_{1a} systems. The collection cost of C_{1a} systems is the time it takes an evaluator from he/she enters the web site till submits the reputation information. Thus it is the time an evaluator needs to browse the web site, from the home page to the page that he/she can write reviews (T_{br}) plus the time it takes him/her to complete the reputation information and submit it to the server (T_{cp}). Hence, $T_{ip,c_{a1}} = T_{br} + T_{cp}$.

It is obvious that T_{br} mainly depends on the design of web sites. Therefore, for practical purposes, all the T_{br} are assumed to be the same. With careful calculations and assumptions, Liu et al. [28] estimated $T_{br} = 67.6$ s.

T_{cp} is dependent to the information format and breadth. For example, it only takes several seconds to give a rating, whereas it can take 10 min or more to write a 100-words text comments. Therefore, T_{cp} can be calculated by:

$$T_{cp} = \sum_{j=1}^{N_{if1}} T_{if1,j} + \sum_{j=1}^{N_{if2}} T_{if2,j} + \sum_{j=1}^{N_{if3}} T_{if3,j}. \tag{4.3}$$

$N_{if1}, N_{if2}, N_{if3}$ denote the number of ratings, text comments and rich media information the reputation system requires. $T_{if1}, T_{if2}, T_{if3}$ are the time for completing the corresponding information respectively. With a number of assumptions, Liu et al. [28] concluded the collection cost of C_{1a} as:

$$T_{ip,c_{1a}} = 67.6 + 1.2 * N_{f1} + 3.16 * \sum_{i=1}^{N_{f2}} W_{pr,i} + \sum_{i=1}^{N_{f3}} T_{ip,3,i}. \tag{4.4}$$

W_{pr} is the number of words of each text reviews. $T_{ip,3}$ denotes the time it takes to make and submit the rich media information.

4.1.3.3.2. Collection cost of C_{1b} systems. When sending invitations to evaluators, C_{1b} systems can provide the link to the exact review page. In this case, there is no T_{br} for C_{1b} systems: $T_{ip,c_{1b}} = T_{cp}$. According to Eqs. 4.3 and 4.4, the collection cost of C_{1b} can be estimated by:

$$T_{ip,c_{1b}} = T_{cp} = 1.2 * N_{f1} + 3.16 * \sum_{i=1}^{N_{f2}} W_{pr,i} + \sum_{i=1}^{N_{f3}} T_{ip,3,i}. \tag{4.5}$$

4.1.3.3.3. Collection cost of C_2 systems. Because C_2 systems collect information from other systems rather than evaluators, their collection cost is much less than the other two kinds of systems. The cost depends on the indexing speed of the system:

$$T_{ip,c_2} = \text{indexing speed}. \tag{4.6}$$

The equations show that for C_{1a} and C_{1b} systems, the information format (Criterion *I5*) and breadth (Criterion *I6*) have a great impact on the costs. Under the current technology conditions, it takes much more time to collect rich media format information than ratings or reviews. However the cost for C_2 systems mainly depends on the performance of their indexing technologies.

4.2. Processing

Processing is a set of activities that transforms the raw information into a more meaningful form. The first three criteria, *target rating algorithms*, *evaluator credibility algorithms* and *feedback aggregation algorithms*, specify the algorithms that reputation systems adopted to calculate the corresponding ratings. The rest criteria, *update frequency*, *robustness*, *algorithm complexity* and *system complexity*, identify the performance of the algorithms.

4.2.1. Criterion P1. Target rating algorithms

This criterion identifies the algorithms that systems use to aggregate the target's ratings. At the moment, most systems choose to use simple algorithms, such as summation, average or percentage. Mean, mode and median are rarely used. However, academic researchers have proposed many complex aggregating algorithms, such as, Bayesian systems [43,18] and fuzzy models [36]. A number of papers have reviewed and compared those academic algorithms [19,31,37].

4.2.2. Criterion P2. Evaluator credibility algorithms

This criterion specifies how the evaluator credibilities are aggregated. Similar to the main rating algorithms, only simple algorithms are used, e.g., average and summation.

4.2.3. Criterion P3. Feedback aggregation algorithms

This criterion identifies how feedback ratings are aggregated. Summation and average are the two most common algorithms used for feedback calculation.
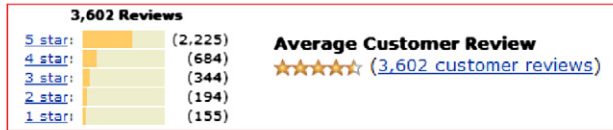
4.2.4. Criterion P4. Update frequency

Update frequency refers to how often a system updates their reputation ratings and other relevant information. In other words, it assesses how often the algorithms run. Some systems update their ratings as soon as a review has been submitted, while others choose to update information on a daily, weekly or even longer basis.

4.2.5. Criterion P5. Algorithm robustness

Algorithm robustness is used to assess the robustness of the three algorithms. According to basic statistical analysis, one simple way to measure the robustness is to check the *breakdown point* of an

Customer Reviews

Fig. 5. Output—aggregated and individual information.

algorithm. The breakdown point (ε) is the proportion of manipulated ratings required to make the algorithm return an arbitrary value. A higher breakdown point indicates a more robust algorithm [25]. For example, assume R is the set of ratings of a target, $R \in \{r_1, r_2, ..., r_n\}$. The overall rating of the target is R_t. If someone wants to change the value of R_t, he/she needs to add a number (m) of manipulated ratings to R. Then, $\varepsilon = \frac{m}{n}$. Therefore, the larger the n, the more robust the algorithm is.

If an algorithm is robust, the new ratings may not easily change the results of the overall rating. In other words, a robust algorithm is not sensitive to the change of new ratings. Reputation systems need to find a balance between the robustness and sensitiveness of the algorithms.

4.2.6. Criterion P6. Algorithm complexity

The *algorithm complexity* refers to the complexity of each algorithm, which relates to the analysis of algorithms [24]. As discussed earlier, most centralized reputation systems use very simple algorithms to calculate the ratings, and under the development of current computing technologies, they all have relatively low complexities.

4.2.7. Criterion P7. System complexity

The *system complexity* is the complexity of the whole system, which is determined by the features a system provides. If a system provides many complicate features, it is then more complex than the system that provides less features.

4.3. Output

In information systems, *output* is the production of useful information, usually in the form of documents and reports [38]. The output of reputation systems is to report and disseminate the reputation information. Therefore the evaluation of the Output component should focus on the dissemination and report of the information.

There are two main kinds of information that a reputation system needs to report: *aggregated information* and *individual information*. The former shows the results of the Processing component, such as the overall rating. The latter presents the individual ratings and reviews that are collected through the Input component (Fig. 5 shows the screen shots of the aggregated and individual information from Amazon.com).

4.3.1. Dissemination

Dissemination measures who can retrieve reputation information and how they can access it. The two criteria are:

4.3.1.1. Criterion O1. The set of end users.
The set of end users refers to who are the end users, i.e., who can retrieve the reputation information. Most websites allow all their site visitors (U_v) to access their published reputation information. Some systems may require end users to register with them (U_r). A few sites reserve some information for restricted users (U_t). For example, IMDb, the Internet Movie Database, shows the Top 250 movies to all Internet users, but offers the Top 500 to their IMDbPro users, who have paid subscription fees.

4.3.1.2. Criterion O2. Access methods.
Unlike distributed systems, all centralized reputation systems publish information on their websites. Therefore, this criterion focuses on whether the system supplies alternative ways for their users to get information. Some systems can send emails to users when a new review has been left for targets of interest. Moreover, systems may provide an RSS (really simple syndication) feed for users to track new reviews of a target or new reviews submitted by a specific evaluator.

It can be imagined that in the future, with the Internet technology development, more access methods will emerge. In fact, with the booming of social networking sites, such as Facebook and Twitter, reputation systems can take advantages of these services to help end users to retrieve information.

4.3.2. Report—aggregated information

The aim of providing the aggregated reputation information is to present the target's overall reputation in a concise and comparable format.

4.3.2.1. Criterion O3. Timeliness.
Sometimes the target's quality may change over time, for example, a hotel may provide better room services than it used to. Therefore, it is important for reputation systems to be able to present the target's overall ratings in different time periods. For example, eBay presents the seller's overall ratings during the last month, 6-months and 12-months.

4.3.2.2. Criterion O4. Descriptive dimensions.
This criterion specifies how many dimensions a system uses to illustrate the target's aggregated reputation information. The aim of the aggregated information is to illustrate the majority evaluators' opinions on the target, which have many different ways to approach. For example, arithmetic

mean, mode and median can all be used to show the central tendency of a set of ratings. End users are considered to have different needs for different measurements. Therefore it is essential that reputation systems can provide multiple dimensions when presenting the aggregated information. For example, IMDb not only shows the arithmetic mean as the overall rating of a target but also shows the median.

4.3.3. Report—individual information

Individual information is the information that provided by each evaluator. It is necessary for the systems to present the raw information as it collected. In addition, reputation systems also need to provide more information on evaluators and the feedback of reviews.

4.3.3.1. Criterion O5. Information filtering.
As noted in Section 4.1.2, reputation systems need a sufficient number of reviews to represent the true reputation of the target. However when more and more reviews come out, information overload can occur. Therefore, reputation systems need filtering and sorting abilities to help end users to retrieve their desired information more effectively. For example, most systems allow end users to sort or filter reviews by the rating scores or the date it has been left.

4.3.3.2. Criterion O6. Evaluator information.
In the real world, a person's identity and their personal character can affect trust [32]. Therefore, reputation systems must provide information about their evaluators such as their rating histories, their credibility (if possible) and even their real names.

4.3.3.3. Criterion O7. Feedback loop information.
This criterion identifies how the Feedback Loop results are presented. When presenting the results, reputation systems can provide the full results (e.g., the number of both helpful and unhelpful votes) or merely the number of helpful votes.

4.3.4. Response time

When evaluating web sites, response time is always a vital factor. It assesses how quickly the systems react to the users inquiries. However as discussed earlier, reputation systems do not exist in isolation. They are integrated into business applications' websites. Thus, it is difficult to measure the reputation systems response time without discussing their websites features. Therefore, we do not list the response time as a criterion in the analysis model.

4.4. Feedback loop

The quality of the review determines whether the reputation systems can work properly. One of the best ways to control the quality is to let users assess the reviews. In other words, a feedback loop works as a simple version reputation system in which the targets are the reviews. Thus, a feedback loop can also be divided into 'input, processing and output'.

- The input of the feedback loop is the collection of feedback information. This part can be evaluated in a similar fashion as the Input component.
- The processing of the feedback has two meanings: calculation algorithms, which were defined as Criterion P3 in Section 4.2.3, and the function (or roles) of the feedback loop.
- The output of feedback is to publish the feedback results, which were measured in the Output component.

Therefore, the criteria of the feedback loop can be grouped into the feedback function and the feedback collection.

4.4.1. Feedback function

4.4.1.1. Criterion F1. Feedback loop function.
The feedback loop function refers to the roles of the feedback loop. The aim of the feedback loop is

to assess the quality of the reviews. In other words, the major role of feedback loop is to detect review spams. There are two main kinds of review spams: *Untruthful reviews*, which are the reviews do not reflect the true opinion of the evaluators and *Non-reviews*, the content of the reviews are not related to the targets at all, such as advertisements [17,16,1]. Thus, the functions of feedback loop are:

1. Reputation systems allow users to rate the reviews as 'helpful' or 'not helpful' to identify the untruthful reviews. Systems can also use the results of this kind of feedback to rank the reviews, so that end users will see the most helpful review first.
2. Non-reviews can be deterred by allowing users to 'report' or 'flag' them. It is understandable that a reputation system may receive a lot of improper information, such as advertisements. Some systems adopt a time-consuming approval mechanism which let editors check each review before they can be published. In contrast, most reputation systems publish the reviews without editor-checking, then allow end users to report the non-reviews. Thus, the systems only need to deal with the reported information, which has a much less amount.

In addition to identifying the review spams, the feedback loop can also be used to provide more information.

3. Usually, there is no need to worry about whether a single review covers all the aspects of the target, as long as there are sufficient reviews. However in some special cases, such as in eBay, when a buyer leaves a negative review on the seller, it would be unfair if the seller does not have the opportunity to provide information from their perspectives. Therefore, the third function of the feedback loop is to provide more information.

4.4.2. Feedback collection

The collection of feedback is much simpler than that of Input. Nearly all feedback is collected through web pages directly (C_{1a}). In addition most systems allow for a wider range of feedback providers to leave simple format feedback (most are ratings). For practical purposes, this paper does not discuss the collection channel, the granularity or the evaluator credibility of the feedback loop as most systems have similar performance on these criteria.

4.4.2.1. Criterion F2. The set of feedback provider.
This criterion defines the set of the feedback providers. Two kinds of people can be feedback providers: end users and 'targets'. End users read the reviews and then leave their feedback as 'helpful' or 'not helpful'. In C2C marketplaces, the buyers and sellers are allowed to rate each other, if we take one party as the evaluator (say, the buyer), which means, the other party (the seller) is the target. Therefore, the rating made by the target (the seller) can be seen as the feedback to the buyer.

Similar to Criterion I2, which was discussed in Section 4.1.2, this criterion can be assessed by the number of feedback providers (U_{fe}).

$$U_{fe} \begin{cases} U_v * p_{fe} & \text{Set 1 : if all system visitors can be feedback evaluators} \\ U_r * p_{fe} & \text{Set 2 : if only registered users can leave feedback} \\ U_t * p_{fe} & \text{Set 3 : if only people have registered with the site} \\ & \text{and are qualified for further restrictions can leave feedback} \\ 1 & \text{if systems only allow the target to leave feedback.} \end{cases}$$

(4.7)

p_{fe} denotes the proportion of people who actually leave feedback. Some C2C sites, where only the targets (buyers/sellers) are allowed to leave feedback, the number of feedback providers is 1. As noted earlier, feedback is only collected through the system web pages; therefore the Sets 4 and 5 in Criterion I2 are not applicable.

4.4.2.2. Criterion F3. Feedback format and breadth.
Most systems only require people to leave ratings (votes for 'helpful' and 'not helpful'
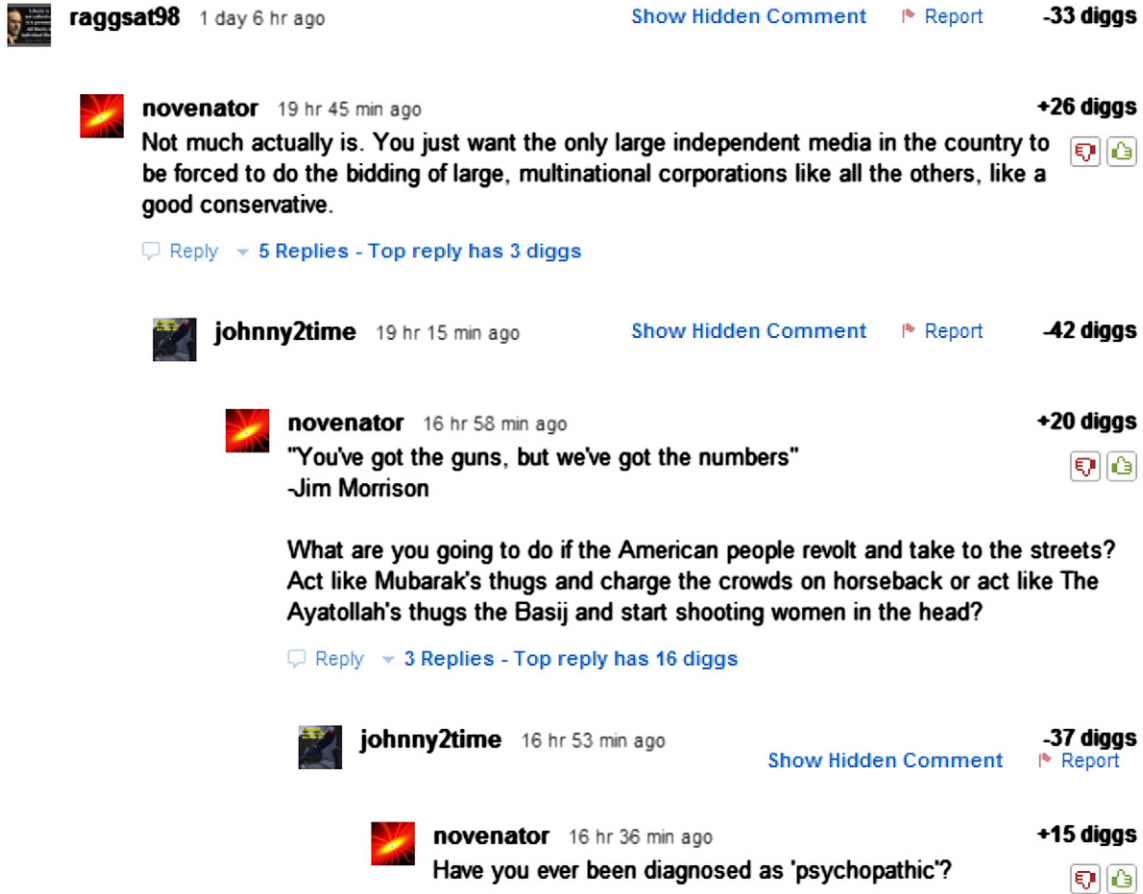
Fig. 6. Digg's multi-feedback level.

can be seen as a special kind of ratings) and text comments as feedback. The breadth of the feedback refers to the number of ratings and text comments.

4.4.2.3. Criterion F4. Feedback loop level. Sometimes reputation systems allow people to reply to the feedback, in other words, they can leave multiple level feedback. Although most systems only accept one level of feedback, Digg allows people to reply to comments on multiple levels. Fig. 6 is a snapshot of Digg's comments, which shows it allows 5 levels. The first feedback was the reply made by 'novenator' to the comment of 'raggsat98'. Then 'johnny2time' and 'novenator' kept replying to each other's comments for 2 more levels. More feedback levels can provide users more opportunities to discuss the details of the targets.

4.4.3. Feedback loop cost

4.4.3.1. Criterion F5. Feedback loop collection costs. The collection costs of feedback loops (T_{fd}), which is similar to the input collection costs (Criterion I7), refers to the time it takes the provider to leave a feedback. Therefore, T_{fd} can be calculated similarly to the input collection costs (T_{ip}). As there is only one collection channel (C_{1a}) for feedback loop, thus,

$$T_{fd} = 67.6 + 1.2 * N_{ff1} + 3.16 * \sum_{i=1}^{N_{ff2}} W_{pf,i}. \qquad (4.8)$$

N_{ff1} and N_{ff2} denote the number of ratings and text comments respectively and W_{pf} is the number of words of the text feedback.

4.5. Storage

The storage stores all the information that has been collected or generated by other components. As discussed in Section 1, reputation systems can store the reputation information at either centralized or distributed locations. Because this paper only concentrates on the centralized systems, there is no need to define the storage location.

Other possible storage measurements, such as data storage speed and capacity, are usually associated with the hardware and software that have been selected by the website. Again, these are not included in this paper. Therefore, the evaluation of storage focuses on the storage costs only. There are three kinds of information that need to be stored: information collected from input, information collected from the feedback loop and the information generated by processing. All three costs can be measured by the data size. Due to the limited space, this paper only discusses the major influential factors of the storages. Liu [27] has provided a more detailed discussion on the assessment of the storage costs.

4.5.1. Criterion S1. Input storage cost

The size of the data is related to the format of the reputation information (*IF*). For example, one rating, which is usually represented by a number, only occupies one byte of storage, whereas a 100-word text review may require more than 600 bytes. A picture or video needs considerably more storage space. The breadth of collected

information (IB), also has an influence on the size of input information. Another factor that can impact the storage cost is the number of total reviews, which is related to the number of evaluators (U_e). Therefore, the storage cost of input information (S_{ip}) is,

$$S_{ip} = f(IF, IB, U_e) \qquad (4.9)$$

4.5.2. Criterion S2. Feedback storage cost

Similar to S_{ip}, the feedback format (FF), breadth (FB) and the number of feedback providers (U_{fe}) also have a great impact on the total size of the feedback (S_{fd}). In addition, the level of the feedback loop (L) is another factor that can be used to calculate the feedback costs.

$$S_{fd} = f\left(FF, FB, U_{fe}, L\right) \qquad (4.10)$$

4.5.3. Criterion S3. Processing information storage cost

The *processing information* refers to the information that is generated from the processing calculations. Usually reputation systems do not make their algorithms or processing public, therefore, the best way to measure the processing information is to assess the aggregated information of the output. It is assumed that all generated information will be published through the output.

Therefore, the storage of processing information (S_p) is decided by the format of aggregated information (AF), the breadth of each format (AB) and the number of total targets (N_{ta}):

$$S_p = f(AF, AB, N_{ta}). \qquad (4.11)$$

4.6. Summary

In total 29 criteria have been defined on the basis of five components. There are 7 criteria for the input, 7 for the processing, 7 for the output, 5 for the feedback loop and 3 for the storage. Within these criteria, some can be used to measure the performance of reputation systems. For example, if a system provides more descriptive dimensions, it is better than the one provides less dimensions. However, some criteria can classify the different types of reputation systems rather than to compare their performance. Take collection channel (Criterion I1) as an example. It cannot be simply said that a system using C_2 is better than the one using C_{1a}.

In addition, within the measurement criteria, some of them refer to the performance of the systems, while the others are focused on the costs. One of the aims of this research is to discuss the cost of reputation systems, which has been long ignored from research. Thus, the research separates the costs criteria out of the measurement criteria.

Therefore, the criteria can be grouped into: *classification criteria*, which are used to classify different types of reputation systems; *measurement criteria* measure the performance of systems; and *cost criteria* assess the costs of systems. Table 1 shows the classification of the defined criteria.

5. Validation

One possible way to validate the analysis model is to compare it with other similar models. However as discussed in Section 2, very few research has proposed similar models which focusing on centralized reputation systems.

Reputation systems essentially are information systems (IS) that are working in the Internet environment. As Delone and Mclean [9] pointed out that the measurements for information systems should not change with an online environment. Thus, the results of comparing the criteria defined in this paper with those of IS can show the

Table 1
Table of criteria.

	Classification criteria	Measurement criteria	Cost criteria
I	I1. Collection channel I5. Information format	I2. Set of evaluators I3. Granularity I4. Evaluator credibility I6. Information breadth	I7. Input collection cost
P	P1. Target rating algorithm P2. Evaluator credibility algorithm P3. Feedback aggregation algorithm	P4. Algorithm robustness P5. Update frequency	P6. Algorithm complexity P7. System complexity
O	O2. Access method	O1. Set of end users O3. Timeliness O4. Descriptive dimensions O5. information filtering O6. Evaluator information O7. Feedback information	
F	F1. Feedback loop function	F2. Set of feedback providers F3. Feedback format and breadth F4. Feedback loop level	F5. Feedback collection cost
S			S1. Input storage cost S2. Feedback storage cost S3. Processing information storage cost

completeness and validity of the model from the information system perspective.

Delone and Mclean [8] surveyed a large number of evaluation papers in the area of IS, and they classified the measures of IS into six categories: system quality, information quality, service quality, intention to use, user satisfaction and net benefits. The first three categories are the foundation of the latter ones. As this paper only focuses on analyzing the intrinsic nature of the reputation system (rather than the whole e-commerce website), only the system quality and information quality measures are relevant. The key dimensions of system quality are: *usefulness, usability, responsiveness, reliability and flexibility* [9], and the dimensions of information quality are: *accuracy, relevance, understandability, completeness, currency, content personalization and variety of information* [42].

5.1. System quality

Usefulness of reputation systems refers to whether the reputation information can reflect the true quality of the targets. It depends on 1) a sufficient number of evaluators (Criterion I2), 2) the granularity (Criterion I3) and credibility of the evaluators (Criterion I4). *Flexibility* refers to whether it provides any customized information or function for users, which can be assessed by Criterion O5 (information filtering).

The *reliability* of reputation systems depends on whether the reputation system can function well. The main factor that decides the success of reputation systems are whether reviews processed and presented in the most meaningful format [10,33]. Criterion O4 (descriptive dimensions) is defined to assess how the aggregated information is presented to the end users. In addition, review spam can also reduce the reliability of the reputation system. By making good use of the feedback loop component, reputation systems can reduce the number of review spams. The feedback function (Criterion F1), which specifies how the system uses the feedback loop, conforms to the reliability as well.

System *responsiveness* and *usability* mainly depend on the performance of the websites, rather than on the performance of the reputation systems. As this paper only focused on the intrinsic nature of

reputation systems, there is no criterion relates to these two dimensions.

5.2. Information quality

The information quality in reputation systems refers to the quality of the ratings, reviews and feedback. The *accuracy* of the reputation information is controlled by the evaluators, the processing and the feedback loop. Thus, Criteria I2 (the set of evaluators), P1 (target rating algorithms), O4 (descriptive dimensions) and F1 (feedback function) match with the accuracy. The Criterion I3 (granularity) conforms to the *relevance*.

Whether the end users can understand the reputation information depends on how the individual ratings are aggregated and presented. If a system uses simple rating algorithms, such as summation or average, it is not difficult for end users to understand the meaning of the aggregated information. Therefore Criteria P1, P2 and P3, which measure all the relevant algorithms in reputation systems and O4, descriptive dimension, which describes how aggregated information is presented reflect the *understandability* of reputation systems.

The input information breadth (I6), feedback breadth (F3) and most criteria regarding the output component (O4, O6 and O7) are focused on how much information is collected and presented to the end users, which reflects the *completeness* of the reputation information. *Currency* refers to when the information is collected and whether it is up-to-date. Reputation systems usually display the individual reviews with details about the submitted time. Furthermore, some systems present the aggregated ratings with different time periods, such as ratings within 3 months or all time. This feature is measured by Timeliness (Criterion O3). Update Frequency (Criterion P5) also matches with the currency, as it describes how often the system updates the information. Criterion O5 assesses whether the system provides any features for end users to filter or sort the individual reviews in the way they prefer. It is well suited with the *content personalization* dimension. *Variety* of information is related to the different formats and descriptive dimensions of the systems that are presented to the users, which have been defined with I5 (input format), F3 (feedback format) and O4 (descriptive dimensions).

5.3. Summary

Table 2 presents the relationship of the defined criteria with the system quality (SQ) and information quality (IQ). It showed that the criteria defined in this paper can cover most system quality and information quality dimensions. System usability and responsiveness are excluded due to the paper focuses on analyzing reputation system itself rather than the whole site which utilizes the system.

It can be found that some criteria in the analysis model are not mapped with Delone and McLean [8]'s model, such as Criteria I1, P4, P6, O1, O2, F2, F4 and all the cost criteria. This is because the Delone and McLean [8]'s model focused on the business information systems, while our model concentrates on the reputation systems. Most of these non-matched criteria are defined based on reputation systems'

distinguished characteristics. For example, Criterion I1, which describes the collection channel is not applicable with traditional information systems. Furthermore, Delone and McLean [8] did not have much discussion on the costs of systems; thus, none of the cost criterion is mapped with their model. However, it should be noted that costs do have influences on the performance of the systems. For instance, the higher collection costs (Criterion I7) may obstruct the evaluator's willingness of leaving reviews, which will cause insufficient number of evaluators (Criterion I2) and affect the accuracy of the reputation information.

6. Conclusion

6.1. Contributions

The first contribution of the paper is that it defined a number of terms that have been widely used by researchers in the area. Furthermore, the paper proposed an analysis model for measuring different centralized reputation systems under the same context. Reputation systems have the same five underlying components: input, processing, output, feedback loop and storage. Therefore, each component can be measured based on its characteristics. Input criteria are defined based on the collection channels, reputation information and information sources. Processing focuses on the calculation algorithms and the efficiency of them. The output component presents the information to the end users. Consequently the defined criteria focus on the system's information interpretability. The criteria concerning the feedback loop concentrate on the collection and function of the feedback. Criteria for the storage component measure the storage costs.

By comparing the model with other information system measurement dimensions, the paper showed that the defined criteria have well covered most system quality and information quality dimensions, which are the most important technical factors of information systems. Furthermore, the paper has defined criteria for assessing the cost of systems, which has been long ignored from previous research.

The analysis model can also be used as a measurement when addressing or tackling the problems of the reputation systems. For example, Criterion P1 (target rating algorithm) can be used to identify the algorithms and Criterion O5 is designated to measure the system's information filtering ability.

The proposed model is believed to have very good extendability and flexibility. For example, although the model was built to analyze different types of reputation systems, with selected criteria, the model can make good measurements for any specific type of system. Moreover, researchers from different disciplines can select or refine the criteria and quantifications to fit their own research needs.

6.2. Limitations and future work

As stated in previous sections, this paper aimed at identifying criteria at theoretical level. Although, the paper does have discussions on the possible quantification and measurements of the criteria, more analysis is still in need. Thus, in the future we will consider to analyze each criterion with more details and discuss more deeply on how to quantify the criteria, so that they can be used to evaluate reputation systems directly. Furthermore, future research will focus on analyzing and testifying the correlation between the criteria. Based on that, specific models, which integrated the costs with the performance criteria, can be proposed for each different type of systems. In addition, although to our best knowledge our model is by far the most comprehensive analysis model of reputation systems, it still can be extended with criteria concerning not only reputation system factors but also the websites performance features, such as responsiveness and usability.

Table 2
SQ and IQ factors and defined criteria.

SQ measures	Relevant criteria	IQ dimensions	Relevant criteria
Reliability	O4, F1	Accuracy	I2, P1, O4, F1
Responsiveness	N/A	Currency	P5, O3
Usability	N/A	Completeness	I6, O4, O6, O7, F3
Usefulness	I2, I3, I4	Personalization	O5
Flexibility	O5	Relevance	I3
		Understandability	P1, P2, P3, O4
		Variety of information	I5, O4, F3

References

[1] C.C. Chen, Y.D. Tseng, Quality evaluation of product reviews using an information quality framework, Decision Support Systems 50 (2010) 755–768.

[2] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: online book reviews, Journal of Marketing Research (JMR) 43 (2006) 345–354.

[3] S. David, T. Pinch, Six degrees of reputation: the use and abuse of online review and recommendation systems, Social Science Research Network Working Paper Series, 2005.

[4] C. Dellarocas, Analyzing the economic efficiency of ebay-like online reputation reporting mechanisms, The 3rd ACM Conference on Electronic Commerce (EC'01), ACM Press, 2001, pp. 171–179.

[5] C. Dellarocas, The digitization of word of mouth: promise and challenges of online feedback mechanisms, Management Science 49 (2003) 1407–1424.

[6] C. Dellarocas, The many faces of reputation: towards a discipline of web 2.0 reputation system design, Presented at the International Conference of Reputation—Theory and Technology, 2009.

[7] C. Dellarocas, M.N. Fan, C.A. Wood, *Self-interest, reciprocity, and participation in online reputation systems,* MIT Sloan Working Papers No. 4500–04, 2004.

[8] W.H. Delone, E.R. McLean, The Delone and Mclean model of information systems success: a ten-year update, Journal of Management Information Systems 19 (2003) 9–30.

[9] W.H. Delone, E.R. Mclean, Measuring e-commerce success: applying the Delone & Mclean information systems success model, International Journal of Electronic Commerce 9 (2004) 31–47.

[10] M. Fan, Y. Tan, A.B. Whinston, Evaluation and design of online cooperative feedback mechanisms for reputation management, IEEE Transactions on Knowledge and Data Engineering 17 (2005) 244–254.

[11] F.R. Farmer, B. Glass, Building Web Reputation Systems, O'Reilly and Yahoo! Press, 2010.

[12] E. Friedman, P. Resnick, R. Sami, Manipulation-resistant reputation systems, in: N. Nisan, T. Roughgarden, E. Tardos, V.V. Vazirani (Eds.), Algorithmic Game Theory, Cambridge University Press, 2007, pp. 677–697.

[13] A. Gutowska, On desideratum for B2C e-commerce reputation systems, Journal of Computer Science and Technology 43 (2009) 57–832.

[14] K. Hoffman, D. Zage, C. Nita-Rotaru, A survey of attack and defense techniques for reputation systems, ACM Computing Surveys 42 (2009) 1–31.

[15] D. Houser, J. Wooders, Reputation in auctions: theory, and evidence from ebay, Journal of Economics & Management Strategy 15 (2006) 353–369.

[16] N. Hu, L. Liu, V. Sambamurthy, Fraud detection in online consumer reviews, Decision Support Systems 50 (2010) 614–626.

[17] N. Jindal, B. Liu, *Opinion spam and analysis*, in: Proceedings of the international conference on Web search and web data mining, WSDM'08, 2008, pp. 219–230.

[18] A. Jøsang, R. Ismail, *The beta reputation system*, in: the 15th Bled Conference on Electronic Commerce, 2002, pp. 324–337.

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

[20] N. Kumar, I. Benbasat, The influence of recommendations and consumer reviews on evaluations of websites, Information Systems Research 17 (2006) 425–439.

[21] C. Lampe, P. Resnick, *Slash(dot) and burn: distributed moderation in a large online conversation space*, in: CHI'04: the SIGCHI conference on Human factors in computing systems, 2004, pp. 543–550.

[22] J. Laudon, K. Laudon, Management Information Systems: Managing the Digital Firm & Multimedia, 10th edition Prentice Hall, 2007.

[23] K. Lerman, *Social networks and social information filtering on digg*, unpublished, 2006.

[24] A. Levitin, Introduction to the Design and Analysis of Algorithms, 2nd edition Addison Wesley, 2001.

[25] M.S. Lewis-beck (Ed.), Basic Statistics, SAGE Publications, 1993.

[26] Z. Liang, W. Shi, Performance evaluation of rating aggregation algorithms in reputation systems, International Conference on Collaborative Computing: Networking, Applications and Worksharing, IEEE Computer Society, 2005, p. 10.

[27] L. Liu, *Systematic Measurement of Centralized Online Reputation Systems*, Ph.D. thesis, Durham University, UK, 2011.

[28] L. Liu, M. Munro, W. Song, *Evaluation of collecting reviews in centralized online reputation systems*, in: 6th International Conference on Web Information Systems and Technologies (WEBIST), 2010, pp. 281–286.

[29] D. Lucking-Reiley, D. Bryan, N. Prasad, D. Reeves, Pennies from ebay: the determinants of price in online auctions, Journal of Industrial Economics 55 (2007) 223–233.

[30] R.A. Malaga, Web-based reputation management systems: problems and suggested solutions, Electronic Commerce Research 1 (2001) 403–417.

[31] L. Mui, M. Mohtashemi, A. Halberstadt, *A computational model of trust and reputation*, in: the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 7, 2002, p. 188.

[32] H. Nissenbaum, Securing trust online: wisdom or oxymoron? Boston University Law Review 81 (2001) 635–664.

[33] P. Resnick, K. Kuwabara, R. Zeckhauser, E. Friedman, Reputation systems, Communications of the ACM 43 (2000) 45–48.

[34] P. Resnick, R. Zeckhauser, J. Swanson, K. Lockwood, The value of reputation on ebay: a controlled experiment, Experimental Economics 9 (2006) 79–101.

[35] S. Ruohomaa, L. Kutvonen, E. Koutrouli, *Reputation management survey*, in: ARES'07: The Second International Conference on Availability, Reliability and Security, 2007, pp. 103–111.

[36] J. Sabater, C. Sierra, *Regret: reputation in gregarious societies*, in: AGENTS'01: the fifth international conference on Autonomous agents, 2001, pp. 194–195.

[37] J. Sabater, C. Sierra, Review on computational trust and reputation models, Artificial Intelligence Review 24 (2005) 33–60.

[38] R.M. Stair, G. Reynolds, G.W. Reynolds, Principles of Information System, 9 edition Course Technology, 2010.

[39] J. Surowiecki, The Wisdom of Crowds, Anchor, 2005.

[40] G. Swamynathan, K.C. Almeroth, B.Y. Zhao, The design of a reliable reputation system, Electronic Commerce Research 10 (2010) 239–270.

[41] S. Utz, U. Matzat, C. Snijders, On-line reputation systems: the effects of feedback comments and reactions on building and rebuilding trust in on-line auctions, International Journal of Electronic Commerce 13 (2009) 95–118.

[42] R.W. Wang, D.M. Strong, Beyond accuracy: what data quality means to data consumers, Journal of Management Information Systems 12 (1996) 5–33.

[43] A. Whitby, A. Jøsang, J. Indulska, *Filtering out unfair ratings in Bayesian reputation systems*, in: the 7th International Workshop on Trust in Agent Societies, 2004, pp. 106–117.

[44] G. Zacharia, P. Maes, Trust management through reputation mechanisms, Applied Artificial Intelligence 14 (2000) 881–907.

[45] W. Zheng, L. Jin, *Online reputation systems in web 2.0 era*, in: Americas Conference on Information Systems (AMCIS) Proceedings, 2009, pp. 296–306.

Ling Liu received her PhD from School of Engineering and Computing Sciences, Durham University. She also received her Msc with Distinction in Internet Systems and E-business from Department of Computer Science, Durham University. Her current research interests include reputation systems, social networking and information systems.

Professor Malcolm Munro is emeritus professor of Software Engineering in the School of Engineering and Computing Sciences at Durham University. His main research focus is in the areas of Software Visualization, Software Maintenance and Evolution. The concern of the research is to establish how Legacy Systems evolve over time and to discover representations (visualizations) of those systems to enable better understanding of change. He has also been involved with research in Web Services, protocols for fair exchange of electronic goods, Phishing, and Reputation Systems.