Bus Inf Syst Eng
DOI 10.1007/s12599-017-0493-1

RESEARCH PAPER

Interactive Reputation Systems

How to Cope with Malicious Behavior in Feedback Mechanisms

Johannes Sänger · Günther Pernul

Received: 19 February 2016 / Accepted: 2 February 2017

Abstract Early reputation systems use simple computation metrics that can easily be manipulated by malicious actors. Advanced computation models that mitigate their weaknesses, however, are non-transparent to the end-users thus lowering their understandability and the users' trust towards the reputation system. The paper proposes the concept of interactive reputation systems that combine the cognitive capabilities of the user with the advantages of robust metrics while preserving the system's transparency. Results of the evaluation show that interactive reputation systems increase both the users' detection ability (robustness) and understanding of malicious behavior while avoiding trade-offs in usability.

Keywords Trust · reputation systems · Information security · Visual analytics

1 Introduction

Reputation systems are an essential component of various online platforms such as electronic marketplaces, e-commerce websites, or file-sharing systems. Since these environments usually involve a huge number of strangers who have not interacted before, users do not have any information about e.g. the trustworthiness of a seller or the quality of a product – *ex ante*. To solve this problem, reputation systems encourage users to provide feedback on past experiences. They then compute one or several reputation values and display the output in an actor's or item's feedback profile. In this way, users can establish trust based on the experiences made by others (Resnick et al. 2000; Awad and Ragowsky 2008).

In the recent years, this simple concept has even generated websites that particularly focus on collecting reviews for different use cases such as tripadvisor[1] or yelp[2]. The success of these websites is hardly surprising, as researchers have been able to provide empirical evidence that reputation systems manage to reduce information asymmetry and increase a market's efficiency (Dellarocas 2001; Yamagishi and Matsuda 2002). On eBay, sellers with better reputation even have an increased number of sales and obtain higher prices (Diekmann et al. 2013). While the concomitant economic value of "good" reputation encourages trustworthy participation, it also offers an enticement for malicious actors to exploit the weaknesses of reputation systems in order to appear more reputable than appropriate. For that reason, the robustness[3] of reputation systems has become a focal issue in the trust management research community (Jøsang 2012).

Since most early reputation systems could easily be manipulated, researchers have started to introduce a large number of computation models that are robust against a variety of different attacks. In most cases, these models are based on advanced mathematics and have thus become

Accepted after two revisions by Prof. Dr. Jarke.

Dr. J. Sänger (✉) · Prof. Dr. G. Pernul
Department of Information Systems, University of Regensburg,
Universitätsstraße 31, 93053 Regensburg, Germany
e-mail: johannes.saenger@wiwi.uni-regensburg.de

Prof. Dr. G. Pernul
e-mail: guenther.pernul@wiwi.uni-regensburg.de

[1] http://www.tripadvisor.com.

[2] http://www.yelp.com.

[3] A robust system always provides a higher reputation value for an honest actor compared to a malicious actor who tries to exploit the system.

Published online: 22 August 2017

quite complex (Marsh et al. 2012). High complexity and resulting non-transparency, however, lower the users' trust towards the reputation systems (Hammer et al. 2013). In practice, in contrast, this trend to more robust computation models have not taken effect. Common electronic marketplaces such as eBay[4] or e-commerce websites like Amazon[5] still use a very simple and easy to understand metric. Such reputation systems, in turn, have been shown to be very weak against several attacks.

To close the gap between robust metrics introduced in literature and transparent and easy to understand user interfaces mostly used in practice, we propose the concept of *interactive reputation systems*. Involving the user in reputation assessment, interactive reputation systems aim at providing enhanced insights through combining the cognitive capabilities and the experience of the users with the advantages of robust reputation metrics. While there has been considerable work on certain building blocks of interactive reputation systems [(e.g. modularizing and configuring the computation (Hillebrand and Coetzee 2015; Sänger et al. 2015c) or displaying reputation data in interactive visualizations (Sänger et al. 2015b)], the work presented in this paper is the first to the best of our knowledge that connects the distinct building blocks to introduce the complete and integrated concept of interactive reputation systems.

The remainder of this paper is based on the design science research paradigm (Hevner et al. 2004) following the phases introduced by Peffers et al. (2007) (problem identification and motivation, objective of the solution, design and development, demonstration, evaluation, communication): first, we describe the background of online trust and reputation systems as well as robust metrics and transparent presentation in Sect. 2 (problem identification and motivation). Based on that, we point out the research gap and define the objectives of our work (objective of the solution). In Sect. 3 we introduce the concept of interactive reputation systems. Here, we propose the process model and describe how the single building blocks can be integrated (design and development). To demonstrate how interactive reputation systems could look in practice we present a software prototype (demonstration). In order to evaluate the quality of our solution, we conducted several experiments as well as an international user study that are described in Sect. 4 (evaluation). Results show that through involving the user in reputation assessment, attacks on reputation systems can be reliably detected while maintaining the system's transparency. In this way, the users' understanding of malicious behavior can be enhanced. Finally we sum up our contribution and discuss the implications of our findings in the Sect. 5 (communication).

2 Background and Objectives

2.1 Online Trust and Reputation Systems

The concept of trust has been studied in research for many years. While the emergence of the Internet and e-commerce is a quite recent phenomenon, the research on trust spans a wide area of further disciplines such as sociology, psychology or economics (Riegelsberger et al. 2005). Consequently, there are just as many definitions and researchers even disagree on basic characteristics (McKnight and Chervany 1996; Gefen et al. 2003). In this work, we refer to the definition of *reliability trust* by Gambetta (1988) that is commonly used in the trust management research community: *"trust is the subjective probability with which an agent assesses that another agent will perform a particular action in a context in which it affects his own action"*.

In search of a value describing Gambetta's subjective probability or what is called trustworthiness, reputation is often used as a measure (Jøsang et al. 2007). Reputation-based trust is deduced from past experiences or behavior of an entity. Since a common online environment may involve millions of actors, it can be difficult to determine a person's reputation manually. For this reason, reputation systems have become popular tools to support reputation assessment. Reputation systems encourage users to leave feedback, e.g., about the performance of a seller or the quality of a product. They then collect all feedback created, aggregate the ratings to compute one or several reputation values and present the output in a feedback profile (Resnick et al. 2000). Figure 1 depicts the generic process of a reputation system. Here, we also added the user who employs the reputation system to come to a decision (e.g. whether to buy or not to buy from a specific seller).

2.2 Exploiting Reputation Systems

Since the early works of Marsh (1994) on trust as a computational concept, many researchers have been able to provide empirical evidence for positive effects of

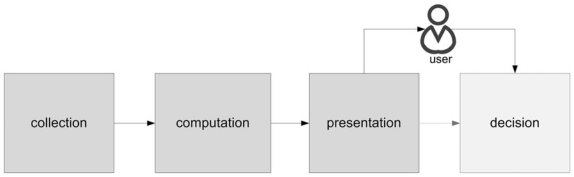


Fig. 1 Generic process of a reputation system

[4] http://www.ebay.com.

[5] http://www.amazon.com.

Table 1 Example attacks on reputation systems

Attack class	Name	Description
Unfair feedback	Ballot stuffing	The attacker provides many high ratings to unfairly push the reputation of an entity
	Bad mouthing	The attacker provides many low ratings to unfairly destroy the reputation of an entity
Inconsistent behavior	Value imbalance exploitation	The attackers gathers good reputation selling cheap items, but at the same time cheats on the expensive ones
	On-off attack	The attackers first acts honestly to build high reputation, than "milks" the good reputation. After a certain threshold value is reached, he behaves honestly again and starts from the beginning
Identity-based	Whitewashing	The attacker behaves maliciously from the beginning. After he has received negative ratings, he opens a new account
	Sybil-attack	The attacker creates many accounts (Sybils) at the same time to increase his influence in a community

reputation systems in online communities (Resnick and Zeckhauser 2002; Houser and Wooders 2006). In electronic marketplaces, for example, reputable sellers generate price premiums and have an increased number of sales (Ba and Pavlou 2002; Resnick et al. 2006; Diekmann et al. 2013). This offers an enticement for malicious actors to exploit the weaknesses of reputation systems in order to unfairly improve their position.

In the recent years, a huge variety of different attacks on reputation systems has been described in literature (Jøsang and Golbeck 2009; Sun and Liu 2012). In order to give a structured overview of weaknesses, Koutrouli and Tsalgatidou (2012) identified three classes of attacks, namely unfair feedback, inconsistent behavior and identity-based attacks. In unfair feedback attacks, malicious peers provide many unfairly high or low ratings to e.g. push the reputation of a friend or destroy the reputation of a competitor (Dellarocas 2000). The second class – inconsistent behavior – involves attacks that benefit from the characteristics of specific computation metrics. In reputation systems that do not consider transaction context, for instance, a malicious actor could show a discriminating behavior in different situations such as selling high quality chewing gum but low quality laptops. Thirdly, identity-based attacks primarily address systems that allow cheap pseudonyms. Table 1 gives an overview of selected example attacks.

To cope with the weaknesses of early reputation systems, a large body of work on the design of new trust models and robust metrics has been published (Artz and Gil 2007; Koutrouli and Tsalgatidou 2012; Sun and Liu 2012). While most early computation models could easily be manipulated by malicious actors, current models (particularly those introduced in literature) are quite resistant against a variety of attacks.

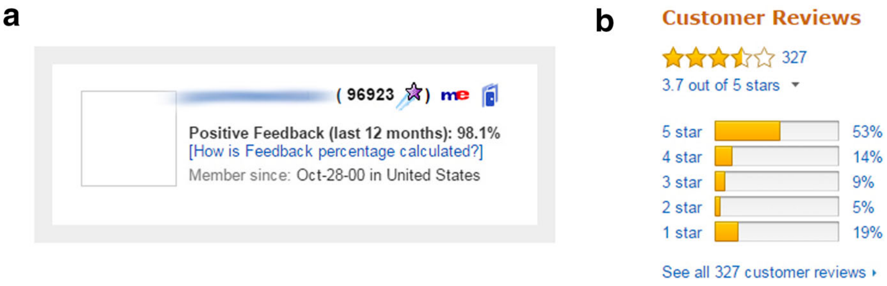
2.3 Robust Metrics vs. Transparent Presentation

A reputation system is deemed robust if it always delivers a higher reputation value for an honest actor than for a malicious actor in case of an attack (Zhang et al. 2012b). In order to increase the robustness of reputation systems, researchers have introduced a large number of complex computation models in the recent years. These are usually based on advanced mathematics using e.g. beta probability density functions, hidden markov models or complex clustering algorithms. Table 2 gives an overview of selected examples.

Admittedly, these sophisticated models are more robust when attacked, however, they are beyond the understanding of the users intended to apply them in most cases (Marsh et al. 2012). One or several numerical values as only outcomes of the computation process are in general not sufficiently transparent to the end-user as they cannot convey any details about the exact input data (e.g. which reviews were considered, which not?) leading to a cognitive gap. A recent user study conducted by Hammer et al. (2013) in which they compared two reputation metrics was able to show that more than half of the participants

Table 2 Example reputation models

Name	Author	Formal basis
Beta reputation system, TRAVOS	Jøsang and Ismail (2002), Teacy et al. (2006)	Beta probability density functions (PDF)
iClub	Liu et al. (2011)	Clustering-based model
Evidential model	Yu and Singh (2002)	Belief model, Dempster-Shafer theory
Web services reputation	Malik et al. (2009)	Hidden Markov Model
REGRET	Sabater and Sierra (2001)	Fuzzy model



(96923) me

Positive Feedback (last 12 months): 98.1%
[How is Feedback percentage calculated?]
Member since: Oct-28-00 in United States

Customer Reviews

327

3.7 out of 5 stars

5 star	53%
4 star	14%
3 star	9%
2 star	5%
1 star	19%

See all 327 customer reviews ›

Fig. 2 **a** eBay's and **b** Amazon's overview

criticized this lack of transparency if they were not able to understand the computation.

When analyzing reputation systems in practice, however, this trend to ever more robust computation models is not reflected (particularly in e-commerce). Most prominent reputation systems, such as eBay's feedback mechanism or Amazon's customer review system, still use a very simple and easy to understand metric to provide an overview of the ratings. eBay's reputation value shows the share of positive feedback of the overall feedback, while Amazon provides the average rating value on a five-star scale. Figure 2 depicts these two examples.

Though these metrics are easy to understand, they are highly vulnerable to different kinds of attacks as pointed out in several works (Kerr and Cohen 2009; Hoffman et al. 2009; Yao et al. 2012).

2.4 Research Gap and Objectives

Based on the problems identified above, the main research question (RQ) addressed in this work is stated as follows:

RQ. How can the robustness of reputation systems be enhanced without reducing both transparency and the user's understanding of reputation data?

Most research conducted to increase the robustness of reputation systems in the recent years focused on the improvement of the computation phase (see Fig. 1). While advanced computation models may provide an effective remedy against malicious behavior, they have been shown to be non-transparent to the end-user. Transparent interfaces using simple computation models, in contrast, are non-robust against attacks. We believe that only through combining both sides of the coin, an effective detecting and understanding of malicious behavior can be achieved (see Fig. 3).

To this end, we take a different path to increase the robustness of reputation systems: instead of designing a further improved computation model we integrate computation and presentation through user interaction. By incorporating the visual-cognitive capabilities of a human user and the computing power of a machine, we think that malicious behavior can be reliably and transparently identified. Our objectives when introducing this concept are:

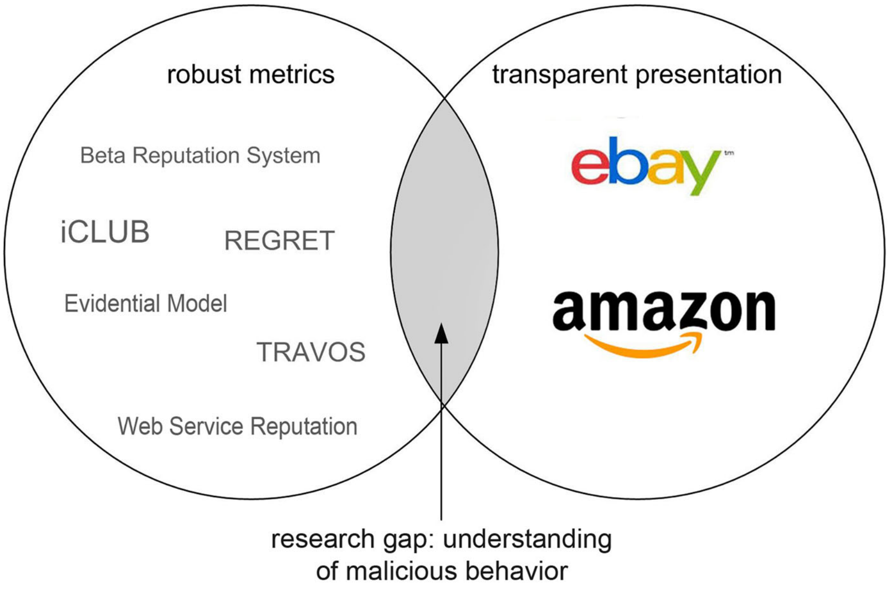


Fig. 3 Robust metrics vs. transparent presentation

1. To increase the user's ability to detect attacks on reputation systems (robustness), and
2. to enhance the user's understanding of reputation data.

2.5 Research Approach

When designing a complex artifact, there may not be a definite final condition at the beginning. In line with the human problem solving approach (Newell 1972), the analysis of a design problem therefore involves the successive decomposition to sub-problems until they seem solvable, followed by a final composition of the single sub-solutions to an overall solution (Chandrasekaran 1990). With this in mind, we first decomposed the overall problem to two sub-problems, each addressing one of the two central phases of the generic process of a reputation system: *computation* and *presentation*. On the one hand, we elaborated on the question of how the computation could be re-designed as an interactive process (Sänger et al. 2015a, c). On the other hand, we analyzed how reputation data could be presented to involve the user in reputation assessment (Sänger and Pernul 2014, 2016; Sänger et al. 2015b). In these previous works, however, the two phases – computation and presentation – were considered as isolated building blocks.

In this paper, we go one step beyond the state of the art by connecting the results of our recent works (sub-solutions) and introduce the overall concept of interactive reputation systems. Only after combining the findings made for the respective blocks, the idea of involving the user in reputation assessment through interaction can take full effect. First, we will introduce and define the notion of an interactive reputation system and describe the process model. Second, based on the findings made for interactive computation and interactive presentation, we will integrate both blocks on a conceptual level. After that, we show what an interactive reputation system could look in a real-world environment by means of a software prototype and evaluate this concept as to the objectives defined above.

3 Interactive Reputation Systems

3.1 Process Model

A classical reputation system delivers static numerical reputation values as output. Though the user interface may provide some filters to read specific reviews (e.g. only negative feedback), the reputation value is not adjusted.

A few novel reputation systems that have emerged in the recent years, however, provide the possibility to adapt single settings and reflect these changes accordingly in the reputation value. Tripadvisor, for example, allows the users to focus on reviews provided by a specific peer group (e.g. families, business travelers, etc.), filters the corresponding reviews and displays an aggregated reputation value that only involves ratings of the selected group. We call such systems *interactive* reputation systems, since unlike static systems these systems encourage the user to interact and adapt the computation levers in order to gain enhanced insights. Based on this, the definition of an interactive reputation system as used in this work is as follows:

Definition *An interactive reputation system is a reputation system that allows the users to dynamically adapt the underlying computation mechanism.*

Reviewing the process model of a classical reputation system as introduced in Sect. 2.1, a user who may interactively adapt the computation is not considered. To depict this coherence for an interactive reputation system, the process model of a static system needs to be extended by a bilateral flow between the user and the computation via the presentation interface as depicted in Fig. 4.

Exactly as in the process model of a classical reputation system (see Fig. 1), data is collected in the first step. Second, both the raw data as well as models of the data calculated in the computation are presented to the end-user. Ideally, a very intuitive metric such as the average rating value is used in the initial computation to provide a transparent starting point for further analyses. On eBay, for instance, a list of the textual reviews (raw data), the share of positive ratings and an aggregation of positive, neutral and negative ratings for specific time-frames (models) are presented in the feedback profile interface. In an interactive reputation system, the user can then "interact" with the visual representation of the reputation profile e.g. through adapting computation settings, thus changing the models that are computed. Typical interaction techniques include the filtering of unnecessary reviews, adding of weighting rules or changing of the entire aggregation metrics. The inherent user-driven definition of the calculation rules will

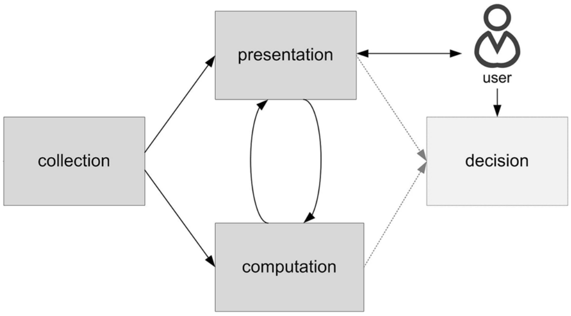


Fig. 4 Generic process of an interactive reputation system

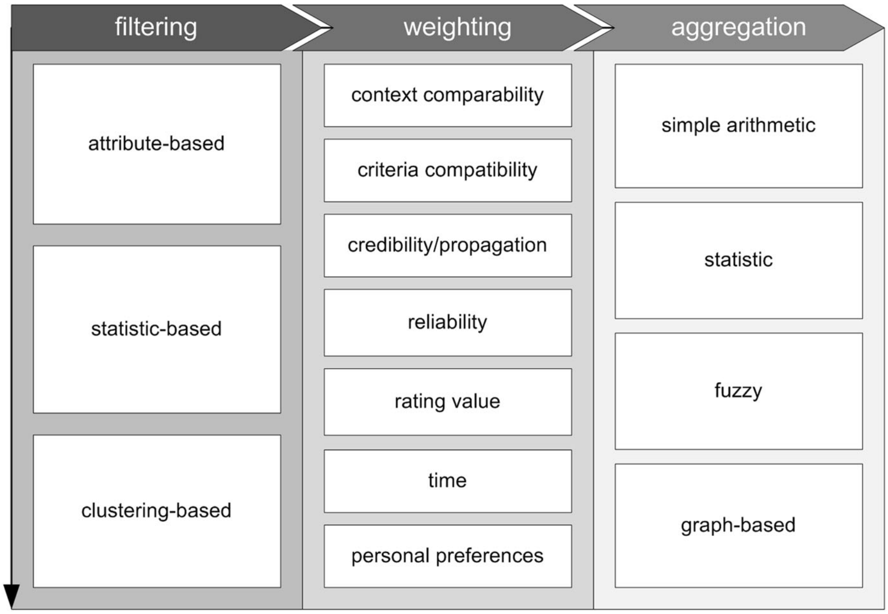


Fig. 5 Classes of filtering-, weighting- and aggregation-techniques (Sänger et al. 2015c)

logically lead to a clearer understanding of the models computed.[6]

3.2 Conceptual Design

When starting to plan an interactive reputation system as introduced above, several decisions need to be made. While there is no need to adapt the collection, the design of the computation and presentation must be completely revised compared to a classical reputation system. Here, two major design decisions were necessary:

1. Computation: To transform the computation mechanism from static to dynamic in a way that it may be interactively adapted by the user, we chose a *modular design* that allows to dynamically add and remove components.
2. Presentation: The user-friendly and transparent presentation of reputation data is a vital requirement to guarantee the proper understanding of the information provided to the end-user. Adding *interactive visualizations* of reputation data to the user interface provides a promising alternative to overcome the shortcomings of a single numerical value and link computation and presentation.

3.2.1 Computation

If the computation mechanism is modularized, a user will be able to dynamically compose a computation model through adding, removing and adapting components that implement specific functionality. To this end, the general building blocks of the computation models used in common reputation systems were identified first. Here, all well-established systems described in literature on trust models were analyzed. Based on their characteristics, a hierarchical taxonomy of computation components was developed as depicted in Fig. 5.

On the primary level it involves the three primary classes “filtering”, “weighting” and “aggregation”. Obviously, the first task when assessing reputation is to filter all information available for what is relevant for the specific situation. Secondly, the significance of single referrals needs to be determined such that e.g. very new feedback might be treated as more relevant than very old feedback (weighting). Finally the input data is aggregated to provide a final output (e.g. reputation value). Based on these three main classes, 14 secondary component classes, 26 components and 36 subsets were identified (Sänger et al. 2015c). Here, the components and sub-sets represent examples of how the functional blocks were instantiated in different trust models. The secondary class “simple arithmetic”, for instance, contains components that implement different simple aggregation techniques such as the “average rating value” or the “share of positive ratings”.

Secondly, we implemented a component repository that stores an example implementation and a structured

[6] This process model is strongly inspired by the visual analytics process model that tries to combine the visual-cognitive capabilities of a human analyst and the computing power of a machine (Keim et al. 2010).

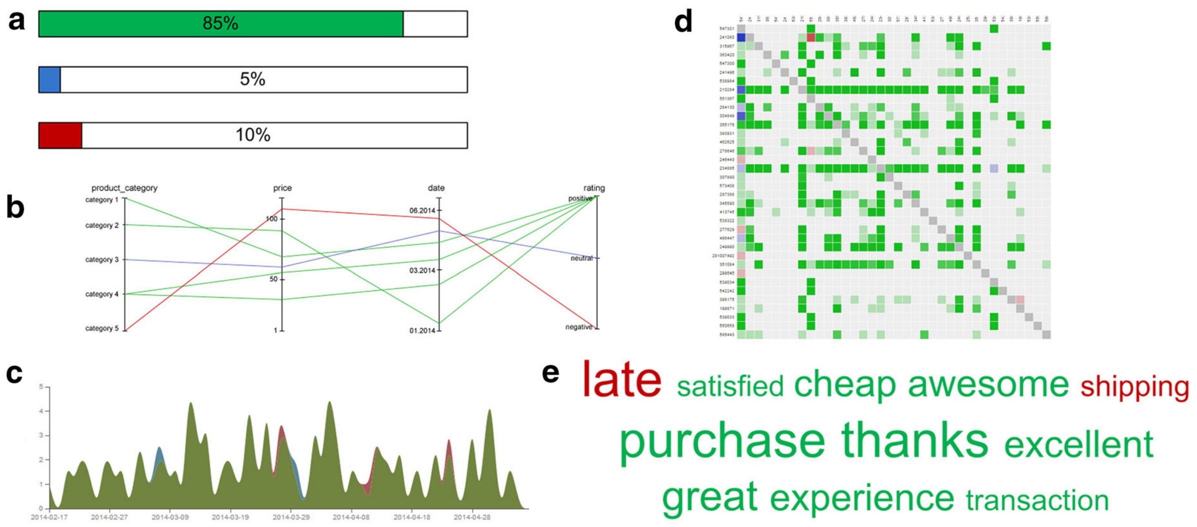


Fig. 6 Example visualizations of reputation data: **a** bar chart showing the relative share of positive, neutral and negative ratings (uni-variate data), **b** parallel coordinates displaying ratings and the corresponding transaction context (multi-variate data), **c** stack graph depicting the distribution of ratings of time (timed-based uni-variate data), **d** reputation matrix revealing the feedback relations (network data) and **e** colored tag cloud giving an overview of feedback texts (textual data)

description (ID, type, URL, parameter, example call, example output) of each component as reusable web services.[7] The component repository (centrally stored) is publicly available and encourages researchers to add novel components and extend existing ideas. All web services implement well-defined interfaces and can easily be called using the "WebserviceCallHelper" – a helper class available on the repository website.

By dynamically combining these components through interaction (adding, removing or adapting), a user can compose his individual computation mechanism on-the-fly. To recreate Tripadvisor's computation mechanism, for instance, only the two components "context-based filter" and "average rating value" of the classes "filtering (primary class), attribute-based filter (secondary class)" and "aggregation (primary class), simple arithmetic (secondary class)" need to be selected and composed. Overall, over 170,000 distinct combinations of the computation components are possible.

3.2.2 Presentation

A numerical value as only outcome of the computation lacks transparency. Therefore, we chose a different path by additionally developing an integrated visual representation of reputation data using interactive visualizations (Sänger and Pernul 2016). Visualizations are very convenient to depict a wide range of information in one integrated view.

To this end, the data structure of reputation data used in electronic marketplaces was analyzed and all different facets of reputation data were classified according to their data type. Here, we could identify three main information blocks – feedback, transaction context and actors. The resulting classification of reputation data within each block builds the basis for the selection of suitable visualization and interaction techniques. The first block "feedback", for instance, usually contains the three items *rating* (uni-variate), *review text* (textual) and *feedback time* (time-based uni-variate). Figure 6 gives an overview of different example visualizations suited for distinct types of reputation data.

3.2.3 Integrating Computation and Presentation

"[...] if complex models are used in the background for whatever reason, the user interface has a role to make the model understandable without losing any of its predictive power" (Marsh et al. 2012). In our recent research, computation and presentation were mostly considered as isolated blocks. However, the models that are computed based on the modular mechanism described above also need to be presented in a way the user may understand. Here, interactive visualizations offer a perfectly suitable technique as the intermediate outcomes of filtering, weighting and the final aggregated reputation value can be represented in the user interface. Due to the standardized interfaces of the computation components implemented as web services, the output of filtering, weighting and aggregation is independent of the functional behavior of the particular component

[7] http://trust.bayforsec.de/ngot/index.php?section=service_repository.

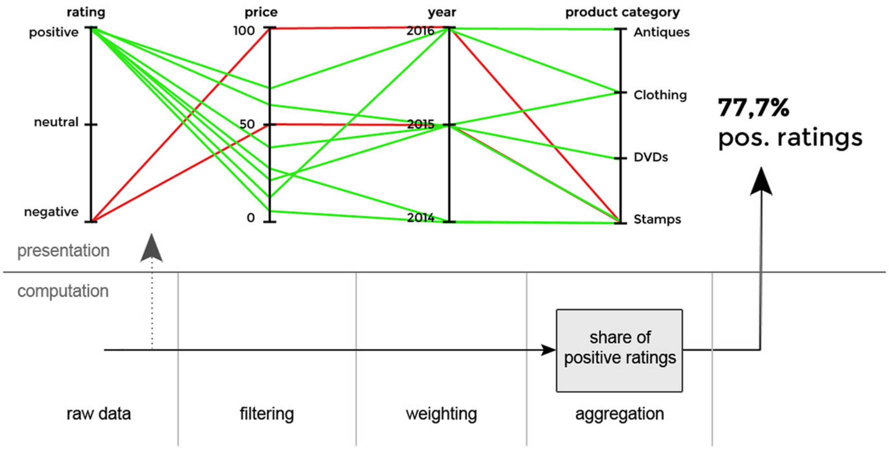


Fig. 7 Starting situation: reputation data is displayed in the GUI using a parallel coordinates visualization accompanied by the share of positive ratings

but exactly alike for components of each primary class. In order to integrate computation and presentation, visualizations must therefore be extended to the ability to represent the outcomes of the respective computation step.

To give an example, let us set the following scenario: A reputation system used in an electronic market place should be adapted to enable the interactive representation of a reputation value with regard to the transaction context. To compute the reputation value, we initially use the "share of positive ratings" aggregation component on the computation layer. To display reputation data on the presentation layer, we choose a parallel coordinates (PA) visualization that shows ratings and the corresponding transaction context. Parallel coordinates is a visualization technique that is perfectly suitable for giving an overview of and revealing correlations within a multi-variate dataset (Heinrich and Weiskopf 2013). Here, four axes (each representing one of the context attributes rating, price, year, and product category) are layed out in parallel side by side (Inselberg 1985). Values on these axes are then connected by colored polylines, each displaying one tuple of the dataset. Figure 7 shows a feedback profile of a seller who received 7 positive and 2 negative ratings leading to a share of 77.7% positive ratings (right). PA (left) display the raw data (ratings and corresponding transaction context).

While Fig. 7 gives a general overview of seller feedback, the user/buyer might want to know what kind of feedback the seller received for the product category "stamps" as he plans to buy stamps. Through clicking on the category "stamps" label on the fourth axis of the interactive visualization, all feedback not attributed to stamps is blinded out. On the computation layer, selecting ranges on the scales is equivalent to adding a context-based filter that filters all feedback but that related to the selected attributes. Adding this component, the reputation value is dynamically updated to 50% positive ratings (see Fig. 8).

Finally, the buyer wants to focus on high price stamps. Through clicking on the "price" label, a weighting factor ranging from [0;1] is added to each rating (0 for the lower end and 1 for the upper end of the scale). Accordingly, the opacity of the polylines is adapted. On computation level, this is equivalent to adding a context-based weighting

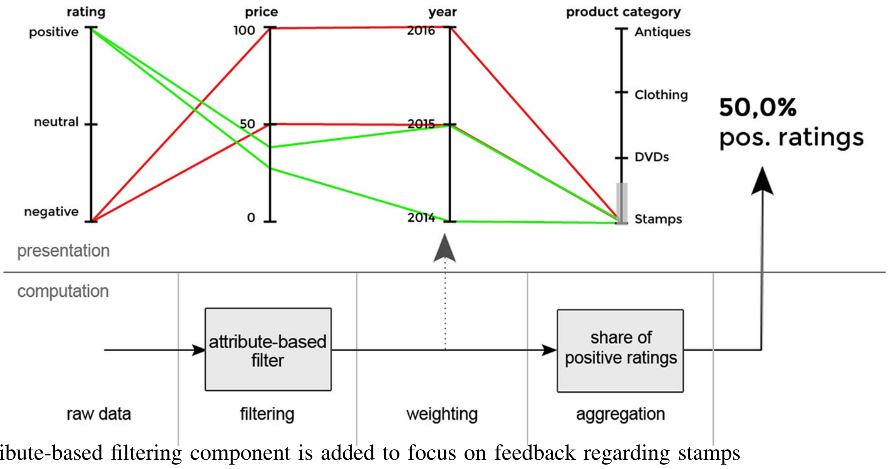


Fig. 8 Step 1: An attribute-based filtering component is added to focus on feedback regarding stamps

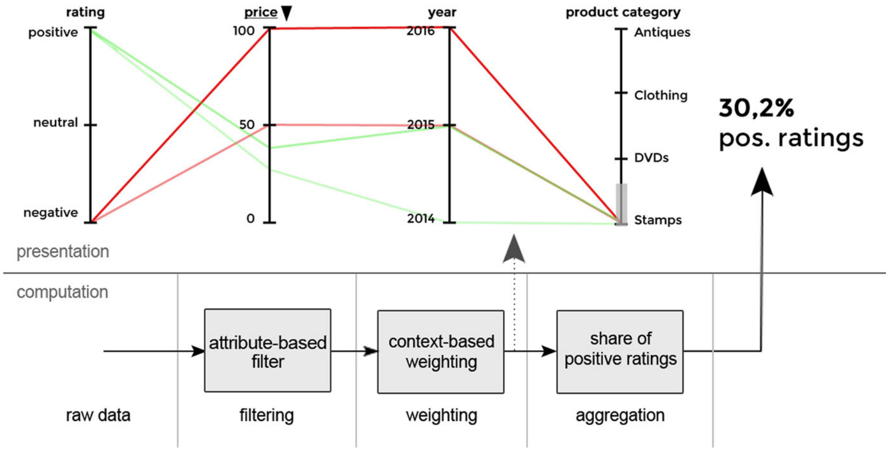


Fig. 9 Step 2: A context-based weighting component is added allocating a higher weighting factor to higher prices

component. The corresponding reputation value is again dynamically updated to 30.2% positive ratings (see Fig. 9).

This example reveals that displaying reputation data in an interactive visualization and encouraging the user to add filtering and weighting rules through user interaction may notably increase the transparency and understanding of a rather complex computation model. If a reputation value of 30.2% based on a context-aware metric was provided initially, in contrast, a huge variety of potentially useful and interesting information would have been omitted. It would neither be clear which reviews were filtered nor which weighting factors were allocated to the different ratings.

3.3 Implementation

3.3.1 A Generic Software Framework

To demonstrate how interactive reputation systems could look in a practical environment, we implemented a software framework that allows to combine the component model on the computation layer (see Sect. 3.2.1) with interactive visualizations on the presentation layer (see Sect. 3.2.2). The framework was designed as a three-tier web application. On the data layer, we employed Couchbase[8] (NoSQL database) and MySQL (both can be used interchangeably, each having its advantages for different analyses). The logic layer was implemented using PHP with the framework Laravel[9] on server side and JavaScript with AngularJS[10] on client side. For the presentation we used HTML5, JavaScript and CSS. The data visualization was realized with the JavaScript library d3.js.[11] d3 is perfectly suitable for creating and manipulating documents based on a huge range of data. Figure 10 gives a schematic overview of the architecture.

Taking a closer look at the detailed architecture of the logic layer, we find two main building blocks. On server side the “dynamic composition webservice” (Laravel) represents the computation phase. On client side, the “dynamic evaluation client” represents the presentation phase. When the website is initially built, all raw reputation data is handed over to the dynamic evaluation client. Here, the analytics handler implemented in AngularJS takes all input data and builds the front-end website involving interactive visualizations (d3). In order to provide an initial reputation value, the “dynamic composition webservice” is called up. It ensures the composition of the computation engine, the calculation of reputation values and the delivery of output data to the client. After receiving a pre-defined or user-defined configuration as well as a list of rating data as input, the workflow engine validates the configuration and processes the defined combination. Here, all necessary computation components (web services) are sequentially called up. In our implementation, the dynamic composition web service as well as all computation components (web services) are stored in the central service repository introduced in Sect. 3.2.1. Once a reputation value is calculated, the output is sent to the client. A data exchange between client and server is carried out using JSON-format. After the initial reputation value is calculated, the website has loaded completely and is presented to the user.

Whenever the user starts to interact with a visualization, the analytics handler directly updates the composition configuration, sends it to the dynamic composition webservice and receives an updated output (reputation value) as an answer. In this way, interactions with the visualization can be directly represented in an updated reputation value.

To give an impression of what interactive reputation systems could look in practice, this paper presents a

[8] http://couchbase.com/.

[9] http://laravel.com/.

[10] https://angularjs.org/.

[11] http://d3js.org/.

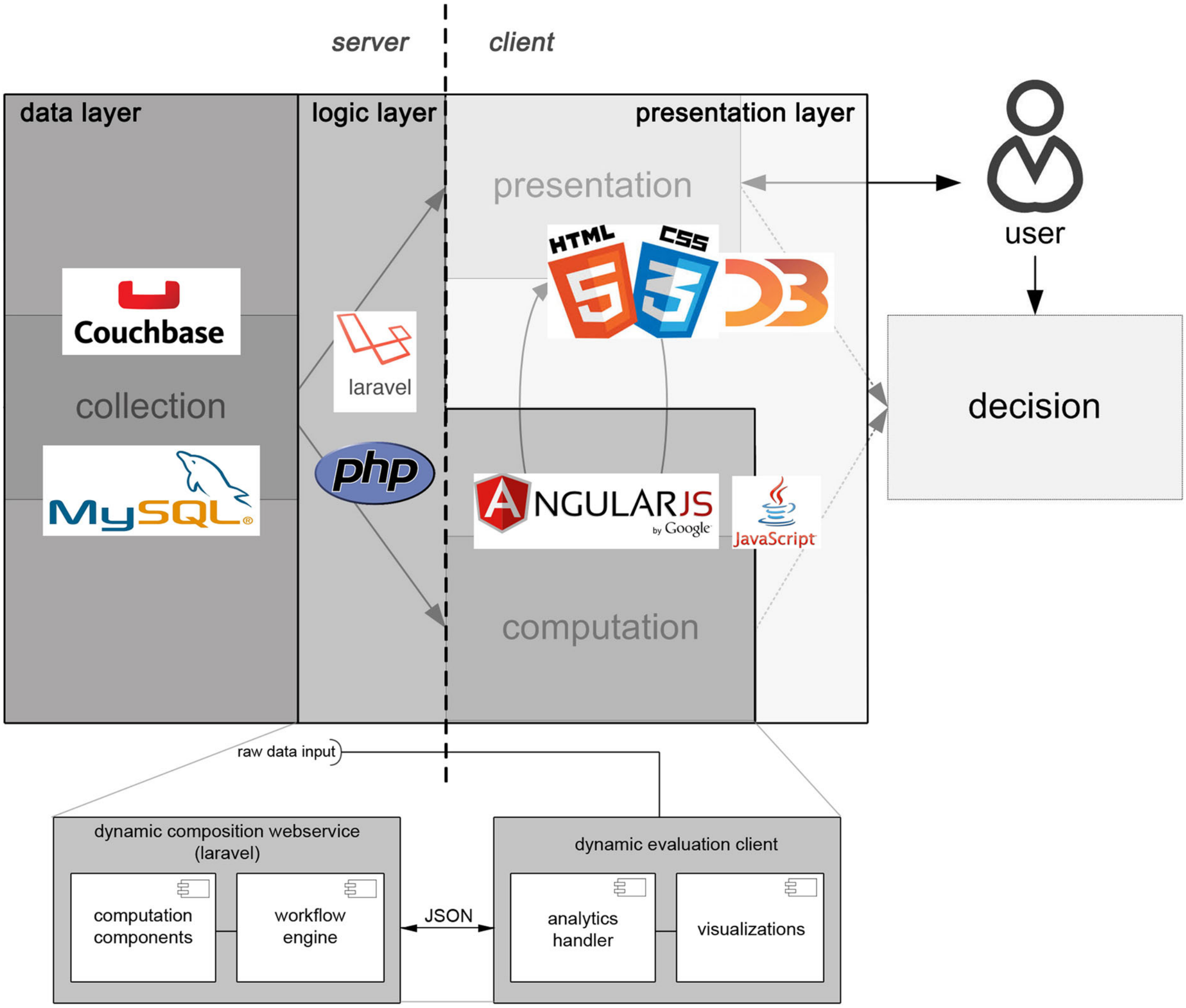


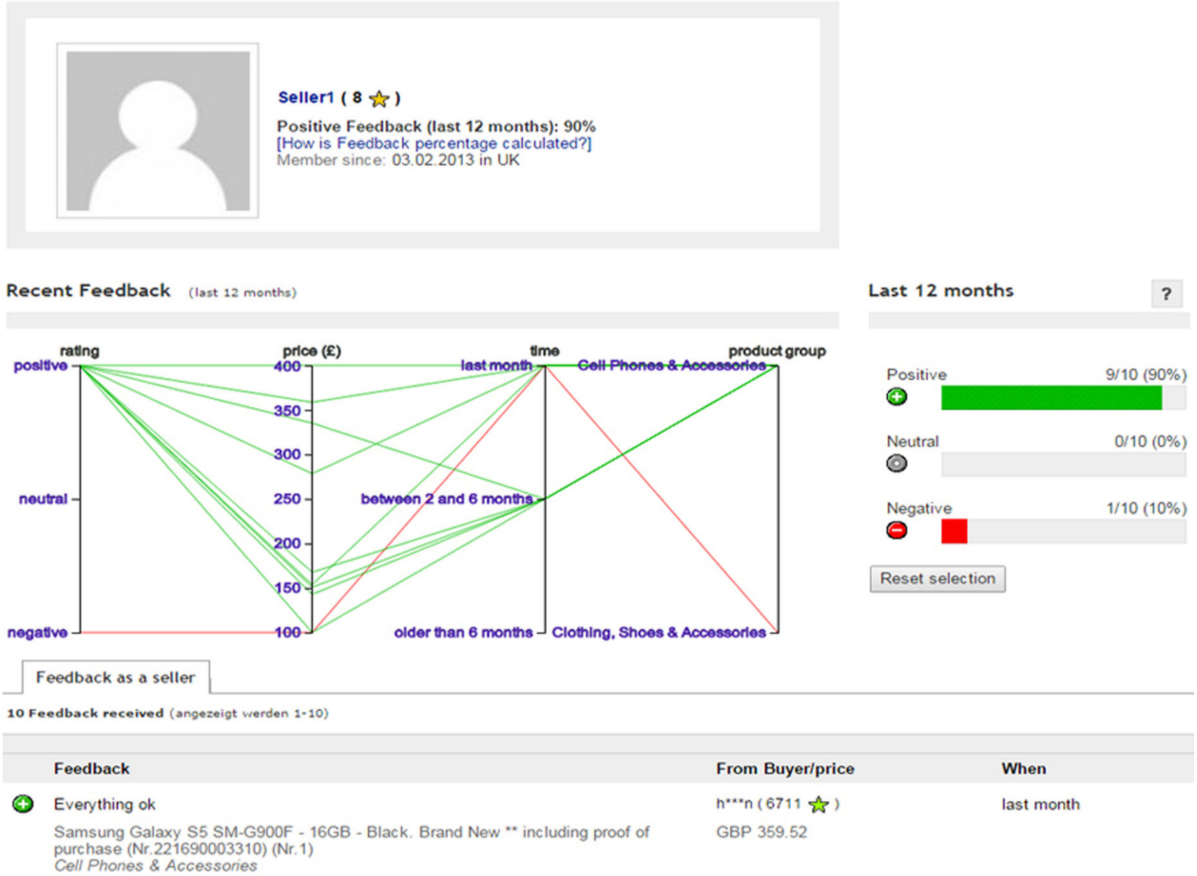
Fig. 10 Schematic view of the architecture

prototype (using real-world data) that was implemented based on the software framework described above. The prototype shows an example of how an existing reputation system could be adapted to enable user interaction.

3.3.2 *Prototype: Extension of an eBay-like Feedback Profile*

eBay's current feedback profile lacks a proper presentation of transaction context. A malicious seller could, for instance, build a high reputation selling a huge number of cheap items but then cheat on the expensive ones (see Sect. 2.2 for the different attack scenarios). This so-called value imbalance problem has not only been theoretically described but could also be observed in several real-world cases. Implementing this attack, a Californian seller managed to deceive victims for over $300,000 (Zhang et al. 2012a). To cope with this weakness, our prototype extends eBay's feedback profile by an interactive parallel coordinates visualization and a bar-chart providing an overview of the relative share of positive, neutral and negative ratings.

Similar to eBay, a summary of seller behavior is given in the top. In the middle, the parallel coordinates visualization (left) as well as the relative share of positive, neutral and negative ratings in a bar-chart (right) provide an overview of feedback and transaction context. Below the detailed feedback is listed. For interactivity, we added a "context-based filtering" component that is called up when parts of the axes are highlighted through "brushing". The reputation values (right) are additionally visualized in a bar-chart. Figure 11 shows an example screenshot of the prototype. Selecting ranges on the axis of the parallel coordinates visualization, the polylines displaying tuples that are not within the specified boundaries as well as the corresponding reviews are hidden. At the same time, all reputation values are dynamically updated. In this way, users can dive into the data and easily reveal correlations



Seller1 (8 ★)

Positive Feedback (last 12 months): 90%

[How is Feedback percentage calculated?]

Member since: 03.02.2013 in UK

Recent Feedback (last 12 months)

Last 12 months ?

Positive 9/10 (90%)

Neutral 0/10 (0%)

Negative 1/10 (10%)

10 Feedback received (angezeigt werden 1-10)

	Feedback	From Buyer/price	When
⊕	Everything ok Samsung Galaxy S5 SM-G900F - 16GB - Black. Brand New ** including proof of purchase (Nr.221690003310) (Nr.1) *Cell Phones & Accessories*	h***n (6711 ★) GBP 359.52	last month

Fig. 11 Screenshot of an example feedback profile for a seller who received 9 positive and 1 negative reviews in the product categories "Cell Phones" and "Clothing"

(Fig. 12 of the case study shows an example where high price transaction are highlighted).

4 Evaluation

To measure the quality of our concept, several evaluations have been conducted that involve both a real world case and an international user study. While the case proves that interactive reputation systems notably support the detection of malicious behavior if correctly applied, the user study provides empirical evidence that this concept is also beneficial for the average end-user.

4.1 Case Study

Actors who profit from the value imbalance problem build a high reputation by selling cheap products while cheating on the expensive ones. For this first case study, real world data from eBay Germany was gathered. The seller feedback profile introduced in this case listed 136 positive, 0 neutral and 8 negative reviews resulting in an average of 94.4% positive ratings in the last 12 months. The profile was found through a thread in eBay's community boards, where buyer complained that he had not yet received any goods although having paid a month ago.

At a first glance, the feedback profile looked okay with only 8 negative reviews and a larger number of positive referrals. After starting to interact with parallel coordinates visualization through brushing of the price axis of prices higher than 400€, however, this revealed that the seller had a very high rate of negative ratings (78%) for high-priced transactions (see Fig. 12). He received negative ratings for 7 of 9 transactions attributed to high prices. All negative ratings except one involve products of a price higher than 400€. Overall, this seller managed to cheat on a total amount of more than 12.000€.

The very fact that the malicious seller managed to cheat on 8 buyers and that his account was still active points out that eBay's current seller profiles are not capable of revealing such attacks. The representation of ratings and corresponding transaction context in an interactive reputation system, however, is a very intuitive way to support the detection of malicious behavior and prevent deception

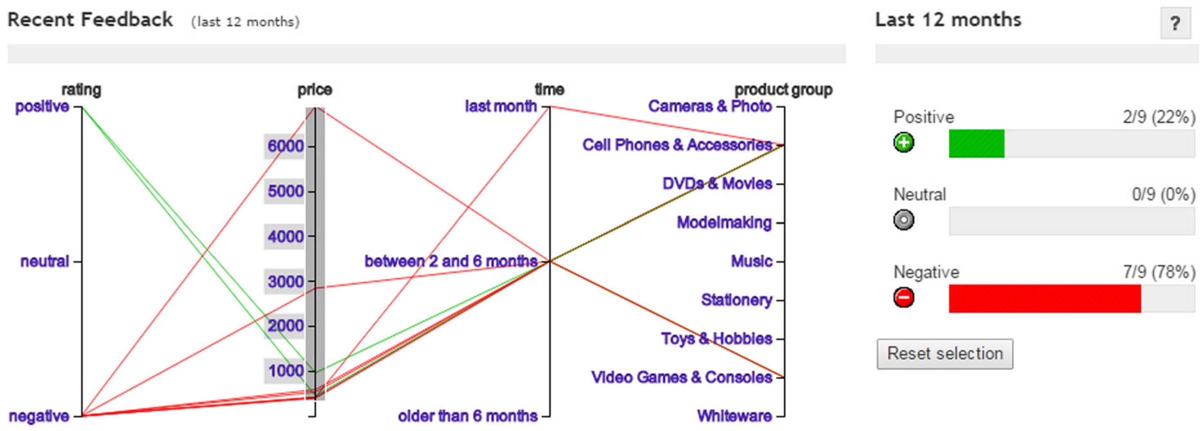


Fig. 12 Seller feedback profile highlighting the review attributed to prices higher than 400€

(objective 1). By instead using a context-sensitive reputation metric instead that provides a low reputation value for high price transaction, malicious behavior might also have been prevented. A single numerical value, though, could not have revealed the overall coherence and hence grant transparency.

4.2 International User Study

The case study described above demonstrates that interactive reputation systems can theoretically support the detection of malicious behavior. However, so far it has not been experimentally verified that our approach indeed increases the understanding of reputation data and can be applied by the average end user. Therefore, a controlled between-subject experiment[12] with 40 UK and 41 German participants (with different backgrounds and experience) in which we analyzed the users' detection ability and understanding of malicious behavior in an electronic marketplace (Sänger et al. 2016) was conducted. The study was carried out at University College London (UCL) and Friedrich Alexander University Erlangen-Nuremberg (FAU). In order to avoid priming the participants for fraud detection, they were asked to take part in a "usability test of online marketplace features". The study was divided into three phases. In the pre-study phase, the participants answered some initial questions regarding their experience with electronic marketplaces. During the main part of the study, participants had to solve four cases. In each case they were asked to buy a specific item (e.g. a mobile phone for 500€), to have a look at the feedback profiles of two sellers who offered the item and to give a preference for one seller. One of two sellers was honest while the other was malicious (participants were not told before to not prime them for fraud detection). The malicious seller showed a discriminating behavior for different context attributes in each case such as a high rate of negative ratings for high price items, whereas the honest seller behaved consistently within the entire context. The treatment group used our prototype (new interface condition) while the control group used an eBay-like interface that did not involve an interactive visualization and a dynamic reputation value (old interface condition). To obtain an overview of the UI, both groups had to watch a short video introducing the basic functions of the respective interface. In the third phase, participants were asked about the perceived usability by means of the System Usability Scale (Brooke 1996). Furthermore participants using the new interface provided details on the perceived usefulness. At the end, all participants were debriefed.

Participants were recruited via a standardized participant recruitment website, an internal mailing list as well as internal flyering. The age of the participants ranged from 18 to 41 years and 35 were female. For each case, participants reported a preference for one of both sellers on a 5-point likert scale with 3 representing no preference. We recoded the scale to +2 (preference for the honest seller) to −2 (preference for the malicious seller). Results show that the detection ability using the new interface was significantly higher[13] (77%) opposed to the old interface (56%), although the same amount of information was presented in both interfaces (objective 1). A positive value on the recoded preference scale was considered a correct detection. Repeating this measure with online buying experience as a further independent variable, we found that this effect was even more pronounced. When using the old eBay-like

[12] The study was conducted in collaboration between Johannes Sänger (University of Regensburg), Norman Hänsch (FAU), Brian Glass (UCL), Zinaida Benenson (FAU), Robert Landwirth (FAU) and M. Angela Sasse (UCL).

[13] $\chi^2(2, N = 324) = 16.44$, $p < 0.001$.

Table 3 Results of the user study

Measure	Old interface	New interface
Correct detection of malicious seller	56%	77%
Preferences for honest (of user who decided to buy)	58%	85%
Preferences for malicious (of user who decided to buy)	30%	7%
No preference (of user who decided to buy)	12%	8%
Sensemaking Score [0; 1]	0.25	0.46

interface, inexperienced participants were not able to detect any malicious behavior in 3 out of 4 cases. Participants reported their lifetime purchases on electronic marketplaces and we used the median cutoff score to categorize them as experienced or inexperienced.

More importantly, we asked the participants if they would indeed buy from one of the two sellers or abstain from buying[14]. In the new interface condition, only 7% of the users who decided to buy selected the malicious seller whereas 85% chose the honest seller. In the old interface condition, in contrast, 30% endorsed that they would buy from the malicious seller opposed to 58% who would buy from the honest seller[15]. The respective remaining shares did not give a preference. Table 3 provides an overview.

Furthermore, participants provided oral and written accounts for their decisions. Here, independent raters checked whether these reasons contained *a priori* specified criteria for determining that one seller is malicious and the other is honest. A "Sensemaking Score" was calculated taking the number of mentioned criteria divided by the number of all possible criteria. We found that users who used the new interface had a significantly higher Sensemaking Score (0.46 opposed to 0.25) indicating that they had a better understanding of seller behavior (objective 2). Finally, we asked participants about the perceived usability and the perceived usefulness of the new interface. Overall, participants reported the new system to be superior regarding a range of characteristics without trade-offs in usability, suggesting that such a system might be quickly adopted by users due to its high perceived usefulness.

5 Discussion and Conclusion

In this work, we introduced the notion of an interactive reputation system that tries to involve the user in reputation assessment to increase the users' detection ability and understanding of malicious behavior. We first delineated the generic process model, described how we designed and implemented the single building blocks, gave details on the integration of both blocks and presented an example prototype. The evaluation showed that both objectives – the improvement of the detection ability (robustness) and the enhancement of the users' understanding could be met without trade-offs in usability.

According to Zhang et al. (2012b) – who measured the robustness of a number of trust models in a recent work – "the more robust the trust model, the larger the transaction volume difference between the honest and dishonest duopoly seller". Reviewing the results of the user study, there is a difference of 0.78 (85–7% for the new interface) compared to 0.28 (58–30% old interface) between participants who decided to buy from one of the sellers. We assumed that users with no preference randomly choose one of the sellers and can thus be left out of consideration. This leads to an increase of 178% in terms of robustness. Measures of the Sensemaking Score also indicate an increase of 84% (0.46 opposed to 0.25) in the sense of understanding of reputation data.

While the concept of interactive reputation systems is in the early stages of development, this work provides some promising insights and a new perspective on how the robustness of reputation systems could be addressed while maintaining their transparency. The use of interactive visualizations for displaying reputation data and involving the user in reputation assessment proved to be valuable not only for expert use but also for the average end-user. Example implementations such as the prototype presented in this work show that current reputation systems do not necessarily need to be completely changed, but could also be extended to interactive functionality. Users reported such an extension to be very useful as demonstrated in the user study.

This work also has some limitations: In the evaluation, the experiments only focused on the usage of one specific prototype implementing one interactive visualization. Also, we analyzed the detection of specific attacks. Future research should therefore involve a larger variety of different prototypes implementing distinct visualization and interaction techniques as well as other facets of malicious behavior (e.g. unfair ratings). As the pool of participants who took part in the user study may not have been a perfectly representative sample (the majority had an academic

[14] There was no significant main effect on the decision to buy between both conditions for each case (ANOVA).

[15] χ^2 on the decision to buy revealed that those in the new interface condition were more likely to select the honest seller $\chi^2(2, N = 188) = 18,85$, $p < 0.001$.

background), further studies using different prototypes as well as a broader pool of participants may lead to insights that are even more substantial. Here, it is particularly interesting if end-users can deal with more complex interactive visualizations as well as more complex attacks. Moreover, participants being asked to "solve" cases may have analyzed the seller profiles more critically than in a real-life situation. Here, field studies outside of a lab setting may be a necessary step. Overall, more extensive prototyping as well as more comprehensive experiments will help to complete the design science cycle in an even more rigorous way.

One further issue that should be covered by future work are novel attacks and weaknesses that result from displaying reputation data in visualizations or involving the user in reputation assessment. As reputation systems are constantly under attack and malicious behavior is constantly adapted to new conditions (new threats may arise), a detailed analysis of possible new weaknesses may help to prevent simple exploits.

Overall, due to the practical relevance of this problem area and the promising outcomes we hope that our work encourages the conduct of further research on interactive reputation systems.

Acknowledgements The research leading to these results was supported by the "Bavarian State Ministry of Education, Science and the Arts" as part of the FORSEC research association.

References

Artz D, Gil Y (2007) A survey of trust in computer science and the Semantic Web. Web Semant 5(2):58–71

Awad NF, Ragowsky A (2008) Establishing trust in electronic commerce through online word of mouth: an examination across genders. J Manag Inf Syst 24(4):101–121

Ba S, Pavlou PA (2002) Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. MIS Q 26(3):243–268

Brooke J (1996) SUS—a quick and dirty usability scale. Usability evaluation in industry. Taylor & Francis, London

Chandrasekaran B (1990) Design problem solving: a task analysis. AI Mag 11(4):59–71

Dellarocas C (2000) Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In: Proceedings of the 2nd ACM conference on electronic commerce, Minneapolis, MN, USA, October 17–20, pp 150–157

Dellarocas C (2001) Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms. In: Proceedings of the 3rd ACM conference on electronic commerce, Tampa, EC '01, pp 171–179

Diekmann A, Jann B, Przepiorka W, Wehrli S (2013) Reputation formation and the evolution of cooperation in anonymous online markets. Am Sociol Rev 79(1):65–85

Gambetta D (1988) Can we trust trust? In: Gambetta D (ed) Trust: making and breaking cooperative relations. Basil Blackwell, Oxford, pp 213–237

Gefen D, Elena K, DW Straub (2003) Trust and TAM in online shopping: an integrated model. MIS Q 27(1):51–90

Hammer S, Kiefhaber R, Redlin M, Andre E, Ungerer T (2013) A user centric study of reputation metrics in online communities. In: Proceedings of the 3rd workshop on trust, reputation and user modeling (TRUM'13), Rome

Heinrich J, Weiskopf D (2013) State of the art of parallel coordinates. In: STAR proceedings of eurographics . Eurographics Association, pp 95–116

Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. MIS Q 28(1):75–105

Hillebrand C, Coetzee M (2015) The design of a configurable reputation service. In: Proceedings of the 12th international conference on trust, privacy and security in digital business (TrustBus 2015), Valencia, pp 60–70

Hoffman K, Zage D, Nita-Rotaru C (2009) A survey of attack and defense techniques for reputation systems. ACM Comput Surv 42(1):1–31

Houser D, Wooders J (2006) Reputation in auctions: theory, and evidence from eBay. J Econ Manag Strateg 15(2):353–369

Inselberg A (1985) The plane with parallel coordinates. Vis Comput 1(2):69–91

Jøsang A (2012) Robustness of trust and reputation systems: does it matter? In: Trust management VI—proceedings of the 6th IFIP WG 11.11 international conference IFIPTM 2012, Surat, pp 253–262

Jøsang A, Ismail R, Boyd C (2007) A survey of trust and reputation systems for online service provision. Decis Support Syst 43(2):618–644

Jøsang A, Golbeck J (2009) Challenges for robust of trust and reputation systems. In: Proceedings of the 5th international workshop on security and trust management, Saint Malo

Jøsang A, Ismail R (2002) The beta reputation system. In: Proceedings of the 15th Bled electronic commerce conference, Bled, pp 41–55

Keim Daniel A, Kohlhammer Jörn, Ellis Geoffrey, Mansmann Florian (2010) Mastering the information age—solving problems with visual analytics. Eurographics Association, Geneve

Kerr R, Cohen R (2009) Smart cheaters do prosper: defeating trust and reputation systems. In: Proceedings of The 8th international conference on autonomous agents and multiagent systems–volume 2 AAMAS '09, Budapest, pp 993–1000

Koutrouli E, Tsalgatidou A (2012) Taxonomy of attacks and defense mechanisms in P2P reputation systems—lessons for reputation system designers. Comput Sci Rev 6(23):47–70

Liu S, Zhang J, Miao C, Theng Y-L, Kot AC (2011) iCLUB: an integrated clustering-based approach to improve the robustness of reputation systems. In: The 10th international conference on autonomous agents and multiagent systems—volume 3 (AAMAS '11), Taipei, pp 1151–1152

Malik Z, Akbar I, Bouguettaya A (2009) Web services reputation assessment using a hidden Markov model. In: Proceeding of the 7th international joint conference on service-oriented computing (ICSOC-ServiceWave '09), Stockholm, pp 576–591

Marsh SP (1994) Formalising trust as a computational concept. PhD thesis, University of Stirling, Stirling

Marsh S, Basu A, Dwyer N (2012) Rendering unto Caesar the things that are Caesar's: complex trust models and human understanding. In: Proceedings of the 6th IFIP WG 11.11 international conference, IFIPTM 2012, Surat, pp 191–200

McKnight DH, Chervany NL (1996) The meanings of trust. Technical report MISRC working paper series 96-04, University of Minnesota, Management Information Systems Research Center

Newell A (1972) Human problem solving. Prentice-Hall, Upper Saddle River

Peffers K, Tuunanen T, Rothenberger M, Chatterjee S (2007) A design science research methodology for information systems research. J Manag Inf Syst 24(3):45–77

Resnick P, Kuwabara K, Zeckhauser R, Friedman E (2000) Reputation systems. Commun ACM 43(12):45–48

Resnick P, Zeckhauser R, Swanson J, Lockwood K (2006) The value of reputation on eBay: a controlled experiment. Exp Econ 9(2):79–101

Resnick P, Zeckhauser R (2002) Trust among strangers in internet transactions: empirical analysis of ebays reputation system. Econ Internet E-commerce 11(2):23–25

Riegelsberger J, Angela Sasse M, McCarthy JD (2005) The mechanics of trust: a framework for research and design. Int J Hum-Comput Stud 62(3):381–422

Sabater J, Sierra C (2001) REGRET: reputation in gregarious societies. In: Proceedings of the 5th international conference on autonomous agents (AGENTS '01), Hakodate, pp 194–195

Sänger J, Hänsch N, Glass B, Benenson Z, Landwirth R, Angela Sasse M (2016) Look before you leap: improving the users ability to detect fraud in electronicmMarketplaces. In: Proceedings of the ACM SIGCHI conference on human factors in computing systems (CHI 2016), San Jose

Sänger J, Pernul G (2014) Visualizing transaction context in trust and reputation systems. In: Proceedings of the 9th international conference on availability, reliability and security (ARES), Fribourg

Sänger J, Pernul G (2016) TRIVIA: visualizing reputation profiles to detect malicious sellers in electronic marketplaces. J Trust Manag 3(1):1–22

Sänger J, Richthammer C, Kremser A, Pernul G (2015a) Personalized composition of trustful reputation systems. In: Proceedings of the 29th Annual IFIP WG 11.3 working conference on data and applications security and privacy XXIX (DBSec 2015), Fairfax

Sänger J, Richthammer C, Kunz M, Meier S, Pernul G (2015b) Visualizing unfair ratings in online reputation systems. In: Proceedings of the 23rd European conference on information systems, ECIS 2015, Münster

Sänger J, Richthammer C, Pernul G (2015c) Reusable components for online reputation systems. J Trust Manag 2(5):1–21

Sun Y, Liu Y (2012) Security of online reputation systems: the evolution of attacks and defenses. IEEE Signal Proc Mag 29(2):87–97

Teacy WT, Patel J, Jennings NR, Luck M (2006) TRAVOS: trust and reputation in the context of inaccurate information sources. Auton Agents Multi-Agent Syst 12(2):183–198

Yamagishi T, Matsuda M (2002) Improving the lemons market with a reputation system: an experimental study of internet auctioning. Hokkaido University

Yao Y, Ruohomaa S, Feng X (2012) Addressing common vulnerabilities of reputation systems for electronic commerce. J Theor Appl Electron Commer Res 7(1):1–20

Yu B, Singh MP (2002) An evidential model of distributed reputation management. In: Proceedings of the 1st international joint conference on autonomous agents and multi-agent systems (AAMAS '02), Bologna, pp 294–301

Zhang L, Jiang S, Zhang J, Keong Ng W (2012b) Robustness of trust models and combinations for handling unfair ratings. In: Trust management VI—proceedings of the 6th IFIP WG 11.11 international conference, IFIPTM 2012, Surat, pp 36–51

Zhang H, Wang Y, Zhang X (2012a) A trust vector approach to transaction context-aware trust evaluation in e-commerce and e-service environments. In: Proceedings of the 5th IEEE international conference on service-oriented computing and applications, Taipeh, pp 1–8