

New Algorithms for Mining the Reputation of Participants of Online Auctions

Mikołaj Morzy

Institute of Computing Science, Poznań University of Technology,
Piotrowo 3A, 60-965 Poznań, Poland
Mikolaj.Morzy@cs.put.poznan.pl

Abstract. The assessment of credibility and reputation of contractors in online auctions is the key issue in providing reliable environment for customer-to-customer e-commerce. Confident reputation rating system is an important factor in managing risk and building customer satisfaction. Unfortunately, most online auction sites employ a very simple reputation rating scheme that utilizes user feedbacks and comments issued after committed auctions. Such schemes are easy to deceive and do not provide satisfactory protection against several types of fraud. In this paper we propose two novel measures of trustworthiness, namely, credibility and density. We draw inspiration from social network analysis and present two algorithms for reputation rating estimation. Our first algorithm computes the credibility of participants by an iterative search of inter-participant connections. Our second algorithm discovers clusters of participants who are densely connected through committed auctions. We test both measures on a large body of real-world data and we experimentally compare them with existing solutions.

1 Introduction

Electronic commerce, or e-commerce for short, is quickly becoming a noticeable market in contemporary economy. In the third quarter of 2004 e-commerce sales in the United States augmented to \$17.6 billion, which made 2% of the total retail sales. Important models of e-commerce include business-to-business (B2B), business-to-customer (B2C), and customer-to-customer (C2C) commerce. The later model describes auctions, which are one of the oldest forms of economic activity known to mankind. Although known for centuries, auctions are experiencing an incredible revival due to the Internet which is providing auction participants with new and unprecedented possibilities. Cautious estimates predict that over 15% of the entire e-commerce market can be attributed to online auctions. Global auction sites, such as www.ebay.com, www.onsale.com, or www.qxl.com, attract millions of users every day. Auction sites differ on the frequency of auctions, closing and bumping rules, or auction types. Most popular auction protocols include English, First Price Sealed Bid, Dutch, Sealed Double Auction, and Vickrey auctions. For example, at www.priceline.com buyers provide bids on commodity items without knowledge of prior bids of other users, and the bids are being immediately accepted or rejected by sellers. Other auction places offer

the possibility to bid on non-commodity items exposing the existing bids and establishing deadlines for auction termination.

Online auctions provide opportunities for malicious participants to commit fraud [11]. Fraudulent practices can occur during bidding process and after the bidding finishes. During the first stage two types of fraud are common: bid shilling and bid shielding. Bid shilling happens when a false identity is created to drive up the bidding process on behalf of a seller. On the other hand, bid shielding aims at discouraging other bidders from participating in an auction because the shielder sets the bid unusually high. At the last moment, the shielder withdraws the bid, leaving the second highest bid of shielder's partner the winner of an auction. After-bidding fraud includes sending no merchandise at all, sending merchandise that is of lower quality or inconsistent with the announcement, or never sending payment for the committed auction.

In order to successfully trade using online auction channel, users must develop trust toward their partners. Reputation ratings of participants of online auctions are both limited and unreliable. Furthermore, the anonymity and geographical dispersion encourages dishonest participants to default and deliberately commit fraud. Unfortunately, most popular auction sites employ only simplest reputation rating mechanisms which are easy to deceive and do not provide satisfactory protection against cheaters. Often, user reputation rating is a simple count of committed auctions that are described by auction parties using brief comments, usually labeled "praise", "neutral", or "complaint". This simple schema of reputation rating can be easily deceived into assigning an unfair rating to a user. Let us assume that a dishonest seller wants to increase the rating. One possibility is to create several artificial buyers who would provide extra positive comments at virtually no cost. Such unfairly high seller ratings are referred to as "ballot stuffing". Inflated seller's rating biases the system by providing the seller with an unearned reputation estimate, which allows the seller to obtain more bids or higher prices from other users [8, 10]. Still, an examination of artificial buyers' ratings would reveal their low credibility as they would not participate in any other auctions except the cheating seller's auctions. A careful cheater could disguise artificial buyers by creating a network of connections between them to form a clique. Such structure of inter-buyer connections would be hard to identify because all involved buyers would pretend to be fairly credible and their reported ratings would be high. On the other hand, a dishonest seller could create a few artificial buyers to provide unfairly low ratings to seller's competitors. Such technique is called "bad-mouthing" and reflects the negative impact of a deceitful gossip. Bad-mouthing is harder to implement because it requires an investment in winning competitor's auctions. Still, such an investment could prove beneficial if the expected gain of driving a competitor out of the market exceeds the initial cost. Sellers can also bias the system by providing discriminatory ratings, both positive and negative, to selected buyers.

In this paper we propose two novel measures of online auction participant trustworthiness. Instead of using simple participation counts, we propose to use data mining to analyze the topology of the network of seller-buyer connections

to derive useful knowledge about each participant. Building upon social network analysis techniques we extract two notions that characterize auction participants: credibility and density. Our original contribution includes the definition of those novel measures, the construction of efficient algorithms to compute them, and experimental evaluation of the proposal. We use a large body of data acquired from www.allegro.pl, a leading Polish online auction site, to empirically prove the feasibility and benefit of our measures.

This paper is organized as follows. In Section 2 we present the related work. Section 3 introduces the concept of credibility and presents an algorithm used to compute it. In Section 4 we define the notion of density of participants. The results of the experimental evaluation of our proposal are presented in Section 5. Finally, the paper concludes with a future work agenda in Section 6.

2 Related Work

Reputation systems [9] are commonly used to build trust on the Internet, where millions of individuals perform commercial transactions without knowing each other. Web-based auction sites typically rely on trust models in which credibility of participants is assessed by counting positive and negative comments received from their trading partners after each transaction [5]. A critical analysis of this simple model can be found in [7]. The author points to the subjective nature of feedbacks, the lack of feedback context, the need to perform feedback aging. Of particular importance is the fact that positive and negative feedbacks are highly asymmetric, because users refrain from providing a negative feedback until the quality of service becomes totally unacceptable.

Several solutions have been recently proposed to address at least some limitations of current feedback-based models. In [1] the authors introduced a complaint-only trust management method. A method presented in [3] evaluates the quality of ratings assuming that a rating is credible if it is consistent with the majority of ratings for a given user. [2] proposed to introduce a trusted third party that could be used to authorize, identify, and establish the reputation of auction participants. A comparison of fraud mechanisms in online auctions and pay-per-call industry in the early 1990s can be found in [11]. The author states that the existing efforts of online auction industry self-regulation are not adequate and would not solve the problem, hence legal action must be undertaken by the government to provide consumer protection. In [12] a trust model called PeerTrust for peer-to-peer e-commerce communities was proposed. The presented model includes several trust parameters, i.e., feedback in terms of satisfaction, number of transactions, credibility of feedback, transaction context, and community context. Interesting idea of trust and distrust propagation was formulated in [4]. The authors present a method to compute trust between any two individuals based on a small amount of explicit trust/distrust statements per individual.

In our approach, rather than trying to solve all the problems with reputation assessment in Web-based auctions, we focus on just one issue that we believe is

the most important, i.e. credibility of feedback. We go a step further than [12] with the goal of discovering networks of artificial auction participants created by cheating sellers and providing reciprocal positive comments. We mine the network of auction participants to derive knowledge on seller's credibility that would be independent of other users' feedbacks. To achieve this, we measure the density of each seller's neighborhood.

The problem of evaluating importance of Web pages by Web search engines can be regarded as similar to the problem of reputation assessment in online auctions. In terms of implementation details our method for credibility assessment in online auctions is similar to an algorithm proposed to evaluate the quality of Web pages, called HITS (hyperlink-induced topic search) presented in [6]. HITS divides the pages into authorities (covering a certain topic) and hubs (directory-like pages linking to authorities devoted to a certain topic). In our method, we apply a similar distinction, dividing auction participants into those that are mainly sellers and those that are mainly buyers. Our second method, i.e. the assessment of seller density, is similar to density-based clustering schemes.

3 Credibility of Participants

Given a set of buyers $B = \{b_1, b_2, \dots, b_n\}$ and a set of sellers $S = \{s_1, s_2, \dots, s_m\}$. Given a set of possible comments $C = \{-1, 0, 1\}$, where each value represents the "negative", "neutral", and "positive" comment, respectively. Given a set of auctions $A = \{a_1, a_2, \dots, a_p\}$. An auction is a tuple $a_i = \langle b_j, s_k, c \rangle$ where $b_j \in B \wedge s_k \in S \wedge c \in C$. Let $S(b_j)$ represent the set of sellers who sold an item to the buyer b_j . We denote the *support* of the buyer b_j as $support(b_j) = |S(b_j)|$. Let $B(s_k)$ represent the set of buyers who bought an item from the seller s_k . We denote the *support* of the seller s_k as $support(s_k) = |B(s_k)|$. According to this formulation, the support of the participant is identical to the reputation rating measure currently employed by leading online auction providers.

Given a $m \times n$ matrix M_S . Each entry in the matrix represents the flow of support from a buyer to a seller in a committed auction. Entries in the matrix M_S are initialized as follows.

$$\forall i \in \langle 1, m \rangle \quad M_S[i, j] = \frac{1}{support(b_j)} \text{ if } \langle b_j, s_i, * \rangle \in A, \quad 0 \text{ otherwise}$$

Given a $m \times n$ matrix M_B . Each entry in the matrix represents the flow of support from a seller to a buyer in a committed auction. Entries in the matrix M_B are initialized as follows.

$$\forall j \in \langle 1, n \rangle \quad M_B[i, j] = \frac{1}{support(s_i)} \text{ if } \langle b_j, s_i, * \rangle \in A, \quad 0 \text{ otherwise}$$

Given a vector $S_C = [s_1, s_2, \dots, s_m]$ of seller credibility ratings. Initially, all sellers receive the same credibility of 1. Analogously, given a vector of buyer credibility ratings $B_C = [b_1, b_2, \dots, b_n]$. Initially, all buyers receive the same credibility of 1. Reputations of sellers and buyers are independent of each other

despite the true identity of participants. In other words, the fact that a person can be a seller and a buyer at the same time is not considered. In our opinion this simplification is justified in practice and it reflects the real behavior of online auction participants who form distinct and well separated clusters of buyers and sellers. Upon the termination of the algorithm vectors S_C and B_C contain diversified credibility ratings for sellers and buyers, respectively. A reputation rating for a buyer b_j is a tuple $R(b_j) = \langle C_-, C_0, C_+ \rangle$. Each component represents the sum of credibilities of sellers participating in transactions with a given buyer and posting a negative, neutral, or positive comment, respectively. Formally, $C_- = \sum_k S_C[k]$ where $\langle b_j, s_k, -1 \rangle \in A$, $C_0 = \sum_k S_C[k]$ where $\langle b_j, s_k, 0 \rangle \in A$, and $C_+ = \sum_k S_C[k]$ where $\langle b_j, s_k, +1 \rangle \in A$. Reputation rating for a seller is defined analogously.

Our method of reputation rating estimation is based on the following recursive definition of credibility. We consider a given buyer to be highly credible if the buyer participates in many auctions involving credible sellers. Analogously, we define a given seller to be credible if the seller participates in many auctions involving credible buyers. Since there is no *a priori* estimation of credibility of participants, we assume that initially all participants have equal credibility. Then, we iteratively recompute the credibility of sellers and buyers in the following way. In each iteration we distribute the current credibility of each buyer among participating sellers. Next, we update the credibility of all sellers by aggregating credibility collected from participating buyers. After the credibility of sellers has been updated, we propagate current credibility of sellers to buyers and we refresh the appropriate ratings. We repeat this procedure several times until the credibility of sellers and buyers converge. Alternatively, the procedure can be repeated a given number of times. Our experiments suggest that in practical applications ten iterations are sufficient to estimate the credibility correctly. After assessing the credibility of all participants the credibility ratings can be used together with the database of past comments to derive the proper reputation ratings by aggregating the credibility of contractors grouped by the type of the comment issued after the transaction.

The algorithm, presented in Fig. 1, works as follows. First, all required structures are initialized as explained above. Next, the algorithm begins to iteratively recompute the credibility for sellers and buyers. The intuition behind the algorithm is that the credibility of “good” buyers quickly aggregates in “good” sellers and *vice versa*. Initial ratings consisting of simple participation counts are quickly replaced by the true credibility which reflects the importance of every participant. Casual auction participants receive significant recommendation ratings only if they trade with highly ranked sellers, otherwise their initial unitary recommendation rating dissolves among lowly rated sellers.

Presented algorithm interestingly safeguards against two popular schemes of reputation rating deception. Ballot stuffing, which is a conspiracy between a seller and a group of buyers in order to unfairly augment the reputation ranking of a seller, is prevented by quick decrease in the reputation of such buyers. This can be attributed to the fact that, accordingly to the algorithm, buyers trading

Require: $A = \{a_1, a_2, \dots, a_p\}$, the set of committed auctions
Require: $B = \{b_1, b_2, \dots, b_n\}$, the set of buyers
Require: $S = \{s_1, s_2, \dots, s_m\}$, the set of sellers
Require: M_S, M_B , matrices representing the structure of the inter-participant network
Require: S_C, B_C , vectors representing the credibility of participants

- 1: Initialize matrices M_S, M_B and vectors S_C, B_C appropriately
- 2: **repeat**
- 3: **for all** $s_k \in S$ **do**
- 4: $S_C[s_k] = \sum_{j=1}^n M_S[j, k] * B_C[b_j]$
- 5: **end for**
- 6: **for all** $b_j \in B$ **do**
- 7: $B_C[b_j] = \sum_{k=1}^m M_B[j, k] * S_C[s_k]$
- 8: **end for**
- 9: **until** vectors S_C and B_C converge
- 10: Output S_C and B_C as credibility ratings
- 11: Compute reputation ratings $R(b_j), R(s_k) \forall b_j \in B, \forall s_k \in S$ using S_C, B_C , and A

Fig. 1. Algorithm for computing the credibility of participants

with a few sellers are generally not considered trustworthy. An attempt to create a clique of participants who try to mutually increase their reputation ratings is also prevented by the algorithm. In such case, the algorithm discovers a subset of participants with constant reputation rating. An artificial clique is indeed a closed system with no inflow or outflow of credibility. The detection of such a closed system quickly leads to the discovery of fraudulent collaboration between dishonest participants.

4 Density of Sellers

The main drawback of the credibility assessment method presented in Sec. 3 is the fact, that the credibility of a user depends on the credibility of other users directly involved in auctions with a given user. This encourages us to propose a novel measure of user credibility that indirectly employs information about participating users. We restrict our measure only to sellers, because the credibility of sellers is more important from the economical point of view (buyers risk financially more than sellers when involved in auctions with unreliable sellers).

We say that two sellers s_i and s_j are *linked* if there exist at least min_buyers who finalized an auction with both sellers, and the final price of each auction was at least min_value . The number n_{ij} of such buyers is called the *strength* of the link and is denoted as $|link(s_i, s_j)|$. We define the *neighborhood* $N(s_i)$ of a seller s_i as the set of sellers $\{s_j\}$ with whom the seller s_i is linked, given user-defined thresholds min_buyers and min_value . We call the cardinality of the neighborhood $N(s_i)$ the *density* of the neighborhood. The threshold min_buyers is used to select only sellers with significant volume of sales. The threshold min_value is used to protect against cheaters trying to impersonate credible sellers. Note that this measure does not take into account the strength of the link between

any two sellers, only the number of other sellers in a given seller's neighborhood. Therefore, we introduce another measure, called *score*, and defined as

$$score(s_i) = \sum_{s_j \in N(s_i)} density(s_j) * \log_{min_buyers} |link(s_i, s_j)|$$

The rationale behind the density measure is the following: a buyer who buys from two different sellers acknowledges the quality of both sellers. Experienced buyers who buy many items are used to link sellers, so we naturally discard less reliable information from unexperienced buyers. The fact that two sellers are linked indicates that either their product offer is very similar (e.g., both sellers sell used books and there are many buyers who buy books from both sellers), or that their offer is complementary (e.g., one seller sells computers and the other seller sells peripherals). Of course, the link between any two sellers may be purely coincidental, but we believe that this is the case of sellers with low density. Clusters with high density represent groups of very credible sellers. The score measure uses density of each seller in the neighborhood of the current seller and multiplies it by the strength of the link between the two. The logarithm is used to reduce the impact of very strong links between sellers.

The main benefit of the density measure is that it is very resistant to fraud. Consider a malicious seller trying to enter a cluster of reliable sellers. Linking to a single seller requires the cheater to create *min_buyers* and investing at least *min_buyers*min_value* in winning auctions of a credible seller. Such investment still links the cheater to only one seller, so in order to receive higher density the cheater has to repeat this procedure several times. This feature of our measure is caused by the fact that it uses other sellers to rate a current seller, rather than using information from buyers. Several studies show that it is much easier for malicious users to create false buyer identities (e.g., to provide artificially high positive feedbacks) than to form cooperating cliques of real sellers.

5 Experimental Results

The data have been acquired from www.allegro.pl, the leading Polish provider of online auctions. Allegro uses a simple auction protocol: each auction has an explicit deadline and all current bids are exposed to all participants, users may use a proxy which performs stepwise bidding until the maximum bid defined by a user has been reached. The dataset contains information on 440 000 participants and 400 000 terminated auctions (with 1 400 000 bids). We have chosen 10 000 different sellers and for this group we have selected all their auctions and participants of these auctions during a period of six months. Analogously, we have selected 10 000 buyers and for this group we have collected information on all auctions and their participants during a period of six months. Therefore, we had access to full information on 20 000 users, and partial information on another 420 000 users. All experiments are conducted on Pentium IV 2.4 GHz with 480 MB of memory. Data are stored and preprocessed using Oracle 10g database.

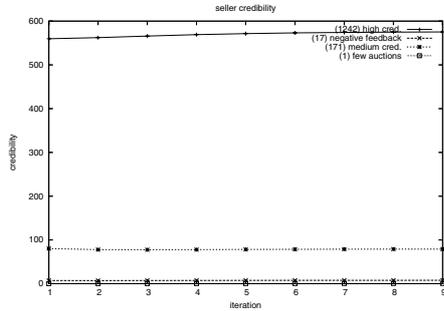


Fig. 2. Seller credibility

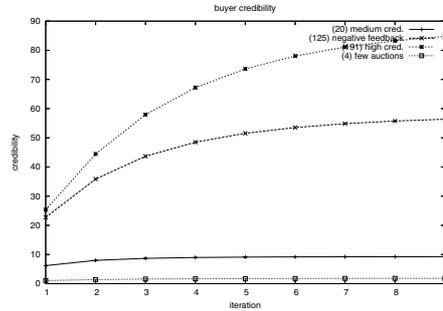


Fig. 3. Buyer credibility

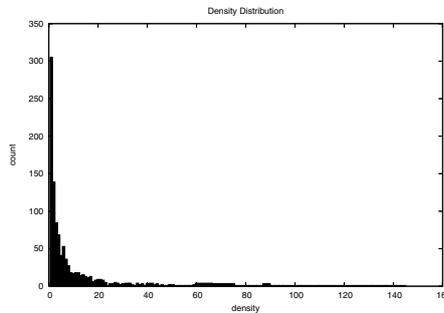


Fig. 4. Density distribution

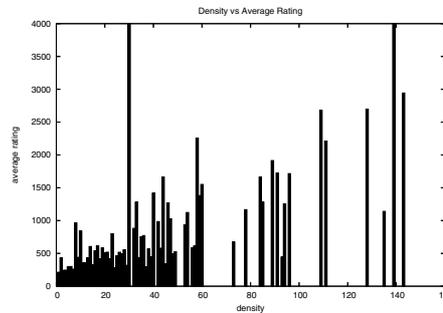


Fig. 5. Avg. rating distribution

Figure 2 depicts the convergence of credibility in subsequent iterations for selected sellers. Numbers in parentheses represent currently used reputation counts for those sellers. Observe that the relative difference in credibility for sellers (1242) and (171) is smaller than for reputation counts. Also note that credibility estimation converges after only a few iterations. The results of credibility estimation for a group buyers are depicted in Fig. 3. One can easily notice that the credibility as defined by us is not a linear function of reputation counts. Again, we observe that the computations converge after only a few iterations.

For testing of density and score measures we used the set of 10000 sellers for whom we had data on all auctions during the period of six months. We choose threshold values $min_buyers = 3$ and $min_value = 50$ PLN (ca. \$15). Figure 4 presents the distribution of density in the examined set. Most sellers have density lower than 6, but we also observe sellers with very high density exceeding 100. Around 10% of all examined sellers turned out to be dense (1026 out of 10000). Figure 5 presents the distribution of average rating with respect to density. This result supports our claim that high density represents reliable and credible sellers, one can easily notice that the average rating grows quickly with the increase of the density.

Figures 6 and 7 show projections of density and score on currently used rating value. Interestingly, we find many sellers with very high rating (above 500 auc-

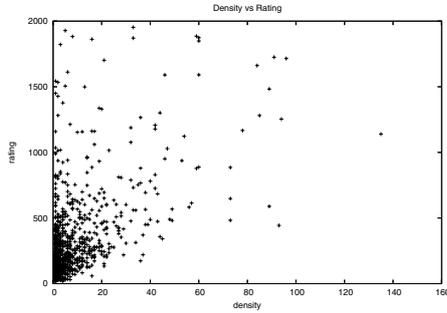


Fig. 6. Density vs Rating

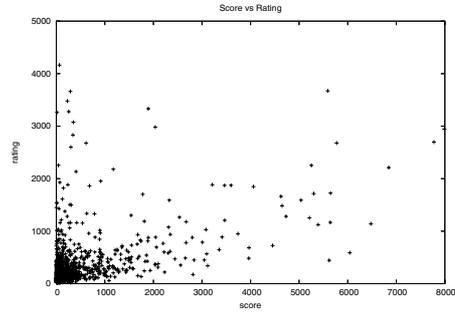


Fig. 7. Score vs Rating

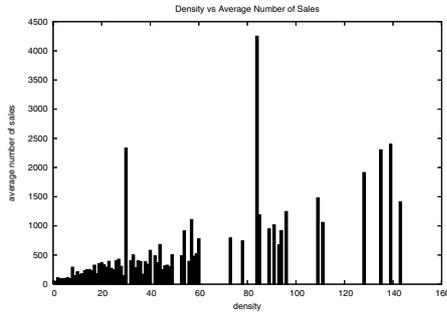


Fig. 8. Avg. number of sales distribution

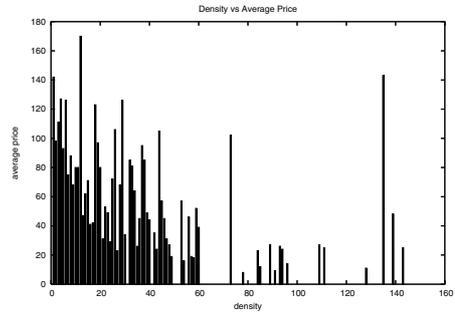


Fig. 9. Avg. price distribution

tions) with very low density. Of course, these could be users who trade low-value goods (e.g. used books, small collectibles, etc.) and they are being punished by relatively high *min_value* threshold. On the other hand, the average price of auctions in the examined set was close to 90 PLN (ca. \$28), so we do not feel that our setting was too prohibitive. Figure 7 reveals a significant shift along the x-axis. This suggests that the sellers with low density and high rating have much higher average strength of the link than densely connected sellers.

Finally, Fig. 8 and 9 show the impact of density on the average number of sales and the average price of items for each seller. We discover that the density is a good predictor of the volume of sales, and sellers with high density enjoy a much higher number of auctions. On the other hand, there is no clear indication whether the density of a seller impacts the average price of offered goods.

6 Conclusions

In this paper we have presented two novel measures for reputation rating in online auction environments. Our measures, credibility and density, evaluate the reputation of auction participants by mining the topology of the network of seller-buyer relationships to retrieve useful knowledge. We believe that the

patterns that we discover in the network provide additional insight into user characteristics and can be used as a predictor of user reliability. The experiments prove the practical usability and applicability of the presented solution. Our work extends previous proposals in the number of ways, namely, it considers the structure of inter-participant relationships and computes the reputation ratings iteratively by simulating the flow of credibility between auction participants. The results of conducted experiments encourage us to follow the research in the area. For the moment, we compute the reputation of a buyer and a seller disjointly, even if they are the same physical person. The next step is to combine the information about these reputation ratings for every distinct individual. An interesting, yet often disregarded, feature of online bidding is the timing of the bid. We believe that the timing carries valuable knowledge about the nature of the bid and can be successfully used to discover fraud. Finally, our experiments were conducted on a fairly small subset of the original data. Due to the immense popularity of online auctions, the volume of data to be analyzed poses a significant challenge. We plan to scale the algorithms to allow for almost real-time analysis of huge amounts of raw data.

References

1. K. Aberer and Z. Despotovic. Managing trust in a peer-2-peer information system. In *CIKM '01: Proceedings of the tenth international conference on Information and knowledge management*, pages 310–317, New York, NY, USA, 2001. ACM Press.
2. S. Ba, A. B. Whinston, and H. Zhang. Building trust in online auction markets through an economic incentive mechanism. *Decision Support Systems*, (35), 2003.
3. M. Chen and J. P. Singh. Computing and using reputations for internet ratings. In *EC '01: Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 154–162, New York, NY, USA, 2001. ACM Press.
4. R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 403–412, New York, NY, USA, 2004. ACM Press.
5. D. Houser and J. Wooders. Reputation in auctions: Theory, and evidence from ebay. Technical report, University of Arizona, 2001.
6. J. M. Kleinberg. Authoritative sources in a hyperlinked environment. In *SODA '98: Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, pages 668–677, Philadelphia, PA, USA, 1998.
7. R. A. Malaga. Web-based reputation management systems: Problems and suggested solutions. *Electronic Commerce Research*, 4(1), 2001.
8. M. I. Melnik and J. Alm. Does a seller's ecommerce reputation matter? evidence from ebay auctions. *The Journal of Industrial Economics*, L(3), September 2002.
9. P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems. *Communications of the ACM*, 43(12), 2000.
10. P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on ebay: a controlled experiment. Technical report, 2004.
11. J. M. Snyder. Online auction fraud: Are the auction houses doing all they should or could to stop online fraud? *Federal Communications Law Journal*, 52(2), 2000.
12. L. Xiong and L. Liu. A reputation-based trust model for peer-to-peer ecommerce communities. In *EC '03: Proceedings of the 4th ACM conference on Electronic commerce*, pages 228–229, New York, NY, USA, 2003. ACM Press.