Reputation Premiums in Electronic Peer-to-Peer Markets: Analyzing Textual Feedback and Network Structure

Anindya Ghose
aghose@stern.nyu.edu

Panagiotis G. Ipeirotis
panos@stern.nyu.edu

Arun Sundararajan
asundara@stern.nyu.edu

Department of Information, Operations, and Management Sciences
Stern School of Business, New York University
44 W. 4th Street, New York, NY 10012-1126, USA

ABSTRACT

Web-based systems that establish reputation are central to the viability of many electronic markets. We present theory that identifies the different dimensions of online reputation and characterizes their influence on the pricing power of sellers. We provide evidence that existing, numeric reputation scores conceal important seller-specific dimensions of reputation and we validate our theory further by proposing a new text mining technique that identifies and quantitatively evaluates further dimensions of importance in reputation profiles. We also suggest that the buyer-seller network contains critical reputation information that we can further exploit to improve the design of a reputation mechanism. Our experimental evaluation validates the predictions of our model using a new data set containing over 12,000 transactions for consumer software on Amazon.com's online secondary marketplace. This paper is the first attempt to integrate econometric methods and text and link mining techniques towards a more complete analysis of the information captured by reputation systems, and it presents new evidence of the importance of their effective and judicious design.

Categories and Subject Descriptors

K.4.4 [**Computing Milieux**]: Computers and Society—*Electronic Commerce*; H.4 [**Information Systems Applications**]: Miscellaneous; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*

General Terms

Algorithms, Measurement, Design, Economics, Experimentation, Theory

Keywords

reputation, electronic markets, econometrics, peer to peer markets, text analysis, text mining, opinion analysis, transaction network analysis

SIGCOMM'05 Workshops, August 22–26, 2005, Philadelphia, PA, USA.

1. INTRODUCTION

Online reputation mechanisms play a central role in the viability of many electronic trading networks in which trade occurs directly between peers. A fairly extensive recent literature (see [1] for a review) has studied the effective and reliable design of online reputation systems for sellers in mediated electronic markets like eBay. In such markets, reputation typically consists of a feedback profile that provides the following information to potential buyers:

- The number of transactions the seller has successfully completed,
- The scores (or ratings) provided by the buyers who have completed transactions with the seller,
- A chronological list of textual feedback posted by buyers who evaluate the quality of the transactions they have conducted with the seller in the past.

An increasing fraction of peer-to-peer exchanges now take place on trading networks that are not mediated by a central authority; the most visible of such networks are those used for file sharing, such as Kazaa and Grokster. As these networks evolve towards being the platform for more complex trade, rather than just for the free exchange of files, online reputation takes on an increasingly important role in these decentralized trading environments. A similar evolution occurred on Usenet groups in the 1990's, many of which were used as electronic secondary markets, and which, in the absence of a central mediating authority, used purely text-based feedback as their reputation mechanism. Currently, there are some simple structured methods of establishing "reputation" of sorts on a peer-to-peer network (for example, by using Altnet's Peer Point Manager on Grokster), and there are proposals for establishing reputation through certification [2], for identifying low performing peers [5], and for the rank-based selection of peers in P2P media streaming [3].

Most studies of online reputation thus far base a trader's reputation on a single number (based on the numerical rating) that characterizes the seller. Our study is based on the conjecture that text-based feedback plays a substantial and important role in establishing reputation in decentralized peer-to-peer trading environments. Our observation of mediated electronic markets suggests that different sellers in these markets derive their reputation from different characteristics: some sellers have a reputation for fast delivery, while some others have a reputation of having the lowest price among their peers. Similarly, while some sellers are praised for their packaging, others get good comments for selling high-quality goods but are criticized for being "rather slow" with shipping. This is potentially useful reputation information, but is often buried in text and its value has not been well-understood in the literature. Moreover, if this feedback can be filtered using a more sophisticated interface, it could improve the efficiency of decentralized electronic markets; summarizing it intelligently based on an automated analysis of its text could benefit traders who do not have the cognitive ability to extract its full value. Our research aims to therefore develop techniques for extracting the information contained in text-based feedback, and then develop and test a model of the value of this information.

Moreover, when each trader transacts with just a small subset of others, there is information in feedback that pertains to repeat transactions, and this is of at least two kinds. First, it is quite likely that a trader who has transacted with another in the past relies not just on the latter's overall reputation, but on his or her own experience with the trader (and it is likely that the latter is given more "weight"). Second, when repeat trade occurs, this might make a trader's prior positive feedback more credible (or negative feedback less credible). Additionally, observing how feedback from transactor pairs evolves over time might give new traders information about the reliability of a specific trader, and enable them to better "benchmark" the feedback from a specific peer. Specific traders may prove to be reliable "indicators" of specific categories of trade (or they may become "hubs"). Clearly, these two issues (the value of text-based feedback, and the way traders use the information in a profile) are related.

This introductory summary frames our research questions, which are summarized below:

- What is an appropriate way of representing the information contained in an online reputation? Are numerical scores sufficient? If yes, then what is the purpose of text-based feedback? If not, how text-based feedback improves a trader's understanding of "true reputation"?

- What is the relationship between an appropriate measure of online reputation (that potentially contains both numerical information and information extracted from text-based feedback) and the "trading premium" it leads to? An example of trading premium might be the incremental price premium associated with an increase in measured reputation.

- Can taking the structure of the network of transactions between peers into account provide superior information about the reputation of a trader?

- How might these results influence the effective design of online reputation systems for decentralized peer-to-peer electronic markets?

The research we have done thus far towards answering these questions makes the following contributions:

- We have developed a novel text analysis technique for extracting customer sentiment from textual feedback, and relating the extracted sentiment to an enhanced measure of reputation.

- We have developed an economic model of the value of online reputation where traders are heterogeneous in the "quality" of their trades, and where peers value differently the various aspects of reputation. This model also accounts for the structure of interactions between peers.

- We have performed a preliminary econometric study using data from the peer-to-peer secondary market of Amazon.com that relates our reputation variables to the price premium a seller enjoys, and found support for these predictions.

The rest of the paper is organized as follows. Section 2 outlines the economic model. Then, Section 3 describes the data that we used for the econometric analysis, which is presented in Section 4. Section 5 describes our text analysis algorithm, and Section 6 shows some preliminary results. Section 7 briefly covers our network analysis method and Section 8 concludes the paper.

2. ECONOMIC MODEL

We model an electronic secondary marketplace in which m competing sellers offer a single product (a generalization to multiple products and varying numbers of sellers is straightforward). There are M buyers of this product and each buyer values n different *fulfillment characteristics*. Examples of these characteristics might be speed of delivery, quality of packaging, post-sale support and so on. Buyers also differ in the extent to which they place importance on each of these characteristics, and each buyer is therefore indexed by a type vector $w = (w_1, w_2, ..., w_n)$, where a higher value of w_i indicates that the buyer places a relatively higher value on fulfillment characteristic i. Each buyer's type is drawn from a common distribution with distribution function $F(w)$, which we assume is symmetric.

Correspondingly each seller is indexed by a characteristics vector $X = (X_1, X_2, ..., X_n)$, where X_i represents the seller's ability to provide the i^{th} dimension of fulfillment. Each X_i is a random variable with mean x_i. For simplicity, we assume that these random variables vary across sellers only in their means. In other words, X has the distribution function $G(\cdot, x)$ where $x = (x_1, x_2, ..., x_n)$ is the vector of mean values that completely characterizes a seller. We also assume that the products sold by each seller are of identical quality (which is consistent with our data set comprising consumer software, though adding an extra characteristic x_0 to represent product quality would not affect our results).

When a buyer with type w purchases a product from a seller of type x, there is a *realized* value of fulfillment (the quality of fulfillment provided by the seller on that specific transaction) $z = (z_1, z_2, ..., z_n)$, which is a random draw from the distribution $G(\cdot, x)$. If the price charged by the buyer is p, the value that the buyer gets from this transaction is

$$u(w, z) - p, \tag{1}$$

where $u(w, z)$ is increasing in each component of its arguments. For example, $u(w, z)$ might be a weighted average of the realized fulfillment values. After each transaction, the buyer posts a feedback set which contains the seller's ID, the buyer's ID, and information about the fulfillment on that transaction. There is consequently a feedback set $t_k = \{s_k, b_k, \phi_k\}$ associated with each transaction k, where the value of s_k identifies the seller, the value of b_k identifies the buyer, and ϕ_k contains information about the quality of fulfillment. In most reputation systems, ϕ_k contains a numerical score rating the overall quality of the transaction, along with unstructured text describing some of the dimensions of the transaction. In our model, we assume that ϕ_k simply contains an accurate report by the buyer of the realized value z of the fulfillment characteristics vector for that transaction, though this is not essential for what follows

At any point in time, prior to engaging in a transaction, each buyer has available the entire set of feedback sets $T = \{t_1, t_2, ...\}$. The buyer can therefore associate the reputation profile S_i of seller i

$$S_i = \{t_k \in T, s_k = i\}$$

that contains the feedback sets for all transactions in which seller i participated. Typically, the reputation system of the electronic marketplace provides this to the buyer, for any seller. The *average reputation* of seller i is simply the vector containing the mean of each component of ϕ_k for those feedback sets $t_k \in S_i$, and the *level of experience* of seller i is the size of its reputation profile $|S_i|$.

We consider two possible equilibrium concepts. The first is a Bayes-Nash equilibrium, in which the choices of price is a specific equilibrium of a game of incomplete information. [1] The second (simpler) concept is one in which sellers choose prices "competitively": that is, prices are chosen so that, in equilibrium, each seller has a *equal expected profit* from their choices. Under either of these equilibrium concepts, our model leads to the following propositions:

PROPOSITION 2.1. *For any two sellers with the same level of experience, the equilibrium price of a seller with a higher average reputation is higher than that of a seller with a lower average reputation.* □

PROPOSITION 2.2. *For any two sellers with the same average reputation, the equilibrium price of a seller with a higher level of experience is higher than that of a seller with a lower level of experience.* □

3. DATA DESCRIPTION

We have compiled a market-level data set on a cross-section of software vendors, encompassing several different categories. Our data are compiled from publicly available information on used software product listings at Amazon. The data are gathered using automated Java scripts to access and parse HTML pages downloaded from the retailer. The data was collected over an 180 day time period from October 2004 to March 2005 and is still ongoing. It includes 280 individual software titles. This panel includes an equal number of software products from each of the major categories. We use software because it is generally of uniform quality, and allows us to separate the trading reputation of a seller from the quality of products the seller offers.

Our marketplace data includes the price, condition, and seller reputation for each used product listed for sale. Condition is self-reported by the seller and can be either "like new," "very good," "good," or "acceptable." The reputation data from Amazon's marketplace, includes a summary of scores (or ratings) provided by buyers who have completed transactions with the seller in the past. In addition to (or instead of) an average over all scores obtained over the seller's life time, an average of scores obtained more recently (30 days, 90 days and 365 days, for example) is reported for each of the three categories: positive, neutral and negative. Thus we are able to see how a seller's feedback profile has changed over time, and also construct the network of buyer-seller pairs. The variables in our dataset consist of sale price, seller ratings over different time periods, product's condition, competitors' prices, competitors' ratings over different time periods, competitors' product conditions, and price premium.

Importantly, we are able to infer using Amazon.com's XML data feed, which listings on the secondary market result in actual trades, when these trades occur, and what the competing prices were at the time. Our sellers consist of both individuals and larger well-established sellers. While there are a number of one-time transactions, our buyer-seller network consists of over 9000 pairs of buyers and sellers who have transacted with each other at least twice, and the degree distribution of this network follows a power law, which is as we anticipated. The weights on this bipartite multigraph correspond to the ratings we associate with each transaction.

The second part of our data set contains the reputation history of each seller who had listed a product for sale during our 180 day window. Each of these sellers has a feedback profile, which as described earlier, consists of numerical scores and text-based feedback left by buyers. The numerical ratings are provided on a scale of 0 to 5 stars. All ratings lower than three are denoted as negative , those ratings above 3 are denoted as positive, and therefore, a rating of 3 is categorized as neutral. These ratings are averaged to provide an overall score to the seller. Amazon also reports similar averages over the last 30 days, 90 days and 365 days, for each of the three categories: positive, neutral and negative.

4. ECONOMETRIC ANALYSIS

We validate the predictions of our theory in two steps. In our first step, we focus on the numeric feedback scores reported by buyers, and ignore all text-based feedback completely. For each transaction, we define the *PricePremium*, which is the difference between the price at which the transaction occurred, and the average price of its competitors. If, as predicted by our theory, a seller with a higher average reputation or a higher level of experience can charge a higher price, then this seller should enjoy a higher price premium. The main variables are the price at which the transaction occurs (*SalePrice*), the average value of the seller's numerical scores (ignoring text-based feedback) over their entire transaction history (*SRating*), and the total number of seller transactions (*SLife*),which measures the seller's level of ex-

[1] Defining this game more precisely and deriving its equilibrium requires a little more simplification. Since the details of this are not central to the rest of the paper, they are omitted.

Variable	Ln[PricePremium]
Constant	$-3.04(0.56)$
Ln[SalePrice]	0.84(0.004)
Ln[SRating]	0.884(0.402)
Ln[SLife]	0.089(0.037)
Ln[Condition]	0.189(0.05)
	$R^2 = 37.8\%$

Table 1: The effect of average reputation and level of experience on pricing power, controlling for unobserved heterogeneity across sellers. Values in parenthesis are the standard errors.

perience. We estimated equations of the following form:

$$Ln[PricePremium] = \alpha + \beta_1 Ln[SalePrice] + \beta_2 Ln[SRating] + \beta_3 Ln[SLife] + \beta_4 Ln[Condition]$$

These OLS regressions were estimated alternatively, with seller and product fixed effects, which enabled us to control for both, unobserved heterogeneity across sellers, and for unobserved heterogeneity across products. Both regressions yielded qualitatively similar results.

The results of these estimations are presented in Table 2. These results strongly support our hypothesis: both average seller reputation (*SRating*) and the seller's level of experience (*SLife*) have a positive and significant effect on pricing premiums. Notice that the average price premium changes much more rapidly with changes in average reputation and experience than the price premium relative to one's nearest competitor, which is interesting. The coefficient of *Ln[SalePrice]* is significant and less than 1 in each case, indicating that while the magnitude of the price premium increases with sales price, it decreases in percentage terms. This is not surprising, and consistent with our model treating the seller's reputation as measuring characteristics that have to do with fulfillment rather than the product itself (therefore, the premium does increase, but not proportionate to the increase in sale price).

5. TEXT ANALYSIS

The text analysis part processes the chronological list of textual feedback. The goal of the text analysis technique is twofold:

- Discover the fulfillment dimensions that contribute to the reputation of each vendor and the weight of the contribution.
- Describe in quantitative terms the textual evaluations given by the users (e.g., "cool packaging" is better than "very good packaging").

The basic idea is to break down the overall reputation of a seller into *"micro-reputations"* for each of the discovered dimensions (delivery speed, packaging, responsiveness) and examine how differences in the micro-reputations are reflected in the price premiums. Formally, we assume that each feedback posting evaluates the seller in each of the n fulfillment dimensions $X_1, \ldots, X_n$ that contribute to the overall reputation, by assigning a score $a_1, \ldots, a_n$ for each one of them. A positive score corresponds to a positive evaluation and a negative score corresponds to a negative evaluation.[2]

In order to discover the different fulfillment dimensions across which the sellers are evaluated, we scan the feedback postings and we keep all the nouns and noun phrases that appear in the textual feedback. These nouns will serve as the initial, expanded set of dimensions across which the sellers are evaluated. We are currently working on methods that identify nouns and noun phrases that are used to describe the same characteristic of a seller (e.g., "shipping" and "delivery" refer to the same fulfillment dimension.) Each of these micro-reputations of the sellers contributes with a given weight $w(X_i)$ to the overall reputation of the seller. For now, we assume a linear combination of weights to create overall reputation.

Of course, in textual feedback the users do not assign explicitly numeric scores. Rather, they use adjectives to evaluate the seller. (e.g., "*fast* delivery," "*slow* delivery" and so on). To illustrate this with an example: suppose dimension X_1 is "delivery", dimension X_2 is "packaging", and dimension X_3 is "service". A feedback set ϕ_1 which contains the posting *"I was impressed by the speedy delivery! Great service!"* is then encoded as

$$\phi_1 = [speedy, NULL, great],$$

while a feedback set which contains the posting *"The item arrived in awful packaging, and the delivery was slow"* is encoded as

$$\phi_2 = [slow, awful, NULL].$$

In our approach we assume that each adjective is used to assign a score to the respective fulfillment dimension with which it is associated. In order to assign a "value" to this reputation profile, we have also developed and implemented a method for inferring the numerical scores that should be associated with each adjective, for each dimension. Our technique exploits the residuals from the OLS analysis presented in Section 4. After eliminating the effect of all factors that can increase the price premium (e.g., condition of the product – see Table 1), we examine how differences in the reputation postings change the price premiums. For example, everything else being equal, a seller with "speedy" delivery charges \$10 than a seller with "slow" delivery. Using this information, we can conclude that "speedy" is better than "slow" and when applied to the dimension "delivery" can increase the price premium by \$10.

6. TEXT ANALYSIS RESULTS

Based on these factors, and an appropriate choice of weights, we assess the reputation value of each element (i.e., noun-adjective pair) that appears sufficiently frequently in our text feedback set. We use the *residuals* from our product fixed-effects regressions to score each of these elements, thereby isolating the price premium associated with the element *after* accounting for the seller's numerical reputation score and level of experience.

Some results from this assessment are summarized in Table 2, for those elements that emerged as having the highest positive and negative impact on a seller's reputation. These results demonstrate that having specific text elements in

[2]Each feedback posting evaluates the seller in only a limited number of fulfillment dimensions; for each non-evaluated fulfillment dimension X_i, we set $a_i = 0$.

[Modifier, Dimension] Pair	Weight
[wonderful, product]	17.99
[perfect, transaction]	10.49
[fast, shipping]	8.16
[friendly, service]	4.30
[excellent, packaging]	4.06
[pristine, condition]	3.42
[never, responded]	0.02
[not, received]	0.11
[wrong, item]	0.15
[took, forever]	0.25
[wrong, address]	0.29

Table 2: Summary of the dimension-modifier pairs in text-based feedback that influence a seller's pricing power most strongly. A weight higher than 1 indicates a positive effect, while a weight lower than 1 indicates a negative effect.

one's feedback profile lead to an economically significant impact on a seller's pricing power. Our work in progress aims to quantify this impact in terms of realized pricing premiums.

7. NETWORK ANALYSIS

We are also extending our model to incorporate ideas from Kleinberg's "hubs and authorities" model [4]. The seller-buyer transaction network is a bi-partite graph, similar in style to the hubs and authorities model. We use the weighted reputation multigraph in which nodes are buyers/sellers and edges are reputation scores. In out case, the "good" sellers serve as authorities and the "good" buyers ("power buyers" might be a preferable term) serve as hubs. Specifically, we assume that the "power buyers" spend more time evaluating sellers and tend to buy more from reputable sellers. (The notion of "authority" here is how "effective" a seller is, which is the seller's reputation is supposed to be, in the first place). Hence, we can use the clues provided by the network structure and derive better reputation score by weighting buyer ratings based on how good a "hub" the buyer is.

8. CONCLUSIONS

We have presented a new approach for identifying and quantifying the dimensions of value from online reputation. Our approach characterizes how both numerical and qualitative measures of reputation affect a seller's pricing power in a mediated electronic secondary market. We have validated the predictions of this theory by combining the results of the estimation of an econometric model with a novel text analysis technique. To the best of our knowledge, this represents the first study of this kind, and the first set of results that establishes the value of information contained in the text-based feedback of an online reputation system.

Our analysis of the information in qualitative text feedback is likely to gain importance as the fraction of peer-to-peer exchanges taking place on trading networks that are not mediated by a central market maker increases. We hope our study will pave the way for future research in this area.

9. REFERENCES

[1] C. Dellarocas. The digitization of word-of-mouth: Promise and challenges of online reputation mechanisms. *Management Science*, 49(10):1407–1424, 2003.

[2] P. Dewan and P. Dasgupta. Pride: Peer-to-peer reputation infrastructure for decentralized environments. In *Proceedings of the 13th International World Wide Web Conference (WWW13)*, pages 480–481, 2004.

[3] A. Habib and J. Chuang. Incentive mechanism for peer-to-peer media streaming. In *12th International Workshop on Quality of Service - IWQoS 2004*, pages 171–180, 2004.

[4] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of ACM*, 46(5):604–632, 1999.

[5] T. Papaioannou and G. D. Stamoulis. Effective use of reputation in peer-to-peer environments. In *Proceedings of IEEE/ACM CCGRID 2004 (Workshop on Global P2P Computing)*, 2004.