An Approach to Implement A Trading Network Visualizing System for Internet Auctions

Masao Kobayashi[†] **Takayuki Ito**[‡]
[†]Department of Computer Science, Nagoya Institute of Technology, Japan
kobayashi@itolab.mta.nitech.ac.jp
[‡]Techno-Business School / Department of Computer Science
Nagoya Institute of Technology, Japan
ito.takayuki@nitech.ac.jp

Abstract

As Internet auctions have increased, so too has auction fraud. This paper describes the design of a system that supports fraud detection and market prediction by visualizing transaction networks on Internet auctions. Our system employs link mining techniques and user information on an Internet auction. At this stage, we successfully showed visualized trading networks by extracting trading histories from an auction site. Using a visualized graph, the system shows suspiciousness with user ID information. Also, the system presents trading relationships as a network structure in various viewpoints. Furthermore, the system possesses an explanation function on a visualized trading network that predicts which buyers and sellers are active and did better behaviors. Our preliminary experiment demonstrates that the graph presentation function is scalable enough against the number of sellers and buyers. In discussion, We considered this system from a point of view of formality and mentioned eigenvalue in particular at this time.

Keywords: Internet Auction, Visualize, Graph, Transaction

1 Instructions

In recent years, as Internet auctions increase, reports of such illegal acts as fraud have also begun to surface in the media. In its monthly report for March 2007, Yahoo! Auction, the largest Japanese Internet auction site, announced that the number of monthly login IDs has surpassed 19 million and the number of exhibitions at auctions is over 12 million cases per day. Furthermore, the number of login IDs exceeds 5 million in only an auction section. The second largest auction site in Japan, Rakuten Auction, has 3 million IDs. Unfortunately, fraud victims are also increasing. In 2006 December, over 990 cases of fraud were reported totaling losses of ￥88 million (about $725,100).

Notice the difficulty of judging how to sell or bid safely and make a better market because in Internet auctions, relations between persons or between a person and an article are immediate. Therefore, we decided to find new values by treating an Internet auction as a single network. We intended to find a similarity with an existing problem by regarding Internet auction as a network structure.

In addition to network structure, we focus on user reputation information provided by other users. Most auction and shopping sites are equipped such reputation mechanisms. Such ratings unfortunately lack exact and correct information. However, we can learn how the player acts based on his/her trading history. This information warns novices to be sensitive and aware of auction participation. We also utilize other extractable information from auction sites, such as the amount of buying, selling, and bidding.

In this work, the transaction relationship on Internet auction sites is treated as a network structure with information, and we present a user-friendly visualization system that has the following three points:

1. This system supports the identification of fraud from transaction relationships and detects suspicious behavior with such user ID information as selling/buying history.

2. This system presents a transaction relationship as a network structure from various viewpoints. Users can view the network from various angles such as seller or buyer.

3. This system has explanation functions about the visualized network. This function

predicts auction markets.

Many researchers have realized the importance of fraud detection and methods and systems. There are some similar works. NetProbe [1] extracts a network from e-bay auction sites and tries to identify possible fraud. They actually only focus on network structure without utilizing such user information as reputation ratings, selling/buying histories, etc. We, however, use such private information to identify suspicious acts. Other related work will be shown in Section 2.

The rest of this paper is organized as follows. We review related work and the present condition of auction fraud in Section 2. Then we describe the detailed system architecture in Section 3. Next, in Section 4, we report experiments that evaluate this system with user IDs on actual auction sites. In this paper, we also evaluated execute time. In Section 5, we discuss future systems and evaluation methods. Finally, we summarize our findings in Section 6.

2 Related Work

In this section, we survey related approaches for fraud detection in auction sites as well as the literature on reputation systems typically used by auction sites to prevent fraud. We also look at related works on trust or authority propagation, reputation systems, and link mining that could be applied to predict network links. In addition, we describe related work on market prediction and theoretical results.

2.1 Trust or authority propagation

Trust propagation is used by TrustRank [2] to detect Web spam. The goal is to distinguish between " good " and " bad " sites. Because web spam pages employ various ways to achieve higher ranking in search engine results, we filter only " good " pages. Authority propagation has been studied in Web searches. PageRank [3] uses hyperlinks as " votes. " In this concept, a page linked by many " good " pages must after all be a " good " page. In effect, the importance of pages is propagated over hyperlinks connecting them. HITS [4] represent this concept as hub and authority pages. Authority indicates abundant information about a specific topic. Hub indicates that links to high value pages have authority, which is abundant. In general, " good " sites about a specific topic have a high evaluation value of authority, and a " good " collection of links has a high evaluation value of hub.

2.2 Reputation system

Auction sites use reputation systems to prevent fraud. Comment of a user is one of the persuasive evaluation indexes. But they are usually simple and can be easily foiled because it is difficult to detect honest behavior and show that a user 's reputation reflects actual intention [5]. Resnick et al. [6] and Melnik et al. [7] show that reputation systems might not be effective to prevent fraud.

2.3 Link mining

Link mining is a kind of data mining and a focus graph structure that consists of nodes and edges. Pattern mining and graph mining resemble link mining [8]. Many kinds of tasks are included such as link prediction, classifying nodes, ranking nodes, and subgraph discovering. [9] The subgraph discovery is a very important task. It is thought that we can use that we find a subgraph having a similar pattern for the detection of similar preference of a user and detection of a fraud group. A link prediction can be defined as: " when a known part of a network structure is given, I predict an unknown part to a clue in this. " This idea can be identified as the presence of a link between two nodes in fraud detection.

2.4 Theoretical results

Many theoretical results are related to false-name proof auction design. Such auctions [10][11][12] guarantee that no player benefits from virtual or dummy players who make false bids. However, these results are quite theoreti-cal. In particular, they focus on VCG [13][14][15], the most theoretically desirable combinatorial auction that has not yet been employed in the practical world due to sever-al application problems. On the other hand, in our pro-posed system, we practically construct a system that can reveal cheats or predict targeted markets.

3 Transactional Relationship Visualization System

3.1 Outline of the system

This system consists of three parts. The first part selects transaction relationship records from Internet auction sites. The second displays graphs consisting of the characteristic elements of the extracted records. The third is an instruction part with information from selected transaction records. We show a conception diagram of the system in Figure 1. On this system, the user queries with an auction ID or a product name. With the query entry, the Java program accesses the Internet auction site and analyzes the HTML sources of the auction site pages. The Java programs select data available for visualizing transaction relationships. Then, user ID information from the auction site is stored in the user database. We can chronologically analyze information by adding acquisition data to the time tags. Finally, a visualization transaction relationship network represents users by graphs. We describe system implementations below.

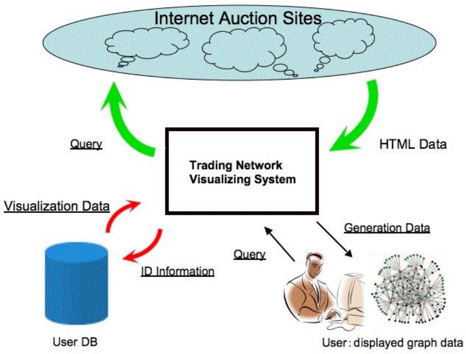


Figure 1. Outline of proposed system

3.2 Visualization of Transaction Relationship

We first have to get transaction data from Internet auction sites to visualize an auction network. We use Yahoo! Auction, the largest auction site in Japan, for implementation. We obtained transaction data and analyzed the HTML source codes of auction pages to exploit user IDs and articles. This system possesses review pages of personal ratings. If this page describes user A, it represents the ratings of A from partners who dealt with A. Therefore, this page indicates A 's transaction history.

Using the Magnetic-spring model [16], we make graphs that consist of nodes and edges. The spring model is one drawing method for a graph. The magnetic-spring model introduces magnetic theory into spring models. Figure 2 represents a graph about certain user IDs written in Java programs.

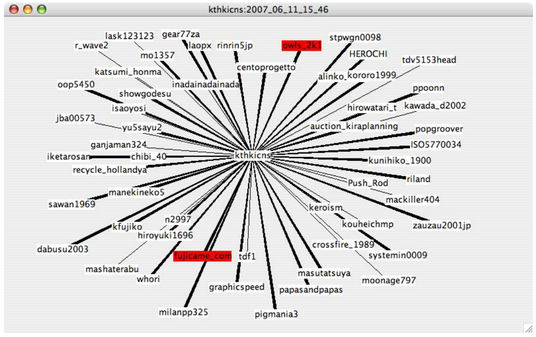


Figure 2. Visualization of transaction relationship

These edges represent seller and buyer relationships as differences of line width. Here, wide edges indicate that the user ID is a " successful bidder " for the trading partner. Narrow edges indicate that the user sold goods to the trading partner. In addition, double-clicking on the node forms a new graph. That is, search depth can be lowered by clicking the node. In the future we aim to visualize the entire Internet auction network. This function is the first step.

User ID rating value is one of the strongest component parts of graphs. Reputation systems are widely used in many Internet auction sites, making rating analysis important. However, the reliability of rating value must be verified in the future.

3.3 Introduction function with transaction information

This function captures the rating value of user IDs and represents the following seven items:

1. rating value of user ID
2. rating value average of traded partners
3. auction type of user ID
4. percentage of user ID was rated as " bad "
5. number of banned IDs traded with user ID
6. data table of traded partners
7. log data

First, the rating value of the user ID score is given by trading partner in Yahoo! Auction. This score is assigned 1point values ranging from -2 to +2. The first item is the total value. Second, the rating value average of trading partner is the total rating value to user IDs divided by trading partners. In other words, the value expresses the height of the grade of a trading partner of the user ID. Next, auction type represents that part occupied by seller or buyer user IDs in all transaction records as a percentage. For example, if there have been ten dealings, and six times the bidder was successful, the auction type is successful, and its numerical value is 60 % . The percentage of the user ID was rated as " bad " is the number of user IDs have been rated " bad " or " very bad " divided by all partners. This value may be calculated highly even if rating value of user ID, the first item, is high. If a value of the fifth item is high, the user ID is suspicious. Next, the number of banned trading partners represents IDs banned by Yahoo or those lapsed for any reason. Some cheats have multiple disposable IDs. In the generated graph, banned ID nodes are indicated in red, as in Figure 3. In a precise sense, a banned ID cannot identify bad users such as fraudsters. The kinds of banned IDs can be estimated [17]. Next, the data table of trading partners includes much partner data such as ID names, rating value, number of transaction times, rating to user IDs, and status like seller or buyer. Finally, the log data is a list of files that were output when visualizing the graph in this system.

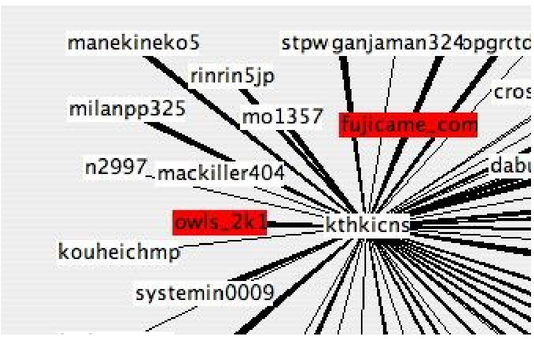
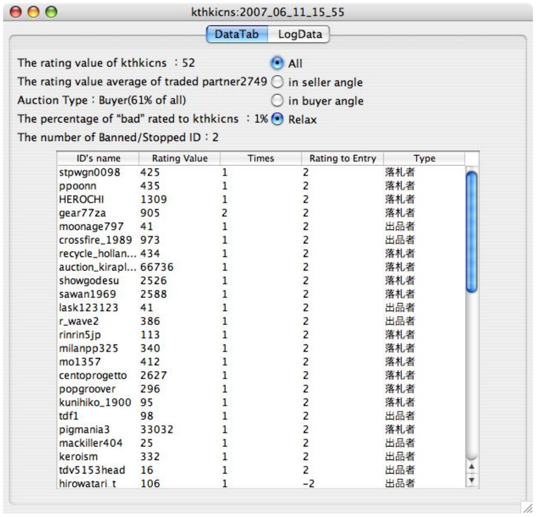


Figure 3. Banned IDs is displayed in red

The above items are displayed in one window that also has a function that switches the view of the graph. There are three types of view: all view, seller angle, and buyer angle. All view displays every transaction relationship, including both seller and buyer stances of user IDs. Seller angle displays the seller stances of user IDs with all wide edges. Buyer angle displays the buyer stances of user IDs with all thin edges. Figure 4 is an output window that queries with certain user IDs. In addition, whenever queries are made with certain IDs, the Java program takes logs with the present time. In the log data window, log files are listed in a text area.


kthkicns:2007_06_11_15_55

DataTab LogData

The rating value of kthkicns ：52
The rating value average of traded partner2749
Auction Type：Buyer(61% of all)
The percentage of "bad" rated to kthkicns ：1%
The number of Banned/Stopped ID：2

All
in seller angle
in buyer angle
Relax

ID's name	Rating Value	Times	Rating to Entry	Type
stpwgn0098	425	1	2	落札者
ppoonn	435	1	2	落札者
HEROCHI	1309	1	2	落札者
gear77za	905	2	2	落札者
moonage797	41	1	2	出品者
crossfire_1989	973	1	2	出品者
recycle_hollan...	434	1	2	落札者
auction_kirapl...	66736	1	2	落札者
showgodesu	2526	1	2	落札者
sawan1969	2588	1	2	落札者
lask123123	41	1	2	出品者
r_wave2	386	1	2	出品者
rinrin5jp	113	1	2	落札者
milanpp325	340	1	2	落札者
mo1357	412	1	2	落札者
centoprogetto	2627	1	2	落札者
popgroover	296	1	2	落札者
kunihiko_1900	95	1	2	落札者
tdf1	98	1	2	出品者
pigmania3	33032	1	2	落札者
mackiller404	25	1	2	出品者
keroism	332	1	2	出品者
tdv5153head	16	1	2	出品者
hirowatari_t	106	1	-2	出品者

Figure 4. Data window

4 Preliminary Experiment

We can use the reliability of user IDs generated by this system as evaluation elements. However, such evaluation remains to be seen because the present phase is experimental. Therefore, we verified response time as an evaluation method. The following are the computer specifications used in our experiment. The operating system was a Mac OS X version 10.4.9. The processor was 2.16 GHz with Intel Core 2 Duo with 1 GB of memory.

This system has functions that display the time required from querying user ID as graph seed that is a word to become the basis of a search to graph beginning drawn on the window. The time depends on ID search in Yahoo! Auction. We experimented with several rating values of user IDs. This point expresses an identical item as the first value of the above data table.

The response time in this system grows almost linearly with the rating value of user IDs because the rating value of user IDs indicates the actual number of transaction partners. In Internet auction sites ratings after trading is customary, but not mandatory.

5 Discussion

5.1 Advantages

The system we present has the following three advantages:

1. Supports identification of cheats from transaction relationships.

2. Visualizes transaction relationships as network structure in various viewpoints.

3. Markets expectations with visualized network data.

The banned ID display function in the present system relates fraud detection in Internet auction networks. Such personal data as rating value and banned transaction partner ID is available for fraud detection. In the research area of large-scale databases, the fraud community is detected by mining rating time [17]. Here, they focus on fraudulent ratings for fraudulent acts that have dealings and get user trust with specified IDs. They represent that the communities between fraudsters and fraudulent raters with this bipartite graph. These communities can be extracted from the data-set in Internet auction sites. This idea is included in our system. The log data of transactions is used to detect clues of fraudulent acts. On other hand, we can read characteristics from the visualizing graph by analyzing the auction network structure.

Google uses the PageRank algorithm to arrange search results. PageRank calculates value with a proving matrix eigenvalue problem that can also be applied to transaction relationships. We can handle hyperlink relations in a website as transaction relationships in Internet auctions. Also, web pages are treated as auction users.

In our second point, we aim to visualize the entire Internet auction network from transaction records with our system, but now we can only search transaction records with user IDs with auction search. A visualization model in this system would become increasingly complex with additional data. Therefore, selecting only useful data can reduce complexity. Other advantages of just exploiting userelated high or characteristic data are that system users can monitor auction networks from various visualization views

Presently, market prediction has not yet implemented this system. The analysis of transaction relationships and relations between persons and items enables market prediction. Generally it is said that recommendation and a web personalization are not effective because it takes costs for effect very much. We must collect a large amount of personal information of a user for recommendation. But in late years the long tail attracting attention points out the effectiveness. The long tail is a law that a total of a low element of occurrence frequency holds a ratio that cannot be ignored for the whole sells. This law is applied in Internet auction sites.

5.2 Generality

Our system can extract information from web sites, but not from the inner databases managed by auction sites. Thus, our system can be easily modified to analyze a variety of auction sites. Such generality is crucial because currently many market systems are available, for example, ad-exchanging portals, insurance markets, etc. On such web-based markets, there must be trading networks. Fortunately, such web-based markets tend to have reputation mechanisms in which participants are rated based on particular measures. Thus, our methodologies can be applied to such web-based markets.

6 Summary

In this paper, we built a system that can extract transaction relationships on Internet auction sites that are visualized as graphs. The proposed visualization system has the following three advantages:

1. Supports the identification of cheats from transaction relationships and detects suspiciousness with user ID information such as selling or buying histories.

2. Presents a transaction relationship as a network structure from various viewpoints. Users can view the network from various angles such as seller or buyer.

3. Possesses explanation functions about visualized networks and predicts auction markets.

These components have been implemented in Java programming language and for some web-based scripts. As future works, we will inspect a method of visualizing Internet auction networks with both personal user information and network structure. Most auction sites have a reputation mechanism in which each participant is rated based on its own activity in past trading. Although one study only used network structure for eBay trading [1], we will utilize reputation data to analyze trading network more effectively. Further we will develop a reasonable evaluation method and try a comparatively large-scale experiment.

References

[1] Shashank Pandit, Duen Horng Chau, Samuel Wang, and Chris tos Faloutsos. Netprobe: A fast and scalable system for fraud detection in online auction networks. *World Wide Web*, 2007.

[2] Z. Gyongyi, H. G. Molina, and J. Pedersen. Combating web spam with trustrank. *VLDB*, 2004.

[3] S. Brin and L. Page. The anatomy of a large-scale hyper-textual web search engine. *WWW7/Computer Networks*, 30, 1998.

[4] J. Kleinberg. Authoritative source in a hyperlinked envi-ronment. *Proc. 9th ACM-SIAM Symposium on Discrete Al-gorithms*, 1998.

[5] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems. *Communications of the ACM*, 43, 2000.

[6] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on ebay. *A controlled experiment*, 2003.

[7] M. Melnik and J. Alm. Does a seller 's ecommerce reputa-tion matter? evidence from ebay auctions. *Journal of Industrial Economics*, 2002.

[8] H. Kashima. Survey of network structure prediction meth-ods. *TJSAI*, 22, 2007.

[9] Getoor L. and Diehl C. P. Link mining: a survey. *SIGKDD Explorations*, 7, 2005.

[10] Matsuo. T., Ito. T., Day. R. W., and Shintani. T. A robust combinatorial auction mechanism against shill bidders. *the Fifth International Joint Conference on Autonomous Agents and Multi Agent Systems*, 2006.

[11] Matsuo. T., Ito. T., Day. R. W., and Shintani. T. A two-stage robust combinatorial auction mechanism against falsename bids. *the IEEE Pacific Rim International Workshop on Electronic Commerce*, 2006.

[12] Yokoo. M., Sakurai. Y., and Matsubara. S. The effect of false-name bids in combinatorial auctions new fraud in in-ternet auctions. *Games and Economic Behavior*, 1, 2004.

[13] Clarke. E. H. Multipart pricing of public goods. *Public Choice*, 11, 1971.

[14] Groves. T. Incentives in teams. *Econometrica*, 41, 1973.

[15] Vickrey. W. Counterspeculation, auctions, and com-petitive sealed tenders. *Journal of Finance XVI*, 1961.

[16] K. Sugiyama and K. Misue. Graph drawing by the magnetic spring model. *J. of Visual Languages and Computing*, 6, 1995.

[17] Y. Hirate, A. Akiyoshizawa, S. O, Y. Ioku, F. Kido, and H. Yamana. Support system for detecting abuse users in internet auction. *DBSJ Letters*, 5, 2006.