Limits of and Enhancements to NetProbe for Analyzing Online Auction Networks

Abhishek Bhattacharyya
abhatta1@stanford.edu

Nathan Howard
ndhoward@stanford.edu

Ashish Kulkarni
ashish87@cs.stanford.edu

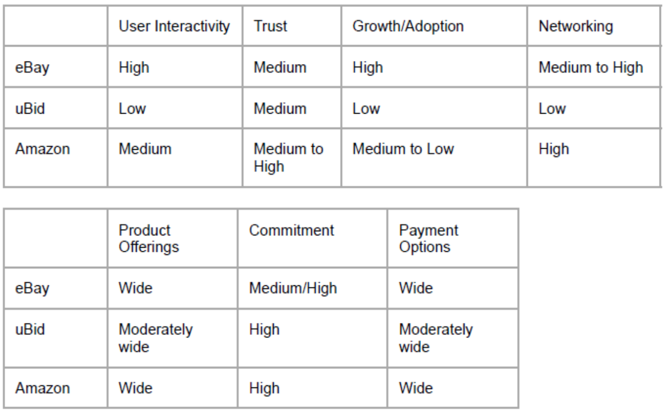
Abstract—**Fraud is an ever present threat in online auctions. The anonymity of the Internet provides a hospitable environment to both buyers and sellers of disrepute. To combat this problem many solutions have been proposed which use belief propagation and social network analysis to identify fraudulent actors. In this paper, we mainly analyze NetProbe which uses belief propagation over Markov Random Fields to classify users in online auctions as honest, fraud and accomplices. However, NetProbe fails to address some properties inherent to online auction networks like eBay. We identify cases where NetProbe might misclassify users or how users can exploit vulnerabilities of the NetProbe algorithm. We then propose a modification to the belief propagation algorithm to address the vulnerabilities to a great extent. Our assumptions on the properties of the original eBay graph and the effectiveness of the modified belief propagation algorithm are validated by running it on an synthetic network of 1000 nodes and approximate 2500 edges and an eBay network of 50,074 nodes and 149,935 edges collected using the eBay API. We observe that there is a 3.8% chance of a fraudster being flipped to an honest user, a 4.4% chance that an accomplice is flipped to an honest user and a 0.1% chance of a honest user being misclassified as an accomplice or fraudster in our synthetic test network.**

Keywords- belief propagation;auction networks;

I. Introduction

Internet auction has developed as one of the primary businesses in the internet space. Online auction is responsible for 29% of all e-commerce in 2002 [1]. EBay, the world’s largest auction site,reported a third quarter revenue of $1,449 billion, with over 212 million registered users [2]. The user growth rate has been reported to be around 26% while the revenue growth has been approximately 31%. However auction fraud remains one of the highly reported crimes in the internet domain. In fact it accounts for 63% of all crimes reported in the Federal Internet Crime Complaint in 2007[2].

As a result, auction network analysis has become an investigative subject in recent years. In fact, auction networks have evolved from simple C2C (Consumer to consumer) and B2C (Business to consumer) models to more dynamic consumer-to-consumer (C2C), business-to-consumer (B2C), business-to-business (B2B), business-to-government (B2G), and government-to public (G2P) models [1]. It would be expected that fraud detection mechanisms would also evolve with time. Unfortunately, auction fraud detection is still far from being perfect. [3] mentions the following factors for the growth of an online auction site: User interactivity, trust,growth/adoption, networking, product offering, commitment and payment options. It bases its ratings of various websites based on user survey and data analysis as in Table 1.



	User Interactivity	Trust	Growth/Adoption	Networking
eBay	High	Medium	High	Medium to High
uBid	Low	Medium	Low	Low
Amazon	Medium	Medium to High	Medium to Low	High

	Product Offerings	Commitment	Payment Options
eBay	Wide	Medium/High	Wide
uBid	Moderately wide	High	Moderately wide
Amazon	Wide	High	Wide

Table 1. Rating of auction websites on various parameters [3]

It needs to be noted that none of the established sites rate highly in the trust factor. This is because most auction sites employ a reputation and recommendation system to prevent fraud. It is not uncommon to defeat these systems by boosting one’s reputation. Some auction sites like eBay do fraud analysis at the backend with stale data. However, it is difficult to predict or detect fraud with perfect confidence. Other efforts to detect fraud has been mostly “common sense” approaches which require online profiles to be verified with law enforcement history of an individual. Unfortunately such basic efforts are unable to solve large scale distributed online frauds. Another possible way to combat online fraud might be in authority propagation system similar to TrustRank which detects online web spam. However, such systems have not been investigated in details in this context.

In [4], the authors present a way to model and analyze auction networks as social graphs. It analyzes eBay markets for digital cameras and liquid crystal display screens. The auctions are marked as nodes and weighted edges between the nodes represent the number of bidders competing between a pair of auctions. It found that the distribution of the number of auctions a bidder participates in follows a power law. There are a small number of bidders who participate in a disproportionate amount of activity.It then identifies communities or strongly connected components in the auction network social graph.However it fails to draw any reasonable generalized conclusion about the community structure of an auction network.Most of the conclusions drawn are ad-hoc and does not fit into any model. It therefore might be worth investigating into the community structure of an auction network as well to understand the trust-distrust dynamics of the network.

Guha et al. [5] claim that the reputation scoring mechanism by major auction sites is overly simple to detect collusive attempts by some sellers to fraudulently increase their own reputation rating. They propose that network structures generated by the past transactions can be used to expose sellers who use collusion to increase their reputation. This paper uses two social network indicators, k-core and center weights algorithms, to create a collaborative-based recommendation system that could indicate collusion between accounts. These social network indicators were tested on real world data of blacklisted accounts from leading auction sites to be able to screen out 76% of blacklisted accounts. Also, these indicators possess the ability to recognize suspicious behavior of accounts months before they are blacklist, thus helping protect against any possible seller collusion.

However, we found out that one of the most effective systems to detect and predict fraudulent behaviors in auction networks is Netprobe[2]. It models online auction as a Markov Random Field and tries to identify suspicious behaviors in the modeled space. It then uses a Belief propagation algorithm to detect fraudsters. Netprobe ran experiments on simulated data with 7000 nodes and 30,000 edges and found suspicious patterns with 90% precision. It also gathered data of eBay auction sales and identified likely suspicious users. However, they had no way to verify the trustworthiness of the identified users except for some negative ratings provided by other users due to lack of real-life ground truth.We base our algorithm on Netprobe simply because it happens to be one of the most powerful tools available at present.

The rest of the paper is organized as follows: In section II we explain some of the models to predict auction frauds including the Netprobe algorithm. Section III explores the vulnerabilities of Netprobe.Section IV presents the enhancement to Netprobe in order to make it more robust and the intuition behind it. The next section deals with our empirical results on eBay data and the simulation results on a synthetic network. We finally conclude in Section VI highlighting some of the additional enhancements which could be made to our degree-aware Netprobe algorithm.

II. Fraud Detection Methodologies

A. Earlier Approaches

One of the earlier papers which analyze empirical eBay data and model trust-distrust dynamics is by Resnick et al [6]. The eBay feedback system assigns a rating of -1 for every negative rating, a rating of +1 for every positive rating and 0 for neutral rating. One of the main problems with this system is that earlier users could easily bias themselves by creating accomplices and faking their own rating. They also observed that the number of negative ratings (0.3%) is far less than the number of positive ratings (51.2%). Hence they proposed a weighted rating system , one in which the negative score assigned will be some function of the present user score. This however has the additional disadvantage that if somehow a fraudulent user gets hold of certain number of accomplices that bias his ratings initially , he will be able to bring down the ratings of his neighbors easily. They also proposed a regression model which predicts the probability of a user being problematic based on his past history as:

$$\ln\left(\frac{\Pr(problematic)}{1-\Pr(problematic)}\right) = -3.9404 + .7712 * \ln(PROBLEMS + 1) - .5137 * \ln(POS + 1)$$

Our experiments with this model exposed a serious flaw of the model. If a fraudulent person has a few accomplices from whom he gets very high positive feedback ratings then this model places his chances of being problematic at a meager 0.2% (assuming 100 positive ratings). However, the fraudulent person might now easily dupe other people and then again open another account connecting with the same accomplices and carry out the same trick again and again, each time having a high positive rating and thereby fooling the reputation system. The trick therefore does not lie solely in the ratings received by a user but in the structure of the network around a fraudulent user. Netprobe bases its algorithm under this belief.

B. Netprobe

Netprobe uses belief propagation on a Markov Random Field to identify bipartite cores in the auction network. Each user is assigned a probability distribution of his chances of being either honest, accomplice or fraudulent based on how he is embedded in the feedback network. The dependency between a node and its neighbor is represented in terms of the propagation matrix ψ. The (i,j)th entry in the matrix represents the probability that a node is in state j given that its neighbor is in state i. Netprobe uses an intuitive distribution for assigning probabilities in the propagation matrix represented in Table 2.

Neighbor state	**Node state**		
	Fraud	Accomplice	Honest
Fraud	ε	1-2ε	ε
Accomplice	0.5	2ε	0.5-2ε
Honest	ε	(1-ε)/2	(1-ε)/2

Table 2. Propagation matrix ψ

Each node transforms its belief to its neighbor by passing messages as follows:

$$m_{ij}(\sigma) \leftarrow \sum_{\sigma'} \psi(\sigma',\sigma) \prod_{n\in N(i)\setminus j} m_{ni}(\sigma')$$

$$b_i(\sigma) \leftarrow k \prod_{j \in N(i)} m_{ji}(\sigma)$$

Here m_{ij} is the message vector sent to node i by node j, $b_i(\sigma)$ is the belief that node i is in state σ, N(i) is the neighbor list and k is a normalization constant.

Netprobe assigns the value of ε to 0.05 in an ad-hoc basis. From Table 2, this basically implies that an accomplice has a very low probability of connecting to an accomplice (0.1) while a fraudster has a very low probability of connecting to a honest node (0.05). We exploit these two properties of Netprobe to explore some of the basic vulnerabilities which arise from such an assignment to the parameter ε .

III. Vulnerabilities of Netprobe

As mentioned in section 2, the NetProbe algorithm works under the assumption that fraudsters are connected to accomplices with high probability (0.9) and it connects to other fraudsters or honest people with a very low probability (each having a probability of 0.05). Also accomplices are connected to accomplices with a very low probability (0.1). Hence the Netprobe algorithm essentially assumes a bipartite graph where fraudsters are disconnected from other fraudsters and honest people while accomplices are disconnected from accomplices. Unfortunately these probabilities of connection are not something empirical but an educated guess.

We start with the sample Netprobe Graph illustrated in fig. 1 and run Netprobe on top of it to figure out with what confidence Netprobe actually classifies fraudulent nodes.

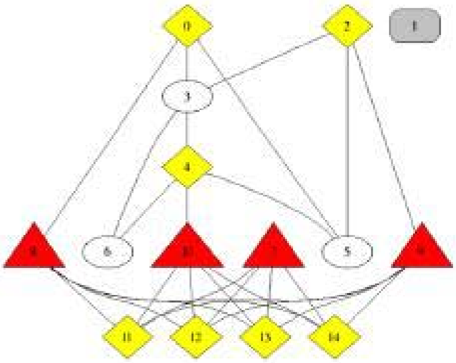


Fig. 1. Original Netprobe graph

NetProbe predicts node 10 to be fraud with a confidence of 65% while most of nodes 11,12,13 and 14 are accomplices with almost 100% confidence.

Now suppose a fraudster in an eBay network knowing about the essential assumption which Netprobe makes, wants to connect to honest users i.e. a fraudster carries out a couple of transactions with honest people on eBay. Then following the Barabasi Albert graph generation model for scale free networks (ebay closely mimics this power-law model and Netprobe simulations are based on this model) the fraudulent nodes are most likely going to attach to the high-degree nodes in the ebay network following the preferential attachment model. As a result our network structure might assume the form in fig. 2.

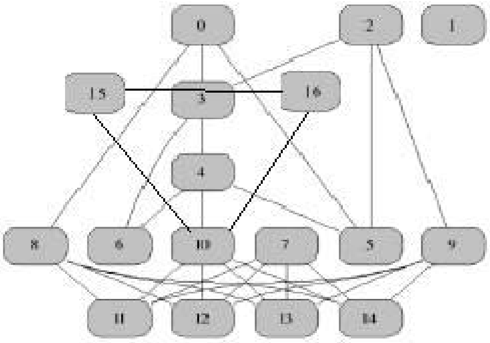


Fig. 2. Attachment of bipartite core to the Barabasi network

Here we see node 10 connected to two high-degree nodes 15 and 16. Since 15 and 16 are high-degree nodes, following the preferential attachment model 15 and 16 might also be connected with a high probability. In this way nodes 10,15 and 16 form a cycle. Now Netprobe believes that accomplices connect with each other with a very low probability. Hence the probability that node 15 and 16 are accomplices gets ruled out to a high extent. Node 15 and 16 are believed to be accomplices with a lesser confidence (56%) and hence the confidence that node 10 is a fraudster decreases to a high extent. In fact, now Netprobe concludes that node 10 is actually an honest node with a confidence of 65%. Thus it is possible to easily flip a fraudster to an honest user by modifying certain connections in Netprobe.

However in an actual eBay network as mentioned in [7], the highly connected nodes are not connected to one another. The intuition behind the absence of the rich club connectivity phenomenon in the eBay network stems from the fact that high-degree nodes in the eBay network are mostly sellers (B2C model [1]) representing companies and they are generally not connected to one another.. As a result the edge between nodes 15 and 16 might not exist in an original eBay graph. The modified graph structure now assumes the form shown in fig. 3.

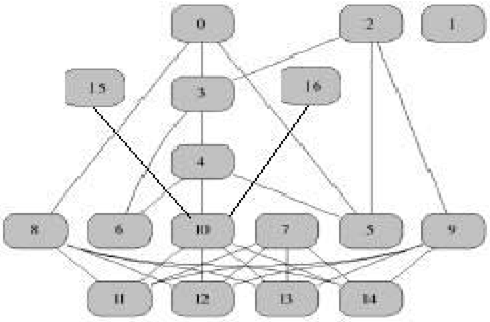


Fig. 3. Attachment of bipartite core to pruned Barabasi graph mimicking eBay network

Now the fact that nodes 15 and 16 are not connected to each other increases the confidence that nodes 15 and 16 are accomplices (70%). This on the other hand decreases the confidence that node 10 is an honest node. However in this case the decrease is not sufficient to convert node 10 from an honest node to an dishonest node. In fact Netprobe concludes that node 10 is honest with a confidence of 55%. This is illustrated in Fig 4. This example clearly demonstrates that it might be possible for Netprobe to mistakenly conclude that an honest node is an accomplice or a fraudster depending on the network topology.

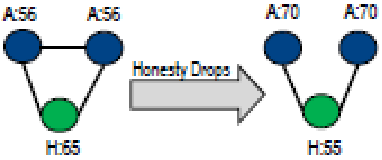


Fig. 4. Honesty reduction due to absence of rich club connectivity phenomenon

Thus we see that it is possible that a fraudster might camouflage himself to dupe Netprobe into believing that it is honest. Similarly, reversing our logic it is possible for Netprobe to wrongly conclude that honest nodes are fraudulent. Hence Netprobe might have both false positives and negatives.

IV. Enhancements on Netprobe

While it is not completely possible to prohibit a node from flipping itself, we suggest a method which might prevent unsolicited node-flipping to a larger extent compared to Netprobe. This algorithm is based on the belief that high-degree nodes in the network to which other nodes get preferentially attached are most probably honest in nature since it does not pay-off making a whole lot of transactions if a person is fraud. Our algorithm biases the Netprobe belief propagation algorithm by passing messages biased by the degree of the attached node. Mathematically,

$$m_{ij}(\sigma) \leftarrow \sum_{\sigma'} \psi(\sigma',\sigma) \prod_{n\in N(i)\backslash j} m_{ni}(\sigma')\, f(\deg(j),\sigma')$$

$$b_i(\sigma) \leftarrow k \prod_{j\in N(i)} m_{ji}(\sigma)\, f(\deg(j),\sigma')$$

where f(deg(j),σ) is the biasing function. The biasing function is constant (set as 1) upto the average node degree in the network after which it increases linearly with the node degree.

V. Results

We extract data using the eBay API and the set of blacklisted nodes specified in [8] as the seed. Our network currently consists of 50,074 nodes and 149,935 edges. The graph of the network connectivity is presented in Fig 5.

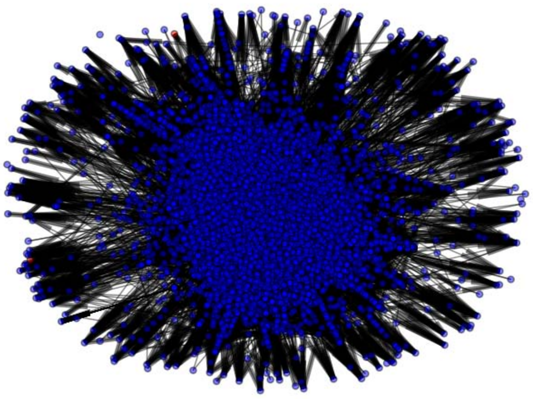


Fig. 5. Structure of the eBay network

We first validate our assumptions of the behavior of the eBay network. We found out that the eBay network is in fact an approximate power-law distributed Barabasi-Albert network with α =1.7. The degree distribution is presented in Fig.6.

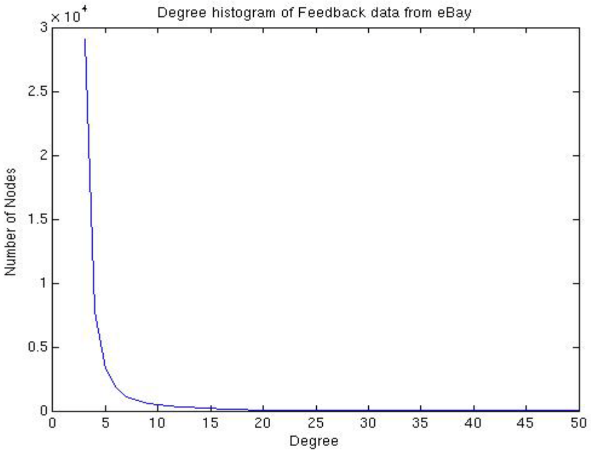


Fig. 6. Degree distribution of the eBay network

We then validate our assumption about the absence of the rich club connectivity phenomenon in the eBay graph among the top 50 high degree nodes in the network. We observed that 0.04% of the total possible connections exist among the high-degree nodes in an eBay network while it stands at 0.26% for the Barabasi-Albert graph. Thus the Barabasi graph demonstrates a 6x times more rich club connectivity than the eBay graph.

Our tests to validate our hypothesis regarding the vulnerabilities of Netprobe also proved to be successful. Both the cases, where fraudsters are flipped to honest nodes and honest nodes are flipped to accomplices or fraudsters were observed. For this purpose we injected 25 bipartite cores similar to Fig.1. into a synthetic Barabasi-Albert graph of 1000 nodes and then connected in turn the fraudulent node (node 10 in Fig.1.), an accomplice node (node 11 in Fig.1.) and an honest node (node 3 in Fig.1.) in each subgraph to a pair of high degree honest nodes in the original Barabasi graph as outlined in section III. Then we ran Netprobe on the produced graph. The results are summarized in Table 3.

	Ideal detection	Honest node connected	Fraudulent node connected	Accomplice node connected
Honest nodes	1075	955	1136	1131
Accomplice nodes	175	204	209	21
Fraudulent nodes	125	115	30	4

Table 3. Netprobe on Synthetic graph structure

We observe that topology affects the detection behavior in Netprobe as specified in our hypothesis. There is thus a 3.8% chance that a fraudster converts itself to an honest node by simply carrying out two transactions with two honest users, a 4.4% chance that an accomplice disguises itself as an honest node and a 0.1% chance that a honest node gets flipped to either an accomplice or a fraudster.

As a next step we verified the bias which our modified algorithm provides to the belief propagation algorithm. It is shown in Fig 7. and Fig 8 . This figure shows that our algorithm biases the network with more honest ratings when a node is connected to a high degree node which in turn resists the flip of an honest node to an accomplice or a fraudster.

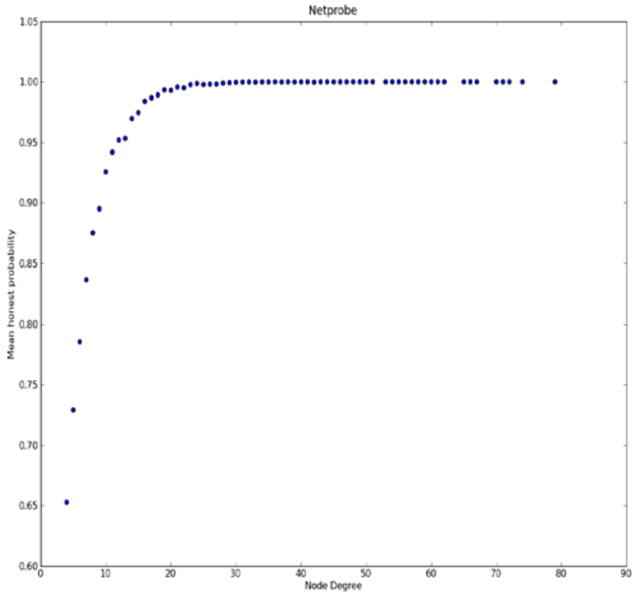


Fig. 7. Mean honesty rating for netprobe

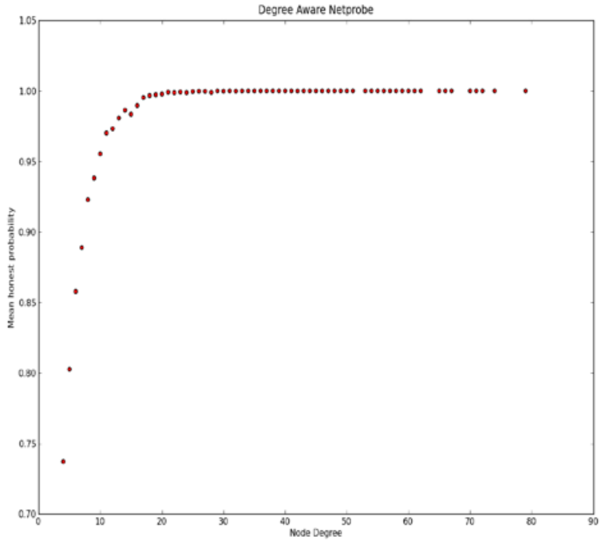


Fig. 8 . Mean honesty rating for modified degree aware Netprobe

VI. Conclusion and Future work

In this paper we have demonstrated that although Netprobe is a highly efficient algorithm there may be cases where it might misclassify nodes or where fraudulent users might easily exploit its assumptions to camouflage themselves. Our proposal to enhance Netprobe is designed to make it more robust in cases where it flips an honest user to an accomplice. However, there is still no guarantee that our algorithm will detect or eliminate all cases where a fraudster converts itself to an honest user. One of the basic understandings is that the value of the parameter ϵ used in Netprobe needs more fine-tuning in order to make it more robust. The assumptions which become inherent from assigning an arbitrary value to ϵ might significantly affect the performance of Netprobe. As a next step, we plan to run precision tests on the eBay data using an optimised degree-aware modified Netprobe algorithm. We plan to include the value of the total amount of transactions into our biasing function in order to make it even more robust.

VII. Links

Source Code and eBay data
hosting: http://code.google.com/p/jure224/

References

[1] Subir Bandyopadhyay, Julie Wolfe ,A critical review of online auction models,Journal of the Academy of Business and Economics, Jan, 2004.

[2] Shashank Pandit, Duen Horng Chau, Samuel Wang, Christos Faloutsos, NetProbe: A Fast and Scalable System for Fraud Detection in Online

Auction Networks,WWW 2007, May 8–12, 2007, Banff, Alberta, Canada

[3] Bressler, Stacey and Grantham, Charles, Communities of Commerce: Building Internet Business Communities to Accelerate Growth Minimize Risk and Increase Customer Loyalty, McGraw Hill, New York, 2000.

[4] R. Kang-Xing, David C. Parkes, and Patrick J. Wolfe. 2007. Analysis of bidding networks in eBay: Aggregate preference identification through community detection. Paper presented at AAAI Workshop on Plan, Activity and Intent Recognition: 66-73.

[5] 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.

[6] Paul Resnick, Richard Zeckhauser (2002), Trust among strangers in internet transactions: Empirical analysis of eBay' s reputation system, in Professor Michael Baye, Professor John Maxwell (ed.) The Economics of the Internet and E-commerce (Advances in Applied Microeconomics, Volume 11), Emerald Group Publishing Limited, pp.127-157.

[7] Beyene, Y., Faloutsos, M., Duen Horng Chau, Faloutsos, C,"The eBay graph: How do online auction users interact?," INFOCOM Workshops 2008, IEEE , vol., no., pp.1-6, 13-18 April 2008.

[8] http://blacklistedebayers.blogspot.com/