The eBay Graph: How Do Online Auction Users Interact?

Yordanos Beyene, Michalis Faloutsos
University of California, Riverside
{yordanos, michalis}@cs.ucr.edu

Duen Horng (Polo) Chau , Christos Faloutsos
Carnegie Mellon University
{dchau,christos}@cs.cmu.edu

***Abstract*— Online auctions have revolutionized the ability of people to buy and sell items without middlemen, and sales reaching more than $57 billion every year on eBay alone. The user interactions at online auctions form a network of interactions akin to a social network. Unlike other online social networks, online auction networks have not been studied so far. In this paper, we model and characterize the structure and evolution of the network of user interactions on eBay. Specifically, we use over 54 million eBay transaction feedback that users leave for each other. A distinguishing feature of the graph is the rich meaning that nodes and edges can have (for example, an edge could be positive or negative, posted by a buyer or a seller) in contrast to other graphs such as the topology of the Internet. First, we provide the vital statistics of the emerging graph, which we call eBay graph. We observe that the graph exhibits both significant differences and similarities to commonly studied graphs. Second, we study the feedback behavior of users: feedback is not always reciprocal, and negative feedback is scarce (less than 1%), but few negative feedbacks could discourage new users. Finally, we develop an intuitive model that captures key properties of the graph in a visual and memorable way. Our work can be seen as a first step in understanding online auction graphs, which could enable us to detect trends, abnormalities and ultimately fraudulent behavior.** [1]

I. Introduction

There has been tremendous growth of online auction activities over the last several years with millions of people buying and selling goods online. eBay, uBid, Bidz, Yahoo are some of the major online auction sites with eBay having the lead. On any given day, eBay has more than a 100 million items available for sale, with 6.4 million new items added every day. eBay users worldwide trade more than $1,812 worth of goods on the site every second.

Despite the growth of online auctions, there does not exist prior work focusing on the network structure of online auctions. We have seen several studies on the graph structure of the Internet [1], [3], [12], WWW [2], and social networks [4], [5]. Recent studies also focus on understanding the evolution of complex graphs [6]–[8]. To our knowledge, this is the first detailed study focusing on the structure and evolution of huge interactions of online auction users with over 54 million eBay transaction feedbacks. There are studies that specifically focus on detecting fraudsters [10], [11].

In this study, we model and characterize the network of interactions of eBay over 7 years. Specifically, we answer the following questions: What are the fundamental characteristics and properties of the graph? How does it evolve? How can we represent this massive data by a meaningful intuitive model?

In this analysis, we use 54 million eBay feedbacks that users leave for each other to reconstruct the network wide behavior of more than 11 million users. The information in the dataset is what eBay uses to rate the trustworthiness of users. A distinguishing feature of the graph is the rich meaning that nodes and edges can have (for example, an edge could have a positive or negative weight, and originate from a buyer or a seller) in contrast to previous graphs. Some users are sellers only, others buyers only and majority buyers and sellers which we refer them as traders. In addition, each transaction feedback is time-stamped which enables us to study the growth of the graph.

The main contributions can be summarized in the following points:

1) **The eBay graph differs from other scale-free networks.** We observe that the graph exhibits both significant differences and similarities to commonly studied graphs such as the Internet topology. Unlike the Internet and many social networks, the rich-club phenomenon does not appear on the eBay graph, though like many networks, it has skewed degree distribution and is disassortative. Another interesting property is that the degree distribution of negative feedbacks is skewed.
2) **The graph becomes denser over time.** We see graph densification and growth of giant component over time. Linear preferential attachment

[1]This work was supported by NSF NETS 0721889 and NSF IDM 0208950 and a CISCO URP grant.

exists partially as we explain later. Furthermore, the negative feedbacks that a user has decrease significantly the "preferential" status of the user.
3) **Feedback is not always reciprocal and negative feedback is rare.** Our analysis shows the rate of retaliatory negative feedback is about 20% and the rate of reciprocating positive feedback is about 51%. Overall, negative feedback is scarce (less than 1%).
4) **We develop an intuitive model**, which captures key properties of the graph in a visual and memorable way. The model is based on graph eccentricity which is key metric to identify central nodes.

The paper is organized as follows. In Section II we describe the data set and the different types of graphs constructed for our analysis. Section III presents the fundamental graph properties. In section IV we study the evolution of the eBay graph. Section V covers the feedback behavior. We developed an intuitive model in section VI. In section VII, we cover related work. We conclude in section VIII.

II. DATA DESCRIPTION

The dataset we use for this study is eBay feedbacks that users leave for each other. eBay sellers and buyers have the opportunity to rate each other (1, 0 or -1) and leave comments after each transaction. It is a key information to compute the feedback score used by the eBay reputation mechanism to rate the trustworthiness of users. The feedback score of a user is computed by taking the number of distinct users who left positive feedback, and subtracting the number of unique users who left negative feedbacks.

Users with good feedback are regarded as trustworthy individuals, and have the benefit of selling goods for a higher price compared to those who have received negative feedbacks or lack previous transaction records. In fact, eBay users have the opportunity to view the breakdown of positive, negative, and neutral feedbacks for the past one month, six months, and one year of a user.

The dataset contains about 54 million eBay transaction feedbacks that involve 11 million users, where 66,130 of them are fully crawled, i.e. all the feedbacks left for the users during seven years (1999 to 2006) are included to the dataset. The transaction feedback contains the user-name of the person who left the feedback, the user-name of the person who received the feedback, the feedback score, the time the feedback is written and the the role of users i.e. who is the seller and who is the buyer. It also contains a user profile which includes the eBay user-name, the date joined, the location, and the status of crawling. The data is imported into a MySQL 5.0.37 server.

The data is collected by crawling the eBay site using a Breadth First Search mechanism. Initially, a seed set of users is inserted into a queue. In the crawling process, the first entry of the queue is popped out, and all feedbacks left for the user are collected. While all the feedback left for the user is added to the dataset, every user seen for the first time is added to the tail of the queue. Once all the user feedbacks are collected, the user is marked as visited, and removed from the queue.

We model the data as a graph where a user is represented by a node and a transaction feedback between users by an edge. The eBay graph has a rich meaning compared to previous graphs. A node can be either a seller only, buyer only, or a trader. An edge could have a positive, neutral or negative weight, and is either a buyer feedback or a seller feedback.

We construct three different types of graphs: the Trust graph, the Transaction graph, and the Undirected graph for understanding the key properties of the eBay transaction feedback.

The **Trust graph** is a directed graph which represents the eBay reputation system. We draw an edge from a to b if a votes for b. The weight of the edge is the value of the vote (+1, 0, or 1)

The **Transaction graph** is a directed graph which represents the role of users. We draw a line from node a to node b if a buys from b.

The **Undirected graph** consists of undirected edges between nodes that at least one node has left feedback for the other.

Based on the method of construction, the three graphs reveal different properties of the eBay transaction feedback. We use the Trust graph to study the node degree distribution, graph growth and evolution whereas the undirected graph to compute fundamental graph metric properties such as the rich club connectivity, node eccentricity, diameter of the graph and other properties. Note that we mention the Transaction graph for completeness, but we do not study this graph here.

III. FUNDAMENTAL GRAPH PROPERTIES

In this section, we focus on understanding the fundamental topological properties of the eBay graph. In the past, several studies examine the graph structure of the Internet [1], [3], [12], WWW [2], and social networks [4], [5]. Most of the studies reveal the power-law degree distribution, small-world, rich club connectivity, assortativity and other graph metrics, which are important to understand and model the graphs structure.

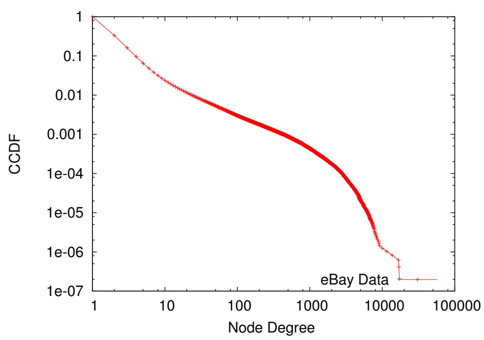

Fig. 1. The in-degree distribution of the Trust graph.

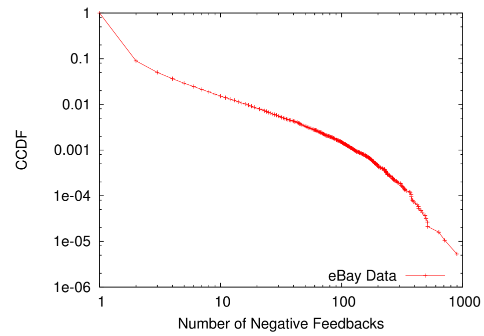

Fig. 2. The distribution of negative feedbacks received by a node (for nodes with negative feedback).

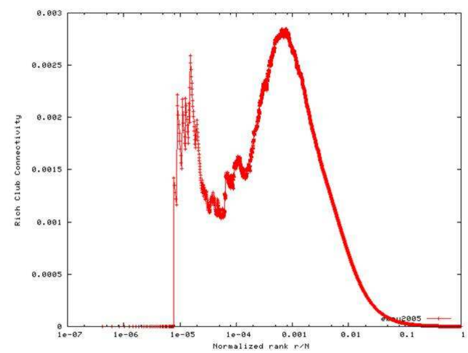

Fig. 3. The rich club connectivity of the graph.

The eBay graph exhibits both significant differences and similarities to commonly studied graphs such as the Internet, Word Wide Web and social networks. It exhibits a skewed degree distribution, disassortativity and reveals graph densification over time. But unlike the Internet and social networks, it does not posses the rich club connectivity phenomenon, and preferential attachment holds partially as we explain in the next section.

A. *Degree Distribution*

The degree distribution of many graphs such as the Internet, the Web and social networks follow a power-law degree distribution [12]. In this study, we used the Trust graph to examine the node in-degree distribution to capture distribution of feedbacks among users. In this analysis, we focus on the in-degree distribution of about 66,000 fully crawled nodes. We also study the distribution of negative feedbacks, which is a primary parameter in determining the trustworthiness of an eBay user.

The Complementary Cumulative Distribution Function (CCDF) shows a skewed degree distribution as can be seen in Fig. 1. The majority of the nodes are of low degree and few nodes have high degree. Similarly, the CCDF of negative feedbacks is skewed (Fig. 2). We see that approximately 99% of the nodes with negative feedback have less than 10 negative feedbacks, while there are some nodes with a large number of negative feedbacks (up to 896).

B. *Rich Club Connectivity*

In the Internet graph at the network level, high degree nodes are very well connected to each other. This property is referred as the *rich club* phenomenon. The eBay graph differs from the Internet topology when it comes to the Rich Club connectivity [1]. The rich club phenomenon does not appear on the eBay feedback graph (see Fig. 3). High degree nodes are hardly connected directly with each other. In order to compute the rich club coefficient, nodes in a network are sorted by decreasing number of links that each node contains. The node rank r denotes the position of a node on this ordered list. The node rank is normalized by the total number of nodes in the graph. The rich-club connectivity of the normalized rank r is defined as the ratio of the actual number of links to the maximum possible number of links among nodes with rank less than or equal to r. The maximum possible number of links between n nodes is n(n-1)/2. In this analysis, we consider all the feedbacks exchanged within one year (2005), and we observe that the top 61 high degree nodes (0.001% of the total) which are attached to 25% of the edges are not connected with each other. We repeated the same set of experiments over several years and observed similar behavior. We attribute this to the fact that high degree nodes are usually sellers, which rarely interact.

C. *Distance*

We examine the hop plot and eccentricity of the undirected eBay graph. The hop plot, also called the basic neighborhood function N(h), of a graph, is defined as the number of pairs of nodes within a specified distance h, for all distances h. Fig. 4 shows

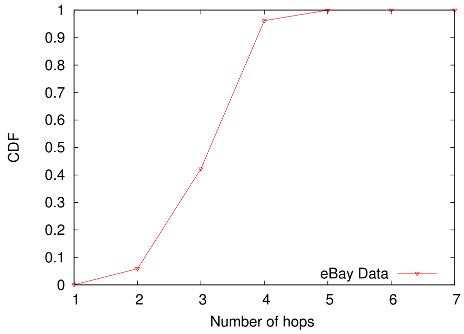

Fig. 4. The hop plot of the graph (CDF).

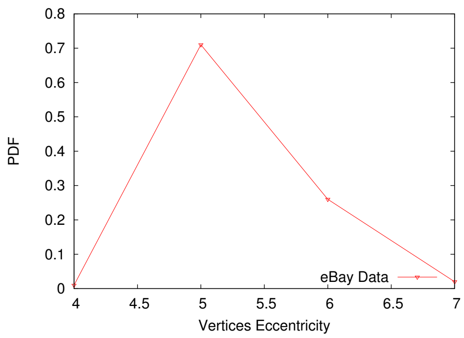

Fig. 5. The node eccentricity distribution of the graph.

the CDF of the hop-plot of the eBay graph, where we see that roughly 97% of the nodes are within 4 hops from each other.

Another distance metric is the eccentricity of a vertex v, which is defined as the maximum distance between v and any other vertex in the graph. Note that the maximum eccentricity over all nodes in a graph is the graph diameter. Intuitively, the set of nodes with maximum eccentricity form the graph periphery, while nodes with minimum eccentricity correspond to the center of the graph. Our results show that the minimum graph eccentricity is 4, and the maximum is 7. Fig. 5 shows the PDF of eccentricity. As we explain later, we use the eccentricity to identify central nodes, when constructing our intuitive model.

IV. Graph Evolution

An in-depth understanding of the graph evolution is useful for modeling the growth of the eBay graph in general and individual users in particular. In addition, understanding the graph evolution is a first step to detect abnormal trends. However here, we focus on the densification and preferential attachment properties of the eBay graph which has been observed in many complex networks [8], [9].

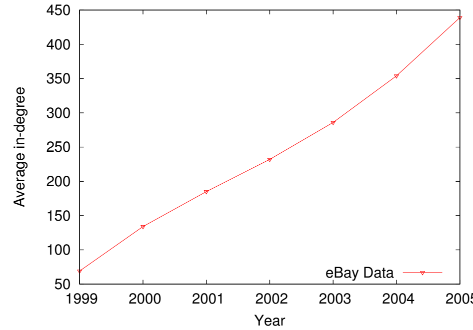

Fig. 6. The average in-degree over time.

A. *Densification*

We study the density by computing the average node in-degree of the eBay trust graph over time. The number of new nodes increases over the years, and new edges appear as new nodes connects to an old node or when two old nodes connect. Our analysis revealed an increase in the average node in-degree in the network over the years, with the number of edges growing super-linearly in the number of nodes which is an indication of graph densification (Fig. 6).

B. *Preferential attachment*

In many evolving networks, new nodes are more likely to connect to nodes that already have a larger number of links, a phenomena widely known as preferential attachment. In fact, preferential attachment explains the emergence of the heterogeneous network structure and skewed degree distribution [8], [9].

Our study of the eBay Trust graph revealed that a linear preferential attachment exists partially. Linear preferential attachment means that the probability of connecting to a node is proportional to its in-degree. It is known that eBay mechanisms “put less weight” on the old behavior of nodes (from the way reputation is calculated and show). Thus, we did the following: We compute the degree of the nodes using the feedbacks left during first 9 months of the year 2005. Then, we observe the new edges during the remaining 3 months of that year. First, we examine good nodes: nodes with less than 10 negative feedbacks and more than 90% positive feedback. Then, we plot the number of new edges as a function of the node degree. Our results show that the average number of new connections is proportional to the degree of the nodes. We observe linear preferential attachment especially for nodes of degree less than 500 (see Fig. 7 and Fig. 8). As nodes obtain higher degrees, say above 500, the correlation between degree and preference weakens. We repeated the same experiment over different years and found similar results.

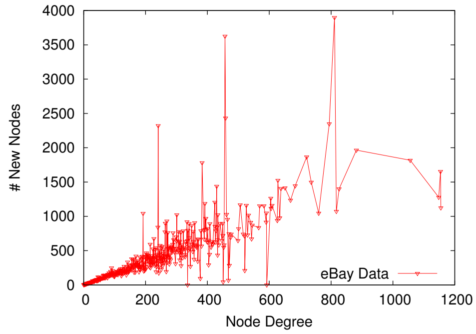

Fig. 7. The average number of degree increase per degree.

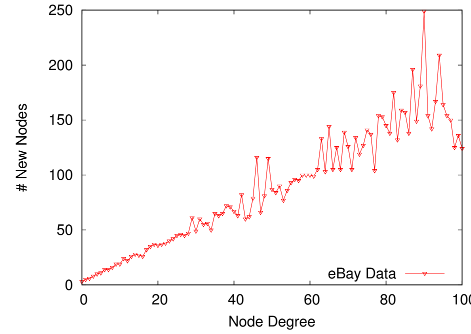

Fig. 8. The average degree increase per degree for nodes of degree less than 500. Nodes are grouped in bins 1-5, 6-10 etc for statistical purposes.

Furthermore, we examine the effect of negative feedbacks and find that users with high negative scores have very low chance of attracting new users. Here, we study nodes with over 10 negative feedbacks and less than 90% positive feedback ratio from year 2005 as above. We find that there are no new edge attached to 50% of these nodes.

We conclude that linear preferential attachment exists for nodes with low negative feedbacks, and is a factor only in low to medium degree nodes ($<= 500$).

V. Feedback and Retaliation

As mentioned earlier, users can leave positive, neutral or negative feedback for a user with whom they transacted. Many comments on the eBay feedback forum claim that the fear of receiving retaliatory negative feedback is a reason for the limited negative feedback. We tried to examine the validity of this conjecture by answering the following questions: Do people reciprocate with a positive feedback, and do people retaliate to a negative feedback?

We study the feedback considering the relative order with which the feedback is written. We examine feedbacks between fully crawled nodes, and we look at the first feedback and analyze the response, if any. Our analysis shows the rate of retaliatory feedbacks is 20%, namely, whenever user 'A' writes a negative feedback for user 'B', then user 'B' retaliates one out of five times (by posting a negative feedback for user 'A'). The remaining 80% get positive, zero or no feedback in return. Thus, we believe retaliatory feedback is not as wide spread as one might think. At the same time, reciprocating to positive feedback is around 50%, which is higher than the retaliation, but again significantly less than a default positive response.

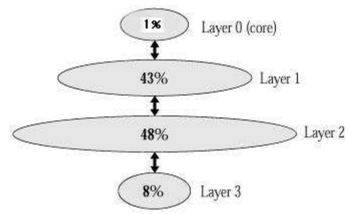

Fig. 9. An intuitive model based on eccentricity.

VI. Intuitive Model

We develop an intuitive model that captures key properties of the graph in a visual and memorable way. Our goal is to generate a simple model that one could draw easily on a board with a few lines. To develop our model, we use the node eccentricity, which we defined earlier. Here, we use the Undirected graph of the fully crawled connected nodes (66K nodes).

We start by using eccentricity to define the "center" of the graph. We define the *core* (*layer 0*) to be the set of nodes with minimum eccentricity, which is 4, as shown in Fig. 9. Then, we iteratively define the next layer as the nodes that are directly connected to at least one node in the upper layer. Given this definition, it is easy to prove that the eccentricity of nodes at layer-i+1 is bounded by the eccentricity of a directly connected node at layer-i plus one.

Interestingly, we have only 4 layers in our graph. The core layer contains 1% of the nodes, the two middle layers contain the majority of the nodes (over 43+48 = 91%) and the last layer accounts for 8% of the nodes. As one would expect, the majority of the nodes in the core layer are high degree nodes.

The model is helpful in visualizing the structure of the graph and some of its properties.

First, the model reveals the "shallowness" or compactness of the graph. These properties appear in many highly skewed, small-world and scale-free networks.

Second, the model provides visually a bound for the eccentricity and diameter of the graph, assuming we know the minimum eccentricity. The eccentricity of the nodes at core is $ecc(core) = 4$ by construction, and the eccentricity of nodes at layer i is bounded by $ecc(layeri) \leq ecc(core) + i$. Thus, the model can quickly point to the fact that the maximum eccentricity of the graph is bounded by 7. Recall that the diameter is 7 in this graph.

VII. Related Work

Several research efforts study the structure of the Internet [1], [3] [12] and WWW [2]. In fact, there are too many to list here exhaustively. Social networks have been analyzed in [4]–[7], where researchers report that the core of such network contains a very large component linked to some smaller but highly connected components. These studies also compare the structure of social networks to that of other types of networks, such as the Internet. Other interesting studies focus on complex network graphs in general [8], [9]. These studies also analyze the evolution of the graph which includes extended discussions on the issues of graph densification, shrinking diameter, and preferential attachment. Another interesting work focuses on detecting fraudsters on eBay [10], [11], and develops techniques to analyze the same data as we have here, namely feedback records from eBay transactions, and identify suspicious user behavior. That work exploits the local graph structure: the collusion of users to artificially boost their score. However, that work does not attempt to characterize the graph structure of transactions.

Our work is different from these research efforts: to the best of our knowledge, this is the first study that focuses on online auction users with the intent to model the network structure and identify network-wide trends.

VIII. Conclusions

In this paper, we model and characterize the network of interactions of eBay users using 10 GBytes data for 54 million transaction feedback entries that users leave for each.

We observe that the graph exhibits both significant differences and similarities to commonly studied graphs such as the Internet topology.

We also observe that linear preferential attachment exists partially: especially for the low-degree nodes (<=500) of good standing. One explanation is that once a user seems sufficiently credible, the exact number of satisfied users becomes less important. However, negative feedbacks seem to hurt the reputation of a user in a significant way.

We find that negative feedback is less than 1% of the total feedback. This could mean that either most people are good, or people are afraid to leave negative feedback. Investigating this further, however, we find that retaliatory negative feedback is less than 20%.

Finally, we develop an intuitive model based on eccentricity. The model groups nodes in 4 layers, with the top layer being the "central" nodes in terms of connectivity. We find that the model captures key properties of the graph such as the shallowness of the graph, bound for graph diameter in a visual and memorable way.

In the future, we want to focus on characterizing the evolution of individual users to create a few typical profiles and develop methods to identify abnormal user behavior and patterns. We also want to understand the interaction of users in terms of communities: in other words, look at clustering properties in the graph.

References

[1] P. Mahadavan, D. Krioukov, M. Fomenkov and B. Huffaker, "Lessons from Three Views of the Internet topology", CAIDA Technical Report TR-2005-02.

[2] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener, "Graph structure in the Web", Proceedings of the Ninth International World Wide Web Conference pp. 309-320, May 2000.

[3] G.Siganos, S.L. Tauro, M.Faloutsos, "Jellyfish: A Conceptual Model for the AS Internet Topology", Journal of Communications and Networks 2006.

[4] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of Topological Characteristics of Huge Online Social Networking Services", The 10th Asia Pacific Physics Conference, Aug 2007.

[5] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks", In Proceedings of the 5th ACM/USENIX Internet Measurement Conference October, 2007.

[6] K.-I. Goh, Y.-H. Eom, H. Jeong, B. Kahng, and D. Kim, "Structure and evolution of online social relationships: Heterogeneity in unrestricted discussions", Phys. Rev. E 73 066123, 2006

[7] R. Kumar, J. Novak, A. Tomkins, "Structure and evolution of online social networks" , ACM SIGKDD, 2006.

[8] A.-L. Barabasi, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek, and A. Schubert, "Evolution of the social network of scientic collaborations", Physica A 311, pp. 590-614, 2002.

[9] J. Leskovec, J. Kleinberg, C. Faloutsos, "Graph Evolution: Densification and Shrinking Diameters", ACM Transactions on Knowledge Discovery from Data, 2007.

[10] S. Pandit, D.H. Chau, S. Wang, and C. Faloutsos, "NetProbe: A Fast and Scalable System for Fraud Detection in Online Auction Networks", Proceedings of the 16th international conference on World Wide Web pp. 201-210 May, 2007.

[11] D. H. Chau, S. Pandit, S. Wang, and C. Faloutsos, "Detecting Fraudulent Personalities in Networks of Online Auctioneers", Proceedings of PKDD, 2006.

[12] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology", ACM SIGCOMM, 1999.