

The eBay Graph: How Do Online Auction Users Interact?

1569100833

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

Duen Horng 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 such as the Internet topology. Second, we study the feedback behavior of users: feedback is not always reciprocal, and negative feedback is scarce (less than 1%), but plays a major role in discouraging 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], 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 that work the authors developed a data mining techniques that analyzes publicly available histories of eBay transactions that identifies suspicious online behaviors and dubious associations among users.

In this study, we model and characterize the network of interactions of eBay over 7 years of period. 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 a 54 million eBay feedback 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 use 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, 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 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 feedback graph, but the graph is has skewed degree distribution and is disassortative. Another interesting property is that the degree distribution of negative feedbacks is skewed and follows a power law.
- 2) **The graph becomes denser over time.** We see graph densification and growth of giant compo-

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

ment over time. Linear preferential attachment exists partially as we explain later. The number of negative feedback that a user possesses has a major impact on preferentiality of other users to connect with.

- 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%. Negative feedback is scarce (less than 1%), but plays a major role in discouraging new users.
- 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, i.e. we define the "center" of the graph to be the nodes with minimum eccentricity, and it represents key properties.

The paper is organized as follows. In Section 2 we describe the data set and the different types of graphs constructed for our analysis. Section 3 presents the fundamental graph properties. Section 4 covers the feedback behavior. We developed an intuitive model that captures key properties of the graph in a visual and memorable way in section 5. In section 6, we cover related work. We study the limitations of our study in section 7. We conclude in section 8.

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 lacks 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) period of time are added to the dataset. The transaction feedback contain username of the person who left feedback, username of the person who received feedback, feedback score, time the feedback is written and status of both users i.e. who is the seller and who is the buyer. We also have user profile such

as eBay username, date joined, location, and status of crawling. The data is imported into MySQL 5.0.37 server.

The dataset was collected using Breadth First Search mechanism where a queue data structure is used to store a list of uncrawled users. 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 by crawling the user profile all feedbacks left for the user are collected. While all the feedback left for the user are 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 modeled the dataset as a graph where a user is represented by a node and a transaction feedback between users by an edge. A distinguishing feature of the eBay graph from the Internet or other social networks is the rich meaning that nodes and edges can have. An edge could have a positive, neutral or negative weight, and is either a buyer feedback or a seller feedback in contrast to previous graphs. Moreover, a node represents a buyer, seller or buyer and seller.

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

Trust Graph: 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)

Transaction Graph: a directed seller/buyer graph where an edge points to a seller. We draw a line from node a to node b if either user a leaves feedback to user b or vice versa, and b is a seller.

Undirected graph: An edge exists from a to b if there is a transaction feedback between a and b.

Based on the method of construction, the three graphs reveal different key properties of the eBay transaction feedback. We use the directed 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. The transaction graph is used to examine the role of users, specifically proportion of users are sellers only, buyers only, and traders.

III. FUNDAMENTAL GRAPH STRUCTURE

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 con-

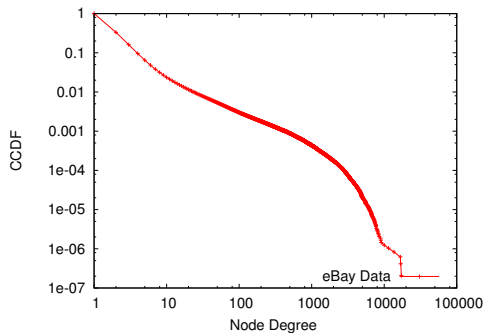


Fig. 1. eBay Graph CCDF.

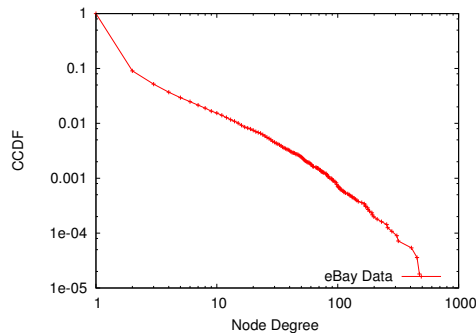


Fig. 2. Negative eBay Feedback CCDF.

nectivity, assortativity and other graph metrics, which are important to understand and model the graphs structure.

Unlike other scale-free networks, we observe that the eBay graph exhibits both significant differences and similarities to commonly studied graphs such as the Internet, World Wide Web and Social Networks. Like the Internet, World Wide Web and other social networks, the eBay graph exhibits a skewed degree distribution, disassortativity and graph densification over time. But unlike the Internet, it does not have the rich club connectivity phenomenon, and preferential attachment holds partially as we explain below.

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 examined the node in-degree distribution of the eBay trust graph. We want to understand the distribution of feedbacks among users. We only considered full crawled nodes in order to avoid wrong conclusion attributed due to missing data. We also looked specifically at the distribution of negative feedbacks, which is a primary factor determining the trustworthiness of a user. As we explain in section 4, negative feedbacks are rare but few negative feedbacks could hurt the trustworthiness of a user.

The Complementary Cumulative degree distribution (Fig 2) shows that the graph has a skewed degree distribution. Majority of the nodes are of low degree and few nodes have high degree. Similarly the CCDF of the negative degree distribution is skewed degree distribution (Fig 2). It is very important to understand the distribution of negative transaction feedbacks as it is a key component to compute the trustworthiness of a user and also to understand the effectiveness of the eBay reputation mechanism.

B. Rich Club Connectivity

The eBay graph differs from the Internet topology when it comes to the Rich Club connectivity [1]. The

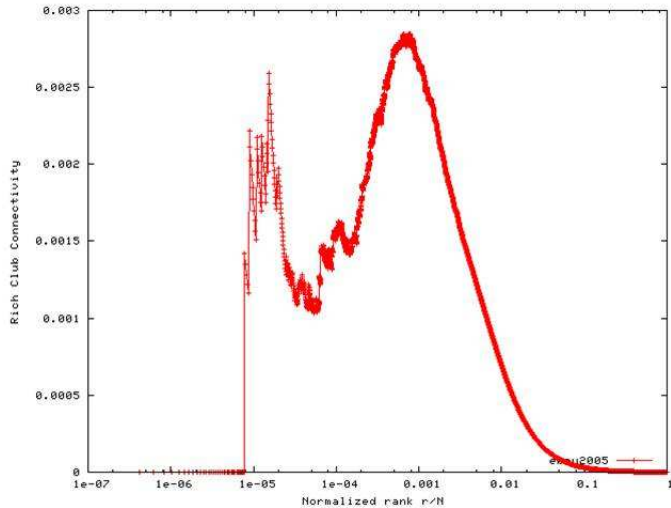


Fig. 3. Rich Club Connectivity.

rich club phenomenon does not appear on the eBay feedback graph (see Fig 3). Unlike the Internet high degree nodes don't interact with each other. In our analysis we consider all the feedbacks exchanged within a one year span of time and we observed the top 61 (0.001%) high degree nodes which are attached to 25% of the edges are disconnected. We repeated the same set of experiments over several years of time and similar behavior exists. We attribute this to the fact that high degree nodes are usually sellers and hence don't 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 . It gives us a general overview of how pairs of nodes are distributed. Fig 4 shows the CDF of the hop-plot of the eBay graph. On the other hand, the eccentricity of a vertex

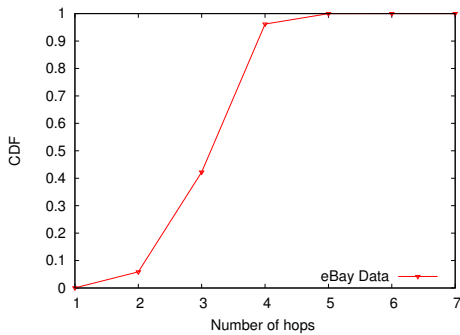


Fig. 4. CDF of Hop plot.

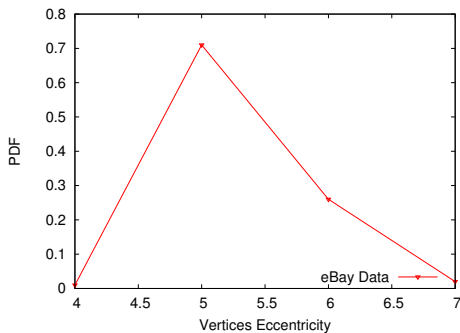


Fig. 5. Node Eccentricity.

v is the maximum distance from v to any other vertex in the graph. It gives us a sense of how far nodes are apart. The maximum eccentricity in a graph is also the graph diameter. The set of nodes with maximum eccentricity forms graph periphery, while nodes with minimum eccentricity belong to graph center. Our results show the minimum graph eccentricity is 4. Fig 5 shows the PDF of Graph eccentricity. As we explain later, we use the set of nodes with minimum eccentricity as central nodes when constructing an intuitive model.

IV. GRAPH EVOLUTION

In the past, many studies have focused on understanding the evolution of various types of graph over time. An in-depth understanding of graph evolution is key to understand and characterize the growth of the eBay graph in general and individual users in particular. Previous study show fraudsters boost their score over short of time by selling low priced goods. Therefore, knowledge of the graph evolution is a first step to detect normal from abnormal trend. However here, we just present the fundamental trends in the ebay graph evolution.

A. Densification

We study the density by computing the average node in-degree of the ebay trust graph over time [8], [9]. 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).

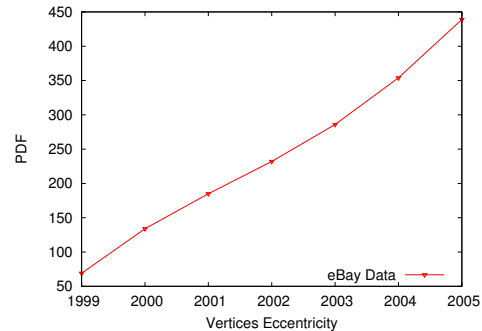


Fig. 6. Graph Densification.

B. Preferential attachment

In scale free 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. That is the probability of connecting to a node is proportional to its current degree. In this analysis, we count the number of new nodes attached to node that have 10 or less negative feedbacks for a period of 3 months, and then we computed the average number of new nodes attached as a function of node degree. our results show the average number of new nodes attached is linearly correlated to the degree especially for low degree nodes. We plot and see preference approximated by a linear relationship to its degree (see Fig 7, Fig 8). We repeated the same experiment over different years and found similar results. Furthermore, we examine the effects of negative feedback and find that users with high negative scores have have a slim chance of attracting new users. We examine nodes with over 10 negative feedbacks and less than 90% positive feedback. No new edge was attached to 50% of the nodes. Thus, linear preferential attachment exists for nodes with low negative score. Negative feedback is rare but few negative ratings can seriously damage the reputation of a user.

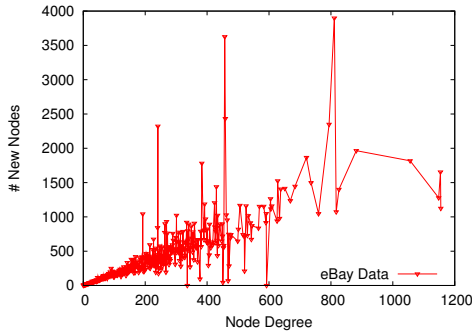


Fig. 7. Preferential attachment.

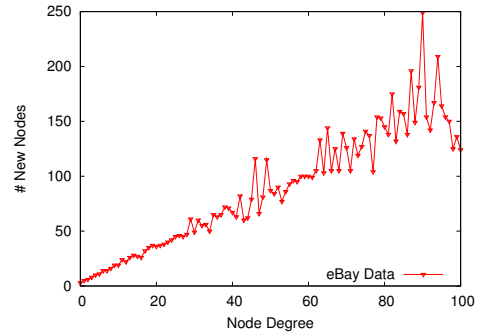


Fig. 8. Preferential attachment for nodes with low degree (less than 100 degree).

V. FEEDBACK AND RETALIATION

As mentioned earlier, users can leave positive, neutral or negative feedback for a person with whom they transacted. Many comments on the eBay feedback forum claim discuss of receiving retaliatory feedback as a reason for few negative feedback. We tried to prove the validity of this perception by answering the following questions. Do people reward a positive feedback, and do people retaliate to a negative feedback?

We study the feedback considering the relative order with which 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%. That is, only 20%, of the people who write negative feedback to a user get negative feedback. The remaining 80% get positive, zero or no feedback. Thus, we believe retaliatory feedback is not as one might think. On the other hand, positive feedback reward rate is around 50%. That is, about half of the users who leave positive feedbacks get rewarded (get positive feedback).

VI. INTUITIVE MODEL

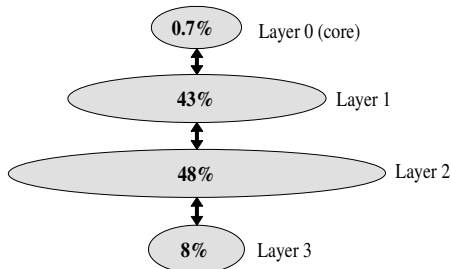


Fig. 9. Eccentricity based Model.

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 define our model, we use the node eccentricity, which is a basic graph metric. The eccentricity of node v in a connected graph G is defined as the maximum distance between v and any other vertex u of G . Note that the maximum eccentricity among all nodes is the graph diameter. Here, we use the undirected graph of the fully crawled nodes with 66K nodes approximately, which is connected as we have mentioned earlier.

To start, we use eccentricity to define a “center” for the graph as shown in Fig. 9. Specifically, we define the *core* (layer 0) to be the set of nodes with minimum eccentricity, which is 4 in our graph. Then, we iteratively define the next layer as the nodes that are directly connected to nodes of the previous layer. Given this definition it is easy to prove that the node of each layer can be at most higher by one compared to the eccentricity of the “parent” layer.

Interestingly, we have only 4 layers in our graph. The core layer contains 0.5% 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 deriving very important graph metric parameters.

First, the model reveals the “shallowness” and compactness of the graph. These two properties appear in many highly skewed and scale-free networks.

Second, the model provides a bound for the eccentricity and diameter of the graph. The eccentricity of core is $ecc(core) = 4$ as we saw, and the eccentricity of nodes at layer i is bounded by $ecc(layer_i) \leq ecc(core) + i$. Given that we have four layers, we find that the maximum eccentricity is bounded by 7, and, thus, the diameter of the graph is bounded by 7. In fact, the diameter is 7 in this graph.

VII. RELATED WORK

Important research efforts quantifying the structure of the Internet can be found in [1], [3] and WWW

[2]. These efforts describe how to sample and generate Internet-scale AS graphs to simulate experiments using large topologies. The social network aspect has been analyzed in [4], [5], where researchers report that the core of such social graphs contains a very dense component linked to some smaller but highly connected components. Similar results have been reported by studies focusing on complex graphs [6]–[8], which feature a giant component and a star structure leading out to the fringes of the network. Another interesting piece of work can be found in [12], where researchers discuss how to provide an implicit reputation score to users by observing the connectivity graph in order to reduce the incentives for collusion. In [13], a similar scheme using distributed agent architectures is described. Some studies focus on detecting fraudsters [10], [11], where they develop data mining techniques to analyze publicly available histories of eBay transactions and identify suspicious online behavior and dubious associations among users. Our work is different from these important research efforts. We analyze feedback-reciprocity among eBay users, extract interesting and representative features of the interaction graph and propose a model based on these metrics. To the best of our knowledge, this is the first detailed study focusing on the structure and evolution of large-scale interactions amongst online auction users with over 54 million eBay transaction feedbacks.

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 find that, unlike the Internet topology, the rich-club phenomenon does not appear on the eBay feedback graph, but the graph has skewed degree distribution and is disassortative.

We also observe that linear preferential attachment exists partially especially for the low-degree nodes. As nodes obtain higher degrees, say above a few hundred, the correlation between preference and degree weakens. One explanation is that once a user seems sufficiently credible, the exact number of satisfied users becomes less important. However, negative feedback 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 feedback is not high: less than 20% and initial negative feedback receives negative feedback in return.

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, pp. 309-320, Proceedings of the Ninth International World Wide Web Conference, 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”, APPC 10: The 10th Asia Pacific Physics Conference (Postech, Pohang, Korea, 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 (IMC’07), San Diego, CA, 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, KDD, 2006
- [8] A.-L. Barabasi, H. Jeong, R. Ravasz, Z. Nda, T. Vicsek, and A. Schubert Evolution of the social network of scientific collaborations, Physica A 311, 590-614 (2002)
- [9] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph Evolution: Densification and Shrinking Diameters, ACM Transactions on Knowledge Discovery from Data (ACM TKDD), 1(1), 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 (WWW’07). May 8-12, 2007. Banff, Alberta, Canada. pp. 201-210
- [11] D. H. Chau, S. Pandit, S. Wang, and C. Faloutsos Parallel Crawling for Online Social Networks To appear at WWW 2007.
- [12] M. Faloutsos, P. Faloutsos, and C. Faloutsos On power-law relationships of the Internet topology, in ACM SIGCOMM, 1999
- [13] B. Yu and M. P. Singh Distributed Reputation Management for Electronic Commerce, in Procs. of IFAA-MAS 2002.