Social Network Analysis of Online Marketplaces

Pushpa Kumar Kang Zhang
Department of Computer Science, University of Texas at Dallas, Richardson, TX 75083
{pkumar, kzhang}@utdallas.edu

Abstract

This paper presents a preliminary effort on visualization and analysis of social networks for online marketplaces. We use one of the most popular e-business models, eBay, as a case study. By extracting relevant data, we generate tree structure displays using 'Prefuse' interactive visualization toolkit, and study the social interaction among sellers and buyers, across geographical boundaries. We also illustrate how this study can improvise our understanding of online customer interaction patterns. Such an analysis could lead to possible enhancements to the current e-business models.

1. Introduction

Social networks and their analysis is becoming a rapidly emerging field of research. Massive quantities of data on large social networks are available from blogs, knowledge-sharing sites, collaborative filtering systems, online gaming, social networking sites, newsgroups, chat rooms, e-businesses and so on [1]. Social networking also refers to a category of Internet applications to help connect friends, business partners, or other individuals using a variety of tools. These applications like "MySpace" and "LiveJournal" are becoming increasingly popular. A social network is a social structure made of nodes, which are generally individuals or organizations. It indicates the ways in which they are connected through various social familiarities ranging from casual acquaintance to close familial bonds.

Recent research has been performed in varied fields like hybrid representations of online activities [13], research community mining [10] and semantic social network analysis [9]. We attempt to utilize concepts of social network analysis for studying networks of online marketplaces. We take eBay, one of the successful online businesses as a case study for our research. A user can sell almost anything through an online auction. The bidding opens at a price the seller specifies and when the listing ends, the buyer with the highest bid wins. With a membership in excess of 135 million users, this model facilitates tremendous interaction between sellers and buyers and is a good candidate for deriving social networks. Currently, there are a few tools readily available to the eBay user like the "eBay toolbar", with many user-friendly customizable features. There is also an eBay Certified Application "biddingBuddies", which is the only social network exclusively for eBay members. In this paper, we attempt to mine eBay customer data from its production environment and construct social networks. The customer profiles retrieved in this manner for sellers and buyers include "userid", geographic location and transaction history. Information visualization techniques facilitate analysis of the network with regards to direct and indirect connections, as well as locations and roles of various participants in the network. Social network graph structure varies with the rate at which selling/buying occurs.

Existing broker models focus on creating markets by bringing buyers and sellers together and facilitating transactions between them. Those can be business-to-business (B2B), business-to-consumer (B2C), or consumer-to-consumer (C2C) markets. Being one of the oldest forms of brokering, auctions have been widely used throughout the world to set prices for such items as agricultural commodities, financial instruments, and unique items like fine art and antiquities. The Web has popularized the auction model and broadened its applicability to a wide array of goods and services [15]. Our aim is to gain insight into customer behaviors with certain patterns, so that these brokerage e-business models could be further enhanced.

The rest of the paper is organized as follows. Section 2 provides a brief overview of social network analysis. Section 3 highlights the tools and methodology relevant to our eBay case study example. Section 4 provides details of visualization and analysis of resultant data. Section 5 presents implementation of the methodology. Section 6 discusses related work, and Section 7 gives the conclusions and future work.

2. Social Network Analysis

Social network analysis or SNA is related to network theory and has emerged as a key technique in modern sociology, anthropology, geography, social psychology, information science and organizational studies, as well as a popular topic of speculation and study [12]. SNA provides statistical tools for examining relational data rather than merely characterizing attributes of individual actors, and focuses on describing patterns of relationships among actors, and analyzing the structure of these patterns. An "actor" is a discrete individual, organization, event or collective social entity that links to others in a network and is represented as a "node" [14]. Elements of a social network can be illustrated in a simple "sociogram" [2]. The nodes in a network are represented as circles, and directional lines represent the links or connections between them. An "Adjacency matrix" is a square matrix, usually consisting of zeros and ones that indicates whether each pair of actors in the network is connected or not [14]. Measuring the network location is finding the centrality of a node [12]. The three most popular individual centrality measures defined below give us insight into the various roles and groupings in a network.

i) *Degree Centrality*: Degree for a node is highest when the node has the maximum possible number of direct connections to other nodes. Degree is thus the number of direct ties to other nodes, and measures activity of nodes in the network. Nodes are 'connectors' or 'hubs' in this network.

ii) *Closeness Centrality*: Closeness for a node is highest when a node can reach all other nodes in the network. Mathematically, closeness is the graph-theoretic distance of a node to all other nodes. Nodes with high centrality closeness are ones most likely to receive and transmit innovations. A node with high betweenness has great influence over what flows in the network.

iii) *Betweenness Centrality*: Betweenness for a node is highest when nodes connecting to other nodes maximally utilize that node. That is, betweenness measures how many paths pass through a node. A node high on betweenness has a high opportunity to play gatekeeper, liaison, or broker role. It is in excellent position and location to monitor the information flow in the network and has the best visibility into what is happening in the network.

For online market businesses, higher values for the three degrees of centrality represent users who are better connected and hence sell more products than others. A few reasons for this may be the following:

- Superior publicity techniques for their products – more cost may be involved
- Selling popular categories of items
- Better location – for example, advertising at the top of the pages for more visibility
- Better visualization techniques – utilizing sophisticated cameras for higher resolution and improved quality of images for added visual effects

An online market business may additionally choose to target advertisements and promotions for these better located users.

Through social network analysis of online market places, we aim at answering the following questions:

- What are the typical seller-buyer relationships? What is the buyer's loyalty (buying the same kind of items from the same seller)?
- What categories of goods are popular?
- What are the correlations among the number of buyers, the number of for sale and the number of sold items?
- Measurements of networks relating to centralities.

3. Case Study - eBay

This section provides a brief overview of the tools and concepts utilized for our case study**.** The eBay Web Services package provides programmatic access to the eBay marketplace and enables third-party applications to build custom applications, tools, and services that leverage the eBay marketplace in new ways [11].

We use an extensible software framework called 'Prefuse' that helps software developers to create interactive information visualization applications using Java. Fundamental to our work on discovering and analyzing social networks on eBay, is the mining and extraction of relevant data from the eBay. To accomplish this task, we utilized eBay provided web services package and the API test tool described above. The overall methodology is depicted in Figure 1.

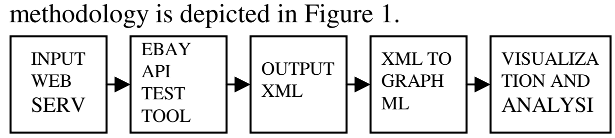

Figure 1. Overview of methodology

eBay related Web service calls in the form of input XML are forwarded to the API test tool in a production environment, which outputs the corresponding resultant XML. This output XML which is information rich, can then be mined for our purposes. The resultant XML was converted into the toolkit requirement of GraphML format, and visualized using both tree-based structures and force-directed layouts in Prefuse. We perform social network analysis (SNA) on the retrieved data.

4. Data Visualization and Analysis

Using Prefuse, we visualize the resultant data that was converted to the required GraphML structures. We aim to study buyer-seller interaction and understand user habits and popular categories of sold items. This kind of analysis can help in enhancing the business strategy and practice, and help market researchers to better predict and hence market their products over time. The overall trends can be seasonal or geographical with some categories of products selling better than others for a few months, or in certain parts of the world.

A screenshot of the resultant hierarchical structure is partially depicted in Figure 2, where the number in square brackets ([]) behind the "userid" represents multiple purchases from the same seller. There are a total of 531 users with one root node "svg2331" and the remaining nodes are distributed among four hierarchy levels. The root node has 17 children nodes.

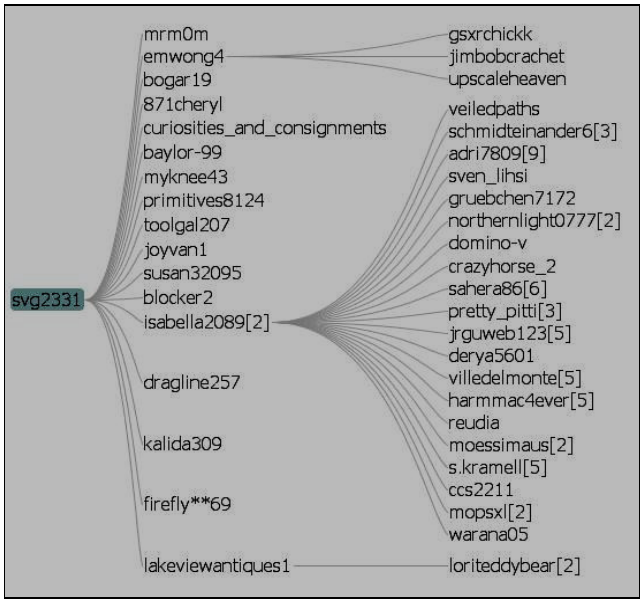

Figure 2. Screenshot of tree visualization

We performed node reduction by removing cross-linking across multiple hierarchy levels for the same node. The aim is to eliminate repetitive nodes and achieve a more efficient tree structure. Figure 3a depicts the original tree structure, while Figure 3b depicts the improved version for tree node 'A'. There were a total of 52 sellers, 483 buyers and 36 redundant cross linked nodes across four levels of hierarchy resulting in a 7% node reduction.

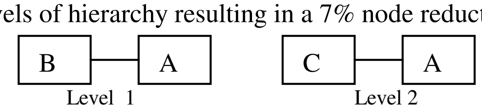

Figure 3a. Original tree structure

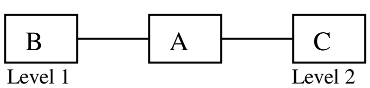

Figure 3b. Reduced tree structure

Using "Parvis" visualization tool, parallel co-ordinates visualization was generated for 52 sellers with regards to the number of buyers, total items sold and number of sold items and is depicted in Figure 4. As can be seen from the figure, there is a direct positive correlation between the total items and number of items sold. The number of unsold items is derived from total number of items and number of sold items. To study buyer loyalty to determine if the buyer is purchasing the same kind of items from the same seller, we extracted information for 15 users over four hierarchies and a quarter of the year (Nov 06 to Feb 07). We observe that except for one user ("northernl") most buyers (93%) bought items from the same category for the same seller over the period of time.

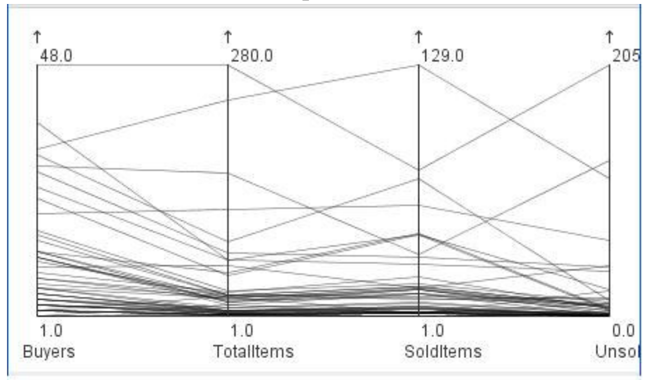

Figure 4. Parallel coordinates for seller activity

To identify popular categories for sold items, we assigned a two letter code for each of the 18 categories obtained for 531 users utilizing the "GetItem" webservice call and extracting <PrimaryCategory> node information from the resultant xml. "GetItem" retrieves details of an item for a particular "itemid". Figure 5 visualizes the category information using a force directed layout with 18 different colors assigned to each category.

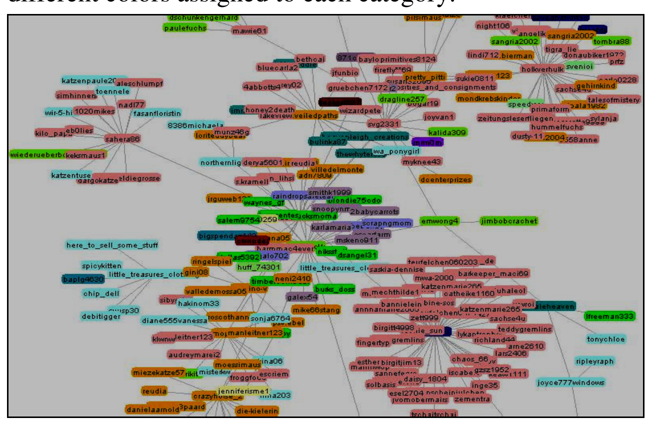

Figure 5. Force directed layout – categories

It is interesting to note that some sellers sold completely different items from what they bought, while for some sellers almost all buyers bought items from the same category. From this analysis, we conclude that the three most popular categories of items sold are "Books (46%)", "Collectibles (18%)" and "Clothing, shoes & accessories (12%)" for the users under our current study.

As defined in Section 2, an "actor" is a discrete individual, organization, event or collective social entity that links to others in a network. We can measure the degree of centrality for the top ten most active and visible actors in the network for studying their social interaction. They have the highest number of non- repetitive buyers and hence highest interactions for the time period of our study. Figure 6 shows the sociogram for these actors, where arrows represent "direct" connections between them. Relevant statistics about the three centrality measures: Degree Centrality, Betweenness Centrality and Closeness Centrality, calculated from the constructed sociogram are depicted as a bar chart in Figure 7.

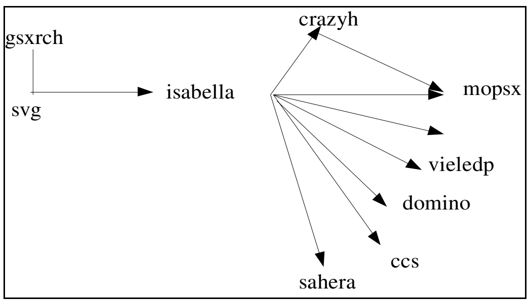

Figure 6. Sociogram for various actors

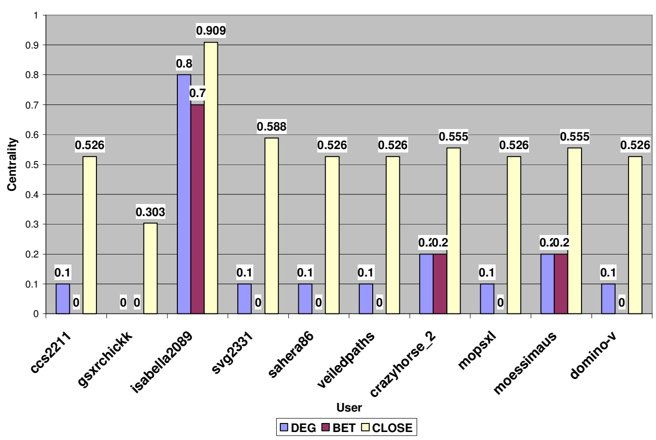

Figure 7. Bar chart for centrality measures

We observe that the most active actor (e.g. ccs) may not necessarily be the socially best-connected actor with regards to degrees, betweenness and closeness centrality measures (e.g. isabella). This analysis provides insight into users with the best location in the network under consideration. "Degree Centality" (DEG in Figure 7) varies between 0 and 80%, "Betweenness Centrality" (BET) varies between 0 and 70% and "Closeness Centrality" (CLOSE) varies between 30.3% and 90.9%.

5. Implementation

This section discusses implementation of the methodology discussed in Section 3. We customized the XML API for web service "GetSellerEvents", by adding "UserID" child element corresponding to the seller in whom we are interested. We also modified the "ModTimeFrom" and "ModTimeTo" element data to the time period we are interested in. The input XML call retrieves a list of the items for which a seller event has occurred. A seller event is an event that is of interest to an item's seller, such as the change of the current bid price or the item's ending time. "GetSellerEvents" returns the data for the item into an ItemArray object [11]. The root node "ItemArray" comprises of many "Item" child elements. Relevant data is stored in various child elements of the "Item" node. We observe that the data values for "userid" and "Site" can be utilized for the construction and analysis of social networks. The "userid" refers to the buyer of the sold item, while "Site" is the geographical location data. Prefuse toolkit requires conversion of XML data to its equivalent XML format consisting of branches and leaves for a tree structure. For social network graphs, we convert the XML data to a GraphML format.

6. Related Work

Most of the related ongoing research is in the areas of social network analysis in varied fields, and web visualization. A new method of pattern-based visualization can enable business owners, application designers, programmers, and operations staff to quickly understand the behavior of complex web services applications [4]. Web visualization framework architecture (WVFA) is a conceptual model for the Web visualization that is closer to human visual perception [6] and a generalized framework has been developed for determining how to choose the right set of technologies for a web-based interactive visualization [7]. A visual data mining system that uses a new algebra to operate on web graph objects helps to discover new patterns and interesting hidden facts in web navigational data [11].

Recent research has been performed on the structure and evolution of online social networks [5] and how we can mine social networks for purposes such as marketing campaigns [1]. The structure and dynamics of social groups reveals the factors influencing an individual to join a particular community [3]. Systems such as the Web Access Visualization (WAV) have been used to visualize the affinities and relationships for large volumes of web transaction data [8]. Semantic social network analysis is facilitated by tools such as "iQuest", which is a software system that is based upon a "grammar" and allows us to understand and visualize patterns of communication [9]. The "Friend of a Friend" (FOAF) project is the Semantic web's largest and most popular which is a vocabulary for describing people and whom they know [1]. In this paper, we have attempted to utilize and apply concepts of social network analysis towards online marketplaces.

7. Conclusions and Future Work

Online marketplaces house vital information for visualizing and analyzing the behaviors of online customers through social networks. As our case study tool, eBay web services package provides a convenient mechanism to mine relevant data and derive social networks. Visualization of tree hierarchies indicates different interaction patterns. Statistics obtained from eBay data reveal interesting insights into roles of various actors in the network and their measure of centrality. A buyer of a product can reside in a geographical location completely different from the seller of the product, yet can be closely connected through the social network. Also, the most visible actor needs not necessarily be the one that has the best location in the network. User behavior analysis can help us to further understand the potential trend. Existing brokerage e-business models can be enhanced by insight into customer relationships and behaviors. This paper presents a preliminary effort. Our future work will involve a more in-depth analysis of SNA tool for user behavior, and employing clustering mechanisms along with visualization to provide more intelligence.

8. Acknowledgment

We would like to thank the US Department of Education GAANN Fellowship program for funding this project.

9. References

[1] S. Staab, P. Domingos, P. Mika, J. Golbeck, L. Ding, T. Finin, A. Joshi, A. Nowak, and R. R. Vallacher, "Social Networks Applied", *IEEE Intelligent Systems*, Vol. 20, No. 1, Jan/Feb, 2005, 80-93.

[2] E. F. Churchill and C. A. Halverson, "Social Networks and Social Networking", *IEEE Internet Computing*, vol. 9, no. 5, 2005, 14-19.

[3] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group Formation in Large Social Networks: Membership, Growth and Evolution", *Proc. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2006, 44-54.

[4] W. D. Pauw, S. Krasikov, and J. F. Morar, "Execution patterns for Visualizing Web services", *Proc. 2006 ACM Symposium on Software Visualization*, 2006, 37-45.

[5] R. Kumar, J. Novak, and A. Tomkins, "Structure and Evolution of Online Social Networks", *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, 2006, 611-617.

[6] S. Eibe and O. Marban, "WVFA: A Web Visualization Framework Architecture", *Proc. 2004 IEEE International Conference on Information Reuse and Integration*, 2004, 103-108.

[7] N. Holmberg, B. Wiinsche, and E. Tempero, "A Framework for Interactive Web-Based Visualization", *Proc. 7th Australasian User Interface Conference*, Vol. 50. Australian Computer Society, Inc. Darlinghurst, 2006, 137-144.

[8] M. C. Hao, P. Garg, U. Dayal, V. Machiraju, and D. Cotting, "Visualization of Large Web Access Data Sets", *ACM International Conference Proceeding Series*, Vol. 22, Eurographics Association, Aire-la-Ville, 2002, 201-ff.

[9] P. A. Gloor and Y. Zhao, "Analyzing Actors and Their Discussion Topics by Semantic Social Network Analysis", *Proc. 2006 Conference on Information Visualization*, IEEE CS Press, Washington DC, 2006, 130-135.

[10] R. Ichise, H. Takeda, and T. Muraki, "Research Community Mining with Topic Identification", *Proc. 2006 Conference on Information Visualization*, Washington DC, 2006, 276-281.

[11] http://developer.ebay.com/support/docs/

[12] http://www.orgnet.com/sna.html

[13] Y. "Eugene" Medynskiy, N. Ducheneaut, and A. Farahat, "Using Hybrid Networks for the Analysis of Online Software Development Communities", *Proc. SIGCHI Conference on Human Factors in Computing Systems*, 2006, 513-516.

[14] M. Kilduff and W. Tsai, "*Social Networks and Organizations*", 1st Ed, Sage Inc, 2003.

[15] http://digitalenterprise.org/models/models.html