

©2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Finding Influential eBay Buyers for Viral Marketing - A Conceptual Model of BuyerRank

Hon Wai Lam, Chen Wu
Digital Ecosystems and Business Intelligence Institute
Curtin Business School
Curtin University of Technology
Perth, Western Australia
{S.Lam, Chen.Wu@cbs.curtin.edu.au}

Abstract—User Generated Content (UGC) plays a key role in the current and future e-commerce at the era of Web 2.0. As an important form of UGC, the online auction site eBay has enjoyed tremendous growth rates since its inception. Many social networks emerge across different communities on eBay. The notion of viral marketing has been proposed in both literature and practice. However, in order to find the “opinion leader” in the social network, marketers need to have a sound analytic tool to rank potential buyers. In order to tackle this issue, this paper propose a BuyerRank, a Social Network Analysis (SNA) model that assists marketers to rank potential buyers based on their future influence estimated from their past auction/purchase behaviour on eBay. The paper also provides a detailed state-of-the-art review of recent work on SNA and viral marketing in the light of the Web 2.0 e-commerce context.

Keywords—eBay; e-Commerce, Social Network Analysis; Viral Marketing; User Generated Content; PageRank; Web 2.0

I. INTRODUCTION

Humans are flooded with a tsunami of data in the digital world. According to [1], 161 billion GB of data was created, captured, and replicated in 2006. This is “more than in the previous 5,000 years” [2]. Gantz [1] further estimate that 75% of the data was actually created or replicated by individuals through User Generated Content (UGC). By 2010, the total amount of data will add up to 988 billion GB, of which 70% will be created by consumers [1]. As an important form of UGC, the online auction site eBay has enjoyed tremendous growth rates since its inception. At present day, eBay has up to “100 million items available for sale, with 6.4 million new items added everyday” [3]. The amount of active users in eBay also enjoys an exponential increase, with up to 82.9 million users within the first quarter of 2007, 10 percent more than the first quarter in 2006 [4].

Within this auction website one can find many social networks. A social network can be loosely defined as a community of people, which have direct contact with one other, most commonly through face-to-face or phone conversations, communication through email, instant messaging and such. Awareness for this in the e-commerce setting often leads to a phenomenon known as word-of-mouth (WOM), a form of viral marketing. Viral marketing is a concept in which one person would tell one or many other people within their social network about products or services that they like or dislike, thereby having the potential to cover the entire network. In recent studies, it is shown “that more than 75 percent of customers

will consult a friend before deciding on the purchase of a certain product or service” [5].

Previous studies in marketing also suggest that the key success factor of WOM-based viral marketing strategy is the identification of the most “influential” person in the potential buyer community [6]. However, it is unclear, based on the existing research, how such an influential person can be found using an analytic tool that doesn’t rely on a marketer’s intuition or heuristics. This gap has motivated us to propose the BuyerRank, a social network analysis model that assists marketers in ranking potential buyers based on their future influence, estimated from their past auction/purchase behaviour on eBay.

The major contribution of this paper is twofold. First, we provide a detailed state-of-the-art review on recent work on Social Network Analysis (SNA) and viral marketing in the light of the Web 2.0 e-commerce context. The summarised table allows future researchers to horizontally compare a majority of the SNA research within the same graph model. Second, we propose a conceptual model for ranking buyers within eBay. The ranking model is divided into two parts, UserRank and BuyerRank. The concept of UserRank borrows essential parameters from the PageRank [7] algorithm used in the search engine Google with new interpretations. The BuyerRank algorithm is an extension of UserRank in order to determine the ultimate ‘influential score’ of a particular eBay buyer.

The remainder of this paper is organised as follows. In section II we discuss recent research into online social network analysis, together with a summary of the graph models used in those studies and also studies pertaining to viral marketing. Section III provides a number of key concepts such as *Prestige*, *PageRank Authority*, and *Hub*. Section IV describes the proposed UserRank and BuyerRank algorithms and provides some examples. Section V provides some discussion and the possible directions of future works and Section VI concludes.

II. RELATED WORK

A. Online Social Network Analysis (SNA)

Social network analysis has been the basis for many forms of marketing efforts. The basic idea is to employ analytic techniques to a specific network of consumers, who can be distinguished from a whole community, together with their

opinion leader/s in order to carry out marketing strategies appropriate to that community.

For example, Kumar and Zhang [8] used eBay as a case study to better understand the social aspects of online marketing networks. Their illustration of a social network consisted of actors (individual users in the network) represented as nodes, connections between two users represented as lines, and a square matrix, also known as an Adjacency Matrix, consisting of 1s and 0s, which indicated the connection between a pair of actors [8]. The study stated 3 centralities that can be used to 'rank' individual node locations in the Network: *Degree Centrality* - the number of direct connections one node has with other nodes in the network; considered the highest when a node has the most number of direct connections possible in the network; *Closeness Centrality* - the distance separating one node from another. This is considered highest when one node is able to interact easily with other nodes since the distance to all other actors is the shortest. High closeness shows that a node will be more likely to receive and transmit innovations; *Betweenness Centrality* - the number of paths that a node is within, or how many paths include that individual node. 'Gatekeeper', 'liaison' and 'broker' roles are most likely played by nodes with a high Betweenness, since that node would have high visibility of the information flowing throughout the network.

A node with a high score in these centralities might indicate that the actor is better connected and thus would be able to sell more products or even influence others into buying a particular product more than other actors. A few reasons for this stated in the article include better publicity, which would incur more costs, the products are of a more popular category, advertisements having better placement and better visualization techniques [8]. The analysis result showed that the most active user/actor in the network was not always the highest ranked in terms of scores based on the 3 centralities [8]. This suggested that users with a high number of transactions do not necessarily belong to the best market segments in the social network for marketing purposes. Also, with today's technological advancements, it might be possible that the two actors can be in completely different geographical locations but still be highly connected in that particular social network.

Social network connections may include emails, telephone calls, text messages and transactions. Doyle [5] proposed a method to identify a social network, which required the identification of opinion leaders, people who have a high influence over the network for reasons such as having the relevant knowledge about a product or service, or who are one of the first few people who have purchased the product or tried out the service [5], like for example the first batch of people to purchase the Sony Playstation 3 console or a developer for Sony. Opinion leaders also have other traits such as good communication skills, being a part of a large social network and are keen to learn new pieces of information [5]. By identifying social networks and their opinion leaders, marketing efforts can be targeted to a few specific customers with the assumption that they would in turn spread the word thereby advertising the product through word of mouth. In the case studied in [5], data was collected and analysed to produce a set of key outputs which described the social effects of each individual in the network. With this data, the analysis team was

able to determine why, when a certain customer stopped purchasing from the store, all other members of the network also stopped. The reason was that the particular customer was an opinion leader in the network; it so happened one day that she was unhappy with the service of a sales assistant and the refund policy of the store when she tried to ask for a refund and hence told all of her friends not to purchase from that store anymore, causing a massive chain reaction [5]. Word of mouth is therefore one of the key factors that should be taken into consideration in marketing strategies, one or many good reviews about a product or service could be exponentially more effective than a well placed advertisement. Similarly, a bad review could be extremely detrimental to the reputation and hence the popularity of said product/service. Interestingly though, it is a common occurrence that one bad review by an opinion leader usually overshadows a multitude of good reviews other reviewers, this should also be taken into consideration when relying on word of mouth marketing strategies.

Relationships that people share with one another also appear to have a substantial effect within a social network. Singla and Richardson [9] attempted to study the relationship between a social group and the personal behaviour of a person within it. The purpose was to determine if two people that talk to each other are more likely to share similar interests. Two data sets were used in this research, referred to as the *Social Network Data* and *Personal Interest Data* [9]. The first set was obtained through message logs that were available from the MSN logs, which provided the researchers with data on the interactions between users of the network, such as the time and length of the chat between two users. The second was made up of keyword searches made by users of the MSN search engine. The subsequent correlation data analysis suggested that people who communicate with each other on instant messaging networks such as MSN would more than likely have more in common, such as categories when browsing through the net, over a random pair of users (e.g. two users from different social networks) [9].

It can be seen that through online communication, geographic location is no longer a barrier in transmitting information; marketing efforts can take the form of electronic brochures not like those in the spam category sent to only a select few users who have a high influence in a social network, on or offline. Singla and Richardson [9] further confirmed that identifying a person within a large social group and his/her interest, would be a huge advantage when targeting customer segments.

B. Graph Based Analysis

One of the more common representations of a social network, online or offline is the graph-based model. In analysing eBay data, Beyene et al. [3] constructed three graphs, namely the trust, transaction and undirected graphs and discovered the characteristics of those graphs and how they evolved over seven years [3]. For the trust graph, the data was gathered from reviews or feedbacks left from one eBay user to another, which comes in the rating of 1, 0 and -1 representing positive, neutral and negative feedbacks respectively. The feedback score is calculated by subtracting the negative feedback score left by each unique user from the total of the

positive score. Thus users with a high feedback score would be more reliable or according to the article, more trustworthy thereby giving said user more privileges over other users [3]. The data was modelled into a way where a user was represented as a node, with the feedback taking the form of an edge, where a node can be classified as a buyer, seller or a trader, which is a user that buys and sells [3]. The subsequent graph analysis revealed several interesting findings. Firstly, the graph differs from commonly used graphs (e.g. Internet Topology) in that the rich club phenomenon (where high degree nodes are well connected) does not occur [3]. Secondly, feedback is not always returned and users seldom leave negative feedback for fear of receiving negative feedback themselves. Finally, users seem to suffer a huge drop off in reputation should they receive feedback that is negative in nature.

item, and is complementary if one item adds value to another [10]. Examples of substitutes can include two different models of printers each with similar features and an example of a complementary item would be the ink cartridges. The authors focused on two markets, which were the branded digital cameras and the Liquid Crystal Display (LCD) screens markets.

The data in [10] was programmably collected by searching closed listings on eBay. The next step involved the construction of a bidding network by creating a graph to represent the information, followed by a series of data refinement steps (e.g. excluding data “noise” from extreme bidding characteristics) to come up with a resulting illustration that according to the article, represented the “largest maximally connected component of the Canon bidding network.” [10]. Using the resulting graph, the researchers identified a community within the network using the “Greedy Q” algorithm developed by [11]. The results showed that substitute and complementary items could be identified from a network by applying community detection methods, which in turn allowed the researchers to determine a collective preference within a bidding network [10]. This work provided very interesting findings on bidding behaviours; certain products that seem to appeal to a network can be marketed together with a complement which could potentially save costs, since the preference throughout the network is known, which could lead to more fruitful results in marketing efforts. On the other hand, this study only focuses on “one-time” purchase items (e.g. digital cameras and televisions), which might be inapplicable to other genres of items, like for example computer games, where there would be many repeat purchases and while two items might be substitutes (two kinds of action games), a bidder might choose to bid on both items and win, thereby giving the impression that those two items are complementary.

A study of a similar nature conducted by Mislove et al. [12] looked at four social networking sites, namely Orkut, YouTube, LiveJournal and Flickr; examining the data gathered from these online sites to provide a deeper understanding of the nature of online social networks and to improve designs in current and future applications that would incorporate them. The data was obtained through crawling the user graphs by gaining access to the public web interface, which was provided by the respective sites, which allowed data to be collected from multiple sites [12]. The constructed networks were compared with each other based on the data and their similarities and differences were also compared and contrasted with the properties of the Web that were already available. The results of the analysis showed several interesting characteristics, one of those points is that in networks where high level nodes (nodes which many other nodes connected/point/linked to) are more inclined to connect to other high level nodes, that network is more susceptible to an “epidemic” or an outbreak of information, viruses etc. while networks with a converse nature are in turn less likely to be subject of one [12]. Other observation is that the networks follow a power-law, in which there are a small number of highly connected nodes and new nodes are more likely to connect to them than other nodes. This somewhat creates a core in which the entire network is held together by a small collection of nodes (about 10% according to the study) with the highest connectivity. Should this core be

Article Title	Node	Edge	Edge Value	Description
Social Networks Analysis of Online Marketplaces	Users in the eBay network	Any connection or link between two actors in the network (e.g. direct or indirect links to another user)	0, 1 in the square matrix. No values are assigned in the undirected graph	A square matrix (Adjacency Matrix), which consists of 1s and 0s indicates the connection between a pair of actors. Graphs are undirected
Yes, There is a Correlation – From social networks to personal behaviour on the web	Users in the instant messaging network	Similarities in interests or keyword searches that two users share with each other/ Correlation of user searches	Determined by defined formula	Formula is used to compute similarities and likewise for correlation of searches made by users. Graphs are undirected
The eBay Graph: How do online auction users interact?	Users on the eBay online auctioning program	Reviews left by one user for another – Trust graph Transactions between users – Transaction graph Any feedback or transaction between any two users – Undirected graph	0,1,-1 for trust graph only	3 different graphs measuring trustworthiness, transactions and connections between users. Trust and transaction graphs are directed
Analysis of Bidding Networks in eBay : Aggregate Preference Identification through Community Detection	Each individual auction is considered a Node	Edges are drawn when there is a common bidder in two separate auctions	Edge value is equal to the number of shared bidders in any two auctions, after some filtering	Graph is used as a base for community detection. Graphs are undirected
Measurement and Analysis of Online Social Networks	Users in a online social network site	Differs from each graph that is used in the study, but mostly represents the path/link/connection from one node to another (e.g. when a user shares/uploads videos or pictures on a social network site such as YouTube)	No values assigned	Several graphs are used in this research to identify the characteristics of online social networks. Graphs are directed
Network-Based Marketing : Identifying likely adopters via consumer network	Consumers within a dataset provided by the firm to the research team	Communication between the consumers’ network neighbours (e.g. talking over the phone, emails, general conversations, etc)	Each edge value range is different for each attribute	Information on analysis is presented in tables and histograms. Graphs are undirected

Table 1: Summary of graph models

Unlike traditional SNA methods which solely treat people as actors, Jin et al. [10] present alternative definition for actors in SNA. The aim was to find out whether substitute and complementary goods could be analysed to automatically predict bidder behaviour. An item is a substitute of another if the one item’s value to the bidder holds the same for another

removed, the network would be dispersed into many loosely connected clusters consisting of low level nodes [12]. This study provided deeper insights into the behaviour of actors in an online social network, which would help immensely in the development of future systems based on social networks, as mentioned by the authors. With the basic knowledge of how networks function, marketing efforts could be adjusted accordingly to suit these characteristics which would bring about a greater effect for any service or product.

In analysing the online SNA literature, we find that graph theory has been frequently used for modelling the social network. A typical graph model often includes: nodes, edges, and edge values. It is very useful to examine and compare how the graph model is constructed in previous studies. Therefore, we provide a summary of the graph models used in the aforementioned studies as shown in Table 1.

C. Viral Marketing

Viral marketing has become a more prominent issue with regards to marketing practices over recent years. Companies are seeking for more effective methods of marketing in favour of the traditional methods such as customer segmentation for niche markets and mass marketing to gain awareness for a certain product. The recent advent of Social Web has further facilitated the adoption of viral marketing.

Domingos [13] first addresses the issue of traditional marketing practices which ignore the network value of customers. The main purpose was to demonstrate the significance of marketing only to a few customers with high network values. A customer's network value is the predicted number of sales to other customers that is a result of marketing to that customer [13]. In other words, how well that one customer can influence others to try a product or service if s/he favours it. Using the proposed model it was possible to predict if a customer is more or less likely to use a certain product or service. The model includes several factors that determine the customer network value. Firstly, the customer has to prefer using that product or service, ideally a lot more than its competitors, since marketing to a customer with a high connectivity in the social network that dislikes that product would instead have negative effects [13]. Secondly, the customer should have more influence over his peers than them on him [13]. Lastly, the people that the customer can influence should in turn have some influence over other customers not directly linked to the first customer [13]. This condition was stated as one of the more important factors in determining the network value of a customer since it can potentially cover the entire network. One interesting finding is that a customer with a low connectivity in a network might in fact have a high network value if he/she can influence another customer who is highly connected [13].

This reinforces the theory that blindly marketing to every customer should be avoided since it is shown that marketing to a select few with high network value would prove more effective. Also, mass marketing could have more negative effects than good should a certain customer with a high network value dislike the product and hence influences others not to use it. Although providing new insights into the basis of viral marketing, an actual illustration of the model was not

provided, but the benefits and capabilities of the model were clearly stated (the model could still be accurate with only 5% of the edges known), though without an example of how the model and data mining system work, it is hard to conduct further studies based on this article alone.

Similar to [13], Subramani and Rajagopalan [14] developed a viral marketing framework based on two key factors: "role of the influencer", which determines if the influence of an actor is passive or actively persuasive, and "network externalities", which are benefits that occur when a product or service is used by a larger number of users within a community [14]. The framework presents a high level view of viral marketing which is split into four quadrants: Awareness Creation and Benefits Signalling (ACBS), Targeted Recommendation (TR), Signalling Use, Group Membership (SGM) and Motivated Evangelism (ME). The framework paves the way in the study of viral marketing since it provides a high level view of consumer behaviours within an online social network. It also shows where these characteristics would be most prominent. However, the framework does not account for abnormalities or extreme behavioural events which would occur in any network. It relies heavily on normal human behaviour which is reliable in most cases but it would have been more interesting if the outliers were taken into account as well.

As mentioned previously, traditional methods in marketing may not be effective in the present day. A study by Hill et al. [15] aimed to see if traditional marketing methods would cause firms to overlook potential customers that fell through the 'gap' of these methods and also to find out if other methods would be able to identify them. The author discussed three modes of marketing that might be complementary to each other: Explicit Advocacy – this is where individuals become publicists for the product of service itself by recommending it to other people. Implicit Advocacy – this mode is where consumers promote a product by using it themselves, thereby influencing the people around them. Network Targeting – this is where marketing efforts are focused on the consumers' network neighbours (people who have direct communication with the consumer). The dataset was derived from a marketing campaign to potential customers of a communications service.

The analysis of the data in [15] revealed three main results. Firstly, consumers who have direct communication with an existing consumer of a firm, also known as network neighbours, are more likely to adopt a product than the marketing list of consumers selected by a firm's marketing team [15]. Secondly, statistical models that are built based on customer information are greatly improved when network information is included. Lastly, with more details about a social network, marketing efforts can be targeted at a small number of consumers with the highest probability of adoption [15]. This work also proves the effect of identifying opinion leaders and their networks in marketing campaigns. Traditional methods that target large customer segments can be likened to fishing vessels casting a large net over the water without actually knowing where the fish are. This study shows that gathering information about social networks, the authorities (opinion leaders or nodes within an online network with a high level of indegree) and their neighbours could potentially cover the entire customer network with minimal resources used.

III. KEY CONCEPTS

A. Prestige

Prestige is a method of measuring the importance of an actor in regards to a social network. In comparison with the degree centrality mentioned by [8], prestige measures the number of in-links (number of other actors that link to an actor) an actor has. Also, the prestige can only be calculated if the relations between the actors are directional and the graphs are directed. Wasserman and Faust [16] define 3 measures of prestige: Degree Prestige, Proximity Prestige and Rank Prestige, the last of which is used in many search engine algorithms (e.g. Google).

Degree Prestige – this is the simplest measure of prestige. Plainly speaking, an actor has a high prestige if he has a high number of in-links. As the following diagram explains:

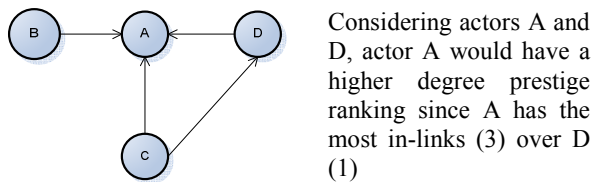


Figure 1: Degree Prestige

Proximity Prestige – this measure considers actors that are adjacent to one actor. In other words, actors that are connected link to another actor through either a direct or indirect link. Note that this is a directional link and is therefore in contrast from the Betweenness Centrality measure in which the links can go either way. The following diagram illustrates:

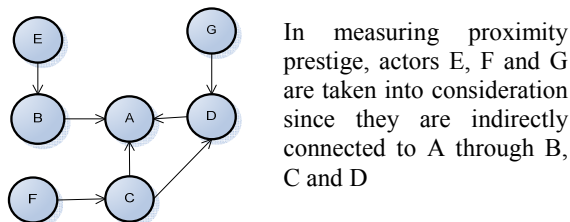


Figure 2: Proximity Prestige

Rank Prestige – this measure considers the prestige of an actor that links to another actor. For example, if actor A, who has a high prestige value, links or votes to/for another actor B, then actor B in turn enjoys a high prestige value. A simple illustration provided by [16] says that within a company, a recommendation from the CEO would carry a much higher weighting than would a good peer review from a regular employee. Therefore an actor's prestige value is directly affected by the prestige value of his network neighbours. As the figure below explains:

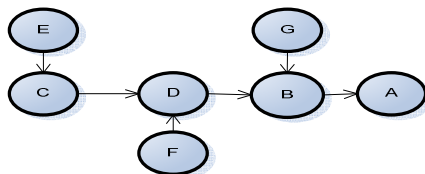


Figure 3: Rank Prestige

In the figure, if the actor D has a highest prestige in the network, then his linking to actor B carries more weight than a link from actor F or E, therefore actor B, receiving an in-link from D would therefore gain prestige from D's prestige rank and in turn, actor A benefits from this as well. High prestige in the context of PageRank is the number of in-links that a page receives which the algorithm uses to determine its quality and credibility. Therefore a link from a page with a high number of in-links would carry more weight than from one with few or no in-links.

B. PageRank

The PageRank [7] algorithm uses the number of links to and from a particular webpage to gauge its quality. If a particular webpage A has a link to another webpage B, it is interpreted that page A has cast a vote for page B and so on. Although a good indication of a page's quality in that administrators of any website cannot easily create hyperlinks to their own sites from other pages, which protects users from spam, this criteria is not the only one within the PageRank algorithm. PageRank also takes into account the issue of who votes for whom, in other words, the quality of the pages that cast a vote for any other page. This is likened to the Rank Prestige logic [16] in social network analysis, where should one party with a high amount of influence or credibility vote for someone or something, then that someone or something would gain a huge amount of credibility based on that one vote. Therefore, in terms of the PageRank algorithm, pages with a high quality would cast a vote with a higher weighting than one of a lower status. The PageRank value of a web page is therefore determined by the number of pages that link to it and also the value of those individual pages. [16]

The main advantage of PageRank is that it is more resistant to spamming. As mentioned above, it is not easy for page owners or administrators to add links to their own pages from other pages; therefore it is difficult to affect the PageRank value by influencing the variables involved in the algorithm. The unique aspect of PageRank is that it is query independent [16], meaning that the values of the pages are saved before a query is entered in the search engine, therefore when a query is entered PageRank just looks up the query and together with algorithms produces a result, making it more efficient.

Being query independent is also one of the disadvantages that PageRank faces. According to [16], the algorithm could not distinguish between two pages, both having a high quality but with only one that was related to the search topic. Another disadvantage is that PageRank does not factor in the age of the web pages it values. For example, a page would have a high value if there are many links to it from other sites, but this also means that the page might have been around for a long period of time and would not have updated information. In other words, pages with a high value 1 year ago might not be relevant at present day, but the PageRank value of it is still high. This also means that newer pages that would be more relevant and of a high quality might be given an unfair value since they have not been on the web long and would hence have fewer in-links to it than an older page [16].

C. Authorities and Hubs

The Hypertext Induced Topic Search (HITS) algorithm is query dependant and uses two primary variables when producing a result for a query, namely a page's authority and hub ranking [16]. A page with a high authority is one with many in-links while a hub, in contrast to an authority is one with out-links to other pages. The logic here is that a page with content that is highly relevant to the query or search topic would have many other pages linking to it due to its credibility and quality of content. A hub therefore serves as a gateway of sorts and users that visit this page can gain access to pages that have authoritative content regarding the users' query. The key concept of this is that good hub would link to authoritative pages and those pages in turn would have many in-links from other hubs, thus resulting in a mutually beneficial or according to [16], a mutually reinforcing relationship.

The main advantage stated by [16] is that the results that are produced are more likely to be highly related to the search query, since it is query dependent, thereby gaining an edge over the PageRank algorithm. The disadvantage though, appears to outweigh the advantages. One of which is that hub scores are easily influenced or manipulated by site owners by adding many hyperlinks to their pages that point to sites that with a high authority, which therefore makes it more susceptible to spam. Since the hub and authority sites are closely connected, the ranking of the authority site is also thus affected. Also, being query dependant, the amount of time it takes to gather the relevant sites and rank them according to the algorithm is rather substantial [16].

IV. USERRANK AND BUYERRANK

In this section we propose a method similar to the PageRank algorithm; the basic principles are kept and extended to be employed in this study. As mentioned in the previous section, the fact that the PageRank algorithm is resistant to fraud is one of the main reasons why it was chosen as the base to expand upon (users in eBay are unlikely to create separate accounts in eBay just to buy items off themselves in order to raise their rankings). Just as PageRank is used to search for web pages, UserRank is used to search and rank Users (in this case, bidders in eBay or any other online auction) based on certain criteria. Rank prestige will be taken into account and graph edges that are used in our study are directional. We first discuss the UserRank concept and algorithm. We then elaborate the BuyerRank conceptual model.

A. User Rank

The concept of UserRank bears the analogy to the PageRank concept when thinking of each user in the social network as the HTML page in the Web graph. The difference is that in the case of UserRank, the nodes are considered to be individual users within an online social network (e.g. Users in eBay, Facebook, Friendster, etc.) and an edge is any form of communication that one user would have with another, such as in the form of emails, transactions and such. The PageRank algorithm [7] is as follows:

$$r_{k+1}(Pi) = \sum_{P_j \in B_{pi}} \frac{r_k(P_j)}{|P_j|}$$

Where B_{pi} is the set of pages pointing into P_i and $|P_j|$ is the number of outlinks from page P [7]. r_{k+1} was added to the equation by [7] to take into account the iterative nature of the equation, with $r_{k+1}(Pi)$ representing the PageRank of Page P_i at any point within the iterations. The above equation is expanded from:

$$r(P_i) = \sum_{P_j \in B_{pi}} \frac{r(P_j)}{|P_j|}$$

This is the root equation for PageRank [7]. Following the PageRank algorithm, the one that UserRank uses adopts a similar fashion:

$$r_{k+1}(U_i) = \sum_{U_j \in B_{ui}} \frac{rk(U_j)}{|U_j|}$$

B_{ui} , in this case represents the number of sellers that sell items to the buyer $|U_i|$ and similarly, $|U_j|$ represents the number of links/votes said user has to other users within the network. Following the PageRank concept, the idea of UserRank assumes that at the start of rank derivation, all users would have the same ranking, which can be presented as: $1/n$ where n is the total number of users in the social network. The diagram below illustrates a possible network structure and, according to the User Rank algorithm, the results that would follow:

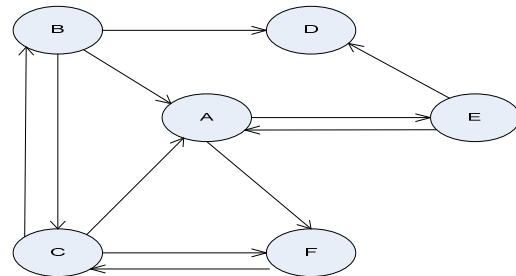


Figure 4: User Rank

Iteration 0	Iteration 1	Ranking
$r(U_A) = 1/6$	$r(U_A) = 7/36$	2
$r(U_B) = 1/6$	$r(U_B) = 1/18$	5
$r(U_C) = 1/6$	$r(U_C) = 2/9$	1
$r(U_D) = 1/6$	$r(U_D) = 5/36$	4
$r(U_E) = 1/6$	$r(U_E) = 1/12$	3
$r(U_F) = 1/6$	$r(U_F) = 5/36$	4

Table 2 : UserRank Results

B. Buyer Rank

Since in online auctions, there are many bids on a single item, the winner of the auction and the bids before should be taken into account. In this section, we present an addition to the User Rank algorithm that takes into account the edge values (in the case of auctions, the bids), such that the equation would become:

$$r_{k+1}(U_i) = \sum_{U_j \in \text{Bui}} \left(\frac{rk(U_j)}{|U_j|} \cdot w_j \right)$$

Where w represents the normalized edge value derived from the following formula:

$$w = \frac{T}{\sum V}$$

Where T represents the winning bid price and V represents the individual bids price that was placed on the item. The reason of introducing a parameter w is to accurately capture the purchase capability or the degree of “willingness to buy” of the winning bidder. The absolute value of each bidding price cannot fairly represent this capability. Therefore, we normalise the bidding price by considering other bidders’ offer. The fundamental rationale is to differentiate buyers who offer a much higher price than others from those who barely just cover the highest current bidding price. It is our belief that many potential buyers simply adopt a “last minute bidding” strategy with a minimum price increase. From a marketer’s point of view, this is not desirable. Such users are not considered to be a good potential “opinion leader” for viral marketing, and therefore their edge values should be manipulated in a way that is in favour of other buyers who are really willing to pay more premium to guarantee the bidding item.

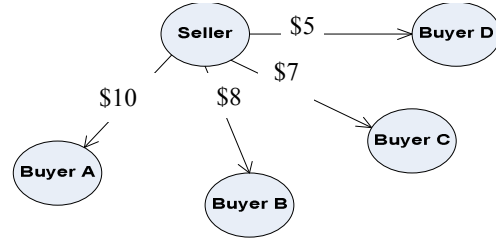


Figure 5: Buyer Rank

Figure 5 illustrates an example, in which four potential buyers bid for an item with different prices. In this case Buyer A is the winning bidder, therefore the value of w is:

$$w = \frac{10}{5 + 7 + 8 + 10} = 33.3\%$$

Therefore in addition to the User Rank algorithm, the following add on applies:

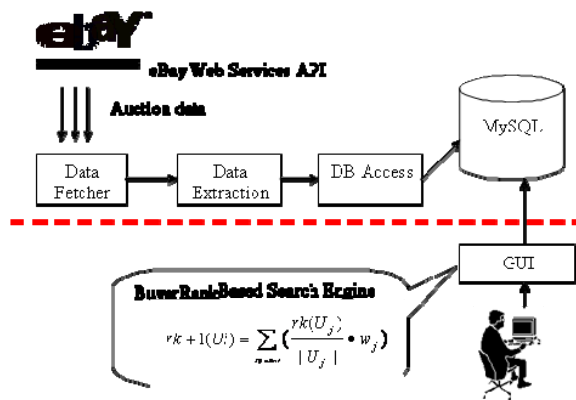
$$r_{k+1}(U_A) = \sum_{U_j \in \text{Bui}} \left(\frac{rk(U_j)}{|U_j|} \cdot 33.3\% \right)$$

The following table provides a brief summary and comparison between the PageRank, UserRank and Buyer Rank concepts:

	Page Rank	User Rank	Buyer Rank
Node	Individual Pages on the Web	Individual Users in an online social network/online auctioning network	Individual Bidders in the online auction
Edge	A hyperlink that points to the webpage or a hyperlink on the webpage that points to other pages	Any form of communication that one user would have with another user (E.g. Email, instant chatting, placing a bid within the same auction)	Transactions between a buyer and a seller
Edge Values	Value of an edge is determined by the quality of the pages that have hyperlinks to a page.	Rank of a User is affected by the network neighbours’ own UserRank score. The UserRank algorithm is based on PageRank.	Buyer rank is computed through an equation and factored into the end score of the ranking system. $w = \frac{T}{V_1 + V_2 + V_3 + \dots + V_n}$ (Where V represents each individual bid in the auction and T represents the winning bid.)

Table 3: Comparison between PageRank, UserRank and BuyerRank.

C. Implementation of BuyerRank



In order to verify the proposed BuyerRank conceptual model, we are developing the proof-of-concept software prototype. The overall architecture of this prototype is shown in Figure 4. We have developed the Data Fetcher that utilises the eBay Web Services API to retrieve auction data essential for the model such as bidding price, product information, and buyer profiles that includes product genre information. The Data Extraction module is then responsible for extract useful information from the eBay data structures and stores them into MySQL database through the DB Access module. We are currently using JUNG 1 (Java Universal Network/Graph Framework) to realise the BuyerRank algorithm in the gaming genre proposed in previous sections. We then employ the Lucene 2 Java search library to carry out the product item search for past Buyers. The implementation work is currently underway, and our future work will report the quantitative experiment result in terms of the performance of BuyerRank model.

V. CONCLUSION

Online auctions give bidders an opportunity to purchase sought after items for a price lower than most retail outlets but for the sellers, it can get slightly frustrating when waiting for their items to be searched out or chanced upon by bidders. Therefore, if there was a way for sellers to identify potential bidders of their items, it might prove to be invaluable since it would help move the auction along and even attract more bidders. Furthermore, taking into consideration the outstanding effects of viral marketing, this could not only be limited to an online auction site, since bidders would have their own social networks offline and might influence their peers into taking up their interests in an item, thereby spreading awareness and maybe the eventual adoption of a particular product or service.

This paper propose a BuyerRank, a Social Network Analysis (SNA) model that assists marketers to rank potential buyers based on their future influence estimated from their past auction/purchase behaviour on eBay. The paper also provides a detailed state-of-the-art review of recent work on SNA and viral marketing in the light of the Web 2.0 e-commerce context. We envision carry out more quantitative experiment to further

refine and validate the BuyerRank conceptual model proposed in this paper.

VI. REFERENCES

- [1] J. F. Gantz, "The Expanding Digital Universe - A Forecast of Worldwide Information Growth Through 2010," IDC 2007.
- [2] M. L. Brodie, "The End of the Computing Era: Hephaestus meets The Olympians," presented at Keynote Speech, 2nd IEEE International Conference on Digital Ecosystems and Technologies, Phitsanulok, Thailand, 2008.
- [3] Y. Beyene, M. Faloutsos, D. H. P. Chau, and C. Faloutsos, "The eBay Graph: How do online auction users interact?," *Computer Communications Workshops*, pp. 1-6, 2008.
- [4] Y. Tu, "An application of web-based data mining: selling strategies for online auctions," *Online Information Review*, vol. 32, pp. 146-162, 2007.
- [5] S. Doyle, "The role of social networks in marketing," *Database Marketing & Consumer Strategy Management*, vol. 15, pp. 60-64, 2007.
- [6] R. D. Blackwell, P. W. Miniard, and J. F. Engel, *Consumer Behaviour (9th Edition)*: Thomson Learning, 2001.
- [7] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks*, 1998.
- [8] P. Kumar and K. Zhang, "Social Networks Analysis of Online Marketplaces" *IEEE International Conference on e-Business Engineering*, 2007.
- [9] P. Singla and M. Richardson, "Yes, There is a Correlation – From social networks to personal behaviour on the web," *Proceeding of the 17th international conference on World Wide Web, Beijing, China* 2008.
- [10] K. X. R. Jin, D. C. Parkes, and P. J. Wolfe, "Analysis of Bidding Networks in eBay : Aggregate Preference Identification through Community Detection," *Proc. AAAI Workshop on Plan, Activity and Intent Recognition (PAIR) (Vancouver, Canada)*, 2007.
- [11] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E* 69:066133, 2004.
- [12] A. Mislove, M. Marcon, K. P. Gummadi, P. Drushcel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks" *In Proceedings of the 5th ACM/USENIX Internet Measurement Conference (IMC'07)*, 2007.
- [13] P. Domingos, "Mining Social Networks for Viral Marketing," *IEEE Intelligent Systems* vol. 20, pp. 80–82, 2005.
- [14] M. R. Subramani and B. Rajagopalan, "Knowledge-Sharing and Influence in Online Social Networks via Viral Marketing," *Communications of the ACM*, vol. 46, 2003.
- [15] S. Hill, F. Provost, and C. Volinsky, "Network-Based Marketing: Identifying Likely Adoptors via Consumer Networks," *Statistical Science*, vol. 21, pp. 256-276, 2006.
- [16] S. Wasserman and K. Faust, *Social Network Analysis*: Cambridge University Press, 1994.

¹ <http://jung.sourceforge.net/>

² <http://lucene.apache.org/>