



Research Article

ELITE - A Novel Ranking Algorithm for Social Networking Sites Using Generic Scoring Function

Khuan Yew Lee and Jer Lang Hong

School of Computing and IT, Taylor's University, Malaysia

Received 20 June 2012; Accepted 18 September 2012; Published 27 February 2013

Academic Editor: Maria José Angélico Gonçalves

Abstract

Social networks need to manage and control the drift of insane amount of information by filtering and ranking everything in order to ensure they are right there for users' viewing pleasure. However, the realization of social networks ranking is currently dictated by fairly straightforward optimization algorithms. Hence, there is a need for a newly enhanced and improved ranking algorithm to be formulated since users have been occasionally seeing what they should not. A composition of a generic score and a collective score that would equate to a whopping new-fangled algorithm called E.L.I.T.E. which comprises of five essential elements - Engagement-U, Lifetime, Impression, Timeframe and Engagement-O in ensuring a more accurate result for users to see more of what they care about, less of what they do not and more of who they are interested in, less of who they are not. Engagement-U is the affinity between users measured by the relationships and other related interests between them, Lifetime is a trace of users' past based on their positive, neutral and even negative interactions and actions with other users, Impression is the weight of each object determined by the number of positive responses from users, Timeframe is the timeline scoring technique in which an object naturally loses its value as time passes and Engagement-O is the attraction of users to objects measured between objects and associated interests of users.

Keywords: Search Ranking, Social Network.

Introduction

The first eleven years of the **twenty-first century** brought a host of new technological innovations. These advancements subsequently led to the creation of the Web 2.0. Content creation, interaction and collaboration with one another and information retrieval become much easier

than before. Given the freedom, it is now essential for almost everyone to search somebody, link somewhere, tag someone or share something.

Social networking sites such as Facebook, Twitter, MySpace and Google+ are by far the most popular applications in the Web 2.0 era. These social networking sites which allow

Copyright © 2013 Khuan Yew Lee and Jer Lang Hong. This is an open access article distributed under the Creative Commons Attribution License unported 3.0, which permits unrestricted use, distribution, and reproduction in any medium, provided that original work is properly cited. Contact author: Khuan Yew Lee E-mail: khuanew.lee@taylors.edu.my

How to Cite this Article: Khuan Yew Lee and Jer Lang Hong, "ELITE - A Novel Ranking Algorithm for Social Networking Sites Using Generic Scoring Function," *Journal of Internet Social Networking & Virtual Communities*, vol. 2013, Article ID 757069, 16 pages
DOI: 10.5171/2013.757069

people to interact through the exchange of multimedia objects such as text, audio and videos have established themselves as very popular sites for finding and making friends of similar interests to share various opinions and experiences between them. People have used the idea of social network to get connected at all scales, from interpersonal to international.

Each user needs to register for an account to create a profile where he or she declares his or her personal information before establishing relationships with other users on the similar network. These relationships outline a massive graph of nodes interconnecting users and their interactions. An interaction is defined by the transmission of information from one user to another user. For instance, user A posts a message or comment on user B's profile. The more content people upload to their profiles, the more time people spend viewing those profiles and the longer people stay on the network.

Social network is not merely a trend today. Some people have become so dependent on social networking sites that they must check them at least on a daily basis, if not more. Did we even stalk our exes, remember our relatives' birthdays and bug our friends before social networks? The ease and flexibility of communication made available by either these low access cost or free multimedia platforms magnified by the freedom that everyone now has in publishing their thoughts and views, has led to a very rapid propagation of enormous and dominant social networking sites.

Real examples of such are Facebook (www.facebook.com) with over 600 million registered users, Twitter (www.twitter.com) with over 175 million registered users, MySpace (www.myspace.com) with over 100 million registered users and Google+ (www.plus.google.com) with over 50 million registered users Wikipedia (2011). Social networks are the new trend on the Internet, they came to stay. In massive social networks like these, vast information resources often flood the entire network. Probing valuable information efficiently is of extensive inquisitiveness. The enormous number of the

outcomes causes the user to only emphasize on the top results.

For example, search engines would normally return thousands of results with a mixture of relevant and irrelevant information Dell (2004). However, only about 65% to 70% of the users will choose the first page, approximately 20% to 25% of them may choose the second page and just 3% to 4% of the users will check the subsequent pages Croft (1980). In short, search engines must give a good result to satisfy their users' request.

For this research, we have chosen Facebook as the center of study because it is the largest and possibly has the richest secure content and complex architecture model. Facebook was founded in 2004 and was initially available only to university students in the United States. It has then been made public and has over millions of users today. Facebook connects hundreds of millions of people and expedite sharing among them by passing and regulating all the incoming information which spontaneously decides which objects we see and which we do not.

No doubt that people believe news and information retrieved from social medias by their close friends are more reliable than those from the common source - World Wide Web Guy et al (2002), Ellison et al (2009). This explains the reason Facebook wants to ensure they are right there for users' viewing pleasure. The key success to this are the news feed and friends list which help to manage and control the drift of insane amount of information by filtering and ranking everything in order, just like Google's search results.

There are two streams of Facebook news feed - recent stories and top stories. All of the news is now in one place with the most interesting stories featured at the top. If users have not visited Facebook for a while, the first things they will see are the top photos and status updates posted while they have been away. On the other hand, if users check Facebook more frequently, they will see the most recent stories first. Both are completely different and ranking algorithm is only applied to top stories. Every item that shows

up in news feed is considered an object, a status update for instance.

Top stories are stories published displaying the most popular content from user's friends which consists of status updates, tagged photos and videos, notes, responses to events, friend requests and more since users last checked news feed that they will find interesting. They may differ depending on how long it has been since users last visited their news feed.

On the other hand, we often see friends list in the left column of users' profiles, search list at the top of any page, tag list in any status updates, comments, photos, videos, places and friend suggestions. Friends list organizes users' friends and displays them in an ordered sequence.

In this paper, we conduct a user survey to find out user needs and expectations on their viewing list in Facebook. We believe that current ranking algorithms in social networking sites need improvement and they do not seem to represent all the factors in the search space. Besides, there have not been any standard conventions to have ranking algorithms applied across all the features in Facebook. The user survey conducted will give a deeper insight into current user needs and expectations, and we hope that this survey will be a guide for future researchers in developing a better and more accurate search ranking algorithm. Using the feedbacks from user survey, we develop a composition of a generic score and a collective score that would equate to a whopping new-fangled algorithm called E.L.I.T.E. which comprises of five essential elements - Engagement-U, Lifetime, Impression, Timeframe and Engagement-O in ensuring a more accurate result for users to see more of what they care about, less of what they do not and more of who they are interested in, less of who they are not. Engagement-U is the affinity between users measured by the relationships and other related interests between them, Lifetime is a trace of users' past based on their positive, neutral and even negative interactions and actions with other users, Impression is the weight of each object determined by the number of positive responses from users, Timeframe is the timeline scoring technique

in which an object naturally loses its value as time passes and Engagement-O is the attraction of users to objects measured between objects and associated interests of users. Preliminary versions of this paper have been accepted in Khuan Yew Lee (2012), Khuan Yew Lee (2012).

Related Work

Ranking algorithms play a vital and crucial role in various traits of mechanisms such as search engines and social networks. Millions of people use these tools every day and research continues on the exploration of these algorithms to discover significant findings. Some of these algorithms merely measure the importance and prominence of objects based on their relations and contents.

However, integrated computation of these algorithms are mostly impractical because individuals are hesitant to share their interaction graphs due to privacy concerns as assured by the regulations Facebook (2010) and restrictions in the terms of service Rights (2010) of social networks. Privacy protection is an obligation as users are typically reluctant to reveal explicit information about their activity Zheleva and Getoor (2009). Preserving privacy of users is challenging as any information exchange involved in this computation should not contain any private information.

On top of that, several personalized ranking algorithms have also been suggested to further enhance results obtained by including numerous types of additional information Micarelli et al (2007). Predicting something which someone may have interest in has lately turn out to be an essential task in social networks. Many different approaches have been proposed for commendation such as content-based filtering Balabanovic and Shoham (1997), collaborative filtering Goldberg et al (1992), component centrality Ilyas and Radha (2010) and graph model Aggarwal et al (1999).

Content-based filtering Zhang et al (2002), Bissus and Pazani (2002), Mooney and Roy (2000), Pazani and Bissus (1997) recommends objects for users based on connections between preferences of users

and the content of the objects. This method creates a profile for users and objects to characterize their nature. Users will then be recommended objects similar to those the users preferred in the past. Users tend to engage with those who share common interests and they are often more concerned with information from close friends than from others.

Likewise, collaborative filtering Breese (1998), Konstan et al (1997), Linden et al (2003), Sarwar et al (2001), Getoor and Sahami (1999), Hoffman (2003), Pavlov and Pennock (2002), Ungar and Foster (1998), Cao et al (2008), Cai et al (2010) is a common approach in recommender systems to excellently predict user preferences for objects.

Conventionally, these recommender systems tend to discover user preferences for objects that match their likings by modeling the relation between them in order to assist the user in selecting objects from an overwhelming set of choices. Typically, the similarities between two users are based on their ratings of objects that both users have given. Two objects are said to be similar if they are both selected by a set of users. Alternatively, two users are similar if they both select the same set of objects. The underlying assumption of collaborative filtering approach is that those who agreed in the past tend to agree again in the future.

On the other hand, the hunch behind the new measure of component centrality is that users who are connected to well-connected users even if they are poorly connected have a more dominant standing. The fundamental justification is rooted in the postulation that in social networks, users (nodes) with more friends (connections) tend to send and receive more messages. Similarly, they will receive more messages from friends that have a lot of traffic than from those who have lesser. The information flow is modeled as an influence process as the computation distribution of component centrality amongst users does not require a central entity to access the friendship graph.

Correspondingly, data collections can be represented in the form of graphs where nodes signify entities and edges symbolize

the relationships between paired entities. Web can be seen as a very enormous graph, where nodes signify webpages and edges symbolize links between those pages Freeman (1979), Kemeny and Snell (1976), Stepheson and Zelen (2011), Wesserman and Fauss (1994). One of the most well-known algorithms for the web search is the Google's PageRank where it is defined recursively depending on the number and metric of all the pages that link to it. A hyperlink to a particular page is considered a vote of support and a page that is linked to by many other pages receives a higher rank itself. In other words, if there is no link to a webpage, there is no support for that page Page (2010). The value of PageRank reflects the idea that a page is important if there are many important pages themselves linking to it.

Likewise, prior techniques for computing top results in social network are typically centralized Goyal et al (2008), Leskovec et al (2007). However, the graph is a directed multi-graph Bollobas (1998) where the nodes characteristically represent individuals while the edges represent the relationships amongst them. Sociologists have proposed various approaches to determine the centrality of a node in a social network Kemeny and Snell (1976). The shortest path in a weighted graph offers an advantageous dimension technique in evaluating components of that graph. The dispersal of weights can hence be termed as the fixed point scheme on the graph.

However, the algorithm to pre-compute all friendship pairs and indexing them is unfeasible as the millions of users in a graph is too large to be stored even though each user has only about a hundred friends because it would still result in a very large index size of friendship distances. Moreover, pre-computing all distances between any pair of users presents no scalability as the number of friendship distances that need to be stored would be overwhelming if the number of users is multiplied ten-fold.

Likewise, an algorithm to calculate all distances on-the-go at scoring time is also impractical. A solution would then be the bi-directional breadth-first search (BFS) Russell and Norvig (2003) by looking for intersections points. The distances can then

be used to rank the results based on the earlier algorithm. Also, the combination of both the preceding algorithms would be faster and more space saving. However, it only captures the relations up to friends-of-friends.

Hence, a more promising approach is to pre-compute a friends-of-friends list for every user and intersect it with the list of friends of the user submitting the query. This would allow up to friendship distances of three to be captured. Yet, it is not ideal to limit friendship distances to three as it would lose precision figures. Also, the high time and space requirements cause it to be unfeasible for huge social networks like Facebook.

As proposed by various researchers Dor et al (2000), Rattigan et al (2006), Thorup and Zwick (2001), Zwick (2001), a system of network landmarks comprises of pre-determining set of seed nodes that serves as navigational beacons in the friendship graph can be used to approximate shortest paths. Beginning from each seed, a breadth-first search (BFS) can be used to reach out to all the nodes in the network. For each node reached, the distance from the start of the search to the seed can be marked. With that, a vector of distances to seeds can be associated with each node of the graph.

On the contrary, keyword-only query languages employed by search engines are too basic to express users' information need accurately. Indeed, users need precise results list that can help them to find what they want as fast as possible. A few words maybe obviously not be detailed enough to judge the real requirements of users. Hence, social annotation system brings a whole new way for improving the effectiveness of information retrieval by suggesting public interests. These technology expanding queries mainly exploit the directivity of interest of users based on social annotations Qinghai (2010) whereby the system may expand its queries based on social annotations automatically when users submit a query.

Generally, there are three different sorts of queries – navigational, informational and transactional. Navigational queries are used

to find specific pages, informational queries are used to find static pages and transactional queries are used to find interactive pages Andrei (2002). In order to classify query types inevitably, a click-through based ranking method can be used Yiqun et al (2006). It has been demonstrated that we can use different algorithms to enhance search performance if we can identify different types of queries. Top results are often significant and they need to be managed explicitly, particularly navigational queries Eugene and Zijian (2006).

Considering that search engine users will only usually see the top 30 results, a searching method which examines search logs, abstracts navigational queries and click-through information for queries to mark the most relevant outcome in finding the most clicked result from historical data has been proposed Yiqun (2007). The benefit of this method is extracting relevant information from users' historical data which meets users need better, whereas the shortcoming is that it only meets the navigational queries (30%) while most of the users' queries are informational (48%) Andrei (2002).

Subsequently, a model of users' web search behavior to host richer information of users' reading and interaction has been proposed so as to return better results Eugene et al (2006). This method is designed to build mapping between users' search behavior and the selected web pages. With that, real query intent of users can be understood easily to mark relevant results automatically by analyzing the large number of users' queries to get higher precision in meeting their information needs.

Above and beyond, there are various techniques in evaluating the significance of the nodes in a network as derived from graph theory and graph-based data mining Washio and Motoda (2003). A lot has been researched on social networks Scott (2002), concentrating on amplifying the differences amongst the nodes to distinguish the prominence and related method such as between-ness ranking, degree ranking, closeness ranking Freeman (2001) and so on.

Based on the definitive theory of fields Landau (2001) on many physical phenomena, a field is a state of the interaction of particles. In other words, particles create a field around itself and a certain force then acts on every other particle in the same field. This constant field independent of the time variable is an essential physical field which can be described by a scalar potential function or a vector field intensity function convertible into one another. Since the computation of scalar functions is more straightforward and down-to-earth, the characteristics of constant fields are commonly termed with the support of scalar potential, which is the function of position coordinates and magnitudes characterizing the field.

Stirred by the notion, the theory of fields is introduced into the network topological structure to define the relationship amongst the nodes being linked by edges and to disclose the overall characteristic of the core importance distribution. Hence, this method offers an overall framework for some typical ranking processes, and by augmenting influence factor, it can also make known the position differences of network structure. This topological ranking algorithm exploits the data field theory to define the interaction of all nodes in the network by describing and computing topological potential score of each node to evaluate their importance.

With that, a more precise universal ranking which can mirror nodes importance in the network can be acquired. When each node only affects its neighbors, the topological ranking is regular with the degree ranking. In other words, as the influence of the node spreads and when the influence extends to the diameter of the network, the topological ranking is adjacent to the closeness ranking.

Problem Statement

A top story is determined based on lots of factors, including user's relationship with the person who posted the story, how many comments and likes it got, what type of story it is and so on. For example, a friend's status update that might not normally be a top story may become a top story after many other friends comment on it.

$$\sum_e^{\infty} (u_e w_e d_e)$$

u_e – affinity score between viewing user and edge creator

w_e – weight of this edge type

d_e – time decay factor based on how long ago the edge was created

Generally, the realization of Facebook top stories is currently dictated by a fairly straightforward optimization algorithm called EdgeRank, which consists of three main components – affinity score, weight and time decay.

Affinity score is the points between the viewing user and object's creator. If you often like your friend's objects or send them messages, then you will have a higher affinity score for that friend of yours as compared to one you have not spoken to for years. The more interaction Facebook sees between you and that user, the higher score they give to your relationship.

Secondly, there is a weight given to each type of object. For instance, a comment would probably have more importance than just a like. Facebook is looking at the strength of those interactions you had with that user.

And finally, time decay - the most important factor of the formula. It disregards the relevance of objects to users. The longer it has been up, the less appealing it is. In other words, the older an object, the less important it becomes.

On the other hand, the ordered sequence of friends list is generally determined based on friends who users view and interact with the most in wall posts, comments and mutually attended events. By default, a changing selection of all users' Facebook friends always appears under the friends heading in the left column of their profile. However, they are not selected based on whose profiles users choose to view or who they interact with over messages and chat.

Alternatively, when we type something into the search bar at the top of page or tag our

friends using the '@' symbol, the most relevant results start populating in a dropdown menu even before we complete our search. These results are determined even based on whose profiles users choose to view or who they interact with over messages and chat.

Likewise, friend suggestions – “People You May Know” helps us find people we are likely to know based on mutual friends, work and education information, networks we are part of, contacts we have imported using friend finder and many other factors.

In a nutshell, we should see more of what we care about and less of what we do not and more of who we are interested in and less of who we are not. Therefore, the motivation of this research is that we have been occasionally seeing what we should not since the current Facebook EdgeRank algorithm is of inadequate aspects. Hence, the goal of this research is to formulate a newly enhanced and improved ranking algorithm in ensuring a more accurate result for users.

A User Study on Ranking Algorithm in Facebook

Objective

The realization of Facebook top stories is currently dictated by a fairly straightforward optimization algorithm called EdgeRank, which only consists of three main components – affinity score, weight and time decay.

The justification of this research is that users have been occasionally seeing what they should not. Hence, the main objective of this research is to formulate an enhanced and improved ranking algorithm in ensuring a more accurate top stories and a well-ordered friends list for user’s viewing pleasure. Users should be able to see more of what they care about, less of what they do not and more of who they are interested in, less of who they are not.

Methodology

The methodology of this research would be an online survey deliberately designed to

discover from a cross-section of social networking sites users on their usage and behavior on those online platform services. The online survey (https://docs.google.com/spreadsheets/viewform?hl=en_GB&formkey=dDIQZHFxX0JQbWVEWmZ6dzk2RW9pbWc6MQ#gid=0) is conducted on 334 local and international users comprising of any age, gender and background classes to gather comprehensive quantitative results.

Data Findings and Results

Question 1 – What is your gender?

The initial impression from Question 1 is that there is a disproportion of sexual category in this research survey, where the survey respondents are of 56% males and only 44% females, which brings a total of 334 individuals. However, the result of this gender inequity will not have any negative impact on the study of this research.

Question 2 – What is your age?

Question 2 gives the age fractions of the 334 survey respondents. The initial impression from Question 2 is that majority of the individuals belong to the age group of 21 – 30 years old, whereas 33% are 13 – 20 years old, 3% are 31 – 40 years old and the remaining 1% are above 40 years old. Apparently none of the survey respondents are below 13 years old since individuals must meet the minimum age required to be eligible to sign up for Facebook, which is 13 years of age or older. On the contrary, it is evidently shown that individuals of the age group 21 – 30 are the main crowds on Facebook today. Correspondingly, high school and college students who are 13 – 20 years old are the succeeding largest age group on Facebook.

Question 3 – Education level

Question 3 specifies the education levels of all the survey respondents. Undoubtedly, most of the individuals who responded to the research survey hold a bachelor degree qualification. Similarly, individuals of the age group of 13 – 20 years old who are also pre-university students constitute 35% of the survey respondents, followed by 26% from

diploma and 21% from high schools. Only 6% of the 334 survey respondents are of professional, master degree and doctorate (PhD) education qualification, where 8 of them are of professional qualification, 10 of master degree and only 3 of doctorate (PhD) qualification.

Question 4 – When did you join Facebook?

Facebook was launched in the year 2004 and from the results of Question 4, it is noticeable that still not many people were connected through Facebook during that time, even on the following year in 2005 as only 9% of the total survey respondents are socially connected during that first two years of debut. However, the number of survey respondents on Facebook significantly increased from the year 2006 until 2008, where another 62% of them joined the mass. Unfortunately, the number of respondents joining Facebook decreased drastically since the year 2009.

Question 5 – How often do you go online?

Question 5 specifies the regularities of the survey respondents going online. The initial impression from Question 5 is that 89% of the survey respondents would go online as frequent as few times in a day whereas only 10% of them would go online few times in a week while just 1% of the survey respondents would go online as infrequent as few times in a month or even few times in a year. The results clearly ratifies that most of the people in this information era today would go online at least once a day, if not more.

Question 6 – Do you check Facebook every time you online?

Question 6 gives the tendencies of the survey respondents checking Facebook each time they go online. The initial impression from the chart above is that 90% of the survey respondents would check Facebook every time they go online whereas only 10% of them would not check Facebook every time they go online. The results undoubtedly ascertains that most of the people in this technology generation today would check Facebook every time they go online.

Question 7 – How long do you usually spend on Facebook each time?

The results from Question 7 give the amount of time spent by the survey respondents each time they check Facebook. A majority of 38% of the survey respondents would spend less than an hour whereas 37% of them would spend a few hours on Facebook each time. On the other hand, 15% of the survey respondents would spend all day long while the remaining 11% would spend half a day on Facebook each time. From the results, it is obvious that individuals would not want to spend too much time on Facebook despite them checking it every time they go online.

Question 8 – What do you usually do on Facebook?

The results from Question 8 give the common activities done by the survey respondents on Facebook. The most common activity by the 334 survey respondents is checking the news feed, followed by liking or leaving comments on statuses, photos, videos or notes, chatting or sending messages, updating statuses or sharing links, viewing other's profiles, uploading photos or videos, playing games and asking questions. From the results, it is clearly shown that it is crucial to have an accurate top stories and a well-ordered friends list on Facebook.

Question 9 – What do you think is the accuracy level of your top stories in news feed?

Question 9 specifies the accuracy level of top stories in news feed by the survey respondents. The initial impression from the chart is that 25% of the survey respondents think that the top stories in news feed is only moderately accurate, with a rating of five (5) from a scale of one (1) to ten (10). Hence, it is undeniable that people are not actually satisfied or pleased with the current top stories in news feed.

Question 10 – How well do you think your friend list is ordered?

The results from Question 10 indicate the orderliness level of the friend list by the survey respondents. The initial impression

from the chart is that 31% of the survey respondents think that the friend list is only reasonably accurate, with a rating of five (5) from a scale of one (1) to ten (10). Hence, it is noticeable that people are not actually satisfied or pleased with the current order of the friend list they have.

Question 11 – Do you wish to have more accurate top stories and a well-ordered friend list?

The results from Question 11 show the responses of all 334 respondents in having a more accurate top stories and a well-ordered friend list. It is rather obvious from the chart above that a majority of 92% of the survey respondents wishes to have more accurate top stories and well-ordered friend lists since most of them are not actually satisfied or pleased with the current top stories in news feed and friend list they have.

Question 12 – Given a choice, what do you want your top stories and friend list to be most affected by?

The results from Question 12 give the factors of which the survey respondents want their top stories and friend list to be most affected by. At a glance, the survey respondents want their top stories and friend list to be affected by those they viewed and interacted most followed by those they recently viewed and interacted, number of friends in common,

most popular stories or friends, relationships between users and interests in common. This evidently ascertains that people want to see more of what they care about, less of what they do not and more of who they are interested in, less of who they are not.

Elite Ranking Algorithm Proposed Solutions

Living in our information age, most of us have social networking sites on our screens throughout the day, escalating the prospect of discovery. Most of us are not enthusiastically looking for content, but when a friend posts something, we often click to check it out even if we were doing something else. When thinking of content prioritization, we usually check out suggestions, view photos and videos and click links from our friends before anything else.

Although the sign-up rate of new users on some social networks may have stalled, the user engagement continues to develop intensely. More users are compelled to respond to content which they have seen on social networks knowing that the engagement rate has increased beyond the average reach of objects which clearly shows that social networks are becoming an even more crucial space for main interaction and communication between users. On the whole, using social networks to engage with others is about purpose and posting quality content and creating effective social voice in communicating with others are both positive ways to engage.

Generic Score + Collective Score = Composite Score

It would be helpful if there were a judgment, assessment or evaluation ranking model which could be used irrespective of the set of circumstances to ensure a more accurate result for users in a social network. A model that is flexible enough to take any local parameters for assessment into account. A model that is easy to be applied despite the properties defined. A model that is completely transparent, compelling and does not counteract accordingly.

Firstly, a computable global numeric point scheme is to be defined where the points can be attributed to a variety of scoring metrics empirically. This universal scoring used for general control is user independent where the points are allocated in fixed expanse. In other words, the submissions from every user are scored on the same scale. It is one that can be applied to a wide range of different conditions as an objective way of measuring and recording the complex social network state in order to compare users and objects.

This generic score is centered on three factors – Engagement-U, Impression and Engagement-O. Engagement-U is the affinity between users measured by the relationships and other related interests between users, whereas Impression is the weight of each object determined by the number of positive responses from users while Engagement-O is the attraction of users to objects measured between objects and associated interests of users.

On the other hand, an accumulated numeric point scheme is to be derived where the points can be accredited to all users explicitly. This historic score formulated is user reliant where the points are stored constantly and endlessly. It is one that can retain the data of users throughout experiencing social space as a subjective way of determining and evaluating the complex social network state between users and objects.

This collective score is focused on two factors – Lifetime and Timeframe. Lifetime is a trace of users' past based on their positive, neutral and even negative interactions and actions with other users whereas Timeframe is the timeline scoring technique in which an object naturally loses its value as time passes. The longer an object has been up, the less appealing it is no matter how relevant it may be to a user. In short, the older an object, the less important it becomes.

To sum up, the composition of the generic score and the collective score would equate to a whopping new-fangled algorithm called E.L.I.T.E. which comprises of five essential elements as mentioned earlier - Engagement-U, Lifetime, Impression, Timeframe and Engagement-O.

ELITE Ranking Algorithm

Engagement-U

In massive social networks like Facebook, the most common action is to look for old friends and make new ones. User interactions in a social network reflect that of their real life. Users tend to interact more with those they really know in person. In other words, they

will most likely want to see people who they are closer to in the social space. For instance, a user A will most likely be pleased if he or she sees a direct friend of him or her rather than someone who he or she does not know and who is not even known by his or her friends.

The high commitment and user engagement statistics of social networks are tied to human longing to sustain and uphold relations with others. Although social networks have been noble for establishing user connections with many different people, many realize just how central social networks have been to the formation of deep and long-lasting friendships. In social networks with a very large number of users where user preference is important, personalized recommendation of people becomes essential. Social relations between users are very useful for personalized recommendation.

Traditional recommender systems attempt to discover user preferences over friends by modeling the relations between them. The aim is to recommend friends that match the fondness (likes or dislikes) of users in order to assist them in ranking from an overwhelming set of friends. Behavior of users often reflects that of others who have similar interests or related information profiles in social networks. Therefore, if interest of users in certain areas can be traced and their preferences in terms of activities can be tracked, significant and reliable information about these people can be pinpointed in a formulated method.

The obvious data of interest is the personal data uploaded onto user profiles. Much of this data is tagged by the user with metadata, making it easy to store and analyze. Therefore, Engagement-U solely measures relations, friends in common, similar education and work, philosophy, arts and entertainment, sports, activities, interests and locations between users. The affinity score for each user is the level of similarity shown by users and this will give a clearer picture of the relationship between them.

Lifetime

When we post or share objects such as status updates, photos, videos and links, we often receive responses from our friends, either a like, a comment or they probably share them. All is magnificent. Our friends love us, but what about the actual click-throughs? The fact about all these objects is that only a handful of people will actually respond to them. Hence, we need to find out the number of users who actually views each object by clicking on them.

Although click-through data will not necessarily indicate what the person did with the object or even if they actually read or watch it, it gives us a much clearer picture of the real impact of every individual object as it is more meaningful. The gap between viewing an object by clicking on it and responding to an object may be filled in when more accurate content or data is introduced.

Click-through data will add more accurate content to social networking objects, thus improve the relevance measurement. Clearly, users do not click on objects random, but make a somewhat informed choice. While click-through data is typically noisy and clicks are not perfect relevance judgments, the clicks are still likely to convey some information.

Click-through data is the number of times an object is viewed, anywhere on Facebook. As a result, click-throughs are usually higher than the number of impressions for each object. For instance, one object may be seen once per person, whereas another may be seen five times per person. User click-through data can be extracted from a large amount of logs accumulated. Although these clicks do not reflect the exact relevancy, they provide valuable indications to the users' intention by associating a set of objects. If a user clicks on an object, it is likely that the object is somewhat interesting and attractive or relevant, or at least related to some extent.

Hence, Lifetime takes into account the click-through data of each object, which is the total number of times an object is actually viewed despite any neutral or negative responses by other users such as like, comment or share. This information will provide a better picture of the feedback or reaction for every object.

Impression

Facebook is neither a website nor a newspaper advertisement. It is all about interaction. From a business perspective, Facebook is less like a sales meeting with new potential clients but more like a networking session to have the opportunity to meet people and build relationships. Hence, very few people are going to like it, comment or share if an object is sales-oriented.

Still, social buttons are undeniably the powerful tools behind the whole lot. The king of Facebook engagement remains the like button which accounts for a massive 84% of all Facebook user reactions. Of the rest, status updates and commenting in response to objects represents 15% with sharing merely 1% of all user engagements.

This thumb up reaction to content has become impulsive. In fact, the attractiveness of the like button definitely does not come as a surprise as it has eventually become a nature for most, involving very little of effort and thought. Liking an object is the top user engagement drawer on the social network as it is spontaneous and instant. The like button is the catalyst that paves the way for more real-time engagement to happen.

On the other hand, commenting evidently involves some thought, understanding and engagement while sharing requires a bit more effort, a little more thought and even more time. Engagement is up for sure but the overall quality of engagement has not exactly as meaningful as it could be as it independently looks at one aspect of potential reach measurement.

Usually, objects of 80 characters or less generate the most user engagement. Likewise, people are three times more likely to click on a tiny URL as opposed to a full-length URL. To be effective at social networks, ask a simple question at the end of each object to generate more user engagement. However, putting a question at the beginning of the object or leave it buried somewhere in the middle are far less likely to be seen.

When you ask, give a tip, or share a joke to your friends, not only it is likely to be shared, mentioned, or liked, but users will give feedbacks and that is instant user-generated content. When users receive a communication from another user while interacting and communicating on social networks, they can make different responses, either positive or negative. Some will probably notice yet ignore it sometimes, hence a negative response.

Therefore, Impression exclusively measures the number of likes, comments, tags and shares an object receives. It does not measure clicks, video plays or similar. The number of impressions for an object is the raw number of impressions shown by users. While impression percentage as high as 0.50 have been seen, the average impression percentage is commonly somewhere around 0.10. This will again give a much better picture of the actual traffic on each object.

Timeframe

In the world of social media, real time is not fast enough. To ensure only the freshest and most current objects appear, old objects drop out of the news feed so newer objects are more likely to appear. One of the most important characteristics of objects is that they have a definite lifespan. All objects decay over time. They cannot be held for 10 days. In fact, the time decay of objects increases exponentially. The value of an object decreases as its expiration date approaches. The less time left on the object, the greater the effects of time decay.

Like Google, Facebook is concerned with the freshness of content. Newer objects and interactions hold more importance and have a higher likelihood of being published in news feed. The older the interactions are with objects, the lower the score and ability to push content to the top of people's news feeds.

Facebook is highly dependent upon the temporal nature of objects. Therefore, to maximize the effectiveness of any object on Facebook, they should be created at the points of time when audiences are most likely

to be using Facebook in order to decrease the time decay and increase the chances of the content reaching news feeds of friends.

Hence, Timeframe is expressed in terms of the score that an object value will lose on a daily basis with reference to how recent an object is by measuring the rate of decline in an object value due to the passage of time. It just has to do with how old an object is. This is almost self-explanatory as it involves the relevancy of objects. With everyone wanting fresh content, rankings are highest with the quickest posts and interactions. The freshness will influence what news you get to see. Fresh objects are worth more than older objects. That is why time sensitive updates are so effective because they involve objects about something and an upcoming or expiration date.

The importance of this factor is that we really have to know when our friends are on Facebook and not at other times. In other words, we may have to make our updates at night or on Sunday mornings or on rainy Saturdays. Just because we may be online during the workday does not mean that is the best time to post content to Facebook.

Engagement-O

Advancement in digital data acquisition and storage technology has led to in the growth of huge databases and it has occurred in all areas of human endeavor, from the mundane (such as supermarket transaction data, credit card usage records, telephone call details, and government statistics) to the more exotic (such as images of astronomical bodies, molecular databases, and medical records). Interest has grown in the possibility of tapping these data, of extracting from them information that might be of value to the owner of the database.

Similarly, social networks hold enormous quantity of data and they are very valuable because users interact and communicate naturally with one other in them. There are also priceless views and thoughts from users about anything and everything on various topics. The conventional technique of revolving information into knowledge relies on manual analysis and interpretation.

Across a wide variety of fields, data are being collected and accumulated at a dramatic pace.

There is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information or knowledge from the rapidly growing volumes of digital data. The art of mining useful information from large data sets or databases is known as data extraction. It is a new discipline lying at the intersection of diverse areas apprehensive with certain traits of data analysis, having much in common yet distinct flavors, emphasizing particular problems and types of solution.

Data extraction is the scrutiny of often enormous information sets to discover hidden connections and to encapsulate the data in unique ways that are both logical and valuable to the data owner. Retrieval by content is a major difficulty for the consideration of large databases, specifically for data types such as images where the algorithms for retrieval by content have boundless possible utility across a variety of applications. The objective of data extraction is to discern and ascertain new information from data, defining patterns across datasets and or separating signal from noise.

However, the relationships and structures found within a set of data must of course, be novel. While novelty is the key property of the relationships we pursue, it is inadequate to label a relationship as being worth finding. In addition, the relationships must also be explicable although direct relationships are more readily understood and preferred than convoluted ones.

Therefore, Engagement-O uniquely compares the profiles of users such as education and work, philosophy, arts and entertainment, sports, activities, interests and locations to data extracted from every object. The resulting score for each object will give a clearer picture of the correlation between users and objects.

Conclusion

Once upon a time, there were webpages. Now, we have close to a thousand diverse social networks and more than hundreds of

million social networkers. These numbers continue to escalate whether it is Facebook, Twitter, MySpace or Google+ and it does not really matter since people are always connected. More connectivity creates more opportunities to be involved in a dissimilar type of social networking experience, one that is much more engaged and interactive. However, current search ranking algorithm in social networking sites lack uniformity in its design, and they do not consider for other factors in search ranking. Our user survey shows that most users would like to see more of what they care about, less of what they do not and more of who they are interested in, less of who they are not. We believe this important observation is useful for future researchers to develop a more accurate and efficient search ranking algorithm. In order to improve on current ranking algorithms, we develop a composition of a generic score and a collective score that would equate to a whopping new-fangled algorithm called E.L.I.T.E. which comprises of five essential elements - Engagement-U, Lifetime, Impression, Timeframe and Engagement-O in ensuring a more accurate result for users to see more of what they care about, less of what they do not and more of who they are interested in, less of who they are not. Engagement-U is the affinity between users measured by the relationships and other related interests between them, Lifetime is a trace of users' past based on their positive, neutral and even negative interactions and actions with other users, Impression is the weight of each object determined by the number of positive responses from users, Timeframe is the timeline scoring technique in which an object naturally loses its value as time passes and Engagement-O is the attraction of users to objects measured between objects and associated interests of users.

References

- Aggarwal, C. C., Wolf, J. L., Wu, K. L. & Yu, P. S. (1999). "Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering," In *Proc. of KDD'99*. pp. 201-212.
- Agichtein, E., Brill, E., Dumais, S. & Rago, R. (2006). "Learning User Interaction Models for Predicting Web Search Result Preferences,"

- [C] Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle: ACM, 3-10.
- Agichtein, E. & Zheng, Z. (2006). "Identifying "Best Bet" Web Search Results by Mining Past User Behavior," In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (SIGKDD)*.
- Balabanovic, M. & Shoham, Y. (1997). "Content-Based Collaborative Recommendation," *Commun. ACM*, Vol. 40(3).
- Billsus, D. & Pazzani, M. J. (2000). "User Modeling for Adaptive News Access," *User Modeling and User-Adapted Interaction*, vol. 10, pp. 147-180.
- Breese, J. S., Heckerman, D. & Kadie, C. (1998). "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," in Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43-52
- Broder, A (2002). "A Taxonomy of Web Search," In *SIGIR Forum*.
- Bollobas, B. (1998). *Modern Graph Theory*, Springer Verlag.
- Cai, X., Bain, M., Krzywicki, A., Wobcke, W., Kim, Y. S., Compton, P. & Mahidadia, A. (2010). "Collaborative Filtering for People to People Recommendation in Social Networks," in *AI 2010: Advances in Artificial Intelligence*, J. Li, Ed. Berlin: Springer-Verlag.
- Cao, B., Sun, J.- T., Wu, J., Yang, Q. & Chen, Z. (2008). "Learning Bidirectional Similarity for Collaborative Filtering," in *Machine Learning and Knowledge Discovery in Databases*, W. Daelemans, B. Goethals, and K. Morik, Eds. Berlin: Springer-Verlag, pp. 178-194.
- Croft, W. B. (1980). "A Model of Cluster Searching Based on Classification," *Information Systems*, Vol.5: 189-195.
- Dell, Z. (2004). 'Semantic, Hierarchical, Online Clustering of Web Search Results [C],' Hangzhou China: Proceedings of the 6th Asia Pacific Web Conference, 69-78.
- Dor, D., Halperin, S. & Zwick, U. (2000). "All-Pairs Almost Shortest Paths," *SIAM Journal on Computing*, 29(5): 1740-1759.
- Ellison, N. B., SteinField, C. & Lampe, C. (2009). 'The Benefit of Facebook Friends: Social Capital and College Student's Use of Online Social Network Sites,' *Journal of Computer Mediated Communication*, August, 29, pp. 1143-1168.
- Facebook Advertising, <http://www.facebook.com/advertising/>, 2010.
- Facebook Help Center, <http://www.facebook.com/help>, 2011.
- Freeman, L. C. (1979). "Centrality in Social Networks: I. Conceptual Clarification [J]," *Social Networks*, 1: 215-239.
- Getoor, L. & Sahami, M. (1999). "Using Probabilistic Relational Models for Collaborative Filtering," in *Working Notes of the KDD-99 Workshop on Web Usage Analysis and User Profiling*.
- Goldberg, D., Nichols, D., Oki, B. M. & Terry, D. (1992). "Using Collaborative Filtering to Weave an Information Tapestry," *Commun. ACM*, Vol. 35(12):61-70.
- Goyal, A., Bonchi, F. & Lakshmanan, L. V. S. (2008). "Discovering Leaders from Community Actions," In *Proceeding of the 17th ACM Conference on Information and Knowledge Management*, pages 499-508. *ACM*.
- Guy, I., Jacovi, M., Shahar, E., Meshulam, N., Soroka, V. & Farrell, S. (2002). 'Harvesting with SONAR: The Value of Aggregating Social Networking Information,' in *Proceedings of CHI, ACM*, pp. 1017-1026.
- Hofmann, T. (2003). "Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 259-266.
- Ilyas, M. U. & Radha, H. (2010). "A KLT-Inspired Node Centrality for Identifying

- Influential Neighborhoods in Graphs," In Conference on Information Sciences and Systems, Princeton, NJ, *Princeton University*.
- Kemeny, J. G. & Snell, J. L. (1976). *Finite Markov Chains, Springer Verlag*.
- Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R. & Riedl, J. (1997). "GroupLens: Applying Collaborative Filtering to Usenet News," *Communications of the ACM*, vol. 40, no. 3, pp. 77-87.
- Landau, L. D. & Lifshitz, E. M. (2001). 'The Classical Theory of Fields,' *Beijing World Publishing Ltd*.
- Lee, K. W. & Hong, J. L. (2012). "A User Survey on Search Ranking Algorithm for Social Networking Sites," 9th IEEE International Conference on Fuzzy Systems and Knowledge Discovery, FSKD.
- Lee, K. W. & Hong, J. L. (2012). "ELITE - A Novel Ranking Algorithm for Social Networking Sites," 9th IEEE International Conference on Fuzzy Systems and Knowledge Discovery, FSKD.
- Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J. & Glance, N. (2007). "Cost-Effective Outbreak Detection in Networks," In *Proceedings of KDD*.
- Linden, G., Smith, B. & York, J. (2003). "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Transaction on Internet Computing*, vol. 7, no. 1, pp. 76-80, 2003.
- Liu, Y., Fu, Y., Zhang, M., Ma, S. & Ru, L. (2007). "Automatic Search Engine Performance Evaluation with Click-through Data Analysis," In *Proceedings of WWW*.
- Liu, Y., Zhang, M., Ru, L. & Ma, S. (2006). "Automatic Query Type Identification Based on Click through Information," *Asia Information Retrieval Symposium (AIRS)*.
- Meng, Q. (2010). 'Research on Personalized Query Expanding Technology Based-on Social Annotation,' *Harbin Engineering University*.
- Micarelli, A., Gasparetti, F., Sciarrone, F. & Gauch, S. (2007). "Personalized Search on the World Wide Web," In *Lecture Notes in Computer Science*, 4321: 195-230.
- Mooney, R. J. & Roy, L. (2000). "Content-Based Book Recommending Using Learning for Text Categorization," in Proceedings of the 5th ACM Conference on Digital Libraries, pp. 195-204.
- Page, L., Brin, S., Motwani, R., Winograd, T. (2010). 'The Page Rank Citation Ranking: Bringing Order to the Web,' *Stanford University, Computer Science Department Technical Report*.
- Pavlov, D. Y. & Pennock, D. M. (2002). "A Maximum Entropy Approach to Collaborative Filtering in Dynamic, Sparse, High-Dimensional Domains," in *Neural Information Processing Systems*, pp. 1441-1448.
- Pazzani, M. & Billsus, D. (1997). "Learning and Revising User Profiles: The Identification of Interesting Web Sites," *Machine Learning*, vol. 27, pp. 313-331.
- Rattigan, M. J., Maier, M. & Jensen, D. (2006). "Using Structure Indices for Efficient Approximation of Network Properties," In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 357-366, New York, NY, USA.
- Russell, S. & Norvig, P. (2003). "Artificial Intelligence: A Modern Approach," *Prentice-Hall, Englewood Cliffs, NJ, 2nd Edition*, 2003.
- Sarwar, B., Karypis, G., Konstan, J. & Reidl, J. (2001). "Item-Based Collaborative Filtering Recommendation Algorithms," in Proceedings of the 10th International Conference on World Wide Web, pp. 285-295.
- Scott, J. (2002). 'Social network analysis: A Handbook (2nd Ed.),' [M]. London: *Sage Publications*.
- Statement of Rights and Responsibilities, <http://www.facebook.com/terms.php>, 2010.

Stepheson, K. & Zelen, M. (2011). "Rethinking Centrality: Methods and Examples," *Social Networks*, 11, pp. 1-37.

Thorup, M. & Zwick, U. (2001). "Approximate Distance Oracles," In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, pages 183-192, New York, NY, USA.

Ungar, L. H. & Foster, D. P. (1998). "Clustering Methods for Collaborative Filtering," in *Proceedings of the AAAI-98 Workshop on Recommender Systems*, pp. 112-125.

Washio, T. & Motoda, H. (2003). "State of the Art of Graph-Based Data Mining [J]," *SIGKDD Explor. Newsl*, 5(1):59-68.

Wasserman, S. & Faust, K. (1994). "Social Network Analysis: Methods and Applications," Cambridge: *Cambridge University Press*.

Wikipedia List of Social Networking Websites,
http://en.wikipedia.org/wiki/List_of_social_networking_websites, 2011.

Zhang, Y., Callan, J. & Minka, T. (2002). "Novelty and Redundancy Detection in Adaptive Filtering," in *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 81-88.

Zheleva, E. & Getoor, L. (2009). "To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles," In *18th International World Wide Web Conference (WWW)*.

Zwick, U. (2001). "Exact and Approximate Distances in Graphs - A Survey," In *Proceedings of the 9th Annual European Symposium on Algorithms*, pages 33-48, London, UK.