ASSESSMENT OF SOCIAL THE NETWORK SERVICES RANKING THAT IS RELIABLE AND VITAL

[1]Dr. K.Venkata Ramana, [2]Narendra Kumar Kotha, [3]P. Mounika, [4]Bezawada Prameela
[1] Professor, [2,3]Assistant Professor, [4]Student, Dept. of Computer Science Engineering, Newton's Institute of Engineering, Macherla, Andhra Pradesh, India.

ABSTRACT:
Social networking services have been prevalent at many online communities such as twitter.com and weibo.com, where millions of users keep interacting with each other every day, One interesting and important problem in the social networking services is to rank users based on their vitality in a timely fashion. An accurate ranking list of user vitality could benefit many parties in social network services such as the ads providers and site operators. Although it is very promising to obtain a vitality-based ranking list of users, there are many technical challenges due to the large scale and dynamics of social networking data. In this project, we propose a unique perspective to achieve this goal, which is quantifying user vitality by analyzing the dynamic interactions among users on social networks. Examples of social network include but are not limited to social networks in micro blog sites and academicals collaboration networks. If a user has many interactions with his friends within a time period and most of his friends do not have many interactions with their friends simultaneously, based on this idea, we develop quantitative measurements for user vitality and propose our first algorithm for ranking users based vitality. Also we further consider the mutual influence between users while computing the vitality measurements and propose the second ranking algorithm, which computes user vitality in an iterative way

1.INTRODUCTION

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks.

Classes: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.

Clusters: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.

Associations: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.

Sequential patterns: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

Extract, transform, and load transaction data onto the data warehouse system.

Store and manage the data in a multidimensional database system.
Provide data access to business analysts and information technology professionals.
Analyze the data by application software.
Present the data in a useful format, such as a graph or table.
Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
Genetic algorithms: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.
Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.
Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k=1). Sometimes called the k-nearest neighbor technique.
Rule induction: The extraction of useful if-then rules from data based on statistical significance.
Data visualization: The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

II. PROBLEM DEFINITION

The motivation of this project is to implement design for the functional minimum-storage regenerating (FMSR) codes. The system FMSR code implementation maintains double-fault tolerance and has the same storage cost as in traditional erasure coding schemes based on RAID-6 codes, but uses less repair traffic when recovering a single-cloud failure.The aim of our project is to minimize the cost of storage repair (due to the migration of data over the clouds) for a permanent single-cloud failure. In this work, we focus on comparing two codes: traditional RAID-6 codes and our FMSR codes with double-fault tolerance.
The main objective of our project is to minimize the cost of storage repair (due to migration of data over the clouds) for a permanent single-cloud failure.we focus on comparing: traditional RAID-6 codes and FMSR codes with double-fault tolerance.

III.LITERATURE REVIEW

The anatomy of a large- scale hyper textual web search engine

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu/. To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day. Despite the importance of large-scale search engines on the web, very little academic research has been done on them. Furthermore, due to rapid advance in technology and web proliferation, creating a web search engine today is very different from three years ago. This paper provides an in-depth description of our large-scale web search engine -- the first such detailed public description we know of to date. Apart from the problems of scaling traditional search techniques to data of this magnitude, there are new technical challenges involved with using the additional information present in hypertext to produce better search results. This paper addresses this question of how to build a practical large-scale system which can exploit the additional information present in hypertext. Also we look at the problem of how to effectively deal with uncontrolled hypertext collections where anyone can publish anything they want.

Expertise identification using email communications.

A common method for finding information in an organization is to use social networks---ask people, following referrals until someone with the right information is found. Another way is to automatically mine documents to determine who knows what. Email documents seem particularly well suited to this task of "expertise location", as people routinely communicate what they know. Moreover, because people explicitly direct email to one another, social networks are likely to be contained in the patterns of communication. Can these patterns be used to discover experts on particular topics? Is this approach better than mining message content alone? To find answers to these questions, two algorithms for determining expertise from email were compared: a content-based approach that takes account only of email text, and a graph-based ranking algorithm (HITS) that takes account both of text and communication patterns. An evaluation was done using email and explicit expertise ratings from two different organizations. The rankings given by each algorithm were compared to the explicit rankings with the precision and recall measures commonly used in information retrieval, as well as the d' measure commonly used in signal-detection theory. Results show that the graph-based algorithm performs better than the content-based algorithm at identifying experts in both cases, demonstrating that the graph-based algorithm effectively extracts more information than is found in content alone.

Mining the network value of customers.

One of the major applications of data mining is in helping companies determine which potential customers to market to. If the expected profit from a customer is greater than the cost of marketing to her, the marketing action for that customer is executed. So far, work in this area has considered only the intrinsic value of the customer (i.e, the expected profit from sales to her). We propose to model also the customer's network value: the expected profit from sales to other customers she may influence to buy, the customers those may influence, and so on recursively. Instead of viewing a market as a set of independent entities, we view it as a social network and model it as a Markov random field. We show the advantages of this approach using a social network mined from a collaborative filtering database. Marketing that exploits the network value of customers – also known as viral marketing – can be extremely effective, but is still a black art. Our work can be viewed as a step towards providing a more solid foundation for it, taking advantage of the availability of large relevant databases.

Measuring user influence on twitter using modified k-shell decomposition.

We survey the several measures that exists in literature to rank influential users in Twitter network. We propose a classification of these measures according to different criteria, such as the kind of metrics, the use of Page Rank algorithm, the use of content analysis, among others. Besides the influential users, we also study measures of activity and popularity. We finish by mentioning some aspects of this topic related with computational complexity and correlation measures. Centrality is one of the most studied concepts in social network analysis. There is a huge literature regarding centrality measures, as ways to identify the most relevant users in a social network. The challenge is to find measures that can be computed efficiently, and that can be able to classify the users according to relevance criteria as close as possible to reality. We address this problem in the context of the Twitter network, an online social networking service with millions of users and an impressive flow of messages that are published and spread daily by interactions between users. Twitter has different types of users, but the greatest utility lies in finding the most influential ones. The purpose of this article is to collect and classify the different Twitter influence measures that exist so far in literature. These measures are very diverse. Some are based on simple metrics provided by the Twitter API, while others are based on complex mathematical models. Several measures are based on the Page Rank algorithm, traditionally used to rank the websites on the Internet.

Some others consider the timeline of publication, others the content of the messages, some are focused on specific topics, and others try to make predictions. We consider all these aspects, and some additional ones. Furthermore, we include measures of activity and popularity, the traditional mechanisms to correlate measures, and some important aspects of computational complexity for this particular context.

A Huber man Influence and passivity in social media

The ever-increasing amount of information flowing through Social Media forces the members of these networks to compete for attention and influence by relying on other people to spread their message. A large study of information propagation within Twitter reveals that the majority of users act as passive information consumers and do not forward the content to the network. Therefore, in order for individuals to become influential they must not only obtain attention and thus be popular, but also overcome user passivity. We propose an algorithm that determines the influence and passivity of users based on their information forwarding activity. An evaluation performed with a 2.5 million user dataset shows that our influence measure is a good predictor of URL clicks, outperforming several other measures that do not explicitly take user passivity into account. We demonstrate that high popularity does not necessarily imply high influence and vice-versa.

Online health communities (OHCs) have become a major source of support for people with health problems. This research tries to improve our understanding of social influence and to identify influential users in OHCs. The outcome can facilitate OHC management, improve community sustainability, and eventually benefit OHC users.

IV. SYSTEM ANALYSIS

Twitter user ranking algorithm was proposed to identify authoritative users who often submit useful information.Page Rank algorithm named Twitter Rank was developed to rank Twitter users based on their influence.K-shell decomposition algorithm is developed to measure the user influence in Twitter

Most of these measurements quantify the influence in an isolated way, rather than in a collective way. The proposed algorithm mainly works based on the user-tweet graph, rather than the user-user social graph. The focus of these methods is on influence, which is still different from the vitality.

Many interactions often keep going on within online social networks over time. Examples of interaction include but are not limited to the retweeting, mention, and sending message. Our goal is to rank user vitality based on all interactions in a time period. Suppose that we have a social network S that contains N users (nodes) denoted as fUjg1_j_N and L links among users denoted as fEjkg1_j;k_N, where j and k are indices. We have recorded all interactions between them within M consecutive time periods Ti(1 _ i_ M).

Our goal is to rank all users from high vitality to low vitality for a time period Tibased on all previously observed interactions. Such a vitality-based ranking list of users may provide a good guidance for the social networking service providers to understand the dynamics of systems. They may directly find the relatively most active users and make better operation and business decisions upon the findings.

The accurate results of both user vitality ranking and prediction could benefit many parties in different social networking services,a user vitality ranking list could help ads providers to better display their ads to active users and reach more audiences.

V.IMPLEMENTATION

OSN MODULE

In the first module, we develop the Online Social Networking (OSN) system module. We build up the system with the feature of Online Social Networking. Where, this module is used for new user registrations and after registrations the users can login with their authentication.

Where after the existing users can send messages to privately and publicly, options are built. Users can also share post with others. The user can able to search the other user profiles and public posts. In this module users can also accept and send friend requests.
With all the basic feature of Online Social Networking System modules is build up in the initial module, to prove and evaluate our system features.

VITALITY RANKING MODULE
The accumulated number of interactions SAiof a node j (1 _ j _ N) in time period i(1 _ i_ M) within a social network I is defined as:

$$SA_j^i = \sum_{k \in \{Nei(j)\}} \theta_{kj},$$

Calculating Average Interaction
The relative increase of interaction of a node j (1 _ j _ N) in time period i(1 _ i_ M) within a social network I is defined as:

$$IA_j^i = \frac{SA_j^i}{SA_j^{i-1}}.$$

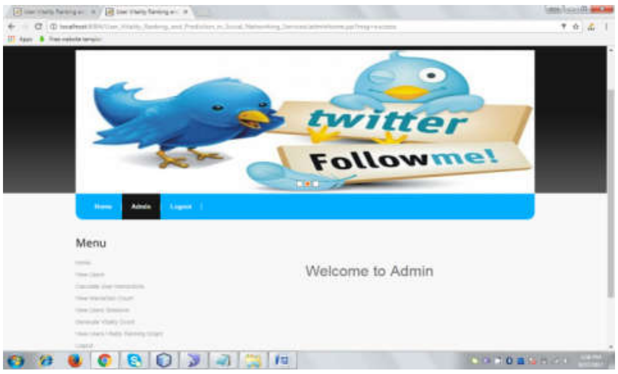
Calculating Vitality Ranking
Given the number of interactions between all pairs of users, we may count the number of all interactions for each user and rank them based on the count using Initial Ranking Algorithm. However, given the number of interactions between two nodes (users), The unified vitality score _i j of a node j (1 _ j _ N) in time period i(1 _ i_ M) within a social network I is defined as:

ALGORITHM

- The accumulated number of interactions SAijof a node j (1 _ j _ N) in time period i(1 _ i M)within a social network I is defined asSAij=Σk2fNei(j)g_kj,wherefNei(j)g denotes the set of users that are connectedto user j.
- First, if the accumulatednumber of interaction of node (i.e., SAij) increases a lotover that in the previous time period.
- The relative increase of interaction of a nodej (1 _ j _ N) in time period i(1 _ i_ M) within a socialnetwork I is defined as:IAij=SAijSAij.
- we can get that the relative increaseof interaction for node A and node C in time period n.
- The average interaction for user Si is definedas:AverageIij=SAijdegreeijwhere degreeijdenotes the number of connected friends foruser j.
- The term AverageIijrepresents the average numberof interactions of User j in period i.
- The unified vitality score _ijof a node j(1 _ j _ N) in time period i(1 _ i_ M) within a socialnetwork I is defined by average of SAijSAij.
- we define the user's vitality score _ijwithtwo terms and combine them in a linear way.
- The first part,SAijSAij, indicates the dynamic vitality level in period i.
- The second part denotes the static vitalitylevel of user in one period.
- By tuning the parameter, we may balancethe impact of the relative increase of interaction and theaverage interaction.

- In theexperiment, we will empirically examine the impact of _ onthe performance of our algorithm. Given a social network, wewill compute the ijfor all nodes (users) for a specified timeperiod, and then rank all nodes according to the value.
-

VI. EXPERIMENTAL RESULTS



VII. CONCLUSION AND FUTURE SCOPE

In this project, we presented a study on user vitality ranking andprediction in social networking services such as microblogapplication. Specifically, we first introduced a user vitalityranking problem, which is based on dynamic interactionsbetween users on social networks. To solve this problem,we developed two algorithms to rank users based on vitality.While the first algorithm works based on the developedtwo user vitality measurements, the second algorithm furthertakes into account the mutual influence among users whilecomputing the vitality measurements. Then we presented auser vitality prediction problem and introduced a regressionbasedmethod for the prediction task. Intensive experimentson two real-world data sets that are collected from differentdomains clearly demonstrate the effectiveness of our rankingand prediction methods. The accurate results of both uservitality ranking and prediction could benefit many partiesin different social networking services, e.g., a user vitalityranking list could help ads providers to better display theirads to active users and reach more audiences.

REFERENCES

1. EytanBakshy, Jake M Hofman, Winter A Mason, and Duncan JWatts. Everyone’s an influencer: quantifying influence on twitter. InProceedings of the fourth ACM international conference on Web searchand data mining, pages 65–74. ACM, 2011.
2. Sergey Brin and Lawrence Page. Reprint of: The anatomy of a largescalehypertextual web search engine. Computer networks, 56(18):3825–3833, 2012.
3. Robert Goodell Brown. Smoothing, forecasting and prediction ofdiscrete time series. Courier Corporation, 2004.
4. Christopher S Campbell, Paul P Maglio, Alex Cozzi, and Byron Dom.Expertise identification using email communications. In Proceedingsof the twelfth international conference on Information and knowledgemanagement, pages 528–531. ACM, 2003.
5. Meeyoung Cha, HamedHaddadi, FabricioBenevenuto, and P KrishnaGummadi.Measuring user influence in twitter: The million followerfallacy.ICWSM, 10(10-17):30, 2010.
6. Pedro Domingos and Matt Richardson.Mining the network value ofcustomers. In Proceedings of the seventh ACM SIGKDD internationalconference on Knowledge discovery and data mining, pages 57–66.ACM, 2001.
7. Philip E Brown JunlanFeng. Measuring user influence on twitter usingmodified k-shell decomposition. 2011.
8. Jian Jiao, Jun Yan, Haibei Zhao, and Weiguo Fan. Expertrank: An expertuser ranking algorithm in online communities. In New Trends in Informationand Service Science, 2009.NISS’09. International Conferenceon, pages 674–679. IEEE, 2009.
9. Jon M Kleinberg. Authoritative sources in a hyperlinked environment.Journal of the ACM (JACM), 46(5):604–632, 1999.
10. Shamanth Kumar, Fred Morstatter, and Huan Liu.Twitter data analytics.Springer, 2014.
11. Wenting Liu, Guangxia Li, and James Cheng. Fast pagerankapproximationby adaptive sampling. Knowledge and Information Systems,42(1):127–146, 2015.
12. RadaMihalcea. Graph-based ranking algorithms for sentence extraction,applied to text summarization. In Proceedings of the ACL 2004 onInteractive poster and demonstration sessions, page 20. Association forComputational Linguistics, 2004.
13. Amin Omidvar, Mehdi Garakani, and Hamid R Safarpour. Contextbasedusr ranking in forums for expert finding using wordnet dictionaryand social network analysis. Information Technology and Management,15(1):51–63, 2014.
14. Tore Opsahl, FilipAgneessens, and John Skvoretz. Node centralityin weighted networks: Generalizing degree and shortest paths. Socialnetworks, 32(3):245–251, 2010.
15. Daniel M Romero, WojciechGaluba, SitaramAsur, and Bernardo AHuberman. Influence and passivity in social media. In Machine learningand knowledge discovery in databases, pages 18–33. Springer, 2011.
16. Yuanfeng Song, Wilfred Ng, Kenneth Wai-Ting Leung, and QiongFang.Sfp-rank: significant frequent pattern analysis for effective ranking.Knowledge and Information Systems, 43(3):529–553, 2015.
17. Stanley Wasserman and Philippa Pattison. Logit models and logisticregressions for social networks: I. an introduction to markovgraphsandp. Psychometrika, 61(3):401–425, 1996.
18. JianshuWeng, Ee-Peng Lim, Jing Jiang, and Qi He.Twitterrank: findingtopic-sensitive influential twitterers. In Proceedings of the third ACMinternational conference on Web search and data mining, pages 261–270.ACM, 2010.

19. Yuto Yamaguchi, Tsubasa Takahashi, Toshiyuki Amagasa, and HiroyukiKitagawa.Turank: Twitter user ranking based on user-tweet graphanalysis. In Web Information Systems Engineering–WISE 2010, pages240–253. Springer, 2010.

20. Yang Yang, Ryan N Lichtenwalter, and Nitesh V Chawla. Evaluatinglink prediction methods.Knowledge and Information Systems,45(3):751–782, 2015.

21. yKang Zhao, John Yen, Greta Greer, BaojunQiu, PrasenjitMitra, andKennethPortier.Finding influential users of online health communities:a new metric based on sentiment influence.Journal of the AmericanMedical Informatics Association, 21(e2):e212–e218, 2014.