International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

User Vitality Ranking and Prediction in Social Networking Services: a Dynamic Network Perspective

Mr.N.Manoj Kumar
Department of CSE
Anurag group of institutions

Mr.Pallam Ravi
Professor
Anurag group of institutions

Dr.G.Vishnu Murthy
Professor,HOD,
Anurag Group of institutions

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 paper, 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 microblog sites and academical collaboration networks. Intuitively, 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, it is very likely that this user has high vitality. 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. Other than user vitality ranking, we also introduce a vitality prediction problem, which is also of great importance for many applications in social networking services. Along this line, we develop a customized prediction model to solve the vitality prediction problem. To evaluate the performance of our algorithms, we collect two dynamic social network data sets. The experimental results with both data sets clearly demonstrate the advantage of our ranking and prediction methods.*

Index Terms—Distributed systems, monitoring data, social networks, user activity, vitality ranking, vitality prediction.

INTRODUCTION With the development of web technology, social networking service has been prevalent at many online platforms. The social networking service facilitates the building of social networks or social relations among users who, for instance, share interest, activities, background and physical connections. Through such service, users could stay connected with each other and be informed of friends' behaviors such as posting at a platform, and consequently be influenced by each other. For instance, in today's Twitter and Weibo (one of the most popular social networking sites in China), a user can get the instant updates about his connected friends' postings and could further retweet or comment the postings. Within a time period, millions of users may take different actions such as posting and retweeting at these social networking sites. One interesting and important problem is how to rank users based on their vitality with historical data [10]. An accurate vitality ranking of users will provide great insight for many applications in most online social networking sites. For instance, online ads providers may make better strategy for delivering their ads via considering the ranked vitality of users; site operators may design better practices for online campaigns (e.g., online survey) via leveraging the ranking list. While it is very promising for many parties to provide a vitality ranking of users, there are many technical challenges to tackle this problem. First, to decide the vitality of a user, we could not only examine his own interaction with others, but also need to look into the interactions of other users collectively. For instance, suppose one user has had many interactions with most of his friends in a time period, we may conclude different vitality of this user when most of his friends also have had many interactions in the same time period versus when most of his friends do not have had many interactions. Second, as the scale of social networks increases, it becomes more challenging to rank the vitality of users because a large number of nodes (users) may influence the vitality of an individual node (user). Third, as the social networks in many online sites evolve over time, the vitality of users may also change over time. Thus efficient methods are needed to dynamically obtain the vitality of users at different times. In the literature, researchers have made some efforts on ranking users in social networking sites. For instance, in [19], a Twitter user ranking algorithm was proposed to identify authoritative users who often submit useful information. The proposed algorithm mainly works based on the user-tweet graph, rather than the user-user social graph. In [18], an extension of PageRank algorithm named TwitterRank was developed to rank Twitter users based on their influence. They first build topic-specific relationship network among users, then apply the TwitterRank algorithm for ranking. In [7], a modified K-shell decomposition algorithm is developed to measure the user influence in Twitter. Furthermore, in [21], [5], [1], some explicit measurements such as retweets and mentions are developed to measure and rank user influence in Twitter. However, most of these measurements quantify the influence in an isolated way, rather than in a collective way. Furthermore, the focus of these methods is on influence, which is still different from the vitality that we address in this paper. To this end, in this paper, we propose two types of node vitality ranking algorithms that analyze the vitality of all nodes in a collective way. First, for a node A that has many interactions with his friends in a time period, if most of his friends do not have many interactions with their friends, it is very likely that the node A has high vitality. Based on this intuition, we define two measurements to quantify the vitality level of each node and propose the first algorithm. Second, by exploiting the mutual dependency of vitality among all users within a social network, we propose the second algorithm that infers the vitality level of users in an iterative way. Through the iteration, all nodes' measurements propagate through the network and affect each other. Thus the second algorithm is able to collectively analyze the vitality score of all nodes by considering the whole network. Furthermore, upon our in-depth understanding about user vitality, we propose an improved model to predict the vitality of users. The successful prediction results will further benefit many applications on social networking sites. Finally, we conduct intensive experiments on both user vitality ranking and prediction with two large-scale real world data sets. The experimental results demonstrate the effectiveness and efficiency of our methods.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

PROBLEM STATEMENT In the section, we introduce the research problem of user vitality ranking in the context of social networking service. 2.1 Vitality Ranking in a Social Network 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 $\{Uj\}1 \le j \le N$ and L links among users denoted as $\{Ejk\}1 \le j,k \le N$, where j and k are indices. We have recorded all interactions between them within M consecutive time periods Ti ($1 \le i \le M$). For instance, we show an example social network in Figure 1, where we have 7 nodes, 10 links with two time periods. For each time period Ti , let us use θ i jk to denote the number of interactions between node j and node k, and SAi j to represent the accumulated number of interactions between node j and all other nodes. In a time period Ti , we can get all interactions between all pairs of nodes, which reflect the vitality of all users in the time period. For instance, in Figure 1, the number 28 above the Node A means this user has 28 interactions with others and indicates the vitality of user A. For simplicity, we use Si to denote all interactions of a social network S within a time period Ti . Consequently, for a social network S, we may have a sequence of Si ($1 \le i \le M$) within M consecutive time periods. Our goal is to rank all users from high vitality to low vitality for a time period Ti based 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. Based on the above description and notations, we formally state the vitality ranking problem as follows. The Vitality Ranking Problem Given: A social network S that includes N nodes Uj ,($1 \le j \le N$), L links Ek,($1 \le k \le L$), and additional information θ i j possibly available for each link. Within each of M time periods Ii ,($1 \le i \le M$), we observe all interactions between all users that are denoted as Si ($1 \le i \le M$). Objective: Ranking all users based on their vitality within each time period i ($1 \le i \le M$). Note that the given social network S in the above vitality ranking problem is a connected graph, which means there is a path between any nodes. Given a social networking system, it is possible that multiple separate social networks may exist, which are completely separated. But we focus on the node vitality ranking in a single social network in this paper. In the following, a social network indicates a connected graph unless specified otherwise. For multiple separate social networks, we may conduct the vitality-based ranking for users in each social network, and then develop a way to merge the multiple ranking lists to obtain a unified ranking list of all users. 2.2 Discussion about the Vitality Ranking Problem First, the social network considered in our problem is an undirected graph and the interaction between two users is also symmetric. Second, 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. However, given the number of interactions between two nodes (users), it is challenging to infer which one contributes how much to all interactions. Thus, it may not be accurate to rank all users based on the accumulated count of all interactions. Third, this problem is different from many existing node ranking problems such as webpage ranking. Most node ranking algorithms could not be directly used for this problem because the goal is to rank nodes based on the dynamic interactions that actually evolve over times.

The Algorithm Time Complexity Analysis In this section, we will discuss the time complexity of our algorithm. For convenience, we can analyze the time complexity with operations of each edge and define the number of edges as Nj , j indicates the period of the data. First, we focus on the time complexity analysis within one iterative. In first iterative, the initial vitality score calculation of each node is linearly related with the number of edges, because each individual operation is allocated the weight of each edge. In one operation, we need to calculate the initial score of two connected nodes and allocated the weight to those two nodes with our algorithm, which indicate the time complexity of each operation is a constant(denotes as a). After allocating all link weights to nodes, one iterative is finished and the time complexity of one iterative is a $*$ Ne, where Ne denotes the number of edges. If we assume the number of iterations is M, it's easy to know that the time complexity of out algorithm is a ∗ M ∗ Ne which can be written as O (M ∗ Ne).

PREDICTING THE USER VITALITY In this section, we introduce and address the problem of predicting the user vitality based on the model and inference of user vitality in a social network. The successful prediction of user vitality could benefit many applications in most social networking sites such as Facebook and Twitter. Particularly, as the number of users in most social networking sites is very large, it is very important to know in advance who will be or will not be very active in the future. First, the site operators may design early and useful strategy to encourage inactive users to interact with others and content. This could help them maintain the global user vitality of a social networking site. Second, the site operators may also decide better ads display strategy by using the future user vitality. For instance, they may deliver and display interesting ads to active users rather than inactive users as the former group has better chance to propogate the ads to others or click the ads directly. This could help them not only save cost for ads display, but also target potential users in a more accurate way, which will consequently help them improve their ads revenue. Particularly, in this paper, we will show the prediction of vitality for those users who are ranked on the top because these users often have high influence in the social networks and could bring more benefit to social networking sites if predicting their vitality successfully. Other than predicting the vitality of individual user, we also address the prediction of vitality for a group of users in this paper. As we know, there are many groups formed in social networking sites. Users in each group often behavior very similarly. For instance, they often chat, tweet and re-tweet with each other. While it may be very challenging to predict the vitality of each single user, it may be easier to predict the aggregate vitality of a group of users. Plus, the successful prediction for a group of users could be beneficial for many parties on social networking sites as well. 4.1 Basic Forecasting Model There are some existing models that could be used for predicting the dynamic vitality score, such as Markov model [17] and exponential smoothing forecast [3]. Based on the characteristic of the data from social networking system, we choose the triple exponential smoothing [3] to predict the dynamic vitality score of each user in this paper. Furthermore, we also

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

propose a way to obtain the vitality score of a group users with the triple exponential smoothing.The Improved Model For Predicting the Ranking Score As we know, in the triple exponential smoothing model, there are two important parameters which affect the prediction result: the smoothing factor α and the initial smoothed value. For the initial smoothed value, we use the average of the vitality scores of users in the last time periods as the initial smoothed value. For the smoothing factor α, it is difficult to get the ideal value. Furthermore, in the social network systems each individual node has independent smoothing factor because each node indicates an individual user who has unique behavior in the system. If we want to predict the vitality score of each node, we have to find the smoothing factor α for each node. In this paper, we use the weight-based method to adjust the factor based on previous information. For example, to predict the vitality score of the user j in time period i + 1, the initial smoothed value is generated by the sequence [tα 1 j ,t α 2 j , ...,t α i−1 j] . Furthermore, the smoothing factor α is generated within time period i and we will use the vitality score of time period i − 1 to predict that in time period i. Then we will use the simulated annealing method to adjustment the smoothing factor. When the simulated annealing method ends, we get an appropriate result of the smoothing factor α that will be used for predicting the vitality score of period i+ 1. As we know, when the value of smoothing factor is close to one, we have less smoothing effect and give greater weight to recent changes. On the other hand, when the vitality score of a user is high and the user has a great change in recent time period, we assume this trend will continue in next time period. Based on this intuition, we can tune the smoothing parameter α with the dynamic vitality score as shown in Equation 23. IMi j = tα i j−tα i−1 j tαi j . (23) Note that, the tα i j and tα i−1 j are positive and the IMi j is less than one. So the improved smoothing parameter α ′ is also less than one and the IM indicates the recent changes of user j. With the improved exponential smoothing model, we can predict the vitality score of users in a networking system. Firstly, we remove lots of low-vitality nodes those average interactions is less than one.Because their behavior is difficult to predict and we only focus on the relatively active nodes. After learned the parameters of prediction model, we can use it to predict the users' future vitality score. Specifically, the input is the computed users' vitality score in the first N periods. The initial smoothed value of each node is calculated by the average vitality score of first 3 periods. Then, we use Equation 23 to generate the smoothing factor α of each node. When we get these two parameters, we can predict the users' vitality score of time period N+1 with Equation 21. We can even further rank users based on the predicted vitality score in time period N+1.

CONCLUDING REMARKS In this paper, we presented a study on user vitality ranking and prediction in social networking services such as microblog application. Specifically, we first introduced a user vitality ranking problem, which is based on dynamic interactions between 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 developed two user vitality measurements, the second algorithm further takes into account the mutual influence among users while computing the vitality measurements. Then we presented a user vitality prediction problem and introduced a regressionbased method for the prediction task. Intensive experiments on two real-world data sets that are collected from different domains clearly demonstrate the effectiveness of our ranking and prediction methods. The accurate results of both user vitality ranking and prediction could benefit many parties in different social networking services, e.g., a user vitality ranking list could help ads providers to better display their ads to active users and reach more audiences.

REFERENCES [1] Eytan Bakshy, Jake M Hofman, Winter A Mason, and Duncan J Watts. Everyone's an influencer: quantifying influence on twitter. In Proceedings of the fourth ACM international conference on Web search and data mining, pages 65–74. ACM, 2011.

[2] Sergey Brin and Lawrence Page. Reprint of: The anatomy of a largescale hypertextual web search engine. Computer networks, 56(18):3825– 3833, 2012.

[3] Robert Goodell Brown. Smoothing, forecasting and prediction of discrete time series. Courier Corporation, 2004.

[4] Christopher S Campbell, Paul P Maglio, Alex Cozzi, and Byron Dom. Expertise identification using email communications.

In Proceedings of the twelfth international conference on Information and knowledge management, pages 528–531. ACM, 2003.

[5] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and P Krishna Gummadi. Measuring user influence in twitter: The million follower fallacy. ICWSM, 10(10-17):30, 2010.

[6] Pedro Domingos and Matt Richardson. Mining the network value of customers. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pages 57–66. ACM, 2001.

[7] Philip E Brown Junlan Feng. Measuring user influence on twitter using modified k-shell decomposition. 2011.

[8] Jian Jiao, Jun Yan, Haibei Zhao, and Weiguo Fan. Expertrank: An expert user ranking algorithm in online communities. In New Trends in Information and Service Science, 2009. NISS'09. International Conference on, 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 pagerank approximation by adaptive sampling. Knowledge and Information Systems, 42(1):127–146, 2015.

[12] Rada Mihalcea. Graph-based ranking algorithms for sentence extraction, applied to text summarization. In Proceedings of the ACL 2004 on Interactive poster and demonstration sessions, page 20. Association for Computational Linguistics, 2004.

[13] Amin Omidvar, Mehdi Garakani, and Hamid R Safarpour. Context based user ranking in forums for expert finding using wordnet dictionary and social network analysis. Information Technology and Management, 15(1):51–63, 2014.

[14] Tore Opsahl, Filip Agneessens, and John Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. Social networks, 32(3):245–251, 2010.

[15] Daniel M Romero, Wojciech Galuba, Sitaram Asur, and Bernardo A Huberman.

Influence and passivity in social media. In Machine learning and knowledge discovery in databases, pages 18–33. Springer, 2011.

[16] Yuanfeng Song, Wilfred Ng, Kenneth Wai-Ting Leung, and Qiong Fang. 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 logistic regressions for social networks: I. an introduction to markov graphs andp. Psychometrika, 61(3):401–425, 1996.

[18] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: finding topic-sensitive influential twitterers. In Proceedings of the third ACM international conference on Web search and data mining, pages 261– 270. ACM, 2010.

[19] Yuto Yamaguchi, Tsubasa Takahashi, Toshiyuki Amagasa, and Hiroyuki Kitagawa. Turank: Twitter user ranking based on user-tweet graph analysis. In Web Information Systems Engineering–WISE 2010, pages 240–253. Springer, 2010.

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

[21] Kang Zhao, John Yen, Greta Greer, Baojun Qiu, Prasenjit Mitra, and Kenneth Portier. Finding influential users of online health communities: a new metric based on sentiment influence. Journal of the American Medical Informatics Association, 21(e2):e212–e218, 2014

Author Details

N.Manoj kumar
Anugar group of institutions (CVSR)
email:manoj.nagelli94.mn@gmail.com

pallam ravi
assistant professor
Anurag group of institutions