J Inf Process Syst, Vol.13, No.4, pp.987~999, August 2017
https://doi.org/10.3745/JIPS.04.0041

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Identifying Influential People Based on Interaction Strength

Muhammad Azam Zia*, Zhongbao Zhang*, Liutong Chen*, Haseeb Ahmad* and Sen Su*

Abstract
Extraction of influential people from their respective domains has attained the attention of scholastic community during current epoch. This study introduces an innovative interaction strength metric for retrieval of the most influential users in the online social network. The interactive strength is measured by three factors, namely re-tweet strength, commencing intensity and mentioning density. In this article, we design a novel algorithm called IPRank that considers the communications from perspectives of followers and followees in order to mine and rank the most influential people based on proposed interaction strength metric. We conducted extensive experiments to evaluate the strength and rank of each user in the micro-blog network. The comparative analysis validates that IPRank discovered high ranked people in terms of interaction strength. While the prior algorithm placed some low influenced people at high rank. The proposed model uncovers influential people due to inclusion of a novel interaction strength metric that improves results significantly in contrast with prior algorithm.

Keywords
Influence, Interaction, IPRank, Micro-Blog, Online Social Network

1. Introduction

Social network is a complex structure comprising of various social entities and the mutual relationship among them. This intricate structure plays a noteworthy role in the world of information broadcasting, advertisement, understanding social behavior, knowledge discovery, and information dissemination. In other words, a social network provides platform to interact with people around the globe in order to share ideas and information. Within online social networks (OSNs) domain, organizations or individuals are tied to interdependencies like fellowship, friendship, and collaboration, etc. [1]. Further, OSNs furnished new aspects for business prototypes and stimulate viral marketing [2]. Pioneer OSNs such as LinkedIn and Facebook have been magically widespread among the people for easily making the connections with friends and well known people throughout the globe. Other leading social networks include instant messaging networks [3], mobile social networks [4] and scientific collaboration networks [5]. Twitter [6] and Sina Weibo [7] are among the most popular blogging service platforms that have been studied for last several years. Under such information sharing

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Manuscript received April 21, 2017; first revision June 16, 2017; accepted June 21, 2017.
Corresponding Author: Muhammad Azam Zia (zia_uaf@hotmail.com)
* State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China ({zia, zhongbaozb, chenliutong, haseeb_da, susen}@bupt.edu.cn)

Copyright© 2017 KIPS

platform, every user is a broadcaster, listener, and messenger within a micro-blog platform. These platforms have revolutionized the world due to their high speed of information sharing and public networking functionality.

To extract influential people for viral marketing, OSNs have been progressed to the highest level with the development of information technologies, which are penetrated in all dimensions of humans' life. By definition, social influence refers to the capability or supremacy of triggering an effect in substantial ways. This fact opens a new dimension of research for finding the most influential people (IP) that can influence to change someone's opinions in a stronger way with respect to other people [8]. Influential people can spread or advertise the news virally or exponentially. The identification of IP in micro-blog is focused by many researchers since last decade and their recommendations provide new insights for viral marketing. In more details, within OSNs domain, researchers mainly focused towards mining the key people [9-11], community leaders [12], trend setters [13], opinion leaders [14] and topic experts [15]. Cha et al. [16] concluded that influential users have substantial influence on variant topics; more influential user is one, having an active audience to re-tweet the tweets or to mention the users but overlooked to consider the significance of re-tweet factor. Similarly, TunkRank algorithm proposed by Tunkelang [17] based on PageRank model, only considers the *following* relationship between users, while there are missing factors that might be considered like re-tweet, mention, and comment, etc. Further, the influence value is distributed evenly rather than proportionally. Albeit of much progress in the underlying field, there still exist many limitations that necessitate to be considered while extracting IP.

We are motivated from the aforementioned studies and their proposed methods overlooked the interaction strength in follower-followee relationships. It is our hypothesis that re-tweet factor has a large impact in the interaction strength (IS) to identify the IP in the micro-blog network. Moreover, other important factors such as comment and mention density are also incorporated in the proposed method. These factors are included to find out the most prestige people. Furthermore, the proposed concept is better evaluated based on prominence gained by the followee people in terms of re-tweets, comments and mentions from the following users in the network. Therefore, our methodology introduces an interactive strength that is measured by three prestigious factors, namely re-tweeting strength, commencing intensity and mentioning density.

The main contributions for the identification of IP in the micro-blog are presented as follows.

First, it is the first attempt to evaluate the IS of each user by considering the communications from perspectives of follower-followee, where the interactive degree is measured by three factors including: re-tweeting strength, commencing intensity and mentioning density.

Second, we present an innovative algorithm named as IPRank that extracts the IP while employing the concept of IS.

Third, we performed extensive experiments and extracted highly ranked IP comparative to other users within micro-blog. Finally, we compared the obtained results with prior TunkRank algorithm. The comparative study highlighted that IPRank outperforms the baseline algorithm while discovering the people having more IS. Further, it is also noted that re-tweet strength of IP is higher than comparative people in micro-blog network.

The rest of the paper is structured into following sections. Related work is discussed in Section 2. Problem definition is presented in Section 3. Section 4 describes our strategy to compute IS weights as well as the detailed insights of the introduced IPRank algorithm. The experimental evaluation and comparative studies are presented in Section 5. Section 6 concludes precise theme of our study.

2. Related Work

The research to find IP has been a hot issue in OSNs domain. Many researchers proposed different techniques to extract the IP. An algorithm was proposed by Lu and Wan [10] to extract the influential people within the micro-blog network. The spreading matrix was constructed in order to rank the each node within the micro-blog network. Further, the nodes ranked based on proposed model to identify the influential people in [10]. Kwak et al. [11] extracted influential people and ranked them within micro-blog network according to the number of followers and concluded that the obtained results were similar that were obtained by PageRank algorithm. Topological characteristics of Twitter were studied and its capability as a new platform of information dissemination was investigated in [11]. An algorithm was proposed by Wang et al. [12] based on PageRank algorithm in order to discover community leaders in micro-blog network. Number of followers, comments and reposts were the three incorporated factors of micro-blog. Using these features and analyzing their relationships, the authors identified the rank of each user within micro-blog. The factors that might be considered like re-tweet and mention were not included. Re-tweets, indegree, and mentions were the three different features of micro-blog discussed by Cha et al. in detail [16]. Using these features, the authors scrutinize the dynamics of a person influence across time and topics. Some interesting observations were also discussed, for instance, large number of influential users can hold strong position on a variety of topics. Moreover, most popular people having extraordinary indegree are not essentially to be called influential with respect to mentions or re-tweets. Further, influence could not be gained accidently or spontaneously but through intensive interaction. TunkRank algorithm [17] incorporated the concept of PageRank model, only by considering the *following* relationship in microblog. But essential factors that might be considered such as re-tweet, mention, and comment, etc. are not incorporated with in TunkRank. Further, the influence value is distributed evenly rather than proportionally.

In order to quantify the user influence based on IS, different models and respective algorithms were presented and applied to different datasets that belongs to Weibo and Twitter etc. The research work presented by Leavitt et al. [18] divided the user's influence into two categories: content based and conversation based. It was concluded that news media is influential at spreading contents, while celebrities are better for induction of conversation. Bakshy et al. [19] considered micro-blog network to find the IP and URLs that remained more interesting and remained a source of spreading optimistic information. Similarly, the most influential user on a specific topic were extracted in [20]. More precisely, the authors employed different algorithms that preprocess the dataset, cluster the similar documents and extract the influential node that could help for influence maximization. Albeit of many nuanced results, either some proposals employed only small datasets or proper weight age is not given to IS. Further, none among the aforementioned proposals provided any information about superiority of employed features with respect to information gain or gain ratio. To best of our knowledge, none of the previous works provided the mathematical formulations in such way to improve TunkRank.

3. Problem Definition

Micro-blog network is defined to be a directed and IS based graph $G = (V, E, IS)$, where V represent users set, the edge set is indicated by E that describes the relationship of all users in the micro-blog network and IS denotes the interaction strength in terms of re-tweets, comments and mentions. $V = \{i \mid i \text{ is a user in the micro-blog}\}$, represents either a followee or a follower, which depends on the flow of information, while $E = \{(i,j) \mid IS(i,j) > 0, (i,j) \in V\}$, where $IS(i \to j)$ denotes IS gained by user i from user j as presented in Fig. 1. Here, i is nominated as followee and j is termed as follower in the micro-blog and the information is flowed from followee towards the follower. It is important to be noted that (i,j) and (j,i) depict two different edges with potentially unequal IS. For instance $IS(i \to j)$ represents the influence of *i* on *j* user, while $IS(j \to i)$ depicts the influence of user *j* on user *i* only if there exists a reciprocal link. First IS is calculated for every user based on the information of re-tweets, comments and mentions. Subsequently, proposed algorithm takes IS as input to find the most influential people. More precisely, the IS can be calculated between two users only if they are connected. The connection between two users is represented by *L*. For instance, L_{ij} means that user *j* follows user *i* and the link value will be *L*=1 otherwise 0. Similarly L_{ji} means that user *i* follows user *j* and here the link value will be *L*=1 otherwise 0. Other terms are based on this connection (*L*) before to consider any other term and its corresponding value. The IS can be calculated between those users whose connection value is *L*=1 otherwise not. Terms and their descriptions are presented in Table 1. The connection between two users within the micro-blog network can be explained in the following way:

$$L_{ij} = \begin{cases} 1 & \text{if } j \text{ follows } i \\ 0 & \text{Otherwise} \end{cases} \tag{1}$$

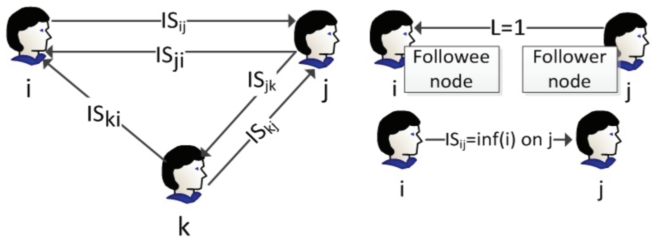

Fig. 1. Interaction Strength.

Table 1. Terms and Descriptions

Term	Description
i,j,k…n	represent the people/users in the micro-blog
L_{ij}	User *i* and *j* are connected while User *i* is followed by *j*.
j←i	*J* is followed by *i* and *j* is a followee
inf(i)	influence value of User *i*
Out (j)	the number of people followed by *j*
followers (i)	followers set of User *i*
followees (i)	followees set of User *i*
RT(i←j)	strength of re-tweet gain by User *i* from *j* User
Retweeted (j)	User *j* re-tweeted value
Tweets (j)	tweets of User *j*
Retweet(j→i)	User *j* re-tweeting *i* User
Retweet (j→m)	User *j* re-tweeting to other people
IS(i→j)	Similarly Comment (CT) and Mention (MN) in Eq. (4) & (5) interaction strength (influence) of User *i* on *j*

4. Influential People Rank Algorithm: (IPRank)

In this section, first we describe TunkRank algorithm and its limitations. Next, we discuss the necessary changes that are essential to evaluate the appropriate influence of each user in the micro-blog network. Subsequently, we provide the mathematical notions of (IS) in terms of re-tweets, comments and mentions in the network. Finally, we explained the outline of IPRank algorithm with its mathematical formulation.

4.1 TunkRank Algorithm: (TunkRank)

TunkRank algorithm is used to calculate every user's influence recursively [17]. The mathematical notation of TunkRank algorithm is presented as follow:

$$inf(i) = \sum_{j \in Followers(i)} \frac{1 + p * Inf(j)}{|Out(j)|} \tag{2}$$

where $inf(i)$ denotes the influence of user i based on probable count of user j who re-tweets the content posted by i. Both readings are counted if the same content is read by a user twice (first directly from tweet and later due to re-tweets). If $j \in followers(i)$ then the probability of j to re-tweet a tweet of i will be $(1/|Out(j)|)$, where $Out(j)$ denotes the number of users followed by j. Besides, if j reads a content posted by i, then p represents a constant probability that the underlying tweet will be re-tweeted by j.

4.2 Limitations of TunkRank

TunkRank works in similar fashion to that of PageRank [21], but this algorithm only considers the following relationship between users, while there are other influential factors that must be considered for finding out more appropriate influential people such as comments and mentions. Also, TunkRank algorithm distributes influence value evenly rather proportionally. We remark that the probability of re-tweeting, commencing and mentioning should be given based on actual influence strength of users because people with high influence strength has high probability of being re-tweeted, commented and mentioned by its followers. In similar fashion, the user with lower influence must be incorporated accordingly.

4.3 Modeling the Interaction Strength

In this section, IS of each user is evaluated based on our novel concept to better analyze the significance of each user. Analysis of IS measure provides a better understanding of the different users' roles in the micro-blog network.

4.3.1 Re-tweet strength

Re-tweets represent the content value of one's tweets and this factor plays a vital role to uplift the influence of a user. In the micro-blog network, re-tweeting is considered as a response of a follower to his/her followee tweets and more re-tweet symbolizes the importance of the followee. For instance, if a follower *j* is interested in tweets of *i* and further wants to introduce the same content to his followers. Thus, *j* is becoming the source to increase the influence of user *i*. Therefore, we count the re-tweet actions of each follower in order to quantify the re-tweet strength gained by followee formulated as follows:

$$RT(i \to j) = \frac{\sum \text{Retweeted}(j)}{\sum Tweets(j)} * \frac{\sum \text{Retweeted}(j \to i)}{\sum \text{Retweet}(j \to m)} \tag{3}$$

where the part of equation prior to multiplication $(*)$ represents the re-tweeting probability p_{RT} of *j*'s tweet to be re-tweeted (which was proposed to be constant (i.e., 0.2) in [17]). In case, if *j* is an influential user, it will have high p_{RT} value. In this way, it will give more weightage to user *i* in order to increase the influence value of underlying user.

4.3.2 Commenting intensity

Another important factor of micro-blog networks is commenting intensity that depicts one's response to the subjects of tweets or re-tweets. Same as *RT*, we define commenting intensity as a measure that counts the commencing strength gained by a followee via its followers. The mathematical formulation of comment intensity is presented as follows:

$$CT(i \to j) = \frac{\sum Commented(j)}{\sum Tweets(j) + \sum \text{Re}tweeted(j)} * \frac{\sum Comment(j \to i)}{\sum Comment(j \to m)} \tag{4}$$

where the part of equation prior to multiplication $(*)$ depicts the commenting probability p_{CT} of j in terms of tweet and re-tweets. In case, if *j* is an influential user, it will have high p_{CT} value. In this way, it will give more weightage to user *i* in order to increase the influence value of underlying user.

4.3.3 Mentioning density

Mention represents the name value of a user marked in order to directly interact with contacts by including @ in the tweets, re-tweets and comments. Similar to Eqs. (3) and (4), we calculate the value of mentioning density to quantify the strength gained by a followee. The mathematical notion of mention density can be expressed as follows:

$$MN(i \to j) = \frac{\sum Mentioned(j)}{\sum Tweets(j) + \sum \text{Re}tweeted(j) + \sum Commented(j)} * \frac{\sum Mention(j \to i)}{\sum Mention(j \to m)} \tag{5}$$

where the part of equation prior to multiplication $(*)$ depicts the commenting probability p_{MN} of j in terms of tweet, re-tweets and comments. In case, if *j* is an influential user, it will have high p_{MN} value. In this way, it will give more weightage to user *i* in order to increase the influence value of underlying node.

Our proposed notions calculate the IS of each user by considering the communications between followers and followees. The interactive degree is measured by three factors including: re-tweeting strength, commencing intensity and mentioning density. However, it obeys the principal that the larger value of IS, the more influential the followee will be. Moreover, IS between two users directly reflects the influential degree as it is depicted by our proposed concept. More precisely, above factors are aggregated into Eq. (6) in order to obtain the influence value of each user. The mathematical notation of influence value expressed as follows:

$$IS(i \to j) = \frac{RT(i \to j) + CT(i \to j) + MN(i \to j)}{3} \tag{6}$$

4.4 IPRank Algorithm: (IPRank)

In this section, we propose an effective IPRank algorithm that considers IS from followee-follower perspectives to extract influential people within micro-blog network. The essence of TunkRank algorithm is incorporated in order to measure the influence of each user by our proposed IPRank algorithm. The mathematical notation of IPRank algorithm is presented as follows:

$$inf(i) = \sum_{j \in Followers(i)} \frac{1 + IS(i \to j) * Inf(j)}{Out(j)} \tag{7}$$

where $IS(i \to j)$ represents the IS of user i has on j while $Out(j)$ denotes the set of people that are followed by j. It is worth to be noted that the varying probabilities of re-tweets, comments and mentions are properly incorporated while calculating IS of each user.

Algorithm 1: IPRank Algorithm

Input social network graph and action file: $G = (V, E)$ and Threshold: T
Output: get each user influence
1. **for** each edge $(i, j) \in E$ do
2. **evaluate** $RT(i \to j), CT(i \to j), MN(i \to j)$
3. **end for**
4. measure each user interaction strength $IS(i \to j) = \frac{RT(i \to j) + CT(i \to j) + MN(i \to j)}{3}$
5. initialize influence value: $1/n$
6. **while** (ratio>T)
7. **for** each i in V do
8. **for** each $j \in followers(i)$
9. get $IS(i \to j)$ and $Inf_t(j)$
10. $Inf_{t+1}(i) += (1 + IS(i \to j) * Inf_t(j)) / Out(j)$
11. **end for**
12. **end for**
13. ratio $= max(Inf_{t+1}(i) - Inf_t(i)) / max(Inf_{t+1}(i))$
14. **end while**

Line 1-3: Calculate the $RT(i \to j), CT(i \to j), MN(i \to j)$ between every two users

Line 4: Calculate the interaction strength between two users based on $RT(i \to j), CT(i \to j), MN(i \to j)$

Line 5: Initialize the influence value of every user

Line 6-14: Update the influence value of every user based on the equation (line 10) until the ratio (the maximum difference of new influence value and old influence value of a user divided by the new value) is smaller than the threshold.

5. Experiments and Results

5.1 Experimental Setup

Sina Weibo [22] has launched in 2009 and got high temptation especially in P. R. China since that time. Presently, millions of users are registered on Weibo, generating multi-million contents every day. KDD Cup 2012 [23] dataset is employed as it is widely adopted due to its comprehension. Further, it includes all the contents of data that are required for our analysis. To ensure accuracy, preprocessing is done to normalize the dataset. In details, the dataset consists of follower-followee relationships along with their actions and reactions in terms of tweets, re-tweets, comments and mentions. The user in the micro-blog acts either as followee or a follower depends upon the flow of information [24].

5.2 Experimental Results and Analysis

IPRank incorporates IS, which is calculated based on Eq. (6) and flow of performed operations are presented in the aforementioned algorithm. The proposed method is applied on the refined dataset in order to mine the influence of each user. Extensive experiments are repeated several times to produce ranks through iterative process. The top ranked people with more IS through the presented model are thus extracted out. We retrieved top 10 most IP within the micro-blog network by IPRank algorithm as depicted in Table 2. In order to evaluate the effectiveness of our proposed algorithm, TunkRank algorithm is also applied on the same dataset. For comparison point of view, we attained top 10 IP by applying well-known baseline algorithm as depicted in Table 3.

Table 2. Top 10 people by proposed IPRank algorithm

Rank	UID	Re-tweeted	Commented	Mentioned
1	1774717	1070699	114470	47558
2	1760378	717321	94405	26073
3	1010402	5128951	46712	75
4	1837210	402877	36537	13407
5	1444067	49419	1050	99650
6	1774766	1130214	16898	5194
7	1675399	341377	37717	1234
8	647356	135350	27678	17711
9	1774505	174078	30408	25096
10	2188303	432275	22166	9409

5.2.1 Comparative analysis

The comparative analysis presents that there are five common users out of top 10 IP, which are retrieved by both algorithms. The new users extracted by IPRank stood at 3rd, 4th, 6th, 7th, and 10th positions are found to be more influential as compared to extracted by TunkRank algorithm for the same ranks. As these influential users own better IS comparative to respective users. In more details, the same users are ranked at 21th, 11th, 47th, 24th, and 12th by TunkRank algorithm as depicted by Table 4. The users with UIDs 1010402, 1837210, 1774766, 1675399, and 2188303 are emerged as new IP in Table 2 by our proposed concept due to their better IS.

The users with UID 1010402 and 1774766 emerged at 3rd and 6th ranks, respectively as these users obtained a better value of IS in terms of exceptional re-tweet actions comparative to the respective users. Similarly, users 1837210, 1675399, 2188303 became IP while acquiring the rank 4th, 7th, 10th score better IS in terms of re-tweet, comment and mention among other people. Besides, the user with UID1444067 ranked at 5th position higher than user with UID 647356, due to exceptional mentioning density in the network. Specifically, we remark that other factors such as comments and mention density play noteworthy role in order to find actual influence of the each user. It is also observed that IPRank achieves much better results than the baseline algorithm, by which less influential people are also retrieved in top 10.

Table 3. Top 10 people by TunkRank algorithm

Rank	UID	Re-tweeted	Commented	Mentioned
1	1774717	1070699	114470	47558
2	1760378	717321	94405	26073
3	647356	135350	27678	17711
4	1419930	103903	11057	4865
5	1760642	109450	22983	10717
6	1444067	49419	1050	99650
7	436442	154317	29592	11428
8	1774505	174078	30408	25096
9	1775024	59693	23511	9927
10	1760350	59907	16181	3187

Table 4. Comparison of top 10 people by IPRank and TunkRank algorithm

S no.	UID	Rank by IPRank	Rank by TunkRank
1	1774717	1	1
2	1760378	2	2
3	1010402	3	21
4	1837210	4	11
5	1444067	5	6
6	1774766	6	47
7	1675399	7	24
8	647356	8	3
9	1774505	9	8
10	2188303	10	12

5.2.2 Implications

The empirical results present that TunkRank relies only on re-tweets by followers, whereas our proposed concept also incorporates their interactive actions in terms of IS as presented in Eq. (6). Therefore, our proposed concept introduces a novel IS that is measured by three effective factors, namely re-tweeting strength, commencing intensity and mentioning density. Our findings provide new insights to discover the IP based on IS in micro-blog network. The experimental results demonstrated that having more followers does not always mean the extracted user is an influential in micro-blog. We have made following observations: first, the boost in the user's influence is more because of active re-tweets than other factors of micro-blog. Additionally, comments and mentions features also play a vital role during the extraction of IP within network; second, influence is not gained accidently but it is because of active role of a user in the network or number of re-tweets, comments and mentions.

5.3 Significance of Factors

In this subsection, we provide the information about superiority of employed factors (re-tweet, comment, mention) upon each other [25]. In order to investigate the effectiveness of each factor, we calculated information gain and gain ratio [26] as depicted in Fig. 2. The outcomes of these measures strengthens our hypothesis that re-tweet factor is more important comparative to remaining two factors of micro-blog. Comment stands at second, while mention ranks at third place. However, all of the factors are important to be included to calculate IS as none among them has got zero information gain or gain ratio value.

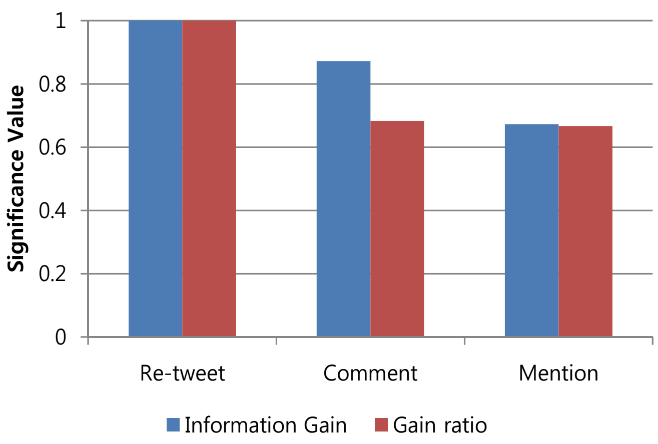

Fig. 2. Significance of factors.

5.4 Computation Cost

We computed time complexity of both algorithms in order to compare the efficiency of IPRank and TunkRank algorithms. The computation complexity of TunkRank is $O(n^2)$, where n denotes the number of nodes. The complexity of proposed algorithm IPRank is also $O(n^2)$, but IPRank converges in less iterations. More precisely, TunkRank converges in 96 iterations, while IPRank converges in 80 iterations for the same dataset. Based on this reason, we can say that IPRank is efficient than TunkRank algorithm.

6. Conclusion

Nowadays, finding influential people in online social networks is burning issue for researchers around the globe. In this paper, we presented a novel concept for finding the influential people in micro-blog network. More precisely, a conception of interaction strength is incorporated that considers different actions of a user in terms of re-tweet strength, commenting intensity and mentioning density from follower-followee perspectives. These real factors are utilized in our innovative IPRank algorithm to get more effective results. We performed extensive experiments whose results demonstrate that our ranking results are more convincing as compared to prior algorithm. Further, it is also noted that influential people own higher re-tweet strength comparative to other users in micro-blog network. However, the remaining two factors are also found to be effective for identifying the actual influential people. We believe that our findings would be useful for viral marketing and advertisement campaigns. The weight factor will be incorporated in order to present the importance of features in our future work.

Acknowledgement

This work was supported in part by the following funding agencies of China: National Natural Science Foundation under (Grant No. 61170274, 61602050 and U1534201).

References

[1] D. H. Lee, "Personalizing information using users' online social networks: a case study of CiteULike," *Journal of Information Processing Systems*, vol. 11, no. 4, pp. 1-21, 2015.

[2] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Canada, 2002, pp. 61-70.

[3] J. Leskovec and E. Horvitz, "Planetary scale views on a large instant messaging network," in *Proceedings of the 17th International Conference on World Wide Web*, Beijing, China, 2008, pp. 915-924.

[4] Y. Wang, G. Cong, G. Song, and K. Xie, "Community based greedy algorithm for mining top-k influential nodes in mobile social networks," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington DC, 2010, pp. 1039-1048.

[5] M. Kimura and K. Saito, "Tractable models for information diffusion in social networks," in *Proceedings of 10th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Berlin, Germany, 2006, pp. 259-271.

[6] M. C. Yang and H. C. Rim, "Identifying interesting Twitter contents using topical analysis," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4330-4336, 2014.

[7] N. Wang, Q. Sun, Y. Zhou and S. Shen, "A study on influential user identification in online social networks," *Chinese Journal of Electronics*, vol. 25, no. 3, pp. 467-473, 2016.

[8] E. Keller and J. Berry, *The Influentials: One American in Ten Tells the Other Nine How to Vote, Where to Eat, and What to Buy*. New York, NY: The Free Press, 2003.

[9] J. Lv, J. Guo, and H. Ren. "Efficient greedy algorithms for influence maximization in social networks," *Journal of Information Processing Systems,* vol. 10, no. 3, pp. 471-482, 2014.

[10] J. Lu and W. Wan, "Identification of key nodes in microblog networks," *ETRI Journal*, vol. 38, no. 1, pp. 52-61, 2016.
[11] H. Kwak, C. Lee, H. Park H and S. Moon, "What is Twitter, a social network or a news media?" in *Proceedings of the 19th International Conference on World Wide Web*, Raleigh, NC, 2010, pp. 591-600.
[12] R. Wang, W. Zhang, H. Deng, N. Wang, Q. Miao, and X. Zhao, "Discover community leader in social network with PageRank," in *International Conference in Swarm Intelligence*. Heidelberg: Springer, 2013, pp. 154-162.
[13] L. L. Yu, S. Asur, and B. A. Huberman, "Artificial inflation: the real story of trends and trend-setters in Sina Weibo," in *Proceedings of International Conference on Social Computing*, Amsterdam, the Netherlands, 2012, pp. 514-519.
[14] W. Zhang, G. Sun, and S. Bin, "An opinion leaders detecting algorithm in multi-relationship online social networks," *International Journal of Hybrid Information Technology*, vol. 9, no. 5, pp. 391-398, 2016.
[15] G. A. Wang, J. Jiao, A. S. Abrahams, W. Fan, and Z. Zhang, "ExpertRank: a topic-aware expert finding algorithm for online knowledge communities," *Decision Support Systems*, vol. 54, no. 3, pp. 1442-1451, 2013.
[16] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring user influence in twitter: the million follower fallacy," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, Washington, DC, 2010, pp. 30-38.
[17] D. Tunkelang, "A Twitter analog to PageRank," 2009 [Online]. Available: http://thenoisychannel.com/2009/01/13/a-twitter-analog-to-pagerank.
[18] A. Leavitt, E. Burchard, D. Fisher, and S. Gilbert, *The Influential: New Approaches for Analyzing Influence on Twitter*. Web Ecology Project, 2009.
[19] E. Bakshy, J. H. Hofman, and J. D. Watts, "Everyone's an influencer: quantifying influence on twitter," in *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, Hong Kong, 2011, pp. 65-74.
[20] N. Nguyen, T. Ho, and P. Do, "Finding the most influential user of a specific topic on the social networks," *Advances in Computer Science: An International Journal*, vol. 4, no. 2, pp. 31-40, 2015.
[21] M. A. Zia, Z. Zhang, X. Li, H. Ahmad, and S. Su, "ComRank: joint weight technique for the identification of influential communities," *China Communications*, vol. 14, no. 4, pp. 101-110, 2017.
[22] Sina Weibo [Online]. Available: http://www.weibo.com.
[23] KDD Cup [Online]. Available: https://www.kaggle.com/c/kddcup2012-track1#description.
[24] Y. Liu and Y. Zhou, "Social media in China: rising Weibo in government," in *Proceedings of the 5th IEEE International Conference on Digital Ecosystems and Technologies Conference (DEST)*, Daejeon, Korea, 2011, pp. 213-219.
[25] M. A. Zia, Z. Zhang, G. Li, H. Ahmad and S. Su, "Prediction of rising venues in citation networks," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 21, no. 4, 2017.
[26] M. A. Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in *Proceedings of SDM-2006 Workshop on Link Analysis, Counter-Terrorism And Security*, Bethesda, MD, 2006.

Muhammad Azam Zia

He received the M.S. degree in computer science from University of Agriculture, Faisalabad (UAF), Pakistan and serving as a faculty member to UAF as well. Currently, he is a PhD candidate in the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. His research interests include social network and data mining.

Zhongbao Zhang

He received the Ph.D. degree in Computer Science from Beijing University of Posts and Telecommunications, China, in 2014. He is an assistant professor in Beijing University of Posts and Telecommunications now. His major is Computer Science. His research interests include network virtualization, social network and big data.

Liutong Chen

He is an undergraduate degree student in the Beijing University of Posts and Telecommunications. His major is Telecommunications Engineering with Management. His research interests include social network and IOT.

Haseeb Ahmad

He received his Master degree in computer science from Virtual University of Pakistan. Currently, he is a PhD candidate in the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. His research interests include data mining and information security.

Sen Su

He received the Ph.D. degree in Computer Science from the University of Electronic Science and Technology, China, in 1998. He is currently a Professor at the Beijing University of Posts and Telecommunications. His research interests include cloud computing and big data.