World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:5, No:12, 2011

TFRank: An Evaluation of Users Importance with Fractal Views in Social Networks

Fei Hao and Hai Wang

Abstract—One of research issues in social network analysis is to evaluate the position/importance of users in social networks. As the information diffusion in social network is evolving, it seems difficult to evaluate the importance of users using traditional approaches. In this paper, we propose an evaluation approach for user importance with fractal view in social networks. In this approach, the global importance (Fractal Importance) and the local importance (Topological Importance) of nodes are considered. The basic idea is that the bigger the product of fractal importance and topological importance of a node is, the more important of the node is. We devise the algorithm called TFRank corresponding to the proposed approach. Finally, we evaluate TFRank by experiments. Experimental results demonstrate our TFRank has the high correlations with PageRank algorithm and potential ranking algorithm, and it shows the effectiveness and advantages of our approach.

Keywords—TFRank, Fractal Importance, Topological Importance, Social Network

I. Introduction

Social networks services are becoming more and more vital for information sharing, and propagation on the web. Many researchers start to pay more attention to social network analysis, such as the features of social networks, and the users' behavior in social networks. One of crucial issues in social network analysis is the problem of extracting the most important nodes, *i.e.*, evaluating the node importance in social networks. Therefore, finding the vital nodes in the networks is helpful to system science, social network and web-based search. For example, in the terrorist network, finding the most vital nodes can help to locate the head of the criminals; in the WWW, evaluating the importance of the pages can help to find the page which is most correlative to the subject [1]; in social marketing, finding the most influential nodes can help company to promote their new products efficiently; in the network of disease and virus, protecting and isolating the vital nodes according to their importance can efficiently restrict the diseases spreading.

There have been a number of related works on evaluation of nodes importance in the network. Katz, Freeman et.al have done a lot about the social network, focusing on enlarging the differences among the nodes to differentiate the importance, related methods such as degree ranking, betweenness ranking, closeness ranking [2], [3]. In web science field, the algorithm PageRank [4], HITS [5] which have been successfully applied to various search engines. In social networking services field, Han et al.[6] proposed a user evaluation algorithm for user-generated video sharing website-YouTube. He et al. [1] defined and calculated the topological potential score of each node with the concept of field, they obtained a more accurate global ranking which can reflect nodes importance in the network. Yang et al. [7] proposed the fractal views to construct a visual abstraction of a large and complex social networks with users selected social actors as focuses. It is the first work to investigate the structure of terrorist social networks. Unfortunately, they did not study the nodes importance. However, we think that modeling the structure of social network and information diffusion in social networks with fractal views is a quite innovative and fresh idea. Hence, we employ this idea and devise the *TFRank*, a novel evaluation approach of nodes (users) importance by considerations of global ranking and local ranking of nodes (users).

In this paper, our contribution are twofold: 1) First, we propose the *TFRank*, an evaluation approach of user importance in social networks. 2) Second, we evaluate proposed algorithm by case studies. Experimental results show the effectiveness and advantages of our approach.

The remainder of this paper is organized as follows: the next section provides the preliminaries and formulation of problem definition; Section III presents the our nodes importance evaluation approach–*TFRank*. Case studies are shown in Section IV. Section V concludes the work.

II. Preliminary

In this section, we firstly formalize the definition of social network and give a specific social network. Then, the problem statement is described.

A. Formulism of Social Network

A social network is modeled as an undirected graph $SN = (V, E)$, V indicates the users in the network and edges E indicates the relationship between users. For example, in our case studies section, we study two social graphs where vertices are club members of karate club network and dolphins in bottlenose dolphin social network. And, there is an edge between two vertices if the two corresponding club members or dolphins have an interaction.

There are many social network-based software applications bring a new way of information propagation and sharing. Flickr, for instance, allows the sharing of photos, del.icio.us the sharing of bookmarks, CiteULike and Connotea the sharing

Fei Hao is with the Department of Computer Science, Korea Advanced Institute of Science and Technology, Daejeon 305-701 Republic of Korea, e-mail: fhao@kaist.ac.kr

Hai Wang is with Department of Computer Science and Engineering, Pohang University of Science and Technology, Pohang 790-784, Republic of Korea, haiwang@postech.ac.kr

Manuscript received April 19, 2005; revised January 11, 2007.

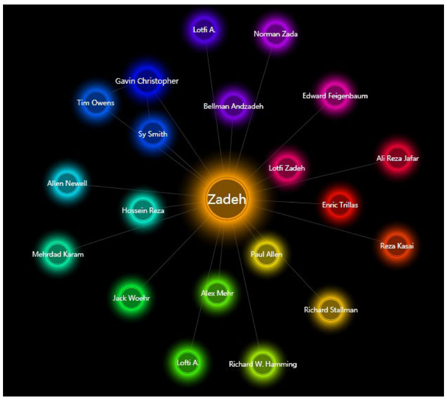

Fig. 1. Social Network–*"Renlifang"*

of bibliographic references and Last.fm the sharing of music listening habits. Particularly, "*Renlifang*" search [1] is a kind of search engine based on social network. In Renlifang social network, each node denotes a user, the relationship is built up according to their shared contents in the web (Shown in Figure 1).

B. Problem Statement

The problem in this paper takes the social network G as the input and rank the nodes in terms of nodes importance. In fact, there are several existing methods to evaluate the nodes importance in social networks. However, they cannot cope with the propagation dynamics and structure of social network. To solve this problem, a solution framework called "*TFRank*" is proposed. *TFRank* contains three technical steps (Figure 2): 1) computation of topological importance of nodes; 2) computation of fractal importance of nodes; 3) ranking the nodes by the multiplication between topological importance of nodes and computation of fractal importance of nodes. The following section studies the *TFRank* in details.

III. TFRank

As we mentioned before, *TFRank* is an evaluation approach with considerations of local importance (Topological Importance) and global importance (Fractal Importance) of Nodes. Figure 2 depicts the basic technical route of *TFRank*. A social graph is our input, then, we calculate the topological importance and fractal importance for each node in the social graph. Finally, we devise the approach of *TFRank* value by multiplying them linearly. In this section, we present the detailed approaches to calculate the topological importance and fractal importance of nodes. Then, *TFRank* is described by combining two values.

A. Fractal Views

Fractal views are information visualization techniques using an information reduction approach. They control the amount of information displayed by focusing on the syntactic structure of the information. The fractal views developed by Koike [8]

[1] http://entitycube.research.microsoft.com/

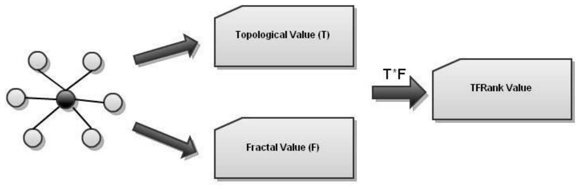

Fig. 2. The Technical Route of TFRank

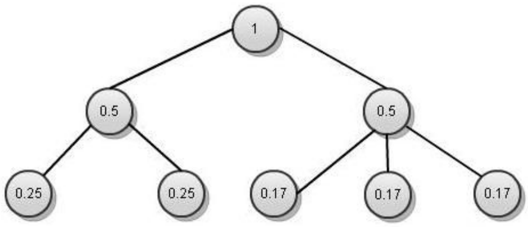

Fig. 3. Computation of fractal values in a tree structure

were applied on tree structures. It controls the number of displayed nodes without relation to the shape of trees [9].

Definition 1: (**Fractal Value of Node in Un-weighted Tree**) Given a tree T, the fractal values of other nodes are propagated from their parents as follows:

$$F_{childofx} = r_x \times F_x \tag{1}$$

where $F_{childofx}$ and F_x are the fractal values of the child of node x and node x,

$$r_x = CN_x^{-1/D} \tag{2}$$

where C is a constant, D is the fractal dimension, and N_x is the number of children of node x.

Figure 3 shows an example of fractal value computation. For simplicity, C and D are set to 1. The fractal value of the root node is 1. The fractal values of the child nodes of the root node are both 0.5. The fractal values of the first two nodes on the third level are 0.25 and the last three nodes on the third level are 0.17.

Definition 2: (**Fractal Value of Node in Weighted Tree**) Given a tree T, a focus node o, the fractal value of the focus o is 1, i.e., $F_o = 1$, the fractal values of other nodes are propagated from their parents as follows:

$$F_{childofx} = r_x \times F_x \tag{3}$$

where $F_{childofx}$ and F_x are the fractal values of the child of node x and node x,

$$r_x = \left(\frac{w_{cp}}{\sum_{c' \in children_of(p)} w_{c'p}}\right)^{-1/D} \tag{4}$$

where c is a child of p; w_{cp} denotes the association weight between c and p. D is the fractal dimension. The fractal values are normalized so that the sum of the fractal values of all the children equals the fractal value of the parent. The association weights are taken into account so that a parent node will propagate more fractal values down to the child nodes which are more strongly associated with the parent.

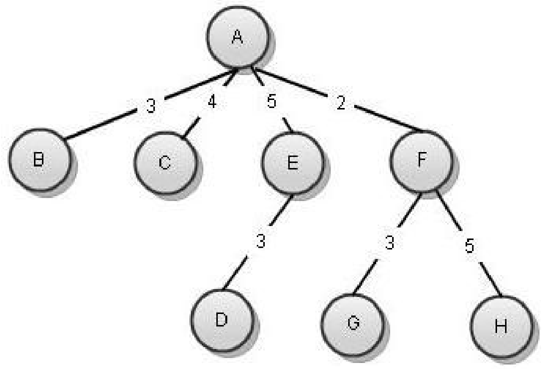

Fig. 4. A weighted tree

Figure 4 presents a weighted tree with node A as a root node (focus).

We can easily to calculate the fractal values of nodes from A to H with node A as the focus.

$F_A = 1$
$F_B = \frac{4+5+2}{3+4+5+2} * \frac{1}{4-1} = 0.26$
$F_C = \frac{3+5+2}{3+4+5+2} * \frac{1}{4-1} = 0.23$
$F_D = \frac{3+4+2}{3+4+5+2} * \frac{1}{4-1} = 0.21$
$F_E = \frac{3+4+2}{3+4+5+2} * \frac{1}{4-1} = 0.21$
$F_F = \frac{3+4+5}{3+4+5+2} * \frac{1}{4-1} = 0.28$
$F_G = \frac{5}{3+5} * \frac{1}{4-1} = 0.175$
$F_H = F_F - F_G = 0.105$

In the un-weighted tree, only the degree of the parent node is considered in the fractal value propagation. In the weighted tree, the weights on the links correspond to the strength of association. As a result, we integrate such factors in the fractal value propagation formulation so that child nodes with stronger weights will be propagated with high fractal values than their siblings with lower weights instead of even propagation of fractal values.

B. Fractal Importance of Nodes

The fractal importance of nodes is a kind of approach with considerations of the fractal values of each children nodes. It is a global node importance approach. The fractal importance of node A ($\widetilde{F}_A$) is defined as follows,

Definition 3: (**Fractal Importance of Node**) Given a tree T, A is the root node of T,

$$\widetilde{F}_A = \sum_{i=1}^{L} \sum_{j=1}^{N(Level_i)} (F_{ChildNodes_i^j}) * (\frac{1}{2})^i \tag{5}$$

where $Level_i$ denotes the i_{th} level in T, $N(Level_i)$ is the number of nodes in the i_{th} level, L is the total level of T. $F_{ChildNodes_i^j}$ means the fractal value of the j_{th} node in the i_{th} level.

Let's take the Figure 4 as an example with A as the focus.

1st level: $(F_B + F_C + F_E + F_F) \times 0.5 = 0.5$
2nd level: $(F_D + F_G + F_H) \times 0.25 = 0.1225$

Consequently, the fractal importance of node A is calculated as follows,

$$\widetilde{F}_A = 0.5 + 0.1225 = 0.6225. \tag{6}$$

C. Topological Importance of Nodes

Roughly, topological value is a kind of measurement for local node importance. The topological importance of node A (T_A) is defined as follows,

Definition 4: (**Topological Importance of Node**) Given a tree T, A is the root node of T,

$$\widetilde{T}_A = \sum_{i=1}^{L} N(Level_i) * (\frac{1}{2})^i \tag{7}$$

where $Level_i$ denotes the i_{th} level in T, $N(Level_i)$ is the number of nodes in the i_{th} level, L is the total level of T.

Let's take the Figure 4 as an example with A as the focus. $\widetilde{T}_A = 4 \times 0.5 + 3 \times 0.25 = 2.75$.

Obviously, the topological importance of node is not related to fractal values. It is just calculated by the structure of the tree. Intuitively, for node A, A can propagate the information to B, C, E, F, G, H. However, the nodes B, C, E, F locate in the first level of the tree, and the nodes D, G, H locate in the second level of the tree. It means as the information propagating from the root node A to its children nodes, the capacity of information will decay.

Up to now, if a node is given, then we can obtain two types of importance of nodes: 1) topological importance of node; and 2) fractal importance of node. They have different measurement results in term of local importance and global importance of nodes. Therefore, we attempt to combine these two types of importance of nodes together and propose the *TFRank* Algorithm for evaluation of node importance in social networks.

D. The Decision Level of a Tree

In the previous section, we have discussed about the root node's fractal importance and the topological importance. But in the real situation, the network is always very complex, hence the shortest distance tree has many levels. Intuitively, in the procedure of information propagation, the person only can influence the one who is near to him, so when consider the importance of a node, there is no necessary for us to consider all the nodes in different levels in a tree. According to the level of a tree, we can decide how many levels we should consider.

Definition 5: (**TF-Level**) Given a tree T,

$$L^{TF} = \begin{cases} L, & \text{L} \leq 3 \\ \lfloor ln(10L+20) \rfloor, & \text{otherwise.} \end{cases} \tag{8}$$

where L denotes the total level number in T, the TF-Level, we denote as L^{TF} which is the number of levels in the tree we should consider when calibrating the fractal importance and topological importance.

To better understand the construction reasons for Definition 5, we study the correlation between the number of levels (L) of tree T, and the number of decision levels (L^{TF}) we should consider in tree T. Figure 5 shows the correlation between L

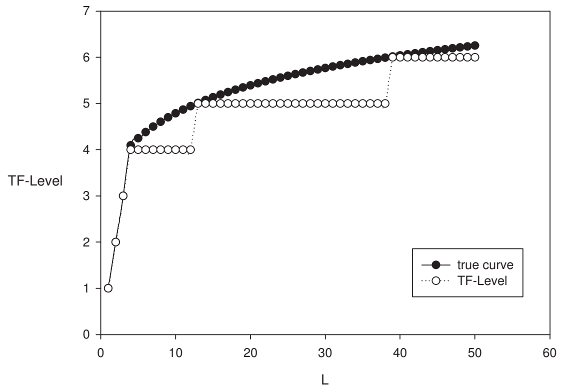

Fig. 5. The correlation between L and L^{TF}

and L^{TF}. Obviously, as the L increases, the L^{TF} falls into a certain range due to our given assumption, i.e., the person only influences the one who is close to him within a given range. Actually, our assumption reduces the complexity when the scale of the social network increases.

E. TFRank Algorithm Description

Basically, the basic idea of *TFRank* is that making the multiplication operation between topological importance of nodes and fractal importance of nodes. However, one critical thing is to convert a social network (social graph) to a tree structure (weighted tree) which is suitable for further calculations. As shown in Figure 2 before, the steps of *TFRank* algorithm (Shown in Algorithm 1) is listed as follows,

- Convert a social network to a weighted tree; We adopt the approach in [10] to generate the tree structure by finding the shortest paths from the focus to every other node in the network using the shortest path algorithm, such as Dijkstra and Floyd algorithms (Line 4-5).
- Calculate the fractal importance (Section 3.2, Line 11) and topological importance (Section 3.3, Line 12) of each node according to a converted weighted tree;
- Make the multiplication operation with topological importance and fractal importance of nodes (Line 13).
  Definition 6: (**TFRank Score of node**) Give a social network G, the converted tree structure T from G, we assume node A as the focus (root node) in T, the TFRank score of A can be described as

$$TF_A = \widetilde{T}_A \times \widetilde{F}_A \tag{9}$$

Algorithm 1 TFRank Algorithm

1: **Input:** Node V, Social Graph G
2: **Output:** Ranking List
3: **while** (V in G)
4: $\quad Floyd(G, V)$
5: $\quad T = Tree(G, i)$
6: $\quad L = GetLevel(T)$;
7: $\quad L^{TF} = Calculate(L)$;
8: $\quad$ **while** (V in T)
9: $\qquad$ F(V);
10: $\quad$ **end while**
11: $\quad \widetilde{F}_V = \sum_{i=1}^{L^{TF}} \sum_{j=1}^{N(Level_i)} (F_{ChildNodes_i^j}) * (\frac{1}{2})^i$
12: $\quad \widetilde{T}_V = \sum_{i=1}^{L^{TF}} N(Level_i) * (\frac{1}{2})^i$
13: $\quad TF_V = \widetilde{T}_A \times \widetilde{F}_A$
14: **end while**
15: Sort(TF_V)

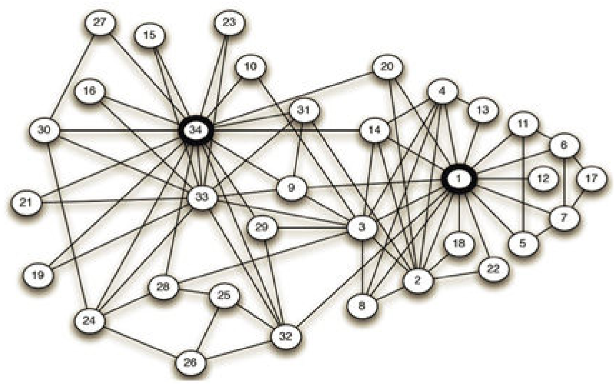

Fig. 6. Visualization of Karate club network

IV. Case Studies

In this section, we evaluate *TFRank* algorithm on two real-world networks and compare it with some other traditional approaches such as degree ranking (DR), betweenness ranking (BR), closeness ranking (CR), PageRank (PR) algorithm and potential ranking (PoR) etc.

A. Zachary's Karate Club Network

Zachary's karate club network is a classic evaluation dataset in social network analysis. In the course of two years in the early 1970s, Wayne Zachary observed social interactions between the members of a karate club at an American university. He constructed networks of ties between members of the club based on their social interactions both within the club and away from it. Figure 6 shows the social networks structure of Karate club.

Table I shows the top 10 vital nodes using different methods for evaluation of users importance. From Table I, the PageRank algorithm, potential rank and *TFRank* essentially have degree bias, roughly in proportion to nodes' degree. But when nodes have the same degree, degree ranking cannot distinguish them at all, e.g, V_9, V_{14}, V_{24} have the same degree 5, and V_4, v_{32} have the same degree 6. However, PageRank algorithm, potential rank and *TFRank* can further analyze this situation, preferring V_{32} to V_4. As for V_9, V_{14}, V_{24}, PageRank algorithm and *TFRank* evaluate V_{24} the most vital one while potential rank evaluate V_9.

B. Bottlenose Dolphins Network

Bottlenose dolphin network is also a very classic social network. The network was compiled by D. Lusseau from

Rank	DR	BR	CR	PR	PoR	**TFRank**
1	34	1	1	34	34	34
2	1	34	3	1	1	1
3	33	33	34	33	33	33
4	3	3	32	3	3	3
5	2	32	9	2	2	2
6	32	9	14	32	32	32
7	4	2	33	4	4	4
8	24	14	20	24	9	24
9	14	20	2	9	14	9
10	9	6	4	14	24	14

TABLE I
TOP 10 VITAL NODES OF THE KARATE'S CLUB NETWORK. DEGREE RANKING (DR), BETWEENNESS RANKING (BR), CLOSENESS RANKING (CR), PAGERANK (PR), AND POTENTIAL RANKING (POR)

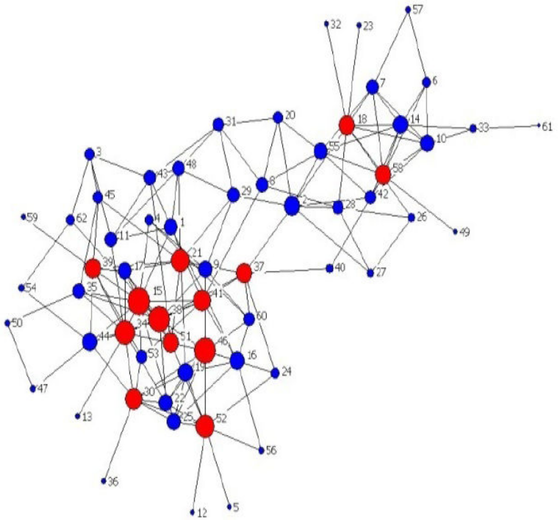

Fig. 7. Visualization of Bottlenose dolphins network

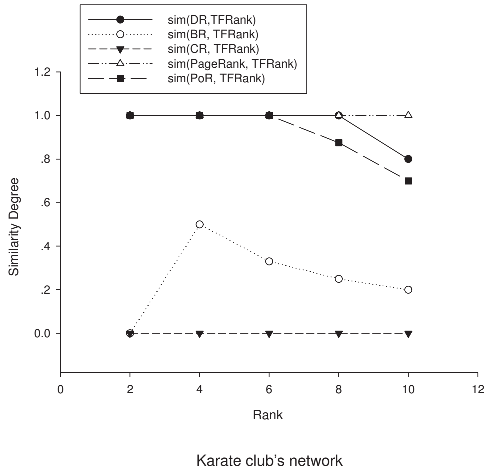

Fig. 8. The similarity degrees of users ranking in Karate club's network

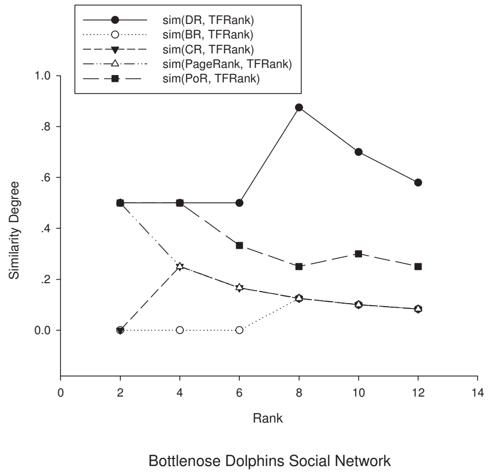

Fig. 9. The similarity degrees of users ranking in Bottlenose dolphins social network

seven years of field studies of the dolphins, with ties between dolphin pairs being established by observation of statistically significant frequent association. Figure 7 is the visualization of bottlenose dolphins network.

Here, we also evaluate our ranking algorithm with other related ranking algorithms. Table II shows the top 13 vital nodes. Table II shows that both PageRank and degree ranking have the similar ranking, only one node different. But the *TFRank* varies a lot, which considers V_{37} and V_{41} more vital. According to Lusseau's observation, V_{37} is a vital connector, as during the absence of V_{37}, two families of dolphins' contact quickly decreased; and when V_{37} was back, the previous close contact recovered. V_{37}'s betweenness and closeness score is the largest, and it doesn't appear in the top 13 lists of degree ranking and PageRank, which may indicate that *TFRank* has its own advantage to reflect the user importance in the network.

Rank	DR	BR	CR	PR	PoR	**TFRank**
1	15	37	37	15	15	15
2	38	2	41	18	38	46
3	46	41	38	52	46	38
4	34	38	21	58	34	34
5	52	8	15	38	21	52
6	18	18	2	46	52	30
7	21	21	8	34	30	21
8	30	55	29	30	41	58
9	58	52	34	14	18	18
10	2	58	9	2	37	41
11	14	40	51	21	58	39
12	39	29	1	39	51	2
13	41	30	46	10	39	14

TABLE II
TOP 13 VITAL NODES OF THE BOTTLENOSE DOLPHINS NETWORK. DEGREE RANKING (DR), BETWEENNESS RANKING (BR), CLOSENESS RANKING (CR), PAGERANK (PR), AND POTENTIAL RANKING (POR)

C. Results Discussions

In this section, we compare the similarity of various ranking algorithms in terms of size of top-k users. We utilize the Jaccard similarity approach to calculate the similarity of various ranking algorithms in terms of k.

Figure 8 shows the similarity degree of users ranking in Karate club's network. Obviously, our algorithm *TFRank* has the highest similarity with PageRank algorithm. It has the smallest similarity with closeness ranking algorithm.

Figure 9 shows the similarity degree of users ranking in Bottlenose dolphins social network. From the Figure 9, we can see that our algorithm *TFRank* has the highest similarity with Degree ranking algorithm. It has the smallest similarity with betweenness ranking algorithm.

V. Conclusions

Finding the key users from social networks is becoming more and more important. It is beneficial to many research applications, such as social marketing, social advertising. It is also a hot research issue in the fields of complex networks, social network analysis and graph-based data mining. This paper mainly proposes a novel users importance evaluation algorithm *TFRank* with fractal views. The *TFRank* considers the global importance and local importance of users together. By defining and computing the *TFRank score* of each node, we can obtain a more accurate global ranking which reflects the users importance in social networks.

References

[1] N. He, W. Y. Gan, D. Y. Li, Evaluate nodes importance in the network using data field thoery, 2007 International Conference on Convergence Information Technology, pp.1225-1230, 2007.
[2] Freeman L. C. Centrality in social networks: I. Conceptual clarification, Social Networks, 1, pp.215-239, 1979.
[3] K. Musial, P. Kazienko, P. Brodka, User position measures in social networks, The 3rd SNA-KDD Workshop, 2009.
[4] S. Brin and L. Page, The anatomy of a large-scale hypertextual web search engine, Computer Networks, 30, pp.107-117, 1998.
[5] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5), 1999.
[6] Y. S. Han, L. Y Kim, J. W. Cha, Evaluation of user reputation on YouTube, LNCS 5621, pp.346-353, 2009.
[7] C. C. Yang, M. Sageman, Analysis of terrorist social networks with fractal views, Journal of Information Science, 35(3), pp.299-320, 2009.
[8] H. Koike, Fractal views: a fractal-based method for controlling information display, ACM Transactions on Information Systems 13(3), 1995.
[9] H. Koike and H. Yoshihara, Fractal approaches for visualizing huge hierarchies. In: Proceedings of IEEE Symposium on Visual Laugnagues, 1993.
[10] J. Xu and H. Chen, Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks, Decision Support Systems, 38, pp.473-487, 2003.

Fei Hao received the B.S. and M.S. degrees in school of Mathematics and Computer Engineering from Xihua University, Chengdu, China, in 2005 and 2008, respectively. He is currently working toward the PH.D degree in the Department of Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. He has published 20 research papers in International and National Journals as well as conferences. His research interests include intelligent information processing, social computing and time series data mining.

Hai Wang was born in 1988 in Hubei,China. In 2011 he entered Pohang University of Science and Technology to pursue a PH.D degree, where he is studying at department of computer science. Prior to this, he studied at Wuhan University in China. In his B.S. period, he published several papers and applied some high quality patents.