Identifying influential users by improving LeaderRank algorithm

Yong Yao [a,*], Wenxiao Wu[a]

[a]*School of Computer Science and Technology, Xi'dian University, China*

Abstract

In large-scale social network, influential users play important roles in public opinion analysis and information promotion. So identifying influential users attract increasing attention in social network studies, and researchers have been trying to find a better algorithm to identify influential users. LeaderRank is an effective algorithm to identify the influential users. The paper analyzes the shortcomings of LeaderRank, proposes new algorithms based on LeaderRank with user's multiple properties, and identifies that the accuracy of the improved algorithms has been effectively improved.

Keywords: Social network, Influential users, Data models, LeaderRank

1. Introduction

In complex network community,each user's activity and his feedback are different,and all users play different roles in the community [1]. Among them, some users would frequently receive information and ideas, and make meaningful information and views which have a great impact on other users [2].These users are called ”opinion leaders”. The information that opinion leaders spread can influence the other parts of the network, and guide their behavior. Therefore, if we use some information to influence the opinion leaders, it may be efficient to affect the other users in the social network. So opinion leader mining has important theoretical and practical significance on the information promotion

*Corresponding author

and public opinion monitoring ,which is important to form a stable community network [3, 4].

In recent years, Zhihu.com has become the largest Chinese online Q&A community. And it has become a professional and universal platform to express public opinion. Among the community, different users play different roles. Some users act as content consumers, and some users act as content producers. Some of these users are the center of the community users. They have a lot of fans, and they can also be considered as leaders. Different from other popular social networks, the definition of opinion leaders in Zhihu.com may contain massive and high qualitied answers. The specific performance of these users is that they receive a number of approvals and thanks from other users. Under their leadership, the discussion about the public opinion will have a tendency. Their personal answers can be considered as the main views, and can be quickly spread through the whole fans, and their approval about other answers can make these answers second spread, so they can make the public opinion develop in a more concentrated direction.

LeaderRank is an effective opinion leader mining algorithm,which is improved based on PageRank algorithm [5, 6]. In this paper, we designed a new algorithm Zhihu-Rank based on LeaderRank,combining with the personal multiple attributes of the users in the social network. The algorithm improved the accuracy of opinion leader mining.

2. Related Theory Knowledge

2.1. Related Research

With the development of social network, more and more researchers concentrate on opinion leader mining. Aral 's et al [7], proved that opinion leaders played a key role in the information dissemination of complex networks. Bai et al [8, 9],thought that we could consider the nodes with the highest degree as the opinion leaders, and could control the information dissemination in the network by controlling these leader nodes. This method has been applied to a variety of complex networks to do opinion leaders mining [3, 10, 11].

Kitsak et at [12],stated that the dissemination ability of nodes in the network had much to do with the location of the nodes. In the core location of the network, even if its degree is small, it also has high impact, and vice versa. At the same time, other researchers in the field of opinion leaders mining have put forward a large number of other indicators one after another [13]. For example, closeness centrality [14], eigenvector centrality [15], degree centrality [16, 17] and between centrality [18]. These indicators are mainly used to measure the dissemination ability of a node in the social network [19]. Song [20] drew lessons from PageRank algorithm, and combined with the analysis of sentiment analysis between users, the implicit relationship of comments, the time decay of comments and other factors to mine the opinion leaders in complex networks. But PageRank is not suitable for opinion leader mining in complex networks with fast changing structure .

In view of the shortcomings of PageRank in the field of opinion leader mining, Lv et at [5]. state LeaderRank algorithm for opinion leader mining. LeaderRank is more accurate than PageRank, and more stable in face of noise and malicious attacks.

At home and abroad, the research of opinion leader mining concentrates more on the relationship between the nodes [21, 22, 23]. And the emphasis of the research is about the improvement based on PageRank, including the weight of edges and nodes. But there are still some improvements on LeaderRank.For example, the algorithm can takes into account the users' personal attributes in social network, especially the impact of users' multiple attributes on opinion leader mining. These factors can also restrict the further enhance of the performance about the algorithm. This paper will make improvements in these areas.

2.2. PageRank

PageRank considers the value of the link as a ranking factor.It uses the node score as a standard of the node influence at different moments. The process can be accumulated by the following formula:

$$PR_i(t+1) = d * \sum_{j \in M} \frac{PR_j(t)}{k_j^{out}} + \frac{(1-d)}{N} \tag{1}$$

In the formula, $PR_i(t)$ is the score of node i at time t, and k_j^{out} denotes the out-degree, i.e. the number of followees,of node j. N is the number of nodes. The parameter d is introduced as damping factor, and it represents the probability that a node randomly selects another node. And it selects a new node to browse with the probability 1-d. Generally, d = 0.85.

Along with the development of social networks, the network structure changes more and more rapidly. And the jump probability value of PageRank c should adjust the changes of the network structure. For a network with particular structure, we should train more times in order to get the best value of the parameter c. This method is not applicable to the network whose structure changes frequently. And, most social networks are unconnected graphs. PageRank can not guarantee the convergence about the unconnected graphs. The above two reasons lead to the limit of PageRank on opinion leader mining.

2.3. LeaderRank

In view of the shortcomings of PageRank, Lv proposed LeaderRank. The main improvement of LeaderRank is that it adds a ground node in the network, and lets the ground node connect to all the other nodes in the network. The network G(N,M) with M edges and N nodes becomes the network G(N+1,M+2N) with M+2N edges and N+1 nodes.

The ground node has the following effects:

a. The ground node connects to all the other nodes in the network, so the network becomes a connected graph, which can ensure the convergence of

LeaderRank algorithm. And the addition of the ground node can reduce the radius of the whole network, and speed up the convergence.

b. LeaderRank algorithm no longer needs the parameter d. Because the number of information sources of a node is inversely proportional to the number of the node i which flows to the ground node.As the network structure changes, different nodes will have different jump probability.

The core formulas of LeaderRank algorithm are as follows:

$$LR_i(t+1) = \sum_{j \in R} \frac{LR_j(t)}{k_j^{out}} \tag{2}$$

$$F\_LR_i = LR_i(t_c) + \frac{LR_g(t_c)}{N} \tag{3}$$

R is the set of the followees of node i. $LR_i(t)$ is the score of node i at time t, t_c is the time of convergence. $LR_g(t_c)$ is the score of ground node at time t_c. F_LR_i is the final score of node i.

Lv et at, showed that LeaderRank algorithm had higher accuracy and stronger stability than PageRank.

3. Improvement Of LeaderRank

3.1. Analysis of the Shortcomings of LeaderRank

In real social networks, the users' influence is not only determined by the their network relationship , but also determined by the attributes which the users accumulate in the development of the social network. Due to the complexity of the social network, users would get a variety of different properties. LeaderRank algorithm doesn't consider the impact of the user' various personal attributes. Therefore, the improved LeaderRank in this paper combined with a variety of users' personal attributes, and used them as calculation factors in algorithm.

For example, in a real social network, user A has fewer fans, and user B has more fans.But user A gets more approvals and thanks when they answer questions in Zhihu.com. These approvals and thanks are not only from their fans, but also other users. It is obvious that user A has more influence than user B. However, if we use traditional LeaderRank to evaluate the user influence, the result must be that user B rank higher than user A, so the result is not consistent with the actual situation.

3.2. Improvement of LeaderRank

3.2.1. NiceRank

In the description above, We can see that, when evaluating the influence of the nodes, PageRank and LeaderRank only consider the topology properties of the nodes in social network, including the nodes' out-degree and in-degree. But the algorithms ignore the nodes' personal attributes. Thus, we improved LeaderRank algorithm by combining with the nodes' personal attributes which are accumulated in the development of social network. In the process of the development of the social network, the number of approvals and thanks can show that the answers of the users are popular, and also show that they have higher influence. So we improve the LeaderRank by combining with users' personal attributes, including the number of users' approvals and thanks, and let the improved LeaderRank be more close to the reality.

In the improved LeaderRank algorithm, we define the user's personal properties as Nice value. We consider that the number of approval and thank is in the same dimension, and there is no direct relationship between these attributes, and both attributes can effect the user's influence, so the Nice values is defined as the sum of the two attributes. In the improved algorithm, we combine two attributes into one property – Nice value, and we call it NiceRank, the specific steps are as follows:

a. Initialization : The $NR_i(t)$ is defined as the score of the node i at time t, and $NR_t(0) = 1$for all nodes i;

b. Iteration : In the NiceRank algorithm, the NR value of a node is based on the number of its followers and the Nice value of the node. Node j is one of the followers of node i. The Nice(i) is the Nice value of node i, and it is the sum of the number of approvals and thanks. U is the set of the followees of node j. So the NR value of node i at time t + 1 can be calculated by the following formula:

$$
\begin{aligned}
NR_i(t+1) = \\
&(1-\alpha) * \sum_{j \in R} \frac{NR_j(t)}{k_j^{out}} + \\
&\alpha * \sum_{j \in R} \frac{Nice(i)}{\sum_{q \in U} Nice(q)} * NR_j(t)
\end{aligned} \tag{4}
$$

c. Termination : If the number of iteration reaches a certain threshold, or the score becomes a stable value, the iteration will finish. Finally, we can use the following formula to calculate the final score:

$$
F\_NR_i = NR_i(t_c) + \frac{NR_g(t_c)}{N} \tag{5}
$$

3.2.2. Analysis of the Shortcomings of NiceRank

In the description above, NiceRank considers not only the topology properties of the nodes in the social network, but also the personal multiple attributes of the nodes. But in the actual social network, the number of approvals and thanks can not really represent the users' influence. For example, if user A answers a large number of questions, and his every answer gets a small amount of approvals and thanks.Because the totality is large, so his Nice value is very large. If we use NiceRank to evaluate the users' influence, user A will rank high. This situation is obviously not consistent with the practical significance.

3.3. Improvement of NiceRank

3.3.1. AveNiceRank

Through the description above, we can see that NiceRank only considers the totality of approval and thank, and it can not meet the real situation of Zhihu, so we need to define a new factor to represent the users' real performance in the social network. By analyzing the characteristics of Zhihu data, we decide to define ANice value, which represents the average-nice value. The value can represent the average performance of the users in each question that they answered, and also represent the actual performance of the user in Zhihu network. The higher the value, the more it can show that the user is getting more approvals and thanks than other users, and it can be a factor in accumulating the influence of the node. In the actual Zhihu data, we use the number of answers and the Nice value to calculate the ANice value. By combining the ANice value, we improve NiceRank algorithm, and we name it AveNiceRank. The specific steps are as follows:

a. Initialization : The $ANR_i(t)$ is defined as the score of the node i at time t, and $NR_i(0) = 1$ for all nodes i;
b. Iteration : In AveNiceRank, ANR value of a node is based on the number of its followers and the ANice value of the node. Node j is one of the followers of node i. ANice(i) is the ANice value of node i, and it is the average value of the nice value of node i. U is the set of the followees of node j. So the NR value of node i at time t + 1 can be calculated by the following formula:

$$\begin{aligned} ANR_i(t+1) = \\ &(1-\alpha) * \sum_{j \in R} \frac{ANR_j(t)}{k_j^{out}} + \\ &\alpha * \sum_{j \in R} \frac{ANice(i)}{\sum_{q \in U} ANice(q)} * ANR_j(t) \end{aligned} \tag{6}$$

c. Termination : If the number of iteration reaches a certain threshold, or the score became a stable value, the iteration will finish. Finally, we can use the following formula to calculate the final score:

$$F\_ANR_i = ANR_i(t_c) + \frac{ANR_g(t_c)}{N} \tag{7}$$

3.3.2. Analysis of the shortcomings of AveNiceRank

Through the description above, AveNiceRank makes full use of the nodes' multiple personal attributes. By designing a reasonable ANice value, AveNiceRank considers not only the topology properties of the nodes in network structure, but also the multiple attributes which can represent the actual performance of the users in the social network. Our algorithm improves based on LeaderRank, so it implies that each point is equally important, and ANR value is divided averagely to each point. In the actual situation of Zhihu network, it is obvious that the importance of every node is not same. In the process of calculation, if we divide the ANR score averagely, a user who has high ANice value, and few fans, he can rank high finally. This can not reflect the actual situation.

3.4. Improvement of AveNiceRank

In the description above, when calculating the actual value of ANRank, ANR value is divided averagely to its followees. However, in the actual social network, the importance of each node is not same, and in order to reduce the impact when calculating the users' personal attributes, we need to design a new distribution rule of ANR value. The more important a node is, the more ANR value can be allocated in each actual calculation. Therefore, we impove ANR algorithm by designing a new distribution rule. We name it ImpAveNiceRank. The specific steps are as follows:

a. Initialization :The $IANR_i(t)$ is defined as the score of node i at time t, and $IANR_i(0) = 1$ for all nodes i;

b. Iteration : In ImpAveNiceRank algorithm, IANR value of a node is based on the number of its followers and ANice value of the node. Node j is one of the followers of node i. ANice(i) is the ANice value of node i, and it is the average value of the Nice value of the node i. U is the set of the followees of node j. So the IANR value of node i at time t + 1 can be calculated by the following formula:

$$
\begin{aligned}
&IANR_i(t+1) = \\
&(1-\alpha) * \sum_{j \in R} \frac{IANR_i(t)}{\sum_{k \in U} IANR_k(t)} * IANR_j(i) + \\
&\alpha * \sum_{j \in R} \frac{ANice(i)}{\sum_{q \in U} ANice(q)} * ANR_j(t)
\end{aligned}
\tag{8}
$$

c. Termination : If the number of iteration reaches a certain threshold, or the score becomes a stable value, the iteration will finish. Finally, we can use the following formula to calculate the final score:

$$
F\_IANR_i = IANR_i(t_c) + \frac{IANR_g(t_c)}{N} \tag{9}
$$

In conclusion, the ImpAveNiceRank algorithm can meet the actual demands in Zhihu network. This algorithm considers not only the topology properties of the nodes in the network structure, but also the personal attributes of the nodes. And it contains a new distribution rule, which can balance the actual impact of the users' personal attributes and users' topology properties on the users' influence.

4. Experimental Evaluation Of The Improved Algorithms

4.1. Data Description

The experimental dataset in this paper is crawled from Zhihu.com, which consists of 5000 users and 155241 relationships. And we crawl the approvals, thanks and answers number of every user. These factors are the users' personal attributes in Zhihu network. In other words, the social network in our experiments consists of 5000 nodes and 155241 edges. And we get every node's mutilple attributes. The experimental dataset and programs can be retrieved on request.

Table 1: Top-20 result ranked by the five approaches

user_id	PageRank	LeaderRank	NiceRank	AveNiceRank	ImpAveNiceRank
88	1	1	8	3	3
50	2	2			
102	3	3			
103	4	5	4		
45	5	4	5		
206	6	6	2		5
25	7	7			
513	8	8			
370	9	9			
6	10	10			
161			1		
248			3		
31			6		
136			7	7	7
46			9		
156			10		
13				1	1
42				2	2
2				4	4
119				5	6
53				6	9
432				8	8
182				9	
814				10	10

Table 2: PageRank Top-10 Result (d = 0.85)

user_id	user_name	followers	answer	Nice	ANice
88	kaifulee	2364	104	80101	770
50	whale	1224	619	12588	20
102	chengyuan	1224	708	7592	10
103	wangxiaofeng	1683	2394	211025	88
45	fu-er	1469	3965	156214	39
206	amuro1230	1514	1140	218653	191
25	hecaitou	1410	267	33575	125
513	lili	815	57	2135	37
370	amonjok	899	387	34542	89
6	oxygen	902	378	7927	20

Table 3: LeaderRank Top-10 Result

user_id	user_name	followers	answer	Nice	ANice
88	kaifulee	2364	104	80101	770
50	whale	1224	619	12588	20
102	chengyuan	1224	708	7592	10
45	fu-er	1469	3965	156214	39
103	wangxiaofeng	1683	2394	211025	88
206	amuro1230	1514	1140	218653	191
25	hecaitou	1410	267	33575	125
513	lili	815	57	2135	37
370	amonjok	899	387	34542	89
6	oxygen	902	378	7927	20

Table 4: NiceRank Top-10 Result ($\alpha = 0.3$)

user_id	user_name	followers	answer	Nice	ANice
161	magasa	1083	801	395059	493
206	amuro1230	1514	1140	218653	191
248	bo-cai-28-7	1199	1103	549660	498
103	wangxiaofeng	1683	2394	211025	88
45	fu-er	1469	3965	156214	39
31	miaomiaomiao	925	785	223593	284
136	dong-ji-zai-hang-zhou	931	208	222946	1071
88	kaifulee	2364	104	80101	770
46	gracelu	795	535	50950	95
156	linaoshi	511	1005	441628	439

Table 5: AveNiceRank Top-10 Result ($\alpha = 0.3$)

user_id	user_name	followers	answer	Nice	ANice
13	DrHow	1096	1	28405	28405
42	shuang-wang	347	13	18710	1439
88	kaifulee	2364	104	80101	770
2	cogito	1060	18	57608	3200
119	zhuxiaobao	511	349	250457	717
53	tao-zi-de-tao	325	211	156849	743
136	dong-ji-zai-hang-zhou	931	208	222946	1071
432	imajin	825	55	147770	2686
182	siyu-yang	314	88	51729	587
814	superhistorical	100	3	32484	10828

Table 6: ImpAveNiceRank Top-10 Result ($\alpha = 0.3$)

user_id	user_name	followers	answer	Nice	ANice
13	DrHow	1096	1	28405	28405
42	shuang-wang	347	13	18710	1439
88	kaifulee	2364	104	80101	770
2	cogito	1060	18	57608	3200
206	amuro1230	1514	1140	218653	191
119	zhuxiaobao	511	349	250457	717
136	dong-ji-zai-hang-zhou	931	208	222946	1071
432	imajin	825	55	147770	2686
53	tao-zi-de-tao	325	211	156849	743
814	superhistorical	100	3	32484	10828

4.2. Analysis of the Top 10 Result

We test these algorithms on the real network data, and the main Top-10 nodes are shown below tables:

By comparing these tables, we can clearly see that there are some points ranking different in the five tables, and the top 10 results of the five algorithms are not same. Compared with PageRank, the main improvement of the Leader-Rank algorithm is increasing the robustness, and speeding up the convergence. We don't talk too much in this paper. The algorithms in this paper mainly improve based on LeaderRank, so we mainly talk about the differences between LeaderRank, NiceRank, AveNiceRank and ImpAveNiceRank.

As can be seen from the results, the result of the LeaderRank algorithm in Table III is mainly consistent of the nodes' followers. However, the LeaderRank algorithm only considers the topology property of the nodes in the social network, so the Nice value of some nodes in the top 10 is not large. Combined with the actual situation of Zhihu.com, the Nice value can represent the degree of recognition of the users in social network, and it can reflect the user' influence, so the results of the LeaderRank algorithm can not match the actual situation.

Different from LeaderRank, the other improved algorithms consider not only the topology property of the node, but also the Nice value. So the nodes' followers and Nice value in the top 10 result in Table IV, V and VI are considerable. In the results of the three algorithms, the nodes which have more followers and high Nice value are added to the top 10 list while the nodes which have less Nice value are deleted from the top 10 results. For example, the node 102 has less Nice value and it ranks high in the result of the LeaderRank algorithm, but it was deleted in other results.

Comparing the results with NiceRank algorithm and AveNiceRank algorithm, we can see that there are more differences between the two algorithms. The main reason is that, NiceRank only considers the Nice value of the nodes, but AveNiceRank considers the average Nice value of the nodes, so some nodes have large node value, but they answer too many questions, the average Nice value will be small. The average Nice value can represent the actual performance of the users in each question they answered, so it can match the actual situation. In the top 10 result of AveNiceRank, there are some special nodes, their Nice value is not large, but their average nice value is large, this is because the number of their answers is small. For example, the node 13 answers only 1 question and its Nice value is 28405, so its average value is 28405. It ranks high. These situations are not normal situation. There are three means to overcome the phenomenon. The first mean is that we can set a threshold about the number of the users' answers to limit the phenomenon whose answer number is small but the average nice is too large. The second mean is that we can set α value to control the impact of the topology properties and the personal attributes on nodes ranking. And the next algorithm—ImpAveNiceRank can solve parts of the problem.

We then talk about the result in the table VI, which is the result of ImpAveNiceRank algorithm. Compared with the result of AveNiceRank, it can be seen that some nodes rank different in the two algorithms. For example, the node 119 ranks 5th in AveNiceRank, but it ranks 6th in ImpAveNiceRank, and the node 206 which has less ANice value ranks higher than the node 119. It is because that,ImpAveNiceRank assigns the node's IANR score to its followees based on its last IANR score, not averagely. The larger the node's IANR score is, the more it can get from its followers. For example, the ANice value of the node 206 is less than the nodes which rank lower than it, but it gets the most IANR value from its followers, and gets high rank in the ImpAveNiceRank algorithm. And we can see that ImpAveNiceRank can solve parts of the problem,that is, a node' answer number is low but it ranks high, but the effect is not great. We should consider the parameter α and the answer number threshold to solve the abnormal situation.

4.3. Analysis of the Top 500 Result

Combined with the actual situation, we can see that the nodes can be divided to three levels, including the zombie users, usual active users and target active users. The target active users is the main objectives in the study, because the proportion of opinion leaders in this group is highest. The zombie users are not only the users who register maliciously, but also the users whose liveness is too low. In the data of Zhihu.com, we define the users whose answer number is 0 as zombie users. And these users make up 10% in the data. Among the other users, we consider that 10% of all nodes, which has high follower number and Nice value as target active users, and we name them nice users. We can define the other nodes as usual active users. We write a program to analyze the data, and set the boundary value of the target active users. The result is that the boundary value of the target active users' follower number is 28, and the boundary value of the target active users' Nice value is 2357. And the nodes whose answer number is 0 are zombie users. The other nodes are usual active users. By analyzing the top 500 result of the five algorithms, we calculate the proportion of these nodes. The result is shown in Table VII.

It can be seen that the improved LeaderRank algorithms can improve the proportion of nice users. And in the result, the number of nice users found by ImpAveNiceRank is the most, so its performance is the best.

Table 7: The Nodes Level in the Top-500

user_level	PR	LR	NR	ANR	IANR
nice users	328	329	373	370	402
usual users	172	171	127	130	98
zombie users	0	0	0	0	0

4.4. The effect of α on the improved algorithms.

In NiceRank, AveNiceRank and ImpAveNiceRank, the parameter α is used to set a proportion of the effect about topology property and the individual attributes on influence evaluation. As we can see from formula 4, 6 and 8, the result is sensitive to the α value. The bigger the α value is, the impact of the user's personal attributes including Nice value and answer number is larger when calculating the users' influence. In order to verify the effect of α value on these algorithms, we define n/L as the differences among the three algorithms, and L value is the top list size of the result, its set is 20,50,100. Based on LeaderRank algorithm, we compare the differences among these algorithms and LeaderRank, respectively, indicating the top-L list is sensitive to the α parameter. Generally speaking, the larger the α value is, the bigger the difference is.

5. Conclusion

LeaderRank can be used to rank users and identify the influential users in social network. Compared with PageRank. It has high accuracy and anti-jamming capability, so it is better. In this paper ,we mainly apply LeaderRank to calculating the nodes' influence in Zhihu social network, and we improve LeaderRank by allowing for the nodes' Nice value. With almost the same convergence speed, we propose three new algorithms based on LeaderRank. Finally, we validate that the accuracy of identifying influential nodes has effectively improved, and we study the effect of parameter α on the improved algorithms.

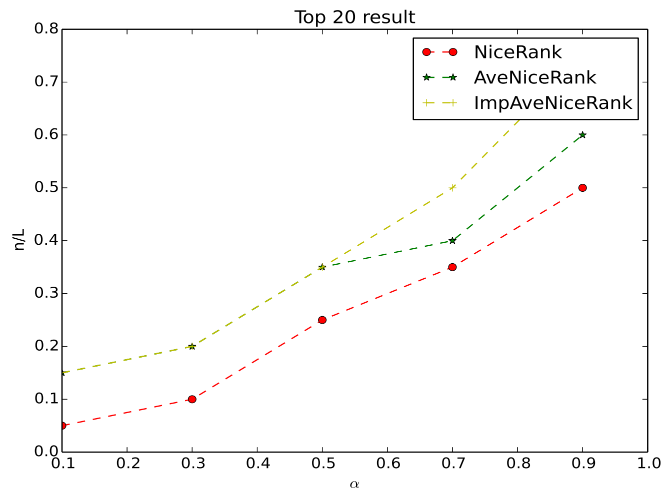

Figure 1: Top-20 Result $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$

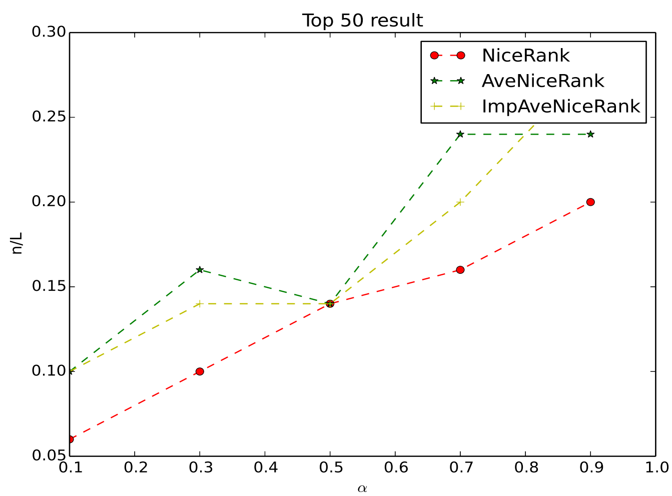

Figure 2: Top-50 Result $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$

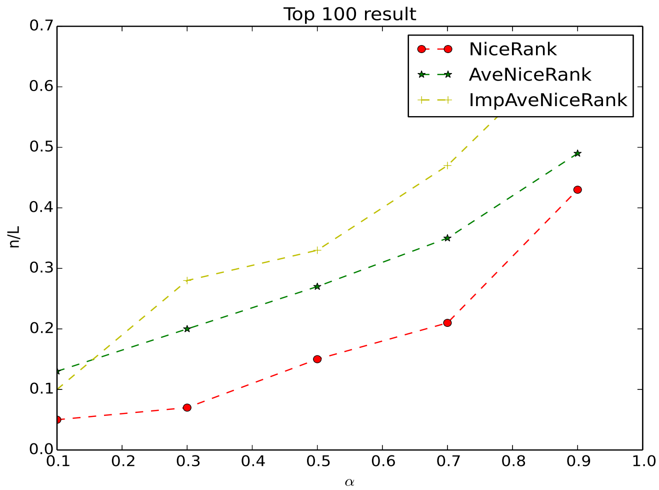

Figure 3: Top-100 Result $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$

References

[1] L. Lu¨, D.-B. Chen, T. Zhou, The small world yields the most effective information spreading, New Journal of Physics 13 (2011) 123005.

[2] Why rumors spread so quickly in social networks, Communications of the ACM 55 (2012) 70–75.

[3] J. Park, M. E. J. Newman, A network-based ranking system for US college football, J. Stat. Mech. (2005) P10014.

[4] R. M. Anderson, R. M. May, B. Anderson, Infectious diseases of humans: dynamics and control, Oxford University Press, 1992.

[5] Lv L,Zhang Y C,Yeung C H,et al. Leaders in social networks,the delicious case[J]. PloS One,2011,6(6).

[6] Page L, Brin S, Motwani R, et al, The PageRank citation ranking: Bring order to the web. Technical Report SIDL WP 1999 0120CA, Stanford InfoLab, 1998.

[7] Aral S,Walker D. Identifying influential and susceptible members of social networks[J]. Science,2012,337:337-341.

[8] Bai W J,Zhou T,Wang B H. Immunization of susceptible- infected model on scale-free networks[J]. Statistical Mechanics and its Applications: Physica A,2007,384(2):656-662.

[9] Hébert-Dufresne L,Allard A,Young J G,et al. Global efficiency of local immunization on complex networks[R]. Scientific Reports,2013.

[10] Huang X,Vodenska I,Wang F,et al. Identifying influential directors in the United States corporate governance network[J]. Physical Review E,2011,84(4).

[11] Zhou Y B,Lv L,Li M. Quantifying the influence of scientists and their publications:distinguishing between prestige and popularity[J]. New Journal of Physics,2012,14(3).

[12] Kitsak M,Gallos L K,Havlin S,et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010,6(11):888-893.

[13] L. C. Freeman, Centrality in social networks conceptual clarification, Social Networks 1 (1979) 215–239.

[14] Katz L. A new status index derived from sociometric analysis[J]. Psychometrika,1953,18(1):39-43.

[15] Dolev S,Elovici Y,Puzis R. Routing betweenness centrality[J]. Journal of the ACM,2010,57(4).

[16] P. Bonacich, Some unique properties of eigenvector centrality, Social Networks 29 (2007) 555–564.

[17] Estrada E,Rodriguez-Velazquez J A.Subgraph centrality in complex networks[J]. Physical Review E,2005,71(5).

[18] Kim H J,Kim J M. Cyclic topology in complex networks[J]. Physical Review E,2005,72:1-4.

[19] C. H. Comin, L. da F. Costa, Identifying the starting point of a spreading process in complex networks, Phys. Rev. E 84 (2011) 056105.

[20] Song K,Wang D,Feng S,et al. Detecting opinion leader dynamically in Chinese news comments[M]. Web-Age Information Management.Berlin Heidelberg:Springer,2012 : 197-209.

[21] D.-B. Chen, L. Lu¨, M.-S. Shang, Y.-C. Zhang, T. Zhou, Identifying influential nodes in complex networks, Physica A 391 (2012) 1777-1787.

[22] Li Q,Zhou T,Lv L,et al. Identifying influential spreaders by weighted LeaderRank[J]. Physica A,2014,404:47-75.

[23] D.-B. Chen, H. Gao, L. Lu¨, T. Zhou, Identifying influential nodes in large-scale directed networks: The role of clustering, PLoS ONE 8 (2013) e77455.