A Voting Approach to Uncover Multiple Influential Spreaders on Weighted Networks

Hong-liang Sun[†,‡]

[†] *NVIDIA Joint-Lab on Mixed Reality, International Doctoral Innovation Centre,*

[‡] *School of Computer Science,*
University of Nottingham, Ningbo, 315100, P. R. China

Duan-bing Chen[*]

Web Sciences Center, The Center for Digital Culture and Media,
Institute of Fundamental and Frontier Sciences and Big Data Research Center,
University of Electronic Science and Technology of China,
Chengdu, 611731, P. R. China

[*] *dbchen@uestc.edu.cn*

Jia-lin He

School of Computer Scicence and Engineering, China West Normal University,
Nanchong, Sichuan, 637009, P.R. China

Eugene Ch'ng

NVIDIA Joint-Lab on Mixed Reality, International Doctoral Innovation Centre,
University of Nottingham, Ningbo, 315100, P. R. China

Abstract

The identification of multiple spreaders on weighted complex networks is a crucial step towards efficient information diffusion, preventing epidemics spreading and etc. In this paper, we propose a novel approach WVoteRank to find multiple spreaders by extending VoteRank. VoteRank has limitations to select multiple spreaders on unweighted networks while various real networks are weighted networks such as trade networks, traffic flow networks and etc. Thus our approach WVoteRank is generalized to deal with both unweighted and weighted networks by considering both degree and weight in voting process. Experimental studies on LFR synthetic networks and real networks show that in the context of Susceptible-Infected-Recovered (SIR) propagation, WVoteRank outperforms existing states of arts methods such as

Preprint submitted to Physica A: Statistical Mechanics and its Applications.February 4, 2019

weighted H-index, weighted K-shell, weighted degree centrality and weighted betweeness centrality on final affected scale. It obtains an improvement of final affected scale as much as 8.96%. Linear time complexity enables it to be applied on large networks effectively.

Keywords: Multiple influential spreaders, Influence maximization, Weighted complex networks
PACS: 89.75.Fb, 87.15.A, 87.23.Ge

1. Introduction

Real complex systems are modelled as complex networks [1, 2] where vital nodes [3] (influential spreaders in propagation process) are of great importance to control the outbreak of epidemics [4, 5, 6, 7], to target potential customers in advertisements, to evaluate the performance of research institutes [8] and etc. Such applications are defined as *influence maximization* [9] to choose a set of initial spreaders which can achieve maximum propagation scale via underlying complex networks.

The influence of a node is closely related to the topological structure of the underlying network and where it locates. Traditional methods usually select a set of highly ranked nodes as influential spreaders because the importance of nodes can be ranked by a specific structural measure. A measure characterizing the importance of a node according to the topological information is usually named *centrality*. Degree centrality is a straightforward approach to count the neighbours of a certain node. It has limitations because only directly linked neighbours are considered while neighbours of neighbours are ignored. Thus Chen et al. proposed an improved method LocalRank [10], which took into accounts of fourth-order neighbours instead of first order neighbours in the case of degree centrality. Furthermore, it was found that the clustering coefficient negatively related to the node influence given the same neighbour number [11]. Thus ClusterRank [12] was proposed considering both neighbour number and clustering coefficient.

Although H-index was proposed to measure a researcher's scientific impacts, it can also be applied to measure the importance of nodes considering the degree information of neighbours [13]. Kitsak et al. argued that the influence of a node was determined by its position of global network structure rather than local neighbours. Coreness [14] and its variants [15, 16] were proposed to iteratively decompose the networks where nodes locating in the central shells were more influential than those in periphery shells. Lü et al. discovered that degree centrality, H-index and coreness centrality can be regarded as the initial, intermediate and final states in the sequences driven by a discrete operator [17]. We mainly have above discussions about centralities on undirected networks. There are many other centralities related to directed networks including PageRank [18], LeaderRank [19], HITs [20] and etc. A number of other measures have also been proposed due to different assumptions, such as betweenness [21], closeness [22, 23], path diversity [24], dynamic sensitive [25], k-core values of neighbors [26], mixed degree decomposition [27], iterative resource allocation [28], gravity formula [29], structural holes [30], spreading probability [31] and etc. Due to the limitation of space, interested readers are suggested to refer to the comprehensive review by Lü et al. [3].

A number of different measures have been proposed to rank nodes in propagation process on complex networks. Those methods can deal with ranking orders of nodes, but it is not effective to select a small number of critical nodes in propagation dynamics. Because top-ranked nodes selected by these methods might be clustered as a rich-club [32]. Initially, some nodes are selected and they contribute to effective information propagation. When more nodes are added to form the rich-club, these nodes have less contributions to effective information propagation. Kempe et al. proposed a hill-climbing heuristic to select a set of spreaders [9]. Chen et al. suggested a degree discount heuristic [33] to select multiple spreaders, which achieved larger propagation scope than degree and other centralities. It also runs in milliseconds on networks with tens of thousands of nodes and edges. Zhao et al. generalized the graph coloring problem from graph theory to identify multiple spreaders on complex networks [34]. Guo et al. proposed a distance-based coloring method [35] which had significant improvement of spreading scale compared with traditional coloring method. Liu et al. proposed a method using local structural similarity (LSS) to identify multiple influential spreaders [36]. Experimental studies on real networks show that LSS outperforms traditional coloring method on final spreading influence. He et al. argued that traditional methods might have limitations to find multiple spreaders clustered in one community[37]. Thus it was discovered that multiple spreaders can be selected from different communities [38, 39] according to measures such as degree centrality, k-core [14] and ClusterRank [12]. Morone et al. mapped this problem onto optical percolation to identify the minimal set of spreaders [40]. This new method Collective Influence

(CI) [41] can discover influencers via optimal percolation in massively large social networks. Hu et al. further discovered that information spread from the seeds and it reached a point of no return and would rapidly reach the percolation cluster. Thus they proposed that local structure could be applied to quantify influential spreaders on the whole network [42].

Existing methods to identify multiple influential spreaders are usually applied to unweighted networks. However, the edges of most real networks are affiliated with link strengths denoting the extent of interactions. Weights and degrees are two important properties of the underlying networks. But there is no obvious evidence to demonstrate they are highly correlated. Thus nodes with high degrees might have small weights because a large number of low weight edges attach to neighboring nodes. It is also possible to see nodes with small degrees but high weights. Weighted networks are wildly applied to economic, trade and traffic networks where weights indicate the trade and traffic flows. Traditional methods on unweighted networks have limitations because spreaders are not only determined by degree but also the weights of edges. Garas et al. extended the k-shell decomposition method from unweighted networks to weight network to solve this problem [43]. Other methods such as degree centrality [3], H-index centrality [17], betweeness centrality [44], closeness centrality [45] also have variants on weighted networks. The common weakness of these methods to select multiple spreaders on weighted networks is that top spreaders ranked by these methods are not accurate due to rich club phenomena [32].

VoteRank is suggested to identify multiple spreaders by voting [46]. The neighboring nodes of selected spreaders have a discount on voting abilities in the subsequent selections. It has high accuracy compared with CI [40] and other methods on large unweighted networks. Thus in this paper, we generalize VoteRank from unweighted networks to both unweighted and weighted networks as WVoteRank. The presented method can detect multiple spreaders on weighted complex networks by voting. All nodes vote for a spreader using edge weight and neighbour number in each turn, and the voting ability of neighbours of selected spreader will be decreased in subsequent turns. For example, a specific candidate attracting a large number of votes leads to a small number of votes given to other candidates. For this reason the voting ability of voters will be decreased after they vote for a specific spreader. This voting method based on VoteRank and is extended from unweighted networks to weighted networks. It is simple and effective compared with existing methods. Experimental studies on four synthetic networks generated by LFR [47] and four real networks show that it has significant improvement against weighted H-index [17], weighted k-shell decomposition [43], weighted degree centrality [3] and weighted betweeness centrality [44]. The maximum improvement of final affected nodes on SIR propagation is as much as 8.96% compared with the second best method weighted H-index.

2. Methods

2.1. VoteRank

Zhang et al. proposed VoteRank to decentralize and vote a set of influential spreaders [46] from underlying network $G = (V, E)$ where E is the set of edges connecting nodes from V. Each node v is attached with a tuple (s_v, va_v) where voting score s_v can be calculated from the neighbors of v by summarizing all its neighbors voting ability scores va_v.

Step 1 *Initialization*: Each node v is initialized with a tuple $(s_v, va_v) = (0, 1)$. p is given to identify the percentage of spreaders from the node set V.

Step 2 *Voting*: Each Node v votes for its directly linked neighbors with voting ability va_v. As a result, each node v obtains voting score s_v by summarizing all voting ability scores from its neighbors in Eq.1. The node with maximum voting score will be selected as candidate spreader in this turn. If the candidate is not the neighbor of existing spreaders, it will be put into the seed set. Selected spreaders also do not participate in voting process in the subsequent rounds ($va_{max} = 0$). Due to the requirement of decentralization, the voting ability scores of the neighbors of the voted spreader will be weaken. A discount d is given to the voting ability, $va_v = va_v - d, (if\ va_v > d, otherwise\ va_v = 0)$. Zhang et al. suggests $d = 1/\langle w \rangle$, where $\langle w \rangle$ is the average degree of the network [46].

$$s_v = \sum_{i \in \gamma(v)} va_i \tag{1}$$

The voting score s_v of node v on an unweighted network is shown in Eq. 1. Given a node v and its neighboring set $\gamma(v)$, the voting score s_v is defined as the summation of the voting ability va_i of each neighbor i.

Step 3 *Iteration*: Step 2 will be repeated until p percent of nodes are selected as spreaders.

2.2. WVoteRank on weighted networks

VoteRank aims to select spreaders on the entire network by voting from its neighbors, selected spreaders don't participate in subsequent elections and voting abilities of their neighboring nodes also decrease afterwards. However, it has limitations that only unweighted networks can be applied by this method. Various real networks are weighted networks where the strength of edges indicate the interactions between nodes. Thus, we extend VoteRank from unweighted networks to weighted networks as WVoteRank. Since unweighted networks are specific weighted networks, it can deal with unweighted networks when the default weight value of each edge is one.

Given a weighted network $G = (V, E)$ where w_{uv} denotes the link strength between node u and node v. The main difference between VoteRank and WVoteRank is in the voting process. Each node v obtains its voting score s_v from its neighbouring nodes. On one hand, the weighted sum of the voting ability scores from the neighboring nodes of v positively influences the voting score s_v. On the other hand, the number of neighbors also has positive impacts. Thus the voting score s_v of node v is defined as square root of the product of such two factors in Eq. 2. Spreader selection method and voting ability updating rule are the same as VoteRank described previously.

$$s_v = \sqrt{|\gamma(v)| \sum_{i \in \gamma(v)} va_i * w_{(v,i)}} \tag{2}$$

It is shown in Eq. 2 to calculate the voting score s_v of node v on a weighted network. Given a node v and its neighboring set $\gamma(v)$, the voting score s_v of node v is determined by three factors, the number of neighbors $|\gamma(v)|$, the voting score va_i of each neighbor i and the edge weight $w_{(v,i)}$ between node v and node i. s_v can be calculated as the square root of the product of two parts, the number of neighbors $|\gamma(v)|$ and weighted sum of va_i of each neighbor i.

2.3. Complexity of WVoteRank

The time complexity of WVoteRank consists of initialization of voting ability scores and voting scores, selection of the node with highest voting score and updating the voting ability and voting scores. It needs $O(n)$ steps to initialize voting ability of n nodes. Because the voting scores of n nodes are initialized by their neighboring nodes, $O(\langle w \rangle * n) = O(m)$ is needed at this step. As a result, the complexity of initialization is $O(n + m)$. It

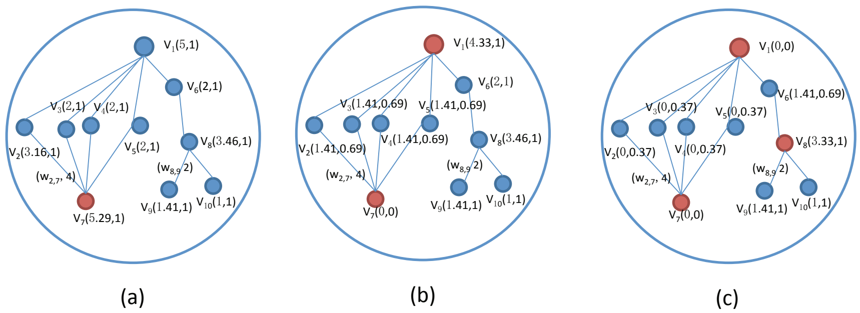

Figure 1: (Color online) A toy sample of weighted network with 10 nodes and 12 edges to demonstrate WVoteRank in (a-c). The weight of edge $e_{2,7}$ is $w_{2,7} = 4$ and $w_{8,9} = 2$. All other edges have default weight value 1. v_7 with the maximum voting score 5.29 is selected as the first spreader in the first round in (a). Then The voting ability of v_7 is set to 0 and it doesn't participate subsequent elections. Similarly, v_1 and v_8 are selected as spreaders in the second and third rounds as shown in (b) and (c) respectively.

takes $O(n)$ to select the spreader in one round. Thus the time complexity is $O(rn)$ to select r spreaders in r round. As the second order neighbours of the selected spreaders are updated after the spreader is selected. Thus $O(r\langle w\rangle^2) = O(rm^2/n^2)$ is needed to update the voting ability and voting scores of selected spreaders. It concludes that $O(n + m + rn + r * m^2/n^2)$ is the total time complexity. If $r << n$ and $O(n) = O(m)$, the time complexity of WVoteRank is simplified as $O(n)$.

Next, we will demonstrate WVoteRank with a toy network shown in Figure 1. The toy network with 10 nodes and 14 edges has the average weighted degree $\langle w\rangle = 0.313$. It is shown in Figure 1 (a) that in the first round, the voting ability scores of all nodes are 1 and each node is voted by its neighboring nodes. As a result, node v_7 is selected as the first spreader with maximum voting score $s_{v_7} = 5.29$. In the second round shown in Figure 1 (b), voting ability scores of the neighbouring nodes of node v_7 is decreased by a discount $\langle w\rangle$= 0.313 because node v_7 is selected as the spreader in the first round. The voting ability of node v_7 decreases as $va_{v_7} = 0$ and it does not participate in the subsequent voting processes. Each node is voted by its neighboring nodes again. v_1 with the largest voting score is selected as the spreader in the second round ($s_{v_1} = 4.33$). Similarly v_8 is selected as the third spreader shown in Figure 1 (c).

3. Experimental Results

In this section, both synthetic and real weighted networks are used to compare WVoteRank with several well-known algorithms.

To evaluate the accuracy of WVoteRank, some well-known methods such as weighted coreness, weighted H-index, weighted degree centrality and weighted betweenness centrality are selected as benchmark methods. One approach is to apply such four methods directly to select top-ranked nodes on weighted networks. However, such a way might not always be effective due to the rich club phenomenon. As a result another greedy approach is applied to such four methods. Given a weighted network, the node with the highest score measured by a specific measure (weighted coreness, weighted H-index, weighted degree centrality or weighted betweenness) is selected as the first node. Then neighboring nodes of the first node cannot be selected in the subsequent rounds to ensure spreaders are not connected directly. The second node is selected according to a specific measure from the left nodes and its neighbors are not considered at the subsequent rounds as well. This process continues until enough nodes are selected using a predefined measure from weighted coreness, weighted degree centrality, weighted H-index or weighted betweenness. The influences of these selected spreaders can be evaluated on specific propagation models. Next, a brief description will be given to explain the main idea of the propagation model SIR.

3.1. Propagation Model

We mainly use Susceptible-Infected-Recovery (SIR) model [48, 49] to compare the performance of various methods. Each node can be in one of the three statuses: susceptible, infected or recovered. At the very beginning, a set of spreaders (p percent nodes) are targeted as spreading sources and other nodes ($1 - p$ percent nodes) are regarded as susceptible nodes. At each time step, each infected node contacts one susceptible neighboring node with transmission rate μ. Meanwhile infected nodes can be recovered with a probability β. Recovered nodes will not be affected by infected nodes again. $\lambda = \mu/\beta$ is the infected rate governing the propagation process in SIR model.

Measure The evaluation of these methods are highly determined by the number of infected nodes. One possible way to measure infection scale is a function $F(t)$ with respect to t where t is the time step in SIR model and $F(t)$ is the number of infected and recovered nodes at step t. At the end of propagation, the number of infected nodes tends to become stable. Final

Table 1: Basic topological features of 4 synthetic networks by LFR model. n is the number of vertices, m is the number of non self-loop edges, c is the average clustering coefficient and r is assortative mixing coefficient of the networks, $\mu_t(\mu_w)$ are the mixing parameters for topology and weight in LFR model.

Net	n	m	$\langle w \rangle$	c	r	$\mu_t(\mu_w)$
$LFR1$	898	2251	6.26	0.08	0.48	0.1
$LFR2$	941	2493	6.99	0.05	0.39	0.3
$LFR3$	4523	10770	5.89	0.06	0.55	0.1
$LFR4$	4699	12204	6.84	0.03	0.35	0.3

affected scale $F(t_c)$ is proposed to measure this effect, which is determined by the number of final recovered nodes $n_{R(t_c)}$ and total node number n in Eq. 3. To evaluate the extent that spreaders are decentralized in the network, *weighted average shortest path length* L_{ws} is introduced as the average weighted distance of every two spreaders u, v in spreader set S from Eq.4.

$$F(t_c) = \frac{n_{R(t_c)}}{n} \tag{3}$$

$$L_{ws} = \frac{1}{|S|(|S|-1)} \sum_{u,v \epsilon S, u \neq v} l_{ws}(u, v) \tag{4}$$

It is shown in Eq. 4 that L_{ws} indicates the weighted average shortest path length to evaluate the extent to which selected spreaders from set S are decentralized distributed. $l_{ws}(u, v)$ indicates the weighted shortest path length from each pair of nodes (u, v) in the spreader set S. The sum of all of them is normalized by a factor $\frac{1}{|S|(|S|-1)}$.

3.2. Synthetic networks

We test the performance of WVoteRank on four synthetic scale-free weighted networks generated by LFR model [47]. In the LFR model, the weight distributions is power laws distribution which is set as $\alpha = 1.5$ in our experiment. The difference for four synthetic networks are the mixing parameter for the topology μ_t, mixing parameter for the weight μ_w, and the number of node n.

It is shown in Table 1 that LFR1 is a small network with small mixing values $\mu_t = \mu_w = 0.1$, LFR2 is a small network with large mixing values $\mu_t =$

$\mu_w = 0.3$, LFR3 is a large network with small mixing values $\mu_t = \mu_w = 0.1$, LFR4 is a large network with large mixing values $\mu_t = \mu_w = 0.3$.

Further experiments with SIR propagation show the spreading scale of selected multiple spreaders by our method WVoteRank and other methods. All methods are evaluated on SIR ($p = 0.05$, $\lambda = 1.2$) and four weighted LFR generated networks.

It is shown in Figure 2 that WVoteRank outperforms other methods in four different LFR generated networks. It obtains an improvement of final spreading scale over the second best method weighted H-index as much as 8.12% on LFR2 network. Similarly, it obtains an improvement of 30.71%, 29.24% and 12.21% compared with weighted coreness, weighted betweenness and weighted degree centrality, respectively. From the study on three other synthetic networks such as LFR1, LFR3 and LFR4, WVoteRank always outperforms other methods as well. Thus it concludes that our experimental studies on synthetic networks demonstrate the effectiveness and accuracy of the presented method WVoteRank.

3.3. Real networks

In this section, we aim to evaluate the performance of WVoteRank and other benchmark methods on 4 real networks. The basic topological features are summarized in Table 2. US airports network is a network of flights between US airports in 2010. The weight of each edge shows the number of flights on that airline[1]. Moreno health network was created from a survey from 1994. Each student was asked to list five best male and five best female friends. An edge indicates two students are friends and the weight of edge represents the strength of interaction between them[2]. Bitcoin alpha network is a trust network among people who trade using Bitcoin on a platform named Bitcoin Alpha. Members rate other members in a scale of -10 (total distrust) and 10 (total trust) [3]. Bitcoin OTC network is another network of Bitcoin traders on a platform named Bitcoin OTC. The weight of an edge follows the same rule of Bitcoin alpha [4].

The first experiment aims to investigate how different methods perform in terms of final affected scale $F(t_c)$ when the percentage of spreaders p

[1] http://konect.uni-koblenz.de/networks/opsahl-usairport
[2] http://moreno.ss.uci.edu/data.html#health
[3] http://snap.stanford.edu/data/soc-sign-bitcoinalpha.html
[4] http://snap.stanford.edu/data/soc-sign-bitcoinotc.html

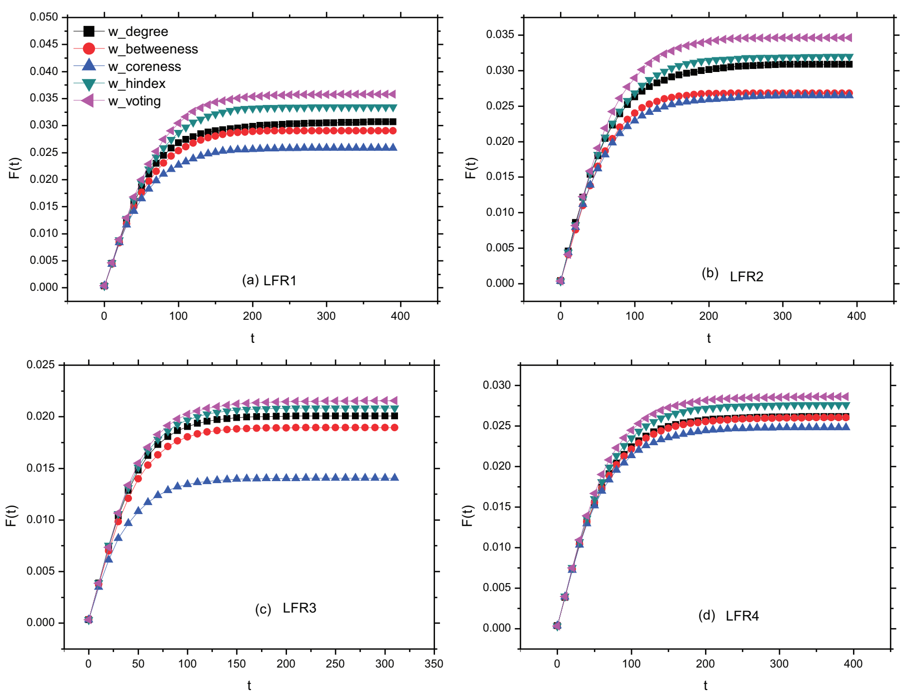

Figure 2: (Color online) The final affected spreading scales of WVoteRank and four benchmark methods. The experimental results of LFR1, LFR2, LFR3 and LFR4 are shown in (a), (b), (c), and (d) respectively. The results are averaged over 200 independent runs.

changes. In SIR model we set affected rate $\lambda = 1.2$ and recovered probability $\beta = \frac{1}{\langle w \rangle}$. It is shown in Figure 3 that weighted H-index is slightly better than WVoteRank when $p \in (0, 0.025]$. $F(t_c)$ of WVoteRank is larger than $F(t_c)$ of weighted coreness and other methods when $p \in [0.025, 0.09]$, which indicates WVoteRank outperforms other methods. WVoteRank obtains an improvement of spreading influence ranging from 1.82% to 4.98% over the second best method weighted H-index on US airports network. Especially on Bitcoin alpha network, WVoteRank outperforms weighted H-index with an improvement from 0.5% to 8.96% when $p \in [0.025, 0.09]$. Meanwhile WVoteRank outperforms weighted K-shell decomposition with an improvement ranging from 0.81% to 12.26%.

Table 2: Basic topological features of 4 real networks. n is the number of vertices, m is the number of non self-loop edges, $\langle w \rangle$ is the weighted average degree, c is the average clustering coefficient and r is assortative mixing coefficient of the networks.

Net	n	m	$\langle w \rangle$	c	r
$US\_airports$	1574	17215	157.578	0.384	-0.113
$Moreno\_health$	2539	10455	9.769	0.05	0.264
$Bitcoin\_alpha$	3719	10363	94.952	0.064	-0.173
$Bitcoin\_OTC$	5875	21218	143.149	0.077	-0.155

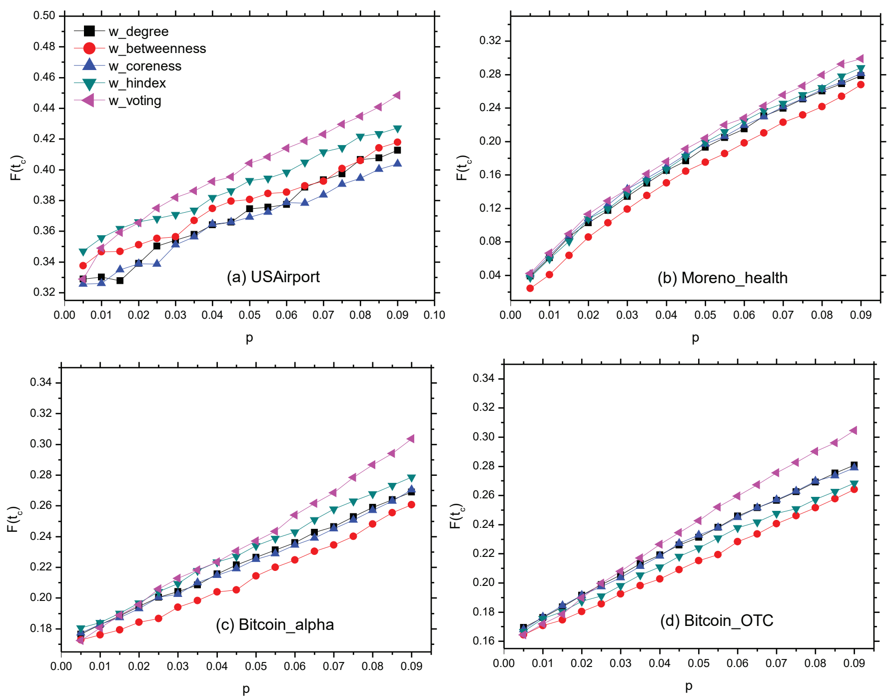

Figure 3: (Color online) The final affected scale $F(t_c)$ changes with respect to the percentage of source influential spreaders p. Experiments are carried out by WVoteRank, weighted H-index, weighted degree centrality, weighted K-shell decomposition and weighted betweeness centrality on four real networks respectively. In (a-d), SIR affected rate $\lambda = 1.2$, SIR recovered probability $\beta = \frac{1}{\langle w \rangle}$. The results are averaged over 200 independent runs.

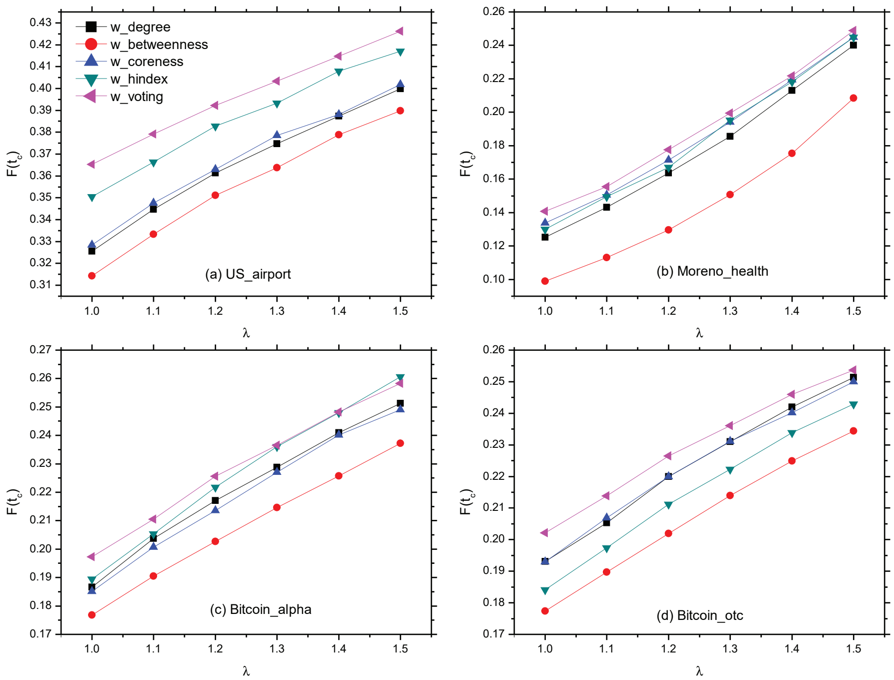

Figure 4: (Color online) The final affected scale $F(t_c)$ with respect to the infected rate λ. Experiments are carried out by WVoteRank, weighted H-index, weighted degree centrality, weighted K-shell decomposition and weighted betweeness centrality on four real networks respectively. In (a-d), the percentage of source influential spreaders $p = 0.05$, SIR recovered probability $\beta = \frac{1}{\langle w \rangle}$. The results are averaged over 200 independent runs.

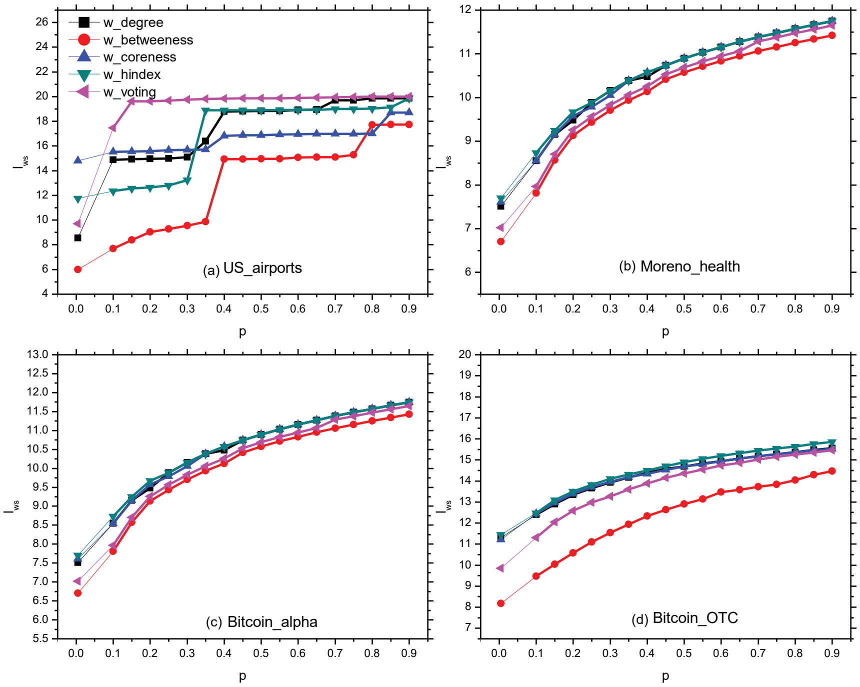

Figure 5: (Color online) The weighted average shortest path length of source spreaders L_{ws} changes with respect to the percentage of source influential spreaders p. Experiments are carried out by WVoteRank, weighted H-index, weighted degree centrality, weighted K-shell decomposition and weighted betweeness centrality on four real networks respectively.

When the percentage of spreaders p is fixed, further experiment is carried out to evaluate how final affected scale $F(t_c)$ changes with respect to infected rate λ. It is shown in Figure 4(c) that when $\lambda = 1.5$, weighted H-index is sightly better than WVoteRank on Bitcoin alpha network. Except this special case, WVoteRank outperforms other methods on 4 networks shown in Figure 4(a-d). Especially on US airports network, it obtains an improvement ranging from 1.68% to 2.54% against the second best method weighted H-index.

Next, we seek to evaluate the weighted average shortest path length L_{ws} over selected spreaders. It is shown in Figure 5 that weighted H-index out-

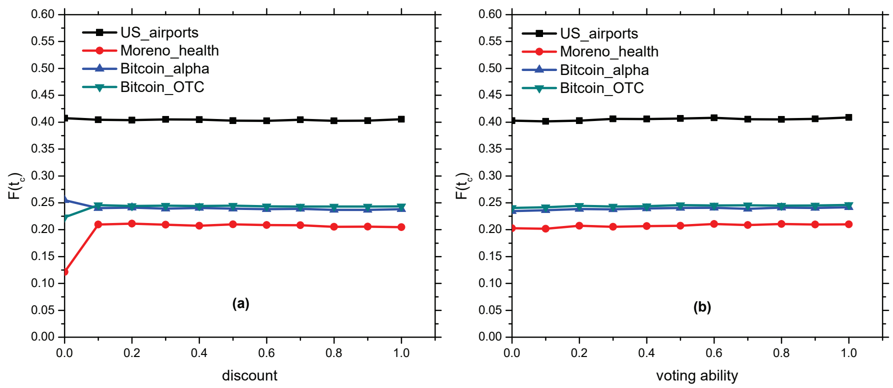

Figure 6: (Color online) The parameters discount and voting ability of WVoteRank method affect final influential scale $F(t_c)$. In (a-b), SIR infected rate $\lambda = 1.2$, SIR recovered probability $\beta = \frac{1}{\langle w \rangle}$, the percentage of source influential spreaders $p = 0.05$. The results are averaged over 200 independent runs.

performs other methods on Moreno health network, Bitcoin alpha network and Bitcoin OTC network. WVoteRank is the best method on US airports network. Considering experimental results on final affected rates shown in Figure 3 and Figure 4, it concludes that weighted average shortest path length L_{ws} is not necessarily positively related to final affected rate. But L_{ws} is a indicator to denote the decentralized locations of selected spreaders.

There are two parameters voting ability and discount in WVoteRank to determine the voting process. Next, we will examine the impact of parameter selection on final results. SIR model is setted with $\lambda = 1.2, \beta = \frac{1}{\langle w \rangle}$ and $p = 0.05$. From Figure 6(a) it is known that final affected scale $F(t_c)$ is relative stable with respect to discount in $[0, 1]$. It is shown in Figure 6(b) that there is an increasing of $F(t_c)$ when voting ability changes from 0.1 to 0.2. But $F(t_c)$ is almost stable when voting ability changes from 0.2 to 0.9. In this paper, discount is suggested as $\frac{1}{\langle w \rangle}$ and the default value of voting ability is 1.

4. Conclusions

Even though great efforts have been contributed to the research of the identification of multiple influential spreaders, it is still an open problem. In this paper, we have suggested a novel method WVoteRank based on VoteRank proposed by Zhang et al [46]. VoteRank has limitations to assign spreaders only on unweighted networks. Thus we extend VoteRank as WVoteRank to deal with both weighted and unweighted networks. Both node degree and edge weights are considered in this new method in voting process.

Further experimental studies on both synthetic networks and real networks demonstrate two advantages compared with other methods. Firstly WVoteRank has been evaluated on four LFR generated networks with up to thousands of nodes and edges. It is also compared with weighted K-shell decomposition, weighted H-index weighted degree centrality and weighted betweeness centrality. The experimental results show that WVoteRank can detect spreaders with maximum final affected spreading scale compared with other methods. It obtains an improvement as much as 8.12% compared with the second best method weighted H-index.

Further experimental studies have been carried out on four real weighted networks in SIR propagation. The first part of the experiments aims to investigate how WVoteRank performs with respect to the percentage of nodes p. It concludes that when p is very small $p <= 2.5\%$, there are no significant differences between WVoteRank and other methods. However when $2.5\% < p <= 9\%$, WVoteRank always outperforms other methods. It obtains an improvement as much as 8.96% over the second best method weighted H-index. Furthermore, we investigate how infected rate λ influences the final affected scale. When $1.0\% <= \lambda <= 1.5\%$ and $p = 0.05$, WVotreRank obtains an improvement ranging from 1.68% to 2.54% over weighted H-index centrality on US airports network.

Due to the impact of rich club phenomenon, a number of methods have been developed to find spreaders decentralized via the underlying network. But higher decentralization does not always lead to higher spreading scale in our experiment. Because weighted H-index can detect spreaders with higher *average shortest path length*, but the spreaders detected by WVoteRank lead to the largest final spreading scale. Thus striking a balance between decentralization and effective spreading scale provides insightful suggestions for future research. In this paper we merely study methods on static weighted homogeneous networks. Thus it is valuable to explore the identification of spreaders on temporal networks [50] and multilayer networks [51] in the near future.

Acknowledgements

This work is jointly supported by the UK Engineering and Physical Sciences Research Council (Grant Number EP/L015463/1), the National Natural Science Foundation of China (Grant Nos. 61433014 and 61673085), the Fundamental Research for the Central Universities (Grant Nos. ZYGX2014Z002 and ZYGX2015J152), and the Sichuan Science and Technology Support Project (Grant Number 2018RZ0081).

References

[1] M. E. J. Newman, Networks: an introduction, Oxford University Press, New York, May 20, 2010 (2010) 18.

[2] X. F. Wang, X. Li, G. R. Chen, Network science: an introduction, Higher Education Press, Beijing, April, 2012 (2012) 10–15.

[3] L. Y. Lü, D. B. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, T. Zhou, Vital nodes identification in complex networks, Physics Reports 650 (2016) 1–63.

[4] Y. Sun, C. Liu, C. X. Zhang, Z. K. Zhang, Epidemic spreading on weighted complex networks, Physics Letters A 378 (7) (2014) 635–640.

[5] K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, A. L. Barabasi, The human disease network, Proceedings of the National Academy of Sciences USA 104 (21) (2007) 8685–8690.

[6] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, Reviews of Modern Physics 87 (3) (2015) 925–979.

[7] W. Wang, M. Tang, H. Eugene Stanley, L. A. Braunstein, Unification of theoretical approaches for epidemic spreading on complex networks, Reports on Progress in Physics 80 (3) (2017) 036603(p1–p16).

[8] J. P. Wang, Q. Guo, e. a. Yang. K, Credit allocation for research institutes, EPL (Europhysics Letters) 118 (4) (2017) 48001 (p1–p6).

[9] K. David, K. Jon, T. Ev a, Maximizing the spread of influence through a social network, Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,Washington, DC, USA, August 24 - 27, 2003 (2003) 137–146.

[10] D. B. Chen, L. Y. Lü, M.-S. Shang, Y. C. Zhang, T. Zhou, Identifying influential nodes in complex networks, Physica A: Statistical Mechanics and its Applications 391 (4) (2012) 1777–1787.

[11] T. Petermann, D. L. R. P., Role of clustering and gridlike ordering in epidemic spreading, Physical Review E 69 (6) (2004) 066116(p1–p14).

[12] D. B. Chen, H. Gao, L. Y. Lü, T. Zhou, Identifying influential nodes in large-scale directed networks: the role of clustering, PLoS ONE 8 (10) (2013) e77455(p1–p10).

[13] J. E. Hirsch, An index to quantify an individual's scientific research output, Proceedings of the National Academy of Sciences of the U.S.A 102 (46) (2005) 16569–16572.

[14] M. Kitsak, L. K. Gallos, S. e. a. Havlin, Identification of influential spreaders in complex networks, Nature Physics 6 (11) (2010) 888893.

[15] Y. Liu, M. Tang, T. Zhou, Y. Do, Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition, Scientific Reports 5 (2015) 9602(p1–p8).

[16] Y. Liu, M. Tang, T. Zhou, Y. Do, Improving the accuracy of the k-shell method by removing redundant links: From a perspective of spreading dynamics, Scientific Reports 5 (2015) 13172(p1–p11).

[17] L. Y. Lü, T. Zhou, Q.-M. Zhang, H. E. Stanley, The h-index of a network node and its relation to degree and coreness, Nature Communications 7 (1) (2016) 10168(p1–p7).

[18] S. Brin, L. Page, The anatomy of a largescale hypertextual web search engine, Computer Networks and ISDN Systems 30 (18) (1998) 107–117.

[19] L. Y. Lü, Y.-C. Zhang, C. H. Yeung, T. Zhou, Leaders in social networks, the delicious case, PLoS ONE 6 (6) (2011) e21202(p1–p9).

[20] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, Journal of the ACM 46 (5) (1999) 604632.

[21] L. C. Freeman, Centrality in social networks conceptual clarification, Social Networks 1 (3) (1979) 215–239.

[22] G. Sabidussi, The centrality index of a graph, Psychometrika 31 (4) (1996) 581–603.

[23] C. Dangalchev, Residual closeness in networks, Physica A: Statistical Mechanics and its Applications 365 (2) (2006) 556–564.

[24] D.-B. Chen, R. Xiao, A. Zeng, Y.-C. Zhang, Path diversity improves the identification of influential spreaders, EPL (Europhysics Letters) 104 (6) (2013) 68006(p1–p6).

[25] J. G. Liu, J. H. Lin, Q. Guo, T. Zhou, Locating influential nodes via dynamics-sensitive centrality, Scientific Reports 6 (2016) 21380(p1–p8).

[26] J. H. Lin, Q. Guo, W. Z. Dong, L. Y. Tang, J. G. Liu, Identifying the node spreading influence with largest k-core values., Physics Letters A 378 (45) (2014) 3279–3284.

[27] A. Zeng, C. J. Zhang, Ranking spreaders by decomposing complex networks, Physics Letters A 377 (14) (2013) 1031–1035.

[28] Z.-M. Ren, A. Zeng, D.-B. Chen, H. Liao, J.-G. Liu, Iterative resource allocation for ranking spreaders in complex networks, EPL (Europhysics Letters) 106 (4) (2014) 48005(p1–p6).

[29] L. ling Ma, C. Ma, H.-F. Zhang, B.-H. Wang, Identifying influential spreaders in complex networks based on gravity formula, Physica A: Statistical Mechanics and its Applications 451 (2016) 205 – 212.

[30] P. Hu, T. Mei, Ranking influential nodes in complex networks with structural holes, Physica A: Statistical Mechanics and its Applications 490 (2018) 624–631.

[31] Z.-K. Bao, C. Ma, B.-B. Xiang, H.-F. Zhang, Identification of influential nodes in complex networks: Method from spreading probability viewpoint, Physica A: Statistical Mechanics and its Applications 468 (2017) 391 – 397.

[32] V. Colizza, A. Flammini, M. Serrano, et al., Detecting rich-club ordering in complex networks, Nature Physics 364 (2) (2006) 110–115.

[33] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France,June 28July 1, 2009 (2009) 199–208.

[34] X. Y. Zhao, B. Huang, M. Tang, et al., Identifying effective multiple spreaders by coloring complex networks, EPL (Europhysics Letters) 108 (6) (2015) 68005(p1–p6).

[35] L. Guo, J.-H. Lin, Q. Guo, J.-G. Liu, Identifying multiple influential spreaders in term of the distance-based coloring, Physics Letters A 380 (7-8) (2016) 837–842.

[36] J. G. Liu, Z. Y. Wang, Q. Guo, L. Guo, Q. Chen, Y. Z. Ni, Identifying multiple influential spreaders via local structural similarity, EPL (Europhysics Letters) 119 (1) (2017) 18001(p1–p6).

[37] J. L. He, Y. Fu, D. B. Chen, A novel top-k strategy for influence maximization in complex networks with community structure, Plos One 12 (10) (2015) e0145283(p1–p10).

[38] H. L. Sun, E. Chng, X. Yong, J. M. Garibaldi, S. See, D. B. Chen, An improved game-theoretic approach to uncover overlapping communities, International Journal of Modern Physics C 28 (9) (2017) 1750112(p1–p17).

[39] H. L. Sun, E. Chng, X. Yong, J. M. Garibaldi, S. See, D. B. Chen, A fast community detection method in bipartite networks by distance dynamics, Physica A: Statistical Mechanics and its Applications 496 (2018) 108–120.

[40] F. Morone, H. A. Makse, Influence maximization in complex networks through optimal percolation, Nature 7563 (524) (2015) 65–68.

[41] F. Morone, B. Min, L. Bo, R. Mari, H. A. Makse, Collective influence algorithm to find influencers via optimal percolation in massively large social media, Sci Rep 6 (2016) 30062(p1–p11).

[42] Y. Q. Hu, S. G. Ji, Y. L. Jin, L. Feng, H. E. Stanley, S. Havlin, Local structure can identify and quantify influential global spreaders in large scale social networks, PNAS 115 (29) (2018) 7468–7472.

[43] A. Garas, F. Schweitzer, S. Havlin, A k-shell decomposition method for weighted networks, New Journal of Physics 14 (8) (2012) 083030(p1–p14).

[44] T. Opsahl, F. Agneessens, J. Skvoretz, Node centrality in weighted networks: Generalizing degree and shortest paths, Social Networks 32 (3) (2010) 245–251.

[45] M. E. J. Newman, Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality, Physical Review E: Statistical Nonlinear Soft Matter Physics 64 (2) (2001) 016132(p1–p7).

[46] J. X. Zhang, D. B. Chen, Q. Dong, et al., Identifying a set of influential spreaders in complex networks, Scientific reports 6 (2016) 27823(p1–p9).

[47] A. Lancichinetti, S. Fortunato, Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities, Physical Review E 80 (1) (2009) 016118(p1–p8).

[48] W. H. Herbert, The mathematics of infectious diseases, SIAM Review 42 (4) (2000) 599–653.

[49] R. Yang, B. H. Wang, J. Ren, W. J. Bai, Z. W. Shi, W. X. Wang, T. Zhou, Epidemic spreading on heterogeneous networks with identical infectivity, Physics Letters A 364 (3) (2007) 189–193.

[50] P. Holme, J. Saramäki, Temporal networks, Physics Reports 519 (3) (2012) 97–125.

[51] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, M. A. Porter, Multilayer networks, Journal of Complex Networks 2 (3) (2014) 203–271.