

Article

A Novel Entropy-Based Centrality Approach for Identifying Vital Nodes in Weighted Networks

Tong Qiao ¹, Wei Shan ^{1,2,*}, Ganjun Yu ¹ and Chen Liu ³

¹ School of Economics and Management, Beihang University, Beijing 100191, China; tqiao@buaa.edu.cn (T.Q.); yuganj@buaa.edu.cn (G.Y.)

² Key Laboratory of Complex System Analysis and Management Decision, Ministry of Education, Beijing 100191, China

³ Business School, University of Shanghai for Science and Technology, Shanghai 200093, China; liuchen.lic@gmail.com

* Correspondence: shanwei@buaa.edu.cn; Tel.: +86-10-8231-4022

Received: 20 March 2018; Accepted: 7 April 2018; Published: 9 April 2018



Abstract: Measuring centrality has recently attracted increasing attention, with algorithms ranging from those that simply calculate the number of immediate neighbors and the shortest paths to those that are complicated iterative refinement processes and objective dynamical approaches. Indeed, vital nodes identification allows us to understand the roles that different nodes play in the structure of a network. However, quantifying centrality in complex networks with various topological structures is not an easy task. In this paper, we introduce a novel definition of entropy-based centrality, which can be applicable to weighted directed networks. By design, the total power of a node is divided into two parts, including its local power and its indirect power. The local power can be obtained by integrating the structural entropy, which reveals the communication activity and popularity of each node, and the interaction frequency entropy, which indicates its accessibility. In addition, the process of influence propagation can be captured by the two-hop subnetworks, resulting in the indirect power. In order to evaluate the performance of the entropy-based centrality, we use four weighted real-world networks with various instance sizes, degree distributions, and densities. Correspondingly, these networks are adolescent health, Bible, United States (US) airports, and Hep-th, respectively. Extensive analytical results demonstrate that the entropy-based centrality outperforms degree centrality, betweenness centrality, closeness centrality, and the Eigenvector centrality.

Keywords: complex network; vital nodes; weighted networks; centrality; entropy-based centrality

1. Introduction

The extensive applications of centrality in complex networks bring considerable value in a large number of scenarios, such as identifying the most influential spreaders in online communities [1], carrying out online precision marketing by identifying the productive and influential bloggers [2], supervising exceptional events [3], detecting the influential criminals [4], predicting essential proteins [5–10], quantifying the academic influence of scientists on the basis of co-authorship networks constructed by their publications and citations [11–13], discovering financial risks (for example the DebtRank algorithm first proposed by Battiston et al. [14] and then developed by Tabak et al. [15]), forecasting career movements that include promotion and resignation by analyzing the characteristics of the employees' networks [16–18], and improving the robustness of power grids in order to prevent catastrophic outages [19–22].

In the literature, the power of an actor in a given network is mostly influenced and mirrored by the topological structure of the network [23–26]. As a result, the vast majority of widely applied

centrality methods exclusively consider the topographic properties of a given network. In other words, the concept of structural centralities is designed with the purpose of detecting structural information and characterizing its influence. As noted by Lü [27], we categorize structural centralities into path-based and neighborhood-based centralities and then describe the state of art.

From the perspective of influence propagation, two significant features that most vital nodes share are propagating speed and propagating range, which should be affected by traffic flows to a great extent. Based on this idea, several classical approaches have been proposed, including betweenness centrality [23], Katz centrality [28], closeness centrality [29,30], and eccentricity [31]. Both eccentricity and closeness centrality count geodesic routes based on the idea that the efficiency of information dissemination can be maximized along the shortest paths. Yet, the value of eccentricity solely relies on the maximum length of the shortest path. By comparison, the closeness centrality takes all the shortest distances into consideration, and the value of closeness centrality can be interpreted mathematically as the inverse of the mean length of information dissemination. The betweenness centrality is a strong indicator that reflects the controllability of traffic flow. That is, most vital nodes often act like bridges that connect various communities. Generally speaking, the concept of closeness centrality indicates accessibility, and the concept of betweenness centrality shows controllability. Katz centrality [28] takes all paths into consideration and assigns less weights on the longer paths. Besides the path-based centralities mentioned above, the information centrality index [32] assumes that the loss of information in the process of propagation depends on the length of the path. Therefore, this approach calculates the quantity of information contained in all potential traffic flows.

The extensive studies of the neighbor nodes lay the foundation for neighborhood-based centralities. The LocalRank algorithm proposed by Chen et al. [33] not only utilized the information contained in the immediate neighbors of a given node but took into account the fourth-order neighbors. Gao et al. [34] extended the algorithm to weighted networks. However, it is clear that LocalRank and its revised version fail to capture the process of influence propagation. As was noted by Petermann [35], the local interconnectedness inevitably affects the process of the information transmission. In light of that thought, Chen et al. [36] analyzed the role of clustering and proposed the ClusterRank algorithm by considering the effect of the clustering coefficient on the spreading speed. As a matter of fact, the nodes with high values of ClusterRank are usually the nodes that belong to distinct communities. Thus, there is acceleration of propagation once information passes through those nodes. The centralities discussed above ignore the significance of the location of a node in a given network. Kitsak et al. [37] believed that coreness could be a more effective index to distinguish the relative importance of nodes. Zeng et al. [38] and Pei et al. [39] applied the k -core algorithm to large real-world networks. However, the original k -core algorithm may result in plenty of indistinguishable nodes with same value of coreness. Moreover, the initial algorithm exclusively makes use of the residual degree. Thus, many researchers [40–43] proposed many modified k -core algorithms from different points of view. The last representative neighborhood-based structural centrality is the H -index [44], a local centrality initially employed to measure researchers' academic influence by counting their publications [45,46]. Recently, Lü et al. [47] proved the convergence of the H -indices.

Besides the structural centralities discussed above, recently entropy theories have been applied to measure the complexity and uncertainty of complex networks. In [48–54], the authors have demonstrated that better results of quantitative analyses of influence can be obtained by using entropy theory. In our previous work [55], the same conclusion has been drawn that defined entropy centrality has proven far superior to other widely used methods, such as degree centrality, betweenness centrality, closeness centrality, Eigenvector centrality, and PageRank. It is clear that the ideal algorithm, free of any limitations or assumptions, does not exist. The previous model was intentionally applicable to undirected, unweighted networks only and not general cases.

In this paper, with the intention of providing a more effective and more general framework to quantify the power of each actor in a directed, weighted network, we first study the properties of directed, weighted networks. Then, by generalizing the features of two-way behaviors between two

actors, we make direct use of directed networks to quantify influence. In particular, the total power of each actor in a given network can be calculated as the result of integrating the direct influence on its one-hop neighbors and the indirect influence on its two-hop neighbors. In terms of the direct influence mentioned above, we divide it into two parts measured by the structural entropy and the interaction frequency entropy, respectively. Generally, the former kind of information entropy reflects the communication activity, popularity, and strength of an actor, and the latter kind of information entropy, which deep mines the information carried by weights corresponding to the interaction frequency, mirrors the closeness among an actor and its neighbors. Moreover, we adopt the two-hop subnetworks to capture the process of influence propagation. In order to evaluate the effectiveness of the entropy centrality, we conduct experiments on four weighted networks including adolescent health, Bible, United States (US) airports and Hep-th. We also compare the performance of the entropy-based centrality with that of degree centrality, betweenness centrality, closeness centrality, and the Eigenvector centrality. Extensive experimental results prove that the proposed method has an obvious advantage in identifying influential nodes.

2. Model Description

As discussed above, in our previous work [55], we proposed a centrality method, which depicts the connections among node pairs by using Shannon's entropy and characterizes the process of influence spread by using two-hop subnetworks. However, the proposed method could be inappropriate in the case of directed and weighted networks. Thus, we modified the algorithm and extended its application scenarios.

Now, let us consider a directed, weighted network graph $G(V, E, W)$, where V denotes the set of vertices in a given network, E represents the set of directed edges from one node to another, and set W corresponds to the weighted values. For instance, set V of vertices corresponds to individual users in online social network, set E represents the traffic flow among users $i-j$, and set W of weighted values indicates the total number of any kinds of messages sent from user $i-j$. Examples abound in real-world networks.

With the purpose of describing the definition of the entropy centrality, we use a simple, directed, weighted network as an example, and the graph can be seen in Figure 1. In this simple network, each node corresponds to an airport. Each directed edge represents an airline from one airport to another. Correspondingly, the weight of each directed edge indicates the total number of flights on that connection in the given direction. The values of weights, denoted as C_{ij} , are listed in Table 1.

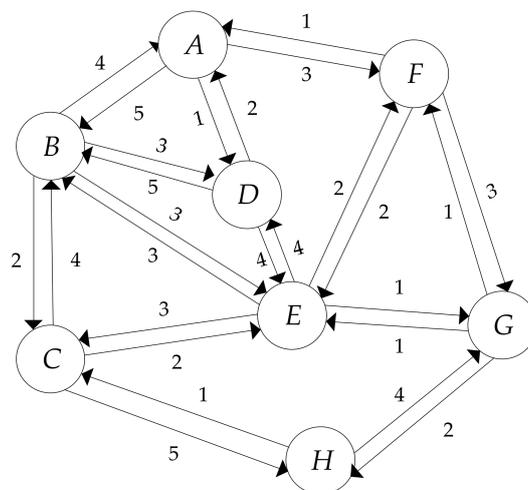


Figure 1. An example of a directed, weighted network.

Table 1. The number of airline connections from one airport to another.

Between Two Airports	The Number of Airlines	Between two Airports	The Number of Airlines
$A \rightarrow B$	5	$E \rightarrow B$	3
$A \rightarrow D$	1	$E \rightarrow C$	3
$A \rightarrow F$	3	$E \rightarrow D$	4
$B \rightarrow A$	4	$E \rightarrow F$	2
$B \rightarrow C$	2	$E \rightarrow G$	1
$B \rightarrow D$	3	$F \rightarrow A$	1
$B \rightarrow E$	3	$F \rightarrow G$	3
$C \rightarrow B$	4	$F \rightarrow E$	2
$C \rightarrow E$	2	$G \rightarrow E$	1
$C \rightarrow H$	5	$G \rightarrow F$	1
$D \rightarrow A$	2	$G \rightarrow H$	2
$D \rightarrow B$	5	$H \rightarrow C$	1
$D \rightarrow E$	4	$H \rightarrow G$	4

In this paper, we deconstruct the global power of a node into two parts, including its local power and its indirect power. The local power of a node indicates its accessibility, activity, popularity, and strength in the small world to which it belongs. Inspired by that thought, first, we deconstruct a complete network into several serval subnetworks centered on certain nodes. For example, Figure 2 shows the subgraph centered on node B , denoted as $subgraph_B$.

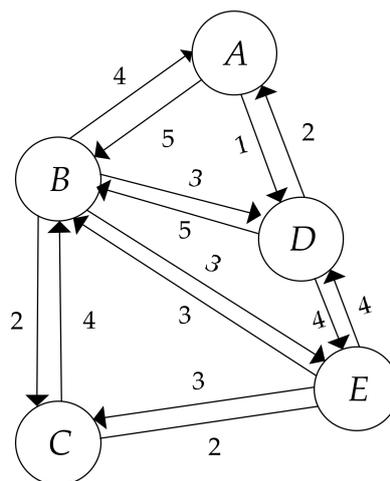


Figure 2. The subgraph constructed by node B and its neighbor nodes.

Then, we observe that each subnetwork contains various kinds of information that may be useful in the process of quantifying the local power, such as the in-degree centrality, out-degree centrality, and weights of edges. For instance, considering a random online social network, if user i follows user j , consequently there will be a directed link from node i to j . Thus, the out-degree centrality of a node reflects its social activity or its authority, and the in-degree centrality of a node interprets its social popularity. So, it seems more suitable to combine multiple information contained in different centralities. In light of that idea, we would like to introduce the following definition of the subgraph degree centrality.

Given a directed weighted graph $G(V, E, W)$, for a vertex $v_i \in V$, the i -centered subgraph represented by G_i can be built by node i and its neighbors. Accordingly, the subgraph degree centrality of node i and its neighbor j , denoted as SDC_i , equals the summation of its in-degree centrality and its out-degree centrality, namely

$$SDC_i = DC_i^{in} + DC_i^{out} \tag{1}$$

where DC_i^{in} represents the in-degree centrality of node i in the given subgraph (i.e., the number of nodes having a directed link pointing to node i in the given subgraph) and DC_i^{out} indicates the out-degree centrality of node i in the given subgraph (i.e., the number of directed edges from node i to other nodes in the given subgraph). Let us take Figure 2 as an example. In this particular subgraph, we count both in-degree and out-degree for each node on the basis of Equation (1). The results can be seen in Table 2.

Table 2. The results of SDC_i of nodes in subgraph G_B .

Node	DC_i^{in}	DC_i^{out}	SDC_i
B	4	4	8
A	2	2	4
C	2	2	4
D	3	3	6
E	3	3	6

Notice that Figure 2 illustrates two-way behaviors between node pairs. Clearly, it is a special case. Then, one may ask what the value of subgraph degree centrality will be if only one-way connection is showed (for example, in the citation network, a node may only have the in-degree centrality or the out-degree centrality). In that situation, the subgraph degree centrality is identical to the in-degree centrality or the out-degree centrality. As for the undirected, weighted networks, the subgraph degree centrality of a specific node should be the number of nodes connecting to the given node in the given subgraph. Notably, Equation (1) is also suitable for other types of networks, such as online social networks, email communication networks, collaboration networks, internet networks, etc.

In order to quantify the local power of a given node, we combine the advantages of both topological structure and information entropy. We believe that more precise influence ranking results will be obtained when information is properly utilized. Therefore, we propose a novel definition of entropy centrality, which takes both structural entropy and frequency entropy into consideration. The structural entropy, which takes advantage of topographic properties of the subgraph, evaluates the activity, popularity, and strength of a given node in specific subnetwork. The frequency entropy, which makes use of information contained in the weights of directed links, depicts the accessibility of a given node.

Based on the concept of subgraph degree centrality, the structural information entropy I_i^s for node i in subgraph G_i is defined as follows:

$$I_i^s = - \sum_{i=1}^{M+1} \frac{SDC_i}{\sum_i^{M+1} SDC_i} \log \frac{SDC_i}{\sum_i^{M+1} SDC_i} \tag{2}$$

where M denotes the number of neighbors of node i in subgraph G_i .

Generally, the weight of directed links acts as an effective indicator that reflects the interaction frequency. We believe that close relationships between node pairs is mainly maintained by their frequent interactions. Motivated by that, the definition of the interaction frequency entropy of node i , denoted as I_i^f , in subgraph G_i is stated as follows:

$$I_i^f = - \sum_{j=1}^M \frac{C_{ij}}{\sum_{k=1}^M C_{ik}} \log \frac{C_{ij}}{\sum_{k=1}^M C_{ik}} \tag{3}$$

where M indicates the total number of node i 's neighbors and C_{ij} denotes the weight of a directed edge in the given direction. Notably, C_{ij} in the Equation (3) should be replaced by the weights of the undirected edges, denoted as W_{ij} , for the undirected weighted networks. As explained above, the interaction frequency entropy of a given node indicates its accessibility to some extent.

In line with Equations (2) and (3), the local influence of node i on its one-hop neighbors, denoted as LI_i , equals the summation of the structural information entropy, denoted as I_i^s , and the interaction frequency entropy, denoted as I_i^f , multiplied by two coefficients respectively, namely

$$LI_i = \omega_1 I_i^s + \omega_2 I_i^f \tag{4}$$

where ω_1 and ω_2 represent the weight coefficients, respectively, and $\omega_1 + \omega_2 = 1$.

In our previous work [55], we were enlightened by the inspiring work done by Christakis [56,57] and other researchers [58–61] who found that meaningful influence can no longer be detectable beyond the boundary of three or four degrees. Consequently, we choose two-hop subnetworks to capture how influence propagates through the whole network. The key assumption of the two-hop influence propagating model is that we might not influence nor be influenced by people at three degrees and beyond. Consistent with analytical results by the empirical study of both artificial datasets and real-world datasets, the superiority of the previous method has been proven in comparison with other widely used approaches, such as degree centrality, betweenness centrality, closeness centrality, the Eigenvector centrality, and PageRank. In addition, the two-hop subnetwork used to measure the indirect power of a given node is concise. Accordingly, in this paper we adopt the same model to describe the process of influence propagation.

Suppose that node p is one of two-hop neighbors of node i , and node j is one of the common neighbors between node i and p . Let N_{ip} represent the number of common one-hop neighbor nodes between i and p . Let us take $N_{ip} = 2$ for example, which can be seen in Figure 3. Notice that two traffic flows exist from node i – p . As discussed above, we have already calculated the local power of each individual node. Consequently, the indirect influence of node i on its two-hop neighbor node p , denoted as II_{ip} , is stated as follows

$$II_{ip} = \frac{LI_i \times LI_j + LI_i \times LI_m}{2} \tag{5}$$

where LI_i , LI_j , and LI_m denote the local power of node i , j , and m , respectively.

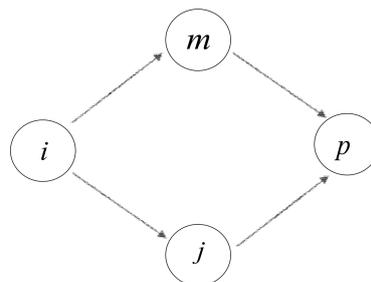


Figure 3. Double path.

Consistent with the above analysis, the indirect influence of node i on its two-hop neighbor p , denoted as II_{ip} , is stated as follows:

$$II_{ip} = \sum_{p=1}^{N_{ip}} \frac{LI_i \times LI_j}{N_{ip}} \tag{6}$$

where N_{ip} represents the total number of the common one-hop neighbors between node i and p .

Evidently, the indirect influence on its two-hop neighbors, denoted as II_i , is stated as follows:

$$II_i = \frac{\sum_{p=1}^{M_i} II_{ip}}{M_i} \tag{7}$$

where M_i corresponds to the number of two-hop neighbors of node i .

Eventually, the total power of node i in the given network $G(V, E, W)$ represented by I_i equals the summation of the direct influence, denoted as LI_i , and the indirect influence, denoted as II_i , multiplied by two coefficients, respectively, namely

$$I_i = \theta_1 LI_i + \theta_2 II_i \tag{8}$$

where θ_1 and θ_2 stands for the weight of local influence LI_i and indirect influence II_i , and correspondingly, $\theta_1 + \theta_2 = 1$.

We use the same network shown in Figure 1 as an example to describe the calculating process of the proposed algorithm. On the basis of Equation (1), the subgraph degree centrality, denoted as SDC , in G_B is listed in Table 2. In light of the values of SDC shown in Table 2, if 10 is the base of the logarithmic function, then the structural entropy of node B is computed as follows:

$$I_B^s = - \sum_{i=1}^5 \frac{SDC_B}{\sum_i^{4+1} SDC_i} \log \frac{SDC_B}{\sum_i^5 SDC_i} = 0.6836 \tag{9}$$

Furthermore, following Equation (3), the interaction frequency entropy is expressed as follows:

$$I_B^f = - \sum_{j=1}^4 \frac{C_{ij}}{\sum_{k=1}^4 C_{ik}} \log \frac{C_{ij}}{\sum_{k=1}^4 C_{ik}} = 0.5898 \tag{10}$$

We purposely applied the same sets of coefficients that were used in [54,55]. The authors have demonstrated that the entropy-based centrality outperforms the classic degree-based centralities and path-based centralities under the conditions of this particular set of parameters. By introducing this particular set of coefficients, the resulting value of influence will always be between zero and one.

Particularly, we set $\omega_1 = 0.4$ and $\omega_2 = 0.6$ and consequently redefined the local influence of node B , represented by LI_B , which is stated as follows:

$$LI_B = 0.4I_B^s + 0.6I_B^f = 0.6273 \tag{11}$$

In accordance with Equations (6) and (7), the indirect influence of node B on its two-hop neighbors, denoted as II_B , is given as follows:

$$II_B = (LI_B \times LI_E + (LI_B \times LI_A + LI_B \times LI_E)/2 + LI_B \times LI_C)/3 = 0.3713 \tag{12}$$

In particular, we set $\theta_1 = 0.6$ and $\theta_2 = 0.4$, and as a result, I_B corresponding to the total influence of node B in the given network is described as follows:

$$I_B = 0.6LI_B + 0.4II_B = 0.5249 \tag{13}$$

As discussed above, based on entropy centrality, the power of each node and the corresponding ranking results are recorded in Tables 3 and 4, respectively.

Table 3. The results of the redefined entropy centrality of each node.

Node	Local Influence	Indirect Influence	Total Influence
A	0.4736	0.2625	0.3892
B	0.6273	0.3713	0.5249
C	0.4955	0.3080	0.4205
D	0.5073	0.3283	0.4357
E	0.6956	0.3619	0.5521
F	0.4930	0.2915	0.4124
G	0.5004	0.2987	0.4197
H	0.3110	0.2323	0.2795

Table 4. The ranking results.

Node	No.
E	1
B	2
D	3
C	4
G	5
F	6
A	7
H	8

3. Performance Evaluation

In order to evaluate the performance of our proposed ranking algorithm, we use four real-world networks, which consist of two directed weighted networks and two undirected weighted networks. The four real-world networks are: (i) Adolescent health [62], a directed network with positively weighted edges. The data was generated from a survey conducted in 1994/1995. In that network, each node corresponds to a student who was asked to list his/her five best male friends and five best female friends. The directed edge between student i and j , denoted as e_{ij} , represents that student i chooses student j as his/her friend. Furthermore, higher weighted values indicate more interactions; (ii) (US) airports [63], a directed network of US infrastructure in 2010. Each node corresponds to an airport and the directed edge represents the airline connection from one airport to another. The weighted value shows the total number of flights on that connection in the given direction; (iii) Bible [63], an undirected network containing nouns (places and names) in the King James version of the Bible and information about their co-occurrences. A node represents one of the above noun types, and an edge indicates that two nouns appeared together in the same Bible verse. The weight of edge denotes how often two nouns occurred together; (iv) Hep-th [64], an undirected, weighted collaboration network with nodes corresponding to scientists posting preprints on the Hep-th and edges indicating collaborations. The assigned weights, which reflect the strength of collaborative ties can be obtained based on the number of preprints that each pair of researchers has done together and the number of other coauthors they worked with on each of those preprints. Table 5 shows the statistics of the four weighted networks mentioned above.

Table 5. The basic statistics of the four weighted networks.

Networks	n	m	c^1	AD^2
Adolescent health	2539	12,969	14.2%	10.216 (overall)
US airports	1574	28,236	38.4%	35.878 (overall)
Bible	1773	9131	16.3%	18.501
Hep-th	8361	15,757	32.7%	3.768

¹ c denotes the clustering coefficient. ² AD represents average degree.

In this paper, we apply the susceptible-infectious model to characterize the dynamic process of influence propagation. In the SI model, all nodes apart from the infected nodes are susceptible initially. A susceptible node will be infected by its infected neighbor nodes with probability β . However, in terms of the weighted networks, the infection probability is not a constant. The information contained in the directed edges should also be considered. In the literature, Yan et al. [65] defined the infection probability, denoted as λ_{ij} , as $\lambda_{ij} = \left(\frac{w_{ij}}{w_{max}}\right)^\alpha$, in which susceptible node i comes in contact with its infected neighbor node j and gets infected. Where α corresponds to a constant with a positive value, w_{ij} is the weighted value of the directed edge e_{ij} , and w_{max} denotes the largest value of w_{ij} . In addition, Wang et al. [66] introduced another kind of infection transmission. The probability that susceptible node i is infected by its neighbor i is stated as $1 - (1 - \beta)^{w_{ij}}$, where β is a positive constant and w_{ij} denotes the weighted value on the edge in the given direction. In this paper, we adopt the latter form of infection transmission proposed by Wang et al. [66].

In order to evaluate the efficiency of proposed method, we use the entropy-based centrality to select k nodes with most influential as seed nodes. In comparison, we also test the performances of degree centrality, betweenness centrality, closeness centrality, and the Eigenvector centrality. Then, the influence spread can be seen as an indicator of the algorithm's effectiveness. In order to obtain that index, we run the SI simulation on four networks 1000 times and select the mean value of the influence spread. Initially, we set the value of k as (10, 20, 30, 40, 50). Correspondingly, the results are illustrated in Figures 4 and 5.

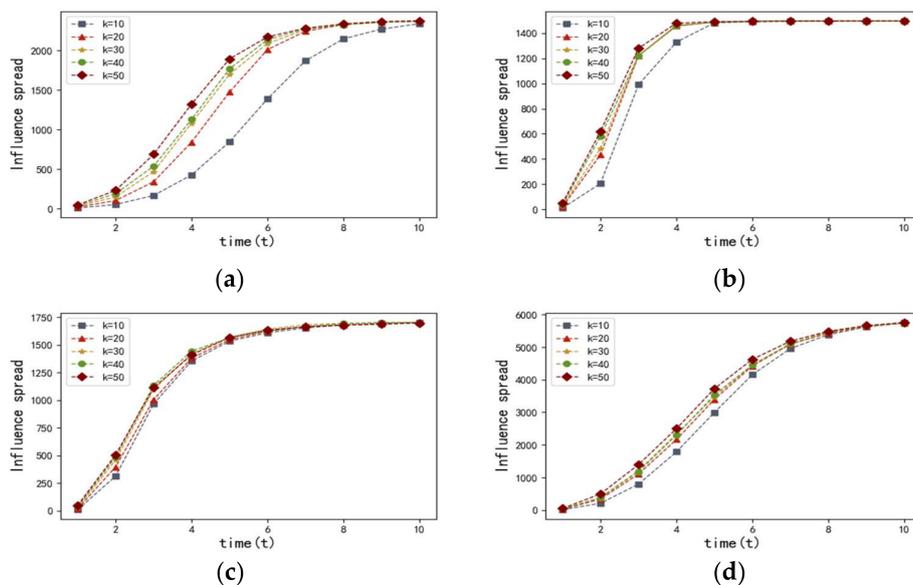


Figure 4. The influence spread with different k at time t . The results are obtained by using the entropy-based centrality in the four weighted networks including: (a) adolescent health; (b) US airports; (c) Bible, and (d) Hep-th, respectively.

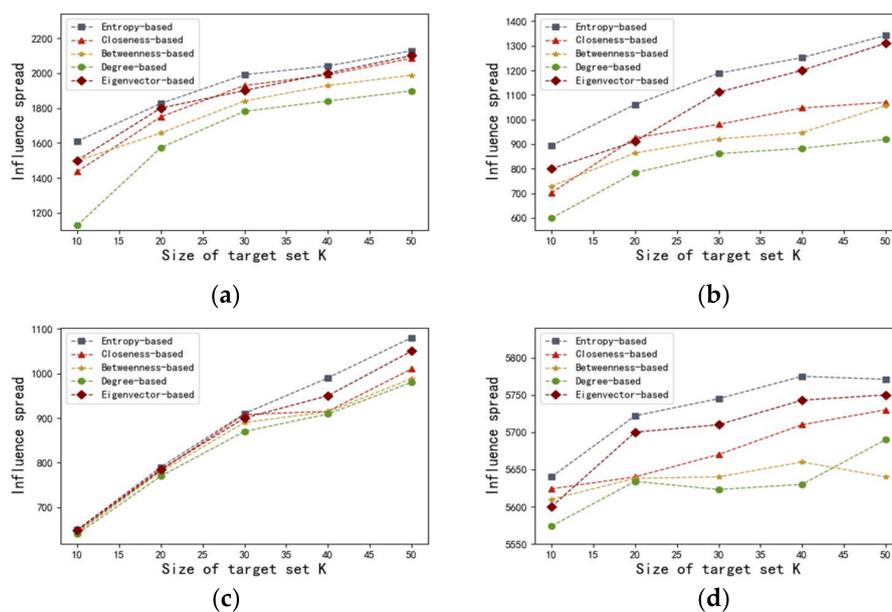


Figure 5. The influence spread with different k in the four weighted networks including: (a) adolescent health, (b) US airports, (c) Bible, and (d) Hep-th, respectively. The results are obtained by using the entropy-based centrality, degree centrality, betweenness centrality, and closeness centrality, respectively.

Figure 4 describes the influence spread of the proposed entropy-based centrality model with different k at time t in the four networks discussed above. The results of the four networks, which are shown in Figure 4, have proven that there is a positive correlation between the influence spread and the value of set k . We observed that the more most influential nodes there are, the more nodes can be influenced. It is also worth noting that the speed of infection transmission accelerates as the value of set k increases.

Figure 5 illustrates the influence spread of various centralities with different k sets of influential seeds for the four networks. As shown in Figure 5, in terms of all the networks, degree centrality, which belongs to the neighborhood-based centralities, performs badly. One possible reason is that degree centrality, which exclusively takes immediate neighbors' information into consideration, is no longer applicable to capture the process of influence propagation. Even though betweenness centrality performs better than degree centrality, the effectiveness of betweenness centrality is significantly inferior to that of closeness centrality and the Eigenvector centrality, as well as that of the entropy-based method we proposed. One plausible explanation is that supposing a large number of nodes are not contained in other node pairs' shortest paths, the values of betweenness should be zero, leading to many indistinguishable nodes with the same betweenness. Compared with closeness centrality, our entropy-based approach can obtain better results in the networks of adolescent health, US airports, and Hep-th, because the same k quantity of initial seeds sorted by entropy centrality eventually infected more nodes. Notice that the curves generated from closeness centrality and the entropy-based centrality are almost overlapping in the Bible network, which indicates that these two centralities show similar efficiency. As the value of k increases from 40 to 50, many more nodes have been infected by the initially infected nodes obtained by applying entropy centrality. Also, it is obvious that our proposed entropy centrality performs better in the networks of adolescent health, US airports, and Hep-th, because the nodes with higher entropy-based centrality, if infected, more quickly infect many more nodes in the networks in comparison with the Eigenvector centrality. In the Bible network, our proposed algorithm achieves better results than the Eigenvector centrality as the value of k increases from 30 to 50. Based on the results illustrated in Figure 5, it can be concluded that the Eigenvector centrality is more effective in identifying vital nodes in comparison with closeness centrality, at least in these four cases.

Figure 5 shows that the influence spread respectively corresponding to the entropy-based centrality and the Eigenvector centrality significantly increases as the value of k varies from 10 to 50. Finally, these two curves come closer and closer. In addition, identical sets of influential nodes (infected seeds) sorted by the entropy-based centrality infect many more susceptible nodes compared with those sorted by the other four centralities. In addition, the influence spread of the proposed model rises more significantly and quickly in comparison with that of other four centralities. In short, it can be concluded that the entropy-based centrality has an advantage in detecting influential spreaders and performs best in all four weighted real-world networks, namely adolescent health, US airports, Bible, and Hep-th. In fact, the performance of the Eigenvector centrality is the second best in the four weighted networks.

4. Conclusions and Discussion

In this paper, with the purpose of identifying the most influential spreaders, we propose a novel definition of entropy-based centrality. By studying the topological properties of the complex network, we introduce structural entropy and interaction frequency entropy. In general, by combining the advantages of the topological structure and information entropy, this entropy-based centrality can not only make full use of the information contained in neighbor nodes but also quantify influence from the perspective of information spreading. In order to verify the efficiency of the entropy-based centrality, we used four weighted real-world networks with varied instance sizes, degree distributions, and densities, including two directed networks and two undirected networks. When using the SI model,

the entropy-based centrality performed best when compared with degree centrality, betweenness centrality, closeness centrality, and the Eigenvector centrality on the basis of extensive analytical results.

It is also noteworthy that there are similarities and differences between the proposed method and other structural centralities. For example, both the proposed approach and degree centrality are based on the idea that the power of a given node can be reflected by its capacity to influence the behaviors of its surrounding neighbors. However, degree centrality fails to capture the process of influence propagation compared with the proposed method. Both ClusterRank [36] and the proposed method take the number of immediate neighbors into consideration. However, the ClusterRank algorithm [36] uses clustering coefficients to describe the information spreading process. Both the proposed method and the information index [32] are built on the assumption that information will be lost during every hop in the network, and therefore, the longer the path, the greater the loss. However, the information index considers information contained in all possible paths from a given node to others. Unlike the proposed method, closeness centrality [29,30], betweenness centrality [23], and eccentricity [31] compute the influence of a given node by measuring the shortest paths. In the literature, the mapping entropy [52], which is fully based on the local information contained in a given node and its immediate neighbors, fails to capture the propagating process. In comparison with the proposed method, which is established on the basis of information entropy, another kind of entropy-based centrality proposed by Fei and Deng [53] takes advantage of relative entropy and the TOPSIS method. They also treat centralities from various measures as multiple attributes in the process of quantifying influence.

However, it is clear that perfect algorithms, free of any limitations or assumptions, do not exist. There is still space to improve our entropy-based centrality. Firstly, in the calculation process, we find that the entropy-based centrality values of some nodes are extremely small and consequently indistinguishable. Secondly, the model of determining direct influence is highly neighborhood-based, resulting in a few indistinguishable nodes with the same entropy-based centrality. Thirdly, the entropy-based centrality is no longer applicable to the networks with negatively weighted edges because the base of the logarithm function, which is the foundation of the information entropy, must be positive. However, there are still plenty of bipartite networks with both positive and negative edges, such as Epinions trust, Wikipedia conflict, Chess, Wikipedia elections, and so on. Accordingly, researchers suggest plenty of algorithms, for instance, the group-based ranking approach proposed by Gao et al. [67] to measure a user's reputation, the correlation-based iterative method [68], the iterative algorithm with reputation redistribution [69], and the HITs method for bipartite networks [70].

As for future work, we expect to further improve the entropy-based centrality. We are also looking forward to applying the entropy-based centrality in the real world, such as in delivering advertisements for companies, predicting career movements, constructing the recommender systems, studying the interdisciplinary knowledge network of China [71] and evaluating airports in China.

Acknowledgments: The authors are very grateful for the insightful comments and suggestions of the anonymous reviewers and the editor, which have helped to significantly improve this article. Furthermore, this research was supported by the National Natural Science Foundation of China (No. 71371025), the Beijing Natural Science Foundation (No. 9182010), and the Aeronautical Science Fund (No. 2016ZG51058).

Author Contributions: Conceptualization, methodology, algorithms, design of experiments, and the writing of the original draft were done by Tong Qiao; topic selection, framework construction, problem formulation, formal analysis, and manuscript revision were done by Wei Shan; code writing, computations, data analysis, and visualization were done by Ganjun Yu and Chen Liu. All the authors equally contributed to proofreading the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rabade, R.; Mishra, N.; Sharma, S. Survey of influential user identification techniques in online social networks. In *Advances in Intelligent Systems and Computing*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 235, pp. 359–370.

2. Akritidis, L.; Katsaros, D.; Bozani, P. Identifying the productive and influential bloggers in a community. *IEEE Trans. Syst. Man Cybern. Part C* **2011**, *41*, 759–764. [[CrossRef](#)]
3. Gonzálezbailón, S.; Borgehothoefer, J.; Rivero, A.; Moreno, Y. The dynamics of protest recruitment through an online network. *Sci. Rep.* **2011**, *1*, 197. [[CrossRef](#)] [[PubMed](#)]
4. Alzaabi, M.; Taha, K.; Martin, T.A. Cisri: A crime investigation system using the relative importance of information spreaders in networks depicting criminals communications. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 2196–2211. [[CrossRef](#)]
5. Ghosh, R.; Lerman, K. Parameterized centrality metric for network analysis. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2011**, *83*, 066118. [[CrossRef](#)] [[PubMed](#)]
6. Min, L.; Zhang, H.; Wang, J.; Yi, P. A new essential protein discovery method based on the integration of protein-protein interaction and gene expression data. *BMC Syst. Biol.* **2012**, *6*, 15.
7. Peng, W.; Wang, J.; Wang, W.; Liu, Q.; Wu, F.X.; Pan, Y. Iteration method for predicting essential proteins based on orthology and protein-protein interaction networks. *BMC Syst. Biol.* **2012**, *6*, 87. [[CrossRef](#)] [[PubMed](#)]
8. Zhang, X.; Xu, J.; Xiao, W.X. A new method for the discovery of essential proteins. *PLoS ONE* **2013**, *8*, e58763. [[CrossRef](#)] [[PubMed](#)]
9. Luo, J.; Qi, Y. Identification of essential proteins based on a new combination of local interaction density and protein complexes. *PLoS ONE* **2015**, *10*, e0131418. [[CrossRef](#)] [[PubMed](#)]
10. Li, M.; Lu, Y.; Niu, Z.; Wu, F.X. United complex centrality for identification of essential proteins from ppi networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *14*, 370–380. [[CrossRef](#)] [[PubMed](#)]
11. Radicchi, F.; Fortunato, S.; Markines, B.; Vespignani, A. Diffusion of scientific credits and the ranking of scientists. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2009**, *80*, 056103. [[CrossRef](#)] [[PubMed](#)]
12. Zhou, Y.B.; Lü, L.; Li, M. Quantifying the influence of scientists and their publications: Distinguish prestige from popularity. *New J. Phys.* **2011**, *14*, 33033–33049. [[CrossRef](#)]
13. Ding, Y. Applying weighted PageRank to author citation networks. *Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 236–245. [[CrossRef](#)]
14. Stefano, B.; Michelangelo, P.; Rahul, K.; Paolo, T.; Guido, C. Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Sci. Rep.* **2012**, *2*, 541.
15. Tabak, B.M.; Souza, S.R.S.; Guerra, S.M. *Assessing the Systemic Risk in the Brazilian Interbank Market*; Working Paper; Banco Central do Brasil: Brasilia, Brazil, 2013.
16. Mossholder, K.W.; Settoon, R.P.; Henagan, S.C. A relational perspective on turnover: examining structural, attitudinal, and behavioral predictors. *Acad. Manag. J.* **2005**, *48*, 607–618. [[CrossRef](#)]
17. Cuadra, L.; Salcedo-Sanz, S.; Ser, J.D.; Jiménez-Fernández, S.; Zong, W.G. A critical review of robustness in power grids using complex networks concepts. *Energies* **2015**, *8*, 9211–9265. [[CrossRef](#)]
18. Cuadra, L.; Pino, M.; Nieto-Borge, J.; Salcedo-Sanz, S. Optimizing the structure of distribution smart grids with renewable generation against abnormal conditions: A complex networks approach with evolutionary algorithms. *Energies* **2017**, *10*, 1097. [[CrossRef](#)]
19. Pagani, G.A.; Aiello, M. From the grid to the smart grid, topologically. *Phys. A Stat. Mech. Appl.* **2016**, *449*, 160–175. [[CrossRef](#)]
20. Omodei, E.; Arenas, A. A network approach to decentralized coordination of energy production-consumption grids. *PLoS ONE* **2018**, *13*, e0191495. [[CrossRef](#)] [[PubMed](#)]
21. Feeley, T.H.; Moon, S.I.; Kozey, R.S.; Slowe, A.S. An erosion model of employee turnover based on network centrality. *J. Appl. Commun. Res.* **2010**, *38*, 167–188. [[CrossRef](#)]
22. Yuan, J.; Zhang, Q.M.; Gao, J.; Zhang, L.; Wan, X.S.; Yu, X.J.; Zhou, T. Promotion and resignation in employee networks. *Phys. A Stat. Mech. Appl.* **2016**, *444*, 442–447. [[CrossRef](#)]
23. Freeman, L.C. A set of measures of centrality based on betweenness. *Sociometry* **1977**, *40*, 35–41. [[CrossRef](#)]
24. Bonacich, P. Power and centrality: A family of measures. *Am. J. Sociol.* **1987**, *92*, 1170–1182. [[CrossRef](#)]
25. Borgatti, S.P. Centrality and network flow. *Soc. Netw.* **2005**, *27*, 55–71. [[CrossRef](#)]
26. Borgatti, S.P.; Everett, M.G. A Graph-theoretic perspective on centrality. *Soc. Netw.* **2006**, *28*, 466–484. [[CrossRef](#)]
27. Lü, L.; Chen, D.; Ren, X.L.; Zhang, Q.M.; Zhang, Y.C.; Zhou, T. Vital nodes identification in complex networks. *Phys. Rep.* **2016**, *650*, 1–63. [[CrossRef](#)]
28. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **1953**, *18*, 39–43. [[CrossRef](#)]

29. Sabidussi, G. The centrality index of a graph. *Psychometrika* **1966**, *31*, 581–603. [[CrossRef](#)] [[PubMed](#)]
30. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1979**, *1*, 215–239. [[CrossRef](#)]
31. Hage, P.; Harary, F. Eccentricity and centrality in networks. *Soc. Netw.* **1995**, *17*, 57–63. [[CrossRef](#)]
32. Stephenson, K.; Zelen, M. Rethinking centrality: Methods and examples. *Soc. Netw.* **1989**, *11*, 1–37. [[CrossRef](#)]
33. Chen, D.; Lü, L.; Shang, M.S.; Zhang, Y.C.; Zhou, T. Identifying influential nodes in complex networks. *Phys. A Stat. Mech. Appl.* **2012**, *391*, 1777–1787. [[CrossRef](#)]
34. Gao, C.; Wei, D.; Hu, Y.; Mahadevan, S.; Deng, Y. A modified evidential methodology of identifying influential nodes in weighted networks. *Phys. A Stat. Mech. Appl.* **2013**, *392*, 5490–5500. [[CrossRef](#)]
35. Petermann, T.; Rios, P.D.L. Role of clustering and gridlike ordering in epidemic spreading. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2004**, *69*, 066116. [[CrossRef](#)] [[PubMed](#)]
36. Chen, D.B.; Gao, H.; Lü, L.; Zhou, T. Identifying influential nodes in large-scale directed networks: The role of clustering. *PLoS ONE* **2013**, *8*, e77455. [[CrossRef](#)] [[PubMed](#)]
37. Kitsak, M.; Gallos, L.K.; Havlin, S.; Liljeros, F.; Muchnik, L.; Stanley, H.E.; Makse, H.A. Identification of influential spreaders in complex networks. *Nat. Phys.* **2011**, *6*, 888–893. [[CrossRef](#)]
38. Zeng, A.; Zhang, C.J. Ranking spreaders by decomposing complex networks. *Phys. Lett. A* **2013**, *377*, 1031–1035. [[CrossRef](#)]
39. Pei, S.; Muchnik, L.; Andrade, J.S., Jr.; Zheng, Z.; Makse, H.A. Searching for superspreaders of information in real-world social media. *Sci. Rep.* **2014**, *4*, 5547. [[CrossRef](#)] [[PubMed](#)]
40. Liu, J.G.; Ren, Z.M.; Guo, Q. Ranking the spreading influence in complex networks. *Phys. A Stat. Mech. Appl.* **2013**, *392*, 4154–4159. [[CrossRef](#)]
41. Hu, Q.; Gao, Y.; Ma, P.; Yin, Y.; Zhang, Y.; Xing, C. A new approach to identify influential spreaders in complex networks. *Acta Phys. Sin.* **2013**, *62*, 99–104.
42. Min, B.; Liljeros, F.; Makse, H.A. Finding influential spreaders from human activity beyond network location. *PLoS ONE* **2015**, *10*, e0136831. [[CrossRef](#)] [[PubMed](#)]
43. Liu, Y.; Tang, M.; Zhou, T.; Do, Y. Improving the accuracy of the k-shell method by removing redundant links: From a perspective of spreading dynamics. *Sci. Rep.* **2015**, *5*, 13172. [[CrossRef](#)] [[PubMed](#)]
44. Hirsch, J.E. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16569–16572. [[CrossRef](#)] [[PubMed](#)]
45. Braun, T.; Glänzel, W.; Schubert, A. A hirsch-type index for journals. *Scientometrics* **2006**, *69*, 169–173. [[CrossRef](#)]
46. Hirsch, J.E. Does the h index have predictive power? *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19193–19198. [[CrossRef](#)] [[PubMed](#)]
47. Lü, L.; Zhou, T.; Zhang, Q.M.; Stanley, H.E. The H-index of a network node and its relation to degree and coreness. *Nat. Commun.* **2016**, *7*, 10168. [[CrossRef](#)] [[PubMed](#)]
48. Cao, S.; Dehmer, M.; Shi, Y. Extremality of degree-based graph entropies. *Inf. Sci.* **2014**, *278*, 22–33. [[CrossRef](#)]
49. Chen, Z.; Dehmer, M.; Shi, Y. Bounds for degree-based network entropies. *Appl. Math. Comput.* **2015**, *265*, 983–993. [[CrossRef](#)]
50. Nikolaev, A.G.; Razib, R.; Kucheriya, A. On efficient use of entropy centrality for social network analysis and community detection. *Soc. Netw.* **2015**, *40*, 154–162. [[CrossRef](#)]
51. Cao, S.; Dehmer, M. Degree-based entropies of networks revisited. *Appl. Math. Comput.* **2015**, *261*, 141–147. [[CrossRef](#)]
52. Nie, T.; Guo, Z.; Zhao, K.; Lu, Z.M. Using mapping entropy to identify node centrality in complex networks. *Phys. A Stat. Mech. Appl.* **2016**, *453*, 290–297. [[CrossRef](#)]
53. Fei, L.; Deng, Y. A new method to identify influential nodes based on relative entropy. *Chaos Solitons Fractals* **2017**, *104*, 257–267. [[CrossRef](#)]
54. Peng, S.; Yang, A.; Cao, L.; Yu, S.; Xie, D. Social influence modeling using information theory in mobile social networks. *Inf. Sci.* **2017**, *379*, 146–159. [[CrossRef](#)]
55. Qiao, T.; Shan, W.; Zhou, C. How to identify the most powerful node in complex networks? A novel entropy centrality approach. *Entropy* **2017**, *19*, 614. [[CrossRef](#)]
56. Christakis, N.A.; Fowler, J.H. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*; Little, Brown: New York, NY, USA, 2011.
57. Christakis, N.A.; Fowler, J.H. Social contagion theory: Examining dynamic social networks and human behavior. *Stat. Med.* **2013**, *32*, 556–577. [[CrossRef](#)] [[PubMed](#)]

58. Brown, J.J.; Reingen, P.H. Social ties and word of mouth referral behavior. *J. Consum. Res.* **1987**, *14*, 350–362. [[CrossRef](#)]
59. Singh, J. Collaborative networks as determinants of knowledge diffusion patterns. *Manag. Sci.* **2005**, *51*, 756–770. [[CrossRef](#)]
60. McDermott, R.; Fowler, J.H.; Christakis, N.A. Breaking up is hard to do, unless everyone else is doing it too: Social network effects on divorce in a longitudinal sample. *Soc. Forces* **2013**, *92*, 491–519. [[CrossRef](#)] [[PubMed](#)]
61. Mednick, S.C.; Christakis, N.A.; Fowler, J.H. The spread of sleep loss influences drug use in adolescent social networks. *PLoS ONE* **2010**, *5*, e9775. [[CrossRef](#)] [[PubMed](#)]
62. Moody, J. Peer influence groups: Identifying dense clusters in large networks. *Soc. Netw.* **2001**, *23*, 261–283. [[CrossRef](#)]
63. Konect. Available online: <http://konect.uni-koblenz.de/networks> (accessed on 25 April 2017).
64. Newman, M.E. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 404–409. [[CrossRef](#)] [[PubMed](#)]
65. Yan, G.; Zhou, T.; Wang, J.; Fu, Z.-Q.; Wang, B.-H. Epidemic spread in weighted scale-free networks. *Chin. Phys. Lett.* **2005**, *22*, 510.
66. Wang, W.; Tang, M.; Zhang, H.F.; Gao, H.; Do, Y.; Liu, Z.H. Epidemic spreading on complex networks with general degree and weight distributions. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2014**, *90*, 042803. [[CrossRef](#)] [[PubMed](#)]
67. Gao, J.; Dong, Y.W.; Shang, M.; Cai, S.M.; Zhou, T. Group-based ranking method for online rating systems with spamming attacks. *Europhys. Lett.* **2015**, *110*, 28003. [[CrossRef](#)]
68. Zhou, Y.; Lei, T.; Zhou, T. A robust ranking algorithm to spamming. *Europhys. Lett.* **2001**, *94*, 1034–1054. [[CrossRef](#)]
69. Liao, H.; Zeng, A.; Xiao, R.; Ren, Z.M.; Chen, D.B.; Zhang, Y.C. Ranking reputation and quality in online rating systems. *PLoS ONE* **2014**, *9*, e97146. [[CrossRef](#)] [[PubMed](#)]
70. Liao, H.; Xiao, R.; Cimini, G.; Medo, M. Network-driven reputation in online scientific communities. *PLoS ONE* **2014**, *9*, e112022. [[CrossRef](#)] [[PubMed](#)]
71. Liu, C.; Shan, W.; Yu, J. Shaping the interdisciplinary knowledge network of china: a network analysis based on citation data from 1981 to 2010. *Scientometrics* **2011**, *89*, 89–106. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).