MDPI

Article

Identifying Influential Nodes in Complex Networks Based on Multiple Local Attributes and Information Entropy

Jinhua Zhang [1], **Qishan Zhang** [1], **Ling Wu** [2] **and Jinxin Zhang** [3],*

1   School of Economics and Management, Fuzhou University, Fuzhou 350108, China;
    zhangjin-hua@foxmail.com (J.Z.); zhangqs@fzu.edu.cn (Q.Z.)
2   College of Computer and Data Science, Fuzhou University, Fuzhou 350108, China; wuling1985@fzu.edu.cn
3   School of Business, Hubei University, Wuhan 430062, China
*   Correspondence: zhangjinxin999@foxmail.com

Abstract: Identifying influential nodes in complex networks has attracted the attention of many researchers in recent years. However, due to the high time complexity, methods based on global attributes have become unsuitable for large-scale complex networks. In addition, compared with methods considering only a single attribute, considering multiple attributes can enhance the performance of the method used. Therefore, this paper proposes a new multiple local attributes-weighted centrality (LWC) based on information entropy, combining degree and clustering coefficient; both one-step and two-step neighborhood information are considered for evaluating the influence of nodes and identifying influential nodes in complex networks. Firstly, the influence of a node in a complex network is divided into direct influence and indirect influence. The degree and clustering coefficient are selected as direct influence measures. Secondly, based on the two direct influence measures, we define two indirect influence measures: two-hop degree and two-hop clustering coefficient. Then, the information entropy is used to weight the above four influence measures, and the LWC of each node is obtained by calculating the weighted sum of these measures. Finally, all the nodes are ranked based on the value of the LWC, and the influential nodes can be identified. The proposed LWC method is applied to identify influential nodes in four real-world networks and is compared with five well-known methods. The experimental results demonstrate the good performance of the proposed method on discrimination capability and accuracy.

Keywords: complex networks; influential nodes; direct influence; indirect influence; information entropy

1. Introduction

Many systems in the real world can be modeled as complex networks to facilitate the study of the systems by complex network analysis. In a complex network, compared with other general nodes, some nodes can greatly affect the function of the whole network. These nodes are called influential nodes, and the quantity of these nodes is small. Ranking and identifying influential nodes in complex networks is one of the most important and basic problems in complex network research.

A variety of measures and methods have been proposed in recent decades for identifying the influential nodes of complex networks [1–3]. The most commonly used measures are centrality measures, such as degree centrality [4], closeness centrality [5,6], betweenness centrality [7,8], Katz centrality [9,10], and k-shell [11]. Degree centrality is the simplest metric that considers only the direct neighbors of a target node. Therefore, the degree centrality has low time complexity. However, its accuracy is also low since it ignores the global structure information of the network. Path-based centralities, such as closeness centrality, betweenness centrality, and Katz centrality can get better results. However, as global metrics, they are not suitable for large-scale networks due to their high computational complexity. Kitsak et al. [11] found that the influence of a node is highly related to

the position of the node in a network, so they proposed to apply the k-shell decomposition method to identify the influential nodes of the network. However, k-shell decomposition is a coarse-grained method; the same k-shell value is assigned to a large number of nodes. Therefore, scholars have proposed a series of improved methods based on k-shell decomposition [12], such as mixed degree decomposition [13], improved neighbors' k-core [14], neighborhood coreness [15], weighted k-shell [16], k-shell iteration factor [17], extended local k-shell sum [18], influence capability [19], hierarchical k-shell [20], integral k-shell [21], improved k-shell hybrid method [22], etc. As for the field of the search engine, there are also some useful methods, such as PageRank [23], LeaderRank [24], and Hits [25]. Other centrality measures include H-index [26,27] and several measures based on the clustering coefficient, such as local centrality [28] and clustered local-degree [29], etc. In addition, information entropy has also been used to evaluate the influence of nodes in complex networks [30–34].

In recent years, some scholars have made practical explorations from new perspectives. Some researchers have found that the influence of nodes includes direct local influence and indirect influence based on neighbors. Therefore, these researchers constructed influential node identification models based on direct influence and indirect influence [30,31,33,35,36]. Some other researchers have found that considering multiple attributes can achieve better node influence evaluation results than considering only a single attribute. Wang et al. [19] applied a multi-attribute ranking method based on a node's position and neighborhood information. Yang et al. [37] applied TOPSIS to identify influential nodes and combined it with grey relational analysis to rank the importance of complex network nodes. Mo and Deng [38] adopted D-S evidence theory to comprehensively consider degree centrality, betweenness centrality, efficiency centrality, and correlation centrality for evaluating the importance of nodes in complex networks. These multi-attributes methods have proved to be promising strategies to evaluate the influence of nodes. However, there are still some limitations in existing multi-attribute methods, such as some methods being based on global attributes during the evaluation, which results in high computational complexity, unsuitable for current fast-growing social networks. In some methods, the contribution of each attribute is viewed as equal, or the weights of different attributes are set manually, which results in limited improvement in evaluating performance.

Degree and clustering coefficient are two critical attributes in node influence evaluation methods. In particular, degree and clustering coefficient belong to local attributes and have low computational complexity. Meanwhile, Liu et al. [39] deeply studied the role of neighbors on the influence of nodes and concluded that considering the one-step or two-step neighborhood of a node can achieve the best performance in most networks. Inspired by the above discussion, we propose a new multi-attribute weighted centrality based on information entropy, combining degree and clustering coefficient; both one-step and two-step neighborhood information are considered. The main contributions of this paper are summarized as follows. First, this paper considers the indirect influence of nodes and proposes a new measure of indirect influence, i.e., the two-hop clustering coefficient. Second, this paper improves the calculation formula of the two-hop degree. In the existing two-hop degree calculation, the degrees of one-hop neighbors are directly added. In our proposed two-hop degree formula, we take a more reasonable process, eliminating duplicate neighbors and then adding the neighbors of one-hop neighbors. Last but not least, this paper proposes a new multi-attribute weighted centrality with a low computational complexity O(n), as it is calculated based on four local attributes, including degree, two-hop degree, clustering coefficient, and two-hop clustering coefficient. Additionally, the entropy weighting method is exploited to weight the contributions of the four local attributes, rather than weight them equally.

The rest of this paper is organized as follows. In Section 2, relevant knowledge is introduced as a preliminary, including the six kinds of well-known centrality measures and the information entropy theory. In Section 3, the proposed multi-attribute weighted centrality is described. In Section 4, the proposed method is applied to four real world

networks and compared with five well-known methods to verify the effectiveness and practicability of our proposed method. Section 5 concludes the paper.

2. Preliminaries

2.1. Typical Centrality Measures

The influence of a node can be referred to as centrality. Scholars have proposed a variety of centrality measures to evaluate and rank the influence of nodes. In this section, we briefly introduce six typical centrality measures, including degree, betweenness centrality, closeness centrality, local centrality, clustering coefficient, and clustered local degree. In the following part of this paper, new centrality measures based on the degree and clustering coefficient are proposed, and some of these centrality measures will be used as comparison methods for experiments.

Definition 1. *(Degree [4]). The degree of node i, denoted as k_i, which is defined as*

$$k_i = \sum_{j}^{N} a_{ij} \tag{1}$$

where N is the node set, node $i, j \in N$, and a_{ij} refers to the connection between node i and node j. The value of a_{ij} is defined as $a_{ij} = \begin{cases} 1, & \text{node i connected with j} \\ 0, & \text{otherwise} \end{cases}$.

Definition 2. *Betweenness centrality [7,8]. The betweenness centrality of node i, denoted as b_i, is defined as*

$$b_i = \sum_{j \neq s \neq i} \frac{n_{js}(i)}{n_{js}} \tag{2}$$

where n_{js} refers to the number of all the shortest paths between node j and node s, and $n_{js}(i)$ denotes the number of those shortest paths which go through node i.

Definition 3. *Closeness centrality [5,6]. The closeness centrality of node i, denoted as c_i, is defined as*

$$c_i = \frac{1}{\sum_{j}^{N} d_{ij}} \tag{3}$$

where d_{ij} refers to the distance between node i and node j. If there is no path between node i and node j, $d_{ij} = \infty$, then $\frac{1}{d_{ij}} = 0$ [5,40].

Definition 4. *Local centrality [28]. The local centrality of node v, denoted as $C_L(v)$, is defined as*

$$Q(u) = \sum_{w \in \Gamma(u)} N(w) \tag{4}$$

$$C_L(v) = \sum_{u \in \Gamma(v)} Q(u) \tag{5}$$

where $\Gamma(u)$ is the set of the one-hop neighbors of node u, $\Gamma(v)$ is the set of the one-hop neighbors of node v, and $N(w)$ is the number of the one-hop and the two-hop neighbors of node w.

Definition 5. *Clustering coefficient [37]. The clustering coefficient of node i, denoted as clc_i, is defined as*

$$clc_i = 2e_i / (k_i(k_i - 1)) \tag{6}$$

where k_i refers to the degree of node i, which can be calculated as Formula (1), and e_i represents the number of edges connected between the neighbors of node i.

Definition 6. *Clustered local degree* [29]. *The clustered local degree of node i, denoted as cld_i, is defined as*

$$cld_i = (1 + cc_i) \sum_{j \in N(i)} k_i \tag{7}$$

where $N(i)$ is the set of the one-hop neighbors of node i, cc_i is the clustering coefficient of node i, which can be calculated as Formula (6), and k_i refers to the degree of node i.

2.2. Information Entropy and Entropy Weighting Method

Information entropy is a famous concept of information theory, which was first proposed by Claude Elwood Shannon in his work "A Mathematical Theory of Communication" in 1948. It is a useful tool for measuring the uncertainty of information. In general, the more uncertain or random an event is, the more information it will contain. Shannon's information entropy has been applied in many areas, such as data mining, statistical inference, machine learning, image processing, and so on.

According to Shannon's information theory, the information entropy is defined as follows:

$$H(X) = H(x_1, x_2, \ldots, x_n) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{8}$$

where $X = \{x_1, x_2, \ldots, x_n\}$ is a set of possible events, and $p(x_i)$ is the probability of event x_i.

The entropy weighting method is one of the most important applications of information entropy in the field of social science. Due to its excellent performance, the entropy weighting method is widely used to determine the weights of different attributes in multi-attribute decision-making problems [41,42], and it is also widely used in node importance evaluation in complex networks. Therefore, in this paper, we also use the entropy weighting method to determine the weight of the four local attributes.

Suppose there are n attributes to be considered. The weight of attribute i, denoted as w_i, is calculated as:

$$w_i = \frac{1 - H_i}{\sum_1^n (1 - H_i)} \tag{9}$$

where $H_i(i = 1, 2, \ldots, n)$ is the information entropy of each attribute [19,43].

3. Method Description

3.1. Direct Influence of a Node with Respect to Local Attributes

As introduced in Section 1, the total influence of a node can be divided into direct influence and indirect influence. Direct influence is the influence of a node on the nodes directly connected to it, that is, the influence exerted on its one-hop neighbors. Considering the degree and clustering coefficient, these two indicators belong to local indicators, which are representative and are simple to calculate. In this study, we selected the degree and clustering coefficient to represent the direct influence of nodes. For the calculation of degree k_i and clustering coefficient clc_i, see Definition 1 and Definition 5, respectively.

3.2. Indirect Influence of a Node on Two-Hop Attributes

The indirect influence of a node is the influence it exerts on nodes that are not directly connected to it. Liu et al. [40] have pointed out that the importance of a node is not only determined by its direct neighbors, but also depends on its multi-step neighbor information. According to their experimental results, for the sake of balancing accuracy and efficiency, considering two-step neighbor information is the best choice. Based on the one-hop neighborhood degree and clustering coefficient, we propose the two-hop degree and two-hop clustering coefficient as indirect influence measures. Wang et al. [19] have also utilized two-hop degree information in their research. However, in their two-hop degree formula, they directly sum the degree of a node's one-hop neighbors, which ignores the overlap of different node's neighbors. Therefore, in our two-hop degree formula, duplicate nodes are eliminated and calculated only once.

Definition 7. *Two-hop degree. The two-hop degree of node i, denoted as thk_i, is defined as*

$$thk_i = NN_i - NS_i - 1 \tag{10}$$

where NN_i is the number of two-hop neighbors of node i, and NS_i denotes the number of same neighbors.

Definition 8. *Two-hop clustering coefficient. The two-hop clustering coefficient of node i, denoted as $thclc_i$, is defined as*

$$thclc_i = \sum_{j \in N(i)} clc_j \tag{11}$$

where clc_j is the clustering coefficient of node j, which is defined in Definition 5; $N(i)$ denotes the neighbors of node i.

3.3. The Multiple Local Attributes Weighted Centrality

In this section, we propose a new multiple local attributes weighted centrality based on the above-mentioned direct influence measures and indirect influence measures for identifying influential nodes in a complex network. For the sake of brevity, we will abbreviate the multiple local attributes weighted centrality as LWC. The basic idea of the LWC is to view each node of the network as a decision scheme and each attribute of the node as an index of the decision scheme. Since different indices have different contributions to the influence of a node, we apply the entropy weight method to calculate the weight of each index; then, the influence of a node is the weighted sum of all indices. The steps of the LWC are described as follows.

Step 1: Construct the multi-attribute decision-making matrix of the node influence. For a complex network with n nodes and the node set $V = \{v_1, v_2, \ldots, v_n\}$, the decision scheme can be expressed as $D = \{d_1, d_2, \ldots, d_n\}$. In this paper, we adopt the above mentioned four local attributes, degree, two-hop degree, clustering coefficient, and two-hop clustering coefficient, as influence evaluation indices, so the evaluation indices can be expressed as $A = \{a_1, a_2, a_3, a_4\} = \{k, thk, clc, thclc\}$. Then, $d_i(a_j)(i = 1, 2, \ldots, n; j = 1, 2, 3, 4)$ represents the jth index of node i, and the multi-attribute decision-making matrix P can be obtained as follows.

$$P = \begin{bmatrix} d_1(a_1) & d_1(a_2) & d_1(a_3) & d_1(a_4) \\ d_2(a_1) & d_2(a_2) & d_2(a_3) & d_2(a_4) \\ \cdots & \cdots & \cdots & \cdots \\ d_n(a_1) & d_n(a_2) & d_n(a_3) & d_n(a_4) \end{bmatrix} \tag{12}$$

Step 2: Normalize the multi-attribute decision-making matrix. Due to each index having a different dimension, the matrix P should be standardized into a normalized matrix. First, set $d_{ij} = d_i(a_j)/max\{d_1(a_j), d_2(a_j), \ldots, d_n(a_j)\}$, then normalize the indices $r_{ij} = d_{ij}/\sum_{i=1}^{n} d_{ij}$, and then the normalized matrix R can be obtained as follows:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & r_{n3} & r_{n4} \end{bmatrix} \tag{13}$$

Step 3: Calculate the information entropy of each evaluation index. According to Formula (8), the information entropy of index j can be calculated by

$$E_j = -\ln(n)^{-1}\sum_{i=1}^{n} r_{ij} \ln r_{ij} (j = 1, 2, 3, 4) \tag{14}$$

Step 4: Determine the weight of each evaluation index. According to Formula (9), the weight of index j can be calculated by

$$w_j = \frac{1 - E_j}{\sum_1^4 (1 - E_j)} \tag{15}$$

Step 5: Calculate the influence of each node. According to the indices normalized in Step 2 and the weight obtained in Step 4, the LWC value of node i can be calculated by

$$LWC_i = \sum_{j=1}^{4} w_j r_{ij} \tag{16}$$

Step 6: Rank the value of the LWC_i in descending order to obtain the node influence ranking list.

Next, the time complexity of the LWC is analyzed. Step 1 needs to calculate degree k, clustering coefficient clc, two-hop degree thk, and two-hop clustering coefficient $thclc$. The complexity of degree k and clustering coefficient clc are O(n). The complexity of two-hop degree thk and two-hop clustering coefficient $thclc$ are O(2n). Step 2 is normalization, and the complexity is O(2n). Step 3 and Step 4 apply the entropy weighting method to weight the four evaluation indices, and the complexity is O(4n). Step 5 calculates the weighted sum to obtain the LWC value, and the complexity is O(n). So, the total time complexity of the proposed LWC algorithm is O(n) + O(2n) + O(2n) + O(4n) + O(n) = O(n).

4. Experimental Evaluation

To verify the performance of our proposed LWC method, we compare it with five well-known ranking methods, i.e., degree [4], betweenness centrality (BC) [7,8], closeness centrality (CC) [5,6], local centrality (LocalC) [28], and clustered local degree (CLD) [29].

Four real-world networks selected from different fields are used in our experiments, which consist of sparse graphs and dense graphs: (1) football; (2) netscience; (3) email; and (4) power. The datasets of the football, netscience, and power networks are provided on http://www-personal.umich.edu/~mejn/netdata/ (accessed on 8 February 2022), and the dataset of the email network is provided on https://networkrepository.com/email-univ.php (accessed on 8 February 2022). The characteristics of these four networks are listed in Table 1, including node number $|V|$, edge number $|E|$, average degree $\langle k \rangle$, max degree k_{max}, and clustering coefficient clc_{ave}.

Table 1. Specific statistical parameters of four real-world networks.

Network	$|V|$	$|E|$	$\langle k \rangle$	k_{max}	clc_{ave}
Football	115	613	10.66	12	0.403
Netscience	379	914	4.823	34	0.741
Email	1133	5451	9.622	71	0.202
Power	4941	6594	2.669	19	0.080

We conducted the experiments on a computer with an Intel Core i5 2.4 GHz processor and 8 GB RAM. All the codes were implemented in Python 3.9. Table 2 shows the average execution time of 100 runs for each method on the four real-world networks.

Table 2. The execution time of six methods on four real-world networks (seconds).

Network	Degree	BC	CC	LocalC	CLD	LWC
Football	0.05	0.12	0.09	0.12	0.05	0.07
Netscience	0.11	0.84	0.28	0.14	0.09	0.17
Email	0.39	8.12	2.92	0.26	0.45	0.52
Power	0.57	115.32	38.92	0.88	0.76	1.39

4.1. Discrimination Capability Analysis

Firstly, we investigate the discriminability of the ranking results obtained by different methods. The monotonicity metric is used for this purpose. The monotonicity metric M [44] is defined as follows:

$$M(S) = \left[1 - \frac{\sum\limits_{s \in S} |V|_s * (|V|_s - 1)}{|V| * (|V| - 1)} \right]^2 \tag{17}$$

where S represents the ranking list of network nodes, $|V|_s$ is the number of nodes with the same rank s of the ranking list S, and $|V|$ represents the total number of nodes of a network. The value of $M(S) \in [0, 1]$. The higher value of M, the better the discrimination capacity the ranking list has. For example, if each node has a unique rank, and the value of $|V|_s$ is 1, then $M(S) = 1$. In this case, the ranking list has a perfect discrimination capacity.

Table 3 shows the monotonicity values of different methods on four real-world networks. In Table 3, the column $M(X)$ refers to the monotonicity metric value calculated based on the ranking list obtained by corresponding method X on each network dataset. Bold numbers represent values that can be achieved by the method with the best monotonicity on each network. The results show that the proposed LWC has the most significant monotonicity values on the football, netscience, and email networks. On the power network, the monotonicity performance of the LWC is inferior only to the CC method. These results indicate that the proposed LWC method has excellent monotonicity capacity.

Table 3. Monotonicity performance of different methods.

Network	M(Degree)	M(BC)	M(CC)	M(LocalC)	M(CLD)	M(LWC)
Football	0.3637	1.0000	0.9488	0.9960	0.9915	1.0000
Netscience	0.7642	0.3390	0.9928	0.9887	0.9793	0.9944
Email	0.8874	0.9400	0.9988	0.9981	0.9974	0.9997
Power	0.5927	0.8319	0.9998	0.9014	0.9001	0.9653

Secondly, in order to obtain a more precise specification of ranking distribution by different methods on the four real-world networks, the complementary cumulative distribution function (CCDF) has been plotted in addition to the monotonicity metric. Figure 1 shows the CCDF plots of different methods on the football, netscience, email, and power networks. The slower slope of the CCDF means the more distinct ranks assigned and the better performance of the ranking distribution. It can be observed in Figure 1 that the CCDF plot of the LWC decreases slowly, yielding the best monotonicity and distribution performance on the football, netscience, and email networks. For the power network, as global information methods, the BC and CC methods have a better ranking distribution performance than the LWC method. However, according to Table 2, the execution time of BC and CC are 115.32 s and 38.92 s, respectively, which is much more than the 1.39 s of the LWC method. Furthermore, if focused on the top 1000 ranked nodes, the LWC method has a similar discrimination performance to the BC and CC methods. Therefore, considering both time cost and discrimination performance, the LWC is a competitive method compared with degree, BC, CC, LocalC, and CLD.
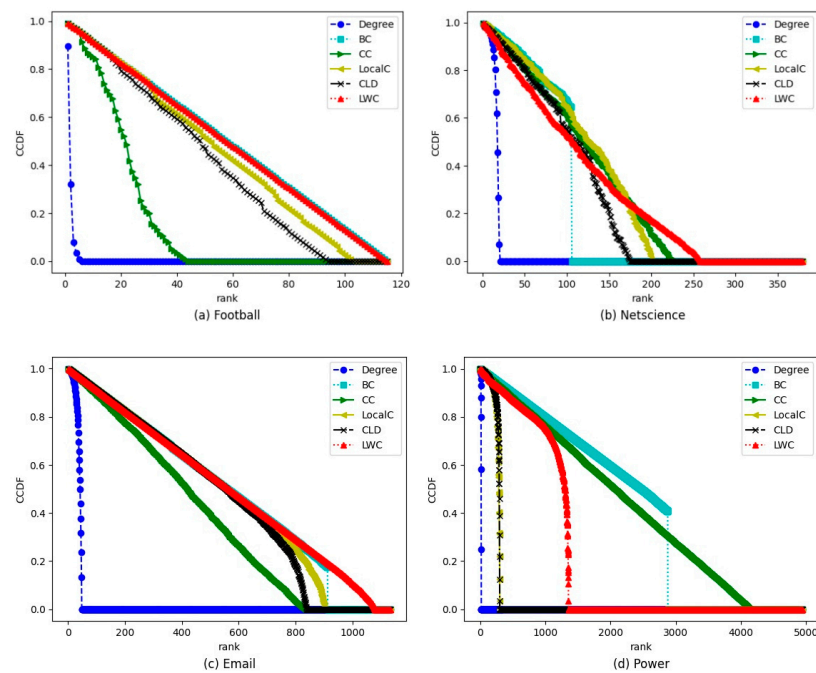
Figure 1. The complementary cumulative distribution function (CCDF) plots for ranking lists offered by different methods on the four real-world networks.

4.2. Accuracy Analysis

This section investigates the accuracy of the performance of the above mentioned methods in evaluating the influence of nodes on the four real-world networks. To this end, the susceptible-infected-recovered (SIR) model is adopted to simulate the spreading influence of each node and extract the ground-truth ranking of node influence. Based on the ground-truth influence ranking, we use the Kendall coefficient, imprecision function, and Jaccard coefficient to evaluate the accuracy of different methods. Firstly, we correlate it with the ranking measure produced by each algorithm. Then, we focus on the top influential nodes to calculate the imprecision function and the Jaccard similarity coefficient.

Researchers have proposed a large number of methods to simulate the spreading process in social networks [45]. The SIR model is one of the most popular spreading models, which has been frequently used to identify influential nodes in complex networks [46]. In the SIR model, each node has three states: susceptible, infected, or removed. To simulate the influence range of node v_i, we first set the state of node v_i as infected, while all other nodes in the network are in a susceptible state. At each timestamp, the infected node tries to infect its neighbors, whose states are susceptible with the probability β, and then the infected node recovers with the probability γ. If a node is in a recovered state, this node is immune and cannot be infected again. Without loss of generality, the probability γ is set to 1. These infecting and recovering processes continue until a steady state is obtained. The simulation experiments are performed 100 times, and the average over these simulations is defined as a node's influence. Finally, the ranking list based on the simulation results is marked as σ for the following accuracy analysis.

Firstly, Kendall's τ correlation coefficient [47] is used to quantify the correlation between ranking list R obtained by each method and the ground-truth ranking list σ obtained by the SIR model. The τ coefficient is calculated as follows

$$\tau(\sigma, R) = \frac{\sum\limits_{i<j} sgn[(x_i - x_j)(y_i - y_j)]}{0.5N(N-1)} \tag{18}$$

where $sgn(x)$ refers to the sign function. N is the total number of nodes of the network, i.e., the length of the ranking list. x_i and x_j represent the ranks corresponding to node i

and node j in the ranking list σ. Similarly, y_i and y_j represent the rank corresponding to node i and node j in ranking list R. Table 4 shows the τ correlation coefficient between the ground-truth influence ranking list σ and one produced by different ranking methods. As seen from Table 4, the correlation performance of the LWC is the best on four networks compared with the other five methods.

Table 4. Correlation between the ranking list obtained by different methods and the ground-truth.

Network	Degree	BC	CC	LocalC	CLD	LWC
Football	0.4089	0.2801	0.3516	0.4781	0.3603	0.4931
Netscience	0.2714	0.0116	0.1835	0.3826	0.4477	0.5048
Email	0.6603	0.4755	0.5644	0.6482	0.7036	0.7217
Power	0.3569	0.1916	0.3695	0.4959	0.5639	0.5670

Since the τ coefficient considers all node pairs in the network, including a mess of insignificant nodes, the value of τ is usually not very large with the increase in the number of nodes. In practice, people are only interested in important nodes in the network. Therefore, in the next experiment, the imprecision function is adopted to quantify the accuracy performance of the most influential nodes. The imprecision function [40] is defined as

$$\varepsilon(p) = 1 - \frac{M(p)}{Meff(p)} \tag{19}$$

where p is the ratio of network nodes we will investigate, $p \in [0,1]$. $M(p)$ is the average spreading influence of top pN nodes in ranking list R, and $Meff(p)$ is the average spreading influence of top pN nodes in ranking list σ. The smaller the value of $\varepsilon(p)$, the more accurate the method is to identify the most influential top pN nodes.

Figure 2 shows the imprecision function results of different methods on the four networks. As we only focus on the most influential nodes, the value of p is set at a range from 0.01 to 0.1. It can be observed that, in the netscience and power networks, the LWC always outperforms the other methods. In the football network, only when $p = 0.07$ and 0.08 is the performance of the LWC inferior to the degree method but better than the other four methods. In the email network, the performance of the LWC is better than the other five methods, except when $p = 0.09$ and 0.1. Therefore, on the whole, the LWC method is more accurate than the other five methods.

The third metric for evaluating the accuracy of the proposed method is the Jaccard similarity coefficient [44]. Here, we consider the top T nodes of the ranking list σ and ranking list R. The Jaccard similarity coefficient J is obtained as

$$J(T) = \frac{|\sigma(T) \cap R(T)|}{|\sigma(T) \cup R(T)|} \tag{20}$$

where $\sigma(T)$ and $R(T)$ are sets containing the top T ranked nodes of ranking list σ and ranking list R, respectively.

Figure 3 shows the Jaccard coefficient of top T influential nodes on the football, netscience, email, and power networks. The value of T is set as 10, 20, 100, and 100, respectively, on the four networks. As can be seen, the LWC is again the top performer in the datasets compared with the other five methods.
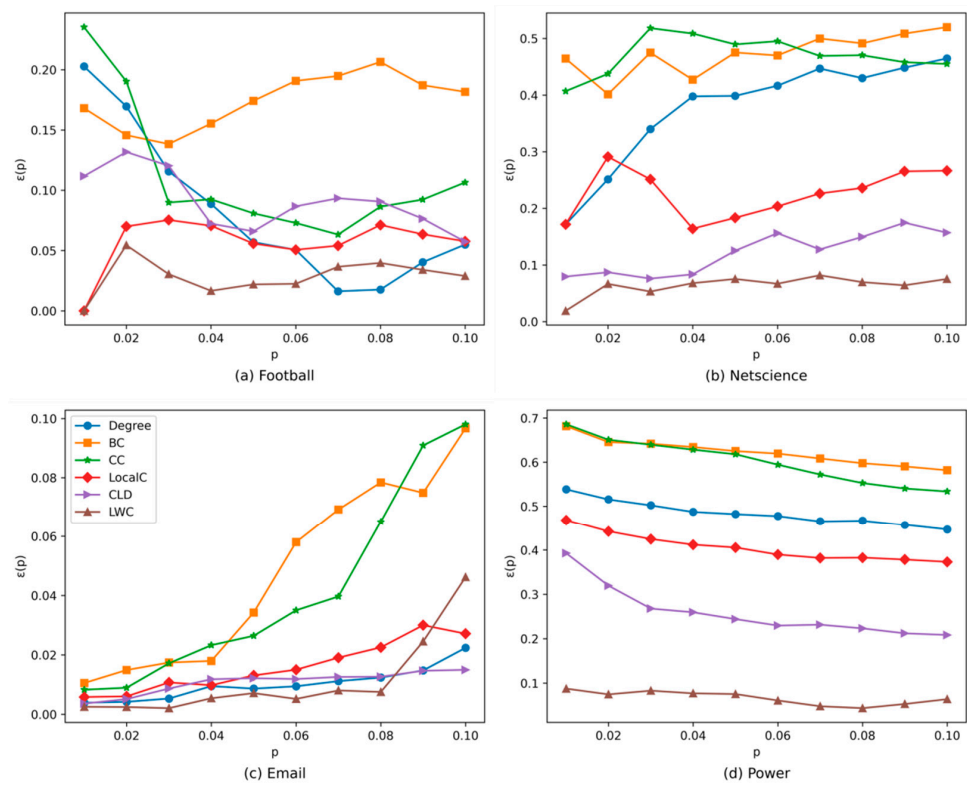
Figure 2. The imprecision function obtained by different methods on the four real-world networks.
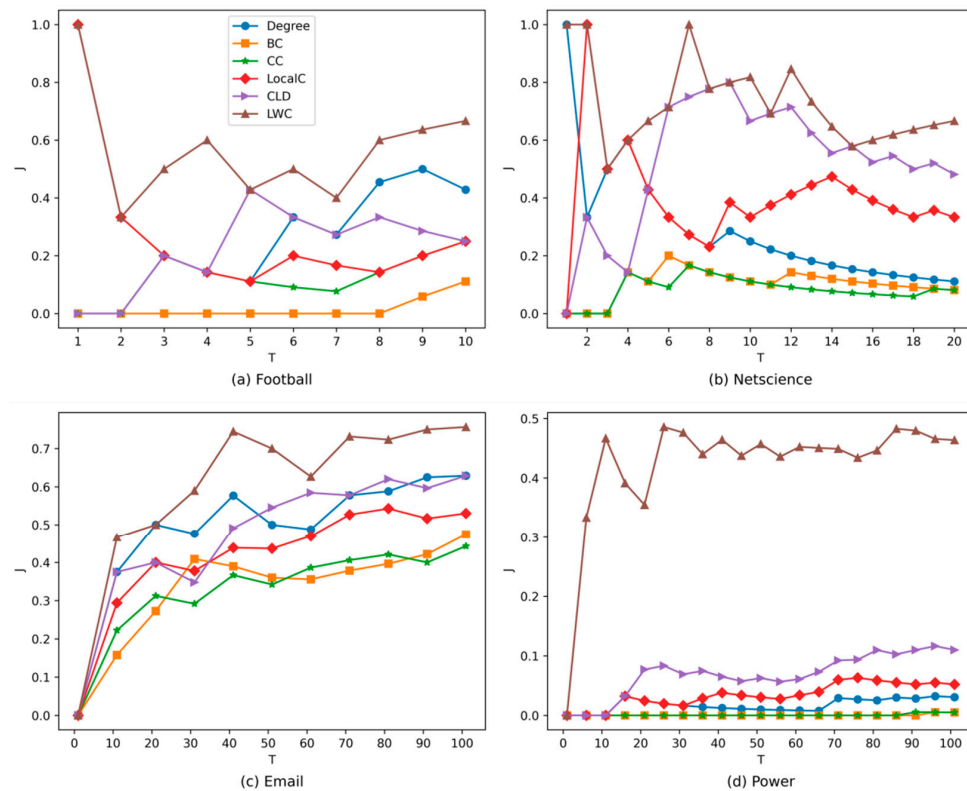

Figure 3. The jaccard similarity coefficient obtained by different methods on the four real-world networks.

5. Conclusions

Identifying influential nodes in a complex network is a fundamental question that has many promising applications in various fields, such as identifying super-spreaders

or opinion leaders in social networks, carrying out viral marketing, controlling disease propagation, stopping the spread of rumors, accelerating information propagation in a network, and so on. However, with the booming of large-scale networks, methods based on a global attribute with high time complexity have become inapplicable. In addition, some studies have pointed out that methods considering multiple attributes can achieve better performance than methods that consider only a single attribute. Therefore, this paper proposed a low time complexity method based on four local attributes, i.e., the multiple local attributes weighted centrality (LWC), to evaluate and identify node influence in complex networks. In the LWC method, we improved the calculation formula of the two-hop degree; proposed a two-hop clustering coefficient; applied the information entropy to determine the weight of the degree, the clustering coefficient, the two-hop degree, and the two-hop clustering coefficient; and then calculated the weighted sum of these four local attributes. Compared with the degree, the betweenness centrality, the closeness centrality, the local centrality, the lustering coefficient, and the clustered local-degree methods, the experiment results on four real-world networks have shown that the proposed LWC can identify influential nodes effectively and accurately. However, there are still some challenges that need to be further overcome to improve our work. Firstly, the performance of our proposed method is better than the comparison method, but it is not very good and still has great potential for improvement. Secondly, this paper only considers the node influence in static networks. However, complex networks, in reality, sometimes change dynamically. Therefore, in future research, we will continue to try to improve the performance of the algorithm and explore dynamic network characteristics.

Author Contributions: Funding acquisition, L.W. and J.Z. (Jinxin Zhang); Supervision, Q.Z.; Writing—original draft, J.Z. (Jinhua Zhang); Writing—review & editing, J.Z. (Jinxin Zhang) and L.W. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

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

References

1. Tulu, M.M.; Mkiramweni, M.E.; Hou, R.; Feisso, S.; Younas, T. Influential nodes selection to enhance data dissemination in mobile social networks: A survey. *J. Netw. Comput. Appl.* **2020**, *169*, 102768. [CrossRef]
2. Hafiene, N.; Karoui, W.; Ben Romdhane, L. Influential nodes detection in dynamic social networks: A survey. *Expert Syst. Appl.* **2020**, *159*, 113642. [CrossRef]
3. Omar, Y.M.; Plapper, P. A Survey of Information Entropy Metrics for Complex Networks. *Entropy* **2020**, *22*, 1417. [CrossRef] [PubMed]
4. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1978**, *1*, 215–239. [CrossRef]
5. Sabidussi, G. The centrality of a graph. *Psychometrika* **1966**, *31*, 581–603. [CrossRef]
6. Du, Y.; Gao, C.; Chen, X.; Hu, Y.; Sadiq, R.; Deng, Y. A new closeness centrality measure via effective distance in complex networks. *Chaos* **2015**, *25*, 033112. [CrossRef]
7. Newman, M.E.J. A measure of betweenness centrality based on random walks. *Soc. Netw.* **2005**, *27*, 39–54. [CrossRef]
8. Prountzos, D.; Pingali, K. Betweenness Centrality: Algorithms and Implementations. *ACM Sigplan Not.* **2013**, *48*, 35–45. [CrossRef]
9. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **1953**, *18*, 39–43. [CrossRef]
10. Zhao, J.; Yang, T.-H.; Huang, Y.; Holme, P. Ranking Candidate Disease Genes from Gene Expression and Protein Interaction: A Katz-Centrality Based Approach. *PLoS ONE* **2011**, *6*, e24306. [CrossRef]
11. 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.* **2010**, *6*, 888–893. [CrossRef]
12. Maji, G.; Mandal, S.; Sen, S. A systematic survey on influential spreaders identification in complex networks with a focus on K-shell based techniques. *Expert Syst. Appl.* **2020**, *161*, 113681. [CrossRef]
13. Zeng, A.; Zhang, C.-J. Ranking spreaders by decomposing complex networks. *Phys. Lett. A* **2013**, *377*, 1031–1035. [CrossRef]

14. Lin, J.-H.; Guo, Q.; Dong, W.-Z.; Tang, L.-Y.; Liu, J.-G. Identifying the node spreading influence with largest k-core values. *Phys. Lett. A* **2014**, *378*, 3279–3284. [CrossRef]

15. Bae, J.; Kim, S. Identifying and ranking influential spreaders in complex networks by neighborhood coreness. *Phys. A Stat. Mech. Its Appl.* **2014**, *395*, 549–559. [CrossRef]

16. Wei, B.; Liu, J.; Wei, D.; Gao, C.; Deng, Y. Weighted k-shell decomposition for complex networks based on potential edge weights. *Phys. A Stat. Mech. Its Appl.* **2015**, *420*, 277–283. [CrossRef]

17. Wang, Z.; Zhao, Y.; Xi, J.; Du, C. Fast ranking influential nodes in complex networks using a k-shell iteration factor. *Phys. A Stat. Mech. Its Appl.* **2016**, *461*, 171–181. [CrossRef]

18. Yang, F.; Zhang, R.; Yang, Z.; Hu, R.; Li, M.; Yuan, Y.; Li, K. Identifying the most influential spreaders in complex networks by an Extended Local K-Shell Sum. *Int. J. Mod. Phys. C* **2017**, *28*, 1750014. [CrossRef]

19. Wang, Z.; Du, C.; Fan, J.; Xing, Y. Ranking influential nodes in social networks based on node position and neighborhood. *Neurocomputing* **2017**, *260*, 466–477. [CrossRef]

20. Zareie, A.; Sheikhahmadi, A. A hierarchical approach for influential node ranking in complex social networks. *Expert Syst. Appl.* **2018**, *93*, 200–211. [CrossRef]

21. Wang, Y.; Chen, B.; Li, W.; Zhang, D. Influential Node Identification in Command and Control Networks Based on Integral k-Shell. *Wirel. Commun. Mob. Comput.* **2019**, *2019*, 6528431. [CrossRef]

22. Maji, G.; Namtirtha, A.; Dutta, A.; Curado Malta, M. Influential spreaders identification in complex networks with improved k-shell hybrid method. *Expert Syst. Appl.* **2020**, *144*, 113092. [CrossRef]

23. Brin, S.; Page, L. The anatomy of a large-scale hypertextual Web search engine. *Comput. Netw. ISDN Syst.* **1998**, *30*, 107–117. [CrossRef]

24. Lü, L.; Zhang, Y.C.; Yeung, C.H.; Zhou, T. Leaders in social networks, the Delicious case. *PLoS ONE* **2011**, *6*, e21202. [CrossRef]

25. Kleinberg, J.M. Authoritative sources in a hyperlinked environment. *J. ACM* **1999**, *46*, 604–632. [CrossRef]

26. 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]

27. Zareie, A.; Sheikhahmadi, A. EHC: Extended H-index Centrality measure for identification of users' spreading influence in complex networks. *Phys. A Stat. Mech. Its Appl.* **2019**, *514*, 141–155. [CrossRef]

28. Chen, D.; Lü, L.; Shang, M.-S.; Zhang, Y.-C.; Zhou, T. Identifying influential nodes in complex networks. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 1777–1787. [CrossRef]

29. Li, M.; Zhang, R.; Hu, R.; Yang, F.; Yao, Y.; Yuan, Y. Identifying and ranking influential spreaders in complex networks by combining a local-degree sum and the clustering coefficient. *Int. J. Mod. Phys. B* **2018**, *32*, 1850118. [CrossRef]

30. 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]

31. Qiao, T.; Shan, W.; Yu, G.; Liu, C. A Novel Entropy-Based Centrality Approach for Identifying Vital Nodes in Weighted Networks. *Entropy* **2018**, *20*, 261. [CrossRef] [PubMed]

32. Li, Y.; Cai, W.; Li, Y.; Du, X. Key Node Ranking in Complex Networks: A Novel Entropy and Mutual Information-Based Approach. *Entropy* **2020**, *22*, 52. [CrossRef] [PubMed]

33. Ni, C.; Yang, J.; Kong, D. Sequential seeding strategy for social influence diffusion with improved entropy-based centrality. *Phys. A Stat. Mech. Its Appl.* **2020**, *545*, 123659. [CrossRef]

34. Guo, C.; Yang, L.; Chen, X.; Chen, D.; Gao, H.; Ma, J. Influential Nodes Identification in Complex Networks via Information Entropy. *Entropy* **2020**, *22*, 242. [CrossRef]

35. 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]

36. Yang, Y.; Wang, X.; Chen, Y.; Hu, M. Identifying Key Nodes in Complex Networks Based on Global Structure. *IEEE Access* **2020**, *8*, 32904–32913. [CrossRef]

37. Yang, P.; Xu, G.; Chen, H. Multi-attribute ranking method for identifying key nodes in complex networks based on GRA. *Int. J. Mod. Phys. B* **2019**, *32*, 1850363. [CrossRef]

38. Mo, H.; Deng, Y. Identifying node importance based on evidence theory in complex networks. *Phys. A Stat. Mech. Its Appl.* **2019**, *529*, 121538. [CrossRef]

39. Liu, Y.; Tang, M.; Zhou, T.; Do, Y. Identify influential spreaders in complex networks, the role of neighborhood. *Phys. A Stat. Mech. Its Appl.* **2016**, *452*, 289–298. [CrossRef]

40. Liu, D.; Nie, H.; Zhao, J.; Wang, Q. Identifying influential spreaders in large-scale networks based on evidence theory. *Neurocomputing* **2019**, *359*, 466–475. [CrossRef]

41. He, D.; Xu, J.; Chen, X. Information-Theoretic-Entropy Based Weight Aggregation Method in Multiple-Attribute Group Decision-Making. *Entropy* **2016**, *18*, 171. [CrossRef]

42. Luo, R.; Huang, S.; Zhao, Y.; Song, Y. Threat Assessment Method of Low Altitude Slow Small (LSS) Targets Based on Information Entropy and AHP. *Entropy* **2021**, *23*, 1292. [CrossRef]

43. Zhao, J.; Song, Y.; Deng, Y. A Novel Model to Identify the Influential Nodes: Evidence Theory Centrality. *IEEE Access* **2020**, *8*, 46773–46780. [CrossRef]

44. Zareie, A.; Sheikhahmadi, A.; Jalili, M.; Fasaei, M.S.K. Finding influential nodes in social networks based on neighborhood correlation coefficient. *Knowl.-Based Syst.* **2020**, *194*, 105580. [CrossRef]

45. Zhang, M.; Qin, S.; Zhu, X. Information diffusion under public crisis in BA scale-free network based on SEIR model—Taking COVID-19 as an example. *Phys. A Stat. Mech. Its Appl.* **2021**, *571*, 125848. [CrossRef]

46. Ullah, A.; Wang, B.; Sheng, J.; Long, J.; Khan, N.; Sun, Z. Identifying vital nodes from local and global perspectives in complex networks. *Expert Syst. Appl.* **2021**, *186*, 115778. [CrossRef]

47. Ullah, A.; Wang, B.; Sheng, J.; Long, J.; Khan, N.; Gambuzza, L.V. Identification of Influential Nodes via Effective Distance-based Centrality Mechanism in Complex Networks. *Complexity* **2021**, *2021*, 8403738. [CrossRef]