Research Article

The Local Triangle Structure Centrality Method to Rank Nodes in Networks

Xiaojian Ma and Yinghong Ma

School of Business, Shandong Normal University, Jinan, Shandong, China

Correspondence should be addressed to Yinghong Ma; yinghongma71@163.com

Received 8 September 2018; Revised 7 December 2018; Accepted 24 December 2018; Published 2 January 2019

Academic Editor: Dimitri Volchenkov

Detecting influential spreaders had become a challenging and crucial topic so far due to its practical application in many areas, such as information propagation inhibition and disease dissemination control. Some traditional local based evaluation methods had given many discussions on ranking important nodes. In this paper, ranking nodes of networks continues to be discussed. A semilocal structures method for ranking nodes based on the degree and the neighbors’ connections of the node is presented. The semilocal structures are regarded as the number of neighbors of the nodes and the connections between the node and its neighbors. We combined the triangle structure and the degree information of the neighbor to define the inner-outerspreading ability of the nodes and then summed the node neighbors’ inner-outerspreading ability to be used as the local triangle structure centrality (LTSC). The LTSC avoids the defect “pseudo denser connections” in measuring the structure of neighbors. The performance of the proposed LTSC method is evaluated by comparing the spreading ability on both real-world and synthetic networks with the SIR model. The simulation results of the discriminability and the correctness compared with pairs of ranks (one is generated by SIR model and the others are generated by central nodes measures) show that LTSC outperforms some other local or semilocal methods in evaluating the node’s influence in most cases, such as degree, betweenness, H-index, local centrality, local structure centrality, K-shell, and S-shell. The experiments prove that the LTSC is an efficient and accurate ranking method which provides a more reasonable evaluating index to rank nodes than some previous approaches.

1. Introduction

Social networks display significant applications in people’s social lives, such as the online social platforms and the communications [1, 2]. The key or the sensitive nodes play important roles in information diffusion. Detecting the vital nodes is a fundamental problem due to its wide application in many areas. For example, the propagation process of news and ideas can be accelerated by the key individuals in networks [3].

A variety of methods to measure the key nodes had been proposed in recent years. The topological location based methods were presented by many researches, such as degree centrality (DC) [4], betweenness centrality (BC) [5], closeness centrality [6], eigenvector centrality [7], and structure hole [8]. Degree centrality is one of the most basic topology attributes with low computational complexity. It is thought that a node with higher degree has much more influence than the node with lower degree. However, the weak connection of nodes proved that the high degree centrality of the node does not always mean the important of it. For example, Granovetter [9] had researched the topic of how people find jobs and discovered that information about jobs that led to employment was more likely to come from the weak connection with acquaintances than from closer friends. Structure hole method, which is somewhat close to the strength of weak ties, searches the bridge joints which are channels spanned by a group of indirectly connected nodes. Betweenness and closeness centralities identify the node’s influences in the global scope. Unfortunately, both of them lost efficiency in large scale networks because of the huge computing complexity in detecting shortest paths between each pair of nodes.

Therefore, some other centrality indexes have been proposed to make up the defects of those measurements. Kitsak and his cooperators [10] found that the most influential
nodes are not the ones with the largest degree but the ones located at the core of the networks with the largest $K$-shell value. Therefore, Kitsak's $K$-shell method was applied to find the influential nodes located at core positions. However, nodes having the same $K$-shell values often have distinct spreading influences. So, researchers extended the $K$-shell decomposition method or took more neighbor information to improve the accuracy. For example, the $S$-shell decomposition method [11], the mixed degree decomposition method [12], and the method based gravity formula [13] had been extended to conquer the weaknesses of $K$-shell. The semilocal centrality (LC) [14], the neighborhood coreness centrality [15], the cluster ranking method [16], and the neighborhood centrality method [17] by encoding many steps of neighbors’ information were proposed to improve the accuracy of the measurement. The path diversity [18] is also taken into account in determining the spreading ability.

$H$-index centrality is extended by $K$-shell to detect the importance of nodes according to the concept of $H$-index [19], where the $H$-index of a node is the maximum value $H$ such that there exist at least $H$ neighbors of degree no less than $H$. The local structural centrality (LSC) [20] is to evaluate the importance of nodes based on the node’s first and second neighbors and their cluster coefficient which is inspired by LC.

In order to find the more effective measure to evaluate the importance of the node, it is needed to capture the natures of the node as much as possible. We take together both the local triangle structures and the degrees of nodes to evaluate the importance of them. The experiments of the presented method and the other seven indexes are achieved on real-world data and synthetic networks, and the results show that the proposed method in this paper is effective and accurate. It is proved that the connections among the node's neighbors and the degree centrality are reasonable factors to evaluate the influence of nodes.

The arrangement of this paper is as follows: Introduction of literatures review and the research problem are given in Section 1. The motivation, the proposed method, and the example analysis are presented in Section 2. The datasets including real networks and synthetic networks are presented in Section 3. In experiments on real data in Section 4 the superior performances are displayed which are verified by the discriminability, the correctness of ranks, and the most influenced nodes. The experimental results of synthetic networks are arranged in Section 5. And, finally, the conclusions and discussions are given in Section 6.

2. The Local Triangle Structure Centrality Method

In this paper, an undirected and simple network without weight on edges or nodes is denoted by $G = (V, E)$, where $V$ and $E$ are the nodes set and edges set, respectively. And denote the size of $V$ and $E$ by $n$ and $m$, respectively.

2.1. The Motivation. In order to find a more effective measure to evaluate the node’s importance, it is needed to capture as many as possible natures of the node. The topology fundamentally affects the dynamics of the networks [21]; therefore, it is important to measure the node’s influence abilities by the local or global information. For example, the clustering coefficient of a node is usually employed as the local structures to evaluate the importance of the node. The topological coefficient of a node’s neighbors indicate the tendency of them to form triangle structures. Once a triangle is formed, the three nodes of this triad would exchange information with each other. Obviously, the ratio of the triangle structures of the node’s neighbors plays an important role in the information propagation. During the process of information dissemination, two adjacent nodes $u$ and $v$ share the same neighbor $k$; they will have more probability to transmit with each other. Therefore, the more the triangle structures formed by the node with its neighbors are, the more likely they will infect each other. Nevertheless, a node with a large clustering coefficient does not mean a large number of triangles of it.

In reality, during the epidemic or information spreading, the degree and the triangle structures of neighbors of the infected node play important roles. The bigger degree of a node means more neighbors. The higher the percentage of triangle structures formed between a node and its neighbors is, the more likely the neighbors infect each other during the spreading. Therefore, the centrality of a node should combine the node’s degree and the triangles of the node with its neighbors.

We are motivated by the semilocal centrality methods, the local centrality, and the local structure centrality [14, 20] and combine both the local triangle structures and the degree of the node to evaluate the importance of it.

The inner-out spreading ability of a node is measured by a linear combination of the local triangle structures and the degrees information of its neighbors, and the sum of the neighbors’ inner and outer spreading ability of the node is defined as the local triangle structure centrality.

2.2. The Method. In networks, the information usually is spread through a node to its neighbors and to the neighbors’ neighbors and so on. So the spreading ability of the node heavily depends on the degree and its neighbors’ structures. The more the triangle structures formed among nodes themselves and their neighbors are, the more likely the node locates in the dense part of the network. The number of triangles and clustering coefficients are not trivially related [22] in networks, where larger density of triangles does not imply the high clustering coefficients. Hence, to avoid the same flaw as the definition of the clustering coefficient and some indexes, the proportion of triangles (TP) instead of the number of triangles is considered as an index to measure how close the nodes neighbors connected to each other.

$$TP(v) = \frac{NTS(v)}{TNTS},$$

where $NTS(v)$ is the number of triangles structures existing in the neighborhood of the node $v$. And $TNTS$ is the sum of triangle structures formed by all the nodes in the networks, $TNTS = \sum_{v \in V} NTS(v)$. In fact, $TP(v)$ is the frequency
of triangles of the node $v$ in networks and is an intuitive measurement of the centrality. So, TP and the degrees might be sensitive indicators to measure the spreading abilities. During the spreading process, the primary infected nodes not only influence the nearest neighbors, but also affect the neighbors’ neighbors.

Therefore, we partition the spreading ability into two parts: one is the inner spreading and the other is the outer spreading. The inner spreading ability is contributed by the degree and the TP which measures how close the nodes neighbors are and reflects the location of a node in a network. The outer spreading ability is measured by the degrees of its neighbors and the degree of itself.

Based on above analysis together with the inner and the outer spreading ability, an inner-outer spreading ability of node $v$ is presented, denoted by $IO(v)$:

$$ IO(v) = d_v (1 + TP(v)) + (D_v - d_v), \quad (2) $$

where $d_v$ is the degree of node $v$, and $D_v$ is the sum of neighbors’ degrees. $D_v = \sum_{u \in \Gamma(v)} d(u)$. $d_v (1 + TP(v))$ and $D_v - d_v$ in (2) represent the inner spreading ability and the outer spreading ability of the node $v$ in the network, respectively.

By definition of the inner-outer spreading ability, a new index to indicate the importance of nodes in a network is defined as the local triangle structures centrality, shortly LTSC.

$$ LTSC(v) = \sum_{u \in \Gamma(v)} IO(u). \quad (3) $$

$LTSC(v)$ is the sum of the inner-outer spreading abilities of the node $v$'s neighbors. Take (2) together with (3); the local triangle structures of the node $v$ are made up of two parts, by its very nature, the neighbors degrees and the triangle structures.

The algorithm pseudocodes of computing the nodes’ LTSC values are presented in Table 1.

The computational complexity for calculating the percentage triangle structures of a node is $O(d^3)$, where $(d)$ is the average degree of the network. The total computing complexity for our centrality measure is $O(n(d^2))$ which grows linearly with the size of a sparse network.

2.3. Example for the LTSC Method. As an example, Figure 1 shows that nodes 7 and 13 are two neighbors of node 6. There are 2 triangles between node 7 and its neighbors more than node 13 and its neighbors. And the spreading abilities, calculated by the SIR model [10], of the two nodes 7 and 13 are 2.28 and 1.61, respectively, which coincide with the number of triangles of nodes. Meanwhile the clustering coefficients of nodes 7 and 13 are 0.333 and 1, respectively. At this point, the clustering coefficients of nodes do not imply the dense edge of their neighbors. On the other hand, nodes 1 and 6 have the same degrees and local centralities, and their spreading abilities are 2.87 and 3.0, respectively. Seen in this way, the number of the node's neighbors, even, and the number of next-nearest neighbors may not be enough to measure the spreading ability of it. It is reasonable to consider the global topological connections. So, we compare the different values of eight indexes on the example of Figure 1.

The spreading ability and the centrality or importance is calculated by DC, BC, H-index, K-shell, S-shell, LC, LSC, and the method we proposed. The centrality values are shown in Table 2. Table 2 shows the values of different measures in Figure 1. DC, H-index, and K-shell encounter the same problem that many nodes are assigned the same centrality value. Some nodes get larger BC value while the spreading influences of them are relatively small. The values of LC show better performance comparing with DC and others. The index’s values of the node are positively correlated with the node's spreading ability. While node 1 and node 6 have the same LC value, the spreading values are quite different. LSC and LTSC take the topology connections into consideration, so they present positive linear correlation trends with the real spreading ability. LSC and LTSC show similar values, and the latter shows better performance because of the percentage of triangle considered.

### Table 1: Computing the node’s LTSC value.

| Algorithm | Input: $G = (V, E), |V| = n$. | Output: $LTSC(v)$ for all $v \in V$. |
|-----------|---------------------------------|----------------------------------|
| Step 1:   | Calculate $NTS(v)$ for all $v \in V$; |                                 |
| Step 2:   | $TNTS = \sum_{v \in V} NTS(v)$; |                                 |
| Step 3:   | Calculate $TP(v)$ by equation (1); |                                 |
| Step 4:   | Calculate $D(v)$ and $d(v)$ for all $v \in V$; |                                 |
| Step 5:   | Calculate $IO(v)$ using formula (2); |                                 |
| Step 6:   | Compute $LTSC(v)$ by formula (3) for all $v \in V$. |                                 |

![Figure 1: An example network consists of 28 nodes and 35 edges. The numeric beside each node is the spreading ability of it. The spreading ability of each node is evaluated with epidemic probability of $\beta = 0.31$ of SIR model [10] by performing 20000 independent runs.](image-url)
### Table 2: The centralities with 8 different measures on the network of Figure 1.

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Sprd</th>
<th>DC</th>
<th>BC</th>
<th>Hin</th>
<th>Ks</th>
<th>Ss</th>
<th>LC</th>
<th>LSC</th>
<th>LTSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.874</td>
<td>218</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>118</td>
<td>35.62</td>
<td>52.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3.127</td>
<td>74</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>79</td>
<td>20.81</td>
<td>31.417</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.063</td>
<td>74</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>79</td>
<td>20.81</td>
<td>31.417</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.128</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>44</td>
<td>11</td>
<td>15.25</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.541</td>
<td>180</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>56</td>
<td>13.82</td>
<td>19.25</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.999</td>
<td>218</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>118</td>
<td>36.52</td>
<td>56.833</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3.278</td>
<td>72.5</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>87</td>
<td>21.86</td>
<td>37.667</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3.276</td>
<td>72.5</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>87</td>
<td>21.86</td>
<td>37.667</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2.128</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2.541</td>
<td>180</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>56</td>
<td>13.82</td>
<td>19.25</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2.574</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>63</td>
<td>15.62</td>
<td>23.333</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>2.585</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>63</td>
<td>15.62</td>
<td>23.333</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2.613</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>63</td>
<td>16.22</td>
<td>23.333</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2.650</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>63</td>
<td>16.22</td>
<td>23.333</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2.194</td>
<td>182</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>29</td>
<td>9.14</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2.191</td>
<td>182</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>29</td>
<td>9.24</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>1.886</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>25</td>
<td>7.57</td>
<td>11.167</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2.262</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>38</td>
<td>10.74</td>
<td>17.25</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>2.298</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>38</td>
<td>10.74</td>
<td>17.25</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2.305</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>38</td>
<td>10.74</td>
<td>17.25</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>2.286</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>38</td>
<td>10.74</td>
<td>17.25</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1.874</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>25</td>
<td>7.57</td>
<td>11.167</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>2.452</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>42</td>
<td>11.34</td>
<td>21.583</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>2.711</td>
<td>3</td>
<td>0.5</td>
<td>2</td>
<td>5</td>
<td>51</td>
<td>14.66</td>
<td>27.5</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>2.440</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>42</td>
<td>11.34</td>
<td>21.583</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>2.443</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>42</td>
<td>11.34</td>
<td>21.583</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>2.709</td>
<td>3</td>
<td>0.5</td>
<td>2</td>
<td>5</td>
<td>51</td>
<td>14.66</td>
<td>27.5</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>2.435</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>42</td>
<td>11.34</td>
<td>21.583</td>
<td></td>
</tr>
</tbody>
</table>

Sprd is the node’s spreading ability got by the SIR model.

Some researches, and the synthetic networks are generated by the famous BA model and LFR model.

#### 3.1. The Real-World Data

Six real-world networks are chosen in the following discussion, including Karate club network (Karate) [23], Polbook network (Polbook) [24], Coauthorship network of scientists (Netscience) [25], Email network of the University of Rovira Virgili (Email) [26], Football [27], Western States Power Grid (PowerGrid) [28]. The basic topological features of these networks are shown in Table 3, where $d_{max}$ and $\langle d \rangle$ are the max and the average degree, respectively, and $C$ is the cluster coefficient of the network.

Zachary's Karate club is a well-known social network of a university Karate club described by Zachary. The network became a popular example of community structure in networks after its use by Michelle Girvan and Mark Newman in 2002 with a paper "Community Structure in Social and Biological Networks". The political books dataset was compiled by Valid Krebs. The nodes represent 105 books about US politics sold by online seller Amazon.com. The edges present the frequent copurchasing of books by the same buyers. Books can be divided according into 3 categories, which are conservative, liberal, and neutral. The Football network represents football games in Division IA colleges. Each node represents a football team and the edge connects the two teams representing games between the two teams during the season. There are 115 teams and 613 games. The teams are divided into 12 "conferences" containing teams around 8 to 12. The Netscience network is a coauthorship network of scientists working on network theory and experiment, compiled by Newman [25]. Here, in this paper, the biggest component with obvious community structures is chosen as the Netscience network with 379 nodes and 914 edges. The Email network was established by professor Alexandre Arenas. It describes the email interchanges between members of the University of Rovira i Virgili (Tarragona). The PowerGrid network is a well-known network showing the scale-free phenomenon, representing the topology of the Western States Power Grid of the United States. The 4941 nodes and 6594 edges represent the substations and transmission lines between the substations, respectively. The national grid is concerned with simple geography dictating that it is always just a few transmission lines from collapse.
3.2. The Synthetic Data. The synthetic networks include networks generated by the Barabasi-Albert (BA) network model [29] and Lancichinetti-Fortunato-Radicchi (LFR) network model [30]. We also use SIR model for examining the performance of our proposed method.

BA model is an algorithm for generating random scale-free networks using a preferential attachment mechanism. Several natural and human-made systems, including the Internet, WWW, citation networks, and some social networks, are thought to be approximately scale-free and certainly contain few nodes with unusually high degree as compared to the other nodes of the network. The BA model tries to explain the existence of such nodes in real networks. The algorithm is named for its inventors Albert Barabasi and Reka Albert [29]. The network begins with an initial connected network of \( m_0 \) nodes. New nodes are added to the network, one node at each time step. Each new node is connected to \( m_t \) existing nodes with a probability \( d_i / \sum_j d_j \) that is proportional to the number of links that the existing nodes \( v_i \) already have. Here \( m_t \leq m_0 \), \( d_i \) is the degree of node \( v_i \) and the sum is made over all preexisting nodes \( v_j \).

The other kind of synthetic networks is the widely used Lancichinetti-Fortunato-Radicchi (LFR) network model [30]. LFR is an algorithm that generates benchmark networks that resemble real-world networks. They have a priori known communities and are used to compare different community detection methods. LFR assumes that both the degree and the community size have power-law distributions with different exponents, \( \gamma \) and \( \lambda \), respectively. \( n \) is the number of nodes and the average degree is \( \langle d \rangle \).

There is a mixing parameter \( \mu \), which is the average fraction of neighboring nodes of a node that do not belong to any community that the benchmark node belongs to. This parameter controls the fraction of edges that are between communities. Thus, it reflects the amount of noise in the network. At the extremes, when \( \mu = 0 \) all links are within community links; if \( \mu = 1 \) all links are between nodes belonging to different communities.

4. The Method of Experiments on Real-World Networks

The feasibility and effectiveness using LTSC to rank the importance of nodes in networks are empirically evaluated through a series of experiments. The experiments on LTSC are compared with other seven well-known measures from the aspects of discriminability, correctness, and so on. These measures include degree centrality (DC), betweenness centrality (BC), H-index method (H-index), K-shell (Ks), S-shell (Ss), local centrality (LC), and local structure centrality (LSC) [20]. DC, H-index, Ks, and Ss of a node heavily depended on the nearest neighbors of it. BC evaluates the node's influence in a global scope. LC of a node is defined by computing the number of neighbors and the neighbors' neighbors of the node. LSC investigates the impact of the topological connections among neighbors. Meanwhile LTSC is to calculate the number of neighbors together with the ratio of triangles of the node. Comparing to LSC and LC, LTSC takes the percentage of triangles (TP) into consideration instead of clustering coefficient to measure the topological connections; LTSC might display better performance.

The comparisons of LTSC with the other seven centrality indexes are analyzed by SIR model in discriminability, correctness of the most influential nodes, and the performances of measures which are evaluated by SI model at the end of this section.

4.1. Discriminability. In general, if nodes \( u \) and \( v \) have much difference of influence abilities, they can be easily distinguished from each other. Complementary Cumulative Distribution Function (CCDF) [15] is often taken as one of the standards to discriminate nodes, where CCDF is a power calculation method and can be performed on size-domain data which is the substitute of time-domain data in statistical study. CCDF is computing the probability density function using a histogram of the measured ranks. The number of nodes histogrammed is determined by the proportion of nodes with same rank and the total nodes. Thus, CCDF as a measurement is taken to evaluate the amount of the discriminability of difference in the spreading capability.

\[
CCDF\left(r\right) = 1 - \frac{\sum_{i=1}^{r} n_i}{n}, \tag{4}
\]

where \( n_i \) is the number of nodes with rank \( i \) on the list, and \( n \) is the number of nodes in the network. \( r \) is the number of ranks. According to the equation 4 of CCDF, it is easy to find that the curve CCDF is slowly going down if \( r \rightarrow n \). That means the discriminability is well defined; nearly each node has a distinct rank. Otherwise, the CCDF curve will decrease rapidly if \( r \rightarrow 1 \). That means all nodes are assigned in a few of ranks.

The CCDF curves by the LTSC and the other seven indexes on Karate, Polbook, Netscience, and Email are plotted in Figure 2. Two reasons of the four networks are chosen.
to verify the discriminability: One is that the size of nodes and edges of the four networks are from tens to thousands of nodes. And the other reason is that there are obvious communities and central nodes in them. LTSC achieves a better rank distribution performance compared with the other indexes. It shows a slowdown trend in the four panels and has similar variant behaviors with LC, LSC, and BC. DC, H-index, and K-shell show poor performances because they only depend on the nearest neighbors without the information of the connections and the neighbors’ of the neighbors. BC, LC, and LSC evaluate nodes with more information, so the three indexes together with LTSC show better performance in distinguishing the influence capability in Karate network. The four indexes agree with each other from the first to the 15th nodes in the rank. BC lost its efficiency from the 21st to the 34th rank. LTSC is better than LC from the 15th to the 29th rank, though some evaluations of LSC are in the overlap LTSC and LC. And all indexes do not distinguish the 30th to the 34th rank, because the 30th to the 34th rank have much similar structures nodes [23]. It might be the reason that LC, LSC, and LTSC are defined by the local topological connections among the node’s neighbors. In Polbook and Email networks, the CCDF curves of LTSC, LSC, LC, and BC almost coincide with each other except a few nodes on BC. In Netscience network, LTSC and LSC almost overlaps from the first to the 268th rank, and LC is a little inferior from the first to 250th rank. LTSC, LSC, and LC show slower slope and more distinct ranks. BC lost efficacy about the 100th to the last rank. The clustering coefficient of Netscience network is 0.7412 more than the other three networks. That is to say, most of nodes’ neighbors form triangles with the node which enlarge
the clustering coefficient. LTSC and LSC are defined by the triangle structures of a node and its neighbors. Thus, LTSC and LSC achieve better ranks.

The interval [0, 1] is uniformly divided into n parts; that is, there are n subintervals, [0, 1/n], [1/n, 2/n], …, [(n − 1)/n, 1]. A rank (x1, x2, …, xn) is defined as a uniform discriminability if there is a evaluation method such that the value of xi is in subinterval [i/n, (i + 1)/n] for all i = 1, 2, …, n which holds. In other words, a rank is a uniform discriminability; the value’s difference of xi and xi−1 is less than 2/n.

In fact, there is a nature of the CCDF curves in Figure 2: a good discriminability curve of CCDF is almost the curve \(y = -x\), where \(x \in [0, 1]\). If \(r = n\), the projection of the CCDF curve onto x-axis is in the interval [0, 1] and is a uniform discriminability. The Polbook and Email networks, LTSC, LC, and LSC are almost uninform discriminability, and some nodes cannot be distinguished clearly in Karate and Netscience. The experiments of the four networks on the rank CCDF show that LTSC displays good performance and LSC is in the same way. The triangle structures formed by nodes and their neighbors promote the ranking quality.

4.2. Correctness of Two Ranks. The rank of nodes in networks generated by an effective method should be as same as the real spreading process. The susceptible-infectious-recovered (SIR) model [10] is employed to simulate the real spreading process. Then, the spreading ability measured by LTSC index is appropriated to be verified by SIR model.

In the SIR model, each node belongs to one of the three states: susceptible (S), infected (I), and recovered (R). In the SIR process, the node to be investigated would be in state I and all other nodes are set to be in state S at the initial time. Then, at each time step, the infected nodes will infect their susceptible neighbors with a probability of \(\beta\) and then the node itself moves to the recovered state. The process is going on until there is no node infected in the network. The number of the recovered nodes denotes the spreading influence of every initially infected node at the end of the epidemic process. Also, we can obtain the ranking list \(\theta\), generated by the SIR model.

The numerical simulations on SIR model are repeated 10000 times if the number of nodes in networks is no more than 100. Otherwise, the simulations are repeated 1000 times. The spreading probability \(\beta\) is set larger than the lower bound probability [15]; that is, \(\beta \geq \beta_h\). The lower bound probability \(\beta_h = \langle d^- \rangle/\langle d^+ \rangle\), where \(\langle d^- \rangle\) and \(\langle d^+ \rangle\) are the average degrees and average second-order degrees of the nodes, respectively. Kendall’s rank correlation coefficient \(\tau\) [31] is often used as a test statistic in a statistical hypothesis test to establish whether two variables may be regarded as statistically dependent. The rank of nodes in networks is used to quantify the extent of similarity of the two ranks: one rank \(\sigma = \{x_1, x_2, …, x_n\}\) is generated by a certain centrality measure; the other rank in this paper \(\theta = \{y_1, y_2, …, y_n\}\) is obtained by SIR simulation. Let \(\{x_i, y_i, \ldots, x_n, y_n\}\) be the pair of ranks \(\sigma\) and \(\theta\). The pair of ranks \(\sigma\) and \(\theta\) are called concordant if \(x_i > x_j\) and \(y_i > y_j\) or \(x_i < x_j\) and \(y_i < y_j\). If \(x_i > x_j\) and \(y_i < y_j\) or \(x_i < x_j\) and \(y_i > y_j\), \(\sigma\) and \(\theta\) are called discordant. The pair of ranks \(\sigma\) and \(\theta\) are neither concordant nor discordant; we call it null if \(x_i = x_j\) or \(y_i = y_j\). Kendall’s coefficient \(\tau\) is defined as follows:

\[
\tau(\sigma, \theta) = \frac{2(n^2 - n_c)}{n(n-1)},
\]

where \(n_c\) and \(n_d\) are the numbers of concordant and discordant pairs in the ranking, respectively. If there is no null of the pairs \(\{x_i, y_i\}\) for all \(i = 1, 2, \ldots, n\) of two ranks \(\sigma\) and \(\theta\), then \(n_c + n_d = (1/2)n(n-1)\), and then \(\tau(\sigma, \theta) = 1 - 4n_d/n(n-1) = 4n_d/n(n-1) - 1\). Therefore, \(\tau \in [-1, 1]\).

The value of \(\tau\) indicates how agreement of the rank is generated by \(\sigma\) with the rank got by the SIR model. Denote Kendall’s \(\tau\) coefficient of the rank generated by one of the six indexes and the SIR model by \(\tau(\sigma, \theta)\). When \(\tau \rightarrow 1\), the two ranks agree with each other perfectly; while \(\tau \rightarrow -1\), the two ranks disagree with each other perfectly. Table 4 shows Kendall’s \(\tau\) coefficients of DC, BC, H-index, LC, LSC, and LTSC with the SIR model.

<table>
<thead>
<tr>
<th>Networks</th>
<th>(\beta_h)</th>
<th>(\beta)</th>
<th>(\tau(DC, \theta))</th>
<th>(\tau(BC, \theta))</th>
<th>(\tau(HC, \theta))</th>
<th>(\tau(KS, \theta))</th>
<th>(\tau(Ss, \theta))</th>
<th>(\tau(LC, \theta))</th>
<th>(\tau(LS\text{C}, \theta))</th>
<th>(\tau(LTSC, \theta))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karate</td>
<td>0.129</td>
<td>0.13</td>
<td>0.7540</td>
<td>0.6522</td>
<td>0.6902</td>
<td>0.6904</td>
<td>0.8278</td>
<td>0.8585</td>
<td>0.8245</td>
<td>0.9001</td>
</tr>
<tr>
<td>Polbook</td>
<td>0.0838</td>
<td>0.09</td>
<td>0.7814</td>
<td>0.3669</td>
<td>0.7588</td>
<td>0.7652</td>
<td>0.7787</td>
<td>0.9017</td>
<td>0.8743</td>
<td>0.9222</td>
</tr>
<tr>
<td>Football</td>
<td>0.0932</td>
<td>0.10</td>
<td>0.7514</td>
<td>0.2659</td>
<td>0.5413</td>
<td>0.1319</td>
<td>0.1319</td>
<td>0.7803</td>
<td>0.7053</td>
<td>0.7937</td>
</tr>
<tr>
<td>Netscience</td>
<td>0.1247</td>
<td>0.13</td>
<td>0.6256</td>
<td>0.3956</td>
<td>0.6061</td>
<td>0.5683</td>
<td>0.7547</td>
<td>0.8127</td>
<td>0.8288</td>
<td>0.9013</td>
</tr>
<tr>
<td>Email</td>
<td>0.0535</td>
<td>0.06</td>
<td>0.7913</td>
<td>0.6308</td>
<td>0.7711</td>
<td>0.8201</td>
<td>0.8628</td>
<td>0.9249</td>
<td>0.9119</td>
<td>0.9341</td>
</tr>
<tr>
<td>PowerGrid</td>
<td>0.2583</td>
<td>0.26</td>
<td>0.5899</td>
<td>0.4183</td>
<td>0.4982</td>
<td>0.5044</td>
<td>0.6452</td>
<td>0.7814</td>
<td>0.7764</td>
<td>0.8077</td>
</tr>
</tbody>
</table>
the spreading ability increases. Then the information would spread much farther away from the initial infected node. The curves of LTSC, LSC, and LC are much similar, while the other three curves display the opposite trends. That might be because of the difference of networks’ structures and the evaluating indexes.

In Figure 3(a), when $\beta$ is smaller than $\beta_{th}$, LSC and DC show better performance, while the two curves decrease gradually with $\beta$ growing more than $\beta_{th}$. BC, KS, and H-index are completely opposite with LTSC and LC. LTSC and LC show constantly growing trends near to 1 with $\beta$ increasing. Thus the number of the concordant ones is more than the number of the discordant ones. That is, Kendall’s $\tau$ coefficients of LC and LTSC increase with the growth of influence probability, which confirmed that the local neighbors’ information is effective in evaluating the influence of a node.

For the Polbook (see Figure 3(b)), the two curves of LTSC and LC are growing close to each other with $\beta$ increasing. Around $\beta_{th}$, the performances of LTSC, LC, and LSC are roughly the same, better than the others. The curves in Football network, Figure 3(c), show similar behaviors as the Polbook. The performances of KS and SS are not well in Football network because almost all of the nodes in Football network are the same K-cores and S-cores; those two methods lost their utility distinguishing them.

In Netscience, the clustering coefficients of nodes are larger than the other networks; that means the connections among the node’s neighbors are dense. It is proper that the topological structure is considered to evaluate the spreading ability. Even though both of the average degree and clustering coefficients of Power network are the smallest, LTSC takes the percentage of triangle structure into consideration which confirms its effectiveness in ranking the spreading ability of nodes. That is the reason of LSC and LTSC achieving better performances in a wide range of $\beta$ than LC in Figures 3(d) and 3(f).

The most weird thing in the six panels is the Email network, shown in Figure 3(e). The curves of DC, BC, H-index, KS, and SS decrease at first and then gradually increase. However, the trends of the three indexes, LTSC, LSC, and LC, show opposite behaviors. LTSC shows the best performance near the threshold $\beta_{th}$ but this is lost when $\beta$ is a little far away from it.
4.3. Correctness of the Most Influential Nodes. The rank of the most influential nodes is more significant in many real applications. People are more interested in the most influential nodes deserves to be discussed in detail, besides Kendall’s $\tau$ coefficient to evaluate each pair of ranks of the network.

Another measurement we introduce to explore the correlation of two ranks is the rank similarity function [33]. The rank similarity function $F(L)$ indicates the similarity between the top $L$ elements of two ranks. $F(L)$ for ranks $\theta$ and $\sigma$ is defined as

$$F(L) = \frac{|\theta(L) \cap \sigma(L)|}{|\theta(L) \cup \sigma(L)|},$$

(6)

where $\theta(L), \sigma(L)$ denote the sets of the first $L$ nodes of the rank $\theta, \sigma$, respectively. Let $\sigma = \{x_1, x_2, \ldots, x_n\}$ and $\theta = \{y_1, y_2, \ldots, y_n\}$ be two ranks, where $x_i$ and $y_j$ are nodes of the network, $1 \leq i, j \leq n$. Choose $L$ as any positive integer in $\{1, 2, \ldots, n\}$. $\theta(L)$ and $\sigma(L)$ denote the element sets of $\theta$ and $\sigma$. $\theta(L) \cap \sigma(L)$ and $\theta(L) \cup \sigma(L)$ are the set intersection and the union of $\theta(L)$ and $\sigma(L)$, respectively. Then $F(L) = k/(2L - k)$ if there are $k$ pairs of $(x_i, y_j)$ such that $x_i = y_j$, where $1 \leq i \leq j \leq n$ and $k = 0, 1, 2, \ldots, L$. Therefore, $0 \leq F(L) \leq 1$ for any fixed $L$. $F(L) = 1$ if $\theta(L)$ and $\sigma(L)$ are the same; that is, the top $L$ nodes are the same. $F(L) = 0$ if no pair of $(x_i, y_j)$ is the same. $F(L)$ is the similarity of the sets of the top $L$ nodes of two ranks; this can be regarded as a special case of Kendall’s $\tau$ coefficient. Similar function is to compare two sets of the top $L$ elements from two different ranks, while Kendall’s $\tau$ coefficient is to compare all ordered pairs of two ranks. On the other hand, the rank similarities $F(L)$ of two ranks are close to the number of $L$. If $L = n$, $F(n) = 1$ because the top $n$ elements are all the nodes of each rank. In fact, for any two ranks with given size $n$, $F(L)$ is a monotonically increasing function when $L \geq (1/2)n$.

If we study the most influential nodes, $L$ should be not too long. One obvious question is how many $L$ should be chosen to be appropriated to evaluate the performance of ranking? In this subsection, two strategies for the selection of the number of the top $L$ nodes are used to compare the rank similarity: one is to take same percentage of the total number of networks; the other is some fixed number. For the first selection strategy, $L$ is set as 10% and 50% nodes of the size of the networks. And for the second selection strategy $L$ is chosen for the top ten nodes.

We calculate the rank similarity function values of LTSC in six real networks and compare them with the other seven indexes, DC, BC, LSC, H-index, KS, Ss, and LC. $\sigma(L)$ is the top $L$ of the rank $\sigma$ obtained by the six different measures, respectively; $\theta(L)$ is the top $L$ of the rank $\theta$ offered by the SIR model simulation. The similarity functions on Karate, Polbook, Football, Netscience, Email, and Power networks are shown in Figures 4 and 5.

The values of $F(L)$ display much diversity of the top 10, 10%, and 50% nodes of Karate, Polbook, and Netscience networks. LTSC in Figure 4(b) is better than Figure 4(a), because the size of Karate is small, so the number of the top 10% of the sizes of Karate is less than the top 10. LTSC and LC reach to 1 alternately. The rank offered by LTSC in Polbook performs the best when $L = 10, 10\%$ and nearly the best when $L = 50\%$, as shown in Figures 4(d)–4(f). In Netscience network, Figures 4(g)–4(i) show that LTSC is the best among the eight indexes when $L = 10$ and better than others except LC when $L = 11$ to about 65 and then the best when $L \geq 65$.

In the Football network, shown as Figures 5(a)–5(c), LTSC displays the best performance of the top 5 nodes, same as the case in the top 10% nodes except three nodes in the six indexes’ ranks. LTSC, LC, LSC, and DC are the best alternating between the 12th and 20th node. LTSC is better than LC from the 21st node to the 60th node, even though there are ten nodes as exception.

LTSC shows almost the best performance in the top 10 except three nodes of Email and PowerGrid networks, as shown in Figures 5(d) and 5(g). In Email network, LTSC exhibits nearly the best of all the nodes in interval of the rank from the 10th to the 100th node. Meanwhile, in PowerGrid, LTSC does not perform as good as LC in the interval of the 75th to 500th nodes, as shown in Figure 5(h). And in the following interval from the 10% to the 50% nodes, they change to the opposite side, as shown in Figure 5(i); LTSC performs better than LC. KS and Ss show good performance when $L = 50\%$ while they lose advantage quickly after that.

Figures 4 and 5 show the good performance of LTSC, especially in evaluating the most influence nodes of networks. LTSC has absolutely superiority in the accuracy and the efficiency of the most top nodes compared with DC, H-index, and BC on entire range of $L$. And it also has great competitiveness in the accuracy for ranking on the whole $L$ length compared with LSC and DC. The results on the six networks illustrate that LTSC can do better than the other indexes in the ranking.

By comparing $F(L)$, the node rank offered by LTSC achieves higher correctness to the rank by the SIR model than the other measurements and also shows better evaluation for the most influence nodes. It is proved once again that the ranking based on the local information, the triangle structures together with the degrees, is more valuable than the indexes without them.
choose the threshold values of Netscience, Email, and Power as $\beta_{th} = 0.13, 0.06,$ and $0.26$. The results are displayed in Figure 6.

LTSC shows better performance than the other indexes in the above experiments, because LTSC takes the percentage of triangles and the number of neighbors of the node into consideration. The spreading ability heavily depends on the number of the neighbors when the spreading probability is smaller. Meanwhile the evaluation of spreading ability concerned not only the network structures, but also the
local neighbor’s structures when the spreading probability is increasing. Comparing with LC and LSC, LTSC employs the percentage of triangle structures rather than the cluster coefficient to denote the topological structure relations among the nodes. LTSC achieves its best performance when the spreading probability is around the threshold $\beta_{th}$. Figure 6 shows that the top nodes ranked by LTSC method have bigger spreading ability in the SI model.

Figure 5: The ranking similarity $F(L)$ for the top $L$ nodes of ranks offered by eight evaluating indexes with the SIR model on Football, Email, and Power Grid networks, respectively, where $L = 10, 10\%$, $50\%$.
5. Experimental Results of Synthetic Networks

Besides the real networks, we also investigate the performance of LTSC method on synthetic networks. The classical synthetic networks are Barabasi-Albert (BA) model [29] and the Lancichinetti-Fortunato-Radicchi (LFR) model [30].

In this section, the BA model is generated with the preferential attachment mechanism: Initially, there are $m_0 = 20$ nodes which are completely connected. At each time step $t$, a new node with $m_t$ links connects the existing nodes with probability proportion to their degrees. When the generation is terminated, the number of nodes is $n = 100, 200, 500$ and 1000 and $m_t = 3, 4, 5, 6$, respectively. We also use SIR model for examining the performance of our proposed method.

In the BA model, the preferential attachment mechanism makes the node's degrees in the network nearly equal; K-shell method may fail to rank nodes. Hence, we compare LTSC with the other six indexes except Ks method with the SIR model in Figure 7. For the fixed $n$ or $m_t$, the correctness curves of the Kendall’s $\tau$ coefficients are influenced by $m_t$ and $n$, where the infection probability $\beta$ is around the epidemic threshold $\beta_h$.

By Figures 7(a)–7(c), for the given size of nodes, it is easy to find that the performance of LC, LSC, and LTSC is better than DC, H-index, and BC methods on a wide range of $\beta$ in those networks. The curves of $\tau$ (LTSC, $\theta$) are getting better with $m_t$ growing, where $m_t$ is positively correlated with the edges of the network. That is, the performance of LTSC method is better than LC and LSC when the density of the network is growing.

In Figures 7(d)–7(f), for the fixed $m_t = 4$, with the growing of the number of nodes of the network, the three correctness curves, $\tau$ (LTSC, $\theta$), $\tau$ (LSC, $\theta$), and $\tau$ (LC, $\theta$), are gradually separated from the other three. And $\tau$ (LTSC, $\theta$) performs better than LSC and LC when the influence probability is around the threshold $\beta_h$.

Figure 7 shows that the proposed LTSC method performs better than the other five indexes when $\beta$ is around the epidemic threshold. And it is more correlated with the real spreading process. Also, even with the size changing, our method still performs better than others on the BA network.

Another kind of synthetic networks used to check the performance of the LTSC method is Lancichinetti-Fortunato-Radicchi (LFR) network model [30].

We fix some parameters, such as the maximum degree and the maximum and minimum community size which are set 50, 50, and 10, respectively. The other parameters in the LFR network model are set as follows: The number of nodes of networks $n = 500$ or 1000; the average degree $\langle d \rangle = 5$ or 10; the power-law exponents for the degree $\gamma = 2$ or 3; the exponent of the distribution of the communities’ size $\lambda = 1$ or 2; the communities mixing parameter $\mu$ is set to $\mu = 0.1, 0.3, 0.5$, respectively.

The LFR networks generated by the above parameters are denoted by $(n, \langle d \rangle, \mu, \gamma, \lambda)$. The cluster coefficients $C$ of those LFR networks are calculated and shown in Figure 8.

Overall, in the nine panels of Figure 8, the performance of LC, LSC, and LTSC is better than the other three in a wide range of $\beta$. In addition, the performance of LTSC method is better than LC and LSC when $\beta$ is around the threshold values.

Comparing the three panels of left column, Figures 8(a), 8(d), and 8(g), the trends of the three tau curves are almost steady. It is showed that the mixing parameter $\mu$ has no much effect of the Kendall $\tau$ coefficients. The two pairs of panels, Figures 8(b) and 8(e) and Figures 8(c) and 8(f), show that the power-law exponents for the degree $\gamma$ have effect on the trends of Kendall’s $\tau$. Because we fixed the parameters, $n, \langle d \rangle$ and $\lambda, \tau$ of LTSC performs better than the LC and LSC when $\gamma$ increases. Figures 8(c) and 8(d) show that Kendall’s $\tau$ is not much affected by the the power of the size of communities distribution $\lambda$ since the trends of those six curves have no much change. The average degree of the network affects Kendall’s $\tau$, shown as Figures 8(e) and 8(f) and Figures 8(h) and 8(i). Here the parameters $(m, \mu, \gamma, \lambda)$ in the two pairs of panels are fixed, respectively. The values of K-shell and S-shell of all nodes in Figures 8(e) and 8(f) are almost the same, so the curve of K-shell in the two panels is omitted. The change...
6. Conclusion

Identifying the most influential nodes is a fundamental problem due to its practical application in many areas, such as information dissemination and epidemic spread control. Thus, constructing efficient methods to detect the influential nodes is a valuable thing.

In this paper, we focus on the problem of finding influential nodes based on the local triangle structures and the degree of the node and present a centrality measure by considering both the neighbors connections and the degree of the node. This centrality definition leverages the proportion of triangle structures instead of local clustering coefficient to quantify the structural characteristics among the neighbors of the node. The property of triangle structures can measure how close its neighbors are connected to each other. The higher the proportion is, the more the connections among the neighbors are. And this can also reflect the location of an node in a network: the higher the proportion is, the more likely it is to be in a dense part of the node.

To evaluate the performance of the proposed LTSC method, we apply LTSC method on both synthetic and real networks. We compared LTSC with other seven centrality measures in terms of CCDF and find that LTSC is effective in assignment of distinct ranks to nodes with different spreading capabilities. Further, the SIR model is employed to simulate the real spreading process. By Kendall's tau rank correlation coefficient, we compute the rank correlation of the average degrees \( \langle d \rangle \) of networks makes the change of coefficients \( \tau \). The two Figures 8(g) and 8(h) show that the number of nodes in networks does not affect Kendall's \( \tau \).

All in all, the experiments on LFR networks show that LTSC method does perform the best when the influence probabilities are around the spreading threshold. Comparing the performance between LSC and our proposed LTSC method, we are convinced that the percentage of triangle structures is effective in measuring the local structural information among the neighbors of a node. It confirms that the local triangle structures together with the degrees among neighbors play an important role in measuring the spreading ability of a node.
between the two ranks generated by the SIR model and the one of centrality measures. The results demonstrate that LTSC method is better correlated with the real spreading process and outperforms the other local and semilocal methods in evaluating the node’s influence in most cases. The SI model also shows the effectiveness of LTSC. Furthermore, other comprehensive experiments also demonstrate that LTSC is more accurate in the identifying the most influential nodes.

Finally, we conduct the experiments on synthetic networks, the BA mode, and the LFR network model in scale-free networks with different sizes and different community size. The results show that LTSC method performs better than
the other centrality measures in evaluating the influence of the node. As further work, it will be necessary to find which measures work better in which type of networks.

Data Availability
The data used to support the findings of this study are included within the article.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Acknowledgments
We show our great appreciation to all the authors who collected and shared the data, such as Karate, Dolphins, Football, Polbook, Email, Netscience, and Facebook, and the methods, such as degree, betweenness, H-index, local centrality, local structure centrality, K-shell, and S-shell, to be used as benchmarks. Finally, we would like to acknowledge the National Natural Science Foundation of China (No. 71471106) that supports this research.

References
