Node Similarity Measure in Directed Weighted Complex Network Based on Node Nearest Neighbor Local Network Relative Weighted Entropy

WANCHANG JIANG AND YINGHUI WANG
School of Computer Science, Northeast Electric Power University, Jilin 132012, China
Corresponding author: Wanchang Jiang (jwchang84@163.com)

This work was supported in part by the Project of the National Natural Science Foundation of China under Grant 61572420, in part by the Research Project of the Education Department of Jilin Province under Grant JJKH20190706KJ, and in part by the Science and Technology Innovation Development Program of Jilin under Grant 20190104140.

ABSTRACT Node similarity is a significant basis for analyzing features in complex network. For complex network with directed weighted edge, the complexity of the relationship among nodes and the diversity of relationship weights make measure node similarity complicated from the huge amounts of nodes. Therefore, a novel node similarity measure is proposed based on the design of node nearest neighbor local network relative weighted entropy. Firstly, degree and strength based directed weighted complex network model is constructed, on the basis of that, the Node Nearest Neighbor Local network is defined. The structural features of each node in the local network are quantized into a set of decision probabilities with multiple indicators. Furthermore, Node Connection Tightness is given to compute the influence of structural complex relationship in local network on node similarity. And then, to evaluate how the outward & inward degree and strength of nodes in the local network affect node similarity, the concept of Node Nearest Neighbor Local Network Relative Weighted Entropy is designed to define Node Relative Difference for measuring the structural difference between any two nodes. Accordingly, a novel node similarity measure is designed to measure the similarity between any pair of nodes in directed weighted complex network, and then the Similar Node Mining Algorithm is proposed to obtain most similar nodes. To clarify the availability and effectiveness of the proposed measure, two sets of experiments were conducted on real-world complex networks. The results show that the measure can not only mine nodes with the most similarity in the same module, but also mine the most similarity nodes from different modules.

INDEX TERMS Complex network, relative weighted entropy, nearest neighbor local network, node similarity measure.

I. INTRODUCTION

The emergence of complex systems has caused widespread concern in many areas. Most complex systems can be abstracted into a complex network of nodes and connected edges. As a typical form of a complex system, there are many researches on complex networks recently, such as the robustness of complex networks [1], [2], attack strategies on network [3], [4], node invulnerability measures [5], [6], important nodes measures [7], [8] and network community division [9].

In the study of complex network, one of the most important directions is node similarity measure, and measuring node similarity is a fundamental work. Many studies on complex network are based on node similarity measures. For example, in complex networks, the classification of nodes [10], importance assessment of nodes [11], link prediction [12] and community division [13]–[15], these studies are based on node similarity measures.

Researchers have proposed various methods to measure node similarity, for example, node similarity is measured based on path and distance [16], including most existing node similarity measures are based on nodes and their neighbor nodes [17], [18]. To simplify measuring, more studies are on
undirected unweighted networks [19]. Although it is faster and simpler to measure, much more node information and complex relationships between nodes are lost, resulting in incomplete decision information of a large number of nodes. This makes the bipartite degree of most node similarity obtained by measures is small, and node similarity cannot be accurately computed. With the development of research, there are new researches on weighted complex networks [20]. For weighted complex networks, the weights of node itself and its neighbors are introduced. Even if the weights are considered, the influence of nodes’ complex relationship information on node similarity in the local network topology cannot be effectively measured.

Since most nodes in the network usually look similar in structure characteristic and are simply to be measured, quantitative structures and internal features have become the focus of measuring node similarity. Compared to undirected unweighted complex networks, directed weighted complex networks have some complex factors, such as the complexity of the relationship between nodes and the diversity of relationship weights, these factors make node similarity measure more complicated.

Thence, in this paper, in order to measure node similarity in complex networks with directed weighted edges accurately, a degree and strength based directed weighted complex network model is constructed. And on the basis of the model, a Nearest Neighbor Local network can be identified, which takes into account both the nodes’ local structure and details of the relationship among nodes. In the stage of measuring node similarity, the weighted entropy and relative weighted entropy are both introduced, and the Connection Tightness is given, Node Nearest Neighbor Local Network Relative Weighted Entropy is proposed to achieve the Node Similarity Measure. And then, Similar Node Mining Algorithm is designed to measure node similarity and obtain most similar nodes in complex network accurately. The processing diagram of Node Similarity Measurement and similar nodes mining in directed weighted complex network is shown in Figure 1.

The rest of paper is organized as follows. Section II presents a directed weighted complex network model. Section III introduces Nearest Neighbor Local network with some primary definitions, and the design of the Node Nearest Neighbor Local network relative weighted entropy is described. In Section IV, with the above work, a novel Similar Node Mining Algorithm is proposed to measure node similarity in complex network. And the experimental evaluation to verify the feasibility and effectiveness of the proposed algorithm will be given in Section V. The conclusion and future work are discussed in Section VI.

II. DEGREE AND STRENGTH BASED DIRECTED WEIGHTED COMPLEX NETWORK MODEL

In order to analyze the node similarity of any two nodes in complex network, a degree and strength based directed weighted complex network model is designed and constructed as $G(V, E, W, K, S)$. In given network, the set of nodes is denoted as $V = \{v_i\}$, $E = \{e_{ij}\}$ denotes the set of directed edges from $v_i$ to $v_j$, namely $e_{ij} = \{(v_i, v_j)\}$, the set of weight can be expressed as $W = \{w_{ij}\}$ correspondingly, and $w_{ij}$ is the weight value of $e_{ij}$. $K = \{k_{in}(v_i), k_{out}(v_i), k(v_i)\}$ indicates the set of three values of each node, which consists of out-degree $k_{out}(v_i)$, in-degree $k_{in}(v_i)$ and degree $k(v_i)$ of $v_i$.

\[
\begin{align*}
    k_{in}(v_i) &= \sum_{j=1}^{N} e_{ji} & i = 1, 2, \cdots, N \\
    k_{out}(v_i) &= \sum_{j=1}^{N} e_{ij} & i = 1, 2, \cdots, N \\
    k(v_i) &= k_{in}(v_i) + k_{out}(v_i)
\end{align*}
\]

\[
S = \{(s_{in}(v_i), s_{out}(v_i), s(v_i))\}
\]

\[
\sum_{j=1}^{N} w_{ji} & i = 1, 2, \cdots, N
\]

\[
\sum_{j=1}^{N} w_{ij} & i = 1, 2, \cdots, N
\]

\[
s(v_i) = s_{in}(v_i) + s_{out}(v_i)
\]

Take one Gene network as an example, its directed weighted complex network model is built as shown in Figure 2. The weight in the model is a correct expression of the actual complex network, and arrow size in this figure represents the weight of the directed edge in the network, which indicates the amount of exchanging chemical signals between genes.

**Definition Node Nearest Neighbor Local Network:** In a degree and strength based directed weighted complex network model, for the structural relationship between each
W. Jiang, Y. Wang: Node Similarity Measure in Directed Weighted Complex Network

FIGURE 2. Degree and strength based directed weighted complex network model of Gene network.

FIGURE 3. The Node Nearest Neighbour Local network of node 360.

III. NODE SIMILARITY MEASURE BASED ON NODE NEAREST NEIGHBOR LOCAL NETWORK

RELATIVE WEIGHTED ENTROPY

As an asymmetrical measure of the difference between two probabilities, the relative entropy [21] is an important conception in the probability theory and information theory.

The relative entropy of two probability distributions $R$ and $T$ can be described in the following form.

$$D_{KL}(R \parallel T) = - \sum_{i=1}^{N} R(x) \cdot \log \frac{R(x)}{T(x)}$$

where $N$ is the number of elements in the probability set $R$ and $T$.

In order to measure node similarity in complex network, Node Nearest Neighbor Local network is defined firstly for each node in Section II, which including the node itself, its direct neighbor nodes and the corresponding relationships among these nodes. With the local network, except for information of nodes and edges, the information of its local structure is also retained for one node.

Therefore, on basic of the nearest neighbor local network of a node, the concept of relative entropy is introduced to propose Node Nearest Neighbor Local Network Relative Weighted Entropy by using the following definitions 1 to 6. And then, based on Relative Weighted Entropy, Node Relative Difference can be designed to define a novel Node Similarity Measure.

**Definition 1:** In directed weighted complex network, the node similarity of one node depends on its outward & inward degree and strength. Hence, the ratio of out-degree to degree, in-degree to degree, out-strength to strength and in-strength to strength of each node are considered simulta-
neously, and Degree & Strength based Node’s own Properties Set can be defined as $P(O_{i}^{\text{Deg}}, O_{i}^{\text{Str}}, I_{i}^{\text{Deg}}, I_{i}^{\text{Str}})$ for node $v_{i}$.

\[
O_{i}^{\text{Deg}} = \frac{k_{\text{in}}(v_{i})}{k(v_{i})} \quad (8)
\]

\[
O_{i}^{\text{Str}} = \frac{s_{\text{out}}(v_{i})}{s(v_{i})} \quad (9)
\]

\[
I_{i}^{\text{Deg}} = \frac{k_{\text{in}}(v_{i})}{k(v_{i})} \quad (10)
\]

\[
I_{i}^{\text{Str}} = \frac{s_{\text{in}}(v_{i})}{s(v_{i})} \quad (11)
\]

Definition 2: For each node in the complex network model, if $v_{1}$ has directly connected edge $e_{ij}$ with $v_{j}$, that is, $v_{j}$ is one direct neighbour node of $v_{1}$. And $v_{1}$ itself and all its $m_{i}$ direct neighbour nodes constitute a Direct Neighbour Node Set $V_{i}^{D}$ of $v_{i}$. Based on this set $V_{i}^{D}$, a node Nearest neighbor local network of $v_{i}$ is defined, which includes the strength of nodes and the corresponding neighbor relationships among nodes.

Definition 3: With degree and strength of static characteristic index of $v_{i}$ itself and its direct neighbour nodes in $V_{i}^{D}$, a probability set of node degree and strength named Neighbourhood Node Information Set is defined as $Q_{i}^{D} = \{(K_{i}^{D}, \text{sum}(K_{i}^{D})), (S_{i}^{D}, \text{sum}(S_{i}^{D}))\}$ for node $v_{i}$, where $K_{i}^{D}$ and $S_{i}^{D}$ are degree set and strength set of $V_{i}^{D}$ respectively. And $\text{sum}(K_{i}^{D})$ is the sum of degree of all nodes in $V_{i}^{D}$, similarly, $\text{sum}(S_{i}^{D})$ is the sum of strength of all nodes in $V_{i}^{D}$.

As $G$ is a directed weighted network, the probability element $q_{\text{out}}(k(v_{i}), \text{sum}(K_{i}^{D})), (s(v_{i}), \text{sum}(S_{i}^{D}))$ can be divided into two parts, namely, $q_{\text{out}}(k(v_{i}), \text{sum}(K_{i}^{D})), (s(v_{i}), \text{sum}(S_{i}^{D}))$ and $q_{\text{in}}(k(v_{i}), \text{sum}(K_{i}^{D})), (s(v_{i}), \text{sum}(S_{i}^{D}))$, which express the node information probability of transmitting node and receiving node respectively. And then, the above divided probability elements of $v_{i}$ can be calculated on the basis of node degree $k(v_{i})$ and node strength $s(v_{i})$ respectively.

\[
q_{\text{out}}(k(v_{i}), \text{sum}(K_{i}^{D})), (s(v_{i}), \text{sum}(S_{i}^{D})) = \frac{k_{\text{out}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} k_{\text{out}}(v_{j}) + 1} \cdot \frac{s_{\text{out}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} s_{\text{out}}(v_{j}) + 1} \quad (13)
\]

\[
q_{\text{in}}(k(v_{i}), \text{sum}(K_{i}^{D})), (s(v_{i}), \text{sum}(S_{i}^{D})) = \frac{k_{\text{in}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} k_{\text{in}}(v_{j}) + 1} \cdot \frac{s_{\text{in}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} s_{\text{in}}(v_{j}) + 1} \quad (14)
\]

With a comprehensive application of formula (12) to formula (14), the set $Q_{i}$ can be obtained and expressed as follows.

\[
Q_{i}^{D} = \left\{ (K_{i}^{D}, \text{sum}(K_{i}^{D})), (S_{i}^{D}, \text{sum}(S_{i}^{D})) \right\} = \left[ \frac{k_{\text{out}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} k_{\text{out}}(v_{j}) + 1} \cdot \frac{s_{\text{out}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} s_{\text{out}}(v_{j}) + 1} \right] + \left[ \frac{k_{\text{in}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} k_{\text{in}}(v_{j}) + 1} \cdot \frac{s_{\text{in}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} s_{\text{in}}(v_{j}) + 1} \right] \quad (15)
\]

Definition 4: In the directed weighted complex network model, the weight values of edges are transformed to reflect the degree of the tightness of connection between nodes in the local network topology. Therefore, the Node Connection Tightness of $v_{i}$ and $v_{j}$ can be defined and calculated by the following formula.

\[
\text{con}_{ij} = \frac{w_{ij}}{s_{\text{out}}(v_{i}) \cdot s_{\text{in}}(v_{i})} \quad (16)
\]

For the weight $w_{ij}$ between $v_{i}$ and $v_{j}$ has symmetry feature, and $s_{\text{in}}$ and $s_{\text{out}}$ are positive values, therefore the node connection tightness $\text{con}_{ij}$ in directed weighted complex network has symmetry $\text{con}_{ij} = \text{con}_{ji}$.

Definition 5: Owing to the complexity of the relationship between one node and its nearest neighbor nodes in different scales, on the basis of definition 3, and the probability of node degree and strength in definition 1 is also used. Node Multiple Indexes Decision Probability Set $R_{i}$ can be defined as follows.

\[
R_{i}^{D} = \left\{ r(v_{i}), r(v_{j}), \ldots, r(v_{i}+1) \right\} \quad (v_{i} \in V_{i}^{D}) \quad (17)
\]

\[
r_{\text{out}}(v_{i}) = \left[ \frac{k_{\text{out}}(v_{i})}{k(v_{i})} \cdot \frac{s_{\text{out}}(v_{i})}{s(v_{i})} \right] + \left[ \frac{k_{\text{in}}(v_{i})}{k(v_{i})} \cdot \frac{s_{\text{in}}(v_{i})}{s(v_{i})} \right] \quad (18)
\]

\[
r_{\text{out}}(v_{i}) = \frac{k_{\text{out}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} k_{\text{out}}(v_{j}) + 1} \quad (19)
\]

\[
r_{\text{in}}(v_{i}) = \frac{k_{\text{in}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} k_{\text{in}}(v_{j}) + 1} \quad (20)
\]

\[
r_{\text{out}}(v_{i}) = \frac{s_{\text{out}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} s_{\text{out}}(v_{j}) + 1} \quad (21)
\]

\[
r_{\text{in}}(v_{i}) = \frac{s_{\text{in}}(v_{i})}{\sum_{v_{j} \in V_{i}^{D}} s_{\text{in}}(v_{j}) + 1} \quad (22)
\]

The determination of the Node Multiple Indexes Decision Probability Set is beneficial to calculate Node Nearest Neighbor Local network Entropy accurately.

Definition 6: To evaluate how outward & inward degree and strength of Nearest neighbor local network of $v_{i}$ affect its node similarity with other nodes in the network, Node Nearest Neighbor Local Network Relative Weighted Entropy $H$ is designed based on the structure of weighted entropy. And the entropy is divided into node outward and inward parts of Node Nearest Neighbor Local Network Relative Weighted Entropy, which can be expressed as $H_{\text{out}}$ and $H_{\text{in}}$ respectively.

\[
H_{\text{out}}(v_{i}) = - \sum_{(v_{i}, v_{j}) \in V_{i}^{D}} \left[ \text{con}_{ij} + 1 \right] \cdot r_{\text{out}}(v_{i}) \quad (23)
\]
With Node Nearest Neighbor Local Network Relative Weighted Entropy value of each node, the difference between any two nodes can be compared by using the two nodes’ Node Relative Difference, which is defined as follows.

**Definition 7:** According to the decision property value of nodes, the complexity of the relationship and the diversity of relationship weights between the two nodes \( v_i \) and \( v_j \) will be defined as Node Relative Difference \( RD_{KL}(H(v_i)||H(v_j)) \), which can be obtained as follows.

\[
RD_{KL}(H(v_i)||H(v_j)) = \sum_{i,j=1}^{N} H(v_i) \cdot \frac{H(v_j)}{H(v_i)} \tag{26}
\]

As can reason from the formula (26), \( RD_{KL}(H(v_i)||H(v_j)) \) and \( RD_{KL}(H(v_j)||H(v_i)) \) are not equal. Therefore, the similarity between the two nodes \( v_i \) and \( v_j \) needs to be calculated jointly by \( RD_{KL}(H(v_i)||H(v_j)) \) and \( RD_{KL}(H(v_j)||H(v_i)) \).

By utilizing the node relative difference between \( v_i \) and \( v_j \), a novel Node Similarity Measure based on Node Nearest Neighbor Local Network Relative Weighted Entropy (LRWE-SNM) can be designed and expressed as follows.

\[
Sim_{ij} = 1 - \frac{RD_{KL}(H(v_i)||H(v_j)) + RD_{KL}(H(v_j)||H(v_i))}{2} \tag{27}
\]

In summary, the above formulas will be used to get the Node Similarity measure of Node Nearest Neighbor Local network Relative Weighted Entropy (LRWE-SNM). And then the similarity between any two nodes in complex network can
be computed, the larger measure value between two nodes means the greater similarity between these two nodes.

### IV. SIMILAR NODE MINING ALGORITHM

To mine high similar nodes in a directed weighted complex network, a Similar Node Mining Algorithm is proposed by using Node Nearest Neighbor Local Network Relative Weighted Entropy based Node Similarity Measure. With the algorithm, the similarity of each pairs of nodes in the network can be measured, and then the most similar nodes in network can be mined.

**Algorithm** Similar Node Mining Algorithm

Input: A directed weighted complex network
Output: Similar nodes set $V_{Sim}$

1. Get the degree and strength based directed weighted complex network model $G(V, E, W, K, S)$;
2. For each $v_i$ in $V$ ($i = 1, 2, \cdots, N$) do
   3. Identify the Node Nearest Neighbor Local network of $v_i$ and get its direct neighbor node set $V_{D_i}$;
4. For each $v_i$ in $V$ do 5 to 8
   5. Calculate the Degree & Strength-based Node’s own Properties Set $P_i$ of $v_i$;
   6. Calculate the Neighborhood Node Information Set $Q_i$ \{\[(K_{D_i}^D, \text{sum}(K_{D_i}^D)), (S_{D_i}^D, \text{sum}(S_{D_i}^D))\]\} by using formula (12) to formula (15);
   7. Calculate the Node Connection Tightness $con_{ij}$ by using formula (16);
8. Obtain Node Multiple Indexes Decision Probability Set $R_i$ with Node Properties Set $P_i$ $(O_{Deg}^{Deg}, O_{Str}^{Deg}, I_{Deg}^{Str})$ and Neighborhood Node Information Set $Q_i$ \{\[(K_{D_i}^D, \text{sum}(K_{D_i}^D)), (S_{D_i}^D, \text{sum}(S_{D_i}^D))\]\};
9. For each $v_i$ in $V$ do 10 to 14
10. Calculate node outward influence measure $H_{out}$ and node inward influence measure $H_{in}$ by using formula (23) and formula (24) respectively, and Node Nearest Neighbor Local network Entropy $H$ can be obtained with Definition 6;
11. For each $v_i (i \neq j)$ in $V$ do 12 to 13
12. Using the above Node Nearest Neighbor Local network Entropy, the Node Relative Difference $RD_{KL}(H(v_i)||H(v_j))$ of nodes $v_i$ and $v_j$ can be get;
13. Calculate node similarity measure value $Sim_{ij}$ between $v_i$ and $v_j$
14. Rank all nodes in $V$ by using measure values in an ascending sort order
15. Obtain the Similar nodes set $V_{Sim} = \{V_{Sim_1}, \cdots, V_{Sim_T}\}$ $(1 \leq t \leq T)$

As can be seen from the Similar Node Mining Algorithm, it is a five-stage process. The first stage is step 1, the directed weighted complex network model is constructed, which will be used for obtaining structural information to measure node similarity in the network. In the second stage, from steps 2 to 3, the Node Nearest Neighbor Local network can be identified for each node on the basis of the above model. And the third stage is steps 4 to 8, where in each node’s Node Properties Set and Neighborhood Node Information Set can be calculated, and the value of Node Connection Tightness of each node can also be obtained. And by considering the complexity of the relationship between nodes within the local network, the Node Multiple Indexes Decision Probability Set can be get. The fourth stage is steps 9 to 10, with the above information, Node Nearest Neighbor Local network Entropy of each node can be calculated. The last stage is steps 11 to 15, the Node Relative Difference and the node similarity measure are computed in order, and then all high similar nodes set $V_{Sim}$ are obtained, which contains top-$t$ similar nodes subsets.
In this section of experiment, to verify the feasibility and rationality of Node Nearest Neighbor Local Network Relative Weighted Entropy based Similar Node Mining Algorithm, the algorithm is applied to two groups of reality complex networks with varying sizes. (i) E.coli transcription network [22] with 95 nodes and 213 edges. In this network, the ‘node’ is operon, and each ‘edge’ is directed from an operon that encodes a transcription factor to an operon that it directly regulates (an operon is one or more genes transcribed on the same mRNA). There are also three gene networks of different sizes. (ii) Gene network with 636 nodes and 3959 edges [23]. In this network, ‘node’ represents gene, and the weight value of ‘edge’ are not much different from each other.

Two sets of experiments were conducted, one set is the overall similar nodes scatter plots obtained under four different similarity measures, which reflects the distribution of similarity value between nodes in complex network. The other set is top similar node tables mined under the different similarity measures, which is analyzed for comparing accuracy between the proposed measure and other three ones.

A. NODES SIMILAR SCATTER PLOTS

In order to evaluate the feasibility of the proposed Similar Node Mining Algorithm in measuring node similarity and mining top similar nodes of network, LRWE-SNM, Local relative entropy (LRE) [19], and two typical measures Eigen-vector (EI) [24] and Weighted Degree (WD) [25] are used for measuring node similarity. And the overall similar nodes scatter plots of E.coli and Gene network under the four measures are shown in Figure 4 to Figure 7 respectively.

As can be seen from Figure 4(a) and 4(b), the node similar scatter plots by measures LRWE-SNM and LRE are showed respectively, and each node has only one most similar node. From Figure 5(a) and 5(b), and different colors represent the different most similar nodes for each node. And these two scatter plots are symmetric about the function $y = x$, which is consistent with the symmetry of the proposed measure. Furthermore, one node may have multiple most similar nodes, which makes the similarity measurement of the nodes lower differentiation degree and less accurate.

As shown in Figure 6(a) and 6(b), although each node corresponds to one of the most similar nodes, from overall view, the nodes in Figure 6(a) are more dispersed and the degree of discrimination is larger. This makes the similarity value between nodes in the network better distinguishable, and each node has the most similar nodes corresponding to itself.

Figure 7(a) shows the disparity in node dispersion under WD, and 7(b) shows the partial enlargement with nodes No. 1 to 100, in which each nodes contains more similar nodes. And the remaining nodes in the Figure 7(a) have fewer similar nodes.

In LRWE-SNM and LRE methods, by comparing the networks of different sizes in Figure 4 to Figure 7, it can be found that the most similar node measured by LRE is close to nodes in its local network, and few most similar nodes are similar to nodes in other parts of the network. With the increase of the scale of the network, this distribution phenomenon becomes more and more obvious. Furthermore, the most similar nodes...
W. Jiang, Y. Wang: Node Similarity Measure in Directed Weighted Complex Network

FIGURE 8. The similar nodes of Gene network with structure characteristics.

TABLE 1. The top 10 node similarity values and corresponding pairs of nodes of Gene networks.

| Node | Most similar node | Node Similarity value |
|------|-------------------|-----------------------|
| 9    | 434               | 0.999999              |
| 50   | 58                | 0.999999              |
| 51   | 57                | 0.999999              |
| 53   | 55                | 0.999999              |
| 72   | 253               | 0.999999              |
| 76   | 110               | 0.999999              |
| 81   | 602               | 0.999999              |
| 104  | 636               | 0.999999              |
| 135  | 297               | 0.999999              |
| 172  | 348               | 0.999999              |

TABLE 2. The last 10 node similarity values and corresponding pairs of nodes of Gene network.

| Node | Most similar node | Node Similarity value |
|------|-------------------|-----------------------|
| 160  | 18                | 0.999989              |
| 19   | 160               | 0.999988              |
| 109  | 18                | 0.999986              |
| 23   | 429               | 0.999984              |
| 25   | 24                | 0.999984              |
| 21   | 22                | 0.999977              |
| 78   | 157               | 0.999971              |
| 616  | 314               | 0.999966              |
| 122  | 52                | 0.999898              |
| 5    | 122               | 0.994657              |

found by LRWE-SNM are more dispersed in the overall network, and the most similar nodes can be mined in the entire structure of network. At the same time, this will not be affected by the network size.

B. TOP SIMILAR NODES UNDER DIFFERENT MEASURES

In order to illustrate the effectiveness of the proposed similarity measure, top similar nodes under different measures of Gene network with 636 nodes and 3959 edges was analyzed in combination with the above scatter plot. First, the values of similarity between each two nodes are arranged in a large to small order. When the similarity of a number of nodes is the same, the first 10 nodes are selected according to the serial number of the nodes, and the selected values are the nodes corresponding to the top 10 and the last 10, which are used as analysis objects. The similarity values between these
TABLE 3. The most similar nodes obtained by LRWE-SNM, LRE, EI and WD.

| Node | LRWE-SNM | LRE | EI (The number of most similar nodes) | WD (The number of most similar nodes) |
|------|----------|-----|--------------------------------------|---------------------------------------|
| 9    | 434      | 76  | /                                    | /                                     |
| 50   | 58       | 76  | /                                    | ⋅⋅⋅ (15)                              |
| 51   | 57       | 76  | /                                    | ⋅⋅⋅ (5)                               |
| 53   | 55       | 76  | /                                    | /                                     |
| 72   | 253      | 508 | ⋅⋅⋅ (148)                            | /                                     |
| 76   | 110      | 80  | /                                    | /                                     |
| 81   | 602      | 85  | /                                    | /                                     |
| 104  | 636      | 557 | 557, ⋅⋅⋅, 636 (148)                  | 105                                   |
| 135  | 297      | 549 | 297, ⋅⋅⋅ (148)                       | /                                     |
| 172  | 348      | 201 | /                                    | /                                     |
| 434  | 9        | 438 | /                                    | /                                     |
| 58   | 50       | 76  | /                                    | ⋅⋅⋅ (15)                              |
| 57   | 51       | 76  | /                                    | ⋅⋅⋅ (15)                              |
| 55   | 53       | 76  | /                                    | ⋅⋅⋅ (15)                              |
| 253  | 72       | 251 | ⋅⋅⋅ (7)                              | /                                     |
| 110  | 76       | 585 | 585, ⋅⋅⋅ (148)                       | /                                     |
| 602  | 81       | 590 | 603                                  | /                                     |
| 636  | 104      | 181 | 104, ⋅⋅⋅ (148)                       | /                                     |
| 297  | 135      | 557 | 135, 557, ⋅⋅⋅ (148)                 | ⋅⋅⋅ (3)                               |
| 348  | 172      | 346 | /                                    | /                                     |
| 19   | 160      | 76  | /                                    | ⋅⋅⋅ (15)                              |
| 109  | 18       | 80  | /                                    | /                                     |
| 23   | 429      | 76  | /                                    | ⋅⋅⋅ (15)                              |
| 25   | 24       | 76  | /                                    | ⋅⋅⋅ (15)                              |
| 21   | 22       | 76  | /                                    | ⋅⋅⋅ (13)                              |
| 78   | 157      | 87  | /                                    | /                                     |
| 616  | 314      | 483 | /                                    | /                                     |
| 5    | 122      | 76  | 122, ⋅⋅⋅ (148)                       | ⋅⋅⋅ (13)                              |
| 160  | 18       | 220 | ⋅⋅⋅ (148)                            | /                                     |
| 18   | 160      | 76  | /                                    | ⋅⋅⋅ (4)                               |
| 429  | 326      | 434 | 326, ⋅⋅⋅ (148)                       | /                                     |
| 24   | 326      | 76  | /                                    | ⋅⋅⋅ (4)                               |
| 22   | 305      | 76  | /                                    | ⋅⋅⋅ (15)                              |
| 157  | 13       | 126 | /                                    | /                                     |
| 314  | 491      | 520 | ⋅⋅⋅ (18)                             | /                                     |
| 52   | 54       | 76  | /                                    | ⋅⋅⋅ (13)                              |
| 122  | 52       | 133 | ⋅⋅⋅ (148)                            | /                                     |

nodes and their corresponding most similar nodes are listed in Table 1 and Table 2 respectively as follows.

Next, the nodes in Table 3 were extracted comprehensively from Table 1 and Table 2, and the duplicate values were removed. And then, the most similar nodes measured by LRWE-SNM, LRE, EI and WD are listed in Table 3 as follows. Columns LRWE-SNM and LRE show the most similar node. Since the number of the most similar nodes obtained by the measures EI and WD is large, so, not all of nodes are listed in the table, columns EI and WD only show the same nodes as the previous two columns showed, and the remaining most similar nodes are indicated by ellipsis. The number in parentheses represents the number of the most similar nodes mined by this measure. “/” means there is no most similar nodes for corresponding node in the first column.

According to the results shown in Table 3, in the column of measure LRE, one of most similar node for many nodes in the first column is 76, and the corresponding nodes in the first column still have more similar nodes or no similar nodes in measures EI and WD.

For the Gene network model, the similar nodes of Gene network with structure characteristics are revealed as shown in Figure 8. Select nodes that the most similar nodes is 76 as example. As can be seen from Figure 8, node 76 is in the red area. And the nodes with the most similar node of 76 mined by LRE measure are node 5, 18, 23, 24, etc,
which are represent in the same area and express as green in the figure. However, the most similar nodes mined by the method designed in this paper is distributed throughout the network. For example, node 76 and node 9 distributed in green area are the most similar nodes to each other by using LRE method, and node 9 and node 434 distributed in yellow area are the most similar nodes to each other by using LRWE-SNM method, this can explain that the method proposed in this paper can not only mined nodes with large similarities in the same module, but also nodes with large similarities between different modules.

VI. CONCLUSION

In order to accurately measure the node similarity in directed weighted complex network, a complex network model with defined Node Nearest Neighbor Local network is designed and constructed. According to this model, the characteristics of nodes and edges in the structure are digitized. With the form of weighted entropy and relative entropy, the Node Nearest Neighbor Local Network Relative Weighted Entropy is designed on the basis of the Node Nearest Neighbor Local network, the difference of structural information between nodes can be evaluated. On this basis, a Similar Node Mining Algorithm based on Node Nearest Neighbor Local Network Relative Weighted Entropy is designed to measure the similarity of any two nodes in a directed weighted complex network. In addition, through two sets of comparison experiments with other three classic measure algorithms on reality complex networks of different scales. The experimental results show that the most similarity node in the same module can be obtained, and the most similarity nodes from different modules even can be mined by the measure.

In future work, to measure the heterogeneity of complex directed network more precisely, the structure of Node Nearest Neighbor Local network are used based on entropy, the difference between local networks is considered to quantify the structural heterogeneity of directed weighted complex network.

REFERENCES

[1] M. Bellingeri and D. Cassi, “Robustness of weighted networks,” Phys. A, Stat. Mech. Appl., vol. 489, pp. 47–55, Jan. 2018.
[2] S. Iyer, T. Killingback, B. Sundaram, and Z. Wang, “Attack robustness and centrality of complex networks,” PLoS ONE, vol. 8, no. 4, Apr. 2013, Art. no. e59613.
[3] M. Bellingeri, D. Cassi, and S. Vincenzi, “Efficiency of attack strategies on complex model and real-world networks,” Phys. A, Stat. Mech. Appl., vol. 414, pp. 174–180, Nov. 2014.
[4] T. Nie, Z. Guo, K. Zhao, and Z. M. Lu, “New attack strategies for complex networks,” Phys. A, Stat. Mech. Appl., vol. 424, pp. 248–253, Apr. 2015.
[5] X. Peng, H. Yao, J. Du, Z. Wang, and C. Ding, “Inviability of scale-free network against critical node failures based on a renewed cascading failure model,” Phys. A, Stat. Mech. Appl., vol. 421, pp. 69–77, Mar. 2015.
[6] X. Peng, Y. Hong, and Z. Zhihao, “Research on cascading inviability of scale-free networks under node-based-attack,” Syst. Eng. Electron., vol. 35, no. 9, pp. 1974–1978, 2013.
[7] T. Qiao, W. Shan, G. Yu, and C. Liu, “A novel entropy-based centrality approach for identifying vital nodes in weighted networks,” Entropy, vol. 20, no. 4, p. 261, Apr. 2018.
[8] T. Qiao, W. Shan, and C. Zhou, “How to identify the most powerful node in complex networks? A novel entropy centrality approach.” Entropy, vol. 19, no. 11, p. 614, Nov. 2017.
[9] X. You and M. Ge, “Community detection algorithm based on belief propagation in complex networks,” J. Comput. Appl., vol. 37, no. 11, pp. 3115–3118, 2017.
[10] X. Zheng, “Communication nodes classification method in directed sensor network,” Comput. Simul., vol. 31, no. 8, 2014, pp. 301–304.
[11] Y. Ruan, S. Lao, J. Wang, L. Bai, and L. Chen, “Node importance measurement based on neighborhood similarity in complex network,” Acta Phys. Sin, vol. 66, no. 3, pp. 371–379, 2017.
[12] K. Zhou, T. P. Michalak, and Y. Vorobeychik, “Adversarial robustness of similarity-based link prediction,” in Proc. IEEE Int. Conf. Data Min-ing (ICDM), Nov. 2019, pp. 1–10.
[13] L. Gan, X. Wang, N. Yan, and L. Wang, “Detecting community structure in complex networks based on node similarity,” Comput. Digit. Eng., vol. 46, no. 2, pp. 213–217, 2018.
[14] Z. Liang, F. Yang, and J. Li, “Community structure detection based on node similarity in complex networks,” J. Comput. Appl., vol. 35, no. 5, pp. 1213–1217, 2015.
[15] Z. Hao, L. Li, and H. Shi, “Multi-label propagation algorithm for community division based on node comprehensive similarity.” Appl. Res. Comput., vol. 36, no. 6, pp. 1705–1708, 2019.
[16] J. Mu, J. Liang, W. Zheng, S. Liu, and J. Wang, “Node similarity measure for complex networks,” J. Frontiers Comput. Sci. Technol., vol. 14, no. 7, pp. 1–11, Jul. 2019.
[17] X. Hu, J. Guo, S. He, and X. Chen, “Community detection algorithm of node similarity and node participation degree in signed networks,” J. Chin. Comput. Syst., vol. 38, no. 10, pp. 2275–2280, 2017.
[18] L. Fu, W. Hao, D. Li, and F. Li, “Community dividing algorithm based on similarity of common neighbor nodes,” J. Comput. Appl., vol. 39, no. 7, pp. 2024–2029, 2019.
[19] Q. Zhang, M. Li, and Y. Deng, “Measure the structure similarity of nodes in complex networks based on relative entropy,” Phys. A, Stat. Mech. Appl., vol. 491, pp. 749–763, Feb. 2018.
[20] J. Lu, Y. Xu, and L. An, “A partitioning method for community structure in weighted networks based on node similarity,” Inf. Control, vol. 41, no. 4, pp. 504–508, 2012.
[21] B. Chen, Z. Wang, and C. Luo, “Integrated evaluation approach for node importance of complex networks based on relative entropy,” J. Syst. Eng. Electron., vol. 27, no. 6, pp. 1219–1226, Dec. 2016.
[22] E. Coli Transcription Network of Uri Alon Lab. [Online]. Available: http://www.weizmann.ac.il/mcb/UriAlon/e coli-transcription-network
[23] SC-TS Biological Networks Data. [Online]. Available: http://networkrepository.com/bio-SC-TS.php
[24] D. Zhong and I. Delfet, “Performance of similarity measures based on histograms of local image feature vectors,” Pattern Recognit. Lett., vol. 28, no. 15, pp. 2003–2010, 2007.
[25] J. Fujisawa, “Weighted degrees and heavy cycles in weighted graphs,” Discrete Math., vol. 309, nos. 23–24, pp. 6483–6495, Dec. 2009.

WANCHANG JIANG was born in Shandong, China, in 1984. He received the bachelor’s degree from Liaocheng University, in 2005, and the master’s and Ph.D. degrees from Yanshan University, in 2008 and 2017, respectively. He was a Teacher with the School of Computer Science, Northeast Electric Power University, from 2008 to 2013. Since 2008, he has been a Teacher with the School of Computer Science, Northeast Electric Power University. He has published ten articles. His research interests include data mining and software security.

YINGHUI WANG was born in Shandong, China, in 1995. She received the bachelor’s degree from Jilin Jianzhu University, in 2017. She is currently pursuing the master’s degree with Northeast Electric Power University. Her research direction is complex networks.

***