Electric Power Communication Network Evaluation Based on Clustering by Fast Search and Finding of Density Peaks

Li DI¹, Zheng ZHENG¹, Min XIA², Kai HU²

¹State Grid Henan Economics Research Institute, State Grid Henan company, Zhengzhou, China
²Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing, China

Abstract — Node evaluation is an important topic in electric power communication network studies. In the view of the problem of single connection weights and single evaluation index for node importance evaluation, the bandwidth and the distance of the electric power communication network is used for the network's weight to compute three evaluating indicators: compactness, node degree and node spread. We use the "fast search and finding of density peaks" method to evaluate the importance of the nodes in power communication network to support the planning of grid communications, and to carry out unsupervised classification when the nodes are divided into four fragile levels. This can effectively improve the unsupervised classification method. Actual data from a province in China is tested to verify the usefulness of this method in electric power communication network.

Keywords - electric power communication networks; node importance; fast density clustering; unsupervised classification; node characteristic

I. INTRODUCTION

Electric power communication network is a special network for electric power industry, which is composed of power plant, transformer substation and power system at different levels. By the continuous development of power system, power communication network carrying capacity has become more and more, the growth rate is amazing. With the continuous expansion of the scale of power communication network, the security and reliability of power communication network are becoming more and more important. The important nodes in the electric power communication network largely determine the stability and security of power communication network. Therefore, how to find the key nodes in the complex power communication network is an important issue in the research of power communication network.

The important nodes in power communication network determine the stability and security of the network. The importance of nodes in power communication network shows that the node can affect the performance of the network under the condition of potential failure. The quantitative evaluation of the network nodes has become a hot spot in the research of electric power communication network. At present, there are some methods to evaluate the importance of the nodes of power communication network. (1) Using node degree as the index of node importance. (2) Evaluation based on node deletion. In this method, the node's importance is analyzed by deleting a node. (3) Evaluation based on node betweenness. The node is more important when node betweenness is more larger. (4) Evaluation based on compactness. The larger compactness, the node is the more important.

At present, the importance of the power communication network node importance assessment mainly used the several methods mentioned above, but these methods have some defects. Taking the degree of nodes as the evaluation index, there is a certain degree of one-sided. For example, the degree of the bridge node is small, but the importance is very strong. If we delete the frontier node in the form of node deletion, it cannot achieve the objective evaluation of the importance of nodes. In addition, because of the definition of node betweenness and compactness are in the global, the method based on node betweenness and compactness are not reasonable.

At present, the evaluation of the importance of power communication network nodes is based on a single connection weight, the distance and bandwidth are not both effective comprehensive evaluation index, and a lot of existing problems in the assessment of power communication network. This paper uses the network bandwidth and distance as a weighted evaluation index to evaluate the importance of nodes, and three evaluating indicators are considered in the model. In this paper, we use the method of multi index to evaluate the importance of nodes, and this classification is a kind of unsupervised classification. At present, the unsupervised classification of power communication network node importance evaluation is mainly K-means method. The K-means method is the main traditional method of unsupervised classification, but because the method is based on the distance, the result is a circular clustering shape, and cannot be a good classification of the complex features. To solve this problem,
this paper adopts a method called "fast search and find of density peaks" which was published in the 2014 issue of the journal 《Science》 [15]. This method not only uses the distance as a measure, but also uses the density as the index to solve the problem of the K-means method. On the basis of weighted network, this paper adopt the method of "fast search and find of density peaks", using the evaluating indicators compactness, precision and node betweenness for power communication network node. Tested by actual data of the provincial power grid communication, it verifies the practicability of this method in power communication network.

II. CORRELATION INDEX ANALYSIS FOR NODE IMPORTANCE

The node number is n, and the number of edges is m. The network of communication network with the power network can be described by a weighted matrix.

\[ W_G = \begin{pmatrix} W_{G1} & \cdots & W_{Gn} \\ \vdots & \ddots & \vdots \\ W_{Gnm} & \cdots & W_{Gnn} \end{pmatrix} \]  

(1)

The elements of the matrix are: If there is a connection between the node i and node j, \( W_{Gij} = B_{ij} \); If there is no connection between the node i and node j, \( W_{Gij} = \infty \). \( B_{ij} \) is the weight of the node i and j. If the distance between the nodes is used as the weight, the greater the weight is, the greater the distance between two points, and the effect between two nodes is weaker. In the present study, the weights of the distance and the bandwidth are 50%.

The importance of nodes in the network has a correlation with many indicators. The influence of nodes include direct influence and indirect influence. Directly influence reflects the ability of nodes to affect other nodes, such as the degree of nodes. The indirect influence reflects the ability of nodes to influence other nodes through the network, such as compactness and node betweenness.

Node degree: the degree \( k_i \) of node i is the number of edges directly connected with the node, which reflects the direct influence of a node to other nodes in the network. In the network with n nodes, the degree of the node will not exceed \( n-1 \). In general, the weighted intensity of node i is defined as \( S_i \):

\[ S_i = k_i / \max(k_j) \], \( j = 1,2,3,\ldots,n \)

(2)

Compactness: \( C \) is used to describe the degree of difficulty of nodes in the network to reach other nodes, and reflect the indirect influence of the nodes. Its value is defined as the reciprocal of the sum of the distance between the node and the distance to all other nodes. In the network with n nodes, the distance of the nodes to all other nodes is not less than \( n-1 \), the normalized compactness index \( C_i \) is defined as:

\[ C_i = \min(d) / \sum_{j=1}^{n} d_{ij} \]  

(3)

For weighted networks, the distance \( d_{ij} \) between two points is defined as the sum of the weights of the shortest path between two nodes.

Betweenness: The calculation of the network topological vulnerability of the network is selected by choosing the appropriate static geometric parameters of the complex network theory. In this paper, betweenness is adopted as an indicator. The number of nodes in each node is represented by a parameter \( b_i \), which reflects the importance of the nodes in the network connectivity. \( b_i \) also reflects the node topological vulnerability, \( V_i \) is the vulnerability of nodes in the entire network \( V_i = \sum_{b_{max}} \frac{b_i}{b_{max}} \). \( b_i \) represents the relative topology of the nodes, measure the extent of the potential impact of failure, can directly find the most brittle weakness.

III. FAST SEARCH AND FIND OF DENSITY PEAKS

In view of some problems existing in the non supervised clustering, this paper uses the method of "Fast search and find of density peaks" (FSFDP) to evaluate the importance of the nodes in power communication network. The core idea of the algorithm is that the cluster center is surrounded by a lower local density neighbor and has a relatively large distance from any point with higher density. Based on this idea, for any data point i, we need to calculate the two quantities: local density values \( \rho_i \) and the distance to a higher local density point \( \delta_i \). The values \( \rho_i \) and \( \delta_i \) are related to the distance between the points. \( d_{ij} \) is the distance between the point i and the point j, and the Euclidean distance is used in this paper. The local density of any data point i is defined as follows:

\[ \rho_i = \sum \chi(d_{ij} - d_c) \]  

(4)

Which \( d_c \) is the truncated distance. If \( d_{ij} - d_c < 0 \), then \( \chi(d_{ij} - d_c) = 1 \). If \( d_{ij} - d_c \geq 0 \), then \( \chi(d_{ij} - d_c) = 0 \). The local density \( \rho_i \) is essentially a number of points within the distance \( d_c \) of the data point i. Experiments show that the classification results have very high robustness, and \( d_c \) have little effect on the classification results. In this paper, the selection rule for \( d_c \) is: the average number of neighbors is about the total number of data points 1-2%. The distance \( \delta_i \) between node i and higher local density point is defined as:
\[
delta_i = \min_{j, \rho_j > \rho_i} (d_{ij}) \tag{5}
\]

In the present algorithm, a point with high \(\rho_i\) and relatively high \(\delta_i\) can be considered as a cluster center, and the remaining points are attributed to a higher density of the nearest neighbor cluster. The cluster assignment can be accomplished by one step, unlike other algorithms to optimize the objective function. The simple example in Figure 1 shows the core idea of the algorithm. Figure 1(a) is the clustering center decision, and Figure 1(b) is clustering results.

Figure 1.(a) Cluster center decision diagram

![Cluster center decision diagram](image)

Figure 1.(b) Clustering results based on Figure 1(a)

![Clustering results](image)

IV. NODE IMPORTANCE CLASSIFICATION RESULTS

In this paper, the actual data of a province in China is tested to verify the usefulness of this method in electric power communication network. In this paper, the evaluation index of node degree, node compactness and betweenness are used as the evaluation index of node importance. The degree of the node is very intuitive, which reflects the node's ability to affect other nodes. In general, the greater the degree reflects more importance, but for some of the bridge node is not correct. Indirect influence on the ability of nodes to exert influence over other nodes through the network, such as compactness and betweenness.

Figure 2 is a compactness analysis of the network nodes. Figure 2 shows that the degree of the node with large compactness is not always large, but the role of the node in the data transmission is very large. Figure 3 is a network node's vulnerability analysis, node's vulnerability represents the node in the network communication ability. If connectivity is strong, then the impact on the network after damage to the network, which is considered a strong vulnerability. As can be seen from Figure 3, the vulnerability of the node is not related to the node's degree distribution and the compactness distribution.

Figure 4 shows the distribution of various factors in this paper. Table 1 gives the correlation between these three factors. From Figure 4 and table 1, we can see that the relationship between these three factors is not entirely positive. Therefore, in the measurement of the importance of nodes, the need to consider these three factors.
Table 1. Correlation between the factors

<table>
<thead>
<tr>
<th>Node degree</th>
<th>Node compactness</th>
<th>Node betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node degree</td>
<td>1</td>
<td>0.57</td>
</tr>
<tr>
<td>Node compactness</td>
<td>0.57</td>
<td>1</td>
</tr>
<tr>
<td>Node betweenness</td>
<td>0.71</td>
<td>0.58</td>
</tr>
</tbody>
</table>

This paper uses FSFDP to evaluate the importance of the nodes in power communication network. The elements in this paper are 3, then the classification space is three-dimensional. Figure 5 gives a cluster center selection graph based on three factors. The four blocks of the graph are the centers of clustering, and all nodes are clustered according to the 4 clustering centers. Clustering results are shown in Figure 6. In Figure 6, the three dimensions represent the three clustering factors, which are node degree, node compactness, and node betweenness. From Figure 5, we can see that the method of this paper can be very good for classification of the importance of nodes. In Figure 6, the nodes of a square are very important. The points "**" represent the less importance than square nodes. The triangle represents the general importance of the nodes. The circular is with weak important. In practical application, the square of the node should be paid enough attention.

Figure 7 shows the importance of nodes in power communication network of a backbone network in a province of China. It can be seen from the graph that we can use the method to classify the important nodes. Owing to the distance and bandwidth as a part of the weight, the general importance nodes' weight is larger than that of the weak important nodes. This is the reason that there is the difference between the general importance nodes and the weak important nodes in Figure 7.
communication network, supporting for the planning of power grid communication. In this paper, FSFDP method is carried out for unsupervised classification. The nodes are divided into 4 levels of importance. The actual power communication network data of a province in China is used for validating the method. The practicability of this method in electric power communication network is verified, and the method can be effectively applied to the evaluation of real power communication network.

ACKNOWLEDGMENTS

This work is supported in part by, the National Natural Science Foundation of PR China (61105115), Six Talent Peaks Program of Jiangsu Province (2014-XXRJ-007), Natural Science Foundation of Jiangsu Province (BK20131002).

REFERENCES


