Acta Technica 62 (2017), No. 4A, 647–656

Identification of vital nodes for the complex network containing various nodes

Wu Shunyu[1], Xu Gang[1]

Abstract. Traditional complex networks theories usually adopt the method of link interruption or cascading failure to study the influence of node failure on the integrity and connectivity of complex network. However, they tend to neglect the fact that industrial complex networks are usually composed of different types of nodes, and different types of nodes have a dependence relationship that steps over the nodes on the middle path. Taking the complex networks with multi types of nodes as research object, the mechanism of network degeneration is studied through analysing the dependence relationships among nodes to establish a dependency efficiency matrix. According to the dependent direction from node to node, the cross-dependency intensity index was proposed. Considering the cross-dependent relationship between the nodes on reachable path, an improved PageRank algorithm was adopted to spread cross-dependency intensity of nodes, and finally realizing the vital node identification. Through the simulation of a typical information system APRA, it is shown that the cross-dependency intensity index proposed can effectively identify the vital nodes which have a significant influence on the robustness and integrity of networks.

Key words. Vital nodes, heterogeneous-interdependent network, dependency efficiency, cross-dependency intensity.

1. Introduction

With the development of technologies in the industrial field, people's lives have been extremely dependent on complex networks, such as, communication network and power grid [1-3]. Complex network structure feature analysis and vital node identification have been two research focuses for researchers. By protecting the vital nodes in complex networks, the integrity, reliability and transmission efficiency of networks can be effectively guaranteed [4-6]. Complex networks in the industrial field are usually composed of various types of nodes and edges with different functional characteristics [7], and they can be regarded as a kind of heterogeneous complex networks. At the same time, most of the industrial complex heterogeneous

[1]Workshop 1 - School of Electrical and Electronic Engineering, North China Electric Power University, Changping District, Beijing, China; e-mail: *wsy817@126.com

networks use fixed paths to transfer directional flow (such as, information and energy) among nodes. Thus, the effectiveness of downstream nodes always depends on the upstream nodes. For example: (1) In a power system, load nodes depend on generator nodes, while the remote load nodes also depend on the load nodes which are closer to generator [8]; (2) In social network, the information acquisition of Internet users depend on certain information disseminators, while the information dissemination users may depend on specific information sources. The complex networks, with the availability of their heterogeneous nodes and nodes dependent on other nodes, are called as Heterogeneous-Interdependent network (HI Net). Due to the interdependency among nodes, the protection for vital nodes of HI Net can effectively decrease the extent of degeneration caused by node failure or attack.

The existing evaluation methods for the relationship between complex networks and nodes failure can be divided into two kinds: link interruption and cascading failure. When using link interruption to evaluate the nodes in complex networks, it usually assumes that the failed node affects the adjacent connected branches only, thus analyzing the impacts of node's failure on complex networks by assessing network structure, shortest path between nodes and node betweenness. This method assumes that complex networks are static. However, due to the significant dynamic characteristics of real networks, network structure changes will affect the existence of non-faulty nodes, resulting in further node chain failure. Cascading failure takes the dynamic characteristics of networks into account. Complex network cascading failure assessment methods can be applied to the networks which have both relatively fixed paths and amount of nodes, such as, logistics network, traffic network and wireless network. However, because the paths of most industrial networks are fixed, cascading failure assessment method cannot calculate the importance of the nodes in an industrial network effectively. Sergey proposed the interdependent networks theory in Nature for the first time, describing the influence of the coupling dependencies between two different networks . He took power system and its information system as an example to demonstrate the chain nodes failure process between power system and its information system.

Actually, nodes in one single network also have interdependence characteristics. In order to identify the degeneration of HI net caused by node failure, we proposed a node importance evaluation method according to the degree of dependence among nodes and the flow path of networks. Considering node failure spread speed, a dependency efficiency matrix is established. The dependence relations between the nodes which have available path between each other are calculated, so as to get the index of cross-dependency intensity for each node. Based on the node dependency relation, the improved PageRank algorithm was used to spread the cross-dependency intensity, which reflects the dependency relations between the nodes on the effective path.

2. CHARACTERISTIC OF HI Network

Heterogeneous network, which refines the relationships among nodes or edges, can be regarded as a kind of complex network with different attributes of objects.

In mathematics, heterogeneous network can be described as a special directed graph.

Definition 1 Heterogeneous Network

Given a directed graph $G =< \mathrm{V}, \mathrm{E}, \mathrm{W} >$, with nodes set $\mathrm{V} = \{v_1, v_2, \cdots, v_N\}$, edges set $\mathrm{E} = \{e_1, e_2, \cdots, e_M\}$ and $\mathrm{W} = \{w_{e1}, w_{e2}, \cdots, w_{e_M}\}$ is an edge weight collection. There is an object mapping function ϕ:V $\rightarrow$ A, where $\phi(v) \in \mathrm{A}(v \in V)$, and an edge mapping function ψ:E $\rightarrow$ R, where $\psi(e) \in \mathrm{R}(e \in \mathrm{E})$. If the number of node types $|\mathrm{A}| > 0$ or the number of edge types $|\mathrm{R}| > 0$, then G is called a heterogeneous network.

Different from traditional complex networks, heterogeneous network should have a clear distinction on the type of internal nodes or edges. The standard for distinguishing the distinction of nodes or edges can be the access degree of nodes or the physical significance of network nodes corresponding to the physical system or information system.

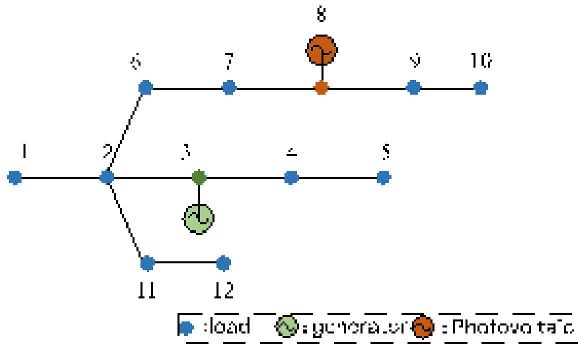

Fig. 1. Examples of node classification methods for heterogeneous network

(a) Heterogeneous node classification based on node access degrees, **(b)**Heterogeneous node classification for power grid.

As shown in *Fig.1*, the nodes in heterogeneous networks can be divided into source node, relay node and end node according to their access degrees, with the source node not dependent on any other nodes. Meanwhile, the nodes in the power grid with multi distribution energy can be divided into load node, DG node and PV node. Generally, the dependency direction is usually opposite to the flow direction. Such as Fig1(b), because of the volatility and randomness of PV node's power output, PV node has no ability to supply power to load only by itself, so the node 8 (PV) is also directly dependent on the node 3 (DG).

Definition 2 Heterogeneous-Interdependent Network (HI Net)

Given a heterogeneous network as directed graph $G =< \mathrm{V}, \mathrm{E}, \mathrm{W} >$.The set of elements T_a and $\mathrm{T}_b$is a subset of the node set V or edge set E, and $\mathrm{T}_a \bigcap \mathrm{T}_b = \emptyset$, $|\mathrm{T}_a| > 0$, $|\mathrm{T}_b| > 0$. If node set T_a's failure causes T_b's failure,denoted by $lost\mathrm{T}_a \Rightarrow lost\mathrm{T}_b$, then Gis considered a heterogeneous-interdependent network. When $|\mathrm{T}_a| = 1$, T_b is completely dependent on T_a. When $|\mathrm{T}_a| > 1$, T_b is partly dependent on T_a.

Heterogeneous-interdependent network describes the existential association between homogeneous elements and heterogeneous elements in a complex network with multiple types of elements (nodes or edges).

A simple directed graph structure is shown in *Fig.2*,where the arrow direction shows the flow in the network, while the edge weight means the flow transfer speed. As *Fig. 2* shown, node 2 and 3 are completely dependent on the source node 1, nodes 4 and 5 are completely dependent on nodes 2 and 3, and node 6 is completely dependent on node 5.When node 2 and node3 in Fig.2 fail, node3, which transfer flow to node5 and node 6, will lead to a faster and wider failure propagation than node 2. Thus, compared with node 2, node 3 is more important to the network. In addition, when node 1 fails, the network failure will propagate through nodes 2, 3 to nodes 4, 5, 6, and eventually lead to a failure and collapse of the nodes in the entire network, the relationship between nodes 4,5,6 and node 1 is indirect dependency.

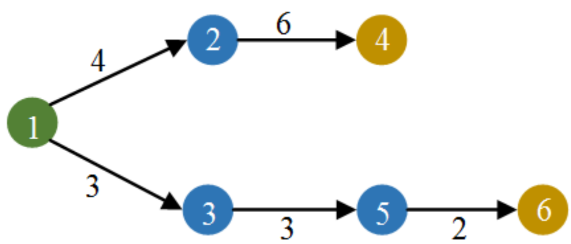

Fig. 2. A simple HI network

Definition 3 Indirect Dependency

Assume that the directed graph $G = <\mathrm{V}, \mathrm{E}, \mathrm{W}>$ is a Heterogeneous-Interdependent network,there are sets of elements T_a, T_b and T_c, and the intersection between the two sets of is $\emptyset$. The path sets between T_a and T_b, T_b and T_c, denoted by E_{ab}&E_{bc}, are non-empty sets. The path set between T_a and T_c is $\mathrm{E}_{ac} = \mathrm{E}_{ab} \bigcup \mathrm{E}_{bc}$. If there is

$$\begin{cases} lost\mathrm{T}_a \Rightarrow lost\mathrm{T}_b \\ lost\mathrm{T}_b \Rightarrow lost\mathrm{T}_c \end{cases} \tag{1}$$

Then it can be said thatT_c has an indirect dependency on T_a.Additional, for a HI Net with multi-source nodes, when a source node fails, the nodes that have connectivity paths with other source nodes will not fail.

3. VITAL NODE RECOGNITION METHOD

3.1. Node cross-dependency intensity

The importanceof nodes in HI Net can be transformed into the influence of node failure on the transmission capacity of the network, the speed of fault propagation and the survivability of other effective nodes. Construct a network dependency efficiency matrix according to network path, heterogeneous node type and inter-node dependency relationships:

$$S = \begin{bmatrix} 0 & s_{12} & \cdots & s_{1N} \\ s_{21} & 0 & \cdots & s_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N1} & s_{N2} & \cdots & 0 \end{bmatrix} \tag{2}$$

Where, the dependency efficiency s_{ij} between nodes is the reciprocal of the shortest path d_{ij},$s_{ij} = 1/d_{ij}$.

According to the definition of heterogeneous network, Sis an asymmetric matrix. The efficiency index is characterized by the direction and velocity of the influence spread among nodes, and also reflects the time of node failure leading tonetwork failure. When $i = j$ or there is no available path from v_i to v_j, $s_{ij} = 0$.

The HI network performance depends on the flow reception ability andthe spread ability of nodes. The higher input of the node is, the more flow it receives from other nodes, the less dependency it has on particular one, and vice versa. Thus, the dependent rate p_i and be-dependent rate q_i of node v_i are respectively

$$\begin{cases} p_i = 1/c_i^{\text{out}} \\ q_i = 1/c_i^{\text{in}} \end{cases} \tag{3}$$

Where, c_i^{in} is the input degree of the node, and c_i^{out} is the output degree of the node. The dependency rate expresses the dependency ratio of node i to its upstream node input. The be-dependency rate indicates the dependent degree of v_i's downstream node to its output. The input and output extent of $v_i$is further stipulatedto realize the correlation between transmission efficiency and dependency ratio among the nodes.

$$P_i = \sum_{j \in \Omega_i^{\text{out}}} s_{ij} q_j^{\text{in}} \tag{4}$$

$$Q_i = \sum_{j \in \Omega_i^{\text{in}}} s_{ji} q_i^{\text{in}} \tag{5}$$

Where, P_i and Q_i are the be-dependent intensity and dependent intensity. s_{ij} is the dependency efficiency from v_i to v_j.s_{ji} is the dependency efficiency from v_j to v_i.Ω_i^{out} is a set of nodes that depend on v_i and Ω_i^{in} is a set of nodes on which$v_i$is depended. Considering the dependency intensity and be-dependent intensity of nodes, the cross dependent intensity H_i is used to evaluate the importance of the nodes in HI Net.

$$H_i = h(P, Q) = \lambda P_i + (1 - \lambda) Q_i \tag{6}$$

Where, λ is the dependent intensity indicator constant,$\lambda \in (0, 1)$. If $\lambda > 0.5$, it represents that the value of node importance mainly depends on the degree of dependence of other nodes on it, while if $\lambda < 0.5$, it represents that the value of node importance mainly depends on the degree of dependence of the node on others.

3.2. CROSS-PROPAGATION STRATEGY

Considering the dependency relationships among heterogeneous nodes in HI Net and their dependency relationships, PageRank algorithm is adopted to realize a secondary node influence spread on dependency path. The average distribution method shown in (7) is used to calculate the influence degree of the end node of the dependent path.

$$PR(x) = \frac{(1-\sigma)}{N} + \sigma \sum_{i=1}^{N} \frac{PR(Y_i)}{c_{Y_i}^{\text{out}}} \tag{7}$$

Where, $PR(x)$ is the PageRank value of web page x, $PR(Y_i)$ is the PageRank value of the page linked to x. N is the total number of nodes, that is, the number of nodes; andσ is the damping coefficient. $c_{Y_i}^{\text{out}}$ is the value of the ith link to page x. While, in the process of heterogeneous network degeneration, the dynamic propagation speed of failure should also be considered after network node failure. Thus, (7) is rewritten to add the network dependency efficiency, the cross-dependency intensity CR_i of node v_i after the influence transmission

$$\text{CR}_i = \frac{(1-\sigma)}{N} + \sigma \sum_{j \in \Omega_i^{\text{in}}} \frac{\text{s}_{ji}}{\text{s}_i^{\text{in}}} \text{H}_j \tag{8}$$

Where, s_i^{in} is the efficiency of node v_i dependent on other nodes, i.e. the sum of the ith row of the network dependency efficiency matrix,$\text{s}_i^{\text{in}} = \sum_{j=1}^{N} \text{s}_{ji}$. In the form of pseudocode, we present the calculation process of HI Net nodes' influence propagation and cross-dependency intensity

Table 2 shows the calculation process of the HI Net node intensity. In the process of calculation, the dependency efficiency matrix S is first established through heterogeneous node classification and dependency relationships. This paper uses the reverse direction of the medium among nodes as the dependency relationships. And then according to the network path, node input and output degree and communication efficiency are used to calculate node cross-dependency intensity. Considering the dynamic process of network decline and the cross-node dependency relationships between heterogeneous nodes, the improved cross-dependency intensity of nodes is used as the initial value, the improved PageRank algorithm is used to propagate the influence between nodes, and the final node cross-dependency level is obtained. During the cycle of improved PageRank algorithm, *sig* is the cyclic stop condition, which can be set according to the change of the cross intensity ranking after the spread of the nodes' influence. In this paper,$sig = 0.01$.

Table 1. Spread strategy ofcross-dependency intensity

Input:	$G = < \mathrm{V}, \mathrm{E}, \mathrm{W} >$
Output:	CD
1	begin
2	node classifications$\leftarrow \phi(v_i)$ //Node classification
3	establish matrix S //Establish a dependency matrix
4	$S \leftarrow sigmoid(S)$ //sigmoidFunction normalization
5	**for**$i$$\leftarrow$1 to N **do** //N=Number of nodes
6	compute P_i and Q_i
7	$H_i \leftarrow h(P_i, Q_i)$
8	**end for**
9	$H \leftarrow [H_1, H_2, \cdots, H_N]$,
10	$CR^0 \leftarrow H$ //Node intensity initialization
11	$\sigma \leftarrow 0.85$ //Damping coefficient
12	**while**$sig < (CD^r - CD^{r-1})$**do**
13	$CD^r \leftarrow Pagerank(CD^{r-1}, S)$
14	**end while**
15	$CD \leftarrow CD^r$

4. Calculations and result analysis

The influence of heterogeneous elements inHI Net is mainly reflected in the network structure integrity, connectivity, robustness and so on. Therefore, the ratio of the remaining nodesLN, the maximum connectivity coefficientMC[20] and the robustness of network$NB$are used to evaluate the impact offaultednodes.

$$LN = \frac{N_{\text{left}}}{N} \tag{9}$$

Where, G is the remaining nodes coefficient, N_{left} is the number of remaining nodes,$N_{\text{left}} = N - N_{\text{lost}}$, and$N_{\text{lost}}$ is the number of the failed nodes.

$$MC = \frac{R}{\frac{1}{2}N(N-1)} \tag{10}$$

Where, R is the remaining nodes connected logarithm after the network recession caused by nodes' failure. If the number of connected subsets in the network becomes smaller as the important nodes fail, the greater impact the failed nodes will have on the HI Net's connectivity.

$$NB = \bar{d}/\frac{N_{lost}}{N} \tag{11}$$

Where, $\bar{d}$ is theaverage path length.

In order to verify the correctness of node dependency index, cross dependency, calculated by the method proposed in this paper, the directional weighted ARPA (Advanced Research Project Agency) network as shown in Fig.3 was used. ARPA network is a kind of early server link structure of the US.

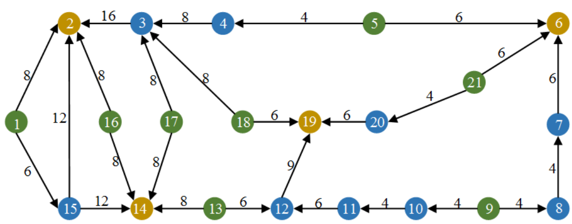

Fig. 3. Structure of ARPA network

The side direction of APRA shows the information transmission direction of the adjacent server. Lowinput degree means that the node server's survival intensity is weak and it's easy to fail due to other nodes or sides' failure. APRA network is a typical HI Net, composed of information source nodes, information relay nodes and information receiving nodes. The normalized calculating results of nodes' crossed dependency (CD) intensity are given in Table.2.

Table 1. APRA status after node fails

No.	NodeID	CD(10^{-3})	LN	MC	NB
1	9	0.9986	0.7619	0.7408	0.3363
2	5	0.6483	0.9048	0.7787	0.6081
3	1	0.5508	0.9048	0.7808	0.6084
4	10	0.5405	0.9048	0.8153	0.6271
5	21	0.4616	0.9048	0.7987	0.6271
6	8	0.347	0.9048	0.8153	0.6271
7	13	0.2569	0.9524	0.8153	0.8682
8	17	0.2273	0.9524	0.8153	0.8682
9	18	0.2221	0.9524	0.8153	0.8682
10	11	0.2115	0.9524	0.8307	0.8812
11	4	0.1813	0.9524	0.8307	0.8812
12	16	0.1529	0.9524	0.8153	0.8812
13	15	0.1529	0.9524	0.8153	0.8812
14	7	0.1248	0.9524	0.8307	0.8812
15	20	0.1111	0.9524	0.8307	0.8812
16	12	0.1111	0.9524	0.8307	0.8812
17	3	0.1057	0.9524	0.8307	0.8812
18	19	0.0689	0.9524	0.8449	0.8930
19	14	0.0689	0.9524	0.8449	0.8930
20	6	0.0689	0.9524	0.8449	0.8930
21	2	0.0689	0.9524	0.8449	0.8930

In Table.2,LN, MC and NB are the network performance indicators of APRA network's decline, which is caused by corresponding node failures. As shown in Tab.2, it can beknown that CD index has a negative correlation with the other three network indicators, indicating that the higher nodes' CD value, the greater the impact of node failure on network-related performance. In particular, for node 10 and node 21, the network's degeneration extent and time-consuming are consistent when one of them fails. However, although $CD_{10} > CD_{21}$, $MC_{10} > MC_{21}$ appears.This is because the MC index is more focused on the number of paths among nodes, while the HI Net is more concerned about the impact of node failure on the survival intensity of other nodes.

5. Conclusion

Aiming at the characteristics of node diversification, relatively fixed transmission path and structure in real industrial networks, the HI Net theory is proposed by analysing the relationship between node types and dependencies in existing networks. The heterogeneous nodes in the network are abstracted into source nodes, relay nodes, and end nodes. In order to solve the problem that the importance evaluation method of existing complex network nodes fails to consider the relationship between inter-node dependency and interlocking faults, the characteristics of HI Net's degeneration are analysed, and the node cross-dependency intensity index is proposed to evaluate the effect of node failure on network structure, and the influence of the survival intensity of other nodes on the propagation path.

References

[1] J. S. Tomar, D. C. Gupta, N. C. Jain: *Exploring complex networks.* nature *410* (2001), No. 6825, 268.

[2] J. S. Tomar, A. K. Gupta: *Power grid vulnerability: A complex network approach.* J Sound and Vibration *98* (1985), No. 2, 257–262.

[3] R. H. Gutierrez, P. A. A. Laura: *Structure and tie strengths in mobile communication networks.* Applied Acoustics *18* (1985), No. 3, 171–180.

[4] R. P. Singh, S. K. Jain: *Structure and tie strengths in mobile communication networks.* Tamkang Journal of Science and Engineering *7* (2004), No. 1, 41–52.

[5] M. N. Gaikwad, K. C. Deshmukh: *Thermal deflection of an inverse thermoelastic problem in a thin isotropic circular plate.* Applied Mathematical Modelling *29* (2005), No. 9, 797–804.

[6] S. Chakraverty, R. Jindal, V. K. Agarwal: *Security Assessment of Electricity Distribution Networks Under DER Node Compromises.* Indian Journal of Engineering and Materials Sciences *12* (2005) 521–528.

[7] N. L. Khobragade, K. C. Deshmukh: *Thermal deformation in a thin circular plate due to a partially distributed heat supply.* IEEE Transactions on Control of Network Systems *30* (2005), No. 4, 555–563.

[8] Y. F. Zhou, Z. M. Wang: *Vibrations of axially moving viscoelastic plate with parabolically varying thickness.* J Sound and Vibration *316* (2008), Nos. 1–5, 198–210.

Received November 16, 2016