IEICE TRANS. INF. & SYST., VOL.E101–D, NO.11 NOVEMBER 2018

PAPER

Critical Nodes Identification of Power Grids Based on Network Efficiency

WenJie KANG[†], **PeiDong ZHU**[††a)], **JieXin ZHANG**[†], *Nonmembers*, *and* **JunYang ZHANG**[†], *Student Member*

SUMMARY Critical nodes identification is of great significance in protecting power grids. Network efficiency can be used as an evaluation index to identify the critical nodes and is an indicator to quantify how efficiently a network exchanges information and transmits energy. Since power grid is a heterogeneous network and can be decomposed into small functionally-independent grids, the concept of the Giant Component does not apply to power grids. In this paper, we first model the power grid as the directed graph and define the Giant Efficiency sub-Graph (GEsG). The GEsG is the functionally-independent unit of the network where electric energy can be transmitted from a generation node (i.e., power plants) to some demand nodes (i.e., transmission stations and distribution stations) via the shortest path. Secondly, we propose an algorithm to evaluate the importance of nodes by calculating their critical degree, results of which can be used to identify critical nodes in heterogeneous networks. Thirdly, we define node efficiency loss to verify the accuracy of critical nodes identification (CNI) algorithm and compare the results that GEsG and Giant Component are separately used as assessment criteria for computing the node efficiency loss. Experiments prove the accuracy and efficiency of our CNI algorithm and show that the GEsG can better reflect heterogeneous characteristics and power transmission of power grids than the Giant Component. Our investigation leads to a counterintuitive finding that the most important critical nodes may not be the generation nodes but some demand nodes.

key words: *network efficiency, giant efficiency sub-graph, the algorithm of critical nodes identification, critical degree, node efficiency loss*

1. Introduction

Over the last decades, many power grids blackouts caused by natural disasters and human factors have occurred, which has a serious impact on network security, economic gain and social stability [1]. For instance, the Western North American blackouts occurred in July and August 1996 [2], the largest blackout in US history took place on 14 August 2003 [3] and the cyber-attack on Ukraine regional grid on 23 December 2015 [4]. It is investigated that large-scale blackouts result from failure of some critical nodes. In practice, the critical nodes are very important for critical infrastructure's security. The failure of a few critical nodes may directly lead to the failure of the entire power grids, which is the major reason causing the blackouts. Therefore, the critical nodes identification is of great practical significance in the field of critical infrastructure protection and has been a hot issue.

Buldyrev et al. [5] laid out the framework for the analysis of catastrophic failures in interdependent networks [6], which breaks through the frontier of complex networks theory that still focuses on a single, non-interacting network [7]–[10].When a single network is cut into multiple components due to a few node failures, the largest component is defined as the giant component in which at least one path can be found to join any two nodes by passing through a certain number of edges [11]–[14]. In previous works, critical size of the giant component was used to represent the functional integrity of the segmented network. This means that those remaining small components are considered invalid and will be removed. Therefore, the concept of the giant component does not apply to power grids, which are composed of different types of nodes (generation nodes, transmission nodes, and distribution nodes) and edges (high and low voltage). The key to maintaining the effectiveness of the network is to transmit power to the distribution area through multiple substations and transmission lines. For instance, when a power grid breaks into several clusters, the smaller clusters are still functional if only they contain the links between generation nodes and demand nodes (transmission nodes and distribution nodes). Since a power grid is composed of small functionally independent grids, we can describe those clusters as those small grids. Therefore, Giant Efficiency sub-Graph (GEsG) is defined to describe and denote small functionally-independent grids, and we can identify critical nodes based on GEsG.

Recently, in order to control the complexity of limited resources, Chen et al. [15] designed three metrics that are used by six algorithms to identify critical nodes for protection or removal. A new method was proposed to identify critical nodes of power systems based on the controllability theories of complex networks [16]. The concept of critical nodes also was regarded as middlemen and was extended to directed networks [17]. Modeling the network as a weighted connected graph, an applicable method of identifying critical nodes and edges was proposed to find a subset in which nodes and edges are removed to cause the largest cost [18]. Wehmuth et al. [19] proposed a methodology to locate the most critical nodes in terms of network robustness in a fully distributed way. Sivakumar et al. [20] suggested multiple methods to find the critical nodes of a network based on residual battery power, reliability, bandwidth, availability and service traffic type. Nagurney et al. [21] assessed critical nodes and links in a financial network by measuring

Manuscript received January 30, 2018.
Manuscript revised May 29, 2018.
Manuscript publicized July 27, 2018.

[†]The authors are with the College of Computer, National University of Defense Technology, Changsha, 410073, China,
[††]The author is with the Department of Electronic Information and Electrical Engineering, Changsha University, Changsha, P.R. China.
a) E-mail: zhupeidong66@126.com (Corresponding author)

DOI: 10.1587/transinf.2018EDP7042

the network performance. The multi-vector viruses were modeled in multi-layered networks to identify critical nodes that allow a better spreading efficiency of these kinds of viruses [22]. The European Union Agency for Network and Information Security (ENISA) developed advice and recommendations on good practice in information security and proposed a methodology for the identification of critical communication networks links and components [23]. Wang et al. [24] proposed electrical centrality metrics to identify the critical nodes in power systems. Multiple vulnerability measures to cascading failure ware proposed to identify the most critical components and evaluate the damage of power grid by removing these identified components [25].

The rest of this paper is organized as follows. Section 2 defines the Giant Efficiency sub-Graph and introduces an efficiency assessment model of power grids. The algorithm of critical nodes identification (CNI) is proposed in Sect. 3. Node efficiency loss is defined to verify the effectiveness and feasibility of CNI algorithm in Sect. 4. Experiments are presented based on the Hainan regional power grids in Sect. 5. Section 6 draws relevant conclusions and presents future work.

2. The Efficiency Assessment Model of Power Grids

The concept of network efficiency was used to measure how efficiently network exchanges information and transmits energy [26]. We represent a real network as a generic weighted graph G that is with N nodes and K edges. The matrix $\{d_{ij}\}$ is used to calculate the shortest path length between node i and j. The efficiency of G depends on two factors: the shortest path length and the maximum possible number of edges. The efficiency ε_{ij} between nodes i and j can be defined to be inversely proportional to the shortest distance: $\varepsilon_{ij} = 1/d_{ij}, \forall i, j$. When there is no path in the graph between vertex i and j, $d_{ij} = \infty$ and consistently $\varepsilon_{ij} = 0$. The average efficiency of G can be defined as:

$$E(G) = \frac{\sum_{i \neq j \in G} \varepsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}} \tag{1}$$

The formula of network efficiency gives a clear physical meaning to the concept of several systems, like neural networks, communication networks, transport networks and power networks, which means how efficiently information is exchanged and energy is transmitted over the network.

The power grids are heterogeneous with different types of nodes and edges. The nodes can be divided into two different groups: the generation nodes, which represent the power plants that are responsible for producing electric power for demand nodes, and the demand nodes, which represent the substations that contain transmission stations and distribution stations. In fact, the power grid forms a directed graph, where electric power is transmitted from power plants to distribution domain via transmission nodes, distribution nodes, and transmission lines. The normal functioning of power grid depends on the generation nodes and the feasible paths that transmit electric power from the generation nodes to the demand nodes.

Traditional approaches consider power grids as homogenous networks, which do not correspond to reality. When some nodes are removed by either intentional attack or natural disaster, the giant component, which is a criterion for assessing the efficiency of networks, is not suitable for power grids. When all generation nodes are successfully attacked, power grid actually fails due to the lack of power supply, but the giant component still exists and is valid. Therefore, we introduce the concept of Giant Efficiency sub-Graph to measure the efficiency of power grids. The GEsG is generated by the shortest path connecting algorithm.

A power grid is considered as a directed graph $G = \{V, E\}$, where V denotes the set of nodes with n elements and E denotes the set of edges with m elements. $V = \{V^G, V^D\}$ can be divided into two sets: the generation node set $V^G = \{v_1^G, v_2^G, \cdots, v_M^G\}$ with M elements and the demand node set $V^D = \{v_1^D, v_2^D, \cdots, v_N^D\}$ with N elements, here $n = M + N$. The superscript G and D represent the generation nodes and demand nodes, respectively. The matrix $\{d_{ij}\}$ denotes the shortest path length from the generation nodes to the demand nodes or between two demand nodes. Because the power grid has several generation nodes and each generation node is responsible for providing some or all of power supply to some demand nodes, the graph G can be divided into several directed sub-graphs G' according to the shortest path of power supply from a generation node. So we define G' as Giant Efficiency sub-Graph for showing the structure of the network clearly and calculating the network efficiency efficiently.

Definition 1 (Giant Efficiency sub-Graph (GEsG)) A given graph $G = \{GEsG_1, GEsG_2, \cdots, GEsG_M\}$ can be divided into M $GEsGs$, where $GEsG_i$ represents a Giant Efficiency sub-Graph of G. In the $GEsG_i = \{V_i, E_i\}$, the node set $V_i = \{v_i^G, v_{i_1}^D, \cdots, v_{i_m}^D\}$ has a generation node v_i^G and m demand nodes v_j^D, and the edge set $E_i = \{e_{i,i_1}, \cdots, e_{i_{m-1},i_m}\}$ contains all reachable shortest paths from a generation node to demand nodes. In particular, $V_i \subseteq V$ and $E_i \subseteq E$. Since each $GEsG$ includes a generation node and some demand nodes, the number of $GEsGs$ is equal to the number of the generation nodes.

The concept of the shortest path connecting algorithm is to divide network into several sub-graphs, which node pairs are connected by the shortest path between them. The algorithm is described as follows: Step one is to calculate the shortest path length d_{ij} of all of the nodes by out-degree. Step two is to identify all generation nodes and demand nodes and to add them to the generation node set V^G and the demand node set V^D, respectively. If the power grid does not provide any information about node types, we have the assumption that nodes with zero in-degree and non-zero out-degree are regarded as the generation nodes, and the other nodes are regarded as the demand nodes. Step three is to traverse all nodes i' in set V^G and to check all nodes j in

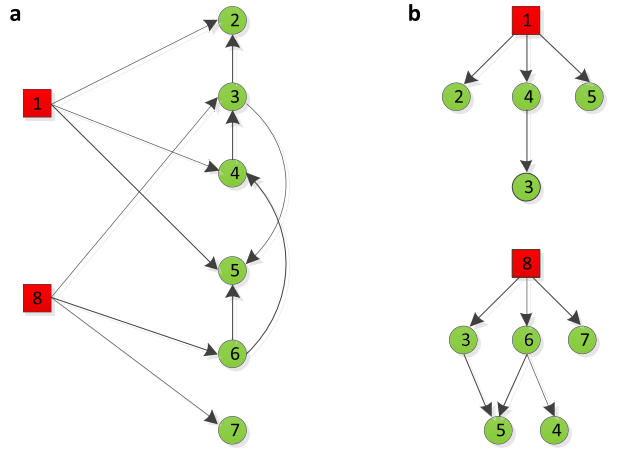

Fig. 1 Modeling a spanning process of giant efficiency sub-graphs by the shortest path connecting algorithm.

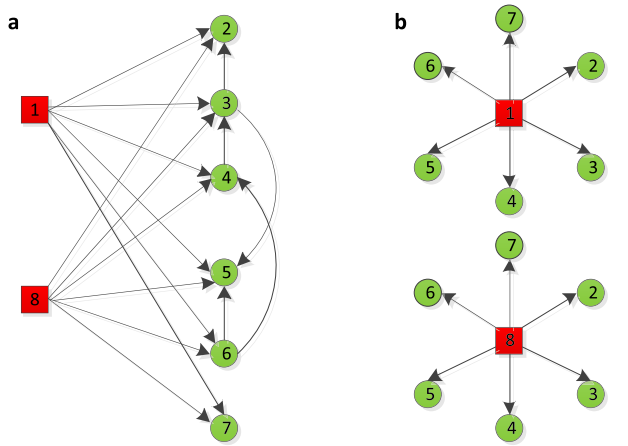

Fig. 2 A power grid and its giant efficiency sub-graph with equivalent network efficiency

set V^D. If $d_{i'j} = 1, j = 1, 2, \cdots, k$, edge $e_{i'j}$ is added to $E_{i'}$ and nodes j are added to $V_{i'}$ until all nodes j are traversed. If $d_{i'j'} = 2$ and $d_{jj'} = 1$, edge $e_{jj'}$ is added to $E_{i'}$ and nodes j' are added to $V_{i'}$. Those steps are iterated until $d_{i'j^*}$ reaches the maximum value. Here, G is decomposed into $GEsG_{i'} = \{V_{i'}, E_{i'}\}, i = 1, 2, \cdots, M$.

Figure 1 shows a spanning process of Giant Efficiency sub-Graphs by the shortest path connecting algorithm. Figure 1 (a) presents a power grids with the generation nodes set {1, 8} and the demand nodes set {2, 3, 4, 5, 6, 7}. Figure 1 (b) presents two Giant Efficiency sub-Graphs that $GEsG_1$ and $GEsG_8$ are generated by the shortest path from generation nodes 1 and 8, respectively. The steps of shortest path connecting algorithm are as follows: Step one is to calculate the shortest path length d_{ij}, in-degree and out-degree. Step two is to identify the generation nodes 1 and 8 with in-degree 0 and out-degree 3. Nodes 1 and 8 are added to $V^G = \{1, 8\}$ and other nodes are added to $V^D = \{2, 3, 4, 5, 6, 7\}$. Step three is to traverse nodes 1 and 8, and to search the shortest path length $d_{12} = d_{14} = d_{15} = 1$, and e_{12}, e_{14}, e_{15} are added to E_1 and nodes 1, 2, 4 and 5 are added to $V_1 = \{1, 2, 4, 5\}$. Similarly, $E_8 = \{e_{83}, e_{86}, e_{87}\}$ and $V_8 = \{8, 3, 6, 7\}$. The node will be searched in order to satisfy the condition that $d_{1j'} = 2$ and $d_{ij'} = 1, i = 2, 4, 5$. Node 3 meets this condition, therefore, it is added to $V_1 = \{1, 2, 4, 5, 3\}$ and e_{43} is added to $E_1 = \{e_{12}, e_{14}, e_{15}, e_{43}\}$, similarly, $E_8 = \{e_{83}, e_{86}, e_{87}, e_{35}, e_{65}, e_{64}\}$ and $V_8 = \{8, 3, 6, 7, 5, 4\}$. In addition, this method divides the graph $G = \{V, E\}$ of the power grid into two Giant Efficiency sub-Graphs $GEsG_1 = \{V_1, E_1\}$ and $GEsG_8 = \{V_8, E_8\}$ by two iterations.

The average efficiency of the network [27] was used to measure the network's performance in [28]–[30]. Similarly, we use the sum of average efficiency of GEsG to measure network efficiency of power grid.

$$E(G) = \frac{1}{N_{V^G} N_{V^D}} \sum_{i \in V^G} \varepsilon(GEsG_i) = \frac{1}{N_{V^G} N_{V^D}} \sum_{i \in V^G} \sum_{j \in V^D} \frac{1}{d_{ij}} \quad (2)$$

Where $GEsG_i$ is Giant Efficiency sub-Graph i that contains one generation node v_i^G and multiple demand nodes v_j^D. N_{V^G} is the number of the generation nodes (e.g., power plant). N_{V^D} is the number of the demand nodes that represent transmission and distribution stations in power grid. $\varepsilon(GEsG_i)$ denotes the sum of the efficiency of $GEsG_i$. d_{ij} is the length of the shortest path between node i and node j.

Figure 2 (a) shows a power grid which network efficiency is equal to 1. Because the power grid can be divided into two GEsGs in which the length of the shortest paths between each generation node and all demand nodes is equal to 1, its network efficiency is equal to 1 according to Formula 2. For instance, the length of the shortest paths between generation nodes 1, 8 and demand nodes 2, 3, 4, 5, 6, 7 is equal to 1 in Fig. 2 (b), and then $\varepsilon(GEsG_1) = \varepsilon(GEsG_8) = (1 + 1 + 1 + 1 + 1 + 1) = 6$. Therefore, the network efficiency of G is calculated by: $E(G) = \frac{1}{2\times6}(\varepsilon(GEsG_1) + \varepsilon(GEsG_8)) = \frac{(6+6)}{2\times6} = 1$.

In order to enhance the network security, we use critical degree to assess the importance of the nodes in power grid. If the critical nodes will be attacked, it may trigger other nodes to fail.

3. Critical Nodes Identification Algorithm

Critical nodes play an important role in exchanging information or transmitting energy in the network. The removal of a small number of critical nodes may cause large network efficiency loss, even results in a breakdown of the network. Therefore, one of the essential things of network security is to identify and protect critical nodes. Critical node identification depends on four factors: the node sharing degree that the common node is shared by different Giant Efficiency sub-Graphs, the distance sum of the shortest path between a certain node i and all of the generation nodes, out-degree of the node i, and in-degree of the node i.

Definition 2 (Node sharing degree): Any power grid can reconstruct Giant Efficiency sub-Graphs with N_{V^G} elements by shortest path connecting algorithm. There are

many common nodes between two GEsG. The node sharing degree $NSD(v_i)$ is measured by the sum of the function $\varphi_j(v_i)$.

$$NSD(v_i) = \sum_{j=1}^{N_{VG}} \varphi_j(v_i) \tag{3}$$

$$\varphi_j(v_i) = \begin{cases} 1, & \text{if } v_i \in V_j \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where v_i denotes the demand node i. N_{VG} denotes the number of GEsG. V_j denotes the node set of the $GEsG_j$.

Definition 3 (Distance sum of the shortest path): The distance sum $s(v_i)$, which is measured by calculating the distance sum of the shortest path between the node i and all of the generation nodes, is defined as follow:

$$s(v_i) = \sum_{v_i \in V^D} \sum_{v_j \in V^G} d_{v_i v_j} \tag{5}$$

where $d_{v_i v_j}$ is the shortest path length between the demand node v_i and the generation node v_j.

Let $G = (V, E)$ and $v \in V$. The in-degree of v is denoted $deg^-(v)$ and its out-degree is denoted $deg^+(v)$. A node with $deg^-(v) = 0$ and $deg^+(v) > 0$ is called a generation node on a power grid. Similarly, a node with $deg^-(v) > 0$ is called a demand node.

In this paper, we propose the concept of the critical degree (CD) to describe the importance level of the nodes on power grids. Critical nodes include two parts: all of the generation nodes that provide power for the power grid, and some demand nodes that have a key role in transmitting power to the user. We use the contribution rate of network efficiency to measure the CD of the generation nodes as follow:

$$CD(v_i^G) = \frac{\varepsilon_{v_i^G \in V_i}(GEsG_i)}{\sum \varepsilon(GEsG_i)} \tag{6}$$

where a node v_i^G belongs to vertices set V_i of giant efficiency subgraph i. $\sum \varepsilon(GEsG_i)$ denotes the sum of network efficiency of all Giant Efficiency sub-Graphs.

Similarly, the CD of a demand node is positively correlated with its out-degree and node sharing degree, yet is negatively correlated with its in-degree and distance sum of the shortest path. Thus, we use four factors to measure critical degree of the demand nodes as follow:

$$CD(v_i^D) = \alpha \times \frac{(deg^+(v_i^D) + \delta) \times NSD(v_i^D)}{(deg^-(v_i^D) + \delta) \times s(v_i^D) + \varphi(v_i^D)} \tag{7}$$

where v_i^D is a demand node i, $deg^+(v_i^D)$ and $deg^-(v_i^D)$ denote out-degree and in-degree of the demand node i,respectively. δ is the number that its limit tends to zero, and δ plus out-degree or in-degree is to avoid the situation that $deg^+(v_i^D)$ or $deg^-(v_i^D)$ equals zero. $NSD(v_i^D)$ and $s(v_i^D)$ denote node sharing degree and distance sum of v_i^D, respectively. $\varphi(v_i^D) = N_{V_G} - NSD(v_i^D)$ denotes penalty factor to punish these nodes that its NSD equals 1. $\alpha = 0.54945$ is a modifying factor to control the algorithm error of critical degree between the generation nodes and the demand nodes.

4. Node Efficiency Loss for CNI Algorithm Verification

The critical degree is used as an evaluation index to assess critical nodes. Similarly, in order to verify the accuracy of the CNI algorithm, we define node efficiency loss (NEL) to quantify the damage of network efficiency after a node fails. The nodes with a larger NEL are more important for protecting the power grid. The importance and the ranking of network components are adopted in terms of the relative drop in efficiency [31], [32]. When a node v_i fails, the downriver nodes can't receive electric energy, which makes a branch of power grids failure. Therefore, we define $branch(v_i)$ to describe the affected component due to the failure of a node v_i.

Definition 4 ($branch(v_i)$): $branch(v_i)$ is a part of GEsG, which is affected by a node i. The nodes of $branch(v_i)$ fail if and only if the node i is removed. If v_i is a generation node and belongs to $GEsG_k$, $branch(v_i) = GEsG_k, k = 1, 2, \cdots, N_{V_G}$. If $\forall v_i \in GEsG_k, \exists branch(v_i) = \{V_b, E_b\} \subseteq GEsG_k$.This means that for any node that belongs to $GEsG_k$, there must be a branch of the node i that also belongs to $GEsG_k$. When $v_j \in branch(v_i)$, v_j need to meets the condition that v_j is the downriver nodes of v_i and does not belong to the branch of other brother nodes of v_i.

There is the shortest path matrix $\{d_{ij}\}$ on $GEsG_k$ and a generation node set $s^G = \{v_1^G, v_2^G, \cdots, v_M^G\}$. A given node v_i is the demand node and the algorithm of generating $branch(v_i)$ is as follow: Step one is to find the node v_j that satisfy the condition: $d_{kv_j} = d_{kv_i}$, where $k \in s^G$ and $v_j \in \{v_{j_1}, v_{j_2}, \cdots, v_{j_k}\}$, obviously, the nodes v_j are brother nodes of v_i. Step two is to find the downriver nodes of v_i that satisfy the condition: $d_{v_i v_l} > 0$, where $v_l \in \{v_{l_1}, v_{l_2}, \cdots, v_{l_m}\}$. Step three is to delete the nodes that satisfy the condition: $d_{v_j v_l} > 0$ and to add the rest nodes to the node set V_b of $branch(v_i)$. For instance, Fig. 1 (b) shows that the nodes set of $branch(1)$ is {1, 2, 3, 4, 5}, and the nodes set of $branch(4)$ is {4, 3} on $GEsG_1$ and {4} on $GEsG_8$, respectively.

We note that $branch(v_i)$ will be invalid after the node i is removed. Therefore, in order to calculate network efficiency loss after the node i fails, the definition of the damage $D(v_i)$ is the node efficiency loss that is caused by the failure of node v_i. We can verify validity and feasibility of critical node identification algorithm based on node efficiency loss with GEsG.

$$\begin{aligned} D(v_i) &= \frac{\Delta E}{\varepsilon(G)} \\ &= \frac{\varepsilon(G) - \sum_{v_i \in GEsG_k} \varepsilon(GEsG_k - branch(v_i))}{\varepsilon(G)} \\ &= \frac{\sum_{v_i \in GEsG_k} \varepsilon(branch(v_i))}{\varepsilon(G)} \end{aligned}$$

$$= \frac{\sum_{v_i \in GEsG_k, v_j \in branch(v_i)} \frac{1}{d_{v_i v_j}}}{\varepsilon(G)} \quad (8)$$

where $\varepsilon(G)$ is the efficiency of the network before the occurrence of any breakdown and ΔE is the efficiency loss that the branch of the node i is disable due to the failure of a node i. $\varepsilon(GEsG_k - branch(v_i))$ is the efficiency of $GEsG_k$ after a breakdown of the branch, when the network reaches a steady state.

5. Case Study

In this section, we first use practical the Hainan Power Grid, which was severely damaged by Typhoon Damrey on September 26, 2005, as an example to evaluate our approach. Hainan is a province in South China with a population of more than 8.26 million. The Hainan Power Grid (HPG), part of the China Southern Power Grid (CSPG), consists of 118 substations and more than 180 transmission lines [33]. For the purpose of security and secret, we only adopt a small part of HPG that is composed of 48 nodes and 63 links in Fig. 3. A directed graph is used to represent the HPG in Fig. 3, where red square nodes represent thermal power plants and green circle nodes represent substations, blue lines represent 220 KV transmission lines and red lines represent 110 KV transmission lines, and the arrows indicate the direction of electric current flow from high voltage substations to low voltage ones or from the generation nodes to the demand nodes.

In Fig. 3, we note that directed graph of HPG power grid can be divided into nine GEsGs by the shortest path connecting algorithm. Electric current flows from each generation node to other demand nodes in different GEsGs. Edges with arrows represent the path of power transmission and the direction of the arrow represents the direction of power transmission in Fig. 3. It is clear that node 211 is responsible for transmitting power for generation nodes 11, 12, 14 and 15. Similarly, power transmission from generation nodes 11, 12, 13, 14 and 15 passes through node 204, and power transmission from generation nodes 11, 12, 13, 14, 15 and 19 passes through node 306. Since failed nodes 211, 204 and 306 will have a great impact on the power grid and may lead to the failure of more demand nodes due to lack of sufficient power supply, the failure of these common nodes (i.e., node 211, 204 and 306) would affect GEsG that contain those nodes. For instance, failed node 306 may trigger 309, 310 and 311 to fail due to lack of power supply.

The GEsG, which is used as a criterion for assessing the performance of exchanging information or transmitting energy, and is better than Giant Component in power grid. For instance, we remove each generation node from the HPG one by one and calculate the network efficiency according to GEsG and giant component, respectively. Figure 4 shows that network efficiency has not significant change if Giant Component is used as the criterion. When all generation nodes are removed, the power grid has failed. The network efficiency is zero for the criterion with GEsG, which can reflect the practice situation of the power grid.

Figures 5 (a)–(d) show the value of four factors for identifying critical nodes, i.e., the node sharing degree, distance sum of the shortest path, in-degree and out-degree. The node sharing degree is calculated according to Formulas (3) and (4), which denotes the number of times that the node is shared by different GEsGs. For instance, NSD of the nodes 319, 215, 318 and 317 is equal to 9 in Fig. 5 (a), which means that these nodes are shared by nine GEsGs and are more important in transmitting electric energy. If these nodes are removed, nine GEsGs will be affected to cause a portion of efficiency loss. Distance sum of the shortest path is calculated by Formula (5). It is easy to find that nodes

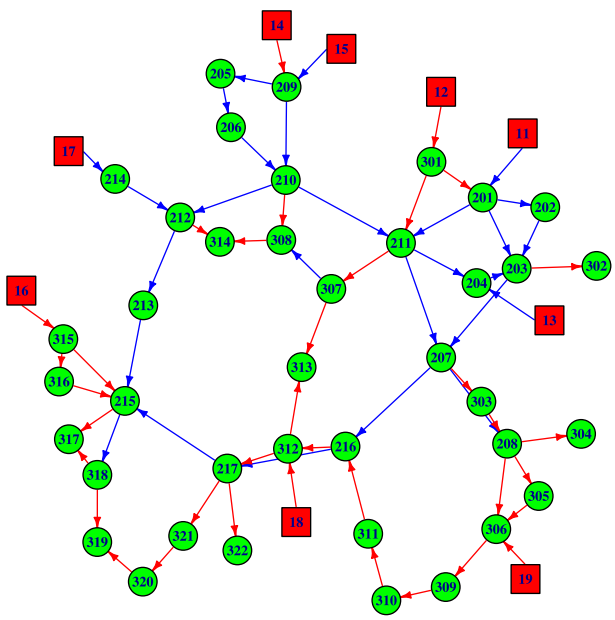

Fig. 3 Directed network structure diagram of Hainan power grid (HPG)

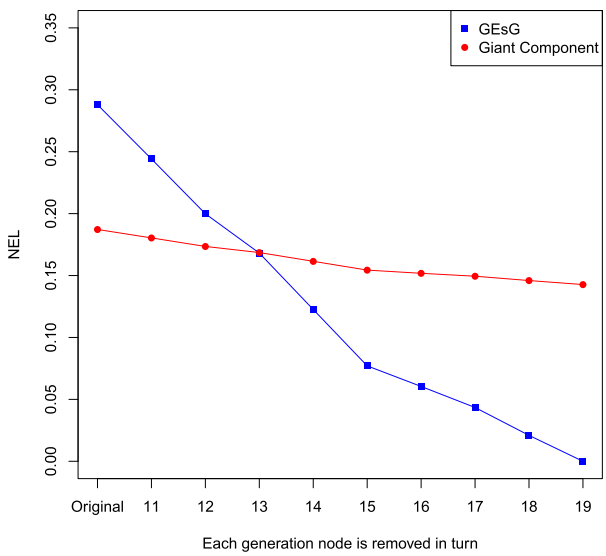

Fig. 4 The comparison of network efficiency (NE) between GEsG and Giant Component

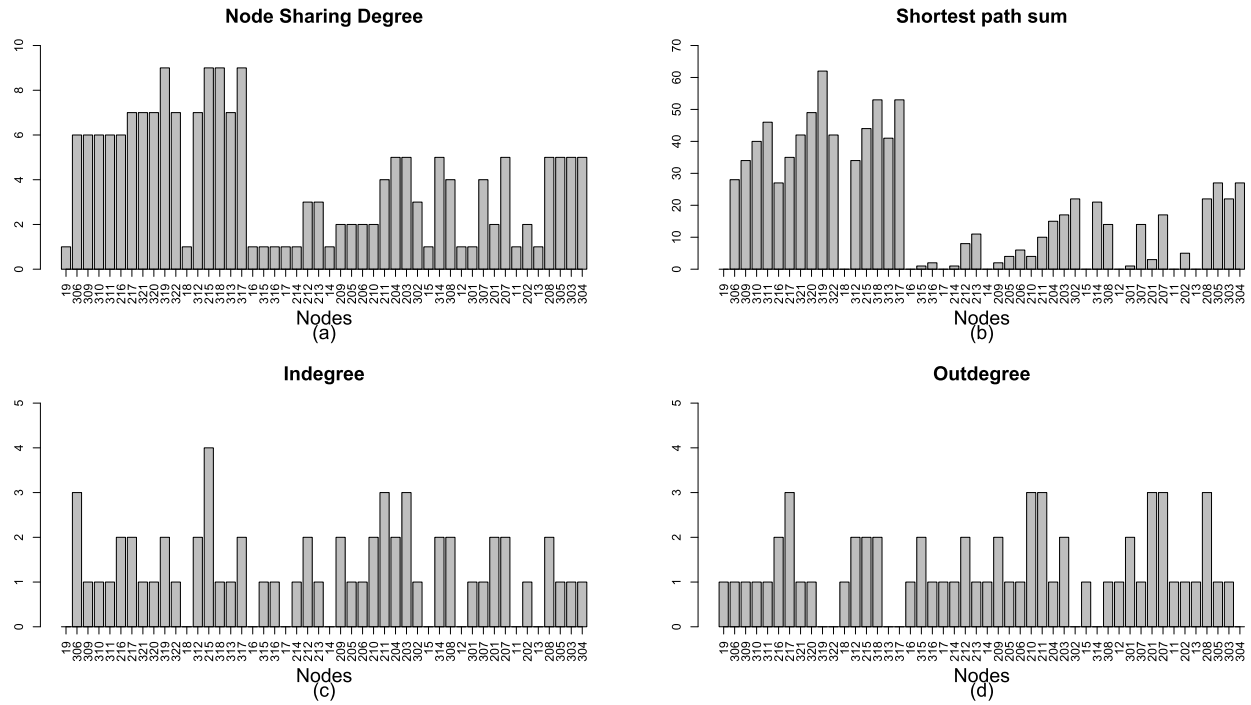

Fig. 5 Four factors of CNI algorithm. (a) Node sharing degree. (b) The distance sum of the shortest path. (c) In-degree. (d) Out-degree.

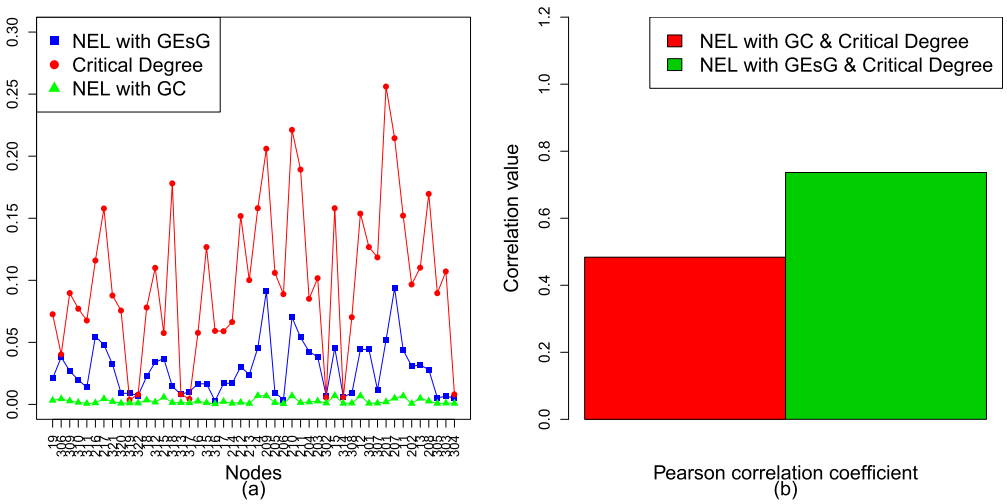

Fig. 6 (a) The trend charts of critical degree, NEL with GEsG and NEL with giant component. (b) The histogram of the correlation comparison between critical degree & NEL with GEsG and critical degree & ENL with GC.

317, 318, 319 and 320 have the larger value of the shortest path sum in Fig. 5 (b). It means that these nodes may not be critical nodes, because there are long distances between them and all generation nodes. The smaller distance sum is, the more important the node may be. Figure 5 (c) also shows that the in-degree of the nodes 11, 12, 13, 14, 15, 16, 17, 18 and 19 is equal to zero, but out-degree of these nodes is greater than 0. There are also some node with out-degree 0 in Fig. 5 (d), such as 319, 322, 313, 317, 302, 314, and 304.

The critical degree (CD) is an important evaluator to identify critical nodes in a network. Node Efficiency Loss is used as an approach to verify the correctness of the algorithm of CNI. Particularly, the larger critical degree of the node is, the more its efficiency loses. Therefore, node efficiency loss with GEsG is in a highly significant correlation with the critical degree, as shown in Fig. 6 (a), a similar tendency is observed between red curve and blue curve, but node efficiency loss with Giant Component (green curve) has no significant change and is not correlated with critical

Table 1 Comparative analysis of critical degree and node efficiency loss

Descending Order	Node	Critical Degree	Node	Node Efficiency Loss
1	201	0.25613484	207	0.09370958
2	210	0.22120737	209	0.09107765
3	207	0.21452099	210	0.07018496
4	209	0.20604396	216	0.054687288
5	211	0.1892552	211	0.053932074
6	318	0.17812376	201	0.051878985
7	208	0.16965106	217	0.048130058
8	14	0.15813568	14	0.045538824
9	15	0.15813568	15	0.045538824
10	217	0.15792155	301	0.044281643
11	12	0.15377007	12	0.044281643
12	11	0.15212119	11	0.04380681
13	212	0.151821870	204	0.042572240
14	301	0.126796280	203	0.038298737
15	315	0.126796280	306	0.038000270

degree (red curve).In order to further verify our conclusion with objective evidence, Pearson correlation coefficient is used to calculate the correlation between Critical Degree & NEL with GEsG and Critical Degree & ENL with GC. Pearson correlation coefficient of samples X and Y is described as:

$$Corr(X, Y) = cov(X, Y)/(\sigma_X * \sigma_Y) \tag{9}$$

where $cov(X, Y)$ represents the covariance between samples X and Y. σ_X and σ_Y represent the variance of samples X and Y, respectively. The closer the correlation coefficient of samples X and Y is to -1 or 1, the stronger the correlation of samples X and Y is; the closer the correlation coefficient is to 0, the weaker the correlation of samples X and Y is. According to Formula (9), $Corr(NEL\ with\ GC,\ Critical\ Degree) = 0.4837831$ and $Corr(NEL\ with\ GEsG,\ Critical\ Degree) = 0.7365024$. The correlation comparison of Critical Degree, NEL with GEsG and NEL with Giant Component is shown in Fig. 6 (b), it is clear that the correlation value of NEL with GEsG and Critical Degree is obviously higher than NEL with GC and Critical Degree. Therefore, the GEsG is more suitable as a criterion to reflect the efficiency loss of nodes in the power grid than Giant Component.

The CD of nodes is calculated by Formula (6) and (7), which veracity can be verified by the NEL according to Formula (8). Approximately 12 percent of the nodes are taken as critical nodes in HPGC power grid. In particular, nodes 201, 210, 207, 209, 211 and 318 have a higher value in descending order of CD, and nodes 207, 209, 210, 216, 211 and 201 have a higher value in descending order of NEL. The top 12% of nodes from CD and NEL has the same nodes 201, 207, 209, 210, and 211, which means that CD has a better effect on identifying critical nodes. In order to verify the effectiveness of critical degree, the veracity of critical node identification (CNI) algorithm is written as:

$$Veracity(p) = \frac{n(CDS_{\lceil N*p \rceil} \bigcap NELS_{\lceil N*p \rceil})}{\lceil N * p \rceil} \tag{10}$$

where p represents the proportion of selected critical nodes and N is the number of all nodes in the power grid. $n(CDS_{\lceil N*p \rceil} \bigcap NELS_{\lceil N*p \rceil})$ represents the number of nodes in the intersection of $CDS_{\lceil N*p \rceil}$ and $NELS_{\lceil N*p \rceil}$. $\lceil N * p \rceil$ represents the ceil function of N*p, which means that $\lceil N * p \rceil$ is equal to integer m when $m - 1 < N * p < m$. $CDS_{\lceil N*p \rceil}$ represents critical node set in which $\lceil N * p \rceil$ nodes are selected in descending order of CD. $NELS_{\lceil N*p \rceil}$ represents NEL set in which $\lceil N * p \rceil$ nodes are selected in descending order of NEL. According to the Table 1 and Formula (10), if we take the top 10 percent of the nodes as critical nodes, the veracity of CNI algorithm can be calculated as follows: $Veracity(10\%) = \frac{c(CDS_{\lceil 48*10\% \rceil} \bigcap NELS_{\lceil 48*10\% \rceil})}{\lceil 48*10\% \rceil} = \frac{c(CDS_5 \bigcap NELS_5)}{5}$ $= \frac{n(\{201,210,207,209,211\} \bigcap \{207,209,210,216,211\})}{5} = \frac{n(\{207,209,210,211\})}{5} = \frac{4}{5} = 0.8 * 100\% = 80\%$. Similarly, if 20 percent of the nodes are taken as critical nodes, the veracity of CNI algorithm can be calculated as follows: $Veracity(20\%) = \frac{n(CDS_{10} \bigcap NELS_{10})}{10} = \frac{n(\{14,15,201,207,209,210,211,217\})}{10} = \frac{8}{10} = 80\%$.

We notice an interesting phenomenon in this experiment that the most critical important nodes may not be the generation nodes but some demand nodes. It has been verified from two different aspects: critical degree and network efficiency loss. Some nodes (i.e., node 207, 209, 210, 211 and 201) hare the larger value on Critical Degree and Node Efficiency Loss than these generation nodes (i.e., node 11, 12, 13, 14, 15, 16, 17, 18 and 19). The reason that leads to this counterintuitive phenomenon is that the failure of those critical nodes does not only lose the efficiency of its GEsG but also causes efficiency loss of other GEsGs, and efficiency losses of these nodes are larger than the one caused by the failure of any a generation node.

Table 1 summarizes that five same nodes are found in the first nodes, as is shown in the results of Critical Degree and Node Efficiency Loss. If we remove those nodes (e.g., nodes 201, 207, 209, 210 and 211), the power grid is cut into five parts, as shown in Fig. 7 above, and it will lead to a serious consequence that 27 percent of the nodes are invalid and 41.26 percent of edges are removed. Although the power grid is divided into five parts, each part is still effective component. Therefore, the total efficiency of the network depends not only on the largest component but also on the smaller one that contains the generation node and the

demand nodes. In particular, this concept of GEsG should be applied to studying the problem of cascading failures in interdependent networks (a power grid and a communication network) for future.

Recently, degree, betweenness, closeness or degree of degree are used as evaluation indicators to identify critical nodes in many research works. The accuracy of these criteria and critical degree are compared in Fig. 8. The results show that the curve of critical degree is much closer to the curve of node efficiency loss than other criteria in Fig. 8 (a). In order to verify the high correlation between critical degree and node efficiency loss, Pearson correlation coefficient is used to calculate and compare the correlation between NEL and degree, betweenness, closeness, degree of degree or critical degree. According to Formula (9), $Corr(Degree, NEL) = 0.5514295$, $Corr(Betweenness, NEL) = 0.6496241$, $Corr(Closeness, NEL) = 0.5665583$, $Corr(Degreeof Degree, NEL) = 0.5578868$ and $Corr(Critical\ degree, NEL) = 0.7365024$. This indicates that critical degree has a high correlation with node efficiency loss in Fig. 8 (b). We take the top five, ten, twenty and thirty percent of the nodes as critical nodes, respectively, the accuracy of critical degree is always the highest in Fig. 8 (c), where p represents the proportion of the selected critical nodes. According to Formula (10), when p is equal to 5, 10, 20 and 30 percent, the veracity of critical degree is 0.6666667, 0.8, 0.8 and 0.7333333, respectively.

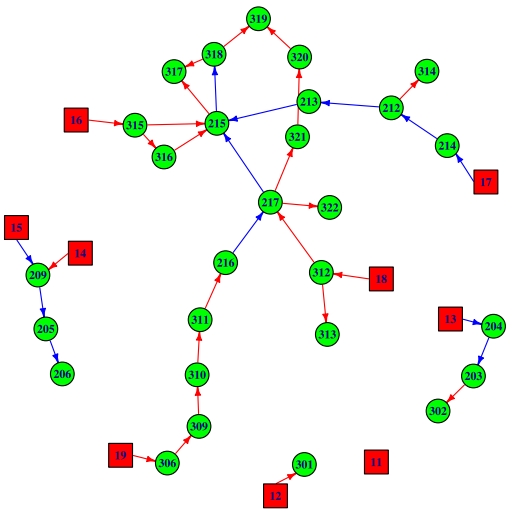

Fig. 7 Directed graph of HPG after five critical nodes fail

From the perspective of the complex network, we study the problem of CNI and the efficiency assessment of power grids. However, a real power grid system has more characters such as voltage, load, and control loops. Genge et al. [34] proposed a cyber attack impact assessment (CAIA) methodology to assess the impact of cyber-attacks on critical infrastructures assets and to rank the assets with a viewpoint of the real network condition. Relative impact is used as an evaluation index in CAIA and can be used to rank the importance of substations. In order to verify the conclusion that CNI methodology is accordant with the actual situation of power grids, we apply our approach on the IEEE 118-bus system [16], [35] that is simulated by the MATLAB PAST toolbox and compare experimental results between critical degree and relative impact. Figure 9 (a) is the graphic presentation of the IEEE 118-bus system, which contains 19 generators (red nodes), 117 links, 99 load nodes (blue nodes). We separately run the algorithms of CNI and CAIA based on the data of IEEE 118-bus test system. In addition, in order to verify the relatively high correlation between relative impacts and critical degree, we separately use degree, betweenness, closeness, and degree of degree as evaluation indicators to calculate the importance of all substations of IEEE 118-bus test system in Fig. 9 (b). Similarly, Pearson correlation coefficient is used to calculate the correlation between relative impact and degree, betweenness, closeness, degree of degree or

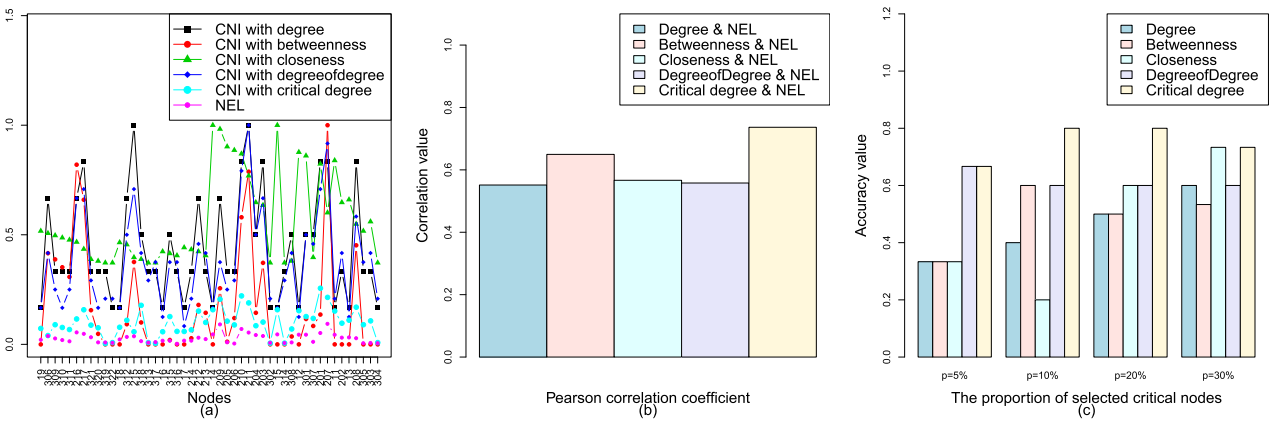

Fig. 8 (a) The comparison of CNI with different evaluation indicators. (b) The correlation value between NEL and degree, betweenness, closeness, the degree of degree, or critical degree. (c) The comparison of the accuracy of different proportions of the selected critical nodes, i.e., 5%, 10%, 20%, and 30%

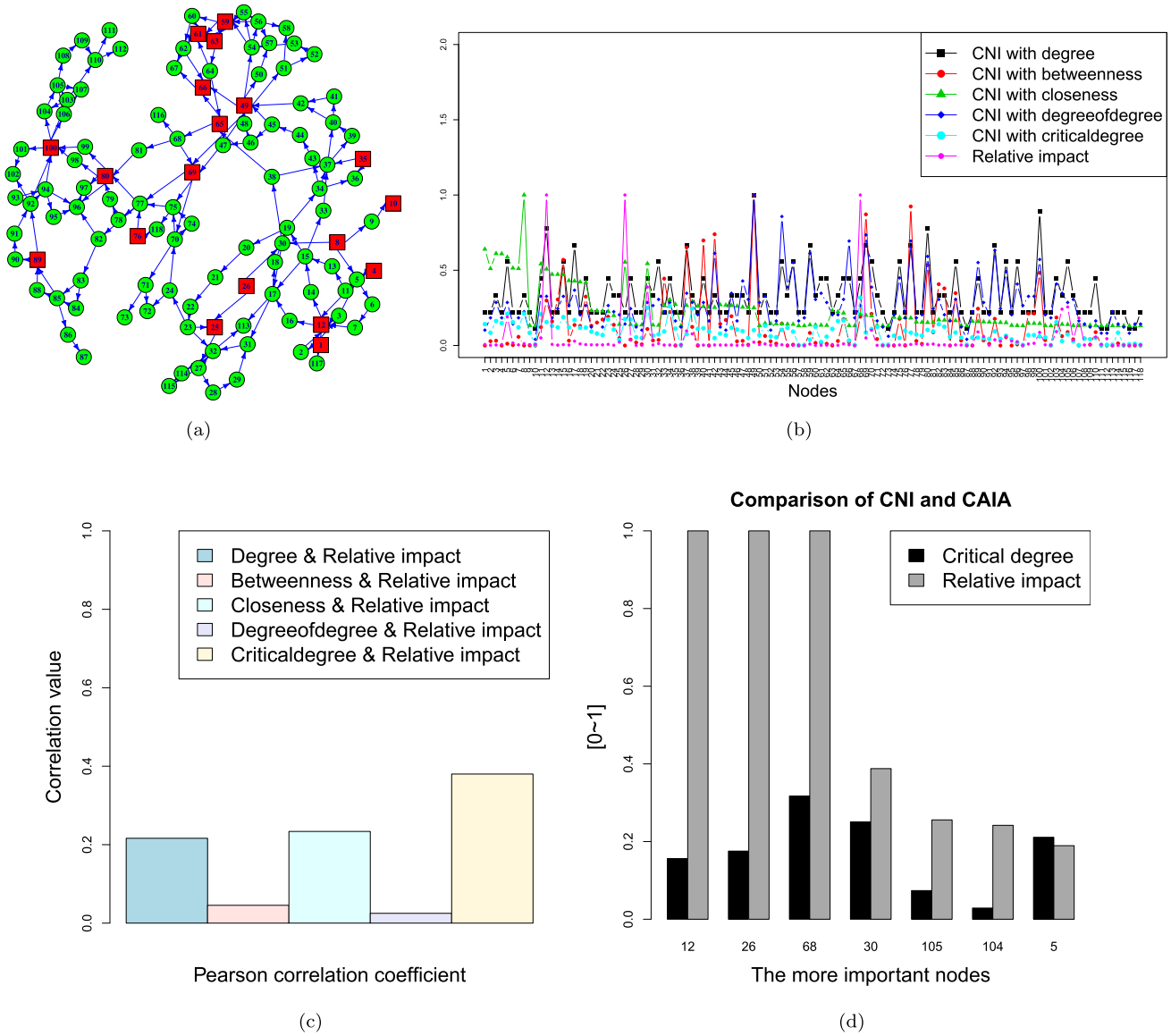

Fig. 9 (a) Network diagram of the IEEE 118-bus test system. (b) Importance assessment value of substations based on different evaluation indicators. (c) The correlation comparison between relative impact and degree, betweenness, closeness, degree of degree, or critical degree. (d) Comparison diagram of CNI and CAIA on the IEEE 118-bus test system.

critical degree. According to Formula (9), the correlation between relative impact and other evaluation indicators is as follows: $Corr(Degree, Relative\ impact) = 0.2163506$, $Corr(Betweenness, Relative\ impact) = 0.04544072$, $Corr(Closeness, Relative\ impact) = 0.2337256$, $Corr(Degree of Degree, Relative\ impact) = 0.02500777$, $Corr(Critical\ degree, Relative\ impact) = 0.3800736$. This indicates that relative impact has a higher correlation with critical degree than degree, betweenness, closeness, and degree of degree in Fig. 9 (c). By comparing the results of critical nodes and relative impact on assets, we also reach the same conclusion that the most critical nodes may not be generation nodes but some demand nodes and relative impact of some demand nodes are larger than generation nodes. We can see that both critical degree and relative impact of the demand node 68 have the largest value in Fig. 9 (d), and the more important critical nodes include two generation nodes (i.e. red nodes 12 and 26) and five demand nodes (i.e. blue nodes 5, 30, 68, 104 and 105). The critical nodes that are identified by the CNI algorithm also have a larger value on relative impact by the CAIA methodology. If these nodes fail due to being attacked, the network will suffer a dramatic power loss and the local failure of power grids may also be triggered.

6. Conclusion and Future Work

In this paper, a novel approach is proposed to identify

critical nodes in power grids, and a counterintuitive phenomenon has been discovered that the most important critical nodes may not be the generation nodes but some demand nodes. We define the Giant Efficiency sub-Graph (GEsG) as an evaluation criterion to assess network efficiency. We also define Node Efficiency Loss to verify the accuracy of CNI algorithm. Experimental results show that the algorithm accuracy is over 80%. Through the comparison of the node efficiency loss between GEsG and Giant Component, GEsG is more suitable for the actual situation of power grids than Giant Component. From the perspective of network structure, the algorithm has a good effect to identify critical nodes. In fact, the smart grid is considered as multi-layer or interdependent networks, in which a fraction of nodes in the power grid is coupled with corresponding nodes in communication network [36]–[38]. Therefore, through identifying critical nodes and edges of interdependent networks, the mechanisms of cascading failure of the cyber-physical systems are worth deeply researching, and developing the cyber-physical attack strategy and defense mechanism will also be my further research.

Acknowledgments

This work was jointly supported by the National Science Foundation of China (NSFC) under Grant 61572514 and Changsha Science and Technology Program (Grant K1705007).

References

[1] Q. Sun, R. Han, H. Zhang, J. Zhou, and J.M. Guerrero, "A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet," IEEE Trans. Smart Grid, vol.6, no.6, pp.3006–3019. 2015.
[2] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the us power grid," Safety Science, vol.47, no.10, pp.1332–1336, 2009.
[3] J.-W. Wang and L.-L. Rong, "Robustness of the western united states power grid under edge attack strategies due to cascading failures," Safety science, vol.49, no.6, pp.807–812, 2011.
[4] The Industrial Control Systems Cyber Emergency Response Team (ICS-CERT), Cyber-attack against Ukrainian critical infrastructure, Alert (IR-ALERTH-16-056-01), 2016. Available at url: https://www.ics-cert.us-cert.gov/alerts/IRALERT-H-16-056-01
[5] S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," Nature, vol.464, no.7291, pp.1025–1028, 2010.
[6] A. Vespignani, "Complex networks: The fragility of interdependency," Nature, vol.464, no.7291, pp.984–985, 2010.
[7] A.E. Motter and Y.-C. Lai, "Cascade-based attacks on complex networks," Physical Review E, vol.66, no.6, 065102, 2002.
[8] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," Reviews of modern physics, vol.74, no.1, p.47, 2002.
[9] C. Song, S. Havlin, and H.A. Makse, "Self-similarity of complex networks," Nature, vol.433, no.7024, pp.392–395, 2005.
[10] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," Physics reports, vol.424, no.4, pp.175–308, 2006.
[11] S.V. Buldyrev, N. Shere, and G.A. Cwilich, "Interdependent networks with correlated degrees of mutually dependent nodes," arXiv preprint arXiv:1009.3183.
[12] J. Gao, S.V. Buldyrev, S. Havlin, and H.E. Stanley, "Robustness of a network of networks," Physical Review Letters, vol.107, no.19, 195701, 2011.
[13] W. Li, A. Bashan, S.V. Buldyrev, H.E. Stanley, and S. Havlin, "Cascading failures in interdependent lattice networks: The critical role of the length of dependency links," Physical review letters, vol.108, no.22, 228702, 2012.
[14] J. Gao, S.V. Buldyrev, S. Havlin, and H.E. Stanley, "Robustness of a network formed by n interdependent networks with a one-to-one correspondence of dependent nodes," Physical Review E, vol.85, no.6, 066134, 2012.
[15] X. Chen, "Critical nodes identification in complex systems," Complex & Intelligent Systems, vol.1, no.1-4, pp.37–56, 2015.
[16] Y.-S. Li, D.-Z. Ma, H.-G. Zhang, and Q.-Y. Sun, "Critical nodes identification of power systems based on controllability of complex networks," Applied Sciences, vol.5, no.3, pp.622–636, 2015.
[17] O. Sims and R.P. Gilles, "Critical nodes in directed networks," arXiv preprint arXiv:1401.0655.
[18] C. Bazgan, S. Toubaline, and D. Vanderpooten, "Critical edges/nodes for the minimum spanning tree problem: complexity and approximation," Journal of Combinatorial Optimization, vol.26, no.1, pp.178–189, 2013.
[19] K. Wehmuth and A. Ziviani, "Distributed location of the critical nodes to network robustness based on spectral analysis," 2011 7th Latin American Network Operations and Management Symposium (LANOMS), pp.1–8, IEEE, 2011.
[20] B. Sivakumar and G. Varaprasad, "Identification of critical node for the efficient performance in manet," International Journal of Advanced Computer Science and Applications, vol.3, no.1, pp.166–171, 2012.
[21] A. Nagurney and Q. Qiang, "Identification of critical nodes and links in financial networks with intermediation and electronic transactions," Computational Methods in Financial Engineering, pp.273–297, Springer, 2008.
[22] S. Cuenda, R. Vida, and J. Galeano, "Identifying critical nodes in multi-layered networks under multi-vector malware attack," Int. J. Complex Systems in Science, vol.3, no.1, pp.97–105, 2013.
[23] K.M. Rossella Mattioli, "Communication network interdependencies in smart grids," Communication network interdependencies in smart grids, 1–53, 2016.
[24] Z. Wang, A. Scaglione, and R.J. Thomas, "Electrical centrality measures for electric power grid vulnerability analysis," 2010 49th IEEE Conference on Decision and Control (CDC), pp.5792–5797, IEEE, 2010.
[25] Z. Wang, A. Scaglione, and R.J. Thomas, "Power grid vulnerability measures to cascading overload failures," 2012 Asia-Pacific Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), pp.1–5, IEEE, 2012.
[26] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," Physical review letters, vol.87, no.19, 198701, 2001.
[27] P. Crucitti, V. Latora, and M. Marchiori, "Model for cascading failures in complex networks," Physical Review E, vol.69, no.4, 045104, 2004.
[28] P. Crucitti, V. Latora, A. Rapisarda, and M. Marchiori, "Efficiency of scale-free networks," Physica A, vol.320, p.642, cond-mat/0205601, 2002.
[29] P. Crucitti, V. Latora, and M. Marchiori, "A topological analysis of the italian electric power grid," Physica A: Statistical Mechanics and its Applications, vol.338, no.1, pp.92–97, 2004.
[30] R. Kinney, P. Crucitti, R. Albert, and V. Latora, "Modeling cascading failures in the north american power grid," The European Physical Journal B-Condensed Matter and Complex Systems, vol.46, no.1, pp.101–107, 2005.
[31] A. Nagurney and Q. Qiang, "A network efficiency measure with application to critical infrastructure networks," Journal of Global Optimization, vol.40, no.1-3, pp.261–275, 2008.
[32] V. Latora and M. Marchiori, "How the science of complex networks

can help developing strategies against terrorism," Chaos, solitons & fractals, vol.20, no.1, pp.69–75, 2004.
[33] L. Chang and Z. Wu, "Performance and reliability of electrical power grids under cascading failures," International Journal of Electrical Power & Energy Systems, vol.33, no.8, pp.1410–1419, 2011.
[34] B. Genge, I. Kiss, and P. Haller, "A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures," International Journal of Critical Infrastructure Protection, vol.10, pp.3–17, 2015.
[35] Z.Y. Li, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "Bilevel Model for Analyzing Coordinated Cyber-Physical Attacks on Power Systems," IEEE Transactions on Smart Grid, vol.7, no.5, pp.2260–2272, 2016.
[36] M. Korkali, J.G. Veneman, B.F. Tivnan, and P.D. Hines, "Reducing cascading failure risk by increasing infrastructure network interdependency," arXiv preprint arXiv:1410.6836.
[37] C. Pu, S. Li, X. Yang, J. Yang, and K. Wang, "Information transport in multiplex networks," Physica A: Statistical Mechanics and its Applications, vol.447, pp.261–269, 2016.
[38] Y. Cai, Y. Cao, Y. Li, T. Huang, and B. Zhou, "Cascading failure analysis considering interaction between power grids and communication networks," IEEE Trans. Smart Grid, vol.7, no.1, pp.530–538, 2016.

WenJie Kang received his Master's degree in computer science from the National University of Defense Technology (NUDT), China, in 2014. He is currently Ph.D. candidate in computer science from the National University of Defense Technology (NUDT). His interests include big data analysis and data visualization, complex network and the cyber-physical system security.

PeiDong Zhu received his Ph.D. degree in computer science from the National University of Defense Technology (NUDT) in 1999. He is currently a professor at the Department of Electronic Information and Electrical Engineering of Changsha University. His research interests include network routing, security of large-scale cyber-physical network, and architecture design of the Internet. Prof. Zhu is now a senior member of IEEE.

JieXin Zhang received his Master' degree in computer science from the National University of Defense Technology (NUDT) in 2016. He is a PhD candidate with College of Computer Science of NUDT. His research is focused on complex network, big data analysis and the cyber-physical system security.

JunYang Zhang received the Master degree in 2013, both from National University of Defense Technology. He is currently a Ph.D. student of National University of Defense Technology. His research interests include Artificial Intelligence, big dada analysis and Deep Learning.