

## Article

# Identification of Key Nodes in a Power Grid Based on Modified PageRank Algorithm

Darui Zhu , Haifeng Wang, Rui Wang, Jiandong Duan  and Jing Bai

School of Electrical Engineering, Xi'an University of Technology, Xi'an 710048, China; 2190321201@stu.xaut.edu.cn (H.W.); 2160320014@stu.xaut.edu.cn (R.W.); duanjd@xaut.edu.cn (J.D.); 2211920048@stu.xaut.edu.cn (J.B.)

\* Correspondence: zhudarui@xaut.edu.cn

**Abstract:** For avoiding the occurrence of large-scale blackouts due to disconnected nodes in the power grid, a modified PageRank algorithm is proposed to identify key nodes by integrating the topological information and node type. The node betweenness index is first introduced based on complex network theory, which is modified to reflect the node topological information in the power grid. Then, according to the characteristics of different node types in the power grid, a modified PageRank algorithm is proposed to rapidly identify key nodes, which takes the generator nodes, load nodes, and contact nodes into account. IEEE 39-Bus system and IEEE 118-Bus system are used for the simulations. Simulation results showed that the network transmission efficiencies of the power grid are reduced from 64.23% to 5.62% and from 45.4% to 5.12% in the two simulation systems compared with other methods. The proposed identification algorithm improved the accuracy, and a provincial power grid simulation system in China is used to verify the feasibility and validity. The identified nodes are removed, which split the power grid according to importance index values. The proposed method in this paper is helpful to prevent the occurrence of cascading failure in the power system, and it can also be used to power systems with renewable energy sources and an AC/DC hybrid power grid.

**Keywords:** complex network; node betweenness; modified PageRank algorithm; transmission efficiency; empirical verification



**Citation:** Zhu, D.; Wang, H.; Wang, R.; Duan, J.; Bai, J. Identification of Key Nodes in a Power Grid Based on Modified PageRank Algorithm. *Energies* **2022**, *15*, 797. <https://doi.org/10.3390/en15030797>

Academic Editors: Abu-Siada Ahmed and Alberto-Jesus Perea-Moreno

Received: 15 December 2021

Accepted: 17 January 2022

Published: 22 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In a large-scale power grid, natural disasters, deliberate assaults, element failures, and other faults may cause large-scale blackouts [1]. Blackouts that occur in the power grid are considered great impact events and may cause large load shedding and even serious social impacts [2,3]. Usually, a blackout develops with one or several failures of power grid elements, which are referred as key elements, such as key generators, transmission lines, transformers, or power load nodes. Generators and power load nodes are at the core of power production and consumption in the power grid, and their failure will have a serious impact on the operation of the power grid. Authors [4] have proposed a new load distribution law to emulate the power grid, where the initial generation of generators and the initial loads of substations are calculated according to the path efficiency and the load of the consumers. The importance of generator nodes and power load nodes is discussed. For avoiding the occurrence of large-scale blackouts, it is important to identify key nodes in the power grid. In the literature, the identification of key nodes in the power grid can be divided into two aspects: dynamic analysis methods and static analysis methods.

For dynamic analysis methods, transmission line faults and load changes are commonly referred to identify key nodes. In [5], a cascading failure model is proposed based on complex network theory by combining the node overload failures and hidden failures of transmission lines in blackouts. In terms of voltage stability analysis, authors in [6]

proposed a new index for identification of vulnerable nodes to improve voltage stability based on reactive compensation. In [7], the influence of different faults is analyzed, and a quantitative coupling degree method is proposed to identify key nodes from the regional power grid in a transient process. From the characteristic analysis of network load, a cascading failure model is established based on the network important assessment index, which is considered the load oscillation degree of the attacked nodes in [8]. In order to avoid only power sources re-energizing the critical loads, authors in [9] proposed a new look-ahead restoration strategy for re-energizing the critical loads. The aforementioned methods can identify the key nodes from the view of operation characteristics in the power grid, but the topological structure is ignored. The key node identification method of the integrated grid topology and operation parameters can fully reflect the importance of nodes in the power grid.

In the past decade, static analysis methods that identify key nodes in the power grid have gradually grown, for example, complex network centrality in [10], topological and controllability features in [11], and electrical betweenness combined with generation rated capacity and load change in [12,13]. Expanded betweenness has been proposed, which considers transmission distribution factors and transmission ultimate capacity [14,15], the network response structural characteristic indexes have been formulated in terms of the Kirchhoff matrix [16], and the bus dependency matrix is established by the maximum power flow of the shortest path and node [17]. Additionally, an improved structure holes theory has been defined by the relative importance between the node and its adjacent node [18]. Because the above indexes and methods only partly consider factors of power grid nodes, the identification results are inaccurate. Based on the topological structure and electrical characteristics, the multi-index evaluation algorithm has been proposed by a different comprehensive method [19]. Authors in [20] proposed a ranking process method, which includes both static (via optimal power flow) and dynamic (via transient stability) performance analyses, to assess deterministic indices. Authors in [21] proposed a Coupling Strength Matrix (CSM) method, which is based on Network Structural Characteristics Theory and the Relative Electrical Distance (RED) between nodes in the network. The basic idea is to establish a power grid model that can depict the actual grid characteristics based on graph theory or complex network theory, and then establish indexes to identify important nodes. The operating characteristics and node types of the power system need to be considered.

Recently, due to the high speed and accuracy for identifying important nodes in a directed network [22], the famous PageRank algorithm has received much attention in many fields. In a power grid, a modified PageRank algorithm is presented to assess the node importance, which considers the characteristics of nodal load properties, transmission ultimate capacity, and model structure [23]. In [24], a simplified connection diagram is constructed to reveal the cascading failure characteristic with hidden failures, and a modified PageRank algorithm is designed to assess ultra-fast, vulnerable transmission lines in large-scale power grids. The modified sorting algorithm PageRank, called hypertext induced topic selection (HITS), is proposed to identify key nodes in the power grid, which is modified based on power flow, load capacity, and power source in [25]. In reference [26], according to the improved PageRank algorithm and optimization coefficient of each node, the important nodes are determined in the distribution network; those algorithms are obtained by the modified PageRank algorithm, which can assess the importance of nodes and identify the key nodes by iteration. However, the node importance varies with node type in the power grid. It is necessary to assess node importance according to the characteristics of different node types. Meanwhile, in a radial distribution network, AC/DC hybrid power system, the modified transfer matrix  $H$  should be defined according to the topology and operation characteristics. Therefore, based on the topological information, the characteristics of node type are introduced in the PageRank algorithm for identifying key nodes of the power grid as follows.

The structure of the remainder of this paper is as follows. Section 2 states the preliminary complex network theory and power grid model. The modified PageRank algorithm is proposed for identification of key nodes by considering node type and state information transfer in Section 3. In Section 4, based on IEEE 39-Bus system, IEEE 118-Bus system, and a provincial power grid simulation system, simulation results are compared and discussed with other approaches in detail. Finally, discussions and conclusions are given in Section 5.

## 2. Preliminary of Complex Network and PageRank Algorithm

### 2.1. Complex Network Theory

Based on complex network theory, a real-world network model can be abstracted into a graph composed of nodes and edges. Let the graph  $G = (V, E)$ ,  $V = \{v_1, v_2, \dots, v_n\}$  be the set of nodes with  $n$  vertices, and  $E = \{e_1, e_2, \dots, e_m\}$  is the set of edges with  $m$  edges. Then, the  $n \times n$  adjacency matrix  $B = \beta_{ij}$  can describe the connection between nodes in graph  $G$ ; that is, when node  $i$  and node  $j$  are connected together,  $\beta_{ij} = 1$ , otherwise,  $\beta_{ij} = 0$ . Several topological properties have been proposed in the past to capture the structural characteristics of various networks, as listed below [27].

#### (1) Degree

Degree is the fundamental character of nodes. It depicts the connection relation that one node is linking with others. The degree of node  $i$  in graph  $G$  is defined as

$$d_i = \sum_{j=1}^n \beta_{ij} \quad (1)$$

where  $\beta_{ij}$  is the element of the adjacency matrix  $B$ , and  $n$  is the number of nodes in graph  $G$ .

#### (2) Betweenness

Betweenness can describe the interaction of the complex network, which is divided into node betweenness and edge betweenness [28]. Node betweenness is defined as the proportion of shortest paths between any pair of nodes that travel through the node, which can be denoted by

$$B_p = \sum_{i,j \in V} \frac{\mu_{ij}(p)}{\mu_{ij}} \quad (2)$$

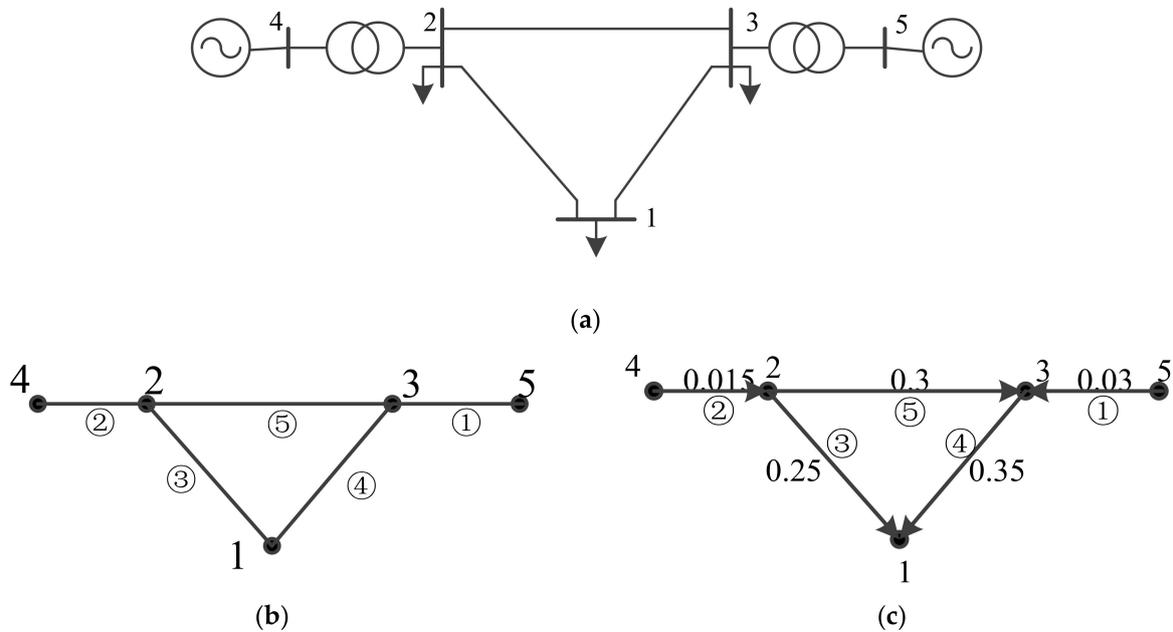
where  $V$  is a set of nodes in network, and  $i$  and  $j$  depict the node in graph  $G$ .  $\mu_{ij}$  means the total number of shortest paths between nodes  $i$  and  $j$ ;  $\mu_{ij}(p)$  means the number that pass thorough node  $p$  in the shortest path between nodes  $i$  and  $j$ .

### 2.2. Power Grid Model

Based on the complex network theory and graph theory, the actual power grid can be considered a large complex network with nodes and edges. In the power grid, the buses can be simplified as the nodes, and the transmission lines and transformer branch can be considered as the edges. Supposing the network is represented as an undirected and unweighted graph  $G = (V, E)$ , with  $n$  vertices in set  $V$  and  $m$  edges in set  $E$ , the adjacency matrix  $B_G$  can be used to define the connectivity of the graph. The elements of the matrix  $B_G$ , i.e.,  $\beta_{ij}$ , denote the weight for the edge connecting the two vertices.  $\beta_{ij} = 1$  means that there is a connection between node  $i$  and node  $j$ ;  $\beta_{ij} = 0$  denotes there is no connection between  $i$  and  $j$ .

The unweighted and undirected graph  $G$  ignores the direction of power transmission in the actual operating power grid. Therefore, according to graph  $G$  and the basic data of the power grid, we can ensure the direction of the edges. At this time, graph  $G$  can be further abstracted as a direct weighted network, denoted by  $D = (V, E, W)$ , where  $W$  is the weight vector that consists of each line reactance. Then, the adjacency matrix  $B_D$  is used to depict the direction of edges linked by two nodes.

An actual network structure of the IEEE 5-Bus system is shown in Figure 1a. The graph  $G$  and network  $D$  are obtained by using the above method, which is shown in Figure 1b,c. From Figure 1b, the degree of nodes is  $[2, 3, 3, 1, 1]^T$ , the edge betweenness is  $[3, 3, 1.5, 1.5, 1]^T$ , and the nodes ②③ and lines ①② are more important than other nodes and lines. From Figure 1c, the out-degree of nodes is  $[0, 2, 1, 1, 1]^T$ , the edge betweenness is  $[2, 3, 0.5, 1.5, 1.5]^T$ , and node ② and line ② are more important.



**Figure 1.** Model of IEEE 5-Bus system: (a) the structure diagram of the IEEE5-Bus system; (b) the graph  $G$  of the system; (c) the network  $D$  of the system.

### 2.3. PageRank Algorithm

PageRank algorithm was firstly proposed by the founders of Google in 1998. It ranks the importance or class of a webpage by the hyperlink structure of the web system [22]. At the heart of Google’s search engine, the PageRank algorithm ranks high importance when a webpage is pointed out by other important pages with high PageRank values. In addition, based on complex network theory, the web system can be abstracted as a directed graph, in which nodes correspond to each webpage and edges correspond to the hyperlinks between two webpages. The PageRank value (denoted by  $PR$ ) of webpage  $P_i$ -th is  $PR(P_i)$ , and the  $PR$  of a page is calculated as follows:

$$PR(P_i) = \sum_{P_j \in D(P_i)} \frac{PR(P_j)}{N(P_j)} \tag{3}$$

where  $D(P_i)$  is the set of pages pointing to  $P_i$ , and  $N(P_j)$  is the number of out-links from page  $P_j$ .

The  $PR(P_j)$  in (3) is unknown. Thus, an iterative process is introduced to solve the problem. The iterative formula is defined as

$$PR_{k+1}(P_i) = \sum_{P_j \in D(P_i)} \frac{PR_k(P_j)}{N(P_j)} \tag{4}$$

Assuming a web system with  $n$  pages, the  $PR$  of a page can be calculated by (3). In order to intuitively view the  $PR$  value of pages before and after iteration, we can introduce an  $n \times n$  matrix  $T$  and an  $n \times 1$  vector  $R$ . Then, (4) is written compactly as

$$R^{k+1} = TR^k \quad (5)$$

where  $R^k$  is the  $PR$  value vector of all pages after the  $k$ -th iteration.  $T$  is a row-normalized hyperlink matrix. If there is a link from  $i$  to  $j$ ,  $T_{ij} = 1/N(P_j)$ ; otherwise,  $T_{ij} = 0$ .

$T$  is a sparse matrix, and it is also a stochastic transition probability matrix. However, the matrix can have zero rows, which means that those nodes have no out-links. In other words, the web system has dangling nodes. To solve the dangling nodes in the web system and ensure convergence, a new matrix  $G$  is proposed, and (5) can be written as

$$R^{k+1} = GR^k \quad (6)$$

where

$$G = \alpha H + (1 - \alpha)ee^T/n \quad (7)$$

$$H = T + ae^T/n \quad (8)$$

In (6)–(8), the  $n \times n$  matrix  $G$  is referred to as *Google matrix*.  $\alpha \in [0, 1]$  is a parameter that controls the proportion of time and follows the hyperlinks to randomly enter a new page, which is generally 0.85.  $e$  is a column vector of all ones. The binary vector  $a$  indicates the dangling node vector. If webpage  $i$  is a dangling node,  $a_i = 1$ ; otherwise,  $a_i = 0$ .  $H$  is a transfer matrix that demonstrates the relationship and information transmission between a webpage and its linked webpages. Matrix  $H$  is the combination of matrix  $T$  and  $ae^T/n$ . Thus, the zero rows of  $T$  are replaced with  $e^T/n$ , indicating that each dangling node is connected to other nodes equally.

To ensure that the iteration process in (6) can still converge to a unique positive vector when the initial value is arbitrarily chosen, the *Google matrix*  $G$  is verified to satisfy the properties of being stochastic, irreducible and aperiodic.

### 3. Key Nodes Identification with the Modified PageRank Algorithm

#### 3.1. PageRank Algorithm Applied to the Power Grid

The PageRank algorithm was originally used to sort internet pages. According to the simplification principle of complex network theory, both internet and power grid can be simplified to a directed-weighted network model. The nodes correspond to buses, edges correspond to the transmission lines, and the connection strength between nodes is represented by the line reactance. The comparison of the directed-weighted network model, internet, and power grid topology is shown in Table 1.

**Table 1.** Comparison internet with power grid.

Directed-Weighted Network	Internet	Power Grid
Node	Web page	Bus or substation
Edge	Hyperlink	Line or transformer
Information value	Link relationship	Power flow
Degree	Number of visits	Generator or load capacity

Based on the above comparison, the power grid can be simplified to the directed-weighted network model and satisfy the application conditions of the PageRank algorithm. The PageRank algorithm can be applied to rank the importance of nodes in the power grid, but it has two shortcomings: (1) the electrical characteristics between nodes are not considered. (2) The transmission power between nodes is not distributed in equal proportion. Therefore, in order to overcome the above disadvantages, the following modified PageRank algorithm is proposed by considering node type and transmission characteristics to assess the node importance.

### 3.2. Modified Transfer Matrix

A power grid can be described as a directed-weighted network model  $D = (V, E, W)$ , the  $n \times n$  adjacency matrix  $B = \beta_{ij}$ . When node  $i$  and node  $j$  are connected together  $j$ ,  $\beta_{ij} = 1$ ; otherwise,  $\beta_{ij} = 0$ . To avoid equivalent power distribution between transmission lines in the power grid, the equivalent impedance is used to describe the connection weight between nodes [8], and it is calculated by

$$Z_{eq-ij} = Z_{ii} + Z_{jj} - 2Z_{ij} \quad (9)$$

where  $Z_{ij}$  is the  $i$ -th row and  $j$ -th column element of the impedance matrix  $Z$ .

Then, in order to describe the node influence from the view of topology, the node betweenness index is modified by combining the node degree and transmission characteristic. The weighted betweenness is obtained as follows:

$$B_w(p) = d_p \sum_{i \in GN, j \in LN} \frac{\delta_{ij}(p)}{\delta_{ij}} \quad (10)$$

where  $\delta_{ij}$  means the total number of all possible paths between generator  $i$  and load  $j$ , and  $\delta_{ij}(p)$  represents the number of through the node  $p$  in all possible paths between generator  $i$  and load  $j$ .  $d_p$  is the degree of node  $p$ .

Different node types assume different responsibilities in the power grid; the node set  $V$  can be divided into three sets: the generator node set  $GN$ , the load nodes set  $LN$ , and the contact node set  $CN$ . The corresponding modified transfer matrix of the PageRank algorithm is proposed in the following.

#### 3.2.1. Generator Nodes

Generator nodes mainly undertake the task of transmitting the power to other adjacent nodes. It is the first node of the power transmission path and guarantees uninterrupted power supply in the power grid. According to the actual power grid topology, the generator node has at least one outlet, and the node voltage and rated active power are known. In other words, the generator node is the PV node type.

To assess the importance of generator nodes, there are two aspects that should be considered: (1) how much output power the generator node transfers to other adjacent nodes; (2) the power weight value of the generator node itself. Therefore, considering the topological importance and power weight value of generator nodes, the modified transfer matrix  $H$  is defined by mapping the relationship of the adjacency matrix, shown as follows.

$$H_{ij} = \begin{cases} \beta_{ij} \frac{Z_{eq-ij}}{\sum_{k \in D_i} Z_{eq-ik}} \left( \frac{P_{Gi}}{\max(P_G)} + B_w(i) \right), & i \neq j \\ B_w(i) \frac{P_{Gi}}{\sum_{g \in GN} P_{Gg}}, & i = j \end{cases} \quad (11)$$

where  $GN$  is the set of generators;  $\beta_{ij}$  is the element of the adjacency matrix  $B$ ;  $D_i$  is the set of out-links of generator node  $i$ ;  $P_{Gi}$  is the rated power of the generator node  $i$ ;  $B_w(i)$  is the weighted betweenness of generator node  $i$ .  $\max(P_G)$  is the maximum capacity of generators in the power grid.

#### 3.2.2. Load Nodes

Load nodes mainly consume the power supplied by generators. The load power, including active power and reactive power, is known. It can be considered that the load node is the PQ node type. Based on the power grid structure and the power flow direction, there are two types of load nodes: (1) the power is transmitted to the load node and all is absorbed; (2) the power is transmitted to load node and only part of the power is adsorbed, the remaining part is transmitted to other load nodes.

For the first type, we know that those nodes only have inflow power to supply the load. Thus, only the load capacity needs to be considered. The element of the modified transfer matrix  $H$  is defined by mapping with the adjacency matrix, shown as follows.

$$H_{ii} = B_w(i) \frac{S_{Di}}{\sum_{l \in DN} S_{Dl}} \quad (12)$$

where  $DN$  is the set of load nodes, and  $S_{Di}$  is the capacity of the load node  $i$ .

For the second type, those load nodes not only consume power, but they also transmit power. Therefore, considering the topological importance and the load power, the element of the modified transfer matrix  $H$  is defined by mapping with the adjacency matrix, as given below.

$$H_{ij} = \begin{cases} \beta_{ij} B_w(i) \frac{Z_{eq-ij}}{\sum_{k \in D_i} Z_{eq-ik}} \frac{S_{Di}}{\sum_{g \in DN} S_{Dg}}, & i \neq j \\ \omega B_w(i) \frac{S_{Di}}{\sum_{g \in DN} S_{Dg}}, & i = j \end{cases} \quad (13)$$

where  $\omega = |U_{\max} - U_i| / |U_i - U_{\min}|$  is the voltage deviation, and  $U_i$  is the voltage of node  $i$ .  $U_{\max}$  and  $U_{\min}$  are the upper and lower voltage limits of nodes.

### 3.2.3. Contact Nodes

Contact nodes mainly undertake the task of power transmission and make the connection between generator and load. The total power of these nodes is always 0, and there is no power consumption. Therefore, introducing the topological information and voltage of contact nodes into the modified transfer matrix  $H$ , the elements are defined as follows.

$$H_{ij} = \begin{cases} \beta_{ij} B_w(i) \frac{Z_{eq-ij}}{\sum_{k \in D_i} Z_{eq-ik}} e^{-(U_i - U_j)}, & i \neq j \\ B_w(i) \left| \frac{U_{\max} - U_i}{U_i - U_{\min}} \right|, & i = j \end{cases} \quad (14)$$

Therefore, the modified transfer matrix  $H$  is defined by (3)–(6). The characteristics, such as the topological importance, the active power of generator, power load, and node voltage, are considered in the modified transfer matrix  $H$ . Thus, it is better and more comprehensive to represent the importance of nodes in the actual power grid.

### 3.3. Convergence of Google Matrix $G_m$

It can be seen that the modified transfer matrix  $H$  is not stochastic. That is, the sum of rows is not equal to 1. For this reason, a background node is added to make the sum of rows equal to 1 in the modified transfer matrix  $H$  [29]. The modified matrix  $M$  is shown as follows.

$$M = \begin{bmatrix} H_{n \times n} & b_{n \times 1} \\ 0_{1 \times n} & 1 \end{bmatrix} \quad (15)$$

where  $b$  is the  $n \times 1$ -dimension vector.

Based on the modified matrix  $M$ , for considering the convergence of Google matrix  $G_m$ , the two vectors  $E$  and  $V$  are defined as follows.

$$\begin{cases} E = [1]_{(n+1) \times 1} \\ V = \frac{1}{n} \begin{bmatrix} 1_{1 \times n} \\ 0 \end{bmatrix} \end{cases} \quad (16)$$

Therefore, the modified Google matrix  $G_m$  can be calculated by

$$G_m = \alpha M + (1 - \alpha) EV \quad (17)$$

Because the nodes consist of two parts, i.e., power grid node and background node, the  $PR$  value of all nodes can be calculated by

$$PR_{k+1} = G_m \cdot PR_k = G_m \cdot \begin{bmatrix} PR_{gk} \\ PR_{bk} \end{bmatrix} \quad (18)$$

where  $PR_{gk}$  is the  $PR$  value of power grid node after  $k$  iterations.  $PR_{bk}$  is the  $PR$  value of background node after  $k$  iterations.  $G_m$  is the  $(n + 1) \times (n + 1)$ -dimension modified Google matrix.

To guarantee the convergence of the modified PageRank algorithm by the modified Google matrix  $G_m$ ,  $G_m$  should satisfy the properties of being stochastic, irreducible and aperiodic.

**Property 1.**  $G_m$  is stochastic.

**Proof.** Stochastic matrix is non-negative, and the sum of each row is equal to 1 [22]. The modified Google matrix  $G_m$  in (9) is clearly satisfying the condition of  $G_m > 0$ . The sum of rows is equal to 1 because the background nodes are added in the matrix  $H$  to guarantee that the modified matrix  $M$  is stochastic, and the two vectors of matrix  $E$  and  $V$  are also stochastic. Therefore,  $G_m$  is stochastic.  $\square$

**Property 2.**  $G_m$  is irreducible.

**Proof.** Based on the conclusion in [22], a matrix is irreducible when the directed graph has strong coupling. Based on the defined (7) and (8), it can be proven that each element of the modified Google matrix  $G_m$  satisfies the condition of  $G_{m-ij} > 0$ . Therefore, the corresponding network of  $G_m$  should have strong coupling, which shows its irreducibility.  $\square$

**Property 3.**  $G_m$  is aperiodic.

**Proof.** According to [22], a primitive matrix is irreducible, which has no less than one positive diagonal element. Meanwhile, a primitive matrix must be aperiodic [22]. Based on the basic PageRank definition, the  $(n + 1)^{th}$  diagonal element of  $G_m$  is 1, and other diagonal elements of  $G_m$  also satisfy the condition  $G_{m-ii} > 0$ . Thus,  $G_m$  is primitive, implying its aperiodicity.  $\square$

## 4. Case Studies

### 4.1. Verification Indexes

With the disconnection of transmission nodes in the power grid, the power transmission path is damaged, and the power transmission capacity is affected. Network efficiency  $E$  is used to assess the network information transmission from the perspective of the entire network. It is defined as follows [30]:

$$E = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \quad (19)$$

where  $n$  represents the number of power grid nodes, and  $d_{ij}$  represents the shortest path distance between node  $i$  and node  $j$ .

When the power grid node is disturbed, the power transmission path and transmission capacity will be affected. To verify the change of transmission capacity in the power grid, the network transmission efficiency of the power grid is defined as

$$\text{Net} = \frac{E_k}{E_0} \times 100\% \quad (20)$$

where  $E_0$  is the network transmission efficiency under the normal running state, and  $E_k$  is the transmission efficiency after the  $k$ -th fault of the power grid.

Based on the modified PageRank algorithm in Equations (11)–(18), the convergence of the algorithm is also verified. According to the process of the modified PageRank algorithm, the process of identification of key nodes in the power grid is shown in Figure 2. Here, the initial PR value of each node is 1.

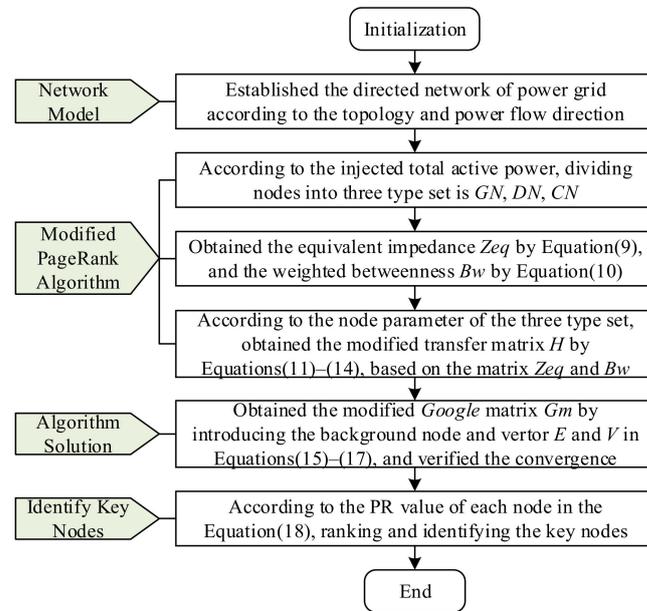


Figure 2. Flowchart of key node identification method.

#### 4.2. IEEE 39-Bus System

In this part, IEEE 39-Bus system is used as the simulation analysis system, which has 39 nodes, including 10 generator nodes and 19 load nodes. According to the capacity of the generator and power load, the 39 nodes can be divided into 9 generator nodes, 20 load nodes, and 10 contact nodes. The corresponding directed-weighted network model is shown in Figure 3.

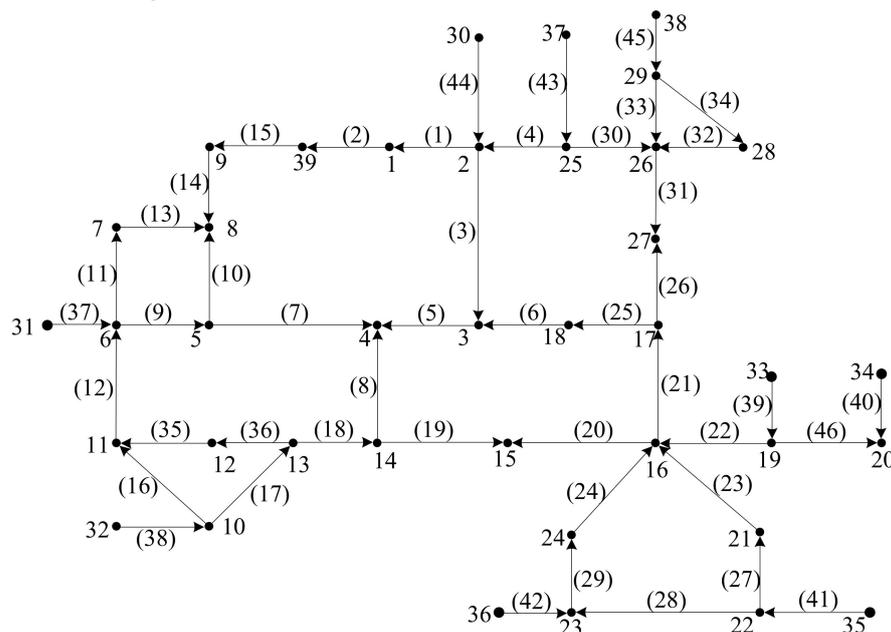


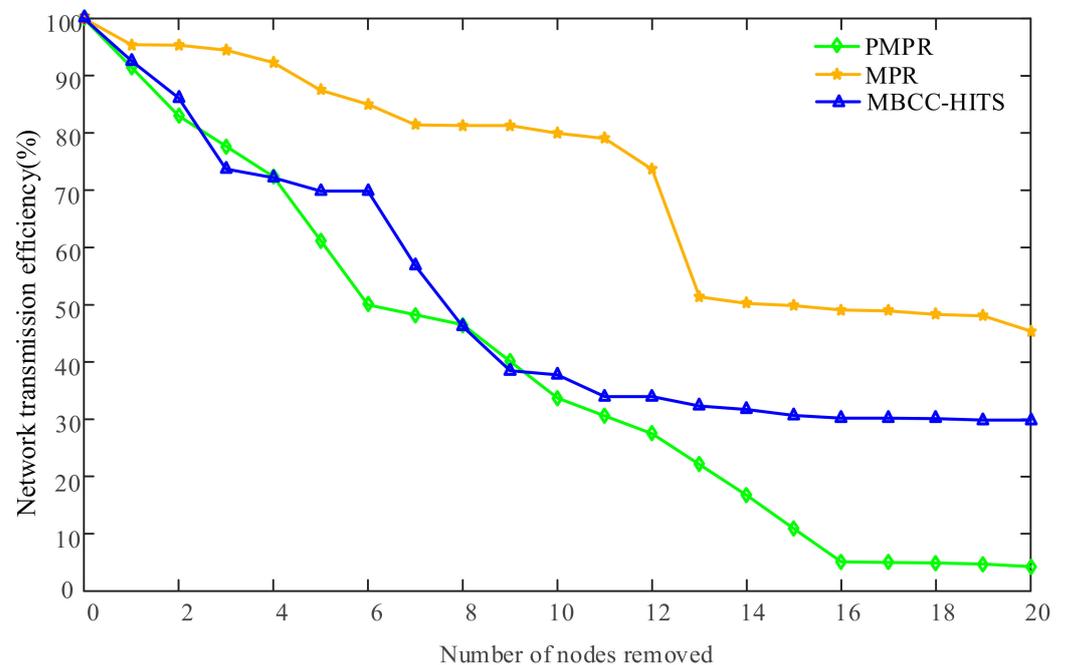
Figure 3. IEEE 39-Bus system directed-weighted network.

For verifying the accuracy of the modified PageRank algorithm, the PR values are calculated, and the key nodes are identified in the simulation test system. According to the sorting of PageRank values, the first 10 key nodes are picked to compare with the proposed modified PageRank algorithm (PMPR), basic PageRank algorithm (BPR), the combination of random matrix and entropy theory (CRMET), and the node importance index method (NII), which are shown in Table 2.

**Table 2.** Key nodes identified by different methods.

Ranking	Node Importance Ranking			
	PMPR	BPR	CRMET	NII
1	6	8	6	16
2	29	4	5	4
3	23	16	11	12
4	10	27	8	26
5	22	15	10	3
6	20	3	7	11
7	19	26	13	15
8	25	17	14	5
9	7	9	4	19
10	21	24	15	14

For verifying the validity of the modified PageRank algorithm for key node identification, an intentional attack is used for simulation in the normal operation test system. The identified first 10 key nodes are removed in turn by PMPR, BPR, CRMET, and the NII method. The network transmission efficiency values are shown in Figure 4.



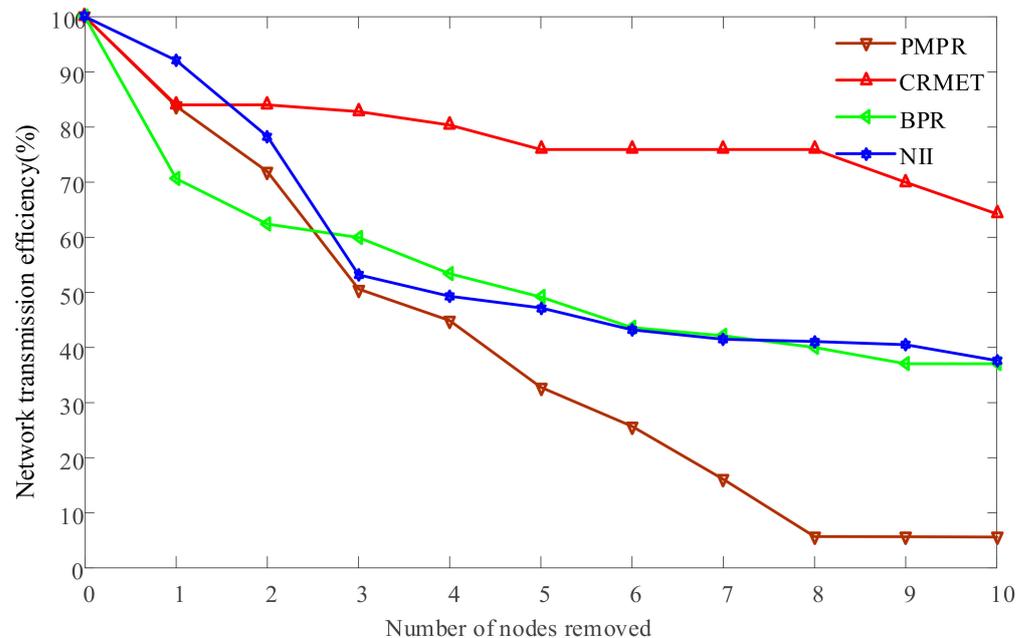
**Figure 4.** Network transmission efficiency with intentional attack in IEEE 39-Bus system.

From the simulation results, when all the 10 key nodes are removed, the network transmission efficiency of the power grid is lower with PMPR than that with CRMET and NII methods. The network transmission efficiency of the grid is 64.23% by the CRMET and 37.06% by NII method. Compared with the network transmission efficiency of the BPR and the PMPR, the network transmission efficiency of BPR is lower than that of PMPR by the first two nodes removed. Because node 4 and node 8 all have three ingoing lines and have 500 MW and 520 MW loads, respectively, these two nodes are removed and

may greatly influence the power grid more than the PMPR method, which removed nodes 6 and 29, from the network transmission efficiency perspective. However, nodes 6 and 29 are connected to generators 31 and 38, which serve as the key nodes for the generator to transmit power, from the perspective of node types; nodes 6 and 29 are more important than nodes 4 and 8. After removing all the 10 identified key nodes by the BPR, the network transmission efficiency was 37.64%. However, after removing all the 10 identified key nodes by the PMPR, the network transmission efficiency was as low as 5.62%. The results show that the identified 10 key nodes by the PMPR have more influence on the transmission capacity than the other three methods.

#### 4.3. IEEE 118-Bus System

In the static deliberate attack mode, the IEEE118-Bus system is also carried out to assess the node importance. The identified key nodes are removed, in turn, by the proposed modified PageRank algorithm (PMPR); the modified PageRank algorithm (MPR), which takes the importance of nodal load, nodal load capacity, and network topology into account [23] but does not consider the node types; and the model based on co-citation (MBCC)-hypertext induced topic selection (HITS) algorithm (MBCC-HITS) [25]. The corresponding transmission efficiency values are shown in Figure 5.

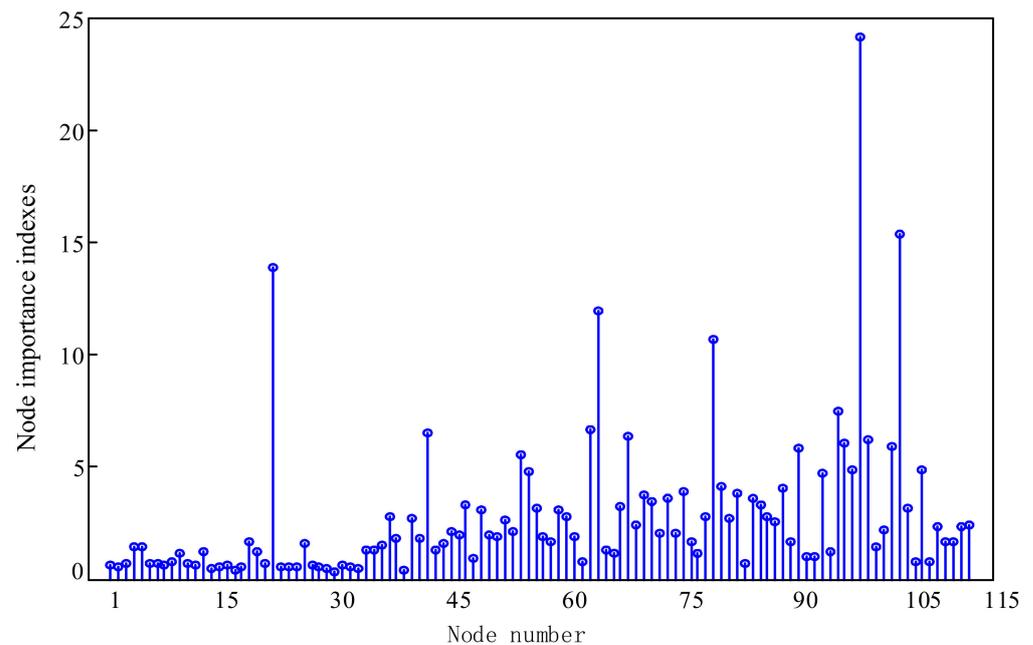


**Figure 5.** Network transmission efficiency with intentional attack in IEEE 118-Bus system.

As shown in Figure 5, the transmission efficiency of the PMPR is larger than the MBCC-HITS method when the first 3 or 9 nodes are removed only. Because the 3rd node is node 49 in the MBCC-HITS method, node 49 is connected to the outgoing line of node 66, which is connected to a load and a generator. The 9th node is node 92, which is connected to the generator node. Removing them may have more influence on the power grid from the network transmission efficiency perspective. However, the network transmission efficiency indexes of the grid reduced to 45.4, 29.84, and 5.12% after all the first 20 identified key nodes were removed by MBCC-HITS, MPR, and PMPR. The reduction in network transmission efficiency with PMPR was greater compared to the MPR and MBCC-HITS method. In other words, the identified key nodes by PMPR had the greatest impact on the power grid. Therefore, compared with the other two methods, the PMPR proposed in this paper takes topological information, node type, and operating characteristics into account comprehensively, and it is more accurate for identifying key nodes in the power grid.

#### 4.4. A Provincial Power Grid Test System

For verifying the practicability of the proposed modified PageRank algorithm, key nodes were identified in an actual operation power grid. The directed-weighted model of a provincial power grid simulation system in China was established, where the corresponding electrical parameters were known. The node importance index values of the provincial power grid were obtained as shown in Figure 6. According to the descending order, the first 15 non-generator nodes were selected as the key nodes of the system, as shown in Table 3.



**Figure 6.** Node importance index value in a provincial power grid test system.

**Table 3.** Key nodes in the provincial power grid test system.

Ranking	Node Importance Ranking		
	Node Number	Substation	Voltage/kV
1	98	YX1	330
2	103	MY1	330
3	64	NL1	330
4	79	LF1	330
5	95	YD1	330
6	63	CY1	330
7	42	YH1	750
8	68	XY1	750
9	99	ZY1	330
10	96	GY1	330
11	102	XS1	330
12	90	XX1	330
13	54	HL1	330
14	106	TY1	330
15	97	HZ1	330

The key nodes shown in Table 3 have two 750 kV substations in the top 15 key nodes. It means that they mainly undertake power transmission and conversion, and they play a pivotal role in the provincial power grid. Removal of the nodes will make the power grid split, and the new energy output power in a certain area cannot be fully consumed, which will reduce the utilization rate of new energy. The remaining key nodes are 330 kV

substations, among them, nodes 103, 64, 79, 63, 54, and 106 are all connection nodes for the power output of thermal power plants. The removal of nodes will lead to the separation of thermal power plants from the main grid and even result in power imbalance in the grid. In conclusion, simulation results verify that the identified key nodes of the proposed modified PageRank algorithm are reasonable and practical.

## 5. Discussion and Conclusions

By analyzing the complex network theory and the famous PageRank algorithm, a modified PageRank algorithm is presented to assess the node importance of a power grid. The proposed method comprehensively takes the following factors into account: network topology, nodal type, and state transmission information. Theoretical analysis and case studies show the necessity of considering these factors to identify the key power grid nodes and the validity of the modified PageRank algorithm. However, it is not comprehensive to construct the modified transfer matrix by the proposed three indexes only. It needs to be analyzed by combining with the operation characteristics of the actual power system. In addition, it is necessary to study the distributed parallel computation to improve the fast solution of the modified PageRank algorithm from a technical point of view. Finally, the evolution process of the key nodes in power system needs further analysis to reveal its mechanism. Nevertheless, the following conclusions are still drawn from our study.

Simulation results indicate that the proposed method, by considering the node type and topological information, is more feasible and unbiased than the PageRank algorithm without considering electrical characteristics. The network transmission efficiency of the grid is as low as 5.62% with the proposed algorithms, compared with 64.23%, 37.06%, and 37.64% for MBCC-HITS, MPR, and PMPR, respectively, in IEEE 39-Bus system. The network transmission efficiency indexes of the grid reduced to 45.4, 29.84, and 5.12% by the MBCC-HITS, MPR, and PMPR, respectively, in IEEE 118-Bus system.

The validity of the key power nodes is verified. The proposed modified algorithm is more accurate and comprehensive, and it can be easily applied to the practical power grid based on the practical application of a provincial power grid test system. Based on the above research and analysis, the proposed method in this paper is helpful to prevent the occurrence of cascading failure in a power system and provide actual reference for power system operators. At the same time, the research can also be used to identify the key nodes of a power system with renewable energy sources and AC/DC hybrid power grids in future research.

**Author Contributions:** Conceptualization, D.Z. and H.W.; methodology, R.W.; software, J.B.; validation, D.Z., H.W. and R.W.; formal analysis, R.W.; investigation, H.W.; resources, D.Z.; data curation, H.W.; writing—original draft preparation, R.W.; writing—review and editing, D.Z.; visualization, H.W.; supervision, D.Z. and J.D.; project administration, D.Z. and J.D.; funding acquisition, D.Z. and J.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by China Postdoctoral Science Foundation, grant number 2020M683685XB; Natural Science Basic Research Plan in Shanxi province of China, grant number 2020JQ-633; and National Natural Science Foundation of China, grant number 51877174.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Arianos, S.; Bompard, E.; Carbone, A.; Xue, F. Power grid vulnerability: A complex network approach. *Chaos Interdiscip. J. Nonlinear Sci.* **2009**, *19*, 013119. [[CrossRef](#)]
2. Andersson, G.; Donalek, P.; Farmer, R.; Hatziaargyriou, N.; Kamwa, I.; Kundur, P.; Martins, N.; Paserba, J.; Pourbeik, P.; Sanchez-Gasca, J.; et al. Causes of the 2003 major grid blackouts in north America and Europe, and recommended means to improve system dynamic performance. *IEEE Trans. Power Syst.* **2005**, *20*, 1922–1928. [[CrossRef](#)]
3. Rampurkar, V.; Pentayya, P.; Harivittal, A.; Faruk, K. Cascading failure analysis for Indian power grid. *IEEE Trans. Smart Grid* **2016**, *7*, 1951–1960. [[CrossRef](#)]
4. Ren, H.P.; Song, J.; Yang, R.; Baptista, M.S.; Grebogi, C. Cascade failure analysis of power grid using new load distribution law and node removal rule. *Phys. A Stat. Mech. Appl.* **2016**, *442*, 239–251. [[CrossRef](#)]
5. Fan, W.L.; Liu, Z.G.; Hu, P.; Mei, S.W. Cascading failure model in power grids using the complex network theory. *IET Gener. Transm. Distrib.* **2016**, *10*, 3940–3949.
6. Moger, T.; Dhadbanjan, T. A novel index for identification of weak nodes for reactive compensation to improve voltage stability. *IET Gener. Transm. Distrib.* **2015**, *9*, 1826–1834. [[CrossRef](#)]
7. Hu, Z.; Li, X.Y. The Degree of Coupling Analysis Method Based on Current Injection. In Proceedings of the 2016 International Symposium on Fundamentals of Electrical Engineering, Bucharest, Romania, 30 June–2 July 2016; pp. 1–6.
8. Li, Z.H.; Duan, D.L. Node Importance of Complex Networks Based on Cascading Failure Dynamic. In Proceedings of the 2016 International Conference on Intelligent Networking and Collaborative System (INCoS), Ostravva, Czech Republic, 7–9 September 2016; pp. 31–35.
9. Ghasemi, S.; Mohammadi, M.; Moshtagh, J. A new look-ahead restoration of critical loads in the distribution networks during blackout with considering load curve of critical loads. *Electr. Power Syst. Res.* **2021**, *191*, 106873. [[CrossRef](#)]
10. Liu, B.; Li, Z.; Chen, X.; Huang, Y.; Liu, X. Recognition and vulnerability analysis of key nodes in power grid based on complex network centrality. *IEEE Trans. Circuits Syst. II Express Briefs* **2018**, *65*, 346–350. [[CrossRef](#)]
11. Li, J.; Dueñas-Osorio, L.; Chen, C.; Berryhill, B.; Yazdani, A. Characterizing the topological and controllability features of US power transmission networks. *Phys. A Stat. Mech. Appl.* **2016**, *453*, 84–98. [[CrossRef](#)]
12. Wang, K.; Zhang, B.H.; Zhang, Z.; Yin, X.G.; Wang, B. An electrical betweenness approach for vulnerability assessment of power grids considering the capacity of generators and load. *Phys. A Stat. Mech. Appl.* **2011**, *390*, 4692–4701. [[CrossRef](#)]
13. Wu, D.; Ma, F.; Javadi, M.; Thulasiraman, K.; Bompard, E.; Jiang, J.N. A study of the impacts of flow direction and electrical constraints on vulnerability assessment of power grid using electrical betweenness measures. *Phys. A Stat. Mech. Appl.* **2017**, *466*, 295–309. [[CrossRef](#)]
14. Bompard, E.; Pons, E.; Wu, D. Extended topological metrics for the analysis of power grid vulnerability. *IEEE Syst. J.* **2012**, *6*, 481–487. [[CrossRef](#)]
15. Yan, J.; He, H.; Sun, Y. Integrated security analysis on cascading failure in complex networks. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 451–463. [[CrossRef](#)]
16. Adebayo, I.; Jimoh, A.A.; Yusuff, A. Voltage stability assessment and identification of important nodes in power transmission network through network response structural characteristics. *IET Gener. Transm. Distrib.* **2017**, *11*, 1398–1408. [[CrossRef](#)]
17. Nasiruzzaman, A.B.M.; Pota, H.R.; Akter, M.N. Vulnerability of the large-scale future smart electric power grid. *Phys. A Stat. Mech. Appl.* **2014**, *413*, 11–24. [[CrossRef](#)]
18. Yu, H.; Cao, X.; Liu, Z.; Li, Y.J. Identifying key nodes based on improved structural holes in complex networks. *Phys. A Stat. Mech. Appl.* **2017**, *486*, 318–327. [[CrossRef](#)]
19. Lin, Z.; Wen, F.; Wang, H.; Lin, G.; Mo, T.; Ye, X. CRITIC-based node importance evaluation in skeleton-network reconfiguration of power grids. *IEEE Trans. Circuits Syst. II Express Briefs* **2017**, *65*, 206–210. [[CrossRef](#)]
20. Da Silva, A.M.L.; Jardim, J.L.; de Lima, L.R.; Machado, Z.S. A method for ranking critical nodes in power networks including load uncertainties. *IEEE Trans. Power Syst.* **2016**, *31*, 1341–1349. [[CrossRef](#)]
21. Alayande, A.S.; Jimoh, A.A.G.; Yusuff, A.A. Identification of critical elements in interconnected power networks. *Iran. J. Sci. Technol. Trans. Electr. Eng.* **2020**, *44*, 197–211. [[CrossRef](#)]
22. Langville, A.N.; Meyer, C.D. *Google's Pagerank and Beyond: The Science of Search Engine Rankings*; Princeton University Press: Princeton, NJ, USA, 2011.
23. Li, C.; Liu, W.; Cao, Y.; Chen, H.; Fang, B.; Zhang, W.; Shi, H. Method for evaluating the importance of power grid nodes based on PageRank algorithm. *IET Gener. Transm. Distrib.* **2014**, *8*, 1843–1847. [[CrossRef](#)]
24. Ma, Z.Y.; Shen, C.; Liu, F.; Mei, S.W. Fast screening of vulnerable transmission lines in power grids: A PageRank-based approach. *IEEE Trans. Smart Grid* **2019**, *10*, 1982–1991. [[CrossRef](#)]
25. Wang, H.; Shan, Z.; Ying, G.; Zhang, B.; Zou, G.; He, B. Evaluation method of node importance for power grid considering inflow and outflow power. *J. Mod. Power Syst. Clean Energy* **2017**, *5*, 696–703. [[CrossRef](#)]
26. Zhou, L.M.; Sheng, W.X.; Su, J.; Liu, W.; Zhang, J.; Shang, Y.W.; Liu, S.P. A calculation method and system of a distribution network health index. China Patent 201911391054.4, 30 December 2019.
27. Guan, W.; Wen, X.; Wang, L.; Lu, Z.; Shen, Y. A service-oriented deployment policy of end-to-end network slicing based on complex network theory. *IEEE Access* **2018**, *6*, 19691–19701. [[CrossRef](#)]

28. Bompard, E.; Wu, D.; Xue, F. The concept of betweenness in the analysis of power grid vulnerability. In Proceedings of the 2010 Complexity in Engineering, Rome, Italy, 22–24 February 2010; pp. 52–54.
29. Buzzanca, M.; Carchiolo, V.; Longheu, A.; Malgeri, M.; Mangioni, G. Black hole metric: Overcoming the pagerank normalization problem. *Inf. Sci.* **2018**, *438*, 58–72. [[CrossRef](#)]
30. Zhou, X.; Zhang, F.M.; Li, K.W.; Hui, X.B.; Wu, H.S. Finding vital node by node importance evaluation matrix in complex networks. *Acta Phys. Sin.* **2012**, *61*, 050201. [[CrossRef](#)]