sustainability

MDPI

Article

Use of Variable Fuzzy Clustering to Quantify the Vulnerability of a Power Grid to Earthquake Damage

Tianhua Li *, Yanchao Du and Yongbo Yuan

Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China; Duyanchao@mail.dlut.edu.cn (Y.D.); yongbo@dlut.edu.cn (Y.Y.)

* Correspondence: litianhua@mail.dlut.edu.cn

Received: 27 July 2019; Accepted: 8 October 2019; Published: 12 October 2019

Abstract: The power grid is a critical component of city infrastructure. If it is damaged by an earthquake, there can be a huge impact on the safety and well-being of society and individuals. Identifying nodes in the grid that are highly vulnerable to earthquake damage is significant for effective pre-earthquake damage prevention, emergency response, and post-earthquake relief. Three indicators, the probability of node disconnection, the node hierarchical level, and the node critical threshold, were chosen, and their combined ability to represent node vulnerability to damage from an earthquake event was analyzed. A variable fuzzy clustering model was used to classify and order the nodes in the grid. The 20-node power grid of a city was used as an example to show how highly vulnerable nodes were identified, and how the reasons for the high vulnerability of these nodes were drawn out of the analysis. Countermeasures were given to reduce network vulnerability. The variable fuzzy clustering method used in this paper offers a new perspective on network vulnerability, and it quantifies the vulnerability of grid nodes more comprehensively than existing methods of assessing grid vulnerability. This research is significant as a baseline reference for future studies of grid vulnerability.

Keywords: power grid; earthquake; vulnerability analysis; variable fuzzy clustering model

1. Introduction

The power grid is an important urban infrastructure that supports the regular operation of a city and ensures the normal functioning of people's daily lives. Urban development increases both dependence on the power grid and the magnitude of loss and damage caused by large-scale failures. Earthquakes pose the greatest threat of all natural disasters to a power grid and can entirely disrupt it. The prevention of damage to the grid due to earthquakes can maintain the safety and sustainability of modern society. Measures to deal with disasters can be divided into pre-disaster damage prevention and preparation, and post-disaster rescue and repair. The academic community believes that the former is superior to the latter and has proposed building a Culture of Prevention. Although post-disaster rescue and repair is always necessary, it is not enough to respond to a disaster only after it happens because the response is costly and its effects are temporary; pre-disaster damage prevention and preparation can contribute to more lasting security. Vulnerability assessment is a major earthquake damage prevention measure for power grids [1].

The concept of vulnerability was first used in international political economy to explain dependency [2]. It was subsequently introduced into the natural sciences and engineering to describe the state of a system and its components that were vulnerable to damage or exposure. Vulnerability is universal. Almost all systems have some degree of vulnerability. The manner of conducting a vulnerability analysis is a key issue when considering how to ensure stability in the operation of a system. We investigated the vulnerability of power grid nodes to earthquake events by finding weak

links in the power grid in order to reduce vulnerability by improving the seismic grade of power facilities and reducing the risk of seismic damage to key nodes before an earthquake event. Our research has further significance in guiding post-earthquake rescue and repair strategies.

There is no single agreed-on concept of power grid vulnerability and no perfect vulnerability assessment method or widely accepted set of indicators. Vulnerability is influenced by internal and external factors. The latter include natural disasters and damage caused by humans, and the former include electrical component failures and convoluted power grid topology. When scholars of different disciplines study the vulnerability of power grid nodes, they often analyze from their own professional perspective, and the research focus varies. Some literature is concerned with the response of mechanical properties of electric power facilities to seismic activity and the capacity of the facilities to resist structural damage [3–5]. Some literature describes the functional modeling of electrical substations as a method of studying how earthquakes damage these structures. The authors found that component damage caused by the earthquake led to a short circuit, propagated inside the substation and to surrounding substations, and could thus cause a system-wide failure [6,7]. Some literature uses the graph theory to analyze the relationship between network topology and reliability from the perspective of complex networks. This kind of research can be divided into a pure model and an extended model according to different abstract forms of the power grid. The former abstracts the power grid into a pure network, focusing on the influence of topological characteristics (such as degree, proximity, betweenness, clustering coefficient, etc.) on the network, which has been applied in studying the failure mechanisms of the U.S. power grid [8–10], the European power grid [11], and the Italian power grid [12]. The latter pays more attention on the electrical performance of the grid, incorporating the characteristics of electrical components, such as impedance, power, and component capacitance, into the network model. Compared with the pure model, the extended model is more in line with the physical characteristics of the grid [13–16], but because of its computational complexity, the extended model is not practical for large networks and complex situations. Some literature has focused on the performance of the power grid. Researchers have analyzed power flow and derived performance indicators that can be used to judge the state of the power grid network [17–19]. The aforementioned research examines vulnerability from different perspectives. However, the vulnerability of a power grid is multi-faceted and complex, so it must be assessed comprehensively. We introduce the variable fuzzy clustering algorithm, which we use to quantify the vulnerability of a power grid by analyzing characteristics of the nodes.

We selected two types of indicators to measure the vulnerability of power grid nodes: Structural vulnerability indicators and functional vulnerability indicators. The former are concerned with vulnerability due to network topology, and the two indicators used are the hierarchical level of each node and the critical threshold. The indicator of functional vulnerability is the service characteristic indicator, also known as the probability of node disconnection. We justify the direction of our research and the choice of indicators as follows.

(1) The common topological indicators used to quantify vulnerability are node degrees and betweenness [20,21], but the research literature [22,23] shows that the conclusions obtained from the use of the degree indicator are one-sided. In general, the greater the degree of a node, the greater its vulnerability. However, some special nodes in the network, such as bridge nodes, although small in degree, are very vulnerable. Use of the betweenness indicator requires a holistic understanding of the network and its information, which is often difficult to obtain. Therefore, we use the critical threshold and the hierarchical level indicators to measure the node vulnerability from a topological perspective.

(2) Under normal conditions, the performance of the power grid should be calculated from the power flow in the network, from which the power distribution, voltage, and other performance indicators can be obtained. However, if there is earthquake damage, the quantitative relationship between the failure probability of high-voltage electrical equipment and the power flow loss is extremely difficult to determine. To facilitate research, this calculation is generally replaced by network connectivity analysis. Node disconnection probability is used instead of the power performance

indicator. If the node is connected, the node power performance is considered to be normal. This substitution is acceptably accurate for earthquake-related vulnerability analysis [24].

(3) The boundary between structural vulnerability indicators and functional vulnerability indicators is not inflexible when indicators are selected. For example, the critical threshold indicator is derived from the cascading failure model of complex networks. This model represents the process of cascading failures caused by load redistribution and includes the functional influence of nodes. Thus, the critical threshold represents a certain functional attribute [22,23]. Therefore, we combined the critical threshold and node disconnection probability for a more comprehensive indicator instead of undertaking power flow analysis of the grid to take into account the performance and topology of the network.

In the past, the research on the power grid was conducted from different angles, and a single index was selected to evaluate the vulnerability of the nodes. For example, the literature [25,26] uses the probability of node failure under earthquakes; the literature [27,28] uses the node degree or the power-based degree; the literature [8,12,29,30] uses the node betweenness or electrical betweenness; the literature [31] uses the node electrical centrality. The above-mentioned indicators assuredly reflect the degree of node vulnerability to some extent, but as vulnerability is a rather complex problem, a single indicator fails to comprehensively measure the node vulnerability. The main contribution of this paper is to use the idea of clustering and select relevant indicators from two aspects (internal factors, i.e., the characteristics of the power grid; external factors, i.e. the impact of earthquake effects) to comprehensively evaluate the vulnerability of power grid nodes.

The rest of this paper is organized as follows. In Section 2, the reasons for using a variable fuzzy clustering model, and the advantages it offers, are presented. In Section 3, the specific calculations for each of the three indicators are given. In Section 4, the algorithm flow of the variable fuzzy clustering model is introduced. In Section 5, the grid of a particular city is used as an example to show how the vulnerabilities of grid nodes are classified and sorted. In Section 6, through a discussion and analysis of the results, the reasons for the high vulnerability of certain nodes are identified, and some targeted measures are proposed to reduce the impact of vulnerabilities at such nodes. In the final section, the research described in this paper is summarized, and the direction of future research is outlined.

2. Methodology

We pioneer the use of a clustering algorithm, which is a data mining technique, to classify vulnerability. This methodology was inspired by the Walmart beer and diaper story. Walmart executives found, by analyzing sales data, that two completely unrelated products, beer and diapers, were often sold at the same time. Research showed that in a family with a newborn, the mother takes care of the baby, and the father is responsible for the purchase of diapers. However, when the father purchases diapers, he often also buys beer. The beer–diaper association is difficult to understand at first sight, but it follows a pattern that can be identified through data mining and analysis.

In a power grid, a single indicator cannot accurately quantify the vulnerability of a node. Only by considering a number of indicators can we comprehensively evaluate the vulnerability. Using a clustering algorithm to find the vulnerability, using data that represents node characteristics, therefore appears to be a fruitful approach. Early clustering algorithms, such as hierarchical clustering and k-means clustering, strictly classify data objects into certain well-defined categories [32], but in many problems that are encountered, the boundaries between categories are vague or ill-defined. This is the case with power grid vulnerability. There is no clear boundary between high and low vulnerability, so it may not be reasonable to use older clustering algorithms for classification. In the 1960s, when Zadeh introduced the concept of a fuzzy set [33], fuzzy set theory was used in clustering problems in fuzzy clustering analysis. In 1984, Bezdek developed the c-means fuzzy clustering algorithm (FCM) [34], which is used extensively. We propose a variable fuzzy clustering model, which improves on FCM.

To study the vulnerability of the power grid to earthquake damage, it is necessary to classify the nodes and to evaluate different types of vulnerability to develop targeted prevention measures. FCM

can only categorize the sample nodes and does not quantify the characteristics of the categories. The variable fuzzy clustering model we introduce in this paper can assess the vulnerability of each category by improving the FCM algorithm.

3. Calculation of Node Vulnerability Indicators

3.1. Node Disconnection Probability

Analysis of power grid connectivity examines the probability of disconnection between a node and the source node. A greater probability of failure indicates a higher vulnerability. The analysis determines whether there is a path between the node and the source node. To calculate the probability of a node being disconnected, the adjacency matrix of the network is first established, and the disconnection probability is obtained using a Monte-Carlo simulation [35,36].

To analyze the reliability of network connectivity, the adjacency matrix A, which describes the connections between nodes in the network, is determined by:

$$A = [a_{ij}] \; a_{ij} \begin{cases} 1, & \text{if node } i \text{ is connected to node } j \\ 0, & \text{if node } i \text{ is not connected to node } j \end{cases} \tag{1}$$

Figure 1 is a schematic of a small network, and its network adjacency matrix A is shown in Equation (2).

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \tag{2}$$

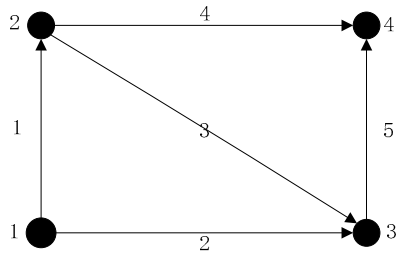

Figure 1. Small network schematic diagram.

After obtaining the adjacency matrix, we construct a judgment matrix that determines whether nodes are connected:

$$M = I + A + A^2 + A^3 + \cdots + A^{n-1} \tag{3}$$

where I is an identity matrix of order n; n is the number of network nodes. If an element in M is nonzero, then $m_{ij} > 0$, and the nodes i and j are connected; if $m_{ij} = 0$, the nodes i and j are disconnected.

The steps for calculating the Monte-Carlo simulation are:

(1) The node failure probability p_i of the node when there is an earthquake is required;
(2) The node generates a random number δ_i in [0, 1] that is then compared with the failure probability p_i of the node. If $\delta_i > p_i$, the node i operates normally; otherwise, node i fails and the edge (or link) connected to i also fails. If i fails, the adjacency matrix A is modified so that the elements in row i and column i become 0; that is, $A[i,:] = 0$, $A[:,i] = 0$;
(3) The modified adjacency matrix A is used to calculate the judgment matrix M. For all elements m_{ij} in M, if $m_{ij} > 0$, the nodes i and j are connected; if $m_{ij} = 0$, the nodes i and j are disconnected;

(4) The connectivity state after each simulation is included in the matrix T,

$$t_{ij}^{(k+1)} = t_{ij}^{(k)} + \begin{cases} 1 & m_{ij} \geq 1 \\ 0 & m_{ij} = 0 \end{cases}; \quad (4)$$

(5) Repeat steps 2–4. When the number of simulations k is large enough, Equation (5) can be used to calculate the node disconnection probability,

$$p_{ij} = 1 - \frac{t_{ij}}{k}. \quad (5)$$

3.2. Hierarchical Level

The hierarchical level G_i indicates the ability of a node to influence the entire transmission network. It is intuitive that when the source node is destroyed, the effect on the whole network is the greatest, and the vulnerability is the highest. Therefore, nodes with different degrees of contact with the source node should have different vulnerability indicator values. We set the hierarchical level of the source node to 1, the hierarchical level of nodes directly connected to it was 2, and so on. The hierarchical levels can be completely calibrated using a step-by-step search of the adjacency matrix A as follows.

(1) The source node is determined. In Figure 1, for example, node 1 is the source node. In the first level search, the first row vector (0, 1, 1, 0), corresponding to node 1 in A, is searched, and elements with value 1 are found at a_{12} and a_{13}. This gives the hierarchical levels of nodes 2 and 3, $G_2 = G_3 = 2$. The elements in the row vector and the column vector corresponding to node 1 are changed to 0, which means that the connection between node 1 and the network is destroyed, and the first level search ends (Figure 2).

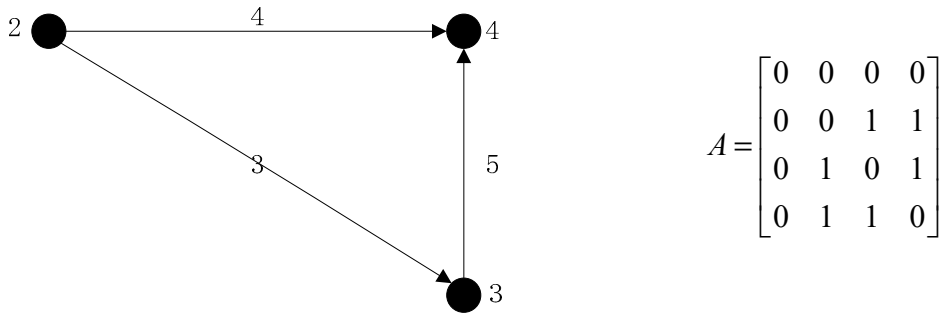

Figure 2. The network after step 1 has been completed.

(2) The second level search starts at the nodes of the second hierarchical level. The row vectors (0, 0, 1, 1) and (0, 1, 0, 1), corresponding to nodes 2 and 3, are searched to find that $a_{23} = 1$, $a_{24} = 1$, $a_{32} = 1$ and $a_{34} = 1$. Thus, the hierarchical level of node 4, $G_4 = 3$, is known. When different hierarchical levels are obtained for the same node, the higher-ranking level is chosen. For example, in the second search $a_{23} = 1$ and $a_{32} = 1$, which indicates that the hierarchical levels of nodes 2 and 3 are 3, but their hierarchical level given by the first search is 2, so the hierarchical level of nodes 2 and 3 is taken to be 2. The elements in the row vector and the column vector corresponding to nodes 2 and 3 are changed to 0, which means that nodes 2 and 3 and the network are disconnected, and the second level search ends (Figure 3).

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 3. The network after step 2 has been completed.

(3) When all the elements in the adjacency matrix A become zero, all nodes have been calibrated, and the search ends.

3.3. Critical Threshold

In a complex network, when a node fails, the load carried by that node is redistributed to the surrounding nodes through the couplings of points and edges. When a nearby node accepts the load, its capacity may be exceeded (capacity is the ultimate load of the node), and it fails. Continuing load redistribution may cause a chain reaction, leading to a partial or complete network collapse. This phenomenon is known as a cascading failure in a complex network. In the literature [23,37], a model for cascading failure based on the local characteristics of nodes was developed that introduced the concept of a critical threshold. The critical threshold is used to measure the probability of a cascading failure when a node fails.

The model referred to use the product of the node degree and its adjacent node degree (the node degree refers to the number of other nodes connected to a certain node) to measure the node load. When a node fails, the load at that node is redistributed according to the proportion of the initial load borne by the surrounding nodes. The critical threshold can be obtained by determining if surrounding nodes will initiate cascading failure. As the critical threshold value increases, the probability of cascading failure increases, and the vulnerability is increased. The steps in the calculation of the critical threshold are as follows.

(1) The initial load of each node is calculated. As shown in Figure 4, the initial load is the product of the node degree and its adjacent node degree (Equation (6)):

$$L_i = (k_i \sum\nolimits_{m \in \Gamma_i} k_m)^{\alpha} \tag{6}$$

where α is a variable parameter that controls the strength of the initial load of the node. In a study of the cascading failures of the North American power grid, Wang and Rong [10] found that when α was 0.7, the nework had the strongest robustnesss against cascading failures, so we used $\alpha = 0.7$ in this study.

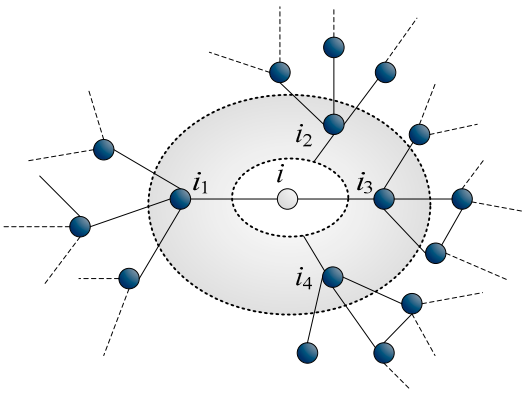

Figure 4. The effect of node local characteristics on the load.

(2) When the initial node i fails, the load is redistributed to the adjacent nodes, as shown in Figure 5. The distribution is governed by Equation (7).

$$\Pi_j = \frac{\left[k_j \sum_{m \in \Gamma_j} k_m\right]^\alpha}{\sum_{n \in \Gamma_i}\left[k_n \sum_{f \in \Gamma_n} k_f\right]^\alpha} \tag{7}$$

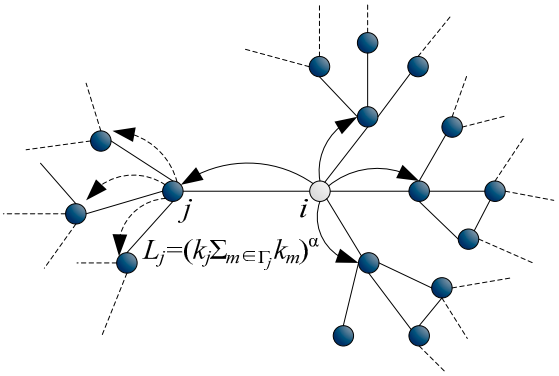

Figure 5. Load redistribution for the failed node.

The incremental load on the adjacent node j is:

$$\Delta L_{ji} = L_i \frac{\left[k_j \sum_{m \in \Gamma_j} k_m\right]^\alpha}{\sum_{n \in \Gamma_i}\left[k_n \sum_{f \in \Gamma_n} k_f\right]^\alpha} = L_i \frac{L_j}{\sum_{n \in \Gamma_i} L_n} \tag{8}$$

(3) To determine whether node j will experience a cascading failure, the node capacity C_j is set to TL_j, where T is the load that will be borne at the node (T is a percentage value somewhere between the initial load and the ultimate load capacity of the node). A greater value of T indicates that a greater maximum load can be carried, but also that the initial investment cost was higher. To avoid cascading failure, the sum of the initial load and the increased load at node j should be less than the ultimate load that the node can carry:

$$C_j = TL_j > L_j + \Delta L_{ji} \tag{9}$$

Equation (8) is substituted into Equation (9) to give:

$$T > 1 + \frac{\left(k_i \sum_{m \in \Gamma_i} k_m\right)^\alpha}{\sum_{n \in \Gamma_i}\left(k_n \sum_{f \in \Gamma_n} k_f\right)^\alpha} = T_C \tag{10}$$

where T_C is the minimum value at which node j does not fail after node i fails. Equation (10) shows that when i fails, the value of T_C is independent of node j. Nodes adjacent to node i have the same T_C value because it depends only on the load on node i and its adjacent nodes. As T_C increases, the value of T for adjacent nodes will increase, and it becomes more likely that cascading failures will occur. Conversely, the probability of cascading failure is reduced as T_C decreases. We used T_C as an indicator of network vulnerability, which was the corresponding threshold after the failure of node i.

4. Variable Fuzzy Clustering Model

4.1. Algorithm Flow of Variable Fuzzy Clustering Model

With a sample size of n, the number of selected metrics is m. The sample index data are listed in matrix X:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \tag{11}$$

where x_{ij} is the data of the ith indicator of sample j, and $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$.

The data are normalized because the dimensions of different indicators differ. The equation to normalize indicators for which larger values indicate better performance is:

$$r_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

The equation to normalize indicators for which smaller values indicate better performance is:

$$r_{ij} = \frac{x_{\max} - x_{ij}}{x_{\max} - x_{\min}} \tag{13}$$

After normalizing X using Equations (12) and (13), the normalized index matrix R is:

$$R == \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{pmatrix} \tag{14}$$

where r_{ij} is the normalized index. The closer r_{ij} is to 1, the greater the vulnerability that is indexed. The purpose of the normalization is to facilitate the sorting of the vulnerability of the cluster center in a later stage (Section 4.2).

The n samples are divided into c clusters, and the c cluster centers can be represented by a matrix, $S = (s_{ih})_{m \times c}$, where s_{ih} is the normalized ith indicator of cluster center h and $0 \le s_{ih} \le 1$, $i = 1, 2, \ldots, m$; $h = 1, 2, \ldots, c$.

The membership matrix $U = (u_{hj})_{c \times n}$ is formed, where u_{hj} is the sample j belonging to the category h. $h = 1, 2, \ldots, c$; $j = 1, 2, \ldots, n$, and the condition $\sum_{h=1}^{c} u_{hj} = 1, 0 \le u_{hj} \le 1$ must be satisfied.

The difference between the sample j and the cluster center h is represented by the distance d_{hj}. The weight vector w = $(w_1\ w_2 \ldots\ w_m) = (w_i)$ is formed, satisfying the condition $\sum_{i=1}^{m} w_i = 1, \quad 0 \le w_i \le 1$, where w_i represent the degree of influence of different indicators on clustering results. The equation for d_{hj} is:

$$d_{hj} = \left\{ \sum_{i=1}^{m} \left[\omega_i \left| r_{ij} - s_{ih} \right| \right]^p \right\}^{\frac{1}{p}} \tag{15}$$

where different p values represent different distance parameters. When $p = 1$, it is the Hamming distance; when p is 2, it is the Euclidean distance.

To create the final membership matrix $U^* = (u_{hj})$ and the cluster center matrix $S^* = (s_{ih})$, the objective function is:

$$\min \left\{ F(u, s, w) = \sum_{j=1}^{n} \sum_{h=1}^{c} u_{hj}^2 d_{hj}^{\alpha} \right\} \tag{16}$$

where α is a variable parameter. When $\alpha = 1$, the function corresponds to the least absolute criterion; when $\alpha = 2$, the function corresponds to the least squares criterion. This model is a conditional extremum problem, which is transformed into an unconditional extremum problem using a Lagrangian multiplier. The final iterative equation is obtained by:

$$u_{hj} = 1/\sum_{k=1}^{c}\left(d_{hj}/d_{kj}\right)^{\alpha}, h = 1, 2, \cdots, c;\ j = 1, 2, \cdots, n \tag{17}$$

$$s_{ih} = \frac{\sum_{j=1}^{n} u_{hj}^{2} r_{ij}}{\sum_{j=1}^{n} u_{hj}^{2}}, \quad i = 1, 2, \ldots, c \tag{18}$$

Using Equations (17) and (18), the final membership matrix U^* and the cluster center matrix S^* are found by looping iterations, where α, p and w_i are the variable parameters. We used $\alpha = 2$ and $p = 2$, and equally weighted parameter combinations were used for clustering calculations.

4.2. Vulnerability Assessment and Level Characteristic Values

To quantify vulnerability, the vulnerability of c cluster centers in S^* is first quantified. The vulnerability of the cluster varies as the vulnerability of the cluster center. In Equations (17) and (18), which normalize the sample data, the vulnerability increases as r_{ij} gets closer to 1. An ideal node $(\underbrace{1, 1, \ldots 1}_{m})$ is one for which all values are 1, which means that all indicators are most vulnerable, so the ideal node has the greatest vulnerability. The vulnerability of the cluster center S_h is quantified based on its distance from the ideal node, with a smaller distance representing a greater vulnerability.

The random number of the initial fuzzy clustering matrix $(u_{hj}^{(0)})$ takes different values for different iterations, so the order of the cluster centers in S^* can differ between iterations. To facilitate sorting, cluster centers are ranked by vulnerability from large to small as 1, 2, ... , c. The membership matrix U^* must be adjusted according to the order of cluster centers in S^*.

After the adjustments to S^* and U^*, the nodes are classified. FCM clustering customarily uses the maximum membership principle, but Chen and Guo [38] explicitly opposed this method, claiming that the classification based on the maximum membership principle lost the global information of membership degree. To give an extreme example: When the sample membership value is equal, the maximum membership principle cannot determine which category the sample belongs to. Therefore, Chen and Guo [38] used the level characteristic value to determine the category. The membership distribution function of u_0 for c categories is $h \sim u_h (h = 1, 2, \cdots, c)$, and the product of the grade variable h and the degree of membership is summed to obtain the level characteristic value:

$$H(u_0) = \sum_{h=1}^{c} u_h h \tag{19}$$

Using Formula (19) to determine the category of samples has a more explicit mathematical and physical meaning: assuming that the unit mass objects are distributed along the horizontal axis, the corresponding masses u_1, u_2, ... , u_c, are concentrated on c points 1, 2, ... , c, on the horizontal axis. As shown in Figure 6, *H(u0)* represents the centroid position of the object.

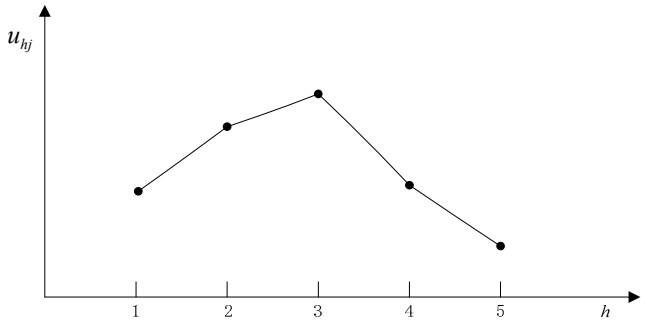

Figure 6. Distribution map of grade variable h and membership degree u_{hj}.

(1) If the membership degree is concentrated at one level point a, there is

$$H(u_0) = a, \begin{cases} u_h = 1, & h = a \\ u_h = 0, & h \text{ is a point other than } a \end{cases} \tag{20}$$

When u_0 is a member of point a, it has a physical meaning that the mass point of the object is at point a.

(2) If the membership degree is evenly distributed at each level point, there is

$$H(u_0) = \sum_{h=1}^{c} \frac{1}{c} h = \frac{1}{c} \frac{c(c+1)}{2} = \frac{c+1}{2}, \; u_h = \frac{1}{c}, h = 1, 2, \cdots c \tag{21}$$

When u_0 is a member of point $(c + 1)/2$, it has a physical meaning that the mass point of the object is at the midpoint of the object.

After $H(u_0)$ is obtained, the grade of u_0 can be determined by:

$$\begin{array}{l} 1.0 \le H(u_0) \le 1.5, \; u_0 \text{ is grade } 1 \\ h - 0.5 < H(u_0) \le h + 0.5, \; u_0 \text{ is grade } h \; (h = 2, 3, \cdots, c-1) \\ c - 0.5 < H(u_0) \le c, \; u_0 \text{ is grade } c \end{array} \tag{22}$$

The level characteristic value given by Equation (19) reflects the global information contained in the u_0 membership degree and can determine more accurately which grade u_0 is.

5. Node Vulnerability Analysis of a Power Grid in a Certain Region Under Earthquake Action

The partial grid of the San Francisco Bay area was selected as the research object in this paper. Due to the limitation of the acquired data, this paper adopted the following principles for the simplification of the power grid: (1) Only power plants and substations are reserved as nodes of the power grid, and transmission lines with voltage above 110 kV are used as edges. (2) Power plants and substations are regarded as indistinctive nodes; regardless of the influence of power flow direction and electrical parameters in the transmission line, it is abstracted as an unweighted and undirected edge. (3) By combining the transmission lines with the same pole, the self-loop and multiple edges in the topology model of the power network are eliminated, and the corresponding diagram becomes a simple one. According to the above principles, the grid in this area was simplified in Figure 7, which contains 20 nodes and 27 edges. Nodes 1 to 5 are power plants, and the rest are substations.

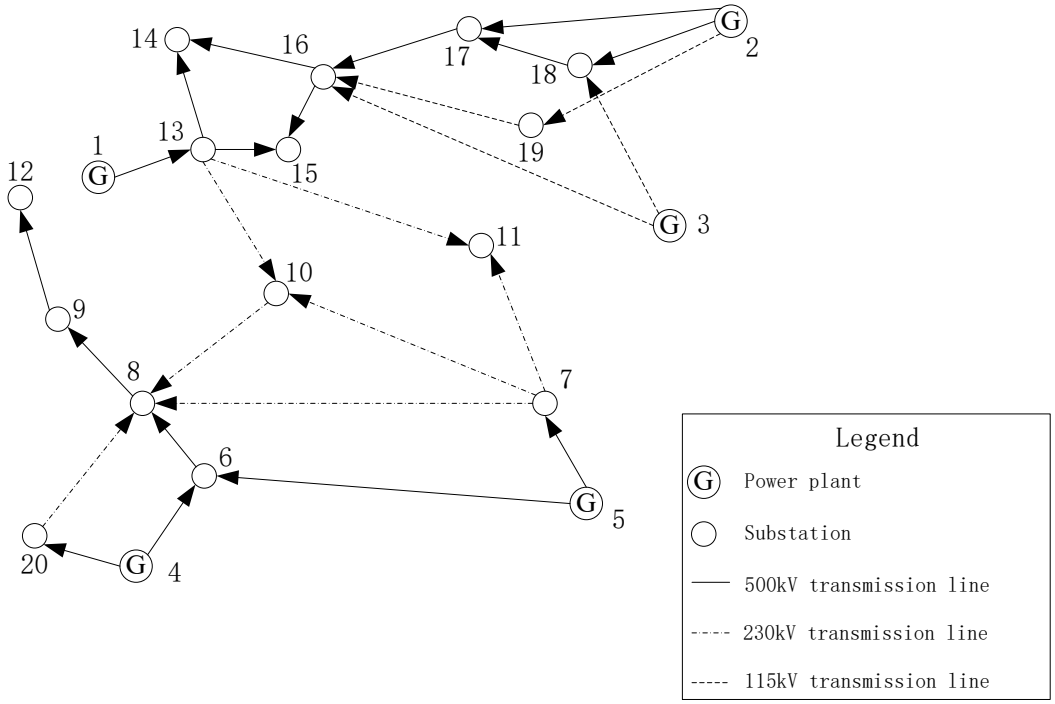

Figure 7. Power grid network in a city.

In this example, the design earthquake was the largest earthquake in the history of this area, the Loma Prieta earthquake in 1989 (Ms = 7.0), with an epicenter 29 km north of Point 4, and the disconnection probability P_i of each node was obtained. The hierarchical level G_i and the critical threshold T_i of each node for cascading failures were obtained. The initial values of these three indicators are listed in Table 1.

Table 1. Indicator data of nodes.

Node	P_i	G_i	T_i	Node	P_i	G_i	T_i
1	0	1	1.198	11	0	3	1.248
2	0	1	1.333	12	0.967	5	1.285
3	0	1	1.27	13	0.446	2	1.381
4	0.706	1	1.306	14	0	3	1.255
5	0	1	1.253	15	0	3	1.255
6	0.265	2	1.324	16	0	2	1.396
7	0	2	1.319	17	0	2	1.33
8	0.79	3	1.385	18	0	2	1.333
9	0.967	4	1.269	19	0	2	1.27
10	0.975	3	1.273	20	0.859	2	1.258

The indicators were normalized to accommodate different dimensions. For the probability of disconnection P_i and the critical threshold T_i, larger values indicate greater vulnerability, so Equation (12) was used for normalization. For the hierarchical level G_i, smaller values indicate greater vulnerability, so Equation (13) was used for normalization. Normalized data values are listed in Table 2.

Table 2. Normalized indicator data of nodes.

Node	P_i	G_i	T_i	Node	P_i	G_i	T_i
1	0	1	0.000	11	0	0.5	0.253
2	0	1	0.682	12	0.992	0	0.439
3	0	1	0.364	13	0.457	0.75	0.924
4	0.724	1	0.545	14	0	0.5	0.288
5	0	1	0.278	15	0	0.5	0.288
6	0.272	0.75	0.636	16	0	0.75	1.000
7	0	0.75	0.611	17	0	0.75	0.667
8	0.81	0.5	0.944	18	0	0.75	0.682
9	0.992	0.25	0.359	19	0	0.75	0.364
10	1	0.5	0.379	20	0.881	0.75	0.303

Node vulnerability in the power grid was divided into three categories, high, medium, and low; that is, $c = 3$. The parameter values were $\alpha = 2$, $p = 2$. Calculations were made according to the weight of each indicator, w = (1/3 1/3 1/3). The cluster center matrix S^* was obtained using Equations (17) and (18).

$$S^* = (s_1, s_2, s_3) = \begin{pmatrix} 0.0256 & 0.9137 & 0.1023 \\ 0.728 & 0.4058 & 0.7867 \\ 0.2969 & 0.4405 & 0.6959 \end{pmatrix}$$

The three cluster centers were s_1 = (0.0256, 0.728, 0.2969)$^{\mathrm{T}}$, s_2 = (0.9137, 0.4058, 0.4405)$^{\mathrm{T}}$, s_3 = (0.1023, 0.7867, 0.6959)$^{\mathrm{T}}$. At this point, it was necessary to determine the preferential order of cluster centers. The ideal node with the greatest vulnerability (1, 1, 1) was used, and the distance from the ideal node to the three cluster centers d and the degree of membership u were calculated using Equations (15) and (17): d = (1.232, 0.8207, 0.9715), u = (0.206, 0.463, 0.331). Using d and u, it was found that the distance between node s_1 and node (1, 1, 1) was the greatest and its degree of membership was the least, which indicated that node s_1 represented the category with the least vulnerability. Node s_2 was at the other extreme and represented the most vulnerable category. Node s_3 was between the two. Based on this result, the optimal cluster center order was adjusted to $S^* = (s_2, s_3, s_1)$. The corresponding membership matrix U^* was also adjusted, and the final result is shown in Table 3.

Table 3. Adjusted membership matrix.

Node	s_1	s_2	s_3	Node	s_1	s_2	s_3
1	0.0830	0.2124	0.7046	11	0.0496	0.1509	0.7995
2	0.0348	0.7709	0.1943	12	0.8106	0.0948	0.0945
3	0.0430	0.3085	0.6486	13	0.1876	0.6246	0.1878
4	0.3815	0.3591	0.2594	14	0.0481	0.1609	0.7911
5	0.0445	0.2344	0.721	15	0.0481	0.1609	0.7911
6	0.0439	0.8094	0.1466	16	0.0637	0.7735	0.1628
7	0.0160	0.8267	0.1573	17	0.0114	0.9054	0.0832
8	0.5833	0.2630	0.1537	18	0.0109	0.9155	0.0737
9	0.9437	0.0278	0.0285	19	0.0055	0.0435	0.9509
10	0.9666	0.0168	0.0165	20	0.7170	0.1382	0.1448

According to the membership degree of 20 nodes in the three cluster centers in Table 3, the level characteristic Equation (19) is applied to obtain the characteristic value H_i of vulnerability level for 20 nodes, which are listed in Table 4:

Table 4. The characteristic value H_i of vulnerability level.

Node	H_i	Node	H_i	Node	H_i	Node	H_i
1	2.6217	6	2.1027	11	2.7498	16	2.0991
2	2.1596	7	2.1413	12	1.2839	17	2.0718
3	2.6056	8	1.5703	13	2.0001	18	2.0628
4	1.8779	9	1.0849	14	2.7429	19	2.9454
5	2.6765	10	1.0499	15	2.7429	20	1.4277

The vulnerability of each node was ranked using Equation (20), giving the following results. Nodes with high vulnerability were 9, 10, 12 and 20; nodes with medium vulnerability were 2, 4, 6, 7, 8, 13, 16, 17 and 18; and nodes with low vulnerability were 1, 3, 5, 11, 14, 15 and 19. The clustering results are shown in Figure 8.

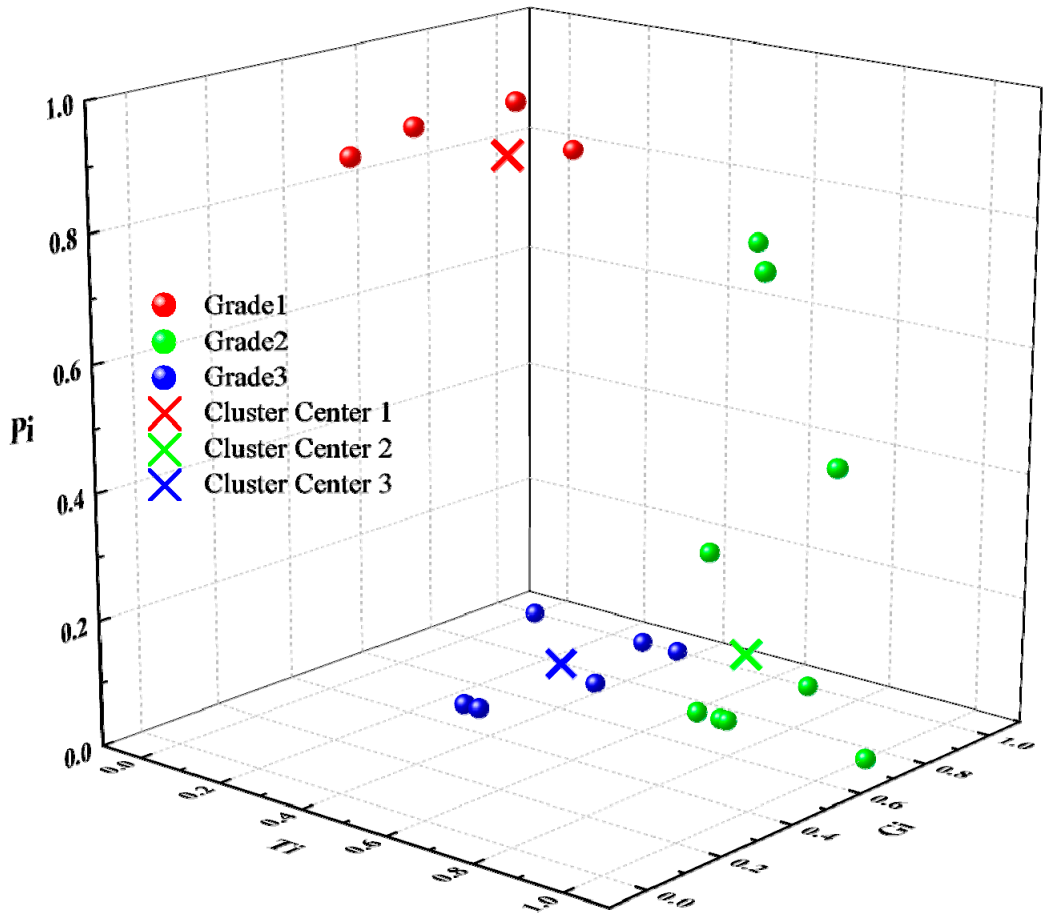

Figure 8. Node cluster map.

6. Result and Discussion

A vulnerability assessment was briefly analyzed. We found that the four nodes with high vulnerability were nodes with the highest probability of disconnection P_i, showing a large discrepancy in node disconnection probability. Thus, the influence of this indicator on the results is significant. This is consistent with the observed characteristic that the fuzzy clustering algorithm is sensitive to outliers with large changes [32]. The seven points with low vulnerability had a probability of disconnection 0, and the critical threshold of each node was also at a low level, which suggests that these nodes not only have a low probability of failure but also have little impact on adjacent nodes after failures. Intuitively they have low vulnerability, which matches everyday experience, showing that the variable fuzzy clustering method can well determine the vulnerability of power grid nodes to earthquake events.

Determining the vulnerability of power grid nodes is of great importance in determining what measures to take to reduce the effects of earthquakes and what post-earthquake emergency responses to initiate. For example, if a node loses its functionality due to structural damage, attention should be paid to improving the seismic grade of the facility, and post-earthquake priority should be given to inspecting and repairing nodes on the main trunk. For nodes with high vulnerability due to the topology of the network, the power grid should be optimized, which may include redundant facilities being added, power sources being better dispersed, and multi-loop power grids being created with each loop having a different power source.

Clustering is a typical algorithm for unsupervised learning. It is intended to explore and discover patterns in data samples and to find similar groups in them [39]. This model can be run without any prior knowledge of the data, which makes it suitable for vulnerability analysis. Nowadays, research on vulnerability is to establish a system performance model. By removing the nodes to simulate the impact on system performance after the failure, this effect actually represents the pattern of vulnerability generation. Therefore, the accuracy of the model becomes a key factor in vulnerability analysis. However, due to the complex mechanism of vulnerability generation and even the lack of an accurate definition to describe it, the accuracy of various system performance models is currently under discussion. In the vulnerability analysis, the conclusions obtained by different system performance models are quite different or even completely opposite. For example, in the literature on the vulnerability of complex networks, the vast majority of research supports the view that nodes with a large load have a great impact on the network. However, Wang and Rong [10] formed a different conclusion after studying the failure mechanism of the power grid in the Western United States. They found that if the parameters of the model meet certain conditions, attacking the node with a small load is more likely to cause a large-scale collapse than if the load is large. Another example is the use of pure models and extended models in the literature [14] to study the vulnerability of the power system, and the conclusions obtained are also divergent. Therefore, the purpose of using the clustering method in this paper is to break out of the limitation of system performance model and analyze the vulnerability from a completely different perspective.

The method described in this paper is more comprehensive in analyzing the impact of earthquake damage on the power grid than were previous studies, which have usually predicted the effects of node failure on the power grid by removing a node from the grid to determine its vulnerability through analysis of power flow and network topology. However, previous studies have assumed that each node has an equal probability of being destroyed, which is contrary to real-world observation. Some researchers have simulated the grid to quantify its vulnerability to earthquake damage by attaching a probability to each node. However, the computation required for this sort of simulation is huge and complex, and this approach has been unsuccessful so far [40,41]. We used the probability of a node being disconnected due to earthquake activity as an indicator of vulnerability and included it in the cluster analysis for a more realistic and reasonable consideration of the effect of an earthquake than the previous study using the same probability assumption.

7. Conclusions

It is difficult to quantify the vulnerability of a power grid to earthquake damage because of its complexity. Most previous studies have used a single indicator of grid vulnerability, which shows that only one perspective of vulnerability is taken into account. The single-index approach fails to represent the vulnerability of the grid comprehensively or accurately. We used three indicators of the vulnerability of the grid to earthquake damage in this study, the probability of disconnection, the hierarchical level, and the critical threshold of the power grid, together with the variable fuzzy clustering model, to obtain a more comprehensive measure.

The use of the indicators and methods proposed in this paper can objectively and accurately assess the vulnerability of grid nodes, but there are still some shortcomings in the research. First, because there is no one agreed-upon precise definition of vulnerability, the choice of an appropriate set

of indicators that accurately reflects vulnerability remains a problem that needs to be studied more deeply than we were able to. In this paper, the pure model is used to calculate the critical threshold, topological metrics identify a first level of vulnerability in the physical structure. However, the flow of electric power in power grid follows Kirchoff's laws, using only topology metrics, ignoring power grid characteristics and technical constraints may lead to inaccurate results. Therefore, in our future work, the influence of technical constraints (voltage, resistance, maximum power, etc.) should be taken into account. Second, power grid performance is also an important indicator of vulnerability. We used the probability of disconnection and the critical threshold as alternatives to functional indicators. This choice is acceptable for analysis of vulnerability to earthquake damage, but functional indicators are also likely to provide accurate measure of the vulnerability of the grid and must be considered in future research. Third, the case power grid used in this paper is small in scale, which is inconsistent with the characteristics of large-scale modern power grid. In future work, the methodology should be tested in a larger-scale power grid. At last, with the development of power grid technology, smart grid has become a new and vibrant research field. In smart grid, the network topology may be frequently changed to optimize its behavior. How to evaluate the impact of structural changes on vulnerability is also an important research field.

Author Contributions: Y.D. collected relevant data; Y.Y. facilitated in funding acquisition and supervised the work; T.L. did most of analysis, contributed extensively to the first draft of the manuscript; Y.D. and Y.Y. reviewed the manuscript and gave many suggestions. All authors read and approved the final manuscript.

Funding: This research received no external funding.

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

Nomenclature:

A	Adjacency matrix
M	Judgment matrix that determines whether nodes are connected
P_i	Disconnection probability of node i
G_i	Hierarchical level of node i
L_i	Load of node i
k_i	Degree of node i
T_i	Critical threshold of node i
m	Number of indicators
n	Number of samples
c	Number of cluster centers
s_{ih}	Normalized ith indicator of cluster center h
$S_{m\times c}$	Cluster center matrix
u_{hj}	Membership degree of the sample j belonging to the category h
$U_{c\times n}$	Membership matrix
d_{hj}	Distance between the sample j and the cluster center h
w_i	Indicator weight
$H(u_0)$	Level characteristic value of the sample u_0

References

1. Scawthorn, C.; Chen, W.F. *Earthquake Engineering Handbook*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2003; pp. 1–18.
2. Samuel, P.H. The Clash of Civilization. In *Power and Interdependence: World Politics in Transition*, 5th ed.; Keohane, R.O., Nye, J.S., Eds.; Routledge: New York, NY, USA, 2017; pp. 32–50.
3. Zhong, M.; Sun, Y.H.; Zhang, Q.; Liu, Z. Study on Seismic Fortification Level of Electrical Equipment. *Adv. Mater. Res.* **2014**, *1065–1069*, 1497–1502. [CrossRef]
4. Guan, Z.; Zhu, Q.; Fan, X.; Cao, M.; Yuan, B.; Wang, H.; Ren, J. Comparison and Research on Seismic Design Practice for Electric Substation Equipment in China, Japan, USA and Europe. In Proceedings of the 2nd

International Workshop on Renewable Energy and Development, Guilin, China, 20–22 April 2018; Volume 153, p. 042017.
5. Kongar, I.; Giovinazzi, S.; Rossetto, T. Seismic performance of buried electrical cables: Evidence-based repair rates and fragility functions. *Bull. Earthq. Eng.* **2017**, *15*, 3151–3181. [CrossRef]
6. Nazemi, M.; Moeini-Aghtaie, M.; Fotuhi-Firuzabad, M.; Dehghanian, P. Energy Storage Planning for Enhanced Resilience of Power Distribution Networks against Earthquakes. *IEEE Trans. Sustain. Energy* **2019**. accept. [CrossRef]
7. Wu, Y.K.; Chang, S.M.; Hu, Y.L. Literature Review of Power System Blackouts. In Proceedings of the 4th International Conference on Power and Energy Systems Engineering, Berlin, Germany, 25–29 September 2017; Bevrani, H., Ed.; Elsevier Science: Amsterdam, The Netherlands, 2017; Volume 141, pp. 428–431.
8. Albert, R.; Albert, I.; Nakarado, G.L. Structural vulnerability of the North American power grid. *Phys. Rev. E* **2004**, *69*, 025103. [CrossRef]
9. Kinney, R.; Crucitti, P.; Albert, R.; Latora, V. Modeling cascading failures in the North American power grid. *Eur. Phys. J. B* **2005**, *46*, 101–107. [CrossRef]
10. Wang, J.W.; Rong, L.L. Robustness of the western United States power grid under edge attack strategies due to cascading failures. *Saf. Sci.* **2011**, *49*, 807–812. [CrossRef]
11. Rosas-Casals, M.; Valverde, S.; Sole, R.V. Topological vulnerability of the European power grid under errors and attacks. *Int. J. Bifurc. Chaos* **2007**, *17*, 2465–2475. [CrossRef]
12. Crucitti, P.; Latora, V.; Marchiori, M. A topological analysis of the Italian electric power grid. *Physica A* **2004**, *338*, 92–97. [CrossRef]
13. Liu, B.; Li, Z.; Chen, X.; Huang, Y.H.; Liu, X.D. Recognition and Vulnerability Analysis of Key Nodes in Power Grid Based on Complex Network Centrality. *IEEE Trans. Circuits Syst. II Exp. Briefs* **2018**, *65*, 346–350. [CrossRef]
14. Hines, P.; Cotilla-Sanchez, E.; Blumsack, S. Do topological models provide good information about electricity infrastructure vulnerability? *Chaos* **2010**, *20*, 033122. [CrossRef]
15. Bompard, E.; Luo, L.G.; Pon, E. A perspective overview of topological approaches for vulnerability analysis of power transmission grids. *Int. J. Crit. Infrastruct.* **2015**, *11*, 15–26. [CrossRef]
16. 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. *Physica A* **2017**, *466*, 295–309. [CrossRef]
17. Yan, J.; Tang, Y.; He, H.; Sun, Y. Cascading Failure Analysis With DC Power Flow Model and Transient Stability Analysis. *IEEE Trans. Power Syst.* **2015**, *30*, 285–297. [CrossRef]
18. Sun, Y.S.; Tang, X.S.; Zhang, G.W.; Miao, F.F.; Wang, P. Dynamic Power Flow Cascading Failure Analysis of Wind Power Integration with Complex Network Theory. *Energies* **2018**, *11*, 63. [CrossRef]
19. Wang, Z.Y.; Chen, G.; Hill, D.J.; Dong, Z.Y. A power flow based model for the analysis of vulnerability in power networks. *Physica A* **2016**, *460*, 105–115. [CrossRef]
20. Ouyang, M. Comparisons of purely topological model, betweenness based model and direct current power flow model to analyze power grid vulnerability. *Chaos* **2013**, *23*, 293–296. [CrossRef]
21. Guan, X.; Liu, J.; Gao, Z.; Yu, D.; Cai, M. Power grids vulnerability analysis based on combination of degree and betweenness. In Proceedings of the 26th Chinese Control and Decision Conference, Changsha, China, 31 May–2 June 2014; IEEE: New York, NY, USA, 2014; pp. 4829–4833.
22. Wang, J.W.; Rong, L.L.; Zhang, L.; Zhang, Z. Attack vulnerability of scale-free networks due to cascading failures. *Physica A* **2008**, *387*, 6671–6678. [CrossRef]
23. Wang, J.W.; Rong, L.L. A model for cascading failures in scale-free networks with a breakdown probability. *Physica A* **2009**, *388*, 1289–1298. [CrossRef]
24. Jie, L. *Earthquake Resistance of Lifeline Engineering—Basic Theory and Application*, 1st ed.; Science Press: Beijing, China, 2005; pp. 172–175.
25. Vanzi, I. Seismic reliability of electric power networks: Methodology and application. *Struct. Saf.* **1996**, *18*, 311–327. [CrossRef]
26. Giannini, R.; Vanzi, I. Seismic reliability of electric networks and interaction with other damage indicators. In Proceedings of the 12th World Conference on Earthquake Engineering, Auckland, New Zealand, 30 January–4 February 2000; p. 2041.

27. Alipour, Z.; Monfared, M.A.S.; Zio, E. Comparing topological and reliability-based vulnerability analysis of Iran power transmission network. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2014**, *228*, 139–151. [CrossRef]
28. Nasiruzzaman, A.B.M.; Pota, H.R.; Mahmud, M.A. Application of centrality measures of complex network framework in power grid. In Proceedings of the 37th Annual Conference of the IEEE Industrial Electronics Society, Melbourne, Australia, 7–10 November 2011; pp. 4660–4665.
29. Bompard, E.; Wu, D.; Xue, F. The Concept of Betweenness in the Analysis of Power Grid Vulnerability. In Proceedings of the International Conference on Complexity in Engineering (COMPENT 2010), Rome, Italy, 22–24 February 2010; pp. 52–54.
30. 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. *Physica A* **2011**, *390*, 4692–4701. [CrossRef]
31. Hines, P.; Blumsack, S. A Centrality Measure for Electrical Networks. In Proceedings of the 41st Annual Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 7–10 January 2008; pp. 1491–1498.
32. Nayak, J.; Naik, B.; Behera, H.S. Fuzzy C-Means (FCM) Clustering Algorithm: A Decade Review from 2000 to 2014. In Proceedings of the 1st International Conference on Computational Intelligence in Data Mining, Burla, India, 20–21 December 2014; Jain, L.C., Behera, H.S., Mandal, J.K., Mohapatra, D.P., Eds.; Springer: India, 2015; Volume 32, pp. 133–149.
33. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]
34. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [CrossRef]
35. Billinton, R.; Li, W.Y. *Reliability Assessment of Electric Power Systems Using Monte Carlo Methods*, 1st ed.; Springer: New York, NY, USA, 1994; pp. 33–73.
36. Ciaponi, C.; Franchioli, L.; Papiri, S. Simplified Procedure for Water Distribution Networks Reliability Assessment. *J. Water Resour. Plan. Manag.* **2012**, *138*, 368–376. [CrossRef]
37. Wang, J.W.; Rong, L.L.; Zhang, L. Effect of Attack on Scale-Free Networks Due to Cascading Failure. *Mod. Phys. Lett. B* **2009**, *23*, 1577–1587. [CrossRef]
38. Chen, S.Y.; Guo, Y. Variable fuzzy sets and its application in comprehensive risk evaluation for flood-control engineering system. *Fuzzy Optim. Decis. Mak.* **2006**, *5*, 4–8.
39. Scrucca, L. Genetic Algorithms for Subset Selection in Model-Based Clustering. In *Unsupervised Learning Algorithms*, 1st ed.; Celebi, M.E., Aydin, K., Eds.; Springer: Cham, Switzerland, 2016; pp. 55–70.
40. Abedi, A.; Gaudard, L.; Romerio, F. Review of major approaches to analyze vulnerability in power system. *Reliab. Eng. Syst. Saf.* **2019**, *183*, 153–172. [CrossRef]
41. Larocca, S.; Johansson, J.; Hassel, H.; Guikema, S. Topological Performance Measures as Surrogates for Physical Flow Models for Risk and Vulnerability Analysis for Electric Power Systems. *Risk Anal.* **2015**, *35*, 608–623. [CrossRef]