

Article

Dynamic Identification of Critical Nodes and Regions in Power Grid Based on Spatio-Temporal Attribute Fusion of Voltage Trajectory

Xiwei Bai ^{1,2}, Daowei Liu ³, Jie Tan ^{1,*}, Hongying Yang ³ and Hengfeng Zheng ³¹ Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; baixiwei2015@ia.ac.cn² University of Chinese Academy of Sciences, Beijing 100049, China³ China Electric Power Research Institute, Beijing 100192, China; liudaowei@epri.sgcc.com.cn (D.L.); yanghongying@epri.sgcc.com.cn (H.Y.); zhenghf0510@163.com (H.Z.)

* Correspondence: jie.tan@ia.ac.cn

Received: 10 January 2019; Accepted: 25 February 2019; Published: 26 February 2019



Abstract: Accurate identification of critical nodes and regions in a power grid is a precondition and guarantee for safety assessment and situational awareness. Existing methods have achieved effective static identification based on the inherent topological and electrical characteristics of the grid. However, they ignore the variations of these critical nodes and regions over time and are not appropriate for online monitoring. To solve this problem, a novel data-driven dynamic identification scheme is proposed in this paper. Three temporal and three spatial attributes are extracted from their corresponding voltage phasor sequences and integrated via Gini-coefficient and Spearman correlation coefficient to form node importance and relevance assessment indices. Critical nodes and regions can be identified dynamically through importance ranking and clustering on the basis of these two indices. The validity and applicability of the proposed method pass the test on various situations of the IEEE-39 benchmark system, showing that this method can identify the critical nodes and regions, locate the potential disturbance source accurately, and depict the variation of node/region criticality dynamically.

Keywords: critical node; critical region; spatio-temporal attribute fusion; node voltage trajectory

1. Introduction

With the development of modern power electronics, an increasing number of loads characterized by high-power, nonlinearity, and randomness are being connected to the power grid. The operating point of the modern power system is approaching its limit [1], which raises the probability of the occurrence of large-scale blackouts. Generally speaking, a power blackout begins with a node failure and gradually radiates through a small area. If there are no appropriate responses to prevent it, the whole regional power system will be affected, causing heavy economic loss. Therefore, it is meaningful to do research on finding critical nodes and understanding their propagation mechanism. This study is closely related to the safety assessment, prevention, and control of power systems, attracting more and more scholars and researchers worldwide in recent years.

Essentially, a regional power grid can be regarded as a graph or network with edges and various types of nodes. Its vulnerability depends largely on the critical nodes with increasing loads, strong local connectivity, and failure-prone characteristics. The effective identification of these nodes requires a proper node importance assessment measure. There have been many studies focusing on finding such metrics. For example, the complex network theory [2] has been applied in different types of network analysis such as social networks, information networks, etc., which more or less share some common characteristics with the abstracted small world or scale-free network models [3], and power

grids are no exceptions. Evidence in [4] shows that although there is no definitive answer to the question for power grids regarding their memberships to typical complex network models, some real world examples do own small world properties to some extent. The complex network theory specializes in analyzing the topological structure of a network [5] and identifying critical nodes using centrality measures like degree centrality, closeness centrality, and betweenness centrality [6]. In [7], a semi-local centrality is proposed to reach a tradeoff among these commonly used centrality measures for large network systems. In [8], the author points out that the PageRank (PR) algorithm for Google search engine is based on the link structure between web pages, and this algorithm can be extended for node rankings in power grid analysis. The correspondences of topology structure between the web and the power grid are analyzed, and an improved PR algorithm application for the power grid is given in [9]. It is noteworthy that—compared with other networks—power grids own some special properties, namely their electrical characteristics (operational parameters, electric distance, etc.). A complex power grid is composed of various electrical equipment, transmission lines, and information equipment. Any connections should follow, and changes in any components may cause large scale impact according to the principle of electrical engineering. Researchers have tried to take these electrical characteristics into consideration [10]. In [11], the power load and supply are combined with degree and betweenness centralities to form an effective assessment index. In [12], in order to evaluate important nodes, some operational characteristics like power flow, load capacity, and power source in the power grid are considered to make modifications for the improved hyper-text induced topic selection algorithm. However, similar problems exist in the aforementioned methods, namely that they are all single factor analyses, whereas the determination of important nodes in a power grid requires information from various aspects. Consequently, information fusion techniques have been adopted by many scholars to implement comprehensive node importance assessment. An example in [6] concerns the combination of different topological indices in complex networks. The degree centrality, efficiency, betweenness, and correlation centralities are fused based on information entropy. A multi-attribute based method proposed in [13] synthesizes three different electrical indices including feasible flow, power transmission efficiency, and electrical betweenness to achieve convincing node importance assessment. This method uses Gini-coefficient to determine the weight of each single index when they are integrated together. In [14], a criteria importance through intercriteria correlation (CRITIC) based method merges seven topological and electrical indices using information entropy and Spearman correlation score. The author compares this method with five existing methods of the same kind. The demonstrated simulations for real power system show better performances than part of the aforementioned single or multi-indices methods on discrimination ability and correctness of key node identification.

Nodes in a regional power grid are not isolated—they are interdependent and correlated [15]. Any abnormalities or failures may propagate along the grid structure. Therefore, it is not enough to merely identify critical nodes; the relevance between them and the other nodes should also be explored. Understanding the interconnection and identifying the sphere of influence of critical nodes or critical regions are necessary for the improvement of power grid vulnerability analysis, situational awareness, prevention, and control capabilities. The so-called relevance can be divided into two categories—topological relevance and attribute relevance [16]. The first measures the spatial relationship while the second measures the temporal relationship of various electrical properties between nodes. For a typical propagation model, events start from a special node and radiate through the whole network. Likewise, all nodes in the critical region should have strong attribute relevance on the basis of fulfilling topological constraints. The critical region identification can be achieved through clustering methods, and the relevance between nodes is equivalent to their similarity measurements. In [17], a k-means clustering strategy is adopted to determine the voltage stability weak area, which is closely related to the node relevance analysis. The node voltage magnitude, severity range index, and clustering error index are used to determine similarities between nodes. In [18], the fast voltage stability index (FVSI) is employed to evaluate line sensitivities in a power grid, and those sensitive

lines are further clustered in a group. In [19], four node sensitivity indices are used to measure similarities, and a fuzzy subtract clustering algorithm is implemented to identify weak voltage regions. Experiments in a real power grid prove the validity of this method.

In summary, researchers around the world have proposed effective methods for identifying critical nodes and regions in a power grid through assessing node importance and relevance. However, most of the above-mentioned assessment indices are static, which means they do not depend on the operating condition of a power grid. In fact, critical nodes and regions shift each time the power grid changes, thus the dynamic identification method should be more suitable. Unfortunately, there are not many studies on it. The problem is how to find suitable dynamic indices from the operating data of key parameters (voltage, phase angle, active and reactive power, etc.) in a power grid. Node voltage trajectories are easily accessible, and we believe that they contain abundant information reflecting the operating condition of the power grid. Based on this hypothesis, a new data-driven method based on spatio-temporal attribute fusion of voltage trajectory is proposed and discussed in detail. We employ three fundamental temporal attributes and three spatial attributes from measured node voltage phasor sequences. Two comprehensive indices are formed by attribute fusion to assess node importance and relevance. The critical nodes are identified by importance ranking, and the critical regions are identified by a density-based spatial clustering of applications with noise (DBSCAN) algorithm with synthesized similarity measurement.

The rest of this paper is organized as follow: Section 2 introduces the main idea and the overall framework of proposed method. Section 3 describes the specific types of spatial and temporal attributes and the way to extract them from node voltage trajectories. Section 4 establishes node importance and relevance assessment indices using information fusion techniques. Section 5 illustrates the critical nodes/regions identification method in detail. Experiments and results analysis are in Section 6. At last, Section 7 concludes this paper.

2. Dynamic Identification of Critical Nodes and Regions in Power Grid

A real power grid is never invariant—the operating conditions, key parameters, or even the grid structure may change at any moment. Therefore, for practical application, any fixed assessment indices for critical nodes and regions identification are inaccurate to varying degrees. Fortunately, with the development of the wide-area measurement system (WAMS), a large number of synchronized phasor measurement units (PMUs) were deployed on a power grid to enable real-time online data acquisition [20]. On the basis of these measured data, we put forward a scheme for dynamic identification of critical nodes and regions. The overall framework is shown in Figure 1.

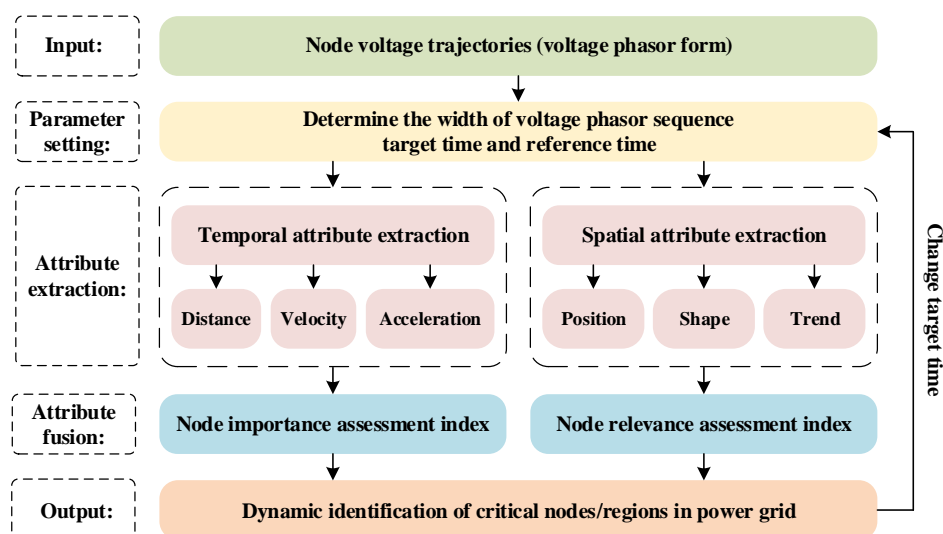


Figure 1. Overall framework of proposed scheme.

Our main idea was to extract useful attributes from the observed node voltage trajectories, merge them into two indices for node importance and relevance assessment, respectively, then further realize the dynamic identification of critical nodes and regions in the power grid. The kinematic properties of node voltages could be drawn from their changing trajectories. Figure 2 demonstrates a numerical example of some varying node voltage trajectories in a small regional power grid in phasor representation. The data were acquired through a continuous power flow (CPF) simulation. The arrow of the time dimension illustrates that voltage phasors move from right to left. Our research was conducted based on the idea that the voltage phasors of critical nodes experience more drastic variations in disturbance environment. Therefore, in Figure 2, considering that a voltage phasor sequence in a time segment may contain more worthy information, we tried to measure the differences between a target and a reference voltage phasor sequence, U_t and U_r , in a different time segment from the identical node voltage trajectory (importance assessment) and between two sequences, U_t and U_c , from two different node voltage trajectories in an identical time segment (relevance assessment). The former determined the importance of each node and the latter determined the relevance between any two nodes. At the beginning stage of the dynamic identification scheme, namely the parameter setting step, the width of the voltage phasor sequence, the target time, and the reference time were set according to experiences and actual data. Next, in the attribute extraction step, the process was divided into two parts. The distance between the center of the target and the reference voltage phasor sequences and their average velocity and acceleration differences for both real and imaginary parts were extracted as temporal attributes. The differences in position, shape, and trend between the target and another corresponding voltage phasor sequence in an identical time segment were extracted as spatial attributes. Then, in the attribute fusion phase, temporal attributes were integrated into a node importance assessment index via Gini-coefficient, and spatial attributes were integrated into a node relevance assessment index via Spearman correlation coefficient. On this basis, the critical nodes were identified by importance ranking, and the critical regions were identified by DBSCAN algorithm with synthesized similarity measurement, which further merged both temporal and spatial assessment indices.

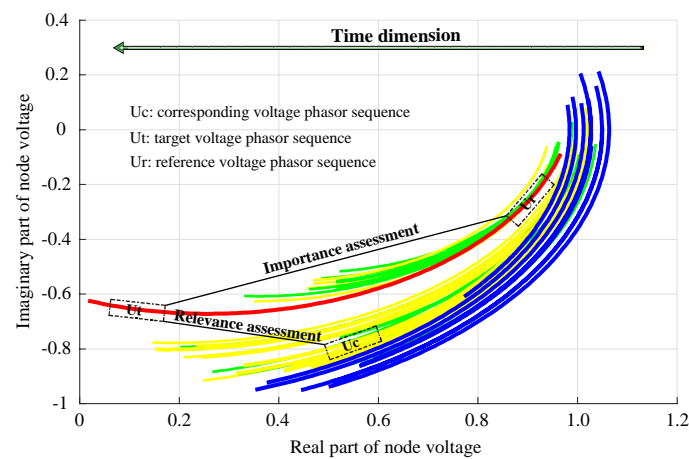


Figure 2. Node voltage trajectories.

The characteristic of a dynamic identification method lies in its capability to update. Therefore, the target phasor sequence will slide along with the time dimension, repeating this scheme and updating the identification results continuously to improve its reliability.

3. Attribute Extraction from Node Voltage Spatio-Temporal Trajectory

To quantify the variations of node voltage phasor sequences over time, their temporal and spatial attributes were determined and extracted online based on the current and historical voltage phasors. The temporal attributes were employed to measure differences between two sequences in two different

time segments of an identical node, while the spatial attributes were employed to measure the similarity between two sequences in an identical time segment of different nodes.

3.1. Node Voltage Temporal Attribute Extraction

Temporal attributes aim to describe the kinematic properties of a moving voltage phasor sequence. Therefore, three fundamental physical quantities—distance, velocity, and acceleration—were selected for attribute characterization. In this paper, a voltage phasor sequence with width W is represented as $\dot{U} = [\dot{u}(1), \dot{u}(2), \dots, \dot{u}(W)]$, where $u(i)$ is the effective value, $\varphi(i)$ is the phase angle, and $\dot{u}(i) = u(i) \angle \varphi(i)$.

The distance attribute is defined as the distance between the center point of the target and the reference voltage phasor sequence, represented as \dot{U}_t and \dot{U}_r with same width W . The real and imaginary parts of a voltage phasor can be computed as $u^R(i) = u(i) \cos \varphi(i)$ and $u^I(i) = u(i) \sin \varphi(i)$. On this basis, the distance attribute A_{dis}^ψ contains two parts:

$$A_{dis}^\psi = |u_t^\psi(m) - u_r^\psi(m)| \quad (1)$$

where ψ represents letter R or I for this and all following attributes, $u_t^R(m)$ and $u_r^R(m)$ are values of real parts belonging to the center points of \dot{U}_t and \dot{U}_r ; likewise, $u_t^I(m)$ and $u_r^I(m)$ are center values of imaginary parts. The value of center point can be computed as:

$$u_t^\psi(m) = \begin{cases} u_t^\psi\left(\frac{W}{2} + 1\right) & W = 2n + 1, n \in \mathbb{N} \\ \left[u_t^\psi\left(\frac{W}{2}\right) + u_t^\psi\left(\frac{W}{2}\right)\right] / 2 & W = 2n, n \in \mathbb{N} \end{cases} \quad (2)$$

The velocity attribute is defined as the average velocity difference between \dot{U}_t and \dot{U}_r . A voltage phasor sequence is composed of W continuous voltage phasors, the norm of the difference value between two adjacent voltage phasors is regarded as velocity value. Same as above, the velocity attribute A_{vel}^ψ can be divided into real and imaginary parts, which are shown in the following equation:

$$A_{vel}^\psi = \frac{1}{W} \sum_{j=1}^W \left| \left[u_t^\psi(j) - u_t^\psi(j-1) \right] - \left[u_r^\psi(j) - u_r^\psi(j-1) \right] \right| \quad (3)$$

The acceleration attribute is defined as the average acceleration difference between \dot{U}_t and \dot{U}_r . On the basis of the aforementioned concepts, acceleration can be obtained by computing the difference value between two adjacent velocity values. Therefore, the acceleration attribute A_{acc}^ψ is represented as:

$$A_{acc}^\psi = \frac{1}{W} \sum_{j=1}^W \left| \left[u_t^\psi(j) - 2 \times u_t^\psi(j-1) + u_t^\psi(j-2) \right] - \left[u_r^\psi(j) - 2 \times u_r^\psi(j-1) + u_r^\psi(j-2) \right] \right| \quad (4)$$

3.2. Node Voltage Spatial Attribute Extraction

Spatial attributes are employed to measure the similarity between two voltage phasor sequences (\dot{U}_t and \dot{U}_c) in an identical time segment. Here in this paper, from three perspectives, we took spatial attributes including position, shape, and trend into consideration. Figure 3 gives a visual illustration of these three attributes.

The position attribute is depicted by the average Euclidean distance among all voltage phasors in \dot{U}_t and \dot{U}_c . Owing to the propagation property of the power grid, the influences from other nodes need to be considered, thus the grey relational degree [21] is introduced in this paper. It is appropriate for both position and shape attributes measurement because of its ability to integrate local distance and global maximum and minimum distance for phasor sequences. Suppose there are $n + 1$ nodes in a

regional power grid, the computation of grey relational degrees $r(\dot{U}_t, \dot{U}_c(i))$ involves one \dot{U}_t and n \dot{U}_c , represented as $\{\dot{U}_c^1, \dot{U}_c^2, \dots, \dot{U}_c^n\}$. We have:

$$A_{pos} = r(\dot{U}_t, \dot{U}_c^i) = \frac{1}{W} \sum_{k=1}^W r(\dot{u}_t(k), \dot{u}_c^i(k)) \quad (5)$$

$$r(\dot{u}_t(k), \dot{u}_c^i(k)) = \frac{\min_{i,j} \|\dot{u}_t(k) - \dot{u}_c^i(k)\|_2 + \xi \max_{i,j} \|\dot{u}_t(k) - \dot{u}_c^i(k)\|_2}{\|\dot{u}_t(k) - \dot{u}_c^i(k)\|_2 + \xi \max_{i,j} \|\dot{u}_t(k) - \dot{u}_c^i(k)\|_2} \quad (6)$$

where A_{pos} is the position attribute, and $r(\dot{U}_t, \dot{U}_c^i)$ is the grey relational degree between \dot{U}_t and \dot{U}_c^i . ξ is the distinguishing coefficient that fulfills $0 < \xi < 1$ and $\xi = 0.5$ generally, W is the width of sequence, and $\dot{u}_c^i(k)$ represents the k^{th} phasor in the i^{th} sequence. The global minimum $\min_{i,j} \|\dot{u}_t(k) - \dot{u}_c^i(k)\|_2$ is computed in two stages. The local minimum $\min \|\dot{u}_t(k) - \dot{u}_c^i(k)\|_2$ is searched for all k firstly, then the global minimum is searched for all i . The global maximum can be acquired in the same way.

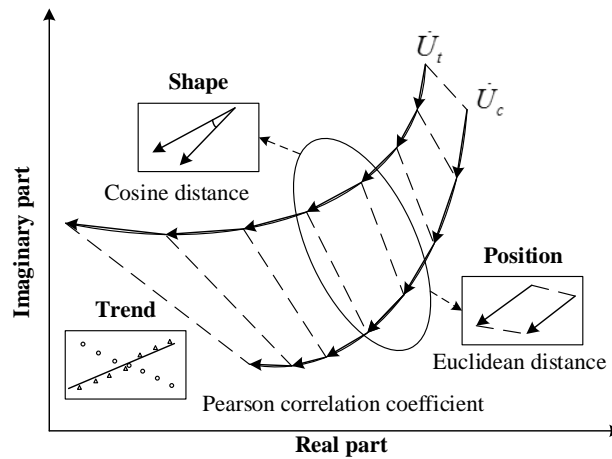


Figure 3. Spatial attribute of node voltage trajectories.

As is shown in Figure 3, the shape attribute A_{sha} is defined as the grey relational degree with cosine distance. It can be obtained directly by replacing the Euclidean distance into cosine distance in Equations (5) and (6). To avoid repetition, their formulas are not given here.

The trend attribute measures the difference of trend variation between \dot{U}_t and \dot{U}_c . In [22], binary direction vectors are employed to measure the trend similarity between 1D time-series. Trends of a variation vector are divided into ascending type and descending type. However, this measurement is rough and has a similar effect as the shape attribute to some extent. To avoid these drawbacks, Pearson correlation coefficient (PCC) is employed to measure the trend difference between two entire voltage phasor sequences. Traditional PCC is not able to depict the trend of 2D series, thus we computed PCCs between \dot{U}_t and \dot{U}_c for their real and imaginary parts, respectively. Results were transformed and multiplied to obtain a final value within the range of 0 and 1. The trend attribute can be obtained through the following equations:

$$A_{tre} = [\rho^R(\dot{U}_t, \dot{U}_c) + 1][\rho^I(\dot{U}_t, \dot{U}_c) + 1]/4 \quad (7)$$

$$\rho^{\psi}(\dot{U}_t, \dot{U}_c) = \frac{\sum_{i=1}^n [u_t^{\psi}(i) - \bar{u}_t^{\psi}(i)] [u_c^{\psi}(i) - \bar{u}_c^{\psi}(i)]}{\sqrt{\sum_{i=1}^n [u_t^{\psi}(i) - \bar{u}_t^{\psi}(i)]^2 \sum_{i=1}^n [u_c^{\psi}(i) - \bar{u}_c^{\psi}(i)]^2}} \quad (8)$$

where $\rho^R(\dot{U}_t, \dot{U}_c)$ and $\rho^I(\dot{U}_t, \dot{U}_c)$ are PCCs for real and imaginary parts, and ψ represents letter R or I .

4. Assessment Index Establishment Based on Attribute Fusion

The proposed temporal attributes and spatial attributes provide criteria for node importance and relevance assessment in different ways with different preferences. To form a unique comprehensive assessment index, both two types of attributes must be integrated. The key to effective attribute fusion is to determine proper weights for all single attributes. On the basis of their characteristics, two information fusion techniques were employed to integrate both temporal and spatial attributes, respectively.

4.1. Temporal Attribute Fusion Based on Gini-Coefficient

Gini-coefficient [13] is initially proposed to measure the inequality in social income distribution based on the Lorenz curve. This coefficient is between 0 and 1, and the bigger the value is, the larger the gap between rich and poor. In computer science, Gini-coefficient is used to measure the impurity of data. There are some famous applications like the classification and regression tree (CART) algorithm. In brief, if the samples of a category of data distribute uniformly, this category is not discriminative. If the distribution is uneven, it indicates that these samples own some sort of aggregation properties and can be separated into several different classes. For temporal attribute fusion, the importance of each kinematic attribute is uncertain, but the discriminative attribute should be given larger weight. The temporal attribute fusion algorithm based on Gini-coefficient is shown in Algorithm 1. An attribute matrix $M^t \in \mathbb{R}^{n \times 6}$ that contains all temporal attributes computed for all nodes should be formed, represented as:

$$M^t = \begin{bmatrix} A_{dis}^R(p) & A_{dis}^I(p) & A_{vel}^R(p) & A_{vel}^I(p) & A_{acc}^R(p) & A_{acc}^I(p) \end{bmatrix}$$

where $p \in [1, n]$, n is the number of nodes, and the number in parenthesis designates the target voltage phasor sequence.

Algorithm 1: Temporal attribute fusion based on Gini-coefficient

1. **Input:** Attribute matrix M^t , the number of attributes n_A
 2. $S \leftarrow \emptyset$
 3. $W^t \leftarrow \emptyset$
 4. **for** $i = 1$ to n_A **do**
 sort $M_{*,i}^t$ in ascending order
 $S_i \leftarrow \sum M_{*,i}^t$
 end for
 5. **for** $j = 2$ to n **do**
 $M_{j,*}^t \leftarrow M_{j,*}^t + M_{j-1,*}^t$
 end for
 6. **for** $k = 1$ to n_A **do**
 $W_k^t \leftarrow 1 - \frac{1}{n} \left(2 \times \sum \frac{M_{*,k}^t}{S_k} - 1 \right)$
 end for
 7. Normalize W^t to the range of 0 to 1
 8. **Output:** weight vector W^t
-

The value of n_A is set to 6. After calculating the weight vector W^t , the node importance assessment index can be given as:

$$I_T(i) = \sum_{j=1}^{n_A} M_{i,j}^t W_j^t, i \in [1, n] \quad (9)$$

4.2. Spatial Attribute Fusion Based on Spearman Correlation Coefficient

Spatial attributes like position, shape, and trend attributes are not disconnected but instead contain parts of information from each other. To achieve effective spatial attribute fusion, their relevance should be reduced by appropriate weight determination. Accordingly, the more independent an attribute is, the larger weight it is given. In this paper, the Spearman correlation coefficient (SCC) [23] is employed to measure the independence of each attribute. Essentially, SCC computed the PCC of the rankings of data between two ordinal variable sequences or transformed continuous variable sequences. For two attribute sequences S_1 and S_2 , they are sorted in ascending order, then the rankings of each element form a two rank sequence. The SCC value $r_{scc}(S_1, S_2)$ between two attribute sequences is equal to the PCC value between these two rank sequences. For m attribute sequences $\{S_1, S_2, \dots, S_m\}$, the weight of S_i is negatively correlated to the sum of the absolute value of SCC between S_i and all others.

The spatial attribute fusion algorithm based on Spearman correlation coefficient is shown in Algorithm 2. Likewise, an attribute matrix $M^s \in \mathbb{R}^{n \times 3}$ is formed as $M^s = \begin{bmatrix} A_{pos}(p) & A_{sha}(p) & A_{tre}(p) \end{bmatrix}$, where the meanings of p and n remain unchanged.

Algorithm 2: Spatial attribute fusion based on Spearman correlation coefficient

1. **Input:** A attribute matrix M^s , the number of attributes n_A
 2. $T \leftarrow \emptyset$
 3. $R \leftarrow \emptyset$
 4. $W^s \leftarrow \emptyset$
 5. **for** $i = 1$ to n_A **do**
 $T \leftarrow M^s$
 delete $T_{*,i}$
for $j = 1$ to $n_A - 1$ **do**
 $R_i \leftarrow r_{scc}(T_{*,i}, T_{*,j})$
end for
 $W_i^s \leftarrow 1 - \frac{1}{n} \sum |R_i|$
end for
 6. Normalize W^s to the range of 0 to 1
 7. **Output:** weight vector W^s
-

The value of n_A is set to 3. After calculating the weight vector W^s , the node relevance assessment index can be given as:

$$I_S(i, j) = \sum_{j=1}^{n_A} M_{i,j}^s W_j^s, i \in [1, n] \quad (10)$$

It is worth noting that values of $I_S(i, j)$ and $I_S(j, i)$ are a bit different due to the change of reference nodes in grey relational degrees computation; their average value is used for the similarity measurement in the next section.

5. Critical Node/Region Identification via Importance Ranking and Synthesized Similarity

After spatio-temporal attribute extraction and fusion, the importance of node and relevance between any two nodes in a specific time segment can be depicted clearly. The critical nodes are defined as those with a large importance score and thus can be identified directly by sorting the node importance in descending order. Generally speaking, we could deem the top-N nodes as critical nodes or set a threshold. For example, if the importance score of a node is larger than 0.7 (the score of the

largest one is 1), it is considered critical. The critical regions are usually critical nodes' centric areas with nodes subjected to relative large and similar influences. Clustering is an unsupervised learning technique that places similar data into identical groups [24]. It is a feasible approach to identifying critical regions. One of the key problems is that there are several special types of nodes in a regional power grid. For example, the voltage phasor of the only slack bus/node and the voltage amplitudes of the generator node keep constant, but the voltage phasors of the critical node usually show stronger mobility than other nodes. A promising clustering algorithm should have the ability to separate these nodes, hence a so-called DBSCAN [25] was selected in this study because of its ability to find arbitrarily-shaped clusters and isolate outliers. Another key problem in achieving effective clustering is similarity measurement. Nodes in a critical region are expected to own large importance values as well as strong linkages. Therefore, their similarities should be an integration of both temporal and spatial attributes, represented as:

$$S(\text{node}_i, \text{node}_j) = |I_T(i) - I_T(j)| \{1 - [I_S(i, j) + I_S(j, i)]/2\} \quad (11)$$

Obviously, if two nodes are similar, the difference in their importance value approaches 0, and the average of their relevance value approaches 1. DBSCAN sets a radius parameter, ε , and a threshold of neighborhood density, $Mpts$. If the number of similar neighborhoods ($S(\text{node}_i, \text{node}_j) < \varepsilon$) of a central node is less than $Mpts$, it is labeled as a special node. These special nodes form new clusters or merge into other clusters according to the following rules:

- If the importance score of a special node is larger than the maximum score in the cluster with the largest average score, it forms a new critical cluster.
- If the importance score of a special node is lower than the minimum score in the cluster with the lowest average score, it forms a new non-critical cluster.
- If the importance score of a special node is between two above limits, it is assigned to the cluster with the most similar average score.

The criteria of the critical region are the average importance scores among all nodes in a cluster. In this paper, the final critical region is comprised of the aforementioned critical cluster and the cluster with the largest average importance score.

The parameter setting of DBSCAN has a direct impact on the final results, and it is subjective. Therefore, different situations should have different settings. The identification problem does not have a fixed evaluation method, thus the knowledge and experience of the operator are important references for acceptable settings. Generally speaking, larger ε and smaller $Mpts$ merges some similar clusters, leading to a large area of the critical region. If the ε decreases and $Mpts$ increases, the area of the critical region shrinks. The main critical region is not affected. In practical application, the operator can set reference values with large ε and small $Mpts$, adjust these parameters, and narrow the area of the critical region from large to small until a satisfying result is obtained.

6. Case Studies and Discussion

The IEEE-39 benchmark system is employed to validate the applicability of the proposed scheme. This system represents a real regional power grid in New England, and its structure is shown in Figure 4. IEEE-39 system contains 10 nodes with generators, 17 nodes with load, and 12 concatenation nodes. The node numbered 31 is a slack bus, which means its voltage amplitude keeps unchanged, and its phase angle is 0.

The continuation power flow (CPF) is used to generate experimental data. To simulate specific changes of operating conditions in this power grid, loads on selected nodes increase at the same rate, their power factors remain constant, and all electric generators contribute jointly to balance the load increment [26].

In this part, three different cases were used for experiments. Samples for each case were simulated using the MATPOWER CPF tool (Version 5.1, Power System Engineering Research Center (PSERC)

at Cornell University, Ithaca, NY, United States) [27]. Voltage amplitude and phase angle for each sample were extracted and represented as voltage phasors. The reference time could be set at any time as a baseline. In the following cases, in order to enhance performance, we set it at the fifth sample (any position near the starting point is acceptable) to maximize the variations. The width of the voltage phasor sequence was determined according to the sampling rate of the data acquisition system. In general, a wider voltage phasor sequence is more accurate but also needs longer computation time. The node was set to 7 in the following case. Operators could select proper parameters according to experiences and practical conditions to improve the accuracy and reduce the time complexity of the proposed scheme.

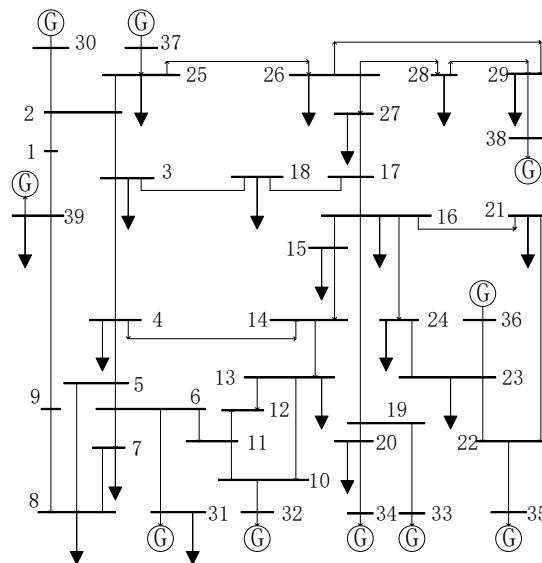


Figure 4. IEEE-39 system.

6.1. Case A: Load Increase on Node 7

In this case, node 7 was selected as the disturbance source. Its load active and reactive power increased from 2.34 MW and 0.84 MVar to 17.04 MW and 6.12 MVar until the static voltage stability margin was reached. A total of 96 samples for each node were simulated. To show the variation of the identified critical nodes and regions, results for three target times T at the 25th, the 50th, and the 80th samples are illustrated in Figure 5. The representative color of critical nodes are shades from red to yellow and then to green. Likewise, clusters are divided into six levels (L1 to L6), from red to cyan green (red for critical cluster and cyan green for non-critical cluster). Critical regions are comprised of L1 clusters and clusters formed by special nodes with large importance scores. Table 1 lists the top 10 critical nodes with their importance scores in case A explicitly.

Generally, load increment in certain node results in voltage reduction and enlargement of phase angle difference. These variations radiate from the center node along with grid structure. The range of radiation is regarded as the critical regions, and the nodes in those regions are critical nodes with different severities, which are measured by their importance scores.

In Figure 5a, at $T = 25$, critical nodes were mainly located on the left side of the grid. It was obvious that the source was node 7, while node 8 and node 9 were critical nodes that suffered from significant impact. Further, at $T = 50$, the rankings of nodes 2, 25, and 30 rose, indicating the propagation direction of disturbances. Finally, nodes 26 and 28 at the upper right corner became two of the top 10 critical nodes at $T = 80$. In Figure 5b, the critical region is identified and marked with light orange background color. Members in the critical region were similar in temporal variation and spatial distribution. It was easy to find out that the range of the critical region expanded over time according to its propagation path from the left side to the right side. Although our method did not

consider the grid structure, the above proved that the variation of the voltage phasor sequence fulfilled these topological constraints naturally.

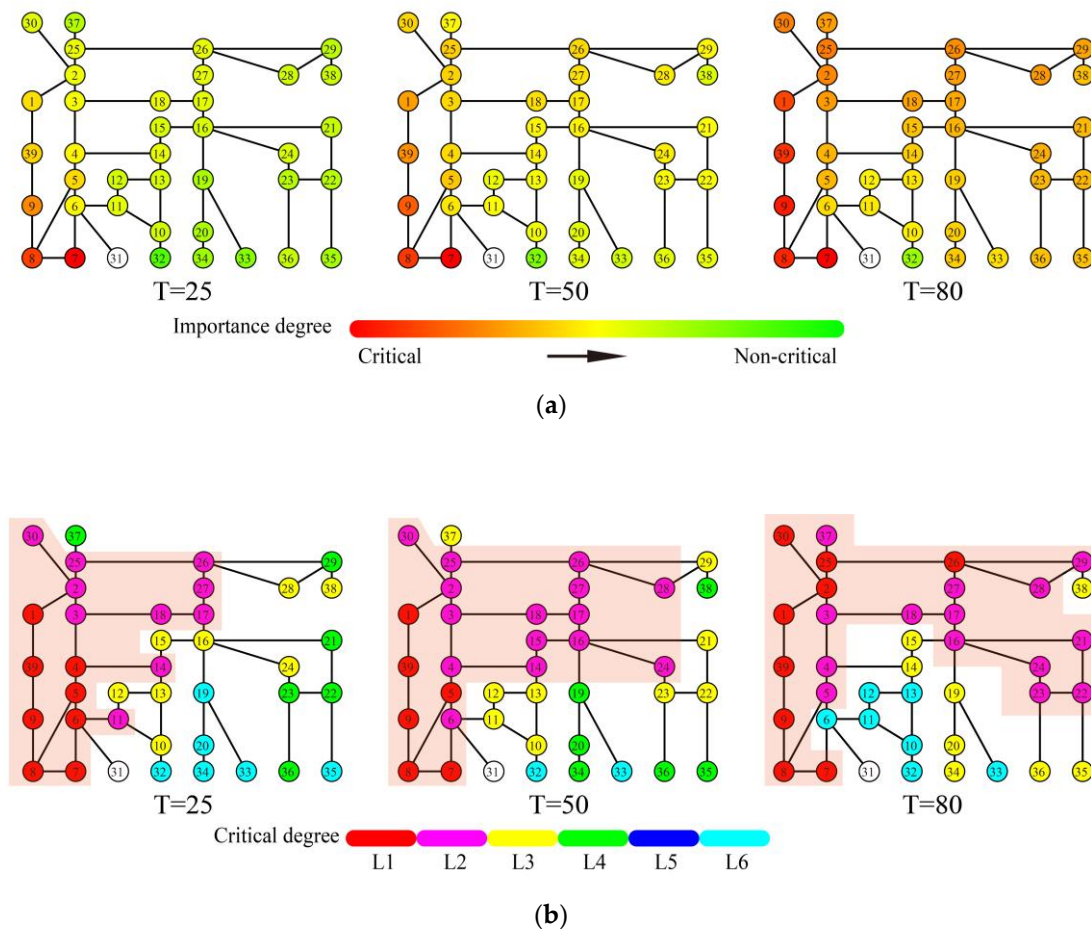


Figure 5. Dynamic critical nodes and regions identification results for case A. (a) the critical nodes; (b) the critical regions.

Table 1. Top 10 critical nodes with their importance score in case A.

Rankings	1	2	3	4	5	6	7	8	9	10
$T = 25$	7(1.00)	8(0.87)	9(0.73)	39(0.58)	5(0.57)	1(0.56)	6(0.51)	4(0.51)	3(0.49)	2(0.47)
$T = 50$	7(1.00)	8(0.89)	9(0.82)	39(0.72)	1(0.69)	5(0.61)	25(0.60)	2(0.59)	30(0.58)	3(0.58)
$T = 80$	7(1.00)	9(0.95)	8(0.93)	39(0.90)	1(0.86)	25(0.76)	30(0.74)	2(0.73)	26(0.72)	28(0.70)

A stable power system always keeps a balance between power supply and demand. Therefore, generators contribute power for demands in disturbance sources. The critical generator nodes, which are marked in bold in Table 1 and all following tables, are the main power suppliers. Apparently, during the early phase, node 39 made major contributions. Then, nodes 30 and 37 joined in after $T = 50$. The final supplier team was mainly comprised of these three nodes.

6.2. Case B: Load Increase on Node 7 And 28

In this case, nodes 7 and 28 were selected as the disturbance sources, and their load demands increased simultaneously. The load active and reactive power of node 7 increased from 2.34 MW and 0.84 MVar to 12.50 MW and 5.79 MVar; the load active and reactive power of node 28 increased from 2.06 MW and 0.28MVar to 8.41 MW and 1.13MVar. A total of 128 samples for each node were simulated. The critical nodes and regions identification results for three target times T at the 20th,

the 60th, and the 120th samples are illustrated in Figure 6. Table 2 lists the top 10 critical nodes in case B.

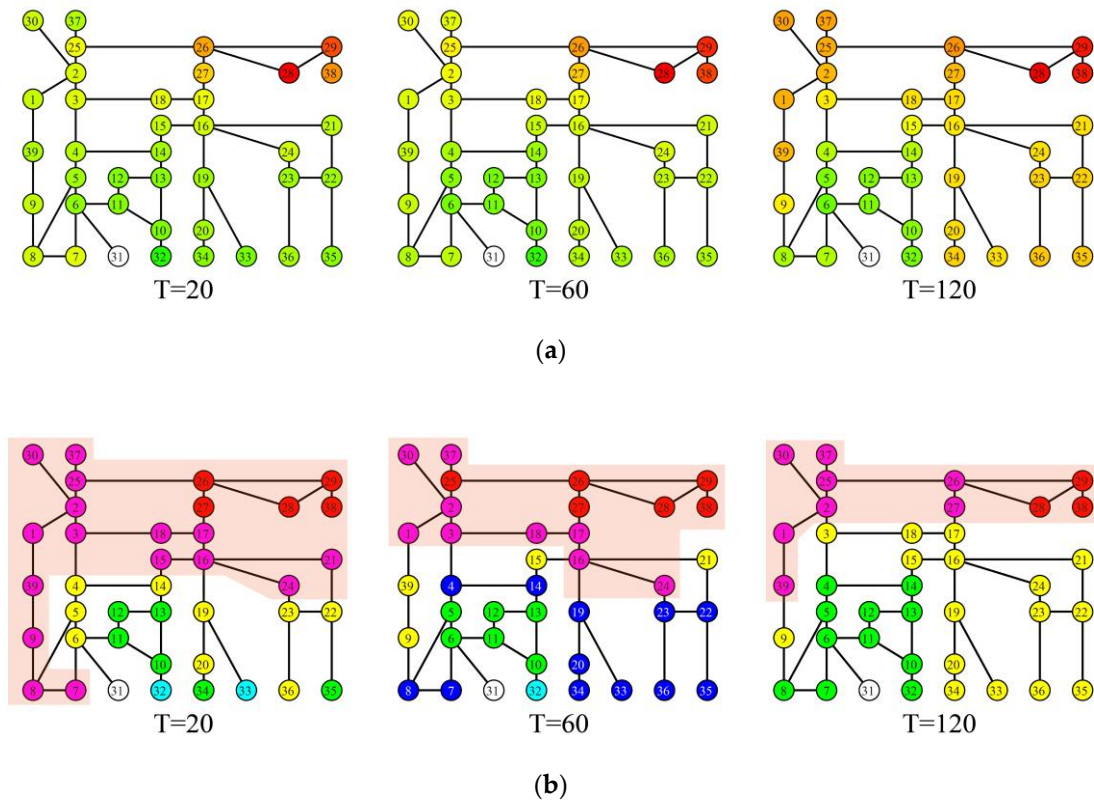


Figure 6. Dynamic critical nodes and regions identification results for case B. (a) The critical nodes; (b) the critical regions.

Table 2. Top 10 critical nodes with their importance score in case B.

Rankings	1	2	3	4	5	6	7	8	9	10
$T = 20$	28(1.00)	29(0.85)	38(0.71)	26(0.68)	27(0.59)	25(0.48)	17(0.47)	18(0.46)	7(0.44)	3(0.44)
$T = 60$	28(1.00)	29(0.92)	38(0.87)	26(0.69)	27(0.60)	25(0.52)	2(0.48)	17(0.48)	30(0.47)	18(0.47)
$T = 120$	28(1.00)	29(0.96)	38(0.95)	26(0.71)	30(0.68)	37(0.66)	25(0.65)	1(0.64)	39(0.64)	2(0.64)

At $T = 20$, from Figure 6a and Table 2, node 28 and its adjacent nodes had high importance scores. The main power supplier was node 38. Although the power demand in node 7 kept increasing, its voltage phasor changed slightly. This illustrated that node 7 was more robust than node 28. The critical region in Figure 6a connected both disturbance sources. Later, at $T = 60$, the influence of node 28 moved left, node 30 supplied more power, and the severity of node 7 fell again. All of them brought an upward shift to the critical region. At $T = 120$, more generator nodes became critical nodes. Although the critical region seemed to shrink, the number of nodes in the L3 cluster grew, thus nodes in the lower right corner tended to become critical.

6.3. Case C: Load Increase on All Load Nodes

In this case, all load nodes were selected as the disturbance sources. A total of 160 samples for each node were simulated. The critical nodes and regions identification results for three target times T at the 40th, the 90th, and the 140th samples are illustrated in Figure 7. Table 3 lists the top 10 critical nodes in case C.

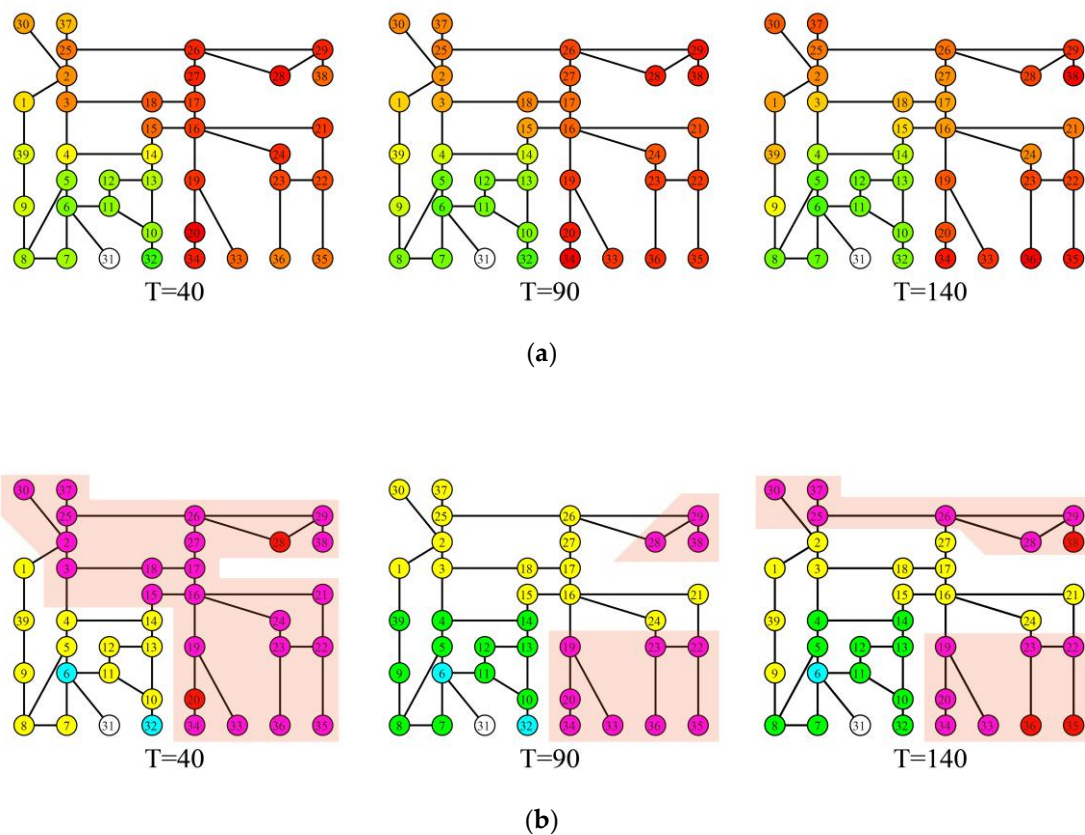


Figure 7. Dynamic critical nodes and regions identification results for case C. (a) The critical nodes; (b) the critical regions.

Table 3. Top 10 critical nodes with their importance score in case C.

Rankings	1	2	3	4	5	6	7	8	9	10
$T = 40$	20(1.00)	28(0.98)	34(0.95)	29(0.94)	27(0.94)	26(0.93)	24(0.92)	19(0.91)	21(0.89)	16(0.89)
$T = 90$	34(1.00)	29(0.96)	28(0.95)	20(0.95)	39(0.95)	36(0.92)	35(0.91)	33(0.90)	19(0.90)	23(0.89)
$T = 140$	36(1.00)	38(0.99)	35(0.96)	34(0.96)	29(0.91)	33(0.91)	28(0.88)	22(0.87)	23(0.86)	20(0.84)

In the initial phase, namely $T = 40$, critical nodes distributed in the right half of the grid. Compared with the above two cases, the downtrend of importance scores in top 10 critical nodes was much slower. Starting from nodes 16 and 17, the severity of nodes rose along two paths—16 to 34 and 17 to 29. These two nodes played parts as concatenation nodes for two critical regions, which is quite distinct in the first chart of Figure 6b. At $T = 90$, two critical regions split up. At $T = 140$, the upper part extended left, while the lower part remained unchanged. The reason was explicable. From Table 3, we know that there were five generator nodes in the top 10 critical nodes. Four of them were located on the lower region. Thus, the demands in the upper region required extra suppliers.

In summary, three cases were tested and analyzed to validate the applicability of the proposed method in different aspects. Case A simulated a situation where only one load node increased its power demand. Results show that the proposed method could identify the source and depict the whole propagation process accurately. In case B, two nodes increased their power demand simultaneously. Their relative robustness could be determined by their critical rankings. In case C, all load nodes increased their power demands. The correctness of identification results is proven by the principle of the supply-demand balance. In short, based on the spatio-temporal attribute fusion of voltage trajectory, our scheme could implement dynamic identification of critical nodes and regions quickly and precisely for different situations.

In practical application, the critical nodes and regions need to be identified online. Therefore, the proposed method should meet the requirements of accuracy and efficiency. Two factors influence the execution time severely—the width of the voltage phasor sequence and the size of the power grid. The growth of the first factor increases the complexity of calculating the grey relational degree in the spatial attribute extraction procedure, and the growth of the second factor enlarges the amount of computation of the similarity matrix for DBSCAN. To investigate the impact of these two factors, the computation time of each procedure should be recorded. Experiments were implemented on a desktop PC with Intel Core i7-6700 CPU, 16GB RAM, and 3.40GHz clock speed for case C. Firstly, the width of the voltage phasor sequence was set to seven. The computation time of temporal attribute extraction, spatial attribute extraction, attribute fusion, and critical nodes/regions identification were 0.00364 s, 1.2936 s, 0.00725 s, and 0.00288 s. Next, if we changed the width from seven to 20, the time became 0.00470 s, 3.3391 s, 0.00747 s, and 0.00254 s. Obviously, the first factor and the spatial attribute extraction procedure were the main computation burdens. Further, the variation trend of the total computation time along with the width of the voltage phasor sequence were explored and are shown in Figure 8.

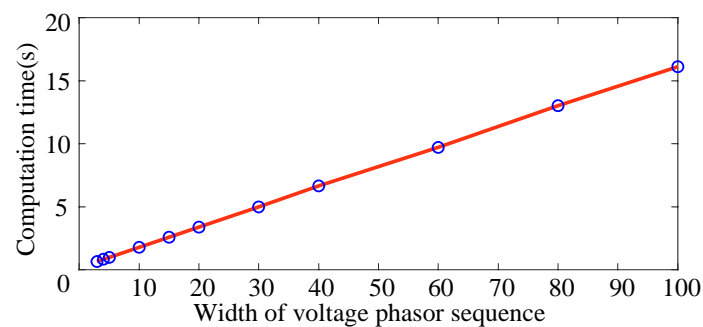


Figure 8. The trend of computation time.

In this figure, we can learn that the computation time grows linearly along with the width of the voltage phasor sequence. Therefore, in real application, the width should be carefully selected to meet the requirements.

7. Conclusions

In this paper, a novel method for dynamic identification of critical nodes and regions in a power grid was proposed. Compared with existing research, this paper focused on the variation of and similarity between node voltage phasor sequences, which is a section of voltage phasor trajectories rather than the topological structure of a power grid. The proposed method could identify and update critical nodes and regions dynamically according to acquired real-time data. Three temporal attributes including distance, velocity, and acceleration were extracted and integrated to assess node importance. Three spatial attributes including position, shape, and trend were extracted and integrated to assess node relevance. The critical nodes were determined by their importance rankings, and the identification of critical regions was realized by clustering nodes with similar importance and high relevance. Case studies on the IEEE 39 benchmark system for different situations were tested and analyzed, and results show that this method could locate disturbance sources, reveal propagation mechanisms, and achieve effective and accurate dynamic identification. Future research will further integrate the topological attributes, establishing a panoramic depiction of both static and dynamic characteristics of all critical nodes and regions in a power grid.

Author Contributions: All authors participated in scheme conception, design and demonstration. The first author completed the experiments and paper writing.

Acknowledgments: This work was supported by the National Nature Science Foundation of China under Grant U1701262.

Conflicts of Interest: Authors declare no conflicts of interest.

References

1. Fan, Y.; Xin, L.; Peng, Z.; Fan, Y.; Xin, L.; Peng, Z. Real-Time Static Voltage Stability Assessment in Large-Scale Power Systems Based on Maximum-Relevance Minimum-Redundancy Ensemble Approach. *IEEE Access* **2017**, *5*, 27281–27291. [[CrossRef](#)]
2. Yan, J.; He, H.; Yan, S. Integrated Security Analysis on Cascading Failure in Complex Networks. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 451–463.
3. Bompard, E.; Di, W.; Fei, X. Structural vulnerability of power systems: A topological approach. *Electr. Power Syst. Res.* **2011**, *81*, 1334–1340. [[CrossRef](#)]
4. Pagani, G.A.; Aiello, M. The Power Grids as a complex network: A survey. *Phys. A: Stat. Mech. Its Appl.* **2013**, *392*, 2688–2700.
5. Strogatz, S.H. Exploring complex networks. *Nature* **2001**, *410*, 268–276. [[CrossRef](#)] [[PubMed](#)]
6. Fan, W.; Liu, Z.; Ping, H. Identifying node importance based on information entropy in complex networks. *Phys. Scr.* **2013**, *88*, 5201.
7. Chen, D.; Lu, L.; Shang, M.-S.; Zhang, Y.-C.; Zhou, T. Identifying influential nodes in complex networks. *Phys. A: Stat. Mech. Its Appl.* **2012**, *391*, 1777–1787. [[CrossRef](#)]
8. Avrachenkov, K.; Litvak, N. The Effect of New Links on Google Pagerank. *Commun. Stat. Stoch. Models* **2006**, *22*, 319–331.
9. Li, C.; Liu, W.; Cao, Y.; Hao, C.; Shi, H. Method for evaluating the importance of power grid nodes based on PageRank algorithm. *IET Gener. Transm. Distrib.* **2014**, *8*, 1843–1847.
10. Huang, J.K.; Liang, D.; Zhang, G.S. Skeleton-Network Reconfiguration Based on Node Importance and Line Optimization. In Proceedings of the Power & Energy Engineering Conference, EPEC 2017, Saskatoon, SK, Canada, 22–25 October 2017; pp. 1–5.
11. Champion, M.; Ranganathan, P. Identification of critical buses based on betweenness-centrality in a smart grid. In Proceedings of the 2017 IEEE Electrical Power and Energy Conference, EPEC, Saskatoon, SK, Canada, 22–25 October 2017; pp. 1–5.
12. Wang, H.; Shan, Z.; Ying, G.; Zhang, B.; Zou, G.; Benteng, H.E. Evaluation method of node importance for power grid considering inflow and outflow power. *J. Mod. Power Syst. Clean Energy* **2016**, *5*, 1–8.
13. Liu, Z.; Fan, W.; Hu, P. A multi-attribute node importance evaluation method based on Gini-coefficient in complex power grids. *IET Gener. Transm. Distrib.* **2016**, *10*, 2027–2034.
14. 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* **2018**, *65*, 206–210.
15. Nguyen, D.T.; Shen, Y.; Thai, M.T. Detecting Critical Nodes in Interdependent Power Networks for Vulnerability Assessment. *IEEE Trans. Smart Grid* **2013**, *4*, 151–159. [[CrossRef](#)]
16. Zaranđi, F.D.; Rafsanjani, M.K. Community detection in complex networks using structural similarity. *Phys. A: Stat. Mech. Its Appl.* **2018**, *503*, 882–891. [[CrossRef](#)]
17. Pérez-Londoño, S.M.; Olivar-Tost, G.; Mora-Florez, J.J. Online determination of voltage stability weak areas for situational awareness improvement. *Electr. Power Syst. Res.* **2017**, *145*, 112–121. [[CrossRef](#)]
18. Musirin, I.; Rahman, T.K.A. Voltage stability based weak area clustering technique in power system. In Proceedings of the Power & Energy Conference, Kuala Lumpur, Malaysia, 29–30 November 2004.
19. Qi, D.; Deng, H.; Di, Y. Weak voltage area recognition method in power system based on fuzzy subtract clustering. In Proceedings of the International Conference on Fuzzy Systems & Knowledge Discovery, Yantai, China, 10–12 August 2010.
20. Hu, F.; Kai, S.; Rosso, A.D.; Farantatos, E.; Bhatt, N.B. Measurement-based real-time voltage stability monitoring for load areas. *IEEE Trans. Power Syst.* **2016**, *31*, 2787–2798.
21. Pakkar, M.S. Multiple attribute grey relational analysis using DEA and AHP. *Complex Intell. Syst.* **2016**, *2*, 243–250.
22. Abughali, I.K.A.; Minz, S. *Binarizing Change for Fast Trend Similarity Based Clustering of Time Series Data*; Springer: Berlin/Heidelberg, Germany, 2015.
23. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1987**, *100*, 441–471. [[PubMed](#)]

24. Aghabozorgi, S.; Shirkhorshidi, A.S.; Wah, T.Y. *Time-series clustering—A decade review*; Elsevier: Amsterdam, The Netherlands, 2015.
25. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the International Conference on Knowledge Discovery & Data Mining, Portland Oregon, OR, USA, 2–4 August 1996.
26. Fan, Y.; Liu, S.; Qin, L.; Li, H.; Qiu, H. A Novel Online Estimation Scheme for Static Voltage Stability Margin Based on Relationships Exploration in a Large Data Set. *IEEE Trans. Power Syst.* **2015**, *30*, 1380–1393. [[CrossRef](#)]
27. Zimmerman, R.D.; Murillo-Sanchez, C.E.; Thomas, R.J. MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education. *IEEE Trans. Power Syst.* **2011**, *26*, 12–19. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).