Electronic Research Archive

ERA, 31(3): 1524–1542.
DOI: 10.3934/era.2023077
Received: 13 September 2022
Revised: 26 December 2022
Accepted: 03 January 2023
Published: 17 January 2023

http://www.aimspress.com/journal/ERA

Research article

Critical node identification in network cascading failure based on load percolation

Hangyu Hu, Fan Wu, Xiaowei Xie, Qiang Wei, Xuemeng Zhai*, Guangmin Hu

School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

* **Correspondence:** Email: zxm@uestc.edu.cn; Tel: +8613808097337; Fax: +862861830209.

Abstract: Identification of network vulnerability is one of the important means of cyberspace operation, management and security. As a typical case of network vulnerability, network cascading failures are often found in infrastructure networks such as the power grid system, communication network and road traffic, where the failure of a few nodes may cause devastating disasters to the whole complex system. Therefore, it is very important to identify the critical nodes in the network cascading failure and understand the internal laws of cascading failure in complex systems so as to fully grasp the vulnerability of complex systems and develop a network management strategy. The existing models for cascading failure analysis mainly evaluate the criticality of nodes by quantifying their importance in the network structure. However, they ignore the important load, node capacity and other attributes in the cascading failure model. In order to address those limitations, this paper proposes a novel critical node identification method in the load network from the perspective of a network adversarial attack. On the basis of obtaining a relatively complete topology, first, the network attack can be modeled as a cascading failure problem for the load network. Then, the concept of load percolation is proposed according to the percolation theory, which is used to construct the load percolation model in the cascading failure problem. After that, the identification method of critical nodes is developed based on the load percolation, which accurately identifies the vulnerable nodes. The experimental results show that the load percolation parameter can discover the affected nodes more accurately, and the final effect is better than those of the existing methods.

Keywords: network vulnerability identification; network cascading failure; critical node identification; percolation theory; network management

1. Introduction

Identification of the network vulnerability is an important part of cyberspace confrontation. As one of the applications of the network topology awareness, network vulnerability identification takes the network structure, nodes and link attributes as inputs, and it identifies the critical nodes or links to support network attacks as the outputs. The network cascading failure is a typical network vulnerability [1]. In a complex network, the failure of one or several node(s) leads to the redistribution of the network load, which in turn leads to the overload of other nodes that makes them also fail, thus causing a chain reaction and, finally, cascading failure [2]. The cascading failure is often found in infrastructure networks such as the power grids, communications, road traffic, etc. The failure of a few nodes may cause devastating damage to the entire complex system. Therefore, studying the identification of critical nodes in cascading failure, along with obtaining an understanding of its internal rules of cascading failure in complex systems, is very important in fully mastering the vulnerability of complex systems, and it is one of the research highlights that has been addressed by academic and industrial circles at home and abroad in recent years.

The existing cascading failure models of complex network structures mainly evaluate the criticality of nodes by quantifying the importance of the nodes in the models. The research on such node importance measurement is mainly divided into two categories [3]. The first category directly calculates the node importance through node centrality, which is directly calculated according to parameters of the network structure, such as the degree, betweenness, PageRank value, Hits value, etc. The second category deletes the critical nodes through optimization strategies, and it evaluates the impact of cascading failure caused by node deletion to measure their importance. The former method based on node centrality often only considers the importance of nodes of the network structure and ignores important attribute factors such as the load and node capacity in the cascading failure model. Thus, the evaluation performance in real networks often has achieved a poor effect. The latter method can enumerate all of the results and find the most influential nodes by means of simulation. However, for medium-sized networks, the time complexity of this second method is very high due to the large number of nodes and complex relationships, which increases the difficulty of identifying critical nodes in the cascading failure model.

In order to resolve the above drawbacks, this paper proposes a novel method for identifying critical nodes based on load percolation. The percolation theory is an innovative method for identifying influential nodes in large-scale complex networks. Different from heuristic methods, such as those using centrality to measure the importance of nodes, the percolation theory simulates the percolation process of the fluid in the network and takes the optimal percolation as the goal to find the minimum number of critical nodes that may damage the network. Therefore, in the proposed cascading failure model, the percolation theory is taken as the structural basis, in combination with the specific attributes of cascading failure, so as to avoid the problem of poor performance of the heuristic centrality algorithm, and to find the most optimal nodes of cascading failure in medium-sized networks in a better way.

This paper first gives a basic overview of network cascading failure, and then it introduces the method for measuring node influence based on the percolation theory. After that, this paper proposes the concept of load percolation, which combines the common influence of the network structure and properties to yield a load percolation model for the cascading failure problem. Finally, based on the load percolation, a method is proposed for the mining of optimal nodes via identification of the critical

nodes that may cause network cascading failures. The obtained experimental results show that the proposed method is superior to the existing methods in terms of the accuracy of node identification and the impact of cascading failure.

The rest of this paper is arranged as follows. Section 2 introduces the related work; Section 3 introduces the basic problems of the network cascading failure. Section 4 gives the critical node identification parameter set algorithm for the load network. In Section 5, experiments with a real network are described to verify the effectiveness of the algorithm. Finally, Section 6 summarizes the work of this paper and discusses the future research scopes.

2. Related work

In order to obtain a relatively complete and accurate network structure, we can further analyze the network structure for obtaining valuable information, including the types of identified nodes, predicted connection relationship and types of identified networks, as well as perform structural vulnerability analysis, etc. The structural vulnerability analysis usually uses the complex network theory to identify vulnerable node sets in complex systems, explain the internal mechanism of complex system collapse and study preventive and controlling measures. The structural vulnerability analysis is closely related to network attacks, involving the network robustness, critical node identification and other related concepts. This paper focuses on the identification of nodes that may destroy the overall function of the network.

For a single node, the centrality of the node is an effective indicator to measure its vulnerability. Accordingly, removing the top-k nodes from the top to bottom by using the centrality indicator is a way to destroy the network function in a scenario [4]. Studies conducted so far have proposed a variety of node centrality, including the degree centrality, betweenness, coreness, closeness, etc. In literature [3,5], Lü et al. and Costa et al. systematically introduced the node centrality. Since the mutual relationship between nodes is not considered, the centrality of nodes suffers from poor performance in finding vulnerable nodes [3]. Thus, a new method for locating vulnerable nodes has emerged. Finding the vulnerable nodes is usually modeled as the influence maximization problem (IMP). Corley and Sha [6] studied the IMP in the maximum flow of the network, and they also studied the IMP related to the shortest path distance. Since then, a variety of heuristic methods have been proposed [7–10]. It is worth noting that Morone and Makse [11] linked the IMP with network percolation and pointed out that the IMP can be solved through the use of network percolation.

Using the above-mentioned method to analyze the node vulnerability of complex networks, researchers have focused on practical networks such as power grids and transportation networks. Panigrahi and Maity [12] used weighted networks to model the power grid network and evaluated the critical nodes and links; Beyza et al. [13] proposed a method for evaluating cascading faults in power grids; Fang et al. [14] proposed a method for identifying the critical links in the power grids based on the maximum network flow; Zhang et al. [15] compared and analyzed the railway networks of China, the United States of America and Japan to identify the vulnerable stations in the system from the perspective of complex networks, and they proposed some effective defense strategies.

In addition to the above-mentioned vulnerability analysis methods for static networks, researchers have paid attention to the vulnerability of dynamic networks. Cascading failure is a typical dynamic network vulnerability, and it refers to the process of failure of a single node that causes failure of other nodes in the network. In the case of modeling, Motter and Lai [2] proposed a classical cascading failure

model fault propagation model by studying the relationship between the node capacity and load. Wu et al. [16] improved the fault propagation model for urban traffic problems. Wang et al. [17] analyzed the behavior of the fault propagation model in scale-free networks. In terms of the vulnerable nodes, they evaluated three removal strategies, pointing out that the attacking effects of different removal strategies on networks are greatly different. Yang et al. [18] identified the nodes vulnerable to cascading failure in a North American power grid, and they pointed out that the set of the identified nodes was composed of a small number of nodes with high topological centrality. From the perspective of network evolution, Holme and colleagues [19,20] studied the cascading failures induced by BA scale-free network edge overload and node overload caused by network growth. Based on the sand pile model, Huang et al. [21] studied the robustness of scale-free networks embedded with weighted grids and found that, in scale-free networks with regional limitations, tight local connections are more likely to cause global network collapse. Dobson et al. proposed the classic OPA model [22] and CASCADE model [23] to solve the cascading failure happening in power grids.

3. Overview of the basic problems of network cascading failure

Real complex systems for analysis can be modeled as complex networks, such as the Internet, power grid systems, aviation networks, transportation networks and social networks. Complex networks are not only a form of data representation, but also a means of scientific research, providing a theoretical basis for understanding the behavioral models of complex systems. Research on complex networks has currently received extensive attention and research.

In reality, the activity of each node in the network is related to the activities of neighboring nodes. In a complex system, the process in which the failure of one or several component(s) causes the failure of other components is called the cascading failure [24]. When one part of the system fails, other parts need to be reallocated to compensate for the failed part, which in turn may overload those nodes to induce failure, causing more nodes to fail successively [2,25]. The whole reallocation process would be repeated until there are no invalid nodes exist.

Cascading failures may occur in many types of complex systems, including the power grid system, computer network, finance system, transportation system, human body and ecosystems [24]. In the Internet, the disconnection of critical routing nodes due to hardware or software failure may severely damage or stop most of the network communication. After the routing nodes go down, the routing protocol needs to re-plan the traffic through another alternative path. Thus, the failure of one router will increase the traffic on other routers, which may cause the alternative path to become overloaded, resulting in cascading failures. In a large-scale power grid network, when one of the components fails completely or partially and transfers its load to nearby components, it may cause those components to exceed their capacity and get overloaded, causing a wider range of overloading that eventually paralyzes the power grid in a short time [26]. For instance, on August 10, 1996, a 1300 MW power line in Oregon was broken, and the current it carried was automatically diverted to two other transmission lines with slightly lower voltages, causing it to become overloaded and fail; the excess current caused 13 generators to fail, which resulted in massive power cuts in 11 states of the USA and two provinces in Canada. Therefore, it is of great significance to study the inherent characteristics of cascading failure to mitigate the effects of network cascading failure and improve the robustness against cascading failure.

Real cascading failure has three common characteristics: first, the initial failure has a very limited

impact on the network structure; second, the initial failure does not stay local, as it spreads along the connection of the network, causing more faults; finally, the propagation of faults makes many nodes unable to function properly. This paper will first explore the patterns of the cascading failure from an empirical perspective and then model it mathematically.

3.1. Cascading failure models

In the recent literature, many models of network cascading failure are proposed based on the initial conditions of a single isolated network. Among these models, at the initial moment, each node is given an initial load and maximum load handling capacity (also known as load capacity); then, the load of the failed node would be redistributed according to the corresponding theoretical model. Through the preliminary analysis of the cascading failure model, it can be found that there are three main factors which have influence on the result of a cascading failure, including the initial load of the node, the capacity of the node and the load allocation strategy.

As mentioned in the related work, although there are many researchers who have done more in-depth research, the reasonable modeling of those three factors is the key initiatives in the exploration of a cascading failure. For example, Wang et al. [27] redefined the initial load of the node based on the degree of the node itself and the degree of the neighbor node, considering the influence of the degree of the neighbor node on the node load. Their method not only avoids the complexity associated with obtaining the node's betweenness centrality, but it also improves the practicability based on the node's own degree. Through the study of multiple real networks, Kim et al. [28] found that there was no linear relationship between capacity and load. And, this conclusion stimulates the consideration of the load-capacity relationship.

Load redistribution is the most important way to alleviate network cascading failure after nodes are attacked or randomly fail, and it is also the last barrier to ensure the normal operation of the network. A reasonable load redistribution strategy can greatly reduce the failure scale of network nodes. The existing load redistribution strategies mainly include the average distribution strategy, random distribution, global distribution strategy based on the shortest path, local optimal distribution and adjustable load redistribution. Obviously, the average distribution and random distribution do not perform load redistribution since they consider little about the specific conditions of the network. So, it is easy to cause further expansion of cascading failures. However, the global allocation strategy based on the shortest path can redistribute the load to the remaining nodes according to the processing capacity of the nodes, but this allocation strategy requires each node to manage the global information, which is difficult to apply to large-scale networks. The local optimal redistribution strategy redistributes the load of the failed node locally according to the load ratio of its neighboring nodes, and it does not require global information, so it can be used in large-scale networks.

3.2. Critical node identification based on the percolation theory

The percolation theory studies the behavior of networks in complex systems after the nodes or edges are removed, which is one of the fundamental theories of network phase transition in complex systems [29], becoming a major concern in network science. Recently, with the development of the theory, the percolation theory not merely focuses on the single-type network, but it also sets the multilayer network in the research category [30–34]. In this paper, we mainly consider the single-type

network in which the existing modeling methods ignore the important load, node capacity and other attributes. The effect of node or edge removal is usually measured in terms of the number of nodes in the largest connected branch. We set p_c as the critical percolation probability of the network and p as the proportion of nodes removed in a network. When p_c is large, the overall robustness of the network is considered better; otherwise, it is worse. In the case of $p > p_c$, the largest connected branch of the network disappears and splits into several small connected branches, resulting in network function failure. Moreover, the percolation theory finds that scale-free networks are robust against the removal of random nodes, but they are fragile if high-degree nodes are removed.

The cascading failure model is a typical independent percolation process [11], and it can be solved by the percolation theory. Therefore, this paper transforms the problem of finding the critical nodes, which is responsible for network cascading failures, into a problem of identifying the critical nodes in the network percolation, and it proposes an effective critical node identification method in the current percolation theory.

In a network, we set the optimal set U^* of the critical nodes as the minimum set that leads to the network function failure after removal. However, solving the problem of U^* is an *N*-body problem that needs to consider the topological interactions between nodes, which makes it an NP-hard problem [3]. It is common to conduct heuristic methods through the centrality of a single node, such as the node degree, betweenness, etc. However, since the centrality of individual nodes is a fuzzy definition of the influence of the nodes, which treats the nodes as isolated entities and ignores the interactions between them, it often achieves poor performance in the solution.

In order to obtain the optimal percolation solution in a large-scale network, Morone and Makse [11] proposed an approximate collective influence (CI) algorithm, which is the most efficient method for solving U^*. The CI value of a node is defined as

$$CI_l(i) = (k_i - 1)\sum_{j \in \partial Ball(i,l)}(k_j - 1) \tag{1}$$

where k_i represents the degree of the node i in a network; $Ball(i, l)$ represents all nodes within the shortest path distance l from node i; $\partial Ball(i, l)$ represents the nodes with the shortest path distance l from node i.

Compared with the node centrality, the CI values contain much more topological information by gathering more information about the neighboring nodes. In this model, the CI value of node i gives the weight value $k_j - 1$ to each neighbor in $\partial Ball(i, l)$, which means that, even if k_i of the node is small, the CI value could be large. Therefore, we can discover the objective nodes in the network through the CI values. For example, in a network with multiple core areas (connected dense area), the node with the largest degree is the most important node from the perspective of a single area, but it will be found from the perspective of the whole network that, although some nodes are not connected much, they are still located in the hub area of several core areas. When these nodes are destroyed, the network will shatter faster.

Thus, the optimal set of nodes U^*, under the influence of the network percolation, can be obtained approximately based on the CI values. The algorithm process can be described as follows:

1) First, we can greedily choose the node with the largest CI value from the remaining nodes of the set as $u \leftarrow \underset{v \in V \setminus U^*}{\arg\max} CI_l(v)$;

2) Second, this node is added to U^* and removed from the network;

3) Third, the CI values of the remaining nodes are recalculated;
4) Finally, the above steps are repeated until the maximum connected branch is destroyed.

4. Identification of critical nodes of a cascading failure in the load network

In a load network, network cascading failures are associated with node load [2]. In an actual network, each node bears a certain amount of traffic. When a node fails, its load will be shared by the surrounding nodes, for which they must have additional load capacity to handle the load of the faulty node. If the shared additional load makes the total load exceed the capacity of node, this node will also fail and continue to redistribute the load to the surrounding nodes, resulting in the cascading failure of the load network. The existing methods for identifying critical nodes do not focus on the actual load failure. In this section, the cascading failure model of the load network is presented first. On that basis, the CI values are improved and the measurement of node importance based on the load percolation is proposed.

4.1. Load network cascading failure model

In Section 3.1, we introduced the existing analysis model for a cascading failure and three main factors of a capacity-overload model, including load initialization, node capacity and load redistribution strategy. The following describes the specific settings of these three factors.

4.1.1. Load initialization

First, the important step is to set the initial load of the edges and nodes in the network $G(E,V)$. In other related literature, the load of nodes is often defined as the betweenness of nodes, the computational complexity of which tends to be relatively high. Moreover, the load of edges is often shown as the bandwidth of communication links and the capacity of roads in real scenes. In a network, the load of a node should be related to the load of its connected edges. Therefore, the load of a node is defined in this paper as the sum of the load of its connected edge, mathematically,

$$F_i = \sum_{n \in N} L_{in} \tag{2}$$

where F_i represents the load of node i, N represents the set of neighboring nodes of i and L_{in} represents the load of an edge.

4.1.2. Node capacity

There is usually a margin that defines the extent of the load that a node can bear in the initial design of a load network. Therefore, the capacity is usually used to describe the maximum load that a node can bear. In this case, common capacity definitions are adopted. For each node i,

$$C_i = (1 + \beta)F_i \tag{3}$$

where β is the margin factor and, usually, $\beta > 0$. When the current load L_i of node i exceeds its capacity C_i, the node would fail.

4.1.3. Load redistribution policy

This paper proposes a local shortest path load redistribution strategy. When a node is destroyed, the load flowing through the node will be redistributed to the local related nodes according to the shortest path principle. Compared with the local optimal redistribution strategy, the redistribution strategy in this paper introduces the shortest path search, which improves the simulated level of the real scene. Compared with the global allocation strategy based on the shortest path, the shortest path search of the reallocation strategy in this study is only carried out between the neighboring nodes of the destroyed node, thus avoiding excessive computational complexity.

Then, considering the load redistribution mode, after removing node *t*, the specific calculation method for load distribution is as follows:

$$\Delta l = \frac{L_{it}+L_{jt}}{d_{ij}(k_t-1)} \quad \forall l \in ES(i,j), (i,j) \in \Lambda_t \tag{4}$$

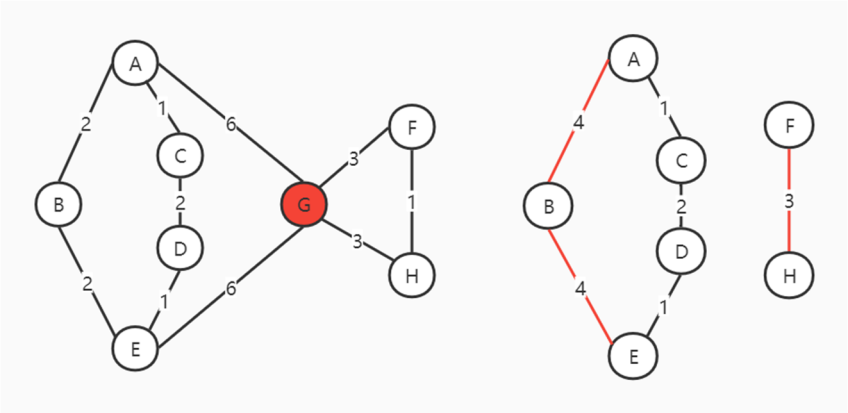
where l represents an edge on the shortest path; Δl_i represents the load variation on the edge l; Λ_t represents the neighbor-node pair of t that remains connected after removing node t, and $ES(i,j)$ represents the edge set of the shortest path between nodes. L_{it} represents the current load of edge *i*-*t*; d_{ij} represents the shortest path length between nodes i and j; k_i represents the degree of node t.

For example, as shown in Figure 1(a), if node G is removed, $k_G = 4$; then, the set of neighbor-node pairs that remain connected to node G, $\Lambda_G = \{(A,E),(F,H)\}$, is traversed, and the load flowing through node G will be allocated to the shortest path of these connected node pairs. Otherwise, for node pairs (A,E), after G is removed, the load on edge AG and edge EG would be redistributed to the shortest path *A*-*B*-*E* between (A,E), which results in $ES(A,E) = \{l_{(A,B)}, L_{(B,E)}\}, d_{AE} = 2$. It is noted that not all of the load on $l_{(A,G)}$ and $l_{(E,B)}$ is redistributed to $ES(A,E)$. We believe that this is because node G is the common hub of nodes A, E, F and H. Thus, the load on $l_{(A,G)}$ not only carries the traffic between (A,E), but it also carries the traffic between (A,F) and (A,H), which is why the denominator $(k_t - 1)$ in Eq (4) exists. After that, $\Delta l_{(A,B)} = \frac{6+6}{2*(4-1)} = 2$, while $\Delta l_{(B,E)} = \frac{6+6}{2*(4-1)} = 2$ and $\Delta l_{(F,H)} = \frac{3+3}{(4-1)} = 2$. As shown in Figure 1(b), after node G has been removed, the load could be reallocated according to Eq (4).

After redistribution of the load as above, the overall load of the network is partially lost, which first affects the load of the shortest path of the node pair. This distribution method is reasonable and more consistent with the change in load in the real network.

Through the overload model, different node attacking strategies have been evaluated, as shown in Algorithm 1, where V_F represents the set of failed nodes; and, Steps (3)–(5) constitute a cycle round in which several operations are performed, such as the removal of nodes, redistribution of the load and identification of failure nodes. The cascading failure parameter is generally a function of the network after the cascading failure and the original network.

Unlike the CI algorithm, which iteratively selects the node with the highest CI value to attack, Algorithm 1 attacks several nodes at the same time. This type of attack method is consistent with the actual network scenario.



(a) (b)

Figure 1. Flow reassignment; (a) initial network, and (b) result of redistribution after node G removal.

Algorithm 1. Algorithm for evaluating the attack strategy based on an overload model.

Input: network $G(V, E)$, initial load on the edges, list A of attacked nodes
Output: cascading failure indicator

(1) Initialize node load $VF \leftarrow A$, $F_i = \sum_{(i,j)\in E} F_{ij}$;
(2) **while** $A \neq \emptyset$ **do**
(3) Remove all nodes from A and redistribute their load by using Eq (4);
(4) $A \leftarrow \{j | L_j \geq C_j, j \in V \backslash V_F\}$;
(5) $V_F \leftarrow V_F \cup A$;
(6) **end while**
(7) **return** cascading failure indicator;

In an actual network, especially in the core attack targets, such as power grid systems, attacks need to be completed in a short time. Thus, the multi-point attack at the same time is mainly due to three considerations. First, the defense mechanism of the target network may prevent further attacks from non-partners. After an attack on the network has been detected, network managers would generally take measures to cut off the external connections, change the authentication methods and enhance the intrusion detection, which makes the iterative attacks difficult to conduct. Second, the overall benefit of a single-point attack is limited, and it is unable to initiate failure in the target network in a short time. Although a single-point attack has a larger failure gain at each time, its overall effect is still not as good as that of multi-point attacks, leaving most of the functions of the target network still available. Therefore, the attack effect is not as ideal as expected. Third, the structure and load sensing process after the attacking process is more time-consuming. As a network manager, although it is possible to collect information about the target network after an attack, the data analysis and processing, structure prediction and other steps also need a certain amount of time, which could make it hard to effectively respond to the rapid changes of a network attack and defense.

4.2. Identification of critical nodes based on load percolation

As mentioned above, the problem of finding the maximum influence in the network node set can

be well solved by using the percolation theory; and, the CI value approximately solves the optimal percolation problem, so we can efficiently find the node sets with strong influence in the network according to the CI value. However, from the formula for the CI value, we can also see that the CI value can identify the hub nodes between multiple core regions, but the density of the core degree of the core region is approximately measured by the size of the node degree. Therefore, when considering the core area, the CI value only considers the structural features in the network, and it does not take the load of nodes into consideration.

The CI value based on the percolation theory can make the network break at the fastest speed, but in a cascade failure of the load network, the load of the node is defined as the sum of the side loads connected to it. The redistribution of the load of a failed node makes the load of the remaining nodes constantly change dynamically, and the CI value does not consider the change of the load, so there is room for improvement in the solution of the cascade failure of the load network.

For the identification of the critical nodes in a cascade failure of load networks, with reference to the CI value, this paper describes the importance of network nodes through the load attributes of the nodes and the network structure; and, it puts forward a load percolation parameter, as shown in Eq (5):

$$CI_l^A(i) = (k_i - 1)\sum_{j\in\partial Ball(i,l)} L_j \tag{5}$$

where L_j represents the load of node j, $Ball(i,l)$ represents all nodes within the shortest path distance l from node i and $\partial Ball(i,l)$ represents the nodes with the shortest path distance l from node i. Similar to the CI-based optimal influence node set solution problem, the approximate solution for the load percolation optimal influence node set can be obtained by greedily choosing the maximum CI_l^A value.

Through the load percolation method, we regard the core area of the network as the high load area, because the high load means that, if a node fails, a large amount of traffic will be distributed to the surrounding nodes, causing a large change in the load of the surrounding nodes and increasing vulnerability to the avalanche effect.

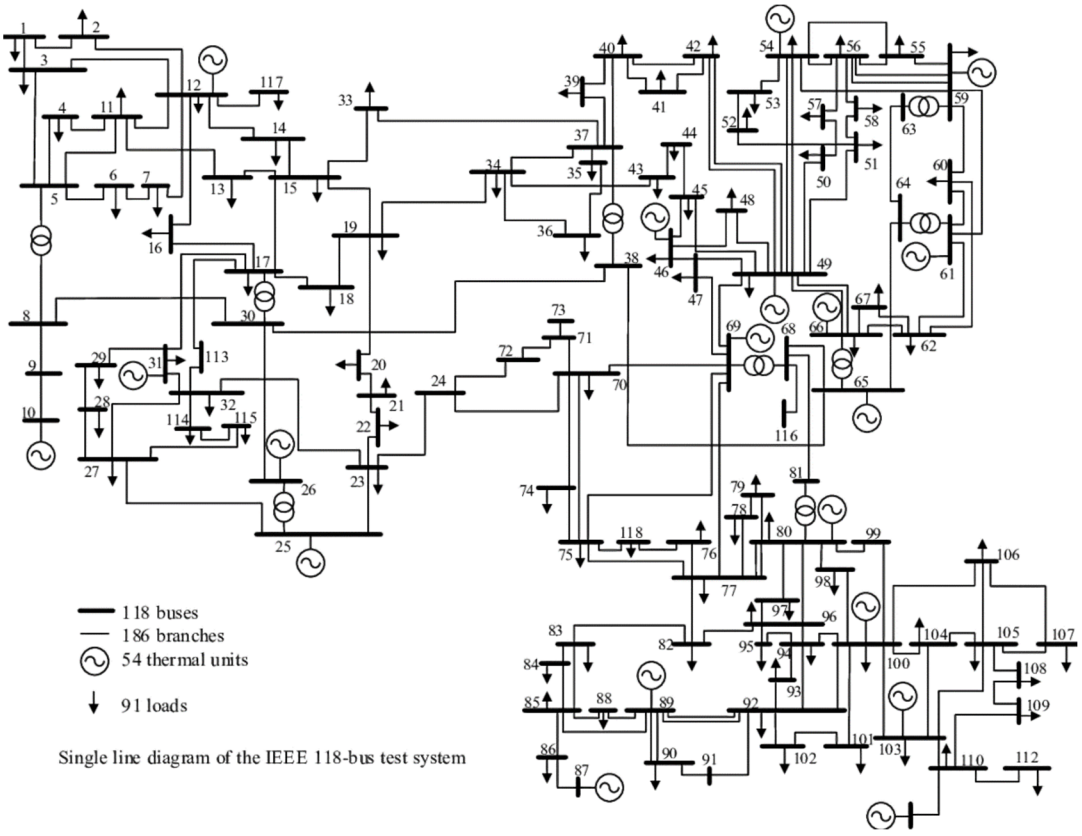
Therefore, through our improvement based on the original CI value, the load percolation method can overcome the limitation of local field of view and find the hub nodes between the high-load core areas in the global network from a larger range. These hub nodes themselves do not have many connections; but, if they are attacked first, with their failure, multiple high-load core areas will be "activated", causing a larger scale of cascading failures. That is to say, we can find those critical nodes in the load network that only have a moderate or less degree, or those that have a load that is small but play a hub role in the load network; the removal will cause a larger range and greater degree of traffic shock. Compared with the maximum degree and maximum load attack strategies, attacking these critical nodes of load percolation will cause a larger scale of cascading failures.

5. Experiment and analysis

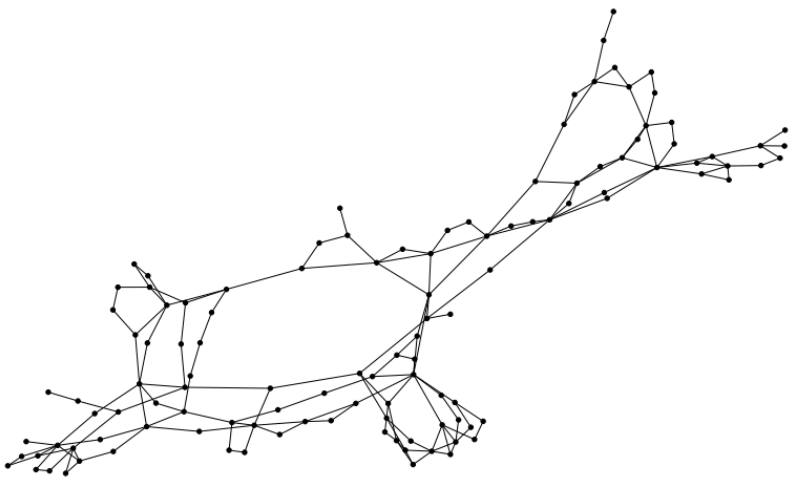
5.1. Data set

In order to evaluate the validity of the load percolation parameter, we performed experiments with two practical networks. One of the considered networks was the IEEE 118-bus network [35]. This network represents the topology and traffic information about the power grid system in the Midwest in December 1962, which included 118 nodes and 179 edges, as shown in Figure 2. The other network

is a simulated routing traffic network, which was built by using the open source network simulation tool NS2. And, this network contains 300 nodes and 403 edges, as shown in Figure 3.



(a)



(b)

Figure 2. IEEE 118-bus power network. (a) Original network [36]; (b) network topology.

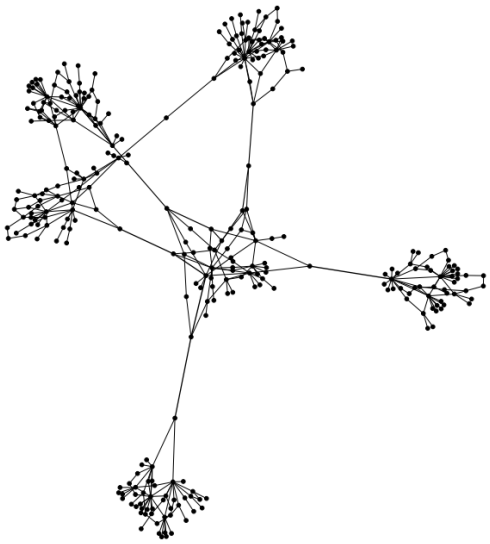


Figure 3. Simulated routing network.

5.2. Experimental setting

5.2.1. Existing attack strategies

This section compares the four key indicators of existing cascading failure and load percolation indicators CI_l^A, including the degree centrality, the CI value, the load centrality and the current betweenness centrality [37]. For these indicators, the load centrality does not consider the structural characteristics, and it can directly use the load values of nodes as the evaluation criterion of node importance. However, the current betweenness centrality replaces the shortest path in the betweenness centrality with the current propagation path, as follows:

$$C_{CB}(i) = \frac{2}{N(N-1)} \sum_{s \neq t \in G} I_i^{st} \tag{6}$$

where I_i^{st} represents the current value that starts at s and ends at t, and it passes through i; N represents the number of nodes in the network G. It is closer to the conditions of real-world load networks since the current betweenness centrality takes the load variation into account.

5.2.2. Evaluation indicator

Here, the attack effect is measured by the indicators of the node failure rate, the residual network load, the network global efficiency [38] and the size of the giant component. The network failure rate is defined as the proportion of failed nodes in the total number of nodes, mathematically,

$$r_F = |V_F|/|V| \tag{7}$$

where V_F represents the set of failed nodes obtained based on Algorithm 1.

And, the residual network load is defined as the ratio of the total load of each node in the network to the total load of the original network structure after the cascading failure occurs, mathematically,

$$r_A = \sum_{i \in V \setminus V_F} L_i / \sum_{i \in V} F_i \tag{8}$$

where L_i represents the load of the nodes in the network after the cascading failure and F_i represents the load of the nodes when the network is in its initial state.

Moreover, the network global efficiency is used to measure the efficiency of network information exchange, mathematically,

$$E_G = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}} \tag{9}$$

where d_{ij} represents the shortest path length from node i to node j.

The size of the giant component indicates the connection proportion of the network after disturbance, which can symbolize the system efficiency in a traffic or communications network.

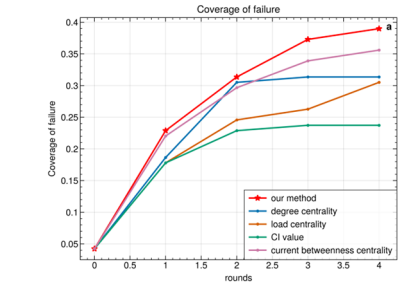
5.3. Effectiveness evaluation for load percolation indicator

Comparisons of the attack effect about the load percolation indicator CI_l^A and the other four indicators on two datasets are shown in Figures 4 and 5, in which the CI_l^A and CI values are uniformly taken as $l = 3$ and the tolerance parameter $\beta = 0.3$.

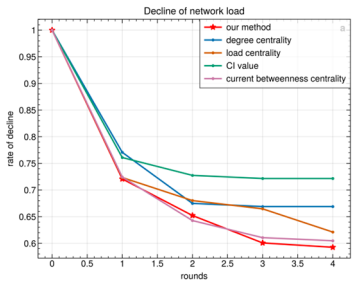
As shown in Figures 4(e),(g) and 5(e),(g), from the perspective of the network global efficiency, the load percolation is inferior to degree centrality and current betweenness in both datasets, but it is slightly better than load centrality. However, from the perspective of the size of the maximum connected slice of the network, for the IEEE 118-bus network, the degree centrality and the current betweenness work best, as shown in Figure 4(f),(h). But, for the simulated routed network, the CI value-based method achieved the best effect at the beginning, since the simulated routed network presents a multi-core network structure compared with the IEEE 118-bus network. Thus, the CI value-based method could exactly find out the hub node connected with the multi-core area.

In terms of the coverage rate of failed nodes, in the case of the IEEE 118-bus network, the nodes selected according to the load percolation parameter achieved the best attack effect, as shown in Figure 4(a),(c), followed by the current betweenness and the degree centrality, load centrality and CI value. This is because the degree centrality and CI value only consider the network topology characteristics, while the load centrality only considers the attribute characteristics, so they did not perform well. The current betweenness simulates the transmission of current in the network and incorporates the traffic transmission characteristics of the network. So, it achieves an excellent attack effect in the load network. The load percolation parameter proposed in our method considers the network topological characteristics and load condition of the nodes comprehensively, finds the global hub of nodes connecting multiple high-load core areas in the load network and, finally, maximizes the number of failed nodes.

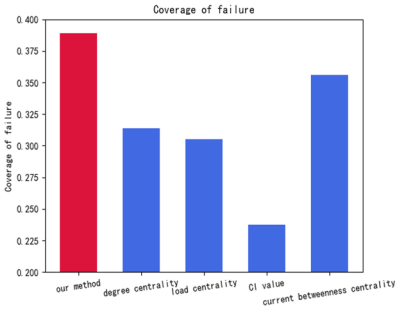
Furthermore, from the perspective of causing the decline of the network load, as shown in Figures 4(b),(d) and 5(b),(d), it was found that the load percolation can also achieve the best attack effect. Moreover, from the perspective of changes in the load of the network, as shown in Figure 5(b), the selection of the node with the largest load for attacking can make the overall network load decline rapidly in a short time. But, the speed would slow down and be overtaken by the load percolation. This further confirms that the local attack strategy, which only considers the degree or load of nodes, does not readily achieve the best attack effect globally.



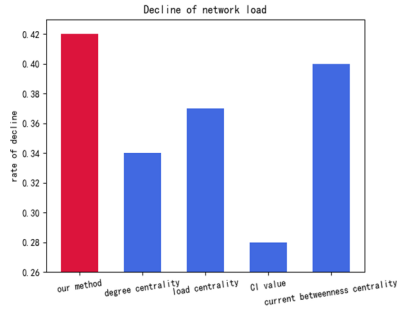
(a)



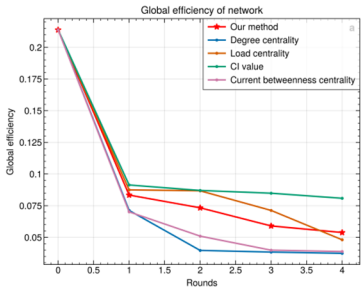
(b)



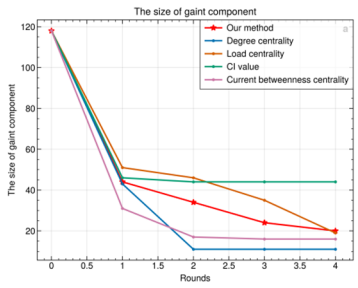
(c)



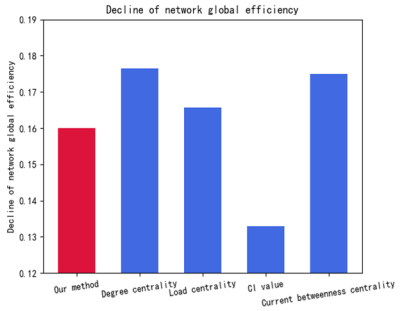
(d)



(e)



(f)



(g)

(h)

Figure 4. Comparison of attack effects on the IEEE 118-bus power network.

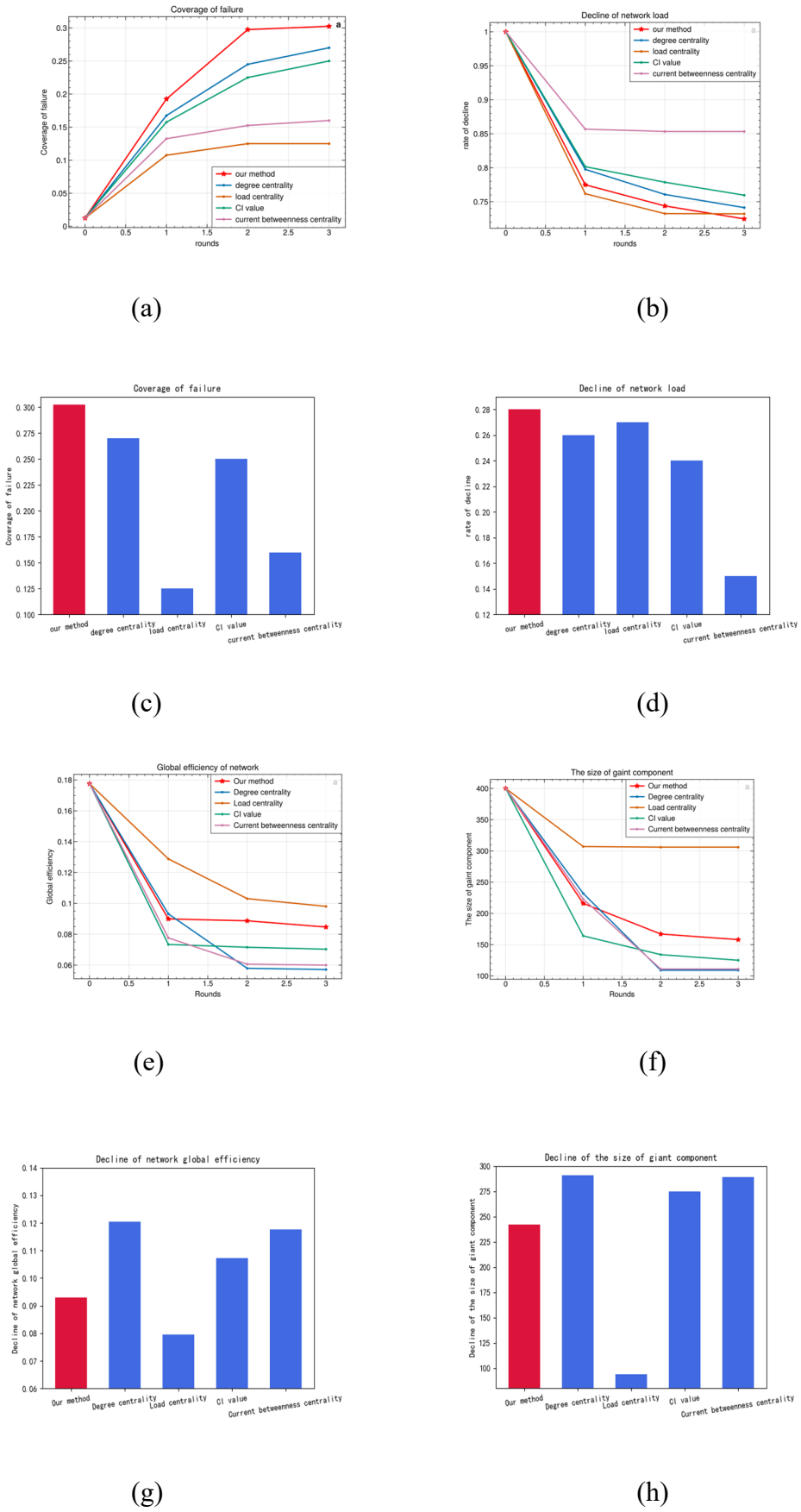


Figure 5. Comparison of attack effects on the simulated routing network.

The global efficiency and the maximum connected component size of the network are two indicators that only describe changes of the network structure characteristics, and do not consider the load. Therefore, the advantage of the load percolation method proposed in this paper, as compared with other methods, is that it can maximize the number of failed nodes in the load network after the cascading failure and maximize the overall load reduction of the network.

6. Conclusions and future work

Under the premise of clarifying the network topology, aiming to overcome the difficulty of identifying critical nodes due to the characteristics of numerous nodes and complex relationships in the target load network, this paper proposes a method for identifying the critical nodes of network cascading failures based on load percolation. First, we studied the mode and modeling method of the cascading failure in the network. Then, based on the percolation theory, the concept of load percolation has been proposed for the cascading failure problem in the load network. After that, we combined the common influences of the network structure with the load and built the load percolation model in the cascading failure scene. Finally, we proposed a novel critical node identifying method for network cascading failure; it can discover the optimal node set for a network cascading failure problem. The experimental results show that the load percolation parameter proposed in this paper is able to more accurately select the set of nodes, and that its final cascading failure effect is better than those of the existing methods.

In the future, research may be carried out from the perspectives of two main aspects. One aspect is to combine specific routing strategies, such as the backup feasible routes in EIGRP, so as to further optimize the load redistribution strategy and improve the accuracy of the overload model. The other aspect is to theoretically study the load percolation parameter in depth. The current parameter has strong intuition, and it is necessary to give theoretical derivation on the basis of the existing percolation theory to further improve the effectiveness of the parameter.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62071095 & No. 62101095) and the Sichuan Science and Technology Program under Grant 2019YFG0456.

The authors would like to express their gratitude to EditSprings (https://www.editsprings.cn) for the expert linguistic services provided.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. D. J. Watts, A simple model of global cascades on random networks, *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002), 5766–5771. https://doi.org/10.1073/pnas.082090499

2. A. E. Motter, Y. C. Lai, Cascade-based attacks on complex networks, *Phys. Rev. E: Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, **66** (2002), 065102. https://doi.org/10.1103/PhysRevE.66.065102
3. L. Lü, D. Chen, X. L. Ren, Q. M. Zhang, Y. C. Zhang, T. Zhou, Vital nodes identification in complex networks, *Phys. Rep.*, **650** (2016), 1–63. https://doi.org/10.1016/j.physrep.2016.06.007
4. R. Albert, H. Jeong, A. L. Barabási, Error and attack tolerance of complex networks, *Nature*, **406** (2000), 378–382. https://doi.org/10.1038/35019019
5. L. D. F. Costa, F. A. Rodrigues, G. Travieso, P. R. Villas Boas, Characterization of complex networks: A survey of measurements, *Adv. Phys.*, **56** (2007), 167–242. https://doi.org/10.1080/00018730601170527
6. H. W. Corley, D. Y. Sha, Most vital links and nodes in weighted networks, *Oper. Res. Lett.*, **1** (1982), 157–160. https://doi.org/10.1016/0167-6377(82)90020-7
7. D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2003), 137–146. https://doi.org/10.1145/956750.956769
8. Y. Chen, G. Paul, S. Havlin, F. Liljeros, H. E. Stanley, Finding a better immunization strategy, *Phys. Rev. Lett.*, **101** (2008), 058701. https://doi.org/10.1103/PhysRevLett.101.058701
9. X. Y. Zhao, B. Huang, M. Tang, H. F. Zhang, D. B. Chen, Identifying effective multiple spreaders by coloring complex networks, *Europhys. Lett.*, **108** (2014), 68005. https://doi.org/10.1209/0295-5075/108/68005
10. X. Zhang, J. Zhu, Q. Wang, H. Zhao, Identifying influential nodes in complex networks with community structure, *Knowledge-Based Syst.*, **42** (2013), 74–84. https://doi.org/10.1016/j.knosys.2013.01.017
11. F. Morone, H. A. Makse, Influence maximization in complex networks through optimal percolation, *Nature*, **524** (2015), 65–68. https://doi.org/10.1038/nature14604
12. P. Panigrahi, S. Maity, Structural vulnerability analysis in small-world power grid networks based on weighted topological model, *Int. Trans. Electr. Energy Syst.*, **30** (2020), e12401. https://doi.org/10.1002/2050-7038.12401
13. J. Beyza, J. M. Yusta, G. J. Correa, H. F. Ruiz, Vulnerability assessment of a large electrical grid by new graph theory approach, *IEEE Lat. Am. Trans.*, **16** (2018), 527–535. https://doi.org/10.1109/TLA.2018.8327409
14. J. Fang, C. Su, Z. Chen, H. Sun, P. Lund, Power system structural vulnerability assessment based on an improved maximum flow approach, *IEEE Trans. Smart Grid*, **9** (2018), 777–785. https://doi.org/10.1109/TSG.2016.2565619
15. J. Zhang, F. Hu, S. Wang, Y. Dai, Y. Wang, Structural vulnerability and intervention of high speed railway networks, *Physica A*, **462** (2016), 743–751. https://doi.org/10.1016/j.physa.2016.06.132
16. J. J. Wu, H. J. Sun, Z. Y. Gao, Cascading failures on weighted urban traffic equilibrium networks, *Physica A*, **386** (2007), 407–413. https://doi.org/10.1016/j.physa.2007.08.034
17. J. Wang, L. Rong, L. Zhang, Z. Zhang, Attack vulnerability of scale-free networks due to cascading failures, *Physica A*, **387** (2008), 6671–6678. https://doi.org/10.1016/j.physa.2008.08.037
18. Y. Yang, T. Nishikawa, A. E. Motter, Small vulnerable sets determine large network cascades in power grids, *Science*, **358** (2017), eaan3184. https://doi.org/10.1126/science.aan3184

19. P. Holme，B. J. Kim，C. N. Yoon, S. K. Han, Attack vulnerability of complex networks, *Phys. Rev. E*, **65** (2002), 056109. https://doi.org/10.1103/PhysRevE.65.056109
20. P. Holme, Edge overload breakdown in evolving networks, *Phys. Rev. E*, **66** (2002), 036119. https://doi.org/10.1103/PhysRevE.66.036119
21. L. Huang, L. Yang, K. Q. Yang, Geographical effects on cascading breakdowns of scale-free networks, *Phys. Rev. E*, **73** (2006), 036102. https://doi.org/10.1103/PhysRevE.73.036102
22. L. Dobson, B. A. Carreras, V. E. Lynch, D. E. Newman, An initial model for complex dynamics in electric power system blackouts, in *34th Hawaii International Conference on System Sciences (HICSS)*, IEEE, Maui, USA, (2001), 710–718. https://doi.org/10.1109/HICSS.2001.926274
23. I. Dobson, B. A. Carreras, D. E. Newman, A probabilistic loading-dependent model of cascading failure and possible implications for blackouts, in *36th Hawaii International Conference on System Sciences (HICSS)*, IEEE, Big Island, USA, (2003), 10. https://doi.org/10.1109/HICSS.2003.1173909
24. L. D. Valdez, L. Shekhtman, C. E. L. Rocca, X. Zhang, S. V. Buldyrev, P. A. Trunfio, et al., Cascading failures in complex networks, *J. Complex Networks*, **8** (2020), cnaa013. https://doi.org/10.1093/comnet/cnaa022
25. R. Parshani, S. V. Buldyrev, S. Havlin, Critical effect of dependency groups on the function of networks, *Proc. Natl. Acad. Sci. U.S.A.*, **108** (2011), 1007–1010. https://doi.org/10.1073/pnas.10084041
26. B. A. Carreras, D. E. Newman, I. Dobson, North American blackout time series statistics and implications for blackout risk, *IEEE Trans. Power Syst.*, **31** (2016), 4406–4414. https://doi.org/10.1109/TPWRS.2015.2510627
27. J. W. Wang, L. L. Rong, D. Wang, Model for cascading failures on complex networks based on local characteristics of nodes, *J. Manage. Sci. China*, **13** (2010), 42–50.
28. D. H. Kim, A. E. Motter, Resource allocation pattern in infrastructure networks, *J. Phys. A: Math. Theor.*, **41** (2008), 224019. https://doi.org/10.1088/1751-8113/41/22/224019
29. M. Li, R. R. Liu, L. Lü, M. B. Hu, S. Xu, Y. C. Zhang, Percolation on complex networks: theory and application, *Phys. Rep.*, **907** (2021), 1–68. https://doi.org/10.1016/j.physrep.2020.12.003
30. S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, et al., The structure and dynamics of multilayer networks, *Phys. Rep.*, **544** (2014), 1–122. https://doi.org/10.1016/j.physrep.2014.07.001
31. Y. Y. Cao, R. R. Liu, C. X. Jia, B. H. Wang, Percolation in multilayer complex networks with connectivity and interdependency topological structures, *Commun. Nonlinear Sci.*, **92** (2021), 105492. https://doi.org/10.1016/j.cnsns.2020.105492
32. R. R. Liu, D. A. Eisenberg, T. P. Seager, Y. C. Lai, The "weak" interdependence of infrastructure systems produces mixed percolation transitions in multilayer networks, *Sci. Rep.*, **8** (2018), 1–13. https://doi.org/10.1038/s41598-018-20019-7
33. L. Zhang, J. Ren, Inhomogeneous percolation on multilayer networks, *J. Stat. Mech: Theory Exp.*, **3** (2019), 033204. https://doi.org/10.1088/1742-5468/ab02ea
34. C. Yang, Z. Chen, Percolation on multi-layer network with joint storage and processing capacities, in *13th International Conference on Computer Modeling and Simulation (ICCMS)*, ACM, Melbourne, Australia, (2021), 114–120. https://doi.org/10.1145/3474963.3474979
35. B. Mirzasoleiman, M. Babaei, M. Jalili, M. Safari, Cascaded failures in weighted networks, *Phys. Rev. E*, **84** (2011), 046114. https://doi.org/10.1103/PhysRevE.84.046114

36. T. Xu, A. B. Birchfield, K. M. Gegner, K. S. Shetye, T. J. Overbye, Application of large-scale synthetic power system models for energy economic studies, in *50th Hawaii International Conference on System Sciences (HICSS)*, IEEE, Waikoloa Village, USA, (2017), 3123–3129.
37. J. Quirós-Tortós, R. Sánchez-García, J. Brodzki, J. Bialek, V. Terzija, Constrained spectral clustering-based methodology for intentional controlled islanding of large-scale power systems, *IET. Gener. Transm. Distrib.*, **9** (2015), 31–42. https://doi.org/10.1049/iet-gtd.2014.0228
38. V. Latora, M. Marchiori, Efficient behavior of small-world networks, *Phys. Rev. Lett.*, **87** (2001), 198701. https://doi.org/10.1103/PhysRevLett.87.198701

AIMS Press