Journal of *Marine Science and Engineering*

MDPI

Article

VGAE-AMF: A Novel Topology Reconstruction Algorithm for Invulnerability of Ocean Wireless Sensor Networks Based on Graph Neural Network

Ying Zhang [1,*,†], Qi Zhang [2], Yu Zhang [1,†] and Zhiyuan Zhu [3]

[1] College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China
[2] Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China
[3] Logistics Engineering College, Shanghai Maritime University, Shanghai 201306, China
* Correspondence: yingzhang@shmtu.edu.cn
† These authors contributed equally to this work.

Abstract: Ocean wireless sensor networks (OWSNs) play an important role in marine environment monitoring, underwater target tracking, and marine defense. OWSNs not only monitor the surface information in real time but also act as an important relay layer for underwater sensor networks to establish data communication between underwater sensors and ship-based base stations, land-based base stations, and satellites. The destructive resistance of OWSNs is closely related to the marine environment where they are located. Affected by the dynamics of seawater, the location of nodes is extremely easy to shift, resulting in the deterioration of the connectivity of the OWSNs and the instability of the network topology. In this paper, a novel topology optimization model of OWSNs based on the idea of link prediction by cascading variational graph auto-encoders and adaptive multilayer filter (VGAE-AMF) was proposed, which attenuates the extent of damage after the network is attacked, extracts the global features of OWSNs by graph convolutional network (GCN) to obtain the graph embedding vector of the network so as to decode and generate a new topology, and finally, an adaptive multilayer filter (AMF) is used to achieve topology control at the node level. Simulation experiment results show that the robustness index of the optimized network is improved by 39.65% and has good invulnerability to both random and deliberate attacks.

Keywords: ocean wireless sensor network; graph neural networks; link prediction; network invulnerability; scale-free networks

Citation: Zhang, Y.; Zhang, Q.; Zhang, Y.; Zhu, Z. VGAE-AMF: A Novel Topology Reconstruction Algorithm for Invulnerability of Ocean Wireless Sensor Networks Based on Graph Neural Network. *J. Mar. Sci. Eng.* **2023**, *11*, 843. https://doi.org/10.3390/jmse11040843

Academic Editors: Nils Morozs and Filippo Campagnaro

Received: 21 February 2023
Revised: 3 April 2023
Accepted: 13 April 2023
Published: 16 April 2023

1. Introduction

In recent years, ocean wireless sensor networks (OWSNs) [1,2] have been widely used in marine monitoring, maritime safety, intelligent shipping, and other fields for their flexibility. OWSNs play a pivotal role in the marine monitoring network, and sensor nodes that carry different types of sensors can achieve the purpose of monitoring different scenarios, such as marine climate monitoring and prediction, monitoring of marine environmental information, and intrusion detection in the marine environment [3]. It can also be used as a relay station between underwater nodes and above-water or terrestrial base stations and used to track underwater targets with underwater sensor networks. However, marine sensor nodes often face the damage of harsh environments, which makes the safety of sensor nodes a great threat. Once some nodes fail, the transmission of other sensor nodes through them is interrupted, and the data cannot be transmitted intact back to the designated terminal, which greatly diminishes the main function of the network. So, the invulnerability of OWSNs, which represents their ability to maintain normal functionality after node failures caused by attacks or excessive load, should be improved [4]. The main research directions for underwater sensor networks are based on fault-tolerant node residual time optimization [5] and high-reliability routing [6]. Depending on the attack intent,

attacks against network nodes can be random, such as software or hardware failures [7] and the effects of bad weather [8], or malicious, such as malicious attacks based on high threat probability.

Due to many natural factors, such as the dynamic features of the ocean climate and the undulating characteristics of waves, the sensor nodes of marine wireless sensor networks often consume a lot of extra energy to communicate with other nodes repeatedly. As the network operation time grows, the lifetime of sensor nodes decreases, which is very likely to cause sudden node failures. Although such failures are predictable, due to their dependence on the amount of data transmission, the node failures are nearly randomly distributed [9]. To address this challenge, we designed the marine wireless sensor network as a scale-free network to resist the random and sudden failure of nodes because of the above factors. A network with scale-free properties is called a scale-free network, i.e., the nodes have a degree distribution conforming to a power-law distribution. In such networks, a few nodes have a large number of connections, while most of the nodes have a handful of connections, making them highly resistant to random failures. The existence of the scale-free feature greatly increases the possibility of high-degree nodes, but scale-free networks are highly vulnerable to malicious attacks. Once the high-degree nodes are attacked, the network will be split into independent subgraphs or even crashed, so the scale-free network exhibits both robustness against random failures and vulnerability against deliberate attacks. In order to ensure the ability of the network to resist random destruction, it is necessary to maintain the scale-free property of the network. Under this condition, how to enhance the network's ability to resist deliberate attacks has become a hot topic and an important problem in the current research on the survivability of scale-free networks. The main contributions of this paper include the following:

(1) For the first time, we combine the domain of scale-free network invulnerability and robustness with graph neural networks to improve their resistance to damage, which can be computed directly on the whole network and is more helpful to extract multi-hop connectivity features of the network. To solve the problem of no attributes on nodes, the characteristic attributes of network nodes are constructed, and unsupervised learning is adopted for network data without truth labels, which is a typical form of encoder and decoder;
(2) We propose the maximum node load based on three indicators: node criticality, edge criticality, and node degree, and use them to construct the feature attributes of the network nodes. We propose an interval grading algorithm for node feature embedding and embed the node features into the graph data, together with the adjacency matrix of the graph, as a dataset for model training, testing, and validation;
(3) In the process of generating a new network topology, we propose an adaptive multi-layer filter. First, we calculate the network connectivity threshold of the new network based on the link generation probability matrix of the first layer filter, and then determine whether there is a link between two nodes through the node hierarchy method to control the degree of each node and maintain the network following a power law distribution.

The rest of the paper is organized as follows: the second part of the paper focuses on related work. The third part mainly focuses on the basic theoretical model of the algorithm. The fourth part introduces the VGAE-AMF algorithm model and the specific process of optimizing the scale-free network topology. The simulation experiments will be indicated in Section 5. The analysis of the results based on the simulation experiment and the analysis of the invulnerability performance will be described later in this chapter. The last part of the article is a summary of the research topic and an outlook for further research.

2. Related Works

In the field of network invulnerability of sensor networks, researchers have focused on the following aspects: (1) dynamic network invulnerability and cascade failure problems [10–12]; (2) load redistribution problems based on network data flow [13,14];

(3) cascade failure problems based on cost and energy efficiency [15,16]; and (4) network topology repair problems based on link reconfiguration.

For scale-free networks, network topology repair by link reconstruction is the most direct and effective method. Tian [17] proposed an adaptive edge reconstruction repair method by combining two reconstruction features: average reconstruction repair and focused repair reconstruction. Hu [18] improved the adaptive edge reconstruction repair method to solve the repair problem of scale-free networks under a random attack strategy. He [19] proposed an edge repair reconstruction method based on neighbor modification, and the algorithm has better prediction in social relation networks. Chen et al. [20] proposed a method based on the discrete artificial bee colony algorithm to repair the damaged network by adding links to the damaged network topology while considering connection cost constraints and network connectivity. Feng [21] proposed a clustering-based edge repair reconstruction method using the idea of link prediction and gave optimal clustering methods for networks with different densities. Victor Martinez et al. [22] proposed an edge repair reconstruction strategy of data mining that predicted the loss or future similarity change relationship based on the currently observed node connection similarity, thereby forming the desired network edge reconstruction strategy. Zhou [23] proposed a node similarity adjustment method based on optimal univariate functions to further improve the prediction accuracy of low similarity scores, which provided a new idea for edge reconstruction repair. On the edge modification strategy, Hu [24] proposed the edge modification strategy TMSE based on the node height operation and degree association operation, which mainly improved the robustness against malicious attacks on scale-free networks. This node-modification-based strategy is not optimized for the overall topology of the network and ignores the multi-hop adjacency characteristics of the nodes to some extent.

When researchers started to consider the specificity of the ocean scenario and the climate dynamics, He [25] optimized the network coverage by adjusting the sensor node positions by simulating fish behavior, repaired the disconnected locations in the network using redundant nodes, eliminated critical nodes, and optimized the network connectivity. Chen et al. [26] proposed an energy-efficiency-based connectivity reconstruction (EECR) scheme. Due to the dynamic nature of the underwater environment, the generation and recovery of dummy nodes in marine wireless sensor networks will also change dynamically, resulting in intermittent interruptions in the network topology. This scheme designs a multiband transmission mode for dummy nodes to reconstruct the topology of underwater acoustic sensor networks in the presence of dummy nodes. A fault prediction, detection, and recovery algorithm based on the Markov Chain Monte Carlo (MCMC) process is proposed in the literature [27]. In the fault prediction algorithm, the error pattern of delay messages is used to identify sensor node faults, and threshold limits are determined based on the temporal probability distribution function. In the fault detection algorithm, faulty sensor nodes are detected using threshold limits and residual energy. In the recovery algorithm, the faulty sensor node is replaced with the nearest sensor node with a higher energy capacity.

There are many methods based on topology reconfiguration strategies to improve network invulnerability, as mentioned above, and good enhancement results have been obtained. However, in practical applications in marine scenarios, research on network invulnerability for marine wireless sensor networks is still sporadic, and topology repair reconfiguration strategies based on damaged networks are a kind of repair method that damages the network greatly, and remediating damaged networks is a high-cost operation. If the network topology can be optimized before the network is damaged instead of designing the reconfiguration strategy after the network is damaged, it is a way to frontload and reduce the damage to the network. Therefore, we predict hidden links in advance based on the idea of link prediction and then decide whether to deploy hidden edges based on the probability of link generation to improve the invulnerability of the network to malicious attacks.

The main traditional methods for link prediction are heuristic methods and embedding methods. Heuristic algorithms compute some heuristic node similarity scores as link probabilities [28], such as common neighbors, preferential dependencies [29], and the Katz index [30], which can be regarded as some predefined graphical structural features. Embedding methods such as matrix factorization (MF) and Node2vec [31] learn free parameters for node embedding in a straightforward way and therefore cannot be generalized to hidden nodes and networks. However, in recent related research, many researchers have achieved good prediction results by integrating graph neural networks with link prediction ideas. Graph neural networks use networks as model inputs and perform operations directly on graph data. Bruna et al. [32] first introduced convolution into graph neural networks, and they developed a graph convolutional network model based on the concept of frequency-domain convolutional operations, which were the first to use learnable convolutional operations on top of graph data. Subsequently, Kipf et al. [33] simplified the definition of frequency-domain graph convolution to enable the operation of graph convolution in the spatial domain, which greatly improved the computational efficiency of the graph convolution model. Due to their convincing performance and high interpretability, graph neural networks have become a widely used method for network analysis in recent years. Currently, there are two main mainstream methods for GNN-based link prediction. One of them is graph auto-encoder (GAE) [34], in which GNN is first applied to the whole network to calculate the representation of each node and then aggregates the representations of source and target nodes to predict the target links. A variant of it is called VGAE, which introduces variational operations into the encoding process of graph auto-encoders. The second method is SEAL [35], which extracts local closed subgraphs around each target link. The nodes in each closed subgraph are labeled differently according to their distances from the source and target nodes. A GNN is then applied to each closed subgraph to learn the link representation for link prediction.

The link prediction method based on graph neural networks is often used in scenarios such as knowledge graph complementation or user recommendation in social networks. However, the potential of link prediction based on a node hierarchical approach to recommend similar nodes by similarity calculation of network nodes is not to be underestimated in the field of marine wireless sensor network invulnerability. At present, there are not only no advanced experiments but also many challenges in applying this method to network topology optimization and anti-attack performance improvement of marine wireless sensor networks with scale-free characteristics, such as graph data processing, feature construction of scale-free networks, and quality evaluation of generated graphs.

3. Theoretical Method

3.1. Basic System Model

The structure of the algorithm model of VGAE-AMF is shown in Figure 1. Due to the fact that scale-free network data are generated through random generation, which results in data without truth tags, it is extremely difficult to directly construct a truth value system, and it is difficult to use effective supervised learning methods to calculate the network. Therefore, we turn to exploring unsupervised learning methods, which mainly use typical encoder and decoder forms.

In the link prediction process based on graph neural networks, the encoding process uses an encoder consisting of two layers of graph convolutional neural networks for graph embedding of scale-free networks, and the model performs best when the number of layers of GNN is set to the radius of the graph [36]. To cover more nodes (when $n \leq 100$), we designed the graph convolutional layers as two as shown in Figure 2.

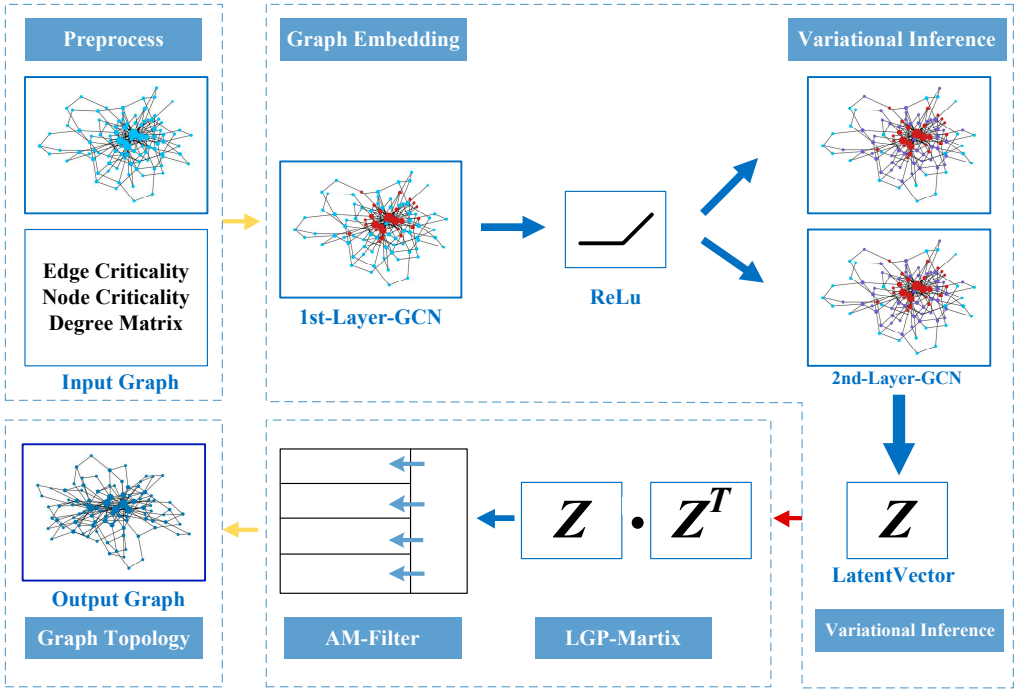
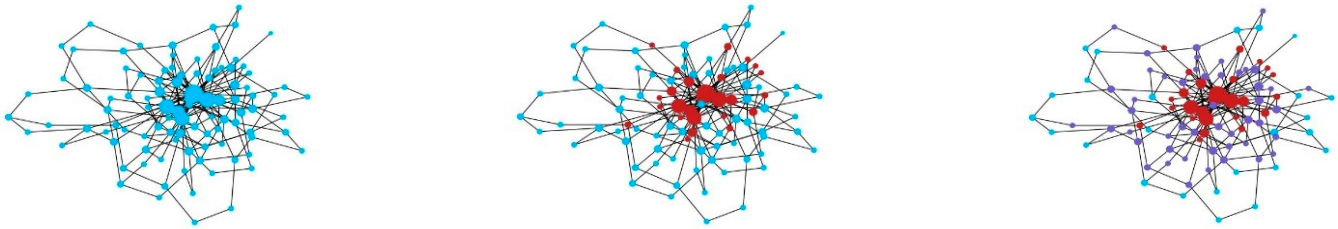


Figure 1. Main structure of the VGAE-AMF algorithm model. (Note that Z is the graph embedding vector of the input network and Z^T is the transpose of Z).



(**a**) (**b**) (**c**)

Figure 2. Number of layers of the graph convolution network versus the number of nodes covered (when $n \leq 100$): (**a**) input graph; (**b**) one-layer graph convolution (red); (**c**) two-layer graph convolution (purple and red).

The first layer of the graph convolutional network receives the adjacency matrix and the feature matrix of the network as input, i.e.,

$$\mathrm{GCN}_{1\mathrm{st}}(\boldsymbol{X}, \boldsymbol{A}) = \widetilde{\boldsymbol{A}}\boldsymbol{X}\boldsymbol{W}_0 \tag{1}$$

where $\widetilde{\boldsymbol{A}} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is the symmetric normalization matrix and $\boldsymbol{W}_0$ is the network weight. The second layer of the graph convolutional network incorporates a variational operation to extract the neighbor features within two hops of the input graph nodes in the form of hidden vectors, i.e.,

$$\mu = \mathrm{GCN}_{\mu}(\boldsymbol{X}, \boldsymbol{A}) = \widetilde{\boldsymbol{A}}ReLU\left(\widetilde{\boldsymbol{A}}\boldsymbol{X}\boldsymbol{W}_0\right)\boldsymbol{W}_1 \tag{2}$$

$$\log\sigma = \mathrm{GCN}_{\sigma}(\boldsymbol{X}, \boldsymbol{A}) = \widetilde{\boldsymbol{A}}ReLU\left(\widetilde{\boldsymbol{A}}\boldsymbol{X}\boldsymbol{W}_0\right)\boldsymbol{W}_1 \tag{3}$$

$$\boldsymbol{Z} \sim \boldsymbol{N}(\mu, \sigma) \tag{4}$$

The hidden vector $\boldsymbol{Z}$ is the Gaussian distribution sampled from the μ and $\log\sigma$, i.e., the graph embedding of the network.

In the decoding process, we used the hidden vector inner product to implement the generation of the link generation probability matrix, i.e.,

$$p(\boldsymbol{A} \mid \boldsymbol{Z}) = \prod_{i=1}^{N}\prod_{j=1}^{N} p\left(A_{ij} \mid \boldsymbol{z}_i, \boldsymbol{z}_j\right) \tag{5}$$

where $p\left(A_{ij} = 1 \mid \mathbf{z}_i, \mathbf{z}_j\right) = \sigma\left(\mathbf{z}_i^{\top}\mathbf{z}_j\right)$, A_{ij} is the element of the adjacency matrix A, and $\sigma(\cdot)$ denotes the logical Sigmoid function. Subsequently, the Adaptive Multilayer Filter (AMF) is accessed instead of the Softmax layer to determine the existence of the link, and thus the new topology of the network can be obtained.

The embedding vector is decoded to generate similar graphs, and the model is trained with the distance between the original graph distribution and the generated graph distribution as the error, i.e.,

$$ELBO = \mathbb{E}_{q(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{A})}[\log p(\boldsymbol{A} \mid \boldsymbol{Z})] - KL[q(\boldsymbol{Z} \mid \boldsymbol{X},\boldsymbol{A}) \parallel p(\boldsymbol{Z})] \tag{6}$$

where $KL[q(\boldsymbol{Z} \mid \boldsymbol{X},\boldsymbol{A}) \parallel p(\boldsymbol{Z})]$ is defined as the KL divergence between $q(\cdot)$ and $p(\cdot)$; $\mathbb{E}_{q(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{A})}[\log p(\boldsymbol{A} \mid \boldsymbol{Z})]$ is the cross-entropy function; and $p(\boldsymbol{Z}) = \prod_i p(\boldsymbol{z_i}) = \prod_i \mathcal{N}(\boldsymbol{z}_i \mid 0, \boldsymbol{I})$ is the Gaussian prior.

3.2. Node Feature Construction

The topology of the network is embedded in the graph by means of a graph convolutional network, where the similarity of nodes is calculated to determine the invulnerability level of nodes, and new edges are established between two nodes with similar levels, which reduces the bridging coefficient of nodes and thus improves the destructive resistance of nodes under malicious attacks. We found that there is no hierarchical relationship between the nodes and no topology-related feature information when we embedded the graph, which means that the scale-free network does not have node features, so we construct the features of the network nodes based on the betweenness centrality and bridging coefficient.

The betweenness centrality [37] mainly reflects the busyness of a node in data transmission, as a way to quantitatively represent the criticality of a node in the network. Assuming any two nodes $i, j \in V$, the path with the least number of hops between node pairs s, t is called its shortest path. Obviously, there may be more than one shortest path between nodes i, j. Let σ_{st} denote the number of shortest paths between nodes i, j, and $\sigma_{st}(V_i)$ denote the number of shortest paths between node pairs s, t passing through nodes V_i, then the betweenness centrality of nodes V_i in a complex information system network G is defined as

$$B(V_i) = \sum_{s \neq v_i \neq t \in V} \frac{\sigma_{st}(V_i)}{\sigma_{st}} = \sum_{s \neq v_i \neq t \in V} \delta_{st}(V_i) \tag{7}$$

The bridge edge [38] is an edge that is used to maintain global connectivity and plays a crucial role in the flow of information in the network. The bridging coefficient mainly reflects the importance of the bridge edge in the network. Therefore, it is a way to quantitatively characterize the criticality of the edge in the network. Assuming that the sets of neighboring nodes v_i' and $v_j\prime$ of two nodes v_i and v_j are directly connected to the edge e_{ij} in the network are M and N, the degrees of nodes v_i and v_j nodes can be expressed as $k_i = |M|$ and $k_j = |N|$, respectively. The ratio of the number of actual connected edges between nodes v_i' and $v_j\prime$, and the number of common neighbor nodes of nodes v_i and v_j to the sum of the total number of possible edges and the total number of nodes is the bridging coefficient, i.e.,

$$E_{ij} = -\log\left(\frac{\sum_{m\in M}\sum_{n\in N} a_{mn} + P + c}{(|M|+1)\times(|N|+1)}\right) \tag{8}$$

where c is the regulator, $\sum_{m\in M}\sum_{n\in N} a_{mn}$ denotes the total number of directly connected edges of the neighbors of node v_i and the neighbors of node v_j, and P denotes the number of common neighbors of the nodes v_i and v_j; the higher the number, the stronger the connection between the nodes v_i and v_j, the weaker the bridge property of the edge e_{ij}, as follows:

$$P = |M \cap N| = \frac{\sum_{m\in M} a_{jm} + \sum_{n\in N} a_{ni}}{2} \tag{9}$$

where $\sum_{m\in M} a_{jm}$ denotes the number of edges connected to the node v_i and the neighbors of node v_j; and $\sum_{n\in N} a_{ni}$ denotes the number of edges connected to the node v_j and the neighbors of node v_i.

3.3. Evaluation Indicators

To quantitatively analyze the invulnerability performance of the network, we use the network connectivity coverage based on the maximum connected subgraph size to evaluate the invulnerability performance of the network topology under different attack types before and after optimization. The connectivity coverage of the network is expressed as the ratio of the maximum connected subgraph of the network to the overall network size and is simply defined by

$$C = \frac{MCS}{N} \tag{10}$$

where MCS denotes the number of nodes of the maximum connected subgraph in the network, and N denotes the number of network nodes in the initial network.

To focus on evaluating the robustness of the network in the presence of malicious attacks, Schneider et al. [39] proposed an evaluation metric, R, which is the number of nodes that count the maximum connected subgraph of the network after each round of attacks, and R will be defined by the following equation:

$$R = \frac{1}{N+1}\sum_{n=0}^{N}\frac{MCS(n)}{N} \tag{11}$$

where N is the number of nodes in the initial network, n is the number of attacks, $MCS(n)$ represents the number of nodes in the maximum connectivity subgraph of the network after the attack n-th, R represents the sum of all nodes in the network after all attacks have failed, and finally, keeping independent of the network size by the normalization operator $1/N+1$.

Moreover, we chose another value to evaluate the invulnerability of the network; the invulnerability entropy measure is calculated by:

$$IE = -\sum_{i=1}^{N} S_i \ln S_i \tag{12}$$

where IE is the invulnerability entropy measure, which is used to measure the invulnerability of the network. It can be seen that the more uniform distribution of S, the greater its invulnerability entropy value, the stronger its ability to resist intentionally attack. S_i is the fusion criticality of nodes and edges, calculated by

$$S_i = \left(\mu \cdot N_i + \alpha \cdot \frac{1}{2|S|}\sum_{j\in S} E_{ij}\right) / \sum_{i=1}^{N}\left(\mu \cdot N_i + \alpha \cdot \frac{1}{2|S|}\sum_{j\in S} E_{ij}\right) \tag{13}$$

where α and μ are the node criticality and edge criticality weights respectively, and $\alpha + \mu = 1$. S is the set of nodes v_i neighbor nodes, $N(v_i)$ is the node criticality, and E_{ij} is the edge criticality.

4. Improving the Invulnerability of Scale-Free Network Based on the VGAE-AMF Algorithm

The VGAE-AMF algorithm mainly uses the idea of link prediction to optimize the network topology. The new topology of the scale-free network generated by the VGAE-AMF algorithm maintains the scale-free property and establishes connections between nodes with similar hierarchy levels to weaken the bridging property of the original multi-hop connections between nodes. Thus, this reduces the bridging factor and improves the invulnerability of the network to deliberate attacks.

4.1. Maximum Node Load and Interval Grading Algorithm

The main difficulty of building scale-free networks using a learning-based approach is in the data aspect. The topology of the network can be represented by the adjacency matrix of the network, but the feature properties of the nodes in the scale-free network are not known. To solve the problem, we constructed the maximum load of the nodes based on the node criticality, edge criticality, and node degree as the feature properties of the network nodes. The larger the maximum load that a node can carry, the more edges the node can accommodate.

In terms of node criticality, we used the betweenness centrality of a node as an important measure of node criticality, which mainly reflects the busyness of a node in data transmission, to quantitatively characterize the criticality of the node in the network. For edge criticality, we introduced the concept of a bridge edge, which is an edge that plays a critical role in the flow of information in the network and is used to maintain global connectivity, also called a critical edge. The critical edges are not only directly related to the nodes at both ends of the edge, but also related to the common neighbor nodes of the nodes to quantitatively characterize the criticality of the edge in the network. To better perform graph embedding for scale-free networks, the edge weights in the network data are set to 1, which will result in the edge criticality not being embedded in the edge weights in an exact match, so we distribute the edge criticality equally to the two endpoints of the link and embed the edge criticality into the node feature. The degree of the network nodes also well reflects the status of the nodes' importance in the network, and the node degree is not well reflected in the above metrics on the node characteristics, so we also embedded the node degree matrix into the node feature. We constructed the node feature of the network using the following definition, i.e., the maximum load of the node:

Definition 1 (maximum node load). *The maximum load that a node can bear is not only influenced by the degree of the node but also related to the criticality of the node and the edges; the more important the node is and the more important the edges connected to the node are, the greater the maximum load of the node, i.e.,*

$$L = \lambda(-\sum_{i=1}^{M} S_i \ln S_i)^{\gamma} + (1-\lambda){k_i}^{\beta} \tag{14}$$

where λ, γ, β are the conditioning parameters, M is the number of network nodes, and k_i denotes the degree of nodes i.

From thegroup of pictures in Figure 3, we can see the obvious hierarchical grading in the performance of the index of the maximum node load. This law happens to help us solve a difficult problem in the construction of the node feature matrix, namely, that the node feature cannot simply be embedded using the value of the maximum load of the node. Therefore, we rely on this rule to design an interval grading algorithm that embeds the maximum load of the node in the feature matrix to deal with the above difficulties.

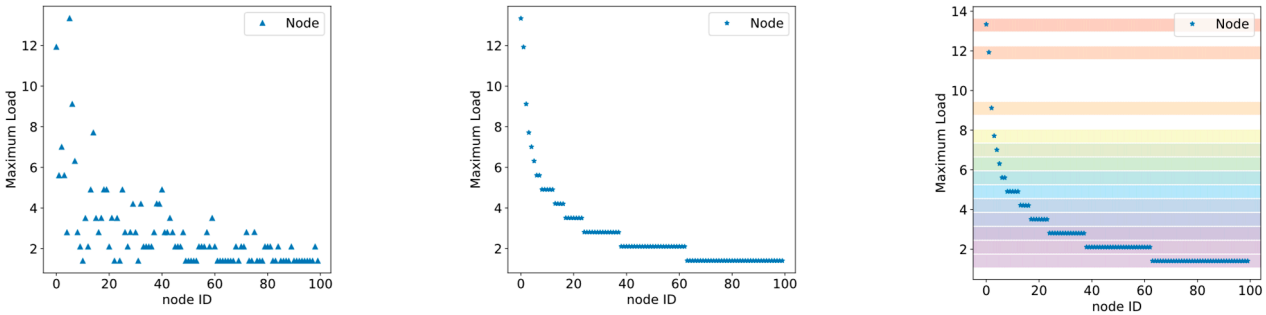


Figure 3. Distribution of the maximum node load when $n \leq 100$: (**a**) unsorted; (**b**) sorted; (**c**) schematic diagram of the interval grading algorithm.

The pseudo code for the interval grading algorithm is shown as Algorithm 1:

Algorithm 1 Interval Grading Algorithm

Input: L_i: The node maximum load matrix, n: number of node
Output: Fea: The node feature matrix
1: Parameter initialization: L_n,Fea,$count$
2: $L_n \leftarrow$ Sorted L_i
3: $Fea \leftarrow \{\}$
4: $count \leftarrow 0$
5: for $i = 0 \rightarrow n$ do
6: if $i \leftarrow n - 1$ then
7: $Fea[count] \leftarrow L_n[i]$
8: else
9: $Distence \leftarrow |L_new[i] - L_new[i] + 1|$
10: if $Distence \leq 0.3$ then
11: $Fea[count] \leftarrow L_n[i]$
12: else
13: $Fea[count] \leftarrow L_n[i]$
14: $count \leftarrow count + 1$
15: end if
16: end if
17: end for
18: Convert Fea to matrix.
19: Compress the matrix into sparse matrix.

The interval algorithm mainly addresses the fact that the features of nodes are based on categories rather than values. We convert the numerical values into categories based on the above-mentioned law exactly, and the category still reflects the hierarchical differences between nodes. Based on the data distribution and our experience, we set the stratification interval at 0.3 to achieve the best grading strategy.

4.2. Adaptive Multi-Layer Filter (AMF)

After obtaining the network embedding vector from the encoder output, the link generation probability matrix is obtained by the inner product operation, but how to design the threshold is crucial to the generation of links or not, so it is not possible to simply fix a threshold to determine the link's existence. Thus, we propose an adaptive multilayer filter based on the link generation probability matrix. This filter mainly controls the degree of each node in the network through macroscopic regulation to ensure that the network maintains the scale-free property.

As shown in Figure 4, the adaptive multilayer filter is divided into two stages, i.e., two layers of filtering. To ensure network connectivity, the first layer is designed as a constant-value filter, and the minimum value of the two links with the highest probability

of link generation among all nodes is used as the threshold of the first layer to perform the layer of connectivity-based filtering on the matrix. To control the degree of each node, the hierarchy of nodes is classified according to the interval grading algorithm in the above definition, and the threshold of the second layer is defined adaptively based on the different grades of the network nodes. Power-law distribution-based filtering is then performed to maintain the network in compliance with the scale-free property.

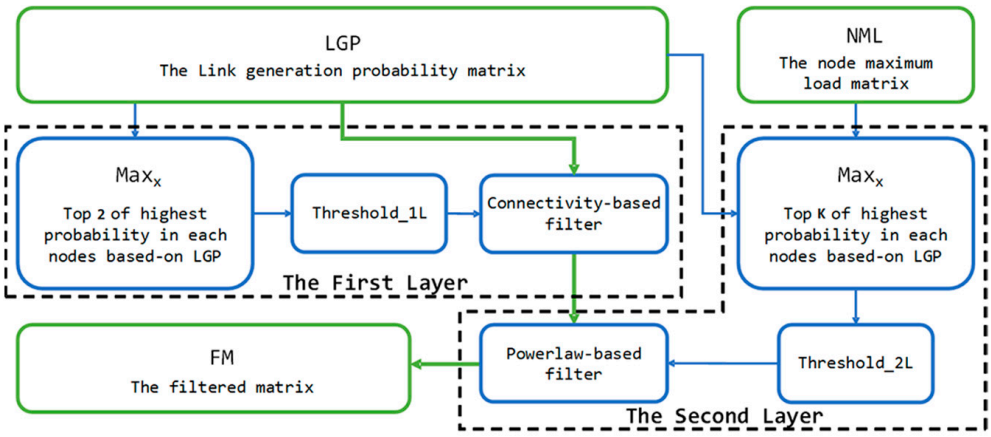


Figure 4. Structure of the adaptive multilayer filter algorithm.

The pseudo code for the adaptive multilayer filtering algorithm is shown as Algorithm 2:

Algorithm 2 Adaptive Multilayer Filter

Input: n: number of nodes
LGP: The link generation probability matrix
L_i: The node maximum load matrix
Output: Fm: The filtered matrix
The First Layer: Connectivity-based filtering
Max_X: Top 2 of highest probability in each node based on LGP
$th_1 \leftarrow$The minimum value of Max_X
for $i = 0 \rightarrow n$ do
  for $j = 0 \rightarrow n$ do
    if $i \neq j$ and $LGP[i,j] \geq th_1$ then
      $Fm[i,j] \leftarrow 1$
    else
      $Fm[i,j] \leftarrow 0$
    end if
  end for
end for

The Second Layer: Power-law-based filtering
k: Node initial level
Max_X: Top k of highest probability in each node based on LGP
$th_2 \leftarrow$The minimum value of Max_X
for $i = 0 \rightarrow n$ do
  for $j = 0 \rightarrow n$ do
    if $i \neq j$ and $LGP[i,j] \geq th_1$ then
      $Fm[i,j] \leftarrow 1$
    else
      $Fm[i,j] \leftarrow 0$
    end if
  end for
end for

4.3. Attack Method

Before evaluating the invulnerability performance of the network, we briefly define the common types of attacks. The two primary types of attacks that cause damage to the network topology are random attacks and malicious attacks. Random attacks (RD—random degree attack) refer to the failure of nodes in the network due to random irresistible factors such as environment and energy. Malicious attacks are based on the local information of the network to precisely attack the weakest part of the network or the most critical nodes, resulting in network collapse. There are various forms of malicious attacks based on several metrics, mainly Initial High Degree node-based (IHD), Recomputed High Degree node-based (RHD), Initial High Betweenness centrality-based (IHB), Recomputed High Betweenness centrality-based (RHB), etc. [24]. We construct a probability matrix of the network under threat and propose a High Threat Degree Node-Based Attack (HTD) based on the specific scenarios of marine sensor network deployment, the initial node degree, and the maximum load of nodes.

Definition 2 (High Threat Degree Attack). *The attack based on high threat degree nodes is not only considering the current high degree node sequence of the network, but also the high degree nodes of the initial network and the sequence of node load, which is calculated by the aggregated network nodes, edge criticality. The probability of threat to each node in the network is calculated as a sequence of high threat-degree nodes, and the fusion of the sequences is calculated as follows:*

$$P = \frac{1}{6} InitialDegree + \frac{1}{3} CurrentDegree + \frac{1}{2} MNL \tag{15}$$

The high threat-degree node sequence calculated by the above definition is shown in Figure 5. In the simulation experiments, for random attacks, we adopted random node removal to simulate the impact of random attacks on the network. In the removal strategy, one node is randomly removed from the network each time, and a total of 20 nodes will be removed. The maximum connected subgraph size of the network will be recorded. For deliberate attacks, we removed high threat-degree nodes to simulate the impact of deliberate attacks on the network. In the removal strategy, the node with the highest threat probability is removed each time, and a total of 20 nodes will be removed. The maximum connected subgraph size of the network will be recorded.

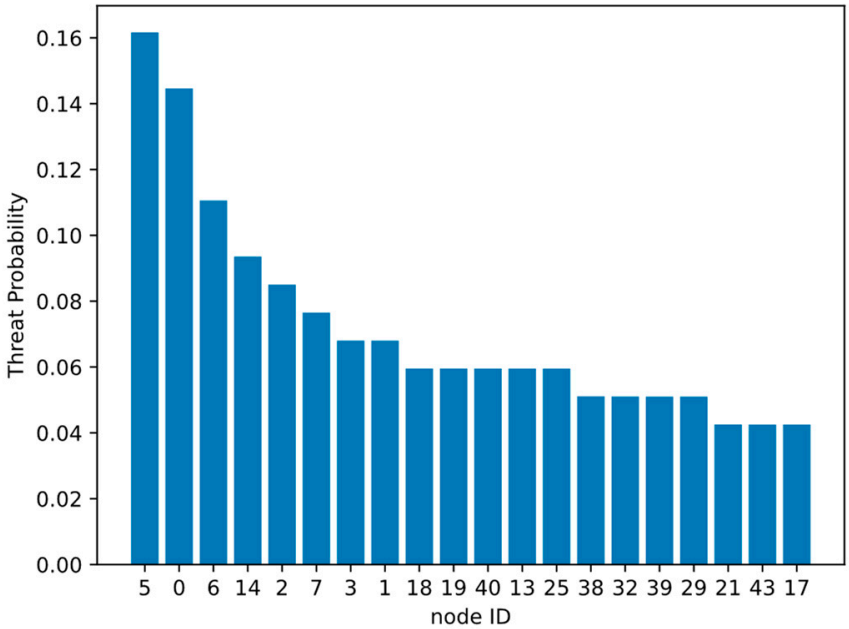


Figure 5. Sequence of high threat-degree nodes (partial).

5. Experimental Detail and Result Analysis

5.1. Dataset Construction

The data used in the experiments are mainly based on the scale-free network model summarized by Albert and Barabasi to construct a marine sensor network topology with scale-free characteristics. A scale-free topology of marine sensor networks is constructed, which consists of randomly placing two nodes in the node distribution area, starting from the third node, generating each new node based on the connection probability according to the preferred connection principle, and selecting two current nodes based on the generation location of nearby new nodes until all nodes are connected, and then the edges of the network are divided by 5% to simulate the accidental disconnection in the node communication.

The network topology based on random mechanisms to generate nodes and meritocratic connections to generate edges has no attributes for nodes and link weights of 1. We construct the characteristic attributes of the network nodes, i.e., the characteristic matrix of the network, based on many of the above definitions. At this point, we have completed the construction of the data required for the experiment, and all the indicators meet the requirements of the experimental application.

5.2. Experimental Design

This section aims to verify the effectiveness and superiority of the above-proposed network topology optimization algorithm and to test the destructiveness of the new network topology based on random and malicious attacks on the constructed network dataset generated by the optimization of the VGAE-AMF algorithm. In this paper, experimental simulations are conducted in Python, based on NetworkX (a network tool library), Pytorch (a deep learning framework), and Pyro (a deep probabilistic programming language). The node distribution area is assumed to be 100×100 m^2, and the parameters used are shown in Table 1.

Table 1. Simulation experiment parameters.

Parameters	Value
Number of network nodes (m)	100
Number of network edges (W)	196
Number of iterations ($epoch$)	200
Learning rate (lr)	0.01
Node key factor weight (α)	0.6
Edge key coefficient regulator (c)	0.7
Edge key coefficient weight (μ)	0.4
Invulnerability entropy measure weight (γ)	1.5
Node degree weight (β)	1
Node maximum load regulator (λ)	0.4

In order to study the structure and characteristics of scale-free networks more quickly, easily, and accurately, we use NetworkX, a mainstream network toolkit based on Python, which is a widely used software for complex network research, visualization, model simulation, and structural feature analysis of complex networks. For variational inference, we use Pyro, a deep probabilistic programming language open-sourced by Uber that shows good flexibility and scalability in the integration of modern deep learning and Bayesian modeling.

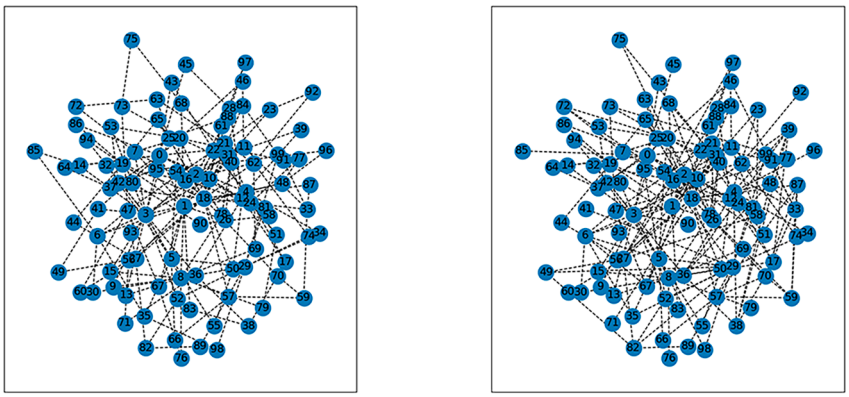
5.3. Simulation Experiment Results and Analysis

5.3.1. Analysis of Invulnerability Based on Evaluation Index

The network scale after optimization by the VGAE-AMF algorithm does not change from the initial network size, which is still 196 edges generated by 100 nodes. We control the number of edges in the network by strictly constraining the link generation matrix

generated by VGAE through AMF and fine-tune it based on the initial graph size to ensure that the network still obeys the power-law distribution.

From Figure 6, we find that a small number of nodes in the newly generated network topology have only one edge link, and by comparing the adjacency matrix, we find that the optimized network topology increases the degree of low-degree nodes in the original topology. The network topology is optimized to increase the destructiveness of the network by removing the links with lower destructiveness gains from the "less important" nodes and giving the quota of these links to the more needed nodes to reduce the number of bridge edges.



(**a**) (**b**)

Figure 6. Network topology before and after algorithm optimization: (**a**) initial network; (**b**) optimized network.

The first aspect we focus on in the invulnerability analysis is the invulnerability entropy measure and robustness metric based on the node and edge critical coefficients described above in this paper. Then, we perform an invulnerability analysis of the network topology before and after optimization based on two different attack strategies.

From Table 2, we can see that after optimization, the invulnerability entropy measure is improved by 5.28%, which indicates that the comprehensive invulnerability performance of the network is improved after optimization by the VGAE-AMF algorithm; and the robustness metric based on malicious attacks against high nodes is improved by 39.65%, which indicates that the invulnerability of the network is significantly improved in the face of malicious attacks. The VGAE-AMF algorithm effectively improves the network's resistance to destruction of the initial network.

Table 2. Network invulnerability entropy measures and robustness metric before and after algorithm optimization.

	Initial Network	Optimized Network	Lifting Value
Invulnerability Entropy Measure	3.43070191	3.61191568	5.28%
Robustness Metrics	0.17980198	0.25108911	39.65%

5.3.2. Analysis of Invulnerability Based on Different Attack Types

Based on the multiple attack types mentioned above, we selected random attacks and malicious attacks based on high threat degrees as the main attack types for experiments. The experimental results are shown in the following figure.

As can be seen from the Figure 7, with the removal of nodes, the size of the maximum connected subgraph of the network before and after optimization shows a decreasing trend, but both networks show a strong resistance when facing random attacks, which reflects the robustness against random attacks obtained by the network obeying the power-law distribution.

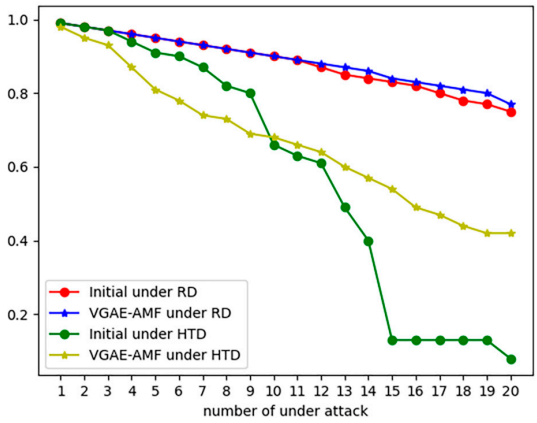


Figure 7. Network connectivity coverage for different types of attacks on the network (RD—Random Attack; HTD—High Threat Degree Attack).

Before and after optimization, the network topology shows different resistance when facing deliberate attacks. As the number of removed nodes increases, the maximum connectivity subgraph size of the initial network tends to shrink more rapidly, leading to network partitioning and loss of network functionality. The optimized network topology shows greater resistance to deliberate attacks but still reduces the connectivity coverage by nearly 50% compared to random attacks, suggesting that deliberate attacks are a form of attack that can seriously damage the functionality of the network.

5.3.3. Effect of Hyperparameters on Network Invulnerability

In constructing the invulnerability entropy, the above algorithm provides three adjustable parameters, i.e., α-node key factor weight, c-edge key coefficient regulator, and λ-regulator of invulnerability entropy, and node degree, to fit the invulnerability performance of the network. These parameters are involved in the construction of the feature matrix and thus have an impact on the optimized network's invulnerability, so we use different combinations of parameters to optimize the initial network and evaluate the damage of the optimized network based on the robustness index.

The simulation results show that the performance of an HTD-based network against deliberate attacks was improved and peaked at $\alpha = 0.6$, but gradually decreased with the increase of the value of α. We analyze the reason for this phenomenon, which is that since the edge criticality will be embedded in the nodes at both ends on average, the node criticality plays a decisive factor when the edge weight is small, which does not represent the contribution of the edge criticality to the network feature construction well and leads to a decrease in the network's invulnerability performance. As the edge weights increase, the role of edge criticality starts to be more prominent, providing better quality features for the graph convolutional network and improving the optimization performance of the model, which results in better invulnerability performance. However, as the edge weights continue to rise in the ratio, the edge criticality coefficients dominate and annihilate the features of node criticality, which leads to the inferior invulnerability performance of the optimized network.

After determining the optimal value of α, we adjusted the values of c. With $c < 0.5$, the simulation experimental results showed that the optimized network was weak against intentional damage, and with the increase of the value of c, the network optimized by the VGAE-AMF algorithm rapidly improved against intentional HTD attacks and $c = 0.9$ reached the optimal value at the time. However, c is still a regulating factor to represent the edge criticality of the network, and the value of c finally set to 0.7 in order to match the edge weights for tuning.

The above experiments determine the optimal values of α and c, and when tuning the values of λ, we found that the invulnerability performance of the optimized network is

extremely sensitive to the values of λ, but performed well in the interval $\lambda \in [0.3, 0.5)$. The invulnerability entropy, as a comprehensive characterization of the node criticality and edge criticality of the network, does not highlight the scale-free property of the scale-free network, but the scale-free property represented by the node degree sequence plays an important role in maintaining the ability of the network to resist random attacks, and should occupy a larger proportion, so the value of λ is 0.4.

From Figures 7–10, we can find that the network optimized by the VGAE-AMF algorithm performs better than the initial network under HTD attacks if we set the values of α, c and λ appropriately, and their performances under RTD attacks are similar, which has achieved our expected goal.

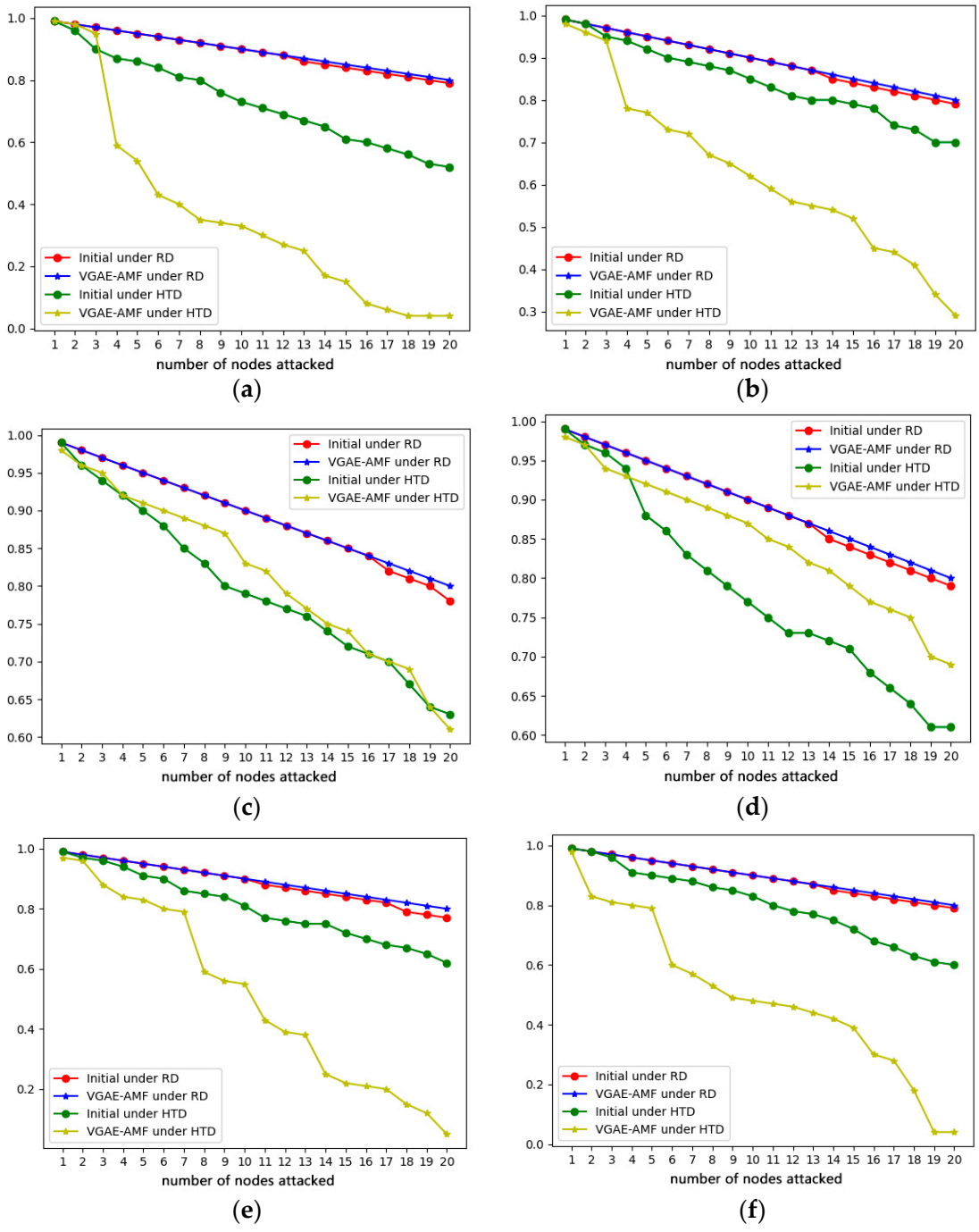


Figure 8. Effect of different values α on algorithm optimization: (**a**) $\alpha = 0.2$; (**b**) $\alpha = 0.3$; (**c**) $\alpha = 0.5$; (**d**) $\alpha = 0.6$; (**e**) $\alpha = 0.7$; (**f**) $\alpha = 0.9$.

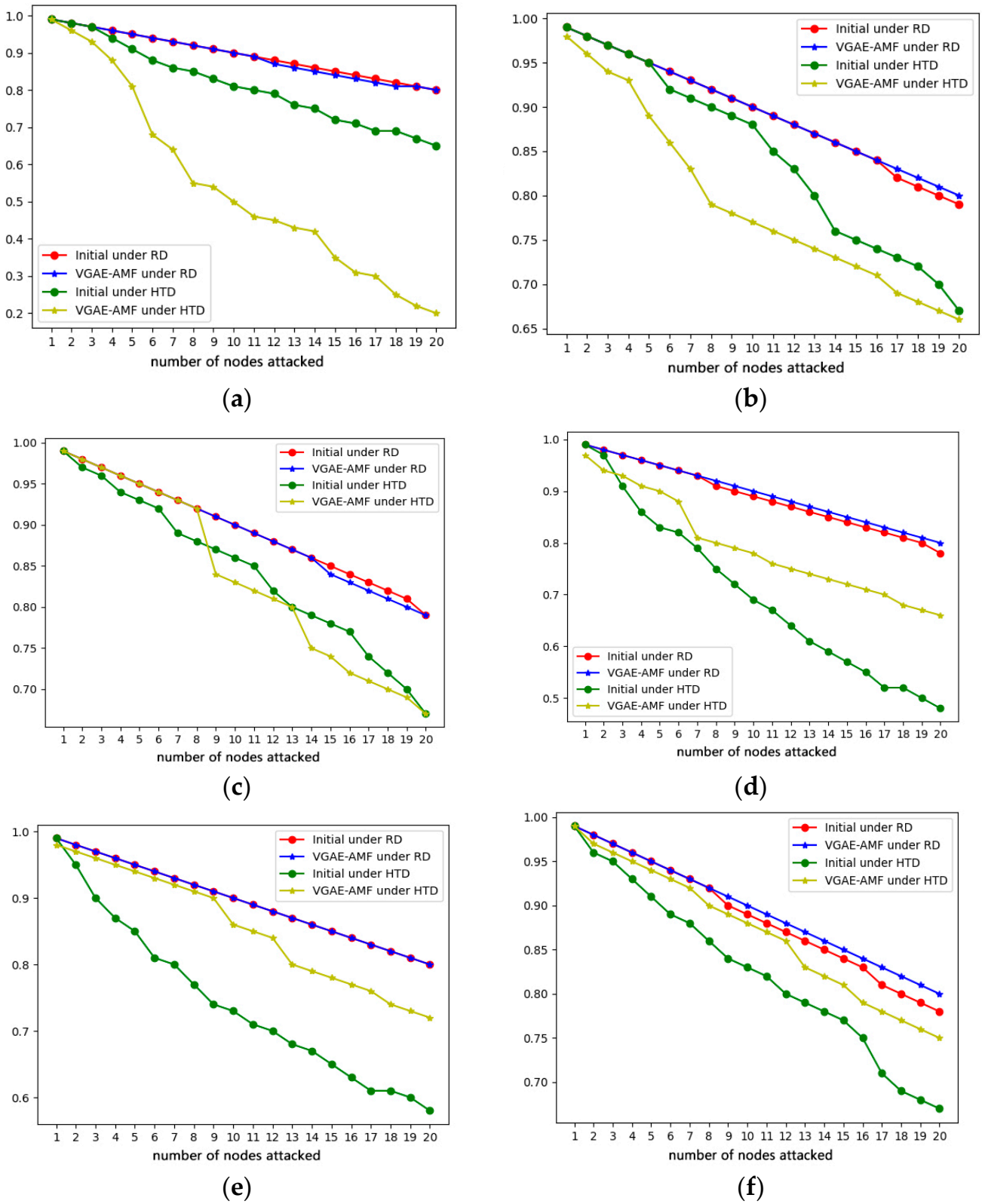


Figure 9. Effect of different values c on algorithm optimization: (**a**) $c = 0.1$; (**b**) $\alpha = 0.3$; (**c**) $\alpha = 0.5$; (**d**) $c = 0.7$; (**e**) $c = 0.8$; (**f**) $c = 0.9$.

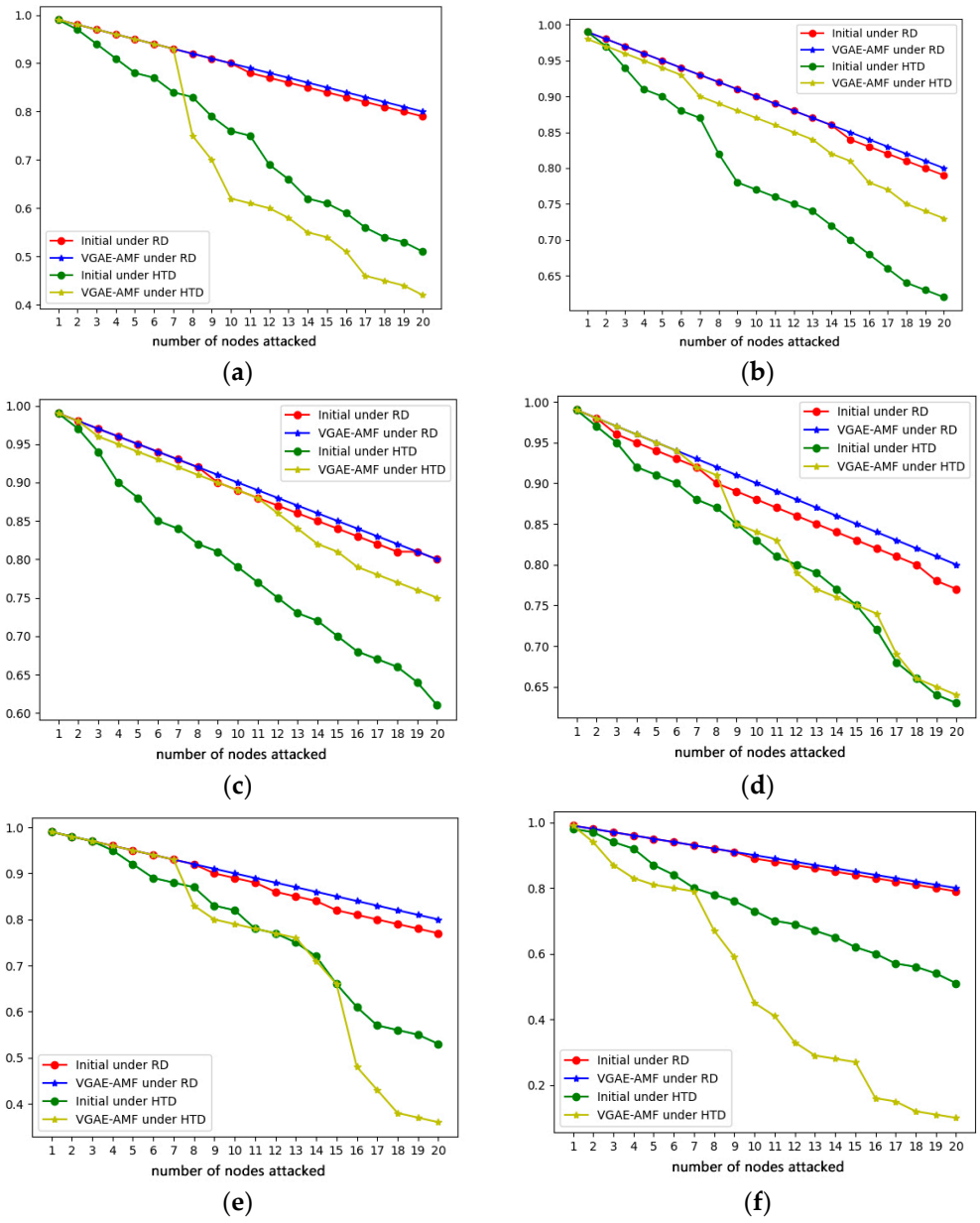


Figure 10. Effect of different value λ on algorithm optimization: (**a**) $\lambda = 0.2$; (**b**) $\lambda = 0.3$; (**c**) $\lambda = 0.4$; (**d**) $\lambda = 0.5$; (**e**) $\lambda = 0.6$; (**f**) $\lambda = 0.7$.

6. Conclusions

Based on the idea of link prediction, this paper proposes the variational graph auto-encoders cascaded adaptive multilayer filter (VGAE-AMF) algorithm, which reduces the number of bridge edges by calculating the similarity of nodes and determining the link existence between nodes based on the link generation probability matrix. Subsequently, the VGAE-AMF algorithm is applied to its marine wireless sensor network topology optimization by constructing a marine wireless sensor network with scale-free properties. The experiments demonstrate that the optimized network has a significant improvement in the invulnerability index, thus indicating the feasibility and superiority of the VGAE-AMF algorithm to optimize the network topology.

In this paper, we mainly focus on the network topology, construct a non-directional topology graph, and work on generating a more invulnerable network topology without considering the network scale too much. We only consider the network topology with 100 nodes or less in this paper and do not do much adaptation testing for larger networks.

In the next work, we will further consider more realistic constraints and further improve our model based on dynamic graphs for real-time changes in the network topology.

Author Contributions: Y.Z. (Ying Zhang) conceived and supervised the research and experiments, contributed as the lead author of the article; Z.Z. wrote the draft of the manuscript; Y.Z. (Yu Zhang) and Q.Z. analyzed and validated the data and the experiments. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (no. 61673259); supported by Shanghai "Science and Technology Innovation Action Plan" Hong Kong, Macao and Taiwan Science and Technology Cooperation Project (no.21510760600); and also supported by Capacity Building Project of Local Colleges and Universities of Shanghai (no. 21010501900).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author.

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

References

1. Fukuda, K.; Takyu, O.; Shirai, K.; Ohta, M.; Fujii, T.; Sasamori, F.; Handa, S. Transmit Control and Data Separation in Physical Wireless Parameter Conversion Sensor Networks with Event Driven Sensors. In Proceedings of the IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet), Anaheim, CA, USA, 14–17 January 2018; pp. 12–14. [CrossRef]
2. Awan, K.M.; Shah, P.A.; Iqbal, K.; Gillani, S.; Ahmad, W.; Nam, Y. Underwater Wireless Sensor Networks: A Review of Recent Issues and Challenges. *Wirel. Commun. Mob. Comput.* **2019**, *2019*, 6470359. [CrossRef]
3. Zhang, Y.; Li, P.; Wang, X. Intrusion Detection for IoT Based on Improved Genetic Algorithm and Deep Belief Network. *IEEE Access* **2019**, *7*, 31711–31722. [CrossRef]
4. Fu, X.; Yao, H.; Yang, Y. Exploring the invulnerability of wireless sensor networks against cascading failures. *Inf. Sci.* **2019**, *491*, 289–305. [CrossRef]
5. Dong, M.; Li, H.; Li, Y.; Deng, Y.; Yin, R. Fault-tolerant topology with lifetime optimization for underwater wireless sensor networks. *Sādhanā* **2020**, *45*, 162. [CrossRef]
6. Zhang, Y.; Zhang, Z.; Chen, L.; Wang, X. Reinforcement Learning-Based Opportunistic Routing Protocol for Underwater Acoustic Sensor Networks. *IEEE Trans. Veh. Technol.* **2021**, *70*, 2756–2770. [CrossRef]
7. Younis, M.; Lee, S.; Senturk, I.F.; Akkaya, K. Topology Management Techniques for Tolerating Node Failure. In *The Art of Wireless Sensor Networks*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 273–311. [CrossRef]
8. Czerwinski, D.; Przylucki, S.; Wojcicki, P.; Sitkiewicz, J. Path Loss Model for a Wireless Sensor Network in Different Weather Conditions. In *The Art of Wireless Sensor Networks*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 718, pp. 106–117. [CrossRef]
9. Wu, H.; Suo, M.; Wang, J.; Mohapatra, P.; Cao, J. A Holistic Approach to Reconstruct Data in Ocean Sensor Network Using Compression Sensing. *IEEE Access* **2017**, *6*, 280–286. [CrossRef]
10. Xing, L. Cascading Failures in Internet of Things: Review and Perspectives on Reliability and Resilience. *IEEE Internet Things J.* **2020**, *8*, 44–64. [CrossRef]
11. Zhang, Y.; Yang, G.; Zhang, B. FW-PSO Algorithm to Enhance the Invulnerability of Industrial Wireless Sensor Networks Topology. *Sensors* **2020**, *20*, 1114. [CrossRef] [PubMed]
12. Fu, X.; Yang, Y. Modeling and analysis of cascading node-link failures in multi-sink wireless sensor networks. *Reliab. Eng. Syst. Saf.* **2020**, *197*, 106815. [CrossRef]
13. Fu, X.; Yao, H.; Yang, Y. Cascading failures in wireless sensor networks with load redistribution of links and nodes. *Ad Hoc Netw.* **2019**, *93*, 101900. [CrossRef]
14. Fu, X.; Yao, H.; Yang, Y. Sink-Convergence Cascading Model for Wireless Sensor Networks with Different Load-Redistribution Schemes. *Complexity* **2019**, *2019*, 7630168. [CrossRef]
15. Wang, Y.; Fu, X.; Yang, Y.; Postolache, O. Analysis on cascading robustness of energy-balanced scale-free wireless sensor networks. *AEU Int. J. Electron. Commun.* **2021**, *140*, 153933. [CrossRef]
16. Ren, W.; Wu, J.; Zhang, X.; Lai, R.; Chen, L. A Stochastic Model of Cascading Failure Dynamics in Communication Networks. *IEEE Trans. Circuits Syst. II Express Briefs* **2018**, *65*, 632–636. [CrossRef]
17. Tian, X.-G.; Zhu, Y.-C.; Luo, K.; Zhang, C.-M. Adaptive reconstruction model for command-and-control system under information age based on complex network theory. *Syst. Eng. Electron.* **2013**, *35*, 91–96.
18. Hu, B.; Li, F. Repair strategies of scale-free networks under multifold attack strategies. *Syst. Eng. Electron.* **2010**, *1*, 43–47.

19. He, X.; Zhao, H.; Cai, W.; Liu, Z.; Si, S.-Z. Earthquake networks based on space–time influence domain. *Phys. A: Stat. Mech. Its Appl.* **2014**, *407*, 175–184. [CrossRef]
20. Chen, G.; Sun, P.; Zhang, J. Repair Strategy of Military Communication Network Based on Discrete Artificial Bee Colony Algorithm. *IEEE Access* **2020**, *8*, 73051–73060. [CrossRef]
21. Feng, X.; Zhao, J.C.; Xu, K. Link prediction in complex networks: A clustering perspective. *Eur. Phys. J. B* **2012**, *85*, 3. [CrossRef]
22. Martínez, V.; Berzal, F.; Cubero, J.-C. A Survey of Link Prediction in Complex Networks. *ACM Comput. Surv.* **2016**, *49*, 1–33. [CrossRef]
23. Zhou, M.-Y.; Liao, H.; Xiong, W.-M.; Wu, X.-Y.; Wei, Z.-W. Connecting Patterns Inspire Link Prediction in Complex Networks. *Complexity* **2017**, *2017*, 8581365. [CrossRef]
24. Hu, S.; Li, G. TMSE: A topology modification strategy to enhance the robustness of scale-free wireless sensor networks. *Comput. Commun.* **2020**, *157*, 53–63. [CrossRef]
25. He, M.; Liang, W.; Chen, Q.; Chen, X.; Chen, J. Optimization method through topology reconfiguration for mobile underwater wireless sensor networks. *J. Commun.* **2015**, *36*, 78–87.
26. Chen, Q.; He, M.; Dai, F.; Zhu, C. Energy-Efficient Connectivity Re-Establishment in UASNs with Dumb Nodes. *IEICE Trans. Inf. Syst.* **2018**, *101*, 2831–2835. [CrossRef]
27. Priyadarshini, R.R.; Sivakumar, N. Failure prediction, detection & recovery algorithms using MCMC in tree-based network topology to improve coverage and connectivity in 3D-UW environment. *Appl. Acoust.* **2020**, *158*, 107053. [CrossRef]
28. Liben-Nowell, D.; Kleinberg, J. The link-prediction problem for social network. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 1019–1031. [CrossRef]
29. Barabási, A.-L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [CrossRef]
30. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **1953**, *18*, 39–43. [CrossRef]
31. Grover, A.; Leskovec, J. Node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.
32. Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral networks and locally connected networks on graphs. *arXiv* **2013**, arXiv:1312.6203.
33. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional network. *arXiv* **2016**, arXiv:1609.02907.
34. Kipf, T.N.; Welling, M. Variational graph auto-encoders. *arXiv* **2016**, arXiv:1611.07308.
35. Zhang, M.; Chen, Y. Link prediction based on graph neural networks. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 5165–5175.
36. Klicpera, J.; Bojchevski, A.; Günnemann, S. Predict then propagate: Graph neural networks meet personalized pagerank. *arXiv* **2018**, arXiv:1810.05997.
37. Brandes, U. A faster algorithm for betweenness centrality. *J. Math. Sociol.* **2001**, *25*, 163–177. [CrossRef]
38. Gao, X.; Li, K.; Chen, B. Invulnerability Measure of a Military Heterogeneous Network Based on Network Structure Entropy. *IEEE Access* **2017**, *6*, 6700–6708. [CrossRef]
39. Schneider, C.M.; Moreira, A.; Andrade, J.; Havlin, S.; Herrmann, H.J. Mitigation of malicious attacks on networks. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3838–3841. [CrossRef]