Working capacity of deep networks with topology attacks

Longjie Zhang

University of Electronic Science and Technology of China

Yong Chen (✉ ychencd@uestc.edu.cn)

University of Electronic Science and Technology of China

Ikram Ali

University of Electronic Science and Technology of China

Research Article

Keywords: deep, attacks, topology, networks, learning, security

Posted Date: November 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1042846/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Working capacity of deep networks with topology attacks

Longjie Zhang¹, Yong Chen^{1,*}, Ali Ikram¹

School of Automation engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

*Corresponding author: ychencd@uestc.edu.cn

ABSTRACT

Deep learning plays an important role in the development of artificial intelligence (AI) technology. The security of deep networks has become the crucial thing to be considered. When the deep learning algorithms are implemented in the hardware platform, the interference for topology structure will appear because of cyber-attacks. We analyze the working capacity of acyclic deep networks under the topology attacks and injection attacks. Considering the topology structure of the deep network, the maximum working capacity is studied under the topology attacks and injection attacks. Furthermore, the robustness of the random networks is researched and the structural robustness index (SRI) is proposed to measure the toleration for the topology attacks. This work supplies some suggestions for building a robust deep network and improving the endogenous safety and security (ESS) of the deep networks.

Introduction

Deep neural network is widely used in many fields. The increasing mature optimal algorithms make the deep networks more practical¹⁻³. To adopt various usage scenarios, the deep neural networks have been extended to various deep networks. such as convolutional neural networks (CNN), recurrent neural networks (RNN), and graph neural networks (GNN), which are good at image data processing^{4,5}, serialized data processing⁶, and structured data processing⁷, respectively. The basic theory architectures of the deep networks have been applied successfully in numerous scenes and tend to be more and more mature. Meanwhile, the industrial applications of the deep networks are concentrated by many researchers for classification, detection, estimation, recognition. To realize the algorithm of the deep networks in practice, the joint work with software and hardware is generally implemented^{8,9}. Communication networks are widely used in the implementation of intelligent algorithms. However, the open network connection will make the deep networks more vulnerable.

In recent years, the attacks for the deep learning and corresponding defense strategy have received extensive attention. The recent work of the adversarial example for deep learning has been summarized by ref¹⁰. The adversarial example is regarded as the perturbation and it can fool deep learning by injecting adverse information into the input samples. The generative adversarial networks (GAN) are the well-known adversarial examples. The adversarial examples consider the perturbation for the inputs of the deep networks, and the corresponding network structures are assumed to be steady. However, the opaque characteristic makes the estimation for the deep network model unreliable because of the impure data inputs and physical uncertainty in the communication device^{11,12}. These uncertainties come from the malicious attack injection through communication channels or the parameter drift of the physical components. In existing work, the topology attacks for the GNNs are researched by ref¹³, the minimax structure optimal method is discussed based on the graph theory. Besides, the node injection poisoning attacks for the GNNs are considered by ref¹⁴. The reinforcement learning method is applied to reduce the performance of the GNNs and more effective attacks for the graph nodes are designed. Motivated by the characteristic of explanation of GNNs (Gnnexplainer), a novel attack framework was proposed and it can attack both a GNN model and the explanations of the GNN in ref¹⁵. However, previous work is mainly concentrated on the research of the attacks on the network inputs and general graph structures, while the acyclic graph is more universal in deep networks. Most artificial neural networks (ANNs) consist of the special acyclic graph structure with forward propagation networks. The research for the structural robustness of the deep neural networks (DNN) with the acyclic graph is the key topic for the security of the deep networks.

Here we explore the topology attacks and injection attacks for the deep networks. Motivated by the topological limits to the parallel processing capability of the multitask networks¹⁶. The deep networks can give play to the maximum work capability by removing the nodes, which is described as dropout¹⁷. But the deep networks are generally considered as the black box, the information bottleneck theory solves this problem to a certain extent¹⁸. Based on the above theory, we analyze the work capability of the deep networks under the topology attack and injection attacks. From the performance trend of the deep networks, we can obtain the extent of the topology attacks that will destroy the work capacity of the deep networks. This will be a meaningful suggestion for designing the optimal dropout rate, which can keep the best network robustness. By arranging as few nodes as possible, the toleration of the deep networks for the injection attacks will be strengthened. The topology attack can be tolerated before the attack rate rises to the extent that has destructive impacts on deep networks. To make the results explicable, an iterative algorithm is described based on the information bottleneck theory here.

54 Results

55 **Deep Neural Networks.** For a classical deep network, we define the activation function $f(\bullet)$. And the l th layer output is

Figure 1 | The topology and weights attacks for the artificial neural networks. (a) The training data with characteristics and labels, the training data input can be images, text, voice, and time series data etc. (b) The neural networks with topology attacks (The nodes with light blue (golden) represent the inputs (outputs), the red nodes represent the nodes under topology attacks), which contains forward propagation and backward propagation. (c) The detail one-layer network model. (d) The one-layer network under the topology attacks (the edges with red) and injection attacks (the edges with light blue). The malicious topology attacks will delete the connected nodes and edges of the forward propagation. The interference sometime will be easier to appear in the networks because of the uncertainty in the application site. (e) The training and validation accuracy for the fashion mnist dataset with full connected five layers DNN. (f) The training and validation accuracy for the five layers DNN with dropout rate 0.3.

$$\mathbf{y}^{(l+1)} = f^{(l+1)}(\mathbf{w}^{(l+1)} \mathbf{y}^{(l)} + \mathbf{b}^{(l+1)}) \quad (1)$$

where $\mathbf{y}^{(l)}, l=0,1,\dots,n$ is the inputs of the $(l+1)$ th layer, $\mathbf{w}^{(l+1)}$ represents the weight of forward propagation and $\mathbf{b}^{(l+1)}$ is the bias. The work process of deep learning is divided into two steps: forward propagation and backward propagation. By inputting the training data to the deep networks, the output of the forward propagation is the predicted labels. We can rely on the cross-entropy or mean square deviation to compare the predicted labels with real labels, and the loss function to be optimized can be built. Based on the forward propagation equations, the error transfer equation will be constructed to build the backward propagation. The gradient descent is applied to update the weights and bias of the forward propagation (1). When the loss function convergence to the threshold, the learning process will be completed and we can obtain the optimized deep networks. Then the model can be applied to perform some functions, such as classification, prediction, evaluation, etc. Over these years, the trained models of machine learning are generally provided to the user service systems, the deep learning models will be exposed to the dangerous cyber environment.

The misguided person will painstakingly look for the system vulnerabilities to steal the model information. The malicious attack will be generated to destroy the structure of the deep networks. For example, $\mathbf{w}^{(l+1)}$ can reflect the nodes and edges information of the networks, attackers will change the network structure by deleting some nodes and edges or injecting stochastic noise to the weight values. Here, our research will consider this situation and find the security situation of the deep networks under the topology attacks and injection attacks.

Attacks for deep networks. To research the influence of topology attacks and injection attacks, we give the attack models for the networks here. Let $w_{ij}^{(l),a} = w_{ij}^{(l)} + \mathcal{A}_{ij}^{(l)}$ represent the attack or interference for the weights of the deep networks, this kind of attack is generated by the malicious attackers or the imprecise physical device.

We take $\mathbf{w}^{(l+1)}$ as the adjacency matrix of the l th layer network, and the topology attack can delete the nodes of the deep network and it can block the propagation process of the corresponding nodes. For each node, the attack probability distribution satisfies the Bernoulli distribution. When the topology attacks are generated, the $(l+1)$ th layer input is:

$$\tilde{\mathbf{y}}^{(l)} = \mathbf{r} \circ \mathbf{y}^{(l)} \quad (2)$$

where $\mathbf{r} = [r_1, r_2, \dots, r_n]$ and $r_k \sim B(1, p)$. Figure 1 shows the topology attacks and injection attacks for the classical artificial neural networks (ANN). In fact, the ANN can be regarded as the acyclic graph. When the nodes are removed with random probability, the training efficacy of the deep networks will be different. According to the percolation theory, the critical exponents p_c for the percolation transition of networks reflect the robustness of the networks, and the robustness of the networks will be better with less value of the critical exponents.

As described in ref¹⁹, the critical point of the regular graph is $p_c = \frac{1}{k_0} - 1$, where k_0 is the degree of every nodes. We can see that the robustness of networks will be stronger with increasing the number of nodes. However, deep networks with numerous nodes will face the problem of overfitting. To find the compromise between the robustness and the overfitting of the network, we should analyze the topology structure of the deep neural networks (DNN) under the errors and the attacks.

Here, the dataset *fashion_mnist* in Keras of TensorFlow 2.3.0 is applied as the inputs of the deep networks. Five layers of DNN with 100 input nodes are designed in our experiments. It is cumbersome to analyze the structure of our designed DNN directly. We give an example of neural networks with 3 layers and 4 inputs nodes to refine the analysis process. DNN is a classical black-box method, the multilayers model (1) is equivalent to the model shown by Figure 1(c). When the nodes of deep networks are removed as Figure 1(b), the corresponding input to output model can be simply expressed by Figure 1(d). We assume that the topology attacks remove the nodes during the training process. The noise injected into the weights of the networks can be represented by Figure 1(d), and the predictive error comes from the noise-induced redundancy $\mathcal{A}_a \cdot \mathbf{x}_i$. Obviously, the influence of predictive error will be less with fewer nodes. By comparing Figure 1(e) and Figure 1(f), the sparsity of the network nodes can effectively suppress the overfitting. In the next part, we will give the index to compromise the topology attacks and the injection attacks.

Compromise for the node removing and the injection noise. A normal fully connected deep neural network is shown in Fig 1 (b). When the weights of the connected edges are suffering interference, the influence of the interference will be accumulated. According to the proposed dropout process¹⁷, some nodes of the networks can be deleted to weak the cumulative phenomenon of interference. Therefore, the work performance of the network can be improved at the same time. When the topology attacks are implemented with a given mass of dropouts for the nodes, the work capacity of networks will be decayed. To defense the deep network from the attacks, we need to have ideas of the scale of the dropouts

Figure 2 | The topology and weights attacks for the classical neural networks. (a) The test accuracy of the deep networks with the topology attack rate (TA_rate) 0.0~0.9 (The percentage of the removed nodes) and the injection attack rate (IA_rate) 0.00~0.10 (The upper bounded of the random injection interference). As the increase of the IA_rate, the test accuracy will decrease monotonously. Differently, the variety of test accuracy isn't monotonous with the increase of TA_rate. The test accuracy will be best when TA_rate equal 0.6 for almost every IA_rate. **(b)** The test loss of the deep networks with TA_rate 0.0~0.9 and IA_rate 0.00~0.10. The test loss will be least when TA_rate equals to 0.5, which is different to the result of (a), the goal of the deep networks described here is a classification problem, the accuracy will be more persuasive in practice. Also, the test loss with TA_rate equaling 0.6 is also keep a small level compare with other curves. **(c)** The average network robustness index for acyclic deep networks with four time experiments.

As the suggestion in ref²⁰, the large degree will lead to the decay of the multitasking ability of the graph. To make a tradeoff between the work capacity and the computing complexity, many works have been done^{16,17}. Here we apply the dropout to represent the action of the attacks to analyze the influence of the topology attacks for the deep networks. As shown in Figure 2, when the number of the edge decays to a certain extent, the capacity of the network will be maximum. From Figure 2(a) and (b), the injection attack can reduce prominently the capacity of the deep network. Differently, as the topology attack rate increasing, the corresponding capacity doesn't decline monotonously but reaches the best when the topology attack rate is 0.5~0.6. From this phenomenon, we can conjecture that the optimal dropout rate for the deep network is around 0.5.

The topology attack will have a destructive effect if the percentage of the removed nodes reaches 0.5~0.6. Here, we assume that the input data is pure without any injected noise or attacks, so the higher robustness index level will reduce the influence of the interference on the weight edges of the networks. In the following work, we will apply the efficiency index of the deep network and information bottleneck theory to validate the above conclusion.

From the aforementioned analysis, the influence of the weight uncertain is reduced when the number of the connection reaches a certain quantity. So there will exist a saddle surface for the network capacity along with the dropout and the uncertain, and it can be regarded as the minimax problems:

$$\min_{p(\mathbf{r}^{(l)})} \max_{p(\mathcal{A}^{(l)})} \mathcal{F}[p(\mathbf{r}^{(l)}), p(\mathcal{A}^{(l)})] \quad (3)$$

Firstly, we can define the capacity function $\mathcal{F}(\bullet)$, which describes the working performance of the deep networks. To make the deep learning work more trustable, we consider improving the robustness of the networks. According to the structural controllability theory of the complex networks studied by ref²¹, the redundant should be reduced as much as possible. Consider the explanation of the neural networks, we apply the information bottleneck²² to solve the optimization problem (3). Equation (3) means that we need to find the best number of the input nodes to improve the robustness of the networks.

To evaluate the robustness of the deep network, we define a network robustness index (NRI) $p_D^{(k)}$, which is represented by the percentage of the driver nodes in deep networks with average degree k . Then, the average network robustness index (ANRI)

along with k is $\bar{p}_D = \frac{1}{N} \sum_k p_D^{(k)}$, where N is the number of the selected degrees. This index is proposed according to the

controllability of the complex networks²¹. We can measure the capacity of the deep networks by analyzing the property of the nodes in directed acyclic networks, which have the same structure as ANN.

Figure 3 | The topology attacks for the general networks. NRI for the (a) SF networks, (b) WS networks, (c) Random regular networks, and (d) ER networks. As the increase of the average degree, the characteristic of the networks will tend to reach the statistical properties, which can more accurately reflect the real properties. The NRI tend to be convergent as the average degree reach to a large enough scale. The structure example of (e) ER networks, (f) Random regular networks, (g) WS networks, and (h) SF networks. For these four kinds of networks, ER networks have more independent driver nodes like the nodes with blue No.1 and No.2, this property weaken the robustness of the networks. The regular network have same degree for every nodes, which reduces the number of the driver nodes and strengthen the robustness. The WS and SF networks have the structure characteristics between ER networks and regular networks.

131 From Figure 2 (c) there exist two extreme points for the robustness of the acyclic graph. The first one is that the attack rate is
 132 around 0.3, the overfitting and underfitting can be avoided. As shown in Figure 1 (f), the training accuracy is consistent with the
 133 validation accuracy as epochs going by. The second one represents the best robustness of the acyclic graphs with TA rate 0.65,
 134 from Figure 2 (a) and (b), the upper-bounded of the test accuracy and lower-bounded of the test loss are that the TA rate is around
 135 0.6, which means that the best robustness characteristic is realized by letting dropout reaches to 0.6. However, the robustness

Figure 4 | The structure robustness of the random networks. (a) The SRI for the random networks. (b) The ANRI for the random networks. The SRI and ANRI suggest that the ER networks have the weakest structure robustness, while the regular networks have the strongest structure robustness, WS and SF networks have the middle structure networks. These results support the discussion in Figure 3.

index of the final points is also high, this unexpected point is caused by the meager number of the network nodes, which arise the percentage of the driven nodes.

To predict the aforementioned two points and optimal network structure, the forward propagation of the deep network is considered as the acyclic digraph. Here we take the propagating process of the neural network as the Markov process which is described by ref²³. Differently, the topology attack process is also regarded as a link of the Markov process, and the dropouts scale can be expressed with conditional probability distributions. According to the information bottleneck iterative algorithm proposed in ref²², we provide an optimal dropout algorithm to find the construction of the networks with the best performance (see Methods).

The topology attacks for the general networks. In the above work, we have researched the robustness of the acyclic network with topology attacks and injection attacks. In recent years, the graph neural network (GNN) has been extensively applied in the structured data processing. The robustness of the random networks is also the critical factor for the structured deep networks under topology attacks. Here, the robustness of Erdős-Rényi (ER) random networks³³, Watts-Strogatz (WS)³⁴ small-world networks, scale-free (SF) networks³⁵, and random regular networks under topology attacks are studied. Consistent with the analysis of the acyclic networks, we assume that the inputs for the networks are pure here.

Different from the acyclic networks, the variety of the robustness index for the general networks monotonically with changing the topology attack (TA) rate. From Figure 3 (e), there exist many isolated nodes ER networks like the nodes with blue NO.1 and NO.2. If the topology attacks happen, the attribute of the nodes will be changed easier. As shown in Figure 3 (d), the change of NRI is remarkable, which means that the robustness of ER network will be influence sensitively. Figure 4 (b) shows that the ANRI $\bar{p}_D$ of ER network has few obvious inflection points compare with another three networks. Differently, the random regular networks have the same degree for every node and the relation between every node is homosexual. As shown in Figure 3 (f), we take any node as an object, they have the same associated node properties for the regular network. Figure 3 (c) shows that when the attack rate rises, the height of the RI is nearly the same, so the regular network will have stronger structure robustness. We can get the same conclusion from the slowly changing ANRI curve, which illustrates that the robustness of regular networks does not change dramatically to a certain extent as the TA rate increasing. The WS networks described in Figure 3 (g) have high aggregation coefficients close to the regular graph. This characteristic determines that the structural robustness is stronger than the general random networks, as shown in Figure 4 (b) and Figure 3 (b) there exists a flat area for the ANRI curve of the WS networks. Moreover, as the special random network [shown in Figure 3 (h)], the scale-free network will have a higher aggregation coefficient compared with general random networks and Figure 3 (a) has shown its characteristics of robustness, and the more regular network structures have stronger robustness compared with random networks as shown in Figure 4 (b).

To express the structural robustness of networks more intuitively, the structural robustness index (SRI) is given and it is defined by the variety rate of the ANRI along with the attack rate changing. According to the analysis for the robustness of four kinds of general networks in Figure 3, if SRI increase, the robustness of the network will be changed more largely and the structure robustness will be reduced as shown in Figure 4 (a). According to this meaning, we can define the SRI as follows:

$$SRI = \frac{1}{M} \sum_i \left| \frac{\bar{p}_D^{(a_i)} - \bar{p}_D^{(a_{i-1})}}{a_i - a_{i-1}} \right| \quad (4)$$

where $a_i, i=1,2,\dots,M$ is the i th topology attack rate and $\bar{p}_D^{(a_i)}$ represents the ANRI for the networks under the TA rate a_i .

According to (4), we can find that the SRI is the discrete derivative of ANRI. When the percentage of the removed nodes of the networks reaches 0.5-0.6. The robustness of the networks will be stronger, the increasing number of the nodes can only change the characteristic slowly. To compare the difference of the network structure robustness among the four kinds of networks, we summarize the results of Figure 4 (a) and Figure 5 is given to show the average SRI (ASRI) of the networks. From Figure 5 we can see that the random regular networks have the least ASRI, which means that the random regular networks have the strongest structure robustness.

Discussion

The robustness of the deep networks under the topology attacks and injection attacks is researched here. The deep networks have been deeply developed these years, for example, the deep quantum neural networks are researched to reply to the increasing amounts of data^{29,30}. Moreover, the individual differences among DNN instances were studied by ref³⁶. To solve the complexity control problem of deep networks, the hidden regularization under the exponential-type loss functions³⁷. However, the structures and the capacity of ANN have great distances with human brains. Considering this, the brain networks³⁸ and deep graph spectral evolution networks³⁹ are deeply researched. Fall all that, the robustness of deep network is the first important meaning of its application The controllability theory of complex networks suggests that the robustness of the networks is related to the property of the nodes²¹. The number of the driver nodes is the important influence factor for the robustness of the networks²¹. Here we apply the percentage of the driver nodes to measure the robustness of deep networks. The acyclic graph is studied to reflect the robustness of the deep networks. By comparing the results shown in Figure 2, the driver nodes can reflect the toleration of the deep networks to a certain extent under the topology attacks and injection attacks. Unexpectedly, the topology attacks will strengthen the toleration of deep networks to the injection attacks when the topology attack rate is below a threshold. To find this threshold, we design an iterative algorithm motivated by the information bottleneck theory, which regards the deep network as a Markov processing and reflects its working processing²².

However, the acyclic graph can't represent all the structures of deep networks, for example, the structure of GNN is designed based on the general networks. As a kind of special graph, the theory of complex networks can be applied to describe the deep acyclic networks. The robustness of the complex network has been researched by many researchers. The static tolerance to errors and attacks of the complex networks is the ability of the system to maintain the connectivity properties when the nodes or edges are removed²⁵. Different removed nodes and edges will bring strange consequences because the nodes and edges will act in different roles and have different influences on the robustness of the networks^{21,24,31,32,40}. Here the four kinds of random networks are considered and the toleration for the topology attacks is researched. We suggest a new structure robustness index to measure the toleration ability of the networks.

Figure 5 | The Average structure robustness index for the networks. The ASRI reflect the overall structure of the four kinds of random networks, which shows the strongest structure robustness of the regular networks.

Nowadays, deep networks are spread to various fields, the network communication is extensively utilized in Artificial intelligence. The widespread deep networks are sensitive to topology attacks. Here, we have researched the structural robustness of the acyclic networks and random networks, which appear in common deep networks. The experiments have shown that the injection attack will generate adverse impacts on the prediction of the deep network models. As the increasing of the injection attack rate, the accuracy of the deep networks will be reduced monotonously. Also, the results suggest that the topology attacks will restrain the influence of injection attacks caused by the attackers or the before the topology attack destroy the function of the deep networks. From the experiments, the toleration of the acyclic deep networks can be evaluated. Furthermore, the toleration for the topology attacks of random networks is researched, the regularized network will have better robustness.

Here our results want to provide some suggestions for improving the endogenous security of the deep networks. If the structures of the deep networks are strengthened, the cyber-attacks can be prevented before they happen. So, endogenous security should be concentrated on by more and more. In conclusion, the results of the research for the security of deep networks will suggest people improve the model security problems. By optimizing the network structures, the endogenous security of the network structure can be improved to deal with various security challenges brought by the network and the operating environment.

Methods

Iterative algorithm for acyclic deep networks. As described in ref²⁶, the mutual information between the input and output of the l th layer network is defined as:

$$\begin{aligned} I(\mathbf{Y}^{(l)}, \mathbf{Y}^{(l+1)}) &= D_{KL}[p(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)}) \| p(\mathbf{y}^{(l)})p(\mathbf{y}^{(l+1)})] \\ &= \sum_{\mathbf{y}^{(l)} \in \mathbf{Y}^{(l)}, \mathbf{y}^{(l+1)} \in \mathbf{Y}^{(l+1)}} p(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)}) \log \left(\frac{p(\mathbf{y}^{(l+1)} | \mathbf{y}^{(l)})}{p(\mathbf{y}^{(l+1)})} \right) \\ &= H(\mathbf{Y}^{(l)}) - H(\mathbf{Y}^{(l+1)} | \mathbf{Y}^{(l)}) \end{aligned} \quad (5)$$

where D_{KL} is the Kullback-Liebler divergence, and H denote the entropy of the variables.

According to the rate-distortion theory, we define a distortion function $d: \mathbf{Y}^{(l)} \times \mathbf{Y}^{(l+1)} \rightarrow \mathbb{R}^+$, which can be minimized to improve the representations from $\mathbf{Y}^{(l)}$ to $\mathbf{Y}^{(l+1)}$ in l th layer. The expected distortion is given by:

Figure 6 | The topology attack rate for the deep networks. Here $r_i = 0$ represents that the i th node will be removed by the topology attacks, otherwise, this node is safe. The probabilities of topology attacks for each node are independent each other. Then, the probability $p(\mathbf{r})$ satisfy multivariate Bernoulli distribution, by analyzing the characteristic of this distribution, we can obtain the dropout rate.

$$\mathbb{E}_{p(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)})}[d(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)})] = \sum_{\mathbf{y}^{(l)} \in \mathbf{Y}^{(l)}, \mathbf{y}^{(l+1)} \in \mathbf{Y}^{(l+1)}} p(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)}) d(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)}) \quad (6)$$

where $d(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)}) = D_{KL}[p(\tilde{\mathbf{y}}^{(l)} | \mathbf{y}^{(l)})p(\tilde{\mathbf{y}}^{(l)} | \mathbf{y}^{(l+1)})]$ represents the distortion. To obtain the optimal structure of the deep networks, we give the following objective function:

$$J[p^{(l)}(\mathbf{r})] = I(\mathbf{Y}^{(l)}, \mathbf{Y}^{(l+1)}) + \beta \mathbb{E}_{p(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)})}[d(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)})] \quad (7)$$

where β is the Lagrange multiplier, the expected distortion satisfies the constraint $\mathbb{E}_{p(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)})}[d(\mathbf{y}^{(l)}, \mathbf{y}^{(l+1)})] \leq D$ and D is upper bound of the expected distortion.

According to the example shown by Figure 6, here we can let $p_i(\tilde{\mathbf{y}}^{(l)} | \mathbf{y}^{(l)}) = p_i(\mathbf{r})$. Then we can obtain the topology optimal algorithm:

$$\begin{cases} p_{i+1}(\mathbf{r}) = \frac{p_i(\tilde{\mathbf{y}}^{(l)})}{Z_i(\tilde{\mathbf{y}}^{(l)}, \beta)} \exp \left[-\beta \sum_{r \in \mathbf{R}} \tilde{\mathcal{A}}_{p_i(r)}^{(l)} p_i(r) p(\mathbf{y}^{(l)}) \log \frac{\sum_{r \in \mathbf{R}} \tilde{\mathcal{A}}_{p_i(r)}^{(l)} p_i(r) p(\mathbf{y}^{(l)})}{\tilde{\mathcal{A}}^{(l)} p_i(\tilde{\mathbf{y}}^{(l)})} \right] \\ p_{i+1}(\tilde{\mathbf{y}}^{(l)}) = \sum_{\mathbf{y}^{(l)} \in \mathbf{Y}^{(l)}} p_{i+1}(\mathbf{r}) p(\mathbf{y}^{(l)}) \end{cases} \quad (8)$$

$$\text{where } Z_i(\tilde{\mathbf{y}}^{(l)}, \beta) = \sum_{\mathbf{y}^{(l)} \in \mathbf{Y}^{(l)}} p_i(\tilde{\mathbf{y}}^{(l)}) \exp \left[-\beta \sum_{r \in \mathbf{R}} \tilde{\mathcal{A}}_{p_i(r)}^{(l)} p_i(r) p(\mathbf{y}^{(l)}) \log \frac{\sum_{r \in \mathbf{R}} \tilde{\mathcal{A}}_{p_i(r)}^{(l)} p_i(r) p(\mathbf{y}^{(l)})}{\tilde{\mathcal{A}}^{(l)} p_i(\tilde{\mathbf{y}}^{(l)})} \right]$$

Experiments setting for the deep networks. The Hungary algorithm and maximum independent set algorithm^{27,28} are applied to find the nodes set with different properties and compute the robustness index of the deep networks. The dataset and the version of the *TensorFlow* can be found in *Results*. We researched the deep networks with four hidden layers and the number of the nodes are 100, 200, 200, and 200, respectively. After the model had been trained, we load the weight of the deep networks. The uniform random with different upper bound 0~0.1 to inject the weights of the deep networks and the topology attacks for the training process with dropout rate 0~0.9 For the simulation of the random networks, the network 2.6.2 is applied. And we searched the robustness of the random networks with 0~0.95 topology attacks rate and 0~100 average degree (*The average degree of WS network is from 10 to 110*).

References

- Poggio, T., Banburski, A. & Liao, Q. Theoretical issues in deep networks. *Proceedings of the National Academy of Sciences* **117**(48), 30039-30045 (2020).
- Liang, T., Poggio, T., Rakhlin, A. & Stokes, J. Fisher-Rao Metric, Geometry, and Complexity of Neural Networks. Preprint at <https://arxiv.org/abs/1711.01530> (2019).
- Banburski, A. Liao, Q. Miranda, B., et al. Theory III: Dynamics and generalization in deep networks. Preprint at <https://arxiv.org/abs/1903.04991> (2019).
- Chrkos, P., Frantzidis, C. A., Gkivogkli, P. T., Bamidis, P. D. & Kourtidou-Papadeli, C. Automatic Sleep Staging Employing Convolutional Neural Networks and Cortical Connectivity Images. *IEEE Transactions on Neural Networks and Learning Systems* **31**(1), 113-123 (2020).
- Protas, É., Bratti, J. D., Gaya, J. F. O., Drews, P. & Botelho, S. S. C. Visualization Methods for Image Transformation Convolutional Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*. **30**(7), 2231-2243 (2019).

6. Chandra, R. Competition and Collaboration in Cooperative Coevolution of Elman Recurrent Neural Networks for Time-Series Prediction. *IEEE Transactions on Neural Networks and Learning Systems* **26**(12), 3123–3136 (2015).
7. Sun, J., Zheng, W., Zhang, Q. & Xu, Z. Graph Neural Network Encoding for Community Detection in Attribute Networks, *IEEE Transactions on Cybernetics*, doi: 10.1109/TCYB.2021.3051021 (2021).
8. Maeda, Y. & Wakamura, M. Simultaneous perturbation learning rule for recurrent neural networks and its FPGA implementation. *IEEE Transactions on Neural Networks* **16**(6), 1664–1672 (2005).
9. Shah, N., Chaudhari, P. & Varghese, K. Runtime Programmable and Memory Bandwidth Optimized FPGA-Based Coprocessor for Deep Convolutional Neural Network. *IEEE Transactions on Neural Networks and Learning Systems* **29**(12), 5922–5934 (2018).
10. Yuan, X., He, P., Zhu, Q. & Li, X. Adversarial Examples: Attacks and Defenses for Deep Learning. *IEEE Transactions on Neural Networks and Learning Systems* **30**(9), 2805–2824 (2019).
11. Blundell, C., Cornebise, J., Kavukcuoglu, K. & Wierstra, D. Weight uncertainty in neural networks. *Proceedings of the 32nd International Conference on International Conference on Machine Learning* **37**, 1613–1622 (2015).
12. Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J. et al. A Survey of Uncertainty in Deep Neural Networks. Preprint at <https://arxiv.org/abs/2107.03342> (2021).
13. Xu, K., Chen, H., Liu, S., Chen, P. et al. Topology Attack and Defense for Graph Neural Networks: An Optimization Perspective. Preprint at <https://arxiv.org/abs/1906.04214> (2019).
14. Sun, Y., Wang, S., Tang, X., Hsieh, T. & Honavar, V. Adversarial Attacks on Graph Neural Networks via Node Injections: A Hierarchical Reinforcement Learning Approach. *Proceedings of The Web Conference 2020*, 673–683 (2020).
15. Fan, W., Jin, W., Liu, X., Xu, H. Jointly Attacking Graph Neural Network and its Explanations. Preprint at <https://arxiv.org/abs/2108.03388> (2021).
16. Petri, G., Musslick, S., Dey, B. et al. Topological limits to the parallel processing capability of network architectures. *Nat. Phys.* **17**, 646–651 (2021).
17. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. et al. Improving neural networks by preventing co-adaptation of feature detectors. Preprint at <https://arxiv.org/abs/1207.0580v1> (2012).
18. Shwartz-Ziv, R. & Tishby, N. Opening the Black Box of Deep Neural Networks via Information. Preprint at <https://arxiv.org/abs/1703.00810> (2017).
19. Molloy, M. & Reed, B. A Critical Point for Random Graphs with a given Degree Sequence. *Random Struct. Algorithms* **6**(2-3), 161–180 (1995).
20. Alon, N., Reichman, D., Shinkar, I., Wagner, T., Musslick, S. et al. A Graph-Theoretic Approach to Multitasking. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2097–2106 (2017).
21. Liu, Y.Y., Slotine, J.J. & Barabási, A.L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
22. Tishby, N., Pereira, F. C. & Bialek, W. The information bottleneck method. *Proc. 37th Allerton Conf. Commun. Comput.*, 368–377 (1999).
23. Tishby, N. & Zaslavsky, N. Deep learning and the information bottleneck principle, *Proc. IEEE ITW*, 1–5 (2015).
24. Nepusz, T., Vicsek, T. Controlling edge dynamics in complex networks. *Nature Phys* **8**, 568–573 (2012).
25. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: Structure and dynamics. *Physics Reports* **424**(4–5), 175–308 (2006).
26. Shwartz-Ziv, R., Tishby, N. Opening the Black Box of Deep Neural Networks via Information. Preprint at <https://arxiv.org/abs/1703.00810> (2017).
27. Tarjan, R. E. & Trojanowski, A. E. Finding a Maximum Independent Set. *SIAM J. Comput.* **6**(3), 537–546 (1977).
28. Robson, J.M. Algorithms for maximum independent sets. *Journal of Algorithms* **7**(3), 425–440 (1986).
29. Huang, H.Y., Broughton, M., Mohseni, M. et al. Power of data in quantum machine learning. *Nat Commun* **12**, 2631 (2021).
30. Beer, K., Bondarenko, D., Farrelly, T. et al. Training deep quantum neural networks. *Nat Commun* **11**, 808 (2020).
31. Ruths, J. & Ruths, D. Control Profiles of Complex Networks. *Science* **343**(6177), 1373–1376 (2014).
32. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. Network motifs: simple building blocks of complex networks. *Science* **298**(5594), 824–7 (2002).
33. Erdos, P.; and Renyi, A.. On the evolution of random graphs. *Publ. Math. Inst. Hungary. Acad. Sci.* **5**, 17–61 (1960).
34. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393** (6684), 440–442 (1998).
35. Barab, A.-L. & Albert, R. Emergence of Scaling in Random Networks. *Science* **286**, 509–512 (1999).
36. Mehrer, J., Spoerer, C.J., Kriesgskorte, N. et al. Individual differences among deep neural network models. *Nat Commun* **11**, 5725 (2020).
37. Poggio, T., Liao, Q. & Banburski, A. Complexity control by gradient descent in deep networks. *Nat Commun* **11**, 1027 (2020).
38. Gu, S., Pasqualetti, F., Cieslak, M. et al. Controllability of structural brain networks. *Nat Commun* **6**, 8414 (2015).
39. Etemadyrad, N., Li, Q. & Zhao, L. Deep Graph Spectral Evolution Networks for Graph Topological Evolution. *Proceedings of the AAAI Conference on Artificial Intelligence* **35**(8) 7358–7366 (2021).
40. Wang, L.Z., Chen, Y.Z., Wang, W.X. et al. Physical controllability of complex networks. *Sci Rep* **7**, 40198 (2017).

289 Acknowledgements

290 This work is supported by the National Nature Science Foundation of China under Grant (61973331), the National Key Research and Development Plan
291 Programs of China (2018YFB0106101). (Corresponding author: Yong Chen).

292 Author contributions

293 L. J. Zhang contributed to the writing, design of the methods and experiments. Y. Chen contributed to the ideas of this works, framework of the paper and guiding
294 the writing of the paper. Ali Ikram analyzed the results and contributed to the implement of simulation. All authors reviewed the manuscript.

295 Competing interests

296 The authors declare no competing interests.

297 Additional information

298 There has no additional supplementary information.