Improving Complex Network Controllability via Link Prediction

Ran Wei[1,2], Weiwei Yuan[1,2(✉)], Donghai Guan[1,2], Asad Masood Khattak[3], and Muhammad Fahim[4]

[1] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
ran_dlml@163.com, {yuanweiwei, dhguan}@nuaa.edu.cn
[2] Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing 210016, China
[3] College of Technological Innovation, Zayed University, Dubai, United Arab Emirates
Asad.Khattak@zu.ac.ae
[4] Institute of Information System, Innopolis University, Innopolis, Russia
m.fahim@innopolis.ru

Abstract. Complex network is a network structure composed of a large number of nodes and complex relationships between these nodes. Using complex network can model many systems in real life. The individual in the system corresponds to the node in the network and the relationship between these individuals corresponds to the edge in the network. The controllability of complex networks is to study how to enable the network to arrive at the desired state from any initial state by external input signals. The external input signals transmit to the whole network through some nodes in the network, and these nodes are called driver node. For the study of controllability of complex network, it is mainly to judge whether the network is controllable or not and how to select the appropriate driver nodes at present. If a network has a high controllability, the network will be easy to control. However, complex networks are vulnerable and will cause declining of controllability. Therefore, we propose in this paper a link prediction-based method to make the network more robust to different modes of attacking. Through experiments we have validated the effectiveness of the proposed method.

Keywords: Network controllability · Link prediction · Complex networks

1 Introduction

Social networking has brought convenience to our life but also brought some negative effects, for example, traffic congestion and large area blackouts. So controlling the state of a complex network is critical.

The research on the controllability [3] of complex networks focuses on controlling the state of the entire network by controlling a few nodes. Due to the large scale of complex networks and the vague information of individuals, traditional control theory can't be directly used to model. Liu et al. [1] put forward the theory of structural

X. B. Zhai et al. (Eds.): MLICOM 2019, LNICST 294, pp. 84–97, 2019.
https://doi.org/10.1007/978-3-030-32388-2_8

controllability [1], and define the concept of driver node which can control the state of other nodes. But structural controllability cannot solve the problem of controllability of networks with all known weights. Subsequently, Yuan et al. put forward the theory of strict controllability [2] and perfected the theory of controllability of complex networks.

Strict controllability has been verified that the controllability of a network is determined by its structure. In most cases, a change in the structure of the network manifests itself in the loss of some links. And usually the dense network structure is easy to control. Our experiments show that the loss of network links in most cases will increase the difficulty of control, that is, the controllability becomes worse. So the most direct way to improve network control is to add links. Due to a lack of purpose, the effect of improvement on controllability is not obvious by adding links randomly in most cases, we can only improve control by restoring the network as much as possible to the pre-lost link structure.

To solve the problems of poor network controllability, in this paper, a link prediction [13] method is proposed to improve the controllability of the network. In our method we extract node properties and local structural features [17] according to current network structure, using these properties and features to train a learning algorithm and then predict the lost edges. Compared to adding edges randomly, link prediction can restore the network as much as possible, and it is simple and the execution efficiency is high. In the case of lost network links, link prediction can effectively improve the controllability of most networks compared to other methods. Our experimental results show that the higher the accuracy of link prediction, the better the controllability improves. In order to improve the accuracy of link prediction, we put forward a Reverse-training method when we don't know the label of test data, this method can effectively adjust learning algorithm and improve accuracy.

The rest of paper is organized as follows. In Sect. 2, we review the related work. In Sect. 3, we introduce our link prediction model. The experimental results are provided and discussed in Sect. 4. Finally, we conclude the paper in Sect. 5.

2 Related Work

2.1 Network Controllability Theory

At present, there are two kinds of controllability theories for complex networks: structural controllability and strict controllability.

If the system matrix [15] A of the complex network system is determined, according to the Kalman controllability criterion, the key to completely control the whole system is to find the appropriate input matrix B and make the matrix C full rank. According to Kalman criterion, Liu et al. put forward the theory of structural controllability: A system is converted into a digraph, if the system is controllable, the directed graph must not contain unreachable nodes and expansion, or it is made of cactus.

The theory of structural controllability [20] can be used to determine whether a network is controllable in most cases, but the foundation of its establishment ignores the edge weight information in the network, and can not exclude that the side weight

combination of the network structure matrix A and the input matrix B happens to be ill conditioned, which causes the network to be controllable theoretically but not controllable in reality. According to Kalman criterion, Liu et al. put forward the theory of structural controllability:

The minimum number of driver nodes required to achieve complete controllability is N_D equals to the largest geometric multiplicity of matrix A:

$$N_D = \max\{\mu(\lambda_i)\} \tag{1}$$

$$\mu(\lambda_i) = N - rank(\lambda_i I_N - A) \tag{2}$$

A is the system matrix, and λ_i is the eigenvalue of the matrix A.

To facilitate the measurement of the difficulty of controlling a network, we define controllable n_D to measure the difficulty of controlling the network:

$$n_D = 1 - \frac{N_D}{N} \tag{3}$$

N is the total node number of the network, and N_D is the minimum number of driving nodes needed to control the network. The less the number of drivers needed to control a network, the better the controllability of the network; the more the minimum number of drivers needed, the worse the controllability.

2.2 Network Attack

The attack modes [4] of complex networks are mainly divided into random attack and selective attack. The random attack is to destroy nodes or edges in a network with some probability. Holme [4] and others have done a more comprehensive study of complex network attacks, divided the attacks into node attacks and edge attacks, each of which contains 4 attack strategies.

① ID (initial degree) attack mode. The nodes (edges) are removed according to the order of their degree in initial network.
② IB (initial betweenness) attack mode. The nodes (edges) are removed according to the order of their betweenness in initial network.
③ RD (recalculated degree) attack mode. The nodes (edges) are removed according to the order of their degree.in current network.
④ RB (recalculated betweenness) attack mode. The nodes (edges) are removed according to the order of their betweenness in current network.

2.3 Link Prediction

Link prediction aims to predict the missing edges or possible links in the future based on the current network structure. The method is divided into local similarity-based approaches and global similarity-based approaches.

Local similarity-based [17] approaches use node neighborhood-related [5] structural information to compute the similarity of each node with other nodes in the

network. These approaches have good results for link prediction and it is very efficient. The existing methods includes Common Neighbors (CN) [6], Jaccard's Coefficient (JC) [7], and Adamic Adar (AA) [8]. CN is represented as the number of common neighbors between two nodes. The more common neighbors two nodes have, the more likely there is a link between them. Compared with the CN coefficient, the JC coefficient takes into account the whole network structure. Adamic and Adar (AA) take into account the correlation when deciding the strong correlation between the two nodes.

3 Model

3.1 Link Prediction Framework

When a complex network is attacked, the network loses some edges. We now consider the problem of predicting the missing edges in the attack in our dataset. For undirected unweighted networks, the essence of link prediction is a classification problem, which can be solved by machine learning classification algorithm [16].

Local similarity-based approaches have good performance for link prediction [19]. In our model, we try to combine these approaches with some attributes of the network as feature for training learning algorithm. The features are divided into two classes. The first class is based on the attribute of the edge. The second class is based on the local approach of the network. For the edge (x, y), the first class we choose is the degree of x and y and the shortest path between x and y; the second class we choose is JC, AA and CN.

We use a logistic regression classifier to combine the evidence from these individual features into link prediction. Logistic regression learns a model of the form

$$\mathrm{P}(+|\mathrm{x}) = \frac{1}{1 + e^{-(b_0 + \sum_i^n b_i x_i)}} \tag{4}$$

where x is a vector of features $(\mathrm{x1}, \ldots, \mathrm{xn})$ and $(\mathrm{b0}, \ldots, \mathrm{bn})$ are the coefficients we estimate based on the training data. For every edge (u, v) with label 1 we sample a random edge with label 0, which ensures that the number of the two labels edges in the data we consider for training and prediction is balanced. Moreover, we also consider two different evaluation measures: the classification accuracy and the area under the ROC curve (AUC). For ease of exposition we focus on classification accuracy on a balanced dataset.

3.2 Reverse-Training Method

The higher the accuracy of link prediction, the better the network recovery. So we now consider the problem of improving the accuracy of link prediction. If we know the labels of training data and test data, we can easily adjust to the best classifier. In addition to the parameters of learning algorithm, the training data which is used to train classifier also affects the accuracy of link prediction. We put forward a reverse-training method here. Compared to the traditional method, this method is more efficient and it shows a good effect in cases that we only know the labels of training data.

This method is divided into two steps, the first step is clearing interfering data and the second step is adjusting the parameters of learning algorithm.

To obtain good classifiers, good training data is required. Some sample features can't reflect the category labels it belongs to, these samples will affect the training of learning algorithms. So we call these samples as interfering data. In filtering the interfering data, we use the idea of data partition [11]. The algorithm is shown in Table 1.

Table 1. Interfering data filtering algorithm.

Algorithm 1 interfering data filtering algorithm
Input: T(training data), n(number of subset), m(division times), H(learning algorithm)
Output: A(detected interfering subset of E)
A←∅
for i = 1, …, m **do**
  form n disjoint almost equally sized subset of E_i, where $\bigcup_i E_i = E$
  for j = 1, …, n **do**
    form $E_t \leftarrow E \backslash E_i$
  for k = 1, …, t **do**
    ues E_t train H to classify E_i
    for every $e \in E_i$ **do**
      if H incorrectly classifies e
        then $A \leftarrow A \cup e$
    end for
  end for
  end for
end for

After clearing up the interfering data [9, 20], the second step is adjusting classifier. The idea of dual-learning is used here [10]. Dual-learning has a good effect on Machine Translation and solves the shortage of parallel training data.

Table 2. Classifier adjustment algorithm.

Algorithm 2 classifier adjustment algorithm
Input: T_1(the training set after filtering interfering data), H(learning algorithm), T_2(the test data), n(the number of T_1)
max = 0
repeat
  count = 0
  use T_1 train H to classify T_2 and get labels of test data, L_1
  use T_2 and L train H to classify T_1 and get labels of training data, $L_{2\backslash}$
  for i = 1, …, n **do**
    if $L_{1i} == L_{2i}$
    then count = count + 1
  end for
  max = count
until convergence

We use the training data to train the classifier, and classify the test data. Conversely, we use the test data and the results as labels for the test data to train the classifier. We get new labels of training data after classifying. By comparing the new labels and the original labels of the training set, we can evaluate the performance of the classifier. If most new labels and the original labels of the training data are identical, it can be proved that the classifier is good. Just like Machine Translation, we convert a message from language A to language B using a translation model. Then we convert the received message from language B back to language A using another translation model. If the message is consistent in language A, we can know whether the two translation model perform well or not. We can adjust the classifier to repeat this process. This process can be iterated for many rounds until the new label of the training set and the original label of the test set have the maximum similarity. We can think the classifier is the best at this time. The Algorithm of adjusting classifier is shown in Table 2.

4 Experiments

In this paper we use four real world networks: Airport, Ant, Jazz, Email, they are all undirected and unweighted. Their network topology properties are shown in Table 3.

Table 3. The information of Network topology.

Network	Node	Edge	Number of driver node
Airport	500	2980	132
Email-Enron	1133	5451	42
Ant	453	2040	27
Jazz	198	2742	7

We remove a certain proportion of edges in the network according to the attack mode, then calculate the controllability of the missing edge network. The ratio of edges removed is called the attack ratio. The attack modes we used in the experiment are: random attack mode, ID attack mode and IB attack mode. ID(IB) attack mode is divided into ID-max(IB-max) mode and ID-min(IB-min) mode, which is defined as removing the edge has the maximum degree (betweenness) in initial network and removing the edge has the minimum degree (betweenness) in initial network. For each network, we adopt different attack ratios and attack modes and analysis the maximal connected subgraph after attacking. Taking into account the randomness of random attack mode, we take the average of the results of the 10 experiments as a reference.

The controllability of each network varies little under random attack. Airport and email have a little decline. When the attack ratio is low, the controllability of Ant has an improvement, but has a decline when the attack ratio is over 15%. Jazz has an improvement in controllability and controllability gradually converges to 1. Even if the attack ratio reaches 40%, the controllability of Email changes very slightly compared to other networks because Email has good controllability initially. Based on the topology

information of the network, the better the controllability of the network is, the smaller the range of variation suffered by random attacks. The experimental results are shown in Fig. 1.

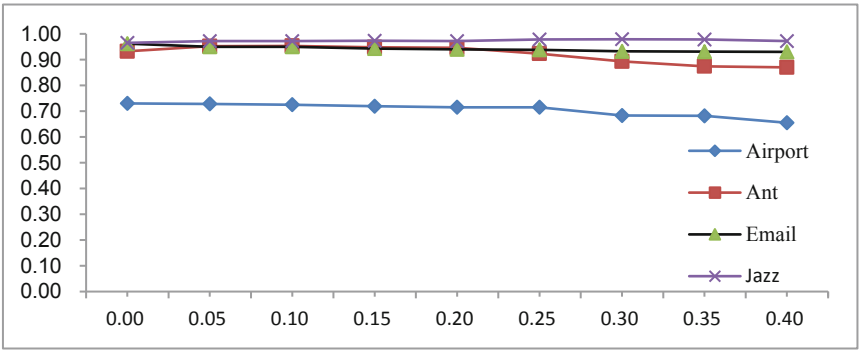

Fig. 1. Controllability of every network under random mode with different attack ratio, the horizontal axis is attack ratio and the vertical axis is controllability.

Depending on the degree of the edge being removed, ID mode is divided into ID-max and ID-min. The controllability of these real world networks has almost no change under ID-max and varies greatly under ID-min, the experimental results are shown in Fig. 2.

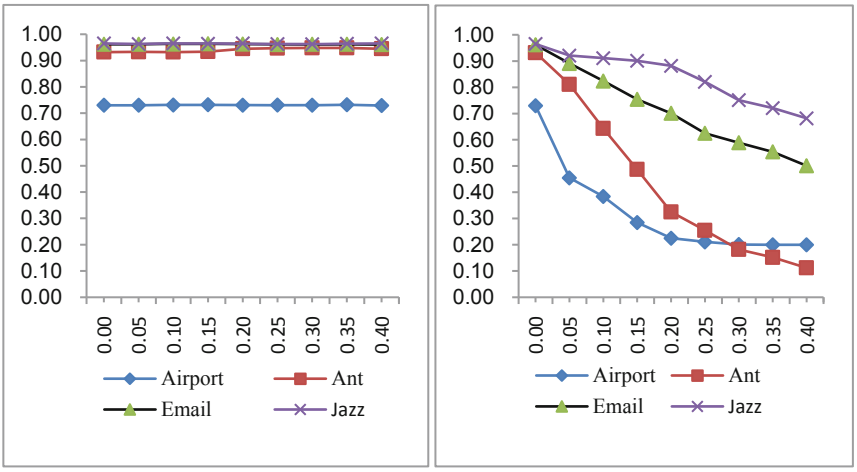

Fig. 2. Controllability of every network under ID mode with different attack ratio, the horizontal axis is attack ratio and the vertical axis is controllability. (a) is ID-max and (b) is ID-min.

There are some edges have no influence in the controllability of network, and these edges are called redundant edge. If a node has no influence on the controllability of the network, it is said to be a redundant node and the degree of redundant node is usually relatively large. Similarly, edges that have no effect on network controllability are called redundant edge. The experimental results show that the controllability of these networks has barely changed under ID-max mode, which prove that the degree of redundant edge is also relatively large. Removing the edges with small degree will have great influence on the network structure and the experimental results show that the controllability of these networks varies greatly even if the attack ratio is low under ID-min mode. It can be proved that the edges which have small degree are crucial for the controllability of network.

Betweenness is also an important topological feature of complex networks. Edge betweenness is defined as the proportion of the number of paths passing through the edge in all shortest paths in the network to the total number of shortest paths. The experimental results of IB mode are similar with which under ID mode, but the controllability of the network under IB-max is not as stable as ID-max. With the increase of attack proportion, the controllability of Airport even improves. It is possible that some edges with large betweenness have certain obstacles to the controllability of the network. Similar to the results under ID-min mode, the controllability of the network under IB-min mode becomes worse with the increase of attack proportion, but the decline rate is not as fast as that under ID-min mode. We can also infer that the small betweenness of edges have an important influence on the controllability of the network. The experimental results are shown in Fig. 3.

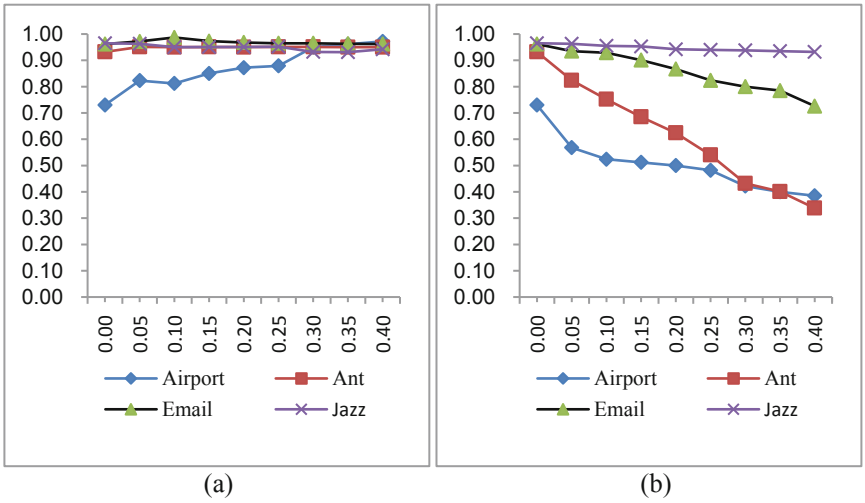

Fig. 3. Controllability of every network under IB mode with different attack ratio, the horizontal axis is attack ratio and the vertical axis is controllability. (a) is ID-max and (b) is ID-min.

The controllability of all networks has a decline under ID-min mode and some networks have a decline under random attack mode. Random addition of edges could not improve the network controllability, so we tried to restore the network structure as much as possible through link prediction. We use the model proposed in this paper to extract attribute features from the network, and take the calculated results of CN, AA and JC as local features. Finally, we use these features to train logistic regression learning algorithm. The link prediction problem can be solved as a binary classification problem. There is an edge between two nodes that belongs to category '1', while there is no edge that belongs to category '0'.

Our model is used to predict the restoration of three networks with reduced controllability through links, and the comparison of the controllability of the network before and after improvement is shown in Fig. 4.

We recorded the controllability curve of the network under different proportions before link prediction as CBLP, and the curve after link prediction as CALP. Experimental results are shown in Fig. 4. The experimental results show that the controllability of the three networks is improved significantly after link prediction under random attacks. When the attack proportion of Ant network is low, the effect does not improve, but decreases. When the attack proportion exceeds 20%, the controllability begins to improve greatly. Compared with random attack mode, the controllability of the network is not significantly improved after link prediction in ID-min mode, and only improves a little when the attack rate is high. From the experimental results we can infer that link prediction is helpful to improve the controllability of the network.

The prediction accuracy of each network through links under different attack modes is shown in Table 4. The structures of every network are seriously damaged under ID-min attack mode because this mode will generate many isolated nodes. So the accuracy of link prediction under this attack mode is low. The function of link prediction is to restore the structure of the network before the attack as much as possible. If the network structure is restored more, the controllability of the network should be improved.

We use the reverse-training method in this paper to improve the accuracy of link prediction. The accuracy after improvement of every network are shown in Table 5.

From the experimental results, we can see that the accuracy of network link prediction using our method has been improved to some extent. We recorded the controllability curve of the network before improving the accuracy rate as CBIA, and the curve after improving the accuracy rate as CAIA. Comparison results of network controllability before and after link prediction accuracy improvement are shown in Fig. 5.

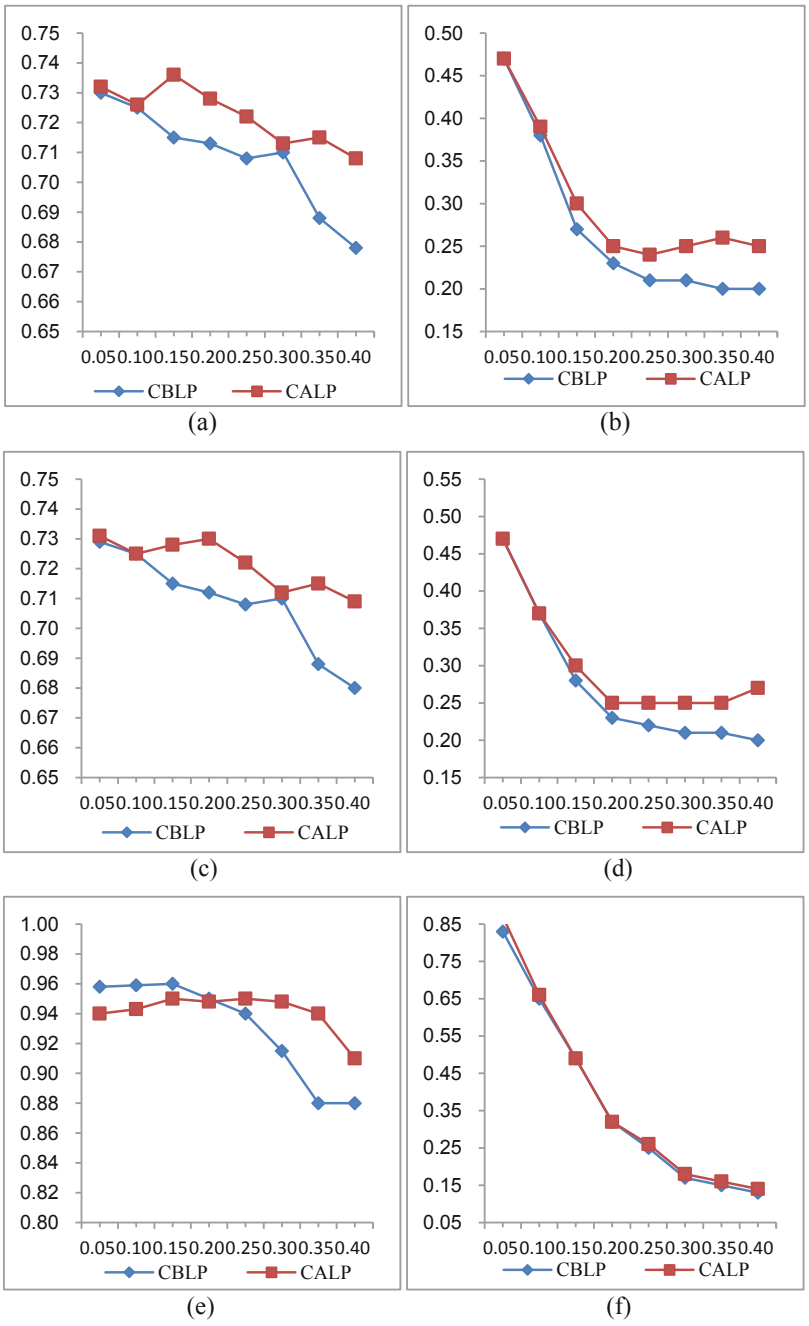

Fig. 4. Comparison of controllability before and after network link prediction. (a) and (b) are comparison of controllability of Airport before and after link prediction in random attack mode and ID-min mode respectively; (c) and (d) are comparison of controllability of Email before and after link prediction in random attack mode and ID-min mode respectively; (e) and (f) are comparison of controllability of Ant before and after link prediction in random attack mode and ID-min mode respectively.

Table 4. Accuracy of link prediction under random and ID-min mode

Mode	Random							
Metric	Accuracy							
Attack ratio	5%	10%	15%	20%	25%	30%	35%	40%
Airport	0.876	0.884	0.869	0.878	0.876	0.867	0.852	0.869
Email	0.810	0.812	0.796	0.801	0.813	0.800	0.800	0.781
Ant	0.831	0.823	0.812	0.825	0.828	0.783	0.803	0.805
Mode	ID-min							
Metric	Accuracy							
Attack ratio	5%	10%	15%	20%	25%	30%	35%	40%
Airport	0.445	0.439	0.445	0.492	0.512	0.512	0.550	0.539
Email	0.502	0.521	0.511	0.513	0.497	0.509	0.498	0.497
Ant	0.472	0.450	0.467	0.469	0.473	0.477	0.482	0.480
Jazz	0.524	0.579	0.550	0.486	0.480	0.485	0.496	0.497

Table 5. Accuracy of link prediction after improvement under random mode and ID-min mode

Mode	Random							
Metric	Accuracy							
Attack ratio	5%	10%	15%	20%	25%	30%	35%	40%
Airport	0.889	0.900	0.896	0.891	0.893	0.884	0.883	0.879
Email	0.842	0.842	0.835	0.828	0.814	0.799	0.800	0.781
Ant	0.896	0.878	0.858	0.855	0.851	0.842	0.838	0.836
Mode	ID-min							
Metric	Accuracy							
Attack ratio	5%	10%	15%	20%	25%	30%	35%	40%
Airport	0.454	0.446	0.447	0.544	0.579	0.582	0.627	0.620
Email	0.647	0.526	0.519	0.528	0.531	0.539	0.509	0.509
Ant	0.637	0.615	0.495	0.469	0.493	0.480	0.486	0.490
Jazz	0.544	0.605	0.562	0.525	0.500	0.506	0.496	0.505

The experimental results show that under random attack mode, the controllability of the three networks has been improved after the improvement of link prediction accuracy. The controllability of the three networks hardly changed when the attack proportion was low in the ID mode, also because the accuracy of the link prediction in the ID mode was low.

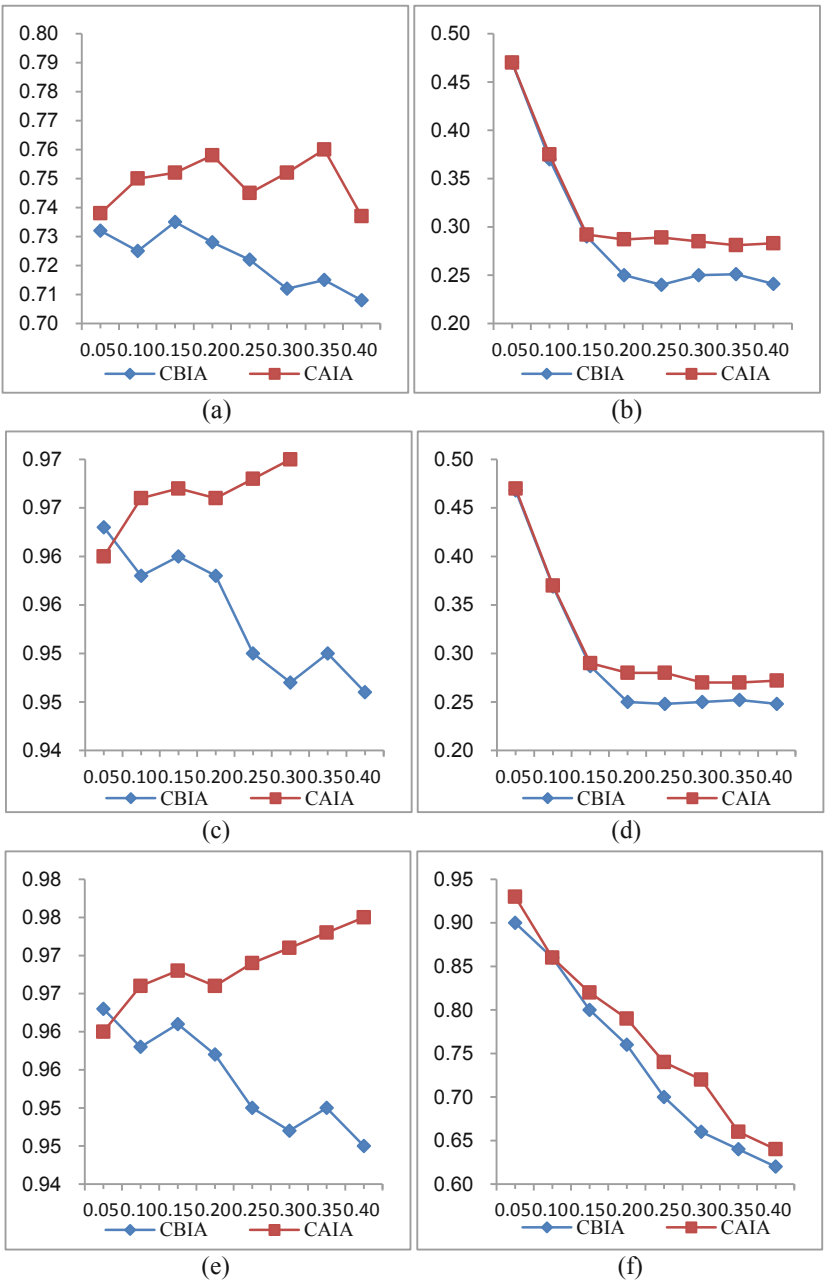

Fig. 5. Comparison of controllability before and after the accuracy of link prediction improved. (a) and (b) are comparison of controllability of Airport before and after improving in random attack mode and id-min mode respectively; (c) and (d) are comparison of controllability of Email before and after improving in random attack mode and id-min mode respectively; (e) and (f) are comparison of controllability of Ant before and after improving in random attack mode and id-min mode respectively.

5 Conclusion

The controllability of complex network is usually reduced after being attacked, which affects the control of the whole network. Therefore, the study of controllability of complex network is very important. For the problem of controllability of network reduction under attack, this paper proposes link prediction model to solve it and the effect is remarkable in most cases. The controllability of most networks is improved under different attack modes. We found that the higher the network restoration degree, the better the controllability improvement effect. In order to improve the accuracy, this paper proposes reverse-training method according to dual-learning algorithm. If the labels of test data are unknown, reverse-training method can adjust the learning algorithm well. Experiments show that the two models both have a good effect. In the future, we will continue to study the controllability of complex networks.

Acknowledgements. This research was supported by Natural Science Foundation of China (Grant no. 61672284), Natural Science Foundation of Jiangsu Province (Grant no. BK20171418), China Postdoctoral Science Foundation (Grant no. 2016M591841), Jiangsu Planned Projects for Postdoctoral Research Funds (No. 1601225C). Meanwhile, this research work was supported by Zayed University Research Cluster Award \# R18038.

References

1. Liu, Y.-Y., Slotine, J.J., Barabási, A.L.: Controllability of complex network. Nature **473**(7346), 167–173 (2011)
2. Yuan, Z., Zhao, C., Di, Z., et al.: Exact controllability of complex networks. Nature Commun. **4**, (2013). Conference 2016. LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016)
3. Rugh, W.J.: Linear System Theory. Prentice Hall, Upper Saddle River (1996)
4. Holme, P., Kim, B.J., Yoon, C.N., Han, S.K.: Attack vulnerability of complex networks. Phys. Rev. E (2002)
5. Davidsen, J., Ebel, H., Bornholdt, S.: Emergence of a small world from local interactions: modeling acquaintance networks. Phys. Rev. Lett. **88**, 128701 (2002)
6. Newman, M.E.J.: Clustering and preferential attachment in growing networks. Phys. Rev. Lett. E **64**, 025102 (2001)
7. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGrawHill, New York (1983)
8. Adamic, L.A., Adar, E.: Friends and neighbors on the web. Soc. Netw. **25**(3), 211–230 (2003)
9. Guan, D., Yuan, W., Ma, T., Lee, S.: Detecting potential labeling errors for bioinformatics by multiple voting. Knowl.-Based Syst. **66**, 28–35 (2014)
10. He, D., Xia, Y., Qin, T., Wang, L., Yu, N., Liu, T.: Dual learning for machine translation. In: Advances in Neural Information Processing Systems 29 (NIPS 2016) (2016)
11. John, G.H.: Robust decision trees: removing outliers from databases. In: Proceeding of International Conference on Knowledge Discovery and Data Mining, pp. 174–179 (1995)
12. Wu, X., Zhu, X., Chen, Q.: Eliminating class noise in large datasets. In: Proceeding of International Conference on Machine Learning, pp. 920–927 (2003)

13. Hasan, M.A., Zaki, M.J.: A survey of link prediction in social networks. In: Aggarwal, C. (ed.) Social Network Data Analytics, pp. 243–275. Springer, Boston (2011). https://doi.org/10.1007/978-1-4419-8462-3_9
14. Bianchin, G., Pasqualetti, F., Zampieri, S.: The role of diameter in the controllability of complex networks. In: CDC 2015, pp. 980–985 (2015)
15. Benavides, P.T., Diwekar, U.M., Cabezas, H.: Controllability of complex networks for sustainable system dynamics. J. Complex Netw. **3**(4), 566–583 (2015)
16. Curiskis, S.A., Osborn, T.R., Kennedy, P.J.: Link prediction and topological feature importance in social networks. In: AusDM 2015, pp. 39–50 (2015)
17. Yu, Z., Kening, G., Feng, L., Ge, Y.: A new method for link prediction using various features in social networks. In: IEEE WISA 2014, pp. 144–147 (2014)
18. Verbaeten, S., Van Assche, A.: Ensemble methods for noise elimination in classification problems. In: Windeatt, T., Roli, F. (eds.) MCS 2003. LNCS, vol. 2709, pp. 317–325. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-44938-8_32
19. Pasqualetti, F., Zampieri, S., Bullo, F.: Controllability metrics, limitations and algorithms for complex networks. IEEE Trans. Control. Netw. Syst. **1**(1), 40–52 (2014)
20. Sun, P.G.: Controllability and modularity of complex networks. Inf. Sci. **325**, 20–32 (2015)