Research on the Improvement of Link Prediction Algorithm and Its Application in Industrial Structure Adjustment

Junhua Rao
Shandong Institute of Commerce and Technology, Jinan, China

Keywords: Complex Network, Link Prediction, Industrial Network, Path Optimization.

Abstract: The adjustment and establishment of a rational industrial structure is aimed at promoting economic development and improving the people's material and cultural living standards. The main indicator of the rationality of the industrial structure is the rational use of resources, so that various industrial sectors can coordinate the provision of products and services needed by society; provide opportunities for full employment of workers; promote the application of advanced industrial technology; get the best economic benefits, etc. The link prediction similarity algorithm is improved, and a new link prediction algorithm is proposed, which is applied to the industrial structure adjustment strategy of China's new normal economy, taking the input-output relationship between departments as the research object, and adjusting and optimizing the network structure based on the link prediction method. In the research of industrial network, this paper first analyzes the economic development of China from a horizontal static perspective by combining geographic information and statistical theory, constructs an interregional industrial network structure model based on the input and output data of the selected main functional area, and finally applies the newly proposed link prediction algorithm to the industrial network model, and obtains the optimization direction of the network structure.

1 INTRODUCTION

In recent years, due to the continuous in-depth development of Internet technology, networking is reflected in all aspects of people's lives. The prominent networking characteristics have prompted people to start studying and analyzing entities from the perspective of networks. The network analysis method provides a methodological and theoretical basis for studying complex network relationships in the real world (Chen, Z. 2021). The economic input-output system can be constructed into an industrial network structure, with nodes representing various industrial sectors in the system, and edges representing the input-output relationship between departments. The network analysis method can predict the demand relationship between departments, propose future development strategies, coordinate and unify the development of industrial networks, promote regional economic restructuring, and enable each region to continuously realize the optimal allocation of resources on the basis of its own resource endowments, so as to continuously improve the overall economic benefits of the region (Jiang, Z. Y. 2021). Under the influence of the new normal economy, industrial structure adjustment is a major strategic direction in China today, the most critical is to form a new engine of innovation-driven economic growth, that is, to explore an effective industrial transfer path, this process shows strong dynamics and complexity (Li, Z 2019).

Link prediction is one of the important tasks in network analysis methods, which can be used for mining and predicting network relationships, and it has important application value in various fields (Ma, Y. X. 2023). For example, biological networks, economic networks, social networks, e-commerce networks, and so on. In biological metabolic networks, link prediction algorithms can predict the interaction between proteins (4). In the economic network, input-output data can be used to construct a complex network of regional industries, reflect the influence and vulnerability of industrial sectors in the industrial chain through network feature values, and analyze industrial transfer according to link prediction theory, so as to provide new ideas for industrial transfer research (Park, J. H.- Sun, W.). The link prediction algorithm can predict the evolution characteristics of different network structures, grasp the dynamic trend of the network, and predict the

Rao, J.
Research on the Improvement of Link Prediction Algorithm and Its Application in Industrial Structure Adjustment.
DOI: 10.5220/0012273600003807

In *Proceedings of the 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (ANIT 2023)*, pages 46-50
ISBN: 978-989-758-677-4

similarity between nodes. There are many similarity measures for link prediction, but they all have different application scenarios (Wu, J. H. 2017). In the process of industrial adjustment, how to determine the optimal industrial transfer path requires the analysis of the overall structural characteristics of the input-output network from the macro and micro perspectives, so as to determine the expression of its optimal similarity measurement (Xu, Z. Q. 2017). Therefore, link prediction can provide a clear reform direction for the adjustment of industrial structure, coordinated development between departments and optimization of industrial path in different regions, and has a scientific role in promoting the implementation of economic policies (Yoon, B.-Yu, Y.).

Link prediction is to use the characteristics of nodes and network structure to predict the similarity between nodes, and then predict the unknown relationship between nodes or the relationship that may occur in the future (Yuan, W. W. 2019). Through literature combing, it is found that the research of link prediction algorithm mainly focuses on algorithm improvement, structure evolution inference and recommendation system application. In the direction of algorithm improvement, predicting unknown edges between nodes and possible future edges based on node similarity has always occupied an important position with low complexity and high accuracy. Node-based similarity is initially studied only based on the similarity of node attributes, that is, if two nodes are close in multiple feature dimensions, the more similar the two nodes are. An important premise of this approach is that the greater the similarity between two nodes, the more likely it is that there will be a link between them (Chen, Z. Y. 2021). Therefore, how to define the similarity of nodes has become a core problem of the method. Although this framework is very simple, the definition of similarity itself is rich in meaning, from a very simple number of common neighbors to a complex mathematical physics such as the average communication time of random walks, or a matrix forest method based on graph theory. So this simple framework offers in fact endless possibilities. Since the local path index only considers the local information of the network, its calculation amount is much smaller than the index based on global information, especially in the case of large and sparse network scale, the advantages of local path index in computational complexity are more obvious, so its application prospect is considerable.

On the one hand, hindered by the difficulty of obtaining the external attributes of network nodes, and on the other hand, benefiting from the rapid development of complex network research, the main research focus of link prediction problems has gradually shifted from the method of relying on node attributes to the method of only using network structure information (Jiang, Z. Y. 2021). Obviously, the latter is also more beautiful and concise in theory. However, research in this area mainly focuses on social networks, and the systematic analysis of the predictive power of a large number of algorithms in various networks is not yet summarized. In addition, there is no in-depth study of the relationship between algorithm performance and network structure features. Discussions of more complex networks, such as rights-based, directed and multipart networks, are rare and unsystematic. Relevant research should be the mainstream in this direction in recent years (Li, Z. 2019).

Network ensemble theory and the associated concept of network entropy and the maximum likelihood estimation method are expected to promote the formation of the theoretical basis of statistical mechanics for complex networks. One problem with this research is that the exact computational complexity of entropy is very large, and it is often not possible for large-scale networks. Some recent link prediction algorithms have applied the concepts of network ensemble and maximum likelihood, but these algorithms are computationally complex and not very accurate, and currently can only process networks with thousands of nodes, and their prediction effect is not good for networks without clear hierarchies. At present, the relevant international research groups are concerned about: first, how to establish a theoretical framework for network link prediction based on network ensemble theory, and produce theoretical conclusions that have a guiding effect on actual prediction, such as estimating the predictable limit through statistical analysis of network structure, guiding the selection of different prediction methods, etc.; The second is how to design efficient algorithms to deal with link prediction problems in large-scale networks.

2 LINK PREDICTION ALGORITHM

Link prediction is one of the important bridges linking complex networks with information science, and it deals with the most basic problem in information science, the restoration and prediction of missing information. The research on link prediction can not only promote the theoretical development of network science and information science, but also has great practical application value, such as knowing protein interaction experiments, conducting online social recommendations, and finding out the connections that play a particularly important role in transportation transmission networks.

The classic triadic closure principle in social network analysis states that if A and B have a common friend C, then the two people are likely to become friends in the future, so that the three nodes form a closed triangle ABC. For general networks, we can generalize this principle as follows: the more neighbors two nodes have, the more similar the two nodes are, and thus the more inclined they are to connect to each other. The simplest node similarity metric based on Common neighbors is defined as follows:

$$s_{xy}^{CN} = |\Gamma(x)| \cap |\Gamma(y)| \quad (1)$$

The advantage of the similarity index based on common neighbor is that the computational complexity is low, but due to the very limited information used, the prediction accuracy is limited, so there are three path-based similarity indicators, which are local path, Katz index and LHN-II index.

A. Local Path Indicators

Consider the factors of the third-order path on the basis of common neighbors, based on the similarity index of the local path, which is defined as：

$$S = A^2 + \alpha A^3 \quad (2)$$

where α are tunable parameters, A represents the adjacency matrix of the network, and α $(A^3)_{xy}$ represents the number of paths with length 3 between nodes Vx and Vy. When α = 0, the LP indicator degenerates into a CN indicator. The CN indicator can also be considered path-based in nature, except that it only takes into account the number of second-order paths. The local path indicator can be extended to higher-order cases, i.e. when considering n-order paths:

$$S^n = A^2 + \alpha \cdot A^3 + \alpha^2 \cdot A^4 + \cdots + \alpha^{n-2} \cdot A^n \quad (3)$$

As n increases, the computational complexity of the local path indicator increases. In general, consider the computational complexity of nth-order paths O(N<k>^n). However, when n tends to infinity, the local path indicator is equivalent to the Katz indicator considering all the paths of the network, and the amount of computation may decrease, because it can be converted into the inverse of the calculation matrix.

B. Katz Indicator

The Katz indicator takes into account the paths of all networks, so it is defined as:

$$S_{xy} = \sum_{l=1}^{\infty} \alpha^l \cdot |paths_{x,y}^{<l>}| = \alpha A_{xy} + \alpha^2 \cdot (A^2)_{xy} + \alpha^3 \cdot (A^3)_{xy} + \cdots \quad (4)$$

where α>0 is a tunable parameter that controls the path weight, and $|paths_{x,y}^{<l>}|$ represents the number of paths of length I in the path connecting nodes Vx and Vy. For the above series to converge, the parameter α should be less than the reciprocal of the maximum eigenvalue of the adjacency matrix, and this definition can also be expressed as:

$$S = (I - \alpha \cdot A)^{-1} - I \quad (5)$$

Obviously, when the parameter α is small, the contribution of higher-order paths is also small, so that the prediction results of the Katz indicator are close to those of the local path indicator.

C. LHN-II Indicator

The LHN-II indicator is another similarity calculation method proposed by Leicht, Holme and Newman, whose basic idea is based on the general equivalence Regular equivalence. Unlike structural equivalence, general equivalence is defined more broadly. Under the definition of general equivalence, if two nodes are connected to similar nodes, then the two nodes are also similar, even if they do not have a common neighbor node between them.

Most of the similarity indicators in the existing link prediction algorithms are single indicators and are only suitable for specific network structures. In this paper, it is proposed to predict the similarity index by mixing links, and the weight in the mixed similarity index is optimized and adjusted by the experimental design method, so as to propose a link prediction algorithm based on the optimal mixed similarity index based on experimental design.

3 OPTIMIZATION OF INDUSTRIAL STRUCTURE

With the deepening of China's economic new normal policy, industrial transfer is a major proposition in China. Adjusting the industrial structure through industrial transfer and eliminating the imbalance in economic development between regions are urgent problems to be solved by China's new economic policy. Therefore, the link prediction method is used to analyze the industrial transfer path, and an industrial transfer path suitable for different regional industrial networks is given, so as to optimize the industrial structure.

The simulation results of different link prediction similarity indicators of the industrial network models in regions A and B are shown in the following table:

Table 1. Simulation table of different link prediction similarity indicators of two industrial network models.

Indicator	A	B
LHN-I	0.470713	0.500014
HDI	0.622898	0.644432
Sorenson	0.679876	0.688796
Jaccard	0.653559	0.754876
Salton	0.699932	0.767789
HPI	0.706543	0.699802
LP	0.712122	0.708758
PA	0.770007	0.711724
CN	0.755432	0.700542
Katz	0.799643	0.699652
RA	0.723117	0.778782

By comparing the accuracy measurement results of the two industrial network models, it can be seen that the AUC accuracy of the RA index considered from the perspective of network resource allocation, considering the number of all paths, the Katz indicator with a large weight for short paths, the CN indicator based on neighbor nodes, and the PA indicator based on the size of the degree value are higher. Among them, the Katz indicator and the RA indicator have a similar idea, and as the path is longer, the lower its weight.

The rules for network optimization according to PA indicators are: priority nodes with large connectivity are reflected in whether the two industrial sectors have greater influence in the same industrial chain. Therefore, the essence of path optimization based on PA indicators is industry chain collaboration. Specifically, according to the industrial chain where the industrial sector is located, priority is given to establishing links between each department and the most influential department, enhancing the inter-departmental connectivity on the network industry chain, and optimizing the allocation of industrial facilities and resource endowments to improve the system cooperation ability of the entire region and promote the closeness of the main functional areas.

The rules for network optimization according to RA metrics are: from the perspective of path, from the x node to the y node to pass part of the resources, their common neighbor becomes the medium of transmission. Assuming that each medium has one unit of resources and distributes them equally to its neighbors, you can get the number of resources that y can receive. The number of resources that can be received is the RA similarity between nodes x,y. In industrial networks, direct input-output relationships are more likely to form between sectors with less influence and those with strong industrial ties. Reflected in the input-output relationship, it also shows that the sectors with less influence in the industrial chain should be the focus of transfer. For example, sectors with less influence should be prioritized as key targets for supply-side reform because of their overcapacity or insufficient technical capabilities. Coordinate the relationship with other departments to form an industrial division of labor, learn from each other's strengths, and promote the productivity of the department.

The path connection tendency of the CN indicator is: there are more neighbor nodes between the two nodes, which is likely to be related. In the industrial network, if two industrial sectors are related to other similar sectors at the same time, they are more likely to have a direct relationship. This idea is embodied in the integration and development of more similar industrial sectors, so that the entire industrial network forms a coordinated whole. For example, two industrial sectors with the same business will produce input-output relationships with the same departments at the same time, and the two departments are more substitutable with each other, and priority should be given to integrating the two departments to develop and grow, strengthen control, achieve capital conservation and technological growth, and make the entire industrial network achieve intensive economic growth.

4 CONCLUSION

With the advancement of science and technology, more and more complex systems have emerged. Data derived from complex networks have increased straightly as well, which have in turn promoted the research process of complex networks. Link prediction is an important research direction of complex networks. It mainly utilizes the known data and their interactions to predict the data that already exists but has not been observed, the data that may appear in the future, and some of the false data.

The hybrid link prediction similarity index can grasp the overall structural characteristics of the industrial network better than the other indicators. The structural adjustment of industrial network through the idea of hybrid link prediction similarity index has the role of coordinated development and overall consideration. The industrial hybrid transfer path is embodied in the convergence of capacity elimination path based on RA indicators, industrial convergence path based on CN indicators, and optimal configuration path based on PA indicators. The three will be mixed in an appropriate proportion, with the production capacity elimination path as the main transfer direction, supplemented by the industrial integration path and the optimal allocation path. This combines the advantages of the three paths to make the migrated network more coordinated.

REFERENCES

Chen, Z. Y., Fan, Z. P., & Sun, M. H. Tensorial graph learning for link prediction in generalized heterogeneous networks (J). *European Journal of Operational Research*, 2021, 290(1), 219-234. http://doi.org/10.1016/j.ejor.2020.05.062

Jiang, Z. Y., Tang, X. K., Zeng, Y., Li, J. K., & Ma, J. F. Adversarial link deception against the link prediction in complex networks (J). *Physica a-Statistical Mechanics and Its Applications*, 2021, 577. http://doi.org/10.1016/j.physa.2021.126074

Li, Z., Liu, Z. L., Huang, J. M., Tang, G. Y., Duan, Y. C., Zhang, Z. Q., & Yang, Y. F. MV-GCN: Multi-View Graph Convolutional Networks for Link Prediction(J). *IEEE Access*, 2019, 7, 176317-176328. http://doi.org/10.1109/access.2019.2957306

Ma, Y. X., Zhao, R., & Yin, N. Application of an Improved Link Prediction Algorithm Based on Complex Network in Industrial Structure Adjustment(J). *Processes*, 2023, 11(6). http://doi.org/10.3390/pr11061689

Park, J. H., Chang, W., & Song, J. W. Link prediction in the Granger causality network of the global currency market (J). *Physica a-Statistical Mechanics and Its Applications*, 2020, 553. http://doi.org/10.1016/j.physa.2020.124668

Sun, W., Li, Q. Y., Wang, J. P., Chen, L. F., Mu, D. M., & Yuan, X. J. A radio link reliability prediction model for wireless sensor networks (J). *International Journal of Sensor Networks*, 2018, 27(4), 215-226. http://doi.org/10.1504/ijsnet.2018.093960

Wu, J. H., Zhang, G. J., & Ren, Y. Z. A balanced modularity maximization link prediction model in social networks (J). *Information Processing & Management*, 2017, 53(1), 295-307. http://doi.org/10.1016/j.ipm.2016.10.001

Xu, Z. Q., Pu, C. L., Sharafat, R. R., Li, L. B., & Yang, J. Entropy-based link prediction in weighted networks (J). *Chinese Physics B*, 2017, 26(1). http://doi.org/10.1088/1674-1056/26/1/018902

Yoon, B., Kim, S., Kim, S., & Seol, H. Doc2vec-based link prediction approach using SAO structures: application to patent network (J). *Scientometrics*, 2022, 127(9), 5385-5414. http://doi.org/10.1007/s11192-021-04187-4

Yu, Y., Tosyali, A., Baek, J., & Jeong, M. K. A Novel Similarity-Based Link Prediction Approach for Transaction Networks (J). IEEE Transactions on Engineering Management, 2022. http://doi.org/10.1109/tem.2022.3146037

Yuan, W. W., He, K. Y., Guan, D. H., Zhou, L., & Li, C. L. Graph kernel based link prediction for signed social networks(J). *Information Fusion*, 2019, 46, 1-10. http://doi.org/10.1016/j.inffus.2018.04.004

Chen, Z. Y., Fan, Z. P., & Sun, M. H. Tensorial graph learning for link prediction in generalized heterogeneous networks (J). *European Journal of Operational Research*, 2021, 290(1), 219-234. http://doi.org/10.1016/j.ejor.2020.05.062

Jiang, Z. Y., Tang, X. K., Zeng, Y., Li, J. K., & Ma, J. F. Adversarial link deception against the link prediction in complex networks (J). *Physica a-Statistical Mechanics and Its Applications*, 2021, 577. http://doi.org/10.1016/j.physa.2021.126074

Li, Z., Liu, Z. L., Huang, J. M., Tang, G. Y., Duan, Y. C., Zhang, Z. Q., & Yang, Y. F. MV-GCN: Multi-View Graph Convolutional Networks for Link Prediction(J). *IEEE Access*, 2019, 7, 176317-176328. http://doi.org/10.1109/access.2019.2957306